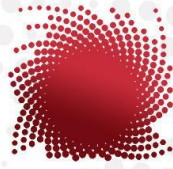




ΕΛΛΗΝΙΚΗ
ΔΗΜΟΚΡΑΤΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΜΑΚΕΔΟΝΙΑΣ



**Business Analytics
and Data Science**

Πρόγραμμα Μεταπτυχιακών Σπουδών στην

ΑΝΑΛΥΤΙΚΗ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ ΚΑΙ ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων

**Πρόγραμμα Μεταπτυχιακών Σπουδών στην Αναλυτική των
Επιχειρήσεων και Επιστήμη των Δεδομένων**

Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων

Διπλωματική Εργασία

**Αναλυτική αθλητικών δεδομένων: στατιστική ανάλυση δεικτών
απόδοσης καλαθοσφαίρισης**

Κούσα Ειρήνη του Οδυσσέα

**Υποβλήθηκε ως απαιτούμενο για την απόκτηση του μεταπτυχιακού διπλώματος
στην Αναλυτική των Επιχειρήσεων και Επιστήμη των Δεδομένων**

Νοέμβριος, 2022

ΠΕΡΙΛΗΨΗ

Με την ταχεία ανάπτυξη της τεχνολογίας της πληροφορίας και του αθλητισμού, η ανάλυση αθλητικών δεδομένων πλέον έχει μεταβεί σε ένα νέο επίπεδο όπου ένας αναλυτής θα πρέπει να είναι να σε θέση να παράγει αποτελέσματα μέσα από ένα τεράστιο σύνολο δεδομένων, να μπορεί να γνωρίζει εσφαλμένες καταχωρήσεις και να μπορεί να παράγει αποτελέσματα για διαφορετικούς τύπους αποδεκτών όπου ο καθένας έχει διαφορετικό αντικείμενο εργασίας και διαφορετικό εκπαιδευτικό υπόβαθρο.

Η πηγή των αθλητικών δεδομένων δεν είναι μόνο το διαδίκτυο αλλά πλέον νέες μέθοδοι άντλησης δεδομένων όπως βίντεο, gps κα έχουν ως αποτέλεσμα την εκθετική αύξηση της πληροφορίας. Τα αθλητικά δεδομένα περιέχουν πλούσιες πληροφορίες για την απόδοση των παικτών και των ομάδων και έχουν αναπτυχθεί εκπληκτικές τεχνολογίες ανάλυσης δεδομένων, οι οποίες επιτρέπουν, μέσα από την ανάλυση τους, την εμφάνιση πληροφορίας πέρα από το κλασικό μέτρο της ευστοχίας, βαθμολογίας κλπ.

Σε αυτή την εργασία παρουσιάζονται κλασικοί και πιο μοντέρνοι μέθοδοι ανάλυσης αλλά και ρεαλιστικών σεναρίων αθλητικών δεδομένων καλαθοσφαίρισης. Εκτός από αυτό σχολιάζονται και οι δυσκολίες που πρέπει να αντιμετωπίσει ένας αναλυτής αθλητικών δεδομένων. Επιπλέον, εξετάζεται η διαχείριση μεγάλων αθλητικών δεδομένων, όπως η απόκτηση αθλητικών μεγάλων δεδομένων, η επισήμανση μεγάλων αθλητικών δεδομένων και η βελτίωση των υπαρχόντων δεδομένων. Τρίτον, παρουσιάζονται μέθοδοι ανάλυσης αθλητικών δεδομένων, συμπεριλαμβανομένης της στατιστικής ανάλυσης. Συμπληρωματικά, περιγράφουμε τις αθλητικές εφαρμογές μεγάλων δεδομένων, όπως η αξιολόγηση και η πρόβλεψη. Αυτή η εργασία θα βοηθήσει τους ερευνητές να αποκτήσουν μια ευρύτερη κατανόηση των αθλητικών μεγάλων δεδομένων και να παράσχει ορισμένες πιθανές ερευνητικές κατευθύνσεις.

Λέξεις κλειδιά : Big data, Ανάλυση, Οπτικοποίηση, Αθλητικά δεδομένα, Καλαθοσφαίριση.

ABSTRACT

With the rapid development of information technology and sports, sports data analysis has now moved to a new level where an analyst should be able to produce results through a huge data set, be able to know wrong entries and to be able to produce results for different types of recipients where each one has a different work-request and different educational background.

The source of sports data is not only the internet but now new methods of extracting data such as video, gps etc. result in an exponential increase in information. Sports data contains rich information about the performance of players and teams, and amazing data analysis technologies have been developed, which allow, through their analysis, the display of information beyond the classic measure of accuracy, score, etc.

In this paper, classical and more modern methods of analysis as well as realistic scenarios of basketball sports data are presented. In addition to this, the difficulties faced by a sports data analyst are also discussed. Additionally, big sports data management is discussed, such as acquiring sports big data, labeling big sports data and improving existing data. Third, sports data analysis methods, including statistical analysis, are presented. Furthermore, we describe sports applications of big data, such as evaluation and prediction. This work will help researchers gain a broader understanding of sports big data and provide some potential research directions.

Keywords: Big data, Analysis, Visualization, Sports Data, Basketball.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ.....	1
ΚΕΦΑΛΑΙΟ 1: Ο ρόλος της ανάλυσης δεδομένων στον αθλητισμό	4
1.1 Εισαγωγή	4
1.2 Ιστορική ανασκόπηση της ανάλυσης αθλητικών δεδομένων.....	5
1.3 Τι είναι η ανάλυση αθλητικών δεδομένων και γιατί εφαρμόζεται.....	7
1.3.1. Γενική περιγραφή	7
1.3.2 Ανάλυση αθλητικών δεδομένων για την κερδοφορία μιας ομάδας	10
1.3.2 Ανάλυση αθλητικών δεδομένων για την βελτίωση μιας ομάδας	15
1.4 Επεξήγηση εννοιών	19
ΚΕΦΑΛΑΙΟ 2: Διαδικασίες διαχείρισης αθλητικών δεδομένων	23
2.1. Τυποποίηση δεδομένων.....	23
2.2. Συγκέντρωση δεδομένων.....	24
2.3. Ποσοτικά δεδομένα	27
2.4. Ποιοτικά δεδομένα	28
2.5. Ενσωμάτωση	31
2.6. Εκτέλεση.....	33
ΚΕΦΑΛΑΙΟ 3: Επισκόπηση γραφικών μεθόδων ανάλυσης	35
3.1. Εισαγωγή	35
3.2. Τα δεδομένα καλαθοσφαίρισης.....	35
3.3. Βασικές στατιστικές αναλύσεις.....	40
3.3.1 Ρυθμός, Βαθμολογίες, Τέσσερις Παράγοντες	40
3.3.2. Ραβδογράμματα και διαγράμματα γραμμής.....	42
3.3.3. Ακτινωτά διαγράμματα	44
3.3.4. Διαγράμματα διασποράς.....	47
3.3.5. Διάγραμμα φυσαλίδας (Bubble plot).....	49
3.3.6. Ανάλυση μεταβλητότητας	51
3.3.7. Ανάλυση ανισοτήτων	54

3.4 Advanced graphics	57
ΚΕΦΑΛΑΙΟ 4: Εύρεση μοτίβων	60
4.1. Δημιουργία μέτρων σχέσεων μεταξύ των μεταβλητών.....	60
4.2. Στατιστική ανεξαρτησία	61
4.3. Effect size	63
4.4. Συσχετίσεις	64
4.5. Ανάλυση δικτύου.....	65
4.6 Pressing.....	68
ΚΕΦΑΛΑΙΟ 5: Επίλογος – Συμπεράσματα.....	78
Βιβλιογραφία	80
Παράρτημα	86
Κώδικας R	86
Γράφημα 1	86
Γράφημα 2	86
Γράφημα 3 και 4.....	87
Γράφημα 5	87
Γράφημα 6	88
Γράφημα 7	88
Πίνακας 5.....	89
Γράφημα 8	90
Πίνακας 6.....	91
Γράφημα 9	91
Γράφημα 11	92
Πίνακας 8.....	93
Συσχετίσεις	93
Γράφημα 13	94
Γράφημα 14	95

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Εκτίμηση αξίας ομάδων NBA Ιανουάριος 2015 (Πηγή: Miller, 2015).....	12
Πίνακας 2. Μισθοί με βάση την εμπειρία επαγγελματιών παικτών του NBA (Πηγή: Adams, 2022)	13
Πίνακας 3. Τυπική ταξινόμηση δεδομένων αγώνων καλαθοσφαίρισης	38
Πίνακας 4. Υπολογισμός τεσσάρων παραγόντων (Πηγή: Kubatko et al. 2007).....	41
Πίνακας 5. Ποσοστά εύστοχων 3ποντων (P3p) και προσπαθειών (P3A) παικτών της ομάδας Oklahoma City Thunder.	52
Πίνακας 6. Παίκτες της ομάδας Oklahoma City Thunder με περισσότερους από 300 πόντους.....	56
Πίνακας 7. Μέτρα αγώνα των Golden State Warriors ανά αντίπαλη ομάδα	62
Πίνακας 8. Μέσοι όροι των μεταβλητών του παιχνιδιού (υπό όρους έως το Playoff) και οι τιμές του 2 της μέσης εξάρτησης των μεταβλητών του παιχνιδιού από την πρόκριση των Playoff.	63
Πίνακας 9. Εξεταζόμενες μεταβλητές.	66
Πίνακας 10. Περιγραφή των δύο συνόλων δεδομένων (Πηγή: Zuccolotto et al. 2018) 70	
Πίνακας 11. Μετατροπή σε κατηγορικές μεταβλητές (Πηγή: Zuccolotto et al., 2018). 73	

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1. Τυπική διάταξη πεντάδας (Πηγή: Abhishek, 2021).....	20
Εικόνα 2. Διοχέτευση δεδομένων χωρίς ενοποίηση.....	30

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Γράφημα 1. Ρυθμός, Επιθετικές/Αμυντικές Βαθμολογίες και Τέσσερις Παράγοντες (διαφορές μεταξύ της ομάδας και του μέσου όρου των υπό εξέταση ομάδων) – Σαιζόν 2017/2018 για τις ομάδες Boston Celtics, Houston Rockets, Cleveland Cavaliers και Golden State Warriors	42
Γράφημα 2. Ραβδόγραμμα γραμμής με αμυντικά στατιστικά των ομάδων του NBA (κορυφή: TOV.O = Σφάλματα των αντιπάλων, PTS.O = Πόντοι που σημειώθηκαν από τους αντιπάλους) - ποσοστά βολής και λεπτά που έπαιξαν οι παίκτες των Houston Rockets (κάτω).	44
Γράφημα 3. Ακτινωτά διαγράμματα επίδοσης παικτών.....	46
Γράφημα 4. Ακτινωτά διαγράμματα επίδοσης παικτών με κανονικοποιημένες (standardized) τιμές.	47
Γράφημα 5. Διάγραμμα διασποράς μεταξύ του αριθμού των λαθών και των ασίστ ανά ποσοστό ευστοχίας.	48
Γράφημα 6. Διάγραμμα φυσαλίδας των εξεταζόμενων ομάδων με βάση με τα ποσοστά επιτυχημένων ελεύθερων βολών και το σύνολο των βολών που επιχειρήθηκαν.....	49
Γράφημα 7. Διάγραμμα φυσαλίδας των παικτών των Golden State Warriors και Cleveland Cavaliers σύμφωνα με στατιστικά στοιχεία που σχετίζονται με την άμυνα (ανά λεπτό παιχνιδιού).....	50
Γράφημα 8. Διάγραμμα φυσαλίδας των παικτών των Golden State Warriors και Cleveland Cavaliers σύμφωνα με στατιστικά στοιχεία που σχετίζονται με την άμυνα (ανά λεπτό παιχνιδιού).....	54
Γράφημα 9. Ανάλυση ανισοτήτων των πόντων που σκόραραν οι ομάδες Brooklyn Nets και Milwaukee Bucks,	57
Γράφημα 10. Κινούμενο – Animated γράφημα συγκεκριμένης επίθεσης στο NBA (Πηγή: Smith, 2016)	58
Γράφημα 11. Shot chart ποσοστού ευστοχίας για τον Kevin Durant.....	59
Γράφημα 12. Πίνακας συσχετίσεων μεταξύ των εξεταζόμενων μέτρων.	65
Γράφημα 13. Διαγράμματα δικτύου της ομάδας Golden State Warriors (Επάνω: Βασικό διάγραμμα. Κάτω: Διάγραμμα με όρια αριθμού σχέσεων =20).....	67
Γράφημα 14. Διάγραμμα δικτύου της ομάδας Golden State Warriors χωρίς τον Draymond Green	68

Γράφημα 15. Μέτρα ορίων των εξεταζόμενων μεταβλητών (πηγή:(Zuccolotto et al., 2018).....	72
Γράφημα 16. Δενδρογράμματα για Serie A (επάνω) και RIO2016 (κάτω) (Πηγή:(Zuccolotto et al., 2018)	74
Γράφημα 17. Επίδοση υπό πίεση (Dataset :Rio 2106, Πηγή: (Zuccolotto et al., 2018)	77

ΕΙΣΑΓΩΓΗ

Σύμφωνα με τους Dmonte & Dmello (2017) υπολογίζεται ότι παράγονται καθημερινά $2,5 \times 10^5$ εκατομμύρια byte δεδομένων. Ο τεράστιος αυτός παραγόμενος όγκος δεδομένων αναφέρεται και ως Big Data και μέσα σε αυτά συμπεριλαμβάνονται οι αναρτήσεις στα μέσα κοινωνικής δικτύωσης, οι φωτογραφίες και τα βίντεο, οι καθημερινές συναλλαγές, τα αρχεία δεδομένων κλπ. Τα ογκώδη δεδομένα που παράγονται πρέπει να αναλυθούν ώστε να αξιοποιηθούν παραγωγικά. Πολλοί οργανισμοί υιοθετούν τεχνολογίες μεγάλων δεδομένων και επιτυγχάνουν πραγματικά αποτελέσματα με αυτήν την προσέγγιση. Τους βοηθά να βρίσκουν μοτίβα, τάσεις στη συμπεριφορά των πελατών και να παρέχουν πιο προσαρμοσμένες υπηρεσίες, να προβλέπουν και να λαμβάνουν έξυπνες και γρήγορες αποφάσεις, για λόγους ασφάλειας και εντοπισμού απάτης, με σκοπό το κέρδος. Έτσι, τα μεγάλα δεδομένα επιτρέπουν στις επιχειρήσεις να κινούνται με μεγαλύτερη ευελιξία και ασφάλεια στον τομέα δραστηριοτήτων τους.

Όπως κάθε δραστηριότητα έτσι και ο αθλητισμός παράγει μεγάλο όγκο δεδομένων που σχετίζονται με τους παίκτες, την απόδοση της ομάδας και το κοινό. Με τα Big Data, είναι δυνατή η ανάλυση μεγάλων ποσοτήτων αυτών των δεδομένων και η αξιοποίησή τους για κερδοσκοπικούς και μη σκοπούς. Δεδομένου ότι υπάρχει μεγαλύτερη ζήτηση για στατιστικά αθλητικών στοιχείων, τα μεγάλα δεδομένα είναι ιδανικά για τον αθλητισμό. Τα Big Data δημιουργούν μια περιεκτική εικόνα στον τομέα του αθλητισμού και οδηγεί τον αθλητισμό αλλά και τους οπαδούς σε υψηλότερα επίπεδα παρακολούθησης της εξέλιξης μιας ομάδας, ενός παίκτη ή ενός αθλητικού γεγονότος. Το βιβλίο του Michael Lewis «Moneyball» (Lewis, 2013) εισήγαγε την έννοια των μεγάλων δεδομένων στον αθλητισμό πολύ πριν χρησιμοποιηθεί πραγματικά. Έδειξε ότι πράγματι τα δεδομένα μπορούν να χρησιμοποιηθούν για τη λήψη σημαντικών αποφάσεων στον χώρο του αθλητισμού.

Ο Maciejewski, (2017) αναφέρει τα Big Data ως game changer, εννοώντας ότι επιτρέπει την μεταβολή-προσαρμογή παραμέτρων στον τομέα των προβλέψεων των αποτελεσμάτων στον αθλητισμό. Το ποδόσφαιρο είναι το άθλημα που έχει κάνει τεράστια και αποτελεσματική χρήση του, ειδικά η Γερμανική Ποδοσφαιρική

Ομοσπονδία. Η νίκη της Γερμανίας στο Παγκόσμιο Κύπελλο του 2014 πιστώνεται στην τεχνολογία Big Data που χρησιμοποίησε εξιδεικευμένο λογισμικό και μεθόδους ανάλυσης δεδομένων για την πρόβλεψη του τρόπου παιχνιδιού των αντίπαλων ομάδων συνυπολογίζοντας και άλλες παραμέτρους όπως; την δυναμικότητα των παικτών, ακόμη και τις καιρικές συνθήκες. Παρόμοιες μέθοδοι εφαρμόζονται και στην Εθνική Ομοσπονδία Καλαθοσφαίρισης (NBA) τόσο σε επίπεδο ομάδας όσο και σε επίπεδο κέρδους – στοιχήματος.

Έχοντας υπόψη τις απεριόριστες εφαρμογές της ανάλυσης των μεγάλων δεδομένων στην παρούσα διατριβή γίνεται προσπάθεια επισκόπησης μεθόδων ανάλυσης μεγάλων δεδομένων και εφαρμογής αυτών. Με αυτό τον τρόπο ο αναγνώστης μπορεί να ενημερωθεί για τις εφαρμογές των μεγάλων δεδομένων και τις μοντέρνες μεθόδους ανάλυσης των δεδομένων αυτών. Στην συνέχεια μπορεί να διαπιστώσει πως αυτές οι μέθοδοι εφαρμόζονται στην πράξη επισημαίνοντας παράλληλα και τα οφέλη των ομάδων που μπορεί να έχουν από αυτού του είδους την ανάλυση. Έτσι, εκτός από την ενημέρωση του αναγνώστη γίνεται παράλληλα και μια προσπάθεια ενίσχυσης παλιότερων μεθόδων στον τομέα του μπάσκετ με νέες που θα μπορούσαν να βοηθήσουν παίκτες και προπονητές στην βελτίωση της ανταγωνιστικότητας τους και παράλληλα να τις ενισχύσουν στο αγωνιστικό τους μέρος – on the field.

Η δομή της διατριβής αυτής χωρίζεται σε δύο κύρια μέρη, το θεωρητικό και το ερευνητικό. Πιο συγκεκριμένα, στο πρώτο κεφάλαιο γίνεται μια εκτενής παρουσίαση της ανάλυση αθλητικών δεδομένων επεξηγώντας με παραδείγματα την αναγκαιότητα χρήσης αυτής τη νέας τεχνολογίας από τις ομάδες. Στο ίδιο κεφάλαιο αναφέρονται και τα κέρδη των ομάδων από την χρήση αυτών των μεθόδων τόσο σε οικονομικό επίπεδο όσο και σε επίπεδο βελτίωσης της ομάδας, παράμετροι, που όπως γίνεται φανερό μέσα από την εργασία, δεν συμβαδίζουν πάντοτε. Στο τέλος του πρώτου κεφαλαίου παρατίθενται βασικές έννοιες στον χώρο της επαγγελματικής καλαθοσφαίρισης για την καλύτερη κατανόηση των αποτελεσμάτων από τον αναγνώστη.

Στο δεύτερο κεφάλαιο δίνεται μια πλήρη περιγραφή της διαχείρισης αθλητικών δεδομένων σε μια ομάδα. Σε αυτό το κεφάλαιο γίνεται μια προσπάθεια ρεαλιστικής

παρουσίασης των συνηθισμένων προβλημάτων διαχείρισης και εκμετάλλευσης αθλητικών δεδομένων περιγραφικά.

Στο τρίτο κεφάλαιο παρουσιάζονται βασικές μέθοδοι ανάλυσης και παρουσίασης αθλητικών δεδομένων με την βοήθεια δεδομένων του NBA για την αγωνιστική περίοδο 2015-2016. Στο τέταρτο κεφάλαιο παρουσιάζονται μέθοδοι αναγνώρισης μοτίβων ή patterns και στο πέμπτο δίνεται ο επίλογος και στα συμπεράσματα της παρουσίασης αυτών των μεθόδων.

Για την εφαρμογή αυτών των μεθόδων χρησιμοποιήθηκε η γλώσσα προγραμματισμού R σε περιβάλλον RStudio 2022.07.1+554 "Spotted Wakerobin" Release (7872775ebddc40635780ca1ed238934c3345c5de, 2022-07-22) for Windows. Ο κωδικός R που παρέχεται στο παράρτημα της εργασίας. Οι γενικές αναφορές στην υλοποίηση των μεθόδων σε R έγιναν να την βοήθεια των Chambers, (2008), Matloff, (2011), και Wickham, (2014).

ΚΕΦΑΛΑΙΟ 1: Ο ρόλος της ανάλυσης δεδομένων στον αθλητισμό

1.1 Εισαγωγή

Τα Analytics είναι ένα σχετικά νέο και ταχέως εξελισσόμενο σύνολο εργαλείων στον επιχειρηματικό κόσμο και τα οποία προσαρμόζονται όλο και περισσότερο στον χώρο των αθλημάτων. Τα Analytics περιλαμβάνουν προηγμένα στατιστικά στοιχεία, διαχείριση δεδομένων, οπτικοποίηση δεδομένων και πολλά άλλα πεδία. Επειδή αυτή η λίστα αλλάζει συνεχώς, η εφαρμογή ενός προγράμματος ανάλυσης για την απόκτηση ανταγωνιστικού πλεονεκτήματος δεν είναι μια απλή διαδικασία. Κάθε αθλητικός οργανισμός αντιμετωπίζει τις δικές του προκλήσεις όσον αφορά την εισαγωγή και την ανάπτυξη αναλυτικών στοιχείων ως μέρος της διαδικασίας λήψης αποφάσεων, αλλά η κατανόηση των στοιχείων ενός προγράμματος ανάλυσης θα βοηθήσει την διοίκηση μιας ομάδας να μεγιστοποιήσει το ανταγωνιστικό της πλεονέκτημα που μπορούν να κερδίσουν από την αναλυτική τους επένδυση.

Η ώθηση στον αθλητισμό —όπως και στις επιχειρήσεις— για τη χρήση μοντέρνων εργαλείων ανάλυσης προέρχεται από την πρόοδο στην υπολογιστική ισχύ και τη διαθεσιμότητα τεράστιων ποσοτήτων δεδομένων τόσο στις ομάδες όσο και στο κοινό, που δημιουργούν μια ευκαιρία για ανταγωνιστικό πλεονέκτημα. Η πρόσβαση σε πληροφορίες που δεν διαθέτουν οι ανταγωνιστές έχει μακρά ιστορία παροχής πλεονεκτημάτων σε ομάδες και επιχειρήσεις. Ομάδες όπως οι Oakland A's, Tampa Bay Rays και San Antonio Spurs έχουν αποδεχθεί την αναγκαιότητα της χρήσης των αναλυτικών στοιχείων και οι τρεις σύλλογοι, αν και βρίσκονται σε μικρές αγορές και επομένως έχουν περιορισμένους πόρους, έχουν σημειώσει τεράστια επιτυχία, εν μέρει λόγω των πληροφοριών που διαχειρίζονται και αναλύουν, πλεονέκτημα που αποκτήθηκε από προγράμματα ανάλυσης δεδομένων που διαθέτουν.

Οι Rays, για παράδειγμα, ήταν μία από τις πρώτες ομάδες που χρησιμοποίησαν δεδομένα από το Pitch F/X¹, το οποίο παρακολουθεί τη διαδρομή της μπάλας σε κάθε γήπεδο για να ενημερώνει καλύτερα την αξιολόγηση των παικτών. Ομάδες που επενδύουν σε

¹ Το PITCHf/x είναι ένα σύστημα που δημιουργήθηκε και συντηρείται από την Sportvision (<https://www.smt.com/>) που παρακολουθεί τις ταχύτητες και τις τροχιές της μπάλας του μπίτζμπολ.

αναλυτικά συστήματα και στην χρήση αναλυτικών εργαλείων θα αποκτήσουν σταθερό ανταγωνιστικό πλεονέκτημα έναντι άλλων ομάδων στο πρωτάθλημα τους.

Οι οργανισμοί κινδυνεύουν να μην αποκομίσουν κανένα πλεονέκτημα από την επένδυση σε ένα πρόγραμμα ανάλυσης δεδομένων εάν δεν επενδύσουν επίσης στην κατανόηση και τον σχεδιασμό του τρόπου ενσωμάτωσης των αναλυτικών στοιχείων στη διαδικασία λήψης αποφάσεων.

Τα Analytics μπορούν να χρησιμοποιηθούν από οργανισμούς σε κάθε τομέα του αθλητισμού. Ο τρόπος χρήσης του θα διαφέρει από επίπεδο σε επίπεδο—οι κολεγιακές ομάδες γυμνασίου προφανώς δεν διαθέτουν τους πόρους των ομάδων NBA και NFL—αλλά οι γενικές ιδέες και στρατηγικές είναι χρήσιμες για τους υπεύθυνους λήψης αποφάσεων σε όλους τους αθλητικούς οργανισμούς.

Στην συνέχεια ο όρος «sports analytics» θα αναφέρεται στη «διαχείριση δομημένων ιστορικών δεδομένων, την εφαρμογή προγνωστικών αναλυτικών μοντέλων που χρησιμοποιούν αυτά τα δεδομένα και τη χρήση πληροφοριακών συστημάτων για την ενημέρωση των υπευθύνων λήψης αποφάσεων και τη δυνατότητα να βοηθήσουν τους οργανισμούς τους. απόκτηση ανταγωνιστικού πλεονεκτήματος στον αγωνιστικό χώρο».

1.2 Ιστορική ανασκόπηση της ανάλυσης αθλητικών δεδομένων

Οι αθλητικές ομάδες έχουν εφαρμόσει αναλυτικά μοντέλα για να τους βοηθήσουν στη λήψη αποφάσεων για περισσότερα από 50 χρόνια. Ωστόσο, μέχρι πρόσφατα, μεγάλο μέρος της προηγμένης εφαρμογής αναλυτικών στοιχείων αφορούσε αποκλειστικά τις στρατηγικές παικτών και ομάδων στον αγωνιστικό χώρο. Για παράδειγμα, σε μια από τις πρώτες μελέτες αθλητικής ανάλυσης, ο Lindsey (Lindsey, 1959) ανέλυσε στρατηγικές παικτών του μπέιζμπολ και της ομάδας τους, όπως η σημασία του μέσου όρου του κτυπήματος, η επιλογή pitcher², οι στρατηγικές προσαρμογές στη σύνθεση και η συνολική αμυντική στρατηγική της ομάδας (Alamar & Mehrotra, 2011)

Ο Lindsey (Lindsey, 1963) ανέλυσε επίσης ιστορικά στατιστικά στοιχεία ομάδων και παικτών για να προσδιορίσει τις ενδεδειγμένες στρατηγικές διαχείρισης του παιχνιδιού.

² Ο παίκτης που πετάει την μπάλα

Ο Lindsey (Lindsey, 1959) προέβλεψε ότι, «για την ερμηνεία των δεδομένων σχετικά με τις προηγούμενες επιδόσεις προκειμένου να επηρεαστούν οι αποφάσεις σχετικά με μελλοντικές καταστάσεις, θα πρέπει να συνυπολογίζονται και άλλες επιμέρους παράμετροι» (σελ. 197). Οι αθλητικές οργανώσεις συνεχίζουν να αναπτύσσουν τις σκέψεις του Lindsey και να χρησιμοποιούν δεδομένα για να επηρεάσουν το προσωπικό των παικτών και τις στρατηγικές αποφάσεις της ομάδας. Επιπλέον, τα τελευταία 10 χρόνια, πολλοί αθλητικοί οργανισμοί έχουν ενσωματώσει τη λήψη αποφάσεων βάσει δεδομένων σε βασικές επιχειρηματικές λειτουργίες, όπως η τιμολόγηση των εισιτηρίων, η διαχείριση βάσεων δεδομένων, η αξιολόγηση εταιρικών συνεργασιών, η αφοσίωση των οπαδών τους και εργαλεία CRM (Troilo et al., 2016a).

Οι αθλητικές επιχειρηματικές οργανώσεις έχουν μιμηθεί στρατηγικές «διαχείρισης εσόδων» και «διαχείρισης απόδοσης» που αναπτύχθηκαν προηγουμένως σε άλλους κλάδους. Για παράδειγμα, από τη δεκαετία του 1990, οι αεροπορικές εταιρείες και οι ξενοδοχειακές βιομηχανίες έχουν εφαρμόσει μεταβλητά και δυναμικά μοντέλα τιμολόγησης εισιτηρίων/δωματίων με βάση τη ζήτηση των καταναλωτών και τη διαθεσιμότητα προϊόντων/υπηρεσιών, επιπλέον της δημιουργίας προγραμμάτων αφοσίωσης και ανταμοιβής πελατών (π.χ. μίλια συχνών επιβατών) μαζί με ανάπτυξη «αναβαθμίσεων» της εμπειρίας πελατών (π.χ. ευκαιρία για τους καταναλωτές να αυξήσουν τις δαπάνες για ένα premium δωμάτιο ξενοδοχείου ή για περισσότερο χώρο για τα πόδια σε ένα αεροπλάνο).

Οι στρατηγικές που βασίζονται σε δεδομένα στον κλάδο των αθλητικών επιχειρήσεων διαφέρουν επίσης από τις επιχειρησιακές στρατηγικές που είναι κοινές στις αεροπορικές και ξενοδοχειακές βιομηχανίες. Οι αθλητικοί οργανισμοί συνεργάζονται επίσης και μαθαίνουν βέλτιστες πρακτικές σε αρκετούς άλλους κλάδους, προκειμένου να αναπτύξουν και να ενσωματώνουν με συνέπεια στρατηγικές ανάλυσης. Για παράδειγμα, ο Davenport (Davenport, 2014) αναφέρει τη συνεργασία των Orlando Magic με την Walt Disney Co.—η οποία έχει μια ισχυρή ομάδα αναλυτικών στοιχείων στο Ορλάντο—για κοινές προωθήσεις σε κατοίκους και επισκέπτες στην περιοχή (σελ. 12).

Ομοίως, οι αθλητικοί οργανισμοί έχουν υιοθετήσει τακτικές ψηφιακού μάρκετινγκ (π.χ. στοχευμένες καμπάνιες ηλεκτρονικού ταχυδρομείου και βελτιστοποιημένα σχέδια ιστοτόπων) καθώς και στρατηγικές αφοσίωσης στην επωνυμία των καταναλωτών (π.χ. προτάσεις προϊόντων, ομαδοποίηση και συνδρομές) από εταιρείες ηλεκτρονικού εμπορίου όπως η Amazon για να αναπτύξουν εξατομικευμένες στρατηγικές βασισμένες σε δεδομένα με βάση τις ιδιαίτερες προτιμήσεις και τα μοντέλα αγορών των καταναλωτών. Οι καινοτόμοι σε άλλους κλάδους συνεχίζουν να δημιουργούν ένα σχέδιο ανάλυσης και να καθιερώνουν βασικούς δείκτες απόδοσης (key performance indicators ή KPIs) για αθλητικούς οργανισμούς.

1.3 Τι είναι η ανάλυση αθλητικών δεδομένων και γιατί εφαρμόζεται

1.3.1. Γενική περιγραφή

Σκοπός της ανάλυσης των αθλητικών επιδόσεων είναι η κατανόηση του πώς διάφοροι παράγοντες συμβάλλουν στην επιτυχία μιας ομάδας στο γήπεδο και αυτή η προσπάθεια ξεκινά με την καταμέτρηση. Πολλοί παράγοντες συμβάλλουν στην επιτυχία μιας ομάδας στον αθλητισμό. Υπάρχουν μέτρα που σχετίζονται με το φυσικό ανάστημα, τη βιοφυσική, την υγεία, τη φυσική κατάσταση του παίκτη και την προετοιμασία. Υπάρχουν μέτρα που αφορούν την ταχύτητα, τη δύναμη, την ευελιξία και την ευκινησία. Υπάρχουν ψυχολογικά μέτρα νοημοσύνης, προσωπικότητας και στάσης. Τέλος, υπάρχουν μέτρα που σχετίζονται με την επάρκεια στον αθλητισμό — γνώση, δεξιότητες και εκτέλεση στην πράξη και στα παιχνίδια.

Πρωταρχικό ερώτημα σε αυτή την περίπτωση είναι «Ποιο είναι το ζητούμενο από τα μέτρα απόδοσης στον αθλητισμό; Τι κάνει ένα μέτρο αξιόπιστο; Τι κάνει ένα μέτρο έγκυρο;». Με τον όρο αξιοπιστία γίνεται αναφορά στην αξιοπιστία ή την επαναληψιμότητα των διαδικασιών μέτρησης. Έτσι, λαμβάνεται υπόψη ο βαθμός στον οποίο επαναλαμβανόμενες μετρήσεις του ίδιου χαρακτηριστικού ταυτοχρόνως συμφωνούν μεταξύ τους. Όταν η αξιολόγηση γίνεται με την έρευνα πολλών στοιχείων-δεδομένων, τότε το ζητούμενο είναι η εσωτερική συνέπεια. Στην περίπτωση της εγκυρότητας, γίνεται αξιολόγηση του βαθμού στον οποίο ένα μέτρο μετρά αυτό που υποτίθεται ότι προσπαθεί να αξιολογήσει. Υπάρχουν υποκειμενικές εκτιμήσεις της

εγκυρότητας προσώπου πχ. ατομικές επιδόσεις ή εγκυρότητας περιεχομένου πχ ομαδικές επιδόσεις.

Μια πιο αντικειμενική προσέγγιση για την αξιολόγηση εγκυρότητας θα ήταν η απόδειξη της προγνωστικής εγκυρότητας δηλαδή της ορθής πρόγνωσης αποτελεσμάτων επιδόσεων. Γνωρίζοντας πώς συσχετίζονται δύο χαρακτηριστικά στη θεωρία, είναι δυνατόν να δημιουργηθούν τα μέτρα αυτών των χαρακτηριστικών και να εξεταστεί ο βαθμός στον οποίο αυτά τα μέτρα σχετίζονται όπως υποδηλώνει η θεωρία. Η έννοια ενός μέτρου ορίζεται από τη σχέση του με άλλα μέτρα. Αυτή είναι η εγκυρότητα κατασκευής ενός δείκτη, δηλαδή μια λογική επέκταση της προγνωστικής εγκυρότητας.

Οι Campbell και Fiske (Campbell & Fiske, 1959) ορίζουν την αξιοπιστία και την εγκυρότητα ως εξής: *«Η αξιοπιστία είναι η συμφωνία μεταξύ δύο προσπαθειών για τη μέτρηση του ίδιου χαρακτηριστικού μέσω όσο το δυνατόν παρόμοιων μεθόδων. Η εγκυρότητα αντιπροσωπεύεται στη συμφωνία μεταξύ δύο προσπαθειών μέτρησης του ίδιου χαρακτηριστικού μέσω όσο το δυνατόν διαφορετικών μεθόδων.»*

Εξ ορισμού ως «δεδομένα» γενικά ορίζονται είτε «γεγονότα ή πληροφορίες που χρησιμοποιούνται για τον υπολογισμό, την ανάλυση ή τον προγραμματισμό μιας διαδικασίας» είτε ως «πληροφορίες που παράγονται ή αποθηκεύονται από έναν υπολογιστή» (Λεξικό Merriam-Webster, Data Definition & Meaning - Merriam-Webster, 2022). Διαφορετικά, τα «δεδομένα» αναφέρονται στον προσδιορισμό, τη συγκέντρωση, την αποθήκευση και τη διαχείριση σχετικών πληροφοριών. Ο συγγραφέας και ειδικός δεδομένων Bernard Marr εξηγεί, «η βασική ιδέα πίσω από τη φράση «μεγάλα δεδομένα» είναι ότι οτιδήποτε κάνουμε στη ζωή μας που έχει ή σύντομα θα αφήσει ένα ψηφιακό ίχνος, το οποίο μπορεί να χρησιμοποιηθεί και να αναλυθεί» (Marr, 2015:9)

Ως Analytics περιγράφει οποιαδήποτε διαδικασία που βασίζεται σε δεδομένα, καθώς και οποιεσδήποτε πρακτικές πληροφορίες που προέρχονται από δεδομένα (Liebowitz, 2014). Η ανάλυση περιλαμβάνει την ανάλυση/αξιολόγηση και την εφαρμογή/παρουσίαση δεδομένων (δηλαδή τη μετατροπή των δεδομένων σε αξιόπιστη πληροφορία και έγκυρα αποτελέσματα). Οι (Ransbotham et al., 2015) περιγράφουν τα αναλυτικά στοιχεία ως «η χρήση δεδομένων και σχετικών επιχειρηματικών γνώσεων που αναπτύχθηκαν μέσω

εφαρμοσμένων αναλυτικών κλάδων (για παράδειγμα, στατιστικά, ποσοτικά, προγνωστικά, γνωστικά και άλλα μοντέλα) για να οδηγήσουν σε προγραμματισμό, αποφάσεις, εκτέλεση, διαχείριση βασισμένη σε γεγονότα, μέτρηση και μάθηση» (σελ. 63). Όπως εφαρμόζεται στον κλάδο των αθλητικών επιχειρήσεων, τα αναλυτικά στοιχεία λειτουργούν ως «εργαλεία για την ερμηνεία και τη χρήση δεδομένων για τη λήψη καλύτερων αποφάσεων...[και για την προσθήκη] πολύτιμης αντικειμενικότητας στη διαδικασία λήψης αποφάσεων» (Rein et al., 2015).

Ο βασικός σκοπός των «sport business analytics» είναι η μετατροπή των ακατέργαστων δεδομένων σε ουσιαστικές και αξιόπιστες πληροφορίες που επιτρέπουν στους επαγγελματίες του αθλητισμού να λαμβάνουν στρατηγικές επιχειρηματικές αποφάσεις, οι οποίες στη συνέχεια έχουν ως αποτέλεσμα βελτιωμένη οικονομική απόδοση της εταιρείας και μετρήσιμο και βιώσιμο ανταγωνιστικό πλεονέκτημα. Ο Ben Alamar, διευθυντής ανάλυσης παραγωγής στο ESPN, προτείνει ότι πρωταρχικός στόχος των αθλητικών επιχειρηματικών αναλυτικών στοιχείων πρέπει να είναι η εξοικονόμηση χρόνου των αρμοδίων για τη λήψη αποφάσεων καθιστώντας τις σχετικές πληροφορίες αποτελεσματικά διαθέσιμες, έτσι ώστε αυτοί που λαμβάνουν αποφάσεις να μπορούν να αναλύουν πληροφορίες αντί να αφιερώνουν χρόνο στη συλλογή πληροφοριών. Επιπλέον, η διαδικασία ανάλυσης που βασίζεται σε δεδομένα θα πρέπει επίσης να παρέχει στους λήπτες αποφάσεων νέα γνώση (Alamar, 2013b).

Ένα αποτελεσματικό σύστημα επιχειρηματικής ανάλυσης για τον αθλητισμό θα πρέπει να έχει ως αποτέλεσμα αυξημένα πρόσθετα έσοδα, μειωμένο κόστος, διαχειριζόμενο ρίσκο, πιο αποτελεσματική χρήση των ανθρώπινων πόρων (π.χ. «αναλύσεις ταλέντων»), βελτιστοποιημένη ανάπτυξη προϊόντων και υπηρεσιών (π.χ., καινοτομία βάσει δεδομένων), βελτιωμένο μάρκετινγκ και εξυπηρέτηση πελατών (π.χ. fan loyalty) και συνολικά πιο ενημερωμένη λήψη στρατηγικών αποφάσεων (Lipscomb, 2014; Troilo et al., 2016b).

Για παράδειγμα, οι αθλητικές επιχειρηματικές αναλύσεις επιτρέπουν στους αθλητικούς οργανισμούς να κατανοήσουν καλύτερα και να εκτελέσουν προωθητικές ενέργειες σε τρέχοντες και υποψήφιους πελάτες μέσω στοχευμένων πρωτοβουλιών διαχείρισης

σχέσεων με τους πελάτες (CRM). Τα αναλυτικά στοιχεία αθλητικών επιχειρήσεων βελτιώνουν και βελτιστοποιούν, επίσης, τις επιχειρηματικές διαδικασίες (και τα επακόλουθα πρόσθετα έσοδα) όσον αφορά την τιμολόγηση των εισιτηρίων, την αναζήτηση και την ενεργοποίηση εταιρικών συνεργασιών και τις δραστηριότητες τροφίμων και ποτών. Σε μια πρόσφατη εμπειρική μελέτη σχετικά με την υιοθέτηση αναλυτικών στοιχείων από επαγγελματικούς αθλητικούς οργανισμούς, η χρήση αναλυτικών στοιχείων συσχετίστηκε με αύξηση εσόδων κατά 7,2% το έτος μετά την υιοθέτηση μιας στρατηγικής επιχειρηματικής ανάλυσης σε σύγκριση με τις γενικές προσδοκίες του κλάδου του αθλητισμού για ετήσιες αυξήσεις 3% έσοδα (Troilo et al., 2016b).

1.3.2 Ανάλυση αθλητικών δεδομένων για την κερδοφορία μιας ομάδας

Κατά την εφαρμογή των νόμων της οικονομίας στον επαγγελματικό αθλητισμό, πρέπει να λαμβάνεται υπόψη η φύση του αθλητισμού και τα κίνητρα των ιδιοκτητών. Τα επαγγελματικά αθλήματα διαφέρουν από άλλες μορφές επιχειρηματικής δραστηριότητας. Υπάρχουν πωλητές και αγοραστές αθλητικής ψυχαγωγίας. Οι πωλητές είναι οι παίκτες και οι ομάδες στα πρωταθλήματα των επαγγελματικών αθλημάτων. Οι αγοραστές είναι καταναλωτές αθλημάτων, πολλοί από τους οποίους δεν πηγαίνουν ποτέ σε αγώνες αυτοπροσώπως, αλλά παρακολουθούν αθλήματα στην τηλεόραση, ακούν ραδιόφωνο και αγοράζουν σύνεργα αθλητικών ομάδων. Ο αθλητισμός ανταγωνίζεται άλλες μορφές ψυχαγωγίας για τον χρόνο και τα χρήματα των ανθρώπων. Και διάφορα αθλήματα ανταγωνίζονται μεταξύ τους, ειδικά όταν οι εποχές τους αλληλεπικαλύπτονται. Οι αθλητικές ομάδες παράγουν ψυχαγωγικό περιεχόμενο που διανέμεται μέσω των μέσων ενημέρωσης και αδειοδοτούν τα εμπορικά σήματα και τα λογότυπά τους σε άλλους οργανισμούς, συμπεριλαμβανομένων των κατασκευαστών αθλητικών ενδυμάτων.

Ακόμη, οι αθλητικές ομάδες δεν είναι ανεξάρτητες επιχειρήσεις που ανταγωνίζονται μεταξύ τους ενώ οι παίκτες και οι ομάδες ανταγωνίζονται στα γήπεδα και τα γήπεδα συνεργάζονται μεταξύ τους ως μέλη των πρωταθλημάτων.

Το βασικό προϊόν του αθλητισμού είναι ο αθλητικός αγώνας, ένα κοινό προϊόν δύο ή περισσότερων παικτών ή δύο ή περισσότερων ομάδων, όμως τα επίπεδα συμμετοχής των οπαδών στον αθλητισμό δεν είναι απαραίτητα δείκτης των επιπέδων ενδιαφέροντος για τον αθλητισμό ως ψυχαγωγία. Οι αθλητικές επιχειρήσεις παράγουν προϊόντα ψυχαγωγίας συνεργαζόμενες μεταξύ τους. Αν και είναι παράνομο για τις επιχειρήσεις των περισσότερων βιομηχανιών να συνεννοούνται για τον καθορισμό της παραγωγής και των τιμών, τα αθλητικά πρωταθλήματα συμμετέχουν σε συνεργατική παραγωγή και τιμολόγηση ως τυπικό μέρος του επιχειρηματικού τους μοντέλου. Ο αριθμός των αγώνων, και μάλιστα ολόκληρο το πρόγραμμα αγώνων σε ένα άθλημα, καθορίζεται από το πρωτάθλημα. Στην πραγματικότητα, οι πτυχές του επαγγελματικού αθλητισμού έχουν μονοπωλιακή εξουσία από την κεντρική διοίκηση (στην Ελλάδα από το υπουργείο πολιτισμού και αθλητισμού).

Κατά την ανάπτυξη ενός μοντέλου για μια τυπική επιχείρηση, ένας οικονομολόγος θα υποθέσει τη μεγιστοποίηση του κέρδους ως κίνητρο, αλλά για μια επαγγελματική αθλητική ομάδα, τα κίνητρα του ιδιοκτήτη μπορεί να μην είναι τόσο εύκολα κατανοητά. Ενώ ένας ιδιοκτήτης μπορεί να λειτουργεί την ομάδα του/της με βάση το ετήσιο κέρδος, ένας άλλος μπορεί να επιδιώξει να μεγιστοποιήσει τα κέρδη παράλληλα με τη συνολική απόδοση της ομάδας. Για παράδειγμα, ενώ η συμμετοχή σε διεθνείς οργανώσεις μπορεί να αποφέρει περισσότερα κέρδη, ένας ιδιοκτήτης μπορεί για μια σεζόν να μην το έχει ως βασικό στόχο με κριτήριο την προετοιμασία και ενδυνάμωση της ομάδας. Κάποιος άλλος μπορεί να αναζητήσει ανατίμηση κεφαλαίου — να αγοράσει και μετά να πουλήσει μετά από μερικά χρόνια.

Ελλείψει γνώσης των κινήτρων των ιδιοκτητών, είναι δύσκολο να προβλεφθεί τι θα κάνουν. Η απόκτηση μεριδίου αγοράς και η κατοχή του βασικού παίκτη είναι στόχος των επιχειρήσεων σε πολλούς κλάδους, όμως όχι τόσο στον επαγγελματικό αθλητισμό. Εάν μια ομάδα είχε εξασφαλίσει τη νίκη σε όλους σχεδόν τους αγώνες της, το ενδιαφέρον για αυτούς τους αγώνες θα μπορούσε να μειωθεί. Μια ομάδα επωφελείται κερδίζοντας συχνότερα από ό,τι χάνει, αλλά το να κερδίζεις συνεχώς μπορεί να είναι λιγότερο ωφέλιμο από το να κερδίζεις τις περισσότερες φορές. Τα επαγγελματικά αθλητικά πρωταθλήματα ισχυρίζονται ότι αναζητούν ανταγωνιστική ισορροπία, αν και υπάρχουν

κυρίαρχες ομάδες σε πολλά πρωταθλήματα. Ο αθλητισμός είναι μεγάλη επιχείρηση όπως φαίνεται από τις αποτιμήσεις και τα οικονομικά των μεγάλων επαγγελματικών αθλημάτων στις Ηνωμένες Πολιτείες και παγκοσμίως. Τα δεδομένα από το Forbes για την Εθνική Ομοσπονδία Καλαθοσφαίρισης (NBA), εμφανίζονται στον πίνακα 1.

Πίνακας 1. Εκτίμηση αξίας ομάδων NBA Ιανουάριος 2015 (Πηγή: Miller, 2015)

Κατάταξη	Ομάδα	Αξία σε εκ \$	Χρέος/Αξία (%)	Κέρδος σε εκ. \$	Έσοδα (μετά την αφαίρεση λειτουργικών εξόδων) σε εκ. \$
1	Los Angeles Lakers	2600	2	293	104,1
2	New York Knicks	2500	0	278	53,4
3	Chicago Bulls	2000	3	201	65,3
4	Boston Celtics	1700	9	173	54,9
5	Los Angeles Clippers	1600	0	146	20,1
6	Brooklyn Nets	1500	19	212	-99,4
7	Golden State Warriors	1300	12	168	44,9
8	Houston Rockets	1250	8	175	38
9	Miami Heat	1175	8	188	12,6
10	Dallas Mavericks	1150	17	168	30,4
11	San Antonio Spurs	1000	8	172	40,9
12	Portland Trail Blazers	940	11	153	11,7
13	Oklahoma City Thunder	930	15	152	30,8
14	Toronto Raptors	920	16	151	17,9
15	Cleveland Cavaliers	915	22	149	20,6
16	Phoenix Suns	910	20	145	28,2
17	Washington Wizards	900	14	143	10,1
18	Orlando Magic	578	17	143	20,9
19	Denver Nuggets	855	1	136	14
20	Utah Jazz	850	6	142	32,7
21	Indiana Pacers	830	18	139	25
22	Atlanta Hawks	825	21	133	14,8
23	Detroit Pistons	810	23	144	17,6
24	Sacramento Kings	800	29	125	8,9
25	Memphis Grizzlies	750	23	135	10,5
26	Charlotte Hornets	725	21	130	1,2
27	Philadelphia 76ers	700	21	125	24,4
28	New Orleans Pelicans	650	19	131	19
29	Minnesota Timberwolves	625	16	128	6,9
30	Milwaukee Bucks	600	29	110	11,5

Οι επαγγελματικές αθλητικές ομάδες ανταγωνίζονται σίγουρα η μία την άλλη στην αγορά εργασίας και η εργατική δύναμη με τη μορφή πρωταγωνιστών είναι ελλιπής. Ορισμένοι υποστηρίζουν ότι τα ανώτατα όρια μισθών είναι απαραίτητα για τη διατήρηση της ανταγωνιστικής ισορροπίας. Τα ανώτατα όρια μισθών βοηθούν επίσης τις ομάδες στον περιορισμό των δαπανών για παίκτες. Τα περισσότερα επαγγελματικά αθλήματα στις Ηνωμένες Πολιτείες έχουν ανώτατα όρια αποδοχών.

Οι ομάδες του NBA έχουν ανώτατο όριο αποδοχών 70 εκατομμυρίων δολαρίων για τη σεζόν 2015-16, με ποινές για τις ομάδες που υπερβαίνουν το ανώτατο όριο. Οι μέγιστοι μισθοί παικτών βασίζονται σε ένα ποσοστό του ανώτατου ορίου και εμπειρία τους (πίνακας 2).

Πίνακας 2. Μισθοί με βάση την εμπειρία επαγγελματιών παικτών του NBA (Πηγή: Adams, 2022)

Έτη εμπειρίας	Μισθός
0	\$1.017.781
1	\$1.637.966
2	\$1.836.090
3	\$1.902.133
4	\$1.968.175
5	\$2.133.278
6	\$2.298.385
7	\$2.463.490
8	\$2.628.597
9	\$2.641.682
10+	\$2.905.851

Τα έσοδα από τα MME είναι σημαντικά για επιτυχημένες αθλητικές ομάδες. Άλλα έσοδα προέρχονται από επιχειρηματικές συνεργασίες, χορηγίες, διαφημίσεις και δικαιώματα ονομασίας σταδίων. Οι ομάδες επιδιώκουν να χτίσουν ένα καλό όνομα, αναπτύσσοντας μια θετική φήμη στο μυαλό των καταναλωτών. Οι παίκτες, όπως και οι οπαδοί, έλκονται από ομάδες με φήμη για σκληρή δουλειά, θάρρος, δίκαιο παιχνίδι, ειλικρίνεια, ομαδική

εργασία και κοινωνική υπηρεσία. Ο χαρακτήρας μιας ομάδας είναι συχνά τόσο σημαντικός όσο και η πιθανότητα να κερδίσει. Οι φίλαθλοι μπορεί να είναι πιστοί σε ένα άθλημα, σε μια ομάδα ή σε μεμονωμένους παίκτες. Οι πολυπαραγοντικές μέθοδοι μπορούν να βοηθήσουν στην κατανόηση του πώς σκέφτονται οι καταναλωτές, αποκαλύπτοντας τις σχέσεις μεταξύ προϊόντων ή επωνυμιών. Από πολλές απόψεις, οι επαγγελματικές αθλητικές ομάδες είναι σαφώς διαφορετικές από άλλες επιχειρήσεις. Βρίσκονται στη δημοσιότητα και έχουν άμεση αμφίδρομη σχέση με τα ΜΜΕ καθώς ένα σημαντικό, αν όχι το σημαντικότερο, μέρος των εσόδων τους προέρχονται από αυτά. (Adams, 2022; Fort, 2011; Leeds & von Allmen, 2014)

Ο αθλητισμός αντιπροσωπεύει, επίσης, ένα εργαστήριο για την έρευνα της αγοράς εργασίας και είναι ένας από τους λίγους κλάδους στους οποίους η εργασιακή απόδοση και η αμοιβή αποτελούν γνώση του κοινού. Οι οικονομικές μελέτες εξετάζουν τα μέτρα απόδοσης των παικτών και την αξία των μεμονωμένων παικτών για τις ομάδες (Kahn, 2000)) και ο Early (Early, 2011) παρέχει μια εικόνα για τις εργασιακές και φυλετικές διακρίσεις στον επαγγελματικό αθλητισμό. Οι αγορές αθλητικών στοιχημάτων έχουν μελετηθεί εκτενώς από οικονομολόγους επειδή παρέχουν δημόσιες πληροφορίες σχετικά με την τιμή, τον όγκο και τα ποσοστά απόδοσης. Επιπλέον, οι ευκαιρίες αθλητικών στοιχημάτων έχουν καθορισμένους χρόνους έναρξης και λήξης και δημοσιεύουν αποδόσεις ή πόντους, γεγονός που καθιστά ευκολότερη τη μελέτη τους από πολλές ευκαιρίες χρηματοοικονομικής επένδυσης. Ως αποτέλεσμα, οι αγορές αθλητικών στοιχημάτων έχουν γίνει ένα εικονικό εργαστήριο πεδίου για τη μελέτη της αποτελεσματικότητας της αγοράς.

Ο Sauer (1998) παρέχει μια ολοκληρωμένη ανασκόπηση των οικονομικών των αγορών στοιχηματισμού. Όταν οι στόχοι διαχείρισης μπορούν να καθοριστούν με σαφήνεια με μαθηματικούς όρους, τότε οι ομάδες χρησιμοποιούν μαθηματικές μεθόδους προγραμματισμού. Επιπλέον, προσπαθούν να μεγιστοποιήσουν τα έσοδα ή να ελαχιστοποιήσουν το κόστος που υπόκεινται σε γνωστούς περιστασιακούς παράγοντες. Έχει γίνει εκτενής δουλειά στα προγράμματα του πρωταθλήματος, για τα οποία ο στόχος μπορεί να είναι να έχουν οι ομάδες να παίξουν μεταξύ τους ίσες φορές, ελαχιστοποιώντας παράλληλα τη συνολική απόσταση που διανύθηκε μεταξύ των πόλεων. Εναλλακτικά, οι

υπεύθυνοι του πρωταθλήματος μπορεί να αναζητήσουν προγράμματα εντός/εκτός, φόρμουλες κατανομής εσόδων ή σχέδια κανόνων λοταρίας που μεγιστοποιούν την ανταγωνιστική ισορροπία. Ο Briskorn (2008) εξετάζει μεθόδους για τον προγραμματισμό αθλητικών αγώνων, βασιζόμενος στον προγραμματισμό ακέραιων αριθμών, τη συνδυαστική και τη θεωρία γραφημάτων. Ακόμη, ο Wright, (2009) παρέχει μια επισκόπηση της επιχειρησιακής έρευνας στον αθλητισμό.

1.3.2 Ανάλυση αθλητικών δεδομένων για την βελτίωση μιας ομάδας

Για τους προπονητές και τους ερευνητές, η συστηματική μοντελοποίηση αθλητικών δεδομένων μπορεί να είναι χρήσιμη επειδή προσφέρει την ευκαιρία να προσδιορίσουν την περιοδικότητα-επαναληψιμότητα των αγώνων και τα τυχαία χαρακτηριστικά των γεγονότων του παιχνιδιού σύμφωνα με το επιθετικό και αμυντικό παιχνίδι. Προφανώς η πληροφόρηση για την απόδοση είναι κρίσιμη για την επίτευξη ατομικής και ομαδικής αποτελεσματικότητας, καθώς αποτελεί βασικό κριτήριο για την προπονητική διαδικασία.

Αρκετοί συγγραφείς έχουν προσπαθήσει να περιγράψουν σημαντικά χαρακτηριστικά τακτικής απόδοσης των ομάδων (Lames & McGarry, 2007; McGarry et al., 2002; Palut & Zanone, 2005). Σε αυτό το εύρος, η μοντελοποίηση έχει χρησιμοποιηθεί για την εύρεση προτύπων μεταξύ των γεγονότων του αγώνα, σύμφωνα με τα χαρακτηριστικά που παρέχουν στους παίκτες και την επιτυχία ή την αποτυχία της ομάδας. Όπως αναφέρθηκε από τους Lames & Hansen (Lames & Hansen, 2001a), είναι σημαντικό να αναρωτηθεί κανείς εάν τα μοντέλα περιέχουν τα βασικά χαρακτηριστικά του αρχικού αθλήματος που παρατηρήθηκε. Αυτός είναι ο λόγος για τον οποίο, πρόσφατα, η έρευνα αθλητικών παιχνιδιών έχει συνειδητοποιήσει ότι μια άλλη πτυχή της διαδικασίας κατασκευής μοντέλων ίσως δεν έχει λάβει αρκετή προσοχή: ο σκοπός του μοντέλου.

Προκειμένου να επιτευχθεί μια βαθύτερη εικόνα του τακτικού παιχνιδιού, είναι απαραίτητο να καταγράφονται οι ουσιαστικές τακτικές ενέργειες με χρονολογική, διαδοχική σειρά, ώστε να μπορεί να αναγνωριστεί η ροή τακτικής συμπεριφοράς, ανάγοντας το πρόβλημα στην ανάλυση και μοντελοποίηση χρονικών σειρών (Pfeiffer & Perl, 2006). Λαμβάνοντας υπόψη την ανάλυση αθλητικών δεδομένων ως το αντικείμενο ανάλυσης σύνθετων αλληλεπιδράσεων, η συστημική προσέγγιση οδηγεί στην εξέταση,

μεταξύ άλλων, δύο κύριων οργανωτικών επιπέδων: τον αντίπαλο και την ομάδα (Gréhaigne & Godbout, 1995).

Ένας αγώνας αποτελεί ένα πολύπλοκο σύστημα καταστάσεων και η κεντρική έννοια του ανταγωνισμού οδηγεί στον υπολογισμό των δύο ομάδων ως αλληλεπιδρώντα οργανωμένα συστήματα (Gréhaigne et al., 1997). Το παιχνίδι μπορεί να θεωρηθεί ότι επιδρά σε περιοχές μετα-σταθερότητας (Kelso, 1995), όπου μεμονωμένες ενέργειες μπορεί να χρησιμεύσουν για την αποσταθεροποίηση ή (επανα)σταθεροποίηση του συστήματος αυτού. Η ευκολία με την οποία ένας επιτιθέμενος ή ένας αμυντικός μπορεί να αποσταθεροποιήσει ή, σε γενικές γραμμές, η ικανότητα μιας ομάδας να αποσταθεροποιήσει ή να (επανα)σταθεροποιήσει ένα σύστημα θα μπορούσε να εξεταστεί σε κρίσιμες στιγμές ενός παιχνιδιού, πχ. σε περίπτωση εμφάνισης μιας απροσδόκητης αλλαγής κατοχής μπάλας.

Η μοντελοποίηση ενός δυναμικού συστήματος σημαίνει χαρτογράφηση όχι μόνο των στοιχείων και της συμπεριφοράς εισόδου-εξόδου, αλλά και ειδικότερα της αλληλεπίδρασης των στοιχείων του (McGarry et al., 2002). Από αυτή την άποψη, οι πληροφορίες σχετικά με τις διαδικασίες αλληλεπίδρασης που δημιουργείται από τη διαδραστικότητα των συμπαικτών και των αντιπάλων δείχνει μια εξαιρετική συνάφεια επειδή η παρατήρηση του τρόπου με τον οποίο συμβαίνει η αλληλεπίδραση σε μια ταυτόχρονη και ανταγωνιστική κατάσταση μπορεί να διευκολύνει το σχεδιασμό μιας συγκεκριμένης και επωφελούς προετοιμασίας.

Η άποψη του Garganta (Garganta, 2009) είναι ότι η έρευνα δεν έχει προχωρήσει σημαντικά περισσότερο από την αρχική εργασία των McGarry & Franks (McGarry et al., 2002) και Hughes et al. (Hughes et al., 1998) για την ανάπτυξη νέων και περιεκτικών μεθόδων δυναμικής ανάλυσης αθλητικών αγώνων. Ωστόσο, οι αναλύσεις δυναμικών συστημάτων μπορεί να κρατήσουν το κλειδί για την εύρεση της «κρυφής λογικής» της αθλητικής απόδοσης και της μεταβλητότητας της. Η δυνατότητα αυτών των μοντέλων να συγκεντρώνουν εξαιρετικά πολύπλοκη συμπεριφορά σε απλές εκφράσεις έχει επιβεβαιωθεί και προσφέρουν σημαντικό πλεονέκτημα έναντι της εντατικής εργασίας και αναποτελεσματικής προσέγγισης που απαιτείται στην παραδοσιακή στατιστική ανάλυση.

Για να περιγράψουν και να ερμηνεύσουν τις ακολουθίες παιχνιδιών σε διαφορετικά αθλήματα, οι Anguera et al. (Anguera & Hernández Mendo, 2013) πρότειναν ένα αξιολογούμενο εργαλείο - τη Μεθοδολογία Παρατήρησης. Σε αυτό το πεδίο, ορισμένοι συγγραφείς έχουν χρησιμοποιήσει τη διαδοχική ανάλυση και την τεχνική πολικών συντεταγμένων στα έργα τους.

Ο Garganta (Garganta, 2008) πρότεινε μια προσέγγιση για την παρατήρηση του παιχνιδιού που βασίζεται σε ένα σχέδιο ανάλυσης διπλού επιπέδου:

- i) τη δημιουργία ενός θεωρητικού χάρτη με σχετικούς δείκτες απόδοσης αγώνα σχετικά με την τακτική οργάνωση
- ii) την παρατήρηση της σειράς των αγώνων και εκμετάλλευση δεδομένων που προέρχονται τόσο από ποιοτική όσο και από ποσοτική ανάλυση της οργάνωσης της ομάδας και του παίκτη

Μια τέτοια πρόθεση είναι πολύ προκλητική λόγω της φύσης και της διαφορετικότητας των περιορισμών που ανταγωνίζονται για την επιτυχία ανάλυσης δεδομένων για τους εξής λόγους:

- i) η πολυπλοκότητα σχετικά με τις άφθονες σχέσεις μεταξύ των παικτών
- ii) το γεγονός ότι τα γεγονότα του παιχνιδιού δεν αντιστοιχούν σε μια προβλέψιμη ακολουθία ενεργειών
- iii) την οξεία ευαισθησία των συμπεριφορών της ομάδας και των παικτών στις αρχικές συνθήκες, λαμβάνοντας υπόψη τον μεγάλο αριθμό μεταβλητών και την αλληλεπίδρασή τους

Για παράδειγμα, σε αθλητικούς κλάδους όπως το ποδόσφαιρο, το μπάσκετ ή το χάντμπολ, οι ομάδες ανταγωνίζονται για την κατοχή της μπάλας, η οποία πρέπει να περάσει μέσα από ένα τέρμα, ενώ στο βόλεϊ, οι ομάδες περνούν τη μπάλα σε μια προσπάθεια να τη φέρουν σε επαφή με περιοχή του αγωνιστικού χώρου των αντιπάλων. Οι ομάδες που συμμετέχουν σε έναν αγώνα συμπεριφέρονται παρόμοια με αυτό-οργανωμένα συστήματα που αναζητούν τάξη και σχήμα σε ένα μακροσκοπικό σχέδιο, σύμφωνα με τις αλληλεπιδράσεις που παράγονται από τους παίκτες.

Η ατομικότητα και οι βαθμοί ελευθερίας της απόδοσης της ομάδας εξαρτώνται από έναν αριθμό παικτών και τις πιθανές αλληλεπιδράσεις τους στο παιχνίδι. Κάθε ομάδα στοχεύει να διαταράξει ή να σπάσει την ισορροπία των αντιπάλων, με πρόθεση να δημιουργήσει αταξία στον οργανισμό του. Από την άλλη πλευρά, οι ομάδες σκοπεύουν να εξασφαλίσουν τη δική τους σταθερότητα και οργάνωση. Με αυτόν τον τρόπο, οι ενέργειες που εκτελούνται κατά μήκος των αγώνων τείνουν να εξασφαλίζουν πλεονέκτημα χώρου και χρόνου έναντι του υποψήφιου, πράγμα που σημαίνει ότι η αναμέτρηση καθορίζει, συνήθως, έναν νικητή και έναν ηττημένο. Επειδή οι ομάδες αντιπροσωπεύουν δυναμικά συστήματα οργανωμένα σύμφωνα με αρχές και κανόνες, η συμπεριφορά των παικτών και της ομάδας δημιουργείται από τη διασταύρωση μεταξύ του συνήθους τρόπου παιχνιδιού και των καινοτομιών που μπορεί να εφαρμόσουν κατά την διάρκεια του αγώνα. Υπό αυτή την έννοια, οι ομάδες προχωρούν ως εξειδικευμένα συστήματα που κυριαρχούνται έντονα από τη στρατηγική και τις ικανότητες για την εύρεση και εφαρμογή νέων μεθόδων παιχνιδιού.

Ο Leon Teodorescu (Teodorescu, 1985) ισχυρίστηκε ότι η ανάλυση αθλητικών δεδομένων δεν μπορεί να περιοριστεί σε οποιοδήποτε μοντέλο αλγορίθμου, επειδή η ομαδική δράση δεν αντιπροσωπεύει προβλέψιμες ακολουθίες. Ο Gréhaigne (Gréhaigne, 1992) κάνει αναφορά στην εύρεση ενός συγκεντρωτικού συλλογισμού και ενισχύει αυτή την ιδέα αναφερόμενος ότι εάν το, ίσως τεράστιο, πλήθος των πιθανών αποτελεσμάτων-αποφάσεων περιοριστεί σε έναν αλγόριθμο δυαδικής επιλογής, τότε ο ερευνητής θα οδηγηθεί σε περιορισμό στην ανάλυση του παιχνιδιού. Οι Lames & Hansen (Lames & Hansen, 2001b) ισχυρίστηκαν ότι η πολυ-αιτιατική δομή της διάγνωσης στην ανάλυση αθλητικών δεδομένων απαιτεί μια ερμηνευτική και όχι αλγοριθμική προσέγγιση. Η σύγκρουση της τακτικής οργάνωσης της ομάδας και του παίκτη δίνει τη δυνατότητα αναγνώρισης γεγονότων του παιχνιδιού, δηλαδή την αναγνώριση κάποιου προτύπου που εκφράζει προτιμητέους τρόπους ή μορφές δράσης και τους διακριτικούς χαρακτήρες που δείχνουν τη μεταβλητότητα των συμπεριφορών και των γεγονότων.

Οι Lames & McGarry (Lames & McGarry, 2007) υποστηρίζουν ότι το ορατό μέρος της παρατήρησης σε ένα αθλητικό παιχνίδι είναι μια δυναμική διαδικασία αλληλεπίδρασης κατά την οποία λαμβάνονται μέτρα και αντίμετρα σε μια προσπάθεια να ξεπεραστεί ο

αντίπαλος. Αυτό σημαίνει ότι η συμπεριφορά που παράγεται δεν είναι πρωτίστως η έκφραση σταθερών ιδιοτήτων των μεμονωμένων παικτών. Σε αυτό το πλαίσιο, η συμπεριφορά λήψης αποφάσεων θεωρείται αποδοτικότερη στο επίπεδο της σχέσης με το περιβάλλον του παίκτη και θεωρείται ότι προκύπτει από τις αλληλεπιδράσεις ατόμων με περιβαλλοντικούς περιορισμούς π.χ. τερέν, καιρικές συνθήκες κα. με την πάροδο του χρόνου, συγκεκριμένους λειτουργικούς στόχους. . Επομένως, η δυσκολία έγκειται στο ότι μια επαρκής ερμηνεία αριθμητικών και οπτικών δεδομένων πρέπει να λαμβάνει υπόψη μεμονωμένες συνθήκες (τακτική, στρατηγική), αλλά και περιστασιακές πτυχές όπως οι φυσικές και γνωστικές διαδικασίες κατά τη διάρκεια του παιχνιδιού, η ποιότητα του αντιπάλου και το επίπεδο προετοιμασίας, δηλαδή προσθέτοντας και τους νόμους της φυσικής μέσα στην ανάλυση αθλητικών δεδομένων .

1.4 Επεξήγηση εννοιών

Σε αυτή την παράγραφο παρουσιάζονται μερικές βασικές έννοιες του παιχνιδιού της καλαθοσφαίρισης που θα επιτρέψουν στον αναγνώστη να κατανοήσει με μεγαλύτερη ευκολία τα αποτελέσματα του ερευνητικού μέρους.

Θέσεις παικτών

Σύμφωνα με τον (Abhishek, 2021) οι θέσεις των 5 παικτών (βλ. και εικόνα 1) σε μια ομάδα καλαθοσφαίρισης είναι:

Center: Ο σέντερ είναι ο ψηλότερος παίκτης σε κάθε ομάδα, παίζοντας κοντά στο καλάθι. Στην επίθεση, ο σέντερ προσπαθεί να σκοράρει σε κοντινές βολές και ριμπάουντ, αλλά στην άμυνα, το κέντρο προσπαθεί να μπλοκάρει τα σουτ των αντιπάλων και να επαναφέρει τα άστοχά τους.

Power Forward: Ο πάουερ φόργουορντ έχει παρόμοιες ευθύνες με τον σέντερ, παίζοντας κοντά στο καλάθι και υπερασπίζεται ψηλότερους παίκτες. Επιπρόσθετα, οι πάουερ φόργουορντ κάνουν μακρύτερες βολές από τους σέντερ.

Small Forward: Ο σμολ φόργουορντ παίζει απέναντι σε μικρούς και μεγάλους παίκτες. Περιφέρονται, μαρκάρουν και σκοράρουν σε όλα τα σημεία του γηπέδου.

Point Guard: Ο πόιντ γκαρντ κινείται στην επίθεση και συνήθως είναι το άτομο που θα κατεβάσει την μπάλα στην αντίπαλη περιοχή καθώς είναι ο καλύτερος τριπλαδόρος και πασέρ της ομάδας. Ο πόιντ γκαρντ αντιμετωπίζει τον αντίπαλο πόιντ γκαρντ και προσπαθεί να κλέψει την μπάλα.

Shooting guard: Ο σούτινγκ γκαρντ είναι συνήθως ο καλύτερος σουτέρ της ομάδας. Ο σούτινγκ γκαρντ μπορεί να κάνει σουτ από μεγάλη απόσταση (τρίποντα) και είναι επίσης καλός τριπλαδόρος.



Εικόνα 1. Τυπική διάταξη πεντάδας (Πηγή: Abhishek, 2021)

Στατιστικά παιχνιδιού

Σύμφωνα με τον Oliver ,(2022), **οι Τέσσερις Παράγοντες ή Four Factors** αποτελούν την βασικό συστατικό για την νίκη μιας ομάδας και χρησιμοποιούνται στους στατιστικούς πίνακες των ομάδων και των παικτών. Αυτοί είναι:

Σουτ με συντελεστή βαρύτητας 40%

Λάθη με συντελεστή βαρύτητας 25%

Ριμπάουντ με συντελεστή βαρύτητας 20% και

Ελεύθερες βολές με συντελεστή βαρύτητας 15%

Οι συντελεστές βαρύτητας στην συγκεκριμένη περίπτωση έχουν προσδιοριστεί αλλά είναι μεταβλητές τιμές καθώς άλλοι εκτιμητές μπορεί να χρησιμοποιήσουν διαφορετικούς συντελεστές.

Αυτοί οι παράγοντες μπορούν να εφαρμοστούν τόσο στην επίθεση όσο και στην άμυνα μιας ομάδας, κάτι που κατά μία έννοια μας δίνει οκτώ παράγοντες. Στην συνέχεια,

παρατίθεται ένα παράδειγμα εφαρμογής των συντελεστών με αριθμητικά δεδομένα από τους Phoenix Suns για την αγωνιστική περίοδο 2004-05.

Σουτ

Ο συντελεστής βολής υπολογίζεται χρησιμοποιώντας το Ποσοστό αποτελεσματικών προσπαθειών γηπέδου (Efficient Field Goal ή eFG%). Ο τύπος τόσο για την επίθεση όσο και για την άμυνα είναι $(FG^3 + 0,5 * 3P^4) / FGA^5$. Για την επίθεση των Suns αυτό είναι ίσο $(3351 + 0,5 * 796) / 7018 = 0,534$ και για την άμυνα είναι $(3328 + 0,5 * 494) / 7485 = 0,478$ (Oliver, 2022).

Λάθη

Ο συντελεστής των λαθών υπολογίζεται χρησιμοποιώντας το ποσοστό λαθών (TOV⁶). Ο τύπος τόσο για την επίθεση όσο και για την άμυνα είναι $TOV / (FGA + 0,44 * FTA^7 + TOV)$. Στην επίθεση, το ποσοστό των Suns ήταν $1125 / (7018 + 0,44 * 2080 + 1125) = 0,124$, ενώ στην άμυνα ήταν $1131 / (7485 + 0,44 * 1775 + 1131) = 0,120$.

Ριμπάουντ

Ο συντελεστής των ριμπάουντ υπολογίζεται χρησιμοποιώντας το ποσοστό επιθετικών και αμυντικών ριμπάουντ (ORB⁸% και DRB⁹%, αντίστοιχα). Ο τύπος υπολογισμού για τα επιθετικά ριμπάουντ είναι $ORB / (ORB + Opp^{10} DRB)$ και για τα αμυντικά ο τύπος άμυνας είναι $DRB / (Opp ORB + DRB)$. Το 2004-05, το ποσοστό επιθετικών ριμπάουντ των Suns ήταν $967 / (967 + 2550) = 0,275$ και των αμυντικών ριμπάουντ ήταν $2652 / (1233 + 2652) = 0,683$.

Ελεύθερες βολές

³ Field goal – Επιτυχημένο σουτ

⁴ Τρίποντο

⁵ Field Goal Attempt – Βολή

⁶ Turnover

⁷ Free throw attempts – Ελεύθερες βολές συνολικά

⁸ Offensive rebounds – Επιθετικό ριμπάουντ

⁹ Defensive rebound – Αμυντικό ριμπάουντ

¹⁰ Opponent – Τα ριμπάουντ των αντιπάλων

Ο παράγοντας ελεύθερων βολών υπολογίζεται ως η αναλογία επιτυχημένων ελεύθερων βολών προς το σύνολο των προσπαθειών ελεύθερων βολών. Ο τύπος υπολογισμού είναι ενιαίος, τόσο για την επίθεση όσο και για την άμυνα και είναι FT / FGA . Ο ρυθμός ελεύθερων βολών των Suns στην επίθεση ήταν $1556 / 7018 = 0,222$ και στην άμυνα ήταν $1320 / 7485 = 0,176$.

ΚΕΦΑΛΑΙΟ 2: Διαδικασίες διαχείρισης αθλητικών δεδομένων

Σε αυτό το κεφάλαιο παρουσιάζονται οι συνήθεις διαδικασίες πραγματικών αθλητικών δεδομένων. Η παρουσίαση αυτή περιλαμβάνει όλα τα στάδια από την συλλογή και τον τρόπο συλλογής ως την διάθεση τους στον υπεύθυνο λήψης αποφάσεων ή την διοίκηση της ομάδας γενικότερα.

2.1. Τυποποίηση δεδομένων

Το πρώτο βήμα των διαδικασιών της αποτελεσματικής διαχείρισης αθλητικών δεδομένων είναι η τυποποίηση των δεδομένων εντός της ομάδας. Η τυποποίηση δεδομένων και η δημιουργία και αποθήκευση δεδομένων σε έναν οποιοδήποτε οργανισμό απαιτεί τη γνώση των πηγών των δεδομένων. Ορισμένες πηγές δεδομένων είναι συνεπείς σε όλες τις ομάδες. Για παράδειγμα, όλες οι ομάδες χρησιμοποιούν βίντεο, διατηρούν δεδομένα πλαισίων και έχουν αναφορές των κυνηγών ταλέντων - scouting.

Οι ομάδες έχουν, επίσης, τα δικά τους μοναδικά σύνολα δεδομένων. Οι Houston Rockets, για παράδειγμα, χρησιμοποιούν μια ομάδα τσάρτερ παιχιδιών που συλλέγουν δεδομένα από κάθε παιχνίδι που παίζουν οι Rockets. Πολλές ομάδες χρησιμοποιούν, ακόμη, όλο και περισσότερο προηγμένη τεχνολογία για να βοηθήσουν στη συλλογή δεδομένων σχετικά με την προπόνηση και την προετοιμασία, όπως μεμονωμένα μόνιτορ καρδιακών παλμών που φοριούνται κατά τη διάρκεια της προπόνησης και βηματόμετρα για την παρακολούθηση της απόστασης και της ταχύτητας που τρέχει ένας παίκτης. Άλλες πάλι ομάδες χρησιμοποιούν λεπτομερή ψυχολογικά προφίλ για να αξιολογήσουν τους παίκτες. Όλες αυτές οι πηγές δεδομένων πρέπει να προσδιορίζονται σε ένα απόθεμα, δηλαδή σε έναν κατάλογο που περιλαμβάνει χρήσιμες πληροφορίες για τους υπεύθυνους λήψης αποφάσεων.

Ο εντοπισμός και η περιγραφή όλων των πηγών δεδομένων δημιουργεί το απόθεμα δεδομένων του οργανισμού. Κατά τη δημιουργία αυτού του καταλόγου, οι υπεύθυνοι λήψης αποφάσεων του οργανισμού πρέπει να λάβουν υπόψη όλες τις λειτουργίες εντός της ομάδας. Η κάθε λειτουργία έχει ένα μοναδικό σύνολο δεδομένων που μπορεί να δημιουργήσει, να αποθηκεύσει και να αποκτήσει πρόσβαση με τον δικό του τρόπο, το

οποίο μπορεί να επιβραδύνει σημαντικά τη διαδικασία λήψης αποφάσεων. Έτσι, θα πρέπει να δημιουργείται ένα τυπικό σύνολο ορισμών για τα διαφορετικά είδη δεδομένων που χρησιμοποιεί η ομάδα.

Ένα απλό παράδειγμα για αυτή την περίπτωση είναι το όνομα του παίκτη. Για κάθε μεταβλητή, ορίζεται ένα τυπικό όνομα για τη μεταβλητή ("Player"), δηλώνεται μια περιγραφή της μεταβλητής και ορίζεται η μορφή για τα δεδομένα σε όλες τις χρήσεις ("Επώνυμο, Όνομα"). Τώρα, κάθε τμήμα μπορεί να ακολουθήσει τους ορισμούς που αναφέρονται στο απόθεμα και θα εισάγει ονόματα παικτών με την ίδια φόρμα. Αυτή η τυποποίηση δημιουργεί ένα πρότυπο σε ολόκληρο τον οργανισμό, έτσι ώστε, ακόμη και χωρίς βελτιώσεις στη διαχείριση δεδομένων της ομάδας, τα δεδομένα από διαφορετικές ομάδες να συνδυάζονται και να αναλύονται τουλάχιστον πιο αποτελεσματικά.

Η διαδικασία τυποποίησης φαίνεται απλή, αλλά στην πραγματικότητα υπάρχουν πολλοί τομείς στους οποίους μπορεί να αποδειχθεί δύσκολη. Για παράδειγμα, από το 1991 έως το 1996 υπήρχαν τρεις παίκτες με το όνομα «Τζέιμς Γουίλιαμς» στο NFL και δύο από αυτούς ήταν ο «Τζέιμς Ε. Γουίλιαμς». Και οι τρεις έπαιζαν διαφορετικές θέσεις για διαφορετικές ομάδες. Η επανάληψη των ονομάτων καθιστά δύσκολο να διασφαλιστεί ότι ο σωστός παίκτης έχει εντοπιστεί (για να περιπλέξει περαιτέρω τα πράγματα, ένας τέταρτος James Williams έπαιξε στο NFL από το 2000 έως το 2003).

2.2. Συγκέντρωση δεδομένων

Κατά την αξιολόγηση ενός παίκτη, οι υπεύθυνοι λήψης αποφάσεων συχνά αναθέτουν συγκεκριμένες εργασίες συλλογής δεδομένων σε οποιονδήποτε έχει το χρόνο να τις ολοκληρώσει. Για παράδειγμα, εάν ο υπεύθυνος λήψης αποφάσεων θέλει να μάθει σε ποιο ποσοστό των επιθετικών κατοχών της ομάδας του ο υποψήφιος αγγίζει την μπάλα, ο υπεύθυνος λήψης αποφάσεων μπορεί να αναθέσει σε έναν ασκούμενο να παρακολουθήσει κάποιο κλιπ του υποψήφιου και να μετρήσει τις κατοχές.

Στις περισσότερες ομάδες, ο ασκούμενος θα ολοκληρώσει αυτήν την εργασία χρησιμοποιώντας ένα απλό υπολογιστικό φύλλο ή στυλό και χαρτί. Όταν κάποιος από τους υπεύθυνους λήψης αποφάσεων στο τμήμα προσωπικού θέλει πρόσβαση στα δεδομένα, πρέπει να βρει τον ασκούμενο και να το ζητήσει. με την προϋπόθεση ότι ο

ασκούμενος βρίσκεται στο χώρο του και ο υπεύθυνος λήψης αποφάσεων έχει τον χρόνο, χωρίς αυτό να είναι ένα υπερβολικό σενάριο. Για παράδειγμα, ο ασκούμενος μπορεί να έχει σταλεί στο αεροδρόμιο για να παραλάβει κάποιους υποψήφιους που έρχονται για προπόνηση. Τότε ο αναλυτής μπορεί να χρειαστεί να περιμένει μερικές ώρες και να καθυστερήσει τη διαδικασία αξιολόγησης.

Αυτό το βασικό παράδειγμα και ρεαλιστικό σενάριο δείχνει τη σημασία της καλής διαχείρισης δεδομένων και πώς μπορεί να εξοικονομήσει πολύτιμο χρόνο. Εάν, αντί να διατηρεί τις πληροφορίες σε προσωπικό φορητό υπολογιστή ο ασκούμενος, ή κάθε υπάλληλος που μπορεί να έχει επιφορτίσει με τον ρόλο συλλογής δεδομένων, μπορούσε να είχε καταχωρίσει τις πληροφορίες στο αρχείο του παίκτη σε μια κεντρική βάση δεδομένων, τότε όλοι οι υπεύθυνοι λήψης αποφάσεων θα είχαν άμεση πρόσβαση στις πληροφορίες.

Σύμφωνα με τον Alamar, (2013), η πρόσβαση στα περισσότερα δεδομένα εξαρτάται από ένα άτομο στο 44 τοις εκατό των ομάδων και η πρόσβαση σε ορισμένα δεδομένα εξαρτάται από ένα άτομο σε πάνω από το 90 τοις εκατό των ομάδων. Η καλή διαχείριση δεδομένων μειώνει τον χρόνο που αφιερώνεται στην αναζήτηση ατόμων που μπορούν να δώσουν στους λήπτες αποφάσεων πρόσβαση στις πληροφορίες που χρειάζονται και παρέχει στην ομάδα ένα σημαντικό ανταγωνιστικό πλεονέκτημα. Όταν όλα τα δεδομένα είναι συγκεντρωτικά, τα στελέχη του προσωπικού μπορούν να αφιερώσουν περισσότερο χρόνο στην αξιολόγηση και οι προπονητές μπορούν να αφιερώσουν περισσότερο χρόνο στη στρατηγική και την καθοδήγηση, παρέχοντάς τους ένα πλεονέκτημα έναντι του ανταγωνισμού.

Αφού δημιουργηθεί το απαραίτητο απόθεμα δεδομένων της ομάδας και τυποποιηθούν τα δεδομένα, τότε μπορεί να γίνει συγκέντρωση των δεδομένων. Αυτό καθιστά τα δεδομένα πιο εύκολα προσβάσιμα στους υπεύθυνους λήψης αποφάσεων. Όταν όλα τα δεδομένα του οργανισμού αποθηκεύονται σε μια κεντρική τοποθεσία, οι υπεύθυνοι λήψης αποφάσεων μπορούν να έχουν πρόσβαση στις πληροφορίες που χρειάζονται όταν τις χρειάζονται.

Πέρα από την πιο αποτελεσματική πρόσβαση, η συγκέντρωση δεδομένων παρέχει πρόσθετα οφέλη όσον αφορά τη συνέπεια και την ακρίβεια των δεδομένων. Ο συγκεντρωτισμός διασφαλίζει ότι όλοι οι υπεύθυνοι λήψης αποφάσεων βλέπουν τα ίδια δεδομένα. Όταν οι υπεύθυνοι λήψης αποφάσεων λαμβάνουν δεδομένα από διαφορετικές πηγές, είναι συχνά πιθανό να βλέπουν διαφορετικά δεδομένα ακόμα κι αν εξετάζουν τις ίδιες μεταβλητές. Για παράδειγμα, εάν δύο στελέχη του NBA αναλύουν την ίδια προοπτική συμμετοχής ενός αμυντικού γραμμής για το ντραφτ και ο καθένας τους παίρνει ύψος, βάρος και χρόνο διασκελισμού από διαφορετικές πηγές, τότε είναι πολύ πιθανό να λαμβάνουν διαφορετικά δεδομένα. Αυτή η ασυμφωνία θα μπορούσε να οδηγήσει σε διαφορετικά συμπεράσματα σχετικά με τις προοπτικές του παίκτη στο NBA. Ενώ οι υπεύθυνοι λήψης αποφάσεων μπορεί σίγουρα να διαφωνούν σχετικά με το πώς ένας παίκτης θα προβληθεί στο επόμενο επίπεδο ανταγωνισμού, αυτές οι διαφωνίες δεν πρέπει να προέρχονται από διαφορετικά δεδομένα. Οι ομάδες θα πρέπει να καθορίσουν ποιες είναι οι καλύτερες πηγές δεδομένων και στη συνέχεια να ζητήσουν από όλους τους φορείς λήψης αποφάσεων να βασίζονται στα ίδια δεδομένα.

Η ύπαρξη ενός συνόλου συνεπών δεδομένων στα οποία μπορούν να βασιστούν όλοι οι υπεύθυνοι λήψης αποφάσεων αναφέρεται συνήθως ως «μία εκδοχή της αλήθειας» (Alamar, 2013a). Όπως αναφέρθηκε προηγουμένως, η ύπαρξη μιας εκδοχής της αλήθειας παρέχει μεγαλύτερη αξιοπιστία και συνέπεια και έχει το πρόσθετο πλεονέκτημα της εξοικονόμησης χρόνου συνάντησης για συζήτηση της ουσίας αντί για το υπόβαθρο.

Μόλις έχει οριστεί, για παράδειγμα, το σύνολο των δεδομένων που απαιτούνται για την ανάλυση ενός αντιπάλου, τότε όλοι όσοι συμμετέχουν σε μια συνάντηση λήψης αποφάσεων μπορούν να έχουν πρόσβαση στις πληροφορίες και να εξετάσουν τα σχετικά δεδομένα εκ των προτέρων εξοικονομώντας χρόνο καθώς όλοι θα έχουν πρόσβαση στα ακριβή δεδομένα και η ουσία της συνάντησης μπορεί να ξεκινήσει αμέσως. Αυτή η αποτελεσματικότητα δίνει στους προπονητές περισσότερο χρόνο για συζήτηση και ανάλυση (τόσο στη συνάντηση όσο και στην προετοιμασία για τη συνάντηση), γεγονός που παρέχει στην ομάδα ένα ανταγωνιστικό πλεονέκτημα.

Ο συγκεντρωτισμός επιτρέπει, επίσης, δεδομένα υψηλότερης ποιότητας. Τα σφάλματα στα οργανωτικά δεδομένα είναι ένα σημαντικό πρόβλημα γενικά. Μια πρόσφατη έρευνα διαπίστωσε ότι περίπου το 59 τοις εκατό των υπολογιστικών φύλλων που χρησιμοποιούνται για σημαντικές επιχειρηματικές πρακτικές περιέχουν σφάλματα (Panko, 1998). Επιπλέον, μόνο το 31 τοις εκατό των ερωτηθέντων είπε ότι τα δεδομένα ελέγχονται πάντα για σφάλματα πριν χρησιμοποιηθούν στη διαδικασία λήψης αποφάσεων (Alamar, 2013).

2.3. Ποσοτικά δεδομένα

Τα ποσοτικά δεδομένα συχνά μπορεί να είναι παραπλανητικά. Ακριβώς επειδή τα δεδομένα παρουσιάζονται με τη μορφή μέσου όρου ή ποσοστού ή αναλογίας δεν σημαίνει ότι είναι χρήσιμες πληροφορίες. Έτσι είναι απαραίτητο, καταρχάς η πληροφορία να μπορεί να αντιπροσωπεύει τις φάσεις του αγώνα. Η Stats LLC, η οποία είναι εταιρεία αθλητικών δεδομένων, παρουσίασε την τελευταίας τεχνολογίας υπηρεσία δεδομένων SportVu. Το SportVu περιλαμβάνει ένα σύστημα έξι σταθερών καμερών που τοποθετούνται πάνω από ένα γήπεδο μπάσκετ (παρόμοια τεχνολογία χρησιμοποιείται, επίσης, στο μπίτζμπολ και στο ποδόσφαιρο) και αυτές οι κάμερες παρακολουθούν κάθε κινούμενο αντικείμενο στο γήπεδο. Το σύστημα δημιουργεί ένα μεγάλο αρχείο δεδομένων για κάθε παιχνίδι, το οποίο παρέχει τη θέση κάθε παίκτη και την μπάλα είκοσι πέντε φορές κάθε δευτερόλεπτο. Για ένα παιχνίδι σαράντα οκτώ λεπτών είναι 72.000 παρατηρήσεις ή 5,9 εκατομμύρια παρατηρήσεις ανά ομάδα ανά σεζόν. Ακόμη όμως και αυτά τα δεδομένα, δεν μπορούν να βοηθήσουν εάν είναι μεμονωμένα. Κανείς δεν μπορεί να κοιτάξει τα ακατέργαστα δεδομένα εκατομμυρίων σειρών και να αντλήσει οτιδήποτε σημαντικό από αυτά.

Έτσι, η Stats LLC επεξεργάστηκε ορισμένα από τα δεδομένα και υπολόγισε το ποσοστό σουτ του Kevin Durant όταν ντρίμπλαρε τη μπάλα τρεις ή περισσότερες φορές και όταν έκανε σουτ δύο ή λιγότερες φορές. Συγκρίνοντας τους δύο μέσους όρους, φάνηκε ότι το ποσοστό σουτ του Durant σχεδόν διπλασιάστηκε όταν ντρίμπλαρε τη μπάλα δύο ή λιγότερες φορές. Ο στόχος της Stats LLC με την παρουσίαση αυτών των πληροφοριών δεν ήταν να παρουσιάσει λεπτομερείς πληροφορίες αναζήτησης για τον Kevin Durant,

αλλά μάλλον να δείξει τις δυνατότητες του συστήματός της. Ένα στέλεχος του NBA παρατήρησε ότι αυτό το σημείο δεδομένων θα μπορούσε να χρησιμοποιηθεί εναντίον του Durant και της ομάδας του, υποδηλώνοντας ότι τα δεδομένα από την Stats LLC ήταν κατά κάποιο τρόπο χρήσιμα στοιχεία.

Το συμπέρασμα που έβγαλε το στέλεχος ήταν ότι αν οι αντίπαλοι ανάγκαζαν τον Ντουράντ να βάλει την μπάλα στο παρκέ και να ντριμπλάρει περισσότερο, τότε η ικανότητα του στο σκοράρισμα θα έπεφτε σημαντικά. Η αντιμετώπιση του «γεγονότος» της Stats LLC ως δεδομένα (όπως προοριζόταν) επιτρέπει την διαπίστωση ότι θα μπορούσε να αποδειχθεί χρήσιμο, αλλά δεν έχει φτάσει στο σημείο να είναι πληροφορία που μπορεί να γίνει πράξη.

Ποιες ήταν οι αποστάσεις των βολών στους δύο μέσους όρους; Ίσως τα σουτ που ήρθαν μετά από δύο ή λιγότερες ντρίμπλες να περιλάμβαναν περισσότερες αποτυχημένες προσπάθειες καρφωμάτων. Αν ο Ντουράντ ντρίμπλαρε λιγότερο επειδή βρισκόταν πιο συχνά στο άκρο του γηπέδου και απλώς έπαιρνε μια πάσα και ντρίμπλαρε μια φορά στο δρόμο για το κάρφωμα, τότε η σύγκριση αυτού του ποσοστού βολής με ένα κανονικό σουτ στην περίμετρο δεν έχει νόημα καθώς οι δύο μέσοι όροι μετρούν εντελώς διαφορετικές δεξιότητες.

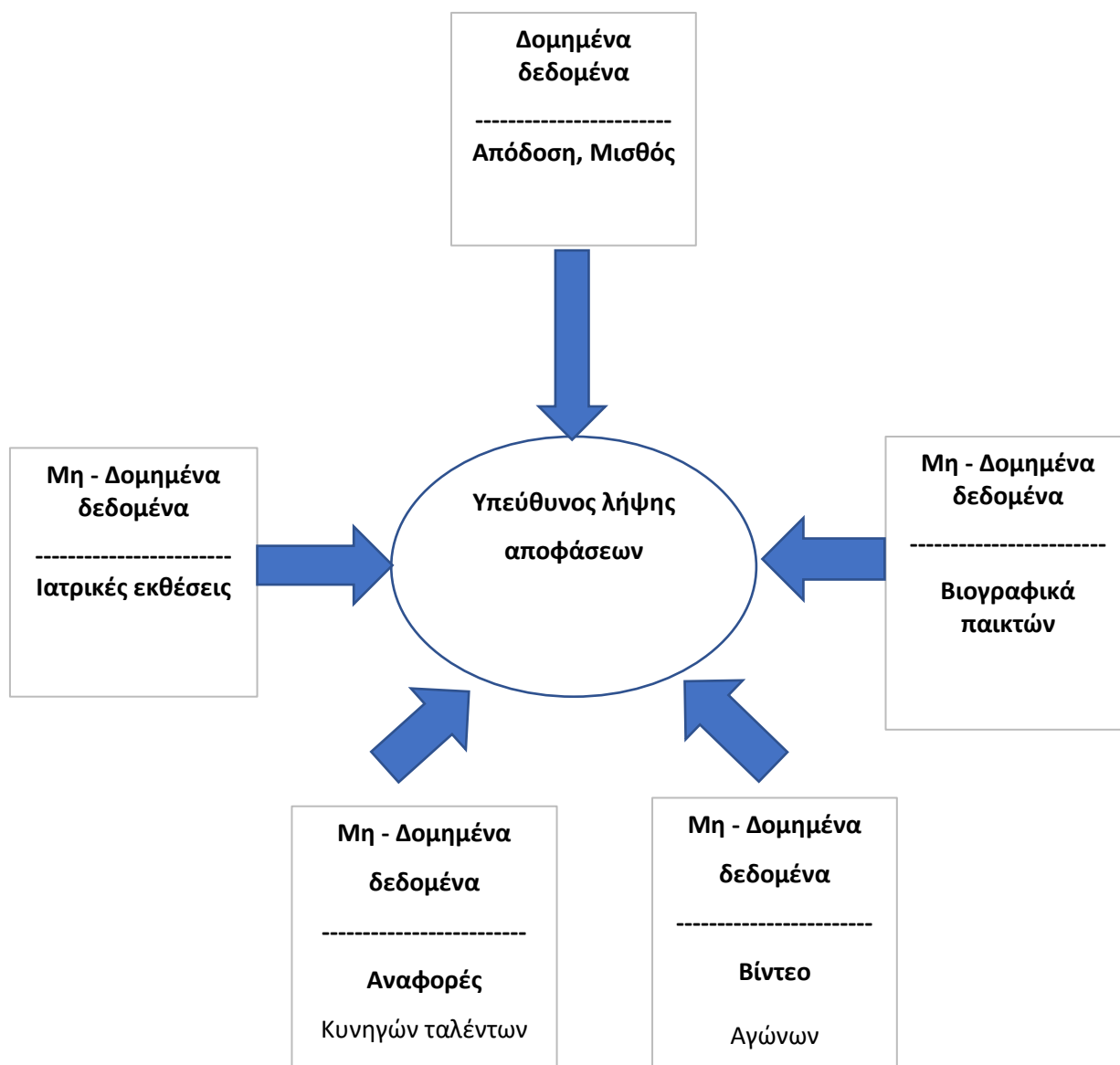
Με αυτό το παράδειγμα γίνεται φανερό ότι τα αριθμητικά δεδομένα δεν έχουν νόημα από μόνα τους. Τα ανεπεξέργαστα δεδομένα δεν παρέχουν στον υπεύθυνο λήψης αποφάσεων και γενικότερα στην διοίκηση χρήσιμες πληροφορίες, επειδή δεν έχουν αναλυθεί επαρκώς. Μόνο αφού γίνει επαρκής ανάλυση σε ακατέργαστα αριθμητικά δεδομένα, παράγονται πληροφορίες που μπορούν να χρησιμοποιηθούν για τη διαδικασία λήψης αποφάσεων. Έτσι είναι σημαντικό για τον αναλυτή όχι μόνο να εκτελέσει μια επαρκή ανάλυση αλλά και να παρέχει την κατάλληλη επεξήγηση αυτών των αποτελεσμάτων.

2.4. Ποιοτικά δεδομένα

Τα περισσότερα ποιοτικά δεδομένα είναι αυτά που είναι γνωστά ως μη δομημένα δεδομένα, που σημαίνει ότι δεν υπάρχουν διακριτά ονόματα μεταβλητών και τα δεδομένα δεν μπορούν να τοποθετηθούν εύκολα και λογικά σε ένα σύνολο γραμμών και στηλών σε ένα υπολογιστικό φύλλο ή σε ένα στατιστικό πρόγραμμα ανάλυσης δεδομένων.

Ορισμένοι οργανισμοί χρησιμοποιούν δομημένες αναφορές για τον εντοπισμό πληροφορίας, στις οποίες οι κυνηγοί (scouters) εισάγουν συγκεκριμένα δεδομένα σε συγκεκριμένα πεδία και αυτά μπορούν να αποθηκευτούν με τον ίδιο τρόπο όπως τα ποσοτικά δεδομένα. Ωστόσο, ακόμη και αυτά συχνά περιλαμβάνουν κάποιο είδος μη δομημένης γραπτής αφήγησης, ενώ όταν τα δεδομένα παίρνουν τη μορφή λέξεων ή εικόνων, τείνουμε να τα σκεφτόμαστε και να τα επεξεργαζόμαστε διαφορετικά από ό,τι με τα ποσοτικά δεδομένα.

Το αποτέλεσμα αυτής της διάκρισης είναι η κατάσταση που απεικονίζεται στην Εικόνα 2. Εδώ, ο υπεύθυνος λήψης αποφάσεων λαμβάνει πληροφορίες από διάφορες πηγές και η ανάλυση είναι ασύνδετη. Τα οφέλη από τη συγκέντρωση και την ενοποίηση δομημένων δεδομένων μειώνονται σημαντικά. Επιπλέον, κάθε είδος δεδομένων αναλύεται χωριστά. Δεν υπάρχει σημείο στη διαδικασία όπου οι διαφορετικοί τύποι δεδομένων και αναλύσεων μπορούν να αλληλοενημερωθούν.



Εικόνα 2. Διοχέτευση δεδομένων χωρίς ενοποίηση

Επειδή τα ποιοτικά δεδομένα μπορεί να μην είναι δομημένα, οι διαφορές στον χειρισμό και την επεξεργασία αυτού του είδους δεδομένων είναι φυσικές, αλλά αυτό δεν σημαίνει ότι τα ποσοτικά και τα ποιοτικά δεδομένα πρέπει να διαχωρίζονται αυστηρά. Τα ακατέργαστα ποιοτικά δεδομένα δεν έχουν μεγαλύτερη σημασία από τα ακατέργαστα ποσοτικά δεδομένα, και πρέπει επίσης να υποβληθούν σε επεξεργασία και να μετατραπούν σε χρήσιμες πληροφορίες. Για παράδειγμα, μια αναφορά εντοπισμού από ένα παιχνίδι μπορεί να παράγει πολλές σελίδες σημειώσεων — ακατέργαστα δεδομένα. Για να είναι χρήσιμα αυτά τα ποσοτικά δεδομένα, πρέπει να συνδυαστούν με άλλες αναφορές όπως πχ. ιατρικές αναφορές, επεξεργασία βίντεο και άλλα είδη δεδομένων που χρησιμοποιεί η ομάδα.

Η γενική στάση απέναντι στα ποιοτικά δεδομένα οδηγεί τους οργανισμούς να τα αποθηκεύουν με πιο απρόσεκτο τρόπο. Δεν είναι ασυνήθιστο μερικά από τα πιο σημαντικά ποιοτικά δεδομένα ενός οργανισμού να βρίσκονται μόνο στους υπολογιστές λίγων ατόμων. Τα ιατρικά δεδομένα, για παράδειγμα, σπάνια οργανώνονται και αποθηκεύονται με την ίδια φροντίδα και δομή με τα δεδομένα μισθών. Συχνά το ιατρικό προσωπικό είναι ο μόνος κριτής για το πού και πώς αποθηκεύονται αυτά τα δεδομένα και ποιος μπορεί να έχει πρόσβαση σε αυτά. Αυτό σημαίνει ότι τα περισσότερα ιατρικά δεδομένα παραμένουν αδόμητα και σπάνια μετατρέπονται σε χρήσιμες πληροφορίες. Το ότι αυτός ο τύπος απρόσεκτης διαχείρισης δεδομένων δημιουργεί προβλήματα είναι ξεκάθαρα εμφανές μέσω της γενικής έλλειψης κατανόησης των μακροπρόθεσμων επιπτώσεων των τραυματισμών στην απόδοση. Τα δεδομένα που θα μπορούσαν να χρησιμοποιηθούν για τον καθορισμό αυτών των επιπτώσεων υπάρχουν σχεδόν σε κάθε αθλητικό οργανισμό, ωστόσο δεν συμβαίνει ποτέ λόγω της φύσης των δεδομένων. Προκειμένου να μεγιστοποιηθεί η απόδοση των αναλυτικών πόρων, όλα τα δεδομένα θα πρέπει να είναι συγκεντρωτικά ώστε να μπορούν να υποβληθούν σε επεξεργασία, να μετατραπούν σε χρήσιμες πληροφορίες και να προσπελαστούν αποτελεσματικά.

2.5. Ενσωμάτωση

Μόλις τα δεδομένα τυποποιηθούν και συγκεντρωθούν, μπορούν να ενσωματωθούν πλήρως. Η ενοποίηση δεδομένων σε όλες τις λειτουργίες εντός της ομάδας επιτρέπει την

απρόσκοπτη πρόσβαση στα δεδομένα κάθε τμήματος. Οι αναφορές του scouter και οι ιατρικές αναφορές συνδέονται με αρχεία βίντεο και οι συνδέσεις συνεχίζονται. Από μόνος του, κάθε τύπος δεδομένων είναι πολύτιμος, αλλά όταν ενσωματώνονται, δημιουργούνται συνέργειες μεταξύ των διαφορετικών πηγών δεδομένων που δεν μπορούν να προκύψουν όταν τα δεδομένα είναι διαχωρισμένα.

Ένας από τους βασικούς τομείς ανησυχίας για την ασυμφωνία των δεδομένων μετά από την ενοποίηση τους είναι οι τραυματισμοί. Όλοι οι υπεύθυνοι λήψης αποφάσεων στον αθλητισμό ανησυχούν για τους τραυματισμούς επειδή είναι σε κάποιο βαθμό ανεξέλεγκτοι και ο αντίκτυπός τους στην καριέρα ενός αθλητή δεν είναι εύκολα κατανοητός.

Αυτός ο τύπος ανάλυσης θα μπορούσε να περιλαμβάνει δεδομένα από εκπαιδευτικά και προπονητικά επιτελεία, δεδομένα απόδοσης και ιατρικά δεδομένα. Ενώ οι περισσότερες από τις απαραίτητες πληροφορίες για αυτό το είδος ανάλυσης υπάρχουν στους περισσότερους επαγγελματικούς αθλητικούς συλλόγους, η απλή συγκέντρωση και οργάνωση όλων των δεδομένων είναι ένα δύσκολο έργο λόγω της έλλειψης κεντρικών και ολοκληρωμένων συστημάτων δεδομένων. Ενώ οι πληροφορίες που θα μπορούσε να παράγει μια τέτοια ανάλυση είναι εξαιρετικά πολύτιμες, τέτοιας, ευρείας κλίμακας, ιστορικές μελέτες σπάνια γίνονται σε αθλητικούς συλλόγους λόγω της τεράστιας προσπάθειας συντονισμού που απαιτείται για να ληφθούν τα δεδομένα σε μια μορφή που μπορεί να αναλυθεί.

Ο Alamar, (2013) αναφέρει ένα παράδειγμα όπου ο γενικός διευθυντής της MLB προσπαθούσε να αναλύσει το προσωπικό των Rays ήξερε ότι χρειαζόταν τα τρία διαφορετικά σύνολα δεδομένων για να πάρει μια απόφαση. Ζήτησε δεδομένα μισθών, αναφορές προσκοπισμού και αναλύσεις από την ομάδα ανάλυσης. Μόλις τα έλαβε, έπρεπε να μπει στη διαδικασία συγχώνευσης των πληροφοριών που του παρουσιάστηκαν. Στην περίπτωση του, περιλάμβανε την ανάγνωση κάθε αναφοράς ξεχωριστά και τη διασταύρωση, λαμβάνοντας τη διακριτή άποψη από κάθε τμήμα.

Ωστόσο, εάν τα δεδομένα ήταν ενσωματωμένα, όλες οι πληροφορίες θα μπορούσαν να παραδοθούν σε μία συνεκτική αναφορά. Αυτή η αναφορά θα μπορούσε να παρουσιάζει τις σχέσεις μεταξύ των διαφορετικών πηγών δεδομένων, τονίζοντας τις διαφορές μεταξύ

των διαφόρων απόψεων κάθε συνάρτησης. Η παρουσίαση των δεδομένων με αυτόν τον ολοκληρωμένο τρόπο επιτρέπει στον υπεύθυνο λήψης αποφάσεων να εντοπίσει και να διερευνήσει τις διαφορές απόψεων με πιο αποτελεσματικό τρόπο.

Για παράδειγμα, εάν τα αναλυτικά δεδομένα δίνουν μια διαφορετική εικόνα ενός παίκτη από τις αναφορές του scouting, οι ενσωματωμένες ιατρικές πληροφορίες μπορεί να μπορούν να εξηγήσουν τις διαφορές. Εάν τα ιατρικά δεδομένα δεν εξηγούν, τότε το ενσωματωμένο βίντεο επιτρέπει στον γενικό διευθυντή να δει τον παίκτη σε δράση και να αποφασίσει μόνος του ποιες πληροφορίες είναι πιο σχετικές. Η ενοποίηση των δεδομένων σημαίνει ότι όλοι οι διαφορετικοί τύποι πληροφοριών παρουσιάζονται μαζί για μια πλήρη εικόνα.

Οι τρεις συνιστώσες της διαχείρισης δεδομένων που συζητούνται εδώ (τυποποίηση, συγκέντρωση και ενοποίηση) παρέχουν τη βάση για ένα αποτελεσματικό σύστημα διαχείρισης δεδομένων που θα παρέχει ανταγωνιστικό πλεονέκτημα εξοικονομώντας χρόνο στους υπεύθυνους λήψης αποφάσεων και δημιουργώντας μια πιο ολοκληρωμένη εικόνα της ομάδας ή του παίκτη. Με ένα αποτελεσματικό και συνεπές σύστημα διαχείρισης δεδομένων, η διαδικασία λήψης αποφάσεων δεν περιλαμβάνει πλέον το άνοιγμα μιας ποικιλίας υπολογιστικών φύλλων και άλλων εγγράφων, καθώς και την πραγματοποίηση μιας σειράς κλήσεων για τη λήψη των απαραίτητων δεδομένων. Όλες οι πληροφορίες είναι διαθέσιμες όταν ο υπεύθυνος λήψης αποφάσεων είναι έτοιμος να ξεκινήσει και είναι λιγότερο πιθανό να χαθεί ένα κομμάτι από τις πληροφορίες επειδή το κατάλληλο άτομο δεν ήταν διαθέσιμο για να το παράγει εγκαίρως.

2.6. Εκτέλεση

Η αξία της ισχυρής διαχείρισης δεδομένων που χρησιμοποιεί τυποποίηση, συγκέντρωση και ενοποίηση είναι αρκετά σαφής. Ωστόσο, η εφαρμογή αυτών των εννοιών μπορεί να είναι πιο περίπλοκη, επειδή απαιτεί επενδύσεις στην τεχνολογία και αλλαγή στη συμπεριφορά όλων των μελών του προσωπικού. Η επένδυση στην τεχνολογία διαχείρισης δεδομένων είναι το πρώτο βήμα και μπορεί να επιτευχθεί είτε μέσω της πρόσληψης προσωπικού για την κατασκευή του συστήματος δεδομένων, της πρόσληψης συμβούλων για την κατασκευή του συστήματος ή της αγοράς λογισμικού "από το ράφι". Μόλις το σύστημα διαχείρισης δεδομένων εγκατασταθεί, πρέπει επιτραπεί η κοινή χρήση

τους σε ολόκληρο τον οργανισμό, κάτι που συνήθως απαιτεί αρκετές γραφειοκρατικές διαδικασίες καθώς τίθενται θέματα ασφάλειας αλλά και προσωπικών δεδομένων.

Η διοίκηση θα πρέπει να θεσπίσει σαφείς κατευθυντήριες γραμμές ως προς το πώς και πού πρέπει να αποθηκεύονται τα δεδομένα, ώστε να μπορεί να επιτευχθεί η πλήρης αξία της επένδυσης. Αυτό μπορεί να είναι μια δύσκολη μετάβαση. Λόγω της συνήθειας και της επιθυμίας για έλεγχο, ορισμένα μέλη του προσωπικού μπορεί να δυσκολεύονται να αλλάξουν τις ροές εργασίας τους.

Προκειμένου να αποφευχθεί αυτό το είδος προβλήματος, τα οφέλη, όσον αφορά την εξοικονόμηση χρόνου και την ελεύθερη πρόσβαση στα δεδομένα, πρέπει να γίνουν σαφή σε όλους τους χρήστες. Μόλις συγκεντρωθούν όλα τα δεδομένα, όλοι οι χρήστες πρέπει να εκπαιδεύονται προσεκτικά για το πώς να χρησιμοποιούν το σύστημα και γιατί πρέπει να χρησιμοποιείται. Μόλις τα μέλη του προσωπικού συνειδητοποιήσουν ότι η χρήση του κεντρικού συστήματος δεδομένων τους εξοικονομεί χρόνο και τον πονοκέφαλο της παρακολούθησης δεδομένων, είναι πολύ πιο πιθανό να το αποδεχθούν.

ΚΕΦΑΛΑΙΟ 3: Επισκόπηση γραφικών μεθόδων ανάλυσης

3.1. Εισαγωγή

Οι βασικές για την ανάλυση των αθλητικών δεδομένων αποτελούνται συνήθως από δείκτες και γραφήματα. Στην περίπτωση του μπάσκετ, είναι διαθέσιμο ένα ευρύ σύνολο τέτοιων αναλυτικών στοιχείων, ξεκινώντας από το πρωτοποριακό έργο του Dean Oliver. Το κύριο μέρος αυτών των αναλυτικών στοιχείων μπορεί εύκολα να υπολογιστεί και να ερμηνευτεί από τους επαγγελματίες. Μεταξύ των αμέτρητων παραδειγμάτων τέτοιων αναλυτικών στοιχείων, γίνεται αναφορά σε εξαιρετικά κοινές αναλύσεις που βασίζονται στις έννοιες της κατοχής και του ρυθμού (Kubatko et al., 2007; Oliver, 2044).

Οι άλλες βασικές στατιστικές με τις οποίες θα ασχοληθεί αυτό το κεφάλαιο είναι εργαλεία περιγραφικών στατιστικών που εφαρμόζονται σε δεδομένα μπάσκετ όπως ραβδογράμματα και ακτινογράμματα (radial plot) που έχουν δημιουργηθεί χρησιμοποιώντας μεταβλητές παιχνιδιού, ποσοστά ή άλλα τυποποιημένα στατιστικά στοιχεία, διαγράμματα διασποράς από δύο επιλεγμένες μεταβλητές, ικανά να αναπαριστούν διάφορα χαρακτηριστικά ομάδων ή παικτών σε ένα μοναδικό γράφημα, εξέταση μεταβλητότητας, διαγράμματα βολών κ.α. Σε αυτό το κεφάλαιο γίνεται επισκόπηση εφαρμογών βασικών και γραφικών μεθόδων που έχουν χρησιμοποιηθεί σε αναλύσεις αθλητικών δεδομένων μαζί με την παρουσίαση αριθμητικών μέτρων και γραφικών απεικονίσεων.

3.2. Τα δεδομένα καλαθοσφαίρισης

Τα δεδομένα είναι ζωτικής σημασίας για την επιστήμη των δεδομένων (Data Science). Ως εκ τούτου, οι διαδικασίες για τη λήψη και την οργάνωση συνόλων δεδομένων πρέπει να είναι δομημένες και επικυρωμένες ώστε να διασφαλίζεται η ποιότητα, η οποία νοείται ως το περιεχόμενο (παρουσία των μεταβλητών που ενδιαφέρουν για τη διεξαγωγή των απαιτούμενων αναλύσεων), ακρίβεια (απουσία ή ελαχιστοποίηση σφαλμάτων), πληρότητα (απουσία ή πιθανή επεξεργασία δεδομένων που λείπουν), συνέπεια (παρουσία μεγάλου αριθμού παρατηρήσεων, απαραίτητων για την ευρωστία των στατιστικών

αναλύσεων), προσβασιμότητα (δυνατότητα εύκολης και γρήγορης εύρεσης σε μορφή αναζήτησης) και επικαιρότητα (συνεχώς ενημερωμένα δεδομένα).

Μια πλήρης συζήτηση σχετικά με τις διαφορετικές προσεγγίσεις για την αξιολόγηση της ποιότητας των δεδομένων παρέχεται από τους Kenett & Shmueli (Kenett & Shmueli, 2016). Ένα άλλο σημαντικό ζήτημα σχετικά με τα δεδομένα είναι το πλαίσιο στο οποίο παρέχονται (Cobb & Moore, 1997), δηλαδή όλες οι πρόσθετες πληροφορίες (ορισμοί, μέθοδοι, τεχνολογίες, συνθήκες, περιβάλλον, κ.λπ.) που περιβάλλουν τα δεδομένα και όποιος τρόπος επηρεάζει τον τρόπο με τον οποίο μπορούμε να τα ερμηνεύσουμε. Τα δεδομένα χωρίς αυτό το γνώρισμα είναι απλώς αριθμοί.

Τα δεδομένα μπορούν να ληφθούν μέσω πολλαπλών καναλιών και πηγών, όπως Εθνικές και Διεθνείς Ομοσπονδίες, αθλητικές οργανώσεις, επαγγελματικές ενώσεις, ενώσεις που σχετίζονται με τον αθλητισμό και άλλες πηγές ειδικού ενδιαφέροντος (Kubatko et al., 2007; Schumaker et al., 2010). Το διαδίκτυο είναι ένας τεράστιος χώρος αποθήκευσης δεδομένων. Τα δεδομένα αυτά μπορεί να είναι επί πληρωμή ή ελεύθερα σε συγκεκριμένους ιστότοπους.

Ως εκ τούτου, οι διαδικασίες συλλογής τους είναι ποικίλες. Η πρόσβαση σε κανάλια ανοιχτού κώδικα απαιτεί μερικές φορές υψηλού επιπέδου δεξιότητες γνώσης μεθόδων άντλησης, όπως εξόρυξη δεδομένων. Επιπλέον, τα σύνολα δεδομένων μπορεί να έχουν διαφορετικά χαρακτηριστικά από άποψη μεγέθους, ποικιλίας και ταχύτητας ενημέρωσης. Μια συνήθης διαδικασία είναι η μετάβαση από μικρά σύνολα δεδομένων σε μεγάλα δεδομένα, έτσι ώστε να είναι ευέλικτα, επεκτάσιμα και καταναμημένα συστήματα. Η ποικιλία των δεδομένων (πχ τύποι δεδομένων, επιλογή παραμέτρων κ.α.) απαιτεί τη διαχείριση σχεσιακών βάσεων δεδομένων και εργαλείων αποθήκευσης δεδομένων (Golfarelli & Rizzi, 2009) προκειμένου να προσαρμόζονται και να χειρίζονται σωστά τα δεδομένα με ένα ευρύ φάσμα διαφορετικών χαρακτηριστικών όπως παραδοσιακά σύνολα δεδομένων διατεταγμένα σε δισδιάστατα πλέγματα (παρατηρήσεις X μεταβλητές), αδόμητα δεδομένα κειμένου, εικόνες ή βίντεο από αισθητήρες και κάμερες, δεδομένα από φορητές συσκευές, κινητά τηλέφωνα, tablet, συντεταγμένες πεδίου

παιχνιδιού, gps δεδομένα με γεωγραφικό πλάτος και μήκος, χρονικές σημάνσεις που δείχνουν χρόνο στο πλησιέστερο χιλιοστό του δευτερολέπτου , κ.α.

Μια, κατά προσέγγιση, ταξινόμηση των δεδομένων σε κατηγορίες, μπορεί να είναι η παρακάτω.

Δεδομένα που καταγράφονται χειροκίνητα, με ή χωρίς τεχνολογικά εργαλεία για σχολιασμό. Αυτή η κατηγορία περιλαμβάνει τα βασικά στατιστικά στοιχεία από τις καταγραφές του αγώνα όπως σκορ, αριθμός φάουλ, αναφορές που συμπληρώθηκαν από τεχνικούς εμπειρογνώμονες και προπονητές κατά τη διάρκεια προπονήσεων, απόψεις και αξιολογήσεις ειδικών που μπορούν να συνδυαστούν με δεδομένα μετρήσεων.

Δεδομένα που ανιχνεύονται από συσκευές τεχνολογίας. Όλο και περισσότερο, η τεχνολογία εισέρχεται τόσο στην προπόνηση όσο και στα παιχνίδια, καθιστώντας διαθέσιμους μεγάλους όγκους δεδομένων. Παραδείγματα είναι τα δεδομένα που καταγράφονται από αισθητήρες GPS ή άλλα συστήματα παρακολούθησης παικτών, τα οποία ανιχνεύουν τις θέσεις των παικτών στο γήπεδο σε πολύ μικρά χρονικά διαστήματα (χιλιοστά του δευτερολέπτου), τα δεδομένα βίντεο που προέρχονται από κάμερες, πλατφόρμες και όλες τις φορητές τεχνολογίες που ανιχνεύουν στάσεις , κινήσεις του σώματος, ζωτικά στοιχεία όπως ο καρδιακός παλμός και η αρτηριακή πίεση.

Δεδομένα από ψυχομετρικά ερωτηματολόγια που χορηγήθηκαν σε αθλητές, με στόχο τη μέτρηση στάσεων και χαρακτηριστικών της προσωπικότητας (δυναμική ομάδα, διαπροσωπικές σχέσεις, κοινωνικές και γνωστικές διαδικασίες, ηγεσία, ψυχική σκληρότητα, προσωπικότητα, στρατηγικές αντιμετώπισης κ.λπ.).

Άλλοι τύποι δεδομένων. Σε αυτή την υπολειπόμενη κατηγορία δεδομένων, συμπεριλαμβάνονται τα δεδομένα που συλλέγονται από τρίτες πηγές όπως πχ. από τα Google Trends και άλλα εργαλεία που μπορούν να παρακολουθούν τις διαδικτυακές αναζητήσεις και τη δημοτικότητα των hashtags (#) .

Τα αρχικά ανεπεξέργαστα δεδομένα (raw data) είναι συχνά αδόμητα, ίσως άναρχα δομημένα και μερικές φορές λείπουν, δηλαδή υπάρχει παρουσία απουσών τιμών ή missing values. Έτσι είναι καθήκον του αναλυτή δεδομένων (Data Scientist) να τα

οργανώσει, να τα καθαρίσει και να τα ολοκληρώσει, επιδιώκοντας την κατασκευή ποιοτικότερων και πιο πληροφοριακών δεδομένων – datasets. Αυτή η προκαταρκτική εργασία χρειάζεται κατάλληλες δεξιότητες και εργαλεία, όπως SQL, Python, R κ.α. Σχετικές αναφορές αυτών των εργαλείων και μεθόδων μπορούν να βρεθούν στους Allison, (2001), Hernández & Stolfo, (1998), Kim et al., (2003) και Little & Rubin, (2014) μέσα από ένα άπειρο πλήθος σχετικών αναφορών.

Μια κατά προσέγγιση ταξινόμηση των δεδομένων σε μακρο-κατηγορίες, είναι η παρακάτω.

1. **Οι βαθμολογίες των ομάδων.** Σε αυτό το σύνολο δεδομένων οι παρατηρήσεις (γραμμές) είναι οι ομάδες που αναλύθηκαν και οι μεταβλητές (στήλες) αναφέρονται στα επιτεύγματα της ομάδας στα εξεταζόμενα παιχνίδια.
2. **Σκορ των αντιπάλων.** Σε αυτό το σύνολο δεδομένων οι παρατηρήσεις (γραμμές) είναι οι ομάδες που αναλύθηκαν και οι μεταβλητές (στήλες) αναφέρονται στα επιτεύγματα των αντιπάλων κάθε ομάδας στα εξεταζόμενα παιχνίδια.
3. **Σκορ των παικτών.** Σε αυτό το σύνολο δεδομένων οι παρατηρήσεις (γραμμές) είναι οι παίκτες που αναλύθηκαν και οι μεταβλητές (στήλες) αναφέρονται στα επιμέρους επιτεύγματα στα εξεταζόμενα παιχνίδια.
4. **Δεδομένα Play-by-play.** Σε αυτό το σύνολο δεδομένων οι παρατηρήσεις (γραμμές) είναι τα γεγονότα που συνέβησαν κατά τη διάρκεια των παιχνιδιών που αναλύθηκαν και οι μεταβλητές (στήλες) είναι περιγραφές των γεγονότων ως προς τον τύπο, τον χρόνο, τους παίκτες που συμμετέχουν, το σκορ, την περιοχή του γηπέδου.
5. **Πρόσθετες πληροφορίες.** Σε αυτό το σύνολο δεδομένων οι παρατηρήσεις (γραμμές) είναι οι ομάδες που αναλύθηκαν και οι μεταβλητές (στήλες) είναι ποιοτικές πληροφορίες όπως διοργάνωση, λίγκα, τελική κατάταξη, πρόκριση για playoffs.

Πίνακας 3. Τυπική ταξινόμηση δεδομένων αγώνων καλαθοσφαίρισης

Μεταβλητή	Περιγραφή	1	2	3	4
Team	Όνομα ομάδας	×	×	×	×
team	Σύντομο όνομα ομάδας				×
Conference	Διοργάνωση				×

Division	Λίγκα				×
Rank	Κατάταξη				×
Playoff	Playoff (Ναι/Όχι)				×
Player	Παίκτης				×
GP	Αριθμός παιχνιδιών	×	×	×	
MIN	Λεπτά παιχνιδιού	×	×	×	
PTS	Πόντοι συνολικά	×	×	×	
W	Νίκες	×	×		
L	Ήττες	×	×		
P2M	Επιτυχίες (Δίποντα)	×	×	×	
P2A	Προσπάθειες (Δίποντα)	×	×	×	
P2p	Ποσοστό επιτυχιών (Δίποντα)	×	×	×	
P3M	Επιτυχίες (Τρίποντα)	×	×	×	
P3A	Προσπάθειες (Τρίποντα)	×	×	×	
P3p	Ποσοστό επιτυχιών (Τρίποντα)	×	×	×	
FTM	Επιτυχίες (Ελ. Βολές)	×	×	×	
FTA	Προσπάθειες (Ελ. Βολές)	×	×	×	
FTp	Ποσοστό επιτυχιών (Ελ. Βολές)	×	×	×	
OREB	Επιθετικά ριμπάουντ	×	×	×	
DREB	Αμυντικά ριμπάουντ	×	×	×	
AST	Ασίστ	×	×	×	
TOV	Λάθη	×	×	×	
STL	Κλεψίματα	×	×	×	
BLK	Μπλοκ	×	×	×	
PF	Ατομικά φάουλ	×	×	×	
PM	Σύν/Πλήν (Διαφορά)	×	×	×	

1: Βαθμολογία ομάδων 2: Σκορ των αντιπάλων 3: Σκορ των παικτών 4: Δεδομένα Play-by-play 5: Πρόσθετες πληροφορίες

Στην συνέχεια παρέχονται παραδείγματα από δεδομένα που αφορούν ολόκληρη την κανονική περίοδο του πρωταθλήματος NBA 2017/2018 (82 αγώνες). Οι βαθμολογίες και οι πρόσθετες πληροφορίες αφορούν όλες τις ομάδες και τους παίκτες, τα δεδομένα play-by-play και αφορούν τους 82 αγώνες που έπαιξαν οι Πρωταθλητές, Γκόλντεν Στέιτ Γουόριορς (Golden State Warriors), κατά τη διάρκεια της κανονικής περιόδου από την Big Data Ball (www.bigdataball.com) και αξιοποιεί μοντέρνες τεχνολογίες για να

εμπλουτίσει και να επεκτείνει τα σύνολα δεδομένων με πολλές μοναδικές μετρήσεις πχ οπτικά μέσα. Οι λειτουργίες της R στηρίζονται στη βιβλιοθήκη BasketballAnalyzeR¹¹

3.3. Βασικές στατιστικές αναλύσεις

Σε αυτή την ενότητα θα συζητηθούν ορισμένα βασικά εργαλεία ανάλυσης αθλητικών δεδομένων με παραδείγματα. Ουσιαστικά, τα αναλυτικά στοιχεία που θα περιγραφούν μπορούν να υπολογιστούν σε οποιαδήποτε ομάδα ή παίκτη για τον οποίο διατίθενται βαθμολογίες σε οποιοδήποτε σύνολο δεδομένων play-by-play, λαμβάνοντας υπόψη έναν αγώνα ή περισσότερα παιχνίδια ομαδοποιημένα σε κάθε στιγμή του πρωταθλήματος ή ένα τουρνουά. Ο κώδικας R για κάθε αποτέλεσμα ή γράφημα παρουσιάζεται στο παράρτημα της εργασίας.

3.3.1 Ρυθμός, Βαθμολογίες, Τέσσερις Παράγοντες

Οι μέθοδοι που εφαρμόζονται σε αυτή την παράγραφο βασίζονται στους Kubatko et al. (Kubatko et al., 2007). Η έρευνα των Kubatko et al. (Kubatko et al., 2007) αποτελεί μια από τις πρώτες έρευνες που καθιέρωσαν ένα κοινό σημείο εκκίνησης για την έρευνα της επιστήμης δεδομένων στον τομέα του αθλητισμού και πιο συγκεκριμένα στην καλαθοσφαίριση. Οι έννοιες που εξετάζονται είναι η κατοχή (POSS) ο ρυθμός (PACE), οι Επιθετικές/Αμυντικές Βαθμολογίες και φυσικά οι Τέσσερις Παράγοντες που έχουν ήδη αναφερθεί στο πρώτο κεφάλαιο.

Ακολουθώντας τους Kubatko et al. (Kubatko et al., 2007) και τους συμβολισμούς του πίνακα 3, τα μέτρα αυτά υπολογίζονται ως

$$POSS = (P2A + P3A) + 0,44 \times FTA - OREB + TOV \quad (3.1)$$

και

$$PACE = 5 \times POSS/MIN \quad (3.2)$$

όπου MIN είναι τα συνολικά λεπτά που έπαιξαν όλοι οι παίκτες. Η αποτελεσματικότητα ανά κατοχή μετράτε ως οι βαθμοί που σημειώνονται ή επιτρέπονται ανά 100 κατοχές,

¹¹ <https://cran.rstudio.com/web/packages/BasketballAnalyzeR/index.html> και <https://github.com/sndmrc/BasketballAnalyzeR>

που ονομάζονται Επιθετική (Offensive Rating ή ORtg) και Αμυντική (Defensive Rating ή DRtg) βαθμολογία – κατάταξη ως εξής ,

$$ORtg = PTS_T / POSS_T \quad (3.3)$$

$$DRtg = PTS_O / POSS_O \quad (3.4)$$

όπου οι δείκτες T(eam) και O(pponent) αναφέρονται στο εάν η μέτρηση αφορά την ομάδα που αναλύθηκε ή τον αντίπαλο. Τέλος, ο τρόπος υπολογισμού των τεσσάρων παραγόντων παρουσιάζεται στον πίνακα 4.

Πίνακας 4. Υπολογισμός τεσσάρων παραγόντων (Πηγή: Kubatko et al. 2007).

Παράγοντας	Ερμηνεία	Επίθεση	Άμυνα
eFG%	Ποσοστό αποτελεσματικών προσπαθειών	$\frac{P_2 M_T + 1.5 \times P_3 M_T}{P_2 A_T + P_3 A_T}$	$\frac{P_2 M_O + 1.5 \times P_3 M_O}{P_2 A_O + P_3 A_O}$
TO Ratio	Ποσοστό λαθών	$\frac{TOV_T}{POSS_T}$	$\frac{TOV_O}{POSS_O}$
REB%	Ποσοστό ριμπάουντ	$\frac{OREB_T}{OREB_T + DREB_O}$	$\frac{OREB_O}{OREB_T + DREB_O}$
FT Rate	Ρυθμός βολών	$\frac{FTM_T}{P_2 A_T + P_3 A_T}$	$\frac{FTM_O}{P_2 A_O + P_3 A_O}$

Με βάση τα μέτρα του πίνακα 4 κατασκευάστηκε το γράφημα 1 με την βοήθεια κώδικα R που παρουσιάζεται στο παράρτημα. Σύμφωνα με το γράφημα 1 διαπιστώθηκαν τα εξής Ρυθμός: ο ρυθμός των αγώνων αυξάνεται μεταβαίνοντας από τους Boston Celtics στους Houston Rockets, Cleveland Cavaliers και Golden State Warriors.

- Βαθμολογίες επίθεσης/άμυνας: Οι Boston Celtics και οι Cleveland Cavaliers έχουν αντίστοιχα χαμηλές και υψηλές τιμές τόσο για ORtg όσο και για DRtg. Οι Golden State Warriors και οι Houston Rockets έχουν την καλύτερη απόδοση, από αυτή την άποψη, καθώς έχουν υψηλό ORtg και σχετικά χαμηλό DRtg.
- Επιθετικοί/Αμυντικοί Τέσσερις Παράγοντες: Οι ράβδοι του γραφήματος 1 αντιπροσωπεύουν, για κάθε ομάδα και κάθε παράγοντα, τη διαφορά μεταξύ της αξίας της ομάδας και του μέσου όρου των τεσσάρων ομάδων που αναλύθηκαν. Η

θετική ή αρνητική τιμή αυτής της διαφοράς και το ύψος της αντίστοιχης ράβδου δίνουν μια ξεκάθαρη εικόνα των δυνατών σημείων και των αδυναμιών των ομάδων σε σχέση μεταξύ τους.



Γράφημα 1. Ρυθμός, Επιθετικές/Αμυντικές Βαθμολογίες και Τέσσερις Παράγοντες (διαφορές μεταξύ της ομάδας και του μέσου όρου των υπό εξέταση ομάδων) – Σαιζόν 2017/2018 για τις ομάδες Boston Celtics, Houston Rockets, Cleveland Cavaliers και Golden State Warriors .

3.3.2. Ραβδογράμματα και διαγράμματα γραμμής

Ένα χρήσιμο γραφικό εργαλείο, που χρησιμοποιείται συχνά για την οπτικοποίηση συγκρίσεων μεταξύ ομάδων ή παικτών σύμφωνα με ορισμένα επιλεγμένα στατιστικά στοιχεία, είναι τα διαγράμματα γραμμής. Στο γράφημα 2, για κάθε ομάδα ή παίκτη που αναλύθηκε, σχεδιάζεται μια ράβδος, της οποίας το ύψος καθορίζεται από μία ή περισσότερες μεταβλητές (στην περίπτωση πολλών ράβδων). Οι ράβδοι μπορούν να κατασκευασθούν σύμφωνα με μια επιλεγμένη μεταβλητή. Επιπλέον, οι σχετικές πληροφορίες μπορούν να προστεθούν στο γράφημα μέσω μιας γραμμής, της οποίας η

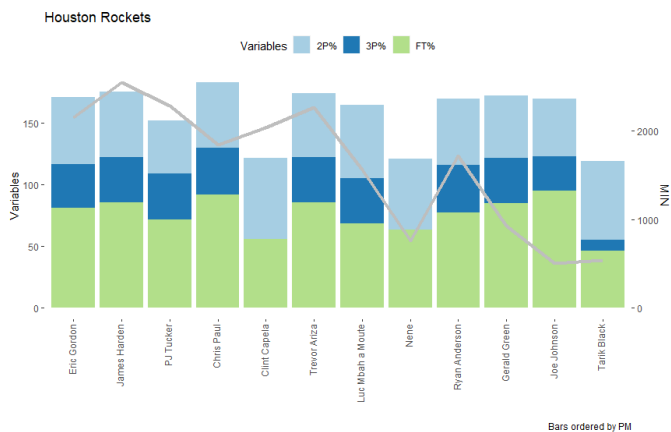
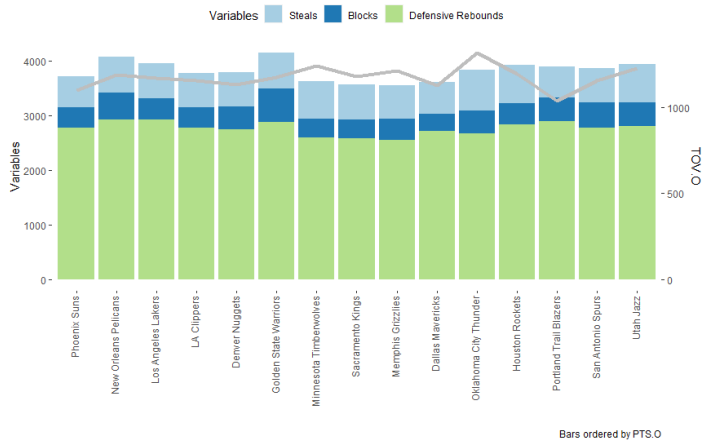
κλίμακα μπορεί να επεξηγηθεί σε έναν άλλο άξονα, συμβατικά τοποθετημένο (συνήθως) στα δεξιά. Δεν υπάρχουν ιδιαίτερα στατιστικά ζητήματα στη σύνταξη αυτού του γραφήματος, αυτό που έχει μόνο σημασία είναι η σωστή επιλογή των μεταβλητών, προκειμένου να επιτευχθεί μια διορατική αναπαράσταση της πραγματικότητας.

Στο γράφημα 2 παρουσιάζονται δύο παραδείγματα για το πώς η επισημαντική γραμμή μπορεί να χρησιμοποιηθεί σε αυτόν τον βαθμό με πολύ ευέλικτο τρόπο. Αντιπροσωπεύει τα κύρια αμυντικά στατιστικά των ομάδων του NBA των δυτικών πολιτειών (West Coast), κλεψίματα, μπλοκ και αμυντικά ριμπάουντ, ταξινομημένα (σε φθίνουσα σειρά) σύμφωνα με τους πόντους που σημείωσαν οι αντίπαλοι (μεταβλητή PTS.Opp). Η γκρίζα γραμμή αντιπροσωπεύει τα λάθη των αντιπάλων (των οποίων η κλίμακα είναι στον δεξιό κατακόρυφο άξονα).

Το γράφημα δεν υπογραμμίζει κάποια εμφανή σχέση μεταξύ των αμυντικών στατιστικών και των πόντων που σημειώθηκαν από τους αντιπάλους της ομάδας ή τα λάθη τους. Στην πραγματικότητα, οι ομάδες στην αριστερή πλευρά του γραφήματος (αυτή με τις υψηλότερες τιμές πόντων που σημείωσαν οι αντίπαλοι, π.χ. Phoenix Suns PTS.Opp = 9290, New Orleans Pelicans PTS.Opp = 9054) δεν είναι απαραίτητα ότι είναι αυτές με τις χαμηλότερες επιδόσεις (μπάρες).

Εναλλακτικά, μπορούν να αναπαρασταθούν στατιστικά στοιχεία παικτών. Ας υποθέσουμε ότι ένας αναλυτής ενδιαφέρεται για τους παίκτες των Houston Rockets που έχουν παίξει τουλάχιστον 500 λεπτά στο πρωτάθλημα, τότε θα πρέπει να σχεδιαστεί μια γραμμή για τα ποσοστά σουτ τους (σουτ 2 και 3 πόντων και ποσοστά ελεύθερων βολών, κατά σειρά στο στατιστικό μέτρο συν-πλην των παικτών) με τις πρόσθετες πληροφορίες των λεπτών που παίχτηκαν (γραμμή όπως εμφανίζεται στο κάτω πλαίσιο του γραφήματος 2). Αυτό το γράφημα δείχνει ότι οι παίκτες στις αριστερές μπάρες (αυτοί με τις υψηλότερες τιμές στατιστικού συν-πλην, π.χ., Eric Gordon PM = 589, James Harden PM = 525) τείνουν να έχουν επίσης τα υψηλότερα λεπτά παιχνιδιού (αντίστοιχα, MIN = 2154 και MIN = 2551). Αυτό όμως δεν σημαίνει απαραίτητα ότι έχουν και την καλύτερη ευστοχία, καθώς υπάρχουν παίκτες με παρόμοια ποσοστά ευστοχίας, αλλά έχουν

σημαντικά χαμηλότερα στατιστικά στοιχεία διαφορά πόντων και λεπτά παιχνιδιού (π.χ. Gerald Green PM = 93 και MIN = 929, Joe Johnson PM = 18 και MIN = 505).



Γράφημα 2. Ραβδόγραμμα γραμμής με αμυντικά στατιστικά των ομάδων του NBA (κορυφή: TOV.O = Σφάλματα των αντιπάλων, PTS.O = Πόντοι που σημειώθηκαν από τους αντιπάλους) - ποσοστά βολής και λεπτά που έπαιζαν οι παίκτες των Houston Rockets (κάτω).

3.3.3. Ακτινωτά διαγράμματα

Ένα άλλο χρήσιμο γράφημα που μπορεί να απεικονίσει τα προφίλ των ομάδων και των παικτών είναι το λεγόμενο ακτινωτό διάγραμμα, όπου οι αριθμητικές τιμές απεικονίζονται ως αποστάσεις από το κέντρο ενός κυκλικού πεδίου στις κατευθύνσεις που ορίζονται από γωνίες που λαμβάνονται διαιρώντας τη γωνία 360 μοιρών. Ενώνοντας τα σημεία που αντιπροσωπεύουν τις αριθμητικές τιμές των μεταβλητών, προκύπτει ένα

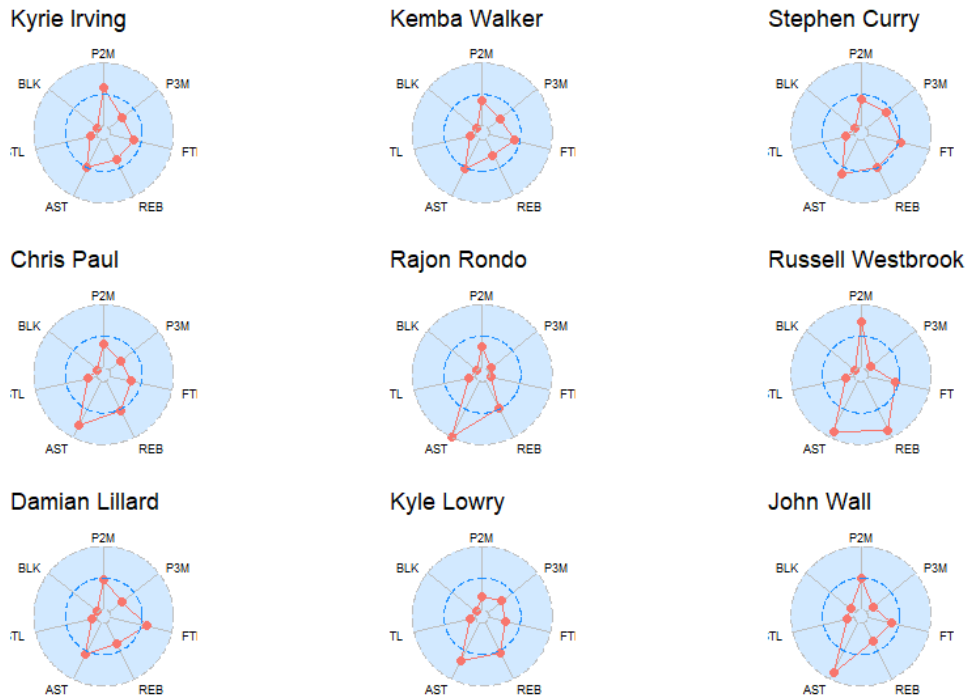
πολύγωνο, που περιγράφει οπτικά το προφίλ του εξεταζόμενου θέματος (ομάδας ή παίκτης).

Ορισμένοι ερευνητές ασκούν κριτική για αυτό το είδος γραφήματος. Σύμφωνα με αυτούς, δυσκολεύει τις συγκρίσεις σχετικά με τα σημεία δεδομένων επειδή τραβάει το βλέμμα στην περιοχή και όχι στην απόσταση κάθε σημείου από το κέντρο. Επιπλέον, η περιοχή καθορίζεται από την αυθαίρετη επιλογή της ακολουθίας των μεταβλητών και υπάρχει γραφική υπερβολική έμφαση σε υψηλούς αριθμούς. Ωστόσο, μπορεί να είναι πολύ χρήσιμο όταν είναι απαραίτητη μια άμεση σύγκριση διαφορετικών προφίλ, όπως, για παράδειγμα, στην ανάλυση συστάδων ή Cluster Analysis η οποία περιγράφεται στις επόμενες παραγράφους.

Ως παράδειγμα αναφέρεται η σύγκριση των προφίλ εννέα point guards, των Russell Westbrook, Stephen Curry, Chris Paul, Kyrie Irving, Damian Lillard, Kyle Lowry, John Wall, Rajon Rondo και Kemba Walker σύμφωνα με σούτ 2 και 3 πόντων που έκαναν, τις ελεύθερες βολές, τα συνολικά ριμπάουντ (επιθετικά και αμυντικά), τις ασίστ, τα κλεψίματα και τα μπλοκ (ανά λεπτό παιχνιδιού).

Η γραφική παράσταση αυτών των δεδομένων παρουσιάζεται στο γράφημα 3 όπου επισημαίνεται ότι, αν και κάθε παίκτης έχει τα δικά του χαρακτηριστικά, υπάρχουν μερικά παρόμοια προφίλ (Irving, Walker και Lillard, Rondo και Wall, Curry και Paul). Παρά τη φαινομενική ευθύτητα του, χρειάζεται μεγάλη προσοχή κατά τη δημιουργία αυτού του γραφήματος. Στην πραγματικότητα, οι άξονες έχουν όλοι την ίδια κλίμακα, που κυμαίνεται από το απόλυτο ελάχιστο έως το απόλυτο μέγιστο (με τη διακεκομμένη μπλε γραμμή στο μέσο μεταξύ τους) όλων των μεταβλητών που αναλύθηκαν. Για το λόγο αυτό, θα πρέπει επιλεγθούν χαρακτηριστικά παιχνιδιού που έχουν περίπου την ίδια κλίμακα. Διαφορετικά, ορισμένες μεταβλητές θα συρρικνωθούν σε άξονες με μια κλίμακα πολύ μεγάλη για αυτές. Αυτό θα είχε ως αποτέλεσμα την ανικανότητα αξιολόγησης των διαφορών μεταξύ των παικτών σε σχέση με αυτές τις μεταβλητές. Κατά κάποιο τρόπο, αυτό συμβαίνει στο γράφημα 2, όπου οι άξονες κυμαίνονται από $\min(X)$ ($\approx 0,005$) έως $\max(X)$ ($\approx 0,313$). Για παράδειγμα, η μεταβλητή STL (κλεψίματα ανά λεπτό

παιχνιδιού) κυμαίνεται από περίπου 0,029 έως 0,052 και οι διαφορές μεταξύ των παικτών σε αυτήν την πτυχή - αν και σημαντικές - δεν είναι οπτικά ανιχνεύσιμες στο γράφημα.



Γράφημα 3. Ακτινωτά διαγράμματα επίδοσης παικτών.

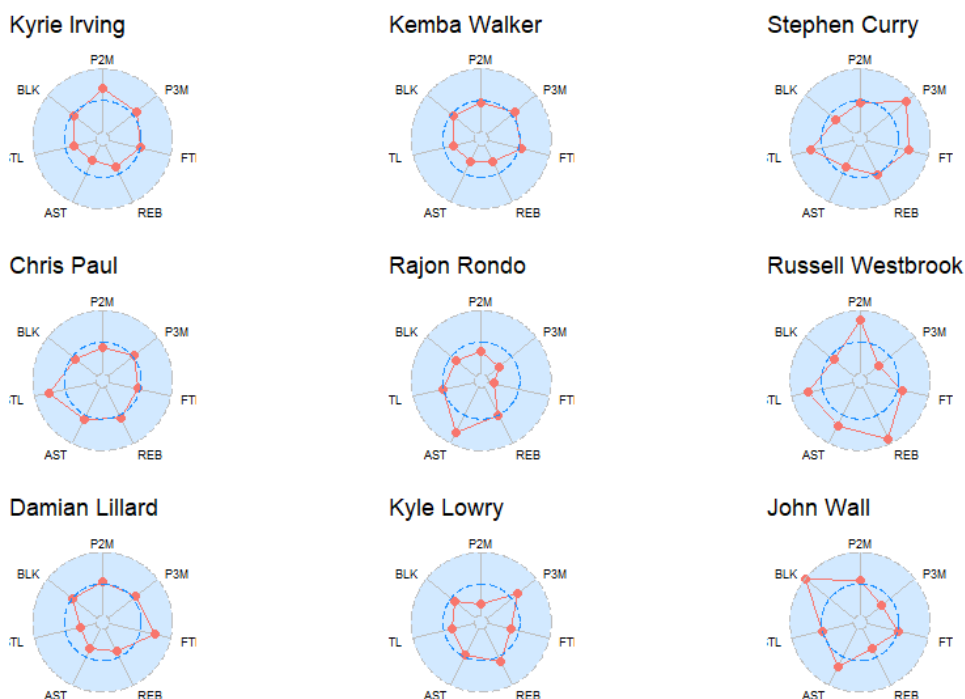
Σε αυτήν την περίπτωση, συνιστάται κατασκευή του γραφήματος με την βοήθεια των τυποποιημένων τιμών τους δηλαδή μετά τον μετασχηματισμό

$$Z = \frac{X - \mu}{\sigma} \quad (3.5)$$

όπου μ η μέση τιμή και σ η τυπική απόκλιση της μεταβλητής X . Το γράφημα που προκύπτει είναι το γράφημα 4 όπου πλέον είναι ορατή η βελτίωση της ευκολίας σύγκρισης των μέτρων των παικτών. Πλέον οι πόιντ γκαρντ Curry και Lowry είναι πιο εύκολα αναγνωρίσιμοι για την ευστοχία τους στα τρίποντα και οι Rondo, Westbrook και Wall για την αστοχία τους. Το ίδιο μπορεί να ειπωθεί για την ευστοχία (Irving και Westbrook) και την αστοχία (Rondo και Lowry) στα σουτ 2 πόντων. Παρόμοια για τα κλεψίματα (υψηλός: Curry, Paul και Westbrook, χαμηλός: Irving, Walker, Lillard και Lowry) και για τις υπόλοιπες αναλυόμενες μεταβλητές.

Λαμβάνοντας υπόψη όλες τις μεταβλητές, το πιο εξαιρετικό προφίλ είναι αυτό του Westbrook, έχοντας όλες τις μεταβλητές πάνω από το μέσο όρο, εκτός από σουτ 3 πόντων και μπλοκ. Δύο σημεία προσοχής αυτής της μορφής του ακτινωτού γραφήματος είναι

1. Ο μέσος όρος υπολογίζεται στους αναλυθέντες παίκτες και όχι σε ολόκληρο το πρωτάθλημα.
2. Θα πρέπει να αποφεύγεται η τυποποίηση για λιγότερους από 5-6 παίκτες.



Γράφημα 4. Ακτινωτά διαγράμματα επίδοσης παικτών με κανονικοποιημένες (standardized) τιμές.

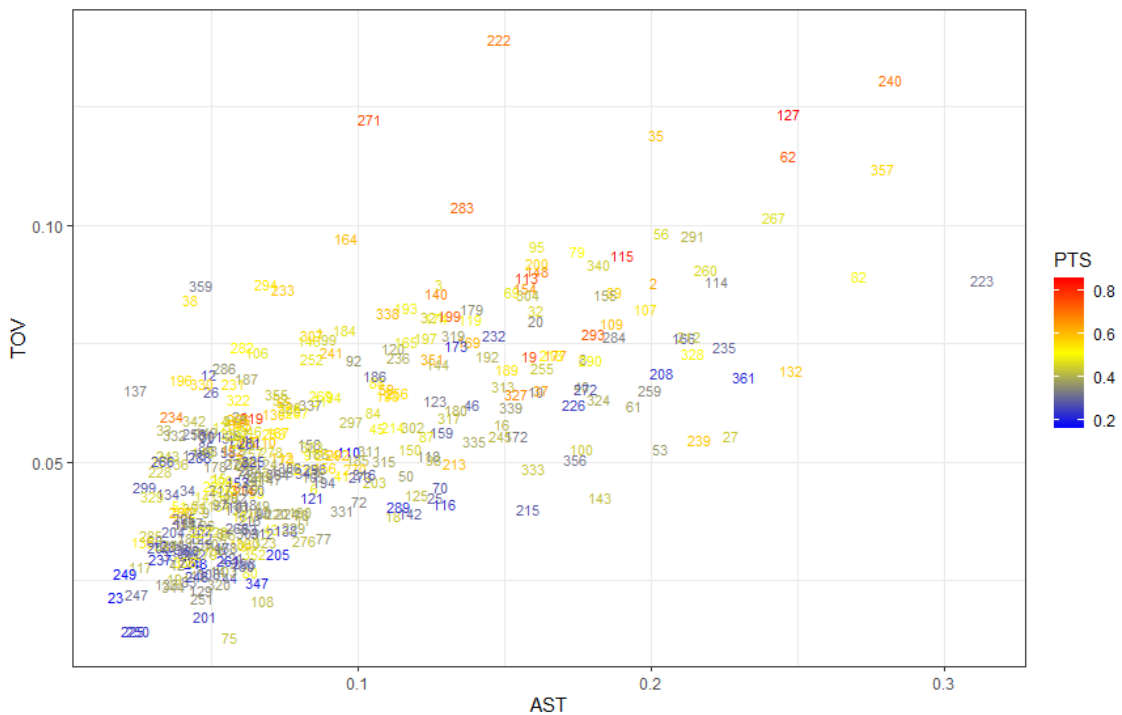
3.3.4. Διαγράμματα διασποράς

Ένα διάγραμμα διασποράς είναι ένα γράφημα που εμφανίζει τιμές για δύο μεταβλητές που παρατηρούνται σε ένα σύνολο δεδομένων χρησιμοποιώντας καρτεσιανές συντεταγμένες. Κάθε παρατήρηση του συνόλου δεδομένων αναπαρίσταται ως ένα σημείο που έχει την τιμή των δύο μεταβλητών ως συντεταγμένες του άξονα x και του άξονα y, αντίστοιχα. Εάν τα σημεία είναι χρωματικά κωδικοποιημένα, το διάγραμμα διασποράς μπορεί να εμφανίσει μία επιπλέον μεταβλητή (ως τη μεταβλητή

χρωματισμού). Τα διαγράμματα διασποράς μπορούν να δώσουν μια γενική ιδέα των σχέσεων μεταξύ των αναλυόμενων μεταβλητών, να προτείνουν διάφορα είδη συσχετισμών και να τονίσουν ανώμαλες περιπτώσεις.

Η συνάρτηση scatterplot επιτρέπει την εμφάνιση διαγραμμάτων διασποράς με ένα σύνολο επιλογών για χρωματική κωδικοποίηση, επισήμανση επιλεγμένων θεμάτων και μεγέθυνση σε ένα υποσύνολο του καρτεσιανού επιπέδου. Ένα παράδειγμα είναι η εξέταση της σχέσης μεταξύ ασίστ και λαθών στο λεπτό όλων των παικτών που έχουν παίξει τουλάχιστον 500 λεπτά κατά τη διάρκεια της κανονικής περιόδου, επισημαίνοντας επίσης, με χρωματικούς κωδικούς, τους πόντους που σημειώνονται ανά λεπτό.

Αυτό το διάγραμμα διασποράς παρουσιάζεται στο γράφημα 5 και δείχνει ότι ο αριθμός των σφαλμάτων τείνει να αυξάνεται με την αύξηση του αριθμού των ασίστ. Αυτό αντικατοπτρίζει μια σαφή εικόνα: όσο περισσότερο ένας παίκτης κινεί την μπάλα, τόσο πιο πιθανό είναι να κάνει λάθη.

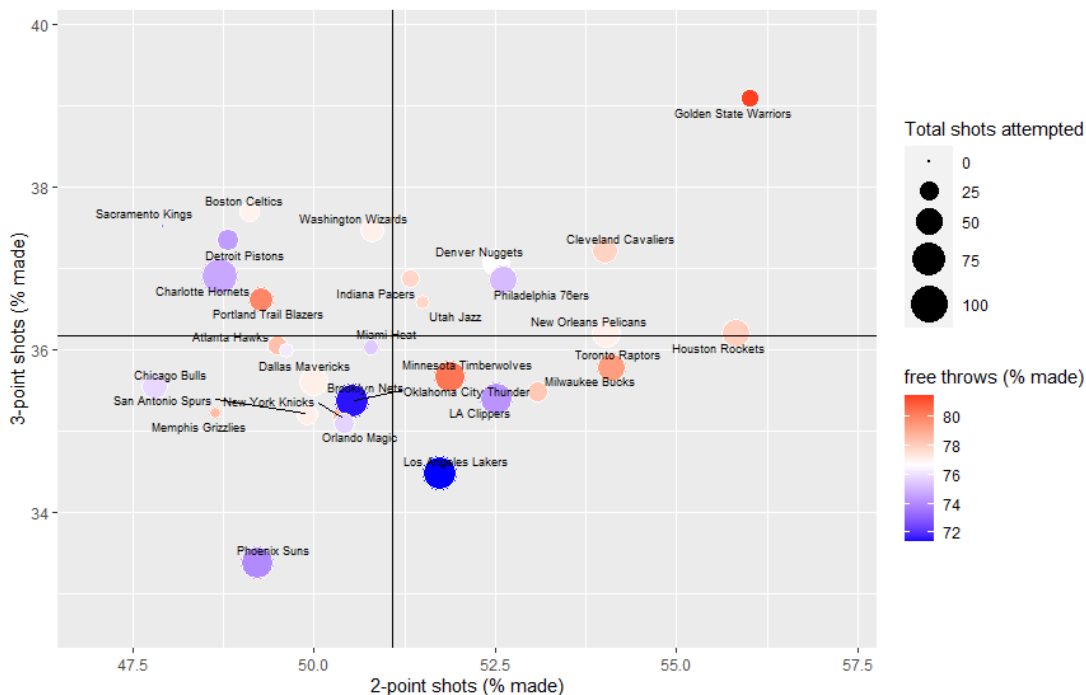


Γράφημα 5. Διάγραμμα διασποράς μεταξύ του αριθμού των λαθών και των ασίστ ανά ποσοστό ευστοχίας.

3.3.5. Διάγραμμα φυσαλίδας (Bubble plot)

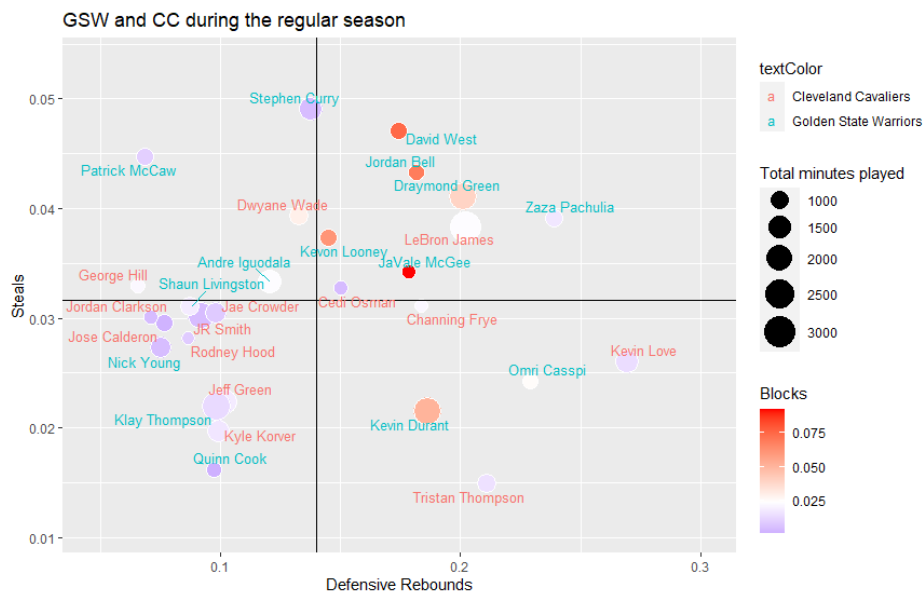
Το διάγραμμα φυσαλίδας αποτελεί μια ειδική περίπτωση του διαγράμματος διασποράς όπου μεμονωμένες περιπτώσεις (ομάδες ή παίκτες) σχεδιάζονται στο επίπεδο μέσω φυσαλίδων αντί σημείων. Το μέγεθος και το χρώμα της φυσαλίδας ορίζονται σύμφωνα με δύο πρόσθετες μεταβλητές, έτσι ώστε η διασπορά των φυσαλίδων να περιλαμβάνει πληροφορίες για τέσσερα χαρακτηριστικά. Μια γραφική παράσταση με φυσαλίδες των ομάδων με μεταβλητές που δίνονται από τα ποσοστά βολής (βολές 2 και 3 πόντων στους άξονες x και y, αντίστοιχα, ελεύθερες βολές σε κλίμακα κόκκινου-μπλε) και τον συνολικό αριθμό των επιχειρούμενων βολών (το μέγεθος των φυσαλίδων) παρουσιάζεται στο γράφημα 6.

Για την βελτίωση της αναγνωσιμότητας, το μέγεθος της φυσαλίδας κλιμακώνεται μεταξύ 0 και 100 (που αντιστοιχεί στον ελάχιστο και τον μέγιστο αριθμό των συνολικών βολών που επιχειρήθηκαν, αντίστοιχα). Από το γράφημα αυτό, γίνεται αμέσως εμφανής η εξαιρετική θέση των Golden State Warriors, οι οποίοι εμφανίζουν τα υψηλότερα ποσοστά αλλά, παραδόξως, χαμηλό αριθμό προσπαθειών.



Γράφημα 6. Διάγραμμα φυσαλίδας των εξεταζόμενων ομάδων με βάση τα ποσοστά επιτυχημένων ελεύθερων βολών και το σύνολο των βολών που επιχειρήθηκαν.

Ένα άλλο παράδειγμα διαγράμματος φυσαλίδας παρουσιάζεται στο γράφημα 7 και αντιπροσωπεύει τους παίκτες δύο ομάδων (εδώ Golden State Warriors και Cleveland Cavaliers), εστιάζοντας μόνο σε αυτούς που έχουν παίξει τουλάχιστον 500 λεπτά στο πρωτάθλημα, σύμφωνα με τα στατιστικά τους σχετικά με την άμυνα: αμυντικά ριμπάουντ, κλεψίματα και μπλοκ ανά λεπτό αγώνα. Στο επάνω δεξιό τεταρτημόριο του γραφήματος που εκφράζει του παίκτες με αμυντικά ριμπάουντ και κλεψίματα πάνω από το μέσο όρο, βρίσκονται κυρίως παίκτες του Golden State Warriors και έχουν επίσης πολύ καλές επιδόσεις στα μπλοκ. Ανάμεσά τους, ο μόνος αξιοσημείωτος παίκτης των Cleveland Cavaliers είναι ο LeBron James με αρκετά υψηλά αμυντικά ριμπάουντ και κλεψίματα, αλλά και μπλοκ κατά μέσο όρο. Αξίζει να αναφερθούν οι David West και Jordan Bell, οι οποίοι επιδεικνύουν εξαιρετικές επιδόσεις, αν και έχουν μικρή διάρκεια συμμετοχής. Στο κάτω αριστερό τεταρτημόριο (παίκτες με αμυντικά ριμπάουντ και κλεψίματα κάτω από το μέσο όρο) υπάρχουν κυρίως παίκτες των Cleveland Cavaliers που τείνουν να έχουν επίσης χαμηλές τιμές για μπλοκ. Ανάμεσά τους βρίσκουμε έναν από τους Golden State Warriors, τον Klay Thompson. Ο Kevin Durant εμφανίζει αμυντικά ριμπάουντ και μπλοκ πάνω από το μέσο όρο, αλλά ο αριθμός κλεψίματος της μπάλας είναι κάτω από το μέσο όρο.



Γράφημα 7. Διάγραμμα φυσαλίδας των παικτών των Golden State Warriors και Cleveland Cavaliers σύμφωνα με στατιστικά στοιχεία που σχετίζονται με την άμυνα (ανά λεπτό παιχνιδιού).

3.3.6. Ανάλυση μεταβλητότητας

Στην περιγραφική στατιστική η μεταβλητότητα ορίζεται ως το μέτρο που δείχνει την απόκλιση των τιμών από τον μέσο όρο, καθώς και το βαθμό στον οποίο τείνουν οι μετρήσεις να διαφέρουν μεταξύ τους. Η μεταβλητότητα αυτή είναι ένας μη αρνητικός πραγματικός αριθμός που υποθέτει την τιμή μηδέν εάν όλα τα δεδομένα είναι ίδια και αυξάνεται καθώς οι τιμές των δεδομένων διαφέρουν-απομακρύνονται μεταξύ τους, και κατά συνέπεια από την μέση τιμή τους.

Για μια δεδομένη μεταβλητή X , υπάρχουν αρκετοί δείκτες μεταβλητότητας αλλά οι πιο συχνά χρησιμοποιούμενοι είναι το εύρος (διαφορά μεταξύ μέγιστων και ελάχιστων τιμών), η ενδοτεταρτομοριακή διαφορά (διαφορά μεταξύ τρίτου και πρώτου τεταρτημορίου) και η διακύμανση σ^2 ή ο αριθμητής της, που ονομάζεται συνολική απόκλιση ή Total Deviance (TD).

Πολύ συχνά, η μεταβλητότητα μετριέται επίσης με την τετραγωνική ρίζα της διακύμανσης, που ονομάζεται τυπική απόκλιση σ . Αναλυτικά, το σ^2 λαμβάνεται με τον μέσο όρο των τετραγώνων αποκλίσεων καθεμιάς από τις N παρατηρήσεις x_i ($i = 1, \dots, N$) της μεταβλητής X από τη μέση τιμή μ_x ως εξής

$$\sigma_x^2 = \frac{\sum_{i=1}^N (x_i - \mu_x)^2}{N} \quad (3.6)$$

και για την τυπική απόκλιση

$$\sigma_x = \sqrt{\sigma_x^2} \quad (3.7)$$

Ένα ακόμη μέτρο μεταβλητότητας είναι ο συντελεστής διακύμανσης (Coefficient of variation ή CV), που υπολογίζεται ως το ο λόγος της τυπικής απόκλισης προς τον μέσο όρο

$$CV = \frac{\sigma_x}{\mu_x} \quad (3.8)$$

και εκφράζει την έκταση της μεταβλητότητας σε σχέση με τον μέσο όρο του πληθυσμού.

Στο μπάσκετ, η μεταβλητότητα μπορεί να αφορά το βαθμό στον οποίο οι παίκτες μιας ομάδας αποδίδουν διαφορετικά μεταξύ τους σύμφωνα με ένα δεδομένο στατιστικό μέτρο. Σε μεταβλητές που σχετίζονται με συγκεκριμένα μέτρα π.χ. ασσίσιτ, ριμπάουντ, κ.λπ., η υψηλή μεταβλητότητα μπορεί να σημαίνει ότι η ομάδα είναι καλά ισορροπημένη σύμφωνα με το εξεταζόμενο μέτρο, δείχνοντας π.χ ότι η ευθύνη του ριμπάουντ κατανέμεται σε όλη την ομάδα και όχι σε λίγους παίκτες. Σε μεταβλητές που σχετίζονται με τη γενική απόδοση π.χ. ευστοχία, η υψηλή μεταβλητότητα σημαίνει ότι η ομάδα εξαρτάται πάρα πολύ από λίγους παίκτες που μπορούν να αποδώσουν καλύτερα, ενώ οι υπόλοιποι είναι κάτω από τα πρότυπα της ομάδας.

Ο Πίνακας 5 δείχνει το ποσοστό των βολών 3 πόντων (P3p) και των βολών 3 πόντων που επιχειρήθηκαν (P3A) για τους επιλεγμένους παίκτες. Ο μέσος όρος, η τυπική απόκλιση, ο συντελεστής διακύμανσης και το εύρος της μεταβλητής P3p ισούνται με 30,41, 10,82, 0,35 και 40,06, αντίστοιχα.

Πίνακας 5. Ποσοστά εύστοχων 3ποντων (P3p) και προσπαθειών (P3A) παικτών της ομάδας Oklahoma City Thunder.

	Player	P3A	P3p
1520	Russell Westbrook	326	29.75460
1320	Paul George	609	40.06568
320	Carmelo Anthony	474	35.65401
1620	Steven Adams	2	0.00000
720	Jerami Grant	110	29.09091
1420	Raymond Felton	230	35.21739
1222	Patrick Patterson	171	38.59649
1104	Alex Abrines	221	38.00905
245	Andre Roberson	36	22.22222
820	Josh Huestis	115	28.69565
1719	Terrance Ferguson	120	33.33333
420	Corey Brewer	67	34.32836

Κατά τη μελέτη της μεταβλητότητας, οι μέσοι όροι και οι τυπικές αποκλίσεις που χρησιμοποιούνται για τον υπολογισμό του συντελεστή διακύμανσης μπορούν προαιρετικά να σταθμιστούν με μια μεταβλητή ενδιαφέροντος. Κάτι τέτοιο γίνεται με την βοήθεια των τύπων

$$w\mu_x = \frac{\sum x_i \cdot w_i}{\sum w_i} \quad (3.09)$$

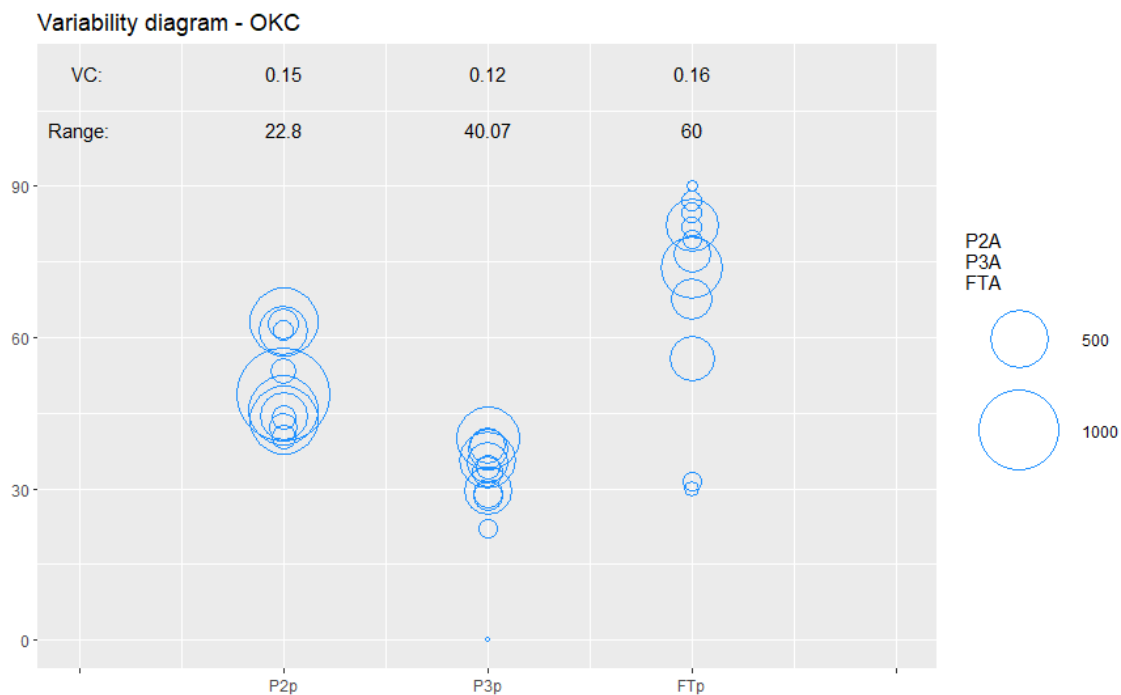
$$w\sigma_x = \sqrt{w\sigma_x^2} = \sqrt{\frac{\sum (x_i - w\mu_x)^2 \cdot w_i}{\sum w_i}} \quad (3.10)$$

Αναλυτικά, ο σταθμισμένος μέσος όρος (3.09) και η σταθμισμένη τυπική απόκλιση (3.10) υπολογίζονται με την βοήθεια του βάρους w_i όπου αντιστοιχεί στο i -ο βάρος, δηλαδή η i -η τιμή της μεταβλητής ώστε κάθε τιμή να σταθμίζεται ξεχωριστά για κάθε παρατήρηση (στο σταθμισμένο μέσο όρο) ή κάθε τετραγωνική απόκλιση από το μέσο όρο (στη σταθμισμένη τυπική απόκλιση), σύμφωνα με τις τιμές της μεταβλητής στάθμισης.

Στο συγκεκριμένο παράδειγμά επιλέχθηκε το P3A ως μεταβλητή στάθμισης και οι νέες τιμές του σταθμισμένου μέσου όρου, της τυπικής απόκλισης και του συντελεστή διακύμανσης ισούνται με 35,35, 4,30, 0,12, αντίστοιχα. Σε αυτήν την περίπτωση, $w_i = P3p = P3M/P3A$.

Στο γράφημα 8 παρουσιάζεται ένα διάγραμμα μεταβλητότητας που μετρά τη μεταβλητότητα των ποσοστών σουτ 2 και 3 πόντων των παικτών της Oklahoma City Thunder που έχουν παίξει τουλάχιστον 500 λεπτά. Το μέγεθος της φυσαλίδας είναι ανάλογο, αντίστοιχα, με τις βολές 2, 3 πόντων και τις ελεύθερες βολές που επιχειρήθηκαν. Οι πιο διάσπαρτες φυσαλίδες σχετίζονται με ελεύθερες βολές που παρουσιάζουν τη μεγαλύτερη μεταβλητότητα, όπως επιβεβαιώνεται από το εύρος και τον συντελεστή διακύμανσης.

Για σουτ 2 πόντων, υπάρχουν δύο ομάδες παικτών με ελαφρώς διαφορετικά επίπεδα απόδοσης. Τέλος, οι βολές 3 πόντων παρουσιάζουν τη χαμηλότερη μεταβλητότητα, όπως επιβεβαιώνεται από τον συντελεστή διακύμανσης. Σε αυτήν την περίπτωση, το εύρος δεν είναι αξιόπιστο μέτρο, καθώς επηρεάζεται σε μεγάλο βαθμό από έναν παίκτη που είναι ακραίος (Steven Adams) με μόνο 2 προσπάθειες. Το γράφημα αυτό επίσης επιτρέπει την εύρεση ακραίων (outliers) και ανώμαλων τιμών που παρουσιάζονται συχνά σε σύνολα δεδομένων. Αυτού του είδους τα δεδομένα μπορούν να επηρεάσουν τις αναλύσεις με διαφορετικούς τρόπους και συνήθως πρέπει να εντοπιστούν και να αντιμετωπιστούν. (Hawkins, 1980). Τέλος, μπορεί να παρατηρηθεί ότι οι παίκτες με τα καλύτερα ποσοστά δεν είναι πάντα αυτοί που επιχειρούν τον μεγαλύτερο αριθμό βολών.



Γράφημα 8. Διάγραμμα φυσαλίδας των παικτών των Golden State Warriors και Cleveland Cavaliers σύμφωνα με στατιστικά στοιχεία που σχετίζονται με την άμυνα (ανά λεπτό παιχνιδιού).

3.3.7. Ανάλυση ανισοτήτων

Η ανισότητα είναι μια έννοια που γεννήθηκε αρχικά στα οικονομικά, για να αντιπροσωπεύσει το εισόδημα ή την κατανομή του πλούτου μιας περιοχής ή μιας χώρας. Σε αυτό το πλαίσιο, η ανισότητα είναι ο βαθμός στον οποίο ένα μικρό τμήμα του πληθυσμού τείνει να κατέχει μεγάλο μερίδιο του συνολικού πλούτου. Τα άκρα αντιπροσωπεύονται από την ίση κατανομή (αν όλοι έχουν το ίδιο μερίδιο πλούτου) ή τη μέγιστη ανισότητα (όταν όλοι δεν έχουν το ίδιο μερίδιο στον πλούτο, εκτός από ένα μόνο άτομο που κατέχει τον συνολικό πλούτο).

Η προηγούμενη ανισότητα μπορεί να παρασταθεί με την καμπύλη του Lorenz. Η καμπύλη Lorenz επινοήθηκε από έναν Αμερικανό οικονομολόγο, τον Max O. Lorenz το 1905 (Lorenz, 1905) ως μέθοδος για τη μέτρηση της συγκέντρωσης του πλούτου.

Αντιπροσωπεύει μια γραφική σχέση μεταξύ της σωρευτικής κανονικοποιημένης κατάταξης του πληθυσμού από τον φτωχότερο στον πλουσιότερο και τον σωρευτικό κανονικοποιημένο πλούτο που κατέχει αυτός ο πληθυσμός από τον φτωχότερο στον πλουσιότερο. Για περισσότερο από έναν αιώνα, η καμπύλη Lorenz χρησιμοποιείται ευρέως για την απεικόνιση των κατανομών του εισοδήματος και του πλούτου και για την εξέταση της ανισότητας σε τέτοιες κατανομές (Sitthiyot & Holasut, 2021).

Η ανάλυση αθλητικών δεδομένων αυτή την έννοια για να διερευνήσει την ανισότητα σε κάποια μεταβλητή του παιχνιδιού μέσα σε μια ομάδα μπάσκετ. Στη συνέχεια, παρουσιάζεται ένα παράδειγμα στην ανισότητα των πόντων, αλλά η ίδια έννοια μπορεί να επεκταθεί σε οποιαδήποτε άλλη μεταβλητή. Πιο συγκεκριμένα, η ανισότητα των πόντων προκύπτει όταν είναι λίγοι οι παίκτες που σκοράρουν μεγάλο μέρος των συνολικών πόντων της ομάδας (με τη μέγιστη ανισότητα να σημαίνει ότι όλοι οι πόντοι σημειώνονται από έναν μόνο παίκτη). Αντίθετα, η ίση κατανομή θα σήμαινε ότι όλα τα μέλη της ομάδας θα σκοράρουν ίδιο αριθμό πόντων. Τα υψηλά επίπεδα ανισότητας υποδηλώνουν μια ομάδα που δεν είναι καλά ισορροπημένη από την άποψη των πόντων που σημειώνονται και τείνει να εξαρτάται πάρα πολύ από λίγους παίκτες.

Στην πράξη, ποτέ δεν παρουσιάζεται μια από τις δύο ακραίες καταστάσεις. Για μέτρηση και την γραφική αναπαράσταση του πόσο απέχει μια πρακτική κατάσταση από την εντελώς ίση κατανομή, χρησιμοποιείται ο συντελεστής Gini και η καμπύλη Lorenz. Ο συντελεστής Gini είναι ένας δείκτης που κυμαίνεται από 0 (ίση κατανομή) έως 100% (μέγιστη ανισότητα) (*Inequality - Income Inequality - OECD Data*, n.d.). Ο μαθηματικός ορισμός του βασίζεται στην καμπύλη Lorenz, η οποία απεικονίζει το κλάσμα y της συνολικής μεταβλητής στον άξονα y που αθροιστικά αναφέρεται στο κάτω κλάσμα x του πληθυσμού. Συνήθως προστίθενται δύο γραμμές στο γράφημα, που αντιπροσωπεύουν τις δύο κατανομές με τέλεια ισότητα και μέγιστη ανισότητα αντίστοιχα.

Στον πίνακα 6 εξετάζονται οι 8 πρώτους παίκτες των Oklahoma City Thunder (που έχουν περισσότερους από 300 πόντους), σε φθίνουσα σειρά πόντων: από τον μικρότερο στον μεγαλύτερο αριθμό πόντων Στο πρώτο βήμα της ανάλυση γίνεται διαίρεση αυτών των μέτρων με τον συνολικό αριθμό των παικτών και τους συνολικούς πόντους για τον

υπολογισμό των αθροιστικών ποσοστών CPI% και CPTS% αντίστοιχα. Αυτό επιτρέπει την εμφάνιση του λόγου των συνολικών πόντων που σημείωσε το i ($i=1:8$) ποσοστό των παικτών. Στο συγκεκριμένο παράδειγμα ότι το 25% των παικτών (1:2) σημείωσαν το 8,39% των συνολικών πόντων, το πρώτο 50% (1:4) σημείωσε το 23,98% των συνολικών πόντων κ.ο.κ.

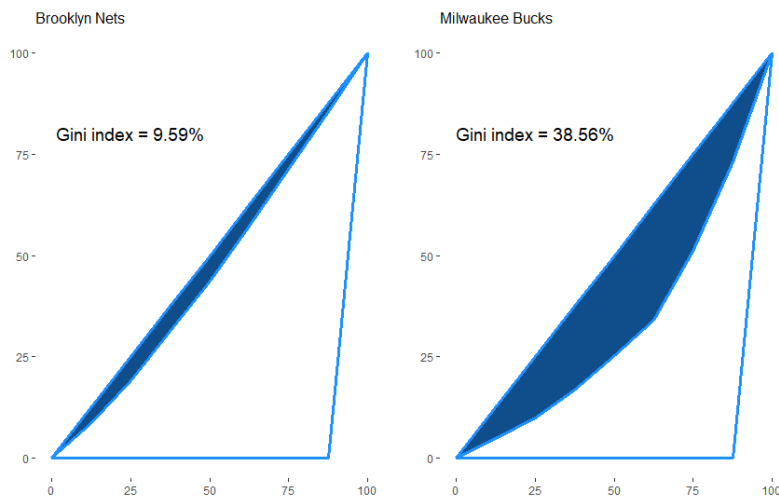
Ο συντελεστής Gini μπορεί στη συνέχεια να ληφθεί ως ο λόγος του εμβαδού της ανισότητας προς το συνολικό εμβαδόν μεταξύ της γραμμής ισότητας και της γραμμής της μέγιστης ανισότητας και υπολογίζεται ως

$$G = \frac{\sum_{i=1}^{N-1} (CPIpct_i - CPTSpcct_i)}{\sum_{i=1}^{N-1} CPI\%_i} \cdot 100 = 29.64\% \quad (3.11)$$

Πίνακας 6. Παίκτες της ομάδας Oklahoma City Thunder με περισσότερους από 300 πόντους

	Player	PTS	CPI	CPTS	CPIpct	CPTSpcct
1222	Patrick Patterson	318	1	318	12.5	3.98
1104	Alex Abrines	353	2	671	25.0	8.39
1420	Raymond Felton	565	3	1236	37.5	15.46
720	Jerami Grant	682	4	1918	50.0	23.99
1620	Steven Adams	1056	5	2974	62.5	37.19
320	Carmelo Anthony	1261	6	4235	75.0	52.96
1320	Paul George	1734	7	5969	87.5	74.64
1520	Russell Westbrook	2028	8	7997	100.0	100.00

Η καμπύλη Lorenz προκύπτει από την ένωση των σημείων με τις συντεταγμένες που δίνονται από CPIpct (άξονας x) και CPTSpcct (άξονας y), όπως φαίνεται στο γράφημα 9, όπου απεικονίζονται επίσης οι γραμμές τέλεισης ισότητας και μέγιστης ανισότητας. Η (μπλε σκιασμένη) περιοχή που βρίσκεται μεταξύ της γραμμής ισότητας και της καμπύλης Lorenz αντιπροσωπεύει τη λεγόμενη περιοχή ανισότητας. Όσο μεγαλύτερο είναι το μέγεθός της, τόσο μεγαλύτερη είναι η ανισότητα.

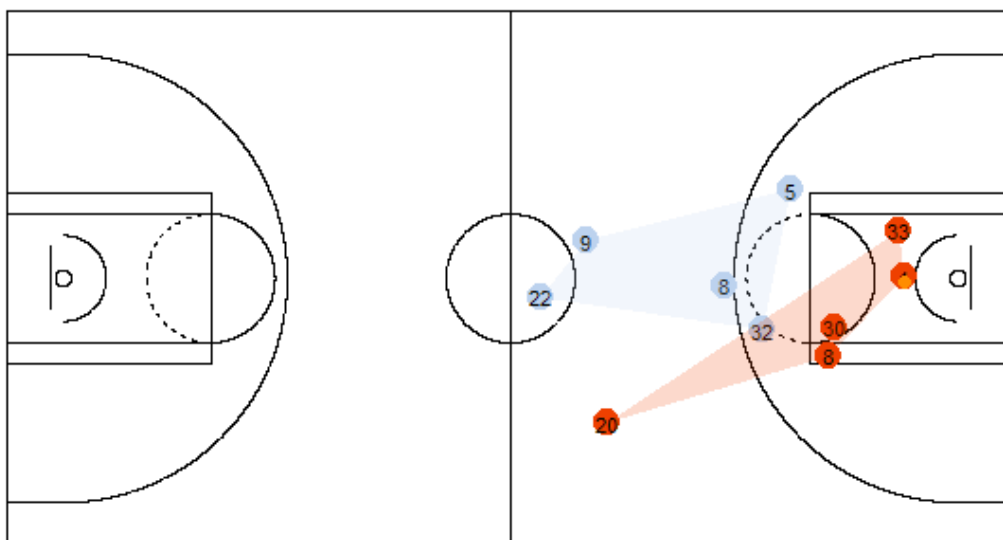


Γράφημα 9. Ανάλυση ανισοτήτων των πόντων που σκόραραν οι ομάδες Brooklyn Nets και Milwaukee Bucks,

3.4 Advanced graphics

Στο τελευταίο μέρος της παρουσίασης βασικών και γραφικών μεθόδων ανάλυσης παρουσιάζονται δύο ιδιαίτερα γραφήματα από πλευράς καινοτομίας στον οπτικό τρόπο παρουσίασης αθλητικών δεδομένων. Το πρώτο παρουσιάζει δεδομένα μέσω κινούμενων γραφικών. Σύμφωνα με τον Smith, (2016), από το 2013, χρησιμοποιείται τηλεμετρία στους επαγγελματικούς αγώνες μπάσκετ NBA στις ΗΠΑ από την SportVU. Έξι κάμερες παρακολουθούν τη θέση των παικτών και της μπάλας εντός γηπέδου, με ανάλυση 25 frames ανά δευτερόλεπτο. Σε συνδυασμό με τα δεδομένα NBA play-by-play (παίκτες, παιχνίδια, φάουλ και βαθμοί που σημειώθηκαν) μπορεί να οδηγήσει σε ένα πλούσιο σύνολο δεδομένων για ανάλυση.

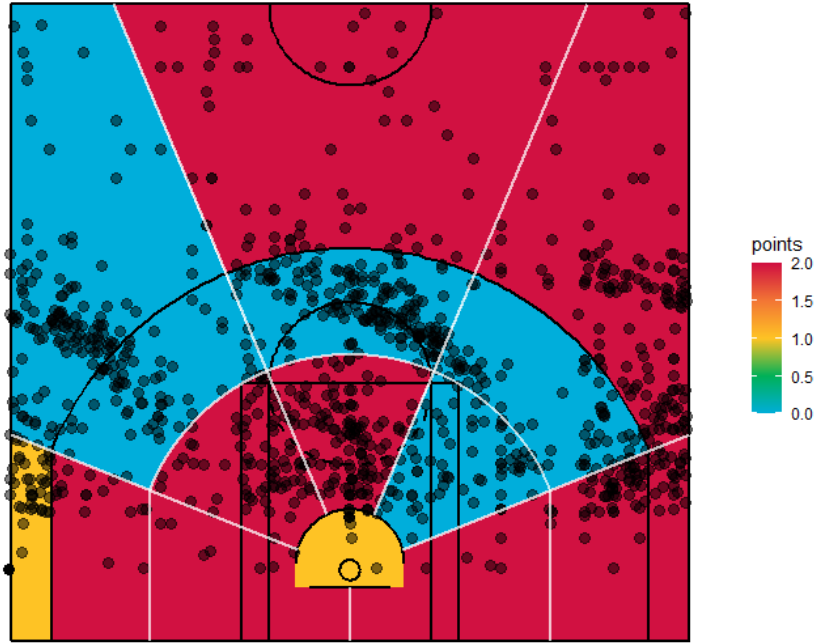
Ο James Curley (Curley, 2016) χρησιμοποίησε τα ίδια δεδομένα και επέκτεινε αυτά τα σενάρια για να ζωντανέψει τα παιχνίδια του NBA μέσα από κινούμενα γραφικά - animation. Στο γράφημα 10 παρουσιάζεται το καλάθι που σημειώθηκε των San Antonio Spurs και των Minnesota Timberwolves τον Δεκέμβριο του 2015. Το πορτοκαλί πολύγωνο είναι ένα μέτρο της απόστασης των παικτών στο γήπεδο. Μέσα από αυτό το απλό γράφημα ένας αναγνώστης μπορεί να έχει μεταβαλλόμενα ή και real time στατιστικά μέτρα για κάθε παίκτη.



Γράφημα 10. Κινούμενο – Animated γράφημα συγκεκριμένης επίθεσης στο NBA (Πηγή: Smith, 2016)

Το δεύτερο advanced graph ονομάζεται shoot chart και είναι ένα χρήσιμο εργαλείο για την εμφάνιση μοτίβων των βολών (ή και άλλων δεδομένων) σε όλο το γήπεδο, προκειμένου να αναλυθούν τα δυνατά σημεία των παικτών στο γήπεδο, με αναφορά σε επιλεγμένες φάσεις του παιχνιδιού.

Το γράφημα 11 δείχνει την θέση και τα ποσοστά ευστοχίας (κλίμακα /2) του παίκτη Kevin Durant. Με την βοήθεια αυτού του γραφήματος ένας προπονητής μπορεί να κατανοήσει ποιες είναι οι θέσεις στις οποίες ο παίκτης είναι πιο εύκολο να σκοράρει δίνοντας χρήσιμες οδηγίες προς την τοποθέτηση του παίκτη κατά την διάρκεια του αγώνα.



Γράφημα 11. Shot chart ποσοστού ευστοχίας για τον Kevin Durant

ΚΕΦΑΛΑΙΟ 4: Εύρεση μοτίβων

Σε αυτό το κεφάλαιο, παρουσιάζονται παραδείγματα εύρεσης μοτίβων (patterns) και διάφορα στατιστικά εργαλεία που μπορούν να αξιοποιηθούν για την ανάλυση δεδομένων. Τέλος, παρουσιάζονται μέθοδοι προσδιορισμού των παραγόντων που καθορίζουν τις καταστάσεις παιχνιδιού πίεσης και επηρεάζουν την πιθανότητα επιτυχούς σκοραρίσματος.

4.1. Δημιουργία μέτρων σχέσεων μεταξύ των μεταβλητών

Με τον όρο εξάρτηση ή συσχέτιση συνήθως εννοείται οποιαδήποτε στατιστική σχέση, είτε αιτιολογική είτε όχι, μεταξύ δύο ή περισσότερων μεταβλητών (Liebetrau, 1983; Merlo & Lynch, 2010). Για να ανιχνευθεί οποιαδήποτε συσχέτιση, όλες οι μεταβλητές πρέπει να μετρούνται με ίδιες κλίμακες. Στο μπάσκετ, η ανακάλυψη συσχετίσεων μέσα σε ένα σύνολο μεταβλητών μπορεί να είναι ενδιαφέρουσα από διαφορετικές οπτικές γωνίες. Για παράδειγμα, μπορεί να αποκαλύψει ποιες μεταβλητές του παιχνιδιού συνδέονται περισσότερο με τους πόντους που σημειώθηκαν, γενικά ή με αναφορά σε μία ομάδα ή ένα παίκτη .

Συνήθως οι σχέσεις μεταξύ των μεταβλητών ταξινομούνται ανάλογα με τη φύση των μεταβλητών που εμπλέκονται. Η διάκριση αυτή γίνεται με βάση τον τύπο της μεταβλητή και είναι κατηγορικές (ή ποιοτικές) και αριθμητικές (ή ποσοτικές) μεταβλητές. Οι κατηγορικές μεταβλητές εκχωρούν κάθε μονάδα παρατήρησης σε μια συγκεκριμένη ομάδα ή κατηγορία με βάση κάποια ποιοτική ιδιότητα. Οι πιθανές τιμές για τις κατηγορικές μεταβλητές είναι κάποια χαρακτηριστικά ή κατηγορίες.

Παραδείγματα κατηγοριοποιημένων μεταβλητών είναι η κατηγορία στην οποία ανήκει μια ομάδα (Ανατολή ή Δύση, πρώτη ή δεύτερη εθνική), το αποτέλεσμα ενός σουτ (επιτυχία/αποτυχία) ή η θέση του παίκτη (πόιντ γκαρντ, σουτ γκαρντ, σμολ φόργουορντ, πάουερ φόργουορντ και κέντρο).

Πιθανές τιμές για τις αριθμητικές μεταβλητές είναι αριθμοί (ακέραιοι, στην περίπτωση διακριτών αριθμητικών μεταβλητών, ή διαστήματα τιμών στην περίπτωση συνεχών αριθμητικών μεταβλητών). Ο συνολικός αριθμός των πόντων που σημειώθηκαν είναι ένα

παράδειγμα διακριτής αριθμητικής μεταβλητής, ενώ ο χρόνος που παίχτηκε στο τρίμηνο είναι μια συνεχής αριθμητική μεταβλητή.

Η διάκριση μεταξύ κατηγορικών και αριθμητικών μεταβλητών είναι σημαντική-γιατί οι στατιστικές μέθοδοι που έχουν σχεδιαστεί για μεταβλητές ενός είδους δεν μπορούν να είναι κατάλληλες για την ανάλυση του άλλου είδους μεταβλητών. Στην πραγματικότητα, υπάρχει μια ιεραρχία, σύμφωνα με την οποία μέθοδοι για κατηγορικές μεταβλητές μπορούν να χρησιμοποιηθούν για την ανάλυση αριθμητικών μεταβλητών (οι τελευταίες έχουν υψηλότερο περιεχόμενο πληροφοριών) και όχι το αντίστροφο. Παρόλα αυτά, λόγω του εύρους φάσματος των διαθέσιμων μεθόδων, είναι πάντα καλύτερη επιλογή να γίνεται ανάλυση με κατάλληλες στατιστικές μεθόδους.

Η ταυτόχρονη εξέταση των σχέσεων μεταξύ πολλών μεταβλητών μπορεί να αντιμετωπιστεί με διάφορες μεθόδους ανάλυσης πολυμεταβλητών δεδομένων (Hardle & Simar, 2015). Εστιάζοντας τώρα στη διμεταβλητή ανάλυση, οι τρεις πιο συνηθισμένοι ορισμοί συσχέτισης μεταξύ δύο μεταβλητών είναι η στατιστική εξάρτηση, η μέση εξάρτηση και η συσχέτιση.

Σύμφωνα με τη φύση των μεταβλητών που αναλύθηκαν, η στατιστική εξάρτηση μπορεί να αξιολογηθεί όταν οι μεταβλητές είναι και οι δύο κατηγορικές ή τουλάχιστον η μία είναι αριθμητική ή και οι δύο είναι αριθμητικές. Η μελέτη της μέσης εξάρτησης απαιτεί τουλάχιστον μία αριθμητική μεταβλητή. Η ανάλυση συσχέτισης απαιτεί και οι δύο μεταβλητές να είναι αριθμητικές.

4.2. Στατιστική ανεξαρτησία

Η στατιστική εξάρτηση μπορεί να μελετηθεί ξεκινώντας από μια αμφίδρομη διασταύρωση και διερευνώντας την ύπαρξη μιας «γενικής» σχέσης μεταξύ των δύο μεταβλητών (Merlo & Lynch, 2010). Ειδικότερα, αυτή η μέθοδος συγκρίνει τον τις παρατηρούμενες και την αναμενόμενες τιμές.

Ο πιο γνωστός δείκτης συσχέτισης με βάση τις διαφορές μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων είναι το X τετράγωνο τεστ ανεξαρτησίας ή X^2 test of independence.

Οι εξεταζόμενες υποθέσεις σε αυτή την περίπτωση είναι

H_0 : Μεταβλητές ανεξάρτητες και

H_1 : Όχι η H_0

Ως παράδειγμα παρουσιάζεται η εξέταση της υπόθεσης

H_0 : Η επίδοση των Golden State Warriors εξαρτάται από την αντίπαλη ομάδα.

Για τον λόγο αυτό υπολογίστηκε ο αριθμός των ελεύθερων βολών, των χαμένων βολών, των ριμπάουντ και των προσπαθειών των Golden State Warriors ανά αντίπαλη ομάδα και τα αποτελέσματα παρουσιάζονται στον πίνακα 7.

Τα αποτελέσματα της εξέτασης ανεξαρτησίας μεταξύ των μεταβλητών έδειξαν ότι η συσχέτιση μεταξύ των δύο μεταβλητών είναι χαμηλή. Πιο συγκεκριμένα, ο δείκτης Cramer's V ισούται με 0,06 ο κλασικός δείκτης του Pearson ισούται με 116,25 και είναι στατιστικά σημαντικός ($p= 0,011$). Τα αποτελέσματα αυτά δείχνουν ότι μέσα από μια χαμηλού βαθμού σχέση, τα στατιστικά ελεύθερες βολές, αποτυχημένες προσπάθειες, ριμπάουντ και βολές εξαρτώνται από την αντίπαλη ομάδα

Πίνακας 7. Μέτρα αγώνα των Golden State Warriors ανά αντίπαλη ομάδα

	free throw	miss	rebound	shot		free throw	miss	rebound	shot
ATL	33	88	81	84	MIL	33	70	74	85
BKN	34	80	98	93	MIN	54	132	142	133
BOS	45	95	90	71	NOP	85	183	175	180
CHA	26	91	90	80	NYK	46	78	79	90
CHI	46	80	98	95	OKC	86	176	179	153
CLE	47	88	95	79	ORL	27	77	92	99
DAL	74	155	188	188	PHI	46	76	88	98
DEN	78	172	164	173	PHX	59	166	178	197
DET	34	75	85	83	POR	60	123	116	125
HOU	56	118	119	131	SAC	76	165	169	159
IND	33	97	90	72	SAS	70	175	169	162
LAC	127	161	166	176	TOR	50	68	66	90
LAL	104	190	202	176	UTA	64	190	167	158
MEM	77	126	128	117	WAS	50	83	89	83
MIA	48	92	92	79					

4.3. Effect size

Το effect size επιτρέπει την σύγκριση του μεγέθους του αποτελέσματος ως προς έναν κατηγορικό παράγοντα με δύο κατηγορίες. Η σύγκριση αυτή γίνεται με το μέτρο η^2 που μετρά το ποσοστό της επιμέρους προς την συνολική διακύμανση της εξαρτημένης μεταβλητής και εξηγεί το ποσοστό της μεταβλητότητας της εξαρτημένης ως προς την ανεξάρτητη μεταβλητή. Παραδείγματα τέτοιου είδους εξέτασης είναι εάν ο μέσος αριθμός πόντων που σημειώνουν όλες οι ομάδες του NBA διαφέρει μεταξύ της Ανατολής και της Δύσης ή μεταξύ των διαφορετικών κατηγοριών ή μεταξύ των προκριθέντων και μη προκριμένων ομάδων στα Playoffs. Σε αυτές τις περιπτώσεις, γίνεται ανάλυση εάν και πώς ο μέσος όρος μιας αριθμητικής μεταβλητής (π.χ. οι βαθμοί που σημειώθηκαν) ποικίλλει μεταξύ των κατηγοριών που ορίζονται από μια άλλη μεταβλητή, η οποία μπορεί να είναι κατηγορική (π.χ. Playoff, με τιμές Ναι ή Όχι).

Στον πίνακα 8, παρουσιάζονται οι μέσες τιμές διαφορών μέτρων παιχνιδιού για τις κατηγορίες (Playoff=Yes) για τις ομάδες που προκρίθηκαν στα playoff και (Playoff=Όχι), δηλαδή των ομάδων που δεν προκρίθηκαν. Οι μεταβλητές του παιχνιδιού που εξετάζονται είναι οι συνολικοί πόντοι, τα 2ποντα, τα 3ποντα, οι ελεύθερες βολές, τα συνολικά ριμπάουντ, οι ασίστ, τα κλεψίματα και τα μπλοκ. Οι τιμές η^2 στον πίνακα 8 στον αριθμό των ριμπάουντ, των ασίστ των ελευθέρων βολών, 2ποντα και 3ποντα δεν εξαρτώνται από την πρόκριση των Playoff ($\eta^2 < 6\%$). Αντίθετα, το σκορ των πόντων που δέχθηκε και έβαλε μια ομάδα εξαρτάται σε μεγάλο βαθμό από τα Playoff ($\eta^2 = 42,53\%$ και $40,25\%$, αντίστοιχα).

Πίνακας 8. Μέσοι όροι των μεταβλητών του παιχνιδιού (υπό όρους έως το Playoff) και οι τιμές του η^2 της μέσης εξάρτησης των μεταβλητών του παιχνιδιού από την πρόκριση των Playoff.

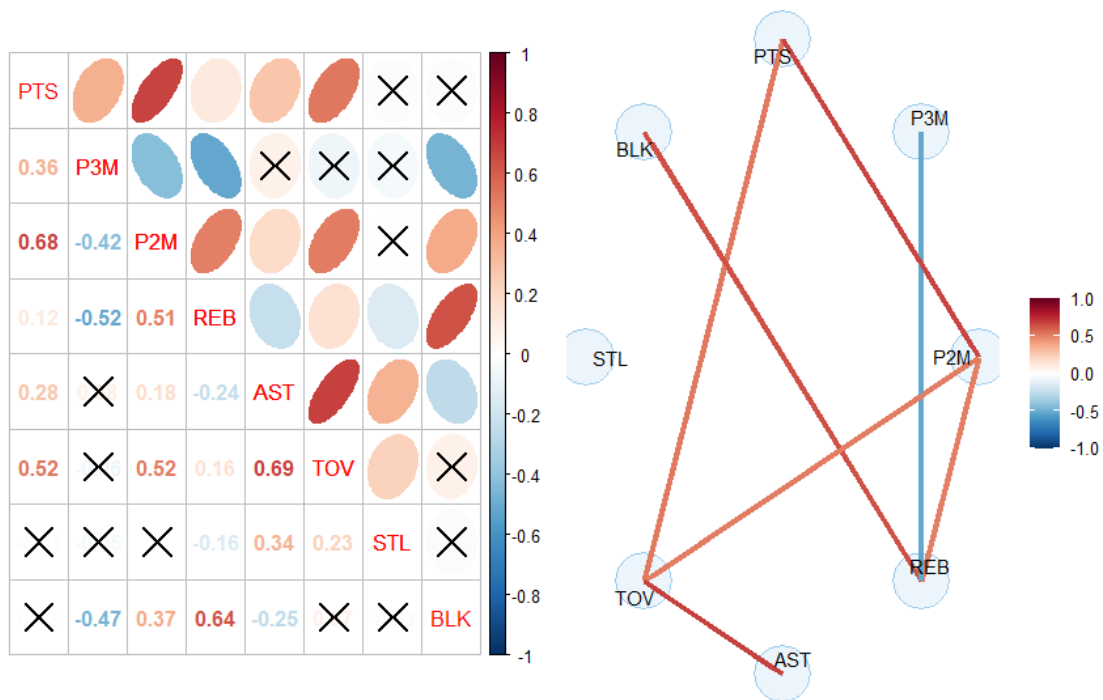
	rownm	No	Yes	eta2
1	DRTg	107.9	104.6	42.53
2	ORTg	104.0	108.1	40.25
3	STL	601.9	659.6	28.77
4	PTS	8576.0	8844.8	19.28
5	BLK	365.6	420.4	18.12
6	FTM	1328.0	1394.4	5.58
7	P2M	2353.7	2417.2	3.28
8	AST	1875.5	1931.6	3.17
9	P3M	846.9	871.9	1.07
10	REB	3558.1	3577.5	0.49

4.4. Συσχετίσεις

Η συσχέτιση αναφέρεται στη γραμμική σχέση μεταξύ δύο αριθμητικών μεταβλητών. Μια πολύ ενδιαφέρουσα ερμηνεία του συντελεστή συσχέτισης Pearson δίνεται όταν χρησιμοποιείται για τη μέτρηση της συσχέτισης μεταξύ της κατάταξης των παικτών. Για παράδειγμα, η εξέταση της κατάταξης των παικτών με βάση τις ασίστ (ανά λεπτό παιχνιδιού). Έτσι η σχέση μεταξύ του αριθμού των ασίστ και των λαθών για παίκτες που έχουν παίξει για περισσότερο από 500 λεπτά ισούται με 0,687 και δείχνει μιας μέτριας έντασης θετική σχέση που δείχνει ότι η πιθανότητα λάθους αυξάνεται με τον αριθμό των ασίστ.

Μια μελέτη της γραμμικής συσχέτισης μεταξύ όλων των ζευγών μεταβλητών όπου "ζεύγος" υποδηλώνει ότι, ακόμη και αν περισσότερες από δύο μεταβλητές αναλύονται από κοινού, ο συντελεστής γραμμικής συσχέτισης παραμένει ένα μέτρο διμεταβλητής συσχέτισης, καθώς αξιολογεί την κατεύθυνση και την ένταση της γραμμικής σχέσης μεταξύ όλων των πιθανών ζευγών μεταβλητών. Έτσι εξετάζοντας τις ακόλουθες μεταβλητές πόντοι, σουτ 3 και 2 πόντων, συνολικά ριμπάουντ (επιθετικά και αμυντικά), ασίστ, λάθη, κλεψίματα και μπλοκ (ανά λεπτό παιχνιδιού), αρχικά υπολογίζεται η γραμμική συσχέτιση μεταξύ των αριθμητικών μεταβλητών και στην συνέχεια παρέχεται η γραφική αναπαράσταση του πίνακα συσχέτισης

Στο γράφημα 12, δίνεται η γραφική περιγραφή του πίνακα συσχέτισης των εξεταζόμενων μεταβλητών. Το δεξί μέρος τους γραφήματος επιτρέπει στον αναγνώστη τον ευκολότερο εντοπισμό των πιο έντονων σχέσεων οι οποίες είναι μεταξύ του των 2ποντων και των λαθών, των δίποντων και των ριμπάουντ, των μπλοκ και των ριμπάουντ, των πόντων και των λαθών και τέλος, μεταξύ των πόντων και των δίποντων.



Γράφημα 12. Πίνακας συσχετίσεων μεταξύ των εξεταζόμενων μέτρων.

4.5. Ανάλυση δικτύου

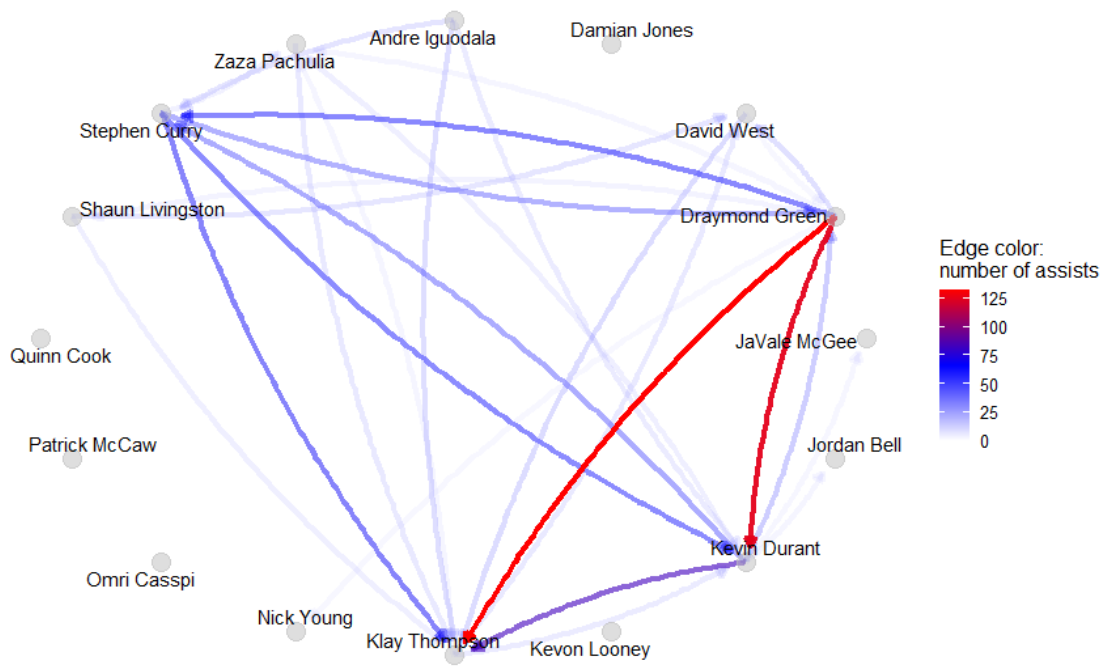
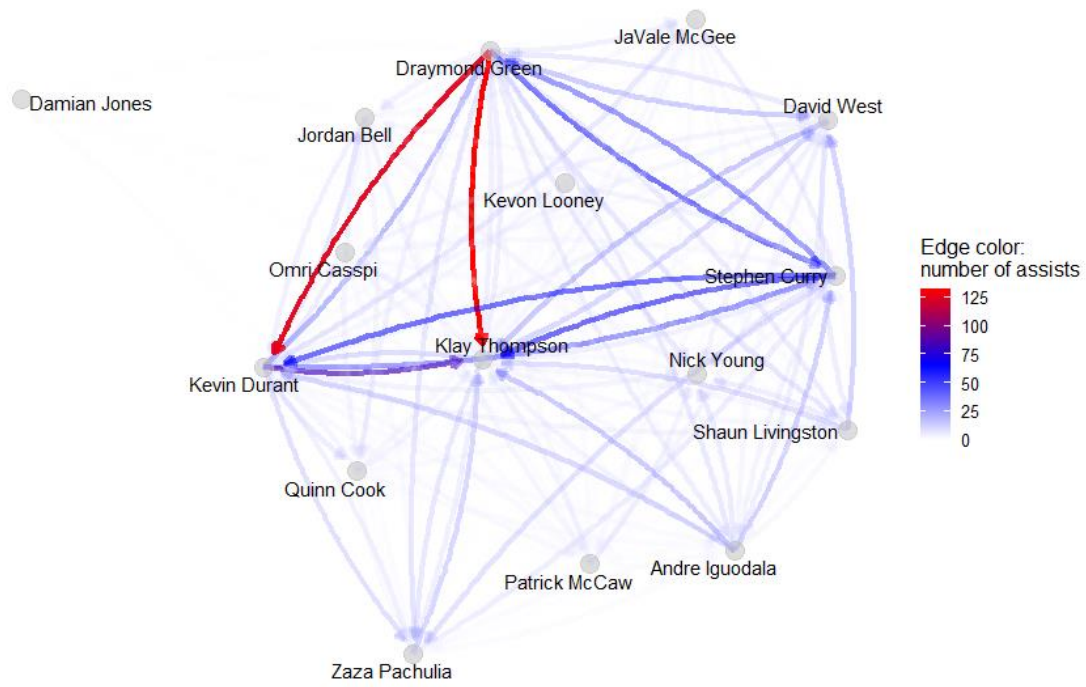
Δεδομένου ότι το μπάσκετ είναι ένα ομαδικό άθλημα, ένα σημαντικό ζήτημα είναι να αναλυθούν οι αλληλεπιδράσεις μεταξύ των παικτών κατά τη διάρκεια του παιχνιδιού. Για αυτόν τον σκοπό, πρέπει να γίνει χρήση δεδομένων της παρακολούθησης ενός παίκτη, που περιέχουν πληροφορίες σχετικά με συμβάντα που συνέβησαν κατά τη διάρκεια του παιχνιδιού. Οι γραφικές αναπαραστάσεις των αλληλεπιδράσεων μπορούν να ληφθούν μέσω της ανάλυσης δικτύου, η οποία επιτρέπει την κατασκευή και ανάλυση γραφημάτων που αποτελούνται από κόμβους και ακμές με δεδομένες ιδιότητες, ικανές να αναπαριστούν συμμετρικές ή ασύμμετρες σχέσεις μεταξύ διακριτών αντικειμένων (μεταβλητών του παιχνιδιού). Τα δεδομένα – ιδιότητες που αναλύθηκαν στο συγκεκριμένο παράδειγμα παρουσιάζονται στον πίνακα 9.

Πίνακας 9. Εξεταζόμενες μεταβλητές.

Μεταβλητή	Περιγραφή
FGM	Προσπάθειες
FGM_AST	Προσπάθειες από ασσίστ
FGM_ASTp	Ποσοστό FGM_AST/FGM
FGPTS	Επιτυχημένες προσπάθειες
FGPTS_AST	Πόντοι από ασσίστ
FGPTS_ASTp	FGPTS_AST/ FGPTS
ASTPTS	Πόντοι συμπαικτών

Τα δύο δίκτυα που εμφανίζονται στο γράφημα 13 δείχνουν ξεκάθαρα ότι, όσον αφορά την τελευταία πάσα και το σουτ, ο βασικός πυρήνας της ομάδας Golden State Warriors για την εξεταζόμενη σαιζόν αποτελείται από την τετράδα Stephen Curry, Klay Thompson, Kevin Durant και Draymond Green. Το γράφημα του δικτύου δείχνει ότι ο Draymond Green είναι ο καλύτερος βοηθός της ομάδας, με σαφή προτίμηση να βοηθάει τους Thompson και Durant. Σε μικρότερο βαθμό, ο Stephen Curry κάνει και δέχεται ασσίστ, κυρίως αλληλοεπιδρώντας με τους άλλους τρεις.

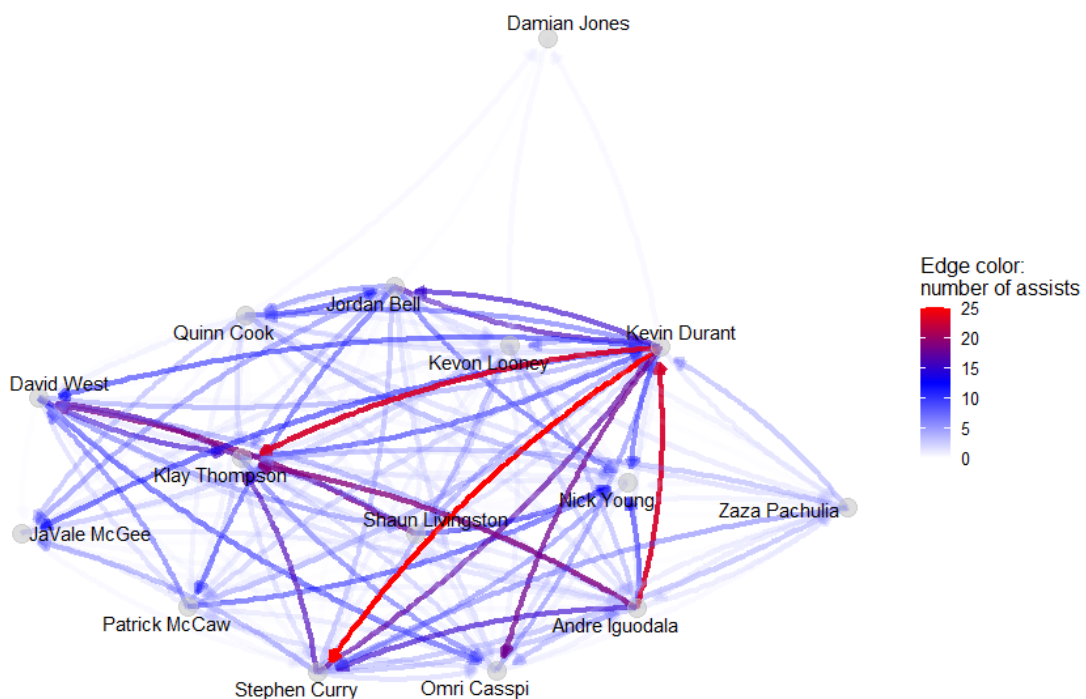
Ο Klay Thompson είναι μακράν ο παίκτης με τις περισσότερες ασσίστ, όχι μόνο από τους άλλους τρεις από την προαναφερθείσα τετράδα, αλλά και από αρκετούς άλλους συμπαίκτες του. Τέλος, ο Kevin Durant ασχολείται κυρίως με τους άλλους τρεις, αλλά παρουσιάζει έναν αξιόλογο αριθμό ασσίστ που έγιναν στους Zaza Pachulia, Jordan Bell και JaVale McGee.



Γράφημα 13. Διαγράμματα δικτύου της ομάδας Golden State Warriors (Επάνω: Βασικό διάγραμμα. Κάτω: Διάγραμμα με όριο αριθμού σχέσεων =20)

Στην συνέχεια, γίνεται εξέταση των προηγούμενων δικτύων χωρίς τον παίκτη Draymond Green που είναι ο παίκτης με τις περισσότερες ασίστ που κατέληξαν σε καλάθι. Αυτό επιτρέπει τη διερεύνηση του τρόπου με τον οποίο η ομάδα εφαρμόζει την επίθεση της στην περίπτωση απουσίας του βασικού οργανωτή της επίθεσης (Γράφημα 14).

Σύμφωνα με το γράφημα 14, το δίκτυο που προκύπτει, δείχνει ότι, απουσία του Draymond Green, οι άλλοι παίκτες τείνουν να μοιράζονται τη δουλειά βοηθώντας τους συμπαίκτες τους. Ο νέος κεντρικός παίκτης είναι ο Kevin Durant και, σε μικρότερο βαθμό, ο Andre Iguodala. Όπως αναφέρθηκε προηγουμένως, αν και είναι πόντ γκαρντ, ο Stephen Curry φαίνεται να έχει λιγότερο σημαντικό ρόλο από την άποψη των ασίστ.



Γράφημα 14. Διάγραμμα δικτύου της ομάδας Golden State Warriors χωρίς τον Draymond Green

4.6 Pressing

Ένα ευρύ σύνολο εξαιρετικά ισχυρών εργαλείων για την ανακάλυψη κρυφών δομών και προτύπων σε δεδομένα προέρχεται από τεχνικές μηχανικής εκμάθησης που καθιστούν διαθέσιμους πολλούς αλγορίθμους τόσο για εποπτευόμενη όσο και για μη εποπτευόμενη

μάθηση. Η φιλοσοφία πίσω από την αλγοριθμική μοντελοποίηση συζητείται από τον Breiman (2001b) και ένας αριθμός αλγορίθμων μηχανικής μάθησης παρουσιάζεται από τους Friedman et al. (2009) και Witten et al. (2016). Σε αυτή την ενότητα, παρουσιάζονται ορισμένα αποτελέσματα που λαμβάνονται μέσω του αλγόριθμου CART (Classification And Regression Trees, Breiman et al., 1984). Ο αλγόριθμος αυτός είναι ικανός να ανακαλύψει τον τρόπο με τον οποίο ένα σύνολο μεταβλητών επηρεάζει ένα δεδομένο αποτέλεσμα, λαμβάνοντας υπόψη πιθανώς μη γραμμικές ή περίπλοκες σχέσεις και αλληλεπιδράσεις μεταξύ των προγνωστικών, που αναφέρεται στην έρευνα των Zuccolotto et al., 2018 που στοχεύει στην περιγραφή της επιρροής ορισμένων καταστάσεων παιχνιδιού υψηλής πίεσης στην ευστοχία των παικτών.

Ο Ettore Messina αναφέρει ότι η απόφαση του προπονητή για το ποιον θα βάλει στο γήπεδο βασίζεται πάνω απ' όλα στα συναισθήματα για τον χαρακτήρα των μεμονωμένων παικτών και την προσωπική τους χημεία με τους συμπαίκτες, αλλά και για την ικανότητά τους να αγωνίζονται υπό πίεση. Ακολουθώντας την ανάγκη μέτρησης της ικανότητας των παικτών να ανταπεξέλθουν στην πίεση, αυτή η μελέτη επικεντρώνεται στην ευστοχία των παικτών και στηρίζεται στην ιδέα ότι οι βολές δεν είναι όλες ίδιες, καθώς μπορούν να συμβούν σε ήρεμες καταστάσεις ή καταστάσεις άγχους.

Ο ορισμός της κατάστασης του παιχνιδιού σε υψηλή πίεση και το ουσιαστικό νόημά του περιέχει και την ψυχολογική πίεση όσο και τον ανταγωνισμό. Η βασική υπόθεση σε αυτή την ανάλυση είναι ότι

H1: Η πρόκληση πίεσης στους παίκτες σχετίζεται με χαμηλότερες πιθανότητες σκοραρίσματος

Δείχνοντας κάτι τέτοιο, αυτά τα ευρήματα μπορούν να χρησιμοποιηθούν για την ανάπτυξη νέων μέτρων απόδοσης βελτιώνοντας τα ήδη χρησιμοποιούμενα στατιστικά αποδόσεων.

Η ανάλυση πραγματοποιήθηκε σε δεδομένα ανά παιχνίδι όλων των αγώνων που έγιναν κατά τη διάρκεια του ιταλικού πρωταθλήματος μπάσκετ δεύτερης εθνικής «Serie A2» 2015/2016. Τα δεδομένα περιέχουν σχεδόν 70.000 εγγραφές, επομένως το μέγεθος του δείγματος είναι αρκετά μεγάλο ώστε να εγγυάται αξιόπιστες εκτιμήσεις των πιθανοτήτων

βαθμολογίας, ακόμη και σε καταστάσεις που μπορεί να συμβούν μόνο περιστασιακά. Επειδή η αντίδραση στην πίεση μπορεί να είναι διαφορετική ανάλογα με το επαγγελματικό επίπεδο των παικτών (Madden et al., 1990, 1995), τα πιο σημαντικά αποτελέσματα έχουν στη συνέχεια ελεγχθεί σε ένα μικρότερο σύνολο δεδομένων από το Olympic Basketball Tournament Rio 2016 (control group). Ο Πίνακας 10 συνοψίζει τα κύρια χαρακτηριστικά των δύο συνόλων δεδομένων.

Πίνακας 10. Περιγραφή των δύο συνόλων δεδομένων (Πηγή: Zuccolotto et al. (2018))

	Serie A2	Rio 2016
Παιχνίδια	480	38
Ομάδες	32	12
Παίκτες	438	144
Δίποντα	33682 (50,9%)	3101 (52,2%)
Τρίποντα	21263 (34,1%)	1780 (33,8%)
Ελεύθερες βολές	14843 (73,5%)	1589 (74,8%)

Οι Zuccolotto et al. (2018) αναφέρουν ότι μια κατάσταση παιχνιδιού προσδιορίζεται ως «υψηλής πίεσης» εάν για κάποιο λόγο είναι πιο προβληματική και απαιτητική, χωρίς διάκριση αν η πίεση προέρχεται από παράγοντες που σχετίζονται με το παιχνίδι, ψυχολογικούς παράγοντες ή και τα δύο. Ακολουθώντας αυτόν τον γενικό ορισμό και λαμβάνοντας υπόψη τα διαθέσιμα δεδομένα, εντόπισαν, ακολουθώντας επίσης κάποιες υποδείξεις ειδικών του μπάσκετ, ορισμένους βασικούς τύπους καταστάσεων που μπορεί να ασκήσουν πίεση στον παίκτη όταν επιχειρείται ένα σουτ: όταν λήγει ο χρόνος των 24 δευτερολέπτων, όταν η διαφορά σκορ σε σχέση με τον αντίπαλο είναι μικρή, όταν η ομάδα ως σύνολο έχει κακή απόδοση κατά τη διάρκεια του αγώνα μέχρι τη συγκεκριμένη στιγμή του παιχνιδιού, όταν ο παίκτης που σουτάρει έχει χάσει το προηγούμενο σουτ του και όταν ο χρόνος που απομένει για την λήξη του παιχνιδιού τελειώνει.

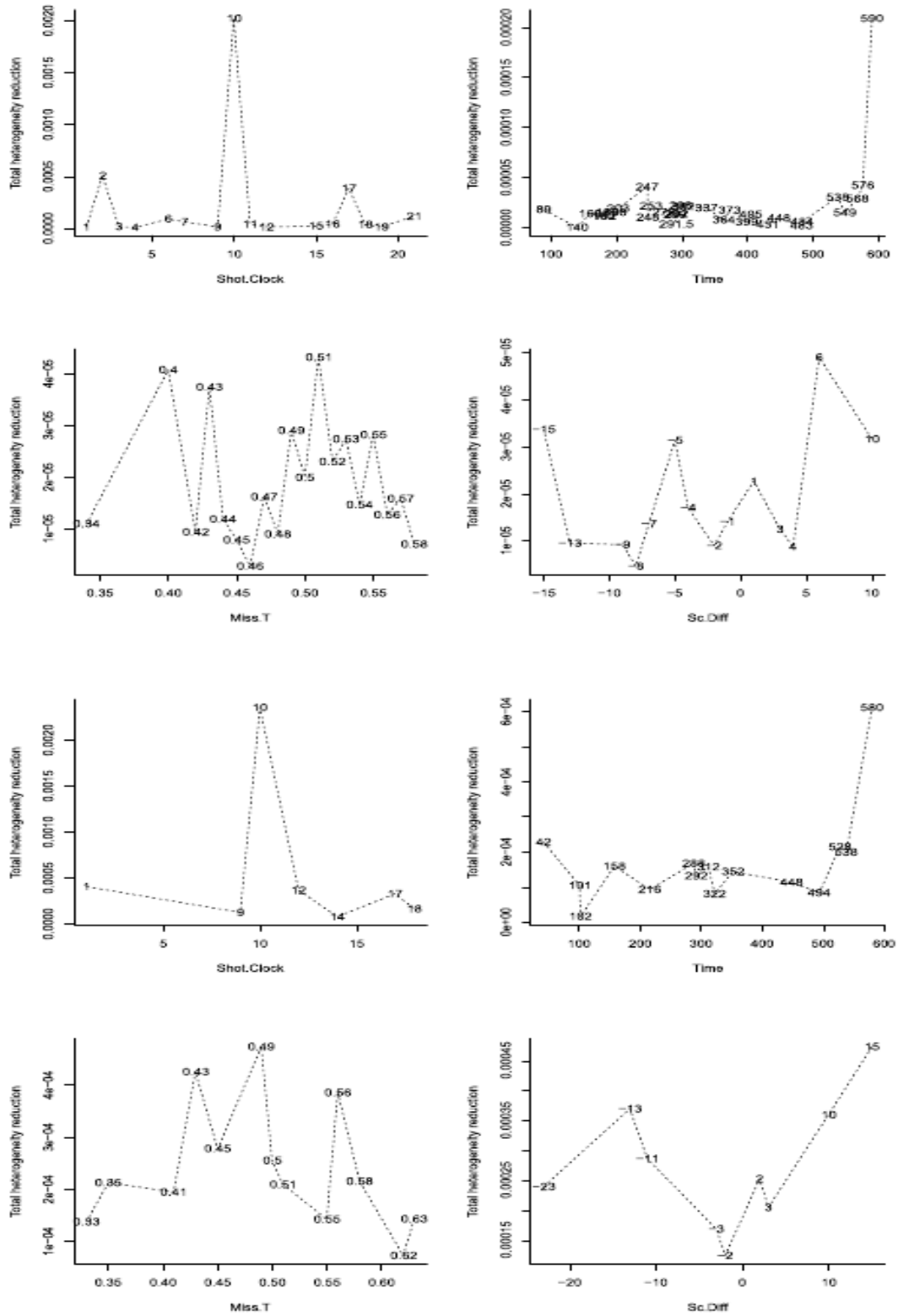
Οι Zuccolotto et al. (Zuccolotto et al., 2018) χρησιμοποίησαν το CART για να αξιολογήσουν τον αντίκτυπο αυτών των καταστάσεων υψηλής πίεσης στην πιθανότητα ευστοχίας λαμβάνοντας υπόψη όλες τις κοινές συσχετίσεις μεταξύ των μεταβλητών. Οι συγγραφείς προτίμησαν να χρησιμοποιήσουν κατηγορικές μεταβλητές αντί για

αριθμητικές, καθώς αυτό εμποδίζει τα δέντρα αποφάσεων (decision trees) να έχουν πολλούς κλάδους (διατηρώντας έτσι την ακρίβεια των αποτελεσμάτων) και επιτρέπει την εστίαση - λαμβάνοντας μόνο τις πιο ενδιαφέρουσες καταστάσεις. Για να μετατρέψουν τις αριθμητικές μεταβλητές σε κατηγορικές, οι συγγραφείς πρότειναν τον προσδιορισμό των τιμών αποκοπής συνδυάζοντας τα αποτελέσματα μιας διαδικασίας μηχανικής μάθησης και τις προτάσεις ορισμένων ειδικών του μπάσκετ. Όσον αφορά τη διαδικασία μηχανικής μάθησης, πρότειναν μια διαδικασία για να ληφθεί ένα όριο (threshold), που υπολογίζεται με την ανάπτυξη ενός προκαταρκτικού δέντρου χρησιμοποιώντας τις αριθμητικές μεταβλητές και στη συνέχεια αθροίζοντας όλες τις μειώσεις του δείκτη ετερογένειας.

Τα αποτελέσματα εμφανίζονται γράφημα 15 για τις μεταβλητές ως εξής: τα δευτερόλεπτα μέχρι τον ήχο του βομβητή 24 δευτερολέπτων (Shot.Clock), τον χρόνο μέχρι το τέλος της κάθε περιόδου (Time), το ποσοστό των χαμένων βολών για όλη την ομάδα έως τη στιγμή που επιχειρείται κάθε σουτ (Miss.T), η διαφορά σκορ σε σχέση με τον αντίπαλο όταν επιχειρείται κάθε σουτ (Sc.Diff).

Ορισμένες παρατηρήσεις μπορούν να διατυπωθούν με βάση το γράφημα 15 είναι

- Η μεταβλητή Shot.Clock παρουσιάζει ορισμένα σαφή όρια στα άκρα και στη μέση (κάτι που συνάδει με τις προτάσεις των ειδικών για εξαίρεση των ακραίων τιμών).
- Η μεταβλητή Time, έχει απότομη άνοδο τα τελευταία 100 δευτερόλεπτα (κάτι που είναι επίσης συνεπές με τις απόψεις των ειδικών σχετικά με τη συνάφεια των τελευταίων 1-2 λεπτών κάθε περιόδου).
- τα δύο σύνολα δεδομένων, αν και πολύ διαφορετικά από την άποψη του επαγγελματικού επιπέδου των παικτών, έχουν δώσει πολύ παρόμοια αποτελέσματα.
- Η μεταβλητή Miss.T παρουσιάζει διακυμάνσεις μεταξύ 0,35 και 0,60, αλλά δεν εμφανίζονται κορυφές που μπορεί να προτείνουν μια πιθανή καλύτερη επιλογή εντός αυτού του διαστήματος, επομένως η διαίρεση μπορεί να γίνει σύμφωνα με εκατοστημόρια.



Γράφημα 15. Μέτρα ορίων των εξεταζόμενων μεταβλητών (πηγή: (Zuccolotto et al., 2018).

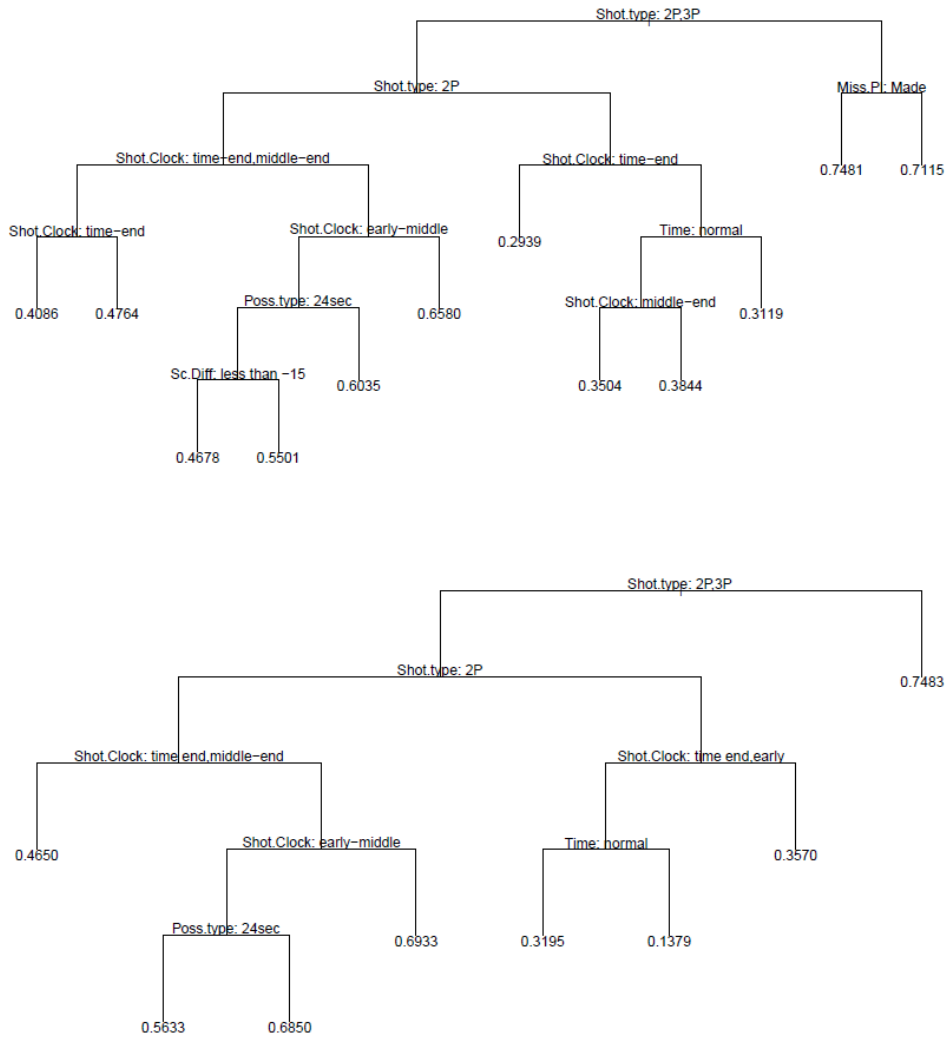
Στην συνέχεια, οι συγγραφείς αποφάσισαν να μετατρέψουν αριθμητικές μεταβλητές σε κατηγορικές μεταβλητές σύμφωνα με τα κριτήρια που συνοψίζονται στον πίνακα 11.

Πίνακας 11. Μετατροπή σε κατηγορικές μεταβλητές (Πηγή: (Zuccolotto et al., 2018))

Shot.Clock	early:Shot.Clock>17
	early-middle:10<Shot.Clock≤17
	middle-end:2<Shot.Clock≤10
	time-end:Shot.Clock≤2
Time	normal:Time≤500
	quarter-end:Time>500
Miss.T	Bad:Miss.T≤25thpercentile
	Medium:25thpercentile<Miss.T≤75thpercentile
	Good:Miss.T>75thpercentile
Sc.Diff	Less than-15:Sc.Diff≤-15
	between-15and-5:-15<Sc.Diff≤-5
	between-5and 1:-5<Sc.Diff≤1
	between1and0.251<Sc.Diff≤6

Αξίζει να σημειωθεί ότι η μεταβλητή Shot.Clock περιγράφει τα δευτερόλεπτα έως ότου ηχήσει ο βομβητής (δηλαδή, η τιμή που εμφανίζεται στο ρολόι λήψης, ανεξάρτητα από το εάν το ρολόι λήψης είχε προηγουμένως επαναφερθεί στα 14 δευτερόλεπτα). Στην πραγματικότητα, η παράταση σε σχέση με τα 24 δευτερόλεπτα μπορεί να τροποποιήσει προσωρινά την πίεση του παιχνιδιού και να επηρεάσει την πιθανότητα σκοραρίσματος. Για το λόγο αυτό, οι συγγραφείς εισήγαγαν στον CART έναν πρόσθετο κατηγορηματικό προγνωστικό δείκτη (Poss.type) που δηλώνει εάν η λήψη έγινε κατά τη διάρκεια των αρχικών 24 δευτερολέπτων στο ρολόι λήψης ή μετά την επαναφορά του ρολογιού λήψης στα 14 δευτερόλεπτα. Το μοντέλο CART με όλες τις κατηγορικές συμμεταβλητές αναπτύχθηκε χρησιμοποιώντας τον δείκτη Gini ως κριτήριο επιλογής διαχωρισμού.

Τα αποτελέσματα εφαρμογής του αλγορίθμου παρουσιάζονται στα δενδρογράμματα του γραφήματος 16.



Γράφημα 16. Δενδρογράμματα για Serie A (επάνω) και RIO2016 (κάτω) (Πηγή:(Zuccolotto et al., 2018)

Το γράφημα 16 παρέχει πολύ ενδιαφέρουσες ερμηνείες σχετικά με τον αντίκτυπο των καταστάσεων παιχνιδιών υψηλής πίεσης στην πιθανότητα να σκοράρουν οι πάικτες. Το πρώτο επίπεδο διαχωρισμού γίνεται σύμφωνα με τη μεταβλητή Shot.Type (βολές 2 πόντων και 3 πόντων στον αριστερό κλάδο και ελεύθερες βολές προς τα δεξιά). Μετά από αυτό, το Shot.Clock έχει τον πιο εξέχοντα ρόλο για τις βολές πεδίου, ακολουθούμενο

από τα Sc.Diff, Time και Poss.type, τα οποία φαίνεται να παίζουν ρόλο στις αλληλεπιδράσεις με το Shot.Clock και μεταξύ αυτών. Για τις ελεύθερες βολές, η σχετική μεταβλητή είναι Miss.PI (που δηλώνει αν το προηγούμενο σουτ του παίκτη που σουτάρει σημείωσε καλάθι ή όχι).

Το δέντρο που βασίζεται στο σύνολο δεδομένων "Serie A2" είναι το πιο πολύ αναπτυγμένο, χάρη στο υψηλότερο μέγεθος δείγματος. Αναλυτικά, αυτό το δέντρο αποκαλύπτει τις ακόλουθες σχέσεις:

- για τις ελεύθερες βολές, η μόνη σχετική μεταβλητή είναι Miss.PI: η εκτιμώμενη πιθανότητα για να σκοράρει ένας παίκτης όταν το προηγούμενο σουτ από αυτόν τον παίκτη σημείωσε καλάθι είναι 0,7481, έναντι 0,7115 όταν η προηγούμενη βολή ήταν χαμένη.
- για βολές 3 πόντων, η πρώτη σχετική μεταβλητή είναι Shot.Clock: η εκτιμώμενη πιθανότητα βαθμολογίας στο τέλος του χρόνου είναι 0,2939. Διαφορετικά, η πιθανότητα να σκοράρει είναι 0,3504 και 0,3844 για μεσαίες και προηγούμενες βολές, αντίστοιχα, με την προϋπόθεση ότι το παιχνίδι δεν είναι στο τέλος της περιόδου, όταν η πιθανότητα σκοραρίσματος μειώνεται στο 0,3119.
- για βολές 2 πόντων, η πιο σχετική μεταβλητή είναι και πάλι Shot.Clock: η εκτιμώμενη πιθανότητα βαθμολογίας είναι 0,4086, 0,4764 και 0,6580 για βολές στο τέλος του χρόνου, στη μέση και στην αρχή, αντίστοιχα.

Πολύ παρόμοιες παρατηρήσεις, αν και λιγότερο λεπτομερείς, καθώς η ανάπτυξη του δέντρου είναι μικρότερη, μπορούν να αντληθούν από το δεύτερο CART, που ελήφθη με το σύνολο δεδομένων "Rio 2016". Σύμφωνα με τα στοιχεία που προσδιορίζονται από τα μοντέλα CART, οι λήψεις δεν είναι όλες ίδιες. Για το λόγο αυτό, οι Zuccolotto et al., (2018) χρησιμοποίησαν τα δέντρα για να αναπτύξουν νέα μέτρα απόδοσης βολών, λαμβάνοντας υπόψη τις συνθήκες κάτω από τις οποίες επιχειρήθηκε κάθε βολή.

Για παράδειγμα, μια βολή 2 πόντων που επιχειρείται στα 2 τελευταία δευτερόλεπτα του ρολογιού βολής έχει πιθανότητα βαθμολόγησης περίπου 40%, σε αντίθεση με μια βολή που επιχειρείται στα πρώτα 7 δευτερόλεπτα, η οποία έχει πιθανότητα σκοραρίσματος

μεγαλύτερη από 65%. Το μέτρο απόδοσης βολής θα πρέπει να λαμβάνει υπόψη αυτά τα στοιχεία και να δίνει μεγαλύτερη αξία σε ένα καλάθι που φτιάχνεται όταν η πιθανότητα βαθμολόγησης είναι μικρότερη. Για κάθε βολή τύπου T (2P: 2-πόντοι, 3P: 3-πόντοι, FT: ελεύθερες βολές), έστω JT το σύνολο των επιχειρούμενων βολών τύπου T και x_{ij} ο δείκτης που παίρνει την τιμή 1 εάν η jη βολή του i του παίκτη ήταν επιτυχής και 0 εάν δεν ήταν. Το νέο μέτρο απόδοσης βολής του παίκτη i για βολή τύπου T δίνεται από τον τύπο

$$P_i(T) = \text{av}_{j \in JT} (x_{ij} - \pi_{ij}) \quad (4.1)$$

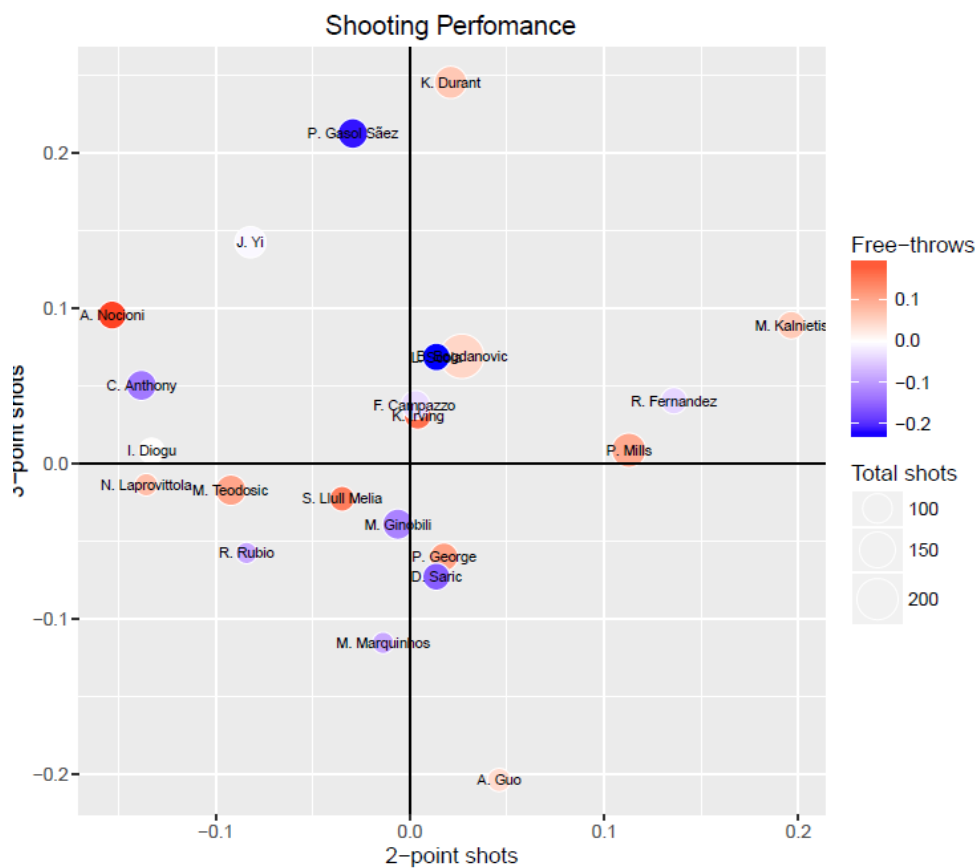
όπου το $\text{av}_{j \in J} ()$ δηλώνει τον μέσο όρο σε όλες τις βολές του τύπου T που επιχειρείται από τον παίκτη i και το π_{ij} είναι η πιθανότητα βαθμολόγησης που εκχωρείται από το μοντέλο CART στο j σουτ του i παίκτη, δηλαδή σε ένα σουτ του ίδιου τύπου και επιχειρήθηκε στην ίδια κατάσταση παιχνιδιού με το j σουτ του i παίκτη.

Για κάθε βολή, η διαφορά $x_{ij} - \pi_{ij}$ μπορεί να χρησιμοποιηθεί ως μέτρο απόδοσης της βολής. Πράγματι, η διαφορά είναι θετική αν το σουτ σημείωσε καλάθι (και όσο μικρότερη είναι η πιθανότητα σκοραρίσματος, τόσο μεγαλύτερη είναι η αξία του) και αρνητική αν ήταν άστοχη (και όσο μεγαλύτερη είναι η πιθανότητα να σκοράρει, τόσο μεγαλύτερη είναι η απόλυτη τιμή του).

Για παράδειγμα, ένα καλάθι αξίζει περισσότερο όταν η πιθανότητα να σκοράρει το αντίστοιχο σουτ είναι χαμηλή. Ενώ όταν σημειωθεί αστοχία, θεωρείται πιο επιζήμιο όταν η πιθανότητα να σκοράρει το αντίστοιχο σουτ είναι υψηλή. Η υψηλότερη ερμηνευτικότητα επιτυγχάνεται όταν οι παίκτες συγκρίνονται μέσω ενός διγραμμάτος φυσαλίδας (bubble plot) με $P_i(2P)$ και $P_i(3P)$ στον άξονα x και στον άξονα y, αντίστοιχα, και το χρώμα που αντιπροσωπεύει το $P_i(FT)$.

Το γράφημα 17 αναφέρεται στο σύνολο δεδομένων "Rio 2016" και εμφανίζει παίκτες που επιχειρήσαν τουλάχιστον 15 βολές για κάθε τύπο βολής. Σε αυτό το διάγραμμα φυσαλίδας, εμφανίζονται εύκολα ποιοι παίκτες είναι καλύτεροι (ή χειρότεροι) από τον μέσο όρο για κάθε τύπο βολής και λαμβάνοντας υπόψη τη συγκεκριμένη κατάσταση παιχνιδιού. Για παράδειγμα, ο Rodolfo Fernandez αποδίδει καλύτερα από τον μέσο όρο τόσο σε σουτ 2 πόντων όσο και σε σουτ 3 πόντων, αλλά χειρότερα στις ελεύθερες βολές.

Το πιο σημαντικό είναι ότι το γράφημα επιτρέπει την εξέταση διαφορών μεταξύ των παικτών που έχουν την ίδια απόδοση, με βάση τα συνηθισμένα ποσοστά επιτυχιών, όπου κάθε σουτ θεωρείται ότι είναι ίδιο με οποιοδήποτε άλλο. Για παράδειγμα, οι Bojan Bogdanovic και Andrés Nocioni έχουν περίπου τα ίδια ποσοστά επιτυχίας σε σουτ 3 πόντων (45% και 45,2%). Ωστόσο, αν αναλογιστούμε τη δυσκολία των αγωνιστικών καταστάσεων στις οποίες επιχειρήσαν τα σουτ τους, ο Nocioni εκτελεί σουτ 3 πόντων ελαφρώς καλύτερα από τον Bogdanovic. Επιπλέον, μπορούμε να βρούμε μερικούς παίκτες με διαφορετικά ποσοστά επιτυχίας στο γήπεδο που λαμβάνουν ίσες αξιολογήσεις όταν η δυσκολία της κατάστασης του παιχνιδιού μπαίνει στο μέτρο απόδοσης. Για παράδειγμα, οι Nicolás Laprovittola, Carmelo Anthony και Ikechukwu Diogu δέχονται παρόμοια μέτρηση απόδοσης σε σουτ 2 πόντων, αν και έχουν διαφορετικά ποσοστά επιτυχίας (36,8%, 38,5% και 39,1%, αντίστοιχα).



Γράφημα 17. Επίδοση υπό πίεση (Dataset :Rio 2106, Πηγή: (Zuccolotto et al., 2018)

ΚΕΦΑΛΑΙΟ 5: Επίλογος – Συμπεράσματα

Με σκοπό την διερεύνηση μεθόδων ανάλυσης και παρουσίασης των αποτελεσμάτων αθλητικών δεδομένων έγινε διερεύνηση απλών και πιο σύνθετων μεθόδων που εφαρμόζονται στο αγώνισμα της καλαθοσφαίρισης. Αν και το πρωτάθλημα καλαθοσφαίρισης των ΗΠΑ ή NBA έχει πρωταγωνιστικό ρόλο στην εξέταση αυτών των μεθόδων, στην εφαρμογή αλλά και στην εκμετάλλευσή τους, η εργασία δεν περιορίστηκε μόνο σε αυτά τα δεδομένα αλλά εξέτασε και άλλες πηγές δεδομένων.

Μέσα από την εφαρμογή αυτών των μεθόδων διαπιστώθηκε ότι ανεξάρτητα της πηγής των δεδομένων, η χρησιμότητα των συμπερασμάτων που προκύπτουν από μια τέτοιου είδους ανάλυση είναι καθολική και πολυδιάστατη. Με άλλα λόγια, μπορούν να χρησιμοποιηθούν για κάθε τύπου δεδομένων αλλά και να εκμεταλλευτούν από κάθε ενδιαφερόμενο. Φυσικά, ο τρόπος παρουσίασης των αποτελεσμάτων θα πρέπει να προσαρμόζεται ανάλογα με τον αποδέκτη. Έτσι, ανάλογα με τον αποδέκτη θα πρέπει να παρουσιαστεί ένα κείμενο που να περιέχει γραφικά που να είναι κατανοητά για τον υπεύθυνο μάρκετινγκ της ομάδας, για τον προπονητή, για τον πρόεδρο και το διοικητικό συμβούλιο της ομάδας και φυσικά για το κοινό.

Μέσα από τα πραγματικά σενάρια που παρουσιάστηκαν μέσα στην εργασία, έγινε φανερή δυσκολία του έργου ενός υπεύθυνου ανάλυσης αθλητικών δεδομένων. Έτσι, πέρα από την φυσική δυσκολία του έργου του, δηλαδή τη συλλογή, επεξεργασία, ανάλυση και παρουσίαση δεδομένων θα πρέπει να αντιμετωπίσει και το να είναι ο ενδιάμεσος κρίκος ή σύνδεσμος πολλών και διαφορετικών αρμοδιοτήτων μέσα στην ομάδα και κυρίως μεταξύ του προπονητή, του διοικητικού συμβουλίου και του προέδρου. Θε πρέπει να μπορεί να παράγει αποτελέσματα που να είναι κατανοητά από άτομα διαφορετικών ειδικοτήτων και να μπορεί να δώσει την ερμηνεία σε κάθε περίπτωση με βάση το αντικείμενο εργασίας τους.

Αυτό όμως το δύσκολο έργο έχει πραγματικά αποτελέσματα καθώς, όπως παρουσιάστηκε στην εργασία, μπορεί να δώσει λύσεις σε τραυματισμούς παικτών, στην απουσία βασικού παίκτη, να προσδιορίσει την βέλτιστη θέση των παικτών, να εξετάσει

παράγοντες απόδοσης της ομάδας και των παικτών, να εξετάσει τις δυνατότητες του ανταγωνισμού πιθανές αδυναμίες της ομάδας κα.

Μέσα λοιπόν από αυτή την παρουσίαση μερικών μόνο μεθόδων ανάλυσης αθλητικών δεδομένων, έγινε φανερό το πόσο σημαντικό είναι το έργο του αναλυτή δεδομένων και η δύναμη της ανάλυσης αθλητικών δεδομένων. Στο σύγχρονο και ανταγωνιστικό περιβάλλον του αθλητισμού όπου πχ ένας προπονητής καλείται να αποκτήσει τον καλύτερο δυνατό παίκτη μέσα από ένα συγκεκριμένο προϋπολογισμό της ομάδας μέσα από διάφορες επιλογές η ανάλυση αθλητικών δεδομένων βοηθά στην ελαχιστοποίηση των λανθασμένων επιλογών.

Βιβλιογραφία

- Abhishek, D. (2021). *Basketball Positions and their Purpose | Chase Your Sport - Sports Social Blog*. <https://www.chaseyoursport.com/Basketball/Basketball-Positions-and-their-Purpose/2579>
- Adams, L. (2022). *NBA Minimum Salaries For 2022/23 | Hoops Rumors*. <https://www.hoopsrumors.com/2022/07/nba-minimum-salaries-for-2022-23.html>
- Alamar, B. (2013a). *Sports Analytics. A guide for coaches, managers and other decision makers*. Columbia University Press.
- Alamar, B. (2013b). *Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers*. . Columbia University Press.
- Alamar, B., & Mehrotra, V. (2011). The rapidly evolving world of sports analytics. *Analytics Magazine*, 33–37.
- Allison, P. D. (2001). *Missing data* (Vol. 136). Sage publications.
- Anguera, M. T., & Hernández Mendo, A. (2013). La metodología observacional en el ámbito del deporte. *E-Balonmano.Com: Revista de Ciencias Del Deporte*, 9, 135–160.
- Briskorn. (2008). *Sports Leagues Scheduling: Models, Combinatorial Properties, and Optimization Algorithms*. . Springer.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Chambers, J. (2008). *Software for data analysis: programming with R*. . Springer Science & Business Media.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, Statistics, and Teaching. *The American Mathematical Monthly*, 104(9), 801. <https://doi.org/10.2307/2975286>
- Curley, J. (2016). *Curley Lab | 2017*. <http://curleylab.psych.columbia.edu/nba.html>

- Data Definition & Meaning - Merriam-Webster.* (2022). <https://www.merriam-webster.com/dictionary/data>
- Davenport, T. H. (2014). What Businesses Can Learn From Sports Analytics. *MIT Sloan Management Review*. <https://sloanreview.mit.edu/article/what-businesses-can-learn-from-sports-analytics/>
- Dmonte, R., & Dmello, A. (2017). Big Data in Sports Leverage Big Data in Sports: An Insight using SAP HANA. *International Journal of Engineering Research & Technology (IJERT)*, 6(1), 380–383.
- Early, G. L. (2011). *A Level Playing Field: African American Athletes and the Republic of Sports*. . Harvard University Press.
- Fort, R. D. (2011). *Sports Economics (third ed.)*. . Prentice Hall.
- Garganta, J. (2008, September). *MODELAÇÃO TÁTICA EM JOGOS DESPORTIVOS: A DESEJÁVEL CUMPLICIDADE ENTRE PESQUISA, TREINO E COMPETIÇÃO*.
- Garganta, J. (2009). Trends of tactical performance analysis in team sports: bridging the gap between research, training and competition. *Revista Portuguesa de Ciências Do Desporto*, 9(1), 81–89. <https://doi.org/10.5628/rpcd.09.01.81>
- Golfarelli, M., & Rizzi, S. (2009). *Data warehouse design: Modern principles and methodologies*. (Vol. 5). McGraw-Hill .
- Gréhaigne, J.-F. (1992). *L'organisation du jeu en football*. ACTIO.
- Gréhaigne, J.-F., Bouthier, D., & David, B. (1997). Dynamic-system analysis of opponent relationships in collective actions in soccer. *Journal of Sports Sciences*, 15(2), 137–149. <https://doi.org/10.1080/026404197367416>
- Gréhaigne, J.-F., & Godbout, P. (1995). Tactical Knowledge in Team Sports From a Constructivist and Cognitivist Perspective. *Quest*, 47(4), 490–505. <https://doi.org/10.1080/00336297.1995.10484171>
- Hardle, W. K., & Simar, L. (2015). *Applied multivariate statistical analysis*. . Springer-Verlag.

- Hernández, M. A., & Stolfo, S. J. (1998). Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. *Data Mining and Knowledge Discovery*. *Data Mining and Knowledge Discovery*, 2(1), 9–37. <https://doi.org/10.1023/A:1009761603038>
- Hughes, M., Dawkins, N., David, R., & Mills, J. (1998). The Perturbation Effect and Goal Opportunities in Soccer. *Journal of Sports Sciences*, 16, 20–21.
- Inequality - Income inequality - OECD Data*. (n.d.). Retrieved October 1, 2022, from <https://data.oecd.org/inequality/income-inequality.htm>
- Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *Journal of Economic Perspectives*, 14(3), 75–94. <https://doi.org/10.1257/jep.14.3.75>
- Kelso, S. (1995). *Dynamic Patterns: The self-organization of brain and behavior*. Human Kinetics.
- Kenett, R. S., & Shmueli, G. (2016). *Information quality (InfoQ): The Potential of Data and Analytics to Generate Knowledge*. . John Wiley & Sons.
- Kim, W., Choi, B.-J., Hong, E.-K., Kim, S.-K., & Lee, D. (2003). A Taxonomy of Dirty Data. *Data Mining and Knowledge Discovery*, 7(1), 81–99. <https://doi.org/10.1023/A:1021564703268>
- Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A Starting Point for Analyzing Basketball Statistics. *Journal of Quantitative Analysis in Sports*, 3(3). <https://doi.org/10.2202/1559-0410.1070>
- Lames, M., & Hansen, G. (2001a). Designing observational systems to support top-level teams in game sports. *International Journal of Performance Analysis in Sport*, 1(1), 83–90. <https://doi.org/10.1080/24748668.2001.11868251>
- Lames, M., & Hansen, G. (2001b). Designing observational systems to support top-level teams in game sports. *International Journal of Performance Analysis in Sport*, 1(1), 83–90. <https://doi.org/10.1080/24748668.2001.11868251>

- Lames, M., & McGarry, T. (2007). On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, 7(1), 62–79. <https://doi.org/10.1080/24748668.2007.11868388>
- Leeds, M. A., & von Allmen, P. (2014). *The Economics of Sports (fifth ed.)*. Pearson.
- Lewis, M. (2013). *Moneyball: The art of winning an unfair game 2nd ed.* Norton.
- Liebetrau, A. M. (1983). *Measures of association*. . Sage .
- Liebowitz, J. (2014). *Business Analytics: An Introduction*. . Taylor & Francis Group, LLC/CRC Press.
- Lindsey, G. R. (1959). Tatistical Data Useful for the Operation of a Baseball Team. *Operations Research*, 7, 197–207.
- Lindsey, G. R. (1963). An Investigation of Strategies in Baseball. *Operations Research*, 11(4), 477–501. <https://doi.org/10.1287/opre.11.4.477>
- Lipscomb, A. (2014). *New York Mets sign SAS® to lead off their analytics lineup - SAS Institute*. <https://news.cision.com/sas-institute/r/new-york-mets-sign-sas--to-lead-off-their-analytics-lineup,c9668526>
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data* (Vol. 333). John Wiley & Sons.
- Lorenz, M. O. (1905). Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association*, 9(70), 209. <https://doi.org/10.2307/2276207>
- Maciejewski, M. (2017). To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences*, 83(1_suppl), 120–135. <https://doi.org/10.1177/0020852316640058>
- Marr, B. (2015). *Big Data: Using SMART Big Data Analytics and Metrics to Make Better Decisions and Improve Performance*. John Wiley & Sons.
- Matloff, N. (2011). *The art of R programming: A tour of statistical software design*. . No Starch Press.

- McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., & Franks, I. M. (2002). Sport competition as a dynamical self-organizing system. *Journal of Sports Sciences*, 20(10), 771–781. <https://doi.org/10.1080/026404102320675620>
- Merlo, J., & Lynch, K. (2010). *Association, measures of*. In *Encyclopedia of Research Design*. SAGE Publications, Inc.
- Miller, W. T. (2015). *Sports Analytics and Data Science: Winning the Game with Methods and Models*. Published by Pearson Education, Inc.
- Oliver, D. (2022). *Four Factors* | *Basketball-Reference.com*. <https://www.basketball-reference.com/about/factors.html>
- Oliver, D. (2044). *Basketball on paper: rules and tools for performance analysis*. . Potomac Books, Inc.
- Palut, Y., & Zanone, P.-G. (2005). A dynamical analysis of tennis: Concepts and data. *Journal of Sports Sciences*, 23(10), 1021–1032. <https://doi.org/10.1080/02640410400021682>
- Panko, R. R. (1998). What We Know About Spreadsheet Errors. *The Journal of End User Computing's Special Issue on Scaling Up End User Development*, 10(2), 15–21.
- Pfeiffer, M., & Perl, J. (2006). Analysis of Tactical Structures in Team Handball by Means of Artificial Neural Networks. *International Journal of Computer Science in Sport*, 5, 4–14.
- Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). Minding the analytics gap. *MIT Sloan Management Review*, 56(3), 63–68.
- Rein, I., Shields, B., & Grossman, A. (2015). *e-Sports Strategist: Developing Leaders for a High-Performance Industry*. Oxford University Press.
- Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature*, 36, 2021–2064.
- Schumaker, R. P., Solieman, O. K., & Chen, H. (2010). *Sports Data Mining*. . Springer.

- Sitthiyot, T., & Holasut, K. (2021). A simple method for estimating the Lorenz curve. *Humanities and Social Sciences Communications*, 8(1), 268. <https://doi.org/10.1057/s41599-021-00948-x>
- Smith, D. (2016). *Analyzing NBA basketball data with R (Revolutions)*. <https://blog.revolutionanalytics.com/2016/09/analyzing-nba-basketball-data-with-r.html>
- Teodorescu, L. (1985). *Contribution to the concept of team sport*.
- Troilo, M., Bouchet, A., Urban, T. L., & Sutton, W. A. (2016a). Perception, reality, and the adoption of business analytics: Evidence from North American professional sport organizations. *Omega*, 59, 72–83. <https://doi.org/10.1016/j.omega.2015.05.011>
- Troilo, M., Bouchet, A., Urban, T. L., & Sutton, W. A. (2016b). Perception, reality, and the adoption of business analytics: Evidence from North American professional sport organizations. *Omega*, 59, 72–83. <https://doi.org/10.1016/j.omega.2015.05.011>
- Wickham, H. (2014). *Advanced R*. CRC Press.
- Wright, M. B. (2009). 50 years of OR in sport. *Journal of the Operational Research Society*, 60(sup1), S161–S168. <https://doi.org/10.1057/jors.2008.170>
- Zuccolotto, P., Manisera, M., & Sandri, M. (2018). Big data analytics for modeling scoring probability in basketball: The effect of shooting under high-pressure conditions. *International Journal of Sports Science & Coaching*, 13(4), 569–589. <https://doi.org/10.1177/1747954117737492>

Παράρτημα

Κώδικας R

```
#####
```

Γράφημα 1

```
#####
```

```
library(BasketballAnalyzeR)
```

```
PbP <- PbPmanipulation(PbP.BDB)
```

```
tm <- c("BOS", "CLE", "GSW", "HOU")
```

```
selTeams <- which(Tadd$team %in% tm)
```

```
FF.sel <- fourfactors(Tbox[selTeams,], Obox[selTeams,])
```

```
FF <- fourfactors(Tbox,Obox)
```

```
listPlots <- plot(FF)
```

```
library(gridExtra)
```

```
grid.arrange(grobs=listPlots[1:2], ncol=1)
```

```
plot(FF.sel)
```

```
#####
```

Γράφημα 2

```
#####
```

```
X <- data.frame(Tbox, PTS.O=Obox$PTS, TOV.O=Obox$TOV,
```

```
          CONF=Tadd$Conference)
```

```
XW <- subset(X, CONF=="W")
```

```
labs <- c("Steals", "Blocks", "Defensive Rebounds")
```

```
barline(data=XW, id="Team", bars=c("STL", "BLK", "DREB"),
```

```
        line="TOV.O", order.by="PTS.O", labels.bars=labs)
```

```
Pbox.HR <- subset(Pbox, Team=="Houston Rockets" &
                MIN>=500)
barline(data=Pbox.HR, id="Player",
        bars=c("P2p", "P3p", "FTp"), line="MIN",
        order.by="PM", labels.bars=c("2P%", "3P%", "FT%"),
        title="Houston Rockets")
```

#####

Γράφημα 3 και 4

#####

```
Pbox.PG <- subset(Pbox, Player=="Russell Westbrook" |
                Player=="Stephen Curry" |
                Player=="Chris Paul" |
                Player=="Kyrie Irving" |
                Player=="Damian Lillard" |
                Player=="Kyle Lowry" |
                Player=="John Wall" |
                Player=="Rajon Rondo" |
                Player=="Kemba Walker")
attach(Pbox.PG)
X <- data.frame(P2M, P3M, FTM, REB=OREB+DREB, AST,
                STL, BLK)/MIN
detach(Pbox.PG)
radialprofile(data=X, title=Pbox.PG$Player, std=FALSE)
```

```
# Για το γράφημα 4
radialprofile(data=X, title=Pbox.PG$Player, std=TRUE)
```

#####

Γράφημα 5

#####


```

Pbox.sel <- subset(Pbox, MIN>= 500)
attach(Pbox.sel)
X <- data.frame(AST, TOV, PTS)/MIN
detach(Pbox.sel)
mypal <- colorRampPalette(c("blue","yellow","red"))
scatterplot(X, data.var=c("AST","TOV"), z.var="PTS",
            labels=1:nrow(X), palette=mypal)
SAS <- which(Pbox.sel$Team=="San Antonio Spurs")

```

#####

Γράφημα 6

#####

```

attach(Tbox)
X <- data.frame(T=Team, P2p, P3p, FTp, AS=P2A+P3A+FTA)
detach(Tbox)
labs <- c("2-point shots (% made)",
          "3-point shots (% made)",
          "free throws (% made)",
          "Total shots attempted")
bubbleplot(X, id="T", x="P2p", y="P3p", col="FTp",
           size="AS", labels=labs)

```

#####

Γράφημα 7

#####

```

Pbox.GSW.CC <- subset(Pbox,

```

```

      (Team=="Golden State Warriors" |
       Team=="Cleveland Cavaliers") &
      MIN>=500)
attach(Pbox.GSW.CC)
X <- data.frame(ID=Player, Team, V1=DREB/MIN, V2=STL/MIN,
               V3=BLK/MIN, V4=MIN)
detach(Pbox.GSW.CC)
labs <- c("Defensive Rebounds", "Steals", "Blocks",
         "Total minutes played")
bubbleplot(X, id="ID", x="V1", y="V2", col="V3",
          size="V4", text.col="Team", labels=labs,
          title="GSW and CC during the regular season",
          text.legend=TRUE, text.size=3.5, scale=FALSE)

```

#####

Πίνακας 5

#####

```

Pbox.OKC <- subset(Pbox, Team=="Oklahoma City Thunder"
                  & MIN>=500)

```

```

new <- Pbox.OKC[,c(2,10,11)]

```

```

res<-matrix(0,2,5)
res[1,2]<-v1<-var(new$P3A)
res[2,2]<-v2<-var(new$P3p)
res[1,3]<-sd1<-sqrt(v1)
res[2,3]<-sd2<-sqrt(v2)
res[1,1]<-m1<-mean(new$P3A)
res[2,1]<-m2<-mean(new$P3p)

```

```

res[1,4]<-cv1<-sd1/m1
res[2,4]<-cv2<-sd2/m2
res[1,5]<-r1<-range(new$P3A)[2]-range(new$P3A)[1]
res[2,5]<-r2<-range(new$P3p)[2]-range(new$P3p)[1]

```

```
res<-data.frame(res)
```

```

names(res)[1]<-"Mean"
names(res)[2]<-"Var"
names(res)[3]<-"SD"
names(res)[4]<-"CV"
names(res)[5]<-"Range"

```

```
res
```

```
#####
```

Γράφημα 8

```
#####
```

```

Pbox.OKC <- subset(Pbox, Team=="Oklahoma City Thunder"
& MIN>=500)

```

```

vrb1 <- variability(data=Pbox.OKC, data.var="P3p",
size.var="P3A")

```

```

vrb2 <- variability(data=Pbox.OKC,
data.var=c("P2p","P3p","FTp"),
size.var=c("P2A","P3A","FTA"),
weight=TRUE)

```

```
plot(vrb2, title="Variability diagram - OKC")
```

#####

Πίνακας 6

#####

```
Pbox.OKC <- subset(Pbox, Team=="Oklahoma City Thunder")
Pbox.OKC <- subset(Pbox.OKC, PTS>200)
Pbox.OKC <- Pbox.OKC[order(Pbox.OKC$PTS),]
Pbox.OKC$CPI <- 1:8
Pbox.OKC <- Pbox.OKC[,c(2,5,23)]
Pbox.OKC$CPTS <- cumsum(Pbox.OKC$PTS)
Pbox.OKC$CPIpct <- 100*Pbox.OKC$CPI/8
Pbox.OKC$CPTSpect <- round(100*Pbox.OKC$PTS/7997,2)
Pbox.OKC$CPTSpect <- cumsum(Pbox.OKC$CPTSpect)
Pbox.OKC
g1 <- sum(Pbox.OKC$CPIpct-Pbox.OKC$CPTSpect)
g2 <- sum(Pbox.OKC$CPIpct)
100*g1/g2
```

#####

Γράφημα 9

#####

```
Pbox.BN <- subset(Pbox, Team=="Brooklyn Nets")
ineqBN <- inequality(Pbox.BN$PTS, nplayers=8)
Pbox.MB <- subset(Pbox, Team=="Milwaukee Bucks")
ineqMB <- inequality(Pbox.MB$PTS, nplayers=8)
library(gridExtra)
p1 <- plot(ineqBN, title="Brooklyn Nets")
p2 <- plot(ineqMB, title="Milwaukee Bucks")
```

```
grid.arrange(p1, p2, nrow=1)
```

```
#####
```

Γράφημα 11

```
#####
```

```
subdata <- subset(PbP.BDB, player=="Kevin Durant")
subdata$xx <- as.numeric(subdata$original_x)/10-25.25
subdata$yy <- as.numeric(subdata$original_y)/10-41.75
shotchart(data=subdata, x="xx", y="yy", z="points",
           num.sect=5, type="sectors", scatter=TRUE)
```

```
#####
```

Πίνακας 7

```
#####
```

```
PbP.GSW <- subset(PbP, team=="GSW")
ev <- c("ejection", "end of period", "jump ball",
       "start of period", "unknown", "violation",
       "timeout", "sub", "foul", "turnover")
event.unsel <- which(PbP.GSW$event_type %in% ev)
PbP.GSW.ev <- PbP.GSW[-event.unsel,]
attach(PbP.GSW.ev)
T <- table(oppTeam, event_type, exclude=ev)
write.csv2(T, "C:\\Users\\user\\Downloads\\table.csv")
detach(PbP.GSW.ev)
library(vcd)
assocstats(T)
```

```
#####
```

Πίνακας 8

#####

```
library(dplyr)
library(lsr)
library(tibble)
FF <- fourfactors(Tbox, Obox)
attach(Tbox)
attach(FF)
X <- data.frame(PTS, P2M, P3M, FTM, REB=OREB+DREB, AST,
               STL, BLK, ORtg, DRtg)
detach(Tbox)
detach(FF)
Playoff <- Tadd$Playoff
eta <- sapply(X, function(Y){
  cm <- round(tapply(Y, Playoff, mean), 1)
  eta2 <- etaSquared(aov(Y~Playoff))[1]*100
  c(cm, round(eta2, 2))
}) %>%
t() %>%
as.data.frame() %>%
rename(No=N, Yes=Y, eta2=V3) %>%
rownames_to_column('rownm') %>%
arrange(-eta2)
column_to_rownames('rownm')
```

#####

Συσχετίσεις

#####

```
data <- subset(Pbox, MIN>=500)
attach(data)
X <- data.frame(AST, TOV)/MIN
detach(data)
cor(X$AST, X$TOV)

cor(X)
```

```
#####
Γράφημα 12
#####
```

```
data <- merge(Pbox, Tadd, by="Team")
data <- subset(data, MIN>=500)
attach(data)
X <- data.frame(PTS, P3M, P2M, REB=(OREB+DREB), AST,
               TOV, STL, BLK)/MIN
X <- data.frame(X, Playoff=Playoff)
detach(data)
corrmatrix <- corranalysis(X[,1:8], threshold=0.5)
plot(corrmatrix)
```

```
#####
Γράφημα 13
#####
```

```
PbP.GSW <- subset(PbP, team=="GSW")
netdata <- assistnet(PbP.GSW)
set.seed(7)
plot(netdata)
plot(netdata, layout="circle", edge.thr=20)
```

```
#####
```

Γράφημα 14

```
#####
```

```
cols <- paste0(c("a","h"), rep(1:5,each=2))
PbP.GSW.DG0 <- PbP.GSW[!apply(PbP.GSW[,cols], 1, "%in%",
                             x="Draymond Green"),]
netdata.DG0 <- assistnet(PbP.GSW.DG0)
set.seed(1)
plot(netdata.DG0)
```