



“Airbnb Pricing: Identifying Primary Determinants through Hedonic Pricing Methods”

Konstantinos Gkovedaros

Master Thesis

Department of economics, University of Macedonia,

Thessaloniki, March 2019

Abstract

In this thesis, we examine what determinants are the most significant in pricing an Airbnb listing, using the city of Athens as a case study and Rome as a testing ground for our findings. As far as we know, Athens has not been tested in empirical studies as a standalone example, so we hope this study will add to the knowledge of Airbnb hosts, policymakers and stakeholders in general. We present certain descriptive statistics that give an intuitive understanding of Airbnb’s presence in Athens. Furthermore, we identify nine core variables that can explain adequately the pricing of a listing and give additional weight to the distance variable, measured with different metrics as to better understand how a listing’s position can increase or decrease its price. Finally we give some views as to what the next steps might be in better comprehending how a listing is priced and how Athens stands out as a case study.

Keywords: Airbnb, Athens, hedonic pricing, host, accommodation industry, LASSO,

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Φουντά για την βοήθεια και την καθοδήγηση του, τους καθηγητές κ. Παναγιωτίδη και κ. Δεργιαδέ για τις πολύτιμες παρατηρήσεις, συμβουλές και ιδέες τους πάνω σε σημεία αλλά και στον προσανατολισμό της εργασίας και τον μεταπτυχιακό φοιτητή και φίλο κ. Λίταινα, για την γενικότερη βοήθεια του και γιατί πάντα μου προσέφερε την σκοπιά του οικονομολόγου. Τέλος, τους γονείς και την αδερφή μου για την πολύτιμη συμπαράσταση και την άνευ όρων στήριξη τους.

Πίνακας Περιεχομένων

1. Εισαγωγή	5
1.1 Το ποσοστό απόκρισης (response rate) και ο χρόνος απόκρισης (response time) του οικοδεσπότη και πως υπολογίζεται.....	5
1.2 Πως λειτουργεί το σύστημα αξιολόγησης της Airbnb;	6
1.3 Τι είναι ένας superhost;.....	7
1.4 Η διαδικασία κράτησης ενός καταλύματος στην πλατφόρμα της Airbnb	8
2. Επισκόπηση της βιβλιογραφίας	8
2.1 Η Airbnb και η οικονομία διαμοιρασμού	8
2.2 Η ηδονική τιμολόγηση.....	11
2.3 Airbnb και τιμολόγηση	12
2.4 Κίνητρο για την επιλογή της Αθήνας	13
3. Δεδομένα.....	16
3.1 Περιγραφή δεδομένων	16
3.2 Περιγραφικά στατιστικά και χάρτες των καταλυμάτων ανά γειτονιά	18
3.3 Ο ρόλος των αξιολογήσεων	34
3.4 Ανάλυση Κειμένου Αξιολογήσεων	36
3.5 Γιατί κάποιοι οικοδεσπότες είναι superhosts.....	38
3.6 Τύπος καταλυμάτων	38
3.7 Τοποθεσία	40
4. Μεθοδολογία.....	42
4.1 Μεταβλητές.....	42
4.2 Αποτελέσματα.....	45
4.3 Εκτιμώντας την απόσταση ως ερμηνευτική μεταβλητή	50
4.4 Έλεγχοι σιβαρότητας μοντέλου	54
5. Συμπεράσματα/Συζήτηση	58
5.1 Μελλοντική έρευνα.....	59
6. Αναφορές	60
7. Appendix.....	68

Πίνακας Σχημάτων

Figure 1: Γειτονιές κατά σειρά καταλυμάτων που βρίσκονται εντός των ορίων τους (μόνο όσες έχουν πάνω από 100 καταλύματα).....	21
Figure 2: Ιστόγραμμα τιμών για τα καταλύματα της Αθήνας	36
Figure 3: Διάγραμμα που δείχνει την αύξηση του αριθμού αξιολογήσεων διαχρονικά. Ημερομηνίες εκκίνησης είναι αυτές που εμφανίζονται πρώτη και τελευταία στα δεδομένα μας	37
Figure 4: Scatter plot του σκορ αξιολογήσεων στα καταλύματα	38
Figure 5: Ιστόγραμμα για τον τύπο καταλυμάτων.....	41
Figure 6: Ιστόγραμμα για την ιδιωτικότητα που προσφέρει το κατάλυμα.....	42
Figure 7: Διάγραμμα του εκτιμώμενου μέσου cross-validated σφάλματος για 50 τιμές του λ ($\lambda.length=50$).....	49
Figure 8: Διάγραμμα των τιμών του μέσου τετραγωνισμένου σφάλματος σε σχέση με τις τιμές του λ που δίνουμε	50
Figure 9: Ιστόγραμμα της τιμής ανά διανυκτέρευση για το σύνολο των καταλυμάτων που απομένουν μετά το φιλτράρισμα στην Ρώμη.....	59

Πίνακας Εικόνων

Image 1: Σήμα superhost	9
Image 2: Χάρτης της Airbnb, δείχνει καταλύματα στο Χαλάνδρι τα οποία δεν εμφανίζονται στο σετ δεδομένων μας.....	17
Image 3: Χάρτης Google (Google Map).....	23
Image 4: Χάρτης ενδιαφέροντος (Heatmap) δείχνει την τοποθεσία των καταλυμάτων εντός του Δήμου Αθηναίων	24
Image 5: Πλάκα - Plaka (Google Maps, 2019).....	25
Image 6: Κουκάκι – Koukaki (Google Maps, 2019)	25
Image 7: Νέος Κόσμος -Neos Kosmos (Google Maps, 2019)	26
Image 8: Παγκράτι –Pangrati (Google Maps, 2019)	26
Image 9: Κολωνάκι – Kolonaki (Google Maps, 2019).....	27
Image 10: Σταθμός Λαρίσης - Larissis (Google Maps, 2019).....	27
Image 11: Αμπελόκηποι – Ambelokipi (Google Maps, 2019).....	28
Image 12: Εξάρχεια – Exarcheia (Google Maps, 2019)	28
Image 13: Ψυρρή – Psyri (Google Maps, 2019).....	29
Image 14: Μεταξουργείο – Metaxourgeio (Google Maps, 2019)	29
Image 15: Νεάπολη – Neapoli (Google Maps, 2019).....	30
Image 16: Θησείο – Thiseio (Google Maps, 2019)	30
Image 17: Μετς - Mets (Google Maps, 2019)	31
Image 18: Πεδίον Άρεως - Pedion Areos (Google Maps, 2019).....	31
Image 19: Πλατεία Αττικής – Attiki (Google Maps, 2019).....	31
Image 20: Πετράλωνα – Petralona (Google Maps, 2019)	32

Image 21: Πατήσια - Patisia (Google Maps, 2019)	32
Image 22: Άγιος Νικόλαος - Agios Nikolaos (Google Maps, 2019)	33
Image 23: Γκάζι – Gazi (Google Maps, 2019)	33
Image 24: Κυψέλη – Kypseli (Google Maps, 2019).....	34
Image 25: Κολωνός – Kolonos (Google Maps, 2019).....	34
Image 26: Ίλisia – Ilisia (Google Maps, 2019)	35
Image 27: Νέφος λέξεων (wordcloud) των σχολίων που αφήνουν οι επισκέπτες στις αξιολογήσεις	39
Image 28: Χάρτης τοποθεσίας του σημείου αναφοράς που επιλέχθηκε	43
Image 29: Google Maps showing minimum walking distance from a random listing to the Parthenon	55
Image 30 Google Maps showing minimum transit distance from the same random listing to the Parthenon	55

Πίνακας Εξισώσεων

Equation 1: Εξίσωση haversine για τον υπολογισμό της απόστασης δύο σημείων	43
Equation 2: Εξίσωση που λύνει το πακέτο glmnet για penalized παλινδρομήσεις	46
Equation 3: Εξίσωση που λύνει το πακέτο glmnet για την lasso παλινδρόμηση	47

Πίνακας Πινάκων

Table 1: Γειτονιές κατά σειρά καταλυμάτων που βρίσκονται εντός των ορίων τους (μόνο όσες έχουν πάνω από 100 καταλύματα).....	22
Table 2: Περιγραφικά στατιστικά για το σύνολο των καταλυμάτων	35
Table 3: Lasso Coefficients for $\lambda=1$ (Athens).....	51
Table 4: Lasso "survivors", R-Squared and Estimate of standard error (Athens)	52
Table 5: OLS for LASSO "survivors", t-stats are in parentheses, and by default * is $p<0.05$, ** is $p<0.01$ and *** is $p<0.001$	53
Table 6: Robust Least Squares Regression with Lasso survivors for both transit time and walking time (both times in minutes), t-stats are in parentheses, and by default * is $p<0.05$, ** is $p<0.01$ and *** is $p<0.001$	56
Table 7: Robust and non-robust statistics for Robust Least Squares Regression with walking time	57
Table 8: Robust and non-robust statistics for Robust Least Squares Regression with transit time	58
Table 9: Lasso Coefficients for $\lambda=1$ (Rome)	60
Table 10: Lasso "survivors", R-Squared and Mean Estimate of standard error (Rome).....	60
Table 11: Robust Least Squares Regression with Lasso survivors for both transit time and walking time (both times in minutes), t-stats are in parentheses, and by default * is $p<0.05$, ** is $p<0.01$ and *** is $p<0.001$	61

Table 12: Robust and non-robust statistics for Robust Least Squares Regression for Rome's listings	62
--	----

1. Εισαγωγή

Η Airbnb είναι μία από τις δημοφιλέστερες πλατφόρμες βραχυπρόθεσμης ενοικίασης (short-term rental) παγκοσμίως. Από την ίδρυση της εταιρείας το 2008 και σύμφωνα με τα στοιχεία που παρατίθενται στον διαδικτυακό τόπο της εταιρείας «...η αγορά της Airbnb προσφέρει πρόσβαση σε περισσότερα από 5 εκατομμύρια καταλύματα, σε πάνω από 81.000 πόλεις σε 191 χώρες» (Airbnb Press Room, 2019). Επιπλέον, η πλατφόρμα έχει πάνω από 150 εκατομμύρια εγγεγραμμένους χρήστες (South EU Summit, 2019) με 640.000 από αυτούς να είναι οικοδεσπότες (hosts). Τα συνολικά οικονομικά μεγέθη της εταιρείας είναι εντυπωσιακά με δεδομένο το σχετικά μικρό χρονικό διάστημα –λιγότερο από 11 έτη- το οποίο λειτουργεί: η συνολική αποτίμηση της εταιρείας εκτιμήθηκε το 2016 στα 31 δισ. δολάρια (TechCrunch, 2019) με έσοδα για το 2018, περίπου 2.6 δισ. δολάρια (Business Insider, 2019) μόλις 8 έτη μετά την ίδρυση της.

Οι οικοδεσπότες της Airbnb μπορούν να καταχωρήσουν διαφορετικά είδη καταλυμάτων – από δωμάτια μέχρι ολόκληρα διαμερίσματα, αγκυροβολημένα σκάφη, ακόμα και κάστρα. Η Airbnb απαριθμεί σαν βασικές προϋποθέσεις για να είναι κάποιος οικοδεσπότης τις εξής:

- την παροχή βασικών ανέσεων (amenities) στις οποίες περιλαμβάνονται χαρτί τουαλέτας, σαπούνι, σεντόνια και σκεπάσματα και τουλάχιστον μια πετσέτα και ένα μαξιλάρι ανά επισκέπτη.
- την καλή και γρήγορη ανταπόκριση: οι οικοδεσπότες θα πρέπει να έχουν έναν υψηλό ρυθμό ανταπόκρισης απαντώντας στα ερωτήματα για κρατήσεις και στα αιτήματα των φιλοξενούμενων εντός 24 ωρών.
- την αποδοχή αιτημάτων κράτησης: η εταιρεία ζητά από τους οικοδεσπότες να αποδέχονται τα αιτήματα φιλοξενίας όποτε είναι διαθέσιμο το κατάλυμα
- την αποφυγή ακυρώσεων: η Airbnb δίνει εξαιρετική σημασία στην ακύρωση κρατήσεων και ζητά από τους οικοδεσπότες να αποφεύγουν κάτι τέτοιο
- την διατήρηση ενός υψηλού ποσοστού αξιολόγησης: το ποσοστό αξιολόγησης είναι η βαθμολόγηση της εμπειρίας ενός φιλοξενούμενου με το πέρας της φιλοξενίας, ωστόσο θα δούμε παρακάτω πιο αναλυτικά από τι αποτελείται και με ποια διαδικασία δίνεται (Airbnb, 2019).

Στις επόμενες σελίδες ξεκαθαρίζουμε κάποιες χρήσιμες έννοιες που θα χρειαστούν για την ερμηνεία των αποτελεσμάτων.

1.1 Το ποσοστό απόκρισης (response rate) και ο χρόνος απόκρισης (response time) του οικοδεσπότη και πως υπολογίζεται

Το ποσοστό απόκρισης και ο χρόνος απόκρισης υπολογίζονται από το πόσο γρήγορα ο οικοδεσπότης απαντά στα ερωτήματα και στα αιτήματα για κράτηση που δέχεται. Το ποσοστό απόκρισης σύμφωνα με την εταιρεία υπολογίζεται από το ποσοστό νέων ερωτημάτων και

αιτημάτων κράτησης που ο οικοδεσπότης απάντησε ή προενέκρινε εντός 24 ωρών, τις τελευταίες 30 ημέρες. Ο χρόνος απόκρισης είναι ο μέσος χρόνος που χρειάστηκε προκειμένου ο οικοδεσπότης να ανταποκριθεί στο αίτημα (ή αιτήματα). Εάν ο οικοδεσπότης έχει λάβει λιγότερα από 10 μηνύματα τις τελευταίες 30 μέρες, το ποσοστό και ο χρόνος απόκρισης θα υπολογιστούν βασισμένα στα 10 πιο πρόσφατα μηνύματα των τελευταίων 90 ημερών. Το ποσοστό απόκρισης προκειμένου να αποφασιστεί εάν ο οικοδεσπότης θα λάβει στάτους superhost, υπολογίζεται διαφορετικά και βασίζεται στην συμπεριφορά του οικοδεσπότη τις τελευταίες 365 ημέρες.

Προκειμένου να βελτιώσουν τον χρόνο και το ποσοστό απόκρισης οι οικοδεσπότες καλούνται να προ-εγκρίνουν ή να αρνηθούν αιτήματα κρατήσεων και να απαντήσουν σε νέα μηνύματα από τους επισκέπτες. Οι απαντήσεις πέραν του εικοσιτετραώρου θεωρούνται αργοπορημένες απαντήσεις και μειώνουν το ποσοστό και τον χρόνο απόκρισης. Η Airbnb προειδοποιεί τους οικοδεσπότες ότι αυτό επηρεάζει το που εμφανίζεται ο οικοδεσπότης στα αποτελέσματα αναζήτησης (Airbnb, 2019). Πέρα από τις δύο αυτές μετρήσεις, υπάρχει και το ποσοστό αποδοχής οικοδεσπότη (host acceptance rate) το οποίο υπολογίζεται ως ο αριθμός αιτημάτων κράτησης που αποδέχεται ο οικοδεσπότης, προς τα συνολικά αιτήματα που αποδέχεται (π.χ. acceptance rate 75% μπορεί να σημαίνει ότι από 4 αιτήματα διαμονής ο οικοδεσπότης αποδέχτηκε τα 3).

1.2 Πως λειτουργεί το σύστημα αξιολόγησης της Airbnb;

Μαζί με τις γραπτές κριτικές, οι επισκέπτες μπορούν να υποβάλλουν μια βασισμένη στο σύστημα των αστεριών συνολική αξιολόγηση, καθώς και επιμέρους αξιολογήσεις για κάποιες κατηγορίες. Οι οικοδεσπότες μπορούν να δουν την αξιολόγηση τους. Οι επισκέπτες μπορούν να υποβάλλουν αξιολόγηση για τις εξής κατηγορίες:

Συνολική εμπειρία (Overall experience): Οι επισκέπτες ρωτούνται για το πως ήταν συνολικά η εμπειρία τους

- Καθαριότητα (Cleanliness): Οι επισκέπτες ρωτούνται αν θεωρούν ότι ο χώρος ήταν καθαρός και τακτοποιημένος
- Ακρίβεια (Accuracy): Οι επισκέπτες ρωτούνται αν η καταχώρηση στην οποία έκαναν κράτηση στην πλατφόρμα ανταποκρίνεται με ακρίβεια στον χώρο που διέμειναν
- Αξία (Value): Οι επισκέπτες ρωτούνται αν η καταχώρηση παρείχε ικανοποιητική αξία σε σχέση με την τιμή (value-for-money). Είναι μια από τις κατηγορίες αξιολόγησης για τις οποίες οι επισκέπτες μπερδεύονται με αυτό που τους ζητείται να αξιολογήσουν
- Επικοινωνία (Communication): Οι επισκέπτες ρωτούνται πόσο ικανοποιητική ήταν η επικοινωνία με τον οικοδεσπότη πριν και κατά την διάρκεια της διαμονής τους
- Άφιξη (Arrival): Οι επισκέπτες ρωτούνται πόσο καλά πήγε η διαδικασία της άφιξης τους στον χώρο (check-in)
- Τοποθεσία (Location): Οι επισκέπτες ρωτούνται πως αισθάνονται για την γειτονία στην οποία βρίσκεται το κατάλυμα στο οποίο διέμειναν.

Σε κάθε κατηγορία, οι επισκέπτες βαθμολογούν από ένα έως πέντε αστέρια, και οι οικοδεσπότες μπορούν να δουν πως βαθμολογήθηκαν οι ίδιοι και κοντινοί επισκέπτες και σε πολλές περιπτώσεις λαμβάνουν οδηγίες με το πως μπορούν να βελτιώσουν το κατάλυμα τους. Ο αριθμός των αστεριών που επιδεικνύεται στην κορυφή μιας σελίδας αναζήτησης ενός καταλύματος είναι το σύνολο των αξιολογήσεων που οι φιλοξενούμενοι έχουν συμπληρώσει για αυτό το κατάλυμα. Στο τέλος της σελίδας αναζήτησης, είναι το σύνολο κάθε μεμονωμένης κριτικής για καθεμία από τις 6 κατηγορίες. Ο οικοδεσπότης θα πρέπει να λάβει τουλάχιστον 3 κριτικές πριν το συνολικό σκορ εμφανιστεί.

1.3 Τι είναι ένας superhost;

Ο superhost είναι ένας οικοδεσπότης (host) με εμπειρία, στον οποίον η Airbnb δίνει ξεχωριστό status. Καθώς ένας οικοδεσπότης γίνει superhost, η εταιρεία εμφανίζει ένα σήμα δίπλα στην φωτογραφία προφίλ του. Η Airbnb ελέγχει 4 φορές τον χρόνο την δραστηριότητα των superhosts ώστε να διασφαλίσει ότι τηρούν τις προδιαγραφές που έχουν τεθεί.

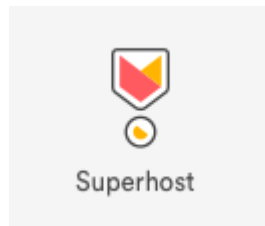


Image 1: Σήμα superhost

Προκειμένου να κερδίσουν το στάτους του superhost, οι οικοδεσπότες πρέπει να πληρούν τις παρακάτω προϋποθέσεις, τις τελευταίες 365 ημέρες:

- Να έχουν συμπληρώσει τουλάχιστον 10 ταξίδια ή να έχουν ολοκληρώσει τουλάχιστον 3 μακροχρόνιες διανυκτερεύσεις συνολικής διαμονής πάνω από 100 νύχτες
- Να έχουν διατηρήσει ποσοστό αξιολόγησης πάνω από 50%
- Να έχουν διατηρήσει ποσοστό απόκρισης πάνω από 90%
- Να έχουν μηδενικές ακυρώσεις με κάποιες εξαιρέσεις που η εταιρεία επισημαίνει

Προκειμένου να κερδίσει κάποιος το σήμα του superhost, δεν χρειάζεται να κάνει κάποια αίτηση αλλά η ίδια η Airbnb ελέγχει αν πληρούνται οι προδιαγραφές. Η επιλογή άμεσης κράτησης (instant book) είναι επίσης ένα στοιχείο που μπορεί να βοηθήσει τους οικοδεσπότες να γίνουν superhost. Τα καταλύματα που έχουν ενεργοποιήσει αυτή την επιλογή, δεν χρειάζονται έγκριση από τον οικοδεσπότη προκειμένου να γίνει κράτηση καθώς αυτή γίνεται άμεσα (Airbnb, 2019).

1.4 Η διαδικασία κράτησης ενός καταλύματος στην πλατφόρμα της Airbnb

Οι επισκέπτες αρχικά καλούνται να εισάγουν στην αρχική σελίδα της Airbnb, τον προορισμό, την ημερομηνία και τον αριθμό των επισκεπτών. Η παρουσία της επιλογής αριθμού ατόμων (number of guests) σε αυτό το στάδιο, τοποθετεί την δυναμικότητα ενός καταλύματος στα κύρια χαρακτηριστικά του, σύμφωνα με τους Gibbs et al. (2017). Έπειτα, οι επισκέπτες επιλέγουν αν αναζητούν ολόκληρο σπίτι (entire home), ιδιωτικό δωμάτιο (private room) σε σπίτι ή διαμοιρασμένο δωμάτιο (shared room) και τιμή. Ένας χάρτης στα δεξιά της οθόνης, δείχνει την τοποθεσία του καταλύματος και επιπλέον επιτρέπει στον χρήστη να «ψάξει» για καταλύματα βάσει αυτής. Το τρίτο και τελευταίο στάδιο επιτρέπει στους επισκέπτες να εισάγουν τα κριτήρια τους για τις παροχές (amenities) του καταλύματος (π.χ. wifi, κουζίνα, πισίνα κτλ), κριτήρια σχετικά με τον οικοδεσπότη (αν έχει το στάτους του superhost), την γειτονιά που επιθυμούν, το μέγεθος του καταλύματος (αριθμό κρεβατιών, μπάνιων), τον τύπο του καταλύματος (διαμέρισμα, loft) και άλλα. Εφόσον γίνει η κράτηση, ο φιλοξενούμενος χρεώνεται και σε 24 ώρες από την άφιξη ο οικοδεσπότης λαμβάνει το συνολικό ποσό της διαμονής. Αφού αποχωρήσει ο φιλοξενούμενος και τα δύο μέρη έχουν 14 ημέρες για να αφήσουν την κριτική τους, ο φιλοξενούμενος για την διαμονή και ο οικοδεσπότης για τον φιλοξενούμενο (Airbnb, 2019).

2. Επισκόπηση της βιβλιογραφίας

2.1 Η Airbnb και η οικονομία διαμοιρασμού

Η Airbnb είναι από τα πιο γνωστά παραδείγματα της οικονομίας διαμοιρασμού (sharing economy ή peer-to-peer economy) παγκοσμίως. Ως αγορά peer-to-peer, ορίζεται το αποκεντρωμένο μοντέλο οικονομίας στο οποίο δύο μέρη μπορούν να αγοράσουν και να πουλήσουν αγαθά ή υπηρεσίες, χωρίς την μεσολάβηση ενός τρίτου μέρους – μεσάζοντα ή χωρίς την χρήση μιας άλλης επιχείρησης (Investopedia, 2019). Οι αγορές αυτές έχουν αναδυθεί ως εναλλακτικοί πάροχοι αγαθών ή υπηρεσιών που παραδοσιακά προσφέρονται από καταξιωμένες βιομηχανίες (Zervas, Proserpio and Byers, 2017). Η ραγδαία μεγέθυνση πλατφορμών peer-to-peer όπως η Uber, το TaskRabbit και η Airbnb, οφείλονται πιθανότατα σε δύο κύριους παράγοντες: τις τεχνολογικές καινοτομίες και την ευελιξία της προσφοράς (Zervas, Proserpio and Byers, 2017).

Προηγούμενες μελέτες πάνω στην οικονομία διαμοιρασμού υποδεικνύουν ότι το σύστημα φήμης (reputation) για την δημιουργία εμπιστοσύνης μεταξύ αγοραστών και πωλητών είναι τεράστιας σημασίας για τις συναλλαγές peer-to-peer (Lee et al., 2015). Ο Belk (2014) αναφέρει ότι όπως οι αγοραστές αξιολογούν τους πωλητές στην πλατφόρμα του ebay μετά από μια αγορά, ή οι χρήστες σε πλατφόρμες διαμοιρασμού ταξιδιών (ride-sharing platforms), έτσι πολλές άλλες

υπηρεσίες χρησιμοποιούν το σύστημα online αξιολόγησης μετά την ολοκλήρωση της συναλλαγής. Επιπλέον, οι θετικές ή αρνητικές αναφορές από χρήστες που έχουν χρησιμοποιήσει ήδη το προϊόν ή την υπηρεσία που προσφέρεται, η ανάρτηση φωτογραφιών και βίντεο των χρηστών που προσφέρουν το προϊόν ή υπηρεσία, όλα αυτά βοηθούν στο χτίσιμο μιας «οικονομίας καλής φήμης» (reputational economy) η οποία βοηθά ώστε οι συναλλαγές μεταξύ αγνώστων να γίνουν πιο ασφαλείς και λιγότερο αβέβαιες (Belk, 2014).

Στην περίπτωση της Airbnb και έχοντας υπόψη της σημασία της διαχείρισης της φήμης, οι αξιολογήσεις των καταναλωτών αποτελούν δημοφιλές αντικείμενο μελέτης για τους ερευνητές που ασχολούνται με την αγορά της Airbnb. Οι Zervas, Proserpio και Byers (2017) βρίσκουν ότι οι αξιολογήσεις στην Airbnb είναι δραματικά πιο θετικές από αυτές που συναντώνται σε καθιερωμένες ηλεκτρονικές πλατφόρμες word-of-mouth όπως για παράδειγμα το TripAdvisor, ενώ οι Lee et al. (2015) χρησιμοποιούν δεδομένα από 4.178 καταλύματα για να καταλήξουν ότι η παλαιότητα ενός μέλους, ο αριθμός των “Wish lists” και το ποσοστό ανταπόκρισης του οικοδεσπότη είναι κοινωνικοί παράγοντες που σχετίζονται σημαντικά με τις πωλήσεις ενός καταλύματος. Στην ίδια έρευνα ωστόσο και σε ότι αφορά την τιμή του καταλύματος ωστόσο, η συνολική αξιολόγηση ενός καταλύματος δεν σχετίζεται σημαντικά με τις πωλήσεις του (Lee et al., 2015). Στον ξενοδοχειακό κλάδο, ο Andersson (2000) χρησιμοποιώντας δεδομένα για ξενοδοχεία και τις online αξιολογήσεις τους, δείχνει ότι μια αύξηση 1% στις αξιολογήσεις οδηγεί σε μια αύξηση 1.42% στα έσοδα ανά διαθέσιμο δωμάτιο. Ακόμη οι Ikkala και Lampinen (2014) δείχνουν ότι μια κοινή τακτική ανάμεσα στους οικοδεσπότες είναι να ανεβάζουν την τιμή του καταλύματος τους, καθώς συσσωρεύουν περισσότερο «κεφάλαιο φήμης» (reputational capital) στην Airbnb, δηλαδή όσο ο αριθμός των θετικών αξιολογήσεων αυξάνεται. Τέλος, οι Gibbs et al. (2017) δείχνουν ότι περισσότερες αξιολογήσεις συνδέονται με μια μείωση της τιμής του καταλύματος. Συμπεραίνουμε έτσι, ότι ενώ ο μηχανισμός της αξιολόγησης και της φήμης είναι σημαντικός για την εξέταση της οικονομίας διαμοιρασμού γενικά και της Airbnb συγκεκριμένα, δεν είναι ξεκάθαρο το πως επηρεάζουν οι αξιολογήσεις την τιμή ενός καταλύματος.

Εξαιρετικής σημασίας είναι και η επίδραση της Airbnb στον ξενοδοχειακό κλάδο. Αρκετές μελέτες έχουν επικεντρωθεί σε αυτήν, εξετάζοντας την επίδραση στα έσοδα των ξενοδοχείων (Zervas, Proserpio and Byers, 2017) και την εργασία στον κλάδο του τουρισμού (Fang, Ye and Law, 2016). Οι Zervas, Proserpio και Byers χρησιμοποιώντας δεδομένα από περίπου 3.000 ξενοδοχεία στην περιοχή του Texas, καταλήγουν ότι τα ανεξάρτητα ξενοδοχεία, τα οποία δεν απευθύνονται σε business επισκέπτες και τα μικρότερα και φθηνότερα ξενοδοχεία, είναι αυτά που επηρεάστηκαν περισσότερο από την Airbnb. Επιπλέον, οι Fang, Ye και Law (2016) βρίσκουν ότι η οικονομία διαμοιρασμού ευνοεί ολόκληρη την βιομηχανία του τουρισμού, δημιουργώντας νέες θέσεις εργασίας καθώς περισσότεροι τουρίστες επισκέπτονται την εκάστοτε περιοχή, λόγω του μικρότερου κόστους φιλοξενίας (Fang, Ye and Law, 2016). Ωστόσο, καθώς τα μικρότερα ξενοδοχεία υφίστανται σοκ και αντικαθίστανται από τα καταλύματα Airbnb, η οριακή επίδραση μειώνεται καθώς το μέγεθος της οικονομίας διαμοιρασμού αυξάνεται (Fang, Ye and Law, 2016).

Η επίδραση της Airbnb ως «διαταρακτική καινοτομία» (disruptive innovation) και η αμηχανία σε ορισμένες περιπτώσεις που η εμφάνιση της έχει προκαλέσει σε πολλές πόλεις ανά τον κόσμο, κυρίως αυτές που δέχονται τις περισσότερες επισκέψεις τουριστών, ξεπερνάει την επίδραση της στον ξενοδοχειακό κλάδο και ως θέμα έχει απασχολήσει αρκετά τους ερευνητές (Quattrone et al., 2016, Nieuwland and van Melik, 2018). Η εμφάνιση της πλατφόρμας συνέπεσε με την αύξηση της τουριστικής κίνησης παγκοσμίως και με το γεγονός ότι πόλεις σε όλο τον κόσμο προσπαθούν να ανταπεξέλθουν στα προβλήματα που προκαλεί η ραγδαία αύξηση των τουριστών προκαλώντας ακόμα και πορείες εναντίον των τουριστών σε πόλεις όπως η Βενετία ή η Βαρκελώνη (Coldwell, 2017). Ο τουριστικός τομέας σημειώνει ανεμπόδιστη πρόοδο στις αφίξεις για τα τελευταία 8 χρόνια συνεχόμενα, με κάθε έτος να αποτελεί και καινούργιο ρεκόρ, ενώ το 2017 μετακινήθηκαν 1.326 εκατομμύρια τουρίστες παγκοσμίως (United World Tourism Organization, 2019). Οι κάτοικοι των περιοχών που δέχονται αυτά τα μεγάλα τουριστικά κύματα σε πολλές περιπτώσεις εκφράζουν τις ανησυχίες τους σχετικά με τον υπερπληθυσμό στα μεγάλα αστικά κέντρα και τα αυξανόμενα ενοίκια που προκαλούνται εκτός των άλλων και από την αύξηση στις κατοικίες βραχυχρόνιας μίσθωσης, για τις οποίες η Airbnb είναι το πιο γνωστό και διαδεδομένο παράδειγμα (Gravari-Barbas & Guinand, 2017).

Η άνοδος της Airbnb, δεν επηρεάζει μόνο τον ξενοδοχειακό κλάδο και την βιομηχανία τουρισμού αλλά και την αγορά κατοικία, με θετικούς και αρνητικούς τρόπους (Ioannides, Röslmaier, & Van der Zee, 2018). Οι δύο μηχανισμοί με τους οποίους συμβαίνει αυτό σύμφωνα με τον Lee (2016) ο οποίος μελέτησε την αγορά κατοικία τους Los Angeles είναι α) η απλή μετατροπή: μια κατοικία στην οποία προηγουμένως διέμενε ένας κάτοικος της πόλης αλλά τώρα έχει καταχωρηθεί στην Airbnb, είναι μια κατοικία που έχει αφαιρεθεί από την αγορά κατοικίας και επί της ουσίας έχει προστεθεί στην ξενοδοχειακή προσφορά του Los Angeles και β) η «ξενοδοχειοποίηση» (hotelization). Όσο ένας ιδιοκτήτης κατοικίας ή ενοικιαστής μπορεί να εκμισθώσει ένα δωμάτιο στην πλατφόρμα για μια τιμή χαμηλότερη από αυτή ενός ξενοδοχείου, ενώ κερδίζει ένα σημαντικό premium σε σχέση με την αγορά κατοικίας, υπάρχει ένα κίνητρο ώστε κάθε μονάδα σε ένα κτίριο να καταχωρηθεί στην Airbnb (Lee, 2016). Αυτοί οι δύο μηχανισμοί δημιουργούν κατά τον ίδιο στρεβλώσεις στην αγορά εργασίας η οποία παραδοσιακά, δεν συνέπεφε με τον ξενοδοχειακό κλάδο. Ο ίδιος υποστηρίζει ότι οι πλατφόρμες βραχυχρόνιας μίσθωσης καταλυμάτων, αυξάνουν τα ενοίκια σε γειτονίες με υψηλή πυκνότητα σε καταλύματα Airbnb, μειώνουν την προσφορά κατοικίας ενθαρρύνοντας τις παράνομες μετατροπές κατοικιών σε Airbnb καταλύματα και τις εξώσεις (Lee, 2016).

Έτσι τα τελευταία χρόνια, μια σειρά από πολιτικές ρύθμισης (regulation) του πλαισίου στο οποίο γίνεται η ενοικίαση καταλυμάτων βραχυχρόνιας μίσθωσης (Ikkala & Lampinen, 2015; (Stors and Kagermeier, 2015). Η εταιρεία προωθεί την εμπειρία που προσφέρει ως μοναδική και αυθεντική φιλοξενία και ταξιδιωτική εμπειρία ενώ συνδέει τους ανθρώπους μεταξύ τους, ικανοποιώντας τις ανάγκες ενός νέου τύπου τουρίστα, του “urban tourist” (Nieuwland and van Melik, 2018). Μένοντας με έναν ντόπιο και έχοντας συζητήσεις με τον οικοδεσπότη (Belarmino et al., 2017) ο προορισμός μπορεί να βιωθεί σαν να ήταν ο τουρίστας ένας προσωρινός κάτοικος

σε αντίθεση με την διαμονή σε ένα παραδοσιακό και συνήθως πιο ακριβό ξενοδοχείο (Ioannides, Röslmaier and van der Zee, 2018).

Στον αντίποδα μιας νέας τουριστικής εμπειρίας, βρίσκονται ακόμα και ανησυχίες ότι, παρά το γεγονός ότι κάτι τέτοιο είναι παράνομο, κάποιες από τις επιδράσεις της «οικονομίας της Airbnb» είναι προσβλητικές για κάποιες πληθυσμιακές ομάδες (Biber et al., 2017). Για παράδειγμα, κάποιες έρευνες έχουν δείξει ότι οι οικοδεσπότες της εταιρείας, ασκούν πολιτική διακρίσεων σε υποψήφιους φιλοξενούμενους το οποίο είναι παράνομο στην ενοικίαση σπιτιών και την αγορά κατοικίας στις περισσότερες χώρες στις οποίες δραστηριοποιείται η εταιρεία, αλλά δεν είναι παράνομο στην βραχυχρόνια μίσθωση (Biber et al., 2017). Για παράδειγμα, οι Edelman και Luca (2014) ελέγχουν για ρατσιστικές διακρίσεις εναντίον οικοδεσποτών στην πλατφόρμα της Airbnb, χρησιμοποιώντας ένα νέο σετ δεδομένων και συνδυάζοντας φωτογραφίες όλων των οικοδεσποτών της Airbnb στη Νέα Υόρκη με τις τιμές και πληροφορία για την ποιότητα των καταλυμάτων τους και δείχνουν ότι οι μη-μαύροι οικοδεσπότες χρεώνουν περίπου 12% παραπάνω από τους μαύρους οικοδεσπότες για την αντίστοιχη πληρωμή διαμονής (Edelman and Luca, 2014).

2.2 Η ηδονική τιμολόγηση

Η ηδονική τιμή (hedonic price) εισάγεται ως όρος από τον Court το 1939 στο έργο «The dynamics of automobile demand» όπου αναφέρει ότι «...οι ηδονικές προτάσεις συνδυάζουν περισσότερες από μια προδιαγραφές ώστε να δημιουργήσουν μια σύνθετη και αντικειμενική μέτρηση χρησιμότητας και ελκυστικότητας ενός προϊόντος...» (Court, 1939, p. 107). Ο Court χρησιμοποιεί τις υποθέσεις αυτές πάνω σε επιβατικά αυτοκίνητα, ενώ το 1974 ο Rossen επεκτείνει τον όρο για να υποστηρίξει ότι μια ομάδα διαφοροποιημένων προϊόντων περιγράφεται πλήρως από ένα διάλυμα αντικειμενικά μετρήσιμων χαρακτηριστικών (Rosen, 1974). Έτσι, σύμφωνα με τον ίδιο, οι παρατηρηθείσες τιμές των προϊόντων και οι συγκεκριμένες τιμές των χαρακτηριστικών που συνδέονται με κάθε αγαθό ορίζουν ένα σετ εγγενών ή «ηδονικών» τιμών (Rosen, 1974). Λίγα χρόνια πριν, ο Lancaster (1966) κάνοντας παρόμοιες βασικές υποθέσεις ότι τα αγαθά διαθέτουν διάφορα χαρακτηριστικά σε καθορισμένες ποσότητες και πως σε αυτά τα χαρακτηριστικά οι προτιμήσεις των καταναλωτών ασκούνται και όχι στο προϊόν καθαυτό, κατασκευάζει ένα μοντέλο ερμηνείας αυτών των προτιμήσεων (Lancaster, 1966).

Η χρήση της ηδονικής τιμολόγησης στην αξιολόγηση της αγοράς κατοικίας είναι επίσης μια μέθοδος που έχει εξεταστεί και χρησιμοποιηθεί κατά κόρον από τους ερευνητές (Chau και Chin, 2002, Kask και Maani, 1992, Li και Li, 2018, Gillard, 1981). Συγκριτικά με άλλα καταναλωτικά αγαθά, η αγορά κατοικίας ξεχωρίζει καθώς επιδεικνύει τα χαρακτηριστικά της αντοχής, της ετερογένειας και της γεωγραφικής σταθερότητας (Chau and Chin, 2002). Έτσι, προκειμένου να μοντελοποιηθεί αυτή η διαφοροποίηση, έχει εισαχθεί η προσέγγιση της ηδονικής τιμολόγησης (Chau and Chin, 2002).

Τα σπίτια έχουν διακριτά χαρακτηριστικά, όπως το εμβαδό, τον αριθμό δωματίων, την τοποθεσία και περιβαλλοντικά χαρακτηριστικά (Champ, Boyle and Brown, 2003). Έτσι, για παράδειγμα, η εγγύτητα μιας κατοικίας σε ανοιχτό χώρο που είναι δημόσιο αγαθό και προσβάσιμο, όπως ένα πάρκο, σε όλους είναι για κάποιους ένα επιθυμητό χαρακτηριστικό και έτσι, όσο πιο κοντά μια κατοικία είναι σε δημόσιο χώρο, τόσο μεγαλύτερη η τιμή της σε σχέση με μια άλλη πανομοιότυπη που είναι πιο μακριά από έναν δημόσιο χώρο (Champ, Boyle and Brown, 2003).

Μερικά προβλήματα τα οποία έχουν αναφερθεί στην βιβλιογραφία, τα οποία σχετίζονται με τα ηδονικά μοντέλα είναι τα σφάλματα εξειδίκευσης, δηλαδή η προσθήκη στο μοντέλο άσχετων ανεξάρτητων μεταβλητών (over-specification) ή όταν σχετικές ανεξάρτητες μεταβλητές παραλείπονται (under-specification) (Chau and Chin, 2002). Ωστόσο, καθώς τα ηδονικά μοντέλα τιμολόγησης έχουν να κάνουν με τις εγγενείς τιμές των ποσοτήτων των χαρακτηριστικών ενός προϊόντος, το πρόβλημα του σφάλματος εξειδίκευσης είναι μάλλον αναπόφευκτο (Chau and Chin, 2002). Ο Butler προτείνει ότι μόνο εκείνα τα χαρακτηριστικά που είναι κοστοβόρα για να παραχθούν και δίνουν χρησιμότητα θα πρέπει να ληφθούν υπόψη για την εξίσωση παλινδρόμησης (Butler, 1982). Ωστόσο, σύμφωνα με τον ίδιο, καθώς όλες οι εκτιμήσεις ενός ηδονικού μοντέλου θα έχουν σφάλμα εξειδίκευσης, τα μοντέλα που χρησιμοποιούν μόνο έναν μικρό αριθμό σημαντικών μεταβλητών, θα αρκούν (Butler, 1982).

Η ηδονική προσέγγιση ωστόσο έχει και πλεονεκτήματα, με το κύριο από αυτά να είναι ότι ο ερευνητής χρειάζεται συγκεκριμένες πληροφορίες όπως η τιμή της κατοικίας, την σύσταση των χαρακτηριστικών της κατοικίας και μια σωστή εξειδίκευση των λειτουργικών σχέσεων (Chau and Chin, 2002). Οι οριακές τιμές των χαρακτηριστικών βρίσκονται εκτιμώντας τις παραμέτρους της εξίσωσης του ηδονικού μοντέλου (Chau and Chin, 2002). Είναι μια αρκετά σαφής μέθοδος καθώς μόνο οι εκτιμήσεις της ηδονικής παλινδρόμησης είναι απαραίτητες για να υποδείξουν την δομή της προτίμησης και δεν χρειάζονται επιπλέον πληροφορίες ή προσωπικά χαρακτηριστικά για τον πωλητή ή τον αγοραστή του σπιτιού (Chau and Chin, 2002).

2.3 Airbnb και τιμολόγηση

Κάθε κατάλυμα Airbnb αποτελείται από μια πλειάδα χαρακτηριστικών που έχουν να κάνουν τόσο με το κατάλυμα καθαυτό, όσο και τα χαρακτηριστικά του οικοδεσπότη που το καταχωρεί και στην βιβλιογραφία αναφέρονται μια σειρά από χαρακτηριστικά τα οποία έχουν μεγάλη επίδραση στις πωλήσεις και στην ζήτηση ενός καταλύματος (Chen and Xie, 2017). Για παράδειγμα, οι Lee et al. (2015) βρίσκουν ότι δύο τύποι χαρακτηριστικών εμφανίζουν έντονη συσχέτιση με τις πωλήσεις ενός καταλύματος. Το ένα είναι τα χαρακτηριστικά του καταλύματος, συμπεριλαμβανομένου της τιμής του, της ελάχιστης διαμονής και των ανέσεων που προσφέρει. Το άλλο είναι τα κοινωνικά χαρακτηριστικά που είναι συγκεκριμένα για κάθε κατάλυμα, συμπεριλαμβανομένου της ταχύτητας απόκρισης του οικοδεσπότη σε αιτήματα διαμονής, τον αριθμό των αξιολογήσεων και το διάστημα του χρόνου κατά το οποίο ο οικοδεσπότης συμμετέχει στην πλατφόρμα (Lee et al., 2015). Επιπλέον, σύμφωνα με τους Chen

και Xie (2017) υπάρχουν εξωτερικοί παράγοντες οι οποίοι επηρεάζουν την τιμή ενός καταλύματος όπως η πυκνότητα των ξενοδοχείων στις γειτονιές όπου βρίσκονται καταλύματα Airbnb (Chen and Xie, 2017).

Μια σειρά από έρευνες έχει επικεντρωθεί και στα αποτελέσματα και τις επιδράσεις των αξιολογήσεων και των φωτογραφιών των οικοδεσποτών. Οι Liang et al. (2017) υποστηρίζουν ότι οι οικοδεσπότες στους οποίους έχει δοθεί το στάτους του superhost, ανεβάζουν τις τιμές των καταλυμάτων τους, ιδιαίτερα όταν λαμβάνουν περισσότερες και καλύτερες (Liang et al., 2017). Επιπλέον, οι Ma et al. (2017) εξετάζουν την αντιλαμβανόμενη αξιοπιστία ενός προφίλ χρησιμοποιώντας θεματικά κωδικοποιημένα προφίλ από 1.200 οικοδεσπότες, δείχνοντας ότι οι μεγαλύτερες περιγραφές ενός οικοδεσπότη γίνονται αντιληπτές ως πιο αξιόπιστες ενώ δείχνουν ακόμα ότι υπάρχουν στρατηγικές (όπως η αναφορά σε ένα μείγμα θεμάτων) που οι οικοδεσπότες χρησιμοποιούν όταν αυτοπεριγράφονται και πως αυτές οι στρατηγικές προκαλούν διαφορές στα σκορ της αξιοπιστίας τους (Ma et al., 2017). Τέλος, οι Gibbs et al. (2017) βρίσκουν, εκτός των αναμενόμενων για το μέγεθος του καταλύματος, την απόσταση από το κέντρο (χρησιμοποιώντας ως κέντρο το δημαρχείο σε καθεμιά από τις 5 πόλεις του Καναδά τις οποίες ερευνούν) και την ιδιωτικότητα, δηλ. κατά πόσο οι φιλοξενούμενοι μένουν σε ολόκληρο σπίτι, ή σε δωμάτιο σπιτιού, πως σημαντικές είναι και οι παροχές της πισίνας και του εξοπλισμού γυμναστικής, τουλάχιστον σε δύο από τις μεγάλες αστικές περιοχές τις οποίες ελέγχουν (Gibbs et al., 2017). Τέλος, σύμφωνα με τους Ert, Fleischer, και Magen (2015) οι online αξιολογήσεις δεν εμφανίζονται να έχουν μεγάλη επίδραση στην τιμή ενός καταλύματος (Ert, Fleischer and Magen, 2015).

2.4 Κίνητρο για την επιλογή της Αθήνας

Σε ότι αφορά την παρουσία της Airbnb στην ελληνική αγορά, τα τελευταία χρόνια σειρά δημοσιευμάτων δείχνουν έναν έντονο προβληματισμό σε κοινωνικό και πολιτικό επίπεδο για την ραγδαία αύξηση των βραχυχρόνιων μισθώσεων. Η κυβέρνηση φαίνεται ότι προσπαθεί να εισάγει περιοριστικές ρυθμίσεις στην χρήση ακινήτων που διαφημίζονται σε πλατφόρμες όπως η Airbnb και το HomeAway (Tsiros, 2019). Οι περιορισμοί αυτοί, αναμένεται να εφαρμοστούν τόσο σε ιδιοκτήτες ενός ή δύο καταλυμάτων, όσο και σε επαγγελματίες οικοδεσπότες και εταιρίες διαχείρισης ακινήτων βραχυχρόνιας μίσθωσης (Tsiros, 2019)

Επιπλέον, σε έρευνα της Grant Thornton για λογαριασμό του ΣΕΤΕ (Σύνδεσμος Ελληνικών Τουριστικών Επιχειρήσεων) υποστηρίζεται ότι «...σε περιοχές τόσο της Αθήνας όσο και του υπολοίπου της χώρας παρατηρείται μια έντονη ανισορροπία μεταξύ του πλήθους των προσφερόμενων ακινήτων για μακροχρόνια μίσθωση σε σχέση με τη βραχυχρόνια. Η αναλογία ακινήτων βραχυχρόνιας μίσθωσης ως προς το σύνολο φτάνει μέχρι το 95%, ενώ η αναλογία των

τιμών αγγίζει το 5:1 μεταξύ βραχυχρόνιας και μακροχρόνιας μίσθωσης...» (Grand Thornton, 2019).

Ακόμα, η εστίαση μας στην Airbnb, προέρχεται από το γεγονός ότι είναι η πιο διαδεδομένη πλατφόρμα sharing economy στην Ελλάδα καθώς και από το γεγονός ότι τα δεδομένα των καταλυμάτων της βρίσκονται σε «κοινή θέα» είναι δηλαδή ανοιχτά σε όποιον θέλει να μελετήσει το φαινόμενο. Η Αθήνα είναι μία από τις πλέον τουριστικές πόλεις στον πλανήτη με ετήσιες αφίξεις περίπου 5.5 εκατομμύρια επισκέπτες για το 2018 (Krinis, 2019).

Συνεπώς η Αθήνα αποτελεί μια αρκετά ενδιαφέρουσα μελέτη περίπτωσης, δεδομένης της ραγδαίας ανάπτυξης τόσο του τουρισμού όσο και των ακινήτων που διατίθενται για βραχυχρόνια μίσθωση. Επιπλέον, έχοντας υπόψη τις προσπάθειες που γίνονται για την χάραξη πολιτικής και ρύθμισης του πεδίου των βραχυχρόνιων μισθώσεων, ευελπιστούμε να προσφέρουμε μια περαιτέρω γνώση πάνω σε αυτό το πεδίο, βασισμένη σε δεδομένα. Τέλος, στην δική μας περίπτωση, δεν ελέγχεται η κάθε γειτονιά ξεχωριστά, ωστόσο χρησιμοποιούμε τρεις μεταβλητές απόστασης, μια συνηθισμένη μέθοδος για τα ηδονικά μοντέλα (Gibbs et al., 2017).

3. Δεδομένα

Τα δεδομένα βρέθηκαν στην σελίδα Inside Airbnb και προσπελάστηκαν στις 25 Ιανουαρίου 2019. Το Inside Airbnb είναι ένα ανεξάρτητο, μη-εμπορικό σετ εργαλείων και δεδομένων που επιτρέπει στους χρήστες να εξερευνήσουν και να αντλήσουν δεδομένα για το πως τα καταλύματα της Airbnb λειτουργούν, σε διάφορες πόλεις του κόσμου. Το σετ δεδομένων δεν περιλαμβάνει όλες τις πόλεις στις οποίες το Airbnb έχει παρουσία, έτσι η αρχική επιλογή της Θεσσαλονίκης ως case study, στάθηκε αδύνατη. Αναλύοντας την διαθέσιμη στο κοινό πληροφορία για τα καταλύματα στο Airbnb μιας πόλης, η σελίδα Inside Airbnb παρέχει φίλτρα και σημαντικές μετρήσεις ώστε να εξαχθούν αξιόλογα συμπεράσματα σχετικά με την κατάσταση της αγοράς και ενοικίασης κατοικίας και του real estate συνολικότερα και τον ανταγωνισμό αυτής με το Airbnb.

3.1 Περιγραφή δεδομένων

Η σελίδα Inside Airbnb παρέχει κάποιες οπτικές αναπαραστάσεις των δεδομένων της Airbnb για την Αθήνα όπου μπορεί κανείς να δει τον τύπο των διαμερισμάτων, την δραστηριότητα, την διαθεσιμότητα και τις κατοικίες ανά οικοδεσπότη για όλα τα καταλύματα της Airbnb εντός του δήμου Αθηναίων. Θα πρέπει να σημειωθεί ότι ελέγχοντας περιοχές της Αθήνας οι οποίες δεν βρίσκονται στον δήμο Αθηναίων, τα καταλύματα της Airbnb λείπουν από το σετ δεδομένων.

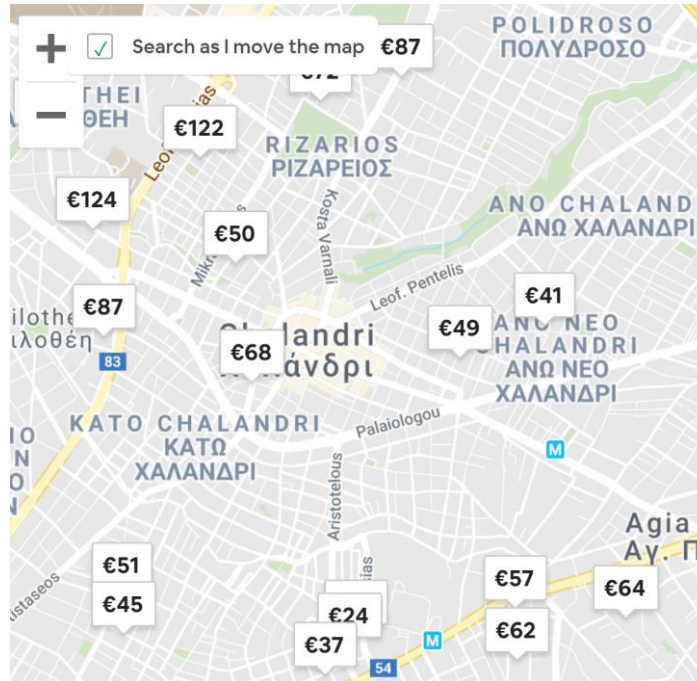


Image 2: Χάρτης της Airbnb, δείχνει καταλύματα στο Χαλάνδρι τα οποία δεν εμφανίζονται στο σετ δεδομένων μας

Για παράδειγμα, καταλύματα στο Χαλάνδρι, δεν περιλαμβάνονται στο σετ δεδομένων μας, ενώ στην αναζήτηση καταλύματος της Airbnb υπάρχουν, οπότε συμπεραίνουμε ότι το σετ δεδομένων μας συμπεριλαμβάνει μόνο τον Κεντρικό Τομέα Αθηνών.

Αφού κατεβάσουμε τα αρχεία “listings.csv.gz” και “listings.csv” εισάγουμε τα δεδομένα μας στην γλώσσα R και στο Excel ώστε να μπορέσουμε να τα διαμορφώσουμε ανάλογα.

Το ακόλουθο αρχείο της R έχει 9.123 σειρές (δηλ. 9.123 καταλύματα και δεδομένα για 96 μεταβλητές. Σύμφωνα με άλλη πηγή, την πλατφόρμα Airdna, η οποία προσφέρει δεδομένα και αναλυτικά στοιχεία για τις πλατφόρμες βραχυχρόνιας μίσθωσης όπως το Airbnb και το Homeaway ανά περιοχή, τα ενεργά καταλύματα στην Αθήνα ανέρχονται στα 8.250, οπότε εκτιμούμε ότι το σετ δεδομένων μας είναι πολύ κοντά στην πραγματικότητα. Από αυτές θα χρησιμοποιήσουμε ορισμένες:

- id: ένας μοναδικός αριθμός ταυτότητας καταλύματος
- listing_url: η ηλεκτρονική διεύθυνση του καταλύματος
- scrape_id: αριθμός αναγνώρισης scraping
- last_scraped: ημερομηνία scraping (στα δεδομένα μας, 15/11/2018)
- name: όνομα καταχώρησης
- summary: συνοπτικές πληροφορίες καταχώρησης
- space: Πληροφορίες για τον χώρο του καταλύματος
- description: Περιγραφή καταλύματος

- experiences_offered: Εμπειρίες που προσφέρονται από τον οικοδεσπότη
- host_neighbourhood: Γειτονιά οικοδεσπότη (συνήθως συμπίπτει με γειτονιά καταλύματος)
- notes: Σημειώσεις
- transit: Συγκοινωνία
- access: Πρόσβαση στο κατάλυμα
- interaction: Διάδραση με τον οικοδεσπότη κατά το check-in
- house_rules: Κανόνες καταλύματος
- picture_url: Ηλ. Διεύθυνση Εικόνας στην αρχική σελίδα της καταχώρησης
- host_id: αριθμός αναγνώρισης οικοδεσπότη
- host_url: ηλ. σελίδα προφίλ οικοδεσπότη
- host_name: όνομα οικοδεσπότη
- host_since: ημερομηνία από την οποία είναι οικοδεσπότης
- host_location: τοποθεσία οικοδεσπότη
- host_about: πληροφορίες για οικοδεσπότη
- host_response_time: χρόνος απόκρισης οικοδεσπότη
- host_response_rate: ποσοστό απόκρισης οικοδεσπότη
- host_acceptance_rate: ποσοστό αποδοχής οικοδεσπότη
- host_is_superhost: εάν ο οικοδεσπότης έχει το στατους του superhost
- host_picture_url: ηλ. διεύθυνση φωτογραφίας οικοδεσπότη
- host_total_listings_count: συνολικός αριθμός καταλυμάτων του οικοδεσπότη
- host_verifications: στοιχεία επαλήθευσης που έχει προσφέρει ο οικοδεσπότης
- host_has_profile_pic: ο οικοδεσπότης έχει εικόνα προφίλ
- host_identity_verified: ο οικοδεσπότης έχει επαληθεύσει την ταυτότητα του
- street: δρόμος καταλύματος
- neighbourhood: γειτονιά καταλύματος
- city: πόλη
- state: πολιτεία
- zipcode: ταχ. κωδικός
- smart_location: έξυπνη τοποθεσία (στην περίπτωση μας Αθήνα, Ελλάδα)
- country_code: κωδικός χώρας
- country: χώρα
- latitude: γεωγραφικό ύψος
- longitude: γεωγραφικό πλάτος
- is_location_exact: ακριβής τοποθεσία
- property_type: τύπος καταλύματος
- room_type: τύπος δωματίου (αν πρόκειται για δωμάτιο)
- accommodates: δυναμικότητα επισκεπτών που μπορεί να φιλοξενηθεί στο κατάλυμα
- bathrooms: αριθμός μπάνιων
- bedrooms: αριθμός κρεβατοκάμαρων
- beds: αριθμός κρεβατιών

- bed_type: τύπος κρεβατιού
- amenities: παροχές/ευκολίες (π.χ. WiFi, πάρκινγκ, κουζίνα, κτλ)
- square_feet: τετραγωνικά μέτρα
- price: τιμή/ημέρα (η προκαθορισμένη τιμή)
- weekly_price: τιμή/εβδομάδα (ενδέχεται να υπάρχουν εκπτώσεις από την ημερήσια τιμή)
- monthly_price: τιμή/μήνα
- security_deposit: κατάθεση ασφαλείας/προκαταβολή
- cleaning_fee: χρέωση καθαρίσματος
- guests_included: επιπλέον επισκέπτες που μπορούν να διαμείνουν
- minimum_nights: ελάχιστος αριθμός διανυκτερεύσεων
- maximum_nights: μέγιστος αριθμός διανυκτερεύσεων
- calendar_updated: τελευταία φορά που το ημερολόγιο ανανεώθηκε
- has_availability: έχει διαθεσιμότητα
- availability_30: διαθεσιμότητα 30 τελευταίων ημερών
- availability_60: διαθεσιμότητα 60 τελευταίων ημερών
- availability_90: διαθεσιμότητα 90 τελευταίων ημερών
- availability_365: διαθεσιμότητα τελευταίου έτους
- calendar_last_scraped: ημερομηνία τελευταίου scraping στο ημερολόγιο
- number_of_reviews: αριθμός αξιολογήσεων
- first_review: πρώτη αξιολόγηση
- last_review: τελευταία αξιολόγηση
- review_scores_rating: συνολική αξιολόγηση καταλύματος (στα 100)
- review_scores_accuracy: αξιολόγηση ακρίβειας (στα 10)
- review_scores_cleanliness: αξιολόγηση καθαριότητας (στα 10)
- review_scores_checkin: αξιολόγηση στο check-in (στα 10)
- review_scores_communication: αξιολόγηση στην επικοινωνία με τον οικοδεσπότη (στα 10)
- review_scores_location: αξιολόγηση τοποθεσίας (στα 10)
- review_scores_value: αξιολόγηση value-for-money (στα 10)
- license: αριθμός άδειας
- jurisdiction_names: ονομασία δικαιοδοσίας
- instant_bookable: άμεση κράτηση
- is_business_travel_ready: είναι φιλικό για επαγγελματικούς ταξιδιώτες
- cancellation_policy: πολιτική ακύρωσης
- require_guest_profile_picture: απαιτεί φωτογραφία προφίλ επισκέπτη
- require_guest_phone_verification: απαιτεί επαλήθευση επισκέπτη
- calculated_host_listings_count:
- reviews_per_month: αξιολογήσεις ανά μήνα

Από τα δεδομένα αυτά θα αφαιρέσουμε εκείνα τα καταλύματα τα οποία έχουν μηδενικά reviews, δηλαδή μάλλον δεν είχαν ποτέ διανυκτέρευση. Η λέξη «μάλλον» αναφέρεται στο γεγονός ότι δεν αφήνουν όλοι οι επισκέπτες κριτική σύμφωνα με την Airbnb (Airbnb, 2019) ωστόσο εφόσον δεν υπάρχει τρόπος να το γνωρίζουμε, αποκλείουμε τα καταλύματα αυτά και απομένουν 7.019 καταλύματα. Στην συνέχεια βλέποντας το παρακάτω ιστόγραμμα της τιμής ανά διανυκτέρευση για το σύνολο των καταλυμάτων, παρατηρούμε ότι υπάρχουν κάποιες ακραίες τιμές η μέγιστη των οποίων εμφανίζεται στα 1.500€ (πρόκειται για βίλα 700 τ.μ. στην Πλάκα, η οποία έχει ανανεωμένη τιμή στα 800€ ανά διανυκτέρευση). Επιλέγουμε να εξαιρέσουμε όλα τα καταλύματα με τιμή μεγαλύτερη των 200€ ανά διανυκτέρευση.

3.2 Περιγραφικά στατιστικά και χάρτες των καταλυμάτων ανά γειτονιά

Στις επόμενες σελίδες παρουσιάζουμε κάποια περιγραφικά στατιστικά για τις επιμέρους γειτονιές του κεντρικού τομέα της Αθήνας για τις οποίες έχουμε δεδομένα. Θα παραβλέψουμε τις γειτονιές εκείνες που έχουν λιγότερο από 100 καταλύματα (αυτές είναι κατά σειρά αριθμού καταλυμάτων οι Άγιος Λουκάς (43 καταλύματα), Βοτανικός (43), Μοναστηράκι (28), Κεραμεικός (27), Σεπόλια (19), Ακαδημία Πλάτωνος (15), Γουδί (13), Προφήτης Δανιήλ (13), Ριζούπολη (13), Ρουφ (2). Στις υπόλοιπες γειτονιές, δίνεται ο χάρτης της περιοχής στο Google Maps, και ένας πίνακας που περιλαμβάνει τις παρακάτω μεταβλητές:

1. την μέση τιμή καταλύματος στην εκάστοτε γειτονιά (Avg. Price)
2. την διάμεση τιμή καταλύματος στην εκάστοτε γειτονιά (Median Price)
3. τον μέσο χρόνο απόστασης περπατήματος κάθε καταλύματος από την Ακρόπολη σε δευτερόλεπτα (Avg. Walking Time)
4. τον μέσο χρόνο transit κάθε καταλύματος από την Ακρόπολη σε δευτερόλεπτα (Avg. Transit Time) δηλαδή συνδυασμού περπατήματος και χρήσης δημόσιας συγκοινωνίας. Αυτός ο χρόνος προέρχεται από το Distance API που δίνει το Google Maps και έχει να κάνει με την ώρα της ημέρας στην οποία καλείται η πληροφορία (π.χ. διαφορετικές τιμές άφιξης-αναχώρησης συρμού μετρό, για την ημέρα και τη νύχτα). Στην δική μας περίπτωση καλέσαμε αυτή την πληροφορία τις πρωινές ώρες της Πέμπτης 13 Φεβρουαρίου 2019.
5. τον μέσο όρο αξιολογήσεων κάθε καταλύματος της γειτονιάς (Avg. Review Rating), όπου σαν αξιολόγηση λαμβάνεται η συνολική αξιολόγηση που περιλαμβάνει τις επιμέρους 6 αξιολογήσεις
6. τον μέσο αριθμό κριτικών ανά μήνα που δέχεται κάθε κατάλυμα (Avg. Reviews/Months), ένας δείκτης που δείχνει πόσο «ενεργή» είναι η γειτονιά αυτή στην Airbnb, δηλαδή πόσο συχνά επιλέγουν επισκέπτες τα καταλύματα της γειτονιάς αυτής για την διαμονή τους
7. τον μέσο αριθμό των καταλυμάτων στην γειτονιά ανά οικοδεσπότη (Avg. N°. Of Listings/Host) που δείχνει κατά πόσο οι οικοδεσπότες σε αυτή την γειτονιά έχουν περισσότερα από ένα καταλύματα. Συνήθως οι οικοδεσπότες με περισσότερα καταλύματα είναι είτε εταιρείες διαχείρισης καταλυμάτων βραχυπρόθεσμης διαμονής, είτε επαγγελματίες της Airbnb και συνεπώς όσο περισσότερα καταλύματα κατά μέσο όρο

διαθέτει ένας οικοδεσπότης στην γειτονιά, τόσο πιο έντονη είναι η παρουσία επαγγελματιών στην γειτονιά αυτή. Παρακάτω είναι το διάγραμμα του αριθμού καταλυμάτων ανά γειτονιά. Πρώτη με διαφορά είναι η Πλάκα και ακολουθούν το Κουκάκι και ο Νέος Κόσμος. Διαισθητικά αυτό ενδέχεται να δείχνει την σημασία που έχει η τοποθεσία των καταλυμάτων σε σχέση με τον Παρθενώνα καθώς και οι 3 αυτές γειτονιές είναι πέριξ του βράχου της Ακρόπολης. Η σημασία της τοποθεσίας θα εξεταστεί στα επόμενα κεφάλαια.

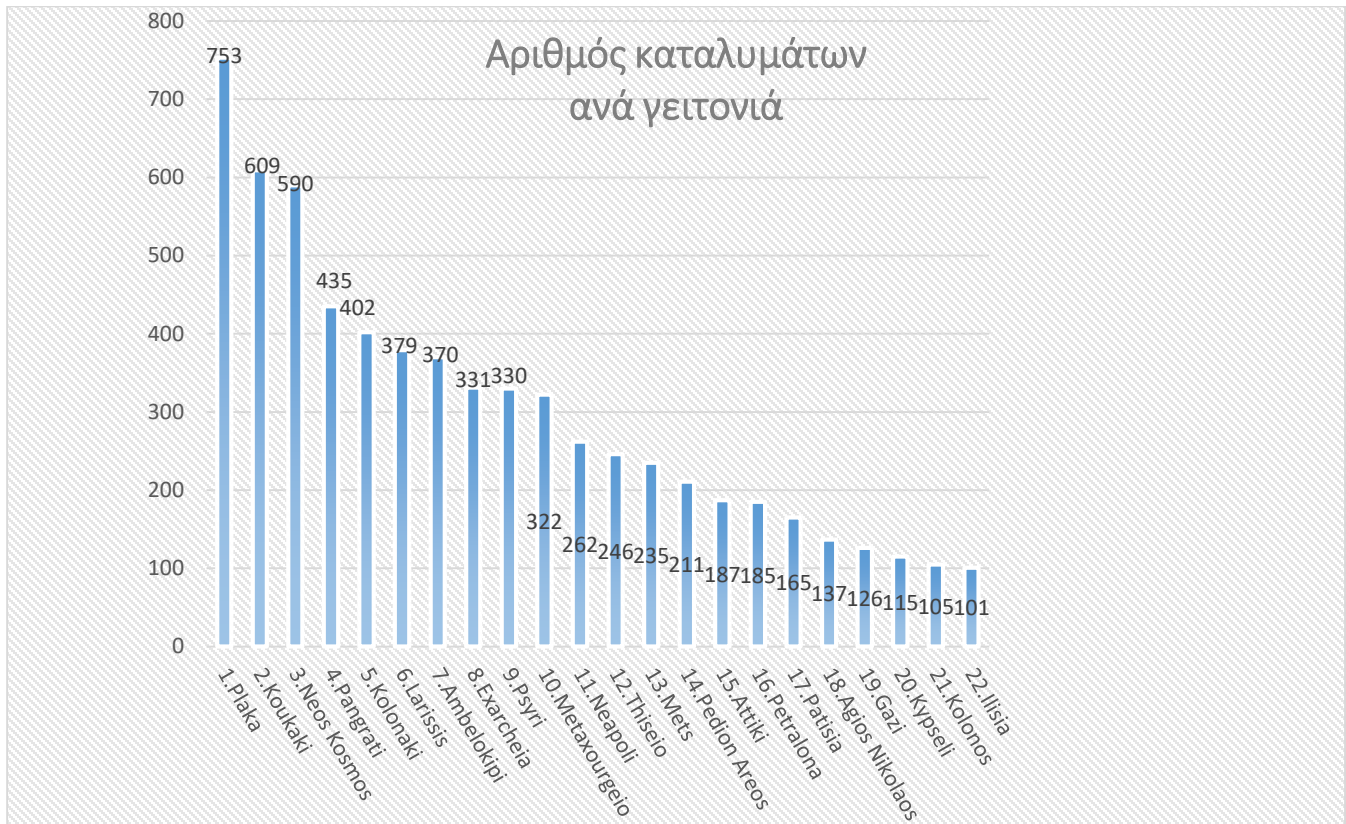


Figure 1: Γειτονιές κατά σειρά καταλυμάτων που βρίσκονται εντός των ορίων τους (μόνο όσες έχουν πάνω από 100 καταλύματα)

Οι γειτονιές ανά συχνότητα καταλύματος είναι οι εξής (επιλέγουμε τις 22 γειτονιές που έχουν πάνω από 100 καταλύματα):

1.Plaka	6.Larissis	11.Neapoli	16.Petralona	21.Kolonos
2.Koukaki	7.Ambelokipi	12.Thiseio	17.Patisia	22.Ilisia
3.Neos Kosmos	8.Exarcheia	13.Mets	18.Agios Nikolaos	
4.Pangrati	9.Psyri	14.Pedion Areos	19.Gazi	
5.Kolonaki	10.Metaxourgeio	15.Attiki	20.Kypseli	

Table 1: Γειτονιές κατά σειρά καταλυμάτων που βρίσκονται εντός των ορίων τους (μόνο όσες έχουν πάνω από 100 καταλύματα)

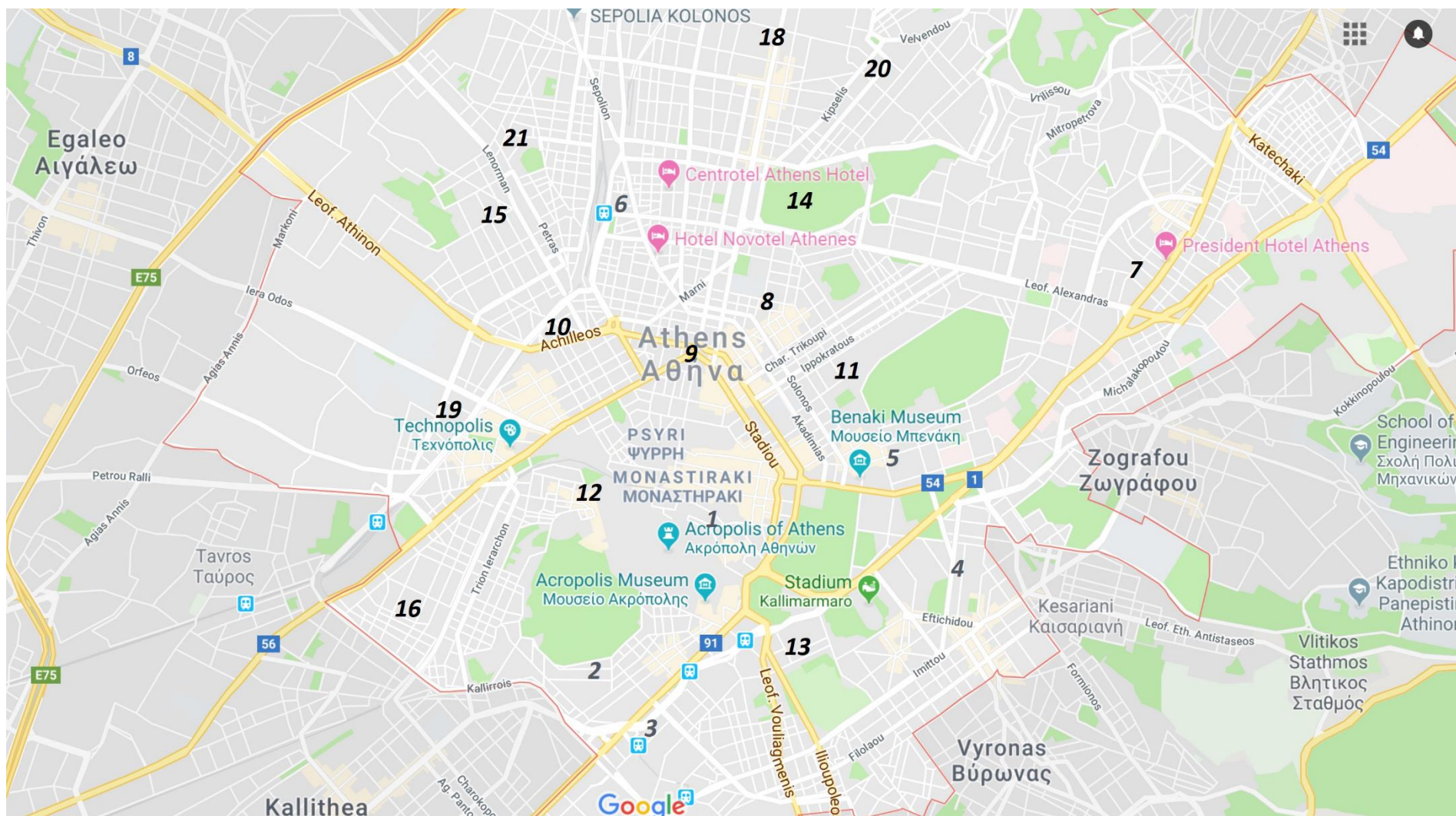


Image 3: Χάρτης Google (Google Map) του κέντρου της Αθήνας δείχνει τις παραπάνω γειτονίες. Οι γειτονίες με υψηλό αύξοντα αριθμό, φαίνεται ότι συνορεύουν με τον βράχο της Ακρόπολης.

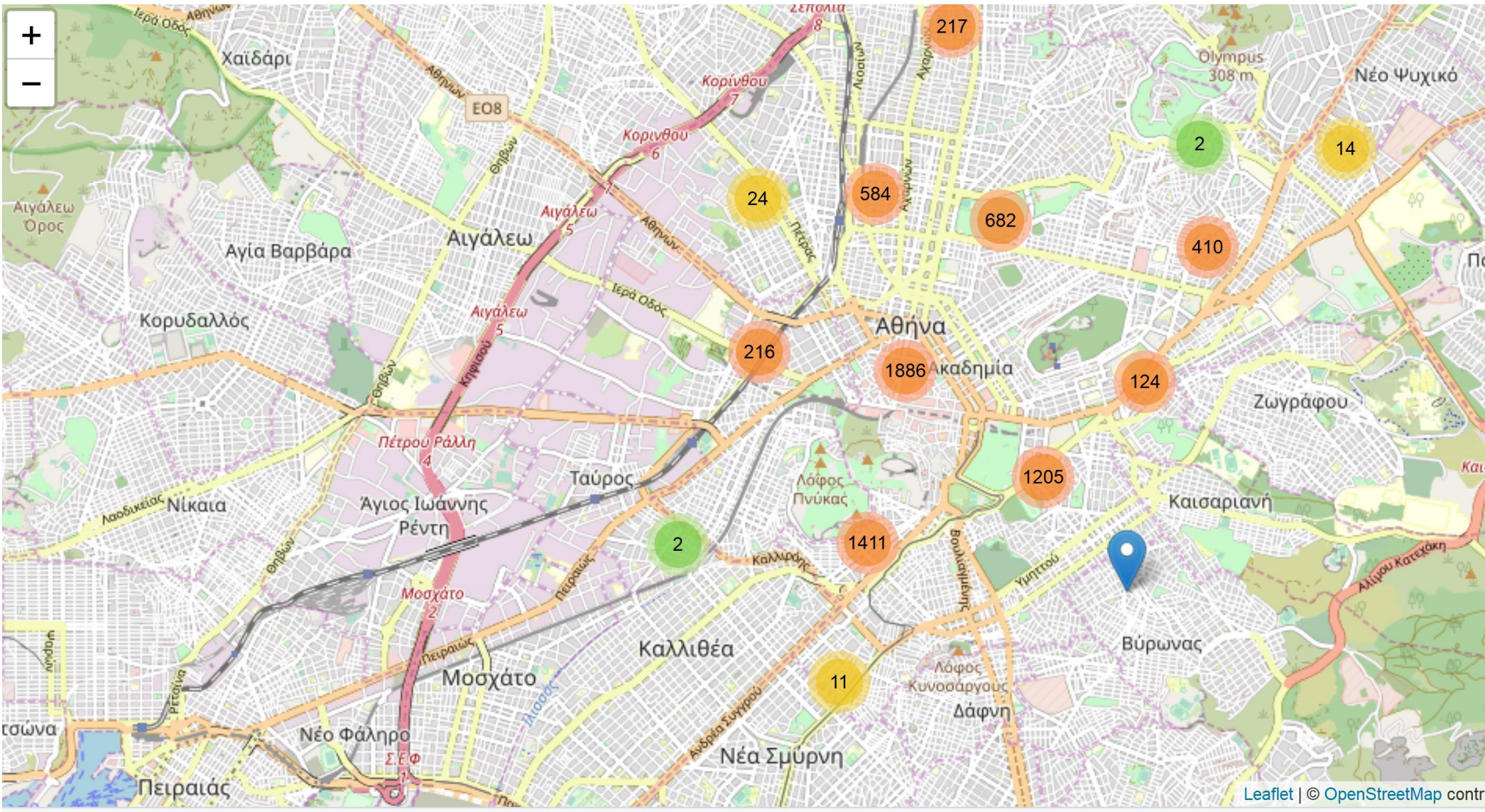


Image 4: Χάρτης ενδιαφέροντος (Heatmap) δείχνει την τοποθεσία των καταλυμάτων εντός του Δήμου Αθηναίων

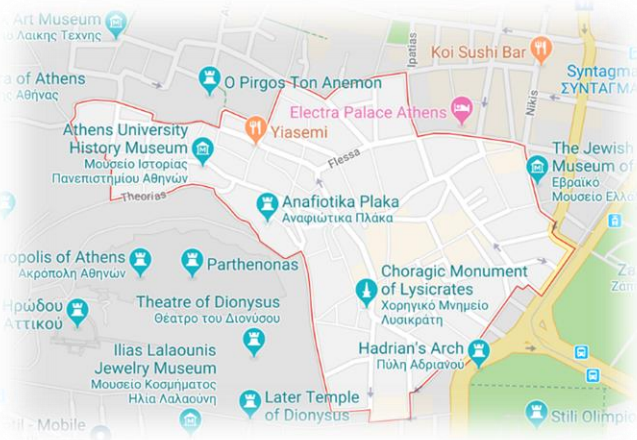


Image 5: Πλάκα - Plaka (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
76.55	70.00
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
803.08	873.69
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
96.07	3.30
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
11.56	4.05

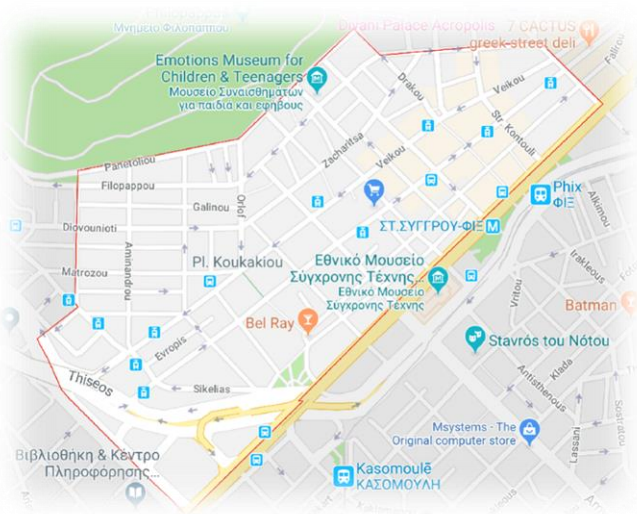


Image 6: Κουκάκι – Koukaki (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
58.79	50.56
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1263.07	1116.22
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.92	3.08
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
9.21	4.09

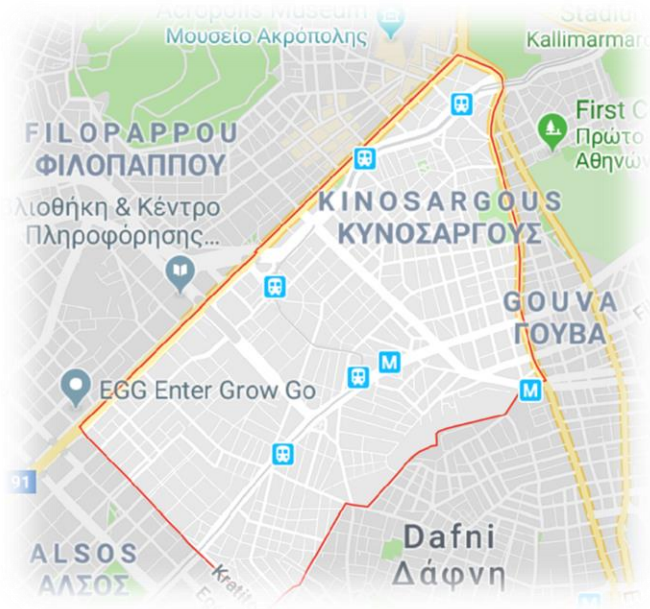


Image 7: Νέος Κόσμος -Neos Kosmos (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
54.11	47.83
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1460.48	1104.06
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.69	3.06
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
8.50	4.05

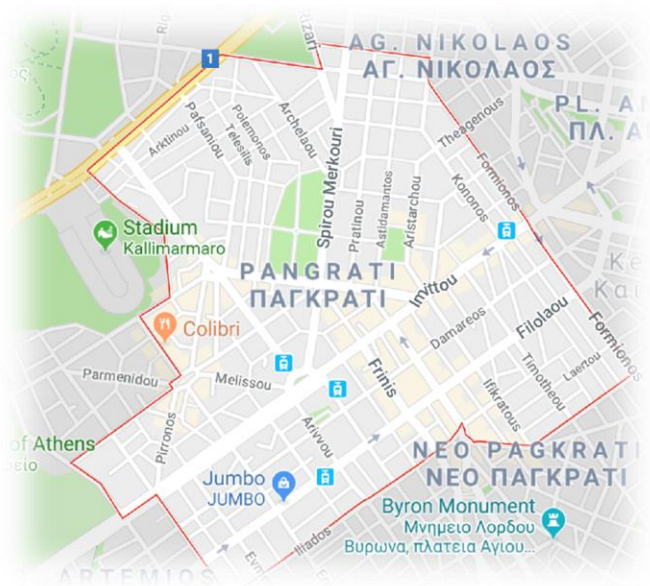


Image 8: Παγκράτι –Pangrati (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
49.80	44.44
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1886.07	1521.90
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.63	2.00
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
6.90	3.83

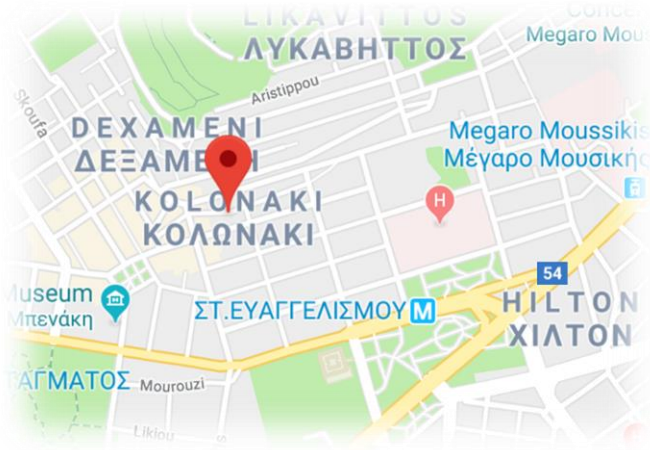


Image 9: Κολωνάκι – Kolonaki (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
74.29	64.56
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1718.04	1340.23
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.90	2.10
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
6.70	3.84

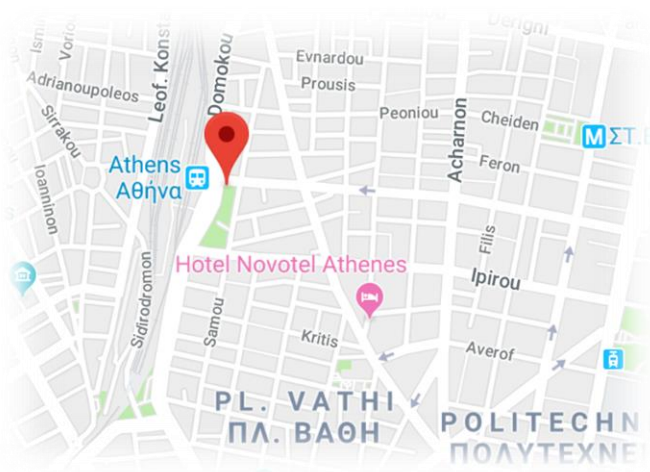


Image 10: Σταθμός Λαρίσης - Larissis (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
44.38	38.33
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
2277.00	1390.58
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
93.64	2.57
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
6.64	4.40

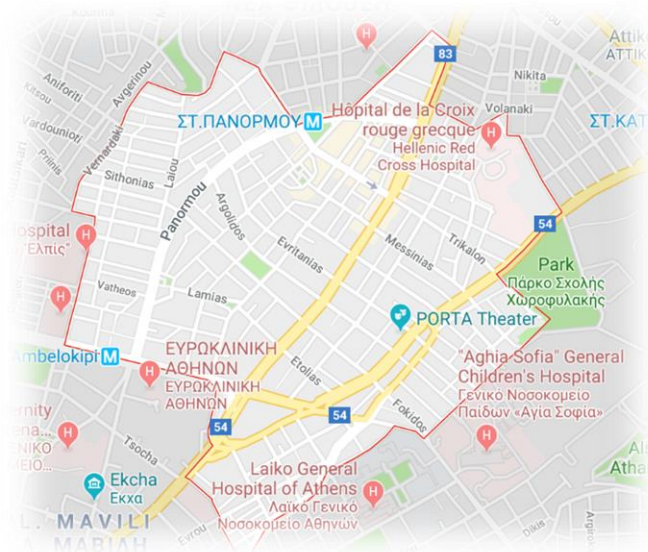


Image 11: Αμπελόκηποι – Ambelokipi (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
44.65	40.00
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
3485.04	1738.11
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.43	1.73
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
5.08	3.58



Image 12: Εξάρχεια – Exarcheia (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
48.41	41.22
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1998.629179	1460.529
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.31	2.43
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
5.05	3.80

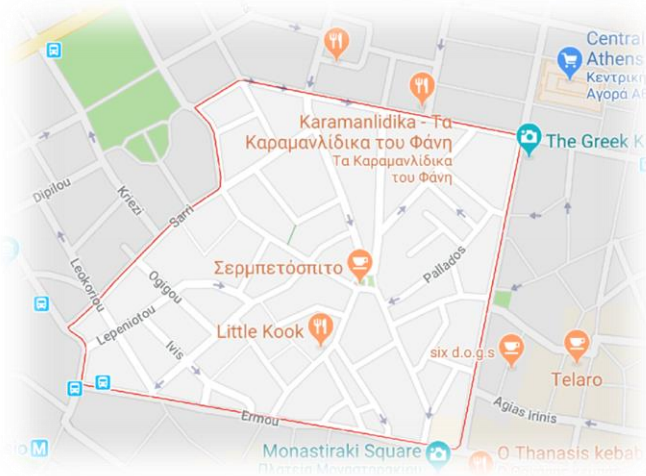


Image 13: Ψυρή – Psyri (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
69.83	59.67
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1219.93	1202.03
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
93.74	3.24
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
13.42	4.35

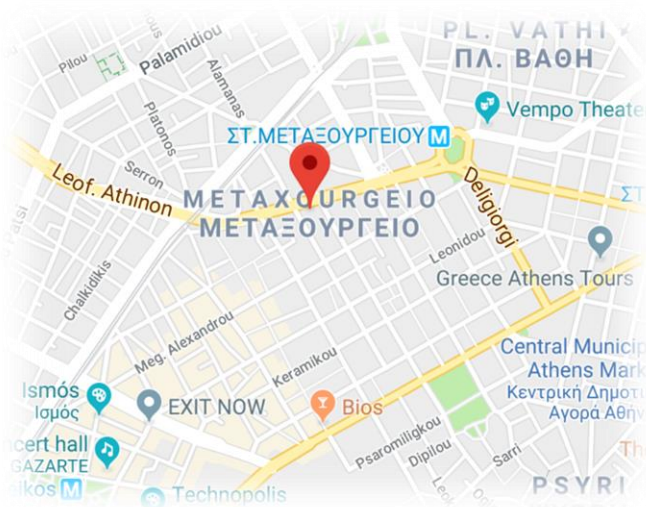


Image 14: Μεταξουργείο – Metaxourgeio (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
50.23	59.67
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1835.11	1297.76
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
92.81	2.87
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
8.31	4.09

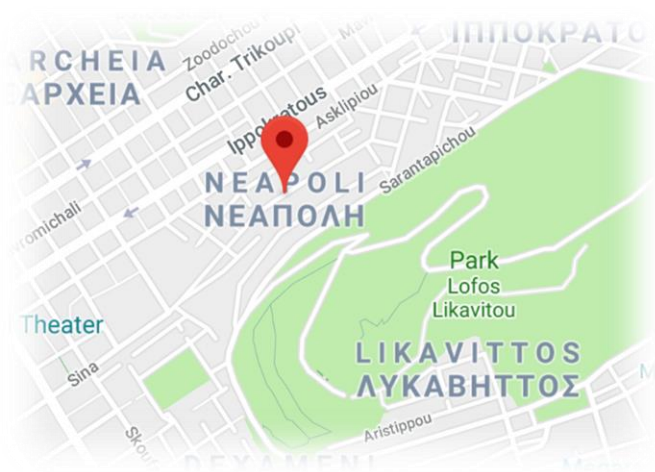


Image 15: Νεάπολη – Neapoli (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
42.90	38.78
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
2177.77	1722.62
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.14	2.00
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
4.60	3.35



Image 16: Θησείο – Thiseio (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
62.74	54.83
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1196.95	1218.92
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
96.57	3.08
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
7.36	4.31



Image 17: Μετς - Mets (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
50.52	43.33
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1562.93	1371.72
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
96.38	2.58
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
6.34	3.68

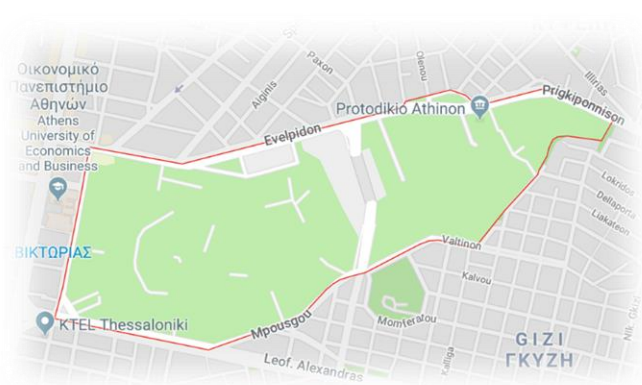


Image 18: Πεδίον Άρεως - Pedion Areos (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
40.19	35.00
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
2907.21	1963.27
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.33	1.70
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
5.46	3.84

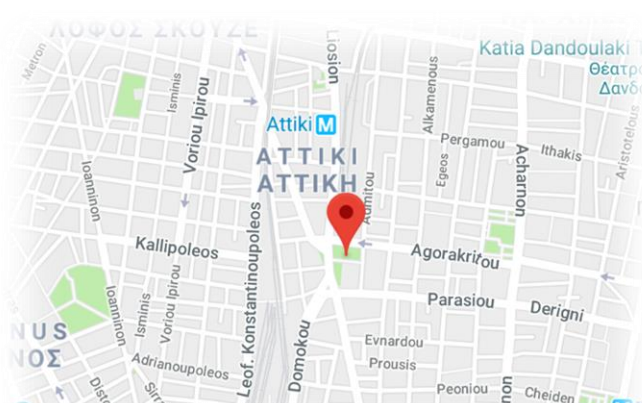


Image 19: Πλατεία Αττικής – Attiki (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
38.26	35.00
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
2869.36	1524.05
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
94.57	2.10
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
6.14	3.90



Image 20: Πετράλωνα – Petralona (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
46.83	42.22
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1641.52	1407.91
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.60	2.27
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
5.07	3.73

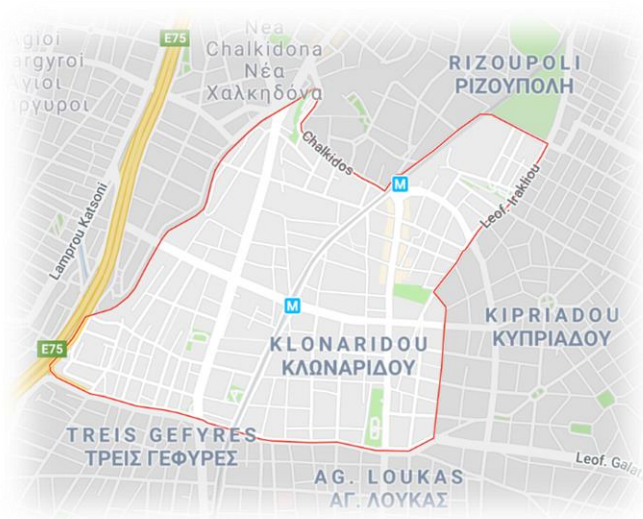


Image 21: Πατήσια - Patisia (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
38.72	32.22
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
4319.70	1755.37
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
94.82	1.73
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
2.88	3.73

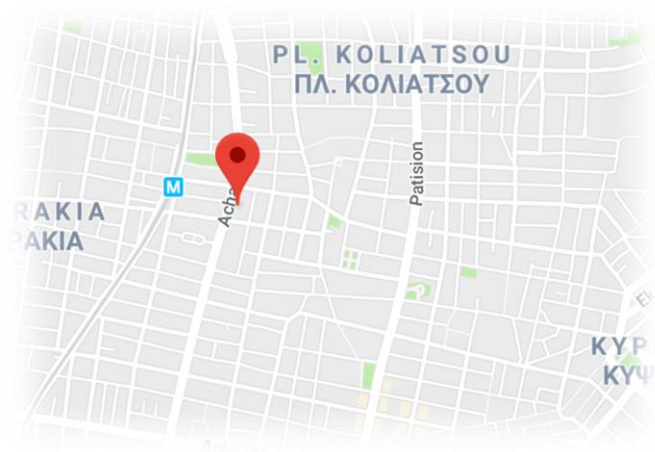


Image 22: Άγιος Νικόλαος - Agios Nikolaos (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
32.19	29.33
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
3443.95	1652.48
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
94.38	2.08
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
4.60	3.54

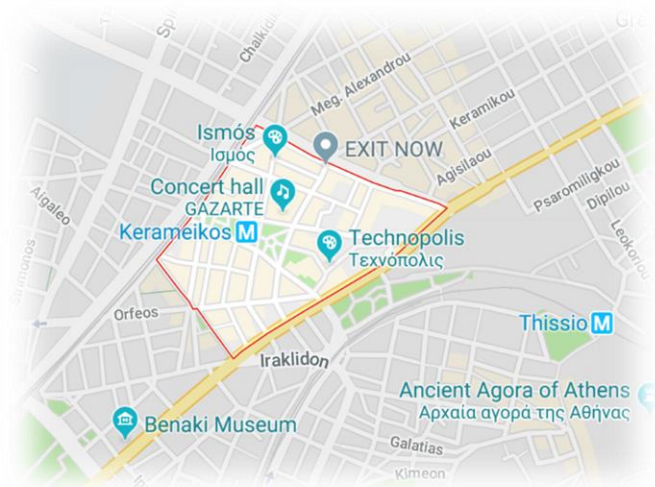


Image 23: Γκάζι – Gazi (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
59.74	51.33
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
1830.87	1304.06
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.76	2.86
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
7.07	4.37

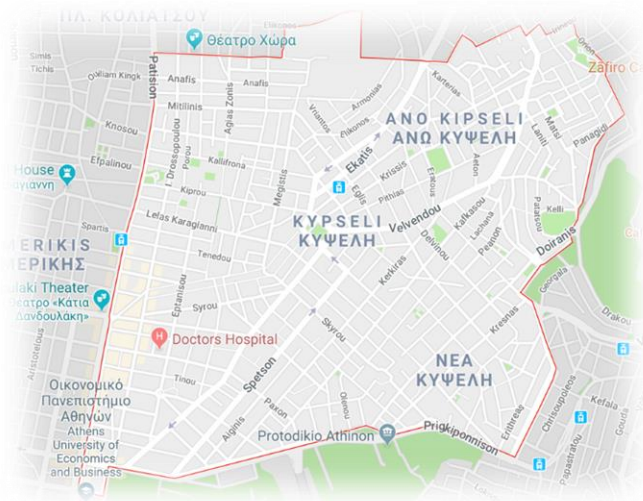


Image 24: Κυψέλη – Kypseli (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
37.75	30.56
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
3460.83	2176.10
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
95.10	1.59
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
4.60	3.53



Image 25: Κολωνός – Kolonos (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
41.38	38.17
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
2715.73	1662.74
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
96.78	2.62
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
3.48	3.53



Image 26: Ιλίσια – Ilisia (Google Maps, 2019)

<u>Avg. Price</u>	<u>Median Price</u>
54.35	48.78
<u>Avg. Walking Time</u>	<u>Avg. Transit Time</u>
2480.13	1524.23
<u>Avg. Review Rating</u>	<u>Avg.Reviews/Month</u>
92.97	1.69
<u>Avg. Nr. of Listings/Host</u>	<u>Avg. People</u>
4.55	4.03

Στο παρακάτω διάγραμμα βλέπουμε πως κατανέμεται η τιμή των καταλυμάτων με την πλειοψηφία αυτών να βρίσκονται κάτω των 200 €. Επιλέγουμε να μην λάβουμε υπόψη τιμές μεγαλύτερες από 200€, δηλαδή θεωρούμε οποιαδήποτε τιμή μεγαλύτερη από αυτό το όριο outlier.

Εν κατακλείδι, εάν η Αθήνα ήταν ένα μοναδικό κατάλυμα, θα ίσχυαν τα παρακάτω:

Avg. New Price 54.25 €	Median New Price 45.55 €
Avg. Walking Time 32.33 minutes	Avg. Transit Time 22.8 minutes
Avg. Reviews Score 95.29	Avg, Reviews/Month 2.52
Avg. Nr. of Listings/Host 11.48	Avg. People Capacity 3.93

Table 2: Περιγραφικά στατιστικά για το σύνολο των καταλυμάτων

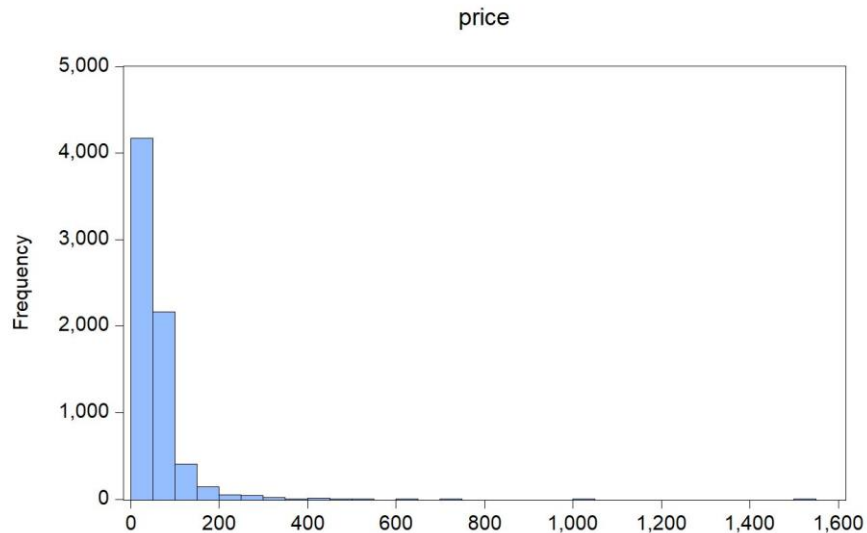


Figure 2: Ιστόγραμμα τιμών για τα καταλύματα της Αθήνας

3.3 Ο ρόλος των αξιολογήσεων

Στην συνέχεια, από τα δεδομένα που έχουμε, θα τυπώσουμε τον συνολικό αριθμό αξιολογήσεων των καταλυμάτων, για κάθε μοναδικό κατάλυμα στον χρόνο. Παίρνουμε την μεταβλητή *number_of_reviews* στον χρόνο, όπου χρόνος είναι τα έτη από το 2010 όπου έχουμε την πρώτη αξιολόγηση (review).

Η πρώτη αξιολόγηση στο σετ δεδομένων μας (και πιθανότατα σε ολόκληρη την Αθήνα), εμφανίζεται στις 8 Ιουλίου του 2010. Η τελευταία αξιολόγηση εμφανίζεται μία ημέρα πριν το scraping των δεδομένων δηλαδή στις 14 Νοεμβρίου 2018. Συνολικά έχουμε 284.742 αξιολογήσεις σε όλα τα καταλύματα της Αθήνας κατά την διάρκεια του διαστήματος που ελέγχουμε. Σύμφωνα με την εταιρεία, περίπου 50% των συνολικών επισκεπτών αφήνουν κριτική, οπότε αν αυτό ισχύει για την Αθήνα θα μπορούσε κανείς (κάπως αυθαίρετα) να συμπεράνει ότι έχουν φιλοξενηθεί συνολικά 569.484 επισκέπτες, δηλαδή πάνω από μισό εκατομμύριο επισκέπτες. Από το διάγραμμα βλέπουμε ξεκάθαρα ένα εποχιακό μοτίβο, όπου οι αξιολογήσεις αυξάνονται από την αρχή του έτους, κορυφώνονται κατά τους θερινούς μήνες για να πέσουν και πάλι προς το τέλος του έτους.

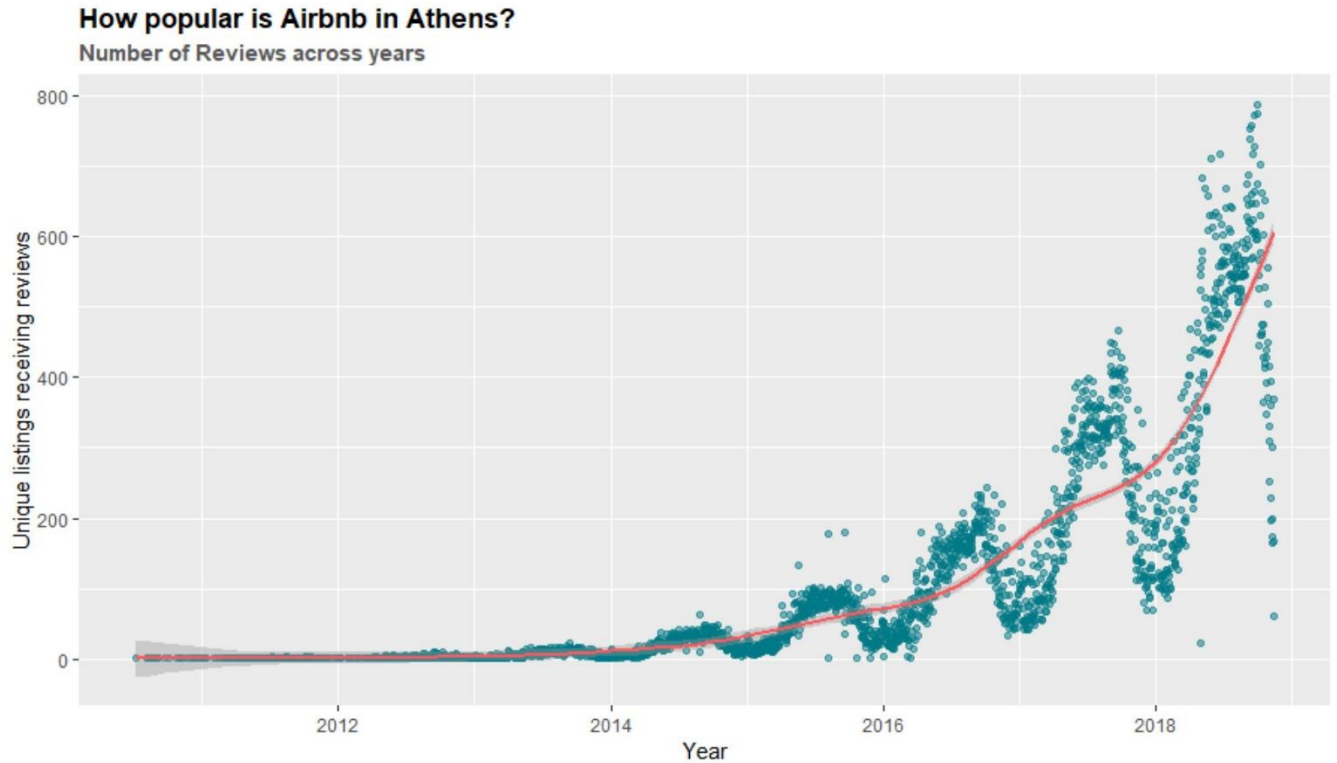


Figure 3: Διάγραμμα που δείχνει την αύξηση του αριθμού αξιολογήσεων διαχρονικά. Ημερομηνίες εκκίνησης είναι αυτές που εμφανίζονται πρώτη και τελευταία στα δεδομένα μας

Το πρόβλημα με τις αξιολογήσεις, φαίνεται στο παρακάτω scatter plot όπου απεικονίζουμε την συνολική αξιολόγηση κάθε καταλύματος (η αξιολόγηση που περιλαμβάνει τις υπόλοιπες έξι και δίνεται από 0 έως 100). Βλέπουμε ότι η συντριπτική πλειοψηφία των συνολικών αξιολογήσεων των καταλυμάτων βρίσκονται από το 90 έως το 100. Οι λόγοι για αυτό μπορεί να είναι η αυστηρότητα της εταιρείας σε καταλύματα που συγκεντρώνουν κριτικές κάτω από ένα όριο, ο τρόπος που ποσοτικοποιούνται οι αξιολογήσεις των επισκεπτών κ.ο.κ. Όπως ήδη αναφέρθηκε και παραπάνω, οι αξιολογήσεις στην πλατφόρμα είναι πολύ υψηλότερες (κοντά στο μέγιστο) σε σχέση με άλλες πλατφόρμες όπως το Tripadvisor (Zervas, Proserpio and Byers, 2017). Αυτό για εμάς σημαίνει ότι η διακύμανση αυτής της μεταβλητής είναι πολύ χαμηλή και δεν θα μας παρέχει ιδιαίτερη πληροφορία εφόσον αποφασίσουμε να την λάβουμε υπόψη σαν ανεξάρτητη μεταβλητή. Ένας άλλος λόγος που ενδέχεται να δημιουργήσει προβλήματα είναι η πιθανή ενδογένεια με την ερμηνευτική μεταβλητή της τιμής, καθώς στην βιβλιογραφία καταγράφεται ότι οι οικοδεσπότες ανεβάζουν τις τιμές τους ώστε να επηρεάσουν τον αριθμό των αξιολογήσεων και κατ'επέκταση και τον βαθμό της συνολικής τους αξιολόγησης. Θα επιλέξουμε για αυτούς τους λόγους να μην δώσουμε μεγάλη βαρύτητα στην επιλογή ερμηνευτικών μεταβλητών σχετικών με την αξιολόγηση των καταλυμάτων

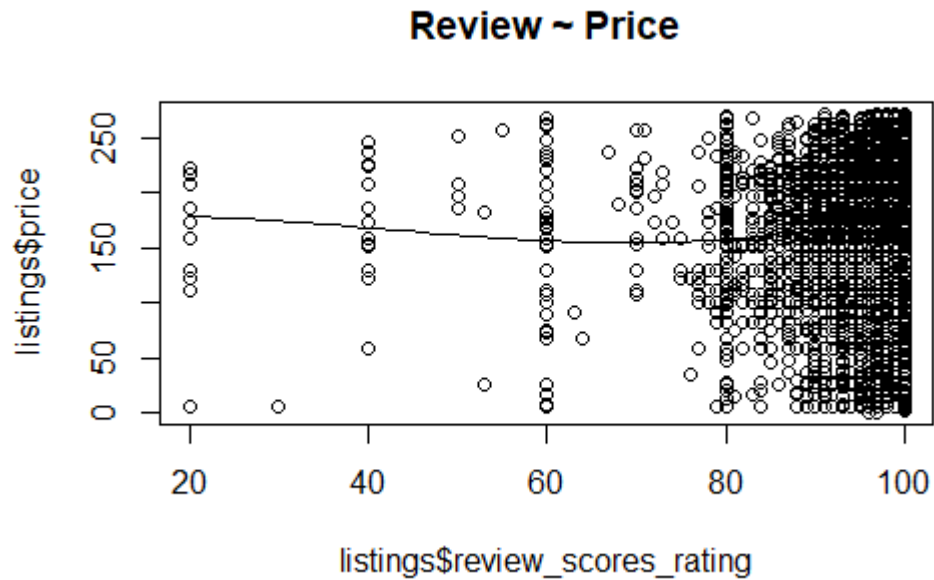


Figure 4: Scatter plot του σκορ αξιολογήσεων στα καταλύματα

Έπειτα, προκειμένου να δούμε τι είδους καταλύματα προσφέρονται στην Αθήνα, δημιουργούμε ανάλογα με το δημοτικό διαμέρισμα στο οποίο βρίσκεται κάθε γειτονιά τις εξής κατηγορίες:

- 1^ο Δημοτικό Διαμέρισμα: Πλάκα, Κουκάκι, Κολωνάκι, Εξάρχεια, Νεάπολη, Θησείο, Γλίσια, Ψυρή, Μοναστηράκι
- 2^ο Δημοτικό Διαμέρισμα: Νέος Κόσμος, Μετς, Παγκράτι
- 3^ο Δημοτικό Διαμέρισμα: Πετράλωνα, Βοτανικός, Μεταξουργείο, Γκάζι, Κεραμεικός, Προφήτης Δανιήλ, Ρουφ
- 4^ο Δημοτικό Διαμέρισμα: Πλατεία Αττικής, Κολωνός, Ακαδημία Πλάτωνος, Σεπόλια
- 5^ο Δημοτικό Διαμέρισμα: Ριζούπολη, Πατήσια, Άγιος Λουκάς
- 6^ο Δημοτικό Διαμέρισμα: Πεδίον του Άρεως, Σταθμός Λαρίσης, Άγιος Νικόλαος, Κυψέλη
- 7^ο Δημοτικό Διαμέρισμα: Αμπελόκηποι, Γουδί

3.4 Ανάλυση Κειμένου Αξιολογήσεων

Αναλύοντας τα σχόλια που αφήνουν οι επισκέπτες κατά την αξιολόγηση της διαμονής τους, μπορούμε να βγάλουμε ορισμένα χρήσιμα συμπεράσματα. Αρχικά, θα επικεντρωθούμε μόνο στα

σχόλια που είναι γραμμένα στα αγγλικά καθώς υπάρχουν κάποιες δυσκολίες επεξεργασίας των σχολίων στις υπόλοιπες γλώσσες και την γλώσσα προγραμματισμού που χρησιμοποιούμε (R), ωστόσο αυτό δεν είναι πρόβλημα καθώς η μεγάλη πλειοψηφία των σχολίων είναι στην αγγλική και κατά δεύτερον, θα εμφανιστούν ακόμα και τα σχόλια σε διαφορετική γλώσσα από την αγγλική. Τα σχόλια αυτά βρίσκονται στο dataset “reviews” στην στήλη “comments”. Αφαιρούμε τα σημεία στίξεως και τα stopwords της αγγλικής. Μια διευκρίνιση: stopwords ονομάζονται συνηθισμένες λέξεις μιας γλώσσας, εδώ για παράδειγμα “the”, “that”, “is”, κοκ. που δεν προσφέρουν κάτι στην ανάλυση που θέλουμε να κάνουμε κι επειδή εμφανίζονται με εξαιρετικά υψηλή συχνότητα, θα εκτόπιζαν τις λέξεις που μας ενδιαφέρουν. Δημιουργούμε λοιπόν ένα νέφος λέξεων (wordcloud) στο οποίο εμφανίζονται οι λέξεις που συναντώνται στα σχόλια των επισκεπτών με την μεγαλύτερη συχνότητα. Όσο μεγαλύτερο είναι το μέγεθος με το οποίο εμφανίζεται μια λέξη, τόσο μεγαλύτερη η συχνότητα εμφάνισης.



Image 27: Νέφος λέξεων (wordcloud) των σχολίων που αφήνουν οι επισκέπτες στις αξιολογήσεις

Βλέπουμε ότι οι πιο συνηθισμένες λέξεις είναι: *everything*, *clean*, *host*, *metro*, *perfect* και με χαμηλότερη συχνότητα: *easy*, *even*, *located*, *amazing*, *flat*, *highly*, *central*, *spacious*, *excellent*, *view* και κάποιες λέξεις σε άλλες γλώσσες όπως *avons*, *pour*, *ist*, *und*, *είναι*, *σε* (τα οποία θα

είχαν φιλτραριστεί αν χρησιμοποιούσαμε φίλτρο για stopwords στις αντίστοιχες γλώσσες) όπως επίσης και κάποια ελληνικά ονόματα όπως: *Maria*, *George* καθώς είναι σύνηθες για τους επισκέπτες να αφήνουν κριτικές αναφέροντας τους οικοδεσπότες τους.

Κάνοντας μια προσπάθεια να ερμηνεύσουμε τα αποτελέσματα των σχολίων των επισκεπτών, βλέπουμε ότι μετράει αρκετά η τοποθεσία (metro, location), το πόσο ευρύχωρο είναι το κατάλυμα (spacious, convenient), η καθαριότητα (clean) η θέα του καταλύματος (view) και ο οικοδεσπότης (host, Maria, George), με την ευρεία έννοια, δηλαδή ο οικοδεσπότης είναι σημαντικό συστατικό σε όλα τα στάδια του ταξιδιού, αλλά δεν γνωρίζουμε που ακριβώς είναι πιο καθοριστική η παρουσία του (επικοινωνία πριν την άφιξη, ευκολία στο check-in, διάδραση κατά την διαμονή, αναχώρηση επισκέπτη).

3.5 Γιατί κάποιοι οικοδεσπότες είναι superhosts

Από τα δεδομένα μας, αφαιρούμε τους οικοδεσπότες εκείνους για τους οποίους η μεταβλητή *host_response_time* δηλ. ο χρόνος απόκρισης οικοδεσπότη, είναι κενή. Υπάρχουν 4 πιθανές τιμές για αυτή την μεταβλητή: λιγότερο από ώρα, λιγότερο από λίγες ώρες, λιγότερο από μέρα, λιγότερο από λίγες μέρες. Όπως είδαμε η Airbnb δίνει μεγάλη βαρύτητα στον χρόνο απόκρισης και είναι μια από τις προϋποθέσεις που θέτει η εταιρεία προκειμένου ένας οικοδεσπότης να αποκτήσει το στατους του superhost. Από τις τιμές που μένουν (5.187) το 46,9% των καταλυμάτων ανήκει σε superhosts. Το πρόβλημα εδώ είναι ότι είναι λογικό να περιμένει κάποιος ότι ένας superhost θα έχει περισσότερα από ένα καταλύματα καθώς ενδέχεται να ασχολείται με την φιλοξενία επαγγελματικά. Επιπλέον, οι οικοδεσπότες στο 85% των καταλυμάτων αποκρίνονται σε λιγότερο από μία ώρα στα αιτήματα κράτησης και τις ερωτήσεις που δέχονται, το 10% εντός λίγων ωρών, το 4% εντός μιας ημέρας και μόλις το 1% σε περισσότερο από μια ημέρα. Αυτό ενισχύει την υπόθεση ότι η εταιρεία επιβάλλει «ποινές» στα καταλύματα των οικοδεσποτών εκείνων οι οποίοι δεν αποκρίνονται γρήγορα στα αιτήματα κρατήσεων.

Τέλος, από το ποσοστό απόκρισης οικοδεσπότη (host response ratio) το οποίο στα δεδομένα μας παίρνει τιμές είτε 0, είτε 1, προκύπτει ότι το 85% των οικοδεσποτών των καταλυμάτων έχουν ποσοστό απόκρισης 1. Εδώ υποθέτουμε ότι ποσοστό απόκρισης 1 σημαίνει μεγαλύτερο του 50% και 0 μικρότερο ή ίσο του 50%.

3.6 Τύπος καταλυμάτων

Ακόμα ένα στοιχείο που προκύπτει από τα δεδομένα είναι ο τύπος των καταλυμάτων που προσφέρονται. Από το φιλτράρισμα των δεδομένων προκύπτουν οι παρακάτω 21 κατηγορίες:

Apartment, Loft, House, Bed and breakfast, Condominium, Townhouse, Serviced apartment, Earth house, Guest suite, Hostel, Guesthouse, Hotel, Aparthotel, Boutique hotel, Villa, Cycladic house (Greece), Cottage, In-law, Tent, Tiny house, Other. Μία από τις αρχικές υποθέσεις μας ήταν ότι τα καταλύματα αποτελούνται κυρίως από διαμερίσματα (Apartments) και πράγματι αυτό δείχνει να επαληθεύεται από τα δεδομένα για περίπου το 90% των καταλυμάτων. Η υπόθεση αυτή είναι σημαντική καθώς η πλατφόρμα της Airbnb είναι επιτρέπει και την καταχώρηση ξενοδοχείων κάτι που πιθανώς θα δημιουργούσε προβληματικά συμπεράσματα καθώς τα ξενοδοχεία παρουσιάζουν διαφορετικά χαρακτηριστικά.

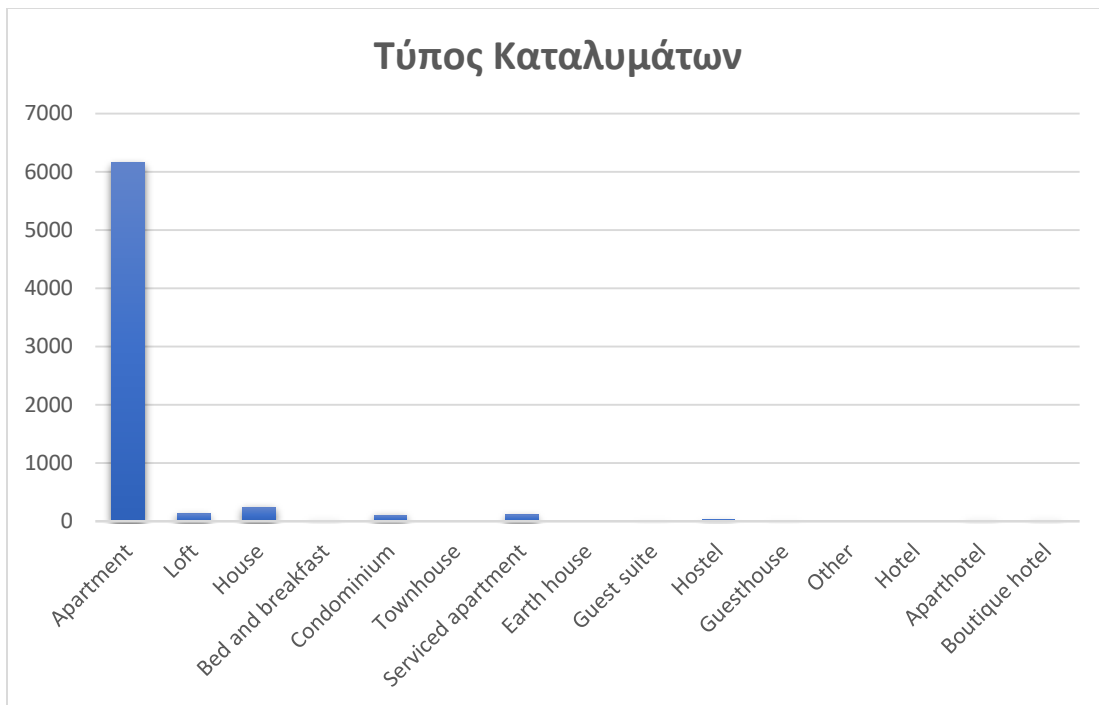


Figure 5: Ιστόγραμμα για τον τύπο καταλυμάτων

Μια σημαντική παράμετρος στην σχετική βιβλιογραφία για την τιμολόγηση ενός καταλύματος είναι το είδος και μέγεθος του χώρου που προσφέρεται για το οποίο έχουμε τις εξής τρεις κατηγορίες: ολόκληρο σπίτι/διαμέρισμα (entire home/apt), ιδιωτικό δωμάτιο (private room) και διαμοιρασμένο δωμάτιο (shared room). Στα καταλύματα της Αθήνας, έχουμε σχεδόν αποκλειστικά ολόκληρα σπίτια σαν καταλύματα κάτι που μας διευκολύνει στις υποθέσεις μας.

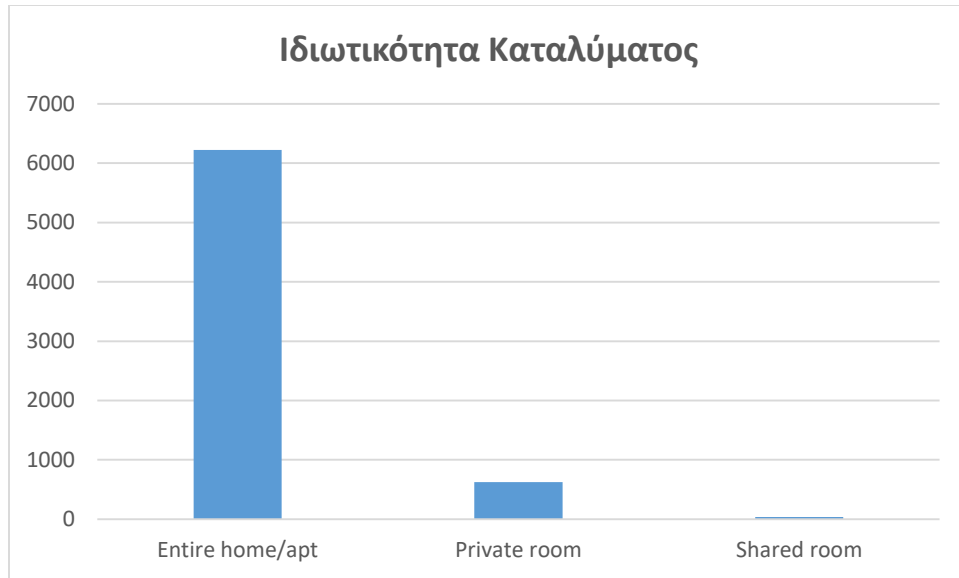


Figure 6: Ιστόγραμμα για την ιδιωτικότητα που προσφέρει το κατάλυμα

3.7 Τοποθεσία

Μια σημαντική παράμετρος για την εκτίμηση της τιμολόγησης ενός καταλύματος είναι η τοποθεσία του (Coenders, Espinet και Saez, 2003, Gibbs et al., 2017) και έχοντας υπόψη την ραγδαία ανάπτυξη των περιοχών πέριξ της Ακρόπολης στην Αθήνα, καθώς και τα περιγραφικά στατιστικά ανά γειτονιά (ακριβότερη μέση τιμή σε πιο κεντρικές γειτονιές) είναι ασφαλές να υποθέσουμε ότι η τοποθεσία θα είναι σημαντική παράμετρος και στο δικό μας μοντέλο. Προκειμένου ωστόσο να υπολογίσουμε την τοποθεσία και γνωρίζοντας την τοποθεσία των καταλυμάτων με ακρίβεια μέτρων ή δεκάδων μέτρων, θα πρέπει να αποφασίσουμε ποιο είναι το σημείο αναφοράς. Παίρνοντας υπόψη τα περιγραφικά στατιστικά, επιλέγουμε τον Ιερό Βράχο της Ακρόπολης και πιο συγκεκριμένα τον Παρθενώνα (γεωγραφικό μήκος: 37.972496, γεωγραφικό πλάτος: 23.725749). Στην επιλογή αυτή συνηγορεί και η σημασία του μνημείου για τους τουρίστες, καθώς και η αναμφισβήτητη θέση του Παρθενώνα ως το πιο σημαντικό αξιοθέατο της Αθήνας (Pretes, 2003).

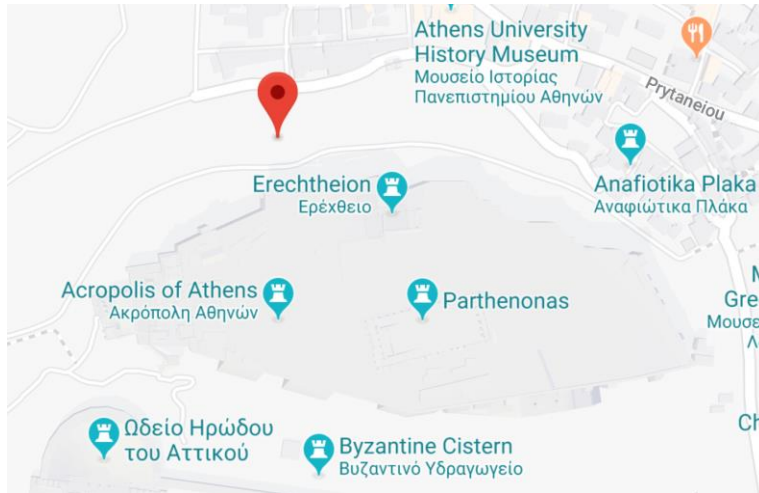


Image 28: Χάρτης τοποθεσίας του σημείου αναφοράς που επιλέχθηκε

Σε κάθε περίπτωση υπολογίζουμε την απόσταση στην ευθεία του χάρτη. Προκειμένου να υπολογιστεί η απόσταση μεταξύ δύο σημείων χρησιμοποιούμε την εξίσωση Haversine (Robusto, 1957) η οποία δίνεται από τον παρακάτω τύπο:

$$d = 2r \arcsin\left(\sqrt{\text{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) \text{hav}(\lambda_2 - \lambda_1)}\right)$$

$$= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

Equation 1: Εξίσωση haversine για τον υπολογισμό της απόστασης δύο σημείων

Η εξίσωση haversine εκτιμά την απόσταση του μεγάλου-κύκλου μεταξύ δύο σημείων σε μία σφαίρα, έχοντας τα γεωγραφικά μήκη και πλάτη των σημείων που μας ενδιαφέρουν. Η απόσταση αυτή είναι η ίδια με την απόσταση που βλέπει ένας χρήστης της πλατφόρμας όταν είναι στην διαδικασία επιλογής καταλύματος. Η υπόθεση εδώ είναι ότι όσο πιο κοντά είναι το κατάλυμα στο σημείο ενδιαφέροντος (εδώ ο Παρθενώνας), τόσο πιο επιθυμητό το κατάλυμα. Μετατρέπουμε την εξίσωση κατάλληλα ώστε να μπορεί να χρησιμοποιηθεί στο Excel και έχουμε υπόψη ότι όπου Earth Equatorial Radius είναι 6378.135 metres από το παγκόσμιο σύστημα γεωδαισίας WGS 84 (Unoosa.org, 2019). Η μέση απόσταση που βρίσκουμε από τον Παρθενώνα για το σύνολο των καταλυμάτων είναι 1.824 χιλιόμετρα.

4. Μεθοδολογία

4.1 Μεταβλητές

Για την κατασκευή του μοντέλου μας έχουμε καταρχήν 41 ερμηνευτικές μεταβλητές πολλές από τις οποίες διαισθητικά φαίνεται να μην είναι σημαντικές. Προκειμένου να μπορέσουμε να κατασκευάσουμε ένα ρεαλιστικό μοντέλο θα ήταν επιθυμητό να επιλέξουμε μόνο κάποιες από αυτές.

Τα τελευταία χρόνια έχουν γίνει πολλές προσπάθειες για την ανάπτυξη μοντέλων παλινδρόμησης που επιβάλλουν ποινές (penalized regression methods) για την ταυτόχρονη επιλογή μεταβλητών και την εκτίμηση των αντίστοιχων συντελεστών (coefficients) (Lockhart et al., 2014). Στην πράξη, ακόμα κι αν το δείγμα είναι μικρό, ένας μεγάλος αριθμός ερμηνευτικών μεταβλητών συνήθως συμπεριλαμβάνεται ώστε να περιοριστεί η μεροληψία του μοντέλου. Με έναν τόσο μεγάλο αριθμό ερμηνευτικών μεταβλητών, όπως στην περίπτωση μας, ενδέχεται να εμφανιστούν προβλήματα στις σχέσεις μεταξύ ερμηνευτικών μεταβλητών, εν προκειμένω προβλήματα πολύσυγγραμμικότητας (Lockhart et al., 2014). Έτσι, είναι επιθυμητό να επιλέξουμε έναν μικρότερο αριθμό ερμηνευτικών μεταβλητών που ωστόσο όχι μόνο έχει την καλύτερη προσαρμογή (fit) με το πλήρες εύρος των μεταβλητών αλλά περιέχει αυτές με την καλύτερη ικανότητα ερμηνείας (Lockhart et al., 2014). Έτσι οι ερευνητές έχουν οδηγηθεί στην δημιουργία διάφορων διακεκριμένων μοντέλων παλινδρόμησης ελαχίστων τετραγώνων (least squares – LS) με διάφορες τιμωρίες (penalties) ώστε να ανακαλύψουν τις πιο σχετικές ερμηνευτικές μεταβλητές και να πάρουν υψηλότερη ακρίβεια στις προβλέψεις τους στην γραμμική παλινδρόμηση (Lockhart et al., 2014).

Μία από αυτές τις μεθόδους είναι και το LASSO (Least Absolute Shrinkage and Selection Operator) που εισήχθη από τον Robert Tibshirani (1996), η οποία σύμφωνα με τον συγγραφέα, ελαχιστοποιεί το τετράγωνο των καταλοίπων (residual sum of squares) υπό τον περιορισμό ότι το άθροισμα της απόλυτης αξίας των συντελεστών είναι μικρότερο από μια σταθερή τιμή (Tibshirani, 1996). Οι κύριες διαφορές της παλινδρόμησης LASSO σε σχέση με την παλινδρόμηση OLS σε ένα πολυμεταβλητό υπόδειγμα εξηγούνται ως εξής: η πρώτη έχει να κάνει με την ακρίβεια της πρόβλεψης της εξαρτημένης μεταβλητής καθώς οι εκτιμήσεις OLS συνήθως έχουν μικρή μεροληψία αλλά μεγάλη διακύμανση και η ακρίβεια μπορεί να διορθωθεί μηδενίζοντας κάποιους συντελεστές. Κάνοντας κάτι τέτοιο, ο ερευνητής θυσιάζει κάποια μεροληψία για να μειώσει την διακύμανση και έτσι να πετύχει καλύτερη ακρίβεια πρόβλεψης του μοντέλου. Η δεύτερη έχει να κάνει με την ικανότητα ερμηνείας του μοντέλου καθώς όταν έχουμε έναν μεγάλο αριθμό ερμηνευτικών μεταβλητών θέλουμε να επιλέξουμε ένα μικρότερο σετ που επιδεικνύει την μεγαλύτερη επίδραση (Tibshirani, 1996).

Συνεπώς, οι αρχικές ερμηνευτικές μεταβλητές που επιλέγουμε είναι οι 41 παρακάτω. Κάποιες από αυτές προέρχονται από διαμόρφωση των αρχικών δεδομένων όπως για παράδειγμα η μεταβλητή 18, `host_is_superhost`, η οποία παίρνει τιμές t (true) ή f (false) οι οποίες

μετατράπηκαν αντίστοιχα σε 1 και 0 όπως και οποιαδήποτε κατηγορική μεταβλητή έχει δύο καταστάσεις.

Ερμηνευτικές Μεταβλητές Καταλύματος:

1. **accommodates:** μέγιστος αριθμός ατόμων που μπορούν να φιλοξενηθούν σε ένα κατάλυμα
2. **bathrooms:** αριθμός μπάνιων
3. **bedrooms:** αριθμός κρεβατοκάμαρων
4. **Distance** (Time_Walking, Time_Transit): οι τρεις αποστάσεις όπως έχουν αναφερθεί παραπάνω. Distance η απόσταση από τον Παρθενώνα σε χιλιόμετρα, Time_Walking: ο χρόνος περπατήματος από τον Παρθενώνα σε λεπτά και Time_Transit: ο χρόνος με τα μέσα μαζικής μεταφοράς από τον Παρθενώνα σε λεπτά.

Ερμηνευτικές Μεταβλητές Οικοδεσπότη:

5. **email:** ο οικοδεσπότης έχει επαληθεύσει το email του
6. **phone:** ο οικοδεσπότης έχει επαληθεύσει το τηλέφωνο του
7. **facebook:** ο οικοδεσπότης έχει επαληθεύσει τον λογαριασμό του στο Facebook
8. **government_id:** ο οικοδεσπότης έχει επαληθεύσει την ταυτότητα του
9. **jumio:** ο οικοδεσπότης έχει επαληθεύσει την ταυτότητα του ηλεκτρονικά
10. **offline_government_id:** ο οικοδεσπότης έχει επαληθεύσει την ταυτότητα του offline
11. **identity_manual:** ο οικοδεσπότης έχει επαληθεύσει την ταυτότητα του με διαφορετικό έγγραφο
12. **reviews:** ο οικοδεσπότης έχει δεχθεί συστατικές αξιολογήσεις από άλλους οικοδεσπότες της πλατφόρμας
13. **work_email:** ο οικοδεσπότης έχει επαληθεύσει το email εργασίας του
14. **kba:** ο οικοδεσπότης έχει επαληθεύσει την είσοδο του με χρήση Knowledge Based Identification
15. **selfie:** ο οικοδεσπότης έχει ανεβάσει selfie φωτογραφία για την επαλήθευση του
16. **google:** ο οικοδεσπότης έχει επαληθεύσει τον λογαριασμό Google Plus του
17. **host_has_profile_pic:** ο οικοδεσπότης έχει φωτογραφία προφίλ
18. **host_is_superhost:** ο οικοδεσπότης έχει το στάτους του superhost
19. **Days_since_host:** ημέρες που ο οικοδεσπότης είναι εγγεγραμμένος στην πλατφόρμα
20. **host_total_listings_count:** συνολικός αριθμός καταλυμάτων του οικοδεσπότη

Ερμηνευτικές Μεταβλητές Υποδομής Καταλύματος (Amenities)

21. **TV:** το κατάλυμα διαθέτει τηλεόραση
22. **Air conditioning:** το κατάλυμα διαθέτει Air-Condition
23. **Elevator:** το κατάλυμα διαθέτει ασανσερ
24. **Kitchen:** το κατάλυμα διαθέτει κουζίνα
25. **Family/kid friendly:** το κατάλυμα είναι φιλικό προς οικογένειες
26. **Parking:** το κατάλυμα διαθέτει πάρκινγκ

- 27. **Wifi:** το κατάλυμα διαθέτει WiFi
- 28. **Heating:** το κατάλυμα διαθέτει θέρμανση
- 29. **Cooking:** στο κατάλυμα επιτρέπεται οι φιλοξενούμενοι να μαγειρέψουν
- 30. **Patio or balcony:** το κατάλυμα διαθέτει αίθριο ή μπαλκόνι
- 31. **Washer:** το κατάλυμα διαθέτει πλυντήριο
- 32. **Smoking allowed:** στο κατάλυμα επιτρέπεται το κάπνισμα
- 33. **Pets Allowed:** στο κατάλυμα επιτρέπονται τα κατοικίδια
- 34. **Smoke detector:** το κατάλυμα διαθέτει ανιχνευτή καπνού
- 35. **Hot water:** το κατάλυμα διαθέτει ζεστό νερό
- 36. **Luggage dropoff allowed:** επιτρέπεται η απόθεση αποσκευών στο κατάλυμα
- 37. **24-hour check-in:** το κατάλυμα έχει την δυνατότητα εικοσιτετράωρου check-in
- 38. **pool:** το κατάλυμα διαθέτει πισίνα

Ερμηνευτικές Μεταβλητές Αξιολόγησης

- 39. **review_scores_rating:** το σκορ της συνολικής αξιολόγησης ενός καταλύματος (σε κλίμακα 100)
- 40. **number_of_reviews:** ο αριθμός αξιολογήσεων που έχει λάβει ένα κατάλυμα

Άλλες ερμηνευτικές μεταβλητές

- 41. **minimum_nights:** ο ελάχιστος αριθμός διανυκτερεύσεων που επιτρέπεται σε ένα κατάλυμα

Οι μεταβλητές που επιλέγονται έχουν κατά κύριο λόγο επισημανθεί από την βιβλιογραφία ως σημαντικές για διαφορετικά σεντ δεδομένων που έχουν ελεγχθεί. Ο αριθμός των κρεβατιών (beds) παρά το γεγονός ότι θεωρείται σημαντική μεταβλητή, δεν συμπεριλαμβάνεται καθώς εμφανίζει υψηλή συσχέτιση με τον αριθμό των κρεβατοκάμαρων.

Στην περίπτωση μας, θα χρησιμοποιήσουμε το lasso αρχικά για να επιλέξουμε ένα υπο-σεντ των πιο «σημαντικών» μεταβλητών και στην συνέχεια α) θα κάνουμε ξανά παλινδρόμηση lasso, για τις μεταβλητές που έχουμε επιλέξει β) θα κάνουμε παλινδρόμηση OLS. Θα χρησιμοποιήσουμε το πακέτο glmnet για την R. Το πακέτο glmnet, λύνει το ακόλουθο πρόβλημα:

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \left[(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1 \right]$$

Equation 2: Εξίσωση που λύνει το πακέτο glmnet για penalized παλινδρομήσεις

Εδώ $l(y, \eta)$ είναι η αρνητική συμβολή της συνάρτησης πιθανοφάνειας για την παρατήρηση i . Το πρόβλημα λύνεται για ένα εύρος των τιμών του λ , καλύπτοντας ολόκληρο το εύρος. Εδώ $\alpha=1$

που σημαίνει ότι επιλέγουμε την lasso παλινδρόμηση (για $\alpha=0$, το glmnet λύνει την παλινδρόμηση Ridge ενώ για οποιαδήποτε ενδιάμεση τιμή, έχουμε την elastic net παλινδρόμηση) και έχουμε

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \alpha [|\beta|_1]$$

Equation 3: Εξίσωση που λύνει το πακέτο glmnet για την lasso παλινδρόμηση

η οποία μπορεί να γραφτεί ως εξής:

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Equation 4: Εξίσωση που ελαχιστοποιεί η παλινδρόμηση lasso

Η παραπάνω εξίσωση είναι αυτή που ελαχιστοποιεί η παλινδρόμηση lasso. Ο πρώτος όρος είναι το άθροισμα των τετραγώνων υπό τον περιορισμό $\sum_{j=1}^p |\beta_j| \leq S$

Σε ότι αφορά την εξαρτημένη μεταβλητή, δημιουργούμε στην R, ένα διάνυσμα y με τις τιμές της τελικής τιμής (new price) η οποία υπολογίζεται ως εξής: η τιμή η οποία δίνεται στα δεδομένα για κάθε κατάλυμα συν το ηλικό της τιμής του κόστους καθαρισμού (cleaning fee) προς τον μέσο αριθμό των ημερών της διανυκτέρευσης. Επειδή, η μέση τιμή της διανυκτέρευσης δεν είναι κάτι που γνωρίζουμε από τα δεδομένα, το κόστος καθαρισμού είναι αρκετά σημαντικό για να μην το λάβουμε υπόψη (σε κάποιες περιπτώσεις έως το 30% της τελικής τιμής) και επειδή χρεώνεται μόνο μια φορά, είτε ο φιλοξενούμενος διαμείνει για μία ημέρα, είτε για δέκα, παίρνουμε ως μέσο αριθμό ημερών διανυκτέρευσης, μια προσέγγιση του αριθμού που δίνει η Airbnb για τα καταλύματα στο Σαν Φρανσίσκο, η οποία είναι 5.5 ημέρες (Airbnb, 2019). Από ότι φαίνεται, ενώ οι περισσότερες διανυκτερεύσεις είναι λιγότερες ημέρες, κάποιες μακροχρόνιες παραμονές ανεβάζουν τον μέσο όρο προς τα επάνω. Στην περίπτωση της Αθήνας χρησιμοποιούμε έναν μέσο αριθμό 4.5 ημερών, κατά μια ημέρα μικρότερο από την εκτίμηση της Airbnb για το Σαν Φρανσίσκο, για να προκύψει η τελική τιμή.

4.2 Αποτελέσματα

Χρησιμοποιώντας το πακέτο *glmnet*, αρχικά εφαρμόζουμε την παλινδρόμηση *lasso* κάνοντας cross-validation για τιμές του λ από 10.000 έως 0.1 για 50 επαναλήψεις (επιλέγουμε δυνάμεις του 10 ξεκινώντας από $\alpha_1 = 4$ και καταλήγοντας στο $\alpha_{50} = 0.01$). Από το μοντέλο που προκύπτει με την μεθοδολογία πλήρους διασταυρωμένης επικύρωσης (cross-validation), μπορούμε να επιλέξουμε το λ (λ_{1se}) που μας δίνει το μικρότερο σφάλμα εντός ενός standard error από το ελάχιστο λ (λ_{min}) το οποίο υπολογίζεται ως η τιμή εκείνη του λ που μας δίνει το ελάχιστο μέσο cross-validated σφάλμα. Η τιμή του α που έχουμε επιλέξει είναι 1, γιατί θέλουμε να χρησιμοποιήσουμε το *lasso* (για $\alpha=0$ θα «τρέχαμε» την παλινδρόμηση *ridge*). Το σφάλμα συγκεντρώνεται και το μέσο σφάλμα και η τυπική απόκλιση των επαναλήψεων υπολογίζεται. Θα προτιμήσουμε ωστόσο μεγαλύτερο λ από το λ_{1se} ώστε να «διώξουμε» περισσότερες μεταβλητές.

Σχετικά με το σφάλμα της παλινδρόμησης *lasso* θα πρέπει να σημειωθεί ότι στις περισσότερες εφαρμογές των *penalized παλινδρομήσεων* (*lasso*, *ridge*) είναι αδύνατο να εκτιμηθεί μια επαρκώς ακριβής εκτίμηση της μεροληψίας καθώς οι υπολογισμοί που βασίζονται σε *bootstrapping* δηλαδή οι έλεγχοι που βασίζονται σε τυχαία δειγματοληψία και αντικατάσταση μπορούν να δώσουν απλά μια εκτίμηση της διακύμανσης των εκτιμήσεων (Goeman, 2010). Αξιοπίστευες εκτιμήσεις της μεροληψίας είναι διαθέσιμες μόνο όταν αξιόπιστοι αμερόληπτοι εκτιμητές είναι διαθέσιμοι, το οποίο δεν ισχύει συνήθως στις *penalized παλινδρομήσεις* (Goeman, 2010). Συνεπώς, ενώ υπολογίζουμε τα μεγέθη R^2 και *Mean Estimate of standard error* (δηλαδή τον μέσο όρο του συνολικού σφάλματος για κάθε λ της σειράς που χρησιμοποιούμε για τις παλινδρομήσεις *lasso*, θα πρέπει να είμαστε προσεκτικοί στην ερμηνεία τους.

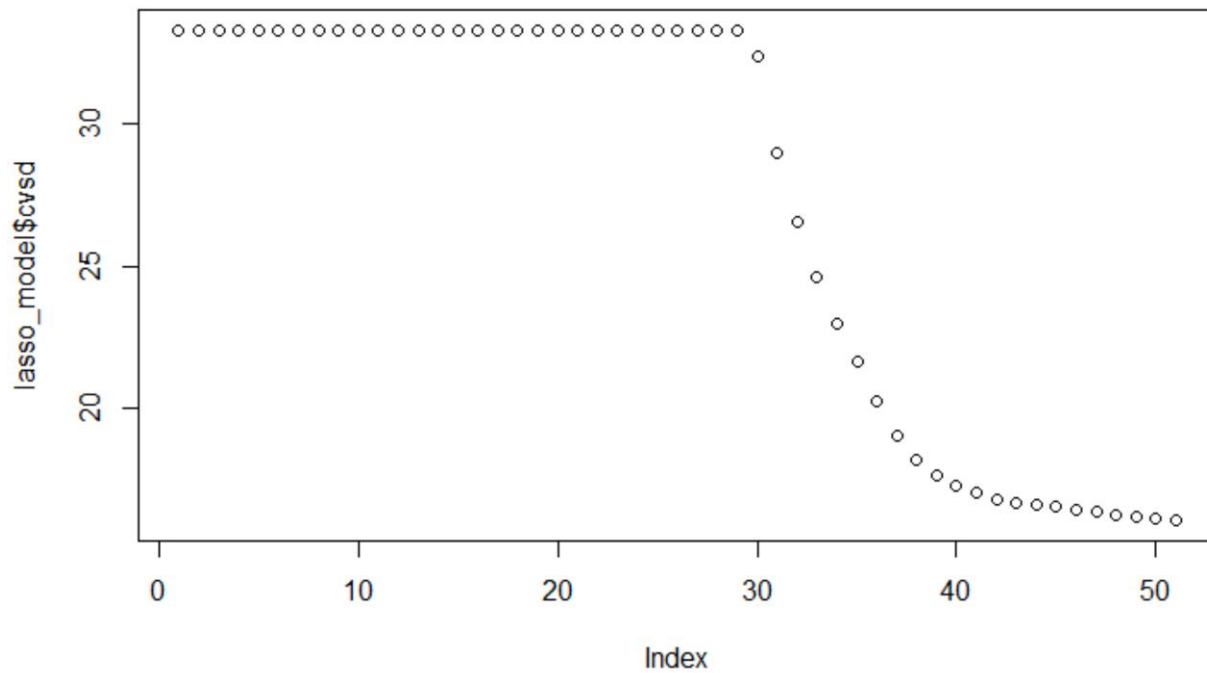


Figure 7: Διάγραμμα του εκτιμώμενου μέσου cross-validated σφάλματος για 50 τιμές του λ ($\lambda.length=50$)

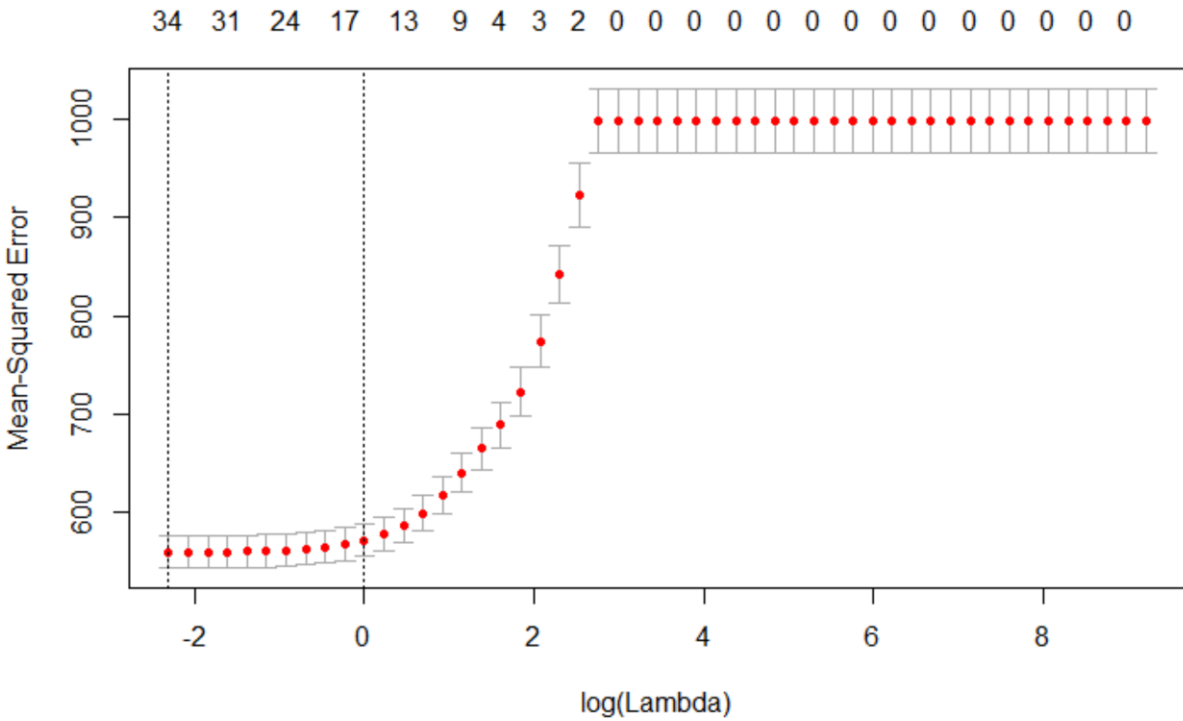


Figure 8: Διάγραμμα των τιμών του μέσου τετραγωνισμένου σφάλματος σε σχέση με τις τιμές του λ που δίνουμε

Οι διακεκομμένες γραμμές μας δίνουν τις δύο διαφορετικές τιμές για το λ . Η πρώτη, το *lambda.min* και η δεύτερη είναι το *lambda.lse*. Εμείς επιλέγουμε το δεύτερο καθώς αποτρέπει το overfitting της παλινδρόμησης. Τα διαστήματα εμπιστοσύνης αντιπροσωπεύουν τις εκτιμήσεις σφάλματος για το μέτρο του λ και υπολογίζονται με cross-validation. Οι αριθμοί στο πάνω μέρος δείχνουν πόσες ερμηνευτικές μεταβλητές παραμένουν στο μοντέλο. Παρατηρούμε ότι καθώς το λ αυξάνεται, ο αριθμός των ερμηνευτικών μεταβλητών μηδενίζεται και έτσι η γραμμή παλινδρόμησης είναι μια γραμμή της μορφής $Y = c \cdot X$. Έτσι, το μοντέλο απλοποιείται όσο αυξάνεται το λ και «τιμωρεί» τις πιο ασήμαντες μεταβλητές, αφαιρώντας αυτές, ωστόσο για μεγάλο αριθμό λ , δεν παραμένει ούτε μία ερμηνευτική μεταβλητή στο μοντέλο.

Δοκιμάζουμε την σειρά $\lambda = 10^i$ για $i = -1, -0.09, \dots, 4$ (με βήμα $j=0.1$) και επιλέγουμε ως βέλτιστο λ , το $\lambda = 10^{1.5} = 31.62$ και παίρνουμε τις παρακάτω εκτιμήσεις.

accommodates	bathrooms	bedrooms
3.6709406	18.85376551	3.681258421
email	phone	facebook

0	0	0
government_id	jumio	offline_government_id
0	0	0
identity_manual	reviews	work_email
0	0	0
kba	selfie	google
0	0	0
host_has_profile_pic	is_superhost	TV
0	0	3.729437624
Air.conditioning	Elevator	Kitchen
4.752341756	0	0
Family.kid.friendly	Parking	Wifi
0	0	0
Heating	Cooking	Patio.or.balcony
0	0	0
Washer	Smoking.allowed	Pets.Allowed
0	-2.17609324	0
Smoke.detector	Hot.water	Luggage.dropoff.allowed
3.244869822	0	0
X24.hour.check.in	pool	Distance
0	33.53340307	-6.503019394
review_scores_rating	minimum_nights	Days_since_host
0	0	0
host_total_listings_count		number_of_reviews
0		0

Table 3: Lasso Coefficients for $\lambda=1$ (Athens)

Αυτό μας αφήνει με **9 μη-μηδενικούς συντελεστές** συνεπώς το μοντέλο μας απλοποιείται. Στην συνέχεια θα δοκιμάσουμε δύο μεθόδους: α) να συνεχίσουμε την ανάλυση μας με παλινδρόμηση LASSO με τις μεταβλητές που «επιβίωσαν» (μη-μηδενικοί συντελεστές) και β) να συνεχίσουμε την ανάλυση μας με OLS.

accommodates	3.677603
Air.conditioning	3.551086
bathrooms	18.410304
bedrooms	3.310354
Distance	-5.820369
pool	27.172566

Smoke.detector	2.066279
Smoking.allowed	-1.578161
TV	2.767273
C	18.936869
R-squared	0.3993866
Mean Estimate of standard error	26.758

Table 4: Lasso "survivors", R-Squared and Estimate of standard error (Athens)

Στην πρώτη περίπτωση, στο μοντέλο που προκύπτει από την παλινδρόμηση *lasso* σε όσες μεταβλητές παρέμειναν, βλέπουμε ότι οι συντελεστές τους δεν απέχουν πολύ από την πρώτη παλινδρόμηση. Δυστυχώς, δεν μπορούμε να βγάλουμε πολλά περαιτέρω συμπεράσματα π.χ. παίρνοντας το R^2 καθώς όπως αναφέραμε, αξιόπιστες εκτιμήσεις της μεροληψίας είναι διαθέσιμες μόνο όταν αξιόπιστοι αμερόληπτοι εκτιμητές είναι διαθέσιμοι (Goeman, 2010).

accommodates	3.657206*** (16.60287)
Air.conditioning	7.132*** (6.816957)
bathrooms	21.02328*** (22.93896)
bedrooms	4.869805*** (8.731773)
Distance	-7.232125*** (-27.0113)
pool	49.88444*** (13.87649)
Smoke.detector	4.701728*** (6.844518)
Smoking.allowed	-4.416073*** (-6.231795)
TV	6.090611*** (7.749145)
C	10.33213 (6.903197)

R-squared	0.412354	Mean dependent var	54.26205
-----------	----------	--------------------	----------

Adjusted R-squared	0.411585	S.D. dependent var	31.60165
S.E. of regression	24.24106	Akaike info criterion	9.215425

Table 5: OLS for LASSO "survivors", t-stats are in parentheses, and by default * is $p < 0.05$, ** is $p < 0.01$ and *** is $p < 0.001$

Παρατηρούμε ότι και στο OLS μοντέλο, οι συντελεστές α) είναι όλοι στατιστικά σημαντικοί, κάτι που ίσως ήταν αναμενόμενο μετά την *lasso παλινδρόμηση* και β) είναι πολύ κοντά σε αυτούς που προέκυψαν τόσο από την αρχική *lasso* για την επιλογή των βασικών ερμηνευτικών μεταβλητών, τόσο και σε αυτές που αναμένονται από την βιβλιογραφία, ωστόσο λείπουν οι ερμηνευτικές μεταβλητές που έχουν να κάνουν με τον οικοδεσπότη.

Τα προβλήματα που προκύπτουν στο OLS μοντέλο έχουν να κάνουν με το σφάλμα εξειδίκευσης το οποίο αναμένουμε από την βιβλιογραφία (ωστόσο στο μοντέλο μας, κάνοντας τεστ RAMSAY με t-statistic 1.079133 απορρίπτουμε την ύπαρξη σφάλματος εξειδίκευσης με p-value 0.2806), τόσο για την εφαρμογή ηδονικών μοντέλων στην αγορά κατοικία (Butler, 1982; Chau and Chin, 2002) όσο και για την εφαρμογή ηδονικών μοντέλων για τα καταλύματα της Airbnb (Gibbs et al., 2017) όσο και ετεροσκεδαστικότητας, η οποία επίσης αναμένεται από την βιβλιογραφία (Gibbs et al., 2017). Πράγματι, κάνοντας το τεστ του White για την ύπαρξη ετεροσκεδαστικότητας προκύπτουν πολύ μεγάλες τιμές για το F-Statistic (F-statistic = 23.51601 για έλεγχο White συμπεριλαμβάνοντας τους cross-terms), έτσι επιλέγουμε να χρησιμοποιήσουμε την *παλινδρόμηση Robust Least Squares*.

accommodates	4.642744*** (28.40484)
Air.conditioning	6.677408*** (8.600772)
bathrooms	17.53849*** (25.78987)
bedrooms	2.015254*** (4.869726)
Distance	-5.518362*** (-27.77625)
pool	50.93833*** (19.09602)
Smoke.detector	3.626719*** (7.115134)
Smoking.allowed	-3.633168*** (-6.909495)
TV	4.444983*** (7.621612)
C	9.06826*** (8.165228)

Robust Statistics			
R-squared	0.263092	Adjusted R-squared	0.262127
Rw-squared	0.528961	Adjust Rw-squared	0.528961
Akaike info criterion	8599.316	Schwarz criterion	8671.245
Deviance	2303256	Scale	16.38154
Rn-squared statistic	7024.167	Prob(Rn-squared stat.)	0
Non-robust Statistics			
Mean dependent var	54.26205	S.D. dependent var	31.60165
S.E. of regression	24.77376	Sum squared resid	4218228

Table 6: Robust Least Squares Regression with Lasso survivors, *t*-stats are in parentheses, and by default * is $p < 0.05$, ** is $p < 0.01$ and *** is $p < 0.001$

4.3 Εκτιμώντας την απόσταση ως ερμηνευτική μεταβλητή

Παραπάνω, ως μέτρο της απόστασης (Distance) χρησιμοποιούμε την απόσταση σε ευθεία γραμμή μεταξύ ενός καταλύματος και ενός σημείου ενδιαφέροντος, στην δική μας περίπτωση τον Παρθενώνα. Προκειμένου να ερευνήσουμε περισσότερο την σημασία της απόστασης, χρησιμοποιούμε δύο ακόμα μέτρα αυτής: τον χρόνο περπατήματος (walking time) από το κατάλυμα στον Παρθενώνα και τον χρόνο μετακίνησης με συνδυασμό περπατήματος και μέσων μαζικής μεταφοράς (transit time). από το κατάλυμα στον Παρθενώνα. Οι δύο αυτοί χρόνοι δεν είναι άμεσα διαθέσιμοι, οπότε συνδεόμαστε στο Distance Matrix API του Google Cloud και παίρνουμε τις πληροφορίες που χρειαζόμαστε δίνοντας για κάθε κατάλυμα τις γεωγραφικές του συντεταγμένες και τις γεωγραφικές συντεταγμένες του Παρθενώνα.

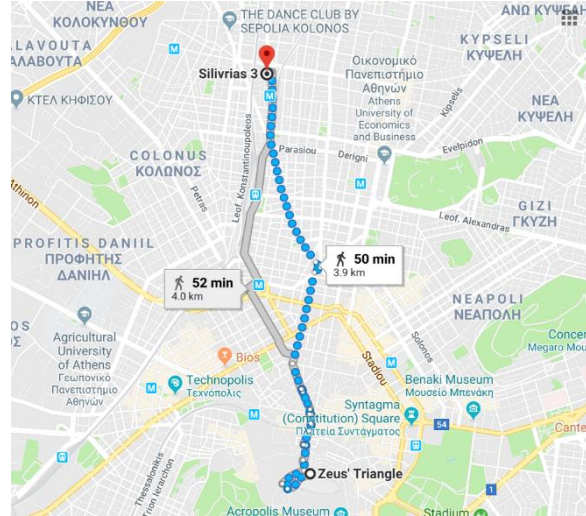
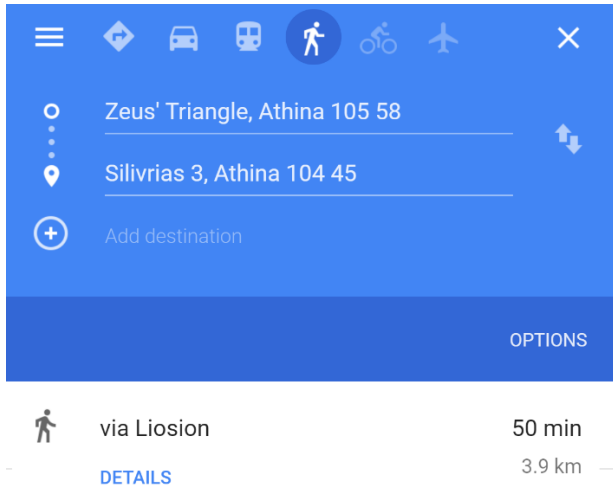


Image 29: Google Maps showing minimum walking distance from a random listing to the Parthenon

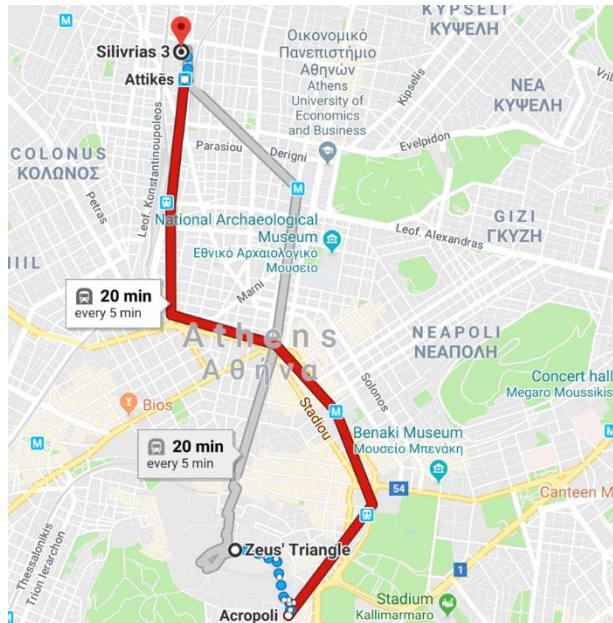
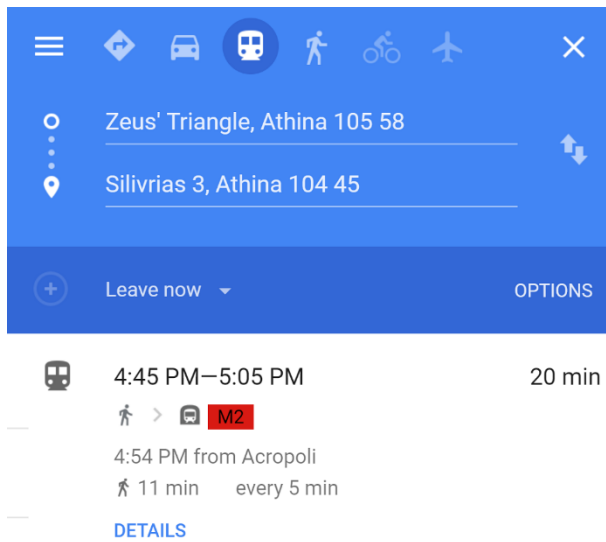


Image 30 Google Maps showing minimum transit distance from the same random listing to the Parthenon

Θα πρέπει να επισημανθεί ότι ενώ ο χρόνος περπατήματος (walking time) είναι κατά πάσα πιθανότητα σταθερός, εκτός από την διεξαγωγή εργασιών και το κλείσιμο κάποιων δρόμων, ο χρόνος μετάβασης με την δημόσια συγκοινωνία (transit time) δεν είναι τόσο σταθερός και διαφέρει ανάλογα με την ώρα της ημέρας που γίνεται η κλήση του αποτελέσματος ενώ μπορεί να επηρεαστεί από απεργίες, αργίες κοκ. Στην δική μας περίπτωση η κλήση των αποτελεσμάτων έγινε το πρωί της Πέμπτης 13 Φεβρουαρίου 2019, ωστόσο προκειμένου να έχουμε πιο ακριβή εκτίμηση του χρόνου transit θα ήταν επιθυμητή η κλήση αποτελεσμάτων διαφορετικές ημέρες

της εβδομάδας και διαφορετικές ώρες καθώς και η κλήση πολλών τιμών χρόνου μετάβασης για το ίδιο κατάλυμα.

Συμπεραίνουμε επιπλέον ότι ο χρόνος περπατήματος έχει εξαιρετικά υψηλή συσχέτιση (0.987) με την απόσταση (Distance) κάτι που είναι αναμενόμενο και διαψεύδει τις αρχικές μας εκτιμήσεις ότι παίρνοντας σημεία περιμετρικά του Παρθενώνα, η απόσταση από τον χάρτη δεν συμφωνεί με τον χρόνο περπατήματος λόγω περιέργου ανάγλυφου του κέντρου της πόλης κάτι που οι τουρίστες, ως προσωρινοί επισκέπτες, δεν μπορούν να γνωρίζουν. Έτσι, προκειμένου τα δύο μεγέθη να είναι συγκρίσιμα, χρησιμοποιούμε μόνο τον χρόνο περπατήματος και τον χρόνο transit στις νέες παλινδρομήσεις ώστε να ερευνήσουμε τι ρόλο παίζουν τα μέσα μαζικής μεταφοράς στην επιλογή καταλύματος.

Μετατρέπουμε τους χρόνους αυτούς σε λεπτά, καθώς το Google Cloud μας τους δίνει σε δευτερόλεπτα και εκτιμώντας την σημασία της απόστασης από τον Παρθενώνα με παλινδρόμηση Robust Least Squares έχουμε τα παρακάτω αποτελέσματα

Variables	Walking Time	Transit Time
accommodates	4.610897*** (28.53512)	4.826473*** (29.04444)
Air.conditioning	6.642961*** (8.650843)	6.714741*** (8.499663)
bathrooms	17.13866*** (0.672806)	17.59968*** (25.4353)
bedrooms	2.0841 (5.093552)	1.65061*** (3.924199)
Smoke.detector	3.61048 (7.162466)	4.171005*** (8.063573)
Smoking.allowed	-3.490986*** (-6.710173)	-3.617636*** (-6.756742)
TV	4.328316*** (7.506078)	3.960515*** (6.678178)
walking_minutes/ transit_minutes	-0.421693*** (-30.07475)	-0.780776*** (-24.26313)
pool	52.63587*** (19.94427)	51.88456*** (19.1123)
C	13.20141*** (11.57505)	16.73178*** (12.7265)

Table 7: Robust Least Squares Regression with Lasso survivors for both transit time and walking time (both times in minutes), t-stats are in parentheses, and by default * is $p < 0.05$, ** is $p < 0.01$ and *** is $p < 0.001$

Robust Statistics			
R-squared	0.269254	Adjusted R-squared	0.268298
Rw-squared	0.53715	Adjust Rw-squared	0.53715
Akaike info criterion	8693.569	Schwarz criterion	8765.354
Deviance	2248479	Scale	16.09756
Rn-squared statistic	7235.122	Prob(Rn-squared stat.)	0
Non-robust Statistics			
Mean dependent var	54.26205	S.D. dependent var	31.60165
S.E. of regression	24.63342	Sum squared resid	4170573

Table 8: Robust and non-robust statistics for Robust Least Squares Regression with walking time

Robust Statistics			
R-squared	0.24986	Adjusted R-squared	0.248878
Rw-squared	0.516586	Adjust Rw-squared	0.516586
Akaike info criterion	8627.972	Schwarz criterion	8699.772
Deviance	2371952	Scale	16.59648
Rn-squared statistic	6666.785	Prob(Rn-squared stat.)	0
Non-robust Statistics			

Mean dependent var	54.26205	S.D. dependent var	31.60165
S.E. of regression	25.12405	Sum squared resid	4338360

Table 9: Robust and non-robust statistics for Robust Least Squares Regression with transit time

4.4 Έλεγχοι στιβαρότητας μοντέλου

Προκειμένου να ελέγξουμε κατά πόσο τα συμπεράσματα και οι συντελεστές που προκύπτουν από τα μοντέλα που χρησιμοποιούμε, έχουν λογικές αποκρίσεις, χρησιμοποιούμε ένα αντίστοιχο σετ δεδομένων για τα καταλύματα της Airbnb στην Ρώμη. Επιλέγουμε την Ρώμη γιατί υποθέτουμε ομοιότητες με την πόλη της Αθήνας στις αφίξεις τουριστών. Στην περίπτωση της Ρώμης τα αρχικά καταλύματα είναι περίπου τα τριπλάσια (29.436), ενώ το προφίλ των καταλυμάτων διαφέρει καθώς το 64% είναι ολόκληρα διαμερίσματα ενώ το υπόλοιπο είναι ιδιωτικά δωμάτια ή διαμοιρασμένα δωμάτια. Επιπλέον, η μέση τιμή στα καταλύματα της Ρώμης είναι 98€/διανυκτέρευση από 67€/ διανυκτέρευση για την Αθήνα. Όπως και στην περίπτωση της Αθήνας απομακρύνουμε τα καταλύματα με κενά στοιχεία στις μεταβλητές που μας ενδιαφέρουν καθώς και όσα έχουν μηδενικές αξιολογήσεις. Όπως και στην Αθήνα, προσθέτουμε στην τιμή κάθε καταλύματος το κόστος καθαρισμού διαιρεμένο με 4.5 ημέρες, το μέτρο της μέσης διαμονής την οποία υποθέσαμε βασισμένοι στα στοιχεία που δίνει η εταιρεία.

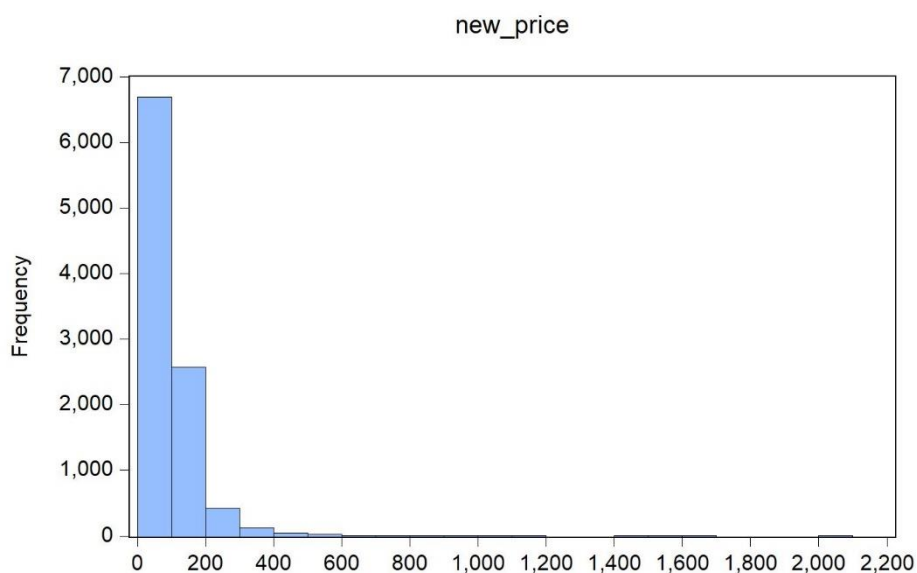


Figure 9: ιστόγραμμα της τιμής ανά διανυκτέρευση για το σύνολο των καταλυμάτων που απομένουν μετά το φιλτράρισμα στην Ρώμη

accommodates	bathrooms	bedrooms
7.264223835	9.336654325	8.446454313
email	phone	facebook
0	0	0
government_id	jumio	offline_government_id
0	0	0
identity_manual	work_email	kba
5.26055109	0	0
selfie	google	host_has_profile_pic
0	0	0
host_is_superhost	TV	Air.conditioning
0	8.355067591	12.6827147
Elevator	Kitchen	Family.kid.friendly
-2.535490077	2.962114599	0
Parking	Wifi	Heating
0	0	0
Cooking	Patio.or.balcony	Washer
0	0	0
Smoking.allowed	Pets.Allowed	Smoke.detector
-4.191148618	0	2.352290376

Hot.water	Luggage.dropoff.allowed	X24.hour.check.in
0	0	0
pool	distance	review_scores_rating
8.991767772	-3.036181971	0
minimum_nights	Days_since_host	host_total_listings_count
0	0	0
number_of_reviews		
0		

Table 10: Lasso Coefficients for $\lambda=1$ (Rome)

Από όπου προκύπτει ο παρακάτω πίνακας με τις **13 μεταβλητές** που «επιζούν» από την παλινδρόμηση lasso.

accommodates	7.264223835
bathrooms	9.336654325
bedrooms	8.446454313
identity_manual	5.26055109
TV	8.355067591
Air.conditioning	12.6827147
Elevator	-2.535490077
Kitchen	2.962114599
Smoking.allowed	-4.191148618
Smoke.detector	2.352290376
pool	8.991767772
distance	-3.036181971
R-Squared	0.4618618
Mean Estimate of standard error	41.47642

Table 11: Lasso "survivors", R-Squared and Mean Estimate of standard error (Rome)

Με αυτές τις μεταβλητές εκτιμούμε την *Robust Least Squares* παλινδρόμηση από την οποία παίρνουμε τα παρακάτω αποτελέσματα

accommodates	7.376577*** (29.09219)
--------------	---------------------------

bathrooms	11.77548*** (16.66701)
bedrooms	9.24284*** (13.52998)
identity_manual	4.676978*** (5.345228)
TV	9.213731 (9.321949)
Air.conditioning	11.53965*** (15.30228)
Elevator	-5.430716*** (-8.677805)
Kitchen	2.934892*** (3.536368)
Smoking.allowed	-5.503709*** (-6.708785)
Smoke.detector	2.050667*** (2.773155)
pool	24.46533*** (7.463501)
distance	-3.058353*** (-32.9273)
C	21.17192*** (15.97365)

Table 12: Robust Least Squares Regression with Lasso survivors for both transit time and walking time (both times in minutes), t-stats are in parentheses, and by default * is $p < 0.05$, ** is $p < 0.01$ and *** is $p < 0.001$

Robust Statistics			
R-squared	0.31051	Adjusted R-squared	0.309623
Rw-squared	0.568964	Adjust Rw-squared	0.568964
Akaike info criterion	11401.07	Schwarz criterion	11498.02
Deviance	8070973	Scale	26.63223
Rn-squared statistic	10764.15	Prob(Rn-squared stat.)	0
Non-robust Statistics			
Mean dependent var	89.37988	S.D. dependent var	50.12502

S.E. of regression	38.12145	Sum squared resid	13550056
--------------------	----------	-------------------	----------

Table 13: Robust and non-robust statistics for Robust Least Squares Regression for Rome's listings

5. Συμπεράσματα/Συζήτηση

Στην ερμηνεία των αποτελεσμάτων, δεν υπήρξε σε καμία περίπτωση κάποιο πρόσημο που ήταν αντίθετο από αυτό που θα περίμενε κάποιος με βάση τις υποθέσεις που κάνουμε για την σημασία που έχουν οι ερμηνευτικές μεταβλητές για τους τουρίστες. Το αξιοσημείωτο όμως είναι στις μεταβλητές που παραμένουν.

Αρχικά, κάτι που δεν ήταν αναμενόμενο είναι η απουσία των χαρακτηριστικών του οικοδεσπότη (*Ερμηνευτικές Μεταβλητές Οικοδεσπότη*) από τις στατιστικά σημαντικές μεταβλητές, ένα εύρημα που δεν συμβαδίζει με μεγάλο μέρος της βιβλιογραφίας. Θυμόμαστε ότι στις λέξεις του wordcloud, δηλαδή στις λέξεις που εμφανίζονται με την μεγαλύτερη συχνότητα στα σχόλια των επισκεπτών βρίσκονται τα ονόματα “George” και “Maria”, τα οποία υποδεικνύουν σε ένα διαισθητικό επίπεδο την σημασία του οικοδεσπότη. Μια πιθανή ερμηνεία για αυτή την απουσία είναι ότι δεν εξετάστηκαν επιπλέον ερμηνευτικές μεταβλητές που έχουν να κάνουν με τον οικοδεσπότη όπως για παράδειγμα ο αριθμός των φωτογραφιών που παρέχουν οι οικοδεσπότες που μελετάται από τους Ert, Fleischer και Magen (2016) ή το μέγεθος και οι θεματικές που χρησιμοποιεί ο κάθε οικοδεσπότης για να περιγράψει τον εαυτό του που μελετάνε οι Ma et al. (2017). Επιπλέον, ενδέχεται η επίδραση του οικοδεσπότη να επηρεάζει έμμεσα την τιμή μέσω των αξιολογήσεων, στην δική μας περίπτωση ωστόσο οι σχετικές με τις αξιολογήσεις μεταβλητές «κόπηκαν» από την παλινδρόμηση *lasso*.

Έπειτα, σε ότι αφορά την δυναμικότητα σε επισκέπτες και το μέγεθος του καταλύματος, όλες οι αρχικές μεταβλητές συμπεριλαμβάνονται (accommodates, bedrooms, bathrooms) εκτός από την μεταβλητή beds, δηλαδή τον αριθμό των κρεβατιών την οποία επιλέξαμε να μην συμπεριλάβουμε ούτε στην αρχική *lasso παλινδρόμηση* λόγω μεγάλης συσχέτισης (0.793) με την μεταβλητή accommodates, δηλαδή τον μέγιστο αριθμό φιλοξενούμενων. Η ύπαρξη ενός επιπλέον μπάνιου, αυξάνει την τιμή ενός καταλύματος κατά περίπου 17.5€, ωστόσο υποθέτουμε ότι αυτό έχει να κάνει με το μέγεθος του καταλύματος καθώς πάνω από ένα μπάνιο μάλλον σημαίνει αρκετά μεγάλο και πιθανώς διώροφο κατάλυμα.

Επιπλέον, από τις μεταβλητές που έχουν να κάνουν με την υποδομή του καταλύματος, παραμένουν και είναι στατιστικά σημαντικές η ύπαρξη τηλεόρασης, air-condition, πισίνας και ανιχνευτή καπνού. Η ύπαρξη πισίνας, όπως είναι λογικό, αυξάνει την τιμή ενός καταλύματος κατά περίπου 51€. Αντίθετα, τα καταλύματα στα οποία το κάπνισμα επιτρέπεται «τιμωρούνται» για αυτό με πτώση της τιμής τους κατά 3.6€.

Τέλος, αναμενόμενη είναι η απουσία των ερμηνευτικών μεταβλητών που έχουν να κάνουν με τις αξιολογήσεις *review_scores_rating* και *number_of_reviews*, καθώς όπως αναφέραμε, με το μικρό εύρος τιμών που παίρνουν και την πιθανότητα ενδογένειας, δεν προσφέρουν ιδιαίτερη πληροφορία.

Σε ότι αφορά την απόσταση, μια ερμηνευτική μεταβλητή στην οποία δώσαμε ιδιαίτερο βάρος, βλέπουμε ότι για κάθε λεπτό απομάκρυνσης περπατήματος από το κέντρο, η «τιμωρία» στην τιμή είναι περίπου 0.42€ ενώ για κάθε λεπτό transit απομάκρυνσης η «τιμωρία» είναι 0.78€. Η διαφορά αυτή, είναι περίπου 46% και είναι ενδεχομένως ένα από τα πλέον ενδιαφέροντα ευρήματα αυτής της εργασίας, καθώς για τον ίδιο χρόνο η απομάκρυνση από την δημόσια συγκοινωνία τιμωρείται 50% παραπάνω σε σχέση με την απομάκρυνση από το καθαυτό σημείο ενδιαφέροντος.

Στα αποτελέσματα της Ρώμης, παρατηρούμε ότι σε γενικές γραμμές δεν υπάρχουν εκπλήξεις, παρά το γεγονός ότι στο μοντέλο εισέρχονται περισσότερες μεταβλητές από αυτό της Αθήνας. Ενδιαφέρον παρουσιάζουν οι επιδράσεις των νέων μεταβλητών, της ύπαρξης ασανσέρ (Elevator) το οποίο «τιμωρείται» κάτι που είναι λογικό καθώς φαίνεται ότι τα καταλύματα που βρίσκονται σε όροφο δεν προτιμούνται. Η ύπαρξη κουζίνας επίσης επηρεάζει θετικά την τιμή όπως και η ύπαρξη ανιχνευτή καπνού και πισίνας, ενώ το κάπνισμα επηρεάζει αρνητικά. Στις «εκπλήξεις» της παλινδρόμησης, η ύπαρξη μιας ερμηνευτικής μεταβλητής που σχετίζεται με τον οικοδεσπότη, της επαλήθευσης της ταυτότητας του οικοδεσπότη, η οποία δεν «επιζεί» στο μοντέλο της Αθήνας.

5.1 Μελλοντική έρευνα

Η Airbnb αποτελεί μια σημαντική πηγή εσόδων για έναν μεγάλο αριθμό ιδιοκτητών στην Αθήνα αλλά και σε πολλές άλλες πόλεις και τουριστικές περιοχές της Ελλάδας. Στον αντίποδα, η ραγδαία ανάπτυξη του φαινομένου της καταχώρησης τουριστικών καταλυμάτων σε πλατφόρμες ενοικίασης η οποία συμβαίνει ταυτόχρονα και ίσως εξαιτίας της ραγδαίας αύξησης των εισερχόμενων τουριστών δημιουργεί αναταράξεις στις κατά τόπους γειτονιές και κοινότητες όπου το φαινόμενο είναι πιο έντονο. Σημαντικά βήματα και περαιτέρω έρευνα πρέπει να γίνουν για να κατανοηθεί περισσότερο το φαινόμενο μέσα από δεδομένα προκειμένου οι αποφάσεις των αρχών για την ρύθμιση του φαινομένου να είναι κατά το δυνατόν προς την βέλτιστη κατεύθυνση. Πεδία έρευνας θα μπορούσαν να είναι οι διαφορές μεταξύ των γειτονιών στην Αθήνα και σε άλλες πόλεις και η επίδραση που είχε η Airbnb στα ενοίκια και στην πιθανή αύξηση τους καθώς και στον ξενοδοχειακό κλάδο.

6. Αναφορές

- Airbnb Press Room. (2019). *About Us - Airbnb Press Room*. [online] Available at: <https://press.airbnb.com/about-us/> [Accessed 17 Feb. 2019].
- Airbnb. (2019). How do I book a place on Airbnb?. [online] Available at: <https://www.airbnb.com/help/article/380/how-do-i-book-a-place-on-airbnb> [Accessed 20 Feb. 2019].
- Airbnb. (2019). How do I book a place on Airbnb?. [online] Available at: <https://www.airbnb.com/help/article/380/how-do-i-book-a-place-on-airbnb> [Accessed 20 Feb. 2019].
- Airbnb. (2019). *What are Airbnb's basic requirements for hosts?*. [online] Available at: <https://www.airbnb.com/help/article/576/what-are-airbnb-s-basic-requirements-for-hosts> [Accessed 17 Feb. 2019].
- Airbnb. (2019). *What is response rate and how is it calculated?*. [online] Available at: <https://www.airbnb.com/help/article/430/what-is-response-rate-and-how-is-it-calculated> [Accessed 17 Feb. 2019].
- AirDNA - Airbnb Data and Analytics. (2019). AirDNA MarketMinder. [online] Available at: <https://www.airdna.co/vacation-rental-data/app/gr/default/athens/overview> [Accessed 21 Feb. 2019].
- Andersson, D. (2000). Hypothesis testing in hedonic price estimation - On the selection of independent variables. *The Annals of Regional Science*, 34(2), pp.293-304.
- Belarmino, A., Whalen, E., Koh, Y. and Bowen, J. (2017). Comparing guests' key attributes of peer-to-peer accommodations and hotels: mixed-methods approach. *Current Issues in Tourism*, 22(1), pp.1-7.
- Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8), pp.1595-1600.
- Biber, E., Light, S., Ruhl, J. and Salzman, J. (2017). Regulating Business Innovation as Policy Disruption: From the Model T to Airbnb. *Vanderbilt Law Research Paper No. 17-24; UCLA School of Law, Public Law Research Paper No. 17-18*.
- Business Insider. (2019). *Airbnb made \$93 million in profit on \$2.6 billion in revenue, but an internal clash sent the CFO out the door*. [online] Available at: <https://www.businessinsider.com/airbnb-profit-revenue-2018-2> [Accessed 17 Feb. 2019].
- Butler, R. (1982). The Specification of Hedonic Indexes for Urban Housing. *Land Economics*, 58(1), p.96.
- Champ, P., Boyle, K. and Brown, T. (2003). *A Primer on Nonmarket Valuation*. Dordrecht: Springer Netherlands.

- Chau, K. and Chin, T. (2002). A Critical Review of Literature on the Hedonic Price Model. *International Journal for Housing Science and Its Applications*, 27(2), pp.145-165.
- Chen, Y. and Xie, K. (2017). Consumer valuation of Airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management*, 29(9), pp.2405-2424.
- Coenders, G., Espinet, J. and Saez, M. (2003). Predicting random level and seasonality of hotel prices: A latent growth curve. *Tourism Analysis*, 8(1), pp.15-31.
- Coldwell, W. (2019). *First Venice and Barcelona: now anti-tourism marches spread across Europe*. [online] the Guardian. Available at: <https://www.theguardian.com/travel/2017/aug/10/anti-tourism-marches-spread-across-europe-venice-barcelona> [Accessed 24 Feb. 2019].
- Court, A. (1939). The dynamics of automobile demand. [New York]: [General Motors Corporation].
- Edelman, B. and Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *SSRN Electronic Journal*.
- Ert, E., Fleischer, A. and Magen, N. (2015). Trust and Reputation in the Sharing Economy: The Role of Personal Photos on Airbnb. *SSRN Electronic Journal*.
- Gibbs, C., Guttentag, D., Gretzel, U., Morton, J. and Goodwill, A. (2017). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 35(1), pp.46-56.
- Gillard, Q. (1981). The effect of environmental amenities on house values: The example of a view lot. *The Professional Geographer*, 33(2), pp.216-220.
- Goeman, J. J., L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal* 52(1), 70–84
- Google Maps. (2019). Google Maps. [online] Available at: <https://www.google.com/maps> [Accessed 20 Feb. 2019].
- Google Maps. (2019). Google Maps. [online] Available at: <https://www.google.com/maps> [Accessed 20 Feb. 2019].
- Gravari-Barbas, M. and Guinand, S. (2017). *Tourism & gentrification in contemporary metropolises*. Routledge.
- Ikkala, T. and Lampinen, A. (2014). Defining the price of hospitality. *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW Companion '14*.
- Investopedia. (2019). Peer-to-Peer (P2P) Economy. [online] Available at: <https://www.investopedia.com/terms/p/peertopeer-p2p-economy.asp> [Accessed 21 Feb. 2019].

- Ioannides, D., Röslmaier, M. and van der Zee, E. (2018). Airbnb as an instigator of 'tourism bubble' expansion in Utrecht's Lombok neighbourhood. *Tourism Geographies*, pp.1-19.
- Kask, S. and Maani, S. (1992). Uncertainty, Information, and Hedonic Pricing. *Land Economics*, 68(2), p.170.
- Krinis, N. (2019). *Athens Expects Record 5.5 Million Tourists in 2018*. [online] GTP Headlines. Available at: <https://news.gtp.gr/2018/10/08/athens-expects-record-5-5-million-tourists-2018/> [Accessed 27 Feb. 2019].
- Lancaster, K. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), pp.132-157.
- Lee, D. (2016). How Airbnb short-term rentals exacerbate Los Angeles's affordable housing crisis: Analysis and policy recommendations. *Harvard Law and Policy Review*, 10, p.229.
- Lee, D., Hyun, W., Ryu, J., Lee, W., Rhee, W. and Suh, B. (2015). An Analysis of Social Features Associated with Room Sales of Airbnb. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing - CSCW'15 Companion*.
- Li, R. and Li, H. (2018). Have Housing Prices Gone with the Smelly Wind? Big Data Analysis on Landfill in Hong Kong. *Sustainability*, 10(2), p.341.
- Liang, S., Schuckert, M., Law, R. and Chen, C. (2017). Be a "Superhost": The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management*, 60, pp.454-465.
- Lockhart, R., Taylor, J., Tibshirani, R. and Tibshirani, R. (2014). A significance test for the lasso. *The Annals of Statistics*, 42(2), pp.413-468.
- Ma, X., Hancock, J., Lim Mingjie, K. and Naaman, M. (2017). Self-Disclosure and Perceived Trustworthiness of Airbnb Host Profiles. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*.
- Nieuwland, S. and van Melik, R. (2018). Regulating Airbnb: how cities deal with perceived negative externalities of short-term rentals. *Current Issues in Tourism*, pp.1-15.
- Pretes, M. (2003). Tourism and nationalism. *Annals of Tourism Research*, 30(1), pp.125-142.
- Quattrone, G., Proserpio, D., Quercia, D., Capra, L. and Musolesi, M. (2016). Who Benefits from the "Sharing" Economy of Airbnb?. *Proceedings of the 25th International Conference on World Wide Web - WWW '16*.
- Robusto, C. (1957). The Cosine-Haversine Formula. *The American Mathematical Monthly*, 64(1), p.38.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), pp.34-55.

South EU Summit. (2019). *Travellers love touring Europe with Airbnb, While Europeans Residents Loathe It*. [online] Available at: <https://www.southeusummit.com/europe/everyone-loves-airbnb-in-europe-except-european-residents/> [Accessed 17 Feb. 2019].

Stors, N. and Kagermeier, A. (2015). Motives for Using Airbnb in Metropolitan Tourism—Why do People Sleep in the Bed of a Stranger?. *Regions Magazine*, 299(1), pp.17-19.

TechCrunch. (2019). *Airbnb closes \$1B round at \$31B valuation, profitable as of 2H 2016, no plans for IPO*. [online] Available at: https://techcrunch.com/2017/03/09/airbnb-closes-1b-round-at-31b-valuation-profitable-as-of-2h-2016-no-plans-for-ipo/?_ga=2.140746043.889344879.1550392910-1757057166.1550392910 [Accessed 17 Feb. 2019].

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), pp.267-288.

United World Tourism Organization. (2019). *UNWTO Highlights Confirm Another Record Year in 2017 / World Tourism Organization UNWTO*. [online] Available at: <http://www2.unwto.org/press-release/2018-08-27/unwto-highlights-confirm-another-record-year-2017> [Accessed 24 Feb. 2019].

Unoosa.org. (2019). World Geodetic System 1984. [online] Available at: http://www.unoosa.org/pdf/icg/2012/template/WGS_84.pdf [Accessed 25 Feb. 2019].

Zervas, G., Proserpio, D. and Byers, J. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, 54(5), pp.687-705.

Επιπλέον, κομμάτι του κώδικα R που χρησιμοποιήθηκε και μορφοποιήθηκε κατάλληλα προέρχεται από τους:

Peshi, Gupta και Agrawal (2018) "Exploratory Data Analysis and Visualization of Airbnb Dataset" source code (Version 1.0) [Source code].
https://github.com/saranggupta94/airbnb/blob/master/code/Airbnb_final_knit.rmd

7. Appendix

R Code

```
x <- model.matrix(lasso_x~.,trainset)
x <- model.matrix(lasso_x~.)
x <- model.matrix(lasso_x~)
x <- model.matrix(lasso_x)
lasso_x
view(lasso_x)
lasso_y = read.csv("final_regression_y.csv")
lasso_y = read.csv("y.csv")
lasso_x = read.csv("x.csv")
y = data.matrix(lasso_y, rownames.force = NA)
x = data.matrix(lasso_x, rownames.force = NA)
view(x)
object <- lars(x,y)
is.na(x)
lasso_y = read.csv("y.csv")
lasso_x = read.csv("final_regression_x.csv")
lasso_y = read.csv("final_regression_y.csv")
lasso_y = read.csv("final_regression_y.csv")
y = data.matrix(lasso_y, rownames.force = NA)
x = data.matrix(lasso_x, rownames.force = NA)
object <- lars(x,y)
library(glmnet)
fit1 = glmnet (x,y)
y = data.matrix (lasso_y)
x = data.matrix (lasso_x)
object <- lars(x,y)
view(lasso_x)
```

```
sum(is.na(y))
sum(is.na(x))
which (is.na(x))
sum(is.na(lasso_x))
which (is.na(lasso_x))
which (is.na(lasso_x$accommodates))
which (is.na(lasso_x$bathrooms))
which (is.na(lasso_x$bedrooms))
batman = lasso_x [192, "bedrooms"]
batman = lasso_x [192, "bedrooms"]
batman = lasso_x [192, "bedrooms"]
batman
lasso_x [192, "bedrooms"] = 1
view(lasso_x)
y = data.matrix (lasso_y)
x = data.matrix (lasso_x)
object <- lars(x,y)
plot(object)
summary(object)
fit1=glmnet(x,y)
print(fit1)
coef(fit1,s=0.01)
set.seed(42)
x.test = x[1:4589,]
x.train = x[1:4589]
x.train = x[1:4589,]
x.test = x[4590:6884,]
x.test = x[4590:6885,]
y.train = y[1:4589,]
y.test = y[4590:6884,]
alpha0.fit <- cv.glmnet(x.train, y.train, type.measure="mse",
```

```

alpha=0, family="gaussian")
alpha0.predicted <-
predict(alpha0.fit, s=alpha0.fit$lambda.1se, newx=x.test)
alpha0.predicted <-
predict(alpha0.fit, s=alpha0.fit$lambda.1se, newx=x.test)
mean((y.test - alpha0.predicted)^2)
alpha1.fit <- cv.glmnet(x.train, y.train, type.measure="mse",
alpha=1, family="gaussian")
alpha1.predicted <-
predict(alpha1.fit, s=alpha1.fit$lambda.1se, newx=x.test)
mean((y.test - alpha1.predicted)^2)
alpha0.5.fit <- cv.glmnet(x.train, y.train, type.measure="mse",
alpha=0.5, family="gaussian")
alpha0.5.predicted <-
predict(alpha0.5.fit, s=alpha0.5.fit$lambda.1se, newx=x.test)
mean((y.test - alpha0.5.predicted)^2)
list.of.fits <- list()
for (i in 0:10) { fit.name <- paste0("alpha", i/10)
list.of.fits[[fit.name]] <- cv.glmnet(x.train, y.train,
type.measure="mse", alpha=i/10, family="gaussian")
}
results <- data.frame()
for (i in 0:10) {
fit.name <- paste0("alpha", i/10)
predicted <- predict(list.of.fits[[fit.name]],
s=list.of.fits[[fit.name]]$lambda.1se, newx=x.test)
mse <- mean((y.test - predicted)^2)
temp <- data.frame(alpha=i/10, mse=mse, fit.name=fit.name)
results <- rbind(results, temp)
}
results
set.seed(1)

```

```
alpha0.fit <- cv.glmnet(x.train, y.train, type.measure="mse",
alpha=0, family="gaussian")
alpha0.predicted <-
predict(alpha0.fit, s=alpha0.fit$lambda.1se, newx=x.test)
mean((y.test - alpha0.predicted)^2)
set.seed(1011)
alpha0.fit <- cv.glmnet(x.train, y.train, type.measure="mse",
alpha=0, family="gaussian")
alpha0.predicted <-
predict(alpha0.fit, s=alpha0.fit$lambda.1se, newx=x.test)
mean((y.test - alpha0.predicted)^2)
cvob1=cv.glmnet(x,y)
plot(cvob1)
coef(cvob1)
cv.glmnet(x, y, alpha = 0)
cvob1=cv.glmnet(x,y, alpha=0)
plot(cvob1)
coef(cvob1)
lasso_survivors = read.csv("lasso_survivals.csv")
x1 = data.matrix(lasso_survivors)
cvob2=cv.glmnet(x1,y)
object <- lars(x1,y)
lasso_survivors = read.csv("lasso_survivals.csv")
x1 = data.matrix(lasso_survivors)
cvob2=cv.glmnet(x1,y)
coef(cvob2)
out <- lm(y,x1)
out <- lm(y,data=x1)
install.packages("lm")
library(lm)
install.packages("stats")
```

```
install.packages("stats")
library(stats)
out <- lm(y,data=x1)
glmmod <- glmnet(x, y, alpha=1, family="binomial")
glmmod <- glmnet(x, y, alpha=1, family="normal")
glmmod <- glmnet(x, y, alpha=1, family="gaussian")
plot(glmmod, xvar="lambda")
cv.glmmod <- cv.glmnet(x, y=asthma, alpha=1)
cv.glmmod <- cv.glmnet(x, y, alpha=1)
plot(cv.glmmod)
(best.lambda <- cv.glmmod$lambda.min)
batman = glmnet(x, y, lambda = 0.1311487)
print(batman)
coef(batman)
view(calendar)
dates1 = as.Date(calendar$date)
calendar$date = as.Date(calendar$date)
view(reviewsNum)
reviewsNum$date = as.Date(reviewsNum$date)
view(reviews1)
reviews1$date = as.Date(reviews1$date)
reviewsNum <- reviews %>% group_by(date = reviews$date) %>%
summarise(number = n())
reviews = read.csv("reviews Athens")
reviews = read.csv("reviews Athens.csv")
view(reviews)
reviews$date = as.Date(reviews$date)
reviewsNum <- reviews %>% group_by(date = reviews$date) %>%
summarise(number = n())
view(reviewsNum)
ggplot(reviewsNum, aes(date, number)) +
```



```

geom_point(na.rm=TRUE, color = "#007A87", alpha=0.5)
+geom_smooth(color = "#FF5A5F")+
ggtitle("How popular is Airbnb?",
subtitle = "Number of Reviews across years") +
labs(x = "Year", y = "Unique listings recieving reviews") +
theme(plot.title = element_text(face = "bold")) +
theme(plot.subtitle = element_text(face = "bold", color =
"grey35")) +
theme(plot.caption = element_text(color = "grey68"))
library(ggplot)
install.packages("ggplot")
library(ggplot2)
ggplot2(reviewsNum, aes(date, number)) +
geom_point(na.rm=TRUE, color = "#007A87", alpha=0.5)
+geom_smooth(color = "#FF5A5F")+
ggtitle("How popular is Airbnb?",
subtitle = "Number of Reviews across years") +
labs(x = "Year", y = "Unique listings recieving reviews") +
theme(plot.title = element_text(face = "bold")) +
theme(plot.subtitle = element_text(face = "bold", color =
"grey35")) +
theme(plot.caption = element_text(color = "grey68"))
ggplot(reviewsNum, aes(date, number)) +
geom_point(na.rm=TRUE, color = "#007A87", alpha=0.5)
+geom_smooth(color = "#FF5A5F")+
ggtitle("How popular is Airbnb?",
subtitle = "Number of Reviews across years") +
labs(x = "Year", y = "Unique listings recieving reviews") +
theme(plot.title = element_text(face = "bold")) +
theme(plot.subtitle = element_text(face = "bold", color =
"grey35")) +
theme(plot.caption = element_text(color = "grey68"))
ggplot(reviewsNum, aes(date, number)) +

```

```

geom_point(na.rm=TRUE, color = "#007A87", alpha=0.5)
+geom_smooth(color = "#FF5A5F")+
ggtitle("How popular is Airbnb in Athens?",
subtitle = "Number of Reviews across years") +
labs(x = "Year", y = "Unique listings receiving reviews") +
theme(plot.title = element_text(face = "bold")) +
theme(plot.subtitle = element_text(face = "bold", color =
"grey35")) +
theme(plot.caption = element_text(color = "grey68"))
lasso_x = read.csv("final_regression_x.csv")
lasso_y = read.csv("final_regression_y.csv")
y = data.matrix (lasso_y)
x = data.matrix (lasso_x)
fit1=glmnet(x,y)
library(glmnet)
fit1=glmnet(x,y)
print(fit1)
coef(fit1,s=0.01)
view(x)
view(y)
z = lasso_y/lasso_x$accommodates
z = data.matrix (z)
fit2=glmnet(x,z)
coef(fit2,s=0.01)
cv.glmmod <- cv.glmnet(x, y, alpha=1)
plot(cv.glmmod)
(best.lambda <- cv.glmmod$lambda.min)
object <- lars(x,y)
library(lars)
object <- lars(x,y)
plot(object)
summary(object)

```

```

fit1=glmnet(x,y, lambda = 0.1198061)
coef(fit1,s=0.01)
fit1=glmnet(x,y, lambda = 0.3)
coef(fit1,s=0.01)
fit1=glmnet(x,y, lambda = 0.6)
coef(fit1,s=0.01)
library(glmnet)
lasso_x = read.csv("final_regression_x.csv")
x = data.matrix(lasso_x)
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_model$lambda.1se
plot(lasso_model)
best.lambda
best_lambda = lasso_model$lambda.1se
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
deviance(lasso_model)
plot(lasso_model)
fit=glmnet(x,y,family="cox")
glmnet.control()
plot.cv.glmnet(lasso_coef)
lasso_survivors = read.csv("lasso_survivors.csv")
z = data.matrix (lasso_survivors)
fit2=glmnet(z,y)
lasso_survivors = read.csv("lasso_survivors.csv")
z = data.matrix (lasso_survivors)
fit2=glmnet(z,y)

```

```

lasso_survivors = read.csv("lasso_survivors.csv")
z = data.matrix (lasso_survivors)
fit2=glmnet(z,y)
coef(fit2,s=0.01)
lasso_model = cv.glmnet(z, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_model$lambda.1se
plot(lasso_model)
best_lambda = lasso_model$lambda.1se
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
_lambda]
lasso_coef
install.packages("penalized")
library(penalized)
help("penfit")
set.seed(1)
fit = penalized(y, x, lambda1=1)
penfit(show)
coefficients(fit, "all")
penalty(fit)
weights(fit)
residuals(fit)
plot(residuals(fit))
fitted(fit)
plot(fitted(fit)fit)
plot(fitted(fit))
model(fit)
lin.pred(fit)
penfit(model)
penalized(fit)
fit = penalized(y, z, lambda1=1)

```

```
penfit(show)
coefficients(fit, "all")
residuals(fit)
plot(residuals(fit))
lambda.1se
lasso_x = read.csv("final_regression_x.csv")
x = data.matrix(lasso_x)
fit = penalized(y, x, lambda1=1)
lasso_1 = read.csv("final_regression_x.csv")
x = data.matrix(lasso_1)
fit = penalized(y, x, lambda1=1)
fit = penalized(y, x, lambda1=2)
fit = penalized(y, x, lambda1=3)
fit = penalized(y, x, lambda1=50)
fit = penalized(y, x, lambda1=100)
coefficients(fit, "all")
library(glmnet)
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_model$lambda.1se
plot(lasso_model)
best.lambda
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
_lambda]
lasso_coef
lasso_model$lambda.min
best_lambda
best_lambda = lasso_model$lambda.min
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
_lambda]
lasso_coef
```

```
lasso_coef > 3.5
best_lambda = lasso_model$lambda.1se
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
install.packages("devtools")
devtools::install_github("ewenharrison/finalfit")
install.packages("rstan")
plot(lasso_model)
plot(lasso_model)
library(gkmnet)
library(glmnet)
plot(lasso_model)
slope <- cor(x, y) * (sd(y) / sd(x))
x
library(stats)
lm_test = lm(y ~ x)
plot(lm_test)
anova(lm_test)
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_model$lambda.1se
plot(lasso_model)
best_lambda
view(lasso_x)
plot(lasso_model)
library(glmnet)
plot(lasso_model)
lasso_model = cv.glmnet(x, y, lambda = 7, alpha=1)
best_lambda = 7
```

```

lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
_lambda]
lasso_coef
best_lambda = 0.5
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
_lambda]
lasso_coef
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_model$lambda.1se
plot(lasso_model)
best.lambda
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
_lambda]
lasso_coef
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best
.lambda]
lasso_coef
install.packages('relaxnet')
install.packages("relaxnet")
library(ggplot2)
library(relaxnet)
view(lasso_1)
x=as.matrix(lasso_1)
y=as.matrix(lasso_y)
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
batman = 10^seq(4, -1, -.1)
batman
plot(batman)
cvm

```

```
lasso_model(cvm)
lasso_model$cvm
plot(lasso_model$cvm)
lasso_model$lambda.min
lasso_model$cvstd
plot(lasso_model$cvstd)
plot(lasso_model)
best_lambda = lasso_model$lambda.1se
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
best_lambda
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==10]
lasso_coef
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==2]
lasso_coef
best_lambda = lasso_model$lambda.1se
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda=2]
best_lambda = 2
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==1.19]
lasso_coef
```



```

best_lambda = lasso_model$lambda.1se
best_lambda
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda=1.05]
best_lambda = 1.05
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef
cvob1=cv.glmnet(x,y)
plot(cvob1)
coef(cvob1)
cvob1=cv.glmnet(x,y, lambda=2)
cvob1=cv.glmnet(x,y, lambda=seq(1,2,0.05))
plot(cvob1)
cvob1=cv.glmnet(x,y, lambda=seq(2,10,1))
plot(cvob1)
cvob1=cv.glmnet(x,y, lambda=seq(3,3,0.05))
cvob1=cv.glmnet(x,y, lambda=seq(3,4,0.05))
plot(cvob1)
plot(cvob1,pch=19)
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==1.2]
lasso_coef
cvob1a=cv.glmnet(x,y,type.measure="mae")
plot(cvob1a)
lasso.1 <- glmnet(y=y.1, x= x.1, family="gaussian")
lasso.1 <- glmnet(x,y, family="gaussian")
lasso.1

```

```

plot(lasso.1) # Plots coefficient path
coef(lasso.1)
write.csv(coef(lasso.1), file = "lasso_coefficients.csv")
lasso_coefficients = as.data.frame(coef(lasso.1))
library(tm)
adtm.df<-as.data.frame(as.matrix(coef(lasso.1)))
write.csv(adtm.df), file = "lasso_coefficients.csv")
write.csv(adtm.df), file = "lasso_coefficients.csv")
write.csv(adtm.df), file = "lasso_coefficients.csv")
write.csv(adtm.df, file = "lasso_coefficients.csv")
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_coefficients = coef(lasso_model)
adtm.df<-as.data.frame(as.matrix(coef(lasso_model)))
write.csv(adtm.df, file = "lasso_coefficients.csv")
write.csv(adtm.df, file = "lasso_coefficients1.csv")
cv.result <- cv.relaxnet(x, y)
predict(cv.result$relaxnet.fit$main.glmnet.fit,
type = "nonzero",
s = cv.result$main.lambda.1se)
coefs <- drop(predict(cv.result, type = "coef"))
coefs
lasso_survivors = read.csv("lasso_survivors.csv")
lasso_survivors = read.csv("lasso_survivors.csv")
x_survivor = data.matrix(lasso_survivors)
lasso_model = cv.glmnet(x_survivor, y, lambda = 10^seq(4, -1, -
.1), alpha=1)
lasso_coefficients = coef(lasso_model)
lasso_coefficients
ln_y = ln(y)
install.packages("SciViews")
library(SciViews)

```

```

ln_y = ln(y)
lasso_model = cv.glmnet(x_survivor, ln_y, lambda = 10^seq(4, -1, -.1),
-.1), alpha=1)
lasso_coefficients = coef(lasso_model)
lasso_coefficients
lasso_model = cv.glmnet(x, ln_y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_coefficients = coef(lasso_model)
lasso_coefficients
r2 <- fit$glmnet.fit$dev.ratio[which(fitnet$glmnet.fit$lambda ==
best_lambda)]
r2 =
lasso_model$glmnet.fit$dev.ratio[,lasso_model$glmnet.fit$lambda==best_lambda]
r2 =
lasso_model$glmnet.fit$dev.ratio[,lasso_model$glmnet.fit$lambda==best_lambda]
r2 =
lasso_model$glmnet.fit$dev.ratio[which(lasso_model$glmnet.fit$lambda==best_lambda)]
r2
best_lambda
lasso_model = cv.glmnet(x, y, lambda = 10^seq(4, -1, -.1),
alpha=1)
lasso_coefficients = coef(lasso_model)
best_lambda = lasso_model$lambda.1se
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef =
lasso_model$glmnet.fit$beta[,lasso_model$glmnet.fit$lambda==best_lambda]
lasso_coef

```

```

r2 =
lasso_model$glmnet.fit$dev.ratio[which(lasso_model$glmnet.fit$lambda==best_lambda)]
r2
lasso_model = cv.glmnet(x_survivor, y, lambda = 10^seq(4, -1, -.1), alpha=1)
best_lambda = lasso_model$lambda.1se
r2 =
lasso_model$glmnet.fit$dev.ratio[which(lasso_model$glmnet.fit$lambda==best_lambda)]
r2
lasso_model$cvm
lasso_model$cvstd
mean(lasso_model$cvstd)
mean(lasso_model$cvm)
lasso_model$glmnet.fit$dev.ratio
rome_x = read.csv("rome_x.csv")
rome_y = read.csv("rome_y.csv")
x_rome = data.matrix(rome_x)
y_rome = data.matrix(rome_y)
lasso_model = cv.glmnet(x_rome, y_rome, lambda = 10^seq(4, -1, -.1), alpha=1)
is.na(rome_x$accommodates)
which(is.na(rome_x))
rome_x = read.csv("rome_x.csv")
rome_y = read.csv("rome_y.csv")
x_rome = data.matrix(rome_x)
y_rome = data.matrix(rome_y)
lasso_model = cv.glmne

```