

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

«Airbnb Beyond Stars: What Reviews Reveal»

ΚΟΥΚΙΑ ΦΡΑΝΤΣ

Διπλωματική εργασία υποβληθείσα προς μερική εκπλήρωση
των απαραίτητων προϋποθέσεων
για την απόκτηση του
Μεταπτυχιακού Διπλώματος Ειδίκευσης

Αθήνα
Δεκέμβριος, 2025



Επιβλέπων καθηγητής: Διοικητόπουλος Ευάγγελος - Οικονομικό Πανεπιστήμιο Αθηνών

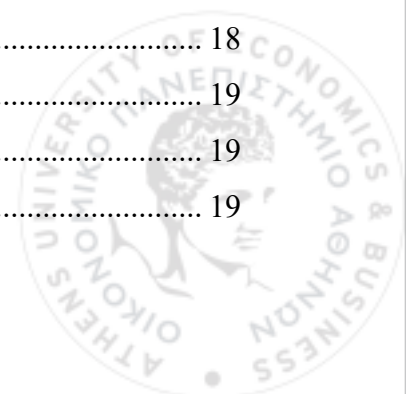
Εξεταστής καθηγητής: Παγκράτης Σπυρίδων, Οικονομικό Πανεπιστήμιο Αθηνών

Εξεταστής καθηγητής: Δενδραμής Ιωάννης, Οικονομικό Πανεπιστήμιο Αθηνών



Table of Contents

Table of Figures	5
Abstract.....	6
Περίληψη	7
Introduction.....	9
Literature Review.....	11
Airbnb and the Sharing Economy.....	11
Determinants of Airbnb Listing Prices	11
Online Reviews, Ratings, and Reputation Systems	12
Sentiment Analysis and Textual Reviews	12
Spatial Heterogeneity in Airbnb Markets	13
Macroeconomic Conditions and Consumer Sentiment.....	13
Contribution of the Present Study.....	14
Data Description	15
Data Sources	15
Airbnb Reviews files (per city).....	15
Airbnb Listings files (per city).....	15
Temporal Coverage.....	15
Macroeconomic Data	16
Data Integration	16
Descriptive Statistics.....	16
Listings Dataset.....	16
Reviews Dataset.....	17
Methodology	18
Configurations.....	18
Preprocessing & Function Setup.....	18
Price Cleaning.....	18
Sentiment Computation	18
Sigmoid Calibration	19
Accessing Macroeconomic indices	19
Defining a regression model	19



Review sample balancing	20
Spatial Clusters and Orientation	21
Main Analysis Pipeline & Price Regression	21
City- year aggregation & Sentiment regression	22
Main execution.....	23
Results and Discussion	25
Price regression	25
Sentiment over time Regression	27
Sentiment inferred ratings against actual ratings	28
Measuring the accuracy of the Sentiment Inferred Rating	30
Sentiment interpretability.....	31
Sentiment Cluster Segmentation.....	34
London clusters further analysis	38
Sentiment dispersion	39
Sentiment Dispersion for London (K = 14)	41
Sentiment Regression.....	42
Macroeconomic indicators analysis	42
Pooled city regression with fixed effects	44
Interpretation of the regression summary	45
Conclusions.....	47
Future work.....	49
References.....	51



Table of Figures

Figure 1: City-year aggregated dataset	23
Figure 2: Price Determinants	27
Figure 3: Sentiment over time for all cities	28
Figure 4: Rating Distributions - Amsterdam.....	29
Figure 5: Rating Distributions - Zurich	30
Figure 6: Top Positive/Negative tokens across cities - Initial approach.....	31
Figure 7: Top 15 Positive/ Negative words aggregated for all cities	33
Figure 8: scatterplot of listing clusters - Berlin	35
Figure 9: Barplot of average sentiment per cluster - Berlin.....	35
Figure 10: Spatial clusters on map - Berlin	35
Figure 11: Top negative words for Berlin cluster 3	36
Figure 12: Scatterplot of listings clusters - London.....	37
Figure 13: Barplot of average Sentiment per cluster - London	37
Figure 14: Spatial clusters on map London	37
Figure 15: Silhouette score for no. of clusters in London.....	38
Figure 16: Clustering using $K = 14$ in London	38
Figure 17: Sentiment dispersion of cities.....	39
Figure 18 : Sentiment dispersion over time - Madrid.....	40
Figure 19: Sentiment dispersion over time - London	40
Figure 20: Sentiment dispersion over time - Berlin.....	40
Figure 21: Inflation over time for all cities	43
Figure 22 : GDP per capita over time for all cities	43
Figure 23: Sentiment regression summary	44



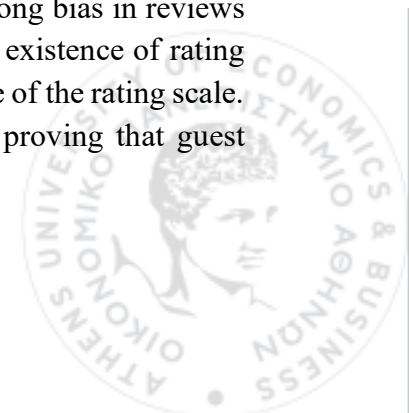
Abstract

Online reviews are a central aspect when it comes to shaping consumer decisions on digital accommodation platforms such as Airbnb. Besides the use of numerical ratings, the textual content of reviews provides much richer qualitative information about user experiences, perceptions, and satisfaction. At the same time, broader economic conditions (GDP per capita, inflation rate, unemployment rate) are affecting everyday life thus influencing user expectations and behavior. The study examined firstly how pricing is associated with different property attributes and secondly how the sentiment that is extracted from the textual part of Airbnb reviews relates over time with spatial clusters within cities, and macroeconomic conditions across multiple European cities.

Using datasets that consist of Airbnb listings (and their characteristics) and reviews from eight European cities, the study applies natural language processing techniques (NLP) to transform the textual part of the review into sentiment scores. To make sure that this process is conducted uniformly across cities and over the same period of time, the review samples are balanced per year and city, and the sentiment is aggregated at both listing level and later on at city-year level. A sigmoid transformation is used to map the raw sentiment scores onto the Airbnb rating scale (bounded scale [1,5]), allowing for direct comparisons between inferred sentiment scores and actual user ratings. Sentences are being split into tokens to support the analysis and understand a city's reviewing tendencies by identifying the words that most strongly contribute to positive and negative sentiment.

The analysis proceeds in several stages. First, regressions are conducted on aggregated listing level to identify price determinants, incorporating property attributes such as accommodation characteristics (capacity of each property and review rating) and host attributes (superhost status, host experience). Second, spatial sentiment analysis is conducted to explore within-city variations that are observed among the different clusters that were created. Finally, in the core section of the study, the sentiment extracted from reviews is tested for association with macroeconomic indicators by constructing a city-year dataset and estimating regression models with city fixed effects.

Among all regressions, it is confirmed that the accommodation capacity is the strongest and most consistent determinant of listing prices, while other factors such as review ratings, superhost status and host experience exhibit heterogeneous or negligible effects on price. A strong bias in reviews is observed by the sentiment analysis that confirms previous literature on the existence of rating inflation, meaning that sentiment scores are more concentrated in the upper side of the rating scale. Geographic clustering suggests that there are differences across city areas, proving that guest experiences are not uniform.



At the macroeconomic level, the pooled regression results indicate that changes in economic conditions are associated with changes in average sentiment over time, although the size of these effects is not highly significant and should be interpreted as associative rather than causal.

Overall, the study demonstrates that integrating sentiment analysis and spatial clustering methods provide additional value that helps us improve our understanding of user behavior that is detected on digital platforms. The findings contribute to the already existing literature on platform economics and computational social science. It is important to note that the sentiment analysis potential is both big, but also has some limitations regarding sentiment measuring methods that are conducted in highly positive reviewed environments.

Περίληψη

Οι διαδικτυακές αξιολογήσεις και κριτικές αποτελούν πλέον έναν από τους βασικότερους παράγοντες που επηρεάζουν τις αποφάσεις των καταναλωτών όταν βρίσκονται στη διαδικασία εύρεσης διαμονής όπως για παράδειγμα στην πλατφόρμα του Airbnb. Εκτός από τη χρήση βαθμολογιών, το λεκτικό περιεχόμενο των κριτικών παρέχει περισσότερη ποιοτική πληροφόρηση σχετικά με τις εμπειρίες, τις αντιλήψεις και την ικανοποίηση των χρηστών. Ταυτόχρονα, μελετάται η επίδραση των μακροοικονομικών παραγόντων (κατά κεφαλήν ΑΕΠ, ποσοστό πληθωρισμού, ποσοστό ανεργίας) στις προσδοκίες και τη συμπεριφορά των χρηστών. Η μελέτη αυτή εξετάζει, πρώτον, πώς η τιμολόγηση συνδέεται με διαφορετικά χαρακτηριστικά των ακινήτων (χωρητικότητα, μέση βαθμολογία καταλύματος, στάτους του οικοδεσπότη, χρόνια εμπειρίας του οικοδεσπότη) και, δεύτερον, πώς το συναίσθημα που εξάγεται από το κείμενο των αξιολογήσεων της Airbnb σχετίζεται με τους ευρύτερους οικονομικούς δείκτες, τις διαφορετικές συστάδες περιοχών μέσα στις πόλεις και τις μακροοικονομικές συνθήκες.

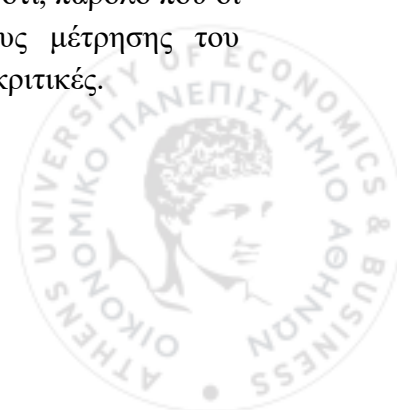
Χρησιμοποιώντας τα δεδομένα των καταλυμάτων και των αξιολογήσεων της Airbnb από οκτώ ευρωπαϊκές πόλεις, η μελέτη εφαρμόζει τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) για να μετατρέψει το λεκτικό μέρος μιας κριτικής στη μέση βαθμολογία συναισθήματος για κάθε κατάλυμα. Για να διασφαλιστεί ότι αυτή η διαδικασία διεξάγεται ομοιόμορφα σε όλες τις πόλεις και στη διάρκεια της χρονική ανάλυσης, τα δείγματα αξιολογήσεων εξισορροπούνται ανά έτος και πόλη, και το εξαγόμενο συναίσθημα συγκεντρώνονται τόσο σε επίπεδο καταλύματος όσο και σε επίπεδο πόλης ανά έτος. Χρησιμοποιείται σιγμοειδής μετασχηματισμός για την αντιστοίχιση των ακατέργαστων βαθμολογιών συναισθήματος που προήλθαν από τη χρήση ενός εξωτερικού εργαλείου με την κλίμακα αξιολόγησης της Airbnb (κλίμακα [1,5]), επιτρέποντας την άμεση σύγκριση μεταξύ του συναισθήματος που εξάγεται και των πραγματικών αξιολογήσεων των χρηστών. Η ανάλυση σε επίπεδο λέξης, χρησιμοποιείται για να βελτιώσει την ερμηνεία των τάσεων αξιολόγησης μιας πόλης, εντοπίζοντας τις λέξεις που συμβάλλουν περισσότερο στη διαμόρφωση θετικού και αρνητικού συναισθήματος αντίστοιχα.



Η ανάλυση εκτείνεται σε διάφορα στάδια. Πρώτον, πραγματοποιούνται παλινδρομήσεις σε συγκεντρωτικό επίπεδο καταλύματος για τον προσδιορισμό των καθοριστικών παραγόντων της τιμής, ενσωματώνοντας ορισμένα χαρακτηριστικά ακινήτων, όπως τα χαρακτηριστικά του καταλύματος, τα χαρακτηριστικά του οικοδεσπότη (καθεστώς superhost, εμπειρία οικοδεσπότη) και τη χωρική ομαδοποίηση με βάση τις γεωγραφικές συντεταγμένες. Δεύτερον, πραγματοποιείται ανάλυση του εξαγόμενου συναισθήματος με βάση τη διαμόρφωση χωρικών ομάδων των καταλυμάτων οι οποίες πραγματοποιούνται ανάλογα με το γεωγραφικό μήκος και πλάτος στο οποίο είναι καταχωρημένο ένα κατάλυμα. Η ανάλυση αυτή σκοπεύει στο να διερευνηθούν οι ενδοαστικές διακυμάνσεις που παρατηρούνται στους διάφορους σχηματισμούς ομάδων και η πιθανή συσχέτιση τους με ευρύτερες περιοχές των πόλεων. Τέλος, το βασικό τμήμα της μελέτης στοχεύει στο να εξετάσει τη συσχέτιση μεταξύ του μέσου συναισθήματος που εξάγεται ανά πόλη και έτος με τους προαναφερθέντες μακροοικονομικούς δείκτες, κατασκευάζοντας ένα συγκεντρωτικό πίνακα πόλης-έτους και εκτιμώντας μοντέλα παλινδρόμησης με σταθερούς παράγοντες πόλης (city-fixed effects).

Τα αποτελέσματα επιβεβαιώνουν ότι η χωρητικότητα των καταλυμάτων είναι ο ισχυρότερος και πιο σταθερός καθοριστικός παράγοντας των τιμών καταλύματος, ενώ άλλοι παράγοντες, όπως οι αξιολογήσεις των χρηστών και η ιδιότητα του Superhost, εμφανίζουν ετερογενείς επιδράσεις στις διάφορες πόλεις. Η ανάλυση του συναισθήματος των χρηστών δείχνει μια ισχυρή θετική προκατάληψη στις κριτικές του Airbnb, με τις βαθμολογίες που προκύπτουν από το συναίσθημα των χρηστών να ευθυγραμμίζονται στενά με τις αριθμητικές αξιολογήσεις, αλλά να είναι πιο συμπιεσμένες στο ανώτερο άκρο της κλίμακας (υπερεκτίμηση των πολύ θετικών κριτικών). Η χωρική ομαδοποίηση ανά πόλη υπογραμμίζει σημαντικές διαφορές στο συναίσθημα που εξάγεται από τις κριτικές μεταξύ των περιοχών της πόλης, υποδηλώνοντας ότι οι εμπειρίες των επισκεπτών δεν είναι γεωγραφικά ομοιόμορφες. Σε μακροοικονομικό επίπεδο, τα συγκεντρωτικά αποτελέσματα της παλινδρόμησης δείχνουν ότι οι αλλαγές στις οικονομικές συνθήκες συνδέονται με αλλαγές στη μέση διάθεση με την πάροδο του χρόνου, αν και το μέγεθος των επιδράσεων δεν είναι ιδιαίτερα σημαντικό και πρέπει να ερμηνεύεται ως συσχετιστικό και όχι αιτιώδες.

Συνολικά, η μελέτη καταδεικνύει ότι υπάρχει πρόσθετη αξία όταν ενσωματώνεται η ανάλυση συναισθημάτων με ευρύτερους μακροοικονομικούς δείκτες και μεθόδους χωρικής ομαδοποίησης, που μας βοηθούν να κατανοήσουμε καλύτερα τη συμπεριφορά των χρηστών που ανιχνεύεται στις ψηφιακές πλατφόρμες. Η έρευνα παρέχει μια πολυεπίπεδη προοπτική σχετικά με τον τρόπο με τον οποίο οι υποκειμενικές αξιολογήσεις αλληλεπιδρούν με δομικούς και συγκείμενους παράγοντες. Τα ευρήματα συμβάλλουν στην αυξανόμενη βιβλιογραφία σχετικά με την οικονομία των πλατφόρμων και τις υπολογιστικές κοινωνικές επιστήμες, ενώ υπογραμμίζουν ότι, παρόλο που οι δυνατότητες είναι υψηλές, υπάρχουν αρκετοί περιορισμοί στις μεθόδους μέτρησης του συναισθήματος που εφαρμόζονται σε περιβάλλοντα με πλεονάζουσες θετικές κριτικές.



Introduction

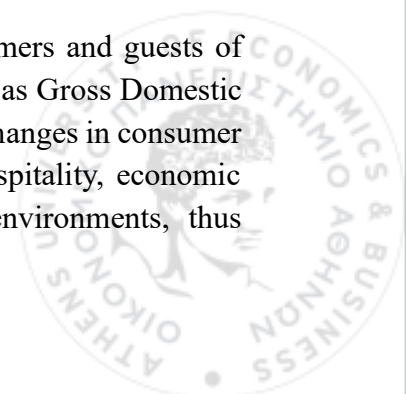
In recent years the rise of the sharing economy has impacted the way customer satisfaction is measured. Airbnb has emerged as a dominant player in the apartment rental market, allowing individuals to rent their property to travelers and natives reshaping the way accommodation is supplied and consumed across countries worldwide. As a result, the need to understand the factors that influence user experience and evaluation has become more crucial for all interested parties.

Airbnb as a case study is suited for this example as it can offer a lot of information given the nature of its evaluation system and the magnitude it operates on (scale and diversity of user base). The platform operates across a wide range of countries and cities, each with distinct cultural, economic, and regulatory environments while the user diversity allows for comparative analysis across markets maintaining a consistent platform structure.

A defining feature of online platforms is the availability of large volumes of user-generated content in the form of textual reviews and ratings. Numerical ratings provide a concise summary of guest satisfaction on a given scale ([1,5]), while textual reviews offer a more detailed and nuanced assessment of the overall accommodation experience and degree of satisfaction. Previous research has shown that textual sentiment extracted from reviews can serve as a valuable insight for user experience. At the same time, the relationship between textual sentiment and numerical ratings is not always straightforward, as ratings tend to be positively skewed and may fail to capture subtle variations in user perceptions. A very good example to understand this is the restaurant sector. Even though the middle point of the reviewing scale for restaurants via google maps is 2.5 out of 5 (same as the Airbnb scale), the ratings are overwhelmingly positively skewed on the right side of the scale that a “mediocre” restaurant is considered to be one with an average rating of 4 or close to that rating. That alone can explain a lot on how the reviewing system works when the ratings are dominated by 5 starred reviews and the tendency that people have, is to review positively.

Numerical ratings offer a convenient way to summarize the evaluations that are left from the users, but sometimes they often offer limited variance and potentially showcase rating inflation, particularly when users may be hesitant to leave negative feedback (users avoiding to review when they’ve had a bad experience). Textual reviews allow users to honestly express themselves, which may lead to identifying the reasons of dissatisfaction, praise, or mixed feelings. Sentiment analysis is proven to be a systematic tool that helps quantify these textual signals, enabling an extra layer of analysis that targets subjective experiences . As such, sentiment scores can complement numerical ratings by capturing nuances that may not be reflected in star-based evaluation systems.

Broad economic conditions affect people in their everyday lives with consumers and guests of these platform being no exception of the rule. Macroeconomic indicators such as Gross Domestic Product per capita, inflation rate, and unemployment rate have been linked to changes in consumer confidence, expectations, and satisfaction. In the context of tourism and hospitality, economic conditions may shape both travelers’ expectations and hosts’ operating environments, thus



indirectly affecting how experiences are perceived and reported on digital platforms. It is important to note that since the majority of the users of the Airbnb platform are travelers and not natives of the cities examined, any potential macroeconomic changes might mainly affect the mix of travelers.

A major aspect of the study is to examine the correlation between property prices and certain attributes that each property has to offer as well as the relationship between user sentiment that is expressed in Airbnb reviews and the overall movement of macroeconomic indices across multiple European cities.

To address these questions, the study adopts a detailed analytical approach. At the micro level, listing characteristics and host attributes are examined to understand pricing behavior within cities. At the mid-level, spatial variation in sentiment is explored to assess whether user experiences differ across areas within the same city. Finally, at the macro level, sentiment trends are linked to country-level economic indicators over time and are tested for any significant association.

The analysis is primarily focused on exploring the nature of the associations and does not aim to establish causal relationships between sentiment, prices, and macroeconomic conditions. Instead, it seeks to identify and potentially document patterns and associations that emerge across cities over time, while highlighting the potential role of economic context in shaping user evaluations on digital platforms.



Literature Review

Airbnb and the Sharing Economy

The rapid increase of businesses that contribute to the sharing economy has generated significant interest among economists due to their potential to change traditional markets and reshape consumer behaviour. Platforms such as Airbnb are a good example of markets that are driven by individuals, where they act as both suppliers and consumers within a digital environment. The fact that the Airbnb business does not operate under centralized ownership structures like traditional hotels, enables private property owners to utilize their unused properties and offer short-term accommodation in return for extra income. In turn that increases the supply and introduces new forms of competition in local tourism markets.

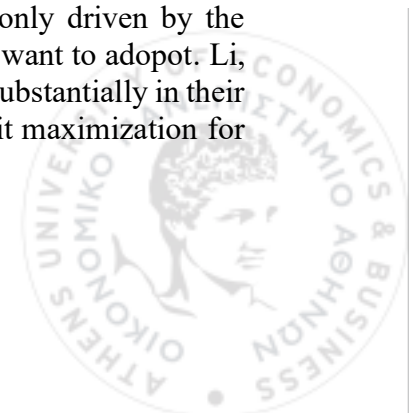
Various types of research has been conducted by economists that focused on the impact of Airbnb in the hotel industry. Specifically, Zervas, Proserpio, and Byers (2017) in one of the first assessments of Airbnb's effects in the hotel industry, come to a conclusion that the bigger the penetration is on the market by Airbnb, the bigger the decline of hotel revenues are, especially when that happens in areas that are linked with high tourism demand and limited hotel capacity. Similarly, Barron, Kung, and Proserpio (2021) point out the effect that Airbnbs can have in housing prices and rental markets (which in turn has an immediate effect on the population living in these areas) and the economic implications that might arise.

Einav, Farronato and Levin (2016) standing from a theoretical perspective, highlight the importance of reputation systems (reviews), in environments where information asymmetry is detected. In these types of markets, the content generated by the users/ guests is an essential tool that helps coordinate both supply and demand becoming a critical function component of the platform.

Determinants of Airbnb Listing Prices

It has been identified by many studies that the property characteristics of a listing are important factors of determining the listing's price. These characteristics have been proved to be the accommodation capacity, the location of the property and the amenities offered. Wang and Nicolau (2017) are highlighting the utility and higher demand that listings located closer to city centers offer and that this very fact allows them to be more expensive.

Host related characteristics also play an important part in deciding prices. Superhosts, that are considered to be high quality and reliable hosts, influence pricing and booking likelihood, although this observation has not been proven to be true for all markets. Gibbs et al. (2018) argue that pricing behavior on Airbnb is highly fragmented, which means that pricing is not only driven by the listings attributes but by the strategy and degree of competitiveness that hosts want to adopt. Li, Moreno, and Zhang (2016) further emphasize that the strategy of hosts differs substantially in their pricing objectives, with some prioritizing occupancy and reputation over profit maximization for given periods of time.



It has also been noted by several studies that the regression models can only explain a small proportion of the variation in price as there are so many factors that cannot be accounted for. That alone reflects the complexity of price formation on these types of platforms, where prices are being determined by many other factors such as local demand conditions, seasonality, regulation, location and host behavior.

Online Reviews, Ratings, and Reputation Systems

Reputation systems are essential nowadays when trying to get information of how guests experienced certain listings. Reviews can remove a part of the uncertainty and offer the user a sense of security that they have made the right choice. On Airbnb, guests are encouraged to leave both numerical ratings and textual reviews following each stay.

However, prior research has identified systematic biases in online rating systems. Dellarocas and Wood (2008) found out that negative feedback is not always reported due to social and other considerations by the guests. That hesitation leads to inflated ratings. Moe and Schweidel (2012) also highlight that there is a bias in positive reviews that leads to average ratings clustering near the upper end of the scale. That is also noted by Filier et al (2021) who state it is considered normal that the expectation from everyone is to review positively.

These findings suggest that numerical ratings alone may fail to capture subtle differences in user experiences. As a result, textual reviews have gained attention as a complementary source of information that may reveal dissatisfaction, nuance, or mixed evaluations that are not reflected in star ratings.

Sentiment Analysis and Textual Reviews

There has been an increase in the use of sentiment analysis methods to extract quantitative information from text, not only in the sharing economy but in financial markets as well. Baker and Wurgler (2006) demonstrate that sentiment can influence asset prices, while Taddy (2013) illustrates how textual sentiment can be systematically measured and incorporated into empirical models.

Lexicon methods are one of the most common tools to review text and extract sentiment, especially in the form of informal text that is similar to reviews. Hutto and Gilbert (2014) introduce the VADER sentiment analysis tool, which is specifically designed to capture sentiment intensity in social media and review-style text. The use of the VADER sentiment analysis tool has proven to be a strong solution to analyze text and has a high degree of correlation between sentiment scores and numerical ratings.



Hu, Zhang, and Pavlou (2009) show that ratings often follow a J-shaped distribution, limiting their ability to differentiate among highly rated products. Liu et al. (2017) further demonstrate that sentiment scores can capture emotional tone and dissatisfaction even when ratings remain high. These findings motivate the use of sentiment analysis as a complementary measure of user evaluations, particularly in settings characterized by rating inflation.

Spatial Variation in Airbnb Markets

Another aspect that has been looked at is the location of the listings and the overall area of Airbnb concentration. Studies show that listings usually are more concentrated in more central and tourist areas and neighbourhoods covering the need of visitors of a place to stay that is near the city center. Quattrone et al (2016) pinpoint the difference that is observed to listings that are located more centrally and the degree of demand they receive compared to other areas.

Clustering techniques are helpful with capturing spatial variation, that is not necessarily linked to administrative areas and neighborhoods. K-means clustering based on geographic coordinates has been used to identify functional urban areas that share similar characteristics. Such approaches are very insightful as they allow for the analysis of spatial variation in prices and sentiment without having restrictions related to predefined neighborhood definitions.

Macroeconomic Conditions and Consumer Sentiment

Studies that are related to macroeconomic conditions and their effect to subjective evaluations of people in their everyday lives have always been a target of research. Di Tella, MacCulloch, and Oswald (2001) show that inflation and unemployment are negatively associated with the general sentiment reported by the public, while Deaton (2008) documents strong associations between income levels and subjective life evaluations (well-being) across countries, though that is not related to short-stay platform reviews. The findings suggest that economic conditions can shape expectations, perceptions, and satisfaction and that the degree of association needs to be explored.

In tourism-related contexts, macroeconomic conditions may influence both who travels and how experiences are evaluated. Economic expansions can increase travel demand and spending, while downturns may alter traveler composition and expectations. Gennaioli, Shleifer, and Vishny (2018) emphasize the role of beliefs and expectations in shaping economic behavior, arguing that sentiment can evolve systematically in response to economic environments.

Despite this, there has not been extensive research conducted to examine how macroeconomic indicators relate directly to sentiment expressed in online reviews. Most existing studies focus on demand or spending rather than subjective evaluations, leaving a gap in understanding how broader economic conditions interact with user-generated sentiment.



Contribution of the Present Study

Drawing from existing literature, this study aims to use sentiment analysis methods to explore possible links and associations with spatial variation, macroeconomic conditions (expressed in the form of indicators) and to confirm or oppose previous claims of price determinants in the Airbnb environment. The analysis is conducted for a number of cities (eight), that are capitals of countries considered as strong economies of Europe, over an extended period of time (with sufficient data to be analyzed) aiming to compare results both across cities but also in relation with time.



Data Description

Data Sources

For the purposes of the analysis conducted for this project, the data (Airbnb platform data & macroeconomic indices) were acquired from two publicly available websites. Inside Airbnb offers many types of data, in our case, the reviews and listings datasets were downloaded for each of the eight (8) cities that were included in the analysis (Amsterdam, London, Madrid, Rome, Berlin, Stockholm, Vienna and Zurich). The corresponding files provide detailed information about accommodation characteristics and user-generated reviews that span over the entire time period of the Airbnb platform operation. The macroeconomic indices were gathered by querying the World Bank website and include measures of economic activity at country level (GDP per capita, inflation, unemployment). The combined datasets cover the period starting from 2009 to 2024, although the analysis focuses on a later subset of this period (2013-2024) due to data insufficiency that was observed at the early years of that period (2009-2012).

Airbnb Reviews files (per city)

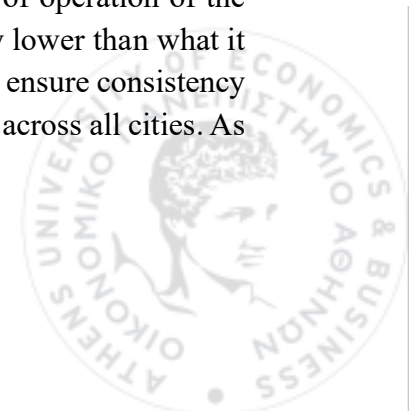
The Reviews dataset used in our analysis contains (among other files) textual reviews provided by guests who stayed at various AirBnB properties/ listings. Each observation corresponds to a single review and includes the textual content of the review, the date it was posted and a unique identifier that links it to the corresponding listing. The reviews themselves are in unstructured format and vary in length.

Airbnb Listings files (per city)

The Listings data comprises detailed information about each property listed on AirBnB. Each observation corresponds to a single listing and includes different variables that describe the pricing, capacity, host characteristics and geographical location. Additionally, an average numerical rating is provided for each listing that has received ratings from guests.

Temporal Coverage

The datasets acquired comprise information about the listings and reviews since the early establishment of the Airbnb (2009) platform, though the review activity on the platform is not evenly distributed over time. That is to be expected as in the very first years of operation of the platform the number of users and thus the number of reviews was substantially lower than what it was a few years later and ultimately than what it has grown to be nowadays. To ensure consistency and data uniformity, the study focuses on years with sufficient data availability across all cities. As a result, the analysis begins from 2013 and onwards.



Macroeconomic Data

Macroeconomic indices were obtained from the World Bank through its public API. These data (GDP, inflation, unemployment) are reported at an annual rate and at country level. Each city selected is associated with a corresponding country, thus matching the observations of each city to these indicators based on each year. These indicators provide a measure of the broader economic environment in which Airbnb activity takes place.

Data Integration

Although the datasets in this study vary in the unit of observation (listings' files are observed at property level while reviews are at individual level), the reviews are linked to listings through the identifiers found within the files (listing id) while the macroeconomic indices are linked to cities through the corresponding country codes.

Descriptive Statistics

To provide context for the empirical analysis, descriptive statistics are presented for both the listings and reviews datasets, illustrating their scale, structure, and key characteristics. The two main files for each city are the Listings csv and the Reviews csv. Both were downloaded and are located within each city's DATA folder and contain information about the listings and the reviews of the city over time.

Listings Dataset

For example, London's listings file contains 96.871 entries having each entry correspond to one listing, and a total of 79 columns. The column section is divided into 4 categories. At first there are identification columns such as :

- id of the listing,
- url of the listing,
- scrape id (unique identifier related to the scraping task)

Then, host related columns:

- host id,
- host url,
- host name,
- host location,
- host is superhost,
- host response time etc.

and several others amounting to a total of 21 columns for that section.

Following there are the property related columns like:

- price



- latitude – longitude
- accommodates (how many people can the listing accommodate)
- bathrooms
- minimum nights etc.

The last section of the columns consists of reviews-related columns such as :

- number of reviews,
- review score rating
- review score's location,
- review score value etc.

Reviews Dataset

Continuing with the same example as in the listings dataset, London's reviews dataset contains 2.097.995 entries each one representing one review. The reviews dataset has a total of 6 columns that include

- the listing id of the given review
- the id of the review,
- the date of the review,
- the id of the reviewer,
- the reviewer's name,
- the textual part of the review

The actual rating of the listing is not available per review but is rather aggregated in a per listing manner, as inside Airbnb does not make individual ratings publicly available.

While the datasets used are publicly available and are often used in similar academic studies, they are not without limitations. The Inside Airbnb data rely on periodic web scraping and may contain missing or outdated information for certain listings. Additionally, not all listings receive reviews, and review activity may vary across hosts over time. These factors are considered in the data preprocessing stage to ensure that the analysis is based on usable and comparable observations.



Methodology

Configurations

After acquiring the data and setting up the folders (per city) that would help the python script access the files of each city, certain necessary configurations were placed. The starting year of the analysis was chosen to be 2013. Even though Airbnb was founded in 2008, in the very first years of the Airbnb platform operation both listings and reviews were sparse and therefore basing the analysis of a whole year on limited data would not satisfy the condition of having a sufficient number of reviews to be included in the analysis (inconclusive or potentially misleading results). The chosen cities for the analysis were Amsterdam, Berlin, London, Vienna, Rome, Madrid, Zurich, Stockholm, all of which are capitals of some of the strongest economies in Europe (Netherlands, Germany, England, Austria, Italy, Spain, Switzerland and Sweden respectively).

Another part of the early processing was based on the idea of creating certain output files as the analysis progressed that would serve as intermediary files (csv files) or as final results (txt files) to be read and interpreted accordingly. Given that, throughout the analysis, several files (txt, csv format) are being created within the city-specific folders that are located in the OUTPUT folder of the study.

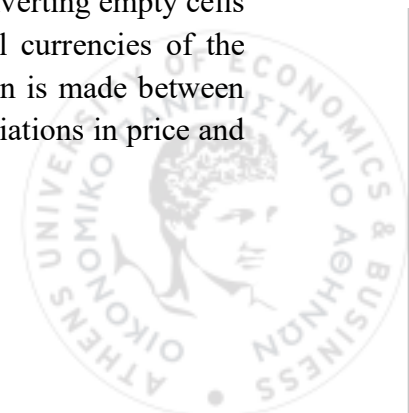
Output Management

To ensure reproducibility, transparency, and effective handling of outputs, all visualizations generated throughout the analysis were systematically saved within city-specific directories in the project's output structure. Each plot was stored in the corresponding city folder under a dedicated plots subdirectory, allowing results to be directly associated with the data and configurations used to generate them. This design choice facilitates traceability between analytical steps and their visual outputs, particularly when multiple cities and alternative model configurations are involved. In addition, this structured storage approach enables efficient referencing of supplementary figures in the appendix, where detailed visual diagnostics and meaningful checks can be presented without interrupting the flow of the main results section.

Preprocessing & Function Setup

Price Cleaning

Since the listing's price for every property could be in different currencies (euro, Swedish corona, GB pounds etc.), a function was created to read the numerical series that is located within the price cells and remove all non-numeric (respective currency symbol) characters, converting empty cells into "NaN" characterization (Not a Number). The choice not to convert all currencies of the countries into the same currency was made given the fact that no comparison is made between countries for pricing in absolute terms and the regression models focus on variations in price and sentiment within each city.



Sentiment Computation

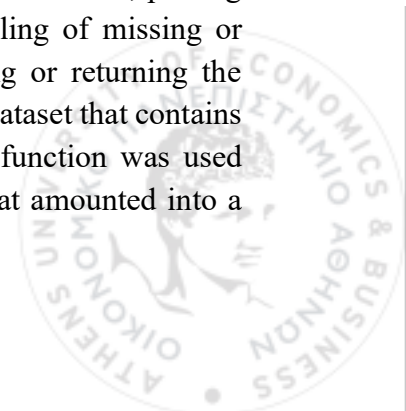
Given that Airbnb users provide numerical ratings alongside their reviews, the study seeks to assess how closely sentiment derived from review text can approximate these evaluations. To convert raw textual reviews into quantitative measures, sentiment analysis was applied to each user review. For this purpose, the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool was selected, as it is specifically designed for short and informal text commonly found in online reviews. VADER assigns a sentiment score to each review on a continuous scale ranging from -1 to $+1$, where positive values indicate positive sentiment, negative values indicate negative sentiment, and values close to zero reflect neutral or mixed sentiment (higher positive/ negative values indicate more positive/ negative sentiment). The resulting sentiment score is stored as an additional variable in the dataset and serves as a key input for the subsequent aggregation and analysis stages.

Sigmoid Calibration

To compare the sentiment score produced by VADER with the actual ratings that users provided, we needed to convert the sentiment scores from the VADER scale into the numerical rating scale used in Airbnb reviews (1 – 5). For that reason, a sigmoid function was defined that would transform the sentiment scores into a bounded scale $[1,5]$ matching the Airbnb rating system. Using a sigmoid function for this purpose benefits the study by preserving the ordering (higher rated reviews remain higher rated reviews even after the conversion) and the relative differences between the sentiment values. Since the actual ratings given by the users were at the listings level, the individual sentiment inferred scores were aggregated to match the format of the actual reviews. The best sigmoid parameters were chosen after the fitting process which also included a grid search to select a plausible value of steepness (k) and midpoint (x_0). The performance of the model was monitored by measuring the MAE (mean absolute error) between predicted and actual ratings. The model searches for multiple parameter pairs updating each time the current best pair, and once the process has finished, the best performing parameters are saved.

Accessing Macroeconomic indices

To get access to the macroeconomic indices needed for the study, we needed to retrieve annual indicators for each country to support the sentiment regression. For that purpose, the World Bank open data API was used to build a URL that specified the country code and the indicator code (macro index to be retrieved each time). The data were requested for a fixed range to match the temporal scope of the analysis, and the process included mechanisms to deal with temporary API/network instability and handle different types of errors (HTTP errors, connection errors, parsing issues) that could occur during the process. The process included the handling of missing or unusable API responses by filtering out entries where the year was missing or returning the expected DataFrames with empty values. The output of the function returns a dataset that contains a column named year and the chosen indicator values for each year. This function was used multiple times (to retrieve data for each country included in the analysis), that amounted into a



total of 24 retrievals (8 countries by 3 indicators per country). The process resulted in generating the final dictionary that comprises one Data Frame per country which consists of the year column and the indicator columns along with their respective values for each year.

Defining a regression model

To effectively examine the questions at hand:

- which are the strongest price determinants of an Airbnb listed property and,
- Is the sentiment captured in reviews associated with overall macroeconomic indicators,

We needed to define a way to measure the potential association between these variables. Running a regression is a very well-known tool that is linked to exposing associations between dependent and independent variables. The OLS regression model is chosen for its capabilities of interpreting results and suitability to work with continuous dependent variables. The OLS (Ordinary Least Squares) regression is used three times, firstly to estimate the relationship between the price (dependent variable) of Airbnb listings with the different attributes of an Airbnb listed property such as:

- accommodates (person capacity of listing),
- review rating (the average rating of the listing),
- superhost status (a binary variable that indicates whether the host is a superhost or not)
- host experience (a variable containing the years of experience each host has had on the platform)

Secondly, a separate OLS regression is used to measure the relationship between the sentiment that is captured from Airbnb reviews and the overall movement of macroeconomic indicators over time.

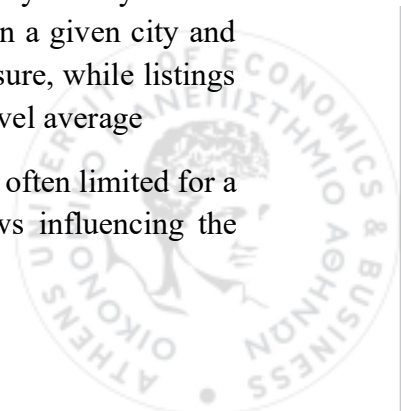
The third regression included a sentiment regression to estimate the association of extracted sentiment over time.

Upon estimation, the model produces a fitted regression output that provides access to the estimated coefficients and their associated measures of statistical significance. For each of the regressions performed a relevant text document was created to be read and interpreted accordingly.

Review sample balancing

To tackle the task of sentiment regression, and since the macroeconomic indicators are analyzed at an annual basis, there had to be an aggregation of sentiment scores into a per year-city format. This aggregation involves averaging the sentiment scores of all reviews within a given city and year. As a result, each review contributes equally to the annual sentiment measure, while listings with a larger number of reviews naturally have a larger influence on the city-level average

Adding to that the fact that in the early years of operation Airbnb reviews were often limited for a single city to sometimes being less than 100 (and therefore extreme reviews influencing the



average a lot more), and the observation that - from a certain point in time and onward – reviews start to become repetitive due to the number of users reviewing and the limited new observations that a new review can bring to be considered new information, we decided to create a balanced review sample per city. The rule decided was that every year, all cities' review samples would have to match the lowest number of reviews observed in a city. That way each city's average sentiment score per year would be calculated based on the same number of reviews, avoiding extreme cases where reviews were very few (very early years) and the later years where review numbers skyrocketed (saving processing time).

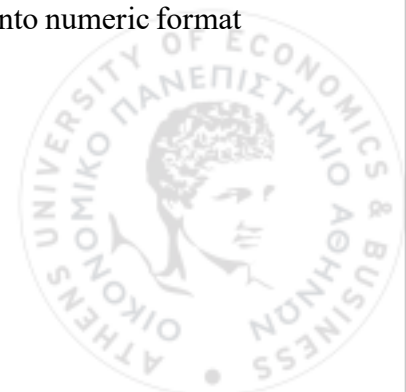
Spatial Clusters and Orientation

The study includes a section where the listings per city are separated into 4 clusters using the longitude and latitude coordinates. These clusters often exhibit common characteristics or similarities and usually receive higher or lower positive evaluations more uniformly. The specific clusters in this study were used to identify groups of areas that are associated with more positive or less positive sentiment. Since these clusters do not correspond to named or administrative areas, another plot was incorporated to improve readability and interpretability of the spatial sentiment analysis results. Each cluster was plotted onto the actual city map thus displaying the exact location that each listing is on and the wider area that each cluster is located in. This extra addition allows for more accurate inspection and interpretation of listings' position relative to the city centers and known urban features.

Main Analysis Pipeline & Price Regression

The main analysis pipeline is executed separately per city creating a dedicated output directory for each city to store the intermediate datasets and the results of the regression models.

- 1) The city listings are loaded, pre-balanced reviews are used as input to emphasize consistency across cities,
- 2) The sentiment scores are computed for each review that belongs in the balanced dataset using VADER. Review dates are parsed to extract the temporal information of each review,
- 3) Reviews are being aggregated at listing level to compute the mean sentiment per listing, which in turn is merged into the new listings dataset,
- 4) All variables intended for numeric use are being cleaned and converted into numeric format to avoid working with string objects,



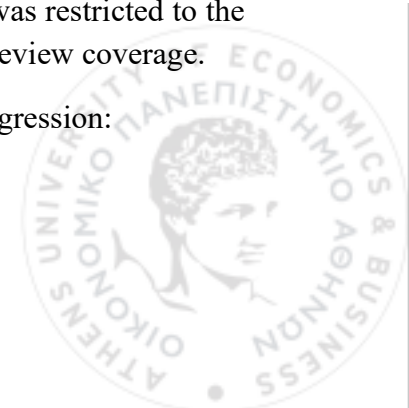
- 5) The superhost status variable is converted into a binary numeric variable to display 1 in case of a superhost and 0 when not. The host experience variable is derived from year of registration and is converted into numeric format,
- 6) Where sufficient paired sentiment – rating observations are available, the sigmoid function defined earlier is applied and the optimal parameters are selected minimizing the MAE between sentiment derived ratings and actual ratings. The transformed sentiment values are stored alongside the raw sentiment scores.
- 7) Listings are clustered into 4 spatial areas using the geographic coordinates (latitude and longitude) and are analyzed to define relative city areas with certain characteristics. Clusters were also converted into dummy variables to be used in the price regression and examine the potential association between the price and each cluster,
- 8) An OLS regression model is estimated to explain the listing prices against the characteristics of an Airbnb listed property. The regression summary results are saved in a text a document for further interpretation.
- 9) A separate OLS regression is used to examine sentiment difference over time to capture long term sentiment trends within each city,
- 10) A final output file is created that contains aggregated data related to the yearly sentiment, merged with the country-level macroeconomic indicators to be used in the analysis that follows and explore any potential associations between sentiment and broader economic conditions.

All intermediate datasets, regression summaries and calibrated parameters were exported in the corresponding files to ensure reproducibility and to be used as reference points.

City- year aggregation & Sentiment regression

To facilitate the last research question of the study: Is the overall captured sentiment associated with movements of macroeconomic indicators, the city-level outputs produced in the previous stage were further aggregated into a pooled city–year dataset that showcases one observation per city-year. That way each row represents the average sentiment, balanced number of reviews, macroeconomic indicators and the city label for a specific year. The analysis was restricted to the period that ranges from 2013 and onwards to exclude early years with sparse review coverage.

The image below showcases the dataset that was produced to be used in the regression:



Year	avg_sentiment	n_reviews	gdp_pc	inflation	unemployment	city
2013	0.835805294	170	52602.29	2.506899	7.241	amsterdam
2014	0.739322862	573	53457.22	0.976035	7.419	amsterdam
2015	0.707732613	1527	45793.81	0.600248	6.872	amsterdam
2016	0.714191744	2689	46808.53	0.316667	6.007	amsterdam
2017	0.681489476	4162	49513.68	1.381459	4.837	amsterdam
2018	0.685446371	6821	53955.08	1.703498	3.832	amsterdam
2019	0.694733724	9130	53554.9	2.633699	3.379	amsterdam
2020	0.575949344	3810	53467.93	1.27246	3.82	amsterdam
2021	0.552074231	5654	60141.99	2.67572	4.209	amsterdam
2022	0.617850377	12333	59123.32	10.00121	3.526	amsterdam
2023	0.579060858	18336	63515.6	3.838394	3.537	amsterdam
2024	0.515833756	12075	67520.42	3.347543	3.599	amsterdam

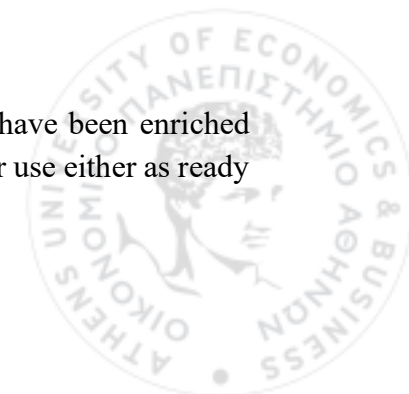
Figure 1: City-year aggregated dataset

Using this aggregated city–year dataset, a pooled Ordinary Least Squares (OLS) regression was estimated to examine the association between average yearly sentiment and macroeconomic conditions. Average city-year sentiment was specified as the dependent variable, while macroeconomic indicators served as the main explanatory variables. City fixed effects were included to control for time-invariant differences across cities, such as structural characteristics or differences in cultural factors. Heteroskedasticity-robust standard errors (HC1) were employed to improve the reliability of statistical inference. Finally, a regression summary was exported for further interpretation and conclusions.

The pooled regression framework was chosen primarily due to data limitations and the structure of the available information. Macroeconomic indicators are observed at an annual frequency, which results in a relatively small number of observations per city, when sentiment is aggregated at the city–year level (12 observations per city ranging from 2013 to 2024). Estimating separate regressions for each city would therefore lead to insufficient statistical power and unreliable coefficient estimates. Pooling observations across cities allows the analysis to further examine the cross-sectional variation but also the temporal differences in the data, while the inclusion of city fixed effects ensures that comparisons are made relative to each city’s own baseline sentiment level.

Main execution

After the analysis has finished, the dedicated city folders in the output path have been enriched with the necessary files. Below is a list of all output folders and their particular use either as ready to read results or as intermediate files:



1. **Listings_model_data** is similar to the initial listings data but has incorporated the sentiment scores generated by VADER and the sentiment scores transformed to the Airbnb rating scale [1,5]. The format of the variables has been converted to numeric where appropriate and a new column containing the cluster label of each listing has been added.
2. **ols_price** is the resulting summary of the price regression with the independent variables (accommodates, review scores rating, host is superhost, host experience),
3. **ols_price_coefficients** file contains the saved coefficients of the price regression with the listings' attributes and is used to plot the price determinants for each city,
4. **ols_sentiment_year** is the summary of the sentiment regression over time showing any links between sentiment and the passage of time,
5. **Reviews_balanced** file contains the reviews per year that matched the lowest number of reviews for a city. All subsequent sentiment computations and regressions were performed using the balanced review samples rather than the raw review data,
6. **Reviews_macros_year** served as an intermediate dataset that stored mean sentiment per city and macroeconomic indicators per year which was then subsequently aggregated to another dataset (**ols_sentiment_macros_pooled_city_fe**) that was in the correct format to be used for the sentiment regression,
7. **Sentiment_caibration** stored the best performing parameters for the sigmoid function and the lowest MAE that was achieved.



Results and Discussion

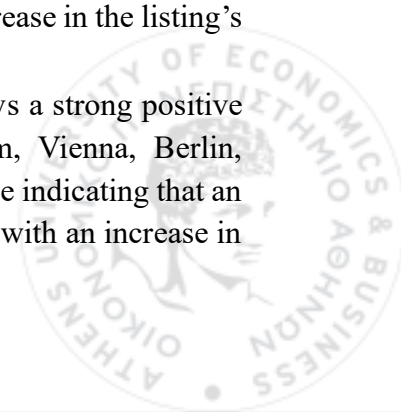
This section presents findings from the major research questions that were posed and discusses the interpretation of the findings. The results are structured around two main components. First, the determinants of Airbnb listing prices are examined at a city level using Ordinary Least Squares regression, focusing on listing characteristics and host attributes. Secondly, as part of the sentiment analysis, several aspects of sentiment are explored (spatial cluster and sentiment, sentiment dispersion, sentiment vs actual ratings and the relationship between the average sentiment per year and city and macroeconomic conditions). Rather than reporting results exhaustively for each city, the discussion section emphasizes general patterns that were identified across cities, highlighting individual cases only when they illustrated meaningful differences or exceptions.

Price regression

The price regression is used to help us understand the way a property's price is associated with unit changes in one of the independent variables. A common pattern observed in all regression summaries is that the explanatory power of the model regarding the variation in listing prices is rather on the lower end. That is to be expected as the price is influenced by many different factors that are not used in this example. As we saw in the data description section, the initial listings dataset consists of ~ 90.000 listings per city and 79 columns, many of which can influence the price of an Airbnb listing. Additionally, there are many other factors that are not associated with the characteristics and attributes of an Airbnb listing that have an immediate effect on the price such as seasonal effects, specific events and safety index of an area. The model with the highest explanatory power is the one used for the city of Amsterdam where it is able to explain 16.7% of the variation in price by the variables used in the regression (**accommodates, ratings, superhost status, host experience**). On the other hand, the lowest achieved explanatory power comes from the model used for the city of Berlin where the R-squared value falls below 1% at 0.6%. The rest of the models' explanatory power is around 1-4%, making Amsterdam stand out as the exception. Amsterdam's pricing structure seems to be more standardized as many listings follow similar pricing strategies whereas for the rest of the cities it can be safely said that the price formation is fragmented and is mostly explained by other parameters such as postcode, seasonality, apartment view, events, and others that are not factored in the regression.

Taking a close look at the individual parameters used in the regression:

- The **accommodates** parameter shows a consistently strong positive association with the price across all cities. Meaning that a unit increase in the accommodation variable (from a number of guests to an addition of an extra guest) is associated with increase in the listing's price (very plausible assumption).
- The **review rating** variable is mostly consistent across cities and shows a strong positive association of the rating of a property with the price. Amsterdam, Vienna, Berlin, Stockholm and Zurich exhibit a positive coefficient of the rating variable indicating that an increase in the rating unit (from a rating of 4 to 5) would be associated with an increase in



the property price, noting though that the unit increase is a substantially large increase (one integer unit in a scale of five). Exceptions to this rule are London, Madrid and Rome that showcase a small negative association of the review rating variable with the listing's price. This shows that a potential improvement in the reputation of a property in these cities is not directly associated with higher prices. A possible explanation for this is that for the cities mentioned, the baseline ratings might already be on the upper scale. Another possible interpretation of this result is that property owners prefer to be competitive and seek to be consistently fully booked rather than achieving short term price maximization.

- The **Superhost Status** variable shows mostly heterogeneous effects across cities, that depend on the context of each city. The pattern observed shows a positive price association with the superhost status in Berlin, Madrid, Rome, Zurich, a negative price association with the superhost status in Amsterdam and Stockholm and a negligible price association for the cities of Vienna and London. The superhost status does not reflect only the quality of an Airbnb listing but the strategy as well. A property owner might want to achieve the goal of being a superhost by being competitive in pricing and thus not charging a lot for what they offer (while also achieving high average ratings), while another property owner might want to provide the best and most luxurious experience. These are two completely different approaches and yet can lead to the same outcome of reaching a superhost status. The superhost status is achieved by maintaining a highly positive overall rating (4.8/5.0 or above) and by being consistent with several other factors (response rate, keeping low cancelation rate, achieving minimum of booked trips). This on its own not indicate setting a specific price range.
- The **Host Experience** variable expectedly shows a weak association with price that is showcased by the very small coefficients of this variable. It is observed that host experience is rarely dominant in any city, it is sometimes positive and other times negative but with no real certainty that a unit increase in the experience of a host would be associated with a price change.



Out of all cities, London stands out as a distinct case, where listing prices appear to be driven primarily by accommodation capacity rather than reputation, reflecting the city's high demand and severe space constraints. Stockholm displays unusually large coefficient magnitudes (for all variables), pointing to a more segmented pricing structure in which listing characteristics translate into substantial price differences (extremely large coefficients). Bar plots of coefficients for both cities are seen below:

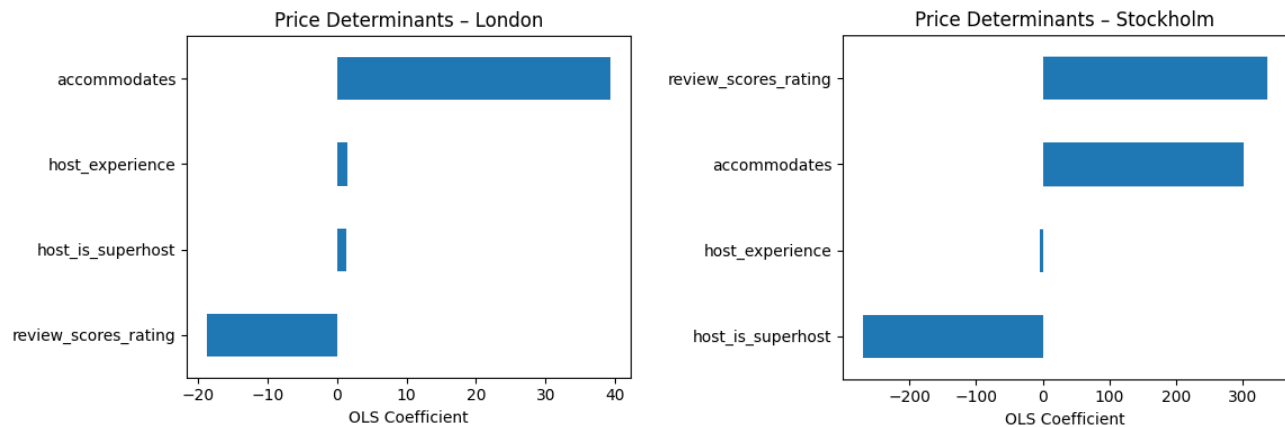
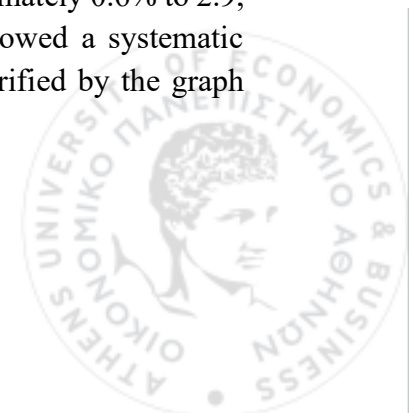


Figure 2: Price Determinants
London & Stockholm

Overall, using a variable that is known to be a price driver (accommodates variable) and a variable that is known to have insignificant effect on the prices (host experience variable) helps us confirm the accuracy with which the regression model operates. The review rating and superhost status variables are open to interpretation and are often times variables that are affected by the strategy of an Airbnb owner and the local market conditions and thus can vary from city to city and culture to culture.

Sentiment over time Regression

The regression results (that is used to examine the evolution of sentiment over time) provide quantitative confirmation of the patterns observed in the graphical analysis of average sentiment across cities. Across all cities, sentiment association with time is negative and statistically significant, indicating a gradual decline in average review sentiment over the study period. That doesn't translate to causal effects (that the sentiment is correlated to the passage of time) but to a broader claim that the sentiment is associated negatively with time. Although the explanatory power of these regressions is low, with R-squared values ranging from approximately 0.6% to 2.9, the consistently significant time coefficients suggest that sentiment has followed a systematic downward trend rather than fluctuating randomly over time. This can be verified by the graph below:



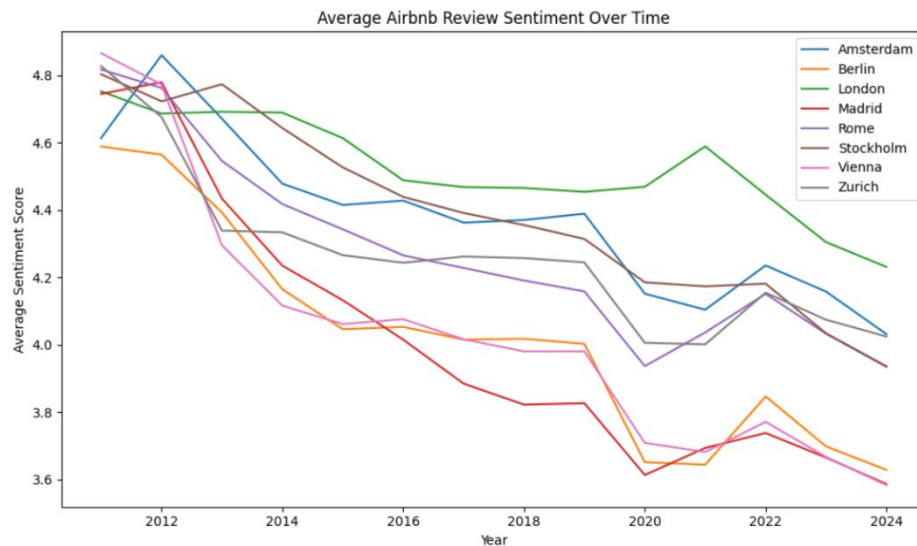


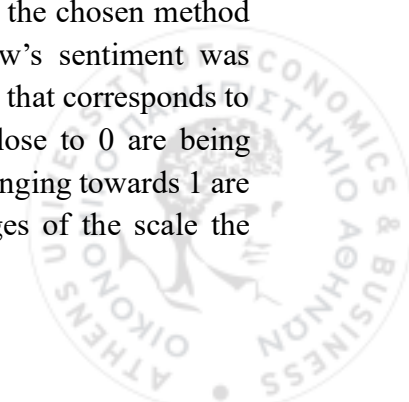
Figure 3: Sentiment over time for all cities

The declining trend observed in the regression summary is visually reinforced by the time-series plot, where most cities display a gradual reduction in average sentiment scores from the early years of the platform toward more recent periods. In 2019 approximately, the review sentiment drops significantly in every city except London. Since the Covid pandemic had already started in China and several traveling restrictions were placed, the composition of Airbnb users at that time might consist mainly of natives. Natives can be stricter when choosing to use the Airbnb platform in their home country thus affecting the overall sentiment decrease that was observed during that time. After 2020 and 2021 an increase or stabilization is observed that reaches a high point in 2022 before decreasing again at the start of 2022 and towards 2024.

It is important to note that the relatively low R-squared values should not be interpreted as a weakness of the analysis. Instead, they reflect the fact that year-to-year changes alone explain only a small fraction of overall sentiment variation, which is primarily driven by listing-level experiences and individual guest perceptions. Nevertheless, the statistical significance and consistent direction of the time coefficients indicate that sentiment on the Airbnb platform has evolved over time in a meaningful way. Taken together, the regression results and the visual evidence confirm that average sentiment is negatively associated with time.

Sentiment inferred ratings against actual ratings

A major part of this study was centered around the question of the approximation of sentiment inferred ratings when compared with user given ratings. As stated previously the chosen method of sentiment extraction was the VADER analysis tool, where each review's sentiment was calculated based on the content of the text and was assigned a numerical value that corresponds to the context of the review (positive or negative). Values ranging from -1 close to 0 are being considered as negative sentiment whereas positive values greater than 0 and ranging towards 1 are considered positive sentiment. The closer the sentiment score is to the edges of the scale the



stronger the positive/ negative sentiment is. These values were then converted using a sigmoid function to replicate the ratings given on the Airbnb platform that range from [1,5]. Given the limitations of this conversion, it is understood that sentiment scores get compressed and distinguishability is reduced. As a result, the moderately positive/ negative reviews and strongly positive/ negative reviews may end up with very similar mapped scores. Given that in the Airbnb reviews structure, most of the reviews are positive, the model may fail to differentiate good from exceptional sentiment.

A common pattern observed in all cities is that the actual ratings cluster heavily between 4.5 and 5.0 scores while fewer listings fall below the 4.0 rating mark. This reflects the well-known rating inflation phenomenon that appears in these types of platforms. Since the mean rating is computed at the listing level, and that the majority of textual reviews are overwhelmingly positive, the overall sentiment is positively skewed while the negative sentiment is diluted. Sentiment-inferred ratings are even more concentrated near the upper bound as positive reviews are being compressed towards the upper level. An example is shown below:

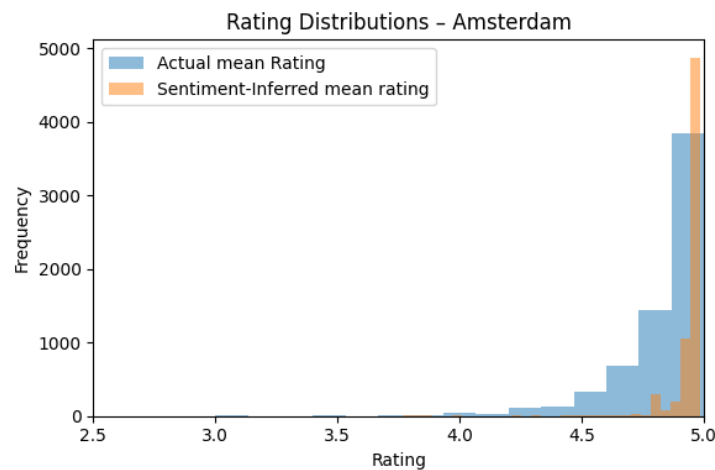
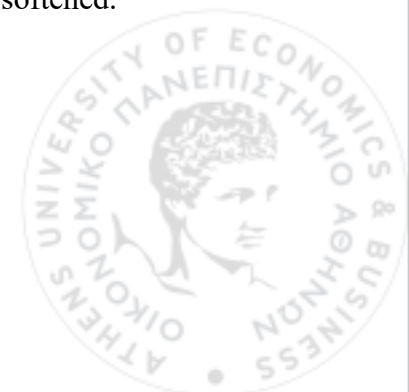


Figure 4: Rating Distributions - Amsterdam

The actual average ratings per listing for Amsterdam are more uniformly spread starting from the 4.5 rating mark and onwards. On the other hand, sentiment inferred ratings are heavily concentrated at the top showcasing some limitations that both the sentiment extraction model and the rating conversion have. The minimal dispersion that is shown in the sentiment inferred ratings can be attributed to (other than the fact that positive reviews get squeezed at the upper scale of the rating) the fact that even negative review text can often be politely phrased not capturing the true sentiment of the user and that negative experiences might be underreported or softened.



One case where the model accurately estimates the frequency of the strongly positive ratings is Zurich as shown below:

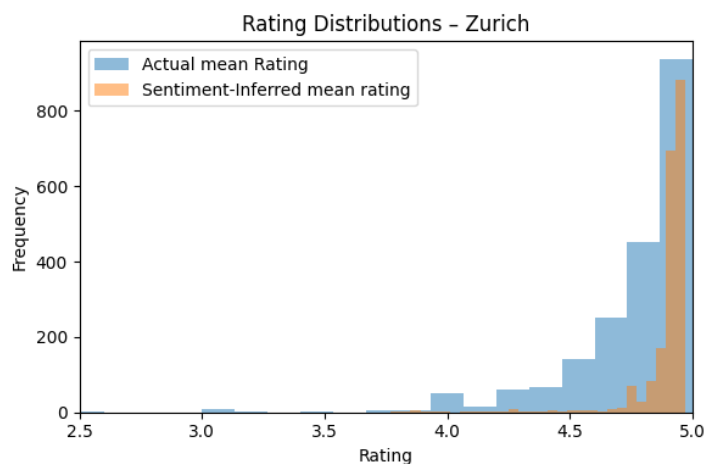


Figure 5: Rating Distributions - Zurich

The same pattern is observed for ratings below the 5.0 mark, where the model underestimates the number of listings that had an average review of 4.5 to approximately 4.8. In the case of the 5 starred reviews, the model is roughly estimating correctly the number of listings that received an average rating of 4.8 or greater.

Measuring the accuracy of the Sentiment Inferred Rating

A good method to measure the accuracy of the sentiment scores is to compare them to the actual ratings in absolute terms. Doing that at the city level means that each city’s average listing sentiment score is being aggregated by computing the average sentiment score of the city. Comparing that metric to the aggregated average rating score of each city produces the mean absolute error of the listings at city level as shown in the graph below:

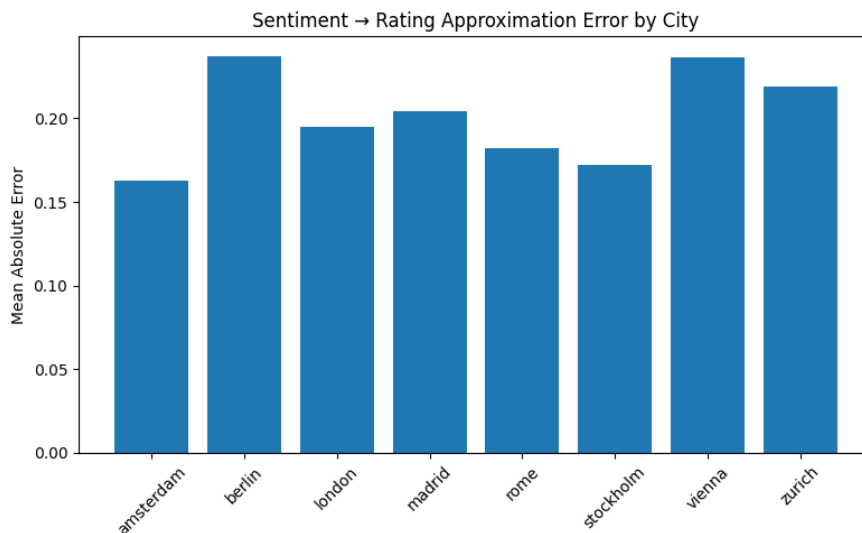


Figure 5: MAE by city at listing level



It is evident that Amsterdam exhibits the lowest mean absolute error (MAE), indicating that, on average, the sentiment-inferred rating deviates from the actual numerical rating by approximately 0.16 units on the five-point rating scale. The highest MAE values are observed in Berlin and Vienna, suggesting a weaker correspondence between textual sentiment and numerical ratings in these cities. This does not imply model failure, but rather reflects differences in how reviewers express sentiment relative to the ratings they assign, potentially due to linguistic, cultural, or stylistic factors. Overall, the relatively narrow range of MAE values across all cities indicates that sentiment-based rating inference performs consistently across markets.

Sentiment interpretability

While aggregate sentiment scores provide a useful quantitative summary of review text, they offer limited insight into the specific linguistic cues driving these evaluations. To enhance interpretability, in this section we examine the contribution of individual words, that are referred to as tokens, to the overall sentiment score that is assigned by the VADER model. By identifying the most influential positive and negative tokens (words) within the balanced review samples for each city, the analysis provides a qualitative perspective on the aspects of the guest experience that most strongly shape sentiment. For our initial approach, VADER computed the top 15 most positive and negative tokens using the valence contribution across all cities. This method, produces the figures below:

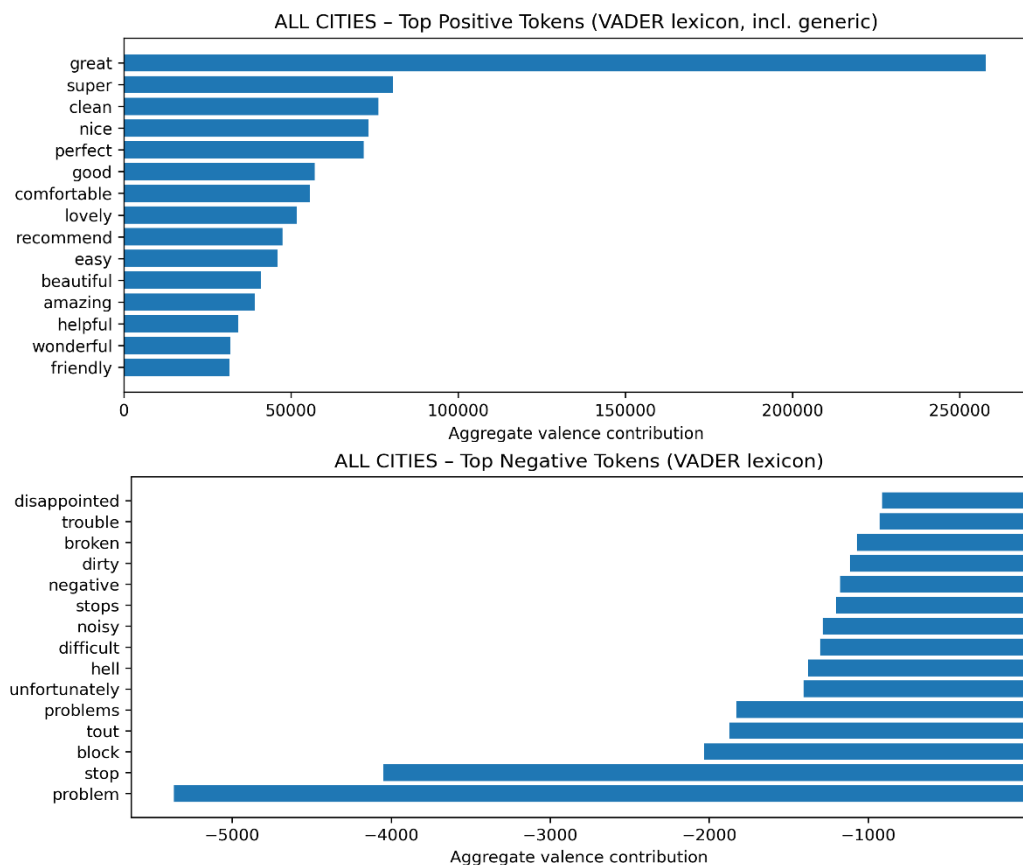
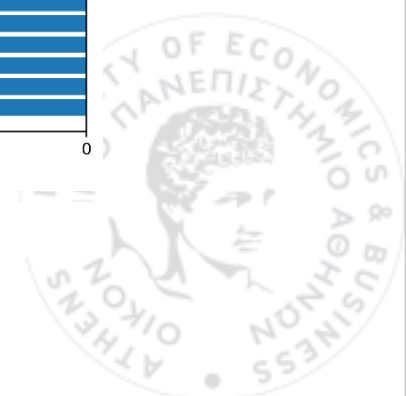


Figure 6: Top Positive/Negative tokens across cities - Initial approach



It is evident that the resulting output is dominated by broadly evaluative adjectives such as great, good, perfect, and amazing on the positive side, and terms like problem, stop, and disappointed on the negative side. This reflects their high frequency usage and strong predefined sentiment weights. While this initial analysis confirmed the overwhelmingly positive tone of Airbnb reviews (as it can be seen by the aggregate valence contribution of each figure) and provided a useful validation of the sentiment extraction process, it offered limited interpretive value with respect to the underlying drivers of guest satisfaction or dissatisfaction. These generic sentiment descriptors convey overall emotional polarity but lack specificity when describing the aspects of the accommodation experience. As a result, although the unfiltered token analysis was informative in demonstrating sentiment skewness and review positivity bias, its lack of explanatory power made it in need of a refinement step that would improve its interpretability.

This step serves to validate the sentiment methodology by ensuring that the computed scores are grounded in meaningful and contextually relevant language, rather than driven by generic or uninformative expressions such as (good, bad, wrong, excellent, miserable, delightful etc.). For that reason, a dictionary of words was created to remove most common positively or negatively bearing words and adjectives that would not be of additional value and are inherently biased.

For that purpose, the analysis performed was based on the balanced review dataset to avoid token counts being dominated by cities with more reviews and ensure comparability across cities. The process included steps like lowercasing the reviews, tokenizing the words and then matching the tokens against VADER's lexicon that assigned a predefined valence score to each token. After filtering for generic words and stopwords (words with length ≤ 3), for each token, an aggregate contribution score was computed by summing the lexicon valence across all occurrences, allowing frequently recurring sentiment cues to be identified.

The output of this process is shown below in the bar plots where the aggregated top 15 positive and negative tokens are displayed for all cities:



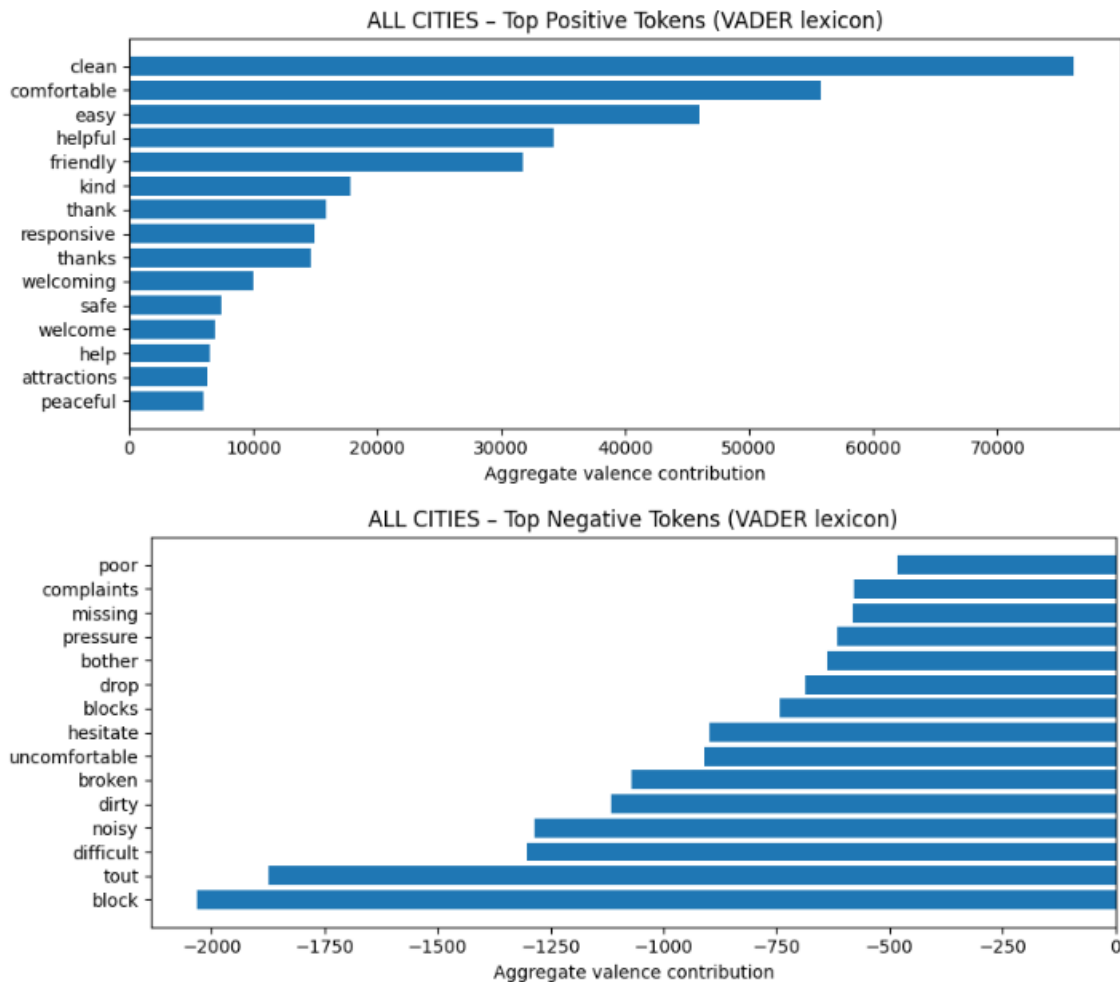
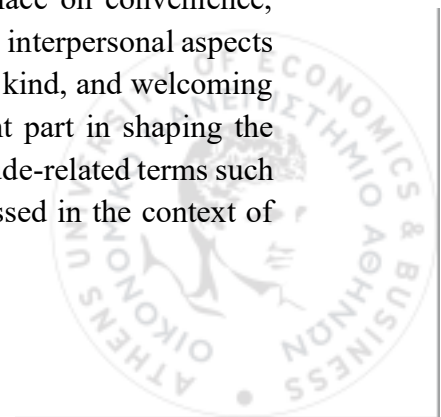


Figure 7: Top 15 Positive/Negative words aggregated for all cities

Inspecting the aggregated token-level results across all cities, a clear and intuitive pattern emerges regarding the aspects of the Airbnb experience that most strongly contribute to positive and negative sentiment. The most influential positive tokens are dominated by words such as clean, comfortable, easy, helpful, friendly, and responsive, all of which are directly related to core dimensions of service quality and guest comfort. The word clean being the strongest positive contributor highlights the central importance of cleanliness in shaping guest satisfaction, a finding that is consistent in both common expectations in hospitality services and prior empirical research. Similarly, words such as comfortable and easy reflect the value guests place on convenience, potential ease of check-in, and overall usability of the accommodation, while interpersonal aspects of the hosting experience are captured through tokens like helpful, friendly, kind, and welcoming noting that the interaction with the Airbnb host always plays an important part in shaping the overall level of satisfaction of the guests. Additionally, the presence of gratitude-related terms such as thank and thanks further suggests that positive sentiment is often expressed in the context of



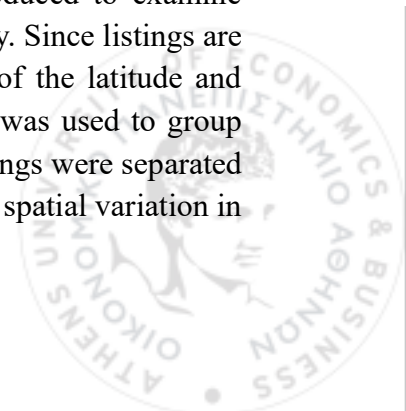
appreciation for host responsiveness and support, rather than solely for physical property attributes. The token safe is also important to keep in mind as it is directly related to the geographical location of the listing and the importance of a guest feeling safe in the area that they live in.

On the negative side, the most influential tokens are strongly associated with issues and disruptions to the guest experience. Words such as dirty, broken, missing, and uncomfortable point to deficiencies in maintenance and cleanliness, directly mirroring the positive emphasis placed on these factors. Tokens like difficult, pressure, and bother indicate environmental and situational stressors that detract from the perceived quality of the stay, while terms such as complaints and poor reflect broader dissatisfaction that is probably related to the host. Additionally, some negative tokens such as noisy, block, tout, and drop may be linked to specific area-related problems such as illegal activities and/ or the feeling of being unsafe that is important for guests. Also, it is important to keep in mind the fact that the geographical location of an Airbnb and the broader area it is located in plays a very important role in opinion forming. The analysis indicates that both positive and negative sentiment largely refer to the same underlying aspects of the accommodation experience, such as cleanliness, comfort, and location. However, while positive sentiment is typically expressed in broader and more general terms, negative sentiment tends to focus on specific problems or inconveniences.

Overall, the results help support the validity of the sentiment analysis approach used in the study. The fact that the sentiment scores are influenced by clear and experience-related words, such as those referring to cleanliness, comfort, location, noise and similar problems, suggests that the scores reflect meaningful aspects of the guest experience rather than random linguistic noise. At the same time, the strong presence of positive terms and the limited contribution of negative ones aligns with earlier findings showing that Airbnb reviews are generally very positive in tone. This helps explain why the sentiment-inferred ratings tend to be concentrated toward the upper end of the scale and show less variation compared to the actual ratings. While this highlights a limitation of sentiment-based measures in environments where reviews are highly skewed toward positive feedback, the consistency of these patterns across cities indicates that guests tend to use similar language when evaluating Airbnb stays. As a result, the token-level analysis provides additional confidence that sentiment analysis captures relevant and interpretable signals related to user satisfaction.

Sentiment Cluster Segmentation

Following the token interpretability section, a spatial component was introduced to examine whether guest sentiment differs systematically between the clusters of each city. Since listings are geographically distributed across areas - which is evident by the existence of the latitude and longitude variables within the initial listings DataFrame -, spatial clustering was used to group listings into broad city areas based on those coordinates. For each city, the listings were separated into four clusters using KMeans,. The resulting graphs provide an overview of spatial variation in



sentiment and allow identification of broad areas associated with relatively more positive or less positive guest evaluations.

Three plots were produced for each city, a scatter plot (figure 8) where each dot represents a listing and has colored clusters accordingly, a bar chart (figure 9) with the mean sentiment score per cluster kept in its original form (-1,+1 scale as it helped with identifying differences in sentiment between clusters) and the same scatterplot (figure 10) of listings placed on the map of the corresponding city.

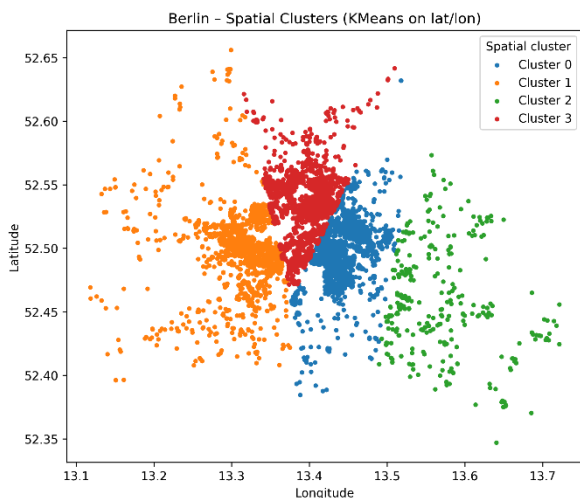


Figure 8: Scatterplot of listing clusters - Berlin

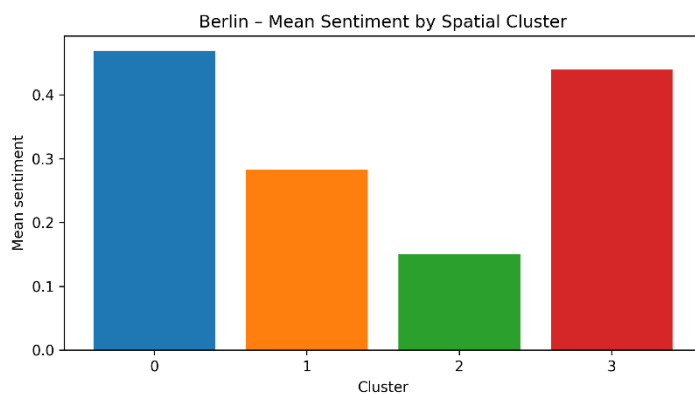


Figure 9: Bar plot of average sentiment per cluster - Berlin

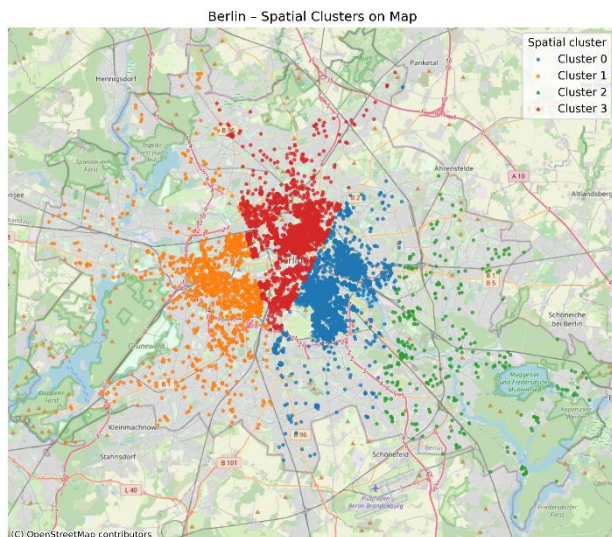


Figure 10: Spatial clusters on map - Berlin

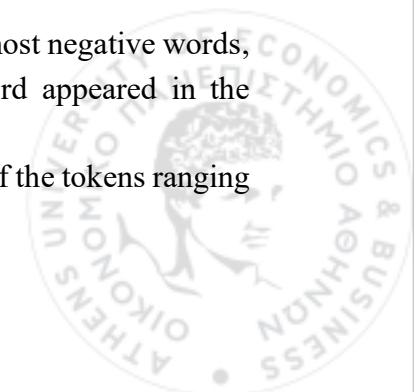


For the example above, the reference city is Berlin as it had the highest observed differences in mean sentiment sentiment between the clusters. Figure number 8 displays the 4 clusters that were created for the listings of Berlin coloring each listing according to the cluster it belonged in. While cluster 0, 1 and 3 appear to be near the city center but more concentrated with few listings on the outskirts of each cluster, cluster 2 is more sparse showing that it included listings from a broader geographical area. That can also be verified by figure 10 where it is evident that cluster 2 of Berlin consists of Airbnb listings that are primarily not located within the central area but are rather concentrated on the wider eastern suburbs of Berlin. That cluster has the lowest average positive sentiment which potentially can be explained by looking at the listings' relative distance from the city center. By examining dispersion (std) among the clusters, the question that arises is whether this is driven by consistently lower positive evaluations or by a mix of very positive and very negative experiences. To explain this observation, a specific process was set up to compute the top 15 negative words specifically for cluster 3 of Berlin to potentially identify tokens that can explain this disparity.

token	count	mean_valence	impact
block	71	-1.9	134.9
tout	207	-0.5	103.5
broken	45	-2.1	94.5
dirty	48	-1.9	91.2
noisy	111	-0.7	77.7
hesitate	64	-1.1	70.4
uncomfortable	34	-1.6	54.4
missing	37	-1.2	44.4
pressure	37	-1.2	44.4
bother	29	-1.4	40.6
complain	24	-1.5	36.0
blocks	39	-0.9	35.1
lack	25	-1.3	32.5
doubt	21	-1.5	31.5
drop	27	-1.1	29.7

Figure 11: Top negative words for Berlin cluster 3

- The token column contains the token-words that appear in the top 15 most negative words,
- The second column (count) displays the number of times each word appeared in the reviews,
- The third column (mean_valence) portrays the valence score for each of the tokens ranging from a scale of -4 to +4,



- The impact column displays the overall impact that the word has on sentiment by computing the product of the valence score and the occurrences of each word.

In this case, the most frequently occurring negative words include word such as block, tout, broken, dirty, noisy, hesitate, and uncomfortable. These terms can broadly be grouped into two categories. The first category relates to property related issues, such as cleanliness and maintenance problems, as reflected by words like dirty and broken. The second category is associated with location- or environment-related concerns, including noisy, block, and tout, which may point to neighborhood conditions or external disturbances affecting the guest experience. Overall, these patterns are consistent with the aggregated negative tokens identified earlier in the study, suggesting that lower sentiment in this cluster is driven by similar types of issues. As a result, the cluster-level analysis reinforces the broader findings without introducing substantially new sources of negative sentiment.

Another example worth showcasing is the uniformity in sentiment with which the clusters of the city of London are formed.

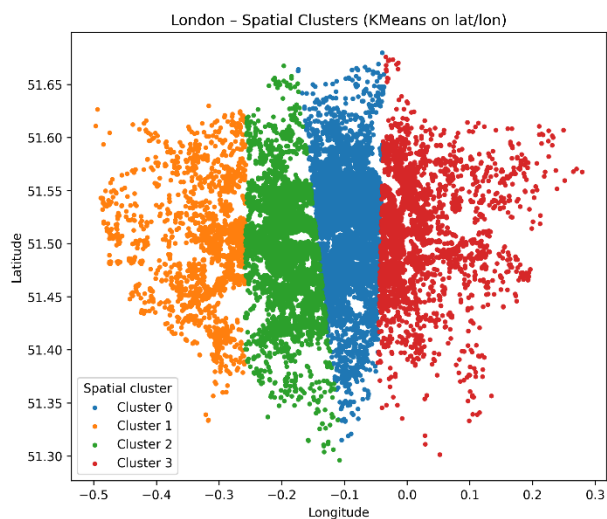


Figure 12: Scatterplot of listings clusters - London

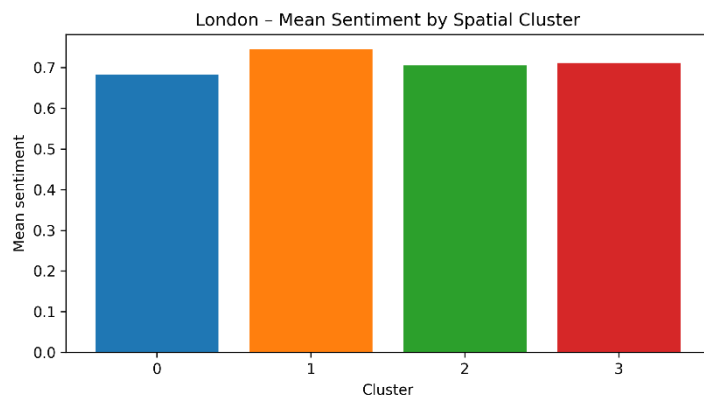


Figure 13: Barplot of average Sentiment per cluster - London

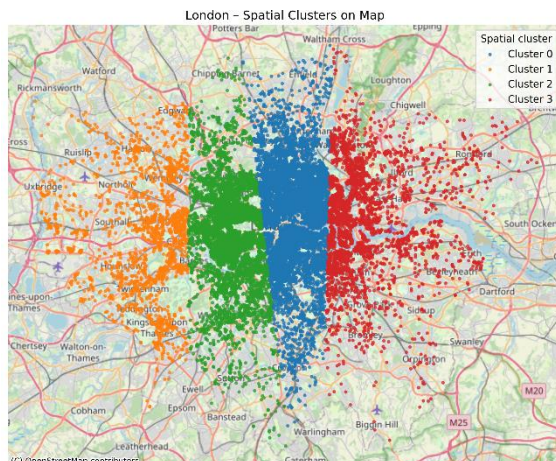


Figure 14: Spatial clusters on map London



The figures above showcase that the mean sentiment per spatial cluster for London is very positive and uniform across clusters. The clusters are placed from the western side (orange cluster), to central western (green cluster) to central area (deep blue) and to central eastern area (red). It seems that guests on average are very pleased regardless of the location of the cluster which might indicate consistent quality in the Airbnb of London for the entire region.

London clusters further analysis

In the spatial analysis, a uniform number of clusters ($K = 4$) was applied consistently across all cities in order to maintain comparability and consistency. This choice was motivated by the objective of identifying broad, city-level spatial patterns rather than fine-grained neighborhood structures. However, London represents a distinct case within the dataset due to the exceptionally high density and large volume of Airbnb listings it contains. As a result, an additional exploratory clustering exercise was conducted specifically for London, allowing for a higher number of clusters and evaluating their suitability using quantitative validation metrics.

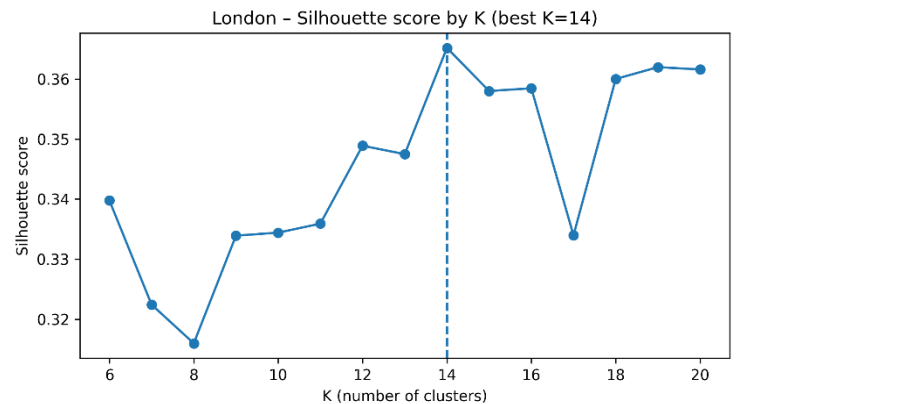


Figure 15: Silhouette score for no. of clusters in London

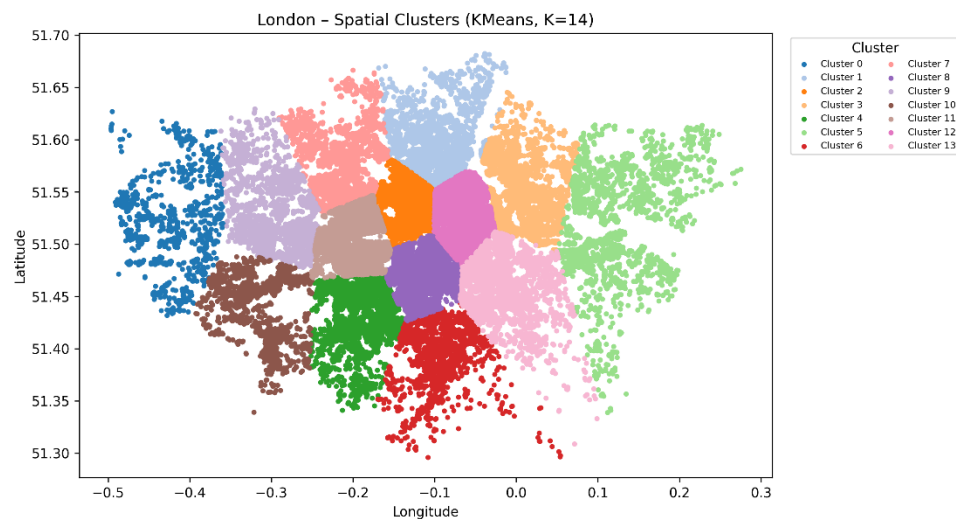


Figure 16: Clustering using $K = 14$ in London



Figure 15 presents the results of this extended clustering analysis for London, using the silhouette-scores as a selection method. The upper panel illustrates how the silhouette score evolves as the number of clusters increases, while the lower panel shows the spatial distribution of listings when the optimal configuration ($K = 14$) is applied. Rather than letting the model select the globally optimal silhouette score—which typically favors very small values of K —the analysis deliberately focuses on values of K greater than five. The restriction reflects the desire to pursue the objective of uncovering meaningful within-city heterogeneity. Within the restricted range, the silhouette score reaches a clear local maximum at $K = 14$, indicating the most favorable balance between cluster cohesion and separation among the tested configurations.

The spatial scatterplot corresponding to the $K = 14$ configuration further supports this modeling choice by revealing a structured and coherent partition of listings across London. The clusters form geographically compact regions that broadly align with spatial zones, indicating that the algorithm is capturing meaningful spatial organization rather than noise. In contrast, lower- K solutions would forcibly aggregate areas with distinct characteristics into the same cluster, limiting the ability to detect localized variation in sentiment and review behavior. This richer clustering structure provides a more appropriate foundation for examining within-city sentiment patterns in a markets as large and heterogeneous as London.

Sentiment dispersion

To further characterize how the review sentiment differs across and within cities, we examined both the average sentiment score and the dispersion of reviews using the balanced review samples. While the mean sentiment provides information on the overall tone of user evaluations, dispersion measures how consistent or varied these evaluations are within each city. Summary statistics were reported using both central tendency (mean) and variability measures (standard deviation), allowing for a more complete description of sentiment patterns. In addition, sentiment dispersion is analyzed over time (figures 18, 19, 20) at the city level to assess whether the variability of guest evaluations remains stable or changes throughout the study period. The table below gives us a first perspective in how the mean sentiment is formed both in the VADER (-1,1) and the common Airbnb scale (1,5).

	city	mean_sentiment	std_sentiment	mean_sentiment_0to5	std_sentiment_0to5	n_reviews
6	vienna	0.391897	0.606060	3.479742	1.515150	77280
1	berlin	0.405067	0.605557	3.512667	1.513894	77280
7	zurich	0.564675	0.508607	3.911687	1.271517	77280
3	madrid	0.367990	0.496225	3.419975	1.240562	77280
4	rome	0.544765	0.488079	3.861913	1.220198	77280
5	stockholm	0.585381	0.486927	3.963454	1.217317	77280
0	amsterdam	0.610814	0.478938	4.027035	1.197344	77280
2	london	0.700884	0.404123	4.252209	1.010308	77280

Figure 17: Sentiment dispersion across cities



Apart from reporting the mean sentiment and standard deviation of each city in VADER values (columns 2 and 3), the mean sentiment and standard deviation were also computed in the Airbnb bounded scale (columns 4 and 5). The highest mean sentiment is achieved in London both in raw sentiment score and in the Airbnb scale (to be expected as the sigmoid function retains the order of the scores) reflecting generally the very positive tone in reviews for the city. In contrast Vienna exhibits the lowest mean sentiment, with an average raw sentiment score of approximately 0.39, corresponding to a mean value of 3.48 on the Airbnb scale.

Examining dispersion, cities such as Vienna and Berlin display relatively higher standard deviation values compared to others, indicating greater heterogeneity in review sentiment. In contrast, cities like London exhibit lower dispersion, suggesting more uniform evaluations across reviews. Overall, standard deviation values range between approximately 0.40 and 0.60 on the raw sentiment scale, implying that while reviews are predominantly positive across all cities, there remains a non-negligible degree of variation in how guests express their experiences. This is not a strange nor a new discovered pattern, as generally speaking, reviewing systems tend to incorporate all different kinds of guests, and observing dispersion within reviews is reasonable.

Below is a graph displaying the mean sentiment dispersion over time for 3 of the cities that are included in the study (Madrid, London and Berlin):

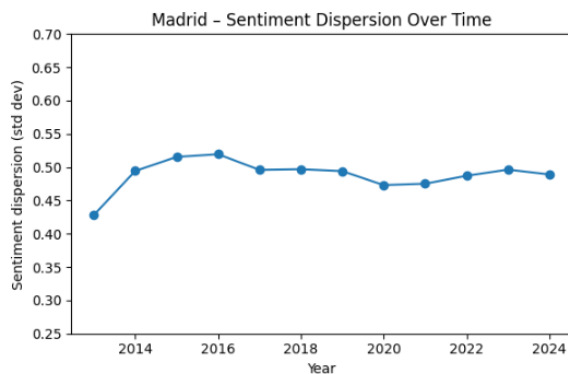


Figure 18 : Sentiment dispersion over time - Madrid

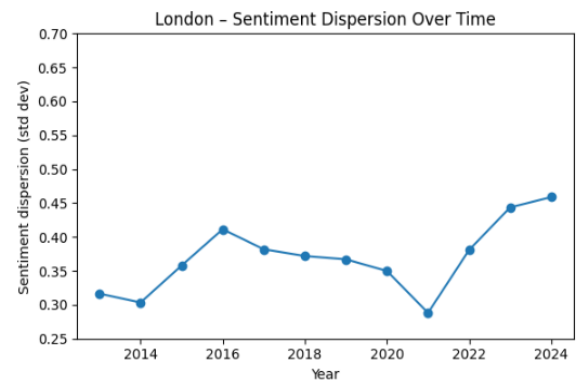


Figure 19: Sentiment dispersion over time - London

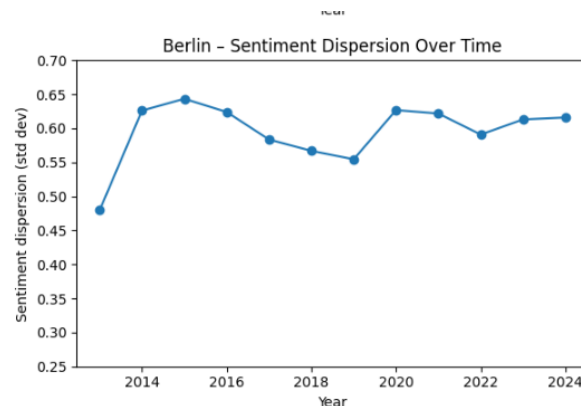


Figure 20: Sentiment dispersion over time - Berlin



Examining the mean sentiment dispersion over time for these cities reveals distinct patterns in how guests evaluate their Airbnb experiences. Madrid exhibits a notably stable pattern of sentiment dispersion. After a modest increase in the early years, dispersion remains largely constant, indicating a relatively steady level of variation in review sentiment and suggesting that the diversity of guest experiences has not changed substantially over the study period. Berlin represents a case of persistently high and relatively volatile dispersion, with standard deviation values remaining close to the upper end of the observed range throughout the period. Although dispersion fluctuates moderately across years, it consistently remains high, suggesting that guest experiences in Berlin are more heterogeneous, with reviews reflecting a wider mix of positive and negative sentiment over time. In contrast, London provides an example of comparatively low dispersion, particularly in earlier years, pointing to more uniform and consistently positive evaluations. However, dispersion increases in later years, indicating a gradual widening in the range of sentiment expressed by guests. Taken together, these three cases illustrate that sentiment variability evolves differently across cities with some markets displaying persistently mixed evaluations, others maintaining stable sentiment patterns, while some experience shifts toward greater heterogeneity over time.

These patterns showcase that sentiment dispersion can also be viewed as an indicator of the consistency of the market. Cities with lower dispersion appear to offer more predictable guest experiences, while higher dispersion reflects greater variation in quality, expectations, or matching between guests and listings. Changes in dispersion over time may therefore signal shifts in market structure or in the composition of guests that use the platform.

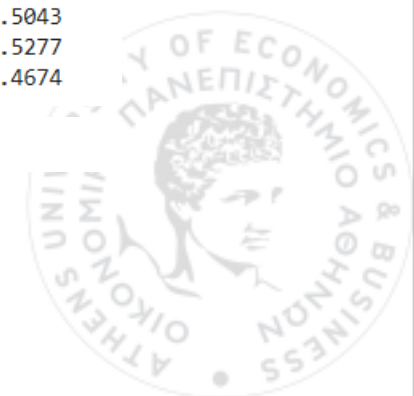
Sentiment Dispersion for London (K = 14)

```

=== LONDON - Mean sentiment by cluster (K=14) ===
  cluster    n  mean_sent  median_sent  std_sent  lon_center  lat_center
0     221    0.6660    0.7668    0.3230   -0.4237    51.5186
1     714    0.7186    0.8378    0.3196   -0.0994    51.5908
2    2379    0.6600    0.7767    0.3423   -0.1312    51.5289
3     894    0.7134    0.8382    0.3253    0.0026    51.5476
4     806    0.7321    0.8635    0.3431   -0.1882    51.4356
5     402    0.6920    0.7964    0.3389    0.1161    51.5100
6     375    0.7164    0.8420    0.3467   -0.0919    51.3947
7     608    0.6876    0.8219    0.3555   -0.2119    51.5716
8    1800    0.6614    0.7867    0.3539   -0.1165    51.4779
9     612    0.7399    0.8486    0.2890   -0.2894    51.5268
10    417    0.7977    0.8731    0.2247   -0.3118    51.4414
11   2838    0.6963    0.8151    0.3290   -0.1945    51.5043
12   2422    0.7042    0.8338    0.3355   -0.0677    51.5277
13    945    0.7187    0.8378    0.3186   -0.0181    51.4674

```

Figure 21: Mean sentiment and dispersion for K = 14 cluster in London



The cluster-level sentiment summary for London when the number of clusters is 14 reveals clear and systematic differences in both average sentiment and sentiment dispersion across the fourteen spatial clusters. Although all clusters exhibit predominantly positive sentiment, the variation in mean sentiment is significant. Mean sentiment values range from approximately 0.66 to 0.80, which is a meaningful spread given the compressed nature of sentiment scores. This indicates that guest experiences are not spatially uniform within London and that location-specific factors play a role in shaping review sentiment.

Several clusters stand out as consistently more positively evaluated. For example, Cluster 10 exhibits the highest mean sentiment (≈ 0.80) alongside the lowest standard deviation (≈ 0.22), suggesting not only more positive but also more consistent guest experiences. This combination is indicative of areas where listings deliver reliably high-quality stays, with relatively few negative deviations. In contrast, clusters such as Cluster 0, Cluster 2, and Cluster 8 display lower mean sentiment combined with comparatively higher dispersion. This pattern suggests areas where guest experiences are more heterogeneous.

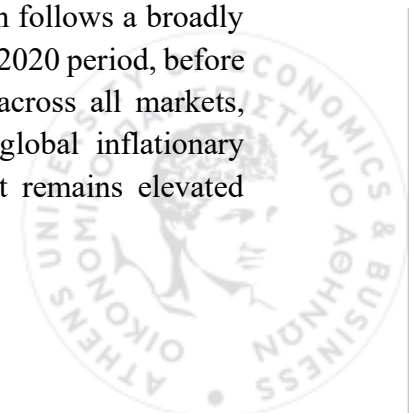
Dispersion plays a crucial role in refining the interpretation beyond mean sentiment alone. Clusters with similar average sentiment levels often differ substantially in their standard deviation, implying different degrees of consistency in guest evaluations. For instance, while several clusters cluster around a mean sentiment of roughly 0.70–0.72, their standard deviations range from approximately 0.28 to 0.36. Higher dispersion indicates that reviews within those areas include a mix of very positive and more critical experiences, whereas lower dispersion suggests more uniform evaluations.

Taken together, these findings provide strong post validation for the extended clustering approach applied to London. Importantly, sentiment was not used as an input in the clustering process, yet clear sentiment patterns emerge across the resulting spatial clusters. This indicates that the clusters capture meaningful geographic structure rather than arbitrary partitions. The results also provide a solid foundation for subsequent analyses that link cluster-level sentiment patterns to linguistic cues, pricing behaviour, or broader spatial characteristics.

Sentiment Regression

Macroeconomic indicators analysis

Figures 22 and 23 present the evolution of inflation and GDP per capita over time for the countries corresponding to the cities included in the study. Across all countries, inflation follows a broadly similar pattern, remaining relatively low and stable throughout most of the pre-2020 period, before increasing sharply around 2021–2022. This spike is observed consistently across all markets, although its magnitude differs, showcasing varying national exposure to global inflationary pressures. In the years following the peak, inflation gradually declines but remains elevated compared to pre-2020 levels.



GDP per capita exhibits more persistent cross-country differences. Zurich consistently displays the highest income levels throughout the period, followed by cities such as Stockholm, Amsterdam, and London, while Madrid and Rome remain at lower levels. Despite these differences in absolute income, most countries follow a similar temporal pattern, with moderate growth interrupted by a decline just before 2020 and a subsequent recovery. These macroeconomic trends provide important contextual background for the sentiment analysis that follows, as they reflect broader economic conditions under which Airbnb experiences were evaluated.

While previous sections focused on listing-level characteristics and within-city sentiment variation, the central objective of this part of the study is to examine whether broader macroeconomic conditions are associated with changes in user sentiment over time. Economic indicators such as income levels, inflation, and unemployment may influence both travel behavior and user expectations, potentially shaping how guests evaluate their Airbnb experiences.

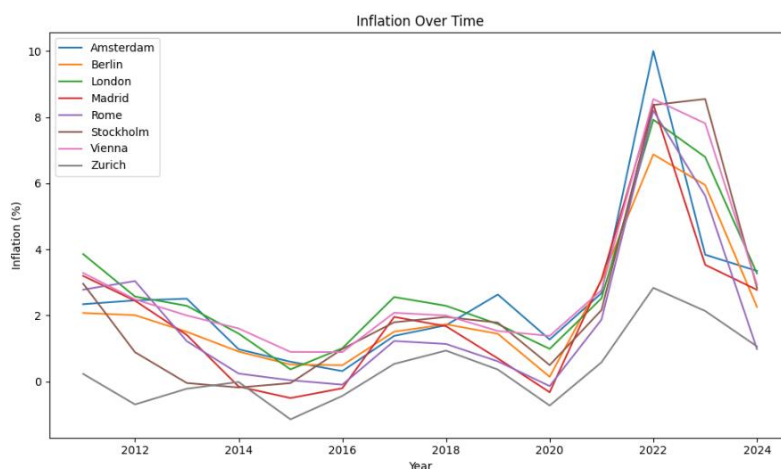


Figure 22: Inflation over time for all cities

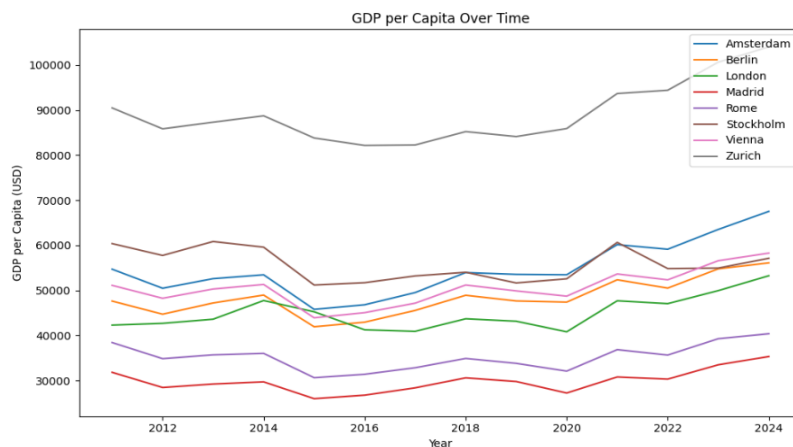


Figure 23 : GDP per capita over time for all cities



Pooled city regression with fixed effects

A key challenge in examining the relationship between sentiment and macroeconomic indicators was the limited number of observations available at the individual city-year level. Since macroeconomic data are observed annually, running separate regressions for each city would be based on very small sample sizes, severely limiting statistical power and making inference unreliable. To address this issue, a pooled regression framework was adopted, combining observations across all cities into a single city-year panel dataset.

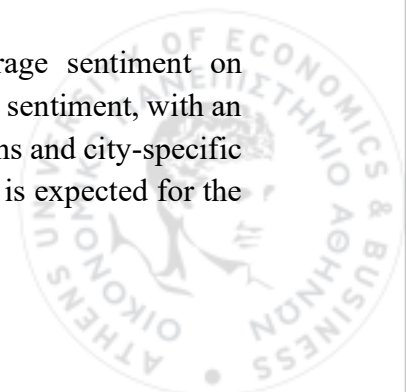
To account for structural differences between cities that are constant over time—such as cultural norms, tourism intensity, housing stock characteristics, or review-writing behavior—city fixed effects were included in the. These fixed effects absorb the heterogeneity across cities, allowing the estimated coefficients on macroeconomic variables to capture within-city variation over time rather than cross-city differences. As a result, the regression identifies how changes in macroeconomic conditions within a country are associated with changes in average sentiment for the corresponding city. The regression summary is shown below:

OLS Regression Results						
Dep. Variable:	avg_sentiment	R-squared:	0.742			
Model:	OLS	Adj. R-squared:	0.711			
Method:	Least Squares	F-statistic:	47.81			
Date:	Wed, 24 Dec 2025	Prob (F-statistic):	1.14e-30			
Time:	01:53:16	Log-Likelihood:	114.53			
No. Observations:	96	AIC:	-207.1			
Df Residuals:	85	BIC:	-178.9			
Df Model:	10					
Covariance Type:	HC1					
	coef	std err	z	P> z	[0.025	0.975]
const	0.8426	0.107	7.866	0.000	0.633	1.053
gdp_pc	-5.213e-06	1.94e-06	-2.693	0.007	-9.01e-06	-1.42e-06
inflation	-0.0065	0.004	-1.526	0.127	-0.015	0.002
unemployment	0.0245	0.003	7.641	0.000	0.018	0.031
city_berlin	-0.2043	0.036	-5.672	0.000	-0.275	-0.134
city_london	0.0403	0.030	1.349	0.177	-0.018	0.099
city_madrid	-0.6513	0.061	-10.691	0.000	-0.771	-0.532
city_rome	-0.3033	0.045	-6.745	0.000	-0.391	-0.215
city_stockholm	-0.0601	0.038	-1.575	0.115	-0.135	0.015
city_vienna	-0.2374	0.035	-6.735	0.000	-0.306	-0.168
city_zurich	0.1055	0.075	1.408	0.159	-0.041	0.252

Figure 24: Sentiment regression summary

Interpretation of the regression summary

Figure 20 reports the results of the pooled city-year regression of average sentiment on macroeconomic indicators. The model explains a large share of the variation in sentiment, with an R-squared value of 0.74, indicating that time-varying macroeconomic conditions and city-specific effects together account for a large share of observed sentiment differences. It is expected for the



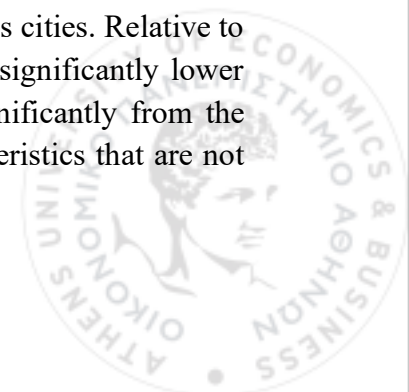
R-squared value to be that high because the analysis is done on city-level yearly averages, not on individual reviews. When thousands of individual opinions are averaged together, random differences in how people write or feel get cancelled out, leaving a much smoother and more predictable pattern. This makes the outcome easier to explain statistically.

In addition, cities tend to have stable characteristics that change very little over time, such as the type of visitors they attract or how reviews are typically written. Once these city differences are taken into account, the larger part of the variation in sentiment is already explained. Finally, both sentiment and macroeconomic indicators change gradually over time rather than randomly, meaning they tend to move together in broad trends. Because the model captures these stable city patterns and shared time trends, it is normal for it to explain a large share of the observed variation, resulting in a high R-squared value.

It is also important to note that Airbnb users predominantly consist of travelers, who may not be representative of the broader population captured by country-level macroeconomic indicators. As a result, changes in GDP or unemployment may influence the population of travelers.

- GDP per capita is estimated to have a negative and statistically significant association with average sentiment. Although the magnitude of the coefficient is small, considering the scale of GDP per capita, the negative sign suggests that increases in the GDP index are associated with slightly lower average sentiment. This result may reflect higher/ lower expectations among customers in wealthier/ poorer economies, where guests may be more demanding/ flexible when evaluating accommodation experiences.
- Inflation exhibits a negative but statistically insignificant coefficient, indicating no strong evidence that short-term changes in inflation are directly associated with changes in sentiment once all other factors are held steady. This suggests that while inflation affects overall economic conditions, its direct influence on how guests evaluate Airbnb stays may be limited or indirect.
- In contrast, unemployment shows a positive and statistically significant association with sentiment. This finding may indicate that during periods of higher unemployment, guests exhibit more positive or lenient evaluations, potentially reflecting shifts in expectations, consumption behavior, or travel patterns. However, this relationship should be interpreted cautiously, as unemployment may also proxy for broader economic stress that influences the composition of travelers rather than the underlying satisfaction.

The city fixed effects reveal systematic differences in baseline sentiment across cities. Relative to the reference city (Amsterdam), Berlin, Madrid, Rome, and Vienna exhibit significantly lower average sentiment, while London, Stockholm, and Zurich do not differ significantly from the reference category. These coefficients capture persistent city-specific characteristics that are not explained by macroeconomic conditions.



Overall, the results suggest that macroeconomic conditions are associated with measurable changes in average sentiment over time, although the direction and magnitude of these associations differ across indicators. Rather than indicating direct causal effects, the findings highlight how broader economic environments may shape user expectations and evaluation behavior on short-term rental platforms.



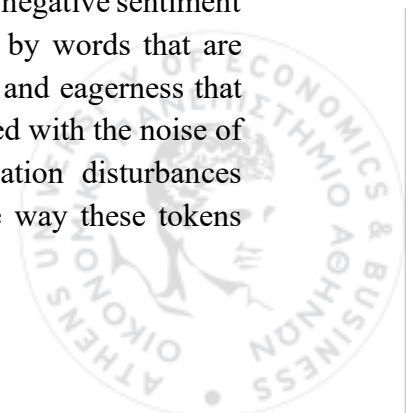
Conclusions

The aim of this study was to firstly explore how the pricing of each property relates to listing characteristics and attributes and secondly how user sentiment that was expressed in Airbnb reviews relates to spatial clustering, dispersion and ultimately to broader macroeconomic conditions. By applying various techniques that exist within the broader field of NLP, with traditional econometric methods, the study tried to provide a comprehensive view of how subjective user evaluations interact with various factors across multiple European cities. The analysis does not focus on a single city or a single dimension, but rather it adopts a comparative approach (where that is allowed) between the cities while maintaining uniformity by balancing the data.

The analysis conducted at the listing level proved and confirmed several insightful and well-established patterns. Accommodation capacity was consistently the strongest determinant of price across all cities, highlighting the importance of property size in pricing strategies on short-term rental platforms, like Airbnb. Review scores were also found to be associated with higher prices in most cities, suggesting that hosts often use their positive reputation to monetize the properties, although the strength and direction of this relationship varied across markets. Superhost status exhibited heterogeneous effects, reinforcing the general idea that the status does not necessarily reflect only quality but also the specific strategies that each host enforces. Host experience, in contrast, showed little to negligible systematic association with price, indicating that experience alone does not translate into higher pricing power once other observable factors are accounted for. Overall, these findings illustrate that even though certain pricing mechanisms are broadly consistent, others cannot be accounted for and usually are shaped by the local markets.

A central contribution of the study lies in its treatment of textual review sentiment. Using VADER sentiment analysis, review text was converted into numerical sentiment scores and aggregated at the listing level. When comparing sentiment-inferred ratings and actual Airbnb ratings we showed that sentiment closely follows and overestimates the overall distribution of numerical ratings, especially by capturing a strong positive skew. The overestimation that sentiment-based measures exhibited reflects the predominantly positive tone of Airbnb reviews and the tendency for negative feedback to be underrepresented or softened in textual form. Despite this limitation, the alignment between the two distributions supports the use of sentiment as a complementary tool to numerical ratings, particularly in contexts where ratings may be missing, delayed, or overly coarse.

To improve the interpretation of sentiment that is hidden within the reviews, token-level analysis was introduced to identify which words contribute most strongly to positive and negative sentiment across cities. The analysis revealed that positive sentiment is mostly driven by words that are related to the cleanliness, comfort, ease of a property and the responsiveness and eagerness that each host exhibited, while negative sentiment is linked with problems connected with the noise of the area, the cleanliness of the house, potential broken facilities, or location disturbances showcasing problems that are related with the location of the property. The way these tokens



appear in all cities is consistent and is proof that the sentiment scores are grounded in meaningful aspects of the guest experience rather than in abstract linguistic terms.

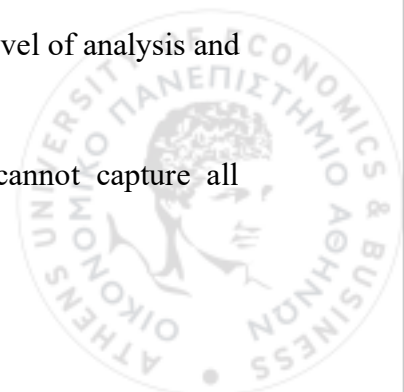
The spatial analysis extended the sentiment investigation by examining within-city differences. By clustering listings (using both $K = 4$ and $K = 14$) based on geographic coordinates, the study identified broad spatial zones that exhibited similar characteristics. While these clusters do not correspond to actual city neighborhoods, they revealed meaningful differences in sentiment across the areas, suggesting that guest experience is not uniform across the clusters. Some clusters exhibited consistently higher sentiment, while others showed lower or more variable evaluations. This highlights the role of location-related factors of an area such as noise, accessibility, safety index of and surrounding environment. The findings emphasize the importance of considering the existence of spatial heterogeneity when analyzing platform-based user evaluations.

One of the main contributions of the study lies in the attempt to associate sentiment to macroeconomic conditions by using a pooled city regression with the necessary adaptations that would allow it to be insightful (city fixed effects). The results suggest that macroeconomic indicators are generally associated with changes in average sentiment over time (statistical significance), although the magnitude of these effects is relatively small. GDP per capita was found to be negatively associated with sentiment, potentially reflecting high expectations in economies that have a higher index of prosperity and lower expectations to countries with a smaller prosperity index. Unemployment showed a slightly positive association with sentiment, while inflation did not display a statistically significant relationship once other factors were controlled for. These results should be interpreted as associations rather than effects that cause major sentiment fluctuations and may partly reflect changes in the composition of travelers rather than direct impacts on user satisfaction.

Limitations

Some limitations that can be highlighted in the study are that:

- Firstly, reviews are generated mostly by travelers (and a smaller proportion of natives) who do not represent the general population that is affected by the macroeconomic indicators. As a result, macroeconomic conditions may influence who chooses to travel rather than how experiences are evaluated.
- Secondly, sentiment analysis is inherently limited by the quality and tone of the review text, especially in environments that are influenced by strong positivity bias.
- Thirdly, the use of annual macroeconomic data restricts the temporal level of analysis and may hide short-term fluctuations or seasonal effects.
- Finally, spatial clustering based solely on geographic coordinates cannot capture all neighborhood-level characteristics that may influence sentiment.



Future work

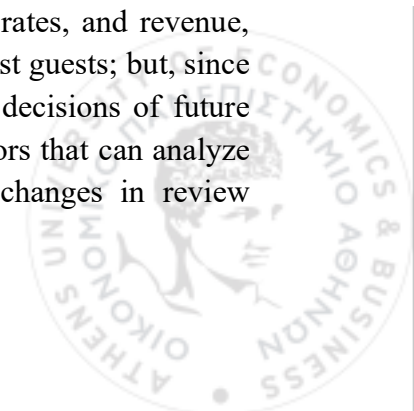
There are several directions in which this research could be extended to further improve the understanding of user sentiment and its relationship with economic conditions. One potential scenario for future work involves the use of higher-frequency and more impactful macroeconomic indicators. The study as is, relies on annual data, which limits the ability to capture short-term economic shocks, seasonal fluctuations, or rapid changes in consumer behavior that may influence user evaluations. Potentially incorporating quarterly or monthly indicators, where available, could prove to gather more information and provide overall a more detailed analysis of how sentiment responds to economic changes over shorter time horizons.

Future research could also focus on how interpretation can also have a causal impact. While the current analysis captures meaningful associations between sentiment and macroeconomic variables, it does not attempt to establish causal relationships. Applying causal identification strategies could provide deeper understanding into whether and how broad economic conditions directly influence user sentiment. Such approaches would help clarify whether changes in sentiment are mostly due to economic reasons or due to changes in the composition of travelers.

Another possible extension would be to broaden the platform scope of the analysis. Comparing Airbnb reviews with reviews from other accommodation platforms or traditional hotels could reveal whether the sentiment patterns observed in this study (and are related to the Airbnb specific reviews) are a phenomenon observed only in the Airbnb platform or a more generalized trend in hospitality services. This could also help assess and measure the extent to which the platform, the review system in place and user expectations shape sentiment expression.

From a methodological perspective, future work could explore the use of alternative sentiment analysis techniques, including supervised models or transformer-based language models, which may better capture context, sarcasm, or nuanced expressions that lexicon-based methods struggle to identify. Combining these types of approaches with explainability techniques that are available and were made to specifically cater the needs of certain models, could improve both the accuracy and interpretability of sentiment measures. Additionally, incorporating multilingual sentiment analysis would allow for a more comprehensive use of reviews written in languages other than English, particularly for cities in which the use of the English language is not common and that have a diverse mix of international visitors.

Finally, future research may look into an extension of this study to include the examination of the link between the sentiment expressed in the historical guest reviews and the behavioural and economic outcomes of future activity, such as booking activity, occupancy rates, and revenue, based measures. On one hand, review sentiment mirrors the experiences of past guests; but, since it is publicly visible on the platform, it may influence the perceptions and decisions of future guests. Hence, the use of sentiment measures along with behavioural indicators that can analyze sentiment and potentially link future purchases and other metrics with changes in review



sentiment. These insights could pinpoint the extent to which future economic performances were a result of previous user evaluations and reviews.



References

- Barron, K., Kung, E., & Proserpio, D. (2021). *The effect of home-sharing on house prices and rents: Evidence from Airbnb*.
<https://doi.org/10.1287/mksc.2020.1227>
- Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). *Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings*.
<https://doi.org/10.1080/10548408.2017.1308292>
- Wang, D., & Nicolau, J. L. (2017). *Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com*.
<https://doi.org/10.1016/j.ijhm.2016.12.007>
- Zervas, G., Proserpio, D., & Byers, J. W. (2017). *The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry*.
<https://doi.org/10.1509/jmr.15.0204>
- Dellarocas, C., & Wood, C. A. (2008). *The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias*.
<https://doi.org/10.1287/mnsc.1070.0743>
- Hu, N., Zhang, J., & Pavlou, P. A. (2009). *Overcoming the J-shaped distribution of product reviews*.
<https://doi.org/10.1145/1562764.1562800>
- Moe, W. W., & Schweidel, D. A. (2012). *Online product opinions: Incidence, evaluation, and evolution*.
<https://doi.org/10.1287/mksc.1110.0662>
- Hutto, C. J., & Gilbert, E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text*.
<https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
- Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). *Big data for big insights: Investigating language-specific drivers of hotel satisfaction with topic modeling*.
<https://doi.org/10.1016/j.ijhm.2016.08.012>

