BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

CENTRAL BANK OF MALTA
WORKING PAPER

Where Does Malta's P2P Rental Market Fit?

A Case Study for 2023

Sarah Spiteri*

WP/4/2025

*Sarah Spiteri is a Senior Research Economist in the Economic Research Office within the Economic Research Department. The author would like to thank Alexander Demarco, Aaron G. Grech, Noel Rapa, and Massimo Giovannini for comments and suggestions. Special thanks go to Brian Micallef for valuable comments and to the Housing Authority for providing the author with rental register data. The views expressed in this report are those of the author and do not necessarily reflect the views of the Central Bank of Malta. Any remaining errors are the sole responsibility of the author.

Any queries or requests should be sent here.

Abstract

This study examines Malta's peer-to-peer (P2P) rental market by comparing rental listings from Facebook Marketplace and an aggregation of real estate agencies' adverts, while the Housing Authority's official register is used as the benchmark characterising the local rental market. Using a combination of hedonic regression models and natural language processing techniques, including Word2Vec, Doc2Vec, and K-Means clustering, the analysis identifies key differences in pricing strategies, property characteristics, and the role of digital platforms in rental advertising. Findings reveal that advertised listings are skewed towards higher-end properties, with a greater prevalence of larger units and higher rental prices compared to the official register. Additionally, real estate agencies exhibit distinct pricing behaviours across platforms, with agency-listed properties on Facebook Marketplace generally featuring advertised rents that are between 1.5% and 2.9% lower than those posted by individual landlords. Moreover, P2P platforms usually contain a broader range of property types compared to those in aggregated agency listings. This suggests that agencies tend to use Facebook Marketplace to reach a wider audience, including lower-budget renters, while agency listings primarily cater to higher-end properties and a wealthier clientele, reflected in its higher average rental prices and greater concentration of premium listings. Linguistic analysis of property descriptions shows that comprehensive, high-quality phrasing increases rental asking prices by 3.7% to 10.2%, while the presence of specific quality-indicative words raises rents by 1.2% to 3.5%, with the range largely depending on the online platform on which an advert is posted. These findings underscore the importance of considering data source biases in rental market analysis and highlight the role of social media and platforms for real estate agencies' listings in shaping rental market dynamics and price-setting behaviours.

JEL Classification: C23, C55, O18, M37, R32

Keywords: Rental Market, Peer-to-Peer Market, Advertised Prices, Hedonic Regressions, Natural Language Processing (NLP), Word2Vec, Doc2Vec, K-Means, Malta

Contents

1 Introduction

The Maltese rental market presents a compelling case study due to its rapid transformation in recent years, driven by legal reforms and shifting demographic trends. The introduction of the Private Residential Leases Act (2020) aimed to professionalise the sector, enhance contract stability, and improve transparency through mandatory lease registration. This reform was particularly significant in addressing Malta's historically dual-market structure, characterised by the coexistence of rigid rent controls and an unregulated rental segment (Micallef, 2021). Additionally, the sector has been significantly influenced by an influx of foreign workers, which has contributed to heightened demand and upward pressure on rental prices (Gauci et al., 2022). Investigating the Maltese rental market provides valuable insights into the effects of policy interventions on housing affordability, market efficiency, and data transparency within a small but highly dynamic economy.

The rapid expansion of the peer-to-peer (P2P) rental market, largely driven by the continued expansion of the tourism sector, an influx of foreign labour and the popularisation of platforms such as Airbnb, has had notable implications for Malta's housing sector. Recent studies, such as Ellul Dimech (2019), have highlighted the impact of short-term rentals on the dynamics of the long-term rental market, particularly in popular tourist districts. This finding has raised concerns regarding housing affordability for residents and has blurred the distinction between traditional long-term rentals and the short-term rental market. Consequently, these developments underscore the need for policies that balance tourism growth with housing availability.

Several studies have explored pricing determinants in the P2P rental market. Fearne (2021) and Camilleri (2020) utilised linear regression models and spatial analysis to demonstrate that factors such as location, amenities, and property ratings significantly influence rental prices. Moreover, regulatory measures have been shown to shape market behaviour, with recent legislative reforms—particularly the Private Residential Leases Act (2020)—aiming to enhance transparency and market stability (Micallef, 2021; Gauci et al., 2022). These findings emphasise the necessity of further research into the P2P rental market, not only to assess its economic impact, but also to inform policy decisions aimed at improving affordability and ensuring market sustainability.

A key limitation of existing studies is their reliance on Airbnb data, which skews findings toward short-term and holiday rentals. In contrast, data from the Maltese Housing Authority indicates that the majority of registered rental contracts pertain to long-term leases (Housing

Authority, 2024). This bias stems from restricted access to comprehensive rental market data. To address this gap, the present study proposes incorporating Facebook Marketplace data to diversify the analysis, thereby capturing a broader spectrum of traditional P2P rental listings and offering a more holistic perspective on the Maltese rental market.

Digital platforms such as Facebook Marketplace have demonstrated significant potential in facilitating landlord-tenant connections. Goodchild and Ferrari (2021) classify Marketplace as a "connector" in the United Kingdom's rental market, providing a platform that reduces information asymmetry while maintaining minimal involvement in transactions. The relevance of Facebook in the Maltese context is supported by Eurobarometer survey data, which indicate that 81% of Malta's population was reachable via Facebook in 2023 (European Commission, 2023). These figures, consistent with previous and more recent surveys, highlight the platform's role in connecting individuals, including landlords and tenants, making it an invaluable resource for analysing rental market features in Malta.

Micallef and Spiteri (2023) leveraged Facebook Marketplace data to compare rental price distributions with those recorded in Malta's official housing register. Their findings indicate that Facebook Marketplace listings typically feature higher rents than those registered, reflecting a distinct market segment. The study estimates that the P2P segment constitutes approximately a quarter of the overall rental market. However, while acknowledging potential differences in pricing behaviour between the P2P market and real estate agencies, the study does not provide a direct comparative analysis. The present research expands upon this by integrating data from a set of aggregated real estate agencies' adverts[1] that consolidates listings from approximately 20 real estate agencies, alongside the official Housing Authority records which are used as a benchmark for overall rental market characteristics.

When analysing rental data from platforms such as Facebook Marketplace or an aggregation of real estate agents' adverts, it is crucial to recognise that these listings typically represent asking prices rather than actual agreed prices. Rental prices are often subject to negotiation between landlords and tenants, as noted by Han and Strange (2016). Furthermore, such datasets do not fully capture market complexities, such as contract renewals, which play a vital role in stabilising administrative rental prices. For instance, 42.65% of registered rental contracts were renewed in 2023, with 92% maintaining the same rental rate (Housing Authority, 2024). In order to account for this and to ensure a meaningful comparison with the other data sources, the Housing Authority rental data for 2023 is cleaned from all contract renewals. In this respect,

[1] This dataset is an aggregation of adverts that are posted online by real estate agents on various platforms but excluding agents' listings on Facebook Marketplace which are treated separately for the purpose of this study.

access to detailed administrative records, such as those maintained by the Housing Authority, are critical in allowing for a meaningful and comprehensive approach to studying the Maltese rental market.

This study compares the official rental market register maintained by the Housing Authority with advertised listings on Facebook Marketplace and an aggregated dataset of real estate agencies' listings for 2023. After applying a rigorous data cleaning process, the final datasets comprise 32,684 new active rental contracts and 16,404 advertised listings from Facebook Marketplace and 11,152 adverts sourced from a collection of aggregated agency adverts. The breadth and depth of these datasets provide a foundation for addressing four key research questions: (i) How do property characteristics differ across the three data sources? (ii) Do real estate agents exhibit different pricing behaviours compared to individual landlords on Facebook Marketplace? (iii) If agents' pricing strategies differ based on their training and their ability to create more compelling property descriptions? (iv) How do advertisements from the P2P rental market compare to those posted by real estate agents across different platforms?

To address these research questions, this study first examines the descriptive statistics of the three datasets, highlighting common patterns and key differences in property attributes. This is followed by a series of hedonic regression models, drawing on the approaches of Micallef and Gauci (2022) and Micallef et al. (2021), to identify the factors influencing rental prices within Facebook Marketplace and across the two advertisement-based datasets. The models incorporate a set of dummy variables to distinguish between listings posted by individual landlords and real estate agents, while also capturing platform-specific effects to assess whether agents adjust their pricing strategies depending on the medium used for advertising.

A key methodological contribution of this study is its use of textual analysis to examine linguistic variations in property descriptions. Two embedding techniques—Word2Vec and Doc2Vec—are applied to transform textual data into a computational vector space, with K-Means clustering employed to derive interpretable indicators for inclusion in the hedonic regression models. These techniques allow for a comparison of different advertising strategies: one approach focuses on the frequency of specific keywords highlighting key property attributes, while the other captures broader messaging strategies by analysing full descriptions.

The empirical analysis reveals distinct differences between the datasets. Advertised rental listings are skewed toward higher-end properties, with larger units and premium locations disproportionately represented in the dataset of agencies' adverts. For instance, 53% of listings in the aggregated agencies dataset feature properties with three or more bedrooms, compared to 44% in the official register. Median rental prices are also significantly higher in advertised datasets, with a median rental rate for new contracts of €850 in the Housing Authority's register, while the median listed rents on Facebook Marketplace and the aggregated agencies dataset reach €1,400 and €1,500, respectively.

Regression results further indicate that real estate agents and individual landlords adopt different pricing strategies on Facebook Marketplace. Agent-listed properties on the platform are advertised at 1.5% to 2.9% lower rents than those posted by individual landlords, suggesting that agencies use the platform to target budget-conscious renters. Conversely, properties listed in the aggregated agencies dataset are priced approximately 2.7% higher on average, reflecting a focus on premium listings by real estate agent platforms.

Another key finding is the impact of descriptive phrasing on rental prices. Listings featuring comprehensive, high-quality descriptive language command rents that are 3.7% to 10.2% higher, whereas the presence of specific quality-indicative keywords correlates with a more modest 1.2% to 3.5% price increase. However, the study finds no evidence that real estate agents use quality-indicative phrasing more frequently than individual landlords. Instead, the effectiveness of these advertising strategies varies depending on the platform, with a greater price premium observed in the aggregated agencies dataset.

The study also highlights discrepancies between advertised and contracted rental rates, emphasising the role of negotiation in price formation. While advertised rents reflect landlords' initial asking prices, actual agreed-upon rents—particularly for long-term contracts—may be lower due to bargaining between tenants and landlords. This suggests that while digital rental listings provide valuable insights into market trends, they do not fully capture the final transaction prices.

These findings underscore the importance of considering both platform selection and advertising strategies when analysing rental market dynamics. By integrating structured datasets with advanced text analysis techniques, this study provides a nuanced perspective on the segmentation within Malta's rental market and the evolving role of digital platforms in property advertising.

The following section presents a comparative analysis of the Housing Authority's official rental register and advertised listings from the two selected databases. The paper then examines the role of Facebook Marketplace as a real estate advertising platform, assessing its use by both independent landlords and real estate agents. The analysis further contrasts the characteristics of P2P rental listings with those posted by agencies, highlighting key differences in pricing, wording, and audience reach. Finally, the paper concludes with a summary of the main findings, offering insights into their implications for market transparency, dataset selection, and rental price dynamics.

2 Comparing Rental Adverts with Registered Contracts

The analysis in this study is based on three distinct datasets. The benchmark dataset is the official rental contracts register maintained by the Housing Authority, which comprises both stock and flow data due to the continuous entry, exit, renewal, and extension of contracts. Unlike previous studies by Micallef and Gauci (2022) and Micallef and Spiteri (2023), this research focuses exclusively on new rental contracts initiated in 2023. This restriction mitigates the anchoring effect of contract renewals and multi-year agreements, as highlighted by Micallef and Gauci (2022), thereby enhancing comparability with advertised rental listings. The official register contains 32,684 contracts commencing in 2023, with 84% classified as long-term leases, indicating that a significant portion of these agreements extended into 2024.

The first advert-based dataset was collected quarterly from a collection of adverts that feature on real estate agency platforms, as outlined by Brincat and Ghigo (2022). To ensure data consistency and reliability, a rigorous preprocessing procedure was applied, including format standardisation, outlier removal, deduplication, and feature extraction from property descriptions. Following these steps, the dataset comprises 11,152 advertisements posted in 2023 by approximately 20 real estate agencies.

The second advertised dataset was compiled from Facebook Marketplace on a bi-weekly basis, as described by Micallef and Spiteri (2023). The differing data collection frequencies were chosen to account for variations in listing persistence and turnover across the two platforms, ensuring that both datasets contain a comparable number of listings. After undergoing a similar preprocessing procedure, the Facebook Marketplace dataset consists of 16,404 advertisements, with 11,656 posted by individual landlords and 4,748 by real estate agents. This classification was derived using textual indicators in property descriptions referencing an affiliation with a real estate agency. However, this method remains susceptible to Type I and Type II classification errors.

Table 1 presents the distribution of property characteristics across the three datasets, with Facebook Marketplace listings further disaggregated into advertisements posted by individual landlords and real estate agents. Real estate agents account for 29% of Facebook Marketplace listings, with textual analysis suggesting that approximately 15 agencies actively advertise on the platform. The figures in Table 1 indicate that the property characteristics within the P2P segment are broadly aligned with those of real estate agents.

Table 1: Characteristics of Rental Listings

	Official Register[2]	Aggregated Real Estate Agencies	Facebook Marketplace P2P Market	Facebook Marketplace Agents' Listings
Observations	32,684	11,152	11,656	4,748
Type of property				
Apartment	68.45%	65.67%	68.62%	69.19%
Penthouse		15.88%	12.43%	14.32%
Maisonette	9.73%	8.34%	7.08%	7.08%
House	6.10%	6.25%	4.99%	3.92%
Villa	0.32%	3.86%	2.63%	2.11%
Room Only[3]	15.41%	0.00%	4.26%	3.39%
Number of bedrooms				
1	20.05%	9.16%	16.57%	14.91%
2	36.16%	37.88%	38.34%	38.18%
3	34.40%	46.68%	40.55%	43.64%
4+	9.39%	6.29%	4.55%	3.26%
Number of bathrooms				
1		28.16%	30.28%	27.84%
2		59.62%	60.68%	63.54%
3		8.97%	6.75%	6.80%
4+		3.26%	2.28%	1.81%
Region				
Northern Harbour	45.50%	64.62%	54.86%	51.05%
Northern	22.30%	15.55%	19.02%	25.38%
South Harbour	10.30%	5.16%	5.95%	5.86%
Southern Eastern	9.59%	4.58%	8.67%	6.68%
Gozo	6.47%	3.05%	2.55%	1.64%
Western	5.83%	7.04%	8.97%	9.39%

Note: The table shows the portions of observations dedicated to the listed categories in the respective datasets.
Source: Own calculations based on the properties of the collected datasets.

The overall composition of the property types is relatively consistent across the three data sources. However, the two advertisement-based datasets exhibit a stronger skew toward apartments and penthouses, which together comprise approximately 82% of listings, whereas the official register contains a higher proportion of maisonettes. A key distinction is the presence of

[2]The official register of the Housing Authority aggregates data differently across property types (such as grouping apartments and penthouses), whilst excluding data for the number of bathrooms and the involvement of agencies. This leads to the missing values in the above table. A comparison between the newly registered contracts presented above and all active contracts in 2023 is available in section A.1 of the appendix.

[3]The definition of shared rental contracts in the official register is based on the legal definition during 2023. The change in legal definition in September 2024 is expected to affect the number of contracts for shared spaces.

shared spaces, classified as "Room Only" on Facebook Marketplace, which account for 4% of total observations. Within the P2P segment, a higher share of properties is listed as houses or shared accommodations, whereas real estate agent listings are more concentrated in the apartment and penthouse categories.

In contrast, the aggregated real estate agencies' listings exclude shared dwellings, while the official register includes a substantial fraction of shared accommodations, comprising 15.41% of new rental contracts in 2023. Furthermore, within the complete register of active rental contracts, the Housing Authority reports that 9.40% of properties were classified as shared spaces in 2023, more than twice the proportion observed on Facebook Marketplace. However, this discrepancy may reflect an underestimation in the advertised dataset, as a single listing on Facebook Marketplace may represent multiple rooms within the same property.

A key distinction among the three data sources is the distribution of property sizes. In the advertised datasets, three-bedroom properties are the most frequently listed, whereas the official register contains a higher proportion of smaller units. Facebook Marketplace exhibits a property size distribution most comparable to the official register for one-bedroom dwellings but records nearly half the proportion of properties with more than four bedrooms. Additionally, the distribution of bathroom counts appears relatively consistent across the two advertisement-based datasets, although Facebook Marketplace has a lower share of properties with four or more bathrooms, mirroring the differences observed in bedroom counts. These discrepancies in property size distributions may be influenced by platform-specific dynamics and the listing behaviours of real estate agencies, which shape the composition of advertised rental properties.

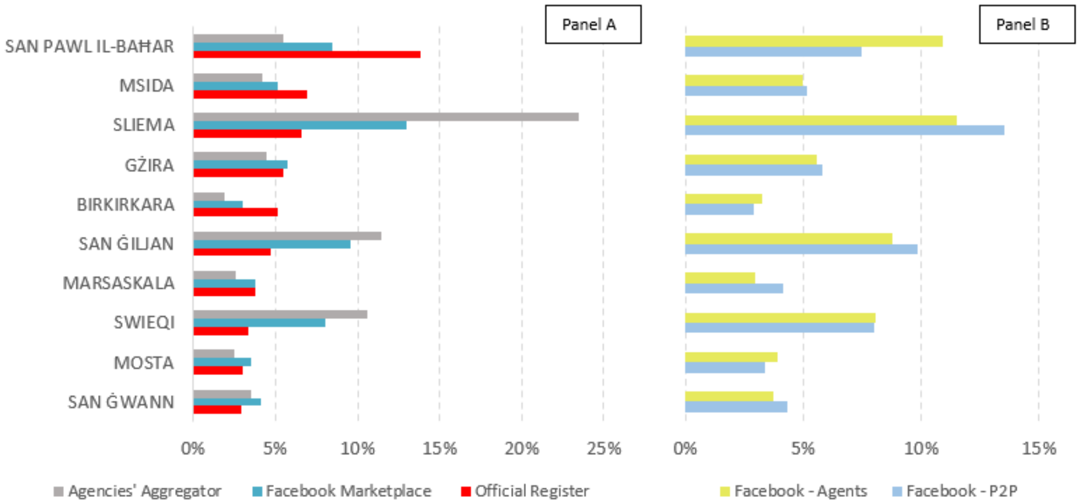
The relationship between property size and time-on-market has been widely studied, although the findings remain inconclusive. Some research suggests that larger properties remain on the market longer due to a smaller pool of potential tenants and higher rental prices (Benefield et al., 2011). An analysis of advertised listings supports this, indicating that larger properties take longer to secure tenants, leading to their accumulation in online datasets, whereas smaller properties are rented more quickly, resulting in higher turnover in listings. This skewness of the data towards properties with above average rental rates persists despite data preprocessing efforts aimed at mitigating biases arising from properties remaining on the platform for extended periods. Additionally, anecdotal evidence suggests that real estate agencies may re-advertise the listings of properties that are being rented to tenants to promote similar properties that have yet to find tenants. Given that agents earn commissions based on property valuations, they may be more inclined to use this strategy for higher-end properties, reinforcing their dominance in advertised datasets. A similar skewness towards larger and higher priced properties is also pronounced when comparing one- and three-bedroom properties. Additional details, such as the number of bathrooms or square meters, would provide a more robust assessment of variations in property size.

Across all three datasets, rented properties are predominantly concentrated in the Northern Harbour region, with the highest concentration observed in real estate agencies' listings (64.62%). The advert-based datasets also show a lower proportion of properties listed in the southern regions and Gozo, areas traditionally associated with lower rental prices. Given that the Northern Harbour region is traditionally the most expensive, landlords and agents may be more inclined to advertise properties in this area, resulting in a greater accumulation of listings. This aligns with existing evidence suggesting that higher-priced properties tend to remain on the market longer and may be advertised by multiple agents and the landlords themselves. When preprocessing the data to account for these behaviours the data remains skewed towards the northern regions. This may also be affected by the promotion tactics of agents that tend to re-advertise or maintain past adverts of accommodations with very specific characteristics that despite not being currently available for rent, tend to increase an agent's chances of being contacted by prospective lessees. In contrast, the lower volume of listings in the more affordable southern regions may reflect a greater reliance on informal rental agreements made outside established platforms. Additionally, the frequency of data collection may not be high enough to fully capture the rapid turnover of properties in these areas, potentially under-representing the true volume of rental activity.

Panel A of Figure 1 shows the top ten localities across the three rental databases. While there are slight differences in the popularity of specific localities across sources, the figure suggests a general convergence in the top ten rankings. The main discrepancies in the localities' popularity are found in Saint Paul's Bay ('San Pawl il-Bahar' in Maltese), Sliema, Swieqi and St. Julian's ('San Ġiljan' in Maltese), while listings in other localities remain broadly aligned. The rent registry shows relatively more entries for Saint Paul's Bay, but relatively fewer entries in the other three localities. These divergences are related to differences in average rent levels. While Saint Paul's Bay exhibits moderate rental prices, Sliema and Swieqi consistently command higher rents, incentivising a greater volume of advertisements without necessarily reflecting a proportional increase in rental agreements. Panel B of Figure 1 focuses on the distribution amongst the top ten localities within Facebook Marketplace. This panel highlights that listings by individual landlords and agents are broadly in line, with prominent discrepancies between the two segments in Saint Paul's Bay and Sliema. Interestingly, the fraction of listings in Saint

Paul's Bay by real estate agents on Facebook Marketplace is the most comparable to the share within the official register, with P2P listings have a higher concentration in Sliema.

Figure 1: Comparison of Localities between the Rental Market Databases

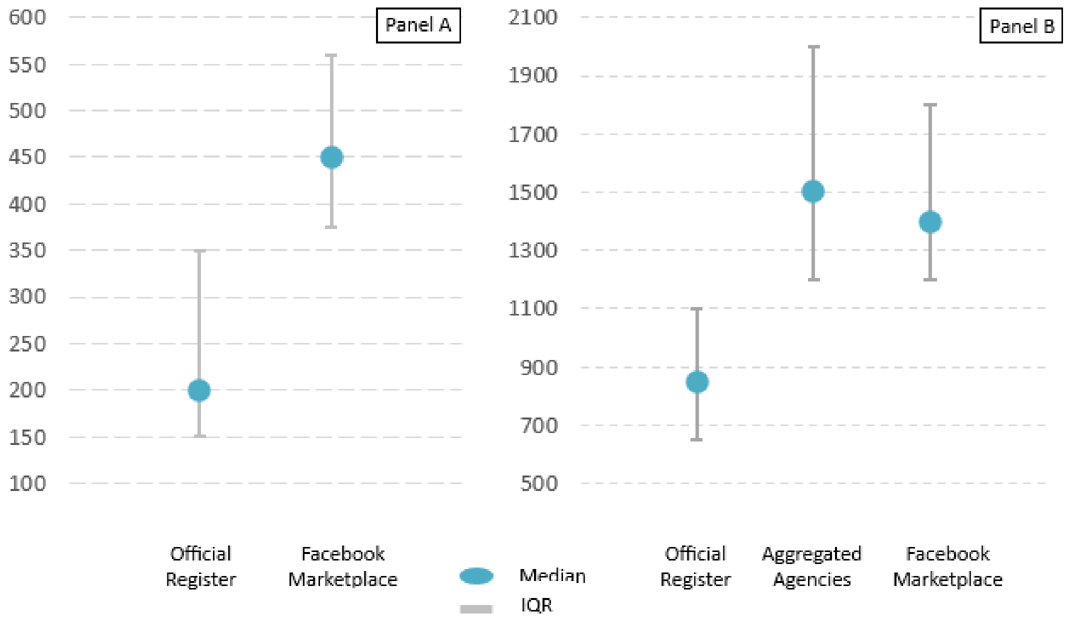


Note: In both panels, the bars represent the share of properties located in the ten local councils with the highest number of rented units. Panel A compares this share across the three data sources, while Panel B focuses on Facebook Marketplace listings, distinguishing between those posted by agents and individual landlords.
Source: Author's calculations.

A further prominent distinction between the Authority's register and Facebook Marketplace advertisements lies in the distribution of rental prices for shared spaces. The 2023 Registered Rental Contracts in Malta report (Housing Authority, 2024) indicates that 68% of registered contracts for shared spaces had a monthly rent between €100 and €299. Panel A of Figure 2 shows that when restricting the official register to new contracts, the median rent for shared spaces remains around €200, whereas on Facebook Marketplace, it stands at approximately €450, with no overlap in the interquartile ranges (IQRs) of the two datasets. Anecdotal evidence suggests that a significant proportion of shared spaces in Malta is advertised on Facebook, primarily through a combination of Marketplace and Groups targeting expatriates. The exclusion of listings from these Groups may partly explain the discrepancy between advertised rents and those recorded in the official register.

However, the discrepancy between advertised and registered prices is not limited to shared spaces. Panel B of Figure 2 illustrates that the median rental rate for new contracts in 2023 was €850, with the overall distribution of registered rental prices peaking between €700 and €999 (Housing Authority, 2024), reflecting the marginal effects of renewals and multi-year contracts. In contrast, the same panel shows significantly higher median rental rates from the aggregated agencies' adverts (€1,500) and Facebook Marketplace (€1,400). This broader divergence between advertised and registered rents is further emphasised by the lack of overlap between the IQRs of advert-based and registered listings. Importantly, negotiation plays a crucial role in shaping these discrepancies—advertised rental rates often serve as starting points for discussions, with final contracted rents potentially settling lower, particularly in a market where landlords and tenants engage in direct bargaining.

Figure 2: Rental Rates Across Datasets



Note: In both panels, points indicate the median monthly rent, while bars represent the IQR, with the lower and upper bounds corresponding to the 25th and 75th percentiles, respectively. Panel A shows rental rates for shared accommodations, whereas Panel B covers all other rental types.
Source: Author's calculations.

The overarching narrative across these datasets indicates that advertised listings are skewed towards the upper segments of the market. This is evident in property attributes such as the number of bedrooms and bathrooms, as well as the concentration of listings in higher-priced localities[4]. These characteristics, in turn, contribute to higher average rent levels for advertised properties. Several other factors contribute to this phenomenon.

One primary factor is that higher-priced properties tend to remain on the market longer due to a more limited pool of potential renters and the distinct characteristics of high-end homes, thereby increasing their representation in advertised listings (Sirmans and Slade, 2010). There-

[4]The 2023 End-of-Year Report on Registered Rental Contracts (Housing Authority, 2024) presents the median rental rates across localities for two- and three-bedroom apartments in Figures 12 and 13, respectively. While these rates likely underestimate advertised rents, they provide valuable insight into price variations between two- and three-bedroom apartments and across different localities. Notably, the figures confirm that Sliema, Swieqi, and St Julian's—the most prominent localities in the advert-based datasets—rank as the top three highest-priced areas.

fore, a portion of the most affordable properties may enter and exit online advertising platforms more rapidly than the data collection intervals allow, resulting in their under-representation. In contrast, landlords are required to register all rental contracts with the Housing Authority, ensuring comprehensive coverage; however, this administrative data is typically subject to a significant time lag. Advertised listings, on the other hand, provide a near real-time snapshot of market activity, albeit with limitations in coverage and representativeness.

When analysing rental data from platforms like Facebook Marketplace or aggregated real estate agencies' listings, it is important to recognise that advertised prices typically reflect initial asking rents rather than final agreed amounts, which are often subject to negotiation. A key contributing factor to this discrepancy is the possibility that advertised rents are subject to downward negotiation. Existing literature suggests that information asymmetries may enable well-informed tenants to negotiate lower rents (Geltner et al., 2013), while landlords' forward-looking rent expectations may diverge from tenants' focus on current market conditions (Gyllenberg and Koppfeldt, 2020). Additionally, prospective tenants may leverage price rigidities by making counteroffers that align more closely with prevailing market rates (Han and Strange, 2016). However, differences in pricing strategies across advertising platforms and between individual landlords and real estate agencies remain an open question.

These factors underscore the importance of using detailed, contract-based administrative records to gain a more comprehensive understanding of the Maltese rental market. For instance, studies on rental affordability and market stability can greatly benefit from key features of registered data, such as the inclusion of contract renewals and multi-year agreements. At the same time, market analysis can also gain insights from the distinctive characteristics of advertised listings sourced from various platforms.

3 Agencies, Semantics & Platforms in the Rental Market

Table 1 illustrates that 29% of the listings on Facebook Marketplace were posted by real estate agencies. It is reasonable to expect that these listings may exhibit distinct characteristics compared to others, as real estate agents are generally assumed to possess more in-depth market knowledge than the average individual landlord, due to their specialised training and diverse experiences across various properties. To test this hypothesis, a time-dummy hedonic regression model, based on the framework of Micallef and Gauci (2022) and Micallef et al. (2021), is employed. This model controls for variables such as the number of bedrooms and bathrooms, amenities (e.g., garage, garden, pool, seafront, and views), property type, and location. These being collectively referred to as a set of housing characteristics for each listing l, denoted by H_l. Additionally, dummy variables for each period (p_m) are used to account for the time dimension, with the first serving as the reference. Hence, the base structure of the hedonic models seeks to explain the rent (r_l) using the functional form:

$$ln(r_l) = \alpha + \sum_{m=1}^{M} \delta_m p_{m,l} + \beta' H_l + \varepsilon_l \tag{1}$$

This base structure is then augmented through three approaches. A key modification from Micallef and Gauci's approach is the inclusion of a binary variable indicating whether the listing was posted by a real estate agent. Furthermore, the model considers the phrasing of the property description, with the hypothesis that the way the description is structured may serve as a quality signal for potential tenants. Lastly, these dummy variables are interacted to test whether quality signalling is predominantly used by agents to a statistically significant degree.

A mix of embedding and clustering algorithms were used to analyse the phrasing of descriptions. With regards to the embedding approaches, these are used to provide an efficient numerical representation of the property descriptions in the adverts such that it captures their meaning and relationships. This paper shortlists two embedding approaches: Word2Vec (Mikolov et al., 2013) and Doc2Vec (Le and Mikolov, 2014). These methods are able to learn numerical vector representations that preserve semantic similarity within the property descriptions of advertisements, but the two algorithms diverge on their approach to derive these vector representations.

Word2Vec and Doc2Vec are neural network-based embedding models designed to represent words and documents as dense, continuous vectors in a lower-dimensional space, thus allowing for the inclusion of property descriptions in numerical modelling and analysis whilst preserving their linguistic and contextual relationships.

Word2Vec focuses on converting individual words into embeddings based on their contextual similarity within a large corpus[5]. Words that frequently appear in similar contexts are mapped closer together in the high-dimensional vector space, enabling the model to capture meaning-based similarities efficiently. This research employs the Continuous Bag of Words (CBOW) architecture of Word2Vec to learn word embeddings. In CBOW, a shallow neural network is trained to predict a target word (w_t) given its surrounding context words within a fixed-size window (z) from the set of vocabulary (V). Formally, given a w_t the model uses unconstrained optimisation to maximise the log-likelihood:

[5]Section A.2 of the appendix includes a more detailed description of the mathematical underpinnings of Word2Vec.

$$\max_{\theta_w} \sum_{t=1}^{T} \log P(w_t \mid w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) \tag{2}$$

where the conditional probability is defined as:

$$P(w_t \mid w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) = \frac{\exp\left(\mathbf{v'}_{w_t}^{T} \cdot \mathbf{h}\right)}{\sum_{w \in V} \exp\left(\mathbf{v'}_{w}^{T} \cdot \mathbf{h}\right)} \tag{3}$$

and h is the average of the vectors for the surrounding context words:

$$\mathbf{h} = \frac{1}{2z} \sum_{-z \leqslant i \leqslant z, i \neq 0}^{z} \mathbf{v}_{w_{t+i}} \tag{4}$$

The softmax function assigns a higher likelihood to words that are semantically aligned with their context. θ_w refers to a set of parameters denoting all learnable parameters—i.e., the input word vectors (v_w), and the output word vectors ($v_w^{'}$)—making them effective representations of word-level semantics.

However, Word2Vec does not incorporate document-level context, which limits its ability to model broader thematic patterns. This limitation motivated the development of Doc2Vec, an extension of Word2Vec that learns fixed-length vector representations for entire documents in addition to words[6]. In this research, the Distributed Memory (DM) version of Doc2Vec is used. It introduces a unique document vector d_D that is trained alongside word vectors to predict w_t given both its surrounding context and the document identity. The objective is to maximise the log-likelihood using unconstrained optimisation:

$$\max_{\theta_d} \sum_{t=1}^{T} \log P(w_t \mid d_D, w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) \tag{5}$$

where the probability function takes the form:

$$P(w_t \mid d_D, w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) = \frac{\exp\left(\mathbf{v'}_{w_t}^{T} \cdot \mathbf{h}\right)}{\sum_{w \in V} \exp\left(\mathbf{v'}_{w}^{T} \cdot \mathbf{h}\right)} \tag{6}$$

and the hidden context vector h now includes the document vector:

$$\mathbf{h} = \frac{1}{2z+1} \left(d_D + \sum_{-z \leqslant i \leqslant z, i \neq 0}^{z} \mathbf{v}_{w_{t+i}} \right) \tag{7}$$

θ_d refers to a set of parameters denoting all learnable parameters—i.e., v_w, $v_w^{'}$, and the vector for each document D in which the target word is present (d_D). This formulation ensures that words

[6]Section A.2 of the appendix includes a more detailed explanation of the mathematical foundations of Doc2Vec.

within the same document share the same document-level representation while also accounting for their local context. The joint optimisation of document and word embeddings allows the model to capture both global semantic themes and fine-grained linguistic patterns.

The embeddings obtained from the two algorithms—Word2Vec and Doc2Vec—are fed into a K-Means clustering model (MacQueen, 1967; Jain, 2010) to reduce the dimensionality of the vector space into features that can be meaningfully incorporated into a linear regression framework[7]. K-Means is an unsupervised machine learning algorithm that partitions a dataset $X = \{x_1, x_2, \ldots, x_n\}$ into k distinct, non-overlapping clusters $C = \{C_1, C_2, \ldots, C_k\}$ by minimising the within-cluster sum of squares:

$$\arg\min_{C} \sum_{\tau=1}^{k} \sum_{x \in C_\tau} \|x - \mu_\tau\|^2 \tag{8}$$

where μ_τ is the centroid of cluster C_τ which is computed as the mean of all embeddings in that cluster.

Nevertheless, the number of clusters k must be specified a priori, with the value of this parameter possibly having significant influence on the quality and interpretability of the resulting clusters. The silhouette score (Rousseeuw, 1987) offers a principled, data-driven approach to guide the selection of the optimal number of clusters k^* by iterating over a range of candidate k values and evaluating how well each data point fits within its assigned cluster. For each vector embedding x, the silhouette score is:

$$s(x) = \frac{b(x) - a(x)}{max(a(x), b(x))} \tag{9}$$

where $a(x)$ is the average distance between x and other points in the same cluster (intra-cluster distance), and $b(x)$ is the average distance to points in the nearest neighbouring cluster (inter-cluster distance). The overall silhouette score for a given k is the mean of all individual scores:

$$S_k = \frac{1}{n} \sum_{\iota=1}^{n} s(x_\iota) \tag{10}$$

The value of k that maximises S_k is selected as optimal. For the current analysis, the silhouette scores computed across the range of k values revealed that the optimal number of clusters for both Word2Vec and Doc2Vec embeddings is two. An analysis of the clusters' properties, based on the distribution of rental rates, showed that one cluster's distribution was skewed toward

[7] A more detailed explanation of the mathematical underpinnings of K-Means clustering and silhouette scores is available in section A.3 of the appendix.

higher rent levels relative to the other. Upon examining the descriptions of properties assigned to the higher-priced cluster, it was found that they contained a greater frequency of phrases and terms associated with quality.

Consequently, a binary indicator for the ”quality indicative phrasing” cluster was introduced, based on each of the embedding techniques. This unsupervised approach to analysing property descriptions eliminates the need for additional measures, such as rent levels, thereby reducing potential issues, such as bias and collinearity, associated with creating such “quality” indicators. The following sub-sections contrast the pricing characteristics of rental properties across various advertising platforms and types of landlords. This analysis is enhanced using the two “quality” indicators derived above, in an effort to assess the use of quality indicative phrasing by real estate agents and individual landlords, and thus better analyse contrasting pricing behaviours across the two advertising platforms.

3.1 Rental Pricing on Facebook Marketplace: Individuals vs. Agents

As a first step, I focus on how pricing behaviour may vary within Facebook Marketplace data. In particular, I regress rental prices found on Facebook Marketplace against a number of controls such as property characteristics, and most importantly in this case, against our quality indicators and against dummies differentiating between agency and individual or P2P adverts. The coefficients of interest are found in the first five rows of Table 2.

The main result indicates that advertised rents listed by real estate agencies on Facebook Marketplace are between 1.5% and 1.8% lower than those posted by individuals. This suggests potential information asymmetries between real estate agents and individual landlords. A comparison across three rental market databases reveals that rents in the Authority's database are generally lower than the advertised rents, implying that real estate agents may be more aligned with prevailing market rates due to their broader access to market information and greater experience. On the other hand, the generally lower rental rates posted by agencies on Facebook Marketplace may reflect the platform's role as a means of targeting lower-income prospective tenants. The magnitude of the difference in pricing behaviours between individual landlords and agents suggests that pricing strategies are the result of market targeting.

Table 2: Facebook Marketplace Data Regression Coefficients

	Agency Indicator	Agency & Doc2Vec Indicators	Agency & Word2Vec Indicators	Agency & Doc2Vec Interaction	Agency & Word2Vec Interaction
Agency	-0.015***	-0.018***	-0.015***		
Quality Phrasing					
Doc2Vec K-Means		0.037***			
Word2Vec K-Means			0.016***		
Agency with Quality Phrasing					
Doc2Vec K-Means				0.011	
Word2Vec K-Means					-0.008
Size					
Beds	0.131***	0.131***	0.131***	0.131***	0.131***
Baths	0.160***	0.159***	0.160***	0.160***	0.160***
Characteristics					
Garden	0.038***	0.035***	0.036***	0.037***	0.038***
Pool	0.326***	0.322***	0.325***	0.327***	0.326***
Seafront	0.093***	0.092***	0.092***	0.094***	0.093***
Has Views	0.107***	0.101***	0.105***	0.104***	0.107***
Garage	0.097***	0.097***	0.097***	0.098***	0.098***
Optional Garage	0.032***	0.028***	0.031***	0.033***	0.032***
Property Type					
Apartment	(b)	(b)	(b)	(b)	(b)
House	0.165***	0.168***	0.166***	0.166***	0.166***
Maisonette	-0.036***	-0.038***	-0.036***	-0.036***	-0.036***
Penthouse	0.118***	0.116***	0.117***	0.118***	0.118***
Room Only	-0.908***	-0.908***	-0.907***	-0.908***	-0.908***
Townhouse	0.079***	0.080***	0.079***	0.080***	0.079***
Villa	0.321***	0.323***	0.320***	0.322***	0.322***
Controls					
Locality Fixed Effects	Yes	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes
Observations	16404	16404	16404	16404	16404
BIC	3555	3514	3555	3563	3562
$* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$					

Note: This table displays the coefficients from a set of regressions of rental prices found on Facebook Marketplace against several controls such as property characteristics, quality indicators and dummies differentiating between agency and P2P adverts.
Source: Author's calculations.

The regression analysis reveals that phrasing techniques in advertisements are used to signal the quality of a property in a statistically significant manner. On average, the use of quality-indicative structuring throughout the entire advert correlates with an increase in advertised rental rates of 3.7%, while quality-indicative keywords alone is related to an increase in requested rents of 1.6%, compared to adverts that do not employ these techniques. This suggests that a comprehensive approach to styling an advert's phrasing has a greater positive correlation to asking prices by creating the perception of higher quality. The Bayesian Information Criterion (BIC) values suggest that the Doc2Vec-based indicator leads to more significant improvements in the overall fit of the regression model when compared to the Word2Vec-based indicator. On the contrary to Doc2Vec-based, a Word2Vec-based indicator takes in consideration the context of the entire advert. In this light, these results provide evidence supporting the idea that a holistic approach to the analysis of an advert is superior to one based solely on individual words that do not take in consideration the context of the wording. While the influence of descriptive text in property adverts on rental prices is well established in the literature (Lin et al., 2022), there is limited evidence on how specific textual phrasing impacts advertised rental prices.

The lower rental rates set by real estate agents remain consistent even when accounting for quality-indicative phrasing in the model. This suggests that the ability of agents to highlight quality is independent of the agencies' broader pricing strategies. This conclusion is further supported by the lack of sufficient evidence suggesting that real estate agencies make significantly more use of quality phrasing techniques, as indicated by the statistically insignificant interaction terms in Table 2[8]. In fact, for the identified quality phrasing clusters, the distribution is similar across both real estate agents and individual landlords for both embedding techniques.

The coefficients for property attributes, including size, characteristics, and property type, are largely consistent in both magnitude and sign across the regression specifications in Table 2. The estimated models suggest that an increase in the number of bathrooms raises rental rates more significantly than an increase in the number of bedrooms. The presence of certain features, such as a pool or views, also leads to substantial increases in rental rates, with pools and views increasing rents by 32.6% and 10.7% on average, respectively. The regressions further indicate that the least expensive and most expensive sole-renter property types are maisonettes and villas, respectively. Shared accommodations generally appear to be significantly more affordable than other property types. The number of shared spaces and their distinct distribution of rental rates suggests the existence of a niche market segment. To test the hypothesis however, a cross-platform analysis is needed to determine whether this finding is specific to the platform in question. Unfortunately, this analysis requires considerably more data on shared spaces than

[8]In a set of regressions with the same framework and regional fixed effects, displayed in section A.5 of the appendix, the interaction term becomes statistically significant. However, for a finding to be sufficiently robust similar coefficients across the two specifications are required.

what is currently available.

3.2 Comparing P2P and Agency Rental Adverts Across Platforms

The second part of the analysis compares the P2P listings on Facebook Marketplace with aggregated listings from real estate agencies. The first three columns of Table 3 present regression coefficients for the P2P segment, following the approach used in Table 2. The next three columns repeat the analysis for agency listings, while the final three pool data from both sources to compare agency behaviour across platforms. The key coefficients are reported in the first four rows of Table 3.

The findings from the previous section on Facebook Marketplace advertisements are largely consistent with the data presented in Table 3. Notably, the coefficients for the binary variables representing real estate agency involvement suggest that there are consistent differences in requested rental rates between platforms. These discrepancies may be attributed to the differences between the target audiences of the platforms, where Facebook Marketplace and the aggregated real estate agencies' listings are believed to cater to the low- and the high-end of the market, respectively. The coefficients in the final three columns of Table 3 indicate that, on average, rental rates requested by agents are approximately 2.5% lower on Facebook Marketplace and 2.7% higher in the aggregated agencies dataset, relative to those of P2P advertisements. These figures highlight the contrasting dynamics between the two platforms from the perspective of real estate agents. While Facebook Marketplace imposes no cost barriers to entry, such as subscriptions and commissions paid to specialised platforms, and attracts a broader audience, which appears to be skewed toward lower rental rates, platforms aggregating real estate agents' listings likely introduce cost barriers and serve a more selective audience of higher-spending viewers.

Table 3: Advertised Databases' Regression Coefficients[9]

	P2P Facebook Data			Aggregated Agencies Data[10]			Combined Advertised Data		
Agency									
Facebook							-0.023***	-0.029***	-0.024***
Aggregated Agencies							0.028***	0.026***	0.027***
Quality Phrasing									
Doc2Vec K-Means		0.051***			0.102***			0.061***	
Word2Vec K-Means			0.012***			0.020***			0.035***
Size									
Beds	0.126***	0.126***	0.126***	0.120***	0.120***	0.121***	0.132***	0.132***	0.132***
Baths	0.162***	0.161***	0.162***	0.143***	0.138***	0.143***	0.145***	0.142***	0.144***
Characteristics									
Garden	0.036**	0.033**	0.035**	0.031**	0.026*	0.029**	0.028***	0.025**	0.025***
Pool	0.304***	0.297***	0.303***	0.348***	0.340***	0.346***	0.337***	0.331***	0.335***
Seafront	0.095***	0.091***	0.095***	0.062***	0.065***	0.060***	0.081***	0.081***	0.080***
Has Views	0.107***	0.102***	0.105***	0.167***	0.155***	0.166***	0.131***	0.122***	0.126***
Garage	0.107***	0.107***	0.106***	0.061***	0.055***	0.060***	0.079***	0.078***	0.078***
Optional Garage	0.025**	0.020	0.024*	-0.015	-0.020	-0.015	0.006	0.001	0.005
Property Type									
Apartment	(b)	(b)	(b)	(b)	(b)	(b)	(b)	(b)	(b)
House	0.171***	0.172***	0.171***	0.169***	0.171***	0.168***	0.164***	0.167***	0.165***
Maisonette	-0.026***	-0.030**	-0.027***	-0.030***	-0.032***	-0.030***	-0.036***	-0.039***	-0.036***
Penthouse	0.114***	0.110***	0.113***	0.152***	0.147***	0.152***	0.132***	0.129***	0.130***
Room Only	-0.911***	-0.911***	-0.910***				-0.907***	-0.907***	-0.903***
Townhouse	0.094***	0.094***	0.094***				0.085***	0.086***	0.084***
Villa	0.353***	0.351***	0.352***	0.509***	0.514***	0.510***	0.405***	0.409***	0.403***
Controls									
Locality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11656	11656	11656	10993	10993	10993	27396	27396	27396
BIC	3381	3329	3386	4256	4003	4257	7916	7698	7850
$* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$									

Note: This table displays the coefficients from an exercise where advertised rental rates are regressed against property characteristics, dummies contrasting between agency and P2P listings, and dummies for adverts with quality indicative phrasing. The first three columns highlight the P2P segment on Facebook Marketplace, the middle three focus on an aggregation of agency listings, whilst the final three combine the two datasets.
Source: Author's calculations.

[9] The official register of rental contracts does not include any indicators distinguishing P2P contracts from those intermediated by real estate agents, and descriptions of the property. Therefore, the data from the official register was excluded from this type of analysis.

[10] The property type labels from the aggregated agencies dataset were grouped differently based on the definition of property types of the agents and property types with less than 100 observations were removed, leading to some incompatibility with the data from Facebook Marketplace.

Another key distinction between P2P listings on Facebook Marketplace and aggregated agency-listed advertisements, is the premium placed on the wording of adverts. For both clustering indicators derived from Word2Vec and Doc2Vec embeddings, the coefficient associated with higher-quality indicative wording is nearly twice as large for aggregated adverts found on agency platforms, compared to those in the P2P market. This suggests that agencies tend to place greater emphasis on descriptive language to enhance the perceived value of their listings. However, this heightened focus on wording may be influenced by unobserved quality attributes and the targeting of adverts towards a higher-end segment of the market. This is further indicated by a larger proportion of properties featuring desirable attributes, such as pools and gardens[11], and a greater share of advertisements using quality-signalling phrasing[12]. Furthermore, the increased use of quality-signalling language used by agencies may reflect the presence of a prescribed standard of writing set by the relevant platforms, as cosine similarity scores across the embedded property descriptions show that the listings of the aggregated agencies are more homogeneous than those in the P2P market[13].

The differing behaviours between the Doc2Vec and Word2Vec quality-signalling indicators observed in Table 2 are also evident in the specifications presented in Table 3. This further supports the intuition that a comprehensive approach to styling an advertisement's phrasing is more positively correlated with boosting requested rental rates by signalling quality in a more comprehensive manner, compared to using individual words to highlight higher-quality attributes. The coefficients for property attributes—such as size, characteristics, and property type—remain largely consistent in both magnitude and sign across the regression specifications in Tables 2 and 3. Key distinctions in the regression coefficients include the fact that listings of the aggregated agencies dataset, which feature a larger share of luxurious properties like penthouses and villas, command a larger premium relative to listings in the P2P market. Similarly, properties with views are valued more highly by the aggregated agencies' listings compared to the P2P market. However, the coefficients for bathrooms, seafront locations, and garages are lower in the aggregated agencies' regressions, likely because such features are more common, more variable in rent, and less strongly associated with rents than on Facebook Marketplace[14].

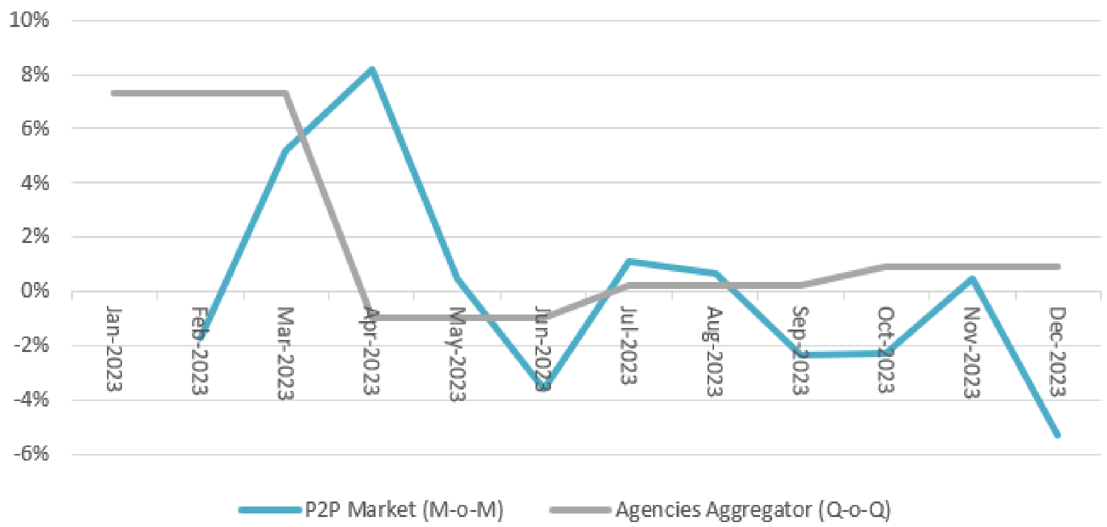
[11] The shares of the listed characteristics in the regressions are displayed in section A.4 of the appendix.

[12] For P2P and aggregated agencies' listings 20.19% and 32.02% of listings were flagged for including quality signalling phrasing when using Doc2Vec embeddings, whilst 74.18% and 80.89% were flagged when using Word2Vec embeddings.

[13] Cosine similarity contrasts vectors by computing the cosine of the angle between them, and score them from 0 (completely dissimilar) to 1 (identical direction). The medians of the similarity matrices for the embedded adverts were 0.55 and 0.94 within the agencies' dataset, versus 0.45 and 0.90 for the P2P data, using Doc2Vec and Word2Vec embeddings respectively.

[14] The regression results shown in section A.6 of the appendix indicate that when changing from locality to regional fixed effects the listed findings are largely consistent.

Figure 3: Comparison of X_t VS X_{t-1} Growth Rates in Advertised Databases



Note: The blue line shows month-on-month rental growth for listings by individual landlords on Facebook Marketplace, while the grey line shows quarter-on-quarter growth for aggregated agent listings.
Source: Author's calculations.

The hedonic regression model for the P2P rental market, excluding quality-related phrasing features, is utilised to construct an index capturing the evolution of rental rates throughout 2023. Figure 3 illustrates the month-on-month (M-o-M) growth rates of the estimated P2P market index, juxtaposed with the quarter-on-quarter (Q-o-Q) growth rates from the aggregated agencies dataset[15]. Both indices exhibit pronounced upward pressures toward the end of the first quarter of 2023, with this momentum extending through April in the P2P market series. From April onwards, rental rates as measured by the aggregated agencies' index stabilise considerably. Rental rates' growth as measured by the P2P index exhibit marginally more volatility with the growth figures fluctuating between positive and negative territories, especially in the second part of the year. The observed volatility in the P2P index is partially attributable to the higher frequency of data collection from Facebook Marketplace, impacting the volume and variability of listings.

The analysis of leader-follower dynamics necessitates further investigation within a robust econometric framework that incorporates a longer time span of multiple years. This could enhance the reliability of these insights and clarify the underlying dynamics of rental price evolution to assess whether the rental rates in the P2P market tend to follow the pricing behaviours of real

[15]The index calculated using the aggregated agencies' dataset is based on the methodology described in Brincat and Ghigo (2022) and is taken from the website of the Central Bank of Malta. Hence, the index for the P2P market was calculated using the time-dummies of the regression without embeddings-based indicators because this hedonic model is the most comparable methodology to that used to calculate the index posted by the Central Bank of Malta.

estate agencies. Given the differing pricing behaviours across different advertising platforms, an aggregation of listings from numerous platforms could also better explain the difference between advertised and contracted rental rates.

4 Conclusion

The findings of this research underscore the importance of considering the source of data when analysing rental markets. The comparative analysis in Table 1 reveals that advertised listings, particularly aggregated adverts posted by agencies on their online platforms, are skewed toward larger, higher-end properties located in Malta's most expensive regions. Specifically, 53% of listings in the aggregated agencies dataset include three or more bedrooms compared to 44% in the Housing Authority's official register, and 65% of aggregated agencies' listings are situated in the Northern Harbour region compared to 46% in the official data. Anecdotal evidence suggests that one factor contributing to this discrepancy is that the adverts of higher-priced and larger properties are posted multiple times in different formats to increase the chances of attracting lessees. As a result, these properties accumulate in the dataset, while other types of listings may enter and exit the platform too quickly to be captured if the frequency of data collection is too low. To mitigate this bias, data collection strategies should be carefully timed to account for the frequency with which listings appear and disappear from a platform, ensuring a more accurate representation of the rental market.

The choice of advertising platforms has important implications for market-oriented analyses. Price differences are also substantial. While the median rent for new contracts in the Housing Authority's register stands at €850, Facebook Marketplace listings and aggregated agencies' adverts report significantly higher medians of €1,400 and €1,500, respectively. Furthermore, the fraction of shared spaces—a more affordable rental segment—is notably under-represented in advertised data: 4% on Facebook Marketplace versus 15.4% in the official register. The exclusion of these properties in the aggregated agencies dataset has significant consequences for studies on rental market affordability in Malta, as regression results in Tables 2 and 3 show that shared accommodations are substantially more affordable than other property types. Nevertheless, this type of accommodation has largely been ignored in Maltese affordability studies, apart from Briguglio and Spiteri (2022) who focus on the experiences of young people. These differences between data sources underscore the potential biases introduced when using digital advertisements as proxies for the broader rental market, particularly for affordability studies.

The results further indicate that real estate agencies adapt their pricing strategies based on the advertising platform. Rental prices of agency-listed properties posted on Facebook Marketplace are between 1.5% to 2.9% lower than those posted on the same platform by individual landlords, suggesting agencies advertising on Facebook tend to target a more budget-conscious audience. Conversely, in the aggregated agencies dataset, listed properties are priced around 2.7% higher on average relative to P2P listings, consistent with a strategy aimed at the premium segment.

A key contribution of this research is the analysis of how property description phrasing influences requested rental prices. The embedding and clustering exercises demonstrate that these techniques effectively capture language intended to signal higher property quality. However, the results vary in magnitude depending on whether Word2Vec or Doc2Vec embeddings are used to construct quality-indicative phrasing indicators. Listings using comprehensive, high-quality descriptive phrasing—identified through Doc2Vec and Word2Vec embeddings—command rent premiums of between 3.7% and 10.2%, depending on the model and platform. Simpler use of quality-indicative keywords alone results in smaller premiums, between 1.2% and 3.5%. Additionally, while real estate agents receive more training on advertising strategies than individual landlords, the analysis finds no evidence that they use quality-indicative phrasing more frequently. Instead, this strategy appears to be more prevalent in more homogeneous markets, where it may reflect the presence of standards of communication. These results highlight the utility of such techniques in capturing previously neglected aspects of rental market advertising. Future research should expand on the analysis of these techniques by contrasting CBOW versus skip-gram rooted embedding approaches, optimising tools to determine the tuning of embedding and clustering approaches and contrasting the behaviours and attributes of other relevant embedding techniques.

The findings of this study are constrained by the limited number of advertising platforms analysed and the relatively short observation period, which restricts the generalisability of the results. Expanding the dataset to cover a longer time frame would provide a more comprehensive understanding of market dynamics and allow for a deeper investigation into the interactions between different market participants. Specifically, a longer time horizon would enable an analysis of potential leader-follower dynamics in pricing strategies, examining whether real estate agents influence the pricing behaviour of individual landlords. Additionally, a broader dataset would facilitate a more rigorous assessment of information asymmetries among landlords, shedding light on how differences in market knowledge and experience affect rental pricing decisions. Future research incorporating multiple platforms and extended time series data could offer valuable insights into the evolution of rental market structures and pricing mechanisms.

By highlighting the influence of data sources, platform selection, and linguistic strategies on rental pricing, this research provides a more nuanced understanding of the factors shaping the rental market. As digital advertising continues to play an increasingly central role in property listings, future studies must account for the biases introduced by different platforms and the evolving strategies of market participants. A more comprehensive and dynamic approach to rental market analysis—enabled by the availability of data and tools such as the Authority's Rental Market Insights Dashboard—will be crucial for formulating evidence-based policies that enhance affordability, transparency, and efficiency in the housing sector.

Bibliography

Benefield, J. D., Cain, C. L., and Johnson, K. H. (2011). On the relationship between property price, time-on-market, and photo depictions in a multiple listing service. *The Journal of Real Estate Finance and Economics*, 43(3):401–422.

Briguglio, M. and Spiteri, G. (2022). Housing affordability: A focus on young people in malta. `https://www.um.edu.mt/library/oar/bitstream/123456789/102031/1/Housing_affordability.pdf`. Accessed: 2025-02-04.

Brincat, E. M. and Ghigo, M. (2022). Constructing a new advertised house price index. Policy Paper 04/2022, Central Bank of Malta.

Camilleri, J. (2020). The impact of regulation on short-term rental markets: Evidence from malta. *Journal of Policy & Economics*, 34(2):412–429.

Ellul Dimech, R. (2019). Short-term rentals in malta: A look at airbnb listings. Policy Paper 05/19, Central Bank of Malta.

European Commission (2023). Eurobarometer: Social media use in malta. `https://europa.eu/eurobarometer/surveys/detail/3153`. Accessed: 2025-06-11.

Fearne, R. (2021). An analysis of the distribution and price determinants of airbnb rentals in malta. *International Journal of Housing Markets and Analysis*, 15(1):231–246.

Gauci, T., Micallef, B., and Fenech, R. (2022). Two years after the 2020 reform: an overview of the rental market. In *The Annual Malta Residential Rental Study: Second Edition*, pages 56–71. Housing Authority.

Geltner, D., Miller, N., Clayton, J., and Eichholtz, P. (2013). *Commercial Real Estate Analysis and Investments*. OnCourse Learning, Mason, 3 edition.

Goodchild, B. and Ferrari, E. (2021). Intermediaries and mediators: An actor-network understanding of online property platforms. *Housing Studies*, pages 1–22.

Gyllenberg, J. and Koppfeldt, M. (2020). Negotiating residential rents: tenant and landlord strategies in different economic contexts. *Journal of Property Research*, 37(2):155–173.

Han, L. and Strange, W. C. (2016). What is the role of the asking price for a house? *Journal of Urban Economics*, 93(C):115–130.

Housing Authority (2024). 2023 h2 update: Registered rental contracts in malta. `https://housingauthority.gov.mt/wp-content/uploads/2024/05/Rent-Report-2023-H2-Update-Housing-Authority.pdf`. Accessed: 2024-07-24.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666.

Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1188–1196.

Lin, Z., Zhang, X., and Zhang, Y. (2022). Online advertising and real estate sales: Evidence from china. *Electronic Commerce Research*, 23(1):1–24.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297.

Micallef, B. (2021). The long-lasting legacy of rent controls: Perspectives on the private rental market in malta within the context of a dual market. *International Journal of Real Estate Studies*, 15(2):43–54.

Micallef, B., Ellul, R., and Debono, N. (2021). A hedonic assessment of the relative importance of structural, locational and neighbourhood factors on advertised rents in malta. *International Journal of Housing Markets and Analysis*, 15(1):203–230.

Micallef, B. and Gauci, T. (2022). Research bulletin. Technical report, Central Bank of Malta, Valletta, Malta.

Micallef, B. and Spiteri, S. (2023). Research bulletin. Technical report, Central Bank of Malta, Valletta, Malta.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. `https://arxiv.org/abs/1301.3781`. arXiv preprint arXiv:1301.3781.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sirmans, C. F. and Slade, B. A. (2010). Sale–leaseback transactions: Price premiums and market efficiency. *Journal of Real Estate Research*, 32(2):221–242.

A Appendix

A.1 A Comparison Across Registered Rental Contracts

Table 4: A Comparison Across Registered Rental Contracts

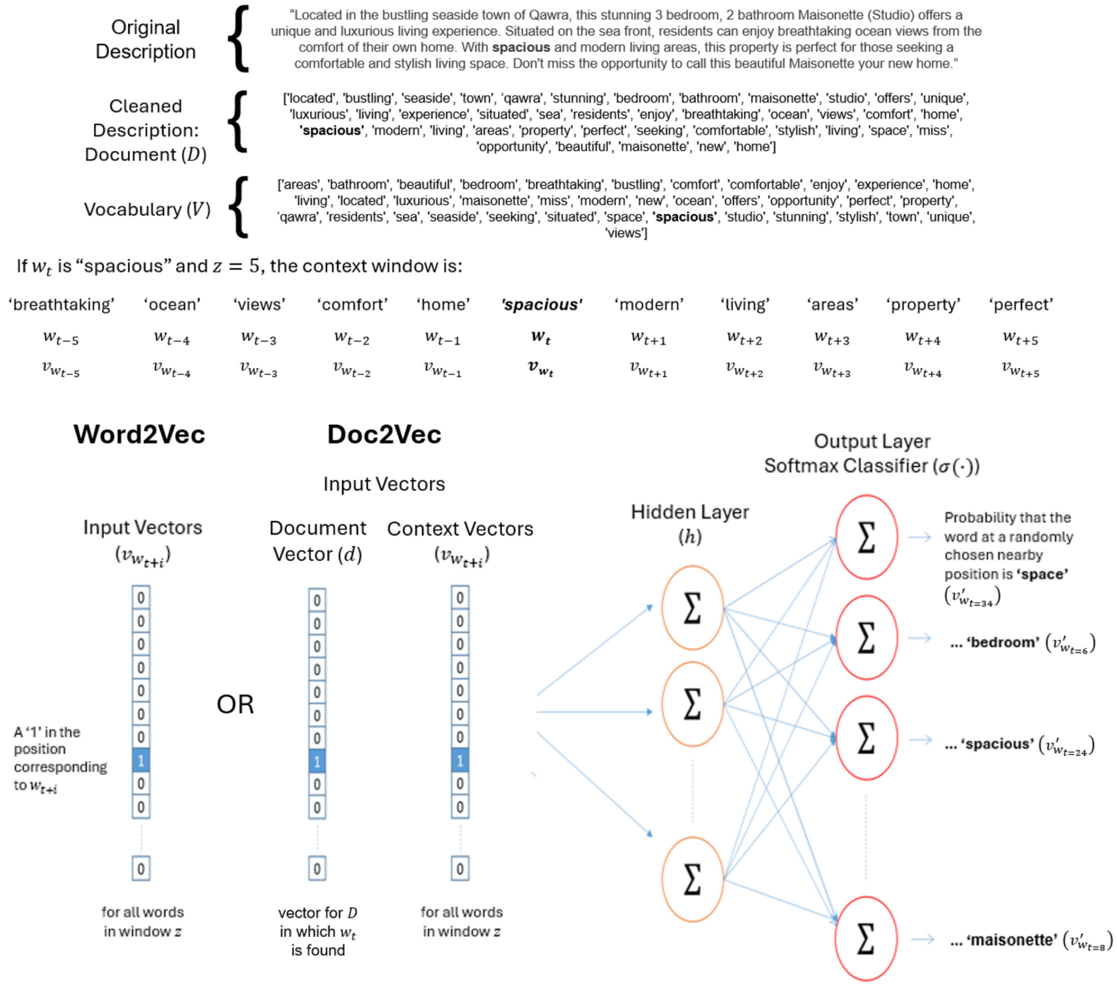
	New Contracts Only	All Active Contracts	Diff.
Observations	32,684	60,339	-27,655
Type of property			
Apartment/ Penthouse	68.45%	72.76%	-4.31%
Maisonette	9.73%	6.72%	3.01%
House	6.10%	10.71%	-4.61%
Villa	0.32%	0.41%	-0.09%
Room Only	15.41%	9.40%	6.01%
Number of bedrooms			
1	20.05%	20.68%	-0.63%
2	36.16%	39.40%	-3.24%
3	34.40%	33.43%	0.97%
4+	9.39%	6.49%	2.90%
Region			
Northern Harbour	45.50%	43.45%	2.05%
Northern	22.30%	23.91%	-1.61%
South Harbour	10.30%	10.53%	-0.23%
Southern Eastern	9.59%	9.62%	-0.03%
Gozo	6.47%	7.06%	-0.59%
Western	5.83%	5.53%	0.30%

Note: The table shows the portions of observations dedicated to the listed categories in the respective datasets. The 'Diff.' column highlights the differences between the characteristics of newly registered contracts relative to the characteristics of all active rental contracts in 2023.
Source: Own calculations based on the properties of the collected datasets.

A.2 Mathematical Context of Word2Vec and Doc2Vec

As indicated in Figure 4, Word2Vec is a neural network-based model that converts words into dense vector representations. These vector representations are referred to as embeddings, where an embedding is a numerical representation of a word, sentence, or document in a high-dimensional vector space, where similar meanings are mapped closer together based on contextual relationships. Therefore, Word2Vec focuses on the similarities between words in a corpus of text to represent words in a vector space, such that similar words have closer embeddings.

Figure 4: Outline of Word2Vec and Doc2Vec



Note: This diagram indicates the procedure of text cleaning, window selection and the creation of textual embeddings undertaken for the corpus of property descriptions.

In this implementation Word2Vec learns word representations using a shallow neural network trained on large text corpora using the CBOW architecture. In this model a word (w_t) is predicted given a sequence of words $w_1, w_2, \ldots, w_T$ by using its surrounding context words $(w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z})$, with the objective function of maximising the log-likelihood of predicting w_t given its surrounding context words within a window of size z without constraints, such that:

$$\max_{\theta_w} \sum_{t=1}^{T} \log P(w_t \mid w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) \tag{11}$$

where:

$$P(w_t \mid w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) = \frac{\exp\left(\mathbf{v'}_{w_t}^{T} \cdot \mathbf{h}\right)}{\sum_{w \in V} \exp\left(\mathbf{v'}_{w}^{T} \cdot \mathbf{h}\right)} \tag{12}$$

$$\mathbf{h} = \frac{1}{2z} \sum_{-z\leqslant i\leqslant z, i\neq 0}^{z} \mathbf{v}_{w_{t+i}} \tag{13}$$

The maximisation implies that the set of parameters—the learned word embeddings represented by $\theta_w = \{\mathbf{v}_w, \mathbf{v}'_w \mid w \in V\}$—are optimised to maximise likelihood while using a log transformation to make the computation numerically stable and turn the product of probabilities into a sum. The probability functions use an exponential form to ensure non-negative probabilities and to assign higher likelihoods to similar word pairs. The numerator represents the unnormalised probability of w_t appearing in the given context, which is calculated using the dot product of the target word vector ($\mathbf{v}_{w_t}$) and the context vector (h), measuring how well w_t aligns with the given context. The denominator takes the sum over the vocabulary of size $\mathbf{V}$ to normalise the probabilities. h is the hidden layer representation, computed as the average of the vectors of the context words ($v_{w_{t+i}}$). This softmax function assigns a high probability to words that are more semantically relevant to the given context. The higher the dot product $v_{w_t}^T \cdot h$, the higher the probability of w_t being the correct word. The denominator scales the probabilities, ensuring that the sum of probabilities over all possible words equals one.

Alternatively, the maximisation function of Word2Vec can be expressed in more detail as:

$$\begin{gathered}\max_{\theta_w} \sum_{t=1}^{T} \ln \sigma \left(\mathbf{v}'_{w_t} \cdot \left(\frac{1}{2z} \sum_{\substack{-z\leq i\leq z\\ i\neq 0}} \mathbf{v}_{w_{t+i}} \right) \right) + \\ \sum_{j=1}^{q} \mathbb{E}_{w_j\sim P_n(w)} \ln \sigma \left(-\mathbf{v}'_{w_j} \cdot \left(\frac{1}{2z} \sum_{\substack{-z\leq i\leq z\\ i\neq 0}} \mathbf{v}_{w_{t+i}} \right) \right)\end{gathered} \tag{14}$$

The more detailed representation highlights that, given a context, the model computes the average of the input vectors $\mathbf{v}_{w_i}$ for each context word and maximises the similarity between this average vector and the output vector $\mathbf{v}'_{w_t}$ of the target word. This similarity is measured by the sigmoid of the dot product $\sigma\left(\mathbf{v'}_{w_t}^{\top} \cdot \mathbf{h}\right)$. To make training computationally efficient, negative sampling is employed as an approximation to full softmax. For each genuine (target, context) pair, the model samples q noise words w_j from a noise distribution $P_n(w)$ and minimises the probability $\sigma\left(-\mathbf{v'}_{w_j}^{\top} \cdot \mathbf{h}\right)$. This distribution P_n is a smoothed unigram distribution—often the word frequency raised to the power of 0.75—which reduces the dominance of extremely common words in the learning process.

Conversely, Figure 4 demonstrates that Doc2Vec extends the architecture of Word2Vec to entire documents, such that the full property description in an advert (D) is represented as a vector (d_D). The document vector is used within the objective function to predict the context words within a DM framework. Therefore, the goal of the model is to learn a fixed-length document vector that encapsulates the semantic meaning of the entire document which is then used to predict the context words in the training corpus. The objective function maximises the log-likelihood of predicting w_t given its surrounding context words within a window of size z with respect to the vector representation of the document, such that:

$$\max_{\theta_d} \sum_{t=1}^{T} \log P(w_t \mid d_D, w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) \tag{15}$$

where:

$$P(w_t \mid d_D, w_{t-z}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+z}) = \frac{\exp\left(\mathbf{v'}_{w_t}^{T} \cdot \mathbf{h}\right)}{\sum_{w \in V} \exp\left(\mathbf{v'}_{w}^{T} \cdot \mathbf{h}\right)} \tag{16}$$

$$\mathbf{h} = \frac{1}{2z+1} \left(d_D + \sum_{-z \leqslant i \leqslant z, i \neq 0}^{z} \mathbf{v}_{w_{t+i}} \right) \tag{17}$$

Within Doc2Vec h represents a hidden vector that is derived from the context, combining d_D and the word vectors of the context words ($v_{w_{t+i}}$), such that words within the same document have the same document-level context from d_D but varying representations of their immediate contexts from $v_{w_{t+i}}$. Unlike the Word2Vec CBOW approach, which only uses context word vectors, the DM Doc2Vec model incorporates a document-specific vector d_D in the softmax function, allowing words to be predicted based on both local context and global document meaning. Averaging ensures that the influence of document and word vectors is balanced. This allows the model to learn parameters $\theta_d = \{d_D, \mathbf{v}_w, \mathbf{v}'_w \mid w \in V\}$.

More formally, the unconstrained optimisation function to maximise the likelihood can be expressed as:

$$\begin{gathered} \max_{\theta_d} \sum_{t=1}^{T} \ln \sigma \left(\mathbf{v}'_{w_t} \cdot \left(\frac{1}{2z+1} \left(d_D + \sum_{\substack{-z \leq i \leq z \\ i \neq 0}} \mathbf{v}_{w_{t+i}} \right) \right) \right) + \\ \sum_{j=1}^{q} \mathbb{E}_{w_j \sim P_n(w)} \ln \sigma \left(-\mathbf{v}'_{w_j} \cdot \left(\frac{1}{2z+1} \left(d_D + \sum_{\substack{-z \leq i \leq z \\ i \neq 0}} \mathbf{v}_{w_{t+i}} \right) \right) \right) \end{gathered} \tag{18}$$

The function shows that the objective remains to maximise the probability of w_t given $\sigma\left(\mathbf{v}'_{w_t} \cdot \mathbf{h}\right)$, while minimising the probability of randomly sampled "noise" words, using negative sampling. This formulation allows d_D to act as a memory of the document's topic or semantic content, improving the model's ability to capture contextual meaning across the text.

A.3 Mathematical Context of K-Means Clustering

Given a dataset of embeddings $X = \{x_1, x_2, \ldots, x_n\}$, with two representations of X being extracted from the embedding models for n property descriptions, K-Means seeks to partition the data into k clusters by minimising the within-cluster sum of squares. Each cluster in a set of clusters $\mathcal{C} = \{C_1, C_2, \ldots, C_{k^*}\}$ represents a partition of X, where each C_i is a set of points assigned to cluster i ($C_i \subset X$), clusters are disjoint ($C_i \cap C_j = \emptyset$ for $i \neq j$), and every point belongs to one cluster ($\bigcup_{i=1}^{k} C_i = X$). Within this mathematical context, the within-cluster minimisation is expressed as:

$$\arg\min_{\mathcal{C}} \sum_{\tau=1}^{k} \sum_{x \in C_\tau} \|x - \mu_\tau\|^2 \tag{19}$$

where μ_τ is the centroid of cluster C_τ, which is computed as:

$$\mu_\tau = \frac{1}{|C_\tau|} \sum_{x \in C_\tau} x \tag{20}$$

The outer sum of the minimisation function aggregates over each cluster C_τ. The inner sum aggregates the squared Euclidean distances from a point x to the cluster centre over all points in cluster C_τ.

The optimal number of clusters k^* is determined by using the silhouette scores calculated by iterating over an expansive grid of k values, $K = \{k_1, k_2, \ldots, k_\eta\}$, such that the number of clusters maximises intra-cluster similarity and inter-cluster separation.

The number of clusters k^* is found by first computing the intra-cluster distance $a(x)$, where for each embedding x this is the average distance between x and all other points in the same cluster:

$$a(x) = \frac{1}{|C_x| - 1} \sum_{\substack{y \in C_x \\ y \neq x}} \|x - y\| \tag{21}$$

where C_x is the cluster to which x belongs, and $\|x - y\|$ is the Euclidean distance between points x and y.

The second step is to compute the inter-cluster distance, which considers the average distance between x and all other points γ in the nearest neighbouring cluster C_ρ:

$$b(x) = \min_{C_\rho \neq C_x} \frac{1}{|C_\rho|} \sum_{\gamma \in C_\rho} \|x - \gamma\| \tag{22}$$

The figures from these two steps are combined to compute the silhouette score for embedding x:

$$s(x) = \frac{b(x) - a(x)}{\max\left(a(x), b(x)\right)} \tag{23}$$

where $s(x)$ ranges from -1 to 1, denoting the likelihood of a point being misclassified or well-clustered, respectively.

For a tested number of clusters K, the overall silhouette score for the entire dataset is obtained by averaging over all n points:

$$S_k = \frac{1}{n} \sum_{\iota=1}^{n} s(x_\iota) \tag{24}$$

Therefore, k^* is determined by selecting the number of clusters k which corresponds to the largest value of S_k.

A.4 Shares of Property Attributes Across Advertised Listings

Table 5: Property Attributes of Rental Listings

	Aggregated Real Estate Agencies	Facebook Marketplace Total	P2P Market	Agents' Listings
Seafront	14.93%	7.21%	7.19%	7.29%
Having Views	31.94%	29.81%	26.11%	38.94%
Gardens	5.01%	3.05%	3.10%	2.93%
Pools	7.17%	4.98%	5.41%	3.94%
Garages	7.60%	5.44%	5.72%	4.74%
Optional Garages	4.42%	3.99%	4.26%	3.35%

Note: The table shows the portions of observations dedicated to the listed categories in the respective datasets.
Source: Author's calculations.

A.5 Facebook Marketplace Dataset with Regional Fixed Effects Regressions

Table 6: Facebook Marketplace Data Regression Coefficients

	Agency Indicator	Agency & Doc2Vec Indicators	Agency & Word2Vec Indicators	Agency & Doc2Vec Interaction	Agency & Word2Vec Interaction
Agency	-0.017***	-0.022***	-0.018***		
Quality Phrasing					
Doc2Vec K-Means		0.045***			
Word2Vec K-Means			0.021***		
Agency with Quality Phrasing					
Doc2Vec K-Means				0.015*	
Word2Vec K-Means					-0.010*
Size					
Beds	0.122***	0.122***	0.122***	0.122***	0.122***
Baths	0.170***	0.168***	0.170***	0.170***	0.170***
Characteristics					
Garden	0.046***	0.042***	0.044***	0.045***	0.047***
Pool	0.337***	0.332***	0.336***	0.339***	0.338***
Seafront	0.134***	0.133***	0.133***	0.135***	0.134***
Has Views	0.122***	0.114***	0.119***	0.118***	0.122***
Garage	0.114***	0.113***	0.113***	0.115***	0.114***
Optional Garage	0.043***	0.038***	0.042***	0.044***	0.044***
Property Type					
Apartment	(b)	(b)	(b)	(b)	(b)
House	0.139***	0.141***	0.139***	0.140***	0.139***
Maisonette	-0.054***	-0.057***	-0.055***	-0.055***	-0.054***
Penthouse	0.102***	0.099***	0.100***	0.101***	0.102***
Room Only	-0.933***	-0.933***	-0.931***	-0.933***	-0.933***
Townhouse	0.060***	0.061***	0.060***	0.061***	0.061***
Villa	0.306***	0.308***	0.304***	0.307***	0.307***
Controls					
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes
Observations	16404	16404	16404	16404	16404
BIC	4601	4523	4595	4610	4611
$* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$					

Note: This table mirrors Table 2, with the distinction of using regions instead of localities for location fixed effects.
Source: Author's calculations.

A.6 Advertised Listings Database Regressions with Regional Fixed Effects

Table 7: Advertised Databases' Regression Coefficients

	P2P Facebook Data			Aggregated Agencies Data			Combined Advertised Data		
Agency									
Facebook							-0.027***	-0.035***	-0.028***
Aggregated Agencies							0.046***	0.043***	0.046***
Quality Phrasing									
Doc2Vec K-Means		0.067***			0.123***			0.076***	
Word2Vec K-Means			0.021**			0.012***			0.041***
Size									
Beds	0.117***	0.118***	0.117***	0.111***	0.113***	0.111***	0.123***	0.123***	0.123***
Baths	0.172***	0.170***	0.173***	0.161***	0.153***	0.161***	0.156***	0.152***	0.156***
Characteristics									
Garden	0.044***	0.040**	0.043**	0.034**	0.027*	0.032**	0.033***	0.029***	0.030***
Pool	0.317***	0.307***	0.315***	0.348***	0.338***	0.347***	0.344***	0.336***	0.341***
Seafront	0.134***	0.129***	0.133***	0.107***	0.108***	0.105***	0.124***	0.123***	0.123***
Has Views	0.125***	0.118***	0.121***	0.198***	0.181***	0.197***	0.151***	0.140***	0.145***
Garage	0.125***	0.124***	0.123***	0.079***	0.072***	0.079***	0.096***	0.094***	0.094***
Optional Garage	0.035***	0.030**	0.034**	-0.002	-0.01	-0.003	0.019**	0.013	0.016*
Property Type									
Apartment	(b)	(b)	(b)	(b)	(b)	(b)	(b)	(b)	(b)
House	0.137***	0.138***	0.137***	0.142***	0.148***	0.142***	0.140***	0.144***	0.140***
Maisonette	-0.048***	-0.052**	-0.049***	-0.058***	-0.059***	-0.058***	-0.057***	-0.060***	-0.058***
Penthouse	0.098***	0.093***	0.096***	0.124***	0.119***	0.124***	0.112***	0.109***	0.109***
Room Only	-0.946***	-0.946***	-0.944***				-0.935***	-0.934***	-0.929***
Townhouse	0.074***	0.074***	0.074***				0.063***	0.065***	0.062***
Villa	0.330***	0.328***	0.329***	0.450***	0.459***	0.450***	0.376***	0.381***	0.374***
Controls									
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11656	11656	11656	10993	10993	10993	27396	27396	27396
BIC	3976	3882	3974	5191	4826	5194	10270	9934	10180
$* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$									

Note: This table mirrors Table 3, with the distinction of using regions instead of localities for location fixed effects.
Source: Author's calculations.