Final Term Project:
The Price of the Location in Athens Airbnb

Course: Data Analysis 2 & Coding 2
Student: Konstantinos Evagorou, 2508673

1. Introduction:

For this final term project, I chose to analyze the "Pricing and Sales" topic, utilizing data from the short-term rental market in Athens, Greece. As a Greek Cypriot who visits Greece and specifically Athens usually, my objective is to investigate the relationship between the listing price of an Airbnb apartment (y) and its distance from the city center (x).

- **Research Question : "**How does the distance from the city center (Syntagma Square) affect Airbnb listing prices in Athens , holding other factors constant?"

The intuition of that, follows the economic "bid-rent" theory, were properties closer to the Central Business District (CBD) or major tourist attraction command higher prices due to demand, so I expect to find a negative relationship between distance and price.

2. Data Presentation and Feutures:

- **Filtration:** The raw dataset in the beginning contained over 15000 observations, but to focus on the typical market I chose to filter out some of the observations like (non-operative listings whos price was 0) or some extreme outliers with (price > 1000 euro per night).

- **Sample Size:** After the cleaning I mentioned above the final dataset consists of 14524 observations providng high statistical power.

- **Outcome Variable (y) price:** From the **plot** (Appendix, Figure 1) produced by the code we can observe that the distribution of prices was highly right-skewed, thus I utilized the natural logarithm of price for my regression analysis.

- **Key explanatory Variable (x) distance:** The raw data that I used provided only cordinated (latidude/longtitude), so in my code I tried to calculate the haversine distance in km from each listing to Syntagma Square.

3. Model:

The strategy I used to answer the research question is by estimating a linear regression and given the log-transformation of the dependent varisable, I used a log-linear specification:

$$\mathbf{Ln(Price) = \beta 0 + \beta 1 Distance + \beta 2 Z + u}$$

The three models I used to ensure the stability of the results:

- **Model 1:** Simple Bivariate regression (ln_price ~ distance)
- **Model 2:** I added some basice property characteristics (accomodates, room_type)
- **Model 3:** Final model that adds quality signals (number_of_reviews, review_scores_rating)

Core Results: The results from the prefered secification (Model 3) are presented inthe Appendix (Table 2). The coefficient for distance is -0.257 and this result is highly statistcally significant with ($p < 0.001$ and $t = -57.8$).

Since this is a a log-level model, the coefficient approximated the percentage change. Holding other factors constant (like capacity room type and review rating), for every additional kilometer a listing is located away from Syntagma Square, the price decreases by approximately 25.7%.

4. **Generalization and External Validity:**
To check the robustness of my finding, I compared distance across all three models wih the results shown as below (Table 3):

- **Model 1 :** -0.298
- **Model 2 :** -0.287
- **Model 3 :** -0.257

This shows that the coefficient remaining stable and negative acroos all specifications.
External Vlidity: This results are highly relevant for Athens which is a monocentric city where most of the tourism is concentrated around specific historical sites. In addition this pattern likely generalizes to other historic European Capitals but might differ in my oppinion in polucentric cities like London or Los-Angeles).

5. **Casual Interpratation:**
In my oppinion, while the association is strong and robust we cannot claim casuality for sure and we must be cautious. There may be omitted variable bias like closer access to transportation systems or specific views (e.g. Acropolis view) which is something we cannot control. However, given the large sample size I used and controls for quality, we can confidently claim that locatoion is a primary casual driver of price differences in this market.

6. **Conclusion:**
In this project I analyzed 14,524 Airbnb listings in Athens to quantify the value of location. The important key findings is that there is a statistically significant negative relationship between distance and price. Also the “Distance penalty” is apx. 25.7% per km and that the relationship is nonlinear as the price drops steeply within the first 1-2km but then flattens out (Appendix Figure 3)

- **Business Recommendations:** For investors properties within 1km of the center yield the highest revenue per night commanding a massive premium and for budget-conscious travelers choosing a listing just 2-3km away can reduce their accommodation cost nearly 50%.

Appendix: Table 1:

Summary Statistics:

	price	distance	accommodates	number_of_reviews
count	14524.0	14524.00	14524.00	14524.00
mean	122.9	1.62	3.74	57.99
std	452.3	0.92	1.94	100.59
min	9.0	0.03	1.00	0.00
25%	54.0	0.98	2.00	3.00
50%	78.0	1.47	4.00	16.00
75%	119.0	2.08	4.00	65.00
max	38000.0	6.20	16.00	1030.00

Figure 1:

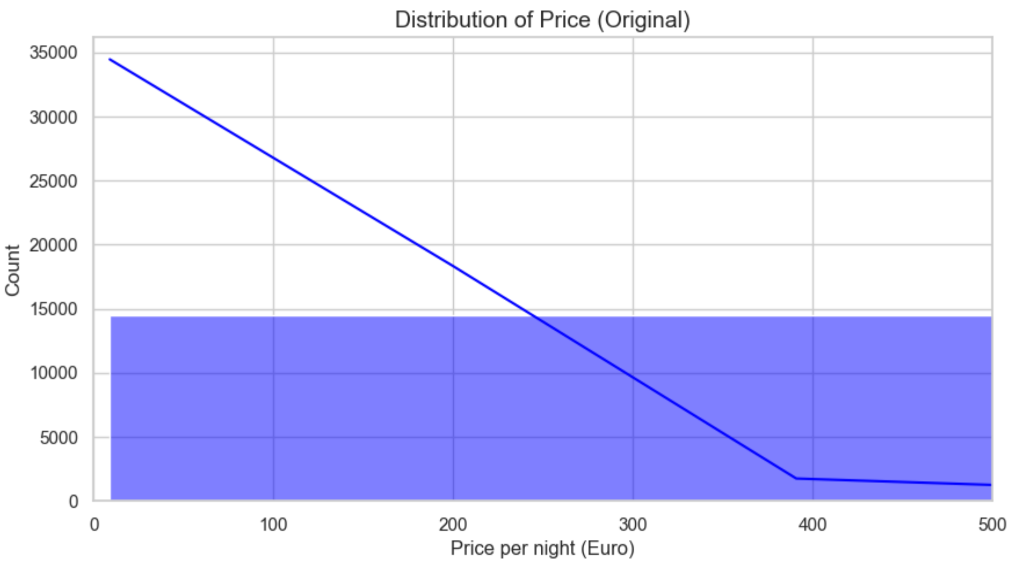

Figure 2:

Table 2:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               ln_price   R-squared:                       0.426
Model:                            OLS   Adj. R-squared:                  0.426
Method:                 Least Squares   F-statistic:                     1312.
Date:                Thu, 11 Dec 2025   Prob (F-statistic):               0.00
Time:                        12:45:01   Log-Likelihood:                -7674.2
No. Observations:               12383   AIC:                         1.536e+04
Df Residuals:                   12375   BIC:                         1.542e+04
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Intercept                     3.3046      0.055     60.148      0.000       3.197       3.412
room_type[T.Hotel room]       0.4043      0.072      5.598      0.000       0.263       0.546
room_type[T.Private room]    -0.2033      0.019    -10.638      0.000      -0.241      -0.166
room_type[T.Shared room]     -0.9633      0.080    -12.092      0.000      -1.120      -0.807
distance                     -0.2572      0.004    -57.821      0.000      -0.266      -0.248
accommodates                  0.1357      0.002     64.113      0.000       0.132       0.140
number_of_reviews            -0.0005   3.88e-05    -12.449      0.000      -0.001      -0.000
review_scores_rating          0.2160      0.011     19.200      0.000       0.194       0.238
==============================================================================
Omnibus:                     1148.782   Durbin-Watson:                   1.746
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1831.742
Skew:                           0.690   Prob(JB):                         0.00
Kurtosis:                       4.282   Cond. No.                     2.48e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.48e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Table 3:

```
--- ROBUSTNESS CHECK: Distance Coefficient ---
Model 1 (Simple): -0.2981
Model 2 (+Controls): -0.2873
Model 3 (+Quality): -0.2572
```

Figure 3: