Statistical Question/Hypothesis:

What factors affect the Airbnb rental prices in Europe?

Variables:

Price: The nightly price of the rental.

Attractions\_rating: The rating of tourist attractions near the rental.

City\_dist: The distance from the city center in km.

Satisfaction\_rating: The satisfaction rating of the guests.

Bedrooms: The number of bedrooms in the rental.

Histogram Analysis:

Price: The histogram shows a right-skewed distribution with outliers on the higher end of the price range.

Attractions\_rating: The histogram shows a normal distribution with no outliers.

City\_dist: The histogram shows a right-skewed distribution with no outliers.

Satisfaction\_rating: The histogram shows a left-skewed distribution with no outliers.

Bedrooms: The histogram shows a right-skewed distribution with outliers on the higher end of the number of bedrooms.

PMF Analysis:

Two scenarios were compared based on the number of bedrooms and the type of room. The first scenario compared private rooms vs. shared rooms based on price, and the second scenario compared the number of bedrooms vs. the room type based on PMF.

CDF Analysis:

The cumulative distribution function was created to show the distribution of the price variable.

Analytical Distribution Analysis:

A normal distribution was applied to the price variable to show how the data fits the normal distribution curve.

Scatter Plot Analysis:

Two scatter plots were created to analyze the correlation between price and attractions\_rating and city\_dist variables. The scatter plot of price vs. attractions\_rating shows a weak positive correlation between the variables, and the scatter plot of price vs. city\_dist shows a weak negative correlation between the variables.

Regression Analysis:

A regression analysis was conducted on the dependent variable, price, and the explanatory variables, attractions\_rating, city\_dist, satisfaction\_rating, and bedrooms. The analysis shows that all variables are statistically significant in predicting the rental prices in Europe. The coefficients of determination (R-squared) are 0.31, 0.46, 0.52 and 0.55 respectively for our Linear, 2, 3 and 4 degree polynomial regressions, which means that our best model being the 4 degree Polynomial regression explains 55% of the variability in rental prices can be explained by the model.

To better understand the distribution of the numerical variables, a logarithmic transformation was applied to the data and histograms were plotted. The resulting histograms revealed that some of the variables were highly skewed, indicating the presence of outliers in the dataset.

To identify and remove outliers, the z-score method was used with a threshold of 3 standard deviations away from the mean. Based on this, values over 1263 were identified as outliers and removed from the dataset. The resulting summary statistics revealed the mean, standard deviation, skewness, and other descriptive characteristics of the remaining variables.

To compare two scenarios using the PMF method, the number of bedrooms was used as the variable of interest, and two different scenarios were created based on whether the room type was private or shared. The resulting PMF plots showed the distribution of the number of bedrooms for each scenario.

A CDF plot was also created for the price variable, which showed the cumulative distribution of prices in the dataset. Additionally, an analytical distribution was plotted using the normal distribution with the calculated mean and standard deviation of the price variable.

Finally, two scatter plots were created to compare the price variable with the attractions rating and city distance variables. The scatter plots showed a positive correlation between price and attractions rating, indicating that properties located near popular attractions tend to have higher prices. However, there was no clear correlation between price and city distance, indicating that other factors may be influencing pricing decisions.

[Title slide with the project title and subtitle] Project Title: Airbnb Prices in Europe Subtitle: An Analysis of the Impact of Variables on Price

[Slide 1: Introduction] Statistical Question/Hypothesis: How do Airbnb prices vary across cities in Europe? Which variables have the greatest impact on the price of a property?

[Slide 2: Variables] Minimum of 5 variables in the dataset were used for analysis. Variables considered that could have an impact on the question are:

* Price: The price per night for a property
* Room Type: The type of room available - private room, shared room or entire apartment/house
* Host Superhost: Whether the host is classified as a superhost or not
* Number of Bedrooms: The number of bedrooms available in the property
* City Distance: The distance of the property from the city center

[Slide 3: Variable Descriptions] The variables in the dataset are described below:

* Price: the price per night for a property
* Room Type: the type of room available - private room, shared room or entire apartment/house
* Host Superhost: whether the host is classified as a superhost or not
* Number of Bedrooms: the number of bedrooms available in the property
* City Distance: the distance of the property from the city center

[Slide 4: Histograms] Histograms of the variables are shown below, outliers are identified and explained for each variable:

* Price: Outliers occur when the price is greater than 1263.905548107905. These outliers were removed using the z-score method.
* Room Type: No outliers were identified.
* Host Superhost: No outliers were identified.
* Number of Bedrooms: Outliers occur when the number of bedrooms is greater than 6. These outliers were not removed.
* City Distance: Outliers occur when the distance from the city center is greater than 11.0376044159004. These outliers were not removed.

[Slide 5: Descriptive Characteristics] Descriptive characteristics of the variables are shown below:

* Price: Mean: 263.370824, Mode: No mode, Spread: 175.846065, Tails: Right-skewed
* Room Type: Mode: Entire home/apt
* Host Superhost: Mode: False
* Number of Bedrooms: Mean: 1.150826, Mode: 1, Spread: 0.618111, Tails: Right-skewed
* City Distance: Mean: 3.194854, Mode: No mode, Spread: 2.396655, Tails: Right-skewed

[Slide 6: Probability Mass Function] A probability mass function (PMF) is used to compare two scenarios using the number of bedrooms:

* Scenario 1: Private rooms
* Scenario 2: Entire home/apt The resulting PMF shows that properties with private rooms have a higher probability of having 1 bedroom than properties that offer an entire home/apartment.

[Slide 7: Cumulative Distribution Function] A cumulative distribution function (CDF) is created for the 'price' variable. The resulting plot shows that around 50% of properties cost less than 200 euros per night.

[Slide 8: Analytical Distribution] An analytical distribution is created using a normal distribution with the mean and standard deviation of the 'price' variable. The resulting plot shows that the 'price' variable has a right-skewed distribution.

[Slide 9: Scatter Plots] Two scatter plots are created to compare the correlation and causation between two variables:

* Price vs. Number of Bedrooms: There is a positive correlation between the two variables, but causation cannot be determined.
* City Distance vs. Price: There is a