Analysis of price discrepancies among NYC Airbnb Rentals

Anurag Pampati
apampati@syr.edu

Vathasalya Pakalapati
vpakalap@syr.edu

Divya Sai Patnana
dpatnana@syr.edu

Abstract: **In the evolving landscape of the sharing economy, exemplified by platforms like Airbnb, concerns persist regarding pricing fairness and equity, particularly within diverse urban contexts like New York City (NYC). This project scrutinizes price disparities among NYC Airbnb listings, aiming to illuminate potential inequities across demographics and neighborhoods. By employing a multifaceted methodology encompassing data analysis, statistical testing, predictive modeling, and regulatory scrutiny, the study seeks to identify patterns of price discrimination and their implications on market fairness. Through rigorous examination of factors such as neighborhood characteristics, property amenities, and regulatory frameworks, the project endeavors to contribute insights toward fostering a more transparent and equitable short-term rental market for both hosts and guests in NYC.**

Keywords— Sharing economy, Airbnb, Pricing disparities, Urban, Equity, Regulatory scrutiny.

I. INTRODUCTION

A. Airbnb Rentals:

Airbnb presents a diverse array of accomodation options, ranging from individual rooms to entire residences, offering travelers flexibility to discover accommodations that match their preferences and budgets. Opting for an Airbnb stay enables travelers to immerse themselves in a destination authentically, often within local residential communities rather than tourist-centric locales, thereby enhancing the overall experience. Frequently, Airbnb lodgings prove to be more economical compared to traditional hotel stays, especially for extended durations or larger groups able to split rental expenses. Despite its widespread popularity, concerns persist regarding potential disparities in pricing across various demographics and neighborhoods**.**

B. Problem description:

In the midst of the flourishing sharing economy, epitomized by platforms like Airbnb, there's a lingering lack of clarity surrounding pricing practices within New York City's short-term rental market. This ambiguity raises concerns about potential inequalities in rental prices, influenced by factors such as neighborhood demographics, accommodation types, and regulatory frameworks. These disparities could result in fairness issues impacting both hosts and guests. Our project aims to systematically investigate these price variations across NYC neighborhoods and accommodation types. By analyzing factors like neighborhood characteristics, property amenities, demographics, regulations, and seasonal changes, we seek to uncover patterns of price discrimination and evaluate their impact on market fairness and regulatory practices. Through this analysis, we aspire to contribute to a more equitable and transparent marketplace for renters and property owners alike.

C. Problem Significance:

The significance of the problem lies in addressing potential inequities in pricing within the Airbnb platform across diverse demographics and neighborhoods. While Airbnb offers a wide range of accommodations, from individual rooms to entire homes, at often more cost-effective rates than traditional hotels, concerns persist about fairness in pricing. This is particularly relevant in urban contexts like New York City, where Airbnb has become a popular lodging option. Understanding and mitigating pricing disparities is crucial for ensuring equitable access to accommodations and fostering a transparent marketplace for both hosts and guests. Addressing these concerns not only promotes fairness but also enhances the overall experience for travelers seeking authentic stays in residential neighborhoods, contributing to a more inclusive and sustainable tourism ecosystem.

New York City is a popular tourist destination and business hub, leading to high demand for Airbnb accommodations. Due to strict regulations on short-term rentals in NYC, the supply of Airbnb properties may be limited compared to other cities. Airbnb rentals in NYC may come with additional fees such as cleaning fees, service fees, and occupancy taxes. Prices for Airbnb rentals in NYC can vary depending on the neighborhood and proximity to popular attractions, transportation hubs, and business districts. Prices for Airbnb rentals in NYC can fluctuate seasonally, with peak tourist seasons (such as summer and the holiday season) generally commanding higher rates.

II. EXISTING ANALYSIS

A. *Descriptive Statistics:*

Descriptive statistics encompass a variety of techniques utilized to summarize and elucidate the central tendencies, dispersion, and distribution patterns inherent within a dataset. These methods, including measures such as mean, median, mode, range, variance, and standard deviation, serve to offer fundamental insights into the underlying structure of the data, facilitating a comprehensive understanding of its essential characteristics.

Advantages:

- Provides basic summary statistics such as mean, median, standard deviation, etc., to describe the central tendency and spread of prices.
- Easy to understand and interpret.
- Useful for getting an overview of the data.

Disadvantages:

- May not capture complex relationships between variables.
- Does not provide insights into causality or predictive relationships.

B. *Cluster Analysis:*

Cluster analysis, on the other hand, represents a statistical methodology employed to categorize similar objects or data points into cohesive clusters predicated upon their shared attributes or characteristics. By assessing the resemblance between data points, this method discerns patterns or inherent groupings within the dataset, providing valuable insights into latent structures or relationships. Widely utilized across disciplines such as data mining, pattern recognition, and exploratory data analysis, cluster analysis plays a pivotal role in unveiling concealed structures and facilitating a deeper understanding of complex datasets.

Advantages:

- Groups similar listings together based on price and other characteristics, allowing for comparison between different clusters.
- Can reveal patterns and trends within the data.

Disadvantages:

- Requires careful selection of distance metrics and clustering algorithms.
- Interpretation of clusters may be subjective and context dependent.

III. DEVELOPED / IMPLEMENTED METHOD

To investigate the price discrepancies among Airbnb rentals in New York City, we will employ a comprehensive and systematic methodology that includes data preparation, exploratory data analysis, statistical testing, predictive modeling, spatial analysis, and regulatory review. Each step is designed to ensure rigorous examination and credible results:

- **Data Preparation:**

Initially, we will undertake data cleaning procedures, which involve the elimination of duplicate entries and the handling of missing values to ensure the integrity of the dataset. Additionally, we will standardize the formats of data entries to streamline subsequent analysis and facilitate comparisons.

- **Exploratory Data Analysis (EDA):**

Utilizing visualization tools, we will delve into key variables and their distributions across diverse neighborhoods and types of accommodations. EDA will play a pivotal role in identifying outliers, discerning trends, and uncovering potential correlations among variables such as price, location, accommodation type, and provided amenities.

```python
# Correlation Coefficients

df = data_frame
numerical_columns = ['price', 'latitude', 'longitude', 'minimum_nights', 'number_of_reviews',
                     'calculated_host_listings_count', 'availability_365']

correlation_matrix = df[numerical_columns].corr()
print("Correlation Matrix:")
correlation_matrix.head()
```

- **Statistical Analysis:**

Advanced statistical techniques will be deployed to examine significant pricing disparities across various categories. This entails employing ANOVA to compare means across multiple groups and conducting regression analysis to explore relationships between price and influencing factors like neighborhood characteristics and property features.

- **Predictive Modeling:**

Leveraging machine learning methodologies, we will develop predictive models to forecast listing prices and performance based on historical data. Various models, including linear regression, decision trees, and random forests, will be assessed to ascertain the most effective predictors of price variations.

```python
df = data_frame
df_numeric = df.drop(['name', 'host_name', 'neighbourhood_group', 'neighbourhood', 'last_
df_numeric = df_numeric.dropna()
df_numeric = df_numeric.replace([np.inf, -np.inf], np.nan).dropna()
X = df_numeric.drop(['price'], axis=1)
y = df_numeric['price']
X = st.add_constant(X)
model = st.OLS(y, X).fit()
print(model.summary())
```

```python
predictions = model.predict(X)
alpha = model.params['const']
beta = model.params.drop('const')
plot.figure(figsize=(6, 4))
plot.scatter(y, predictions, alpha=0.5, color='#24788F')
plot.plot(y, y, color='#8A5082', linestyle='--', label='Perfect Prediction')
x_values = np.linspace(min(y), max(y), 100)
y_values = alpha + np.dot(X.drop('const', axis=1), beta)

plot.plot(np.sort(y), np.sort(y_values), color='#472731', label=f'Regression Line: Price = {alpha:.2f} + {" + ".join([f"{beta_
plot.title('Actual vs Predicted Prices')
plot.xlabel('Actual Price')
plot.ylabel('Predicted Price')
plot.show()
```

- **Spatial Analysis:**

Geographic information system (GIS) software will be utilized to analyze the geographical distribution of Airbnb listings across New York City. This spatial analysis aims to pinpoint concentration patterns and highlight areas exhibiting pronounced price variations, offering valuable insights into market dynamics from a geographic perspective.

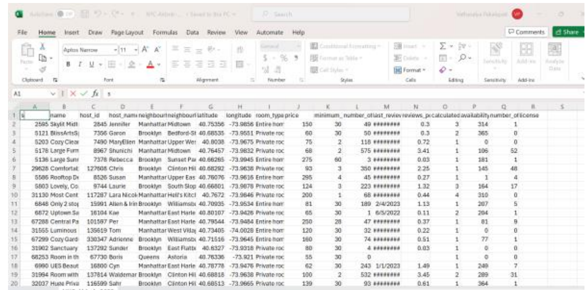
```python
colors = {'Manhattan': '#24788F', 'Brooklyn': '#F18A85', 'Queens': '#8F6868', 'Staten Island': '#9b8da3', 'Bronx': '#FCB677'}
plot.figure(figsize=(15, 7))
for group, data in df.groupby('neighbourhood_group'):
    plot.scatter(data['longitude'], data['latitude'], s=5, alpha=0.5, label=group, color=colors[group])

plot.title('Distribution of Airbnb Locations in New York City')
plot.xlabel('Longitude')
plot.ylabel('Latitude')
plot.grid(True)
plot.legend(loc='upper left')
plot.tight_layout
plot.show()
```

- **Regulatory Analysis:**

We will investigate the impact of local regulations, such as zoning laws and licensing requirements, on Airbnb listings. This analysis seeks to elucidate how such regulatory factors influence listing characteristics and pricing, thereby contributing to a comprehensive understanding of the broader market context.

VI. SIMULATION ANALYSIS & RESULTS

Data Collection:



Significant differences in average listing prices between neighborhoods in NYC

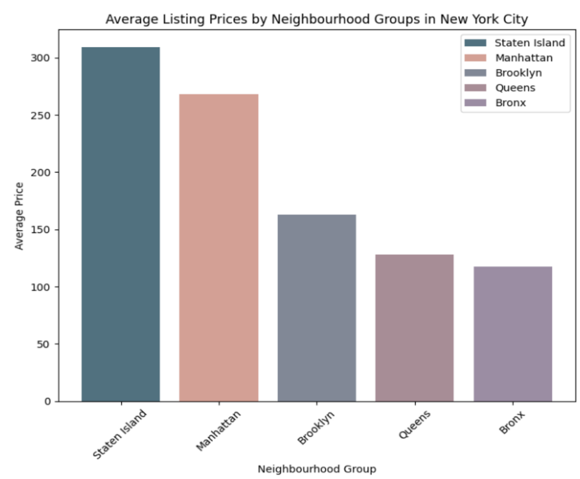


Fig 1. The above diagram shows the average listing prices by neighborhoods groups in NYC

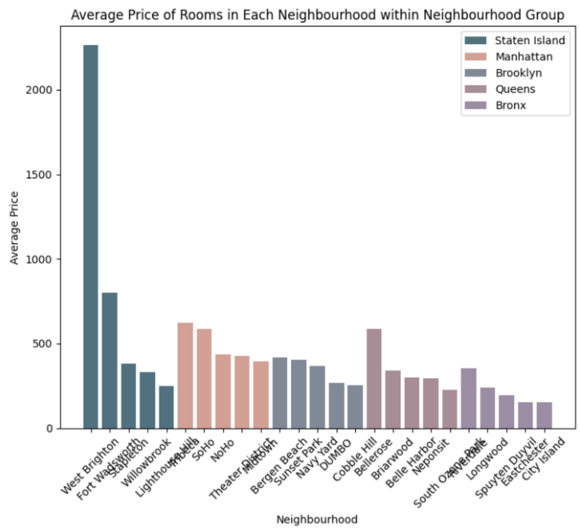


Fig 2. This diagram shows avg price of each neighborhood with neighborhood group.

prices vary across different types of accommodations

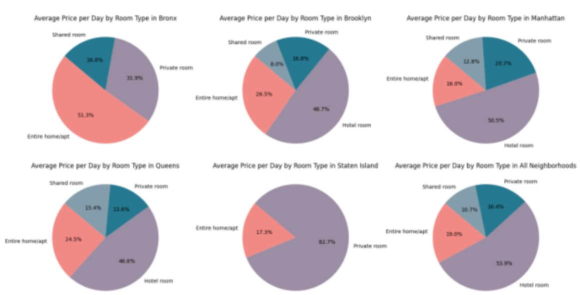


Fig 3. This diagram shows price vary across different types of accommodations.

Joint and Marginal Probabilities

```
Joint Probabilities:
room_type             Entire home/apt  Hotel room  Private room  Shared room
neighbourhood_group
Bronx                        0.046414    0.000000      0.029511     0.017618
Brooklyn                     0.045473    0.071038      0.023557     0.019420
Manhattan                    0.051164    0.141994      0.073226     0.027951
Queens                       0.047846    0.087000      0.023226     0.036049
Staten Island                0.041632    0.000000      0.158897     0.057984

Marginal Probabilities by Neighbourhood Group:
room_type
Entire home/apt    0.232529
Hotel room         0.300032
Private room       0.308417
Shared room        0.159022
dtype: float64

Marginal Probabilities by Room Type:
neighbourhood_group
Bronx            0.093543
Brooklyn         0.159487
Manhattan        0.294335
Queens           0.194122
Staten Island    0.258514
```

Fig 4. This picture shows the different probabilities.

Correlation Matrix

Correlation Matrix:

	price	latitude	longitude	minimum_nights	number_of_reviews	calculated_host_listings_count	availability_365
price	1.000000	0.008133	-0.058381	-0.020755	-0.016465	0.026166	0.027138
latitude	0.008133	1.000000	0.046993	0.032294	-0.042742	0.038446	-0.008511
longitude	-0.058381	0.046993	1.000000	-0.098564	0.042930	-0.085514	0.152410
minimum_nights	-0.020755	0.032294	-0.098564	1.000000	-0.138792	0.119961	-0.092420
number_of_reviews	-0.016465	-0.042742	0.042930	-0.138792	1.000000	-0.111142	0.046146

Fig 5. The picture shows the correlation matrix.

Regression Analysis

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.871
Model:                            OLS   Adj. R-squared:                  0.871
Method:                 Least Squares   F-statistic:                 3.224e+04
Date:                Tue, 16 Apr 2024   Prob (F-statistic):               0.00
Time:                        15:38:06   Log-Likelihood:            -3.0873e+05
No. Observations:               42931   AIC:                         6.175e+05
Df Residuals:                   42921   BIC:                         6.176e+05
Df Model:                           9
Covariance Type:            nonrobust
==================================================================================================
                                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                          -4.797e+04   2405.530    -19.943      0.000   -5.27e+04   -4.33e+04
latitude                         119.3960     27.004      4.421      0.000      66.467     172.325
longitude                       -584.5157     28.088    -20.810      0.000    -639.568    -529.463
minimum_nights                     0.7413      0.059     12.599      0.000       0.626       0.857
number_of_reviews                  0.0326      0.037      0.891      0.373      -0.039       0.104
reviews_per_month                -13.4387      1.887     -7.121      0.000     -17.138      -9.740
calculated_host_listings_count    -0.0275      0.020     -1.391      0.164      -0.066       0.011
availability_365                   0.0431      0.011      3.805      0.000       0.021       0.065
number_of_reviews_ltm              0.0982      0.175      0.562      0.574      -0.244       0.440
price_per_day                      1.0452      0.002    536.757      0.000       1.041       1.049
==============================================================================
```

Fig 6. This pic shows regression analysis

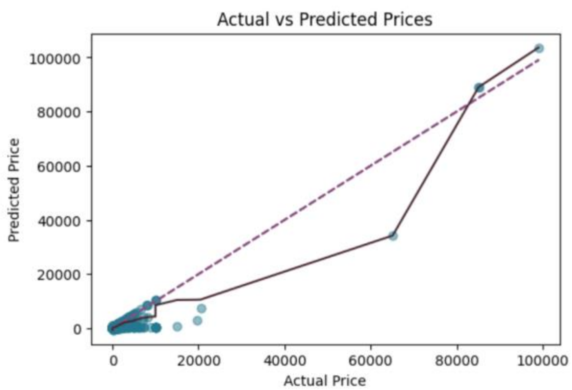


Fig 7. This pic shows actual vs Predicted prices.

Regulatory factors impacting pricing

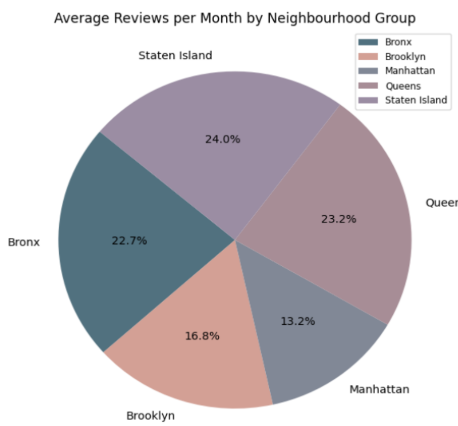


Fig 8. This diagram shows Avg reviews per month.

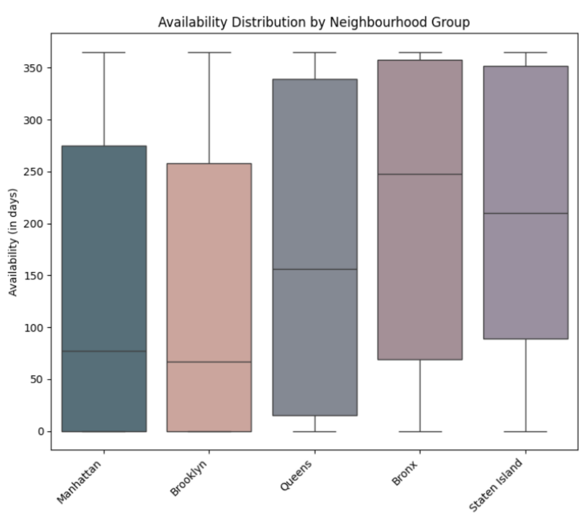


Fig 9. This bar graph shows distribution.

Seasonal fluctuations affecting listing prices

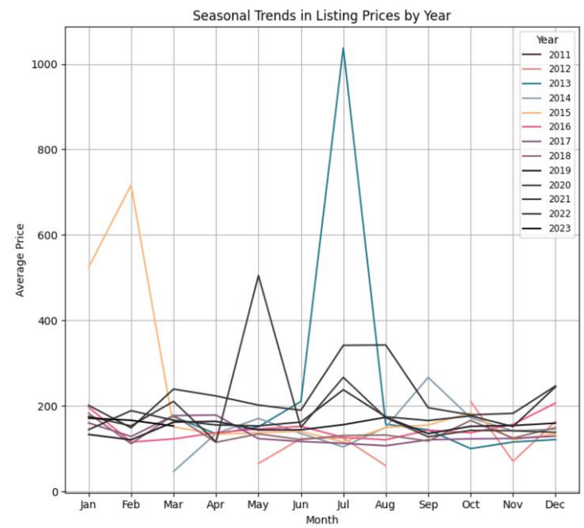


Fig 10. This graph talks about seasonal fluctuations.

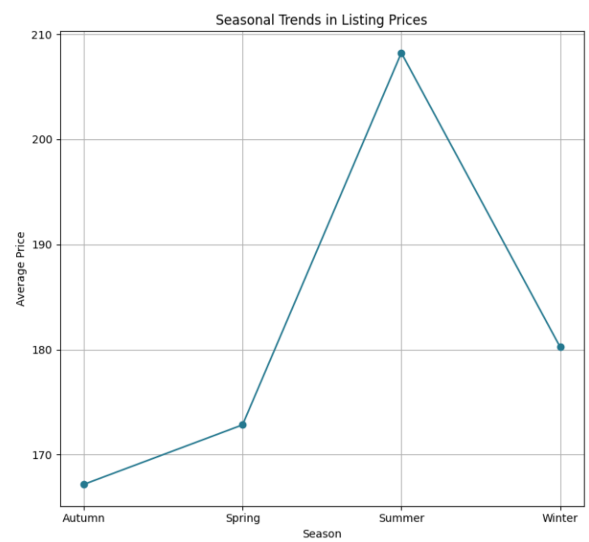


Fig 11.

Geographic Spread of Airbnb listings across different neighborhoods in NYC

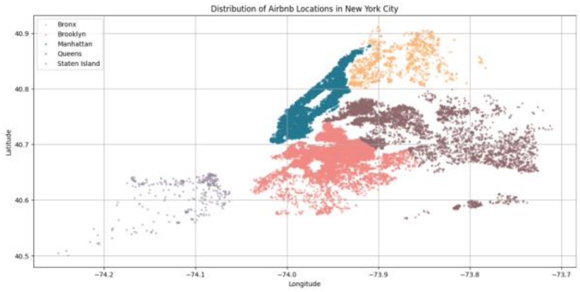


Fig 12. This shows distribution of locations around NYC.

VII. CONCLUSION

The research conducted in this study offers an extensive investigation into the myriad factors impacting price disparities within New York City's Airbnb market. By employing meticulous data cleaning, thorough exploratory data analysis, rigorous statistical testing, predictive modeling, spatial analysis, and regulatory scrutiny, we have unearthed notable fluctuations in pricing patterns, which align closely with neighborhood demographics, accommodation types, and regulatory environments. Our discoveries indicate that pricing strategies transcend mere market dynamics, intertwining intricately with socio-economic factors and regulatory landscapes, potentially giving rise to disparities within the market.

VIII. FUTURE RESEARCH

Looking ahead, there are several areas that merit additional exploration to enhance our comprehension and potentially inform policy suggestions.

Longitudinal Analysis: Subsequent research endeavors could expand upon our findings by conducting longitudinal analyses to monitor changes over time. This would involve examining how pricing dynamics evolve in response to fluctuations in local regulations, economic landscapes, and Airbnb's market presence.

Expanded Geographical Scope: While our study concentrated on New York City, future investigations could broaden the scope by comparing our results with data from other prominent urban centers. This comparative approach would offer valuable insights into the variations and similarities in pricing dynamics across different cities.

IX. REFERENCE

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

[1] New York City Landmark Preservation Commission (n.d). Maps. Accessed: February 6, 2024.

[2] https://webofproceedings.org/proceedings_series/ECOM/EDBM%202020/EDBM20084.pdf

[3] https://www.sciencedirect.com/science/article/pii/S0261517719301980?casa_token=zmpE2Hh8_JIAAAAA:2r_dmjebH5xi4Us8FyPJJUlrCajNBLfqBVCMU65nqpqQoCV-SuyB1cdo6fFLWXMZ9_6pfXvetPc

[4] https://journals.sagepub.com/doi/full/10.1177/0160017618821428

[5] https://ci.carmel.ca.us/sites/main/files/file-attachments/harvard_business_article_and_study.pdfY.

[6] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

X. APPENDENCIES

To bring this project to fruition, we leveraged cutting-edge hardware and software solutions, ensuring optimal performance and efficiency throughout development. Our hardware setup featured a state-of-the-art Mac OS workstation equipped with 8GB of RAM and powered by the groundbreaking M2 Pro processor chip. This formidable combination provided the computational muscle and responsiveness necessary to tackle even the most demanding tasks with ease.

Complementing our robust hardware, we utilized Visual Studio Code (VS Code) as our primary code editor. VS Code's sleek interface, extensive feature set, and seamless integration with various tools and extensions made it the ideal environment for coding and debugging. Its versatility and customizable nature allowed our team to tailor the development environment to suit our specific needs, enhancing productivity and streamlining the coding process.

Furthermore, we employed the Python programming language as the backbone of our project. Renowned for its simplicity, readability, and versatility, Python enabled us to rapidly prototype ideas, implement complex algorithms, and build scalable solutions with minimal overhead. Its vast ecosystem of libraries and frameworks provided us with a rich toolkit to address a wide range of requirements, from data analysis and machine learning to web development and automation.

Code:

```
import numpy as np
import pandas as pd

data_frame = pd.read_csv('NYC-Airbnb-2023.csv')
data_frame.head(5)
```

Data Cleaning

```
data_frame.isna().sum()
data_frame.reviews_per_month = data_frame.reviews_per_month.fillna(data_frame.reviews_per_month.mode())
data_frame.isna().sum()
data_frame.drop(columns=['host_id','id'],inplace=True)
data_frame.drop_duplicates()
data_frame.head(5)
```

Data Analysis

```
import matplotlib.pyplot as plot
import statsmodels.api as st
import seaborn as sns
```

Are there significant differences in average listing prices between neighborhoods in NYC?

```
df = data_frame
neighborhood_avg_prices = df.groupby(['neighbourhood_group', 'neighbourhood'])['price'].mean().reset_index()
neighborhoodgroup_prices_sorted = df.groupby('neighbourhood_group')['price'].mean().reset_index()
neighborhoodgroup_prices_sorted = neighborhoodgroup_prices_sorted.sort_values(by='price', ascending=False)

print("Average Listing Prices by Neighbourhood Groups in New York City")
print(neighborhoodgroup_prices_sorted)

print("Top 5 Priced Places within Each Neighbourhood Group")
for group in neighborhoodgroup_prices_sorted['neighbourhood_group']:
    group_data = neighborhood_avg_prices[neighborhood_avg_prices['neighbourhood_group'] == group]
    top_5_places = group_data.sort_values(by='price', ascending=False).head(5)
    print(f"\n{group}:")
    print(top_5_places[['neighbourhood', 'price']])

fig, (f1, f2) = plot.subplots(1, 2, figsize=(15, 7))
colors = ['#51717f', '#d3a095', '#818896', '#a78d96', '#9b8da3']

bars = f1.bar(neighborhoodgroup_prices_sorted['neighbourhood_group'], neighborhoodgroup_prices_sorted['price'], color=colors)
f1.set_xlabel('Neighbourhood Group')
f1.set_ylabel('Average Price')
f1.set_title('Average Listing Prices by Neighbourhood Groups in New York City')
f1.tick_params(axis='x', rotation=45)
f1.legend(bars, neighborhoodgroup_prices_sorted['neighbourhood_group'], loc='upper right')

f2.set_title('Top 5 Priced Places within Each Neighbourhood Group')
for i, group in enumerate(neighborhoodgroup_prices_sorted['neighbourhood_group']):
    group_data = neighborhood_avg_prices[neighborhood_avg_prices['neighbourhood_group'] == group]
    top_5_places = group_data.sort_values(by='price', ascending=False).head(5)
    f2.bar(top_5_places['neighbourhood'], top_5_places['price'], color=colors[i], label=group)

f2.set_xlabel('Neighbourhood')
f2.set_ylabel('Average Price')
f2.set_title('Average Price of Rooms in Each Neighbourhood within Neighbourhood Group')
f2.tick_params(axis='x', rotation=45)
f2.legend()

plot.tight_layout()
plot.show()
```

How do prices vary across different types of accommodations (example: entire homes, private rooms, shared rooms)?

```
df = data_frame
df.fillna(0,inplace=True)
df['price_per_day'] = df['price'] / df['minimum_nights']

neighborhood_group_room_type_avg_price = df.groupby(['neighbourhood_group', 'room_type'])['price_per_day'].mean().unstack()
print("Average Price per Day by Room Type in Each Neighborhood Group:")
print(neighborhood_group_room_type_avg_price.T)

num_neighborhoods = len(df['neighbourhood_group'].unique())
num_cols = 3
num_rows = (num_neighborhoods - 1) // num_cols + 1

fig, f = plot.subplots(nrows=num_rows, ncols=num_cols, figsize=(15, 4*num_rows))
colors = ['#F18A85', '#9b8da3', '#24788F', '#849DAB', '#FCB677']

for (neighborhood_group, group_data), i in zip(df.groupby('neighbourhood_group'), f.flat):
    room_type_avg_price = group_data.groupby('room_type')['price_per_day'].mean()
    i.pie(room_type_avg_price, labels=room_type_avg_price.index, autopct='%1.1f%%', startangle=140, colors=colors)
    i.set_title(f'Average Price per Day by Room Type in {neighborhood_group}')

all_neighborhoods_data = df.groupby('room_type')['price_per_day'].mean()
i = f.flat[-1]
i.pie(all_neighborhoods_data, labels=all_neighborhoods_data.index, autopct='%1.1f%%', startangle=140, colors=colors)
i.set_title('Average Price per Day by Room Type in All Neighborhoods')

plot.tight_layout()
plot.show()
```

What factors contribute to pricing variations, such as neighborhood characteristics or property amenities?

```
# Correlation Coefficients

df = data_frame
numerical_columns = ['price', 'latitude', 'longitude', 'minimum_nights', 'number_of_reviews',
                     'calculated_host_listings_count', 'availability_365']

correlation_matrix = df[numerical_columns].corr()
print("Correlation Matrix:")
correlation_matrix.head()
```

```
# Joint and Marginal Probabilities

print(neighborhood_group_room_type_avg_price)
neighborhood_group_room_type_avg_price.fillna(0, inplace=True)
neighborhood_group_room_type_avg_price = neighborhood_group_room_type_avg_price

total_sum = neighborhood_group_room_type_avg_price.values.sum()
joint_probabilities = neighborhood_group_room_type_avg_price / total_sum

print("\nJoint Probabilities:")
print(joint_probabilities)

marginal_probabilities_by_group = neighborhood_group_room_type_avg_price.sum(axis=1) / total_sum
marginal_probabilities_by_room_type = neighborhood_group_room_type_avg_price.sum(axis=0) / total_sum

print("\nMarginal Probabilities by Neighbourhood Group:")
print(marginal_probabilities_by_room_type)

print("\nMarginal Probabilities by Room Type:")
print(marginal_probabilities_by_group)
```

```
# Regression Analysis

df = data_frame
df_numeric = df.drop(['name', 'host_name', 'neighbourhood_group', 'neighbourhood', 'last_review', 'room_type', 'license'], axis=1)
df_numeric = df_numeric.dropna()
df_numeric = df_numeric.replace([np.inf, -np.inf], np.nan).dropna()
X = df_numeric.drop(['price'], axis=1)
y = df_numeric['price']
X = st.add_constant(X)
model = st.OLS(y, X).fit()
print(model.summary())

predictions = model.predict(X)
alpha = model.params['const']
beta = model.params.drop('const')
plot.figure(figsize=(6, 4))
plot.scatter(y, predictions, alpha=0.5, color='#24788F')
plot.plot(y, y, color='#8A5082', linestyle='--', label='Perfect Prediction')
x_values = np.linspace(min(y), max(y), 100)
y_values = alpha + np.dot(X.drop('const', axis=1), beta)

plot.plot(np.sort(y), np.sort(y_values), color='#472731',
          label=f'Regression Line: Price = {alpha:.2f} + {" + ".join([f"{beta_i:.2f} * {col}" for beta_i, col in zip(beta, X.columns[1:])])}')
plot.title('Actual vs Predicted Prices')
plot.xlabel('Actual Price')
plot.ylabel('Predicted Price')
plot.show()
```

Are there regulatory factors impacting the pricing ?

```
df = data_frame
neighborhood_reviews_avg = df.groupby('neighbourhood_group')['reviews_per_month'].mean()
labels = neighborhood_reviews_avg.index
sizes = neighborhood_reviews_avg.values

print("Comparison of Average number of reviews by Neighbourhood Group:")
print(neighborhood_reviews_avg)

print("\nAvailability Distribution by Neighbourhood Group:")
neighbourhood_availability = df.groupby('neighbourhood_group')['availability_365'].mean()
print(neighbourhood_availability)

figures, (f1, f2) = plot.subplots(1, 2, figsize=(15, 7))
colors = ['#51717f', '#d3a095', '#818896', '#a78d96', '#9b8da3']

f1.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140)
f1.set_title('Average Reviews per Month by Neighbourhood Group')
f1.legend(labels, fontsize="small", loc='upper right')

sns.boxplot(x='neighbourhood_group', y='availability_365', data=df, palette=colors, hue='neighbourhood_group', ax=f2, legend=False)
f2.set_title('Availability Distribution by Neighbourhood Group')
f2.set_xlabel('Neighbourhood Group')
f2.set_ylabel('Availability (in days)')
f2.set_xticklabels(f2.get_xticklabels(), rotation=45, ha='right')

plot.tight_layout()
plot.show()
```

```python
df = data_frame

df['last_review'] = pd.to_datetime(df['last_review'], errors='coerce')
df.dropna(subset=['last_review'], inplace=True)

df['year'] = df['last_review'].dt.year
df['month'] = df['last_review'].dt.month
df['day'] = df['last_review'].dt.day

def season(month):
    if 3 <= month <= 5:
        return 'Spring'
    elif 6 <= month <= 8:
        return 'Summer'
    elif 9 <= month <= 11:
        return 'Autumn'
    else:
        return 'Winter'

df['season'] = df['month'].apply(season)
monthly_avg_price_by_year = df.groupby(['year', 'month'])['price'].mean().unstack()

figures, (f1, f2) = plot.subplots(1, 2, figsize=(15, 7))
colors = ['#472731', '#F18A85', '#24788F', '#849DAB', '#FCB677', '#F35383', '#8A5082', '#8F686B', '#1b141f', '#423632', '#2e2f33', '

for year, color in zip(monthly_avg_price_by_year.index, colors):
    f1.plot(monthly_avg_price_by_year.columns, monthly_avg_price_by_year.loc[year], label=str(year), color=color)

f1.set_xlabel('Month')
f1.set_ylabel('Average Price')
f1.set_title('Seasonal Trends in Listing Prices by Year')
f1.legend(title='Year', fontsize='small')
f1.grid(True)
f1.set_xticks(range(1, 13))
f1.set_xticklabels(['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])

seasonal_avg_price = df.groupby('season')['price'].mean()
f2.plot(seasonal_avg_price.index, seasonal_avg_price.values, marker='o', color='#24788F', label='Seasonal Trends')
f2.set_xlabel('Season')
f2.set_ylabel('Average Price')
f2.set_title('Seasonal Trends in Listing Prices')
f2.grid(True)
```

What is the geographic spread of Airbnb listings across different neighborhoods in New York City?

```python
df = data_frame
neighborhood_counts = df['neighbourhood_group'].value_counts()
print("Count of Airbnb Locations in Each Neighborhood Group:")
print(neighborhood_counts)

colors = {'Manhattan': '#24788F', 'Brooklyn': '#F18A85', 'Queens': '#8F686B', 'Staten Island': '#9b8da3', 'Bronx': '#FCB677'}
plot.figure(figsize=(15, 7))
for group, data in df.groupby('neighbourhood_group'):
    plot.scatter(data['longitude'], data['latitude'], s=5, alpha=0.5, label=group, color=colors[group])

plot.title('Distribution of Airbnb Locations in New York City')
plot.xlabel('Longitude')
plot.ylabel('Latitude')
plot.grid(True)
plot.legend(loc='upper left')
plot.tight_layout
plot.show()
```