CS-746
FINAL PROJECT REPORT
By
KABALI

Team information

CS746F23Project_KABALI

S.No	Full Name	WSU ID	WSU Email
1	Lakshmi Kiranmai Guduru	B848P394	lxguduru@shockers.wichita.edu
2	Yaswanth Panguluri	T434K326	yxpanguluri@shockers.wichita.edu
3	Raghu Vamsi Anem	U627S967	rxanem@shockers.wichita.edu
4	Rahul Thirukovela	P993K463	rxthirukovela@shockers.wichita.edu
5	Justin Martin Zephaniah Srikotla	B592C567	jxsrikotla@shockers.wichita.edu

Today's Agenda

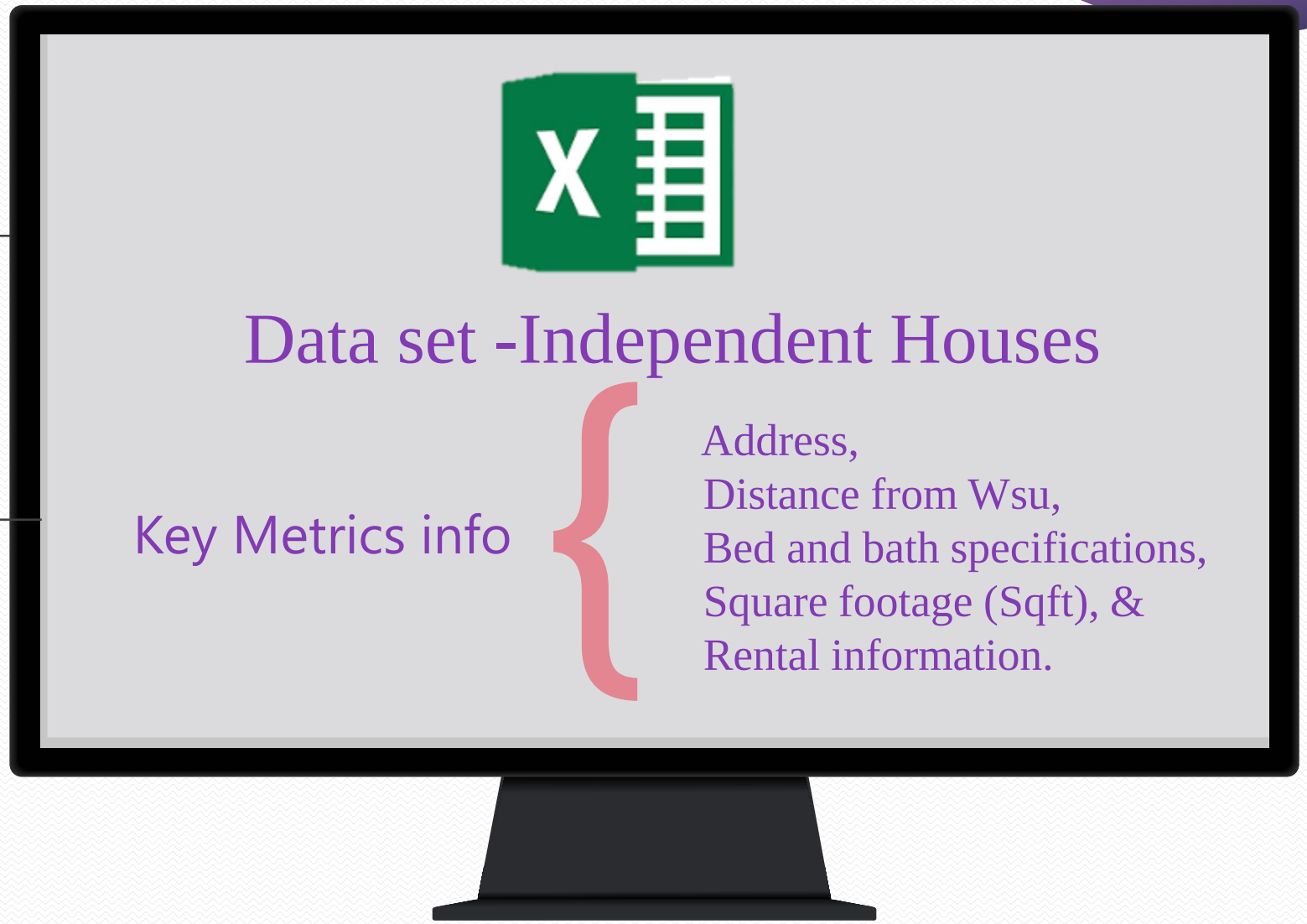
- What's the Chosen Data set.?
- EDA
- Data Cleaning
- Data Visualization
- Linear Regression
    - Performing Simple linear regression
    - Performing Multiple Linear Regression

DATA SET CHOOSEN

Describing the Data

37 Columns

5 Rows



EDA

Exploratory Data Analysis

It is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

Steps in EDA:

- Describing data set.
- Handling Missing values.
- Data Visualization.
- Correlation Matrix.
- Outilers

DATA CLEANING

Here, We dropped 2 Key features i.e., “Address’, 'Bed And Bath and checking the missing values.

Data Cleaning

```
df = df.drop(columns=['Address','Bed And Bath'])
# Display the modified DataFrame
print(df.head())
```

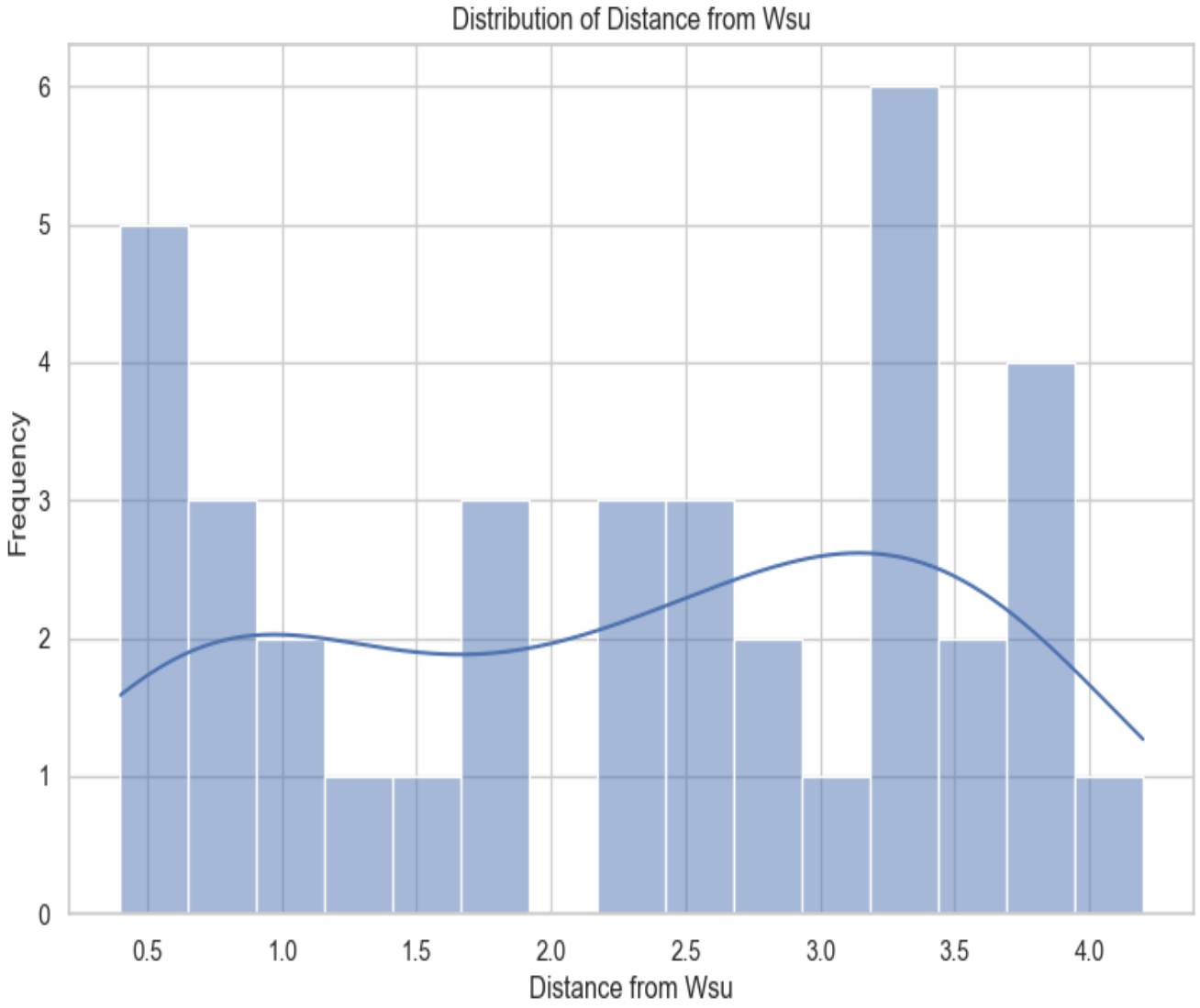
```
   Distance from Wsu    Sqft  Rent  Bedrooms  Bathrooms
0                0.7  2200.0  1475         4        2.0
1                2.7  1617.0  1025         4        2.0
2                0.6  1200.0  1350         4        2.0
3                3.2  1650.0  1650         4        2.5
4                3.8  1748.0  1595         4        1.5
```

```
# Check for missing values
missing_values = df.isnull().sum()
print("Missing Values:\n", missing_values)

# Handle missing values if needed (replace with mean, median, etc.)
df['Sqft'].fillna(df['Sqft'].median(), inplace=True)
```

```
Missing Values:
 Distance from Wsu    0
Sqft                  2
Rent                  0
Bedrooms              0
Bathrooms             0
dtype: int64
```

DATA VISUALIZATION FOR DISTANCE FROM WSU

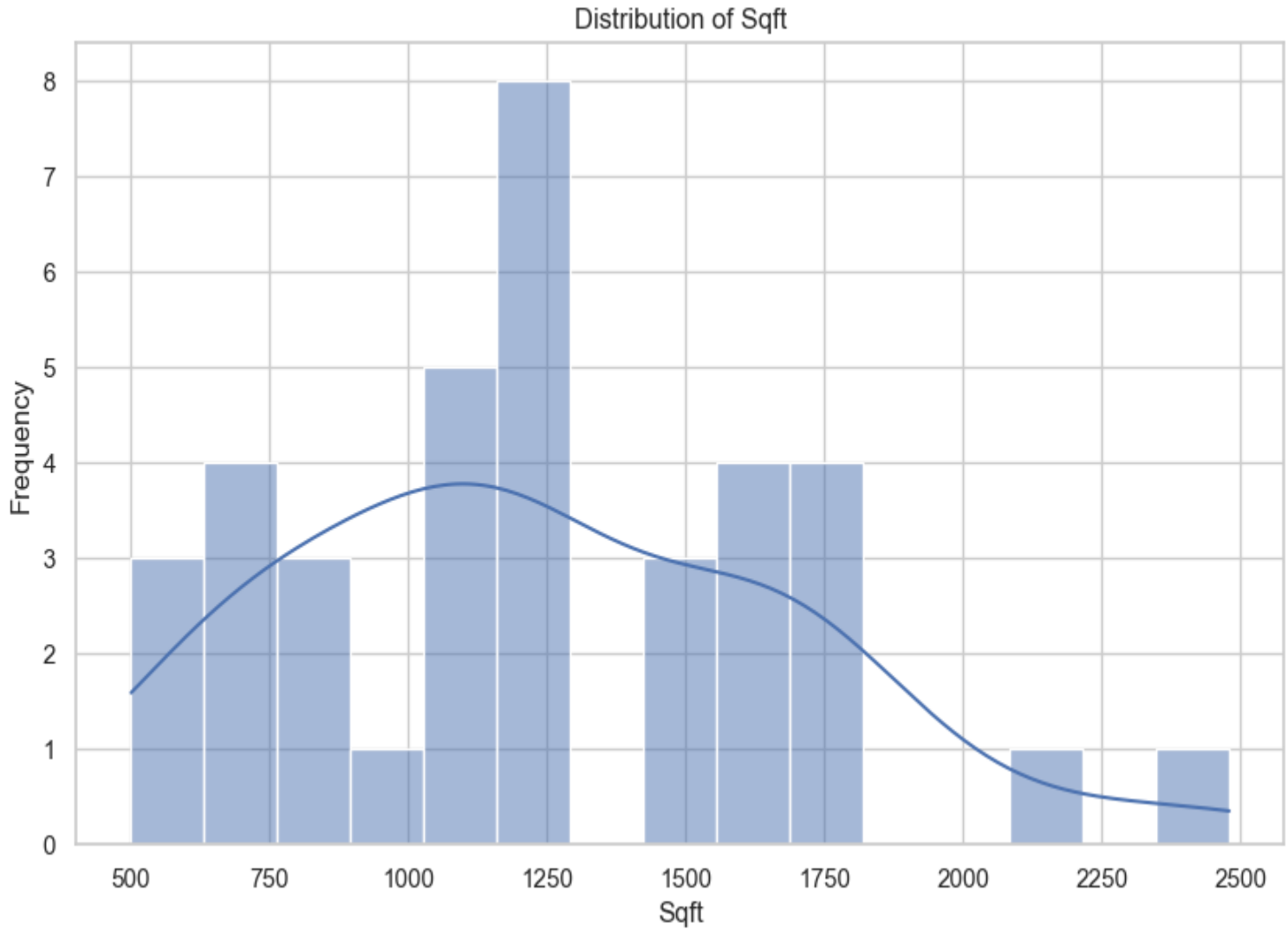
Data visualization is the graphical representation of information and data.



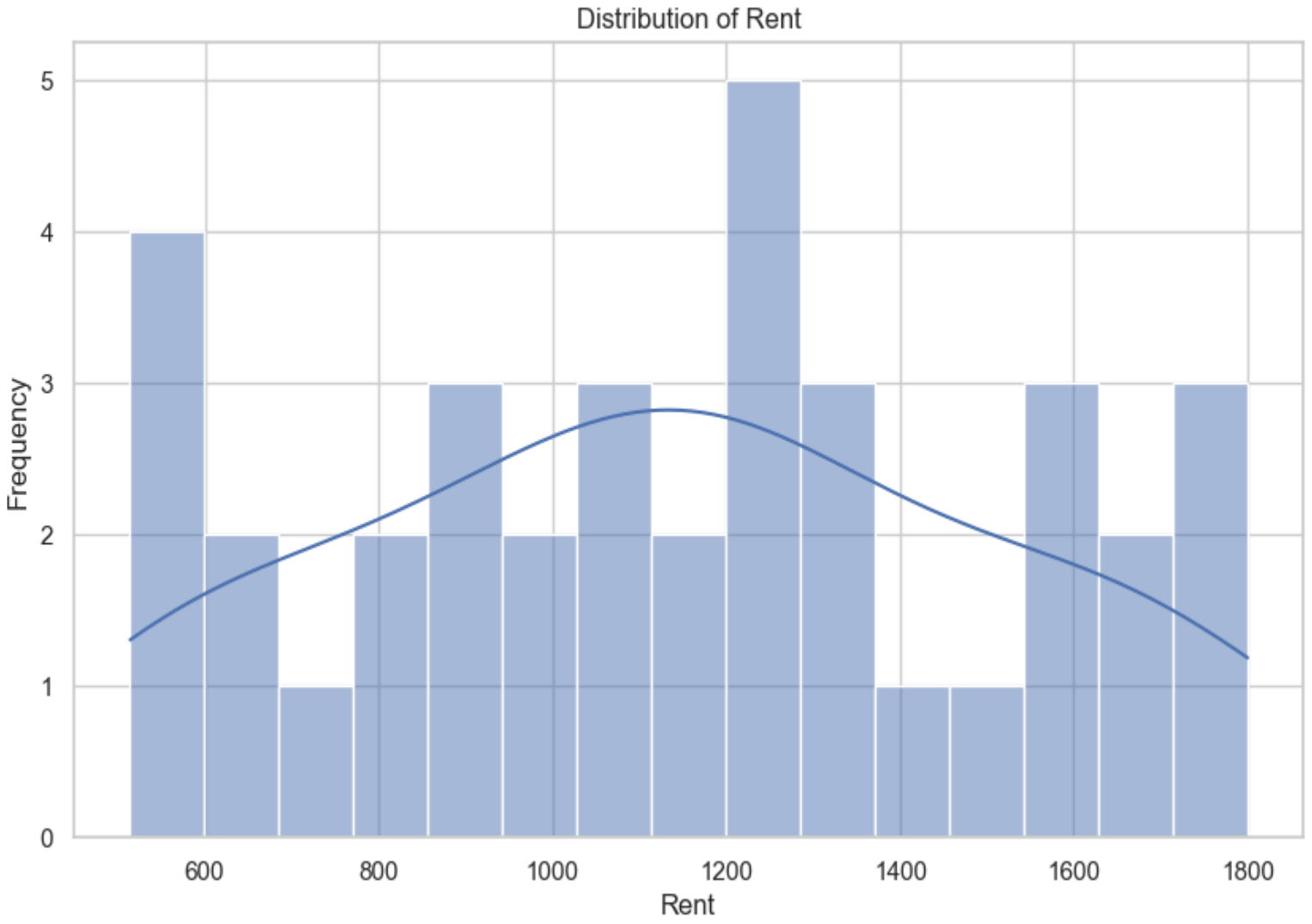

```python
import matplotlib.pyplot as plt
import seaborn as sns

# Set the style for seaborn
sns.set(style="whitegrid")

# Visualization for 'Distance from Wsu'
plt.figure(figsize=(10, 6))
sns.histplot(df['Distance from Wsu'], bins=15, kde=True)
plt.title('Distribution of Distance from Wsu')
plt.xlabel('Distance from Wsu')
plt.ylabel('Frequency')
plt.show()
```

VISUALIZATION FOR SQFT

```
# Visualization for 'Sqft'
plt.figure(figsize=(10, 6))
sns.histplot(df['Sqft'], bins=15, kde=True)
plt.title('Distribution of Sqft')
plt.xlabel('Sqft')
plt.ylabel('Frequency')
plt.show()
```



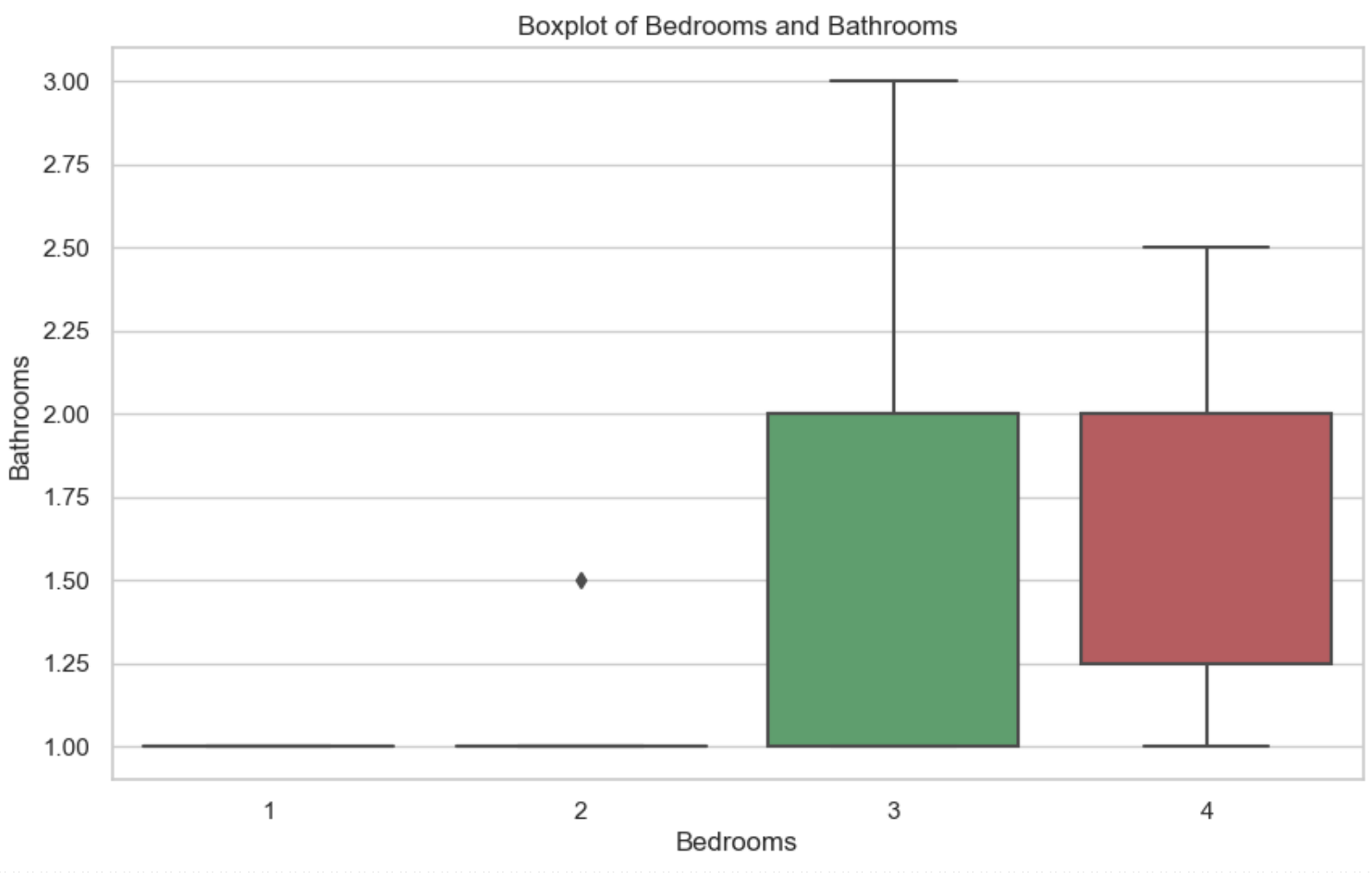
VISUALIZATION FOR RENT



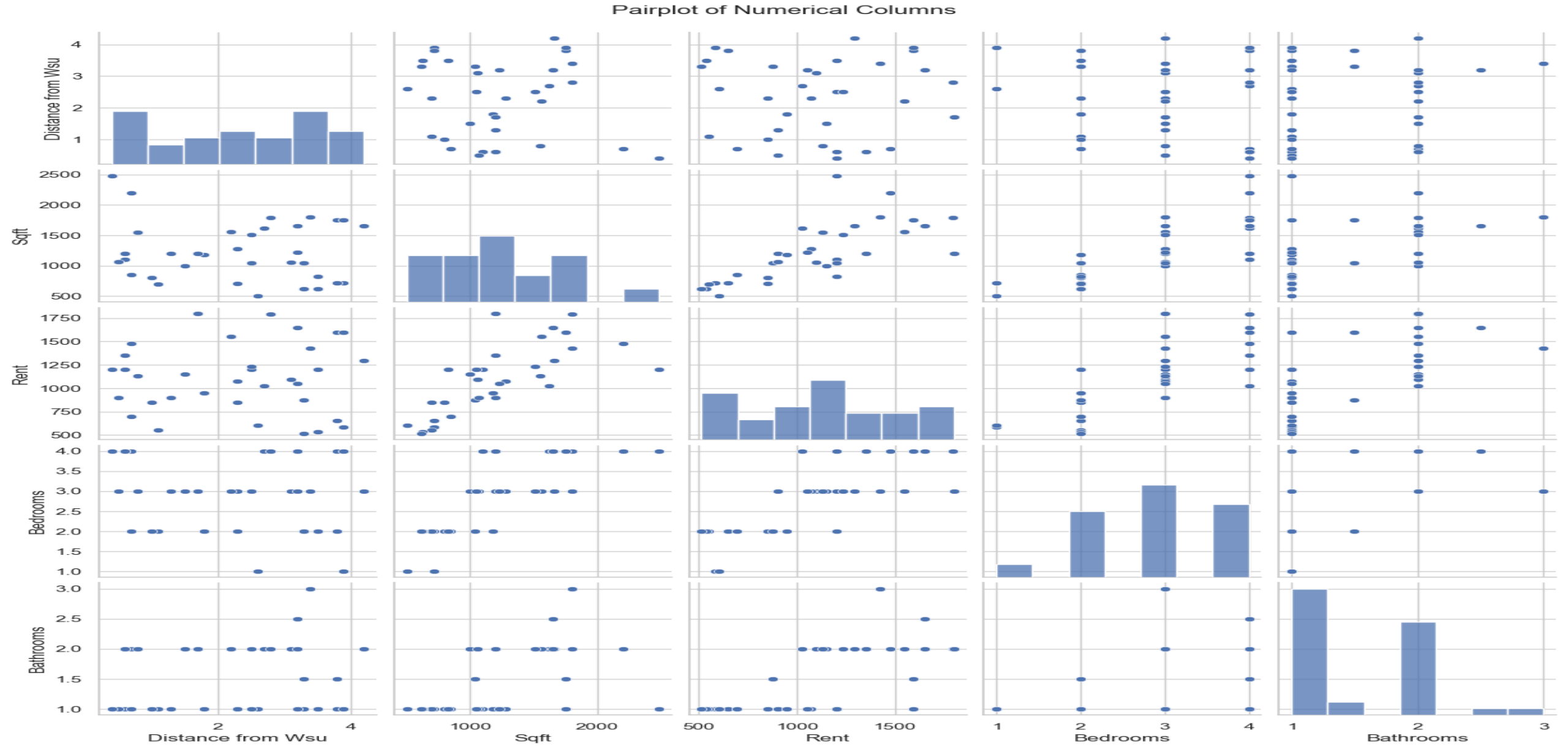

```python
# Visualization for 'Rent'
plt.figure(figsize=(10, 6))
sns.histplot(df['Rent'], bins=15, kde=True)
plt.title('Distribution of Rent')
plt.xlabel('Rent')
plt.ylabel('Frequency')
plt.show()
```

BOX PLOT FOR BEDROOMS & BATHROOMS

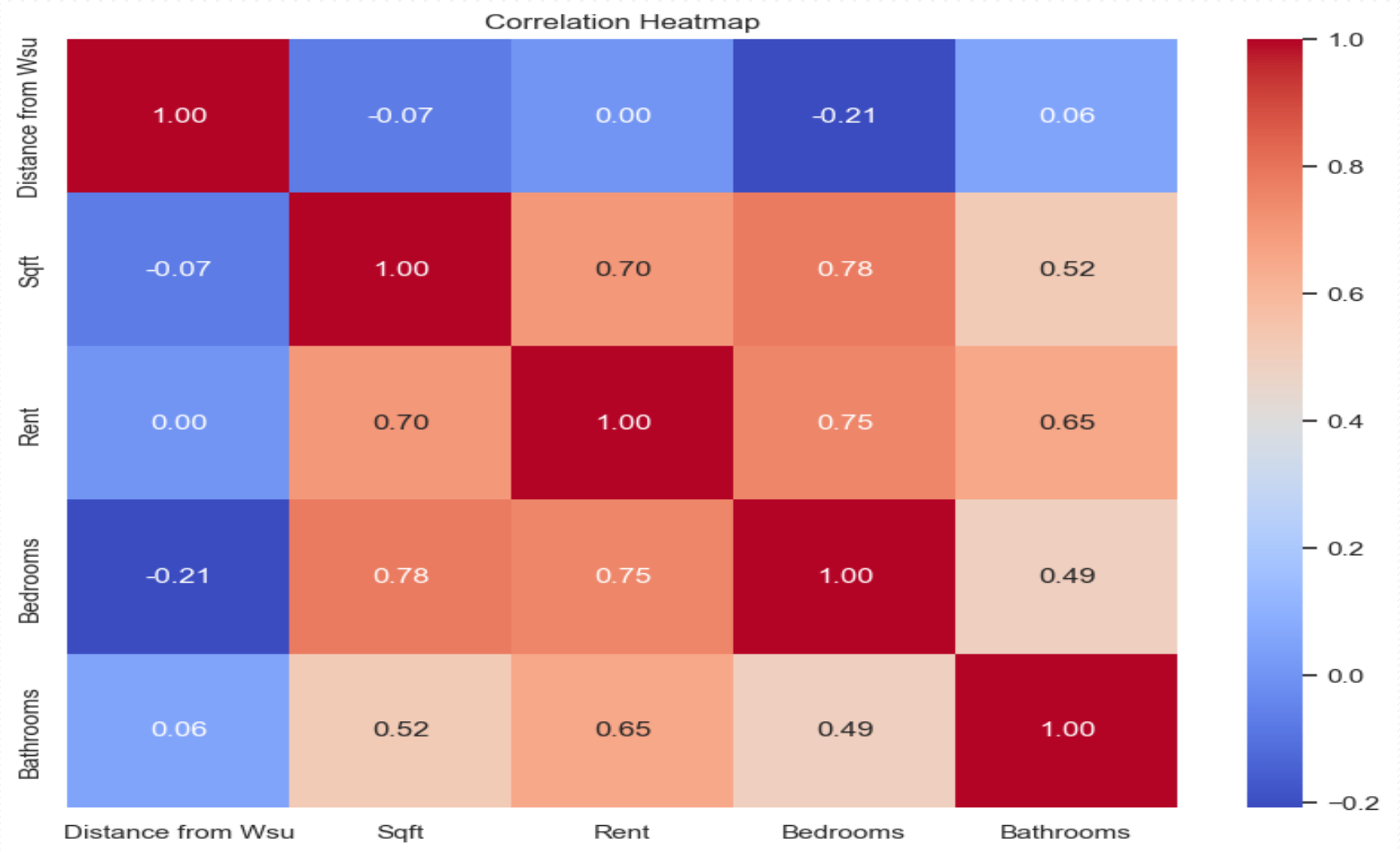
```
sns.boxplot(x='Bedrooms,
y='Bathrooms', data=df)
```



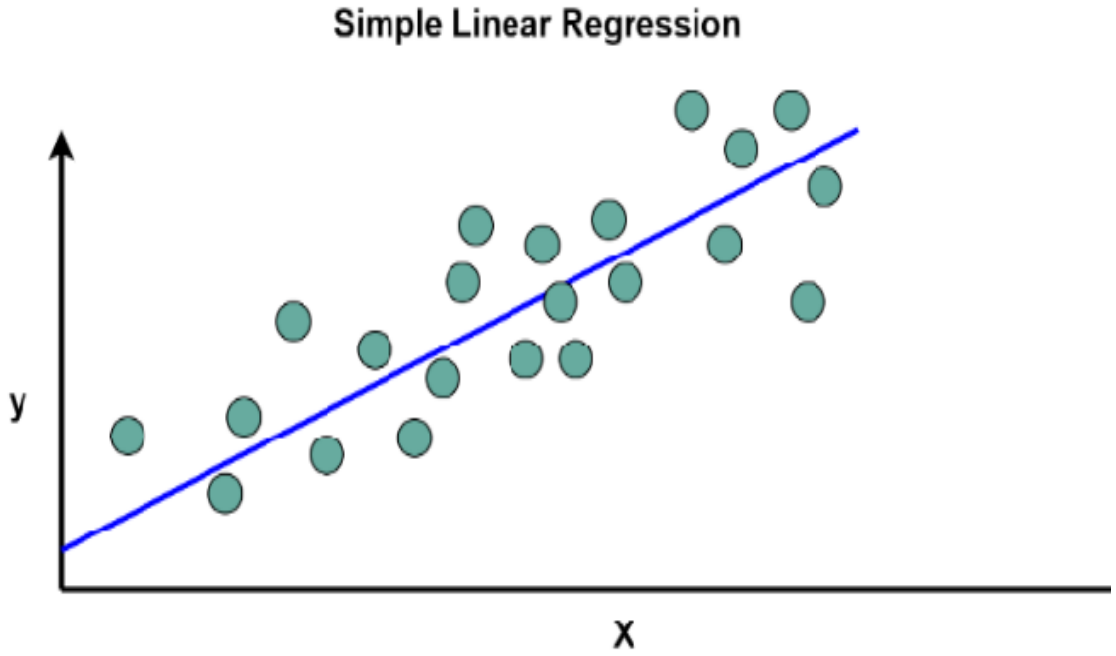
PAIRPLOTS (Distance from Wsu', 'Sqft', 'Rent', 'Bedrooms', 'Bathrooms)

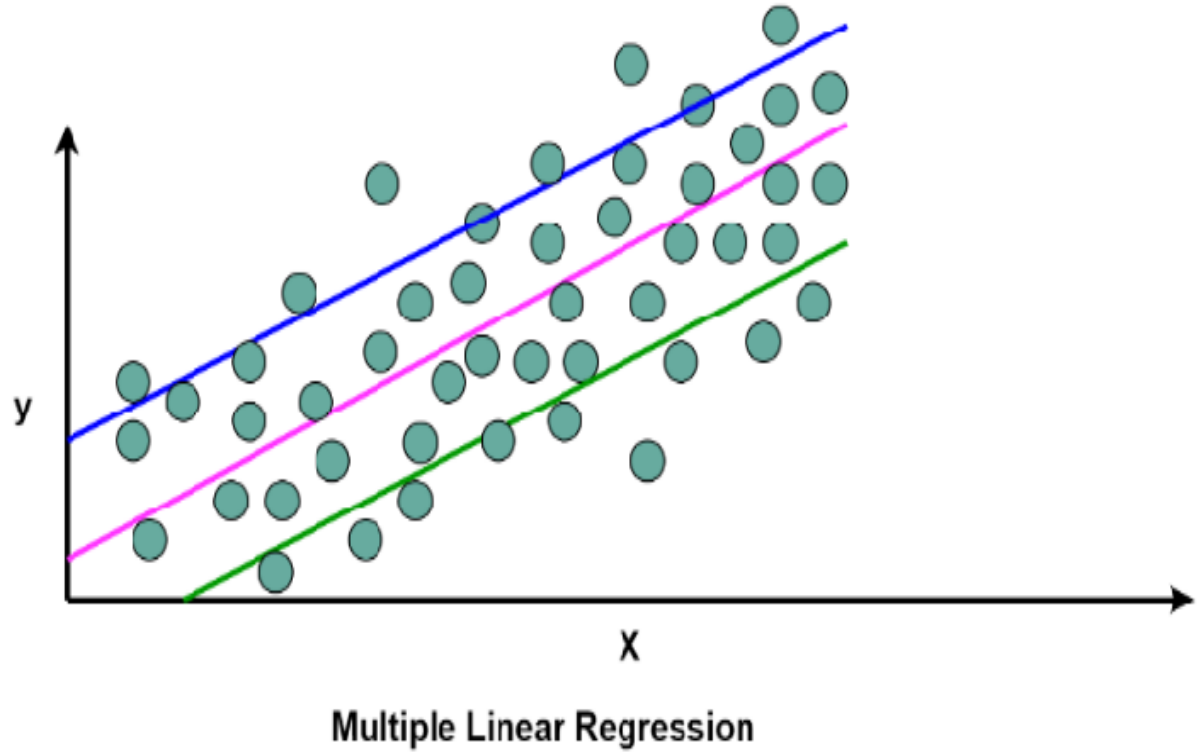


CORRELATION HEAT MAP



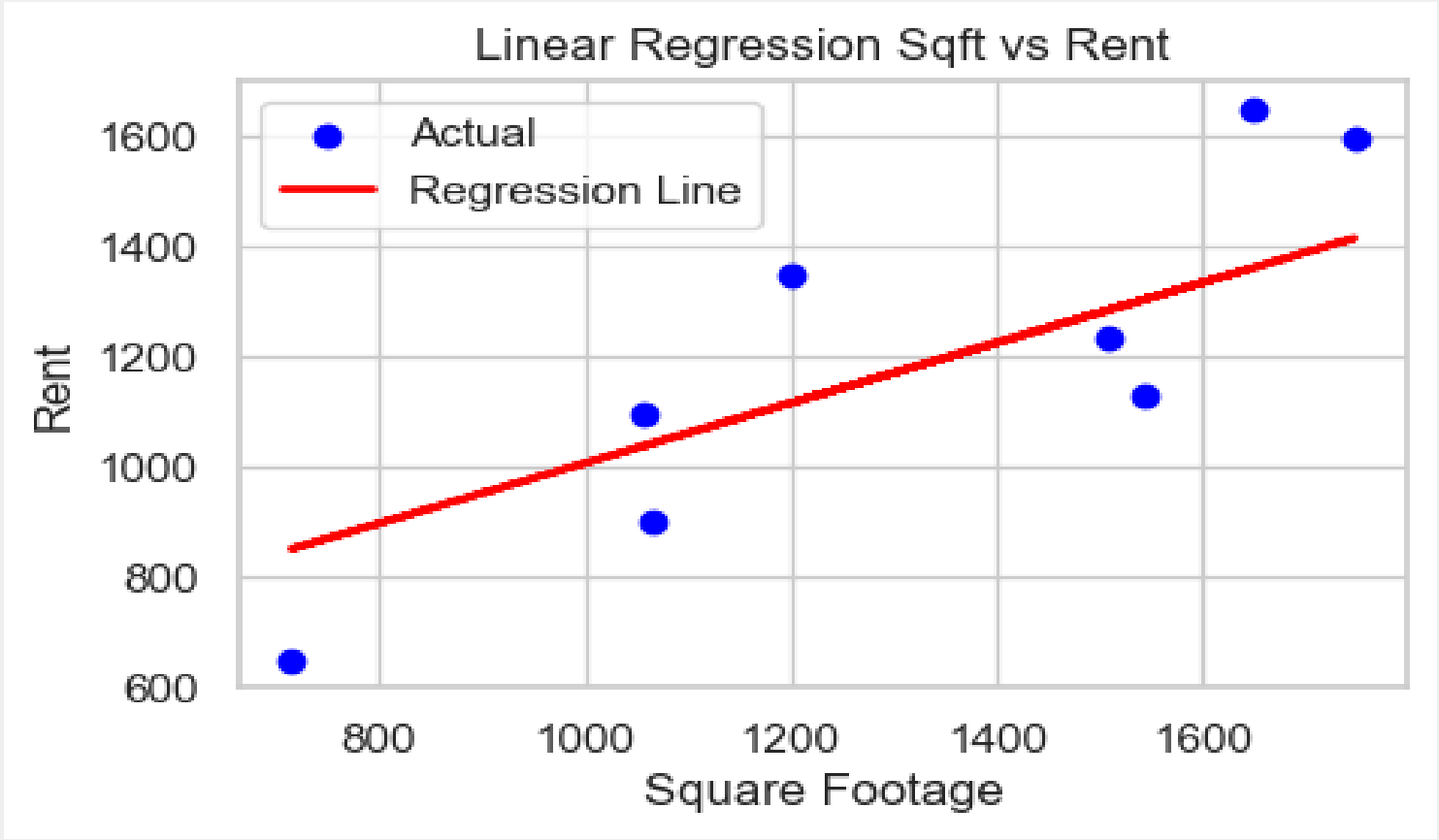
LINEAR REGRESSION & MULTIPLE LINEAR REGRESSION





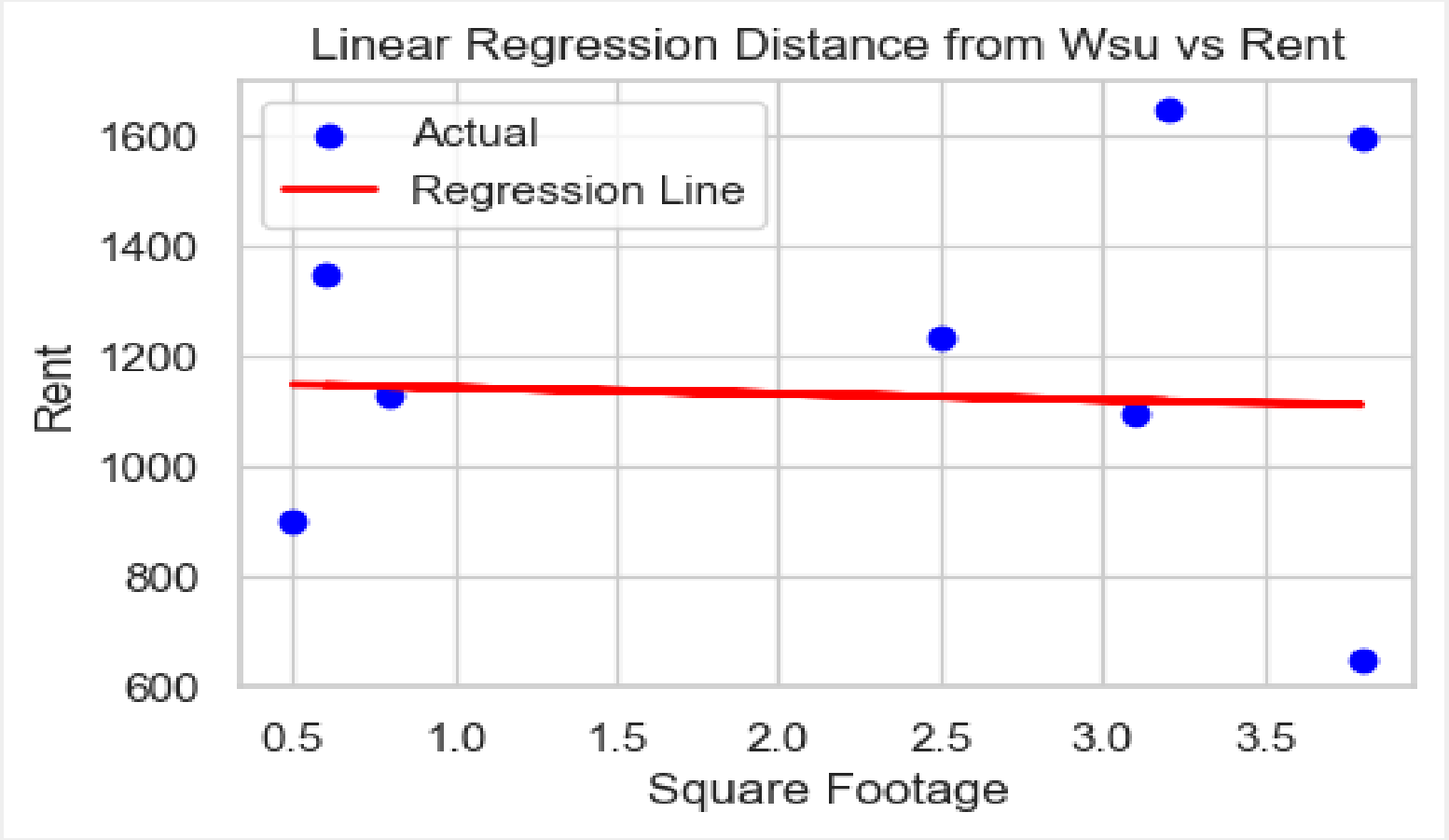
DATA DRIVEN COLLECTION

LINEAR REGRESSION SQFT VS RENT



DATA DRIVEN COLLECTION

LINEAR REGRESSION DISTANCE FROM WSU VS RENT



THANK YOU