**DATA 607 : FINAL PROJECT BY TEAM CUNY VERSE**

**UNDERSTANDING THE EFFECT OF AIRQUALITY ON REAL ESTTE PRICES IN NEW YORK CITY .**

**May 7, 2019**

**TEAM MEMBERS**

**1/ MIA SIRACUSA**

**2. JAVERN WILSON**

**3. KLEBER PEREZ**

**4. YOHANNES DEBOCH**

**Motivation**

It is believed that air pollution is responsible for one third of deaths from stroke, lung cancer and heart disease which is the motivation behind this project, we want to determine the impact of air quality on the price of houses in New York City.

**Data Description**

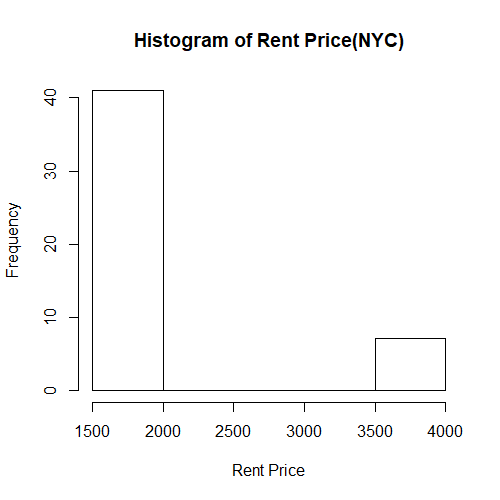
The data for this project were obtained from two sources, the first source is the NYC OpenData where we obtained the data on average benzene concentration which is a measure of air quality while the second data on average rent price was scrapped from Zillow, an online real estate database company that was founded in 2006. Our variables of interest in this project are average rent price and average benzene concentrations in the New York City neighborhoods.

**Data Cleaning and Preparation**

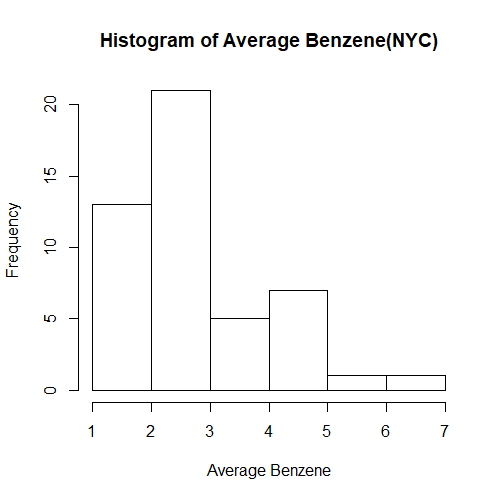
Here we loaded the data into R and Perform data cleaning and transformation. We extracted the variable needed in the air quality data (average benzene concentration) which consist of 48 observations and combined it with average rent dataset. We did random selection of the observations in both datasets to avoid bias in our analysis.

**Data Visualization**

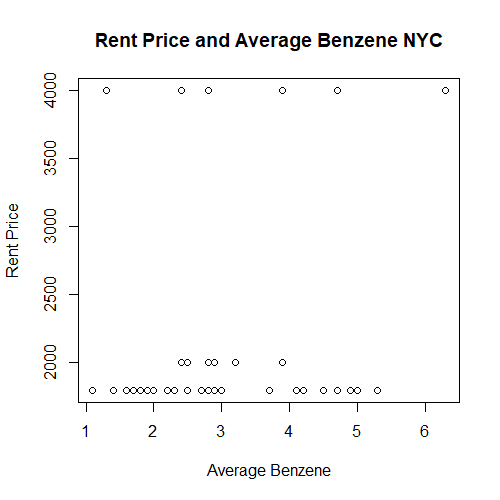
We explore the data to determine the distribution of the variables.



The histogram plot of rent price above shows that most houses in NYC have rent prices between $1500 and $1800 per month. The distribution of the rent price in NYC shows that the rent prices in our data falls between the range of $1500 and $4000 per month.



Looking at the Histogram plot of Average Benzene concentration in NYC, the data points fall between the range of 1 and 7. The distribution is skewed to the right and not normally distributed.



The scatterplot of Rent Price and Average Benzene concentration shows that there is no significant relationship between these two variables.

**Data Analysis**

To determine the effect of benzene concentration on the rent price in NYC, we fitted a linear regression model. The result of the regression model is prevented below.

|  |  |
| --- | --- |
|  | |
|  | *Dependent variable:* |
|  |  |
|  | price |
|  | |
| Benzene | 132.026 |
|  | (96.085) |
|  |  |
| Constant | 1,765.750\*\*\* |
|  | (300.842) |
|  |  |
|  | |
| Observations | 48 |
| R2 | 0.039 |
| Adjusted R2 | 0.019 |
| Residual Std. Error | 768.469 (df = 46) |
| F Statistic | 1.888 (df = 1; 46) |
|  | |
| *Note:* | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |

From the results obtained above, the p-value of benzene is greater than 0.05, this indicates that there is no significant effect of air quality (Benzene Concentration) on the price of houses in NYC. The adjusted R-squared value is obtained to be 0.019, indicating that the model fitted was only able to explain 1.9% variation in average rent price in NYC.

**Conclusion**

In this project, we examine the effect of air quality (Average Benzene concentration) on the rent price in NYC. Data were obtained from two sources, NYC OpenData and Zillow. Based on the analysis done on this data, we therefore conclude that there is no significant relationship between the rent price of houses on NYC and air quality (average benzene concentration).