**PROBLEM STATEMENT**

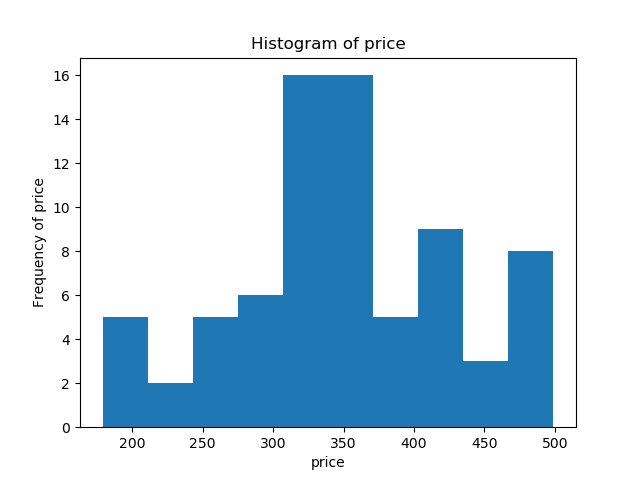
After completing graduation, some of your friends want to go abroad and pursue further studies. However, they’re concerned about the accommodation process and prices. You volunteered to take the responsibility of showing them the average price of rooms in different cities.

Your Task is divided into 2 sub-tasks:

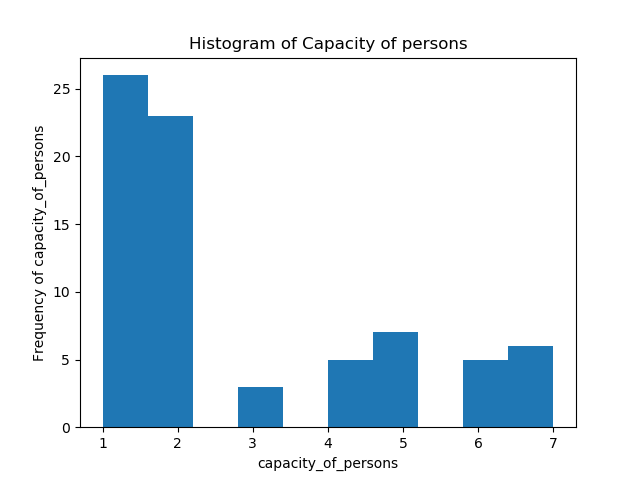
1. Data Collection
2. Data Analysis

**SCREENSHOTS OF FIGURES**

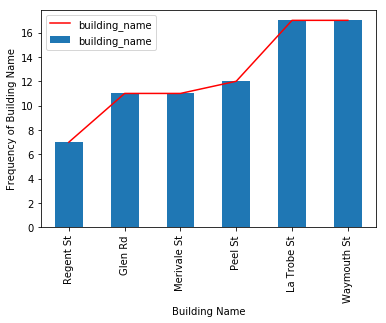
**Figures for Univariate analysis:**



**Figure 1: Histogram of price**

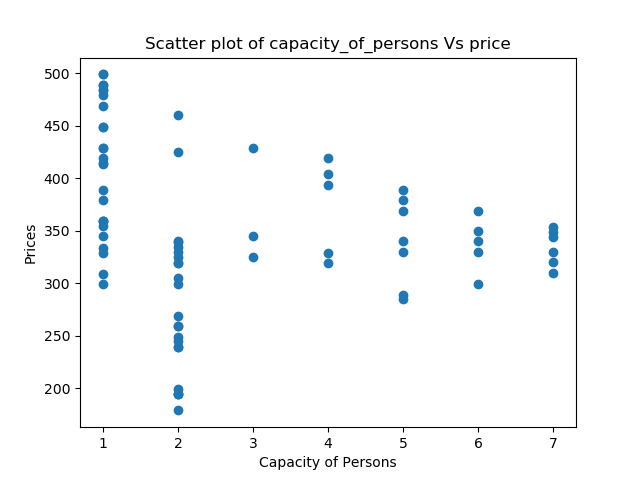


**Figure 2: Histogram of capacity\_of\_persons**

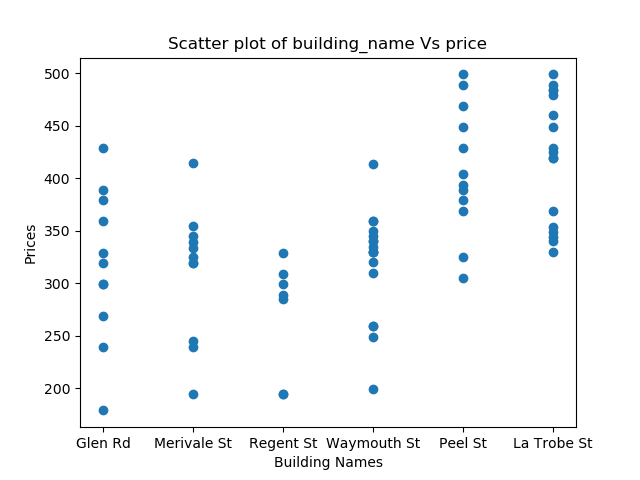


**Figure 3: graph of building\_name**

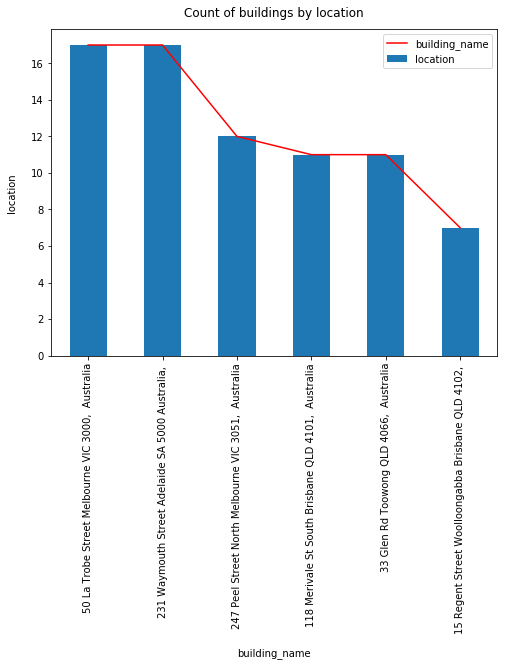
**Figure for Bivariate analysis:**



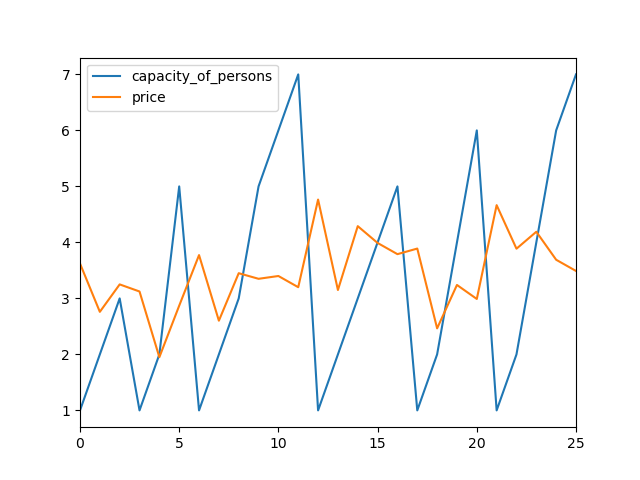
**Figure 4: Scatter plot for capacity\_of\_persons Vs price**



**Figure 5: Scatter plot for building\_name Vs price**



**Figure 7: Plot for building\_name Vs location**



**Figure 7: Line plot for capacity\_of\_persons Vs price**

**TECHNOLOGY STACK USED**

I used python language for the data scraping and data analysis because if you want to code and scrape the data yourself , Python is the best and the recommended as it has lot of web scraping and data analysis libraries and frameworks helping in extracting tons of data like:

* Selenium (for scraping dynamic data from web)
* Beautiful soup (for scraping static data)
* Requests (to make connection with web)
* Matplotlib (for plotting the data)
* Pandas (for dealing with dataframes)

I think a lot of depends on what you actually need from your web scraper. So then it would be more clear on what features are important to discuss and put into the account. From my perspective, the best and most useful web scraping libraries for this task are Selenium and BeautifulSoup. Even if they have some differences, they have more similarities with each other. Both are pretty easy to. Both use Python language. Both generate very detailed results so it’s pretty easy to do great analysis afterward.

**TOTAL HOURS FOR THE COMPLETION OF TASK**

|  |  |  |
| --- | --- | --- |
| **TASK NO.** | **TASK NAME** | **TOTAL HOURS TO COMPLETE** |
| 1 | Data Scraping | 4 |
| 2 | Data Analysis | 1 |
| 3 | Documentation | 0.5 |
| 4 | Full task | 5.5 |

**Task 1: Total Hours for the completion of task**

**FURTHER IMPROVEMENTS**

I would make some improvements if I had more time. Below are some expected improvements:

1. I would convert the ‘features’ column into ‘number of features’ and ad this column into existing dataset because I think number of features also correlates with price of room. So, in this way, student can easily see which room will be best suited for him/her.
2. I would scrape the duration of stay at room (Annual or semester-wise) and price accordingly because it makes sense for the student to check the price of room according to the duration of stay at room with other features as well.
3. I would scrape the floor of the room and afterwards student could see the average price of low floor rooms available at a specific location or so on.