#### **TASK 1 (Business analytic question)**

Pricing strategies: How does pricing vary across different neighbourhoods, and how is it related to the number of reviews and the room type?

#### **TASK 2 (Data analysis)**

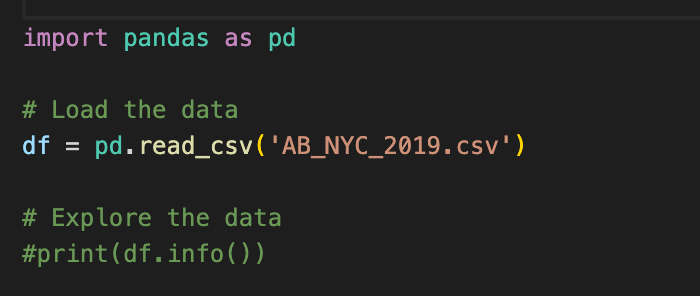
1. Data Loading and Exploration

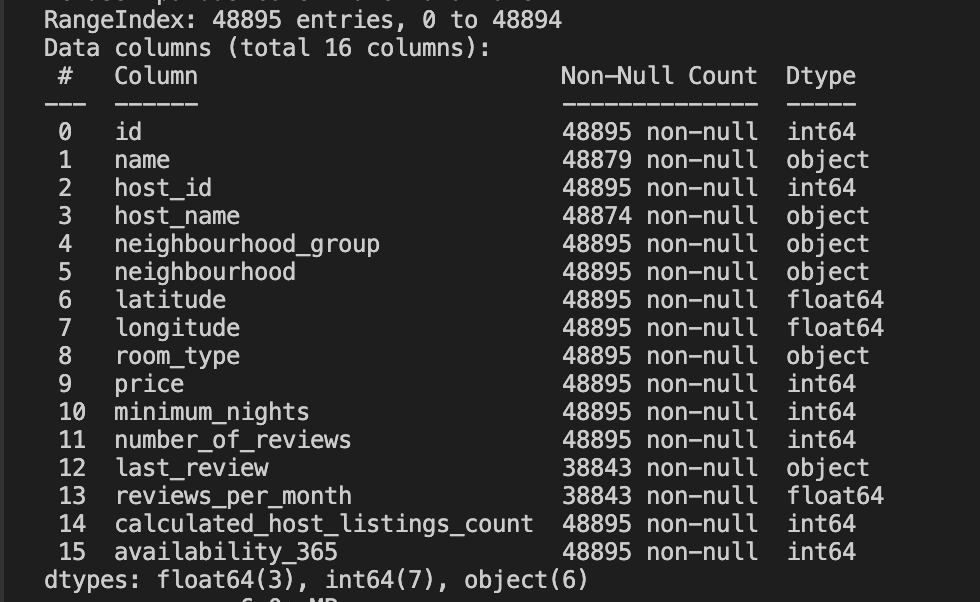
2. Data Preprocessing

3. Exploratory Data Analysis and Data Visualisation

4. Statistical Analysis

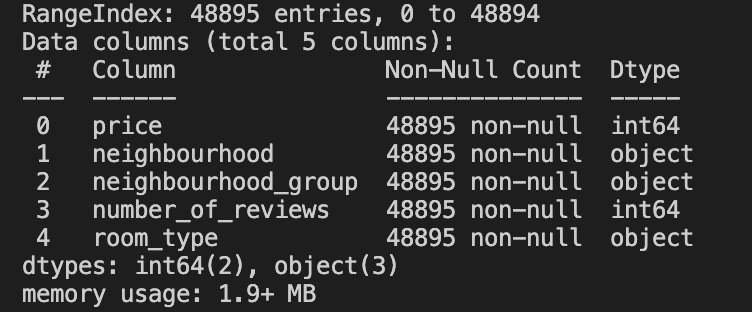
**1. Data Loading and Exploration**

****

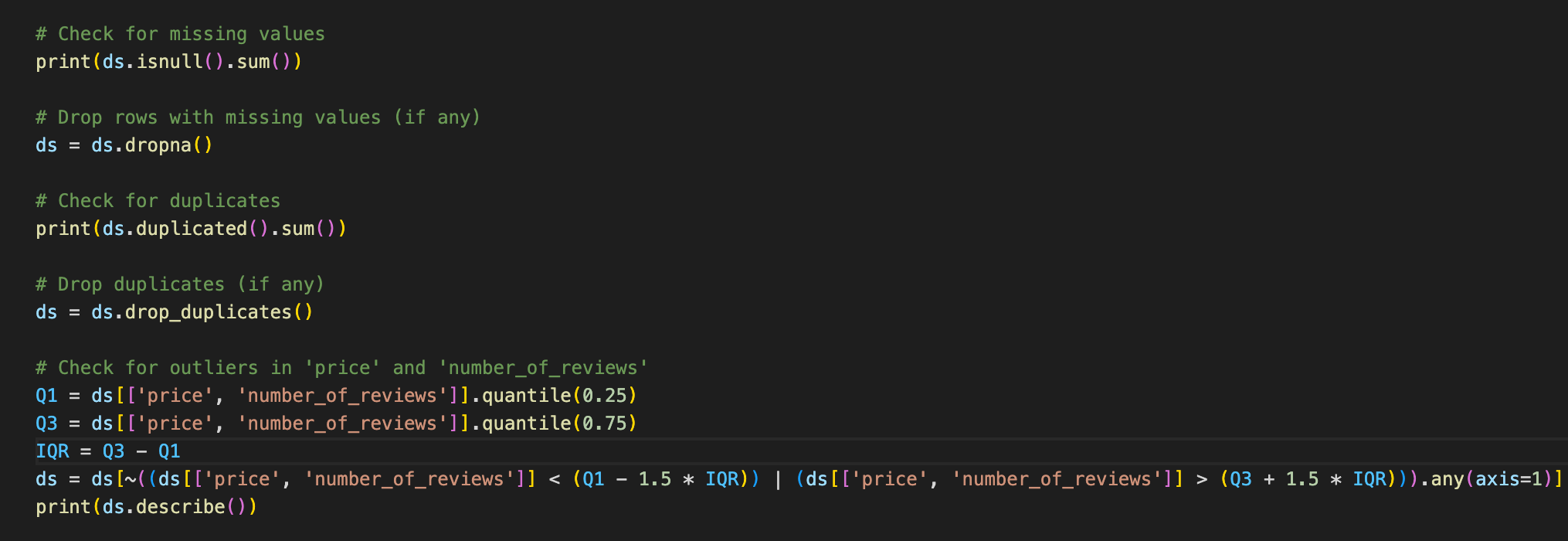
****

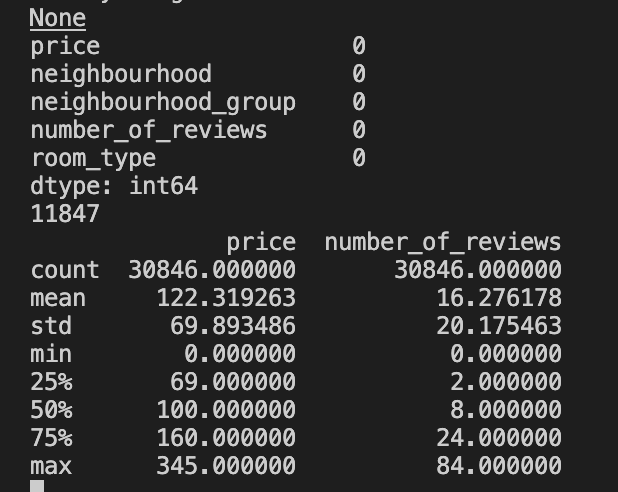
- According to our business analysis question, we will analyse based on these columns **price, neighbourhood, neighbourhood\_group, number\_of\_reviews,** and **room\_type**



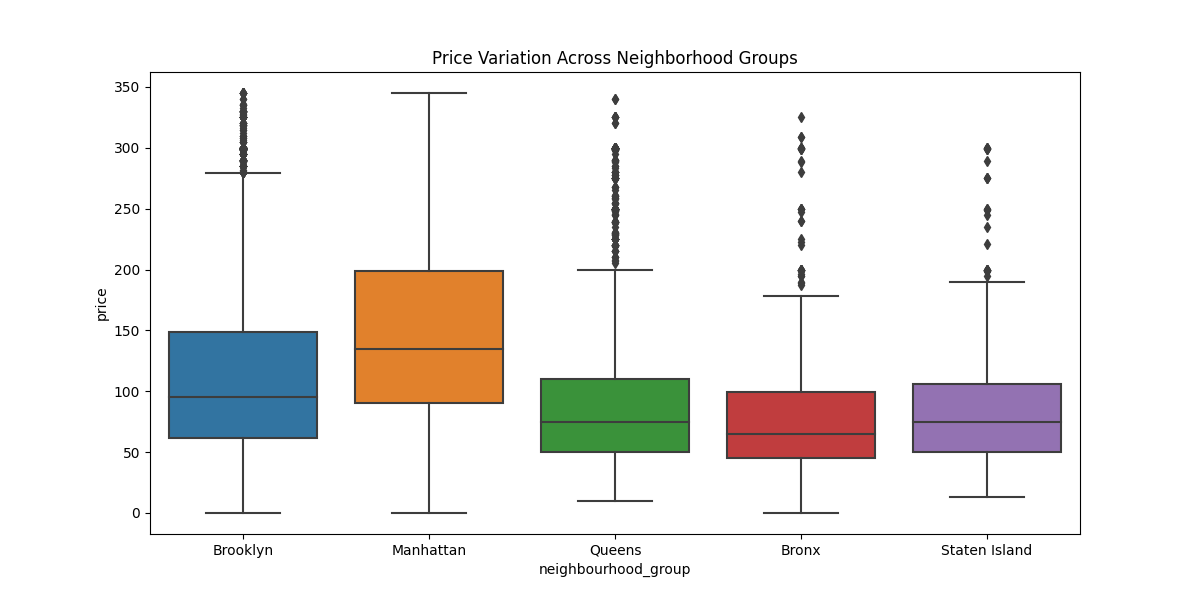
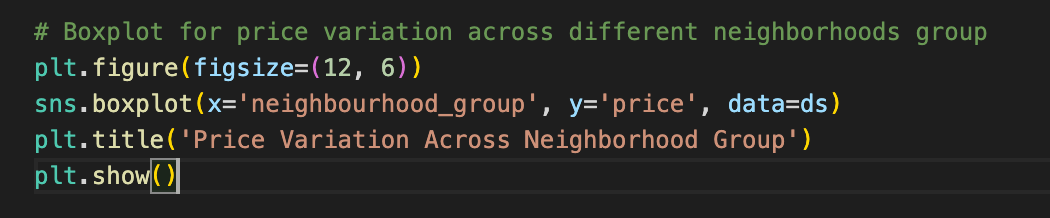


**2. Data Preprocessing**

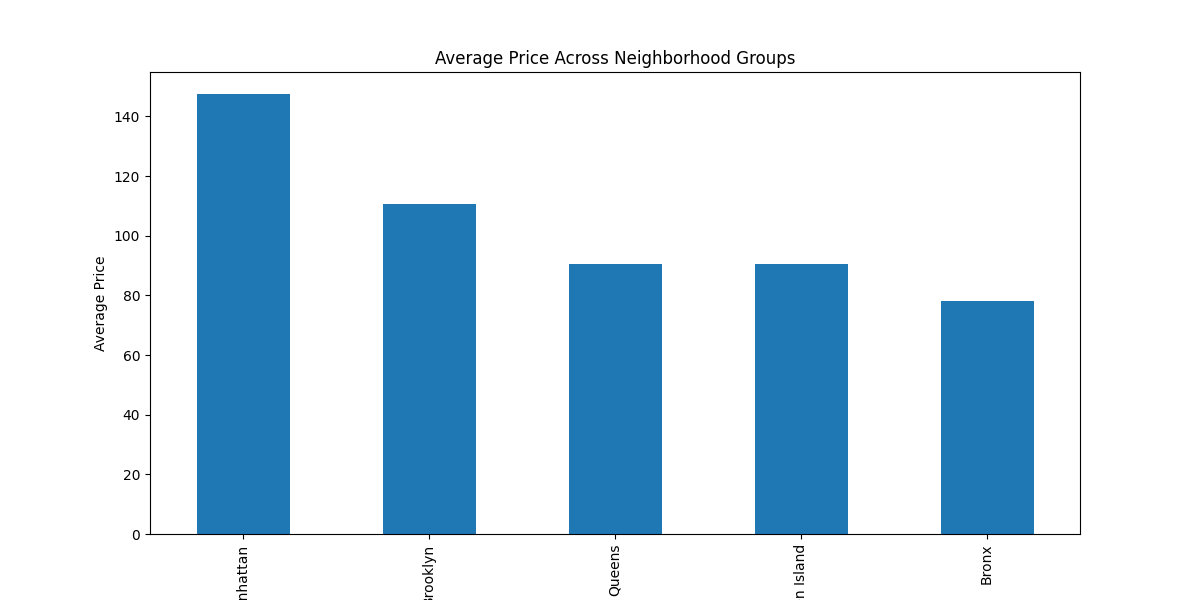




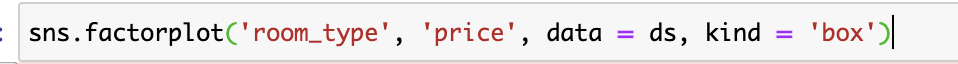
**3. Exploratory Data Analysis and Data Visualisation**

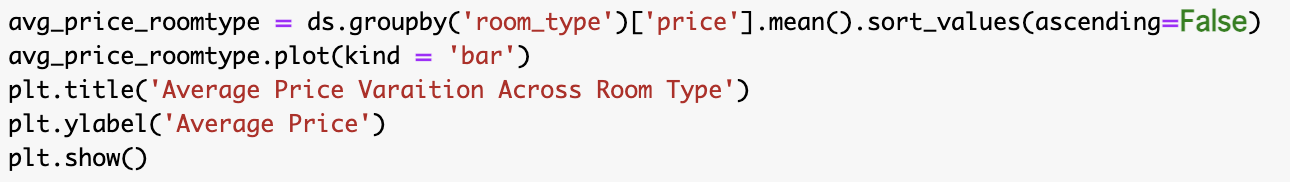


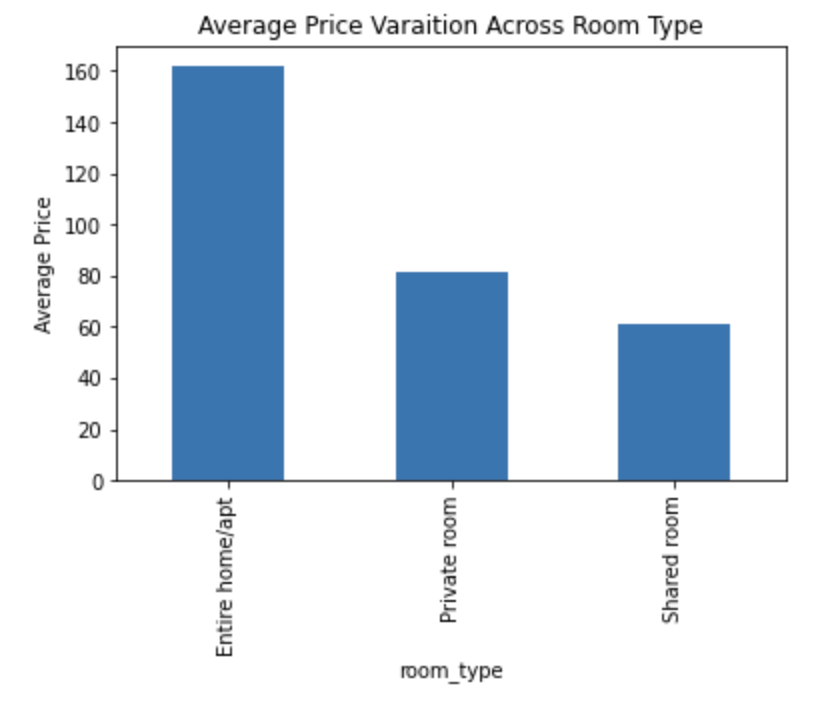


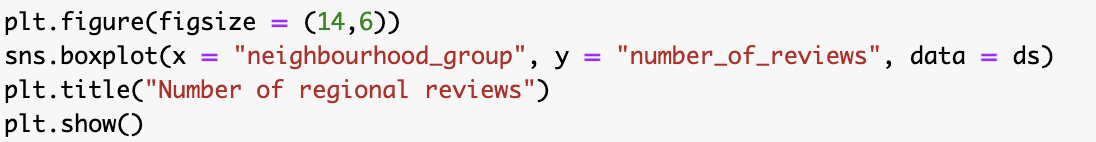


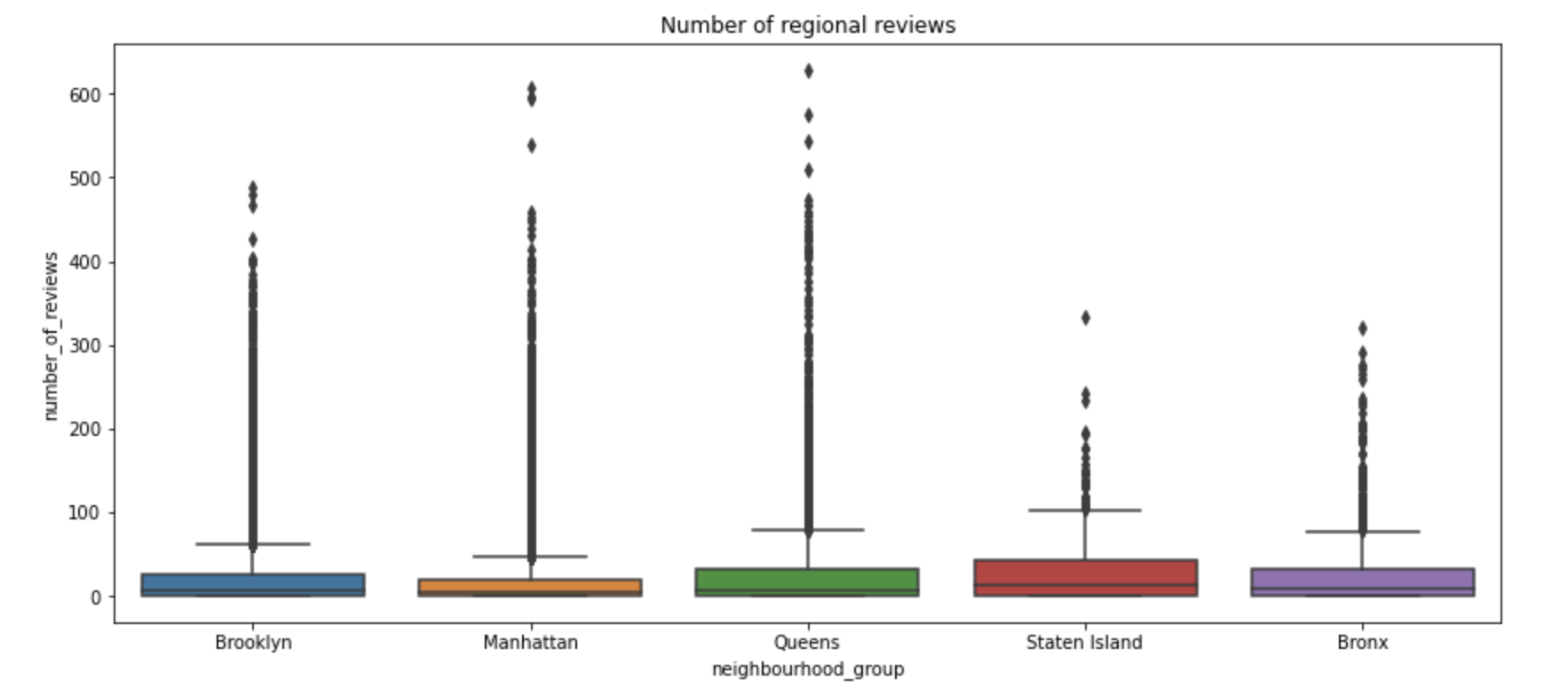
A picture containing screenshot, diagram, text, line

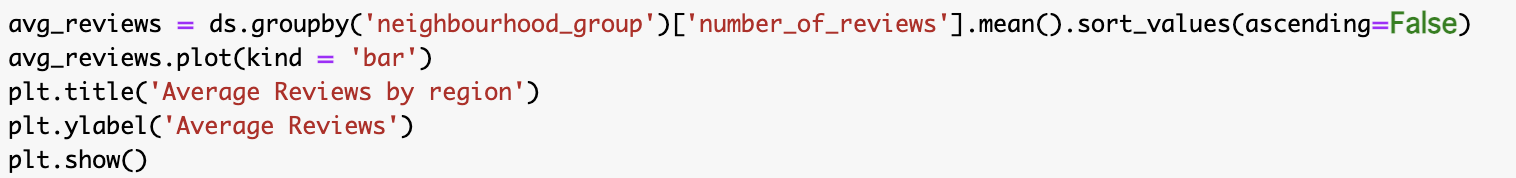
Description automatically generated

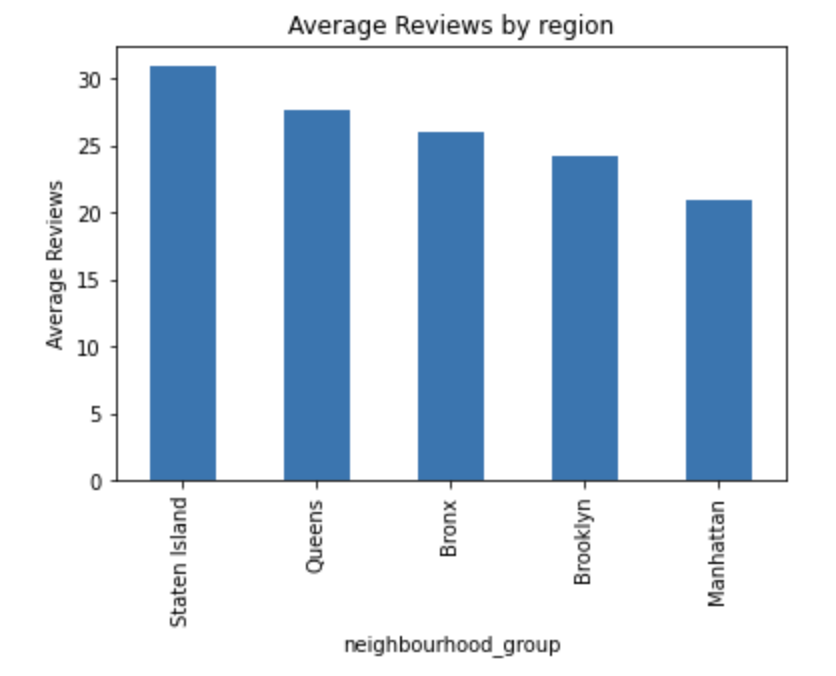














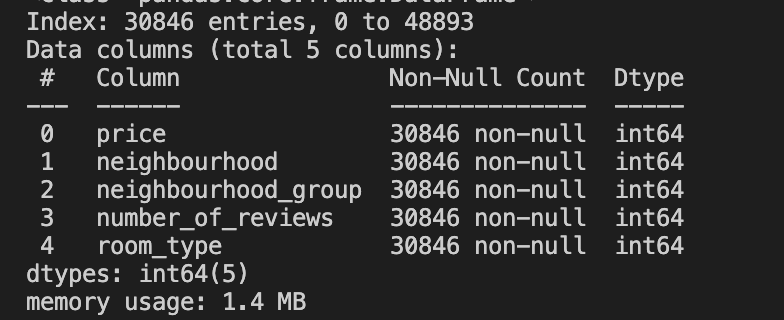
A picture containing text, screenshot, aqua, line

Description automatically generated

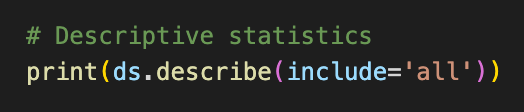
**4. Statistical Analysis**

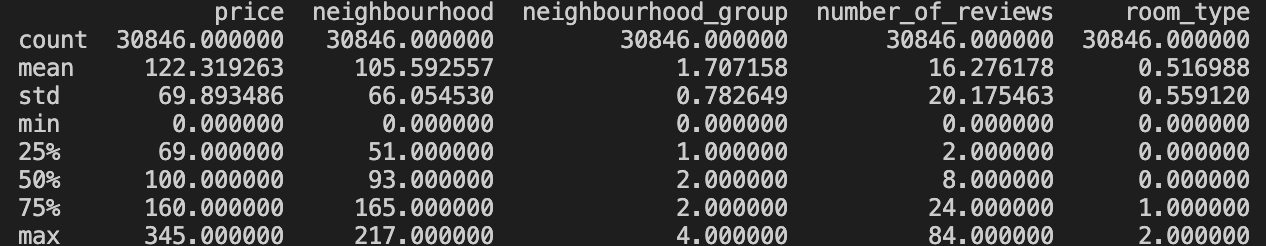
**Transform categorical data into numeric form**



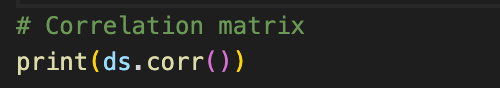


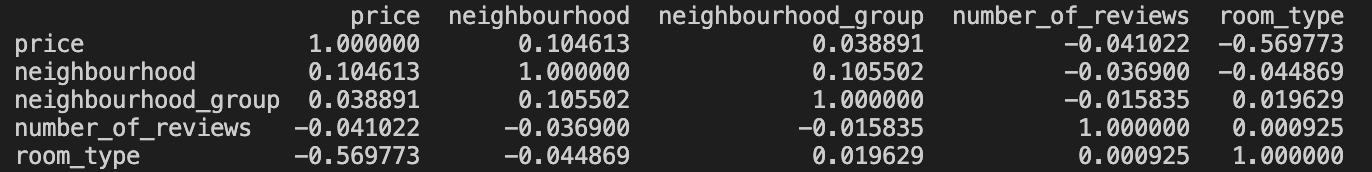
**Descriptive Statistic**

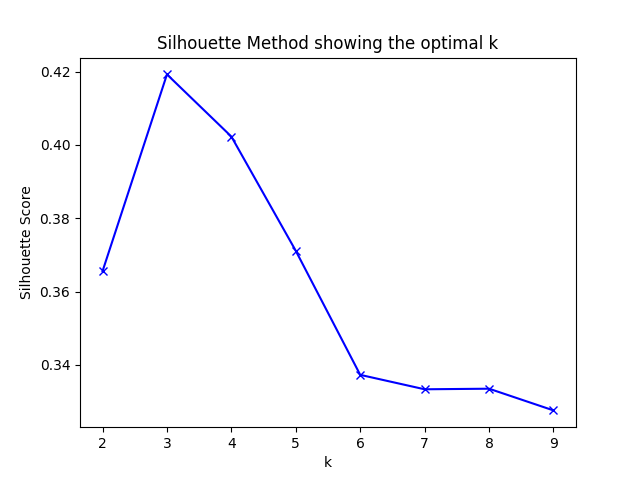
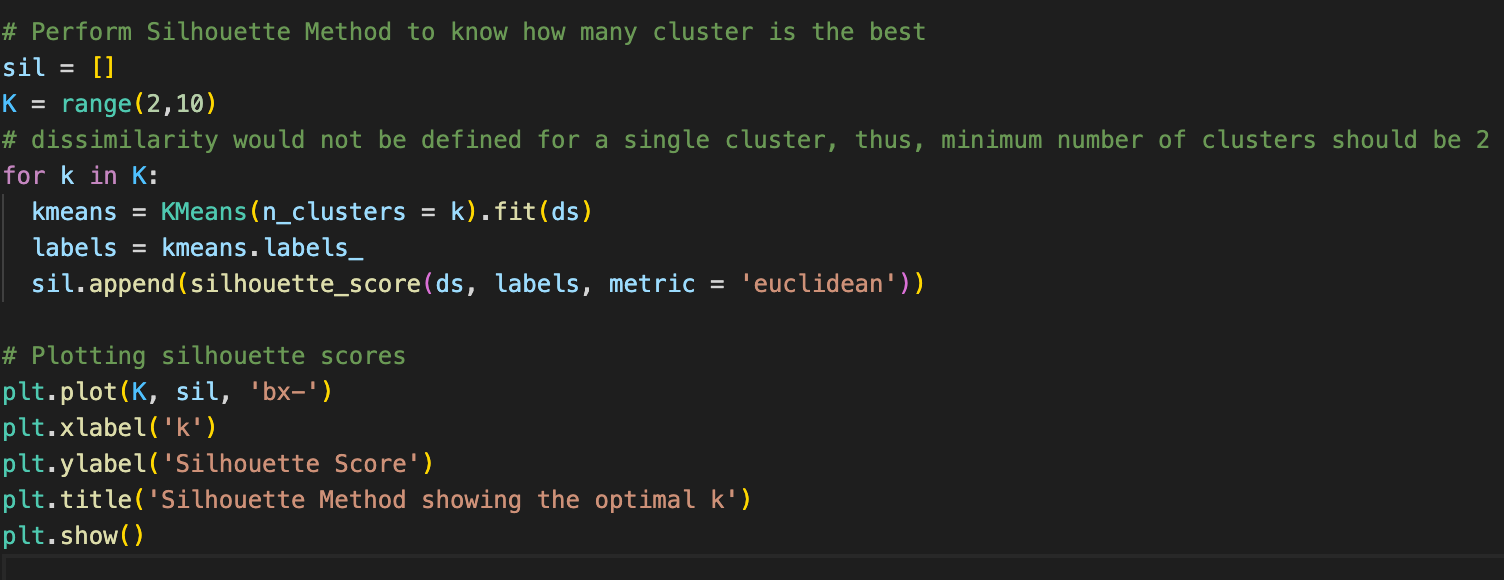




**Correlation Matrix**

****

****

**Silhouette**

**K-mean Clustering**

- According to the result of the Sihouette method, n\_cluster = 3 is the best.

