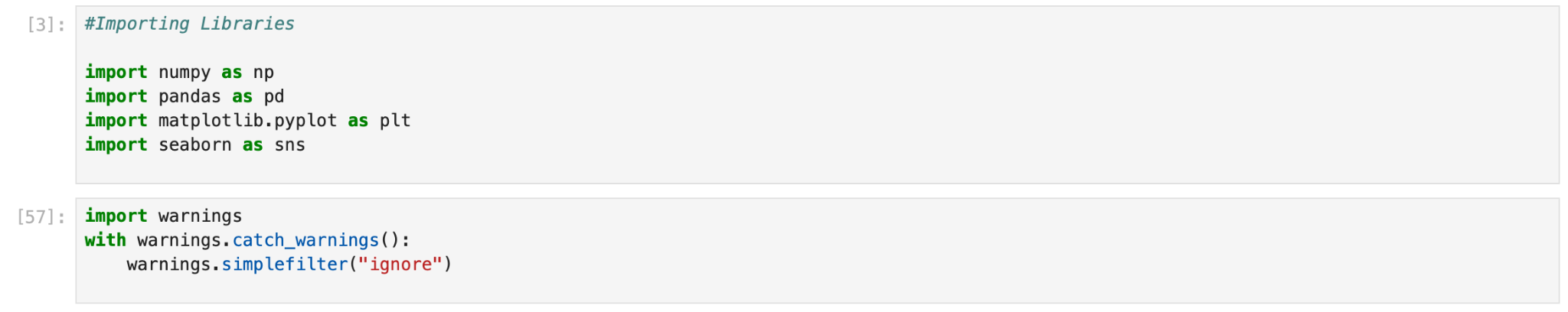
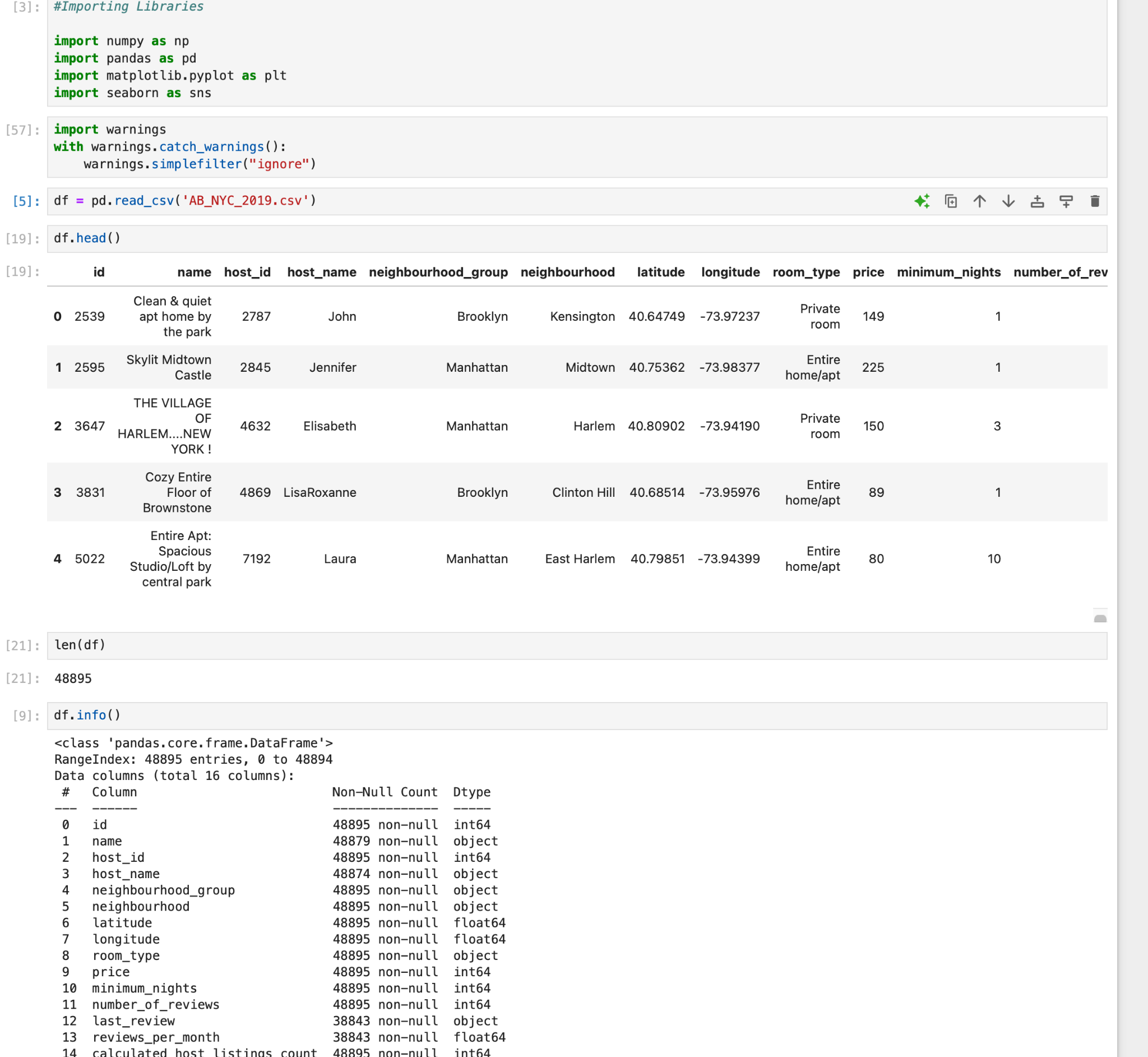
**AirBnb Case Study**

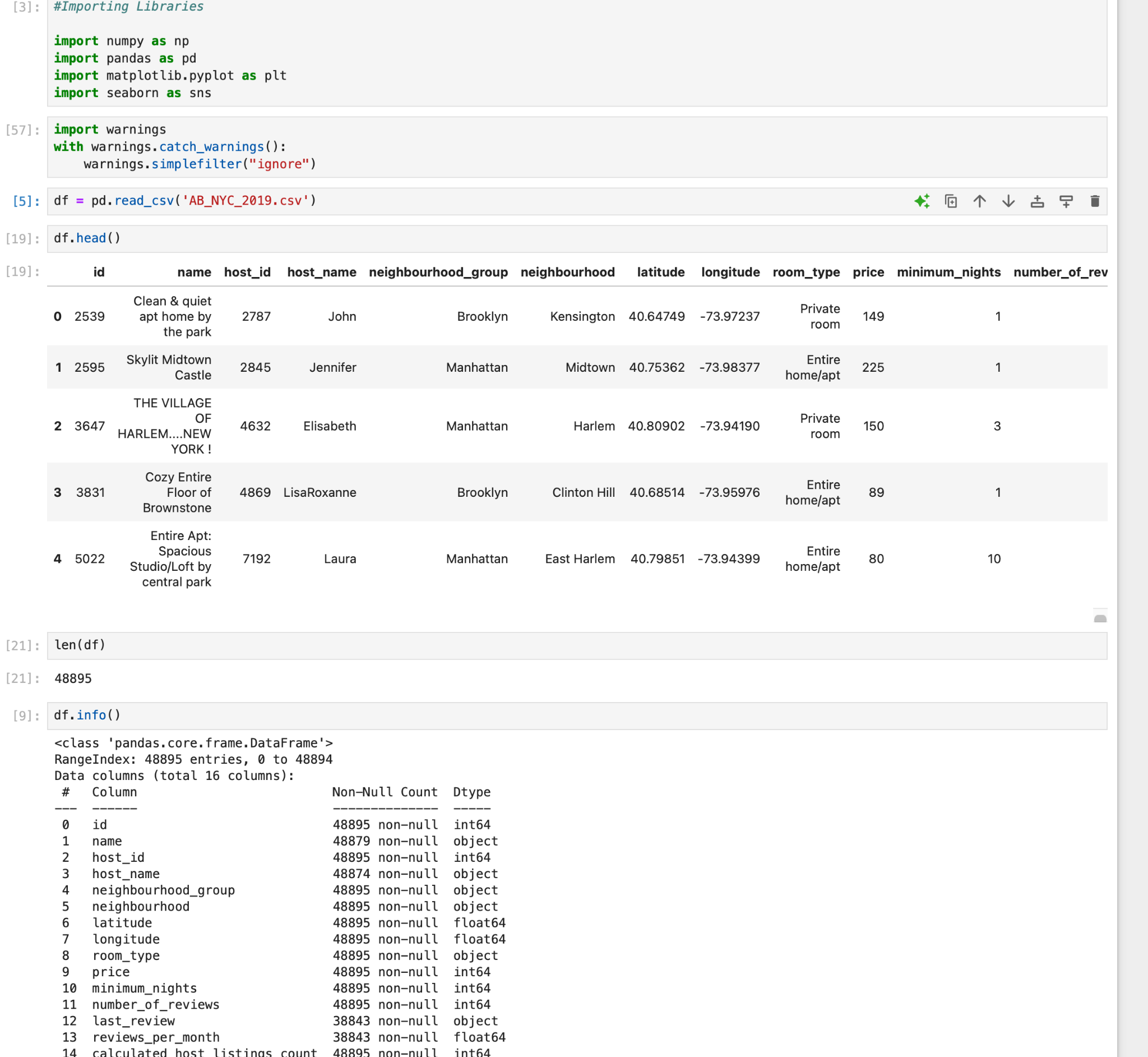
#### **1. Dataset Inspection**

The dataset was loaded into Python for initial analysis, ensuring a thorough understanding of its structure and content:

* **Data Overview:** Displayed the first few rows using head() and reviewed the dataset's shape to understand the number of rows and columns.
* **Data Types and Summary:** Used info() and describe() to inspect data types, identify numeric and categorical variables, and review statistical summaries.
* **Missing Values:** Checked for missing values across all columns using isnull().sum() and calculated the percentage of missing values to prioritize necessary cleaning steps.







#### **2. Handling Missing Values**

Missing values were addressed with appropriate imputation strategies based on the type and context of each column:

* **Text Columns:** Missing values in name and host\_name were filled with the placeholder "Unknown" to maintain data consistency.
* **Date Columns:** Missing values in last\_review were imputed with a placeholder date of "2000-01-01" to avoid null values while preserving data format.
* **Numeric Columns:** The reviews\_per\_month column was imputed with a value of "0" to indicate no reviews in cases where data was missing.

Post-imputation, missing values were rechecked to confirm successful handling.



**3. Data Cleaning and Optimisation**

Steps were taken to standardize and clean the dataset for accuracy and usability:

* **Categorical Data:** Cleaned text formatting in the room\_type column by converting all text to lowercase and removing extra spaces to ensure consistency.
* **Data Type Correction:** Converted the last\_review column to datetime format using pd.to\_datetime() for accurate handling of date-related analyses.



#### **4. Saving and Loading**

The cleaned and optimized dataset was saved for reproducibility and visualization:

* Exported the cleaned dataset to a CSV file for storage and accessibility.
* Loaded the cleaned dataset into Tableau for creating visualizations and conducting further exploratory data analysis.
* Ensured compatibility between Python outputs and Tableau for seamless integration.



#### **5. Visualization and Insights (Using Tableau)**

After cleaning, the dataset was loaded into Tableau to create interactive dashboards and visualizations. Key visualizations included:

**PPT I - Insights for Data Analysis Managers and Lead Data Analyst**

1. **Top Earning Hosts:** Visualized top-performing hosts based on revenue to identify high earners.
2. **Review Activity by Neighborhood:** Analyzed review counts across neighborhoods to understand customer engagement.
3. **Room Types Distribution:** Displayed the proportion of listings for each room type.
4. **Review Distribution by Neighborhood Group:** Highlighted review activity differences among neighborhood groups.
5. **Monthly Review Trends:** Showed review activity trends over time to analyze seasonality and popularity.

**PPT II - Head of Acquisitions & Operations and the Head of User Experience**

1. **Listings by Neighborhood:** Visualized the concentration of listings in different neighborhoods.
2. **Pricing Preferences by Room Type:** Explored price ranges associated with different room types to identify user preferences.
3. **Property Availability Distribution:** Analyzed the distribution of property availability to assess supply across neighborhoods.
4. **Review Count and Price Relation:** Investigated the correlation between review activity and pricing for insights into demand dynamics.
5. **Neighborhood Popularity and Pricing:** Examined price variations across neighborhoods to identify high-demand locations.