**CAPSTONE PROJECT DRAFT REPORT**

Logo

Description automatically generated

**Northeastern University, Boston**

**College of Professional Studies**

**ALY6140 70584 Analytics Systems Technology**

**Professor:**

Richard He

**Project Members:**

Sakshi Grover

Durga Bhanu Nayak

Rohith Mallula

Abhay Somani

**Date of Submission:** 12/02/2021

# INTRODUCTION

The era that we all are living is being driven by the data. And it is being generated in enormous volume and speed.

We can collect the data that is required using various tools and methodologies and use it to extract meaningful results and can make well informed decisions.

And to share our findings and insights to those who need to take data driven decisions, data visualization comes to the picture. A compelling and meaningful visualization of data will help the audience or stakeholders to capture the gist of the analysis.

The purpose of this draft is to explore a public dataset from Airbnb for Amsterdam location using basic data exploratory techniques. Cleaning of dataset by dropping unused variables, converting the variables as per the required data types, checking for null values etc. are some of the processes that are included in the data cleaning process.

After the data cleanup we will try to understand the cleaned data more by applying the visualization techniques and plot graphs to answer our questions to the dataset.

**Why we chose Airbnb Amsterdam?**

AirBnb is an online platform actively working in the short-term rental property sector. It facilitates people around the globe to rent out space for temporary accommodations, which presents an excellent choice for those who are commuting to new cities and looking for short stays. We have chosen this dataset to understand and analyze the current statuses of the company's data with respect to its available features and prices provided to its customers. This dataset lists the company's activities in Amsterdam for the year 2021. As it includes various categorical (such as amenities of the listed accommodations, neighborhood overview, location, hostname, host response time, etc.)  as well as numerical variables (rating/ review, pricing, number of bedrooms, number of bathrooms, dates of availability, and many more) which makes it apt for performing and presenting our analysis on this dataset.

# EXPLORATORY DATA ANALYSIS

For this project, we have used Jupyter notebook IDE or Integrated Development Environment with Python programming language to curate our primary python notebook (capstone\_group10.ipynb) and, we have also used Spyder IDE to prepare the script file (capstone\_group10.py) which contains the self-defined functions which is further called in the main python notebook.

## DATA EXTRACTION

To get the data for this project, we have used Airbnb Amsterdam data that is publicly available by Airbnb under appropriate license. However, prior to loading the data using Python, required libraries for reading the data, visualizing it and preparing models over it are needed to be imported.

* Loading python libraries

Text

Description automatically generated

**Pandas** and **NumPy** are libraries used for the data analysis

**Seaborn** and **matplotlib** are the libraries used for the data visualizations

**Capstone\_group10** utility file is being used to invoke the self-defined functions

* Loading the dataset

Chart

Description automatically generated with low confidence

Here, we have used the function from the utility file to read the csv file by passing the path of the data file in parameter.

Along, with loading the file, we have invoked the **head()** function to check if the data is loaded properly or not to the data frame.

Graphical user interface, text

Description automatically generated

## DATA CLEANING

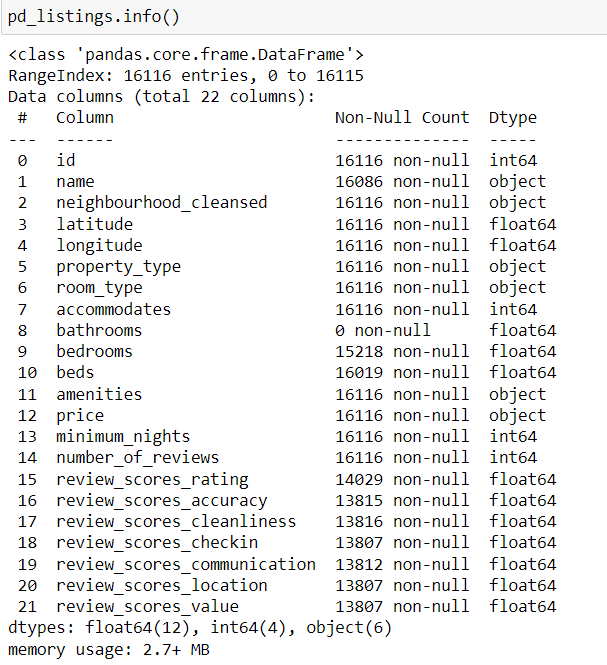
Under this section, we will be performing the data cleaning process by analyzing the dataset.

* Understanding the data
* Shape of the data



From the above command, we get to know that our dataset contains 16116 rows and 74 columns.

* We have dropped the variables which are not required and out of the required variables, looking at the information of the variables.



* Converted the ‘Price’ variable from object to decimal type using the function from utility file.



* We have also replaced the null values wherever present in the numerical variables from the dataset as 0.

Text

Description automatically generated

## DESCRIPTIVE ANALYSIS

## DATA VISUALIZATION

# PREDICTIVE MODELS

# CONCLUSION

# BIBLIOGRAPHY

# CONTRIBUTIONS