**Academic year 2023-24**

**PROJECT-VA**

**NAME: Aman Singh Bhogal and Mausmi Sinha**

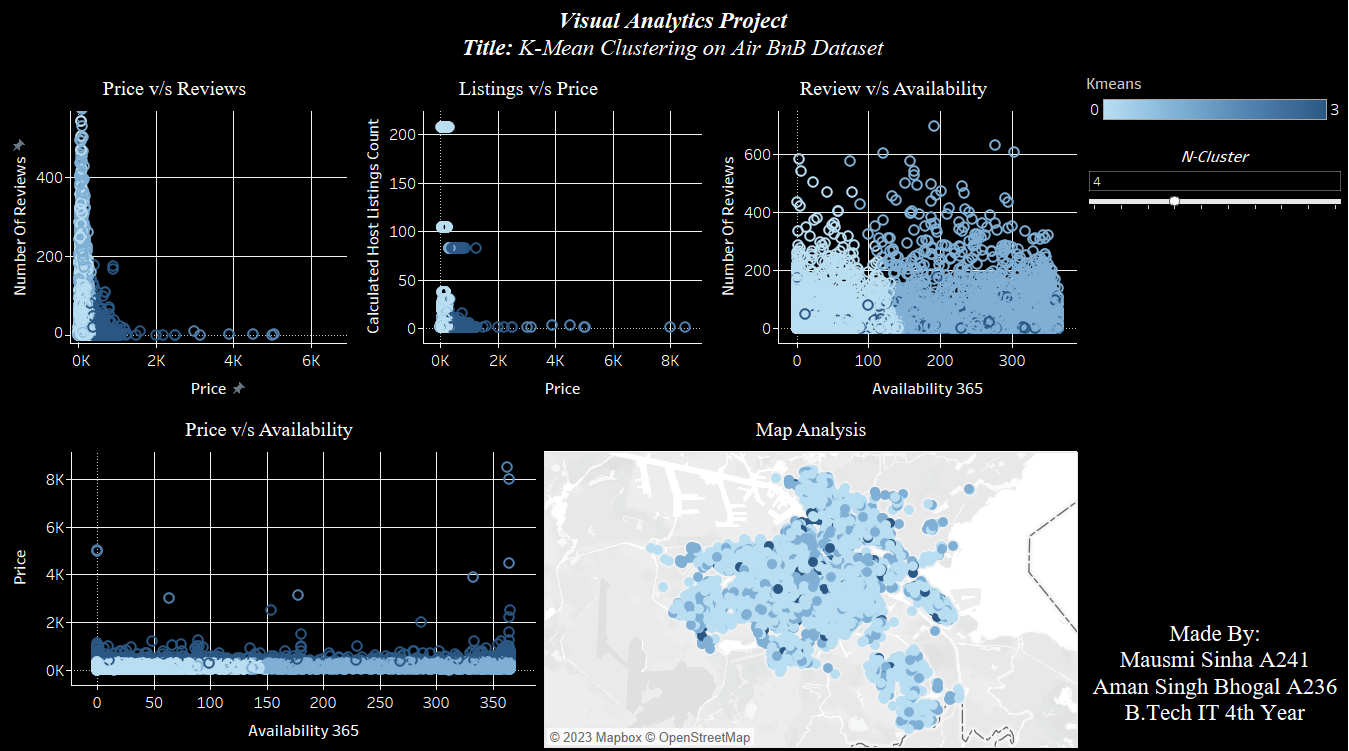
**ROLL-NO: A236 & A241**

**B.TECH IT-Final year.**

Apply any Statistical Analysis Technique for any real-world problem with respect to any dataset in platforms like R, PYTHON etc Implement a ML Model. Provide an Interface of ML model to any BI tool to show the visualizations of the Project you have implemented.

**Title:** *K-Mean Clustering on Air BnB Dataset*

**Output Screenshots:**

****

**A screenshot of a computer screen

Description automatically generated**

**A screenshot of a computer screen

Description automatically generated**

**Individual Charts:**

**A black screen with white lines

Description automatically generated**

**A screen shot of a computer

Description automatically generated**

**A screen shot of a computer screen

Description automatically generated**

**A screen shot of a computer

Description automatically generated**

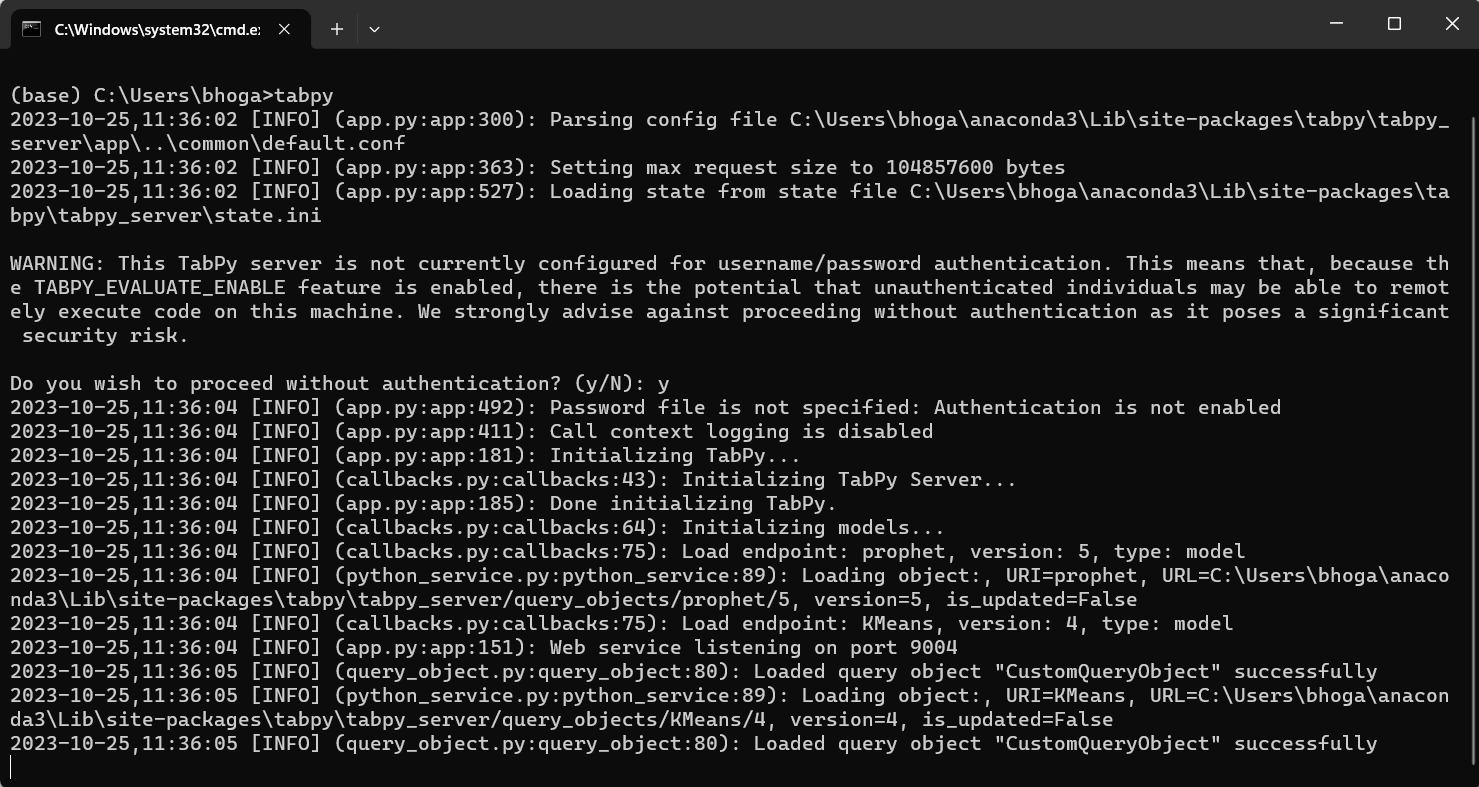
**A map of a city

Description automatically generated**

**Method Used:**

In this project we have used TabPy library that allows us to connect our tableau environment with Python Server.

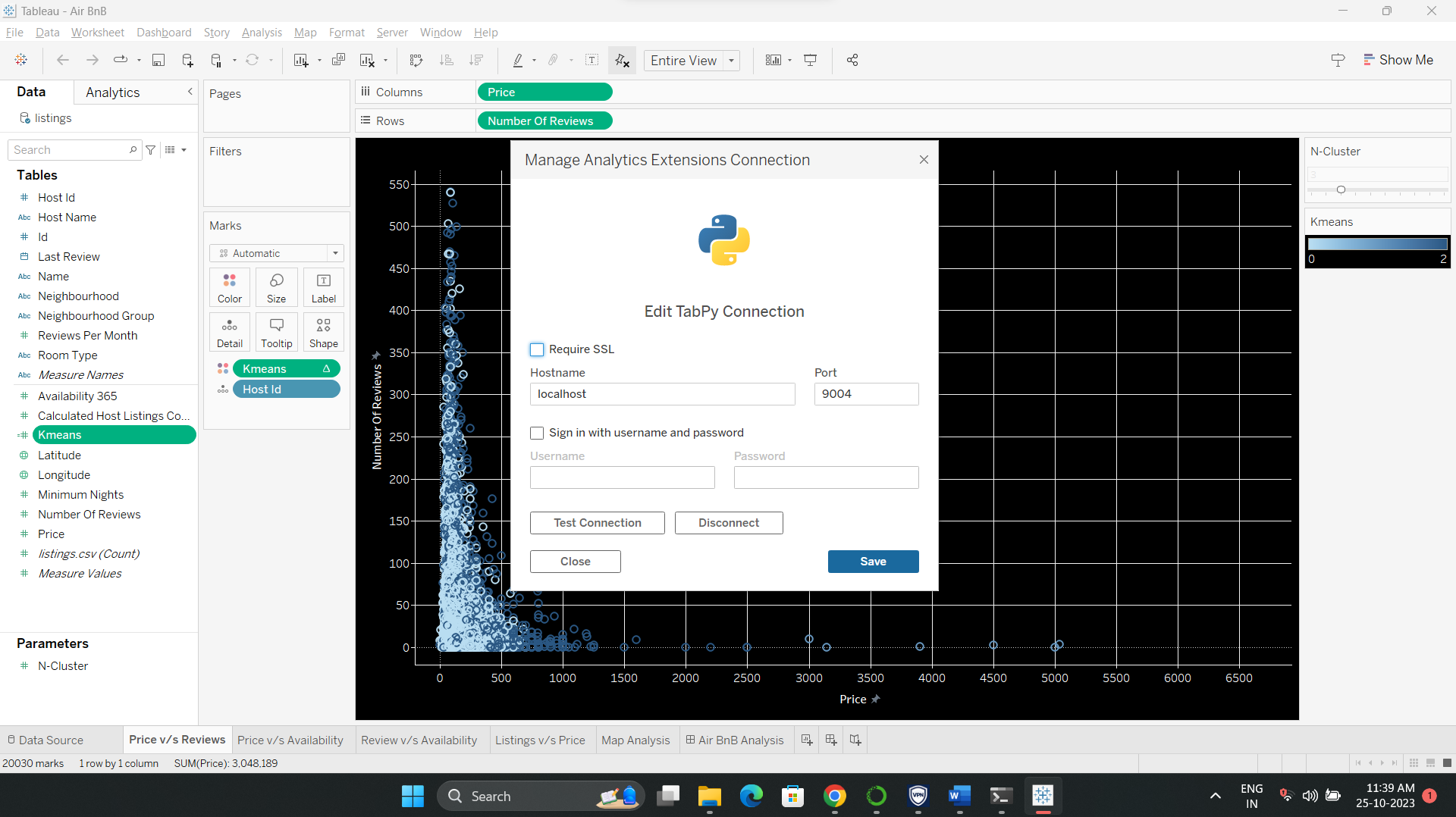
**Starting TabPy server:**



A screenshot of a computer

Description automatically generated

**Connecting Tableau to our TabPy Server:**

****

**A screenshot of a computer

Description automatically generated**

**Writing Script in Tableau using SCRIPT\_INT:**

We created a calculated field using SCRIPT\_INT() that takes data attributes and N-Cluster Parameter as an input and passes it to our python code for calculating the clusters. We used K Means Clustering algorithm provided by sklearn library.

Code:

SCRIPT\_INT("from sklearn.preprocessing import LabelEncoder

from sklearn.cluster import KMeans

import numpy as np

LE = LabelEncoder()

neighbourhood = LE.fit\_transform(\_arg1)

room\_type = LE.fit\_transform(\_arg2)

price = \_arg3

minimum\_nights = \_arg4

number\_of\_reviews = \_arg5

availability\_365 = \_arg6

calculated\_host\_listings\_count = \_arg7

N = \_arg8[0]

X = np.column\_stack(

[

neighbourhood,

room\_type,

price,

minimum\_nights,

number\_of\_reviews,

availability\_365,

calculated\_host\_listings\_count,

]

)

kmeans = KMeans(n\_clusters=N, random\_state=35)

return kmeans.fit\_predict(X).tolist()

",

ATTR([Neighbourhood]),

ATTR([Room Type]),

AVG([Price]),

MEDIAN([Minimum Nights]),

SUM([Number Of Reviews]),

AVG([Availability 365]),

AVG([Calculated Host Listings Count]),

[N-Cluster]

)

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

**Conclusion:**

In summary, our lab experiment showcased the seamless integration of a K-means prediction model into Tableau using the TabPy library within the Anaconda environment. This integration resulted in the creation of dynamic dashboards that adapt their visualizations based on user input, significantly enhancing the clarity and accessibility of the information presented. The collaborative effort between data scientists and visualization experts proved instrumental in not only developing a robust prediction model but also translating its output into actionable insights. This interdisciplinary approach highlights the potential of combining advanced machine learning techniques with interactive data visualization, demonstrating how such integration can revolutionize the way we analyze and comprehend complex datasets, paving the way for informed decision-making in various domains.