<u>Academic year 2023-24</u>

PROJECT-VA

NAME: Aman Singh Bhogal and Mausmi Sinha

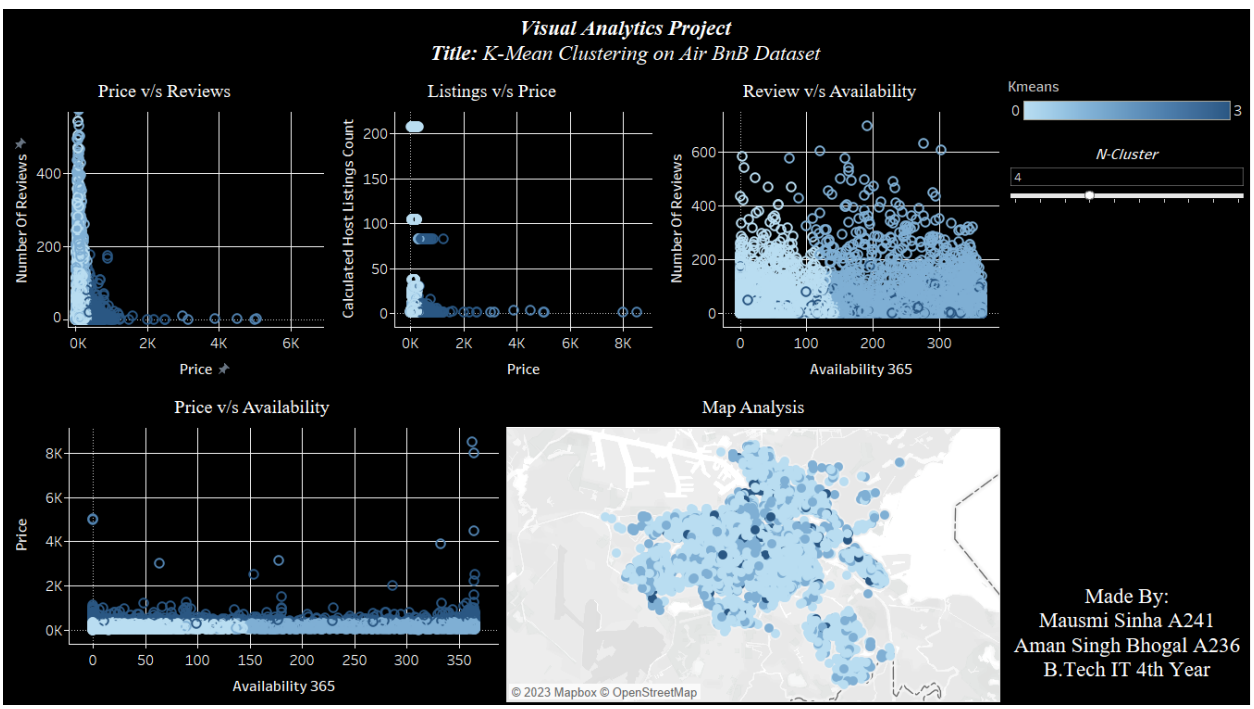
ROLL-NO: A236 & A241

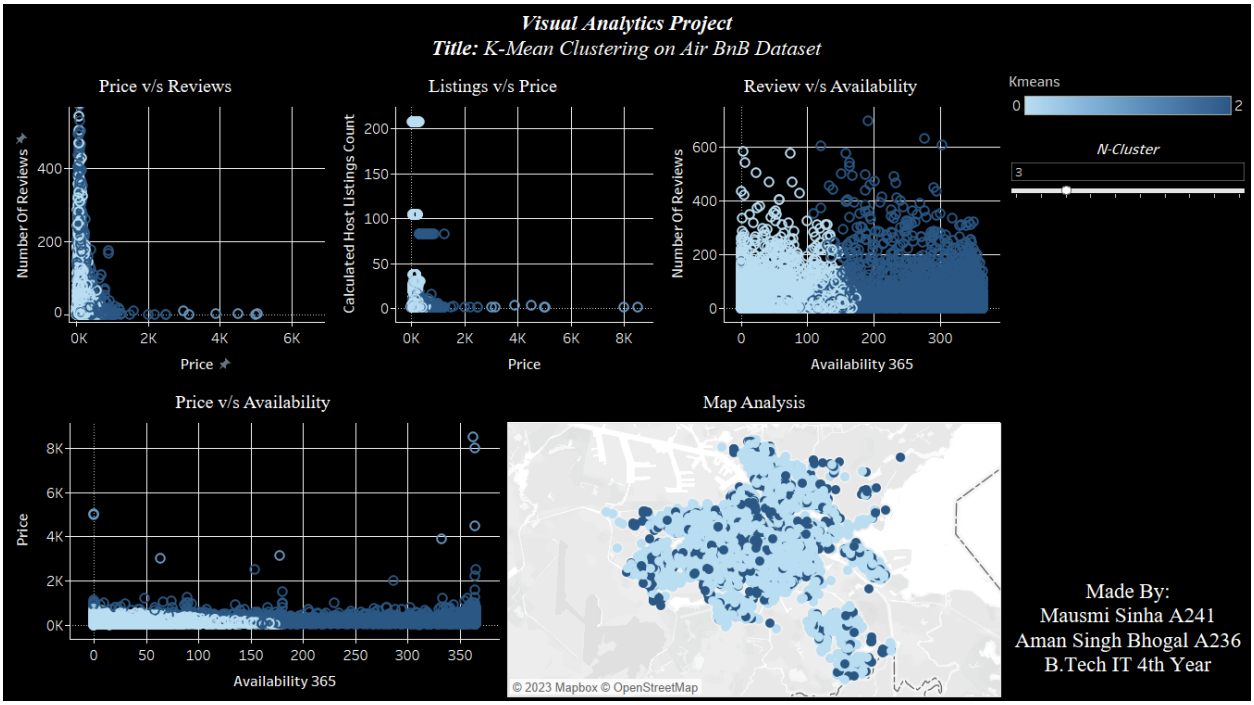
B.TECH IT-Final year.

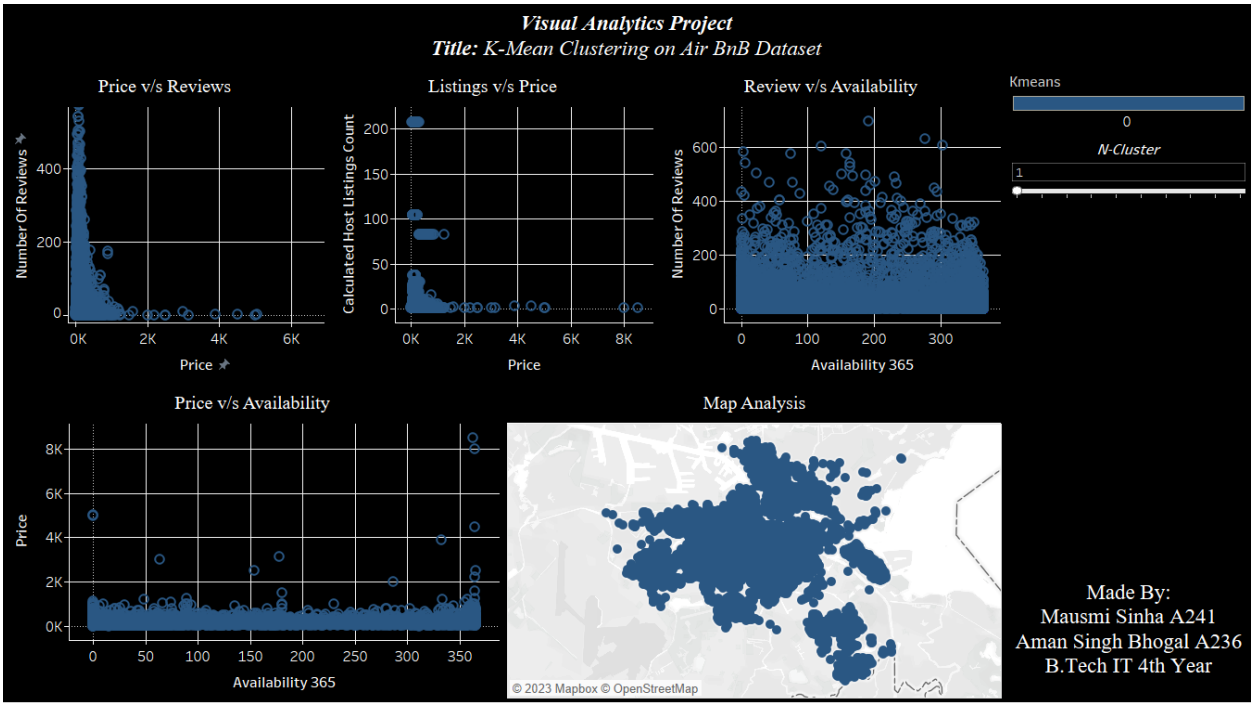
Apply any Statistical Analysis Technique for any real-world problem with respect to any dataset in platforms like R, PYTHON etc Implement a ML Model. Provide an Interface of ML model to any BI tool to show the visualizations of the Project you have implemented.

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

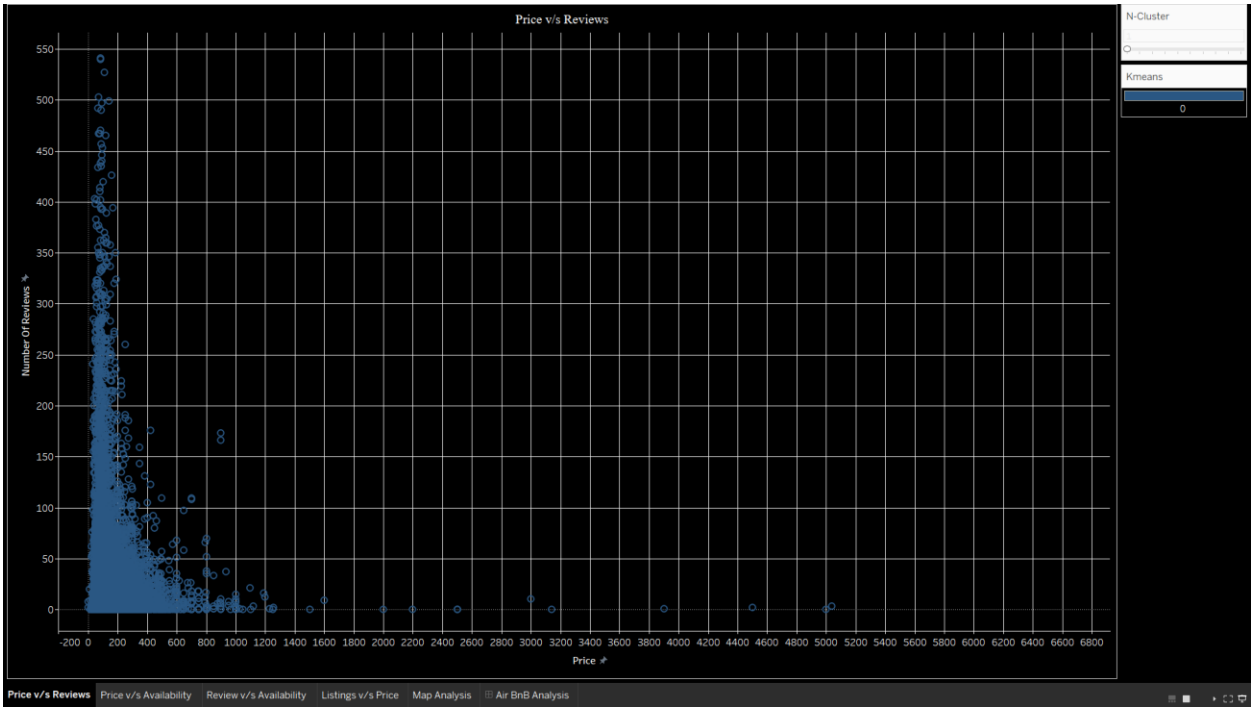
Output Screenshots:

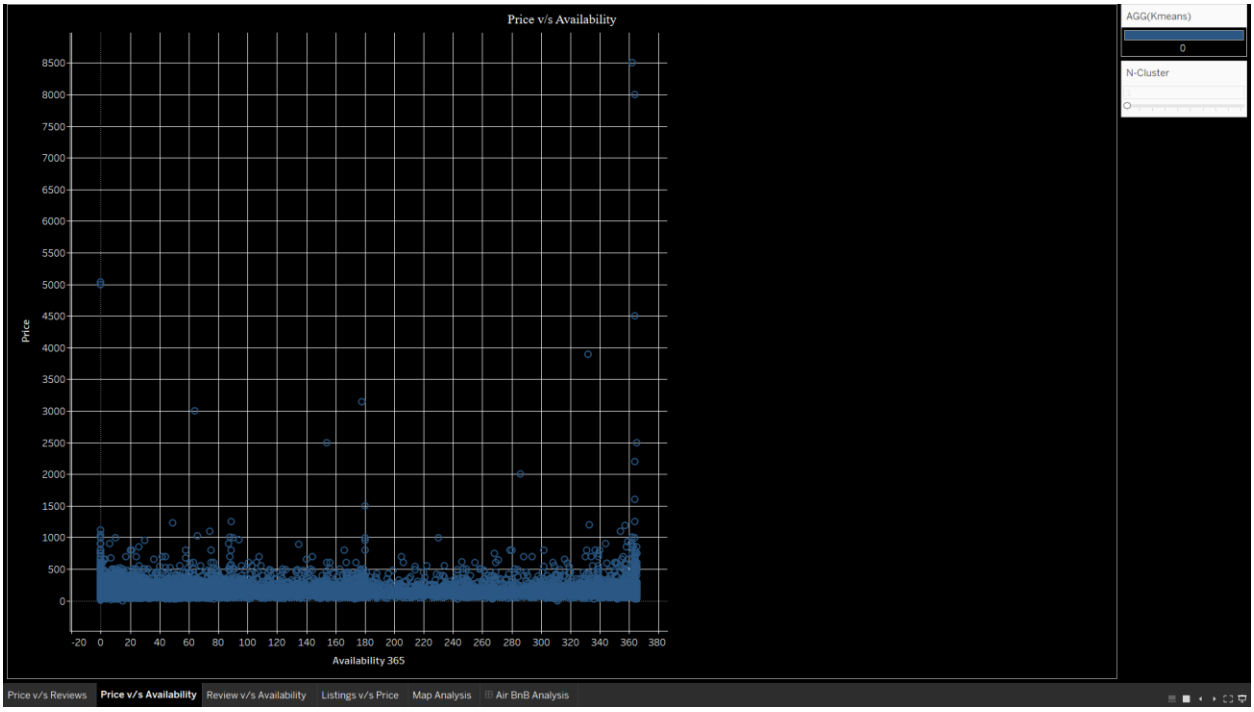


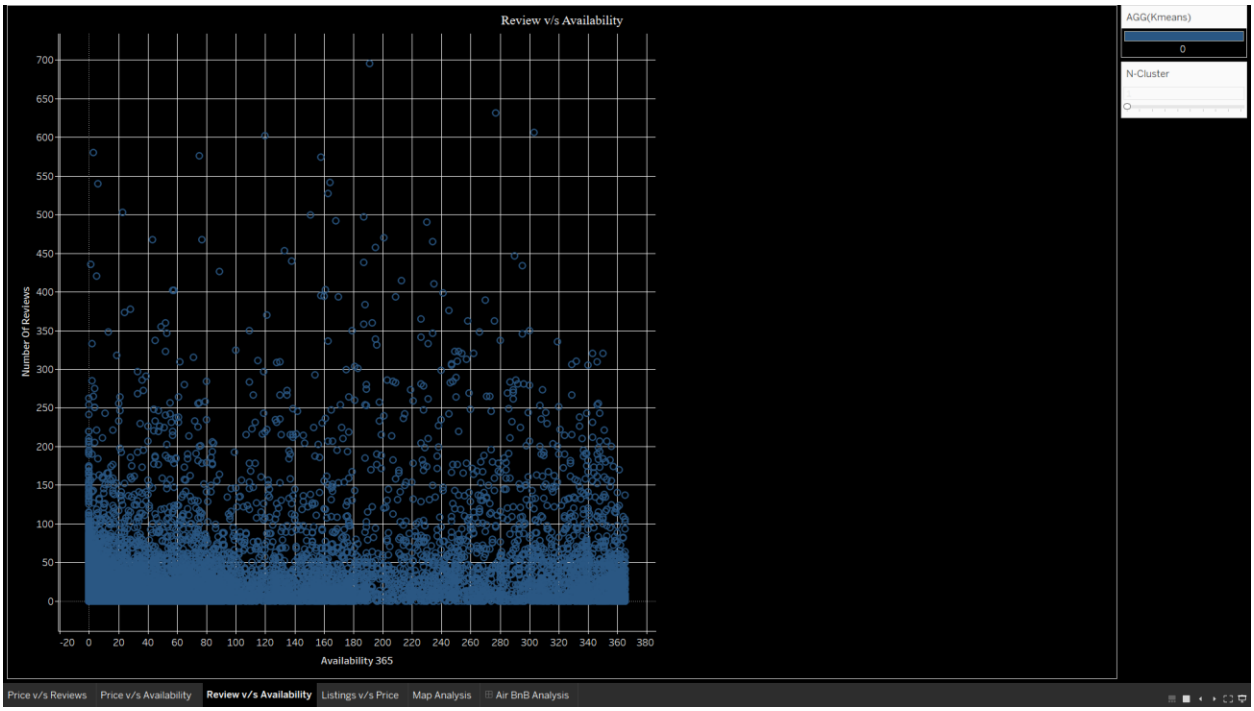

Visual Analytics Project
Title: K-Mean Clustering on Air BnB Dataset
Price v/s Reviews
Number Of Reviews
Price
Listings v/s Price
Calculated Host Listings Count
Price
Review v/s Availability
Number Of Reviews
Availability 365
Kmeans
0
2
N-Cluster
3
Price v/s Availability
Price
Availability 365
Map Analysis
© 2023 Mapbox © OpenStreetMap
Made By:
Mausmi Sinha A241
Aman Singh Bhogal A236
B.Tech IT 4th Year

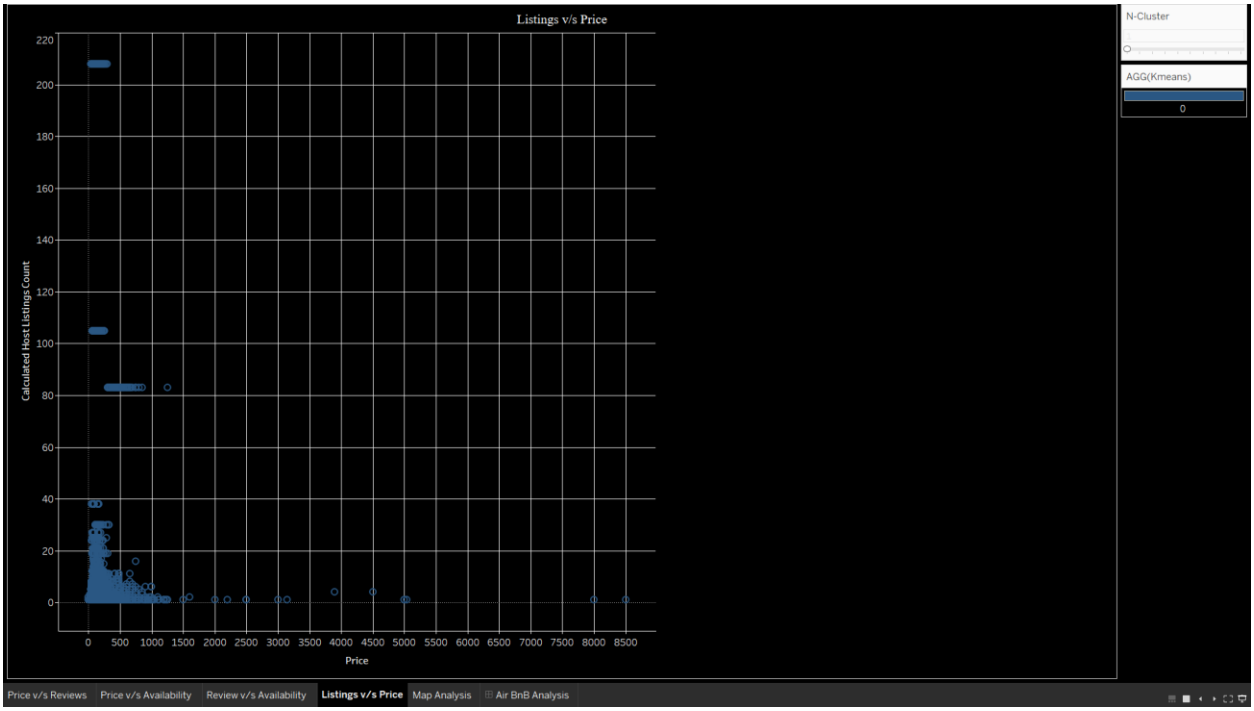

Visual Analytics Project
Title: K-Mean Clustering on Air BnB Dataset
Price v/s Reviews
Number Of Reviews
Price
Listings v/s Price
Calculated Host Listings Count
Price
Review v/s Availability
Number Of Reviews
Availability 365
Kmeans
0
N-Cluster
1
Price v/s Availability
Price
Availability 365
Map Analysis
© 2023 Mapbox © OpenStreetMap
Made By:
Mausmi Sinha A241
Aman Singh Bhogal A236
B.Tech IT 4th Year

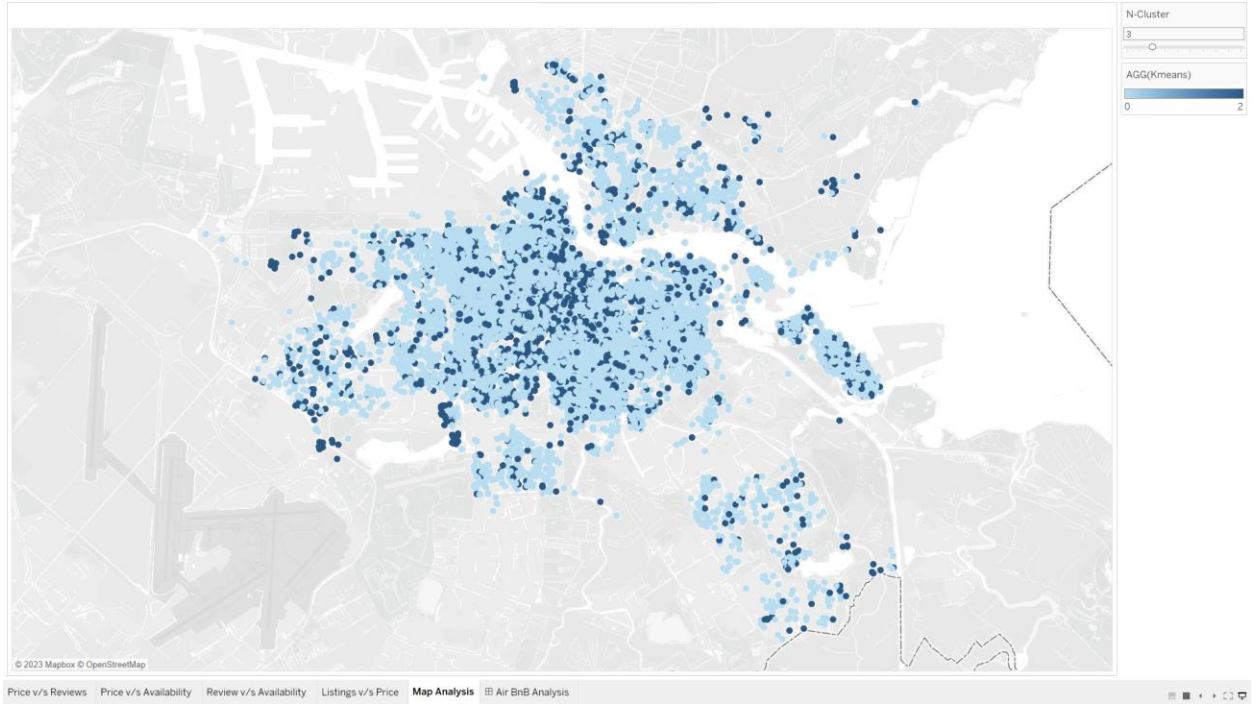
Individual Charts:






Review v/s Availability
Number Of Reviews
Availability 365
AGG(Kmeans)
0
N-Cluster
Price v/s Reviews
Price v/s Availability
Review v/s Availability
Listings v/s Price
Map Analysis
Air BnB Analysis

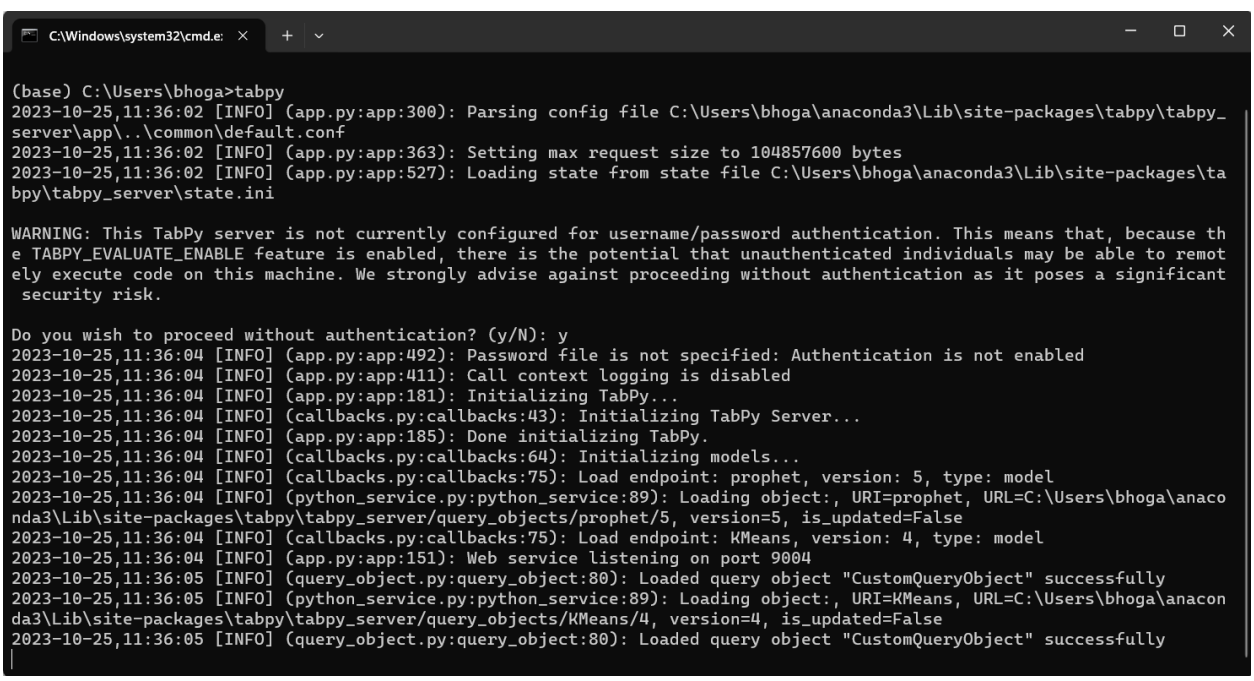

Listings v/s Price
Calculated Host Listings Count
Price
N-Cluster
AGG(Kmeans)
0
Price v/s Reviews
Price v/s Availability
Review v/s Availability
Listings v/s Price
Map Analysis
Air BnB Analysis



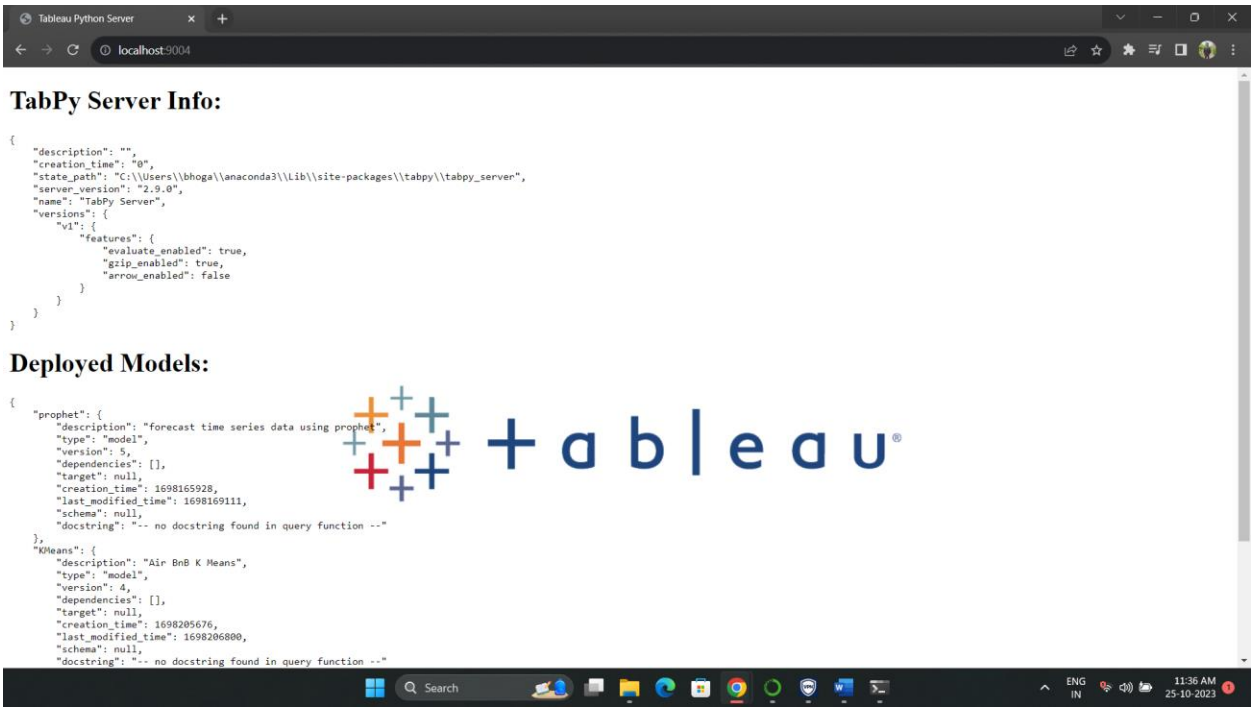
Method Used:

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

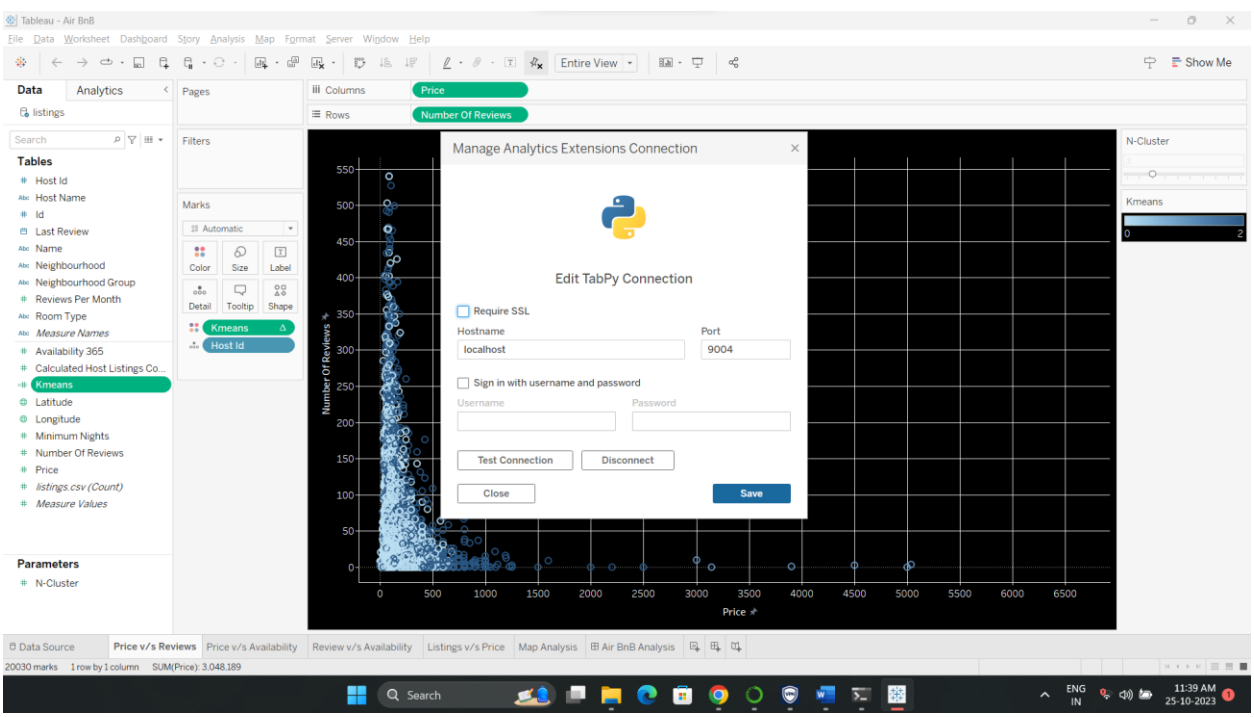
Starting TabPy server:

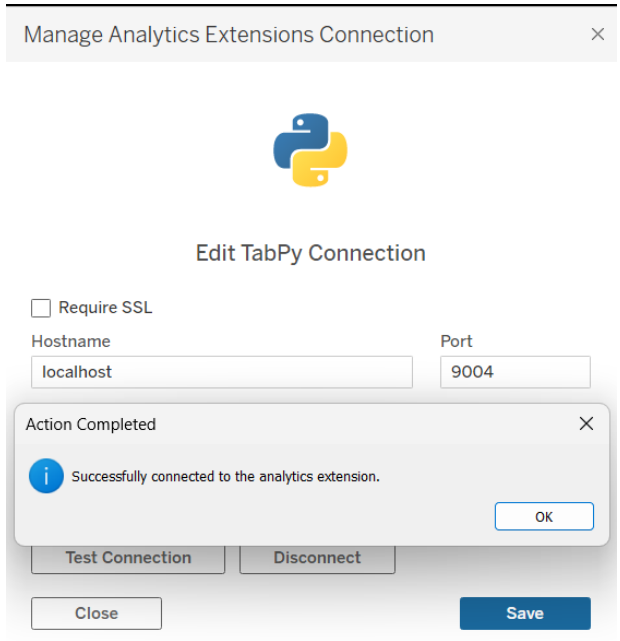


```
C:\Windows\system32\cmd.e:

(base) C:\Users\bhoga>tabpy
2023-10-25,11:36:02 [INFO] (app.py:app:300): Parsing config file C:\Users\bhoga\anaconda3\Lib\site-packages\tabpy\tabpy_
server\app\..\common\default.conf
2023-10-25,11:36:02 [INFO] (app.py:app:363): Setting max request size to 104857600 bytes
2023-10-25,11:36:02 [INFO] (app.py:app:527): Loading state from state file C:\Users\bhoga\anaconda3\Lib\site-packages\ta
bpy\tabpy_server\state.ini

WARNING: This TabPy server is not currently configured for username/password authentication. This means that, because th
e TABPY_EVALUATE_ENABLE feature is enabled, there is the potential that unauthenticated individuals may be able to remot
ely execute code on this machine. We strongly advise against proceeding without authentication as it poses a significant
 security risk.

Do you wish to proceed without authentication? (y/N): y
2023-10-25,11:36:04 [INFO] (app.py:app:492): Password file is not specified: Authentication is not enabled
2023-10-25,11:36:04 [INFO] (app.py:app:411): Call context logging is disabled
2023-10-25,11:36:04 [INFO] (app.py:app:181): Initializing TabPy...
2023-10-25,11:36:04 [INFO] (callbacks.py:callbacks:43): Initializing TabPy Server...
2023-10-25,11:36:04 [INFO] (app.py:app:185): Done initializing TabPy.
2023-10-25,11:36:04 [INFO] (callbacks.py:callbacks:64): Initializing models...
2023-10-25,11:36:04 [INFO] (callbacks.py:callbacks:75): Load endpoint: prophet, version: 5, type: model
2023-10-25,11:36:04 [INFO] (python_service.py:python_service:89): Loading object:, URI=prophet, URL=C:\Users\bhoga\anaco
nda3\Lib\site-packages\tabpy\tabpy_server/query_objects/prophet/5, version=5, is_updated=False
2023-10-25,11:36:04 [INFO] (callbacks.py:callbacks:75): Load endpoint: KMeans, version: 4, type: model
2023-10-25,11:36:04 [INFO] (app.py:app:151): Web service listening on port 9004
2023-10-25,11:36:05 [INFO] (query_object.py:query_object:80): Loaded query object "CustomQueryObject" successfully
2023-10-25,11:36:05 [INFO] (python_service.py:python_service:89): Loading object:, URI=KMeans, URL=C:\Users\bhoga\anacon
da3\Lib\site-packages\tabpy\tabpy_server/query_objects/KMeans/4, version=4, is_updated=False
2023-10-25,11:36:05 [INFO] (query_object.py:query_object:80): Loaded query object "CustomQueryObject" successfully
```



Connecting Tableau to our TabPy Server:





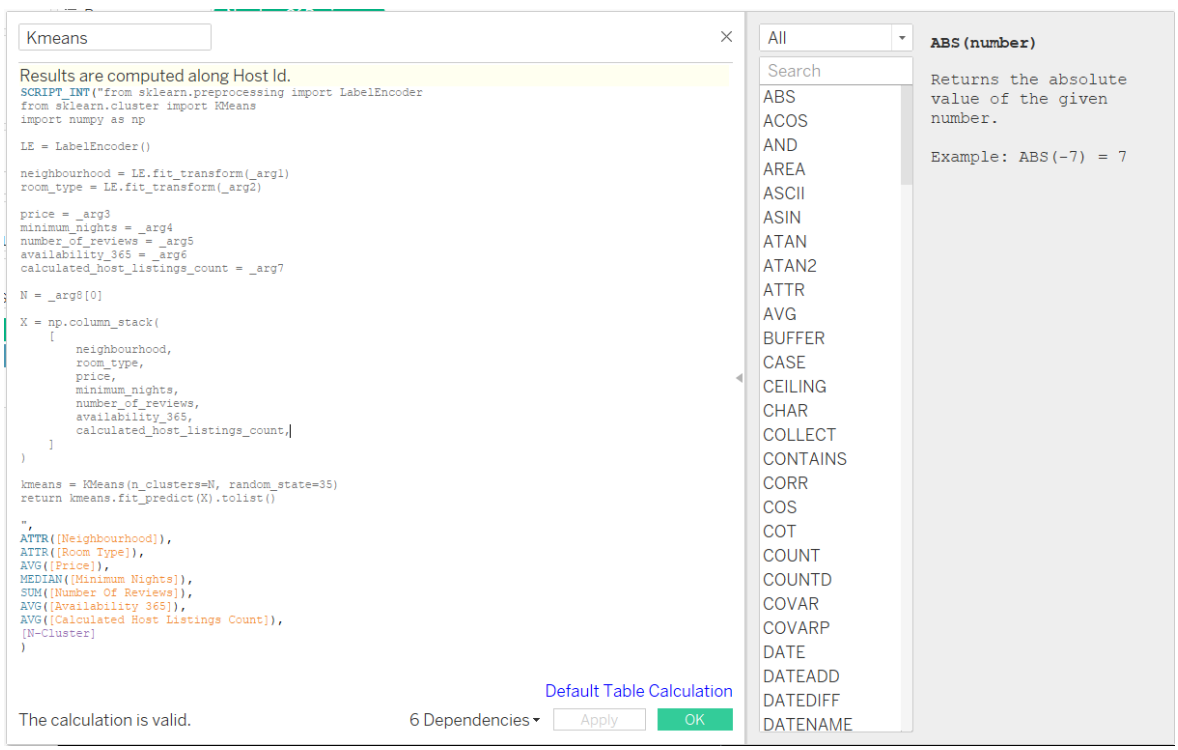
Writing Script in Tableau using SCRIPT_INT:

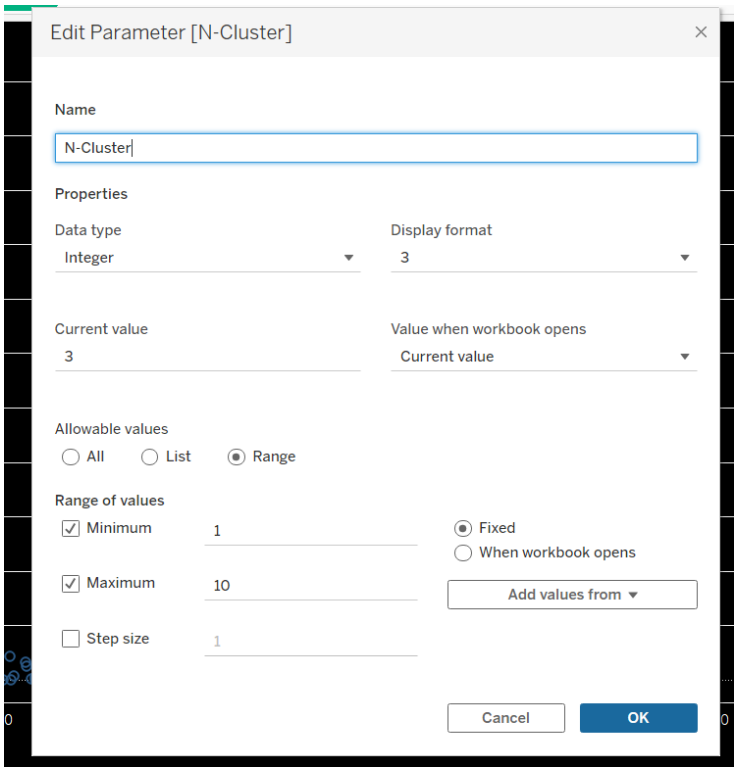
We created a calculated field using SCRIPT_INT() that takes data attributes and N-Cluster Parameter as an input and passes it to our python code for calculating the clusters. We used K Means Clustering algorithm provided by sklearn library.

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]
)
```





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.