ANALYSING NEIGHBOURHOODS OF MANCHESTER FOR STARTING A NEW RESTAURANT

Saurabh Mahalle

Applied Data Science Capstone

Introduction

- Mumbai: One of the biggest cities in India
- **India densely populated city on India's west coast.**
- The population comprises of people of various ethnicities from all over the world

Business Problem

- Start a Hotel Business

- Already have a so many Hotel business in various area of a Mumbai

- Uniqueness in business, that will generate more revenue

Data

- Data of a various hotel business, which is already exist, in various region and there speciality.

➢Through Data scraping from various website(Wikipedia, etc) and survey

- Geographical coordinates of the neighbourhoods

➢Using GeoPy library.

- Venue data from FourSquare

➢Using FourSquare API

Methodology

- Feature Extraction

➢One Hot Encoding

```python
man_1hot = pd.get_dummies(explore_man[['Venue Category']], prefix="", prefix_sep="")


# Add neighbourhood column back to dataframe
man_1hot['Neighbourhood'] = explore_man['Neighbourhood']


# Move neighbourhood column to the first column
fixed_columns = [man_1hot.columns[-1]] + man_1hot.columns[:-1].values.tolist()
man_1hot = man_1hot[fixed_columns]

man_1hot.head()
```

Methodology

- Unsupervised Learning

➢ K-Means Clustering

```
max_range = 15 #Max range 15 (number of clusters)

from sklearn.metrics import silhouette_samples, silhouette_score


indices = []
scores = []

for man_clusters in range(2, max_range) :

    # Run k-means clustering
    man_gc = man_grouped_clustering
    kmeans = KMeans(n_clusters = man_clusters, init = 'k-means++', random_state = 0).fit_predict(man_gc)

    # Gets the score for the clustering operation performed
    score = silhouette_score(man_gc, kmeans)

    # Appending the index and score to the respective lists
    indices.append(man_clusters)
    scores.append(score)
```
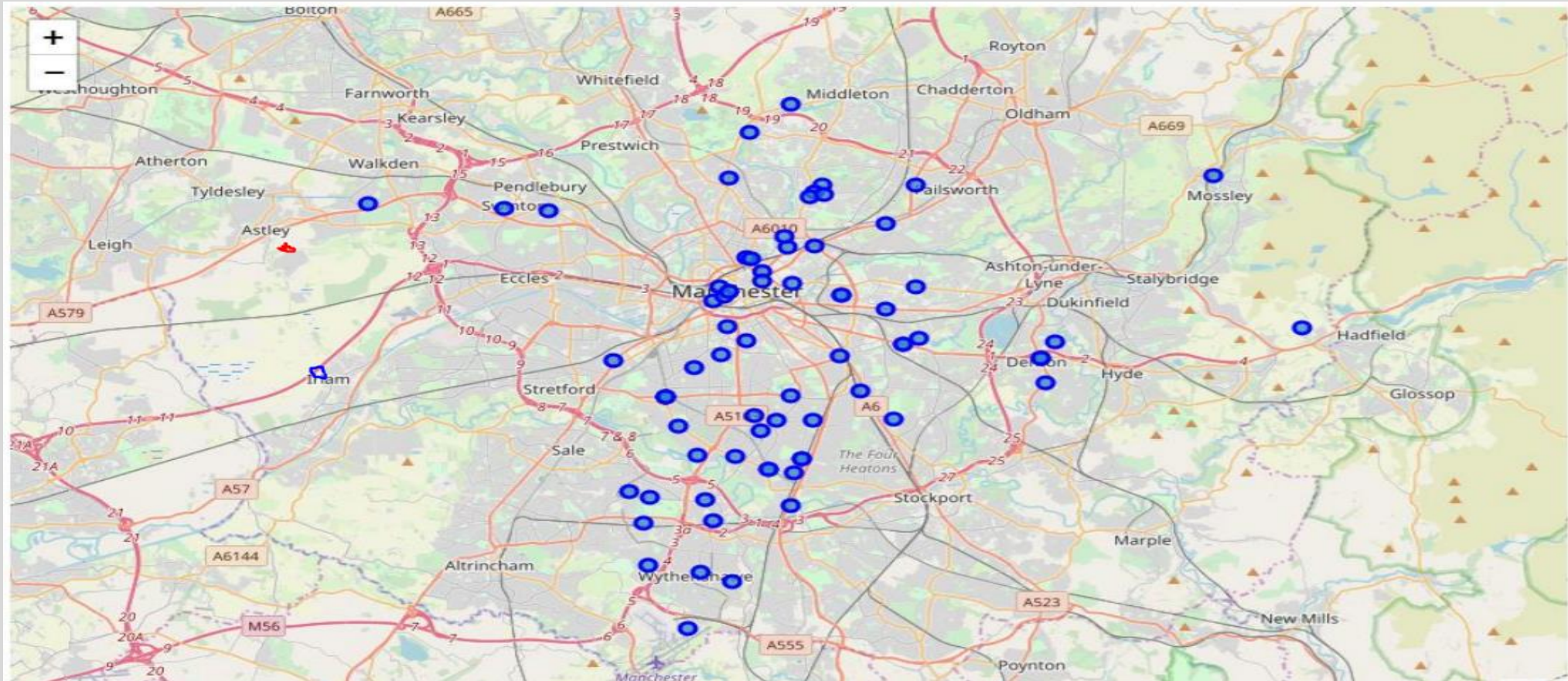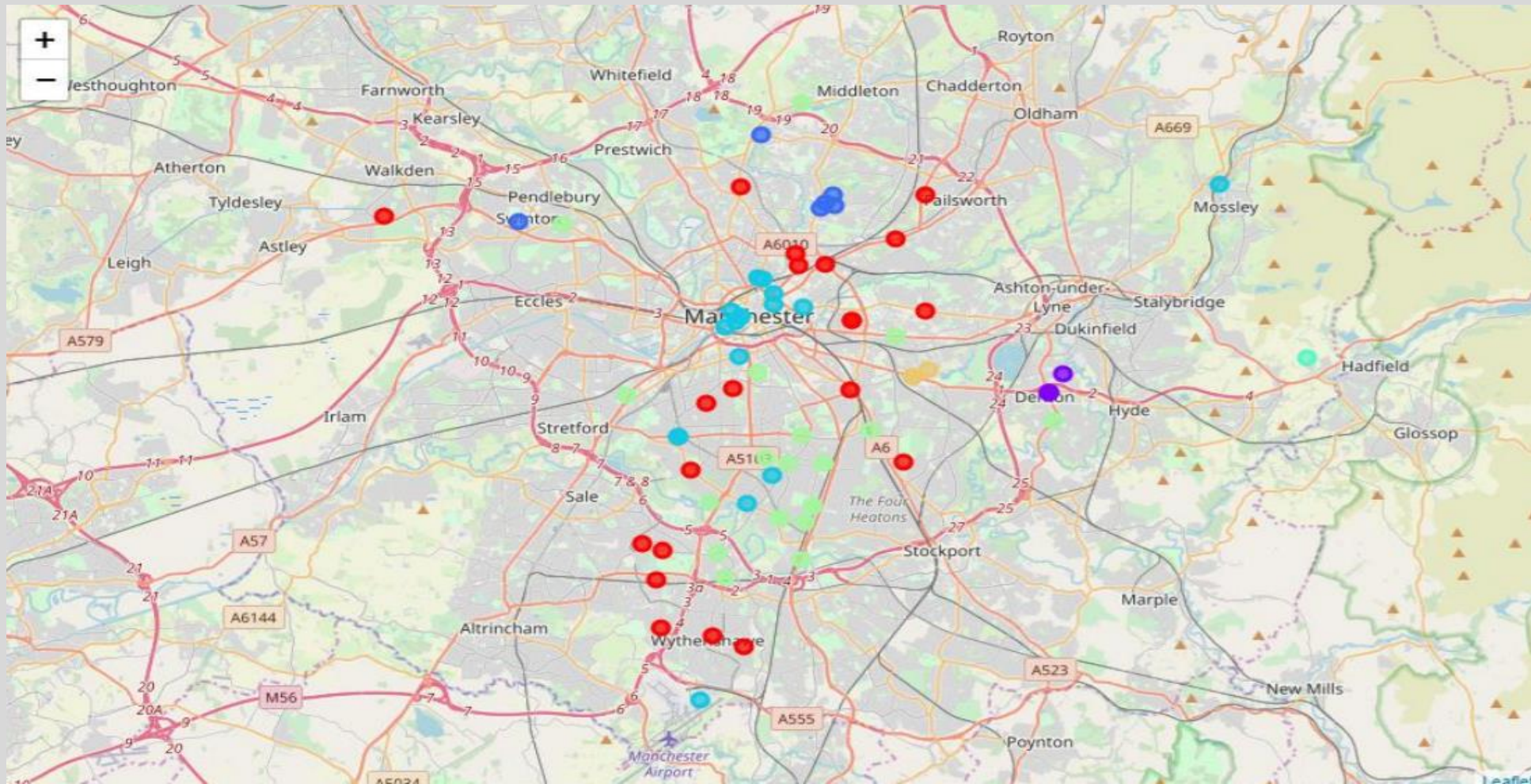
- Plotting
  - Folium

# Results

- Visualization of clusters

Discussion

- Most suitable regions for starting the hotel business are present in the cluster number 3.

- Our K-Means model worked perfectly and successfully clustered similar regions together.

- After studying all four clusters, it is recommended to the client that regions such as india gate, jaitala and hingna that fall in cluster 4 look like good locations for starting their hotel business.

- The client can go ahead and make a decision depending on other factors like availability and legal requirements that are out of scope of this project.

Conclusion

- Data analysis and machine learning techniques used in this project can be very helpful in determining solutions of certain business problems.

- Python's inbuilt libraries such as GeoPy, Folium and BeautifulSoup make it very easy and effective to analyse a geographical location.

- In this project we studied the regions of mumbai city and came up with a recommendation of regions where our client can start their hotel business.
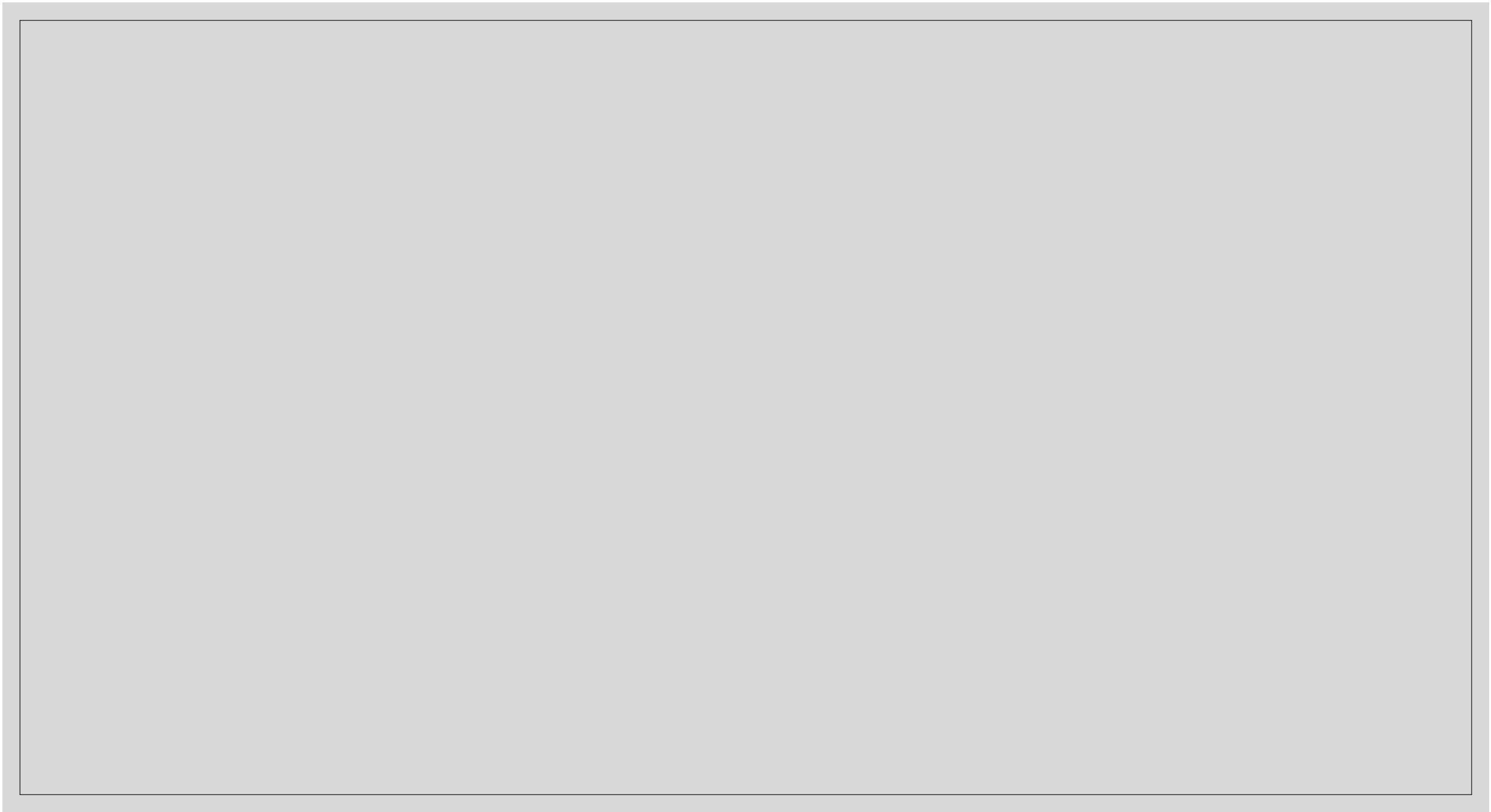
Title Lorem Ipsum

LOREM IPSUM DOLOR SIT AMET, CONSECTETUER ADIPISCING ELIT.

NUNC VIVERRA IMPERDIET ENIM. FUSCE EST. VIVAMUS A TELLUS.

PELLENTESQUE HABITANT MORBI TRISTIQUE SENECTUS ET NETUS.