

A Presentation on:

Immigrant Segmentation for Appropriate Accommodation using EDA and Clustering Analysis

Presented By: **Group 4**

Name of group members:

Devika Dhumal

Sunny Khade

Sanskriti Patole

Omkar Takale

Guided By: **Prof. P. P. Dandavate**



AGENDA

- Introduction
- Overview
- Motivation
- Objective
- Literature Survey
- Problem Statement
- Requirement and Feasibility
- Future Scope
- Approach
- UML
- Current Progress
- References
- Conclusion



INTRODUCTION

Imagine a scenario where a person has newly moved into a new location. They already have certain preferences, certain tastes. It would save both the student and the food providers a lot of hassle if the student lived close to their preferred outlets. Convenience means better sales, and saved time for the customer.

We will be performing Exploratory Data analysis on Geolocational Data. Further we shall use Clustering Analysis (eg. K-Means Clustering etc) for immigrant segmentation, i.e., dividing the immigrants into groups according to their individual Point-of-Interests (POI).

We are using the most optimized clustering algorithm for segmenting the areas (in the city we have chosen for analysis) according to features like restaurants and other amenities in proximity to the desired location.

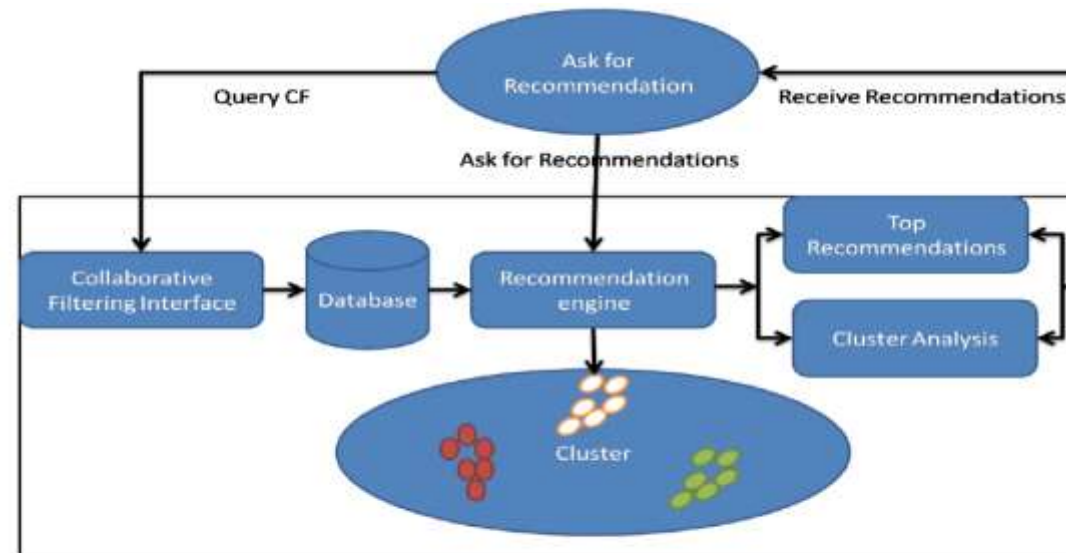
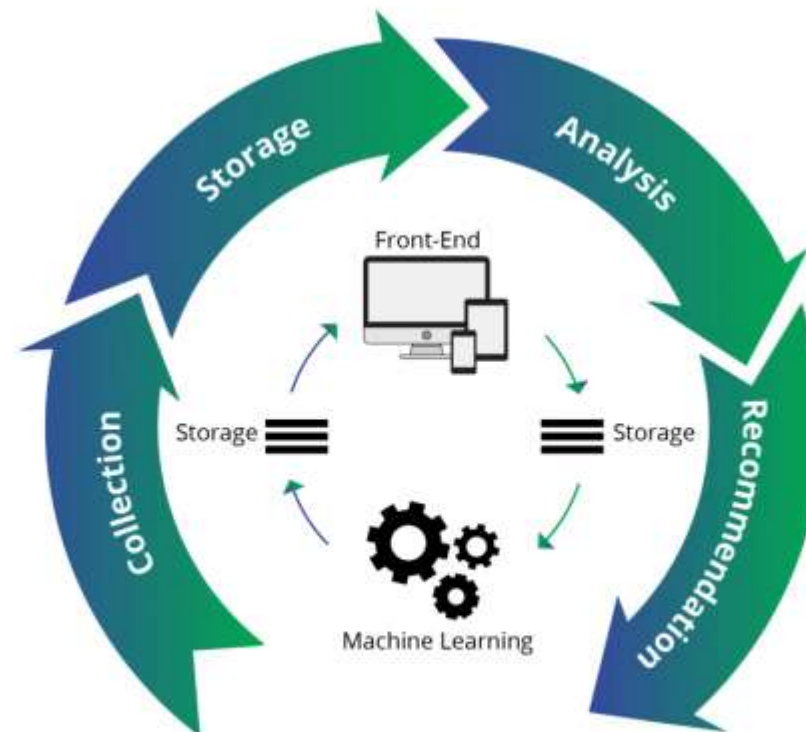


Fig 1: Architecture diagram for collaborative filtering interface with database cluster analysis.



OVERVIEW

- Every human is different and has his own preferences and likings based on which they take their daily life decisions which might include what food to eat, what place to visit, what clothes to wear, etc.
- Our project mainly focuses on Immigrants (e.g., students who are new to the city, the people who have moved to a different city or country) have certain choices in terms of the amenities they want, their food choices, their budget and so on.
- To help immigrants we present them a personalized system. And to build such a system we are performing Exploratory Data Analysis and Cluster Analysis on 'geo-locational' and 'Immigrant Preference' datasets to help them find their suitable accommodation.
- This project will help immigrants choose an area which will be most convenient for them based on their personal opinions and choices.





Motivation

MOTIVATION

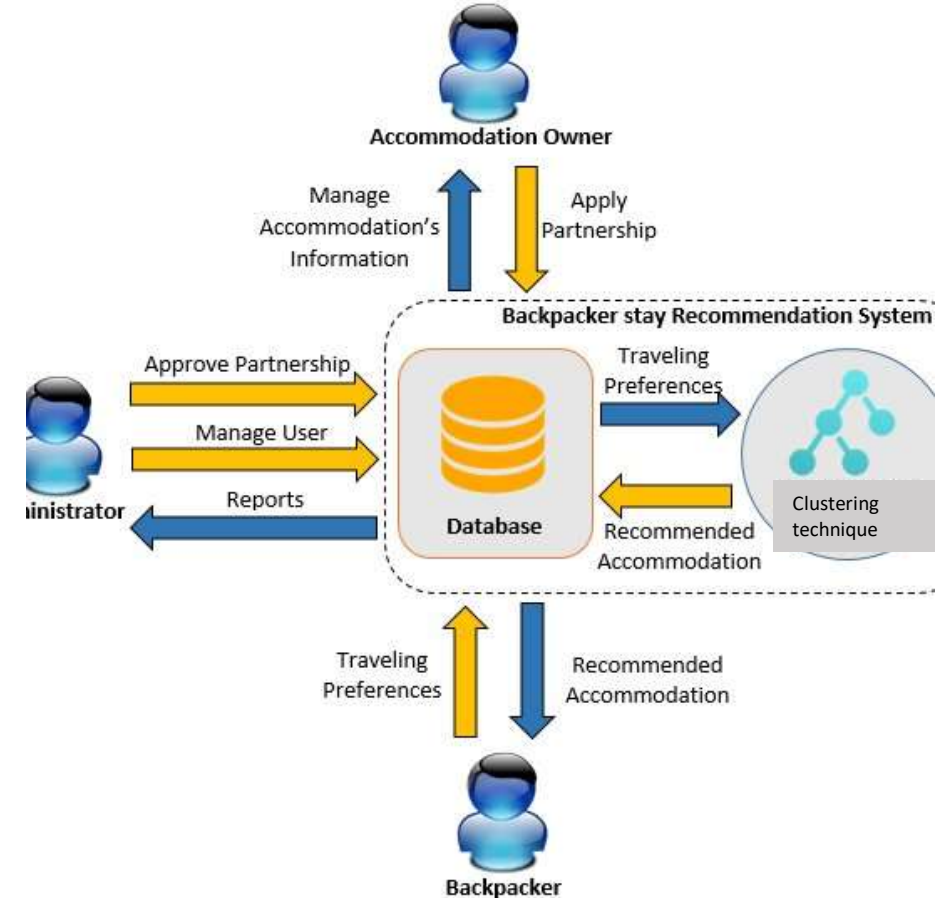
Helping immigrants find the most convenient and suitable place to live.

Segmenting the city/region of choice according to the amenities, cost of living and other factors.

Beneficial not only for the immigrants but also for the food providers and other service providers in the area.

Why Accommodation Recommendation systems?

- There has been a rapid growth of immigrants all over the country. To find a convenient place to live is the first step and the most crucial step for any person new to a city.
- Providing a hassle free method to find such a place is the objective of this project.
- Various factors (immigrant preferences) can be taken into consideration to figure out the suitable location
 - To identify the most relevant Interests for each user
 - Showcase personalized content to each user
 - Suggest top offers and discounts to the right user
- It increases relevance for search through increased consumption





1.Literature Survey

Paper	Author	Methodology	Drawback	Overcome
Improving Accuracy of Recommender System by Item Clustering.	KhanhQuan Truong,Fuyuki Ishikawa, Shinichi Honiden	In this paper, Recommender System (RS) predicts user's preference , and then recommends highly-predicted items to user.	The number of attributes is often very large and so is the diversity amongst them, users who have similar preference in one category may have totally different judgement on attributes of another kind.	We propose a method to cluster attributes, so that inside a cluster, similarity between users does not change significantly from one attribute to another .



2.Literature Survey

Paper	Author	Methodology	Drawback	Overcome
Recommend er Systems: Algorithms, Evaluation and Limitations.	Mubaraka Sani Ibrahim and Charles Isah Saidu	This paper also discusses the state of art machine learning based recommendati o n models including Clustering models and Bayesian Classi fiers. Further, we discuss the widespread ap plication of recommender systems.	It may provide inaccurate results if data entered incorrectly. Since the feature representatio n of the attributess are handengineere d to some extent, many technique requires a lot of domain knowledge.	Good Knowledge of model can capture the specific interests, and can recommend niche items that very few other users are interested in.



3.Literature Survey

Paper	Author	Methodology	Drawback	Overcome
Point-of-interest lists and their potential in Recommendation systems	Stamatelatos, G., Drosatos, G., Gyftopoulos, S. et al.	This paper proposes a new idea on personalized interest based recommender system which is based on lists of end user point of interests.	The model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.	Recommend new experiences and attributes to attract the end users interest .



Problem Statement

“Targeting Individual Point-Of-Interests (POI) to recommend appropriate accommodations with the help of exploratory data analytics using clustering algorithm”



Requirements and Feasibility

1. Hardware Interfaces

- Graphic Card
- 512 SSD
- 8 GB RAM

2. System Specifications

Language: Python3

- IDE: Jupyter Notebook
- Dataset: Kaggle website
- Libraries used:
 1. Data Fetch: Pandas
 2. Data processing: Numpy
 3. Visualization: Matplotlib/ Tableau Desktop s/w
 4. Clustering Analysis: ScikitLearn
 5. Map Display: Folium



Future Scope

Clustering Analysis: Market Research, Pattern Recognition, Image Processing.

Exploratory Data Analysis: is used to perform initial investigation on data to derive predictions and information.

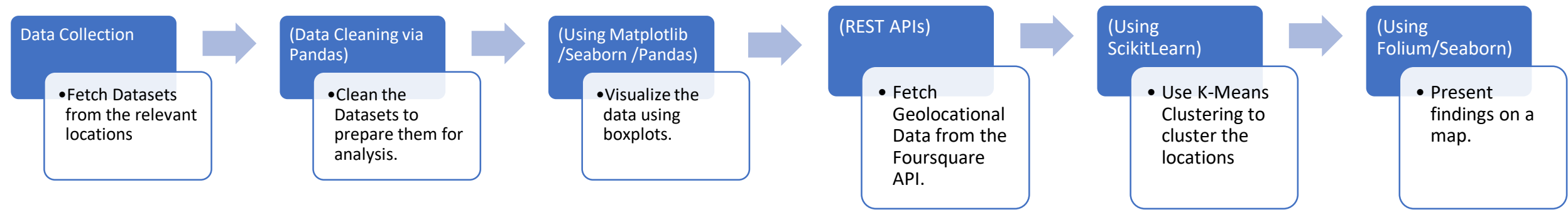
In General: Governments can use the same data work flow to predict appropriate accommodations for mass refugees in their countries.

Analytics will play an important role in data security. Analytics are already transforming differential privacy, intrusion detection, digital watermarking and malware countermeasures.

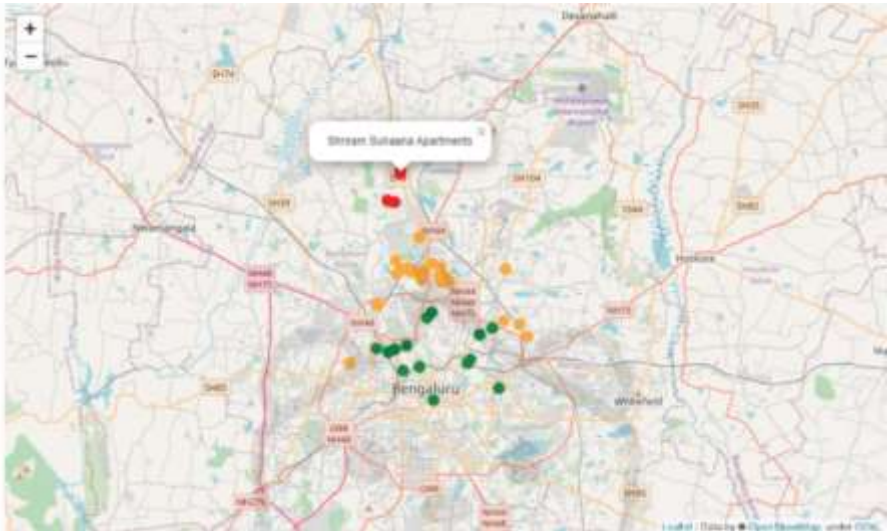
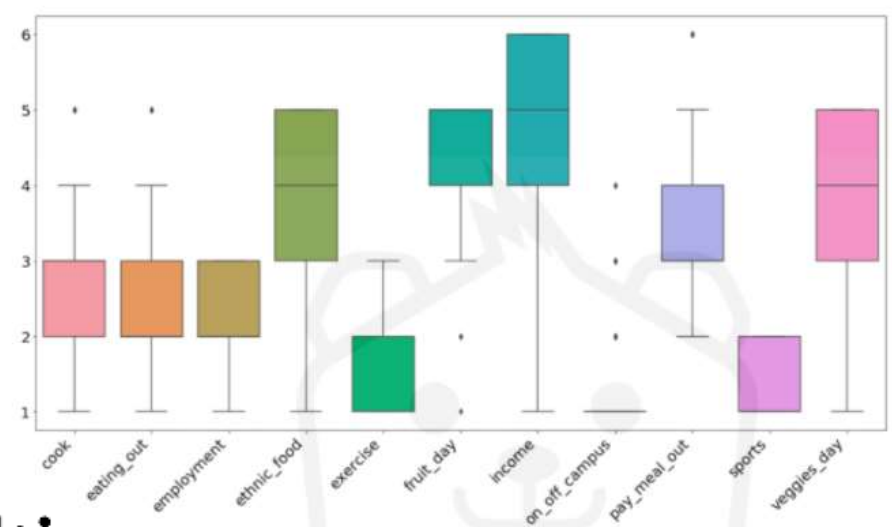


APPROACH

HIGH LEVEL APPROACH



EXPECTED VISUALIZATION/OUTCOME

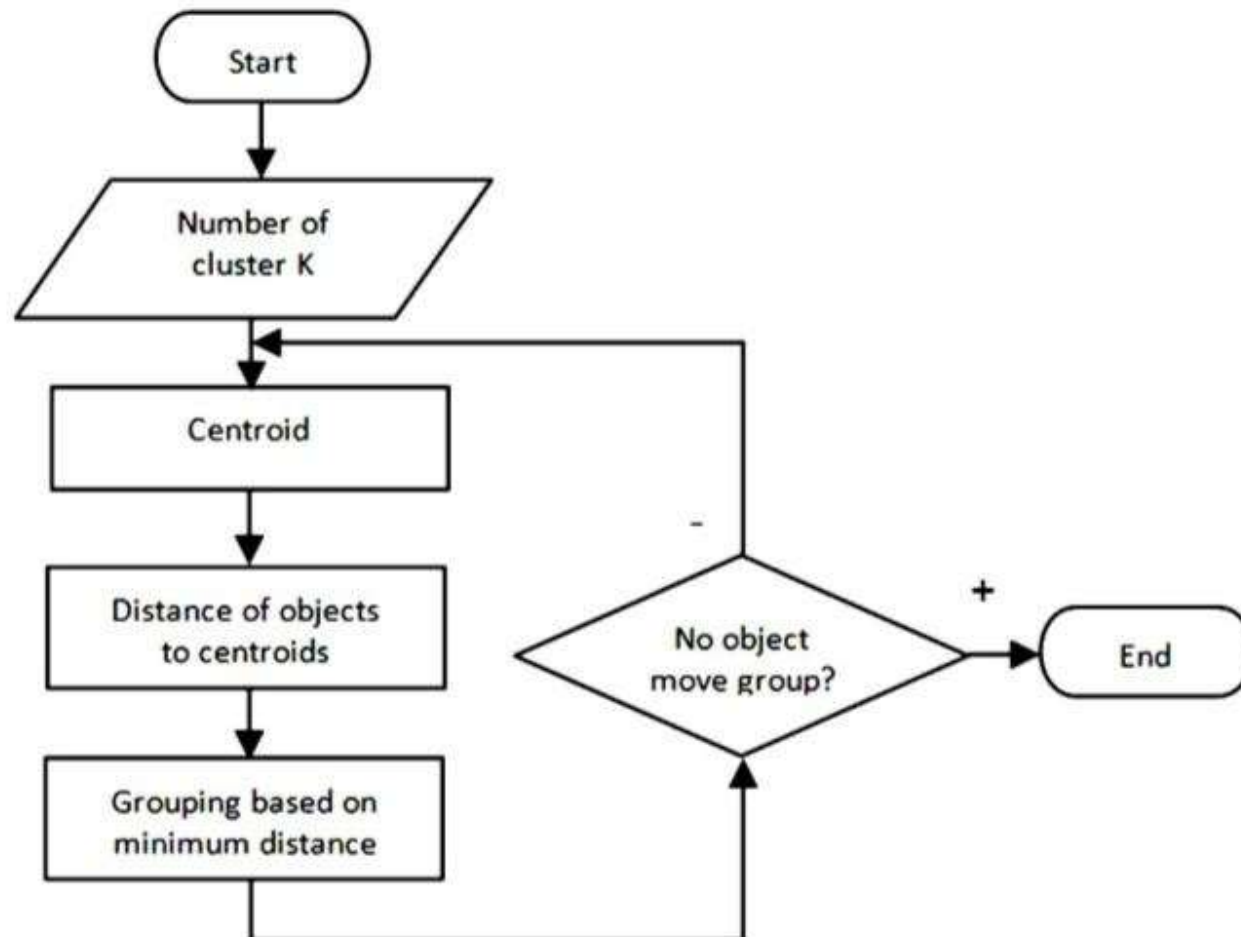


BOXPLOT for Visualization



Cluster Representation on Locational outcome

APPROACH (ALGORITHM)

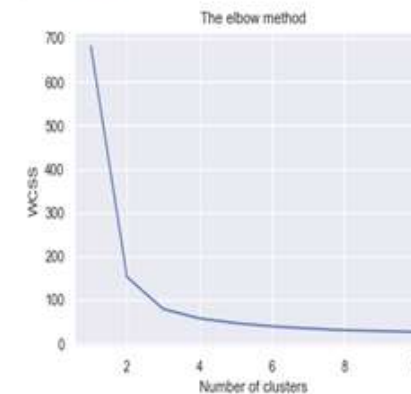


```
In [25]: #Finding the optimum number of clusters for k-means classification

wcss = []

for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)

#Plotting the results onto a line graph, allowing us to observe 'The elbow'
plt.plot(range(1, 11), wcss)
plt.title('The elbow method for optimal k')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS') #within cluster sum of squares
plt.show()
```



We can clearly see why it is called 'The elbow method' from the above graph, the optimum clusters is where the elbow occurs (K=3)

This is when the within cluster sum of squares (WCSS) doesn't decrease significantly with every iteration.

Now that we have the optimum amount of clusters, we can move on to applying K-means clustering to the Iris dataset with K=3.



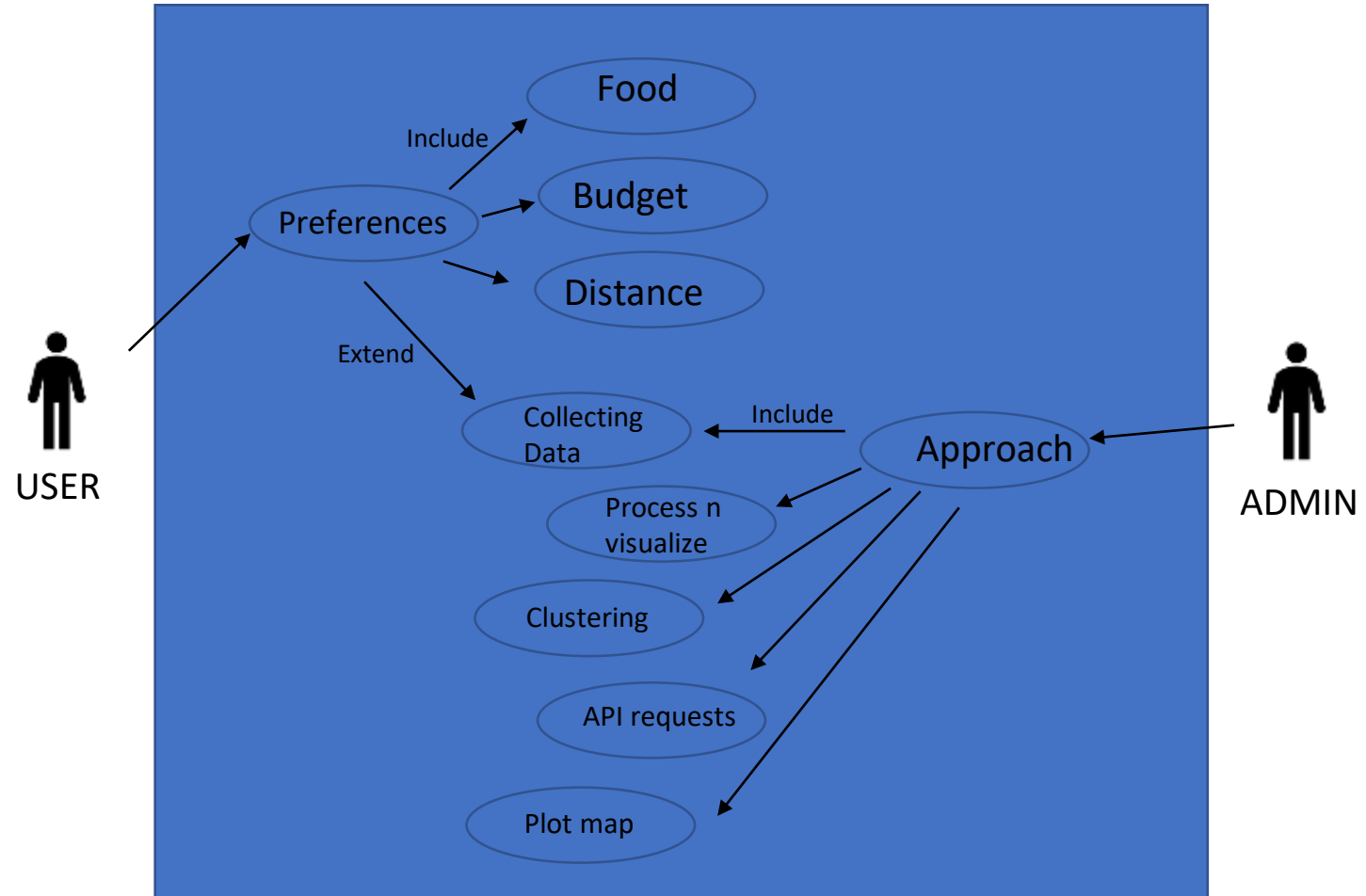
UML- Use Case Diagram

Actor: USER

- Distance: from desired location
- Food: In Budget with the right preference
- Budget: Total Capital allocated
- Recommendation: Accordingly

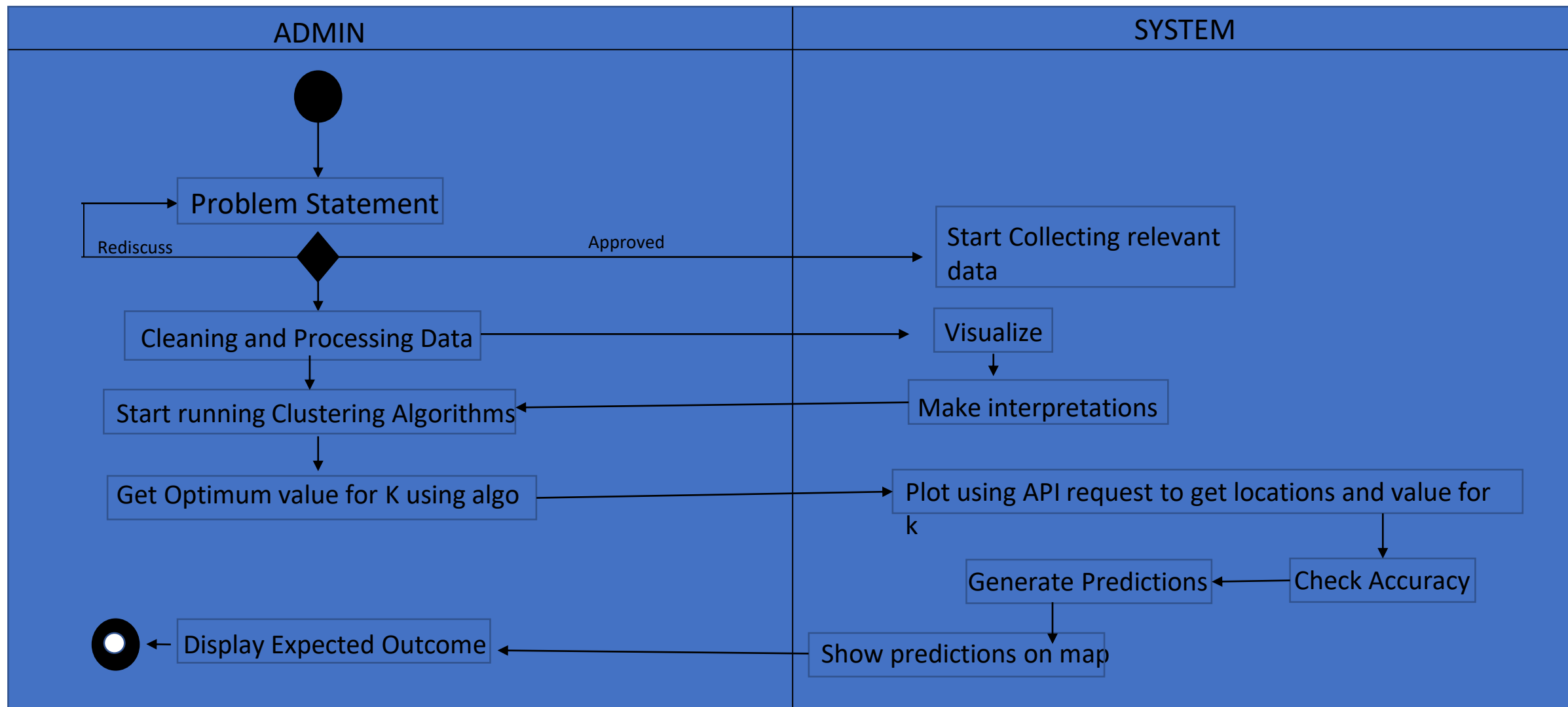
Actor: ADMIN

- Gathering: Collecting and processing relevant data.
- Analysis and cluster: analyze and run algorithms.
- Outcome: Show Results



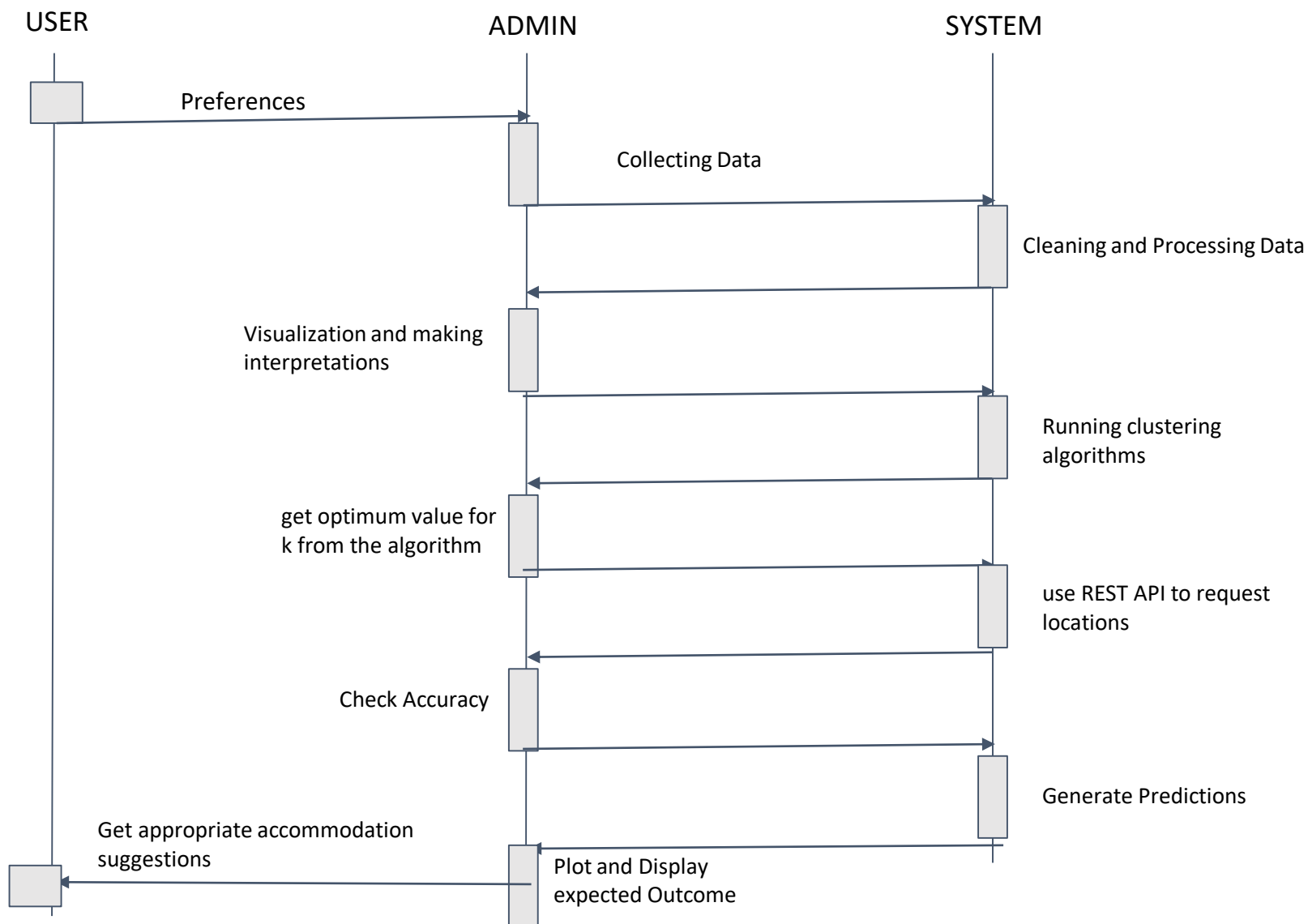


UML – Activity Diagram



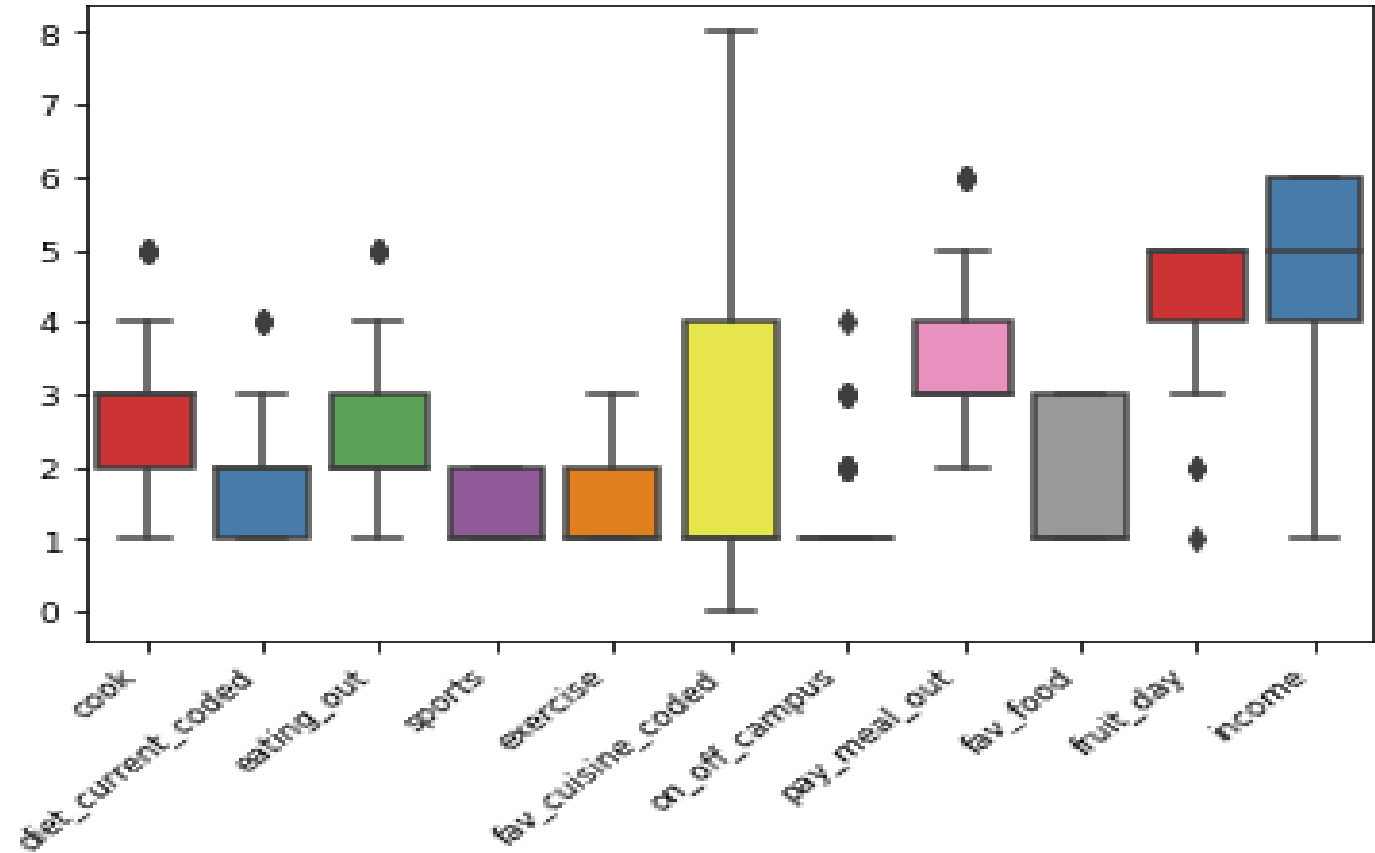


UML – SEQUENCE Diagram



Current Progress

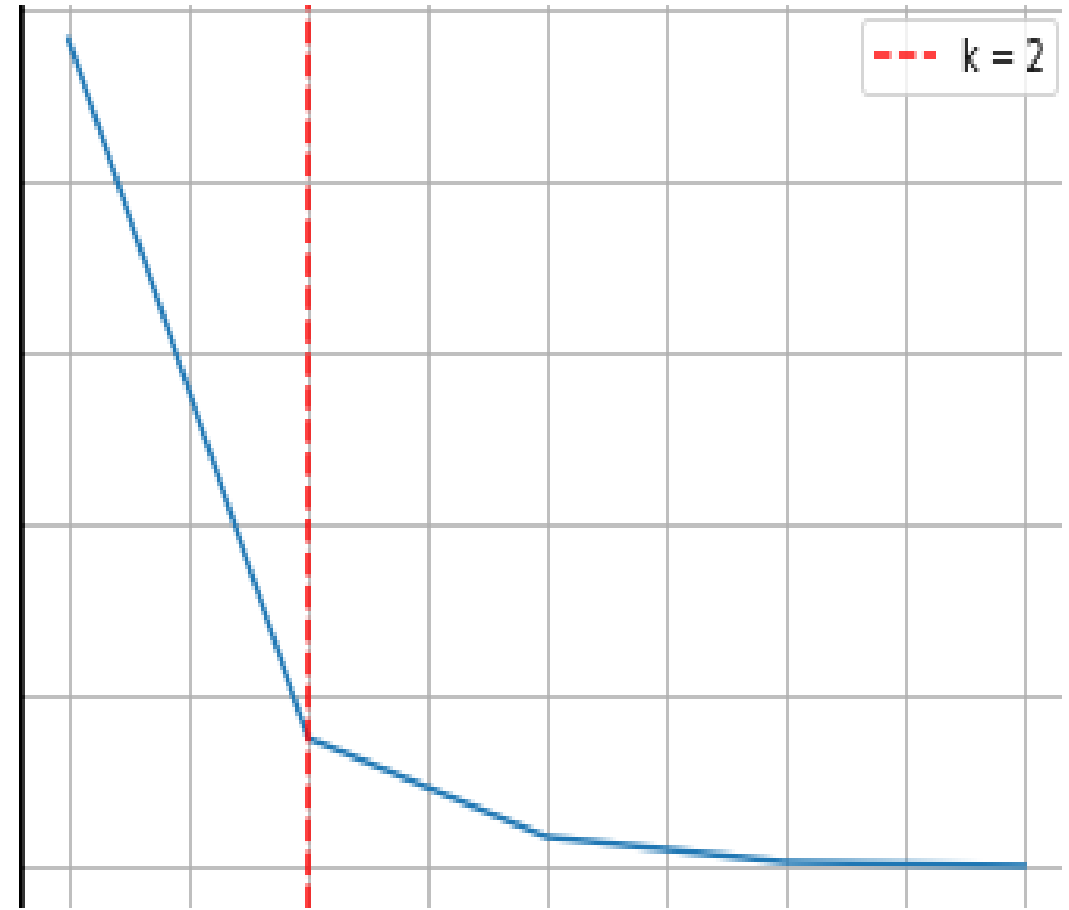
Box plot showing the spread of data in the food preferences dataset used in our algorithm.



Current Progress

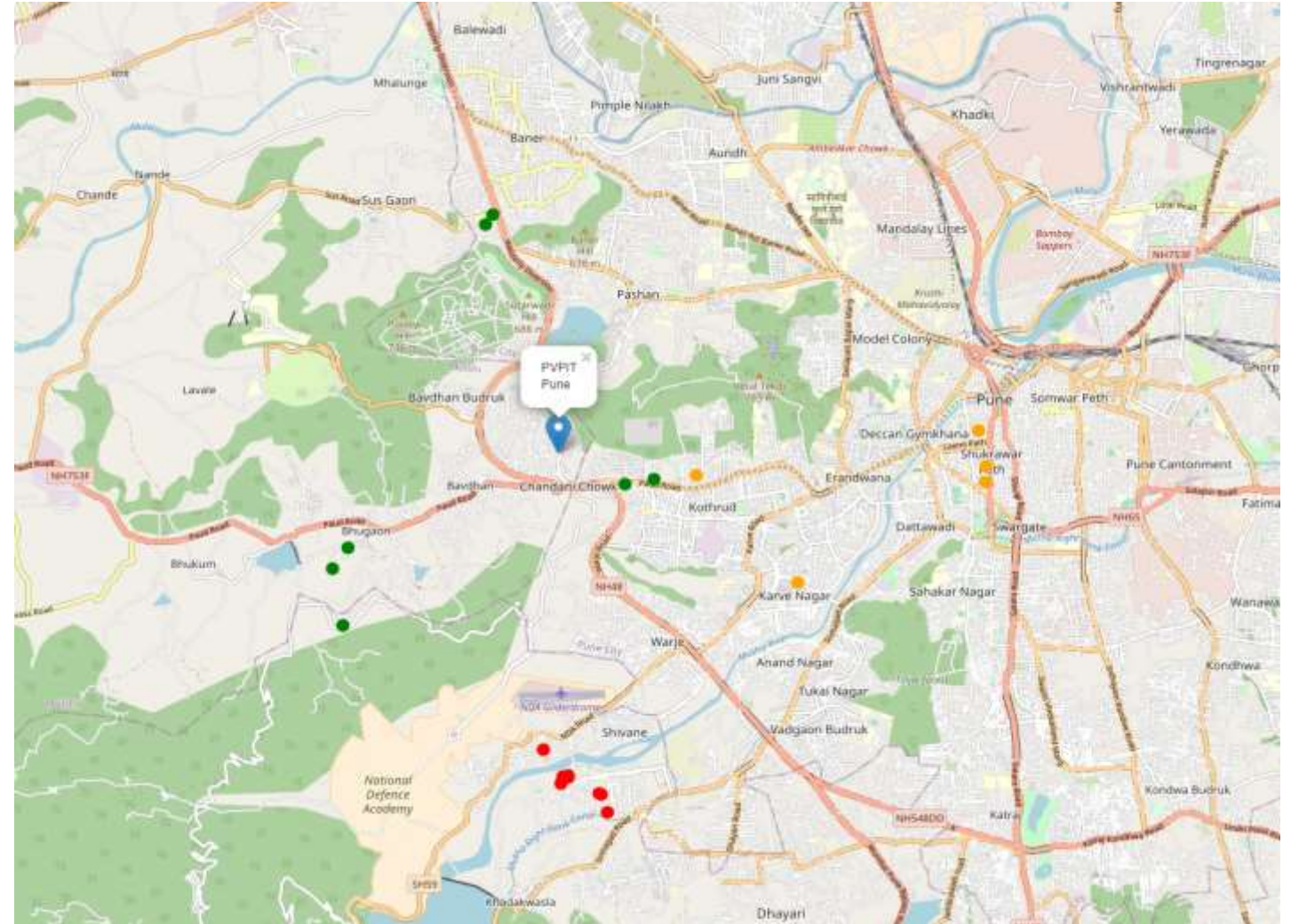
Graph of the elbow method that is used to determine the ideal number of clusters for clustering (on the food preferences dataset used in our algorithm).

The Clustering algorithm used is the K-Means Clustering Algorithm.



Current Progress

Ideal locations for immigrants plotted on the map. The immigrants were divided into three groups/clusters during cluster analysis and each colour represents a cluster. The location for analysis is PVPIT Bavdhan, Pune and we obtained the geolocational data for the above mentioned location with the help of APIs.





References

1. K. Gao et al., "Exploiting Location-Based Context for POI Recommendation When Traveling to a New Region, " in IEEE Access, vol. 8, pp. 52404-52412, 2020, doi: 10.1109/ACCESS.2020.2980982.
2. Stamatelatos, G., Drosatos, G., Gyftopoulos, S. et al. Point-of-interest lists and their potential in recommendation systems. Inf Technol Tourism 23, 209-239 (2021). <https://doi.org/10.1007/s40558-021-00195-5>
3. Dietz, L.W., Sen, A., Roy, R. et al. Mining trips from location-based social networks for clustering travelers and destinations. Inf Technol Tourism 22, 131-166 (2020).
4. K. Gao et al., "Exploiting Location-Based Context for POI Recommendation When Traveling to a New Region, " in IEEE Access, vol. 8, pp. 52404-52412, 2020, doi: 10.1109/ACCESS.2020.2980982.
5. Singh, Ashwani & Soundarabai, Paulsingh. (2017). Collaborative filtering in movie Recommendation System based on Rating and Genre. IJARCCCE. 6. 465-467. 10.17148/IJARCCCE.2017.63107.
6. Xu, Chonghuan & Ju, Chunhua & Qiang, Xiaodan. (2015). Efficient Collaborative Filtering Using Particle Swarm Optimization and K-Harmonic Means Algorithm. Journal of Computational and Theoretical Nanoscience. 12. 6334-6342. 10.1166/jctn.2015.4675.
7. Qing Li and Byeong Man Kim, "Clustering approach for hybrid recommender system," Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), 2003, pp. 33-38, doi: 10.1109/WI.2003.1241167.
8. Sahu, Satya & Nautiyal, Anand & Prasad, Mahendra. (2017). Machine Learning Algorithms for Recommender System - a comparative analysis. International Journal of Computer Applications Technology and Research. 6. 97-100. 10.7753/IJCATR0602.1005.
9. Basu, C., Hirsh, H., and Cohen, W. (1998). Recommendation as Classification: Using Social and Content-based Information in Recommendation. In Recommender System Workshop '98. pp. 11-15.
10. Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, Guangquan Zhang, Recommender system application developments: A survey, Decision Support Systems, Volume 74, 2015, Pages 12-32, ISSN 0167-9236



Conclusion

Recommender structures are proving to be a useful device for addressing a part of the records overload phenomenon from the internet. Its evolution has followed the evolution of the internet. Latest research shows the use of Clustering Analysis in developing of more accurate recommender system. These types of methods are commonly used in e-commerce business.



Thank You