Helping business owners to find potential competitors

Suraksha S. S.
surakshass.is18@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Darunya B. C.
darunyabc.is17@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Priya D.
priyad@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Anisha B. S.
anisha@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Abstract–Opening a business in any place is not an easy task as it needs lots of things to be considered. One should consider its competitors before opening a business. This project is a one attempt to help those business people to find their competitors. In this project a taco palace is planned to open in the city of Monterrey, Mexico. Here the attempt is to find a neighborhood with not too much competitors with enough customers. To find the neighborhood, considering the latitude and longitude values of the city of Monterrey, Mexico and can be able to locate the best place to open the restaurant. In this project Foursquare API is used to explore the neighborhoods and get the most common restaurants near that place and using this function clusters can be grouped. For clustering K-means clustering algorithm is used. Clusters is used to know the similar business in those areas. Clusters group the similar business and list out all its names. To visualize the neighborhoods Folium library is used.

Keywords: *K-means Clustering, Folium library, Foursquare API.*

I. INTRODUCTION

Finding the potential competitors is the basic need for any business establishment. Competitors are those who are in the same business, targeting same customers, serving the same products. In order to be in the business, it is important to know all the details about their ambience, setup, pricing, offers etc. lack of knowledge about the competitors might keep business out of place and order. Establishing a new business is always a challenge. To tackle the challenge, a thorough knowledge about the existing business competitors is important. Without which it is not possible to be in par with or to excel in the same business. Knowing what products and services your competitors offer might help you boost your chances of developing new business with the same clients.

Customers are the main target of any business generally Customers are used to or accustomed to the already existing businesses providing the products and services. Either customers will be happy or will have complains. in order to attract the same customers, it is necessary to know the experiences and grievances of the customers which will help to offer better products and services at a competitive price. So finding the business competitors is a must in establishing and running any business without which business will suffer in many ways. Hence a detailed knowledge about the potential competitors helps to prosper in the business.

Cluster analysis is a technique for grouping data elements based on their commonalities. This method is widely used in business analytics and can assist you in achieving your objectives. The k-means algorithm can also be used to enhance data point resemblance across clusters while reducing familiarity between clusters. As before said, it is an unsupervised algorithm that would not require class labels or a training dataset. This approach is useful when dealing with categorical variables.(e.g. grouping based on category, subcategory and brand). For example, you might utilise this data to classify products based on sales to help your customers.

one must choose the number of groups you want to work with when using the k-means clustering technique. It's critical to choose the right amount of clusters because it'll fall somewhere between full localisation and standardisation (i.e. a store-specific or mass-market approach).

Working with the right amount of clusters for your retail data and market environment can help you make more efficient and effective use of your resources. When employing the k-means algorithm, you can choose the number of clusters by applying industry-specific knowledge or three distinct statistical methodologies.

The elbow technique plots the cost covariance matrix produced by various k values. When you might expect, as k rises, average distortion falls, each cluster has fewer element

examples, and the instances are closer to their distinct centroids. As k grows larger, however, the typical distortion improves less. The elbow is the value of k at which the progress in distortion diminishes the most, and at that must cease separating the data into more clusters.

The Silhouette coefficient is a number that represents how much of a silhouette there is You must assess the quality of the clusters formed in order to establish the ideal number of clusters. This value indicates how close each data point is to the cluster's centroid. Clusters with a high average silhouette coefficient are likely to be effective.

This method examines the silhouette coefficient for a variety of k values. As a result, the maximum number of clusters is the best number. The k-means technique locates the data's mean points, also known as centroids. It then gives a centroid to each data point to create the initial clusters. The method will calculate the distances between each point and the centroids and allocate every point to the point with the shortest distance.

The algorithm then repeats the process to reinvest the points to an unique centroid, with identical data points in each cluster and heterogeneous data points from various clusters.
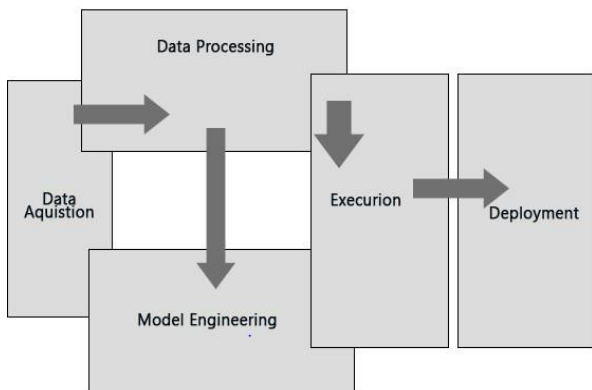

Fig 1. Architecture

II. STUDY ON RELATED WORK

Clustering is a term that refers to a collection of approaches for selection is based of occurrences in a data set. The owner want information in the same group to be similar and observations in other groups to be distinct when clustering them. This is an unsupervised approach since there is no dependent variables. It aims to uncover correlations between both the data instances without being trained by a predictor variables. Clustering allows us to see which events are similar and label them accordingly. For partitioning a dataset into a collection of k groups, K-means clustering is the basic and widely used clustering method.

G M Daiyan and Md. Ataur Rahman Khan[3] in 2014 introduced a method called flood fill which is utilized for bunching. With this method the precision will also increase.

Aswan Supriyadi Sung and Yaya Heryadi [6] in 2014 published a paper where they used Information from Potential Village 2014 in Gorontalo territory was divided into five groups using the K-Medoids calculation method and a distance calculation strategy. It has been compiled Potential Village 2014 information in the Gorontalo area into five town status with the obtained number of towns for each various group relying on distancing.

Shaoyu Qiao and Xinyu Geng, [12] in 2019 published a paper

in which the investigation, everyone look at how the three calculations are presented, as well as how the dataset to be broken down is mixed in with 'commotion' data. Under ideal conditions, the three calculations may produce the best bunching. As can be seen from the analysis, the revised k medoids calculation yields the best result of the three.

Eyke Hullermeier and Maria Rifqi[11] Apart from summarising the Rand file, it also laid the groundwork for dealing with other similarity estimates that are defined in terms of similar fundamental amounts, such as the numbers a, b, c, and d of concordant and grating article sets. Despite the fact that our findings allow for such techniques to be expanded to include fluffy parts, Which of the metrical qualities will be saved by this augmentation is definitely not sufficient. Therefore interested in analysing the typical features of fluffy expansions of explicit measures such as the Jaccard list.

III. DATASET AND PREPROCESSING

In this project the dataset is collected from online which contains the information of City of Monterrey, Mexico. In that place user wants to open a Taco Place and wants a find a potential competitor in that city and to find the best place to open the restaurant. The dataset has latitude and longitude of each and every shop in the city.the dataset contains 3782 rows with 6 columns.

	PostalCode	Neighborhood	State	County	Latitude	Longitude
0	66600	Santa Rosa II	Nuevo Leon	Apodaca	25.7862	-100.1859
1	66600	Futuro Apodaca	Nuevo Leon	Apodaca	25.7814	-100.1800
2	66600	Manuel Villarreal	Nuevo Leon	Apodaca	25.7853	-100.1816
3	66600	Homero Sepulveda	Nuevo Leon	Apodaca	25.7765	-100.1910
4	66600	Moderno Apodaca II	Nuevo Leon	Apodaca	25.7823	-100.1846

Fig 2. Image of Dataset

The dataset has so much of repeated values in the column of postal code. Preprocessing is done to remove those values. The removed values is merged in a new dataframe. As the dataset is taken from online so much of preprocessing is not required.

IV. DESIGN OF THE SYSTEM

Implementation starts from getting the coordinates for Monterrey that is latitude and longitude of each and every location. Those location are mapped in the map using Folium. The map has the points where hese locations are located. The blue dots in the map indicates the locations of the shops.
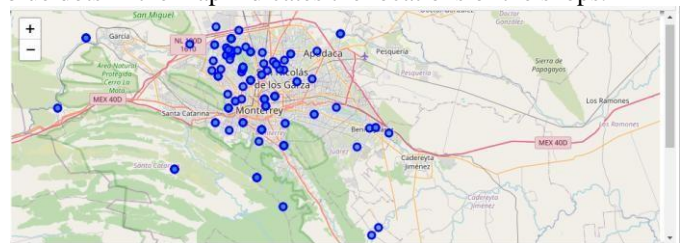
Fig 3. Points indicating the places of business within 500 meters

After getting the latitude and longitude of all locations. The dataset is prepared to connect for Foursquare API with the client ID, client secret ad version. After that first country is explored that is Monterrey with its latitude and longitude. Then the top 100 venues for this neighborhood with a radius of 500 meters is displayed. The result will be generate in the form of url then it is converted to get the human readable form which

displays all the information. Here json_normalize function is used to clean up the results in json and put it in a dataframe. Then 4 venues were returned by Foursquare.

The next step is to explore the neighborhoods in Monterrey. The above same function is used to repeat the same to all process to all the neighborhoods in Monterrey. Then checking for the venues returned by the API. The venues are displayed in the graph.
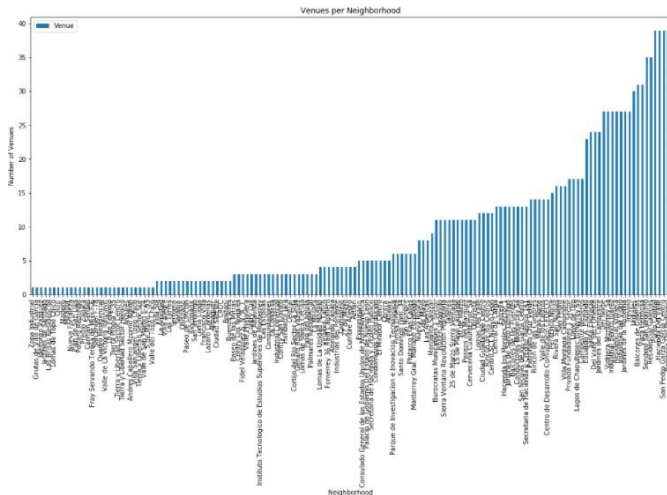

Fig 4. Venues per neighborhood

In dataset contains all categories of business. There are some repeated business so unique categories can be curated from all the returned venues. The each neighborhood is analyzed. Grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Then each neighborhood along with the top 5 most common venues is displayed. In order to put the results into a Pandas dataframe. So for that those venues are sorted in a descending order. Then new dataframe is created. After all these processing clustering is applied for the neighborhood. In order to implement this find the optimal k value using the elbow method. In k-means clustering, the elbow approach is used to estimate the ideal number of groups.

The elbow technique plots the cost covariance matrix produced by various k values. When you might expect, as k rises, average distortion falls, each cluster has fewer element examples, and the instances are closer to their distinct centroids. As k grows larger, however, the typical distortion improves less. The elbow is the value of k at which the progress in distortion diminishes the most, and at that must cease separating the data into more clusters. Then k means clustering algorithm is executed to cluster the neighborhood into 4 clusters which is the optimal value. One more dataframe is created that includes the cluster as well as the top 10 venues for each neighborhood and the result is mapped in a map.
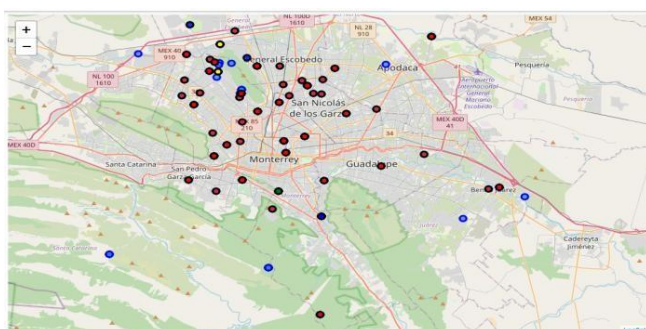

Fig 5. Clusters

Each cluster is examined and review the category that distinguish each other. The category may contain all the business like Dance studio, Design Studio, Barber shop, Furniture shop, Convenience store, restaurants, etc.

So there will be lot of competitors and the cost to open the small restaurant may be high. Four clusters are formed and in each one of the cluster contain different category of business. One may include high end venues where might be not ideal to open a Taco place. One cluster may not have a any restaurants so it might be ideal to open a Taco place which might help the owner to gain more profit and can face less competitors. One cluster includes many Taco Place which is not ideal to open a restaurant there because the owner may face lot of competitions.

So the the best cluster is the one which has no restaurant. According this clusters are suggested to the owner to open their business. The same algorithm can be applied to other business to find the competitors which will be very helpful to open their dream business with less competitors.

IV. RESULTS AND DISCUSSION

This project has made a most effect on business owners who need help to find their potential competitors. This saves the time and effort to find the competitors without surveying the place. With the one click users can find the competitors. Finding competitors is most important aspect in order to open a business.

In Figure 6 Cluster 0 is shown which has wide variety of Venues, including Taco Places and high end venues, so there will be a lot of competitors and the cost to open a small restaurant might be high. So this cluster may not help the owner to open a Taco Place. Choosing this cluster might be the worst decision.

	Neighborhood	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common V...
0	Amway	-100.3321	0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Wings Joint	Design Studio	Department Store	Deli / Bodega	
1	Consulado General de los Estados Unidos de Nor...	-100.3321	0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Wings Joint	Design Studio	Department Store	Deli / Bodega	
2	Secretaria de Hacienda y CrÉdito Publico	-100.3321	0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Wings Joint	Design Studio	Department Store	Deli / Bodega	
3	Palacio de Gobierno Del Estado de Nuevo Leon	-100.3321	0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Wings Joint	Design Studio	Department Store	Deli / Bodega	
4	Obrera	-100.3321	0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Wings Joint	Design Studio	Department Store	Deli / Bodega	
...	...	...	...	...	...	...	...	...	...	...	...	...	
142	Nueva Rosita	-100.2057	0	Pharmacy	Clothing Store	Big Box Store	Mexican Restaurant	Convenience Store	Bakery	Taco Place	Fast Food Restaurant	Deli / Bodega	D...
143	Valle de La Victoria Infonavit	-100.0845	0	Department Store	Restaurant	Wings Joint	Cosmetics Shop	Dessert Shop	Design Studio	Deli / Bodega	Dance Studio	Cycle Studio	Conver...
144	Anzures	-100.0845	0	Department Store	Restaurant	Wings Joint	Cosmetics Shop	Dessert Shop	Design Studio	Deli / Bodega	Dance Studio	Cycle Studio	Conver...
145	Jardines de Santiago	-100.1424	0	Harbor / Marina	Wings Joint	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Conver...
146	Potrero Redondo	-100.1647	0	Campground	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosi...

103 rows × 13 columns

Fig 6. Cluster 0

Cluster 1 includes high end venues, where might not be ideal open a Taco place because high end venues are those who are looking for high-quality products and are unconcerned with the price: a high-end retail store/high-end video equipment High-end buyers seek premium-quality products and are willing to pay a high price for them. So the owner may not meet those qualities which is why it is not suitable for opening a Taco Place in hat cluster.

	Neighborhood	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Obispado	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
12	Loma Linda	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
20	Bortoni	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
23	CROC	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
28	La Amistad	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
30	Francisco Villa	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
43	Lazaro Cardenas	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
58	Sarabia	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
64	Almaguer	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
70	San Jeronimo	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
78	Buenos Aires	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
91	Las Torres	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
93	Villa Sol	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop
94	Ciudad SatElite	-100.311	1	Café	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop

Fig 9. Cluster 1

Cluster 2 has no Taco Place so the neighborhood in this cluster should be considered. Because this will be ideal as it has no competitors. People will have very less option to choose the restaurant. This cluster will help the owner to find the right place to open a business.

	Neighborhood	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
79	Valle Primavera	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
80	Industrial Comercial	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
81	Contry Lux	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
82	Jardines de Altavista	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
83	Instituto Tecnologico de Estudios Superiores d...	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
84	Contry San Juanito	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
85	Contry Los Nogales	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
86	La Condesa	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
87	Palmares Residencial	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
88	Cortijo del Rio Sector La Silla	-100.2667	2	Convenience Store	Fruit & Vegetable Store	Cosmetics Shop	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Wings Joint
101	Fomerrey 36 Raul Caballero	-100.3428	2	Burrito Place	Convenience Store	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Wings Joint
102	Andres Caballero Moreno Agrop	-100.4012	2	Convenience Store	Wings Joint	Cycle Studio	Diner	Dessert Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cosmetics Shop

Fig 10. Cluster 2

Cluster 3 has a wide variety of Venues, including Taco Place which means there will be lot of competitors which is not ideal. So the owner may not choose to open their business here. Because it may lead to loss which is not the objective of our project.

	Neighborhood	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Paseo de las Mitras	-100.3727	3	Pharmacy	Pizza Place	Convenience Store	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Cosmetics Shop	Comfort Food Restaurant
11	Fidel Velazquez S. N. A. T.	-100.3727	3	Pharmacy	Pizza Place	Convenience Store	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Cosmetics Shop	Comfort Food Restaurant
14	Paso del aguila	-100.3727	3	Pharmacy	Pizza Place	Convenience Store	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Cosmetics Shop	Comfort Food Restaurant
15	Genaro Rojas Vazquez	-100.3727	3	Pharmacy	Pizza Place	Convenience Store	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Cosmetics Shop	Comfort Food Restaurant
16	Lomas de Santa Cecilia	-100.3727	3	Pharmacy	Pizza Place	Convenience Store	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Cosmetics Shop	Comfort Food Restaurant
22	Central	-100.3727	3	Pharmacy	Pizza Place	Convenience Store	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Cosmetics Shop	Comfort Food Restaurant
31	San angel F-78	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
32	Fray Servando Teresa de Mier F-6	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
33	18 de Febrero	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
34	Tierra y Libertad Sector Heroica	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
35	Predio Zapata	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
36	Tierra y Libertad Sector Sur	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
37	Cnop	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
38	Tierra y Libertad Sector Centro	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
39	Plan de San Luis	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
40	Madero	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
41	Carmen Serdan	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner
42	Valle del Topo Chico	-100.3705	3	Pharmacy	Food Stand	Cosmetics Shop	Design Studio	Department Store	Deli / Bodega	Dance Studio	Cycle Studio	Convenience Store	Diner

Fig 11. Cluster 3

V. CONCLUSION

The task entails locating similar firms in other places. one usually start by grouping related businesses together. Once the establishing of clusters is done, one can look into the cluster to which a certain eating establishment belongs in order to display eateries in this cluster that are close to the location. The enterprises were clustered using the K- medoids computational technique. One wants a more diverse selection of difference scores because not all qualities are quantitative. Studies with different input settings, difference scores, and characteristics are carried out, and the rand index is reported on a "gold standard" data set. If a company decides to create a branch in a new location, this data set will assist them in identifying possible competitors.

Large corporations may use this information to not only identify possible competitors, but also gain a sense of how often their company will be received in the new area.

The results obtained and the analysis of the results. This section also focuses on the implementation and monitoring of established system performance. The outcomes of the results and objectives are also listed in this section. A comparative analysis is made to verify the compare the results between the developed system and the existing system. The comparison between the present system and the existing system is made and the outcomes are recorded. This is an assessment to determine how much the findings are consistent with the previously established objectives in order to confirm how much the project has gone.

REFERENCES

[1] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M (2016) Tensorflow: a system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016), pp 265-283

[2] Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, Casper J, Catanzaro B, Cheng Q,

Chen G, Chen J (2016) Deep speech 2: End-to-end speech3 recognition in English and Mandarin. International Conference on Machine Learning, pp 173-182

[3] J. Han, M. Kamber "Data Mining: Concepts and Techniques," Second Edition, Elsevier Inc. Rajkamal Electric Press, 2006.

[4] Ester M., Frommelt A, Kriegel H, P., and Sander "Spatial data mining: database primitives, algorithms and efficient DBMS support. Data Mining and Knowledge Discovery", Kluwer Academic Publishers, 4, 2-3, 193-216, 2009

[5] Cadez, I., Smyth, P., and Mannila, "Probabilistic modeling of transactional data with applications to profiling, Visualization, and Prediction", In Proceedings of the 7th ACM SIGKDD, 37-46, San Francisco, CA, 2010

[6] Cooley, R., Mobasher, B., and Srivastava "Data preparation for mining world wide web browsing. Journal of Knowledge Information Systems", 1, 1, 5-32, 2016

[7] H. et Al, "Data Mining : Concepts and Techniques", 2011.

[8] Ada Desa tak Berpenghuni Tapi Dapat Dana Desa Kok Bisa", 2019, [online] Available: https://www.cnbcindonesia.com

[9] A. Singh, A. Yadav and A. Rana, "K-means with Three different Distance Metrics", Int. J. Comput. Appl., vol. 67, no. 10, pp. 13-17, 2013.

[10] E. Hullermeier, M. Rifqi, S. Henzgen and R. Senge, "Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures," in IEEE Transactions on Fuzzy Systems, vol. 20, no. 3, pp. 546-556, June 2012, doi: 10.1109/TFUZZ.2011.2179303.

[11] G. M. Daiyan, F. B. A. Abid, M. A. R. Khan and A. H. Tareq, "An efficient grid algorithm for faster clustering using K medoids approach," 2012 15th International Conference on Computer and Information Technology (ICCIT), 2012, pp. 1-3, doi: 10.1109/ICCITechn.2012.6509704.

[12] A. S. Sunge, Y. Heryadi, Y. Religia and Lukas, "Comparison of Distance Function to Performance of K-Medoids Algorithm for Clustering," 2020 International Conference on Smart Technology and Applications (ICoSTA), 2020, pp. 1-6, doi: 10.1109/ICoSTA48221.2020.1570615793.