**REPORT ON**

# CAPSTONE PROJECT-BATTLE OF NEIGHBOURHOOD

## Introduction to Business Problem:

### ***Opening of a MNC Bank branch in one of the HNI residency location in Bangalore***

###### Bangalore is the capital city of the state of Karnataka.

The economy of Bangalore contributes about 87% to the state of Karnataka. The majority of the city of Bangalore lies in the Bangalore Urban district of Karnataka and the surrounding rural areas are a part of the Bangalore Rural district. The region comprising the Bangalore Urban and Rural districts is known as the Bangalore (region). The Government of Karnataka has carved out the new district of Ramanagaram from the old Bangalore Rural district. The metro area population of Bangalore in 2020 was 12,327,000, a 3.74% increase from 2019. Bangalore has grown by 438,403 since 2015, which represents a 3.56% annual change. Bangalore has an estimated population of 12.34 million in its urban area in 2017, up from 8.5 million in 2011. It is now the 24th most populous city in the world.

Bengaluru is the third wealthiest city in India according to a study. Bengaluru boasts of a total wealth of USD 320 billion. The city is a home to 7,500 millionaires and 8 billionaires.

Millionaires or high net worth individuals (HNWIs) refer to individuals with net assets of USD 1 million or more. Billionaires refer to individuals with net asset of USD 1 billion or more.

HNI's in Bengaluru prefer to invest in various investment products, real estate, digital currency etc., which gives huge opportunities to financial institutions like banks.

Looking at the current situation most of these ultra HNI's will prefer to work from home. So opening bank branch in one of these residency area will provide bank more opportunities to onboard such customers and cross sell multiple banking products.

## Description of the data used to solve the problem:

**Introduction to location data and Foursquare:**

Location Data: Location Data is data describing places and venues such as their geographical location, their category, working hours, full address and so on. **Location data** helps to create an accurate representation of your interests, and this can be used to bring more targeted and relevant ads to potential customers.

**FOURSQUARE:**

Foursquare is a technology company that built a massive dataset of location data. What is interesting about Foursquare is that they were very smart about building their dataset. They actually crowd sourced their data and had people use their app to build their dataset add venues and complete any missing information they had in their dataset.

***Foursquare API:***

The **Foursquare** Places **API** provides location-based experiences with diverse information about venues, users, photos, and check-ins. The **API** supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag.

**Data used for analysis:**

We need neighbourhood name for various HNI residency location of Bangalore so that we can find out the best neighbourhood location to open a bank branch.

For this we will be using Foursquare location data to get our required data for execution as we don't get any particular website which can give location data for only HNI residency in Bangalore.

Firstly, we will get the geographical coordinates for Bangalore using Foursquare location data and then use it to get neighbourhood information for various HNI's residency location in Bangalore.

With the help of Foursquare location data, we create our dataframe.

## **METHODOLOGY**

***What is Machine Learning?***

**Machine learning** is a method of data analysis that automates analytical model building. It is a branch of artificial **intelligence** based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

**We use K-Nearest Neighbour (KNN) classification algorithm to do our analysis.**

**K-Nearest Neighbour (KNN):** KNN is one of the simplest machine learning algorithms based on Supervised learning. This algorithm classifies cases based on other similar cases.

How does KNN works?

The K-NN working can be explained on the basis of the below algorithm:

* **Step-1:** Select the number K of the neighbours
* **Step-2:** Calculate the Euclidean distance of **K number of neighbours**
* **Step-3:** Take the K nearest neighbours as per the calculated Euclidean distance.
* **Step-4:** Among these k neighbours, count the number of the data points in each category.
* **Step-5:** Assign the new data points to that category for which the number of the neighbour is maximum.
* **Step-6:** Our model is ready

## **How to select the value of K in the K-NN Algorithm?**

Below are some points to remember while selecting the value of K in the K-NN algorithm:

* There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.
* A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
* Large values for K are good, but it may find some difficulties.

Since the most preferred value of k is 5, so value of k for our analysis is also 5. We have clustered all our data points in 5 clusters.

**ANALYSIS:**

As we have taken the value of K as 5 the entire dataset consisting of **51** locations is divided into 5 cluster i.e., Cluster1, Cluster2, Cluster3, Cluster4 and Cluster 5.

On examining this clusters, we get the following:

|  |  |
| --- | --- |
| **Cluster no** | **no. of location in each cluster** |
| **1** | **10** |
| **2** | **12** |
| **3** | **14** |
| **4** | **6** |
| **5** | **9** |
| **Total** | **51** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

We can depict the above findings in a bar chart as follows.

**CONCLUSION:**

**From the above findings we can say that going with Cluster3 is more feasible. As it covers the maximum no. of HNI residency location.**

**Locations in Cluster3 are as follows:**



The exact location of opening the branch would involve availability of space in one of these locations and also, we have look at the cost involved. This could be done by doing a survey of the above-mentioned location.

**ACKNOWLEDGEMENT**

Data required for this project is based on Foursquare location data.