The background of the slide is a dense, 3D-rendered field of blue numbers. The numbers, including 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, are of various sizes and are oriented in different directions, creating a sense of depth and movement. They appear to be floating or rising from the bottom, with some numbers being more prominent than others.

Data Science Capstone Project- Bangalore Neighborhood exploration

Bibhu Ashish Panda

Problem Statement

The problem statement that we are trying to address is that a location needs to be found in Bangalore to open a Vegan restaurant.

Why this is important- With a current population of 12,327,000** Bangalore gives ample opportunity to any entrepreneur to start thinking of opening a business. Food being a necessity, and with the aspiration of crowd to be healthy, vegan restaurant proposition holds a lot of promise for any entrepreneur

A small brief about the

** Source- <https://www.macrotrends.net/cities/21176/bangalore/population>

Methodology

- ☐ Business And Data Understanding
- ☐ Data preparation
- ☐ Model Building
- ☐ Model Validation
- ☐ Model Deployment

Business and Data Understanding

Data science is the art of uncovering the insights and trends that are hiding behind data. In a nutshell Data science is the study of data. It is more about data than it is about science.

This project is aimed at analyzing the various locations in Bangalore and the popular venues in these localities. This will be used to find a locality where we can establish a vegan restaurant.

The success criteria for the project will be defined by the an appropriate outcome of the analysis of the localities to give a set of localities where we can establish a vegan restaurant.

To get the data, we will use APIs which are reliable and which will be able to cover the maximum localities that we can find in Bangalore. With this objective, I analyzed the APIs from Zomato, geonames and Wikipedia. Though all these APIs were valuable resources for getting a lot of data about Bangalore, I choose Wikipedia over others as it was more clear and was available free without any constraint. Once these localities were obtained, the API from the foursquare was overlapped with the localities to get the details required for analysis.

Data Preparation

The data was fetched from the Wikipedia with the help of an API and the scrapper libraries available in Python.

```
data = requests.get("https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore").text
soup = BeautifulSoup(data, 'html.parser')
# create a list to store neighborhood data
neighborhoodList = []
```

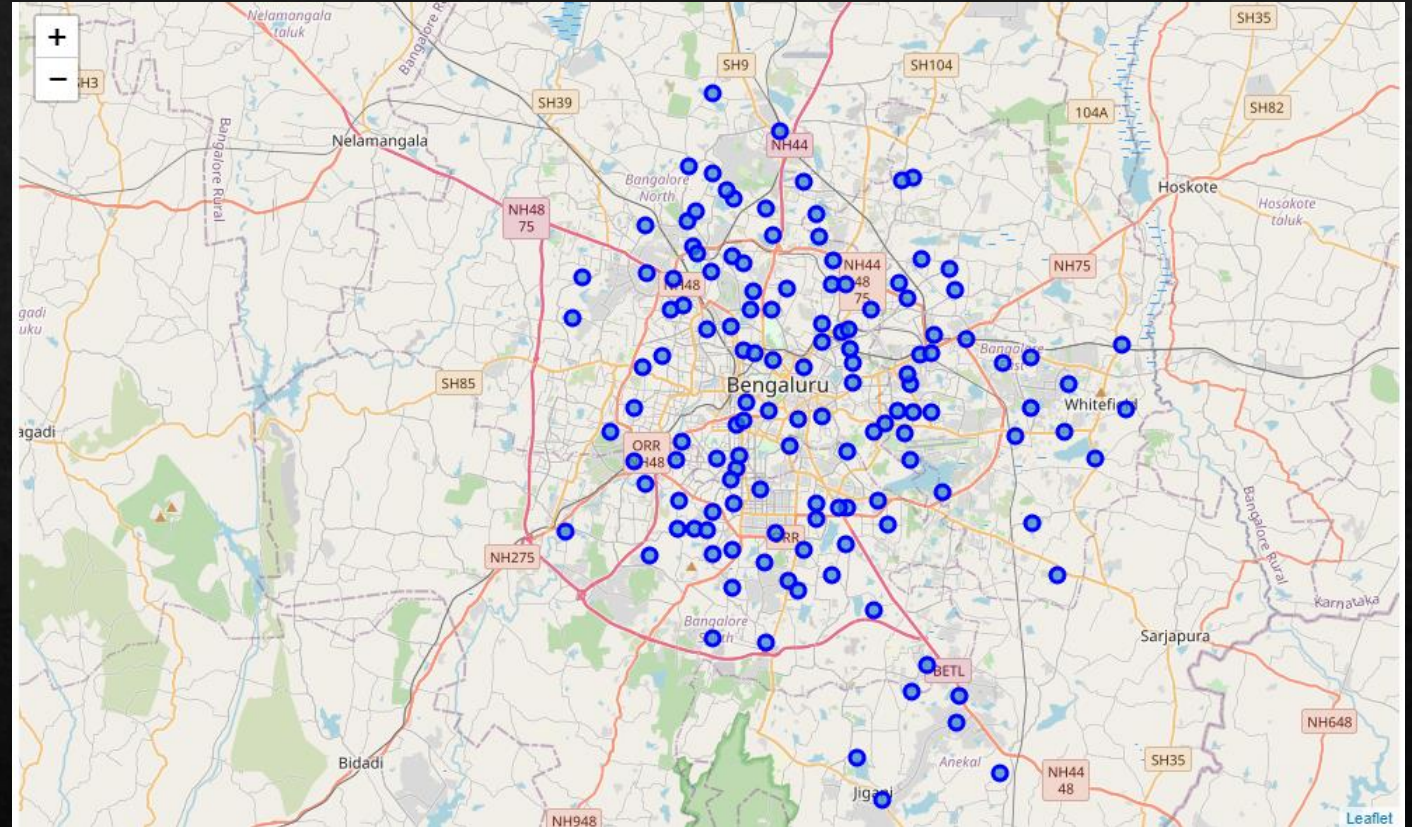
The data was loaded to a data frame and this was cleaned from all the null values

Clean the data of the null values

```
86]: k1_df = k1_df[k1_df['Latitude'].notna()]
      k1_df = k1_df[k1_df['Longitude'].notna()]
```

Data Preparation- Visual exploration

Data visualization- We used the folium library to visually explore the data to find if any anomalies are there in the data or not



Modelling

As I was dealing with a set of unlabeled data and as this requirement dealt with the assessment of human behavior, I used Clustering Association approaches. Clustering which is an unsupervised machine learning algorithm is, grouping of data points or objects that are similar and helps in discovering structure, summarization, and anomaly detection of the data sets. There are various types of clustering algorithms such as partitioning, hierarchical, or density-based clustering.

As part of this approach, k-means clustering was used to segregate the localities to various clusters. k-means is a type of partitioning clustering that is, it divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels. This means it is an unsupervised algorithm. Objects within a cluster are similar and objects across different clusters are quite different or dissimilar.

I decided to create all the localities of group into 5 clusters

Modelling

The top 10 venues were found across all the localities

Listing the top 10 most common venues for all the locality

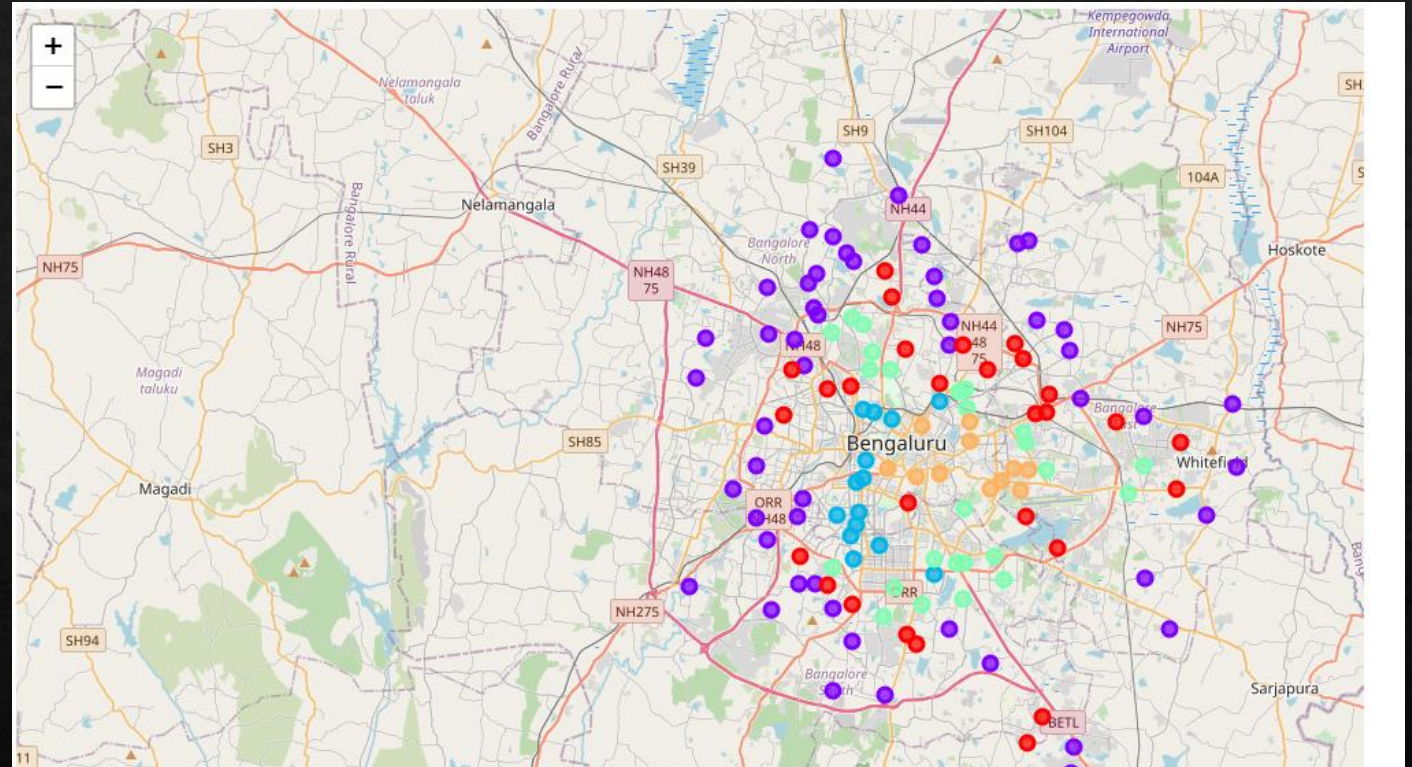
```
] : neighborhoods_venues_sorted
```

```
] :
```

	Neighborhood	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
0	Adugodi	ATM	Coffee Shop	Pool	Women's Store	Dim Sum Restaurant	Farmers Market	Falafel Restaurant	Event Space	Event Service	Electronics Store
1	Agara, Bangalore	Indian Restaurant	Café	Ice Cream Shop	Pizza Place	Department Store	South Indian Restaurant	Clothing Store	Fast Food Restaurant	Shopping Mall	Chinese Restaurant
2	Ananthnagar	Hotel	Indian Restaurant	Pub	Café	Ice Cream Shop	Brewery	Tea Room	Lounge	Bar	Breakfast Spot
3	Anjanapura	Indian Restaurant	Café	BBQ Joint	Korean Restaurant	Coffee Shop	Lounge	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Ice Cream Shop
4	Arekere	Pizza Place	Brewery	Coffee Shop	Gym / Fitness Center	Mediterranean Restaurant	Convenience Store	Sandwich Place	Badminton Court	Sporting Goods Shop	Stadium
5	Austin Town	Indian Restaurant	Ice Cream Shop	Punjabi Restaurant	Park	Pizza Place	Café	Department Store	Juice Bar	Hotel	Bakery
6	BTM Layout	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Bakery	Coffee Shop	Sandwich Place	Movie Theater	Asian Restaurant	Hotel	Restaurant

Modelling

A model was created to segregate the localities as per the most common venues



Modelling

Explore the neighborhoods for the Vegan restaurants

```
: len((kl_grouped[kl_grouped["Vegetarian / Vegan Restaurant"] > 0])) #So there are 55 vegan restaurant in Bangalore (which it very high)  
#So now we want to select a good location where the no of shopping malls are less so that our chances of setting up a shopping mall at that location should be good
```

```
: 55
```

```
: kl_vegan = kl_grouped[["Neighborhood", "Vegetarian / Vegan Restaurant"]]  
kl_vegan=kl_vegan[kl_vegan['Vegetarian / Vegan Restaurant']!=0]  
kl_vegan.sort_values(['Vegetarian / Vegan Restaurant'], ascending=False)
```

Deployment and Test

The number of cluster was decided to be four

```
# set number of clusters
kclusters = 4

kl_clustering = kl_vegan.drop(["Neighborhood"], 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(kl_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]

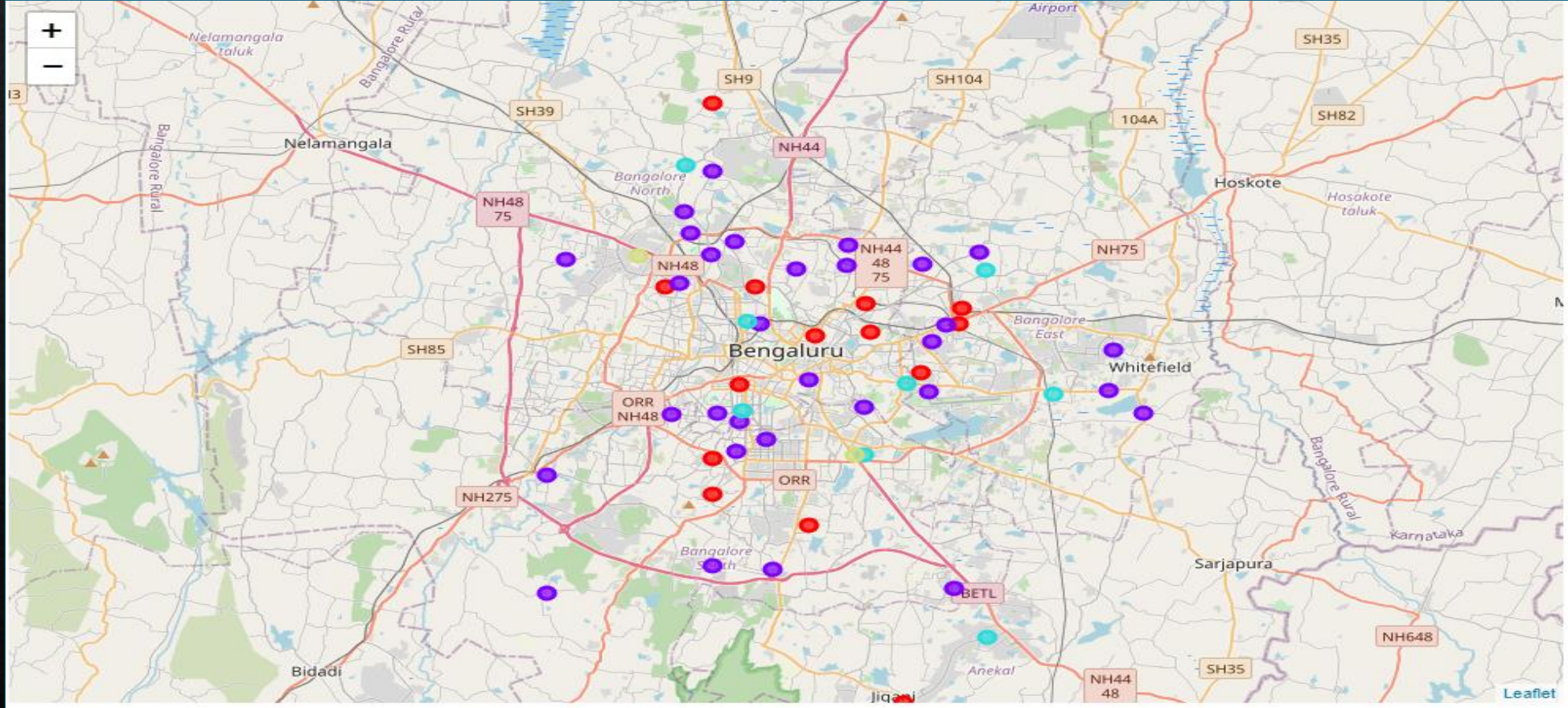
array([1, 1, 1, 1, 1, 1, 2, 2, 2, 1])

# create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.
kl_merged = kl_vegan.copy()

# add clustering labels
kl_merged["Cluster Labels"] = kmeans.labels_

kl_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)
kl_merged.head(10)
```


Model building- Visual representation



Deployment and Test

The localities were divided into 4 clusters depending on the number of vegan restaurants available.

Cluster 1

```
kl_merged.loc[kl_merged['Cluster Labels'] == 0]
#Len(kl_merged.loc[kl_merged['Cluster Labels'] == 0])# -132 neighbourhoods/places in this cluster 0
```

	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels	Latitude	Longitude
103	Ramagondanahalli	2	0	13.117880	77.559210
62	Jigani	2	0	12.781130	77.642580

Cluster 2

```
kl_merged.loc[kl_merged['Cluster Labels'] == 1]
#Len(kl_merged.loc[kl_merged['Cluster Labels'] == 1])
```

	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels	Latitude	Longitude
110	Seshadripuram	1	1	12.993550	77.57988
106	Richmond Town	1	1	12.962340	77.60123
130	Vvalikaval	1	1	13.037920	77.61863

Deployment and Test

The localities were divided into 4 clusters depending on the number of vegan restaurants available.

Cluster 3

```
: kl_merged.loc[kl_merged['Cluster Labels'] == 2]
#len(kl_merged.loc[kl_merged['Cluster Labels'] == 2] )
```

	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels	Latitude	Longitude
96	Palace Guttahalli	3	2	12.995508	77.574772
14	Bellandur	3	2	12.960530	77.643850

Cluster 4

```
: kl_merged.loc[kl_merged['Cluster Labels'] == 3]
#len(kl_merged.loc[kl_merged['Cluster Labels'] == 3] ) #-There are only 2 neighborhoods with 4 restaurants
```

	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels	Latitude	Longitude
81	Madiwala	4	3	12.92052	77.62090

Conclusion

The analysis shows that there are a good number of vegan restaurants spread across the Bangalore city. Though the number is scattered across Bangalore city, couple of places where the Vegan restaurants can be targeted to open are places where there is one restaurant, but they have a good number of IT folks. These places can be Electronic City, BTM Layout, Indira Nagar. It is also surprising to see that Peenya like place having maximum number of Vegan restaurants which is considered to be at the outskirts of Bangalore



Thank You