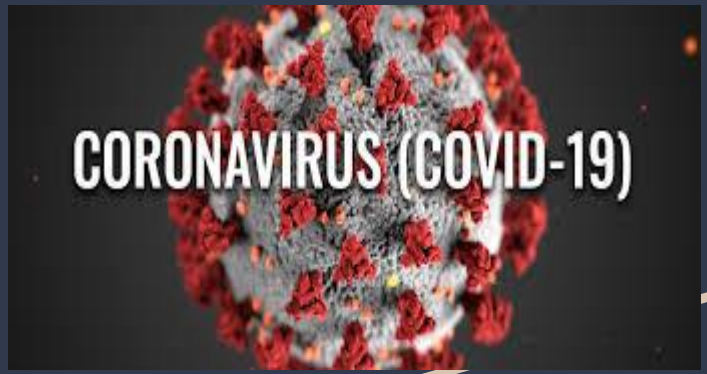
Capstone Project - The Battle of Neighborhoods (Week-2) Final

Shasank M. Dave
Covid-19 Mumbai

Table Of Content:

Introduction:Business Problem

COVID-19



- The outbreak was identified in Wuhan, China in december 2019
- WHO Declared Pandemic on 11 March 2020
- As of 16 April 2020, more than 2.1 million cases of COVID-19 have been reported in 210 countries and territories, resulting in more than 140,000 deaths
- he deaths per diagnosed cases varies significantly between countries

- **Problem**: Fortunately we are in times of social media and internet to keep us update with all the information, as social media has negative side to, with all fake news making rounds on social media which ill-informs manys
- **Interest:** To plot & visualize correct information for common man, sourced from geuine sources
- Neighbourhood selected is “Mumbai” as it is currently eSSpicenter for the COVID-19 in india

Data Acquisition:

Following are the main data sources Identified :

- https://www.kaggle.com/sudalairajkumar/covid19-in-india
- Information tracked on daily basis https://api.covid19india.org/data.json
- Data for mumbai covid-19 cases is sourced from http://stopcoronavirus.mcgm.gov.in/ https://github.com/shasaankdave/Coursera_Capstone/blob/master/Containment_Zones_BMC_Mumbai.pdf
- Details about mumbai Pincodes, Area,Borough(Called BMC Wards) are sourced from https://data.gov.in/resources/all-india-pincode-directory-along-contact-details & only mumbai's pincodes & details are extracted

Data Acquisition:2

- We decided to use folium maps with latitude & longitude to display covid-19 zones, cluster & testing facilities Latitude & Longitude based on Pincodes of mumbai are obtained using Google Maps API geocoding

- Further Details about Hospitals/testing labs are got from Foursquare API using categoryId parameter** in the request URL to select hostipal as venue category example venue categoryId for hospitals : “4bf58dd8d48988d196941735” Refer for more details : https://developer.foursquare.com/docs/build-with-foursquare/categories/

Data Pre-Processing

Data for mumbai city is extracted from an pdf on http://stopcoronavirus.mcgm.gov.in/ Mumbai's pincodes are selected from all india pincode directory. latitude & Longitude data for hospital data was reported to be wrongm hence we use Geolocator API as as follows to get co-ordinates:

Found Latitude on Longitude reported Above as not correct.

Hence Use GeoLocator API to get authentic Latitude & Longitude

[35]:
```
lat=[]
long=[]
pincd=[]

#address = 'Mumbai, india'
for pincode in mumbai_labs_ll['Pincode']:
    add= str(pincode) + ',india'
    address = add
    geolocator = Nominatim(user_agent="can_explorer")
    location = geolocator.geocode(address,timeout=15)
    latitude = lat.append(location.latitude)
    longitude =long.append(location.longitude)
    pincd.append(pincode)
    print('The geograpical coordinate found for {}.'.format(add))
```

```
The geograpical coordinate found for 400012,india.
The geograpical coordinate found for 400034,india.
```

Data Pre-Processing: Four Square API

details on venue category Ids Category Id for Hospitals – 4bf58dd8d48988d196941735

In [41]:
```
#FourSquare API Parameter SetUp

LIMIT = 100

CLIENT_ID = 'ZVLUP1PU41P2XVZN1R15EWMRBMOUU5MA3HTHOYKFKZTHWMYJ' # your Foursquare ID
CLIENT_SECRET = 'ST0MKT1KNGXTV5PXYGRDHNUHA5XRYKTM5V4NGUQG4FHEP245' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
CATEGORYID='4bf58dd8d48988d196941735' #venueid for hospitals

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
lat='19.25023195'
lng='73.16017493'
radius=500
```

```
Your credentails:
CLIENT_ID: ZVLUP1PU41P2XVZN1R15EWMRBMOUU5MA3HTHOYKFKZTHWMYJ
CLIENT_SECRET:ST0MKT1KNGXTV5PXYGRDHNUHA5XRYKTM5V4NGUQG4FHEP245
```

In [42]:
```
 # create the API request URL
radius=500
latitude=19.0025 #Mumbai
longitude=72.8421 #Mumbai
#for latitude,longitude in zip(mumbai_covid_clusters['Latitude'],mumbai_covid_clusters['Longitude']):
url = 'https://api.foursquare.com/v2/venues/search?categoryId={}&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit=
results = requests.get(url).json()["response"]#['groups'][0]['items']
    #results = requests.get(url).json()
results
```

Methodology & Analysis:

Methodology:

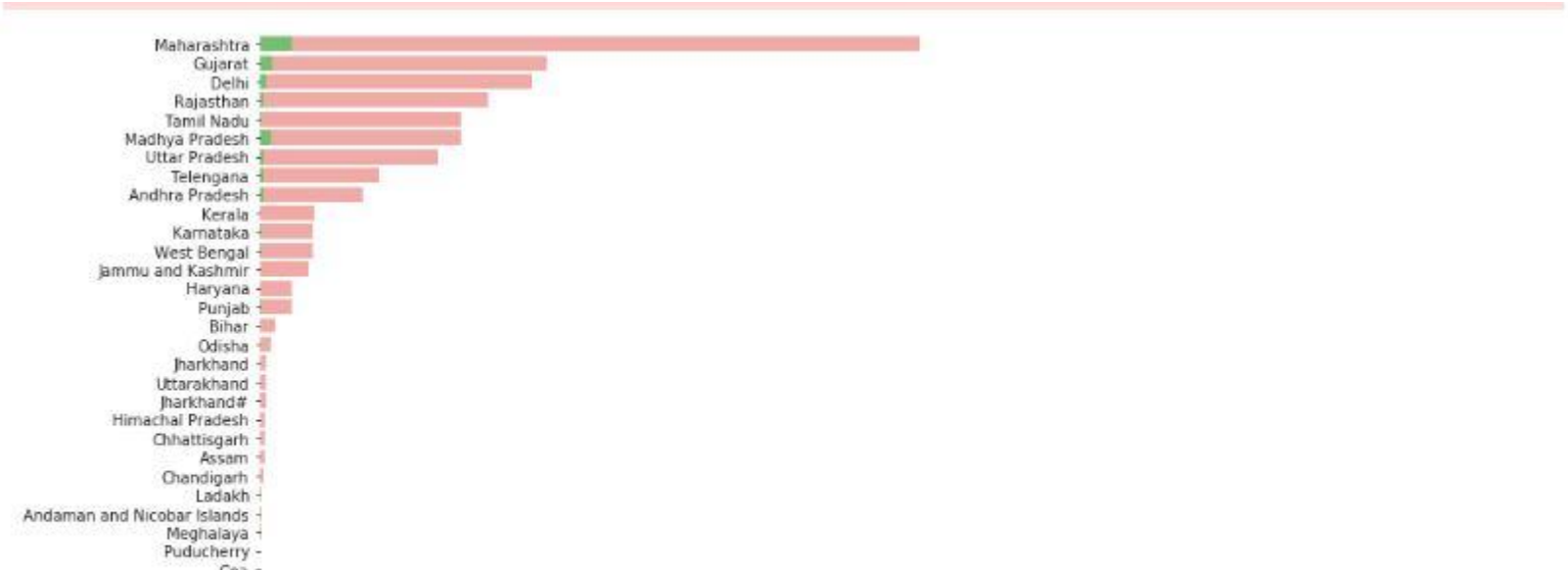
In this project we will direct our efforts on finding covid-19 zones in mumbai and further find details and plot folium maps for:

- Covid-19 zones in Mumbai, India
- Covid-19 testing facilities in the neighbourhood
- Coivd-19 Clustered for mumbai neighbourhood.

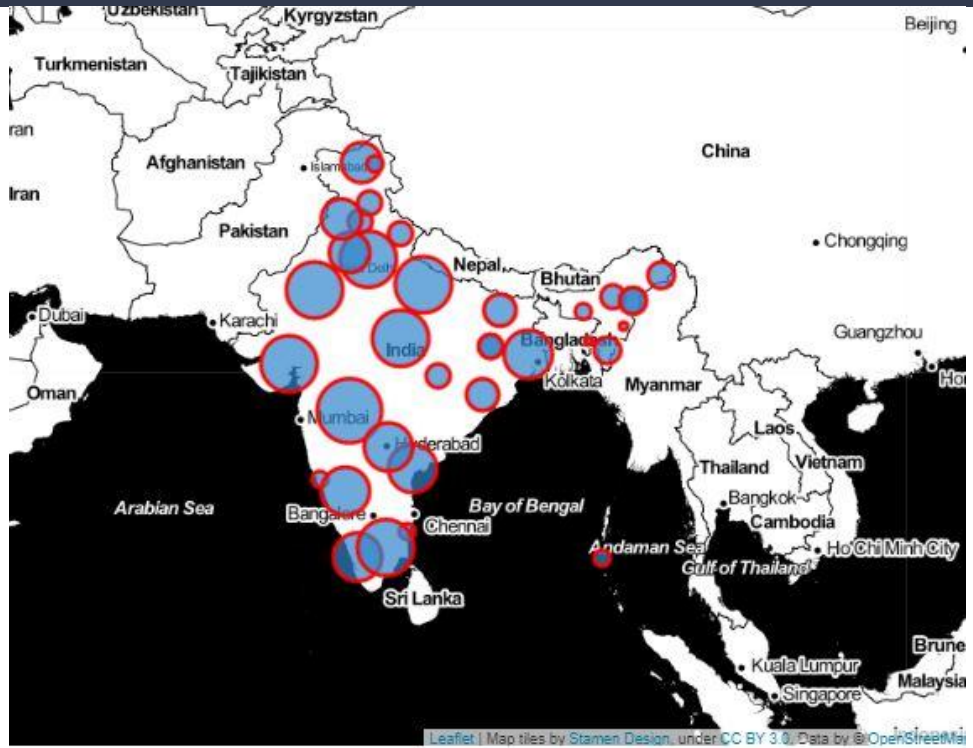
Methodology section is divided into 4 parts.

Methodology:Step 1

In first step with "COVID-19:INDIA" above we tried to analyse covid-19 situation in INDIA with bar plot, line plot & folium maps to get the top effected city, after analysis we found Mumbai, Maharashtra, India is the most effected city in india, Hence we selected Mumbai as our neighbourhood for this capstone project



India COVID-19 Map:

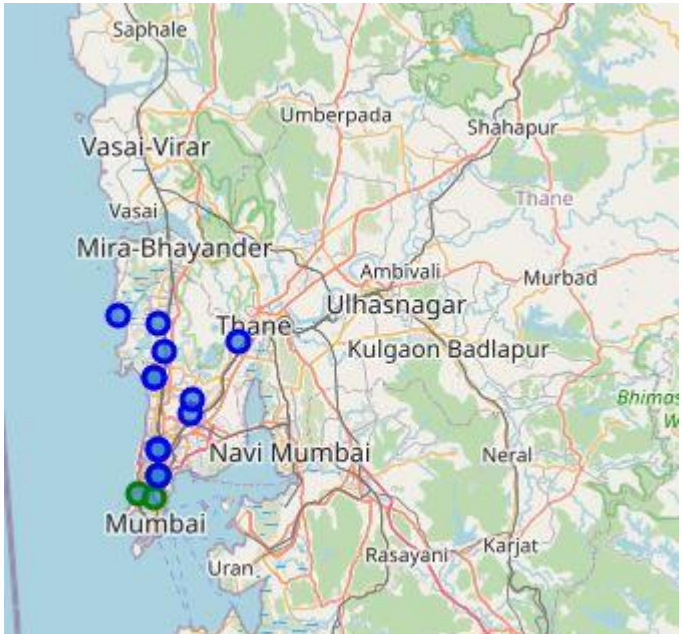


Methodolody: Step2

In Second step we tried get data for hospitals/testing labs in mumbai from source & tried to get details using foursquare API and all information for latitude & longitude was obtained using Geolocator API

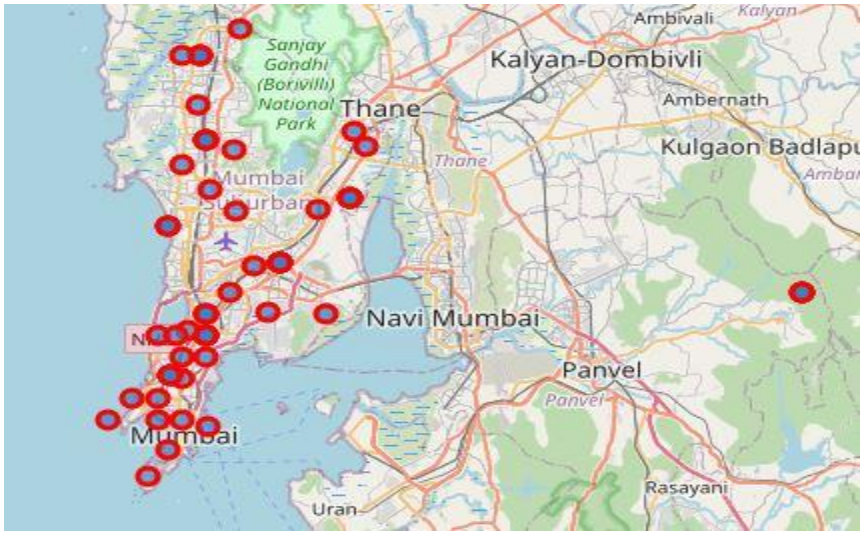
Color Code:

- Blue Non-Govt Lab
- Green Govt Lab



Methodology: Step 3

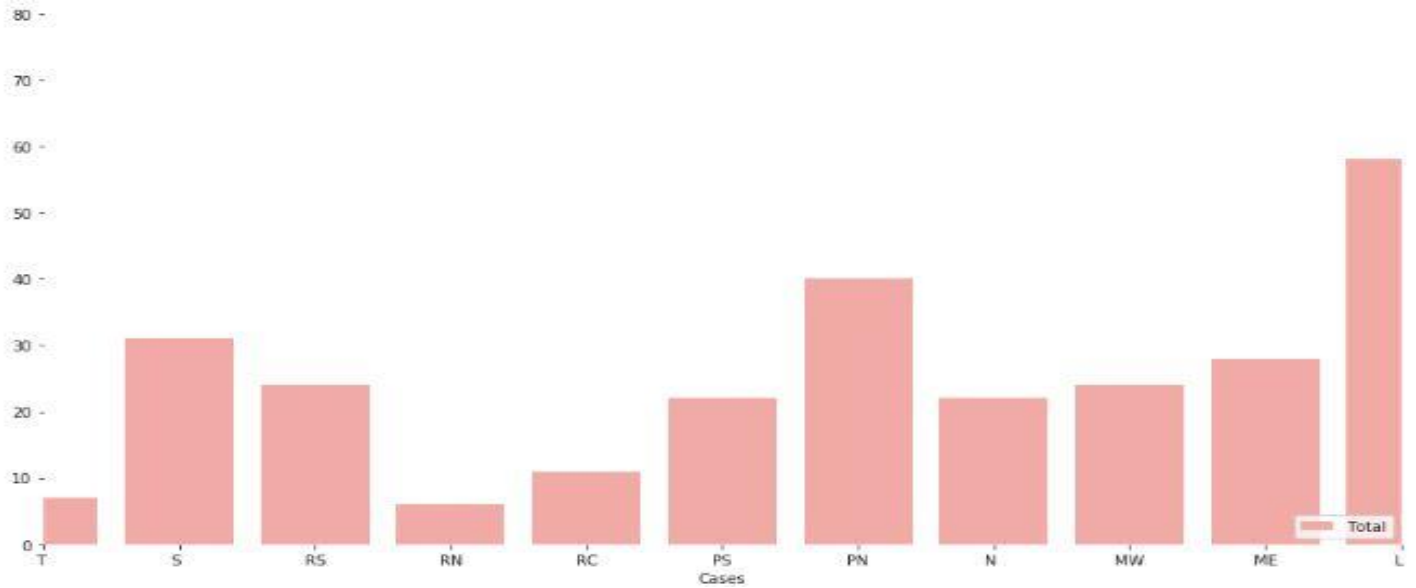
In Third step we started with getting information regarding covid-19 zones in mumbai, information is sourced from mumbai municipal corportation website, other needed information regarding postal code was found from government of india site and to have a clear picture of cases in mumbai we grouped mumbai covid zones as per boroughs(Called Wards in Mumbai) and plotted covid zones



Methodology:Step 3

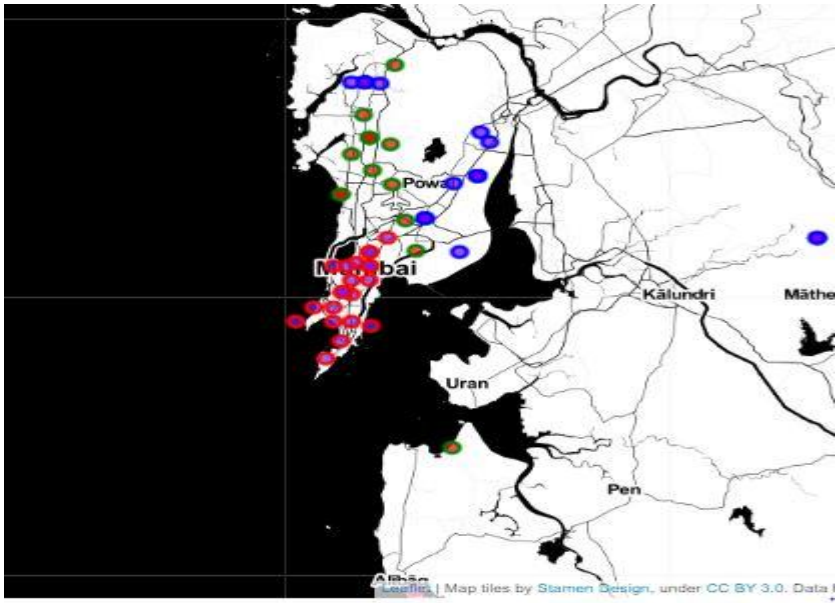
```
            label= 'Total' , color= 'r' )


# Add a legend and informative axis label
ax.legend(ncol=2, loc="lower right", frameon=True)
ax.set(xlim=(0, 10), ylabel="",
       xlabel="Cases")
sns.despine(left=True, bottom=True)
plt.show()
```



Methodology:Step 4

In fourth and final step we tried to cluster mumbai covid-19 zone using k-mean clustering, to have clear label on clustering we plot a folium maps for clusters color-coded by cluster numbers., k-value for clustering was 3 as we have in total 21 wards.



Result & Discussion

After running the K-means clustering we can access each cluster created to see which ward was assigned what cluster[0,1,2]

final result set for cluster data has following columns:

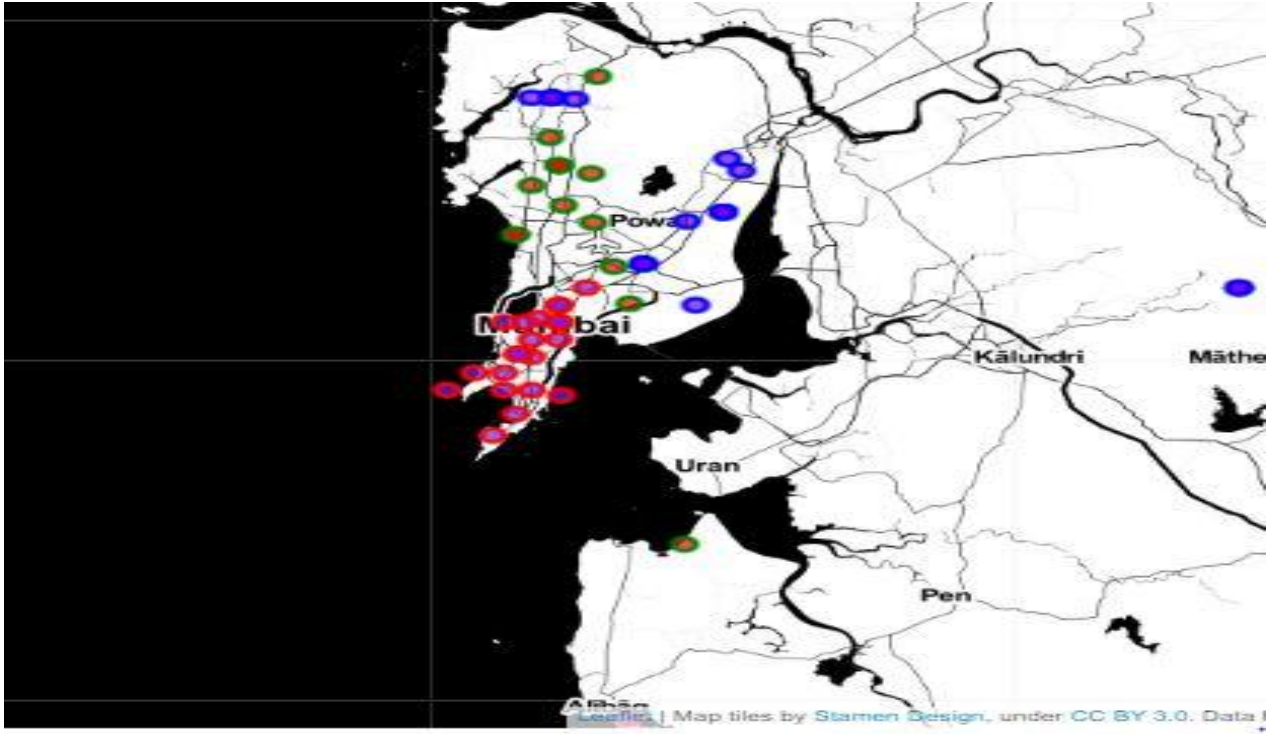
- Pincode
- Latitude
- Longitude
- Cluster_label

```
In [101]: mumbai_covid_clusters=pd.DataFrame({"Pincode": pincd_lst,
                                              "Latitude": covid_lat_lst,
                                              "Longitude":covid_long_lst,
                                              "Cluster_label":cluster_lst,
                                               })
          mumbai_covid_clusters.head()
```

Out[101]:

	Pincode	Latitude	Longitude	Cluster_label
0	400001	18.6291	72.8919	2
1	400002	17.0509	73.2910	2
2	400003	18.9500	72.8333	2
3	400004	18.9500	72.8167	2
4	400005	18.9069	72.8106	2

Results: Clusters Map



Discussion:

Total Number of covid-19 cases in India : 27892

● Total Number of covid-19 cases in mumbai: 5407 src: https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Maharashtra

● Around 20% of totals india's cases are in mumbai with around 17 Hospitals/testing labs(3 Government/14 Non-Government Labs)

- Mumbai City is most populated city in india with population of 20,411,274 Source:https://worldpopulationreview.com/world-cities/mumbai-population/ , with very densily populated slums around and as Mumbai is epicenter for COVID-19 in india currently with around 20% cases of india's number, it is highly on the risk of getting into phase 3(Community transmission). Mumbai city has around 1000 diagnostic centre and only 17 are catering for COVID-19 which is very few in number & for such pandemic situation and for epicenter like mumbai testing facilities should be increased for quick assessment, rapid testing and stop the spread of virus.

Conclusion

Purpose of this project was to start analysing COVID-19 situation in India & pickup the city which has the most number of corona positive cases, Further to put the analysis & visuzlation sourced from authentic sources to avoid fake information & inform people well.

Mumbai was identified as epicenter city for current corona pandemic with around 20% share alone by this city for total number of case in india, further information about testing center was mapped with cases to know relation between number of cases vs testing labs.

Based upon this analysis it was found that for common man testing facilities for COVID-19 in the city of mumbai should be increases looking at the current number.

Finally I would say!!
#StayHome#StaySafe