Capstone Project : The Battle of Neighbourhoods

1 Capstone Project — The Battle of Neighbourhoods

1.0.1 Introduction

New Delhi is the capital city of India. It is a part of the city of Delhi's 11 districts. The city itself has a population of 257,803. However, the much larger metro area has a population that exceeds 26 million.

New Delhi are used interchangeably to refer to the National Capital Territory of Delhi (NCT), these are two distinct entities, with New Delhi forming a small part of Delhi. The National Capital Region is a much larger entity comprising the entire NCT along with adjoining districts in neighboring states.

<p>The official language of New Delhi and the one that is most widely spoken is Hindi. However Over last decadesit is continuously grow because of the citys important role in governmentan

With it's diverse culture , comes diverse food items. There are many restaurants in New Delhi City, each belonging to different categories like Chinese , Italian , French etc.

So as part of this project , we will list and visualise all major parts of New Delhi City .

Questions that can be asked using the above mentioned datasets - What is best location in New Delhi City for Chinese Cuisine ? - Which areas have large number of Chinese Resturant Market ? - Which all areas have less number of resturant ? - Which is the best place to stay if I prefer Chinese Cuisine ? - What places are have best restaurant in New Delhi?

1.0.2 Data

For this project we need the following data :

New Delhi Resturants data that contains list Locality, Resturant name,Rating along with their latitude and longitude.

Data source : <a href="https://www.kaggle.com/shrutimehta/zomato-restaurants-data">Zom

Description : This data set contains the required information. And we will use this data set to explore various locality of new delhi city.

Nearby places in each locality of new delhi city.

Data source : Fousquare API
 Description : By using this api we will get all the venues in each neighborhood.

1.0.3 Approach

- Collect the new delhi city data from Zomato kaggel dataset
- Using FourSquare API we will find all venues for each neighborhood.
- Filter out all venues that are nearby by locality.
- Using aggregative rating for each resturant to find the best places.
- Visualize the Ranking of neighborhoods using folium library(python)

```
In [2]: import pandas as pd
        import numpy as np
        import requests # library to handle requests
        from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe #
        Matplotlib and associated plotting modules
        import matplotlib.cm as cm
        import matplotlib.colors as colors
        # import k-means from clustering stage
        from sklearn.cluster import KMeans

        !conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't
        import folium # map rendering library
        ! pip install geocoder
        import geocoder
```

```
Collecting package metadata (repodata.json): done
Solving environment: done

# All requested packages already installed.

Requirement already satisfied: geocoder in /home/zettadevs/anaconda3/lib/python3.7/site-packag
Requirement already satisfied: requests in /home/zettadevs/anaconda3/lib/python3.7/site-packag
Requirement already satisfied: future in /home/zettadevs/anaconda3/lib/python3.7/site-packages
Requirement already satisfied: ratelim in /home/zettadevs/anaconda3/lib/python3.7/site-package
Requirement already satisfied: click in /home/zettadevs/anaconda3/lib/python3.7/site-packages
Requirement already satisfied: six in /home/zettadevs/anaconda3/lib/python3.7/site-packages (f
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /home/zettadevs/anaconda3/lib/python3.
```

1.1 Read the zomato resturant data from csv file

```
In [18]: df =pd.read_csv('zomato.csv',encoding='ISO-8859-1')
         df_india =df[df['Country Code'] ==1]
         ## NewDelhi
         df_NDLS =df_india[df_india['City'] =='New Delhi']
```

```
2       Average      45
3       Average      11
4       Average     238

[5 rows x 21 columns]
```

1.2 Data Cleaning

remove the unwanted columns and rows from dataset

```
In [32]: df_Res= df_NDLS[df_NDLS.Longitude !=0.000000][['Restaurant Name','Locality','Longitud
```

```
In [344]: df_Res =df_Res[df_Res['Aggregate rating'] !=0.0]
```

```
In [358]: df_Res.head()
```

Out[358]:

	Restaurant Name	Locality	Longitude	Latitude	\
1	Burger.in	Adchini	77.196923	28.535382	
2	Days of the Raj	Adchini	77.197475	28.535493	
3	Dilli Ka Dhaba	Adchini	77.198033	28.537547	
4	Govardhan	Adchini	77.196924	28.535523	
5	Mezbaan Grills	Adchini	77.198122	28.538134	

	Cuisines	Aggregate rating	Rating text	Votes	\
1	Fast Food	3.2	Average	46	
2	North Indian, Seafood, Continental	3.4	Average	45	
3	South Indian, North Indian	2.6	Average	11	
4	South Indian, North Indian, Chinese	3.4	Average	238	
5	Mughlai	3.1	Average	8	

	Cluster
1	0
2	0
3	0
4	0
5	0

1.2.1 created map to show the restaurant cluters

```
In [346]: New_Delhi_Rest =folium.Map(location=[28.52, 77.25], zoom_start=12)

          X= df_Res['Latitude'] Y
          =df_Res['Longitude']
          Z=np.stack((X, Y), axis=1)

          kmeans=KMeans(n_clusters=5, random_state=0).fit(Z)

          clusters =kmeans.labels_
          colors =['red', 'green', 'blue', 'yellow','orange']
```

```
df_Res['Cluster'] = clusters

for latitude, longitude, Locality, cluster in zip(df_Res['Latitude'], df_Res['Longit
    label =folium.Popup(Locality, parse_html=True)
    folium.CircleMarker(
        [latitude, longitude],
        radius=5,
        popup=label,
        color=' black' ,
        fill=True,
        fill_color=colors[cluster],
        fill_opacity=0.7).add_to(New_Delhi_Rest)

New_Delhi_Rest
```

```
/home/zettadevs/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:11: SettingWithCop
Avalue is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] =value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.htm #
  This is added back by InteractiveShellApp.init_path()
```

Out[346]: <folium.folium.Map at 0x7f0ed56960f0>

In [461]:
```
df_Res.head()
```

Out[461]:

	Restaurant Name	Locality	Longitude	Latitude	\
1	Burger.in	Adchini	77.196923	28.535382	
2	Days of the Raj	Adchini	77.197475	28.535493	
3	Dilli Ka Dhaba	Adchini	77.198033	28.537547	
4	Govardhan	Adchini	77.196924	28.535523	
5	Mezbaan Grills	Adchini	77.198122	28.538134	

	Cuisines	Aggregate rating	Rating text	Votes	\
1	Fast Food	3.2	Average	46	
2	North Indian, Seafood, Continental	3.4	Average	45	
3	South Indian, North Indian	2.6	Average	11	
4	South Indian, North Indian, Chinese	3.4	Average	238	
5	Mughlai	3.1	Average	8	

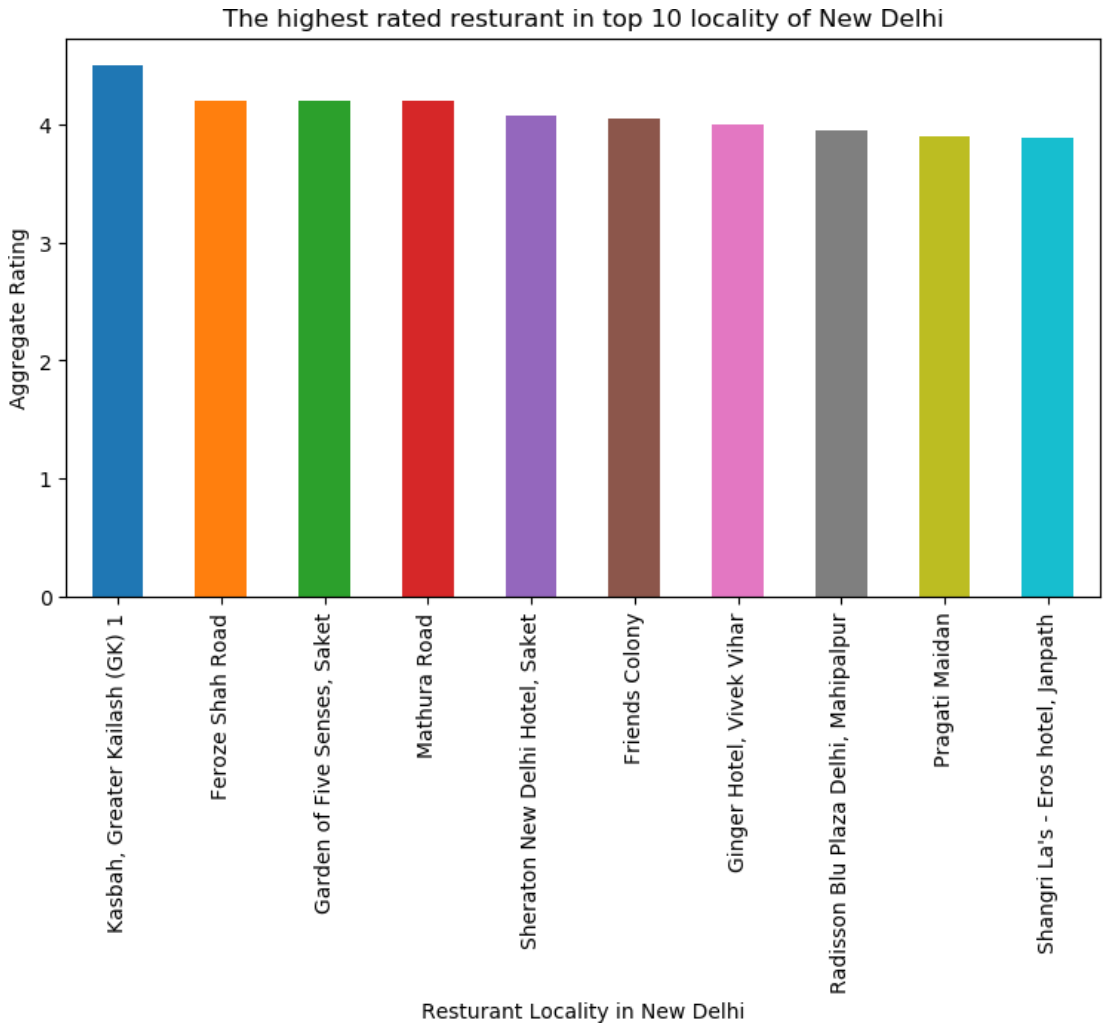
	Cluster
1	0
2	0
3	0
4	0
5	0

1.3 What places are have best restaurant in New Delhi?

In [575]:
```
import matplotlib.pyplot as plt
plt.figure(figsize=(9,5), dpi =100)
# title
plt.title('The highest rated resturant in top 10 locality of NewDelhi')
#Onx-axis

#giving a bar plot
df_Res.groupby('Locality')['Aggregate rating'].mean().nlargest(10).plot(kind='bar')

plt.xlabel('Resturant Locality in NewDelhi')
#Ony-axis
plt.ylabel('Aggregate Rating')
#displays the plot
plt.show()
```



In [468]:
```
df_final =df_final[df_final['Aggregate rating'] != 0.000000]
df_final.columns =['Locality','Lat','Lng', 'No_of_Restaurant','Cusines', 'Agg_Rating
df_final.head()
```

Out[468]:

	Locality	Lat	Lng	No_of_Restaurant	\
0	ARSSMall, PaschimVihar	28.668945	77.101544	1	
1	Adchini	28.537063	77.197808	13	
2	Aditya MegaMall, Karkardooma	28.656131	77.301266	4	
3	Aerocity	28.553077	77.104270	2	
4	Aggarwal City Mall, Pitampura	28.690020	77.134650	3	

	Cusines	Agg_Rating	\
0	North Indian, South Indian, Chinese, Mithai, F...	3.100000	
1	Fast Food, North Indian, Seafood, Continental,...	3.292308	
2	Finger Food, North Indian, Mughlai, Pizza, Fas...	3.275000	
3	Fast Food, Italian, Pizza, North Indian, Conti...	3.200000	
4	North Indian, Chinese, Street Food, Mithai, No...	3.033333	

	Comments	No_of_Votes
0	Average	117
1	Average, Good, Poor, Very Good	1560
2	Average, Good	434
3	Average	59
4	Average	126

In [469]:
```
df_final.shape
```

Out[469]: (240, 8)

1.9 Define Foursquare Credentials and Version

In [593]:
```
## Define Foursquare Credentials and Version
CLIENT_ID ='ClientId' # Foursquare ID
CLIENT_SECRET ='SecretID' # Foursquare Secret
VERSION ='20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' +'XXXXXXXXXXXXXXXXXXXXXXX')
print('CLIENT_SECRET:' + 'XXXXXXXXXXXXXXXXXXX')
```

```
Your credentails:
CLIENT_ID: XXXXXXXXXXXXXXXXXXXXXX
CLIENT_SECRET:XXXXXXXXXXXXXXXXXXX
```

1.10 create a function to repeat the same process to all the Locality in New Delhi

In [484]:
```
## create a function to repeat the same process to all the Locality in NewDelhi
```

```
def getNearbyVenues(names, latitudes, longitudes, radius=500,LIMIT = 100):

    venues_list=[]
    for name,lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url ='https://api.foursquare.com/v2/venues/explore?&client_id={}&client_sec
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results =requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues =pd.DataFrame([item for venue_list in venues_list for item in ven
    nearby_venues.columns =['Locality',
                    'Locality Latitude',
                    'Locality Longitude',
                    'Venue',
                    'Venue Latitude',
                    'Venue Longitude',
                    'Venue Category']

    return(nearby_venues)
```

1.11 find the venues in all New Delhi Locality

In [485]:
```
# find the venues in all NewDelhi Locality
new_Delhi_venues = getNearbyVenues(names=df_final['Locality'],
                                   latitudes=df_final['Lat'],
                                   longitudes=df_final['Lng']
                                  )
```

In [491]:
```
New_Delhi_grouped.shape
```

Out[491]: (239, 216)

In [492]:
```
## print each Locality along with the top 5 most common venues

num_top_venues = 5

for hood in New_Delhi_grouped['Locality']:
    print("----"+hood+"----")
    temp = New_Delhi_grouped[New_Delhi_grouped['Locality'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_

    print('\n')
```

```
        return row_categories_sorted.index.values[0:num_top_venues]
```

In [494]:
```
## create the new dataframe and display the top 10 venues for each Locality.

num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Locality']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
Locality_venues_sorted = pd.DataFrame(columns=columns)
Locality_venues_sorted['Locality'] = New_Delhi_grouped['Locality']

for ind in np.arange(New_Delhi_grouped.shape[0]):
    Locality_venues_sorted.iloc[ind, 1:] = return_most_common_venues(New_Delhi_group

Locality_venues_sorted
```

Out[494]:

	Locality	1st Most Common Venue	\
0	ARSS Mall, Paschim Vihar	Indian Restaurant	
1	Adchini	Café	
2	Aditya Mega Mall, Karkardooma	Pizza Place	
3	Aerocity	Hotel	
4	Aggarwal City Mall, Pitampura	Indian Restaurant	
5	Aggarwal City Plaza, Rohini	Fast Food Restaurant	
6	Alaknanda	Chinese Restaurant	
7	Ambience Mall, Vasant Kunj	Coffee Shop	
8	Anand Lok	Café	
9	Anand Vihar	Café	
10	Andaz Delhi, Aerocity	Hotel	
11	Ansal Plaza Mall, Khel Gaon Marg	Performing Arts Venue	
12	Asaf Ali Road	Indian Restaurant	
13	Ashok Vihar Phase 1	Indian Restaurant	
14	Ashok Vihar Phase 2	Pizza Place	
15	Ashok Vihar Phase 3	Pizza Place	
16	Barakhamba Road	Indian Restaurant	
17	Basant Lok Market, Vasant Vihar	Café	
18	Bellagio, Ashok Vihar Phase 2	Pizza Place	
19	Best Western Taurus Hotel, Mahipalpur	Hotel	
20	Bhikaji Cama Place	Lounge	
21	Chanakyapuri	Café	

```
New_Delhi_merged = df_final.head(239)
New_Delhi_merged['Cluster Labels'] = kmeans.labels_

# merge New_Delhi_grouped with df_Chinese to add latitude/longitude for each Localit
New_Delhi_merged = New_Delhi_merged.join(Locality_venues_sorted.set_index('Locality'

New_Delhi_merged.head()
```

```
/home/zettadevs/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:3: SettingWithCopy
Avalue is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.htm
  This is separate from the ipykernel package so we can avoid doing imports until
```

Out[496]:

	Locality	Lat	Lng	No_of_Restaurant
0	ARSSMall, PaschimVihar	28.668945	77.101544	1
1	Adchini	28.537063	77.197808	13
2	Aditya MegaMall, Karkardooma	28.656131	77.301266	4
3	Aerocity	28.553077	77.104270	2
4	Aggarwal City Mall, Pitampura	28.690020	77.134650	3

	Cusines	Agg_Rating
0	North Indian, South Indian, Chinese, Mithai, F...	3.100000
1	Fast Food, North Indian, Seafood, Continental,...	3.292308
2	Finger Food, North Indian, Mughlai, Pizza, Fas...	3.275000
3	Fast Food, Italian, Pizza, North Indian, Conti...	3.200000
4	North Indian, Chinese, Street Food, Mithai, No...	3.033333

	Comments	No_of_Votes	Cluster Labels
0	Average	117	0
1	Average, Good, Poor, Very Good	1560	3
2	Average, Good	434	2
3	Average	59	1
4	Average	126	0

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Indian Restaurant	Sandwich Place	Market
1	Café	Pub	Indian Restaurant
2	Pizza Place	Indian Restaurant	Shopping Mall
3	Hotel	Rental Car Location	Airport Lounge
4	Indian Restaurant	Department Store	Food

	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Multicuisine Indian Restaurant	Pizza Place	Coffee Shop
1	French Restaurant	Food Truck	Food Court
2	Multiplex	Café	Hotel

```
3                  Airport Terminal            Coffee Shop  Fast Food Restaurant
4                      Pizza Place     African Restaurant                 Airport

  7th Most Common Venue 8th Most Common Venue 9th Most Common Venue  \
0     Chinese Restaurant                   ATM       Airport Service
1      Food &Drink Shop                  Food          Flower Shop
2            Food Court       Food &Drink Shop                  Food
3     Falafel Restaurant      French Restaurant             Food Truck
4     Frozen Yogurt Shop    Fried Chicken Joint     French Restaurant

  10th Most Common Venue
0       Fish &Chips Shop
1              Flea Market
2              Flower Shop
3              Food Court
4              Food Truck
```

In [498]:
```
# create final map
map_clusters =folium.Map(location=[latitude, longitude], zoom_start=10)

# set color schemefor the clusters
x =np.arange(kclusters)
ys =[i+x+(i*x)**2 for i in range(kclusters)]
#colors_array =cm.rainbow(np.linspace(0, 1, len(ys)))
#rainbow =[colors.rgb2hex(i) for i in colors_array]
colors =['red', 'green', 'blue', 'yellow','orange']

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(New_Delhi_merged['Lat'], New_Delhi_merged['Lng'],
    label =folium.Popup(str(poi) +' Cluster ' +str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color='black',
        fill=True,
        fill_color=colors[cluster],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[498]: <folium.folium.Map at 0x7f0ef434ee10>

In [499]:
```
## Examine Clusters

## Cluster 1

New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] ==0, New_Delhi_merged.colum
```

79	Fish & Chips Shop
82	Food
86	Food
91	Movie Theater
100	Gym / Fitness Center
102	Metro Station
108	Food Truck
123	Food Truck
127	Food
143	Diner
159	Flower Shop
171	Food Court
172	Donut Shop
175	Coffee Shop
180	Flower Shop
181	Discount Store
182	Dessert Shop
191	Fast Food Restaurant
201	Gym / Fitness Center
203	Market
208	Nightclub
214	Coffee Shop
219	Food Court
221	Food & Drink Shop
226	Food Truck
227	Nightlife Spot
228	Flea Market
229	Pet Store
230	Food Court

In [503]:
```
## Examine Clusters

## Cluster 5
New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] ==4, New_Delhi_merged.colum
```

Out[503]:

	Lat	Agg_Rating	Comments \
5	28.700516	3.040000	Average, Good, Poor
6	28.527088	3.117391	Average, Good, Poor
7	28.541298	3.425000	Average, Good, Very Good
11	28.562580	3.750000	Average, Good, Very Good
20	28.568193	2.755556	Average, Poor
24	28.649658	3.800000	Average, Excellent, Good, Very Good
27	28.716874	3.400000	Average
29	28.681233	3.300000	Average, Good, Very Good
31	28.632091	3.779832	Average, Excellent, Good, Very Good
34	28.720602	3.600000	Good
37	28.541903	3.000000	Average
38	28.702961	3.225000	Average, Good

184	American Restaurant	Hotel	Multiplex
185	Food Truck	Food Court	Food &Drink Shop
186	Fried Chicken Joint	French Restaurant	Food Truck
187	Food Truck	Food Court	Food &Drink Shop
188	French Restaurant	Food Truck	Food Court
192	Shopping Mall	Multiplex	Hotel
195	Hotel	Donut Shop	Burger Joint
196	Fast Food Restaurant	Snack Place	Miscellaneous Shop
198	Food Court	Food &Drink Shop	Food
200	Breakfast Spot	Airport	Dumpling Restaurant
202	Food Truck	Garden	Park
206	Restaurant	Beer Garden	Bar
207	Chinese Restaurant	Lounge	Beer Garden
212	BBQJoint	Arts &Crafts Store	Light Rail Station
216	Spa	Mediterranean Restaurant	Bistro
218	Hotel	Indian Restaurant	Australian Restaurant
222	Indian Restaurant	Donut Shop	Discount Store
224	Multiplex	Food &Drink Shop	Food Court
225	Food Truck	Food Court	Food &Drink Shop
232	French Restaurant	Food Truck	Food Court
234	Coffee Shop	Hotel	Pizza Place
237	Café	Lounge	Bakery

1.11.1 Conclusion

- Chanakyapuri, Pitampura, Safdarjung are some of the best neighborhoods for Chinese cuisine.
- Pancsheel park, Nehru place have the best Chinese Resturant.
- Cannaught place, Rajouri garden, Malviya nagar are the best places for edible person.
- Greater kailash, Feroze shah road, Saket have best resturants in New Delhi. #### Cluster 1: It is most recommended for Indian Restaurants. #### Cluster 2: It is most recommended for Hotels and nightclub. #### Cluster 3 and Cluster 5: It is most recommended for Fast food. ####Cluster 4: It is most recommended for the cafe and pizza.

In [ ]: