Analyzing City Specific Data to Predict Hotel Density around Airports

PREPARED BY: ZAKI ANSARI

Problem Statement

A travel agency which mainly deals with hotel bookings has multiple locations across the United States is expanding. The management team is looking into the options and trying to decide which city would be best suitable to target for business opportunities.
Good starting point is the explore the cities which have following characteristics:

- City is a commercial hub
- City has tourist attractions
- City has airports
- City has big volume of business and leisure travelers

Data Used

- Annual enplanement information of Top 30 US airports
  - Scraped from Wikipedia
- GDP information corresponding to cities with high annual enplanements
  - Scraped from Wikipedia
- Number of Hotels/Motels/Resorts around the airport.
  - Fetched using Foursquare API

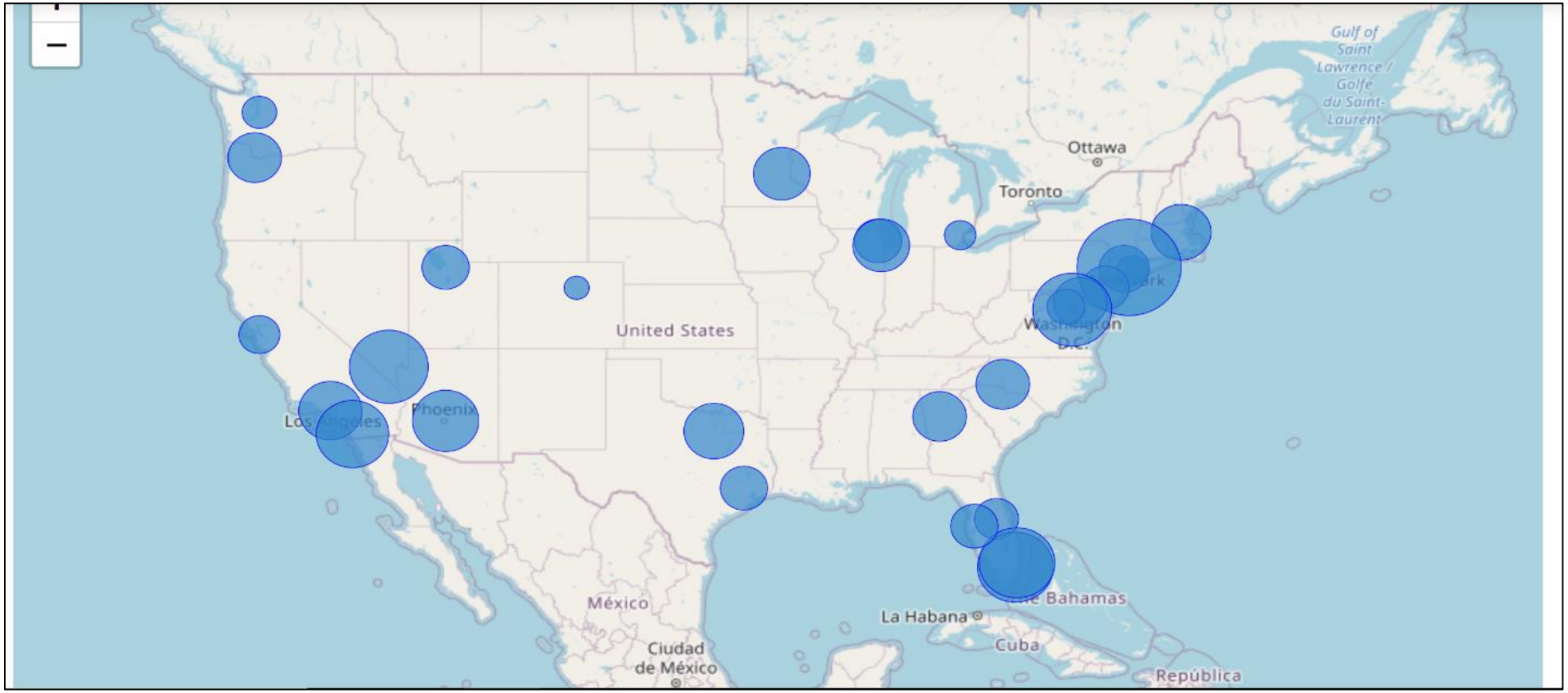
Top Rows of Data Frame Used in Analysis

```
gdp_airport = gdp_airport.join(top_airports['IATACode'])
gdp_airport
```

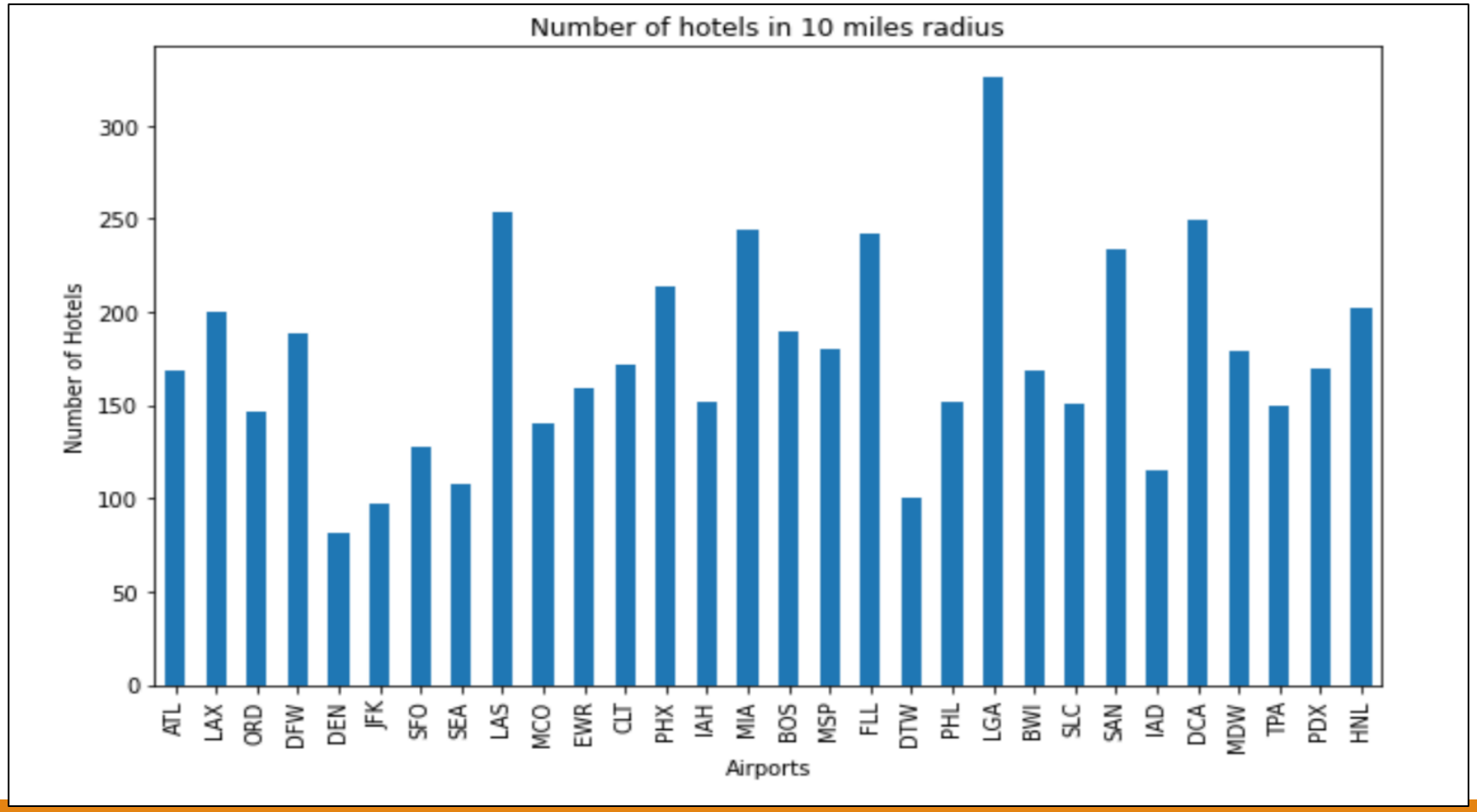
Out[24]:

	Airport	City	Area	AirTraffic	GDP	Latitude	Longitude	NumHotels	IATACode
0	Hartsfield–Jackson Atlanta International Airport	Atlanta	Atlanta-Sandy Springs-Roswell, GA (Metropolita...	51866464	397261	33.637799	-84.429271	169	ATL
1	Los Angeles International Airport	Los Angeles	Los Angeles-Long Beach-Anaheim, CA (Metropolit...	42626783	1047661	33.942168	-118.421376	200	LAX
2	O'Hare International Airport	Chicago	Chicago-Naperville-Elgin, IL-IN-WI (Metropolit...	39874879	689464	41.977985	-87.909321	146	ORD
3	Dallas/Fort Worth International Airport	Dallas	Dallas-Fort Worth-Arlington, TX (Metropolitan ...	32800721	512509	32.896519	-97.046522	188	DFW
4	Denver International Airport	Denver	Denver-Aurora-Lakewood, CO (Metropolitan Stati...	31363573	214157	39.850188	-104.675328	81	DEN
5	John F. Kennedy International Airport	New York	New York-Newark-Jersey City, NY-NJ-PA (Metropo...	30541459	1772319	40.642948	-73.779373	97	JFK
6	San Francisco International Airport	San Francisco	San Francisco-Oakland-Berkeley, CA (Metropolit...	27794154	548613	37.622452	-122.384072	128	SFO
7	Seattle–Tacoma International Airport	Seattle	Seattle-Tacoma-Bellevue, WA (Metropolitan Stat...	24894338	392036	47.447567	-122.308016	108	SEA
8	[illegible]	[illegible]	Las Vegas-Henderson-Paradise, NV	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Airport Vs Number of Hotels in 10 Miles Radius on a Map

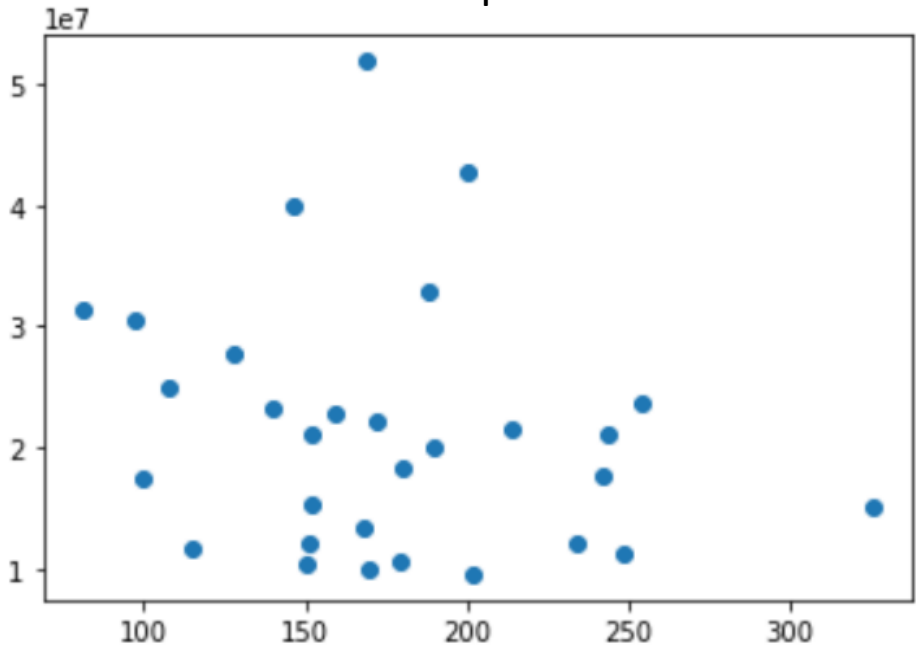


Airport Vs Number of Hotels in 10 Miles Radius – Bar Plot

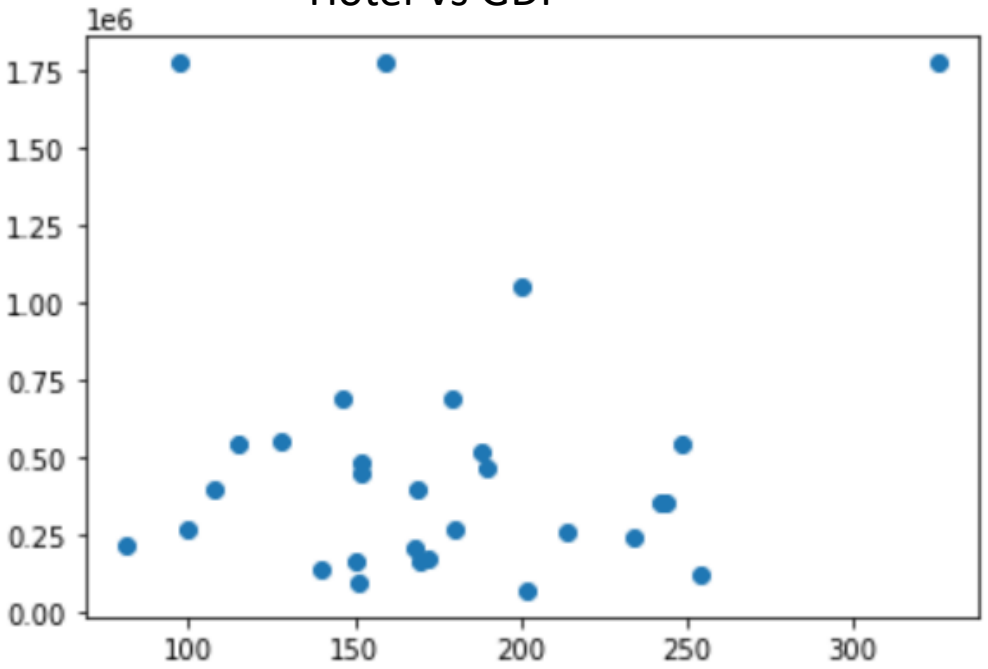


Building Multi Linear Regression Model – Scatter Plots of Independent Vs. Dependent Variables

Hotel Vs Enplanements



Hotel Vs GDP



Building Multi Linear Regression Model

```
X = pd.DataFrame([gdp_airport['GDP'], gdp_airport['AirTraffic']])
X = X.T
y = pd.DataFrame([gdp_airport['NumHotels']])
y = y.T

lm = LinearRegression()
lm.fit(X,y)
lm.score(X,y)
```

```
2]: 0.07478136274087 21
```

Results

It can be seen that R^2 value is pretty low in this case which confirms that the relation between dependent and independent variables is not linear at all.

The analysis of Top-30 US airports shows that GDP & Annual Enplanements is not a good measure for desired prediction. This could be explained by following reasons:

- In some cities the nearest major airports are within the city limits and in others they are in outskirts. If airport is in outskirts then hotels are probably going to be far from the airport even if the airport is busy and city has strong GDP.
- The high enplanement rates could also mean that the airport is only serving as a hub, and a lot of passenger traffic is only connecting passengers.
- Sometimes a single city can have multiple airports. In our data set New York has two major airports; John F. Kennedy (JFK) and La Guardia (LGA). JFK has high annual enplanements as compared to LGA but Foursquare data shows that LGA has 326 hotels around it while JFK only has 97.

Conclusion

The travel agency should explore other criteria like number of big companies in the area, employment rates of the city, GDP per capita, major tourist attractions, number of universities, conference centers etc. to be used as good predictors to estimate hotel density around any airport.