Siemens India Office Locations Analysis (using Machine Learning)

Introduction

Siemens Limited has a vast presence in India with its sales offices being located in 17 prominent cities of the country. Being an employee of Siemens, it was quite tempting for me to do my final Capstone Project of IBM Data Science Specialization on a topic which is related to Siemens Ltd. in one way or the other. Since I am a part of the Strategic Sales department of the organization, interacting with the 17 offices on day to day basis is a crucial aspect of the job. Naturally, I wanted to walk the last leg of the Data Science journey by performing a location analysis on the 17 locations.

The objective of this exercise was to obtain the geographical coordinates of the 17 Siemens sales offices, and identify the venues in a radius of 2 kilometers around these locations using the **Foursquare API**. Based on the information obtained, I used a **Machine Learning** algorithm to draw interesting parallels between the 17 sales offices. Not only that, I selected an office which has the highest amount of venues and analyzed it.

Data

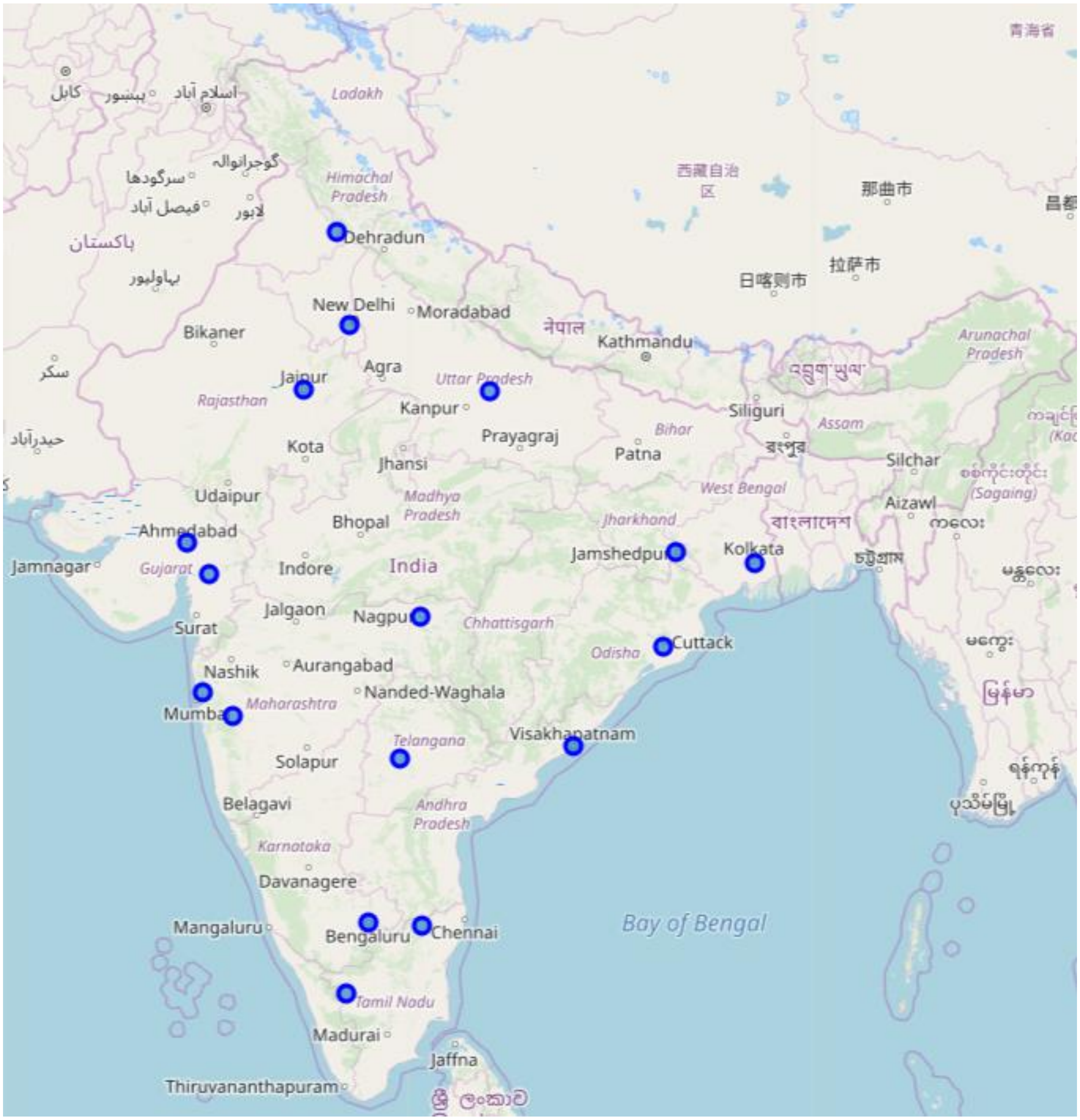
The most crucial part of any data science assignment is....well, data. For this data science project, I began with a simple CSV file which contained the name of the cities and the places where Siemens offices are located,

	Office City	Office Location
0	Bangalore	St Marks Road, Bangalore
1	Hyderabad	Saifabad, Hyderabad
2	Visakhapatnam	Visakhapatnam
3	Chennai	Anna Salai
4	Coimbatore	Avarampalayam
5	Mumbai	Airoli
6	Ahmedabad	Navrangpura
7	Vadodara	Makarpura
8	Pune	ICC Trade Tower, Pune
9	Nagpur	Ramdaspeth, Nagpur
10	Delhi	Sector 18, Gurgaon
11	Chandigarh	Sector 17c, Chandigarh
12	Jaipur	Ashok Nagar Jaipur
13	Lucknow	Gomti Nagar
14	Kolkata	Acropolis Mall, Kolkata
15	Bhubaneswar	Patia, Bhubaneswar
16	Jamshedpur	Bistupur

Of course this isn't sufficient to perform the kind of analysis that I set out to perform. With the help of Python libraries I identified the geographical coordinates of all the locations. I could have googled the same for sure, but then what's the fun in that? It loses its sense of **automation**, doesn't it? With the basic dataset in place, I continued with the analysis.

Methodology

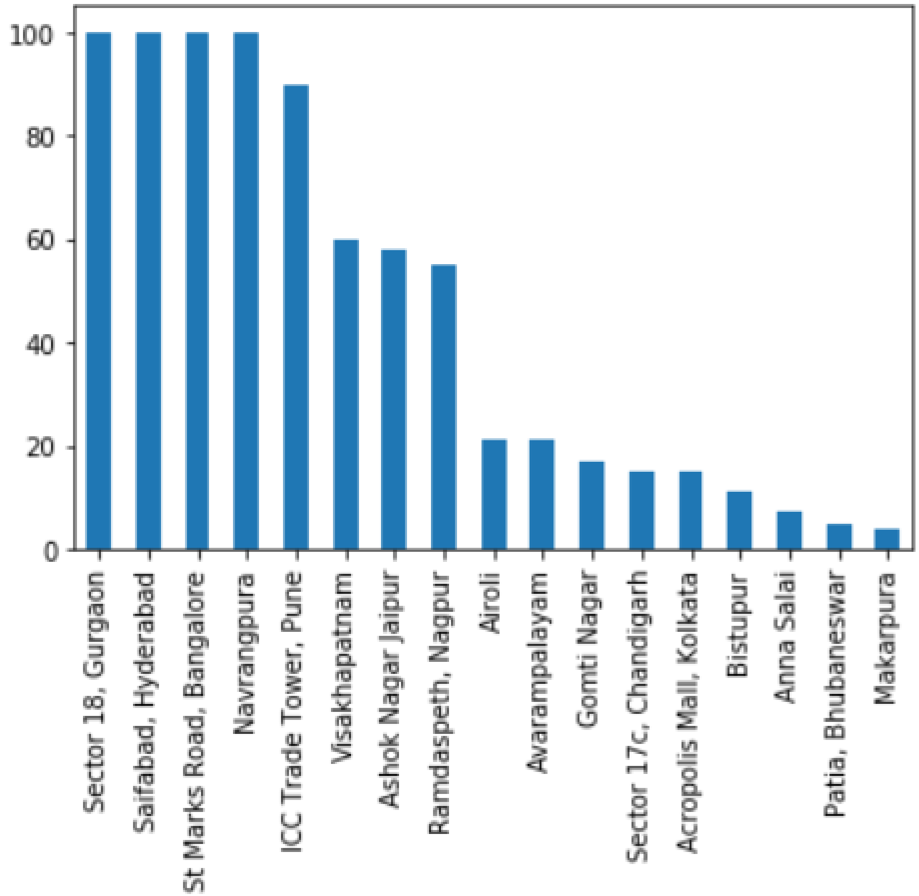
The first step of the analysis was locating these 17 offices on a map of India. Using Nominatim and Folim libraries in **Python**, I obtained the following result:



Since all the locations were mapped properly, I proceeded to the next phase of the analysis. It was time to utilize the Foursquare API to get an idea about the venues (restaurants, cafes, ATMs, hotels, airports, schools, etc.) in a radius of 2 kilometers around each of these 17 offices.

	Office Location	Office Location Latitude	Office Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	St Marks Road, Bangalore	12.977906	77.602201	Hard Rock Cafe Bengaluru	12.976389	77.601468	American Restaurant
1	St Marks Road, Bangalore	12.977906	77.602201	M. Chinnaswamy Stadium	12.978144	77.599223	Cricket Ground
2	St Marks Road, Bangalore	12.977906	77.602201	Church Street Social	12.975559	77.602579	Pub
3	St Marks Road, Bangalore	12.977906	77.602201	M.G Road Boulevard	12.975771	77.603979	Plaza
4	St Marks Road, Bangalore	12.977906	77.602201	Blossom Book House	12.975042	77.604813	Bookstore

The table shown above gives a glimpse of the sample data obtained from the Foursquare API. The complete dataset obtained contained 769 rows. That's is a lot of venues!! To have a quick look at the number of locations that are around each office, I created a bar chart:



It is quite clear that certain office locations like Gurgaon, Hyderabad, Bangalore, and Ahmedabad have umpteen venues around them (>=100) whereas, locations like Baroda, Bhubaneswar, Chennai, etc. don't seem to have many venues around them. If you are curious to know which office I belong to, it is Airoli (Mumbai). Around 143 categories of venues were identified by the Foursquare API around each location. These were:

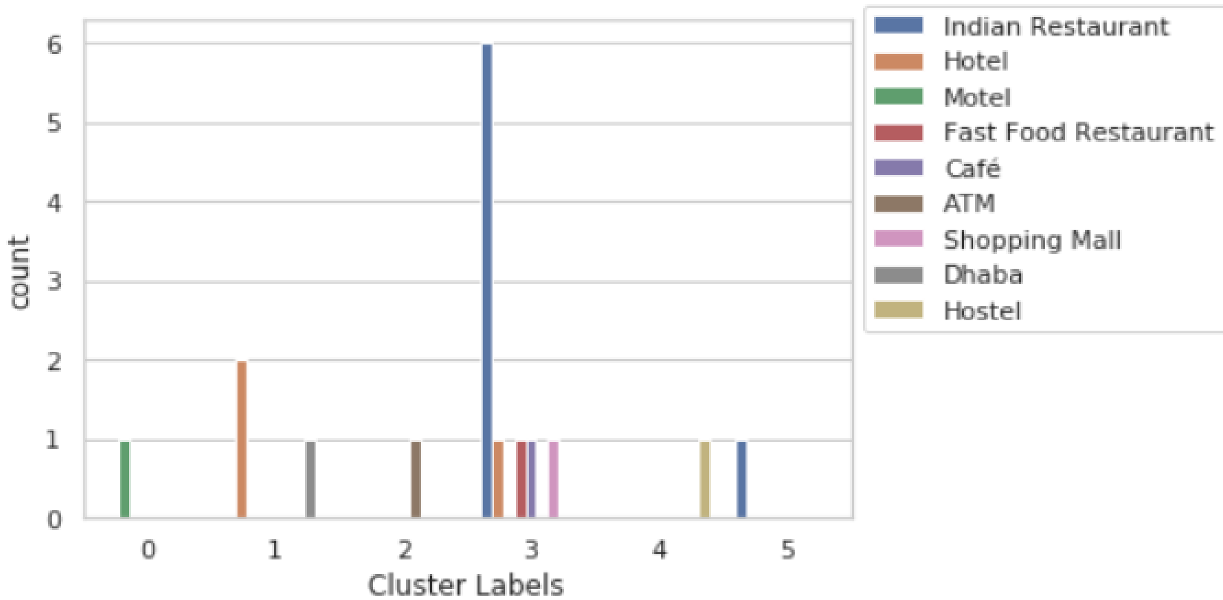
```
['Office Locatio
 'Afghan Restaur
 'Airport',
 'American Resta
 'Andhra Restaur
 'Arcade',
 'Art Gallery',
 'Art Museum',
 'Arts & Crafts
 'Asian Restaura
 'Athletics & Sp
 'BBQ Joint',
 'Bakery',
 'Bar',
 'Bed & Breakfas
 'Bistro',
 'Bookstore',
 'Bowling Alley'
 'Breakfast Spot
 'Brewery',
 'Bubble Tea Sho
 'Burger Joint',
 'Burmese Restau
 'Bus Station',
 'Café',
 'Cantonese Rest
 'Chaat Place',
 'Chinese Restau
 'Clothing Store
 'Cocktail Bar',
 'Coffee Shop',
 'Concert Hall',
 'Convenience St
 'Cricket Ground
 'Convenience St
 'Cricket Ground
 'Cupcake Shop',
 'Dance Studio',
 'Deli / Bodega'
 'Department Sto
 'Dessert Shop',
 'Dhaba',
 'Diner',
 'Donut Shop',
 'Dumpling Resta
 'Eastern Europe
 'Electronics St
 'Fabric Shop',
 'Falafel Restau
 'Farmers Market
 'Fast Food Rest
 'Food Court',
 'Food Truck',
 'French Restaur
 'Fried Chicken
 'Furniture / Ho
 'Gaming Cafe',
 'Garden',
 'Gastropub',
 'Gym',
 'Gym / Fitness
 'Historic Site'
 'Hookah Bar',
 'Hostel',
 'Hotel',
 'Hotel Bar',
 'Hyderabadi Res
 'Ice Cream Shop
 'Indie Movie The
 'Indie Theater',
 'Italian Restaur
 'Japanese Resta
 'Jewelry Store',
 'Juice Bar',
 'Korean Restaura
 'Lake',
 'Lounge',
 'Maharashtrian F
 'Market',
 'Mediterranean F
 "Men's Store",
 'Mexican Restaur
 'Middle Eastern
 'Miscellaneous S
 'Mobile Phone Sh
 'Moroccan Restau
 'Motel',
 'Motorcycle Shop
 'Mughlai Restaur
 'Multicuisine In
 'Multiplex',
 'Museum',
 'Neighborhood',
 'New American Re
 'Nightclub',
 'Optical Shop',
 'Outdoors & Recr
 'Park',
 'Performing Arts
 'Pizza Place',
 'Platform',
 'Plaza',
 'Recreation Center',
 'Residential Building (
 'Rest Area',
 'Restaurant',
 'Road',
 'Sandwich Place',
 'Scenic Lookout',
 'Science Museum',
 'Seafood Restaurant',
 'Shoe Store',
 'Shop & Service',
 'Shopping Mall',
 'Smoke Shop',
 'Snack Place',
 'Soccer Stadium',
 'South Indian Restauran
 'Spa',
 'Stadium',
 'Steakhouse',
 'Street Food Gathering'
 'Sushi Restaurant',
 'Tapas Restaurant',
 'Tea Room',
 'Theater',
 'Theme Park',
 'Toy / Game Store',
 'Trail',
 'Train Station',
 'Turkish Restaurant',
 'Vegetarian / Vegan Res
 'Wine Bar',
 "Women's Store",
 'Zoo']
```

To make things simpler and easier to understand, I selected the top 10 most frequently occurring venues in each office location:

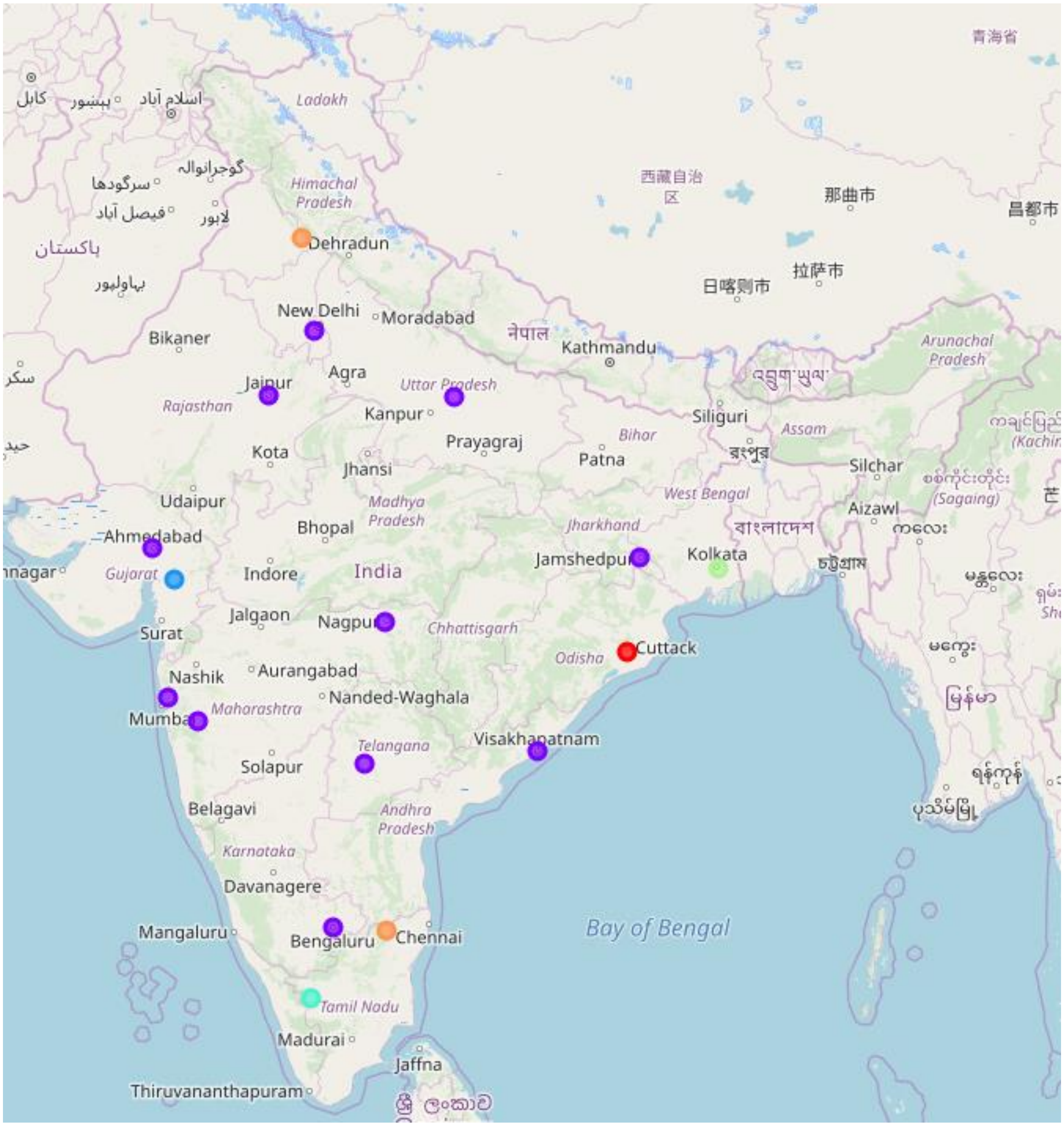
	Office Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Acropolis Mall, Kolkata	Dhaba	Hotel	Shopping Mall	Dessert Shop	Fried Chicken Joint	Mughlai Restaurant	Café	Multiplex	Sandwich Place	Chinese Restaurant
1	Airoli	Fast Food Restaurant	Gym	Platform	Burger Joint	Café	Department Store	Sandwich Place	Restaurant	Chinese Restaurant	Cocktail Bar
2	Anna Salai	Motel	Ice Cream Shop	Indian Restaurant	Pizza Place	Café	Motorcycle Shop	Supermarket	Farmers Market	French Restaurant	Food Court
3	Ashok Nagar Jaipur	Hotel	Indian Restaurant	Historic Site	Market	Coffee Shop	Café	Fast Food Restaurant	Zoo	Ice Cream Shop	Park
4	Avarampalayam	Indian Restaurant	Bakery	Fried Chicken Joint	Farmers Market	Ice Cream Shop	Vegetarian / Vegan Restaurant	Pizza Place	Coffee Shop	Tea Room	Motel
5	Bistupur	Hotel	Indian Restaurant	Bakery	Restaurant	Sandwich Place	Café	Market	Pizza Place	Italian Restaurant	Hookah Bar
6	Gomti Nagar	Indian Restaurant	Hotel	Café	Fast Food Restaurant	Shopping Mall	Bakery	Department Store	Pizza Place	Plaza	Clothing Store
7	ICC Trade Tower, Pune	Indian Restaurant	Coffee Shop	Lounge	Restaurant	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Café	Snack Place	Asian Restaurant
8	Makarpura	ATM	Restaurant	Tapas Restaurant	Electronics Store	Dhaba	Diner	Donut Shop	Dessert Shop	Eastern European Restaurant	Fruit & Vegetable Store
9	Navrangpura	Café	Indian Restaurant	Hotel	Fast Food Restaurant	Dessert Shop	Pizza Place	Sandwich Place	Coffee Shop	Tea Room	Asian Restaurant
10	Patia, Bhubaneswar	Hostel	Pizza Place	Café	Snack Place	Dhaba	Diner	Donut Shop	Dessert Shop	Eastern European Restaurant	Fried Chicken Joint
11	Ramdaspeth, Nagpur	Shopping Mall	Coffee Shop	Ice Cream Shop	Indian Restaurant	Snack Place	Restaurant	Clothing Store	Fast Food Restaurant	Café	Italian Restaurant
12	Saifabad, Hyderabad	Indian Restaurant	Hotel	Café	Multiplex	Bakery	Fast Food Restaurant	Chinese Restaurant	Ice Cream Shop	Restaurant	Hotel Bar
13	Sector 17c, Chandigarh	Indian Restaurant	Pizza Place	Hockey Field	Café	Miscellaneous Shop	Sandwich Place	Residential Building (Apartment / Condo)	Shop & Service	Gym	Arcade
14	Sector 18, Gurgaon	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Shopping Mall	Bar	Coffee Shop	Hotel	Asian Restaurant	Donut Shop
15	St Marks Road, Bangalore	Indian Restaurant	Hotel	Lounge	Pub	Ice Cream Shop	Brewery	Chinese Restaurant	Italian Restaurant	Breakfast Spot	Shopping Mall
16	Visakhapatnam	Hotel	Indian Restaurant	Café	Indie Movie Theater	Platform	Restaurant	Train Station	Italian Restaurant	Pizza Place	Movie Theater

Based on the data of 143 categories obtained earlier, I went to perform k means clustering on the data. K means clustering is a type of Machine Learning algorithm which is succinctly explained by Andrey Bu - ***“the objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.”***

I chose the value of k as 6, which means that the algorithm that I have used here will try to identify 6 different clusters in the data. What does this really mean? Well, our 17 office locations will be divided in 6 different groups based on the data of 143 categories of venues received from Foursquare API. But what will be the criteria of this clustering? To be honest, I can't predict until and unless the analysis is performed. Based on certain underlying patterns which are not easily observable to a human, the machine learning algorithm clusters the data. So offices with similarities will be clubbed in a single cluster while the remaining will be put under other clusters. Let's see what really happens:



Based on the data we have obtained till now, our algorithm has classified the 17 offices into 6 clusters. As you can see, 3rd cluster has the maximum number of venues associated with it, whereas the remaining don't seem to have many venues. Let's see the clustering on the map of India:



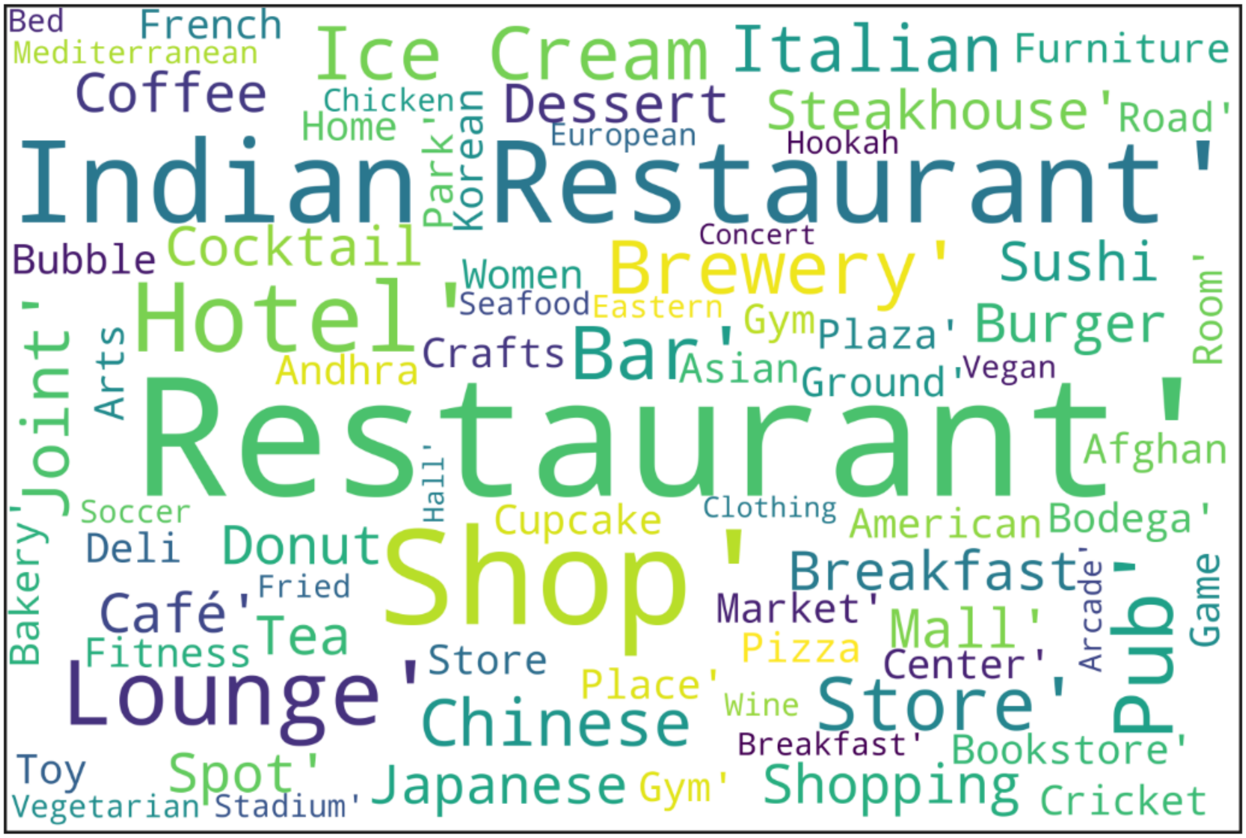
Results

Pretty interesting, isn't it? To save you from the trouble of peering over huge chunks of data, let me just summarize the observations of the clustering algorithm:

1. **Chennai** - It seems to be the only place where the 1st most common venue is a Motel. As a result, it is present in a separate cluster

2. **Jaipur, Kolkata, and Jamshedpur** - If you observe clearly, these 3 offices seem to be in locations where there are venues other than just restaurants and cafes, like - Market, Multiplex, Shopping Mall, Zoo, Park, etc.
3. **Vadodara -** Ah, this is my 2nd favorite cluster! Why, you may ask. Well, it is because this is the only location where the 1st most common venue is an ATM.
4. **Bhubaneswar -** It seems to be the only place where the 1st most common venue is a Hostel. As a result, it is present in a separate cluster
5. **Coimbatore** - It seems to be the only place which has a Farmer's market and a Vegan restaurant. As a result, it is present in a separate cluster
6. **All remaining offices** have a heavy presence of restaurants, diners, and cafes as a result of which they were placed in a separate cluster. This cluster seems to be highly populated and has 10 offices locations in it.

Let's dig deeper into Foursquare data. I chose Bangalore for this analysis since our data clearly shows that it has the maximum number of venues around the office. Wouldn't it be easier to just construct a **Wordcloud** of all the venue categories which are around this office?



Woah!! Clearly Bangalore is all about Restaurants, Breweries and Hotels :)

But a wordcloud wouldn't be sufficient in identifying the locations of these amazing places. So I located these places (50 in number) which are in a radius of 1 km from Siemens office on the map of Bangalore:

I was working on this analysis at around 9:30 am on a Sunday morning which happens to be the breakfast time for me. So I thought I would like to find some trending venues in Bangalore around Siemens' office at this time in the morning. Let's see what happened:

With much anticipation, all I got from the Python code was a:

```
'No trending venues are available at the moment!'
'Cannot generate visual as no trending venues are available at the moment!'
```

Discussion

Our journey began with identifying the geographical coordinates of the 17 sales offices of Siemens India. Then we identified the top 100 venues in a radius of 2 kms around these locations using the Foursquare API. This enabled us to perform the k means clustering so that we can identify the similarities and differences between the offices. It was interesting to see that certain offices which had unique venues like Hostels, Motels, ATMs, etc. were placed in different clusters as compared to the locations where most common venues were Restaurants, Cafes, etc.

From clustering, we moved on to analysing one particular office location - Bangalore. We identified the nearest 50 venues and the most common ones were restaurants, pubs, breweries, etc. which is not surprising as Bangalore is famous for such outlets. However, it was quite interesting to see that there isn't much traction to these venues on a Sunday morning. So is there a business opportunity for a breakfast joint near Bangalore? That's something to ponder over.

Conclusion

It has been quite a long journey from a small 2 column dataframe to identifying nearest 50 venues of an office location. We have identified that there are certain office locations which have as high as 100 venues around them, and unfortunately, there are certain locations which have only 4-5.

One thing that needs to be kept in mind is that no form of machine learning or analysis is fool proof. A lot is left to the way things are interpreted. I sincerely hope that I have done a decent job in collecting data, analysing and dissecting it, and concluding the analysis.

I genuinely thank you for reading this article of mine. I would love to hear from you about your viewpoints on the analysis that I performed. You can access my code from my GitHub profile: https://github.com/eshanrock/Applied-Data-Science-Capstone-Coursera/blob/master/Capstone%20Project_Siemens.ipynb