Capstone Project - State & Capitals

—

Shubham Mahidhar

Overview

Urbanization is important for the benefits it brings and due to increase in commercial activity that creates demand and growth for many more industries. It not only contributes to more economic activities but also generates more employment, revenues, and plays a significant role in development.

Business Problem

1. The benefits of urbanization tend to reflect on the employment opportunity which it gives to the people of that country. The objective of this project is to analyze the different cities of India and try to find the group of cities that have the same amenities, business etc.

2. The target audience for this project include people who are interested in find similarities between cities in India
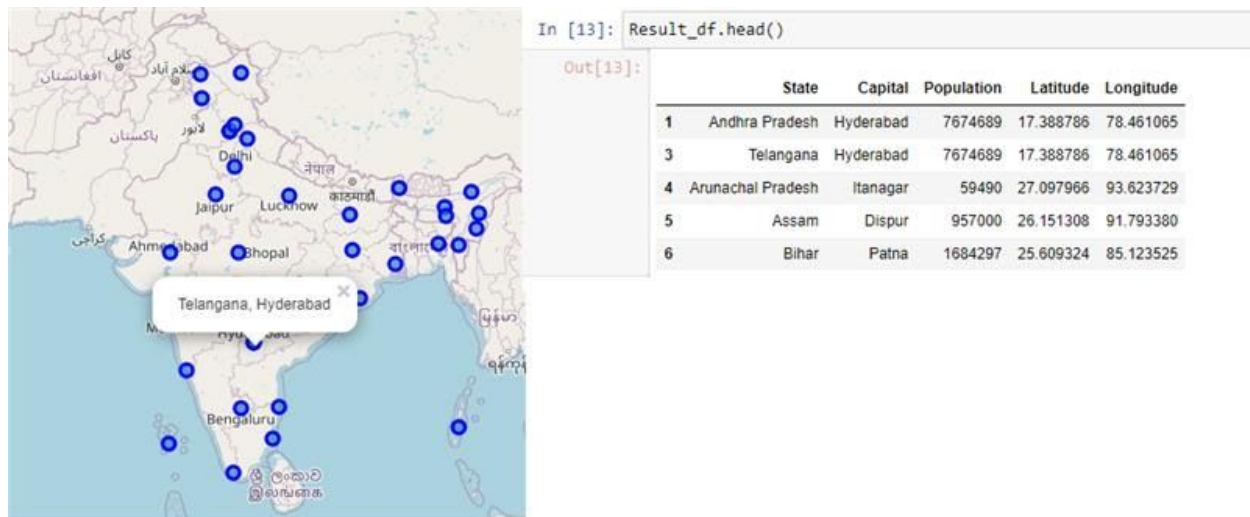
Data Description

To solve the problem we need

1. State and their capitals
2. Latitude and Longitude of all capitals

Python Geopy

To fetch longitude and latitude of each state capital.

```
In [13]: Result_df.head()

Out[13]:
```

	State	Capital	Population	Latitude	Longitude
1	Andhra Pradesh	Hyderabad	7674689	17.388786	78.461065
3	Telangana	Hyderabad	7674689	17.388786	78.461065
4	Arunachal Pradesh	Itanagar	59490	27.097966	93.623729
5	Assam	Dispur	957000	26.151308	91.793380
6	Bihar	Patna	1684297	25.609324	85.123525

The output shows cities plotted on map. Data frame has five columns State, Capital, Population, Latitude and Longitude.

Methodology

I.   Data collection

State , capital and population was gathered and a geopy API was used to fetch latitude and longitude of each capital. Total of 37 capitals along with State and Union territories names constitute data.

Population in each capital
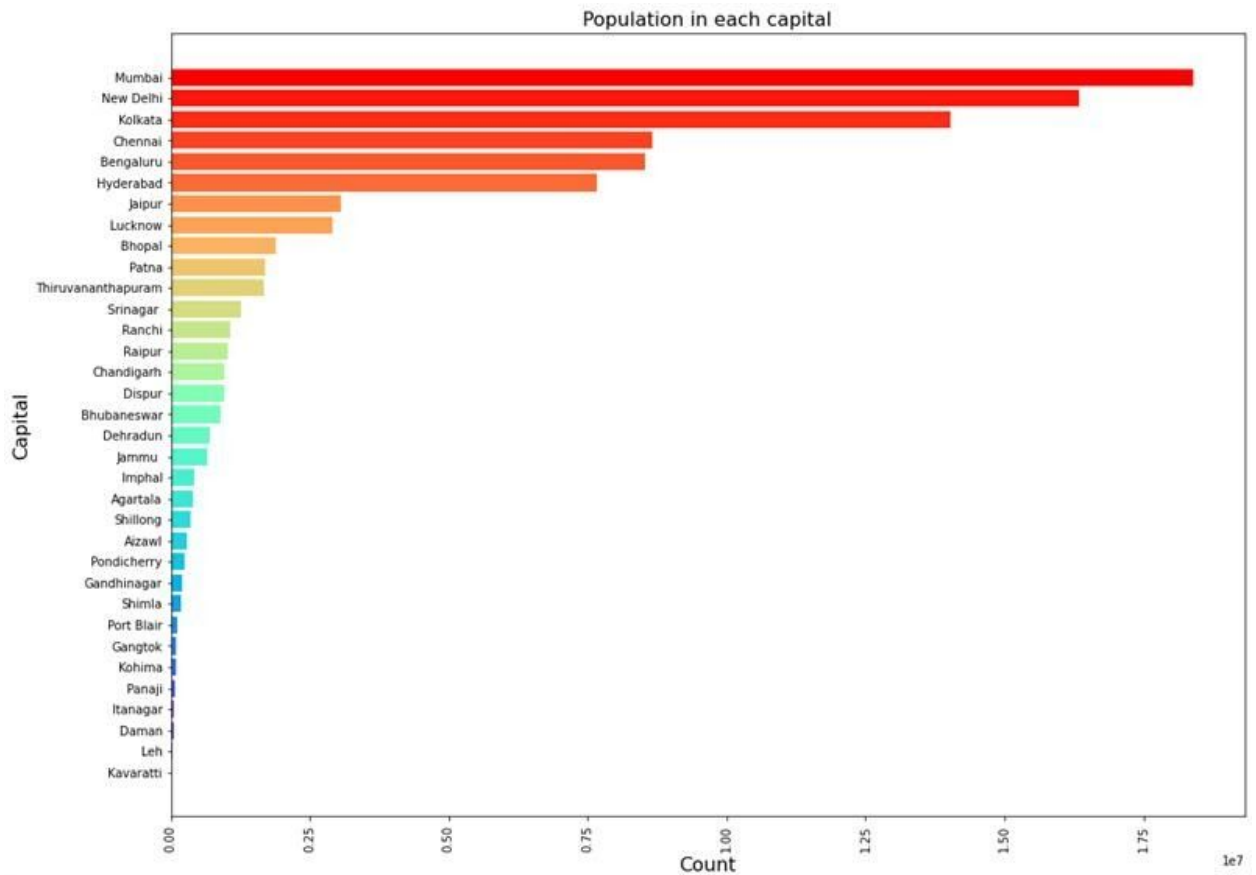
Population chart shows Mumbai,New Delhi and Kolkata are top populous capital.

II.   FourSquare API

Using the FourSquare API we data related to tourist and business spots with 50km radius of capital.

```
tourist_venues.head()
```

Out[24]:

	Capital	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Leh	34.164203	77.584813	Leh Palace	34.166287	77.586390	Palace
11	Leh	34.164203	77.584813	Shey Palace	34.159815	77.633323	Palace
13	Daman	20.420005	72.863763	Daman	20.421682	72.839065	Beach
14	Daman	20.420005	72.863763	The Deltin	20.412848	72.856197	Casino
17	Daman	20.420005	72.863763	Jetty	20.412949	72.832805	Arcade

```
business_venues.head()
```

Out[25]:

|	Capital	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kavaratti	10.559050	72.636569	kavaratti	10.549820	72.626635	Boat or Ferry
1	Leh	34.164203	77.584813	The Grand Dragon	34.156612	77.580351	Hotel
3	Leh	34.164203	77.584813	Rendezvous Cafe	34.165095	77.584020	Bakery
6	Leh	34.164203	77.584813	Coffee Culture	34.165123	77.584145	Coffee Shop
7	Leh	34.164203	77.584813	Pangong Hotel	34.161511	77.582676	Hotel

```
In [29]: print('The recors obtained for tourist venue are - ' + str(tourist_venues.shape[0]))
         print('The recors obtained for business venue are - ' + str(business_venues.shape[0]))

The recors obtained for tourist venue are - 111
The recors obtained for business venue are - 966
```

We got **111** tourist and **966** tourist spots for all the capitals.

No of Tourist Venue in each Capital
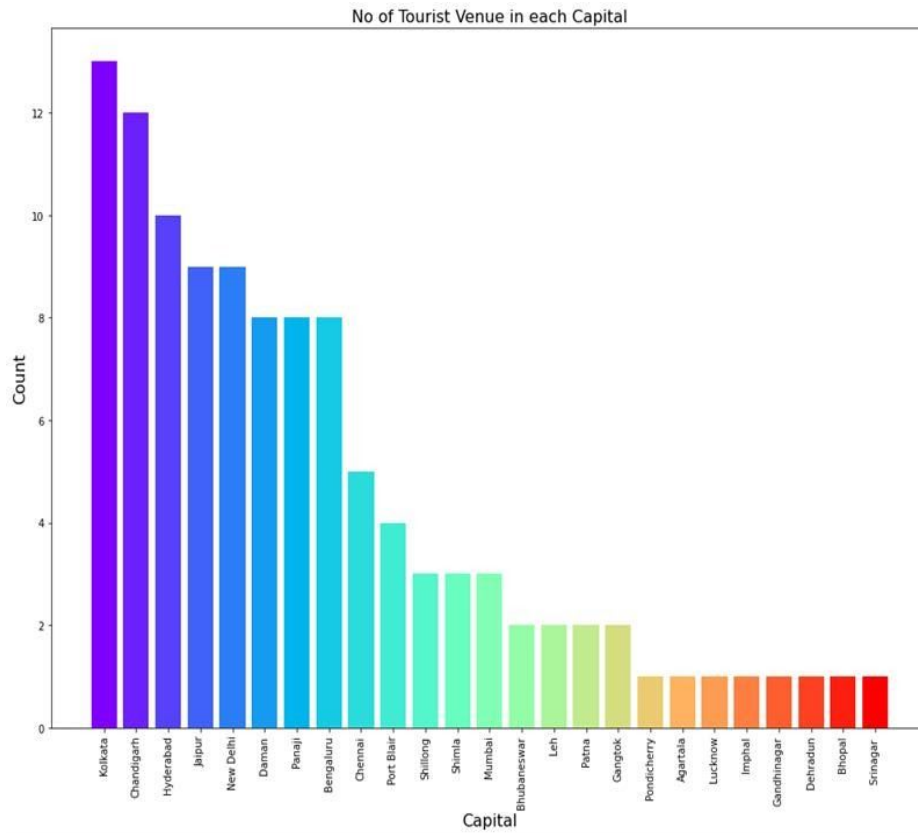
Kolkata had the highest tourist spot among all capitals. Whereas **Hyderabad** has more business venues among all.

No of Business Venue in each Capital
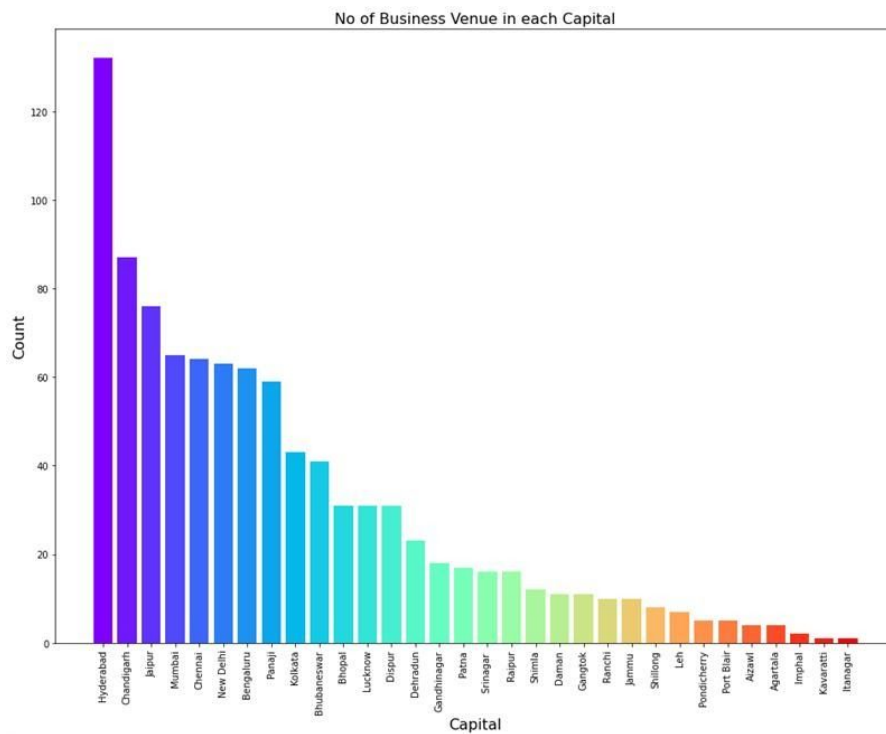
Using the one hot encoding we found the common business and tourist venues categories around capitals.

1. Tourist Venue categories

	Capital	1st Most Common Venue Category	2nd Most Common Venue Category	3rd Most Common Venue Category	4th Most Common Venue Category	5th Most Common Venue Category	6th Most Common Venue Category	7th Most Common Venue Category	8th Most Common Venue Category	9th Most Common Venue Category	10th Most Common Venue Category	11th Most Common Venue Category	12th Most Common Venue Category	13th Most Common Venue Category
0	Agartala	Park	Trail	Island	Art Gallery	Athletics & Sports	Beach	Brewery	Bridge	Campground	Casino	Concert Hall	Field	Garden
1	Bengaluru	Brewery	Arcade	Art Gallery	Plaza	Park	Nightclub	Historic Site	Athletics & Sports	Beach	Bridge	Campground	Casino	Concert Hall
2	Bhopal	National Park	Trail	Island	Art Gallery	Athletics & Sports	Beach	Brewery	Bridge	Campground	Casino	Concert Hall	Field	Garden
3	Bhubaneswar	Park	Trail	Island	Art Gallery	Athletics & Sports	Beach	Brewery	Bridge	Campground	Casino	Concert Hall	Field	Garden
4	Chandigarh	Sculpture Garden	Scenic Lookout	Plaza	Garden	Trail	Historic Site	Art Gallery	Athletics & Sports	Beach	Brewery	Bridge	Campground	Casino

2. Business Venue categories

	Capital	1st Most Common Business Category	2nd Most Common Business Category	3rd Most Common Business Category	4th Most Common Business Category	5th Most Common Business Category	6th Most Common Business Category	7th Most Common Business Category	8th Most Common Business Category	9th Most Common Business Category	10th Most Common Business Category
0	Agartala	Hotel	Indian Restaurant	Coffee Shop	Video Store	Eastern European Restaurant	Deli / Bodega	Department Store	Dessert Shop	Diner	Donut Shop
1	Aizawl	Hotel	Shopping Mall	Café	Video Store	Deli / Bodega	Department Store	Dessert Shop	Diner	Donut Shop	Dumpling Restaurant
2	Bengaluru	Indian Restaurant	Hotel	Asian Restaurant	Department Store	Café	Italian Restaurant	Shopping Mall	Breakfast Spot	Steakhouse	French Restaurant
3	Bhopal	Indian Restaurant	Hotel	Bakery	Department Store	Coffee Shop	Food Court	Pizza Place	Restaurant	Market	Fast Food Restaurant
4	Bhubaneswar	Hotel	Coffee Shop	Clothing Store	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Shopping Mall	Pizza Place	Bakery	Café

Aggregate results against Capital and Venue Category

1. Tourist

```
In [36]: venue_venue_category = tourist_venues.groupby(['Capital', 'Venue Category'], as_index=False).agg(lambda x: ", ".join(x))
         venue_venue_category.head()
```

Out[36]:

	Capital	Venue Category	Venue
0	Agartala	Park	Heritage Park
1	Bengaluru	Arcade	The Lotus Pavillion
2	Bengaluru	Art Gallery	Chitra Kala Parishad
3	Bengaluru	Brewery	The Biere Club, Arbor Brewing Company, Communiti
4	Bengaluru	Nightclub	Pebble

2. Business

```
In [55]: business_venue_category = business_venues.groupby(['Capital', 'Venue Category'], as_index=False).agg(lambda x: ", ".join(x))
         business_venue_category.head()
```

Out[55]:

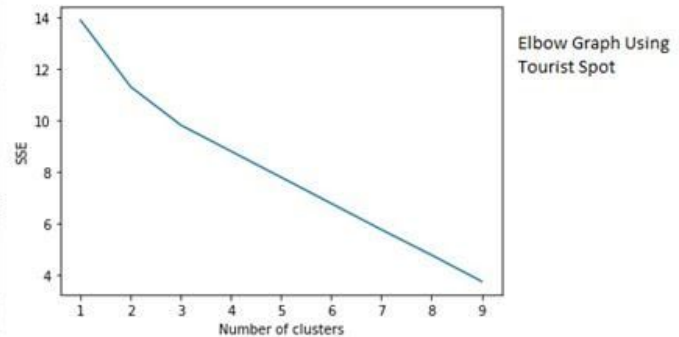|	Capital	Venue Category	Venue
0	Agartala	Coffee Shop	Cafe Coffee Day
1	Agartala	Hotel	Ginger Agartala, Ginger Hotel Agartala
2	Agartala	Indian Restaurant	Bawarchi
3	Aizawl	Café	The Coffee Place, Zarkawt, Aizawl
4	Aizawl	Hotel	Hotel Regency, Hotel Floria

We then use the K-Means clustering algorithm and using elbow method to find optimal value of k .
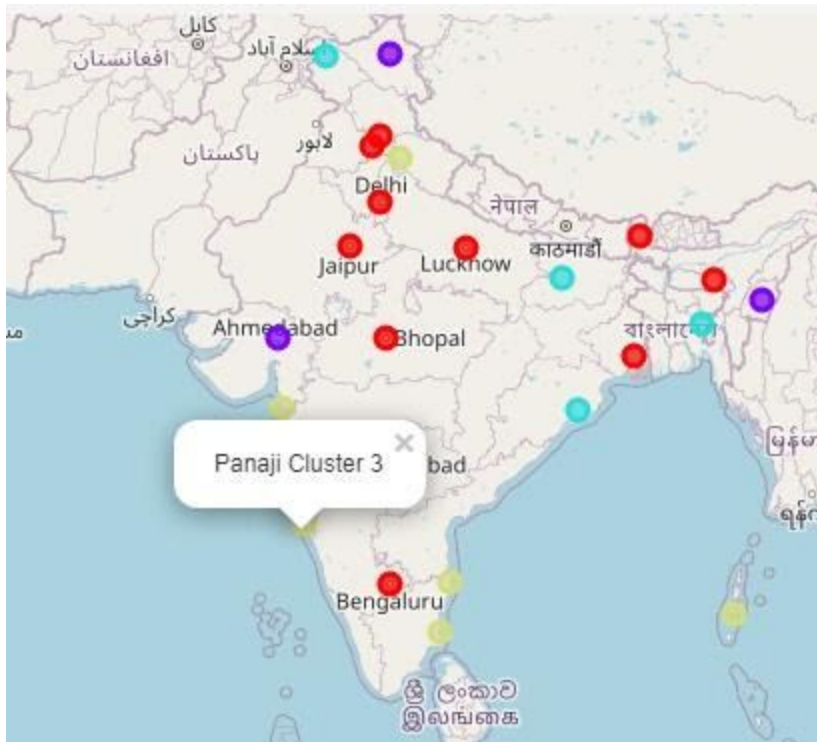
Results

1. Tourist
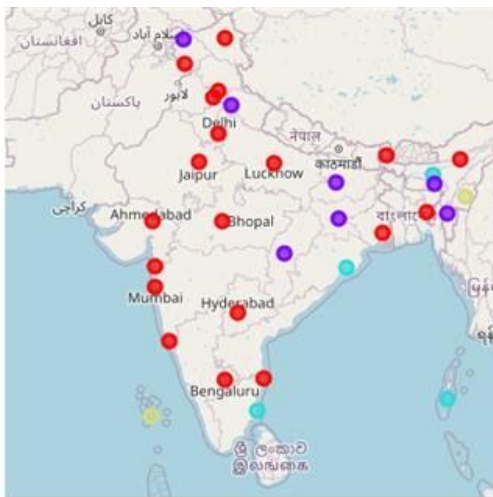

Elbow Graph Using Tourist Spot

Capital Cluster Using Tourist Spot Data

We see diverse plot where cities along coastline have favourite tourist spot as beach cluster label=3

```
capital_tourist_merged.loc[capital_tourist_merged['Cluster Labels'] == 3, capital_tourist_merged.columns[[0] +
```
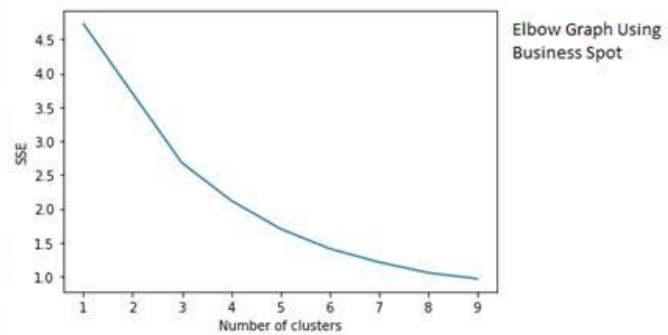
	Cluster Labels	Capital	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	3	Chennai	Marina Beach | மெரினா கடற்கரை (Marina Beach), ...	Island Grounds	Fort Museum							
6	3	Daman	Jetty, Hotel Sandy Resort	Daman, Devka Beach	Daman-Gujarat Border	Hotel Mirasol Resort & Water Park	The Deltin	Daman Fort				
7	3	Dehradun	Forest Research Institute									
18	3	Panaji	Miramar Beach, Coco Beach, Vaiguinim Beach	Deltin Jaqk, Chances	Dr.Shyama Prasad Mukherjee Indoor Stadium	Haunted Pilerne Riverlet	Kala Academy					
20	3	Pondicherry	Karaikal Beach									
21	3	Port Blair	Corbyn's Cove Beach, North Bay	Andaman & Nicobar	Samudrika Marine Museum							

2. Business

Capital Cluster Using Business Data

As per business spot majority have the same pattern (cluster=0 Red).

```
capital_business_merged.loc[capital_business_merged['Cluster Labels'] == 0, capital_business_merged.columns[[0] + li
```

	Cluster Labels	Capital	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
0	0	Agartala	Ginger Agartala, Ginger Hotel Agartala	Bawarchi	Cafe Coffee Day					
2	0	Bengaluru	Masala Klub, Ujwal Bar & Restaurant, MTR 1924,...	JW Marriott Hotel Bengaluru, Taj West End, ITC...	J W Kitchen, Yauatcha	Foodhall, Thom's Bakery and Stores	Matteo, cafe cerrise	Toscano, Sunny's	UB City, Westside	Airlines Hotel, Brahmins Coffee Bar
3	0	Bhopal	Manohar Dairy & Restaurant, Wind n Waves, Indi...	Jehan Numa Palace Hotel Bhopal, Courtyard Bhopal	Bake n shake, Eclair-The Bake Shop	D.B City Mall, HyperCITY	Cafe Coffee Day, Cafe Coffee Day	Top N Town Main Icecream Joint, D.B. City Food...	Pizza Hut, Domino's Pizza	Milan Restaurant N
5	0	Chandigarh	Cafe Nomad, Garlic and Greens	Nik Bakers, Nik Bakers, Monica's	Ghazal, Pal Dhaba (Since 1960)	JW Marriott Hotel Chandigarh	The Coffee Bean & Tea Leaf, Indian	Backpackers Cafe, Subway	Ovenfresh, Brooklyn Central	Zara, Zara,

Food joints and hotels seem to be the most common business among cities.

Conclusion

As per tourist data we see capitals are more diverse in nature whereas from a business standpoint majority capital behave the same.