Review-1

CSE3020-Data Visualization
Slot: D1
Faculty: Lydia Jane G

Indian Startups-2021

Submitted by:
Satwik Tanwar
(20BCE2945)

Dataset Details:

Dataset: Indian Startups – 2021
The Dataset contains the list of Private Limited companies registered with Ministry of Corporate Affairs (MCA) across India in 2021 (January to April). The data is anonymised and unique id is generated for the companies registered.

URL: https://www.kaggle.com/bhararthshiviah/indian-startups-2021
Cleaned Data URL: https://drive.google.com/file/d/1_ST97Tc2dILzk-w_JaUsw9k9k_bSd4XE/view?usp=sharing

Number of Items (Rows) : 54,894

Number of Attributes (Columns) : 12

Language used: Python

Attributes Definition:
1. **company_uid:** Unique identification number generated for each registered company.

2. **date_of_registration:** Date when the company was register with Ministry of Corporate Affairs (MCA).

3. **month_name:** Which month the company was registered.

4. **state:** In which region/location that company was registered.

5. **roc:** This field informs in which Registrar of Companies (RoC) the company was registered.

6. **category:** Weather the entity is limited by shares or guarantee.

7. **class:** Whether the entity is PRIVATE LIMITED (OPC) or PRIVATE LIMITED.

8. **company_type:** Whether it is run by Government or Private.

9. **authorized_capital:** Authorized capital is the maximum amount of capital a company is authorized to raise from its shareholders.

10. **paidup_capital:** Paid-up capital is the amount paid by the shareholders for the shares held by them in the company.

11. **activity_code:** This provides the id of the nature of business the entity is legally registered for.

12. **activity_description:** This provides the nature of business the entity is legally registered for.

Task Abstraction:

Q1. What were the trends of registration of new companies during January-2021 and April-2021?
Attributes Required: month_name, date_of_registration

Q2. How did COVID-19 second-wave affected the number of companies registered monthly?
Attributes Required: montt_name, date_of_registration

Q3. What are the trends in authorized and paid-up capital?
Attributes Required: authorized_capital, paidup_capital, activity_description

Q4. Which sector has the most potential for future opportunities throughout the country?
Attributes Required: authorized_capital, paidup_capital, activity_description

Q5. Does the business trends at the national level reflects in the trends at states and RoC levels?
Attributes Required: state, roc, authorized_capital , paidup_capital

Q6. What are the sectors in which foreign countries are investing in India?
Attributes Required: company_type, activity_description

Q7. How are the foreign companies performing in India?
Attributes Required: company_type, activity_description, authorized_capital, paidup_capital

Q8. What is the preferred category of company in India and why?
Attributes Required: category, company_type, activity_description

The first action is to identify the plots that can represent the data in the best possible way and then understand those plots to achieve the target of answering these question through visualizing the data and getting insights about the new businesses and their trends in India.

Data Abstraction:

S.no.	Attributes	Datatypes
1.	company_uid	Categorical
2.	date_of_registration	Quantitative Ordered
3.	month_name	Categorical
4.	state	Categorical
5.	roc	Categorical
6.	category	Categorical
7.	class	Categorical
8.	company_type	Categorical
9.	authorized_capital	Quantitative Ordered
10.	paidup_capital	Quantitative Ordered
11.	activity_code	Categorical
12.	activity_description	Categorical

Encoding:

Code:

In[1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
get_ipython().run_line_magic('matplotlib', 'inline')
```

In[2]:

```python
companies=pd.read_csv('2021_registered_companies.csv')

# ### Number of Companies Registered in each month.
```

In[3]:

```python
ax=plt.figure(figsize=(10,8)).add_axes([0,0,1,1])
sns.set_style('white')
sns.countplot(x='month_name',data=companies)

for rect in ax.patches:
    ax.text (rect.get_x() + rect.get_width() / 2,rect.get_height(),"%i"% rect.get_height()
        ,fontsize=14 )

plt.title('Registered Companies',fontsize=14, weight='bold')
plt.xlabel('Month',fontsize=14)
plt.ylabel('Number of Companies Registered',fontsize=14)
plt.show()


# ### Companies Registered by Date
```

In[4]:

```python
companies_date=pd.DataFrame(companies['date_of_registration'].groupby
                (companies['date_of_registration'].iloc[:]).count())
companies_date.rename(columns={"date_of_registration":"count"},inplace=True)
companies_date.sort_values('date_of_registration')
companies_date.reset_index(inplace=True)
```

In[5]:

```python
df=companies.pivot_table(index=['month_name','date_of_registration'])
df['count']=np.array(companies_date['count'])
df.drop('activity_code',axis=1,inplace=True)
df.sort_values('date_of_registration',inplace=True)
df.reset_index(inplace=True)


# In[6]:

ax=plt.figure(figsize=(10,8)).add_axes([0,0,1,1])
sns.set_style('white')
sns.boxplot(x='month_name',y='count',data=df)
plt.title('Registered Companies in a day',fontsize=14, weight='bold')
plt.xlabel('Month',fontsize=14)
plt.ylabel('Number of Companies Registered in a day',fontsize=14)
plt.show()


# ### Companies in each Sector

# In[7]:

ax=plt.figure(figsize=(11,8)).add_axes([0,0,1,1])
sns.set_style('white')
sns.countplot(y='activity_description',data=companies)
for rect in ax.patches:
    ax.text (rect.get_width(), rect.get_y() + rect.get_height() / 2,
        "%i"% rect.get_width(),fontsize=13 )

plt.title('Sectors of Companies',fontsize=14, weight='bold')
plt.xlabel('Number of Companies Registered',fontsize=14)
plt.ylabel('Sectors',fontsize=14)
plt.show()


# In[8]:

companies_sector=pd.DataFrame(companies['activity_description'].groupby
                (companies['activity_description'].iloc[:]).count())
companies_sector.rename(columns={"activity_description":"count"},inplace=True)
companies_sector.sort_values('count',ascending=False,inplace=True)
companies_sector.loc['Others'] = [ companies_sector.iloc[5:,0].sum()]
companies_sector.sort_values('count',ascending=False,inplace=True)
companies_sector=companies_sector.iloc[:6,0]


# In[9]:
```

```python
plt.figure(dpi=1000)

pie, ax = plt.subplots(figsize=[10,7])
labels = companies_sector.keys()
plt.pie(x=companies_sector, autopct="%.1f%%",labels=labels, pctdistance=0.5)
plt.title("Registered Companies by Sector", weight='bold',fontsize=14);
plt.show()


# ### Comparision of Companies on the basis of Authorized and Paid-up capital

# In[10]:

companies_activity=companies.pivot_table(index='activity_description',values=['paidup_capi
tal','authorized_capital'])
companies_activity.reset_index(inplace=True)


# In[11]:

plt.figure(figsize=(17,8))
df = companies_activity.melt('activity_description', var_name='capital',  value_name='vals')
sns.pointplot(x="activity_description", y="vals", hue='capital', data=df)
plt.xticks(rotation = 90)
plt.title('Comparision between sectors',fontsize=14, weight='bold')
plt.xlabel('Sectors',fontsize=14)
plt.ylabel('Capital',fontsize=14)
plt.show()


# ### Companies registered in states

# In[12]:

companies_state=pd.DataFrame(companies['state'].groupby
                (companies['state'].iloc[:]).count())
companies_state.rename(columns={"state":"count"},inplace=True)
companies_state.sort_values('count',ascending=False,inplace=True)
companies_state.reset_index(inplace=True)


# In[13]:

plt.figure(figsize=(17,7))
sns.set_style('whitegrid')
sns.barplot(x='state',y='count',data=companies_state)

plt.xticks(rotation = 90)
```

```python
plt.title('Registered Companies in States',fontsize=10, weight='bold')
plt.xlabel('States',fontsize=14)
plt.ylabel('Number of Companies Registered',fontsize=14)
plt.show()


# #### The graph of some of the graphs at last are not visibile so they are seperately plotted
#here

# In[14]:

plt.figure(figsize=(17,7))
sns.set_style('whitegrid')
sns.barplot(x='state',y='count',data=companies_state[21:])

plt.xticks(rotation = 90)
plt.title('Registered Companies in States',fontsize=10, weight='bold')
plt.xlabel('States',fontsize=14)
plt.ylabel('Number of Companies Registered',fontsize=14)
plt.show()


# ### Comparision of trends in top 5 states with national trends

# In[15]:

companies5=companies[(companies['state'] == 'Maharashtra') |
             (companies['state'] == 'Uttar Pradesh') |
             (companies['state'] == 'Delhi') |
             (companies['state'] == 'Karnataka') |
             (companies['state'] == 'Telangana')
            ]


# In[16]:

companies5_sector=pd.DataFrame(companies5['activity_description'].groupby
                (companies5['activity_description'].iloc[:]).count())
companies5_sector.rename(columns={"activity_description":"count"},inplace=True)
companies5_sector.sort_values('count',ascending=False,inplace=True)
companies5_sector.loc['Others'] = [ companies5_sector.iloc[5:,0].sum()]
companies5_sector.sort_values('count',ascending=False,inplace=True)
companies5_sector=companies5_sector.iloc[:6,0]


# In[17]:

plt.figure(dpi=1000)
```

```python
pie, ax = plt.subplots(figsize=[10,7])
labels = companies5_sector.keys()
plt.pie(x=companies5_sector, autopct="%.1f%%",labels=labels, pctdistance=0.5)
plt.title("Registered Companies by Sector in top 5 states", weight='bold',fontsize=14);
plt.show()


# In[18]:


companies5_activity=companies5.pivot_table(index='activity_description',values=['paidup_c
apital','authorized_capital'])
companies5_activity.reset_index(inplace=True)


# In[19]:


plt.figure(figsize=(17,8))
df = companies5_activity.melt('activity_description', var_name='capital',  value_name='vals')
sns.pointplot(x="activity_description", y="vals", hue='capital', data=df)
plt.xticks(rotation = 90)
plt.title('Comparision between sectors in top 5 states',fontsize=14, weight='bold')
plt.xlabel('Sectors',fontsize=14)
plt.ylabel('Capital',fontsize=14)
plt.show()


# ### Companies registered in various RoC

# In[20]:


companies_roc=pd.DataFrame(companies['roc'].groupby
                (companies['roc'].iloc[:]).count())
companies_roc.rename(columns={"roc":"count"},inplace=True)
companies_roc.sort_values('count',ascending=False,inplace=True)
companies_roc.reset_index(inplace=True)


# In[21]:


plt.figure(figsize=(17,7))
sns.set_style('whitegrid')

sns.barplot(x='roc',y='count',data=companies_roc)

plt.xticks(rotation = 90)
plt.title('Registered Companies at RoC',fontsize=10, weight='bold')
plt.xlabel('RoC',fontsize=14)
```

```python
plt.ylabel('Number of Companies Registered',fontsize=14)
plt.show()


# ### Foreign Investments in India

# In[22]:

companies_foreign=companies[companies['company_type']=='Subsidiary of Foreign
Company']


# In[23]:

ax=plt.figure(figsize=(11,8)).add_axes([0,0,1,1])
sns.set_style('white')
sns.countplot(y='activity_description',data=companies_foreign)
for rect in ax.patches:
    ax.text (rect.get_width(), rect.get_y() + rect.get_height() / 2,
        "%i"% rect.get_width(),fontsize=13 )

plt.title('Sectors of Companies',fontsize=14, weight='bold')
plt.xlabel('Number of Companies Registered',fontsize=14)
plt.ylabel('Sectors',fontsize=14)
plt.show()


# ### Capital Raised by Foreign Countries in India

# In[24]:

ax=plt.figure(figsize=(17,10)).add_axes([0,0,1,1])
sns.set_style('whitegrid')
sns.stripplot(x='activity_description',y='paidup_capital',data=companies[companies['paidup_
capital']<1.5e9],
        jitter=True,hue='company_type',palette='hls',size=10)
plt.xticks(rotation = 90)
plt.title('Paid-up Capital',fontsize=14, weight='bold')
plt.xlabel('Sectors',fontsize=14)
plt.ylabel('Capital',fontsize=14)
plt.show()


# In[25]:

fcompanies_activity=companies_foreign.pivot_table(index='activity_description',
                        values=['paidup_capital','authorized_capital'])
fcompanies_activity.reset_index(inplace=True)
```

```python
# In[26]:

plt.figure(figsize=(17,8))
foreign = fcompanies_activity.melt('activity_description', var_name='capital',
value_name='vals')
sns.pointplot(x="activity_description", y="vals", hue='capital', data=foreign)
plt.xticks(rotation = 90)
plt.title('Comparision between sectors',fontsize=14, weight='bold')
plt.xlabel('Sectors',fontsize=14)
plt.ylabel('Capital',fontsize=14)
plt.show()


# In[27]:

ax=plt.figure(figsize=(10,8)).add_axes([0,0,1,1])
sns.set_style('white')
sns.countplot(x='category',data=companies)

for rect in ax.patches:
    ax.text (rect.get_x() + rect.get_width() / 2,rect.get_height(),"%i"% rect.get_height()
        ,fontsize=14 )

plt.title('Category of Registered Companies',fontsize=14, weight='bold')
plt.xlabel('Categories',fontsize=14)
plt.ylabel('Number of Companies Registered',fontsize=14)
plt.show()


# In[28]:

plt.figure(figsize = (5,5))
sns.scatterplot(x='company_type',y='category',data=companies)
plt.xticks(rotation = 90)
plt.show()


# In[29]:

ax=plt.figure(figsize=(7,5)).add_axes([0,0,1,1])
sns.set_style('white')
sns.countplot(y='activity_description',data=companies[companies['category']=='Company
Limited by Guarantee'])
for rect in ax.patches:
    ax.text (rect.get_width(), rect.get_y() + rect.get_height() / 2,
        "%i"% rect.get_width(),fontsize=13 )
```
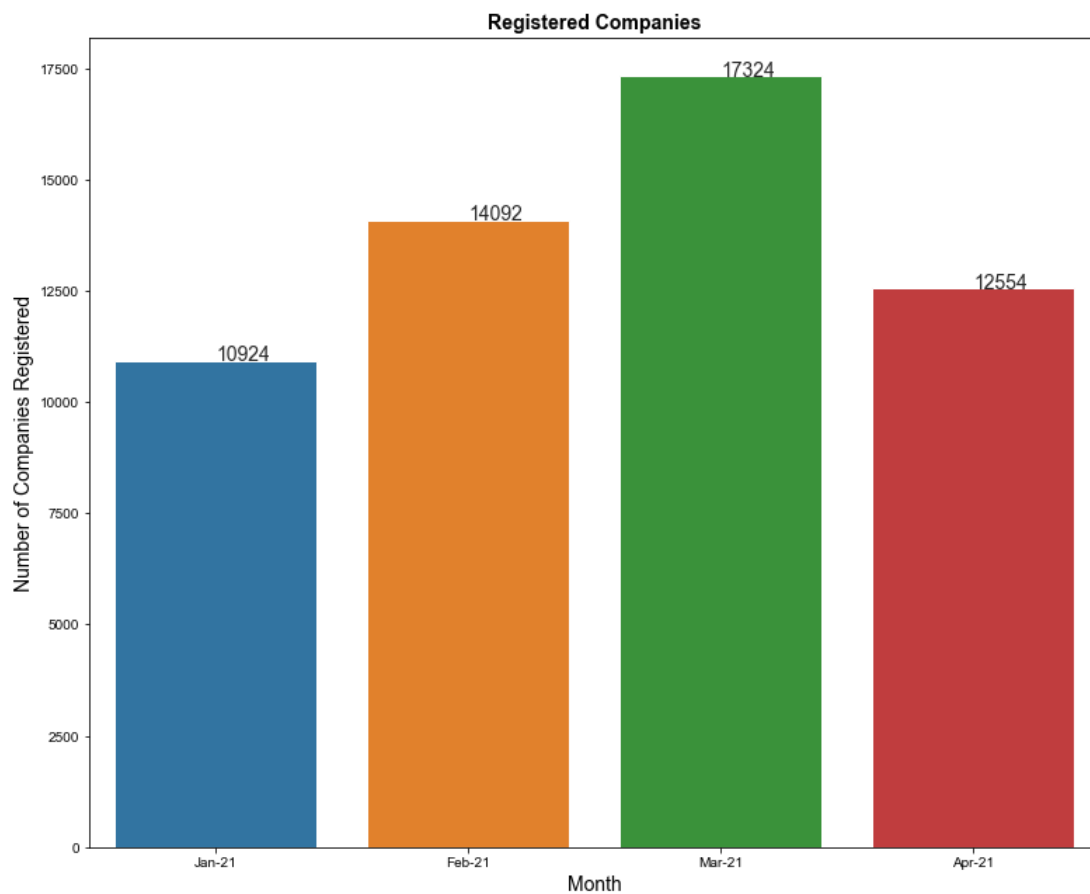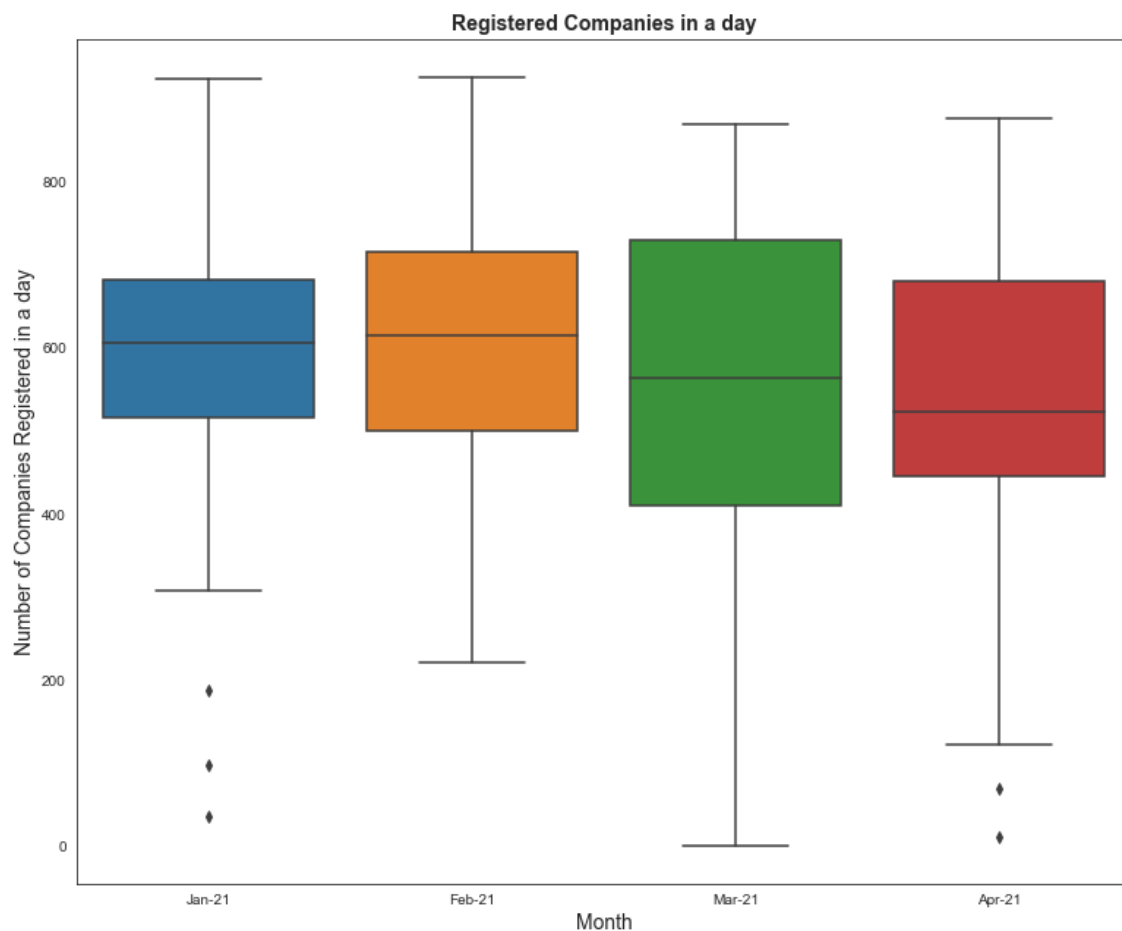
```
plt.title('Sectors of Companies Limited by Guarantee',fontsize=14, weight='bold')
plt.xlabel('Number of Companies Registered',fontsize=14)
plt.ylabel('Sectors',fontsize=14)
plt.show()
```
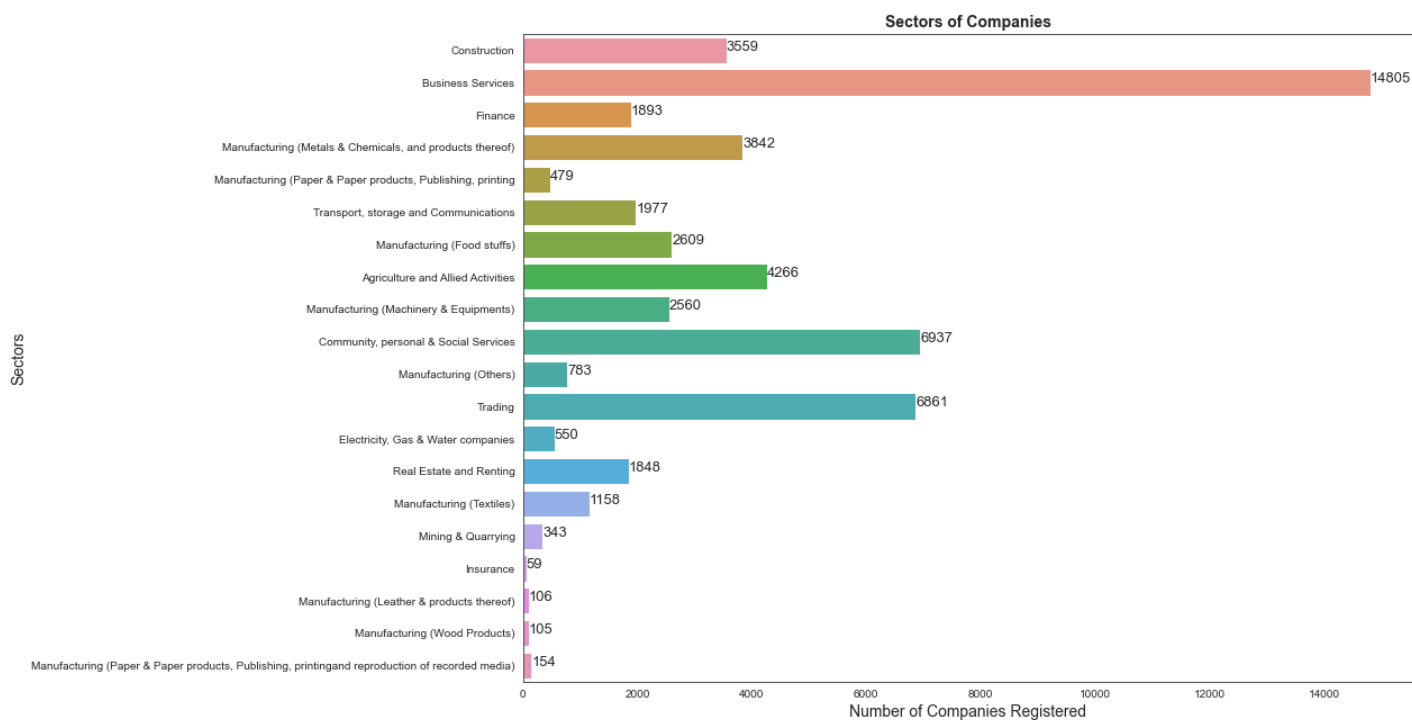
Graphs:

This graph shows the total number of companies that were registered in each month in the given period of time.

Registered Companies in a day

During January 2021 the condition of daily COVID-19 cases was just getting better and the number of companies registered in most of the days of January was between 550 to 650 approximately but there were some days when it was much lower.
As the condition was getting better during February the number of companies registered each day increased and the consistency was good, the value even got better in March but in April the second wave of COVID-19 hit us and the number of companies registered each day decreased and the consistency also went down.
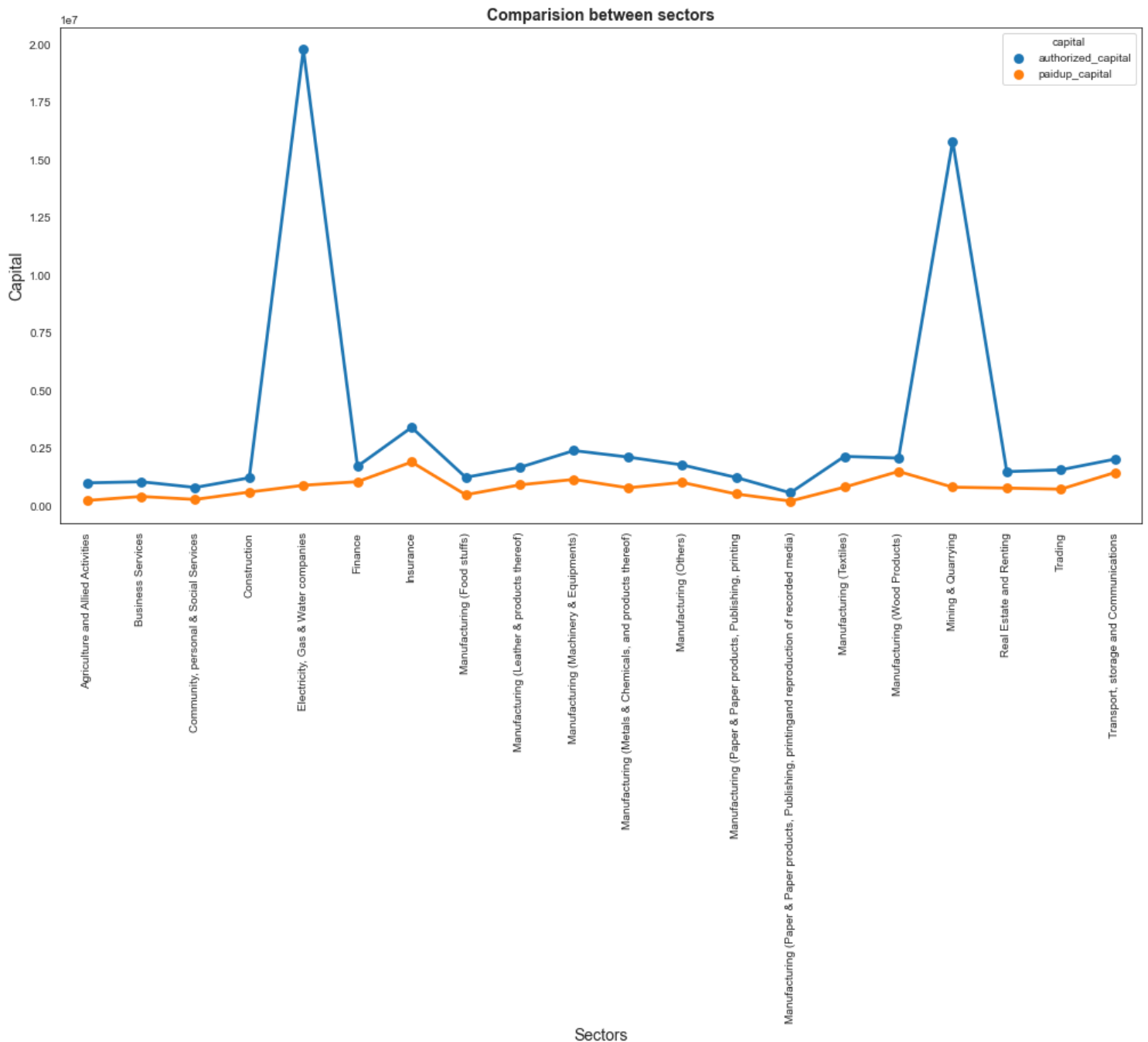
Sectors of Companies

Sector	Number of Companies Registered
Construction	3559
Business Services	14805
Finance	1893
Manufacturing (Metals & Chemicals, and products thereof)	3842
Manufacturing (Paper & Paper products, Publishing, printing	479
Transport, storage and Communications	1977
Manufacturing (Food stuffs)	2609
Agriculture and Allied Activities	4266
Manufacturing (Machinery & Equipments)	2560
Community, personal & Social Services	6937
Manufacturing (Others)	783
Trading	6861
Electricity, Gas & Water companies	550
Real Estate and Renting	1848
Manufacturing (Textiles)	1158
Mining & Quarrying	343
Insurance	59
Manufacturing (Leather & products thereof)	106
Manufacturing (Wood Products)	105
Manufacturing (Paper & Paper products, Publishing, printingand reproduction of recorded media)	154

This graph shows the comparison between the number of companies registered with different activity descriptions i.e. in different sectors.

Registered Companies by Sector

- Others — 33.1%
- Business Services — 27.0%
- Community, personal & Social Services — 12.6%
- Trading — 12.5%
- Agriculture and Allied Activities — 7.8%
- Manufacturing (Metals & Chemicals, and products thereof) — 7.0%

It shows the percentage of companies with different activity description that were registered.
We can clearly observe the sector in which maximum number of companies were registered from the above graphs (Barplot and Pie Chart)

Comparision between sectors

We can observe from the above graph that apart from two sectors (Electricity, Gas & Water Companies and Mining & Quarrying), the difference between authorized and paid-up capital is no that much which means that they have already generated most of the capital that they can but there is a lot of capital available that can be generated in Electricity, Gas & Water Companies and Mining & Quarrying.

If we look at the above three plots we can observe that only 1% and 0.6% of total companies registered were in Electricity, Gas & Water Companies and Mining & Quarrying sectors respectively but the amount of capital left to be generated in these sectors is enormous, therefore these sectors have a lot of potential for future opportunities provided people can come up with new innovative ideas to make proper use of this potential.

Registered Companies in States
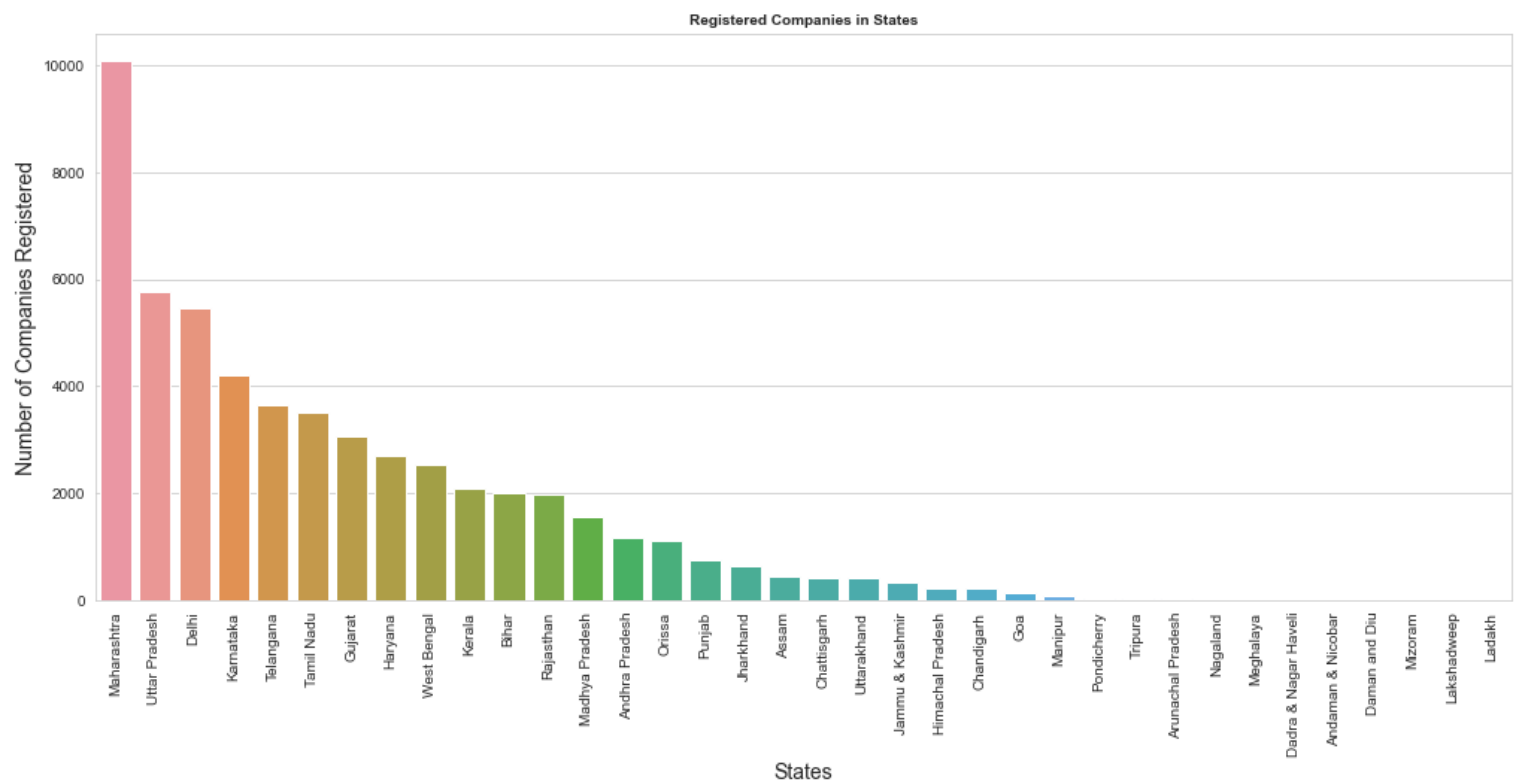
This graph shows the number of companies registered in each state and Union Territory with the top 5 states being Maharashtra, Uttar Pradesh, Delhi, Karnataka and Telangana.

Registered Companies in States
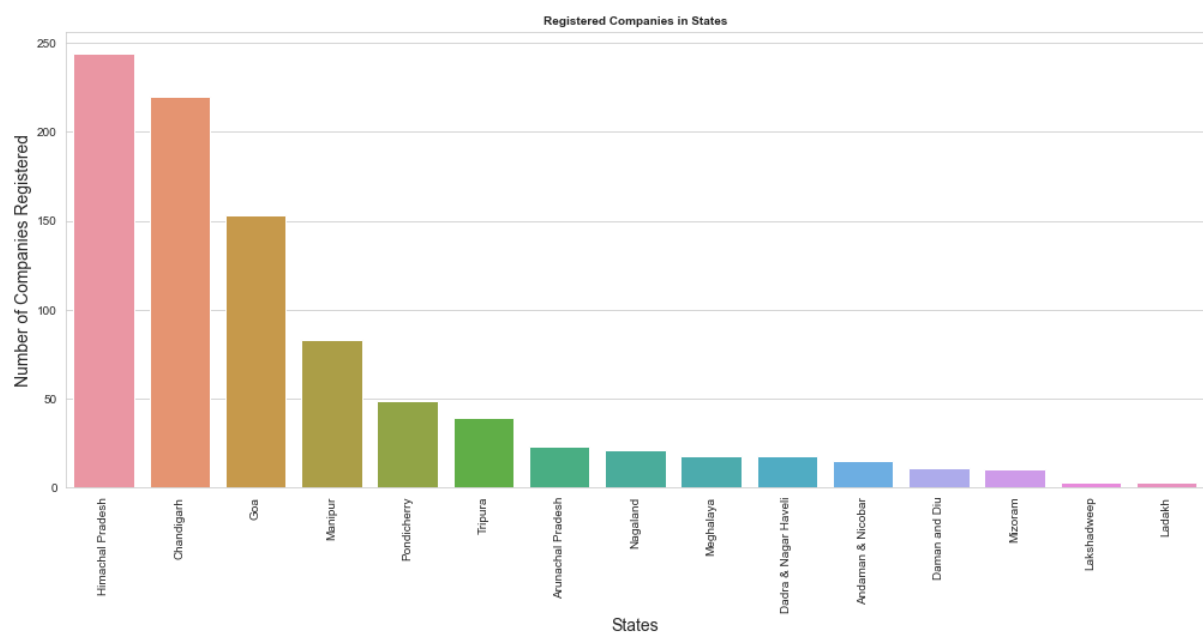
This is a zoomed in version of the States and Union Territories with least registered companies.
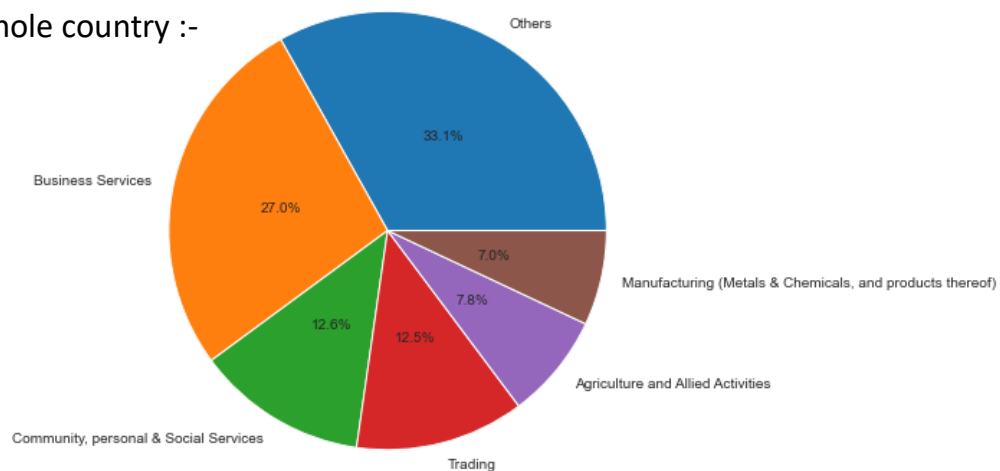
Registered Companies by Sector in top 5 states

Companies Registered in TOP 5 States:-
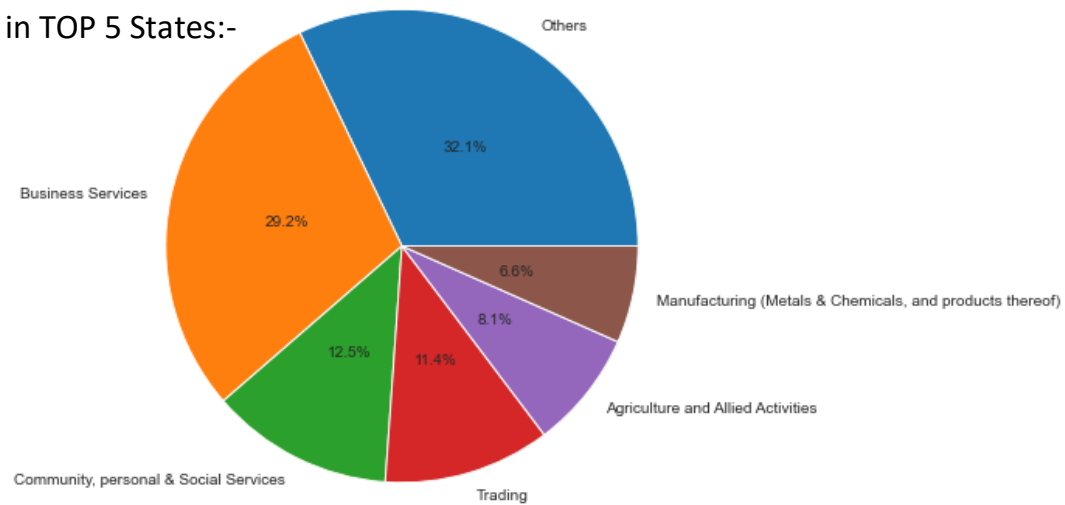
This shows the percentage of companies registered in different sectors in the states with top 5 number of companies registered.

Registered Companies by Sector

Companies registered in whole country :-

If we compare the pie chart for top 5 states with the similar pie chart that we created for the whole country, we can observe that the national trends are reflected in these states while talking about the number of companies registered in each sector.

Comparision between sectors in top 5 states
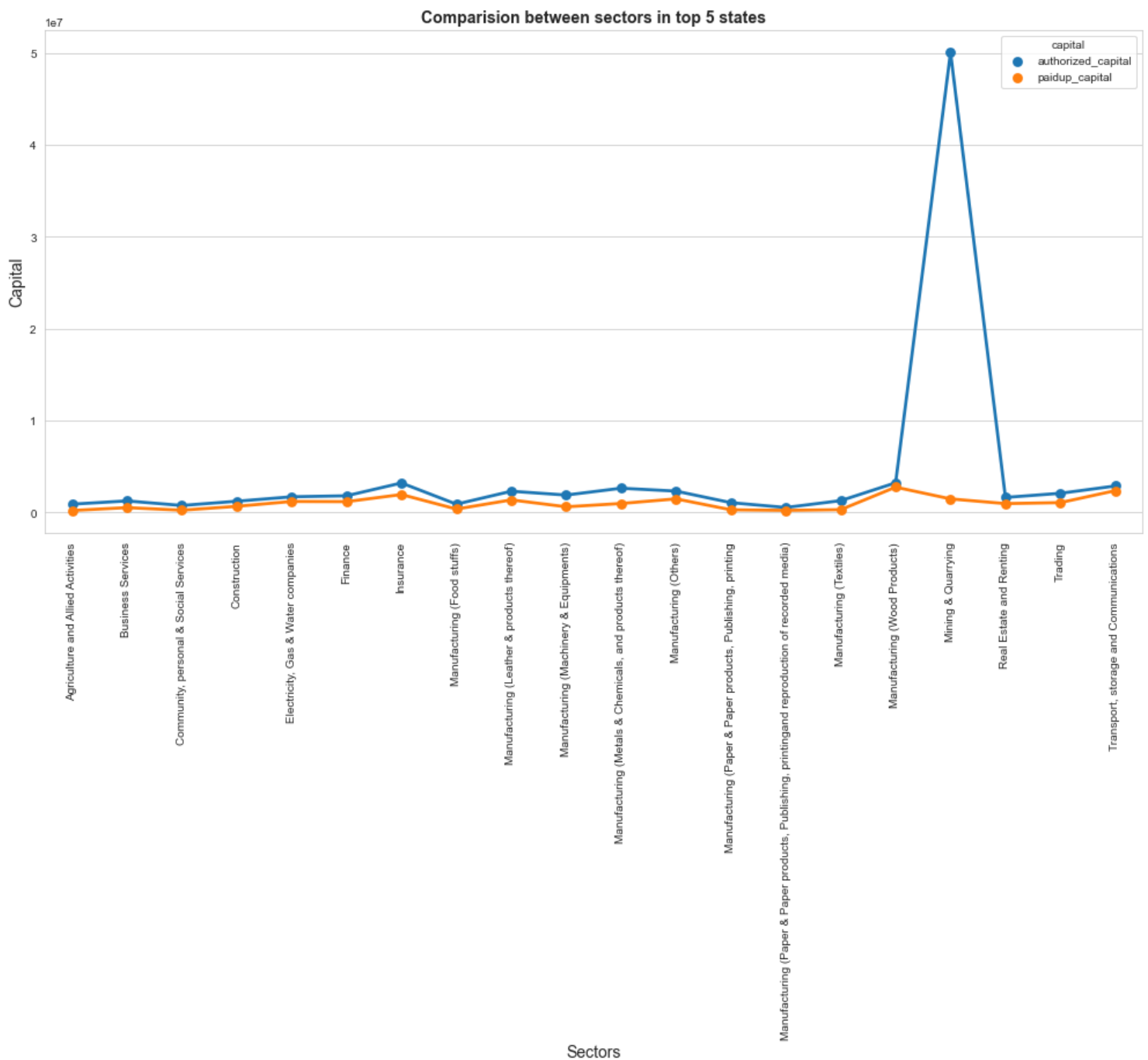
This graph shows the trends of authorized and paid-up capitals of companies in the top 5 states. We can observe that there is a very less difference between authorized and paid-up capitals in almost every sector as compared to the data of all the states and also the difference for Electricity, Gas & Water Companies is significantly lower. This is due to the fact that most of the big companies are located in these places and also there is a lot of competition in these places.

Mining and Quarrying still have a lot of potential capital to be generated in these states as compared to other sectors.

Registered Companies at RoC
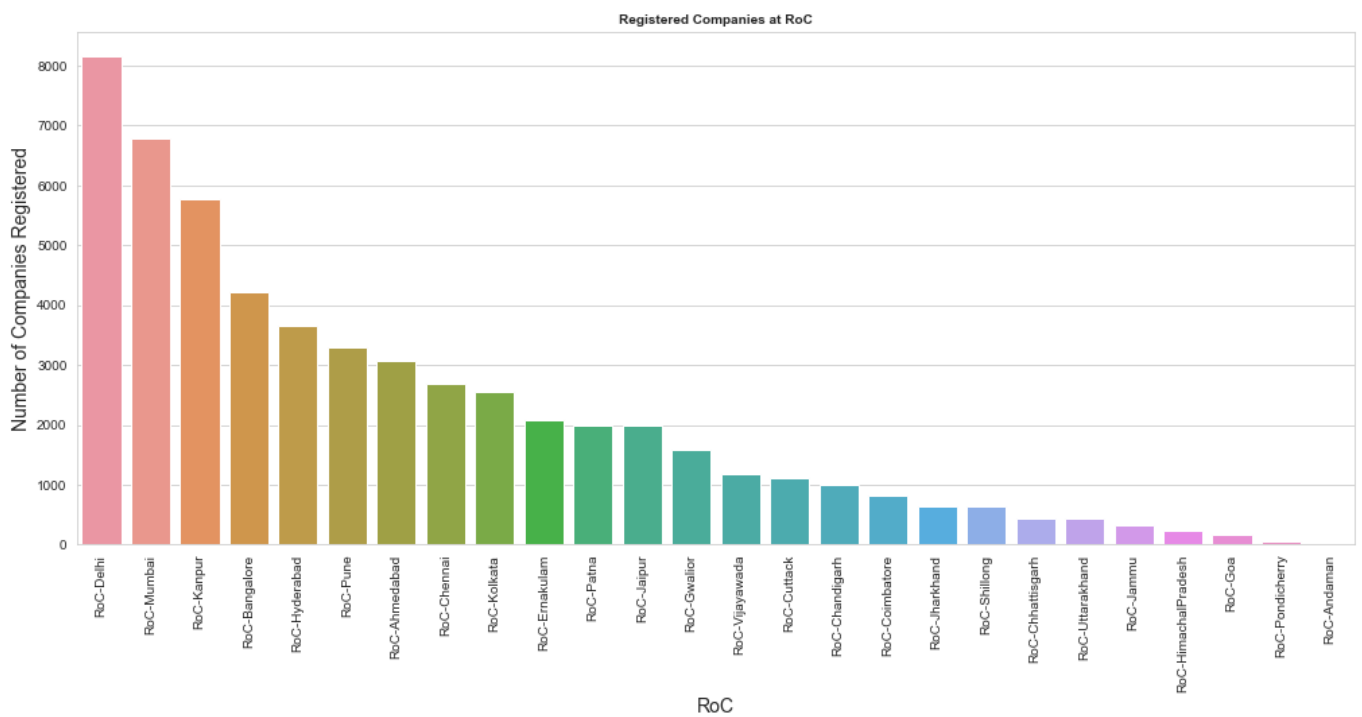
This barplot compares the number of companies registered in each RoC.
We can observe that the top five RoC are from the top 5 states and therefore, we can conclude that the trends that we observed for these stats are reflected on these RoC too.
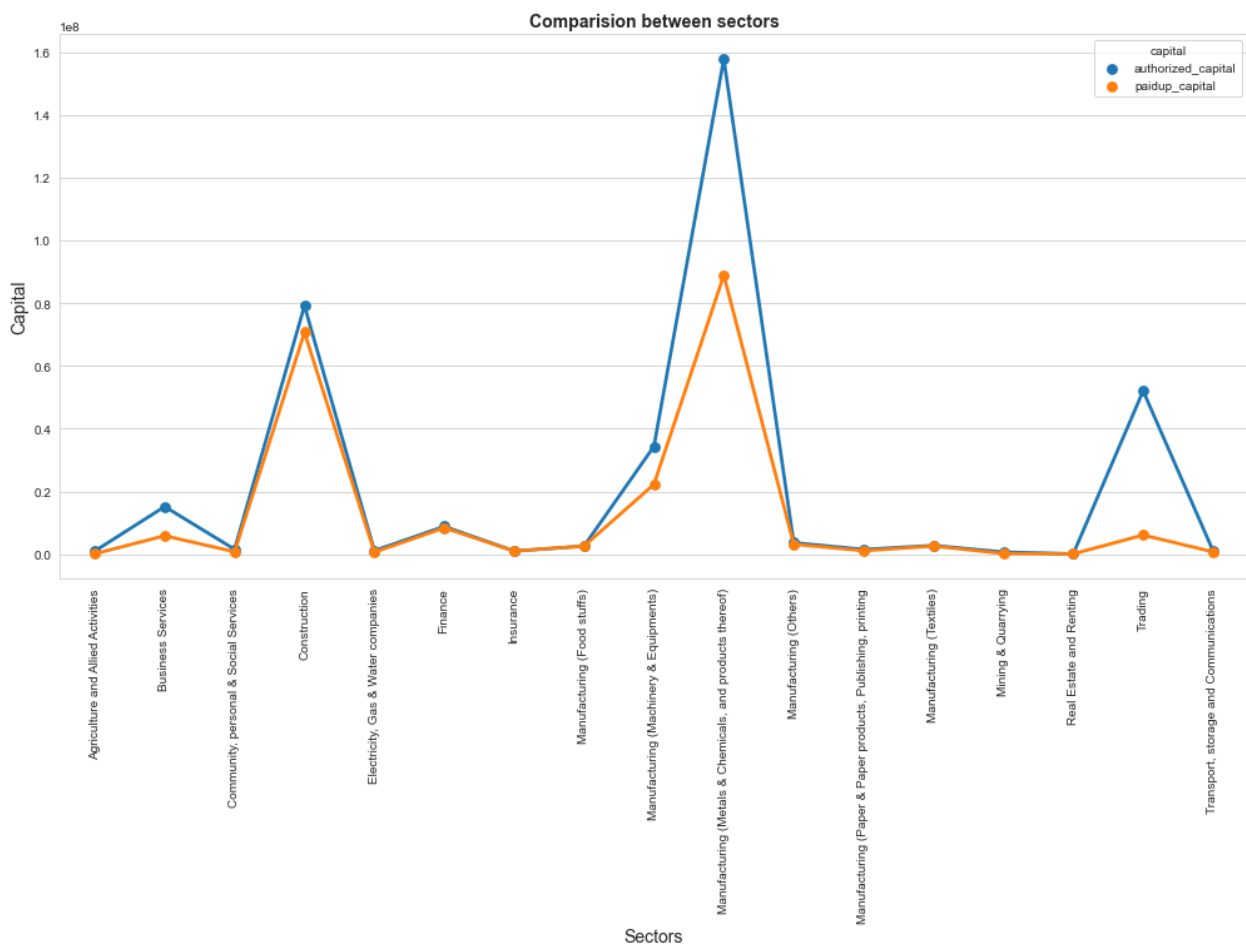

Sectors of Companies
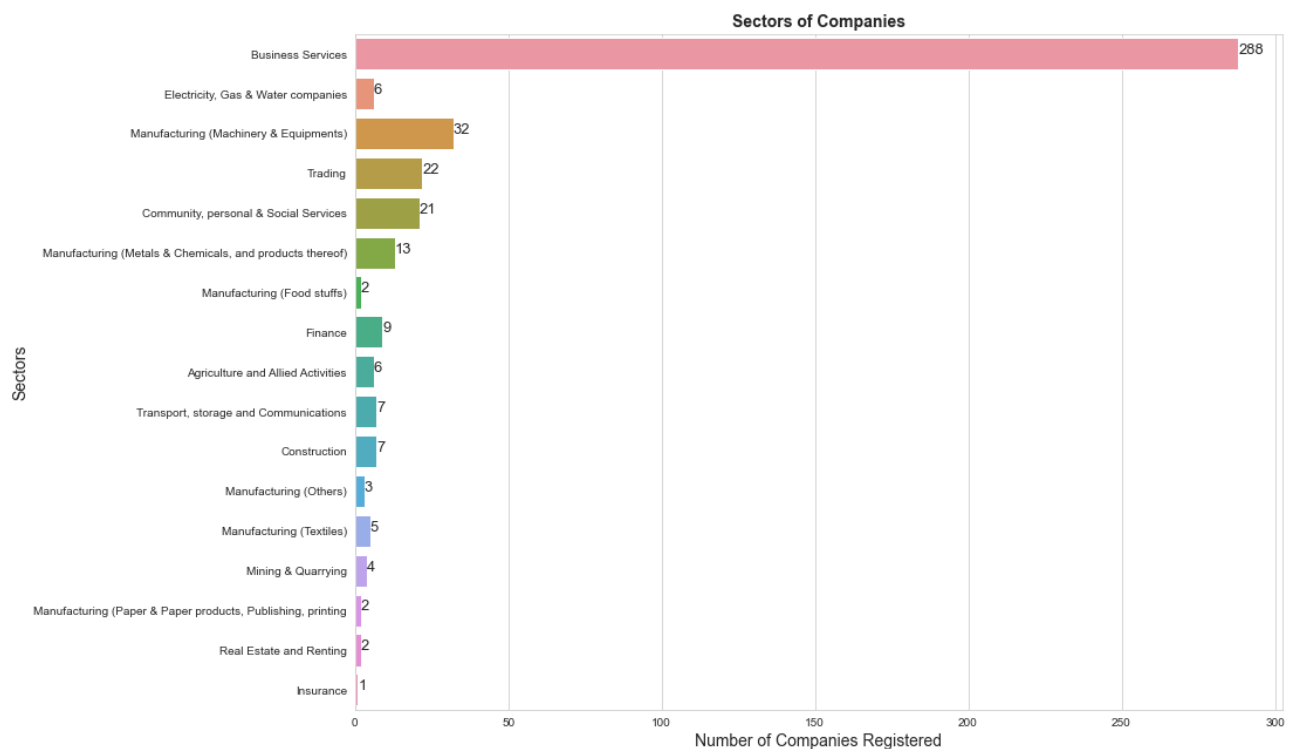
This plot shows all the sectors in which new subsidiaries of foreign companies have been registered. We can observe that the maximum investments are made in the Business Services Sectors by foreign companies.
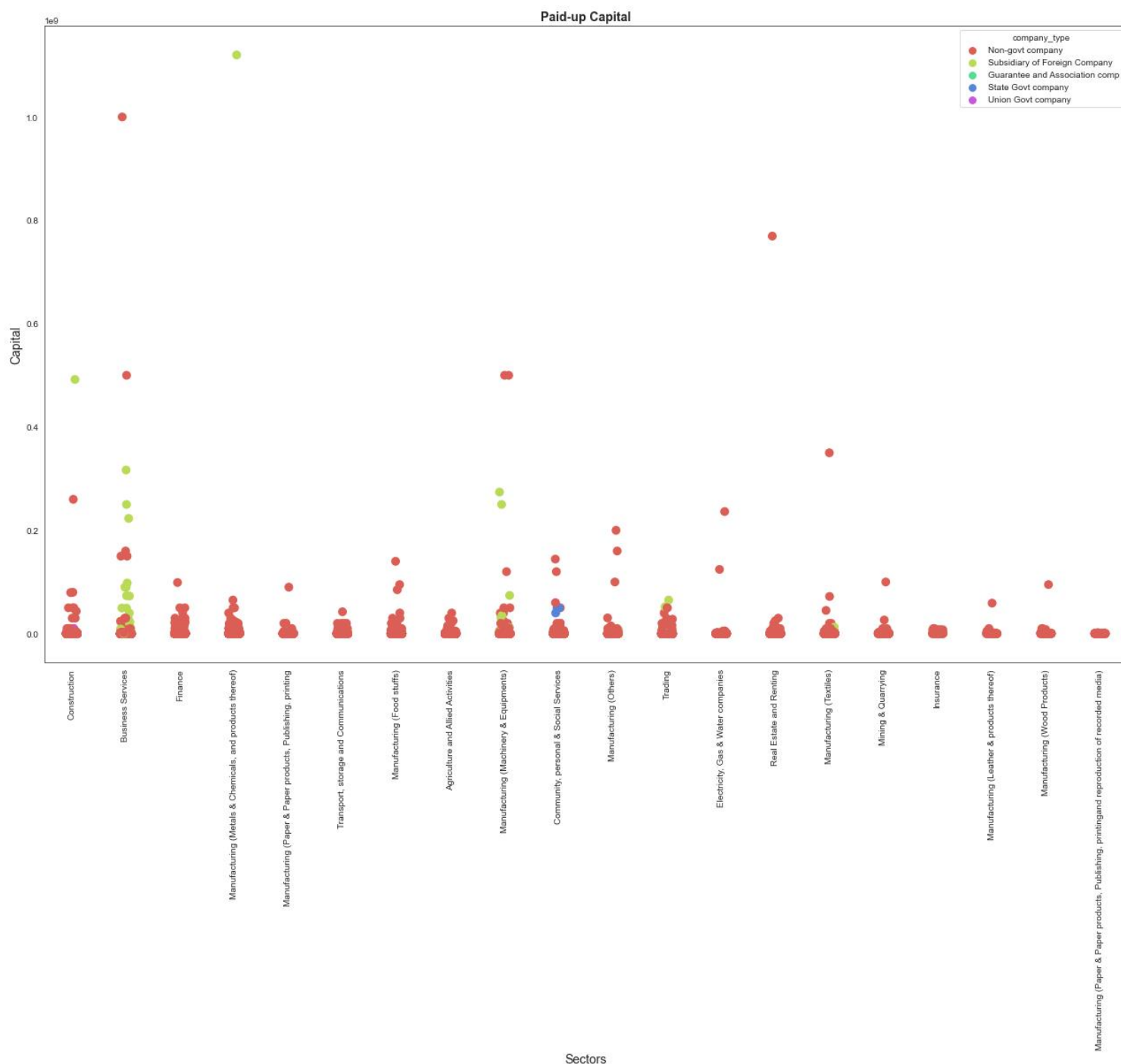
Paid-up Capital

This Jitter Plot shows the paid-up capitals of all the companies of the dataset with different colours for different types of companies.
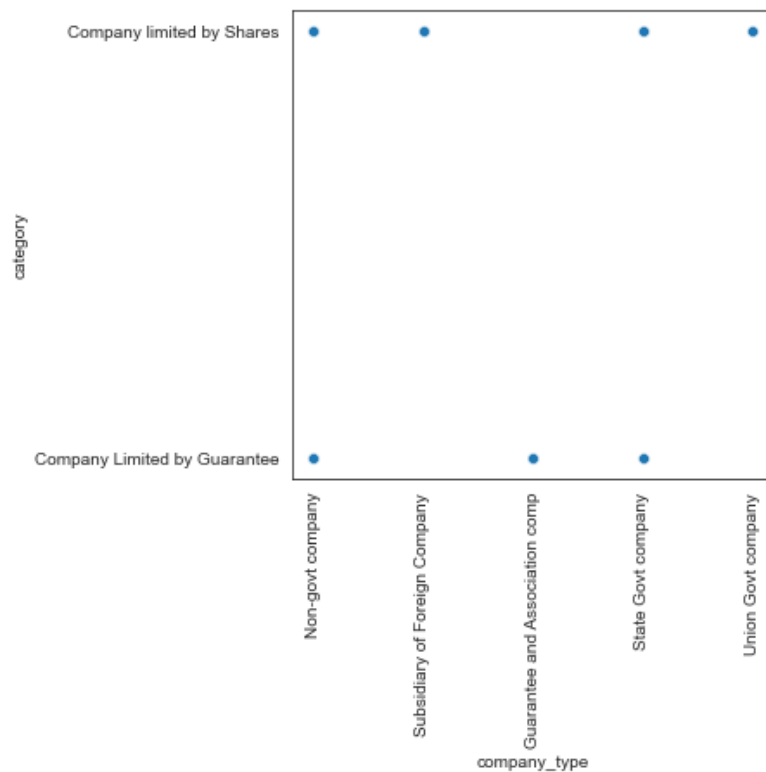
We can observe that most of the foreign companies have more paid-up capital than Indian companies, since more that 65% of foreign companies are in Business Services sector and their paid-up capitals are more than most of the other companies, we can conclude that foreign companies are performing quite well in India.

Comparision between sectors

This shows the difference between authorized and paid-up capitals of foreign companies. We can observe that for most of the sectors, authorized and paid-up capitals are same but in Business services there is a good difference between them indicating potential of growth in this sector for foreign companies.

Category of Registered Companies
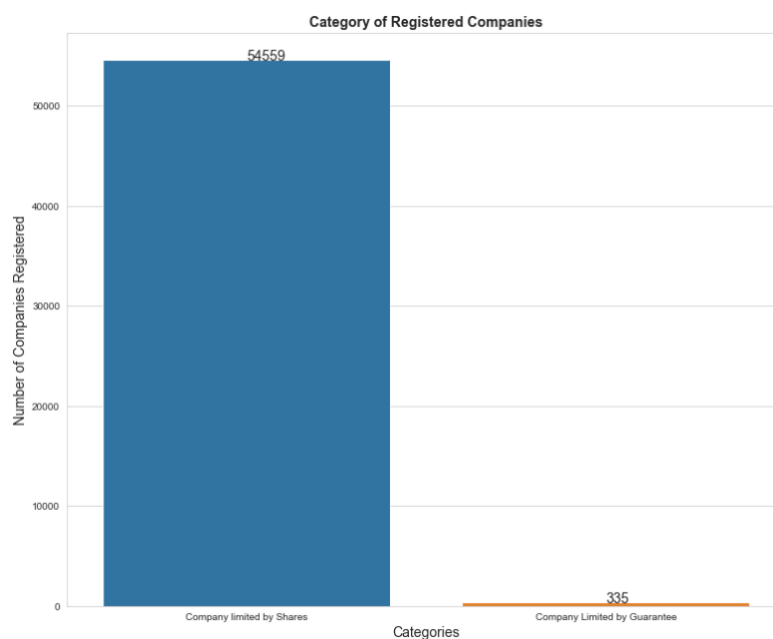
This plot shows that more than 99% of the companies are limited by share in Indian. This is because of the fact that these companies make profits but a company limited by guarantee is a non-profit company.

This scatter plot shows that none of the subsidiaries of foreign companies and Union Government Companies are limited by guarantee.
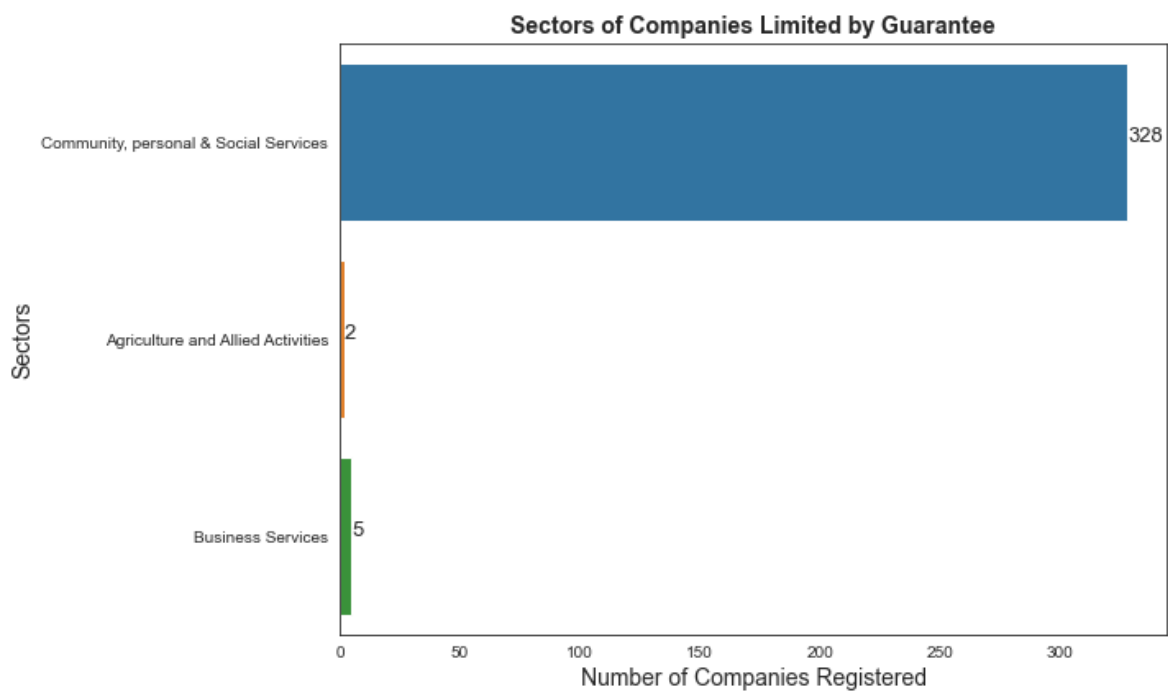
From this plot, we can observe that more than 99% of the companies limited by guarantee are in the Community, personal & Social Services sector because these companies are generally non-profit companies.