

AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of companies (RoC)

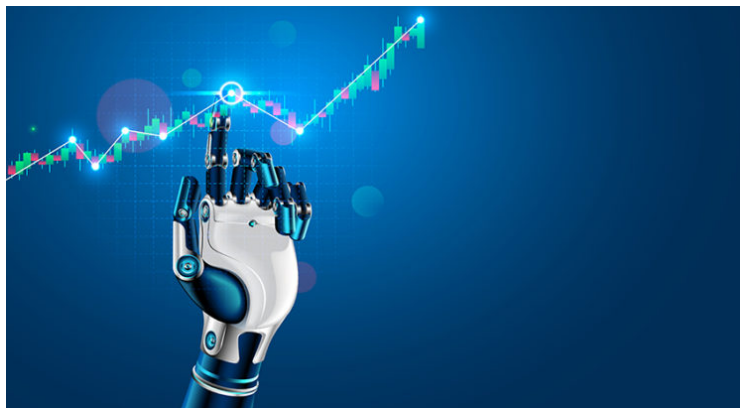
TEAM MEMBER

AU620121243057 :

VIGNESH .G.A

Phase 2 submission Document

Project: AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of companies (RoC)



Introduction:

- In today's fast-paced business landscape, the Registrar of Companies (RoC) plays a pivotal role in tracking company registrations. With the advent of Artificial Intelligence (AI), companies are now harnessing its power to gain insights, streamline processes, and predict trends related to RoC registrations. This AI-driven approach revolutionizes how businesses interpret and anticipate company registration data, offering valuable insights for informed decision-making and regulatory compliance. In this discussion, we will delve into the exciting realm of AI-driven explanations and predictions of company registration trends with RoC, exploring its significance and potential impact on various industries.

- Certainly, let's delve deeper into the topic of AI-driven explanation and prediction of company registration trends with the Registrar of Companies (RoC):

1. ***Data Analysis and Insights:*** AI algorithms are capable of processing vast amounts of historical data from RoC registrations. By analyzing this data, AI can identify patterns, trends, and correlations that might not be immediately apparent to human analysts. This allows businesses and regulatory bodies to gain deeper insights into the dynamics of company registrations.
2. ***Forecasting Trends:*** AI can predict future registration trends based on historical data and current market conditions. This forecasting capability is invaluable for businesses to make strategic decisions, such as identifying potential growth areas or anticipating market shifts.
3. ***Compliance and Risk Management:*** AI can help companies stay compliant with regulatory requirements by flagging discrepancies or anomalies in registration data. This proactive approach reduces the risk of non-compliance and associated penalties.
4. ***Efficiency and Automation:*** AI-powered systems can automate many registration-related tasks, reducing administrative burdens. This not only saves time and resources but also minimizes errors in the registration process.
5. ***Customer Insights:*** Companies can use AI to analyze registration data to better understand their customer base. This knowledge can inform marketing strategies, product development, and customer relationship management.
6. ***Fraud Detection:*** AI can detect fraudulent registrations by identifying unusual patterns or discrepancies in registration data. This is crucial for maintaining the integrity of the registration system.
7. ***Predictive Analytics:*** AI can provide predictive analytics on factors such as industry-specific registration trends, geographical variations, and economic indicators. This information is valuable for businesses looking to expand or enter new markets.

8. ***Government and Policy:*** Governments and regulatory authorities can use AI to monitor economic trends and make data-driven policy decisions. This can lead to more efficient allocation of resources and better economic planning.

9. ***Investor Confidence:*** Accurate prediction and explanation of registration trends can boost investor confidence by providing transparent and reliable information about market conditions.

10. ***Challenges:*** While AI offers immense potential, there are challenges, such as data privacy concerns, the need for high-quality data, and the requirement for skilled AI professionals to develop and maintain these systems.

- In conclusion, AI-driven explanation and prediction of company registration trends with the RoC represent a transformative leap in data analysis and regulatory compliance. The ability to extract meaningful insights, forecast future trends, and enhance efficiency has the potential to revolutionize how businesses and governments interact with registration data, ultimately leading to more informed decisions and improved economic outcomes.

Content for phase 2:

We consider as if this AI-Driven prediction has to be performed by the method of “Ensemble Learning” by the corresponding dataset in many company details.

Data source:

The datasets consists of several of the company details to predict the following of company registration trends in the outcome of the following no of company details are there in the dataset by their company names, status, registration, state these all are included.

(<https://www.kaggle.com/datasets/rowhitsuwami/all-indian-companies-registration-data-1900-2019>)

Steps involved in AI-Driven Company Trends (RoC):

2. ***Data Cleaning*:** Clean the data to handle missing values, outliers, and inconsistencies. This step is crucial as the quality of your data directly impacts the accuracy of your prediction.

- ***Descriptive Statistics***: Calculate summary statistics like mean, median, and standard deviation.

- ***Data Visualization*:** Create plots and charts (e.g., time series plots, histograms, scatter plots) to visualize the data's distribution and patterns.

- ***Correlation Analysis***: Determine relationships between variables using correlation matrices to understand which factors might impact ROC.

4. ***Feature Engineering***: Create new features or transform existing ones to better represent the relationships in your data. For example, you could calculate financial ratios or moving average.
5. ***Model Selection***: Choose an appropriate machine learning or statistical model for predicting ROC. Common models for time series data include ARIMA, LSTM, or even regression models.
6. ***Data Splitting***: Split your data into training and testing sets to evaluate the model's performance accurately.
7. ***Model Training***: Train the chosen model using the training data, considering appropriate hyperparameters.
8. ***Model Evaluation***: Evaluate the model's performance on the testing data using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).
9. ***Model Interpretation***: Analyze the model's coefficients or feature importances to understand which variables are most influential in predicting ROC trends.
10. ***Deployment and Monitoring***: If the model performs well, deploy it to make predictions for future data. Continuously monitor the model's performance and retrain it as needed.

Remember that predicting financial trends is a complex task, and the accuracy of your prediction will depend on the quality and relevance of the data you have and the model you choose. Additionally, consider external factors like economic events or changes in company strategy that can impact ROC.

Exploring the Data:

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: df = pd.read_csv('../input/all-indian-companies-registration-data-1900-2019/registered_companies.csv')
```

The corresponding libraries are imported.

```
/opt/conda/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3063: Dtype
Warning: Columns (10) have mixed types.Specify dtype option on import or set low_me
mory=False.
    interactivity=interactivity, compiler=compiler, result=result)
```

```
In [3]: df.head()
```

```
Out[3]:
```

#to get the number of columns by the following dataset in the corresponded company-RoC driven data.

```
In [4]: len(df)
```

```
Out[4]: 1992170
```

```
In [5]: df.info()
```

The datasets consists of several of the company details to predict the following of company registration trends in the outcome of the following no of company details are there in the dataset by their company names, status, registration, state these all are included.

[illegible]

company details

...

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
CORPORATE	COMPANY NAME	COMPANY STATUS	COMPANY CLASS	COMPANY COMPANY	DATE OF REGISTRATION	REGISTERED STATE	AUTHORIZED OFFICER	CARNICULUS	PRINCIPAL REGISTERED OFFICE ADDRESS	REGISTRAR	EMAIL ADDRESS	LATEST YEAR	YEAR FINANCIAL STATEMENT					
1	P01400	NKO RESOURCES LIMITED	ACTV	WORLDWIDE RESOURCES HOLDING		Guarat	0	0	Agriculture BATHK ROAD LANDMARRAC COURSE RD DELH	0	SANAYAN@NODIA.COM							
2	F0189	WORLDWIDE RESOURCES HOLDING	ACTV		16-07-1998	Guarat	0	0	Agriculture ROAD 1 & 3/22 MAHARAJA HOUSE DELH	0								
3	F01579	ATKINS OCEANIC PACIFIC UALY	ACTV			Guarat	0	0	Agriculture Roadant 1 Post Number 18, 2nd Floor SROK DELH	0	info@nalgina.com							
4	F01802	SOLVAY S A BELGIUM	NAEF		25-04-2001	Guarat	0	0	Agriculture Plot No. 3224 43177600 Industrial SROK DELH	0	amit.kumar@shivay.com							
5	F01889	TATE AND LYLE INDUSTRIES UALY	ACTV		21-07-2000	Guarat	0	0	Agriculture B2 & 1 MALPOTI NAGARELLA RAIN, ROC DELH	0								
6	F01847	OVERSEAS TRADING AND SHIPACT	ACTV		8/07/2001	Guarat	0	0	Agriculture B28 S-140 WARD 12 AOPP BANAR SROK DELH	0	mp@pco@delh.com							
7	F01881	BALLIST MADAM INTERNATIONAL	ACTV		31-12-2001	Guarat	0	0	Agriculture CHAMADHAR B34 URM SOCIETY PVRIC DELH	0								
8	F01958	OL AND GAS EXPLORATION COMACTY	ACTV		4/6/2000	Guarat	0	0	Agriculture B05 SAMUDRA ANNEKT OFF C GROUND DELH	0								
9	F02039	PRECISION DRILLING CYPRUS UALY	ACTV			Guarat	0	0	Agriculture B3RD FLOOR LANDMARR BURESSA SROK DELH	0								
10	F02048	OPRI, OL AGAS LIMITED	ACTV			Guarat	0	0	Agriculture B3rd Floor, 7 NING, Sagar Plaza SROK DELH	0	taras@taraneem.com							
11	F02090	OL AND GAS DRILLING COMPANY	ACTV			Guarat	0	0	Agriculture MA-8 BROADWAY, 40, WEST HAINA SROK DELH	0	og@noda@nafa.net.pl							
12	F02149	MAJKA ENTERPRISES PTE UALY	ACTV			Guarat	0	0	Agriculture B8 41 CORPORATE HOUSE, AUGE SIBURK DELH	0								
13	F02164	JCE NAFINARA	ACTV			Guarat	0	0	Agriculture B6 48 INTA DEPENT APARTMENT SIBURK DELH	0								
14	F02170	MAJKA ENERGY LLC	ACTV			Guarat	0	0	Agriculture B22 FLEASURE TERRACE, 18-VISVAL SROK DELH	0	p.j.nangse@agm.com							
15	F02249	ASENCE INC	ACTV			Guarat	0	0	Agriculture BARABAI CAMPUS D VIKRAMAR SROK DELH	0	dd@delh@barabai.com							
16	F02289	DURANT & CO	NAEF		19-01-2004	Guarat	0	0	Agriculture B27 DEWTH COMPLEX LA BUNGROK DELH	0	anash@starlin-durant.com							
17	F02389	QUAD TECH INC	ACTV			Guarat	0	0	Agriculture B22 FLEASURE TERRACE, 18-VISVAL SROK DELH	0	aashita_83@yahoo.com							
18	F02776	OLEX LTD	ACTV		12/5/2006	Guarat	0	0	Agriculture B3RD FLOOR RADH ARCADE, GANDH SROK DELH	0	vik@agale@ole.com.au							
19	F02833	ANDERSON KALOY ASIA THACTY	ACTV		4/10/2006	Guarat	0	0	Agriculture B33A PINACOLE, OPP ROYAL ORCHARD DELH	0	m.sangh@shivay.com							
20	F02902	NUOVO PINONE S.P.A.	NAEF		22-12-2006	Guarat	0	0	Agriculture PIAON Pignone Ahmedabad PIGNAR SROK DELH	0	netra.babbar@tgc.com							
21	F02902	XO SOFTWARE LLC	NAEF		01-07-2007	Guarat	0	0	Agriculture B202, PRESIDENT HOUSE MAMBAVA SROK DELH	0	info@xosoftware.com							
22	F02916	HEATRE SOLUTIONS INC	NAEF			Guarat	0	0	Agriculture B02, AMBAPAL COMPLEX, NR 45 MAHIC DELH	0	dm@tgm@delh.com							
23	F03003	OLEX NL HOLDINGS (INDIA) UALY	ACTV		26-04-2007	Guarat	0	0	Agriculture B3RD FLOOR, BLOCK CHAMRADE ARCADE SROK DELH	0	jayaram@ambales.com							
24	F03019	KST INTERNATIONAL LTD	ACTV		14-08-2007	Guarat	0	0	Agriculture EPR, 7th Floor, THE CHAMBERS S H SROK DELH	0	dipankar@delh.com.my							
25	F03139	ONIT STOCK COMEX ZANGACTY	ACTV		01-09-2007	Guarat	0	0	Agriculture B4 & 1 GOC Estate Sector - 25 ROC DELH	0	supriyanti@delh.com							
26	F03149	GRADA TEXTILE OMBI INC	ACTV		21-09-2007	Guarat	0	0	Agriculture EPR76/72, Near Narayan Tias Sander NARIC DELH	0	rajuluna@beaondea.de							
27	F03200	CETIC GLOBS HOLDINGS INC	NAEF		14-11-2007	Guarat	0	0	Agriculture K2ND FLOOR, NEW DELHY VIMVAY COMPLEX SROK DELH	0	taras@delh.de							
28	F03230	SHANDONG TIE LAIN ELECTRIC FACTY	ACTV		18-10-2007	Guarat	0	0	Agriculture BADAN PONSER LTD PONSER S23 SIBURK DELH	0	mnp@pco@delh.com							
29	F03320	HYPLAS MACHINERY CO LIMITED	ACTV			Guarat	0	0	Agriculture B01 SHAPATH 1/N, OPPOSITE KARMA SROK DELH	0	uday@noda@delh.com							
30	F03340	PRIMETALS TECHNOLOGIES PRIVATE	ACTV		27-05-2008	Guarat	0	0	Agriculture KCI Naina Park Mills Limited 2m Km BOC DELH	0	administration.pco@primetals.com							

In this clear column has to represent the following detail about the company RoC by the assumption of 35 company details.

```
In [6]: df.describe().transpose()
```

Out[6]:

	count	mean	std	min	25%	50%	75%	max
AUTHORIZED_CAP	1992170.0	4.238508e+07	2.960562e+09	0.0	100000.0	500000.0	1500000.0	1.850
PAIDUP_CAPITAL	1992170.0	2.434621e+07	2.313154e+09	0.0	100000.0	100000.0	502000.0	1.699

In this output value has 1 means the following company to goes the range of prediction as the higher same as the output value has 0 means the prediction can lower in the trends.

Regression:

```
In [7]: df['DATE_OF_REGISTRATION']
```

Out[7]:

```
0      NaN
1    16-07-1998
2      NaN
3    25-06-2001
4    25-07-2001
...
1992167  25-07-2000
1992168   09-07-1998
```

```
In [10]: df['DATE_OF_REGISTRATION'].dropna(axis=0, inplace=True)
```

```
In [11]: df['DATE_OF_REGISTRATION'].count()
```

Out[11]:
1989645

In the next stage of process has to be used for the regression in machine learning model that the Exploratory Data Analysis can be performed.


```
In [12]: df['DATE_OF_REGISTRATION']=pd.to_datetime(df['DATE_OF_REGISTRATION'], errors = 'coerce', format='%d-%m-%Y')
```

```
In [13]: df['YEAR_REG']=pd.DatetimeIndex(df['DATE_OF_REGISTRATION']).year
```

```
In [14]: number_of_companies = df[['YEAR_REG', 'CORPORATE_IDENTIFICATION_NUMBER']]
```

```
In [16]: number_of_companies=number_of_companies.groupby(by='YEAR_REG').size().reset_index(name='No_of_companies')
```

```
In [17]: number_of_companies
```

Output:

Out[17]:

	YEAR_REG	No_of_companies
0	1857.0	1
1	1863.0	3
2	1871.0	3
3	1872.0	3
4	1873.0	3
...
146	2016.0	93876
147	2017.0	107635
148	2018.0	117924

Import Mathplotlib:

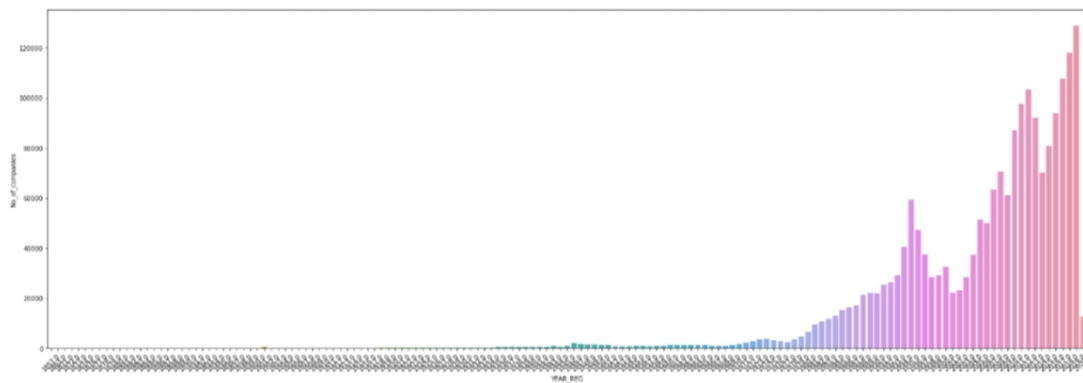
```
In [18]: import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In this matplotlib is used to show our predicted output in a graph.

```
In [19]: plt.figure(figsize=(30,10))
plt.xticks(rotation=45)
sns.barplot(data=number_of_companies,x='YEAR_REG',y='No_of_companies')
```

```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2c8c5c6190>
```

In this following output as,



Training model:

```
In [20]: no_of_companies_state = df[['YEAR_REG', 'CORPORATE_IDENTIFICATION_NUMBER', 'REGISTERED_S  
TATE']]
```

```
In [21]: no_of_companies_state.dropna(axis=0, inplace=True)
```

It measures the year of register in the company (RoC).

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarn  
ing:  
A value is trying to be set on a copy of a slice from a DataFrame  
  
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/  
user\_guide/indexing.html#returning-a-view-versus-a-copy  
"""Entry point for launching an IPython kernel.
```

```
In [22]: no_of_companies_state=no_of_companies_state.groupby(by=['YEAR_REG', 'REGISTERED_STAT  
E']).size().reset_index(name='no_of_companies')
```

```
In [23]: no_of_companies_state
```

Output:

Out[23]:

	YEAR_REG	REGISTERED_STATE	no_of_companies
0	1857.0	Rajasthan	1
1	1863.0	Maharashtra	1
2	1863.0	West Bengal	2
3	1871.0	Maharashtra	2
4	1871.0	West Bengal	1
...
2931	2020.0	Telangana	788
2932	2020.0	Tripura	12
2933	2020.0	Uttar Pradesh	1260
2934	2020.0	Uttaranchal	138

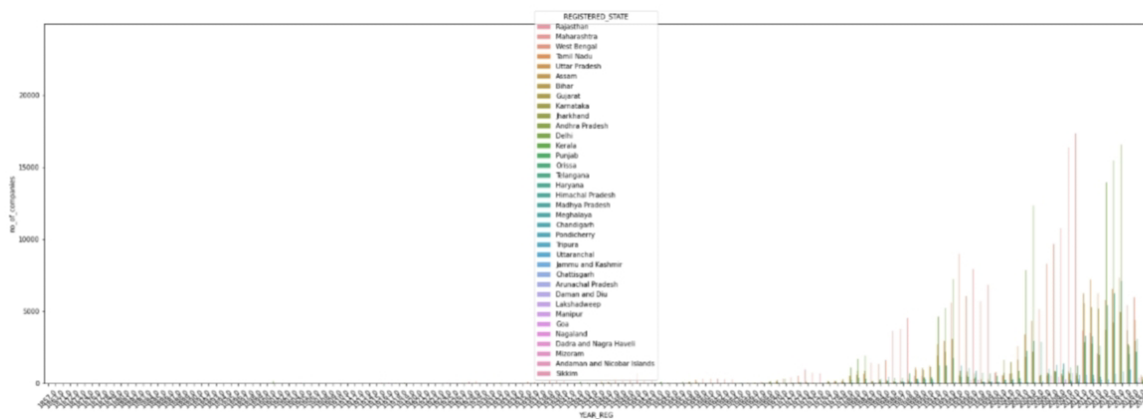
In [24]:

```
plt.figure(figsize=(30,10))
plt.xticks(rotation=45)
sns.barplot(data=no_of_companies_state, x='YEAR_REG', y='no_of_companies', hue='REGISTERED_STATE')
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2c8bdf9d90>
```

Output Graph:



```
In [25]:
state='Tamil Nadu'
plot_data = no_of_companies_state[no_of_companies_state['REGISTERED_STATE']==state]
plt.figure(figsize=(30,10))
plt.xticks(rotation=45)
sns.barplot(data=plot_data,x='YEAR_REG',y='no_of_companies',hue='REGISTERED_STATE')
```

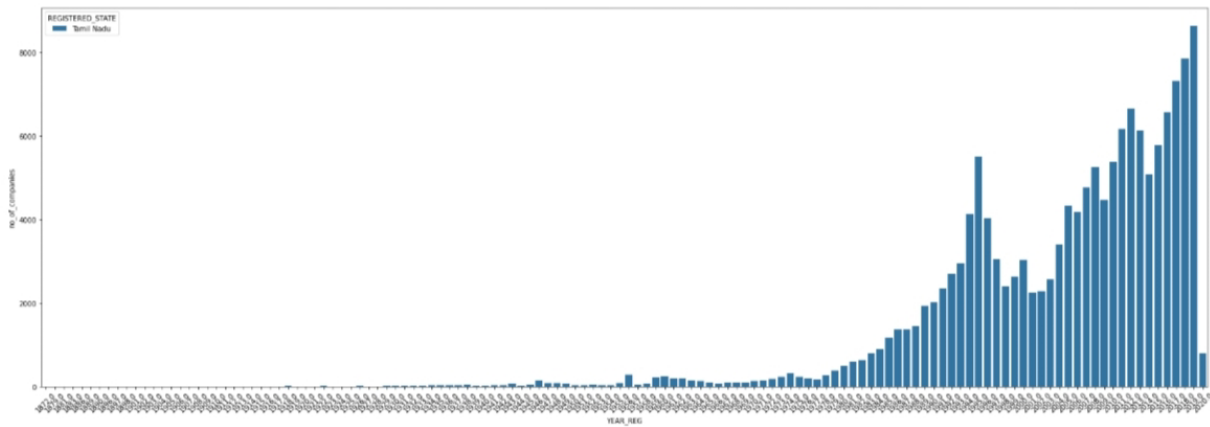
Out[25]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f2c83b0ba50>

REGISTERED_STATE

In this plot can be the following “TamilNadu” to the registered state of the company RoC.

Output Plot:



```
In [26]:
df['REGISTERED_STATE'].value_counts()
```

In this case the following data's of all states as,

```
Out[26]:
```

Maharashtra	395282
Delhi	348230
West Bengal	207005
Tamil Nadu	150871
Karnataka	125779
Uttar Pradesh	116337
Telangana	112401
Gujarat	105554
Rajasthan	61370
Kerala	56098
Haryana	51039
Madhya Pradesh	41318
Bihar	35920
Andhra Pradesh	33050
Punjab	32440
Orissa	27023
Jharkhand	15078
Chandigarh	15018
Assam	12078
Chhattisgarh	10703
Goa	8942
Uttaranchal	8924
Himachal Pradesh	6591
Jammu and Kashmir	5978
Pondicherry	3454
Meghalaya	1115
Manipur	1056
Tripura	645
Arunachal Pradesh	629
Nagaland	620
Dadra and Nagra Haveli	550
Andaman and Nicobar Islands	480
Daman and Diu	397
Mizoram	175

The plotted place in types of companies:

```
In [27]: pie_plot=df['PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN'][df['REGISTERED_STATE']==state].value_counts().rename_axis('Type of companies').reset_index(name='Count')
#pie_plot.columns=['Type of Companies','count']
#pie_plot.reset_index(name='Count')
```

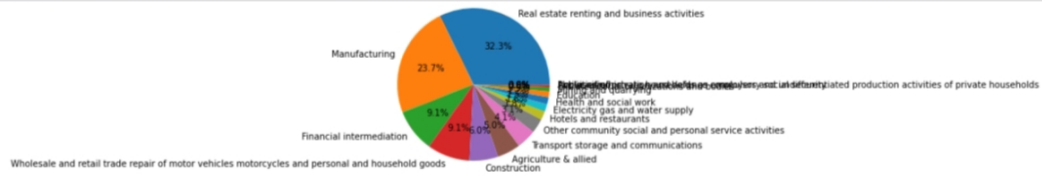
```
In [28]: pie_plot
```

Types of companies in all the states:

Out[28]:

	Type of companies	Count
0	Real estate renting and business activities	48697
1	Manufacturing	35757
2	Financial intermediation	13772
3	Wholesale and retail trade repair of motor veh...	13681
4	Construction	9079
5	Agriculture & allied	7496
6	Transport storage and communications	6231
7	Other community social and personal service ac...	4725
8	Hotels and restaurants	2673
9	Electricity gas and water supply	2459
10	Health and social work	2270
11	Education	1822
12	Mining and quarrying	1377
13	Extraterritorial organizations and bodies	781
14	Public administration and defence compulsory s...	27
15	Activities of private households as employers	19

Pie chat calculation for the above companies:

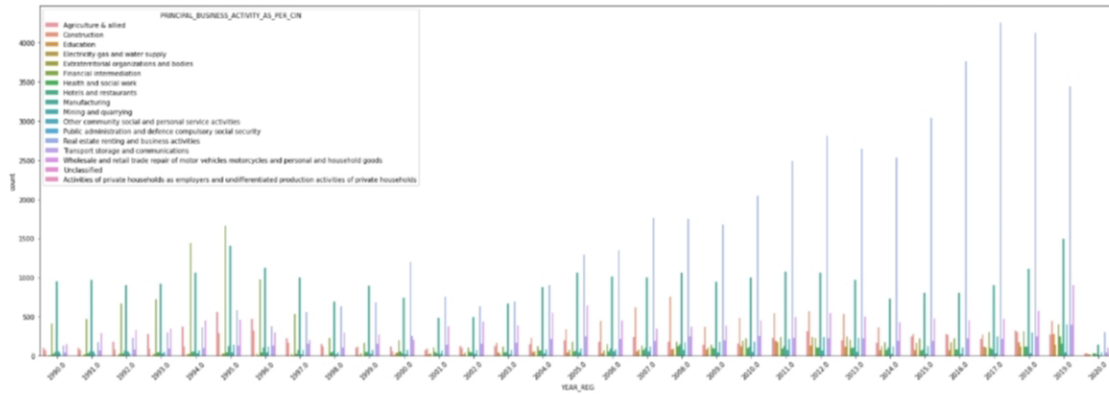


```
In [30]: state_details=df[['COMPANY_NAME', 'COMPANY_SUB_CATEGORY', 'PRINCIPAL_BUSINESS_ACTIVITY_A', 'S_PER_CIN', 'YEAR_REG']][df['REGISTERED_STATE']==state]
```

```
In [34]: plt.figure(figsize=(30,10))
plt.xticks(rotation=45)
sns.color_palette("hls", 8)
sns.barplot(data=state_details[state_details['YEAR_REG']>=1990], x='YEAR_REG', y='count', hue='PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN')
```

```
Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2c82a307d0>
```

The output Predicted company trends:



CONCLUSION:

1. In the Phase 2 conclusion, we will summarize the findings and insights from the advanced regression techniques. We will reiterate the impact of these techniques on improving the AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of companies (RoC).

Future Work: We will discuss potential avenues for future work, such as incorporating additional data sources (e.g., real-time economic indicators), exploring deep learning models for prediction, or expanding the project into a web application with more features and interactivity.