INDIAN COMPANIES REGISTRATION DATA [1857 – 2020]

WEB ADDRESS :

https://www.kaggle.com/datasets/rowhitswami/all-indian-companies-registration-data-1900-2019/code

ABOUT DATSET:

The dataset provides details of any company registered with Registrar of Companies (RoC). Data contains various information like Corporate Identification Number(CIN), Company Name, Company Status, Company Class, Company Category, Authorized Capital in INR, Paid-up Capital in INR, Date of Registration, Registered State, Registrar of Companies, Principal Business Activity, Registered Office Address and Sub Category.

DATA CONTENT:

- The registered_companies.csv(592.9 MB) file contains the details of all Indian companies ever registered in India from 1857 to 2020.
- This dataset is a structured dataset containining 1992170 rows and 17 columns.

HIVE :

- Apache Hive is a data warehouse software built on top of Apache Hadoop for providing data query and analysys.
- Uses HQL(Hive Query Language) used to communicate with databases.
- Here we use apache hive for data analysis.

ANALYTICAL QUESTIONS:

1) Top 5 states with company regisration?
2) Count the statuses of companies?
3) Which business activity have most number of company registarions?
4) Find the comapnies registerd as manufacturing companies in kerala and also their current status?
5) Find the highest authorized capital per state?

CREATE TABLE:

- A table is created from the dataset based on the available columns.

Query:

create table companies(CORPORATE_IDENTIFICATION_NUMBER string,COMPANY_NAME string,COMPANY_STATUS string,COMPANY_CLASS string,COMPANY_CATEGORY string,COMPANY_SUB_CATEGORY string,DATE_OF_REGISTRATION string,REGISTERED_STATE string,AUTHORIZED_CAP float,PAIDUP_CAPITAL float,INDUSTRIAL_CLASS float,PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN string,REGISTERED_OFFICE_ADDRESS string,REGISTRAR_OF_COMPANIES string,EMAIL_ADDR string,LATEST_YEAR_ANNUAL_RETURN string ,LATEST_YEAR_FINANCIAL_STATEMENT string)ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'WITH SERDEPROPERTIES ('separatorChar' = ',','quoteChar' = '\"')tblproperties('skip.header.line.count'='1');

Load data:

*load data local inpath '/home/edwindavid/data/companyregproject/archive(23)/registered_companies.csv' into table companies;

1).TOP 5 STATES WITH COMPANY REGISRATION?

Partition:

- Partition are dividing table into some parts based on values of columns.
- The advantage of using partition is to make queries faster.
- Here the parttiton is based on the column Registered state.

Query:

*create table companies_newpartition(CORPORATE_IDENTIFICATION_NUMBER string,COMPANY_NAME string,COMPANY_STATUS string,COMPANY_CLASS string,COMPANY_CATEGORY string,COMPANY_SUB_CATEGORY string,DATE_OF_REGISTRATION string,AUTHORIZED_CAP float,PAIDUP_CAPITAL float,INDUSTRIAL_CLASS float,PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN string,REGISTERED_OFFICE_ADDRESS string,REGISTRAR_OF_COMPANIES string,EMAIL_ADDR string,LATEST_YEAR_ANNUAL_RETURN string ,LATEST_YEAR_FINANCIAL_STATEMENT string)partitioned by(REGISTERED_STATE string)row format delimited fields terminated by ',';

*set hive.exec.dynamic.partition=true;

*set hive.exec.dynamic.partition.mode=nonstrict;

*insert into companies_newpartition partition(REGISTERED_STATE) select CORPORATE_IDENTIFICATION_NUMBER,COMPANY_NAME,COMPANY_STATUS,COMPANY_CLASS,COMPANY_CATEGORY,COMPANY_SUB_CATEGORY,DATE_OF_REGISTRATION,AUTHORIZED_CAP,PAIDUP_CAPITAL,INDUSTRIAL_CLASS,PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN,REGISTERED_OFFICE_ADDRESS,REGISTRAR_OF_COMPANI

ES,EMAIL_ADDR,LATEST_YEAR_ANNUAL_RETURN,LATEST_YEAR_FINANCIAL_STATEMENT,REGISTERED_STATE from companies;

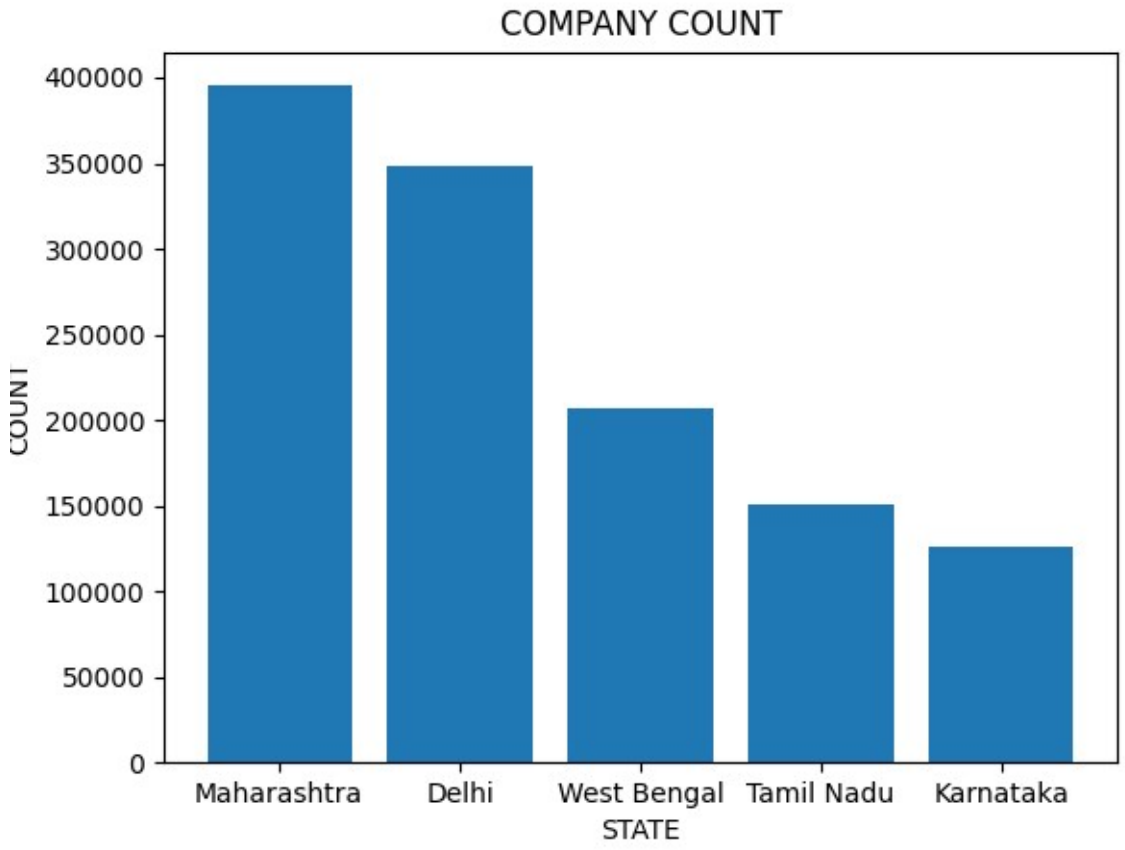
*select registered_state,count(registered_state) as total_count from companies_newpartition group by registered_state order by total_count desc limit 5;

Result:

registered_state	total_count
Maharashtra	395282
Delhi	348230
West Bengal	207005
Tamil Nadu	150871
Karnataka	125779

Conclusion:

- This means that the state with the highest number of registered companies (Maharashtra) is listed first, followed by Delhi, West Bengal, Tamil Nadu, and Karnataka.
- This result provides valuable insights into the distribution of registered companies across different states, with a focus on the states with the highest company counts.



2).FIND THE NUMBER OF NON GOVERNMNET COMPANIES WHICH ARE STILL FUNCTIONING IN EACH STATE ?

Query:

*select registered_state,count(COMPANY_STATUS) as active_comapanies from companies where COMPANY_STATUS='ACTV' and COMPANY_SUB_CATEGORY='Non-govt company' group by registered_state;

Result:

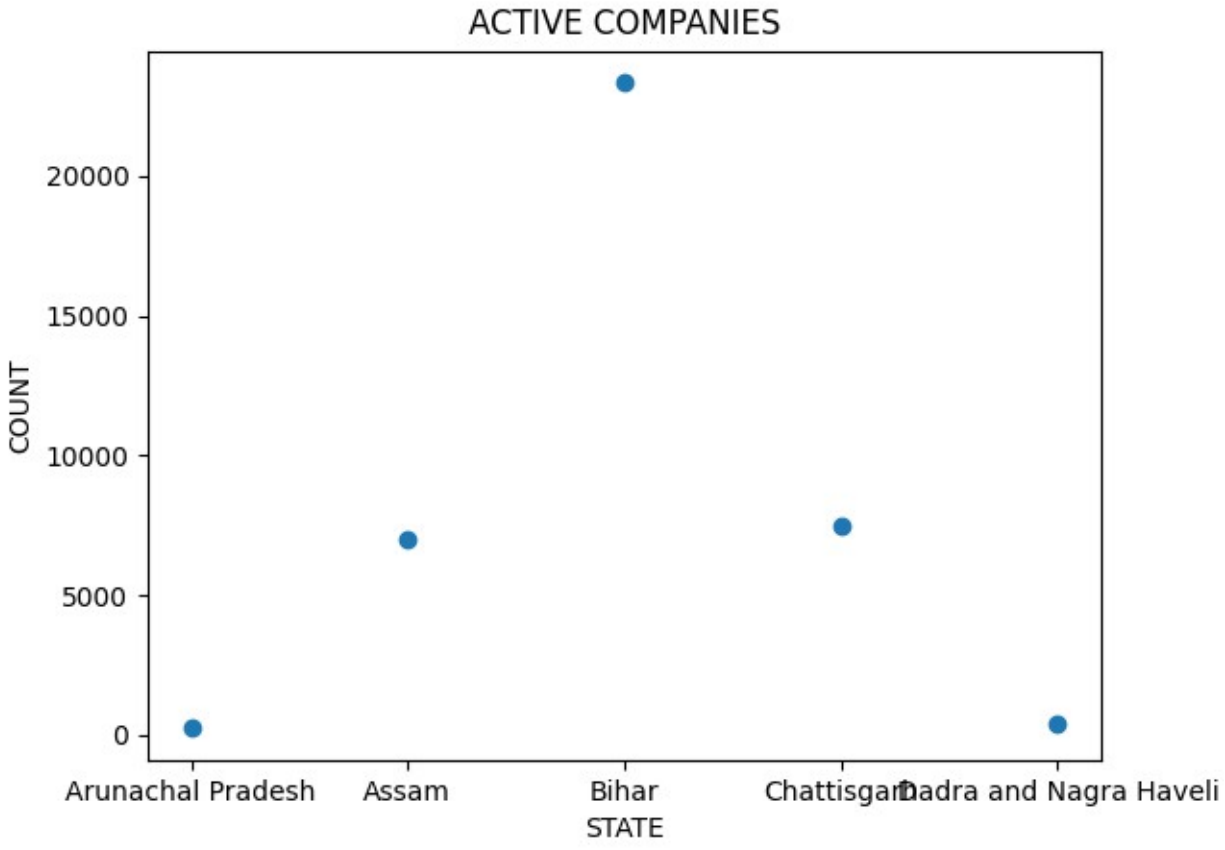
registered_state	active_comapanies
Daman and Diu	249
Jharkhand	10509
Karnataka	75931
Kerala	33276
Lakshadweep	8
Orissa	16252
Pondicherry	1317
Rajasthan	37313
Sikkim	2
Tripura	443
Andaman and Nicobar Islands	351
Andhra Pradesh	19962
Chandigarh	6974
Gujarat	63391
Haryana	34817
Jammu and Kashmir	3164
Madhya Pradesh	24302
Meghalaya	532
Telangana	64056
Uttar Pradesh	78001
West Bengal	125788
Arunachal Pradesh	247
Assam	7019
Bihar	23321
Chattisgarh	7488
Dadra and Nagra Haveli	390
Delhi	200284
Goa	4297
Himachal Pradesh	3524
Maharashtra	228094
Manipur	744
Mizoram	94
Nagaland	236
Punjab	16711
Tamil Nadu	77381
Uttaranchal	6104

Conclusion:

- The table reflects the economic activity in different states and union territories, with higher counts of active non-government companies in regions like Maharashtra,Delhi and TamilNadu.

- Urban areas, especially major cities in states like Delhi, Mumbai (in Maharashtra), Chennai (in Tamil Nadu), Bengaluru (in Karnataka), and Ahmedabad (in Gujarat), contribute significantly to the overall count of active companies.

- Regions with lower counts may represent areas with potential for growth or areas that are actitvely working to attract more business investments.



3).WHICH BUSINESS ACTIVITY HAVE MOST NUMBER OF COMPANY REGISTARIONS?

Query:

*select PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN,count(PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN) as count from companies group by PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN order by count desc;

Result:

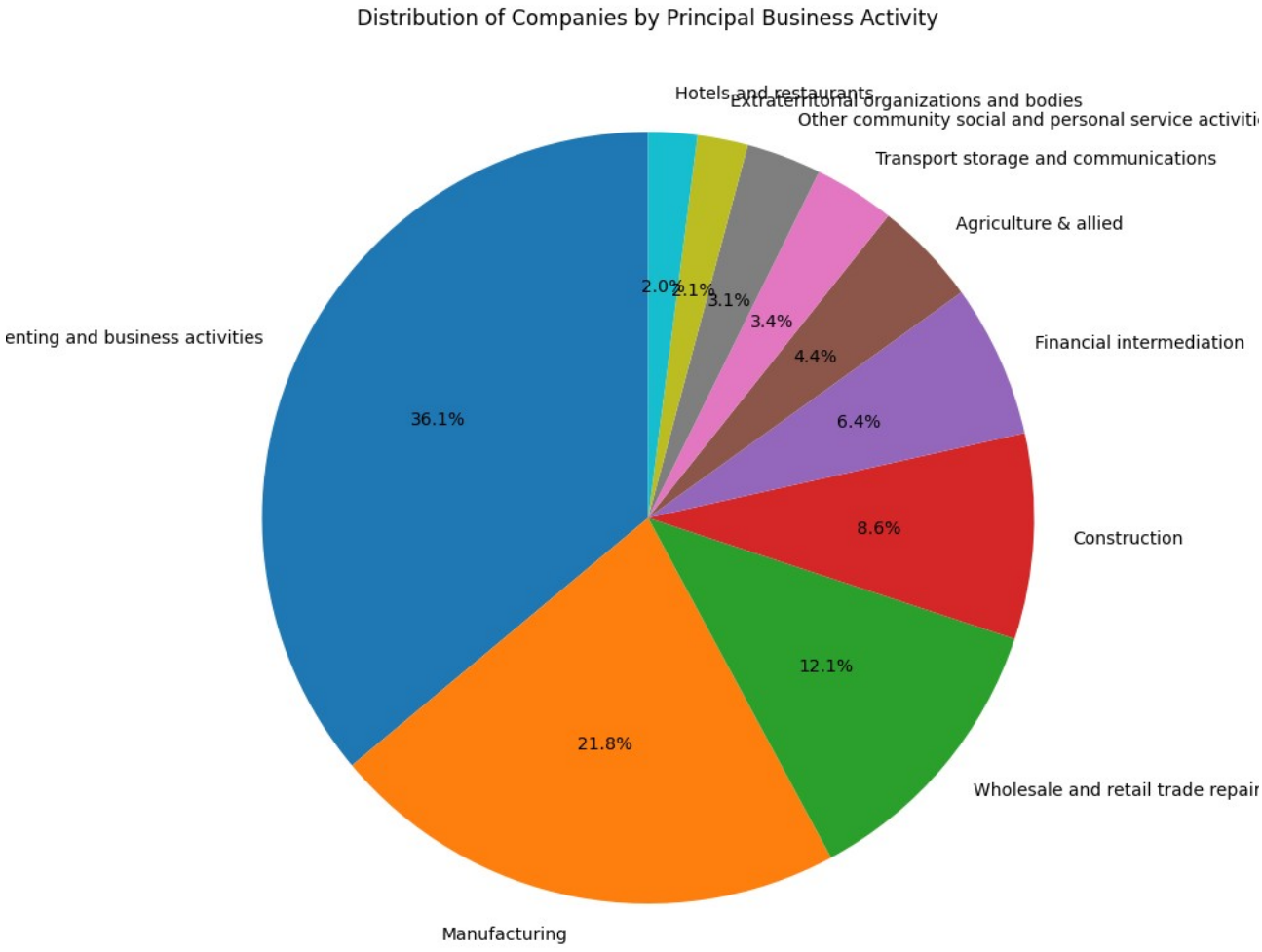
principal_business_activity_as_per_cin	count
Real estate renting and business activities	679955
Manufacturing	410402
Wholesale and retail trade repair of motor vehicles motorcycles and personal and household goods	227333
Construction	162125
Financial intermediation	120714
Agriculture & allied	82572
Transport storage and communications	63738
Other community social and personal service activities	59282
Extraterritorial organizations and bodies	39897
Hotels and restaurants	38518
Health and social work	34489
Education	28532
Electricity gas and water supply	22258
Mining and quarrying	20310
Public administration and defence compulsory social security	850
Unclassified	787
Activities of private households as employers and undifferentiated production activities of private households	396

Conclusion:

- The data reflects the diversity and complexity of the business landscape in India, with companies involved in a wide array of economic activities.
- The most common principal business activity is "Real estate renting and business activities," with a substantial count of 679,955 companies.
- "Real estate renting and business activities" emerge as the most common principal business activity, indicating a significant presence of companies engaged in real estate-related services and business activities.

- The dataset reflects a diverse economic landscape, with companies involved in various sectors, including manufacturing, wholesale and retail trade, construction, financial intermediation, agriculture and more.

- Certain categories, such as public administration and defense, activities of private households, and extraterritorial organizations, have lower counts, suggesting a specialization or lower prevalence in these areas.



4).FIND THE COMAPNIES REGISTERD AS MANUFACTURING COMPANIES IN KERALA AND ALSO THEIR CURRENT STATUS?

Query:

*select company_name,company_status from companies where registered_state='Kerala' and PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN='Manufacturing' limit 10;

Result:

company_name	company_status
LAKSHMI ACETYLENE PVT LTD	ACTV
SOUTHERN ACETYLENE PVT LTD	ACTV
ARABIAN GULF GASES PRIVATE LIMITED	STOF
YESBEE CARBONIC PRIVATE LIMITED	ACTV
COSMOS OXYGEN PRIVATE LIMITED	STOF
MANORAMA OXYGEN PRIVATE LIMITED	ACTV
SAJEV AIR PRODUCTS PRIVATE LIMITED	STOF
BASIC GASES PRIVATE LIMITED	STOF
VIGHNESWARA OXYGEN PRIVATE LIMITED	ACTV
C.V.P. GAS COMPANY PRIVATE LIMITED.	STOF

Conclusion:

- The companies listed with the status "ACTV" are currently active and operational.
- The companies with the status "STOF" are in the process of being struck off.
- There is a mix of active and strike-off companies, indicating a dynamic business environment with both ongoing operations and companies that might be undergoing closure processes.
- The companies listed are involved in manufacturing activities, as indicated by the specified principal business activity. This suggests a presence of manufacturing industries in Kerala.

5).FIND THE HIGHEST AUTHORIZED CAPITAL PER STATE?

Query:

*select registered_state,max(authorized_cap) as highest_capital from companies group by registered_state;

Result:

registered_state	highest_capital
Daman and Diu	973500000.0
Jharkhand	999900.0
Karnataka	9999999.0
Kerala	99999.0
Lakshadweep	500000.0
Orissa	9900000.0
Pondicherry	9900000.0
Rajasthan	999990.0
Sikkim	100000.0
Tripura	990000.0
Andaman and Nicobar Islands	900000.0
Andhra Pradesh	9999999.0
Chandigarh	99000000.0
Gujarat	999999.0
Haryana	999990.0
Jammu and Kashmir	99900.0
Madhya Pradesh	99990.0
Meghalaya	9900000.0
Telangana	99999990.0
Uttar Pradesh	999990.0
West Bengal	999900.0
Arunachal Pradesh	9600000.0
Assam	99300000.0
Bihar	999990.0
Chattisgarh	99000000.0
Dadra and Nagra Haveli	9000000.0
Delhi	999999999990.0
Goa	9900000.0
Himachal Pradesh	99900.0
Maharashtra	99999999000.0
Manipur	9900000.0
Mizoram	70000000.0
Nagaland	95000.0
Punjab	999990.0
Tamil Nadu	999990.0
Uttaranchal	950000000.0

Conclusion:

- The highest authorized capital varies significantly across different states, ranging from relatively small values to extremely large ones.
- States like Delhi, Maharashtra, and Tamil Nadu have companies with exceptionally high authorized capital, which could be indicative of their strong economic presence.
- The authorized capital values can be influenced by the nature of industries and businesses operating in each state.