

# REGISTRARS OF COMPANIES DATA ANALYSIS - steps followed

## Step 1 : Importing the data

- The file was in csv format
- Imported the file into SSMS.

## Step 2: Understanding the data

- The data consisted of 1,88,085 rows and 16 columns related to companies and their registered details.
- Below is the information related to the data:
  1. CIN - It is a unique 12 - digit alphanumeric code assigned by the Registrar of Companies (ROC) to every company registered in India
  2. CompanyName: Official legal name that is approved and registered by ROC
  3. Company ROC Code: It identifies the jurisdiction of ROC office where the company is registered
  4. Company Category : Categorised based on liability structure of the company
    - Company limited by guarantee: Members' liability is limited to a fixed amount they promise to contribute in case the company is wound up
    - Unlimited Company: If the company's assets fall short in settling debts, members' personal assets can be used
    - Company limited by shares: Members' liability is limited to the amount unpaid on their shares
  5. CompanySubCategory: The data had the following sub categories -
    - Non - Government Company
    - Guarantee and association company
    - Union Government company
    - Subsidiary of company incorporated outside India
    - State government company
  6. CompanyClass: Whether the registered company is public, private or One person company
    - Private Company: Has as minimum of 2 and maximum of 200 shareholders. Shares are not publicly traded
    - Public Company: Minimum of 7 shareholders. Shares are traded publicly
    - One person Company: One shareholder and one director
  7. Authorized Capital :The maximum capital a company is legally allowed to raise from shareholders.
    - The maximum share capital a company is legally allowed to issue.
    - Set during incorporation and can be increased by resolution.
    - Determines the upper limit of shareholder equity.
    - Helps gauge the company's funding potential.
  8. Paid up Capital : The actual amount received by the company from shareholders against shares issued.
    - Cannot exceed authorized capital.
    - Reflects real equity investment into the company.
    - Important for analysing company's financial strength.
    - Determines voting rights and shareholding structure.
  9. Company Registration date: The official date the company was incorporated under the Companies Act.
    - Indicates how long the company has been in existence.
    - Useful for analyzing business maturity.
    - Helps track growth trends in different years.

10. Registered Office Address: The official location of the company's head office.

- All legal communication is sent to this address.
- May offer insight into company location (urban/rural).

11. Listingstatus: Indicates if the company's shares are listed on a stock exchange.

- Unlisted: Privately held company.
- Influences investor access and transparency.
- Unlisted companies dominate in small and mid-size segments.

12. Company Status: Current operational/legal status of the company.

- Common statuses Active, Strike Off, Dissolved, Amalgamated.
- *Active* means it's operational.
- *Strike Off* is due to non-compliance or inactivity.

13. CompanyStateCode: The Indian state where the company is registered.

- Indicates regional business activity.
- Useful for state-wise company mapping and policy analysis.
- Helps in jurisdictional tax filings and Roc tracking.
- Essential for local market studies.

14. Company Indian/Foreign Company: Specifies whether the company is Indian or foreign.

- Indicates if the company is Indian or a Foreign entity.
- All in the sample are Indian.

15. Nic Code : Categorize businesses and industries based on the nature of their economic activity.

16. Company Industrial Classification: Company Industrial Classification refers to the broad industry or sector in which a company operates, based on its primary business activity. It is typically derived from or aligned with the NIC Code (National Industrial Classification) but presented in a human-readable format.

### **Step 3 : Checked for abnormalities in the data**

An analysis was conducted on a dataset comprising 19,900 rows, including duplicates. Although all columns initially had correct data types, several data quality issues were identified and subsequently resolved.

1. Rename the Column: Checked and fixed column names to make them clear and consistent
2. Missing Values: Checked each column for null or missing entries and handled them appropriately
3. Incorrect Data Types: Verified that each column had the correct data type (e.g., dates, numbers, strings).

### **Step 4 : Cleaning the data**

1. Deleted Duplicate rows.
2. Filled the null values in the date column by propagating non-null values both forward and backward using forward fill and backward fill methods.
3. Null values in the Authorized Capital and Paid-up Capital columns were imputed using the mean values grouped by company class.
4. Renamed columns
5. Renamed data values that contain 'Manufacturing' along with various classifications, grouping them into a single 'Manufacturing' category
6. Extracted a new table comprising companies whose registration location is outside Bangalore, yet their state code is recorded as Karnataka.

**Step 5 : Analyzing the data**

Formulated a set of analytical questions and executed SQL queries to explore and analyse the dataset.

**Step 6 : Visualized the data**

1. Established a connection between SSMS and Power BI.
2. Utilized DAX functions for data analysis and calculations.
3. Designed and implemented interactive dashboards using suitable chart types.

**Step 7 :** Reported key findings and provided actionable recommendations based on the data analysis.

