A

PROJECT REPORT ON

"A STUDY OF STOCK MARKET IN INDIA"

(CASE STUDY OF INDUSTRIAL MACHINERY SECTOR)

SUBMITTED TO

ALL INDIA MANAGEMENT ASSOCIATION CENTRE FOR MANAGEMENT EDUCATION

MANAGEMENT HOUSE, 14 INSTITUTIONAL AREA, LODHI ROAD, NEW DELHI-110003

BY

MISS. SAKSHI GUPTA

ENROLLMENT NO: P55222000

GUIDED BY

MR. ASHISH KAPASIYA

POSTGRADUATE DEGREE IN DATA SCIENCE

FOR THE PARTIAL FULFILMENT OF

POST GRADUATE CERTIFICATE IN MANAGEMENT – DATA ANALYTICS

LIST OF CONTENTS

Content	Page No
List of Tables	2
List of Figures	3
Acknowledgement	4
Preface	5
Introduction, About the Sector, Statement of Problem, Research Objective	6-7
Research Design and Methodology	8-11
Presentation and Analysis	12-29
Conclusion	30
Recommendation	31
Limitation	32
Appendices	33
Reference	34

LIST OF TABLES

Table	Page No
Questionnaire Survey Result (Primary Research)	11

LIST OF FIGURES

Figure	Page No
Data Cleaning in MySQL – Importing Dataset	12
Data Cleaning in MySQL – Renaming Field Name	12
Data Cleaning in MySQL – Treating Null Values	13 - 19
Data Cleaning in MySQL – Checking Duplicate Values	19
Extracting Insights with Python – Importing Dataset	20
Extracting Insights with Python – Exploratory Data Analysis (EDA)	20 - 21
Extracting Insights with Python – Basic Visualization	22 - 24
Extracting Insights with Python - Calculation of Financial Ratios	25 - 26
Data Cleaning in MySQL (Part 2) – Importing Dataset	27
Data Cleaning in MySQL (Part 2) – Treating Null Values	27
Data Visualization with Tableau	28-29

ACKNOWLEDGEMENT

I have a great pleasure in representing this project report entitled

"A STUDY OF STOCK MARKET IN INDIA"

(CASE STUDY OF INDUSTRIAL MACHINERY SECTOR)

and I grab this opportunity to convey my immense regards towards all the distinguished people who have their valuable contribution in the hour of need.

I take this opportunity to thank Mr. Ashish Kapasiya, Assistant Manager (Product and Student Outcome), for giving me an opportunity to complete this project and the most needed guidance throughout the duration of the Programme.

I am extremely grateful to my project guide Ms. Rachana Bajaj for her valuable guidance and necessary support during each phase of the project. She was the source of continuous encouragement as each milestone was crossed.

I would like to express my sincere gratitude to the All India Management Association (AIMA) for entrusting me with the opportunity to undertake this project as an integral component of the academic prerequisites for the Post Graduate Certificate in Management – Data Analytics.

Sakshi Gupta

PREFACE

In the world of finance, understanding the intricacies of the Industrial Machinery sector in India becomes a crucial pursuit. This research paper aims to shed light on the financial intricacies of the Indian Industrial Machinery sector, guided by clear objectives and supported by the search for insights that deeply resonate with its core mission. The sections that follow present survey responses, data-driven analysis, and visual representations that together portray data-driven insights for finance enthusiasts, fund management houses, and other investors. This project not only study the diverse perspectives but also analyses the challenges faced during data analysis and the aspiration to identify top-performing Industrial Machinery companies in financial sector. The dynamic Tableau dashboard emerges as a tool of empowerment, shedding light on insights for those navigating financial insights. Meanwhile, the acknowledgment of limitations in end serves to define the study's boundaries and underscores the authenticity of the exploration

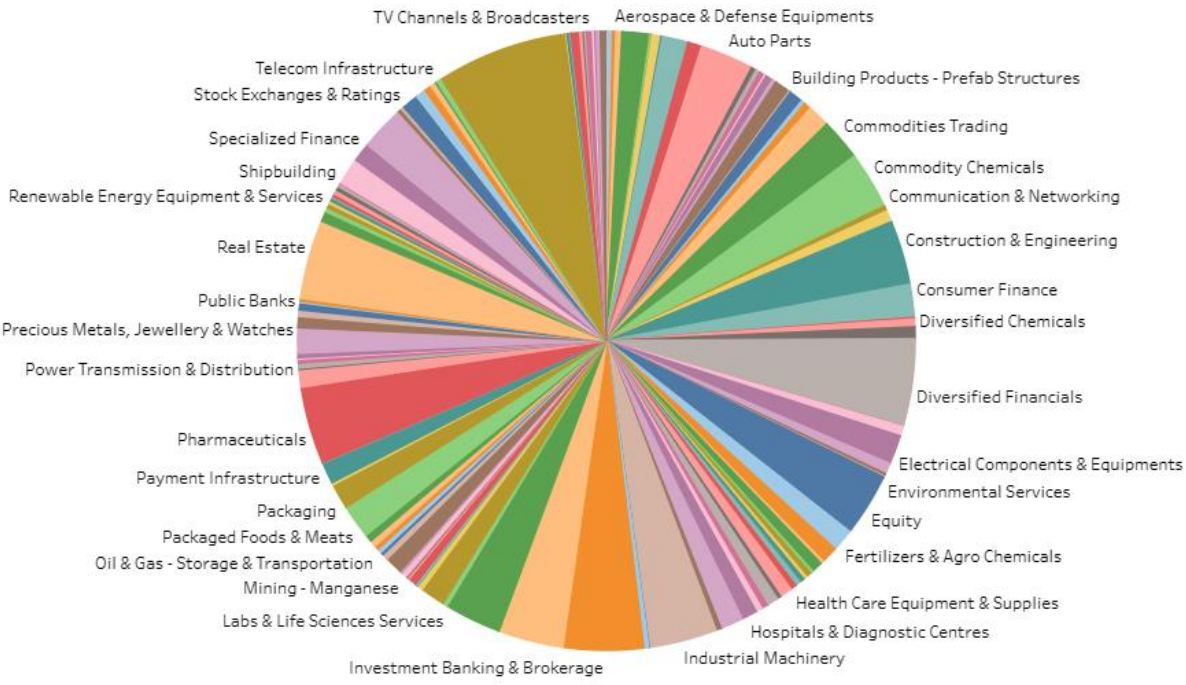
PROJECT OVERVIEW

INTRODUCTION

The stock market in India is one of the largest and most active in the world. It is a vital part of the Indian economy, providing a forum for companies to raise capital and for investors to participate in the growth of the economy.

The Indian stock market is divided into a number of different sectors, each of which is home to a variety of companies. Some of the major sectors in the Indian stock market include:

Diversified Financials, Investment Banking & Brokerage, Real Estate, Pharmaceuticals, Industrial Machinery, Iron & Steel, Construction & Engineering, Commodity Chemicals, IT Services & Consulting, Auto Parts, Specialty Chemicals, Packaged Foods & Meats etc.



This case study serves as a compelling illustration of synergizing SQL, Python, and Tableau to refine stock selection within a vast dataset. With over 4500+ stocks, the objective is to adeptly pinpoint and advocate 5 optimal choices within the industrial machinery subsector.

This endeavor underscores the seamless amalgamation of these three tools, empowering comprehensive exploration, meticulous analysis, and dynamic presentation of intricate stock market data.

By seamlessly integrating SQL's querying prowess, Python's analytical capabilities, and Tableau's visual prowess, this study demonstrates a holistic approach to inform strategic stock decisions, illuminating the potential of cross-disciplinary collaboration in data-driven decision-making.

About the Sector - Industrial Machinery

The Industrial Machinery Sector refers to the segment of the market that involves the manufacturing, distribution, and maintenance of machinery and equipment used in various industries. This sector plays a vital role in supporting the overall industrial development and infrastructure growth of a country.

The growth and performance of the Industrial Machinery Sector are closely linked to the overall economic conditions and industrial activities. It serves as an indicator of industrial health and market demand. Factors such as technological advancements, government policies, global trade, and market trends significantly influence the dynamics of this sector.

Given the significance of the Industrial Machinery Sector in driving economic growth and industrialization, understanding its stock market performance and trends becomes crucial for investors, analysts, and policymakers.

Through in-depth research and analysis, studying the stock market within this sector can provide valuable insights into market dynamics, investment opportunities, and the overall health of the industrial machinery industry.

Statement of Problem

The Indian industrial machinery sector is facing a number of challenges, including high volatility and uncertainty in the stock market, lack of liquidity in the market, lack of transparency in the market, high cost of trading, and lack of financial education among retail investors.

The Indian industrial machinery sector is facing a number of challenges, but it also has a number of opportunities for growth. By understanding the challenges and opportunities, stakeholders can develop strategies to promote the growth of the sector.

Research Objective

The objective is to identify the top 5 companies' stocks within the industrial machinery sector, presenting a compelling investment opportunity for potential investors

To provide insights for investors, businesses, and policymakers on the Indian industrial machinery sector and its potential for growth.

To assess the impact of the stock market on the Indian industrial machinery sector and to develop recommendations for how to mitigate any negative impacts.

Research Design and Methodology

Research methodology provides a framework for conducting research and helps to ensure that the research is conducted in a systematic and objective manner. It also helps to ensure that the research is reproducible and that the findings can be verified by others.

In the case of this project, the research focuses on analysing the structure of the industrial machinery. A qualitative exploratory research method will be used to examine financial parameters and important factors in the industrial machinery industry.

The findings of this research will be valuable to a variety of stakeholders, including investors, industry analysts, and government regulators. The findings will help these stakeholders to better understand the financial performance of industrial machinery companies and to make informed decisions about their investments.

Data Collection Method

Data is collected using both the method Primary Data Collection Method and Secondary Data Collection Method.

- **Primary Data Collection Method**

  Surveys are a good way to collect data from a large number of people. We used surveys to ask people about their investment experiences, their views on the industrial sector in India, and their thoughts on the future of the industry. The total sample of 60 respondents were contacted who responded to the questionnaires.

- **Secondary Data Collection Method**

  Online datasets can provide you with information on a variety of topics, including the performance of stocks in the industrial sector, the trends in the industrial sector, and the factors that are impacting the industrial sector.
  Site - https://datahub.io/collections

QUESTIONNAIRE (IF ANY)

Dear Respondents,
I am an PGCM student currently working on a research project titled "A Study of the Stock Market in India: Case Study of the Industrial Machinery Sector." Your assistance in completing this project is crucial, and I kindly request your participation by filling out the following questionnaire. Your input will greatly contribute to the success of my research.
Rest assured, all responses will be kept confidential.
Thank you for your valuable time and support.

Respondent Information:

Name : _____

Age : _____

Address : _____

Gender : _____

Contact No. : _____

1. Do you prefer to invest in individual stocks?

    - Yes
    - No

2. How much risk are you comfortable with?

    - Low risk
    - Medium risk
    - High risk

3. What is your investment horizon?

    - I plan to invest for the long term.
    - I plan to invest for the medium term.
    - I plan to invest for the short term

4. Do you think investors should focus on companies with strong fundamentals?

    - Yes
    - No

5. What are your investment concerns for the industrial machinery sector in India?

    - The global economic slowdown.
    - The rising cost of raw materials.
    - The increasing competition from China.

6. Do you think investors should look for companies that are well-positioned to benefit from the growth of the Indian economy?

    - Yes
    - No

Sr. No.	Questions	A	B	C
1	Do you prefer to invest in individual stocks? (A) Yes (B) No	45	15	0
2	How much risk are you comfortable with? (A) Low Risk (B) Medium Risk (C) High Risk	25	20	15
3	What is your investment horizon? (A) I plan to invest for the long term. (B) I plan to invest for the medium term. (C) I plan to invest for the short term	20	10	30
4	Do you think investors should focus on companies with strong fundamentals? (A) Yes (B) No	38	22	0
5	What are your investment concerns for the industrial machinery sector in India? (A) The global economic slowdown. (B) The rising cost of raw materials. (C) The increasing competition from China	22	15	23
6	Do you think investors should look for companies that are well-positioned to benefit from the growth of the Indian economy? (A) Yes (B) No	48	12	0

Presentation and Analysis

Data Engineering

- **Data Cleaning**

Data Cleaning in MySQL involves the process of identifying and rectifying inaccuracies, inconsistencies, and errors within a dataset to enhance its quality and reliability for analysis and decision-making.

- Create Database and import the Data in MySQL Workbench and check it shape.

```
CREATE DATABASE AIMA;
USE AIMA;
DESCRIBE STOCKMARKET;
SELECT * FROM STOCKMARKET;
# ************************************************ #
SELECT COUNT(*) AS total_columns
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_SCHEMA = 'Aima'
  AND TABLE_NAME = 'STOCKMARKET';
SELECT COUNT(*) FROM STOCKMARKET;
SELECT * FROM STOCKMARKET;
```

- Renaming the Columns

```
ALTER TABLE STOCKMARKET
    RENAME COLUMN `Name` TO `Company_Name`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `Sub-Sector` TO `Sub_Sector`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `Market Cap` TO `Market_Cap`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `1Y Forward Revenue Growth` TO `1Y_Forward_Revenue_Growth`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `Close Price` TO `Close_Price`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `PE Ratio` TO `PE_Ratio`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `Percentage Buy Recoâ€™s` TO `Percentage_Buy_Records`;
ALTER TABLE STOCKMARKET
    RENAME COLUMN `5Y Historical Revenue Growth` TO `5Y_Historical_Revenue_Growth`,
    RENAME COLUMN `Total Revenue` TO `Total_Revenue`,
    RENAME COLUMN `Taxes & Other Items` TO `Taxes_And_Other_Items`,
    RENAME COLUMN `DII Holding ChangeÂ â€“Â 3M` TO `DII_Holding_Change_3M`,
    RENAME COLUMN `FII Holding ChangeÂ â€“Â 3M` TO `FII_Holding_Change_3M` ,
    RENAME COLUMN `MF Holding ChangeÂ â€“Â 3M` TO `MF_Holding_Change_3M`,
    RENAME COLUMN `Promoter Holding ChangeÂ â€“Â 3M` TO `Promoter_Holding_Change_3M`,
    RENAME COLUMN `No. of analysts with buy reco` TO `No_Of_Analysts_With_Buy_Reco`;
    SELECT * FROM STOCKMARKET;
```

- Treating Null Values

(A) Count of Null Values

```sql
SELECT
    SUM(CASE WHEN Company_Name IS NULL THEN 1 ELSE 0 END) AS Company_Name_null_count,
    SUM(CASE WHEN Ticker IS NULL THEN 1 ELSE 0 END) AS Ticker_null_count,
    SUM(CASE WHEN Sub_Sector IS NULL THEN 1 ELSE 0 END) AS Sub_Sector_null_count,
    SUM(CASE WHEN Market_Cap IS NULL THEN 1 ELSE 0 END) AS Market_Cap_null_count,
    SUM(CASE WHEN Close_Price IS NULL THEN 1 ELSE 0 END) AS Close_Price_null_count,
    SUM(CASE WHEN PE_Ratio IS NULL THEN 1 ELSE 0 END) AS PE_Ratio_null_count,
    SUM(CASE WHEN 5Y_Historical_Revenue_Growth IS NULL THEN 1 ELSE 0 END) AS `5Y_Historical_Revenue_Growth_null_count`,
    SUM(CASE WHEN `1Y_Forward_Revenue_Growth` IS NULL THEN 1 ELSE 0 END) AS `1Y Forward Revenue Growth_null_count`,
    SUM(CASE WHEN Total_Revenue IS NULL THEN 1 ELSE 0 END) AS Total_Revenue_null_count,
    SUM(CASE WHEN PBT IS NULL THEN 1 ELSE 0 END) AS PBT_null_count,
    SUM(CASE WHEN Taxes_And_Other_Items IS NULL THEN 1 ELSE 0 END) AS Taxes_And_Other_Items_null_count,
    SUM(CASE WHEN DII_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS DII_Holding_Change_3M_null_count,
    SUM(CASE WHEN FII_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS FII_Holding_Change_3M_null_count,
    SUM(CASE WHEN MF_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS MF_Holding_Change_3M_null_count,
    SUM(CASE WHEN Promoter_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS Promoter_Holding_Change_3M_null_count,
    SUM(CASE WHEN Percentage_Buy_Records IS NULL THEN 1 ELSE 0 END) AS Percentage_Buy_Records_null_count,
    SUM(CASE WHEN No_Of_Analysts_With_Buy_Reco IS NULL THEN 1 ELSE 0 END) AS No_Of_Analysts_With_Buy_Reco_null_count,
    SUM(CASE WHEN `Percentage Upside` IS NULL THEN 1 ELSE 0 END) AS `Percentage Upside_null_count`
FROM STOCKMARKET;
```

(B) Treating Null Values

- ✓ Calculate Mean and Median of each variable to find existence of outliers.
- ✓ If outliers exist than fill null values with median else use mean and for some company specific variables value fill null values with zero.
- ✓ Check whether the null values has been treated or not.

```sql
## Market Cap Null Value is 129 (total)
SELECT COUNT(Market_Cap) AS total_null_Market_Cap
FROM STOCKMARKET
WHERE `Sub_Sector` LIKE 'Industrial Machinery' AND Market_Cap IS NULL;
## Since the count of null value of Market Cap of `Sub_Sector` = 'Industrial Machinery' is 0 lets treat other null values
SELECT * FROM STOCKMARKET;
####### ************ LETS TREAT NULL VALUES ************** #######
### MARKET_CAP Mean = 6572.563931 and Median = 107.6594081

SELECT
    AVG(`Market_Cap`) AS average_value_Market_Cap,
    (SELECT `Market_Cap`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `Market_Cap`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `Market_Cap` IS NOT NULL
        ORDER BY `Market_Cap`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `Market_Cap` IS NOT NULL
     )
    ) AS median_value_Market_Cap
FROM STOCKMARKET;
```

```sql
## Disable safe update mode
SET SQL_SAFE_UPDATES = 0;
-### Calculating the median value
SELECT `Market_Cap`
INTO @median
FROM (
    SELECT `Market_Cap`, @rownum := @rownum + 1 AS rownum
    FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
    WHERE `Market_Cap` IS NOT NULL
    ORDER BY `Market_Cap`
) AS ranked
WHERE rownum = (
    SELECT COUNT(*) / 2 + 1
    FROM STOCKMARKET
    WHERE `Market_Cap` IS NOT NULL
);
### Updating null values with the calculated median
UPDATE STOCKMARKET
SET `Market_Cap` = IFNULL(`Market_Cap`, @median);

### After completing the deletion, it's recommended to re-enable safe update mode
SET SQL_SAFE_UPDATES = 1;

SELECT COUNT(Market_Cap) AS total_null_Market_Cap
FROM STOCKMARKET
WHERE Market_Cap IS NULL;
```

```sql
# ************************************** #
### Close_Price count of null value is 79
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SET SQL_SAFE_UPDATES = 0;

UPDATE STOCKMARKET
SET `Close_Price` = 0
WHERE `Close_Price` IS NULL;

SELECT COUNT(Close_Price) AS total_null_Close_Price
FROM STOCKMARKET
WHERE Close_Price IS NULL;
## **************************************
### PE_Ratio count of null value is 188
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
UPDATE STOCKMARKET
SET `PE_Ratio` = 0
WHERE `PE_Ratio` IS NULL;

SELECT COUNT(PE_Ratio) AS total_null_PE_Ratio
FROM STOCKMARKET
WHERE PE_Ratio IS NULL;


### 5Y_Historical_Revenue_Growth count of null value is 576
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
UPDATE STOCKMARKET
SET `5Y_Historical_Revenue_Growth` = 0
WHERE `5Y_Historical_Revenue_Growth` IS NULL;

SELECT COUNT(`5Y_Historical_Revenue_Growth`) AS total_null_5Y_Historical_Revenue_Growth
FROM STOCKMARKET
WHERE `5Y_Historical_Revenue_Growth` IS NULL;

### 1Y_Forward_Revenue_Growth count of null value is 3851
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
UPDATE STOCKMARKET
SET `1Y_Forward_Revenue_Growth` = 0
WHERE `1Y_Forward_Revenue_Growth` IS NULL;

SELECT COUNT(`1Y_Forward_Revenue_Growth`) AS total_null_1Y_Forward_Revenue_Growth
FROM STOCKMARKET
WHERE `1Y_Forward_Revenue_Growth` IS NULL;
```

```sql
### Total_Revenue count of null value is 185
## Median = 99.7, Mean= 3258.580460617225
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SELECT
    AVG(`Total_Revenue`) AS average_value_Total_Revenue,
    (SELECT `Total_Revenue`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `Total_Revenue`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `Total_Revenue` IS NOT NULL
        ORDER BY `Total_Revenue`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `Total_Revenue` IS NOT NULL
     )
    ) AS median_value_Total_Revenue
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `Total_Revenue` = 0
WHERE `Total_Revenue` IS NULL;

SELECT COUNT(`Total_Revenue`) AS total_null_Total_Revenue
FROM STOCKMARKET
WHERE `Total_Revenue` IS NULL;

 ### PBT of null value is 178
 ## Mean= 317.2151203887795
 ## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
 SELECT
     AVG(`PBT`) AS average_value_PBT,
     (SELECT `PBT`
      FROM (
         SELECT @rownum := @rownum + 1 AS rownum, `PBT`
         FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
         WHERE `PBT` IS NOT NULL
         ORDER BY `PBT`
      ) ranked
      WHERE rownum = (
         SELECT COUNT(*) / 2 + 1
         FROM STOCKMARKET
         WHERE `PBT` IS NOT NULL
      )
     ) AS median_value_Total_Revenue
 FROM STOCKMARKET;

 UPDATE STOCKMARKET
 SET `PBT` = 0
 WHERE `PBT` IS NULL;

 SELECT COUNT(`PBT`) AS total_null_Total_Revenue
 FROM STOCKMARKET
 WHERE `PBT` IS NULL;

### Taxes_And_Other_Items of null value is 178
## Mean= 97.76544033111092
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SELECT
    AVG(`Taxes_And_Other_Items`) AS average_value_Taxes_And_Other_Items,
    (SELECT `Taxes_And_Other_Items`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `Taxes_And_Other_Items`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `Taxes_And_Other_Items` IS NOT NULL
        ORDER BY `Taxes_And_Other_Items`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `Taxes_And_Other_Items` IS NOT NULL
     )
    ) AS median_value_Taxes_And_Other_Items
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `Taxes_And_Other_Items` = 0
WHERE `Taxes_And_Other_Items` IS NULL;

SELECT COUNT(`Taxes_And_Other_Items`) AS total_null_Taxes_And_Other_Items
FROM STOCKMARKET
WHERE `Taxes_And_Other_Items` IS NULL;
```

```
### DII_Holding_Change_3M of null value is 536
## Mean= 0.007831795397296388
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SELECT
    AVG(`DII_Holding_Change_3M`) AS average_value_DII_Holding_Change_3M,
    (SELECT `DII_Holding_Change_3M`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `DII_Holding_Change_3M`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `DII_Holding_Change_3M` IS NOT NULL
        ORDER BY `DII_Holding_Change_3M`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `DII_Holding_Change_3M` IS NOT NULL
     )
    ) AS median_value_DII_Holding_Change_3M
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `DII_Holding_Change_3M` = 0
WHERE `DII_Holding_Change_3M` IS NULL;

SELECT COUNT(`DII_Holding_Change_3M`) AS total_null_DII_Holding_Change_3M
FROM STOCKMARKET
WHERE `DII_Holding_Change_3M` IS NULL;
```

```
### FII_Holding_Change_3M of null value is 536
## Mean= 0.08576215258907524
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SELECT
    AVG(`FII_Holding_Change_3M`) AS average_value_FII_Holding_Change_3M,
    (SELECT `FII_Holding_Change_3M`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `FII_Holding_Change_3M`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `FII_Holding_Change_3M` IS NOT NULL
        ORDER BY `FII_Holding_Change_3M`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `FII_Holding_Change_3M` IS NOT NULL
     )
    ) AS median_value_FII_Holding_Change_3M
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `FII_Holding_Change_3M` = 0
WHERE `FII_Holding_Change_3M` IS NULL;

SELECT COUNT(`FII_Holding_Change_3M`) AS total_null_FII_Holding_Change_3M
FROM STOCKMARKET
WHERE `FII_Holding_Change_3M` IS NULL;
```

```
### MF_Holding_Change_3M of null value is 536
## Mean= 0.0193463581284866
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SELECT
    AVG(`MF_Holding_Change_3M`) AS average_value_MF_Holding_Change_3M,
    (SELECT `MF_Holding_Change_3M`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `MF_Holding_Change_3M`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `MF_Holding_Change_3M` IS NOT NULL
        ORDER BY `MF_Holding_Change_3M`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `MF_Holding_Change_3M` IS NOT NULL
     )
    ) AS median_value_MF_Holding_Change_3M
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `MF_Holding_Change_3M` = 0
WHERE `MF_Holding_Change_3M` IS NULL;

SELECT COUNT(`MF_Holding_Change_3M`) AS total_null_MF_Holding_Change_3M
FROM STOCKMARKET
WHERE `MF_Holding_Change_3M` IS NULL;
```

```sql
### Promoter_Holding_Change_3M of null value is 536
## Mean= -0.3770983374848735
## we cannot fill null values with mean or median values as it belong to a specific company so lets replace it with 0
SELECT
    AVG(`Promoter_Holding_Change_3M`) AS average_value_Promoter_Holding_Change_3M,
    (SELECT `Promoter_Holding_Change_3M`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `Promoter_Holding_Change_3M`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `Promoter_Holding_Change_3M` IS NOT NULL
        ORDER BY `Promoter_Holding_Change_3M`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `Promoter_Holding_Change_3M` IS NOT NULL
     )
    ) AS median_value_Promoter_Holding_Change_3M
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `Promoter_Holding_Change_3M` = 0
WHERE `Promoter_Holding_Change_3M` IS NULL;

SELECT COUNT(`Promoter_Holding_Change_3M`) AS total_null_Promoter_Holding_Change_3M
FROM STOCKMARKET
WHERE `Promoter_Holding_Change_3M` IS NULL;
```

```sql
### Percentage_Buy_Records of null value is 3791
## Mean= 73.88344743369703, Median = 81.25
## we cannot fill null values with mean or median values as there are two many null values
## so either to drop them or replace them  with 0 but not allowed to drop so lets replace it with 0
SELECT
    AVG(`Percentage_Buy_Records`) AS average_value_Percentage_Buy_Records,
    (SELECT `Percentage_Buy_Records`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `Percentage_Buy_Records`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `Percentage_Buy_Records` IS NOT NULL
        ORDER BY `Percentage_Buy_Records`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `Percentage_Buy_Records` IS NOT NULL
     )
    ) AS median_value_Percentage_Buy_Records
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `Percentage_Buy_Records` = 0
WHERE `Percentage_Buy_Records` IS NULL;
SELECT COUNT(`Percentage_Buy_Records`) AS total_null_Percentage_Buy_Records
FROM STOCKMARKET
WHERE `Percentage_Buy_Records` IS NULL;
```

- ✓ `No_Of_Analysts_With_Buy_Reco` is a categorical variable we should replace null values with the value having highest frequency i.e we should calculate mode.

```
### No_Of_Analysts_With_Buy_Reco of null value is 52
## Mean= 1.1267
## we cannot fill null values with mean or median values as it categorical data
SELECT
    AVG(`No_Of_Analysts_With_Buy_Reco`) AS average_value_No_Of_Analysts_With_Buy_Reco,
    (SELECT `No_Of_Analysts_With_Buy_Reco`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `No_Of_Analysts_With_Buy_Reco`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `No_Of_Analysts_With_Buy_Reco` IS NOT NULL
        ORDER BY `No_Of_Analysts_With_Buy_Reco`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `No_Of_Analysts_With_Buy_Reco` IS NOT NULL
     )
    ) AS median_value_No_Of_Analysts_With_Buy_Reco
FROM STOCKMARKET;
##  Calculate Mode (mode value = 0, frequency = 3779
SELECT No_Of_Analysts_With_Buy_Reco AS mode_value, COUNT(*) AS frequency
FROM STOCKMARKET
GROUP BY No_Of_Analysts_With_Buy_Reco
ORDER BY COUNT(*) DESC
LIMIT 1;
```

```
##  Calculate Mode (mode value = 0, frequency = 3779
SELECT No_Of_Analysts_With_Buy_Reco AS mode_value, COUNT(*) AS frequency
FROM STOCKMARKET
GROUP BY No_Of_Analysts_With_Buy_Reco
ORDER BY COUNT(*) DESC
LIMIT 1;

UPDATE STOCKMARKET
SET `No_Of_Analysts_With_Buy_Reco` = 0
WHERE `No_Of_Analysts_With_Buy_Reco` IS NULL;

SELECT COUNT(`No_Of_Analysts_With_Buy_Reco`) AS total_null_Percentage_Buy_Records
FROM STOCKMARKET
WHERE `No_Of_Analysts_With_Buy_Reco` IS NULL;
```

```
### Percentage Upside of null value is 3789
## Mean= 20.228940825066207, Median = 16.44671177
## we cannot fill null values with mean or median values as there are two many null values
## so either to drop them or replace them  with 0 but not allowed to drop so lets replace it with 0
SELECT
    AVG(`Percentage Upside`) AS `average_value_Percentage Upside`,
    (SELECT `Percentage Upside`
     FROM (
        SELECT @rownum := @rownum + 1 AS rownum, `Percentage Upside`
        FROM STOCKMARKET, (SELECT @rownum := 0) AS dummy
        WHERE `Percentage Upside` IS NOT NULL
        ORDER BY `Percentage Upside`
     ) ranked
     WHERE rownum = (
        SELECT COUNT(*) / 2 + 1
        FROM STOCKMARKET
        WHERE `Percentage Upside` IS NOT NULL
     )
    ) AS `median_value_Percentage Upside`
FROM STOCKMARKET;

UPDATE STOCKMARKET
SET `Percentage Upside` = 0
WHERE `Percentage Upside` IS NULL;

SELECT COUNT(`Percentage Upside`) AS total_null_Percentage_Buy_Records
FROM STOCKMARKET
WHERE `Percentage Upside` IS NULL;

SELECT * FROM STOCKMARKET;
SET SQL_SAFE_UPDATES = 1;
```

- Checking Duplication of Data

```
## Lets start treating duplicated rows
SHOW COLUMNS FROM STOCKMARKET;
SELECT *
FROM STOCKMARKET
GROUP BY `Name`,`Ticker`, `Sub_Sector`, `Market_Cap`,`Close_Price`, `PE_Ratio`,
 `5Y_Historical_Revenue_Growth`,`1Y_Forward_Revenue_Growth`, `Total_Revenue`,
`PBT`, `Taxes_And_Other_Items`, `DII_Holding_Change_3M`, `FII_Holding_Change_3M`,
`MF_Holding_Change_3M`, `Promoter_Holding_Change_3M`,`Percentage_Buy_Records`,
 `No_Of_Analysts_With_Buy_Reco`, `Percentage Upside`,`Net_Profit`,`Profit_Margin`,
 `PS_Ratio`,`Revenue_Growth`,`Total_Institutional_Holding_Change`
HAVING COUNT(*) > 1;
##################         Its Good That we don't have duplicated value
```

```
## Here we are done with Data Cleaning
## Sub_Sector_null_count = 591 lets not treat it as we are not allowed to drop them lets ignore them
################################## Here we are done with Data Cleaning lets export our cleaned dataset in csv format
 ################## by using the MySQL Workbench GUI
```

- **Data Transformation**

Retrieve refined SQL data as CSV. Load to Python via Pandas for manipulation. Utilize NumPy, Matplotlib/Seaborn for calculations. Derive insights e.g., PE ratio, Profit after Tax, Margin, gaining deep stock market understanding.

- Python: A versatile programming language widely used in data analysis, scripting, and software development, known for its simplicity and extensive libraries.
- Pandas: A powerful data manipulation library in Python, offering data structures and tools to efficiently clean, transform, and analyze structured data, making it a cornerstone for data analysis tasks.
- NumPy: The fundamental package for scientific computing in Python, providing support for large, multi-dimensional arrays and matrices, along with a plethora of mathematical functions to operate on these arrays.
- Matplotlib: A widely-used 2D plotting library in Python, enabling creation of a variety of static, animated, and interactive visualizations for effective data representation.
- Seaborn: Built on Matplotlib, Seaborn is a statistical data visualization library, simplifying the creation of informative and aesthetically pleasing visualizations for exploring and understanding complex datasets.
- Importing required libraries and cleaned dataset in Python and check Head and Tail of the dataset.

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import numpy as np
4 import seaborn as sns
5 import statsmodels.api as sm
```

```
1 stock = pd.read_csv("Clean_StockMarketF.csv")
```

```
1 stock.head()
```

	Company_Name	Ticker	Sub_Sector	Market_Cap	Close_Price	PE_Ratio	5Y_Historical_Revenue_Growth	1Y_Forward_Revenue_Growth	Total_Revenue
0	Reliance Industries Ltd	RELIANCE	Oil & Gas - Refining & Marketing	1.778090e+06	2712.20	29.290670	17.936498	30.535512	718061.00
1	Tata Consultancy Services Ltd	TCS	IT Services & Consulting	1.241655e+06	3397.35	32.396353	9.886696	13.295058	195772.00
2	HDFC Bank Ltd	HDFCBANK	Private Banks	8.917948e+05	1597.85	23.435750	14.242193	-27.600817	167695.40
3	Infosys Ltd	INFY	IT Services & Consulting	6.806679e+05	1629.05	30.785523	11.618708	18.128913	123936.00
4	ICICI Bank Ltd	ICICIBANK	Private Banks	6.533383e+05	946.75	26.018932	6.796088	-49.983642	157536.32

```
1 stock.tail()
```

	Company_Name	Ticker	Sub_Sector	Market_Cap	Close_Price	PE_Ratio	5Y_Historical_Revenue_Growth	1Y_Forward_Revenue_Growth	Total_Rever
4522	Tapi Fruit Processing Ltd	TAPIFRUIT	Packaged Foods & Meats	107.659408	61.80	0.0	0.0	0.0	15
4523	Ishan International Ltd	ISHAN	Industrial Machinery	107.659408	45.00	0.0	0.0	0.0	21
4524	Coastal Corporation Ltd Partly Paidup	COASTPP	NaN	107.659408	93.50	0.0	0.0	0.0	0
4525	Anzen India Energy Yield Plus Trust	ANZEN	Power Transmission & Distribution	107.659408	0.00	0.0	0.0	0.0	0
4526	Keystone Realtors Pvt Ltd	RUSTOMJEE	Real Estate	107.659408	543.05	0.0	0.0	0.0	1302

- Checking Shape and Null Values

```
1 stock.shape
```

```
(4527, 18)
```

```
1 stock.isna().sum()
```

```
Company_Name                    0
Ticker                          0
Sub_Sector                    590
Market_Cap                      0
Close_Price                     0
PE_Ratio                        0
5Y_Historical_Revenue_Growth    0
1Y_Forward_Revenue_Growth       0
Total_Revenue                   0
PBT                             0
Taxes_And_Other_Items           0
DII_Holding_Change_3M           0
FII_Holding_Change_3M           0
MF_Holding_Change_3M            0
Promoter_Holding_Change_3M      0
Percentage_Buy_Records          0
No_Of_Analysts_With_Buy_Reco    0
Percentage Upside               0
dtype: int64
```

- Check Column Names

```
1 stock.columns
```

```
Index(['Company_Name', 'Ticker', 'Sub_Sector', 'Market_Cap', 'Close_Price',
       'PE_Ratio', '5Y_Historical_Revenue_Growth', '1Y_Forward_Revenue_Growth',
       'Total_Revenue', 'PBT', 'Taxes_And_Other_Items',
       'DII_Holding_Change_3M', 'FII_Holding_Change_3M',
       'MF_Holding_Change_3M', 'Promoter_Holding_Change_3M',
       'Percentage_Buy_Records', 'No_Of_Analysts_With_Buy_Reco',
       'Percentage Upside'],
      dtype='object')
```

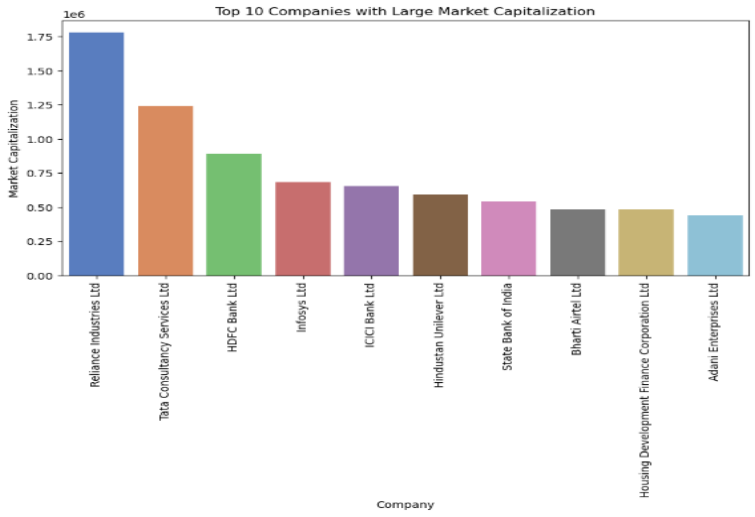
- Sub – Sector

```
1 print("Total Count of Sub-Sector: ",stock.Sub_Sector.nunique())
2 print("Sub-Sector Category: ")
3 print(stock.Sub_Sector.unique())
```

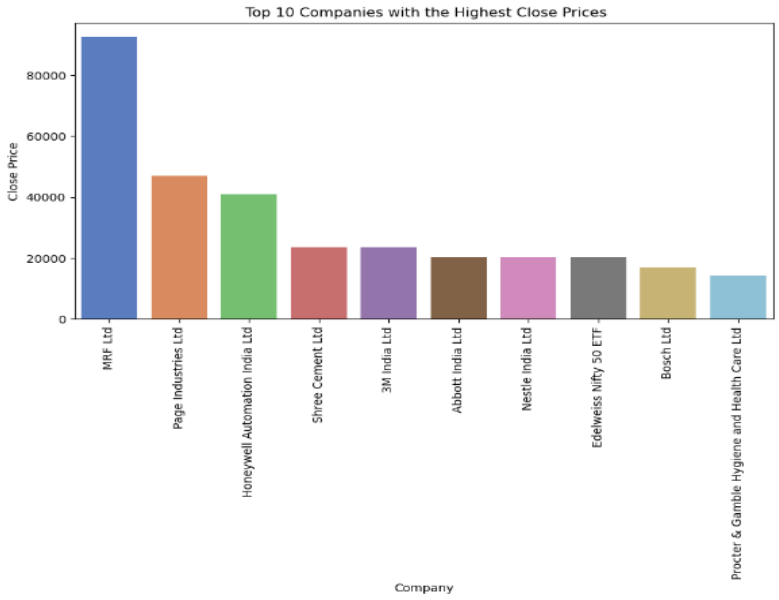
```
Total Count of Sub-Sector:  140
Sub-Sector Category:
['Oil & Gas - Refining & Marketing' 'IT Services & Consulting'
 'Private Banks' 'FMCG - Household Products' 'Public Banks'
 'Telecom Services' 'Home Financing' 'Commodities Trading'
 'FMCG - Tobacco' 'Consumer Finance' 'Insurance' 'Gas Distribution'
 'Renewable Energy' 'Power Infrastructure' 'Paints'
 'Construction & Engineering' 'Four Wheelers' 'Retail - Department Stores'
 'Pharmaceuticals' 'Precious Metals, Jewellery & Watches' 'Cement'
 'FMCG - Foods' 'Ports' 'Oil & Gas - Exploration & Production'
 'Iron & Steel' 'Power Generation' 'Power Transmission & Distribution'
 'Mining - Coal' 'Diversified Chemicals' 'Mining - Diversified'
 'Metals - Diversified' 'Two Wheelers' 'FMCG - Personal Products'
 'Conglomerates' 'Real Estate' 'Metals - Aluminium' 'Trucks & Buses'
 'Aerospace & Defense Equipments' 'Labs & Life Sciences Services'
 'Electronic Equipments' 'Soft Drinks' 'Payment Infrastructure'
 'Electrical Components & Equipments' 'Tea & Coffee' 'Airlines'
 'Asset Management' 'Hospitals & Diagnostic Centres'
 'Heavy Electrical Equipments' 'Alcoholic Beverages'
 'Fertilizers & Agro Chemicals' 'Online Services' 'Telecom Infrastructure'
 'Cycles' 'Apparel & Accessories' 'Retail - Apparel' 'Wellness Services'
 'Auto Parts' 'Logistics' 'Hotels, Resorts & Cruise Lines'
 'Packaged Foods & Meats' 'Specialized Finance' 'Software Services'
 'Tires & Rubber' 'Specialty Chemicals' 'Industrial Machinery'
 'Building Products - Pipes' 'Restaurants & Cafes' 'Mining - Iron Ore'
 'Commodity Chemicals' 'Biotechnology' 'Diversified Financials' 'Textiles
 'Oil & Gas - Storage & Transportation' 'Plastic Products'
 'Business Support Services' 'Tractors' 'Home Electronics & Appliances'
 'Cables' 'Stationery' 'TV Channels & Broadcasters' 'Footwear' 'Equity'
 'Stock Exchanges & Ratings' 'Investment Banking & Brokerage'
 'Building Products - Ceramics' 'Shipbuilding' 'Advertising' 'Batteries'
 'Technology Hardware' 'Power Trading & Consultancy' 'Home Furnishing'
 'Renewable Energy Equipment & Services' 'Wood Products' 'Sugar'
 'Tour & Travel Services' nan 'Theatres' 'Mining - Copper'
 'Telecom Equipments' 'Health Care Equipment & Supplies' 'Agro Products'
 'Paper Products' 'Debt' 'Outsourced services' 'Movies & TV Serials'
 'Roads' 'Infra REIT' 'Employment Services' 'Theme Parks & Gaming'
 'Housewares' 'Rail' 'Packaging' 'Gold' 'Communication & Networking'
 'Education Services' 'Building Products - Laminates'
 'Building Products - Granite' 'Heavy Machinery' 'Cable & D2H'
 'Environmental Services' 'Mining - Manganese' 'Seeds' 'Publishing'
 'Animation' 'Metals - Lead' 'Oil & Gas - Equipment & Services'
 'Agricultural & Farm Machinery' 'Water Management' 'Retail - Speciality'
 'Dredging' 'Building Products - Prefab Structures' 'Radio'
 'Three Wheelers' 'Metals - Coke' 'Building Products - Glass' 'Airports'
 'Retail - Online' 'Metals - Iron' 'Metals - Copper'
 'Academic & Educational Services' '0']
```

- Some basic visualization

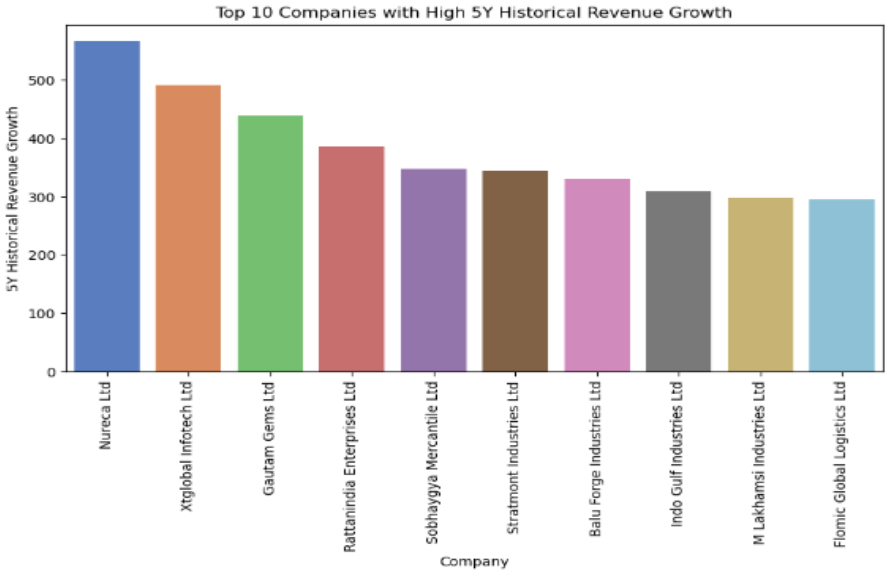
```
stock_M = stock
stock_M = stock_M.sort_values('Market_Cap', ascending=False).head(10)
x = stock_M['Company_Name']
y = stock_M['Market_Cap']
plt.figure(figsize=(10, 5))
sns.barplot(x=x, y=y, palette = "muted")
plt.title('Top 10 Companies with Large Market Capitalization')
plt.xlabel('Company')
plt.ylabel('Market Cap')
plt.xticks(rotation=90)
plt.show()
```



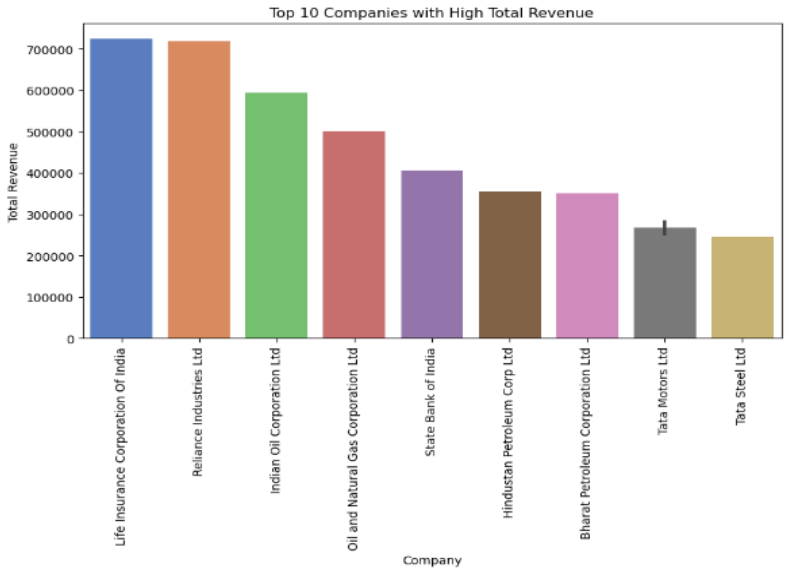

```
stock_M = stock
stock_M = stock_M.sort_values('Close_Price', ascending=False).head(10)
x = stock_M['Company_Name']
y = stock_M['Close_Price']
plt.figure(figsize=(10, 5))
sns.barplot(x=x, y=y, palette = "muted")
plt.title(' Top 10 Companies with the Highest Close Prices')
plt.xlabel('Company')
plt.ylabel('Close Price')
plt.xticks(rotation=90)
plt.show()
```



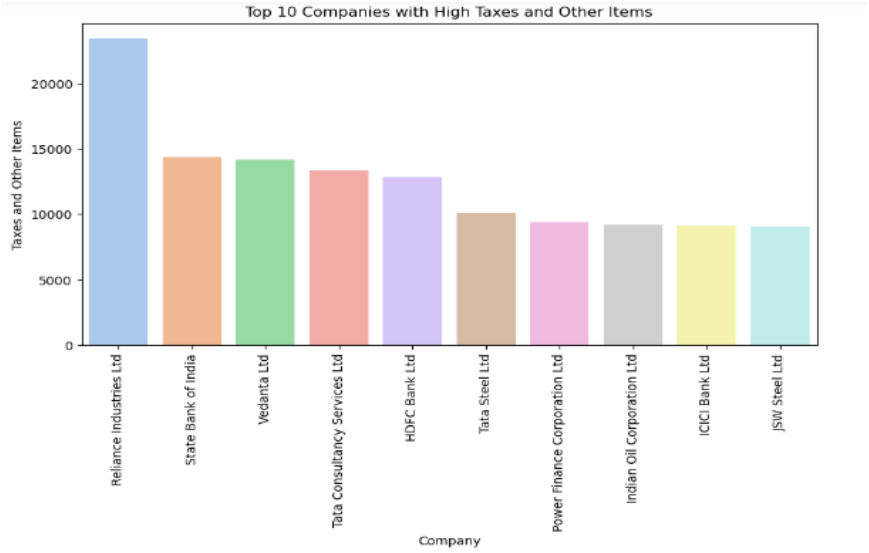
```python
stock_M = stock
stock_M = stock_M.sort_values('5Y_Historical_Revenue_Growth', ascending=False).head(10)
x = stock_M['Company_Name']
y = stock_M['5Y_Historical_Revenue_Growth']
plt.figure(figsize=(10, 5))
sns.barplot(x=x, y=y, palette = "muted")
plt.title('Top 10 Companies with High 5Y Historical Revenue Growth')
plt.xlabel('Company')
plt.ylabel('5Y Historical Revenue Growth')
plt.xticks(rotation=90)
plt.show()
```



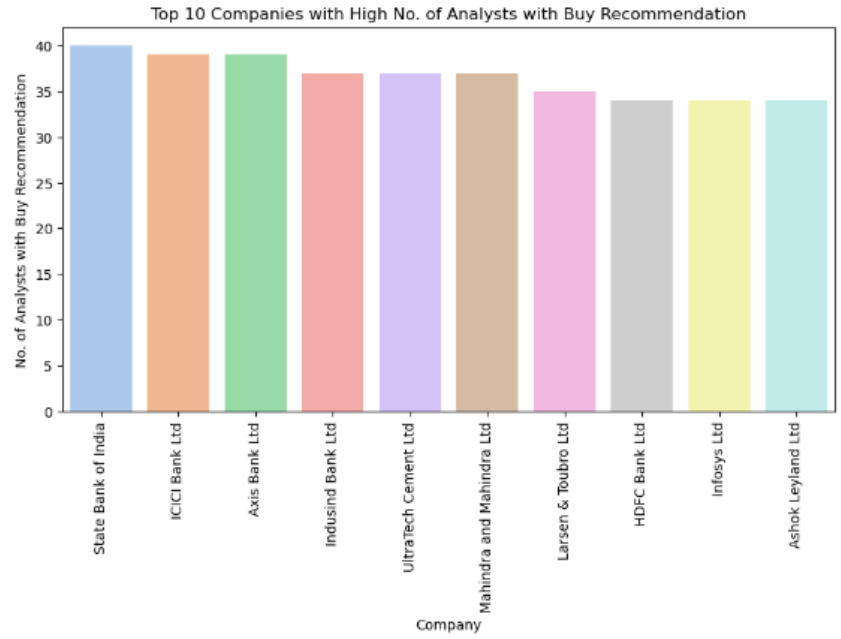

```python
stock_M = stock
stock_M = stock_M.sort_values('Total_Revenue', ascending=False).head(10)
x = stock_M['Company_Name']
y = stock_M['Total_Revenue']
plt.figure(figsize=(10, 5))
sns.barplot(x=x, y=y, palette = "muted")
plt.title('Top 10 Companies with High Total Revenue')
plt.xlabel('Company')
plt.ylabel('Total Revenue')
plt.xticks(rotation=90)
plt.show()
```



```
stock_M = stock
stock_M = stock_M.sort_values('Taxes_And_Other_Items', ascending=False).head(10)
x = stock_M['Company_Name']
y = stock_M['Taxes_And_Other_Items']
plt.figure(figsize=(10, 5))
sns.barplot(x=x, y=y, palette = "pastel")
plt.title('Top 10 Companies with High Taxes and Other Items')
plt.xlabel('Company')
plt.ylabel('Taxes and Other Items')
plt.xticks(rotation=90)
plt.show()
```




```
# Number of Analysts who have given a "Buy" Recommendation for a Particular Company's Stock
stock_M = stock
stock_M = stock_M.sort_values('No_Of_Analysts_With_Buy_Reco', ascending=False).head(10)
x = stock_M['Company_Name']
y = stock_M['No_Of_Analysts_With_Buy_Reco']
plt.figure(figsize=(10, 5))
sns.barplot(x=x, y=y, palette = "pastel")
plt.title('Top 10 Companies with High No. of Analysts with Buy Recommendation')
plt.xlabel('Company')
plt.ylabel('No. of Analysts with Buy Recommendation')
plt.xticks(rotation=90)
plt.show()
```



- Creating new DataFrame and CalculationField

```
stock_add=stock
```

```
stock_add.head()
```

	Company_Name	Ticker	Sub_Sector	Market_Cap	Close_Price	PE_Ratio	5Y_Historical_Revenue_Growth	1Y_Forward_Revenue_Growth	Total_Revenue
0	Reliance Industries Ltd	RELIANCE	Oil & Gas - Refining & Marketing	1.778090e+06	2712.20	29.290670	17.936498	30.535512	718061.00
1	Tata Consultancy Services Ltd	TCS	IT Services & Consulting	1.241655e+06	3397.35	32.396353	9.886696	13.295058	195772.00
2	HDFC Bank Ltd	HDFCBANK	Private Banks	8.917948e+05	1597.85	23.435750	14.242193	-27.600817	167695.40
3	Infosys Ltd	INFY	IT Services & Consulting	6.806679e+05	1629.05	30.785523	11.618708	18.128913	123936.00
4	ICICI Bank Ltd	ICICIBANK	Private Banks	6.533383e+05	946.75	26.018932	6.796088	-49.983642	157536.32

✓ Net_Profit - PAT stands for Profit After Tax or Net Profit. It is a very important measure in the stock market that helps an investor to indicate the profit of a company. PAT is also called Net Profit.
PAT (Profit after tax) = PBT (Profit before tax) – other items & taxes.

```
stock_add['Net_Profit'] = stock_add['PBT'] - stock_add['Taxes_And_Other_Items']
```

```
data = stock_add.sort_values('Net_Profit', ascending=False)
```

✓ Profit_Margin - Net Profit Margin is a financial ratio used to calculate the percentage of profit that a company produces from its total revenue. Profit margin is used to compare the company in the same Sub-Sector.
Profit margin = (Net profit / total revenue) *100

```
stock_add['Profit_Margin'] = (stock_add['Net_Profit'] / stock_add['Total_Revenue'])*100
```

```
data = stock_add.sort_values('Profit_Margin', ascending=True)
```

✓ PS_Ratio - PS Ratio is used to measure the value of a company for each rupee of revenue generated. PSRatio is used to compare the company in the same sub-sector. It is also called Sales Multiple.
PS(Price-to-sales) ratio = (market capitalization / total revenue)

```
stock_add['PS_Ratio'] = stock_add['Market_Cap'] / stock_add['Total_Revenue']
```

```
data = stock_add.sort_values('PS_Ratio', ascending=False)
```

✓ Change Retail Holding - Defines the amount of change in the holding of the retail investor in their portfolio.
Retail Holding change = - (sum of DII Holding Change, FII Holding Change, MF Holding Change, Promoter Holding Change)

```
stock_add['Change_Retail_Holding'] = -(stock_add['DII_Holding_Change_3M'] + stock_add['FII_Holding_Change_3M']
                                       + stock_add['MF_Holding_Change_3M'] + stock_add['Promoter_Holding_Change_3M'])
data = stock_add.sort_values('Change_Retail_Holding', ascending=False)
data.head()
```

```
df=stock_add
df.head()
```

	Company_Name	Ticker	Sub_Sector	Market_Cap	Close_Price	PE_Ratio	5Y_Historical_Revenue_Growth	1Y_Forward_Revenue_Growth	Total_Revenue
0	Reliance Industries Ltd	RELIANCE	Oil & Gas - Refining & Marketing	1.778090e+06	2712.20	29.290670	17.936498	30.535512	718061.00
1	Tata Consultancy Services Ltd	TCS	IT Services & Consulting	1.241655e+06	3397.35	32.396353	9.886696	13.295058	195772.00
2	HDFC Bank Ltd	HDFCBANK	Private Banks	8.917948e+05	1597.85	23.435750	14.242193	-27.600817	167695.40
3	Infosys Ltd	INFY	IT Services & Consulting	6.806679e+05	1629.05	30.785523	11.618708	18.128913	123936.00
4	ICICI Bank Ltd	ICICIBANK	Private Banks	6.533383e+05	946.75	26.018932	6.796088	-49.983642	157536.32

5 rows × 22 columns

- Create a new Categorical Field with name `MarketCapCategory` in which Market Capitalization is divide into 3 section “Large Cap, Mid Cap, Small Cap”.

```
def MarketCapCategory(Market_Cap):
    if Market_Cap > 50000:
        return 'Large Cap'
    elif 20000 <= Market_Cap <= 50000:
        return 'Mid Cap'
    else:
        return 'Small Cap'

# Apply the function to create a new 'MarketCap_Category' column
df['MarketCap_Category'] = df['Market_Cap'].apply(MarketCapCategory)

df.head()
```

nge_3M	Percentage_Buy_Records	No_Of_Analysts_With_Buy_Reco	Percentage Upside	Net_Profit	Profit_Margin	PS_Ratio	Change_Retail_Holding	MarketCap_Category
0.048111	77.419355	24	3.163484	60705.00	11.717946	2.476238	0.000778	Large Cap
0.000000	51.219512	21	1.549737	38327.00	26.401630	6.342352	0.258454	Large Cap
0.087276	91.891892	34	14.215978	38052.75	30.336777	5.317944	-0.299827	Large Cap
0.000644	80.952381	34	2.974126	22110.00	24.294797	5.492092	-8.934226	Large Cap
0.000000	97.500000	39	14.496963	25110.11	21.735483	4.147223	-0.285631	Large Cap

- While creating new fields there were null values also observed. Let’s check it out

```
stock_f=df
stock_f.columns
```

```
Index(['Company_Name', 'Ticker', 'Sub_Sector', 'Market_Cap', 'Close_Price',
       'PE_Ratio', '5Y_Historical_Revenue_Growth', '1Y_Forward_Revenue_Growth',
       'Total_Revenue', 'PBT', 'Taxes_And_Other_Items',
       'DII_Holding_Change_3M', 'FII_Holding_Change_3M',
       'MF_Holding_Change_3M', 'Promoter_Holding_Change_3M',
       'Percentage_Buy_Records', 'No_Of_Analysts_With_Buy_Reco',
       'Percentage Upside', 'Net_Profit', 'Profit_Margin', 'PS_Ratio',
       'Change_Retail_Holding', 'MarketCap_Category'],
      dtype='object')
```

```
stock_f.isna().sum()
```

```
Company_Name                    0
Ticker                          0
Sub_Sector                    590
Market_Cap                      0
Close_Price                     0
PE_Ratio                        0
5Y_Historical_Revenue_Growth    0
1Y_Forward_Revenue_Growth       0
Total_Revenue                   0
PBT                             0
Taxes_And_Other_Items           0
DII_Holding_Change_3M           0
FII_Holding_Change_3M           0
MF_Holding_Change_3M            0
Promoter_Holding_Change_3M      0
Percentage_Buy_Records          0
No_Of_Analysts_With_Buy_Reco    0
Percentage Upside               0
Net_Profit                      0
Profit_Margin                 180
PS_Ratio                       27
Change_Retail_Holding           0
MarketCap_Category              0
dtype: int64
```

As we can observe while creating a calculation field we have null Values Lets Treat them First Before starting Data Visualization

- Lets Export Final Dataset for treating null values

```
stock_viz = pd.DataFrame(stock_f)
```

```
stock_viz.to_csv("stock_vizF.csv")
```

Data Cleaning – Part 2

- Import the new dataset in MySQL Workbench.

```
USE AIMA;
DESCRIBE STOCKMARKETFINAL;
SELECT * FROM STOCKMARKETFINAL;
```

- Check Count of Null Values

```
SELECT
    SUM(CASE WHEN Company_Name IS NULL THEN 1 ELSE 0 END) AS Company_Name_null_count,
    SUM(CASE WHEN Ticker IS NULL THEN 1 ELSE 0 END) AS Ticker_null_count,
    SUM(CASE WHEN Sub_Sector IS NULL THEN 1 ELSE 0 END) AS Sub_Sector_null_count,
    SUM(CASE WHEN Market_Cap IS NULL THEN 1 ELSE 0 END) AS Market_Cap_null_count,
    SUM(CASE WHEN Close_Price IS NULL THEN 1 ELSE 0 END) AS Close_Price_null_count,
    SUM(CASE WHEN PE_Ratio IS NULL THEN 1 ELSE 0 END) AS PE_Ratio_null_count,
    SUM(CASE WHEN 5Y_Historical_Revenue_Growth IS NULL THEN 1 ELSE 0 END) AS `5Y_Historical_Revenue_Growth_null_count`,
    SUM(CASE WHEN `1Y_Forward_Revenue_Growth` IS NULL THEN 1 ELSE 0 END) AS `1Y Forward Revenue Growth_null_count`,
    SUM(CASE WHEN Total_Revenue IS NULL THEN 1 ELSE 0 END) AS Total_Revenue_null_count,
    SUM(CASE WHEN PBT IS NULL THEN 1 ELSE 0 END) AS PBT_null_count,
    SUM(CASE WHEN Taxes_And_Other_Items IS NULL THEN 1 ELSE 0 END) AS Taxes_And_Other_Items_null_count,
    SUM(CASE WHEN DII_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS DII_Holding_Change_3M_null_count,
    SUM(CASE WHEN FII_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS FII_Holding_Change_3M_null_count,
    SUM(CASE WHEN MF_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS MF_Holding_Change_3M_null_count,
    SUM(CASE WHEN Promoter_Holding_Change_3M IS NULL THEN 1 ELSE 0 END) AS Promoter_Holding_Change_3M_null_count,
    SUM(CASE WHEN Percentage_Buy_Records IS NULL THEN 1 ELSE 0 END) AS Percentage_Buy_Records_null_count,
    SUM(CASE WHEN No_Of_Analysts_With_Buy_Reco IS NULL THEN 1 ELSE 0 END) AS No_Of_Analysts_With_Buy_Reco_null_count,
    SUM(CASE WHEN `Percentage Upside` IS NULL THEN 1 ELSE 0 END) AS `Percentage Upside_null_count`,
    SUM(CASE WHEN `Net_Profit` IS NULL THEN 1 ELSE 0 END) AS `Net_Profit_null_count`,
    SUM(CASE WHEN `Profit_Margin` IS NULL THEN 1 ELSE 0 END) AS `Profit_Margin_null_count`,
    SUM(CASE WHEN `PS_Ratio` IS NULL THEN 1 ELSE 0 END) AS `PS_Ratio_null_count`,
    SUM(CASE WHEN `Change_Retail_Holding` IS NULL THEN 1 ELSE 0 END) AS `Change_Retail_Holding_null_count`
FROM STOCKMARKETFINAL;
```

- Start treating Null Values.

```
## LETS TREAT NULL VALUES ###################################
SET SQL_SAFE_UPDATES = 0;
## Profit_Margin NULL VALUE COUNT 274
UPDATE STOCKMARKETFINAL
SET `Profit_Margin` = 0
WHERE `Profit_Margin` IS NULL;

SELECT COUNT(`Profit_Margin`) AS total_null_Profit_Margin
FROM STOCKMARKETFINAL
WHERE `Profit_Margin` IS NULL;

## PS_Ratio NULL VALUE COUNT 274
UPDATE STOCKMARKETFINAL
SET `PS_Ratio` = 0
WHERE `PS_Ratio` IS NULL;

SELECT COUNT(`PS_Ratio`) AS total_null_PS_Ratio
FROM STOCKMARKETFINAL
WHERE `PS_Ratio` IS NULL;



SET SQL_SAFE_UPDATES = 1;
## Here we are done with FINAL Data Cleaning
```

Data Visualization

Data Visualization is the art and science of representing data in visual formats such as charts, graphs, and maps to make complex information more understandable and accessible.

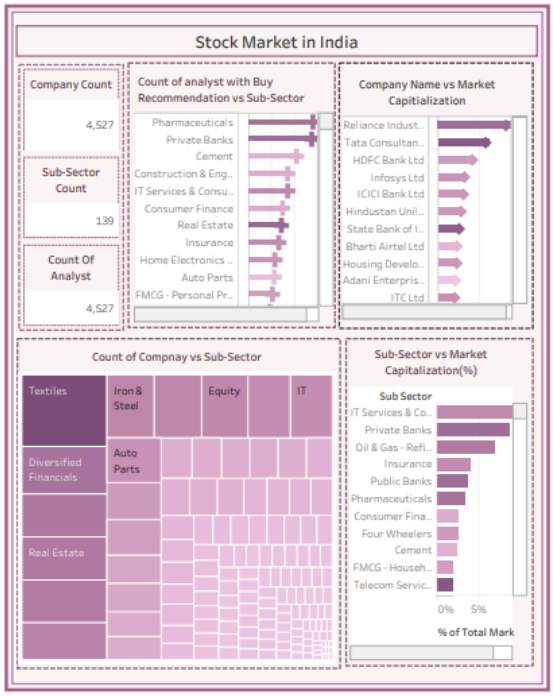
By transforming raw data into meaningful visuals, it allows patterns, trends, and insights to be quickly recognized and communicated. This process aids decision-making, facilitates storytelling, and enables the identification of correlations and outliers that might not be apparent from the data alone.

Effective data visualization enhances comprehension and facilitates data-driven communication across various fields, from business and science to journalism and education.

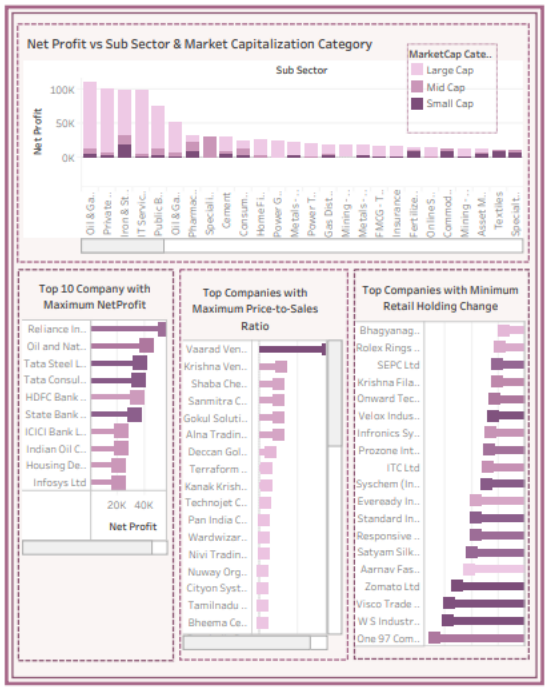
Tableau

- Tableau, a powerful data visualization tool, empowers users to transform raw data into insightful visuals. With an intuitive drag-and-drop interface, it creates interactive dashboards and charts, enabling data-driven decision-making.
- Users can connect to various data sources, perform real-time analysis, and share dynamic reports. Tableau's versatility in handling diverse data types and its user-friendly features make it an essential tool for professionals seeking to uncover trends, patterns, and insights within their data, ultimately enhancing business strategies and communication.

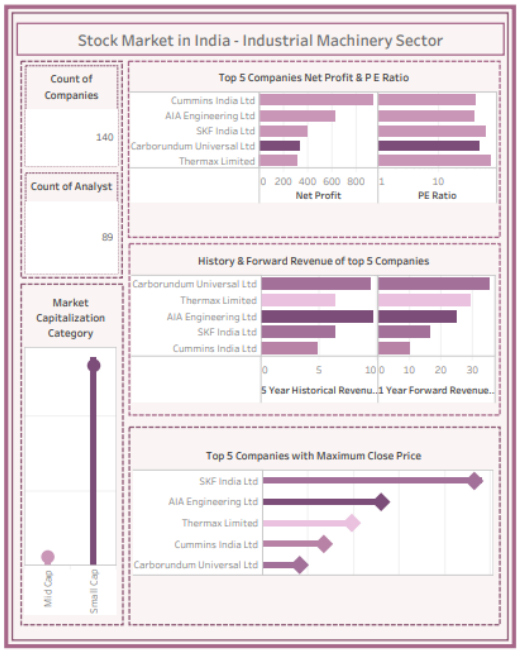
Page 1



Page 2



Page 3



Page 4

Link -
https://public.tableau.com/shared/54SM25M5R?:display_count=n&:origin=viz_share_link

Conclusion

The industrial machinery sector accounts for only 3.56% of the dataset, with a market capitalization of 1.133%. It is further divided into three categories: small capitalization (134 stocks), medium capitalization (6 stocks), and large capitalization (0 stocks).

After detailed analysis, we have identified five stocks in the industrial machinery sector that we believe are the best for investors. These stocks met all of our expectations and outperformed other stocks in the sector.

The five stocks are:

- Cummins India Limited - Cummins in India designs, manufactures, distributes and services diesel and alternative fuel engines from 2.8 to 95 liters, diesel and alternative-fueled power generator sets of up to 3000 kW (3750 kVA), as well as related components and technology.

  Link - https://www.cummins.com/en/in/company/india-companies/cummins-india-limited

- AIA Engineering Limited - AIA Engineering (NSE: AIAENG), Established in the year 1991, a certified ISO 9001 company, specialises in the design, development, manufacture, installation and servicing of high chromium wear, corrosion and abrasion resistant castings used in the cement, mining and thermal power generation industries.

  Link - https://aiaengineering.com/

- SKF India Limited - SKF started its operations in India in 1923 and today provides industry leading automotive and industrial engineered solutions through its five technology-centric platforms: bearings and units, seals, mechatronics, lubrication solutions and services.

  Link - https://www.skf.com/in/organisation/about-skf-india/skf-india-ltd

- Carborundum Universal Limited - Carborundum Universal Ltd, a part of Murugappa Group, is one of the largest and oldest conglomerates in India. CUMI is the leading manufacturer and developer of abrasives, ceramics, refractories, aluminium oxide grains, machine tools, polymers, adhesives and electro minerals in India.

  Link - https://www.cumi-murugappa.com/emd/company-profile/

- Thermax Limited - Thermax Ltd (Thermax) is an engineering company. It provides sustainable solutions for energy and the environment. The company's systems, products and services assist customers in achieving better resource productivity and improving cost efficiency, besides maintaining a cleaner environment. We believe that these five stocks offer investors the potential for attractive returns. They are all well-positioned to benefit from the growth of the industrial machinery sector.

  Link - https://www.thermaxglobal.com/about-us/

We believe that these five stocks offer investors the potential for attractive returns. They are all well-positioned to benefit from the growth of the industrial machinery sector.

Recommendation

We recommend that investors consider investing in the industrial machinery sector. The sector is expected to grow in the coming years, and the five stocks identified in this report offer investors the potential for attractive returns.

Limitation

This report is for informational purposes only and should not be construed as investment advice. Past performance is not indicative of future results. Investors should carefully consider their own investment objectives and risk tolerance before making any investment decisions.

The analysis is based on historical data, which may not be indicative of future results. The analysis does not take into account all factors that could affect the performance of the stocks, such as changes in the macroeconomic environment or the introduction of new technologies. The analysis is based on a limited dataset, which may not be representative of the entire industrial machinery sector. The analysis is subjective and may not be shared by all investors.

APPENDICES

PROFILE OF PROJECT GUIDE

Name: Ashish Kapasiya

Educational Qualification: Postgraduate Degree in Data Science

Professional Experience: 5 Years in Electrical Engineering, 1.5 years in Data Science

Organization: IMS Proschool Pvt Ltd

Current Designation: Assistant Manager – Product and Student Outcome

Brief Profile: An electrical engineer turned data scientist, Ashish Kapasiya is a data science trainer and mentor at IMS Proschool, Delhi. He is also the Assistant Manager of Product and Student Outcome. He conducts data science, machine learning, and database systems workshops across schools in Delhi and also online nationally.

REFERNCE

1. https://www.investopedia.com/articles/stocks/09/indian-stock-market.asp
2. https://m.economictimes.com/
3. https://economictimes.indiatimes.com/
4. https://trendlyne.com/equity/industry/62/industrial-machinery/
5. https://www.thestreet.com/topics/stock/top-rated-equity-machinery
6. https://www.statista.com/statistics/808747/india-investment-in-the-industrial-machinery-sector
7. Challenges and Bottlenecks for Implementing Industry 4.0 in India (isa.org)
8. India's Advanced Manufacturing Sector (trade.gov)
9. https://www.ibef.org/industry/manufacturing-sector-india
10. https://www.mordorintelligence.com/industry-reports/india-agricultural-machinery-market
11. https://www.maiervidorno.com/industry-expertise/machinery-industry-in-india
12. https://www.statista.com/topics/9402/machinery-industry-in-india
13. https://data.gov.in
14. https://www.indiastat.com/data/industries
15. https://www.india.gov.in/topics/industries
16. https://www.crisil.com/en/home/our-businesses/crisil-market-intelligence-and-analytics/india-research/crisil-industry-research.html/