Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

Experiment Design for Creating Visualizations using D3.js on a Finance Dataset

1. Objectives

- To explore and visualize a dataset related to **Finance/Banking/Insurance/Credit** using **D3.js**.
- To create **basic visualizations** (Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot) to understand data distribution and trends.
- To create **advanced visualizations** (Word chart, Box and Whisker plot, Violin plot, Regression plot, 3D chart, Jitter) for deeper insights and complex relationships. ● To perform **hypothesis testing** using the **Pearson correlation coefficient** to evaluate relationships between numerical variables in the dataset.

2. Steps to Perform the Experiment

1. **Choose the Dataset:**
    - Select a dataset related to Finance/Banking/Insurance/Credit. Example datasets:
        - Bank loan data
        - Insurance claims data
        - Credit card transaction data
        - Stock market historical data
2. **Set Up Development Environment:**
    - Install necessary tools:
        - **D3.js**: A JavaScript library for data visualizations.
        - Text editor/IDE (e.g., VS Code).
        - Web server (e.g., Live Server plugin for VS Code) to view D3.js projects.
3. **Data Preprocessing:**
    - Clean and preprocess the data for visualization. Handle missing values, format date-time fields, and categorize data where needed.
4. **Create Basic Visualizations (Using D3.js):**
    - **Bar Chart:** Show the distribution of a categorical variable like loan types, insurance claims by category, etc.
    - **Pie Chart:** Show proportions for categories like credit card transaction types. ○ **Histogram:** Display the distribution of numerical data like loan amounts. ○ **Timeline Chart:** Visualize trends over time (e.g., stock prices over a year). ○ **Scatter Plot:** Display relationships between two numerical variables like income and loan amount.
    - **Bubble Plot:** Visualize multi-variable relationships (e.g., customer age, loan amount, and loan duration).
5. **Create Advanced Visualizations (Using D3.js):**
    - **Word Chart:** Analyze the frequency of words in text data (e.g., claims descriptions).

- **Box and Whisker Plot:** Visualize the spread of financial data (e.g., claim amounts).
- **Violin Plot:** Show the distribution of a variable (e.g., distribution of interest rates)
- **Regression Plot (Linear/Nonlinear):** Explore the relationship between two variables (e.g., loan amount and interest rate).
- **3D Chart:** Visualize multivariate data in 3D (e.g., loan amount, interest rate, and credit score).
- **Jitter Plot:** Display scatter plot points with jitter to avoid overlap, especially for categorical data.

6. **Perform Hypothesis Testing (Using Pearson Correlation Coefficient):** ○ **Step 1:** Formulate a hypothesis (e.g., "There is a positive correlation between customer income and loan approval amount").
   ○ **Step 2:** Calculate the Pearson correlation coefficient to test the hypothesis.
   ○ **Step 3:** Interpret the results (if the coefficient is close to 1, there is a strong positive correlation).
7. **Write Observations:**
   ○ Analyze each chart for trends, patterns, and outliers.
   ○ Identify insights that help in making business decisions (e.g., "Higher loan amounts are correlated with higher credit scores").

3. Programs and Code Examples

Below are some code snippets for creating basic and advanced charts using D3.js:

2. Advanced Charts

2.1 Word Chart (Word Cloud)

For this, you will need a word cloud library for D3.js like d3-cloud.

2.2 Box and Whisker Plot

Box plots are custom in D3.js and require some manual calculations to build them.

2.3 Violin Plot

Violin plots are complex but can be implemented using kernel density estimation.

2.4 Regression Plot

You can implement linear regression by calculating the regression line using D3's line generator and least squares regression.

2.5 3D Chart

For 3D visualizations, you may need to combine D3 with **three.js** for 3D capabilities.

2.6 Jitter Plot

Jitter plots can be implemented by randomly offsetting points slightly to avoid overlapping points.

Dashboard:
This dashboard visualizes financial and performance data for various companies, potentially part of a stock or investment analysis. Here's a breakdown of each chart:

1. Top-Left Line Chart:
   - Objective: Shows the growth trend in revenue or market capitalization for each company over the months.
   - Explanation: Each line represents a company, and the y-axis might display financial metrics such as revenue or market cap in millions or billions. From April to December, this chart visualizes the changes over time, with most companies showing an upward trend, indicating growth.

2. Top-Right Bar Chart:
   - Objective: Displays a specific financial metric (likely revenue, profit, or market cap) for each company at a particular time.
   - Explanation: Each bar represents a company, with the height indicating the value of the metric. This chart is useful for comparing the financial strength or size of each company. For instance, Volkswagen and Deutsche Telekom seem to have high values compared to others.

3. Bottom-Left Scatter Plot:
   - Objective: Compares Return on Assets (ROA) and Return on Equity (ROE) for each company.
   - Explanation: Each point represents a company, plotted based on its ROA (x-axis) and ROE (y-axis). This chart helps analyze the relationship between these two performance indicators, with some companies showing high ROE relative to ROA.

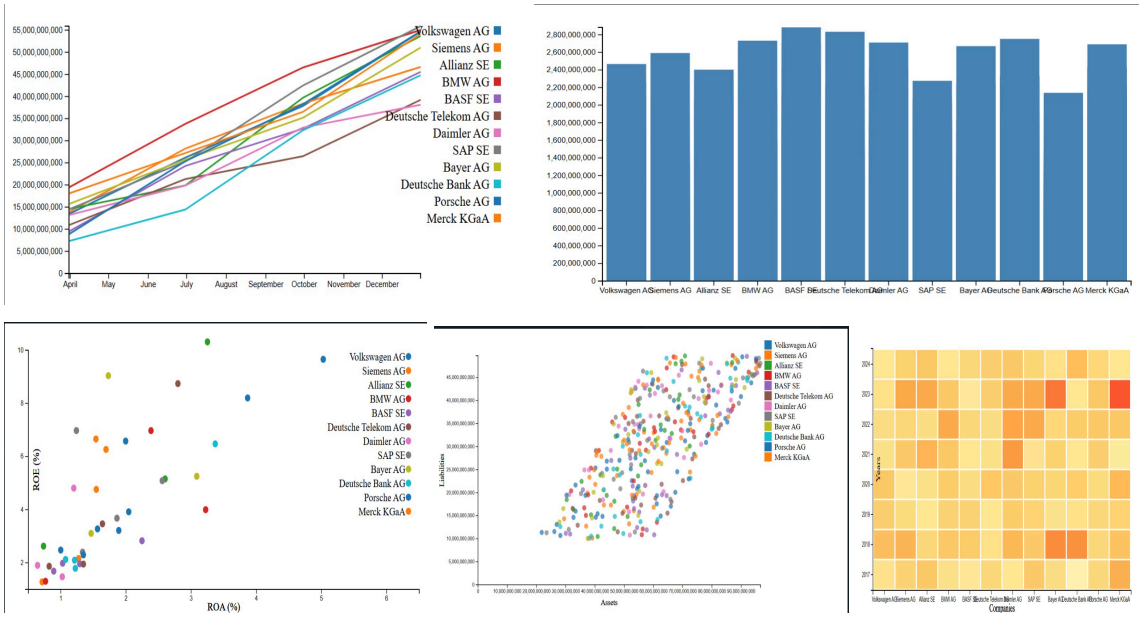
4. Bottom-Middle Scatter Plot:
   - Objective: Visualizes the relationship between two continuous financial variables (e.g., Assets and Revenue).
   - Explanation: Each point represents a company, showing the distribution and correlation between the two metrics. A positive correlation is likely, as many points follow a linear trend.

5. Bottom-Right Heatmap:

- Objective: Represents the correlation or frequency of certain attributes for each company.

- Explanation: Each cell's color indicates the strength of the correlation or magnitude of a value between different attributes or metrics of the companies. Darker shades represent higher values, helping identify which companies or metrics have stronger relationships.

In summary, this dashboard provides insights into the financial growth, performance, and relationships between various metrics of different companies. It's useful for analysts to quickly assess and compare financial health, profitability, and growth trajectories across the companies.



5. FAQs

1. **What is D3.js?**
   - D3.js is a JavaScript library used to create dynamic, interactive data visualizations in web browsers.
2. **What datasets can I use for this experiment?**
   - Financial datasets like bank loans, insurance claims, or stock market data.
3. **What is the purpose of a Pearson correlation coefficient?**
   - To measure the strength and direction of the linear relationship between two continuous variables.
4. **Why should I perform hypothesis testing?**
   - To determine if there is statistical evidence supporting a relationship between variables in your dataset.
5. **What is the advantage of advanced charts like violin plots?**

    - Advanced charts like violin plots provide deeper insights into the data distribution and highlight more complex patterns than simple bar or line charts.

6. Resources

- **Websites:**
    - D3.js Official Documentation
    - Financial Dataset Sources: Datasets related to loans, credit, banking, and insurance.
- **References:**
    - "Data-Driven Documents" by Mike Bostock
    - "Interactive Data Visualization for the Web" by Scott Murray

7. Conclusion

- This experiment enables an understanding of both basic and advanced data visualizations using **D3.js**, giving insights into the finance domain through visual exploration.
- By performing hypothesis testing, you can statistically confirm relationships between variables, which is critical for data-driven decision-making in Finance/Banking/Insurance/Credit sectors.