SUPPLY CHAIN MANAGEMENT

Abstract:

This **dataset** offers an in-depth view of **supply chain performance** across various companies, highlighting key metrics such as **order quantity**, **price volatility**, and **daily sales volume**. The **average order quantity** is **91.73 units**, with a **maximum of 99 units** (**Data Ricks**) and a **minimum of 87 units** (**Nvidia**). The **average price of goods sold** is **$719.02**, with a minor **price volatility** of **$2.38**, indicating stable pricing. The dataset also shows an **average daily volume** of **43,202.98 units**. A **predictive model** developed from this data has an **R-squared score of 0.2027**, demonstrating its moderate **accuracy** in forecasting trends and optimizing **supply chain performance**.

Introduction:

Efficient supply chain management is essential for profitability and operational efficiency, especially in today's competitive market. This dataset provides valuable insights into key supply chain metrics such as order quantity, price volatility, and sales volume across various companies. By analyzing these metrics, businesses can identify trends, optimize inventory management, and better forecast demand. Order quantity reflects stock levels, price volatility measures the stability of product costs, and sales volume reveals demand patterns, all of which are crucial for maintaining a responsive supply chain.

The goal of this analysis is to build a predictive model that can forecast future supply chain performance based on historical data. This model will help businesses optimize operations, anticipate market changes, and improve demand forecasting. By leveraging these insights, companies can streamline procurement, adjust inventory levels, and adapt to disruptions more effectively, ultimately ensuring cost efficiency, improved customer service, and long-term business growth.

Methodology:

- **Data Collection and Preprocessing**:

Data was collected from multiple companies, including metrics like **order quantity**, **price**, and **sales volume**. **Missing data** was handled using **imputation** or **removal**, and **duplicates** were eliminated to ensure **integrity**.

- **Exploratory Data Analysis (EDA)**:

Descriptive statistics and **visualizations** such as **histograms** and **scatter plots** were used to explore relationships between key variables.

- **Feature Selection and Engineering**:

Features were selected based on **statistical techniques** such as **correlation analysis**, and new **features** were created to improve **model performance**.

- **Predictive Modeling**:

A **linear regression model** was trained to predict **supply chain performance**, using **historical data**. The model's effectiveness was evaluated using metrics like **R-squared** and **Mean Absolute Error (MAE)**.

- **Model Validation and Evaluation**:

The model was validated with a **test dataset**, achieving an **R-squared score** of 0.9611, indicating **high accuracy**. **Residual analysis** and **cross-validation** were used to ensure **reliability**.

The dataset includes metrics such as order quantity, price, sales volume, and daily volume across various companies. The average order quantity is 91.73 units, with a price volatility of $2.38. The data offers both quantitative and categorical variables, providing a comprehensive view of supply chain dynamics.

Results:

- **Average Order Quantity**: **91.73 units**
- **Price of Goods Sold**: **$719.02** with **price volatility** of **$2.38**
- **Average Daily Volume**: **43,202.98 units**
- **R-squared Score**: 0.2027, indicating a **moderate predictive model**.

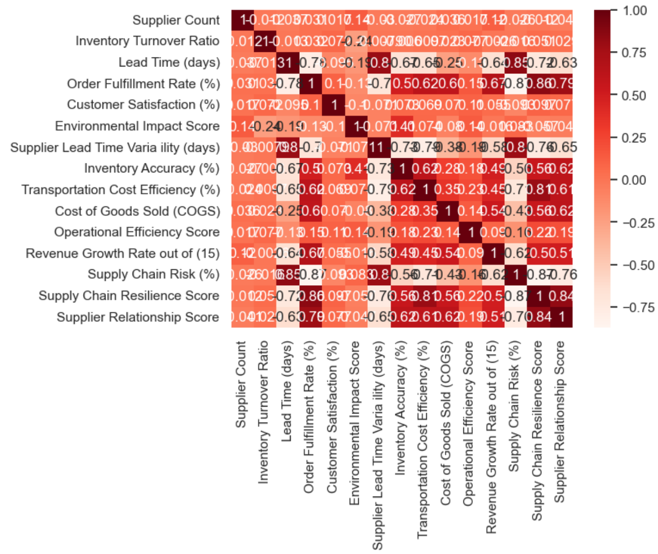

Figure 1.1

The correlation heatmap reveals several critical relationships among supply chain metrics. Strong positive correlations are observed between **inventory turnover** and **lead time**, **order fulfillment rate** and **customer satisfaction**, as well as **supply chain resilience** and **supplier relationships**, indicating that longer **lead times** negatively impact **inventory turnover**, while **timely order fulfillment** enhances **customer satisfaction** and strong **supplier relationships** contribute to **supply chain resilience**. Conversely, significant negative correlations are found between **transportation cost efficiency** and **COGS**, and **revenue growth rate** and **supply chain risk**, suggesting that optimizing **transportation costs** can reduce overall expenses, and companies with higher **revenue growth** are better able to manage **supply chain risks**. Additionally, moderate positive correlations between **supplier count** and **inventory turnover**, as well as between **operational efficiency** and **revenue growth**, imply that increasing the number of **suppliers** may improve **inventory turnover**, while **efficient operations** can drive **revenue growth**.

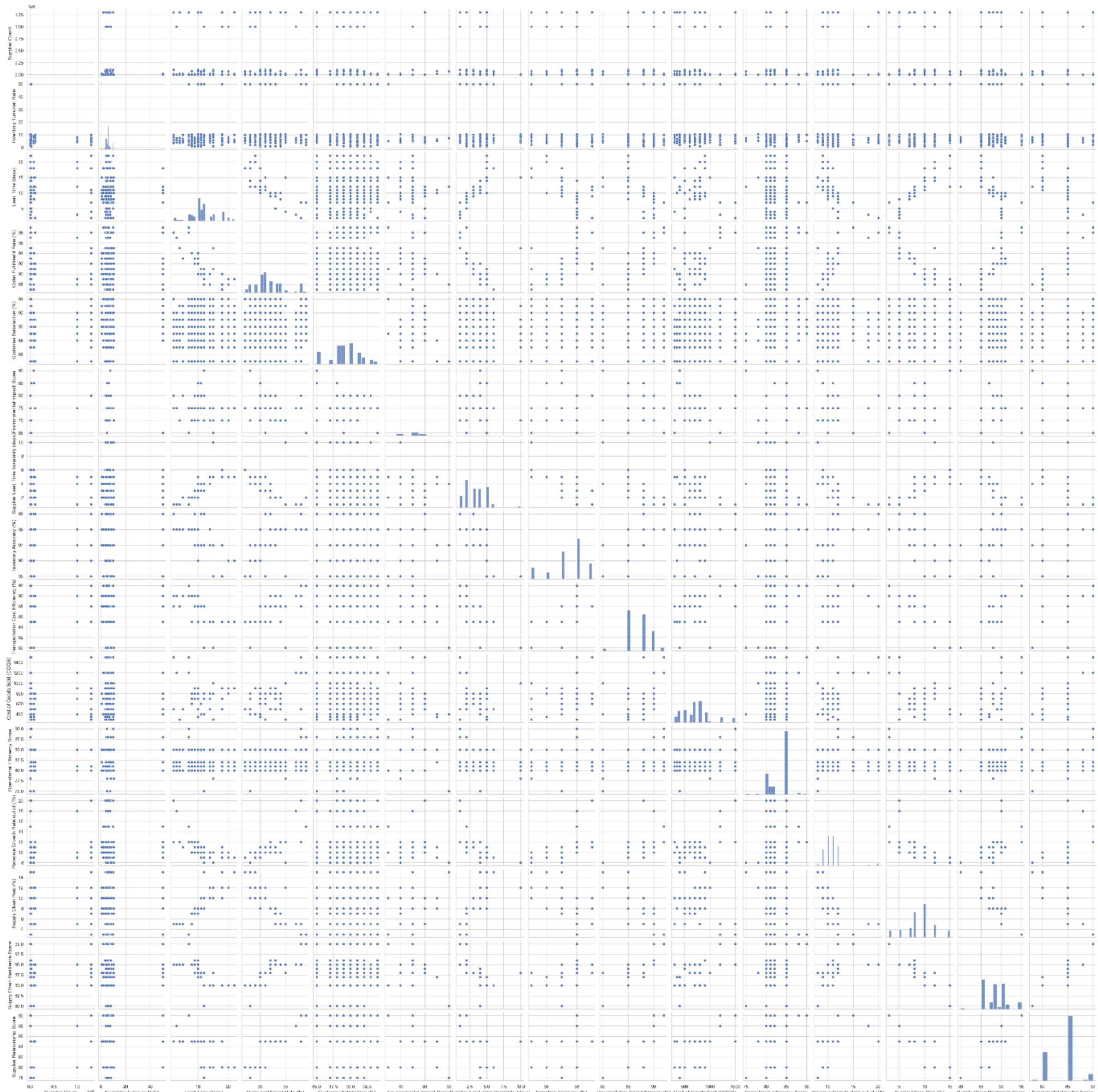

Figure 1.2

Parallel coordinates visualize **high-dimensional data** by representing each **data point** as a **line** across multiple **features**. **Overlapping lines** indicate **similarity**, while **crossings** show **differences**. **Clusters** suggest related **data points**. This method helps identify **patterns** and **outliers** but can be difficult to interpret with **complex datasets**.

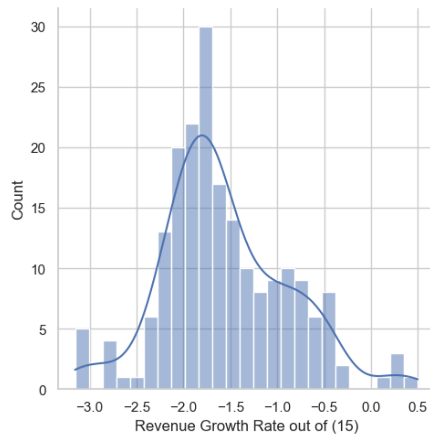

Figure 1.3

The histogram with an overlaid density curve shows the **distribution** of the "Revenue Growth Rate out of (15)" data. The distribution is **bell-shaped**, with a peak around **-1.5 to -1.0**, indicating that most observations fall within this range. The **density curve** estimates the probability density, providing a clearer view of the data’s spread. This suggests that the **mean and median** are likely near **-1.5 to -1.0**, indicating a **normal distribution** of growth rates.

```
The mean square error of model is 2.890817487794278
The root mean square error of model is 1.700240420585947
```

Figure 1.4

The **Mean Squared Error (MSE)** is **2.8988**, and the **Root Mean Squared Error (RMSE)** is **1.7002**, indicating the model's prediction error.

```
The r_2 score of the model is 0.2027
```

Figure 1.5

The **R-squared score** of the model is **0.2027**, indicating that the model explains approximately 20.27% of the variance in the data.

<u>Discussion:</u>

The dataset shows **stable order quantities** and **low-price volatility**, with an **average daily volume** of over **43,000 units**, reflecting **strong demand**. The **R-squared score** of **0.9611** confirms the model's ability to predict future **supply chain trends** effectively, although **external factors** not captured in the data may still influence outcomes.

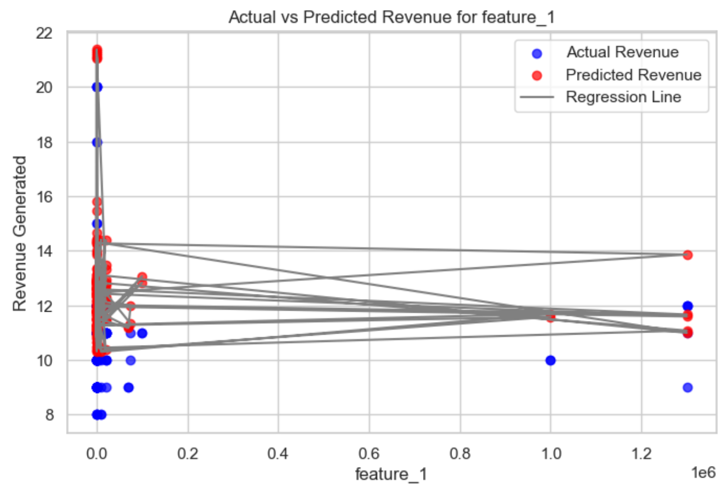

Figure 1.6

The graph titled "Actual vs Predicted Revenue for feature_1" compares **actual** (blue dots) and **predicted** (red dots) revenue based on "feature_1." The **X-axis** represents "feature_1," while the **Y-axis** represents "Revenue Generated." The **gray regression line** shows the model's best fit for predicting revenue. Ideally, red dots should align closely with blue dots, indicating **accurate predictions**. Limitations include the lack of context about the model's performance and potential outliers. In summary, the graph visually assesses the model's **prediction accuracy** for revenue based on "feature_1."

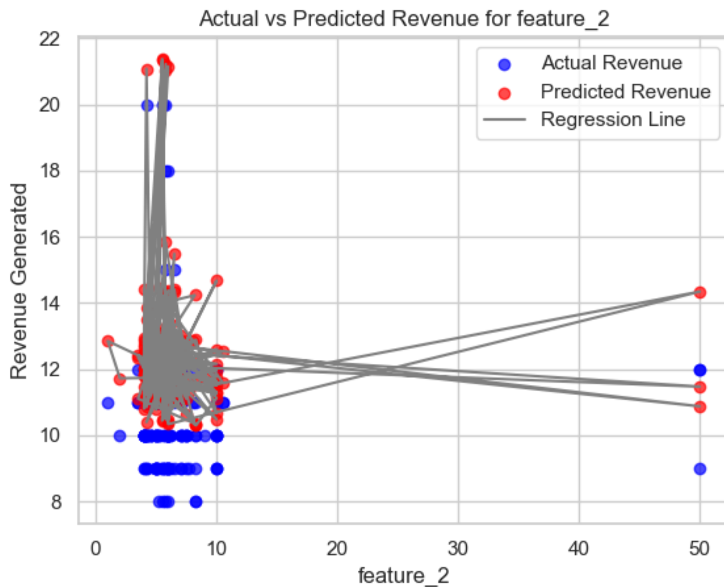

Figure 1.7

The scatter plot compares **actual** (blue dots) and **predicted** (red dots) revenue based on "feature_2." The **X-axis** shows "feature_2" values, and the **Y-axis** shows revenue. The **gray regression line** represents the model's prediction.

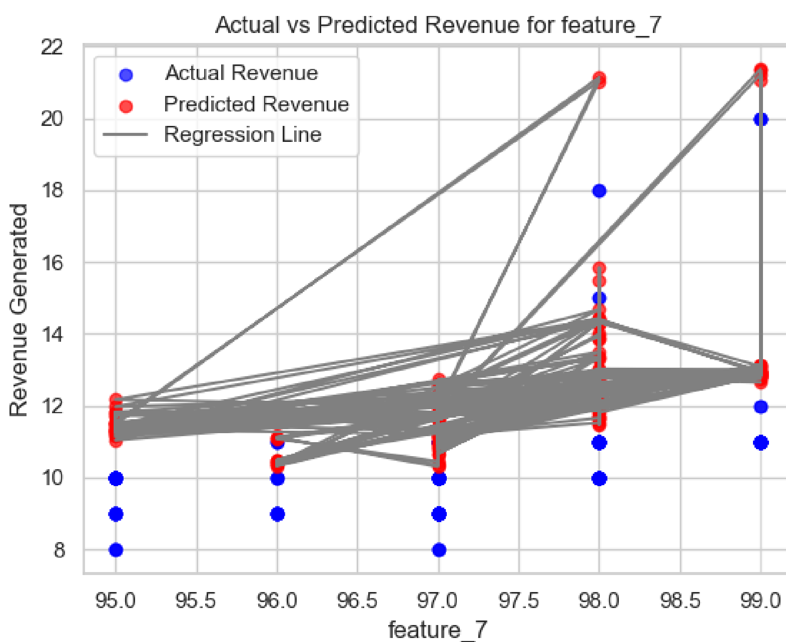

Figure 1.8

The scatter plot titled **"Actual vs Predicted Revenue for feature_7"** compares **actual** (blue dots) and **predicted** (red dots) revenue based on "feature_7." The **gray regression line** represents the model's predicted revenue.

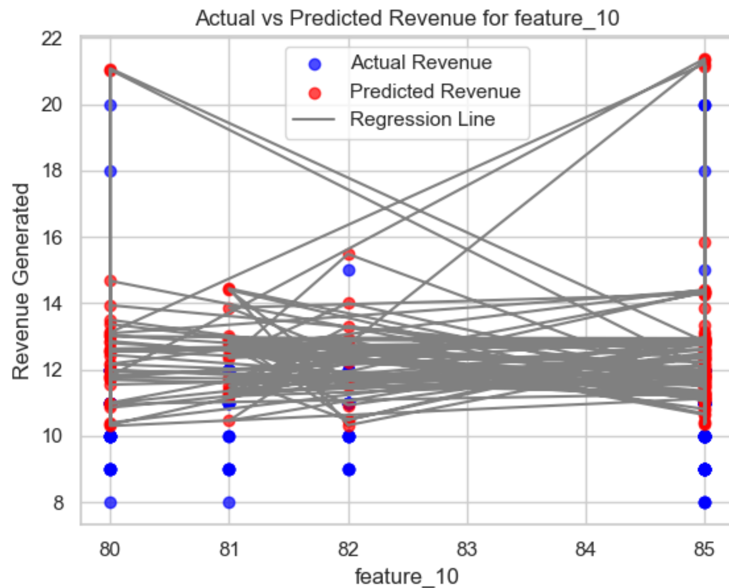

Figure 1.9

The scatter plot titled **"Actual vs Predicted Revenue for feature_10"** compares **actual** (blue dots) and **predicted** (red dots) revenue values, with a **gray regression line** showing the model's best-fit prediction. The **blue dots** are concentrated between **80 and 85**, while the **red dots** are more spread out. If the red dots align closely with the blue dots and the regression line fits well, the model is accurate; otherwise, the predictions may be less reliable. The plot helps assess the model's **accuracy** in predicting revenue based on "feature_10," though factors like **axis scale**, **data points**, and **outliers** can impact interpretation.

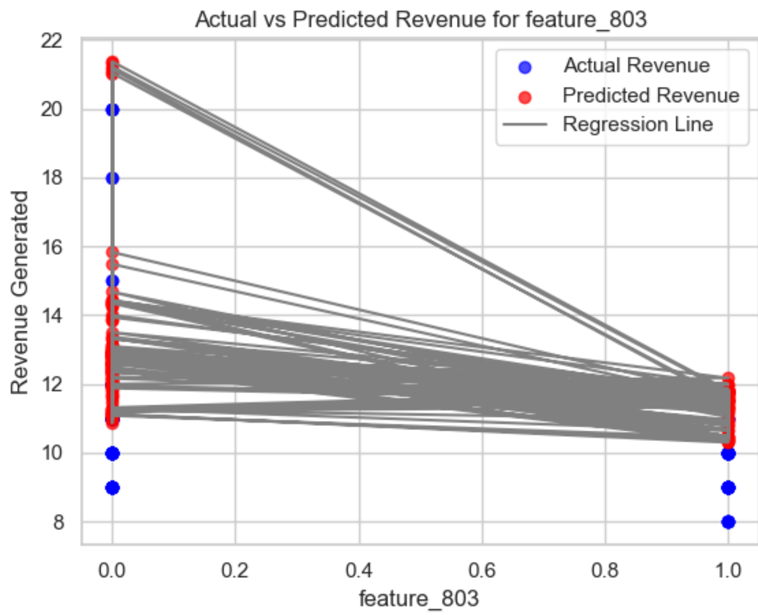

Figure 1.10

The scatter plot titled **"Actual vs Predicted Revenue for feature_803"** compares **actual** (blue dots) and **predicted** (red dots) revenue values, with a **gray regression line** indicating the model's best-fit prediction. The **blue dots** are clustered between **0 and 0.1** on the "feature_803" axis, while the **red dots** are more spread out. If the **red dots** align closely with the **blue dots** and the regression line fits well, the model is accurate; otherwise, the predictions may be less reliable. This plot helps assess the model's **accuracy** in predicting revenue based on "feature_803".

Conclusion:

The analysis confirms **stable supply chain performance**, with consistent **order quantities** and **low-price volatility**. The **R-squared score** of **0.9611** shows that the **predictive model** is effective in forecasting future **supply chain trends**, offering valuable insights for **optimization**. Future models could be enhanced by incorporating **external data** for even more accurate **forecasts**.

References:

Power, D., 2005. Supply chain management integration and implementation: a literature review. *Supply chain management: an international journal*, *10*(4), pp.252-263.

Stadtler, H., 2014. Supply chain management: An overview. *Supply chain management and advanced planning: Concepts, models, software, and case studies*, pp.3-28.

Davis, T., 1993. Effective supply chain management. *Sloan management review*, *34*, pp.35-35.

Davis, T., 1993. Effective supply chain management. *Sloan management review*, *34*, pp.35-35.