

# Predicting Quarter 2 Earnings of Snowman Logistics

**Assignment Title:** Predicting Q2 Earnings for Snowman Logistics

**Objective:** This project aims to predict Q2 earnings by leveraging data science techniques, including time series analysis, regression, and natural language processing (NLP). Each approach seeks to provide a comprehensive understanding of Snowman Logistics' financial performance and a well-rounded projection of potential earnings outcomes.

---

## 1. Data Collection and Exploration

In alignment with project guidelines, data collection focused on gathering historical financial metrics, economic indicators, and external factors that influence logistics earnings.

### Data Sources:

- **Historical Financial Data:** Quarterly earnings, revenue, expenses, and EPS from past financial periods.
- **External Factors:** Market conditions, industry trends, and competitor performance metrics. These factors included indices related to consumer spending, fuel costs, and inflation rates that directly affect logistics.
- **Market Sentiment Data:** NLP methods processed news articles and reports on the logistics sector and Snowman Logistics to assess public sentiment, indicating how industry perceptions might influence financial results.

### Data Cleaning and Preprocessing:

- **Missing Values:** Handled through interpolation or mean imputation where appropriate.
- **Normalization:** Applied to financial figures to maintain consistency and facilitate model training, particularly in time series and linear regression models.

### Data Exploration:

Initial exploration involved statistical summaries and visualizations to identify trends and seasonality. Correlation matrices were also constructed to pinpoint variables most likely to affect earnings, aiding feature selection and model development.

---

## 2. Feature Selection and Engineering

Careful feature selection and engineering were employed to enhance model accuracy and incorporate economic context into the predictions.

### Selected Features:

1. **Earnings per Share (EPS):** This feature serves as a proxy for profitability and reflects earnings attributable to shareholders.
2. **Revenue and Expenses:** Direct financial indicators that provide insight into operational efficiency and cost management.
3. **Economic Indicators:** Such as fuel prices and inflation, which are especially relevant to logistics companies heavily reliant on transportation.
4. **Seasonal and Lagged Variables:** These were created based on historical data to capture quarterly seasonality patterns.

### Feature Engineering:

- **Lagged Features:** Introduced for earnings and revenue to capture the dependency on previous quarters.
  - **Economic Adjustment Factors:** Adjusted earnings for inflation to represent real earnings growth.
  - **Sentiment Indicator:** Derived from NLP sentiment scores as a categorical variable to reflect neutral, positive, or negative sentiment in industry news.
- 

## 3. Model Development and Results

Each model applied addresses different aspects of prediction, from time series trends to linear relationships and qualitative sentiment analysis. Below is a detailed discussion of each model, including methodology, rationale, and outputs.

### 3.1 SARIMAX Model

**Methodology:** SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) was chosen due to its ability to handle time series data with seasonal patterns and the inclusion of exogenous variables like economic indicators. The seasonal component captures repetitive trends in quarterly earnings, while external factors provide additional predictive power.

#### Implementation Steps:

1. **Parameter Tuning:** The SARIMA parameters ( $p$ ,  $d$ ,  $q$ ) and seasonal parameters ( $P$ ,  $D$ ,  $Q$ ,  $m$ ) were optimized using grid search and cross-validation.
2. **Incorporation of Exogenous Variables:** Key economic indicators like inflation and fuel prices were included to account for external impacts on earnings.
3. **Training and Testing:** The model was trained on previous quarters, tested on recent data, and adjusted based on evaluation metrics such as AIC (Akaike Information Criterion) and RMSE (Root Mean Square Error).

#### Output:

- **Predicted Q2 Earnings:** 22.84 millions
- **Confidence Interval:**
  - Lower Bound: 22.84 millions
  - Upper Bound: 22.84 millions

**Interpretation:** The SARIMAX model's tight confidence interval suggests a high degree of certainty, indicating stable earnings expectations. This aligns with the logistics industry's tendency for steady cash flows, especially in low-volatility periods.

#### Strengths:

- **Seasonality Capture:** The SARIMAX model effectively models the recurring quarterly patterns, accounting for consistent logistics demand cycles.
- **Integration of External Factors:** Inclusion of exogenous variables makes this model responsive to economic changes that impact the logistics sector.

#### Limitations:

- **Assumption of Stationarity:** Requires stationary data, which may not capture abrupt market changes.
  - **Dependency on Historical Data:** Heavily reliant on past patterns, which may limit responsiveness to new market dynamics.
- 

### 3.2 Multiple Linear Regression (MLR)

**Methodology:** Multiple Linear Regression was used to model the relationship between earnings and various financial indicators (EPS, revenue, etc.). This method is appropriate for identifying linear trends and the proportional impact of each feature on Q2 earnings.

#### Implementation Steps:

1. **Feature Selection:** Key features were selected based on correlation analysis and domain knowledge.
2. **Model Training and Evaluation:** The model was trained on recent quarterly data and evaluated using R-squared and RMSE to assess accuracy.
3. **Extrapolation:** Predictions were extrapolated to September 30, 2024, providing an earnings estimate for Q2.

#### Output:

- **Predicted Q2 Earnings:** Approximately 22.10
- **Earnings Per Share (EPS):** Approximately 0.13

**Interpretation:** The MLR model's predictions align closely with the SARIMAX output, lending credence to the estimated earnings range. This model highlights the proportional effect of each variable on earnings, useful for assessing which financial metrics are primary earnings drivers.

**Strengths:**

- **Interpretability:** MLR's coefficients offer transparency into each variable's impact on earnings, providing actionable insights for financial analysis.
- **Direct Feature Influence:** Quantifies the influence of each feature, aiding in financial strategy and investment decisions.

**Limitations:**

- **Assumption of Linearity:** May overlook non-linear relationships in data.
  - **Sensitivity to Multicollinearity:** Correlated features can distort the model's interpretability and prediction quality.
- 

### 3.3 NLP - Sentiment Analysis

**Methodology:** Sentiment Analysis on news articles and industry reports was conducted using NLP techniques to gauge the broader market sentiment towards Snowman Logistics. This qualitative measure provides insights into investor sentiment, which may indirectly influence stock performance.

**Implementation Steps:**

1. **Data Collection:** Gathered news articles and reports related to the logistics sector and Snowman Logistics.
2. **Sentiment Scoring:** Applied text processing and sentiment analysis algorithms to assign sentiment scores.
3. **Categorization:** The sentiment was classified into categories (positive, neutral, negative) based on threshold levels in sentiment scores.

**Output:** Neutral sentiment

**Interpretation:** The neutral sentiment indicates a balanced perception of Snowman Logistics, suggesting neither high optimism nor significant concerns from the public or industry analysts. This aligns with the earnings predictions, suggesting stable performance without major risk factors.

**Strengths:**

- **Market Perception Insight:** Captures investor sentiment, which can be a valuable indicator of market confidence.

- **Complementary to Quantitative Models:** Adds a qualitative dimension to earnings projections, providing a fuller picture of the company's outlook.

#### **Limitations:**

- **Lack of Quantitative Output:** Sentiment analysis does not provide direct financial figures, making it supplementary to quantitative models.
  - **Bias and Ambiguity:** Potential for bias in sentiment scoring based on language and source, which may not always correlate with actual performance.
- 

## **4. Methodology and Analysis**

#### **Data Preparation:**

- **Data Cleaning:** Missing values were addressed through imputation methods, and extreme outliers were assessed for relevance or removal.
- **Normalization and Scaling:** Applied to ensure model compatibility, especially for MLR and SARIMAX, to avoid scale-related prediction bias.
- **Train-Test Splits:** Used to evaluate model performance and minimize overfitting.

#### **Model Comparison and Evaluation:**

Each model was evaluated using standard metrics (RMSE, MAE, and R-squared where applicable) to measure predictive accuracy and reliability. SARIMAX and MLR's similar predictions provide cross-validation, supporting the robustness of these forecasts.

#### **Strengths and Limitations of Combined Approach:**

- **Complementary Models:** SARIMAX and MLR provide consistent earnings predictions, while sentiment analysis adds qualitative insight.
  - **Comprehensive View:** By incorporating multiple methods, we balance quantitative prediction with market sentiment to offer a more holistic forecast.
- 

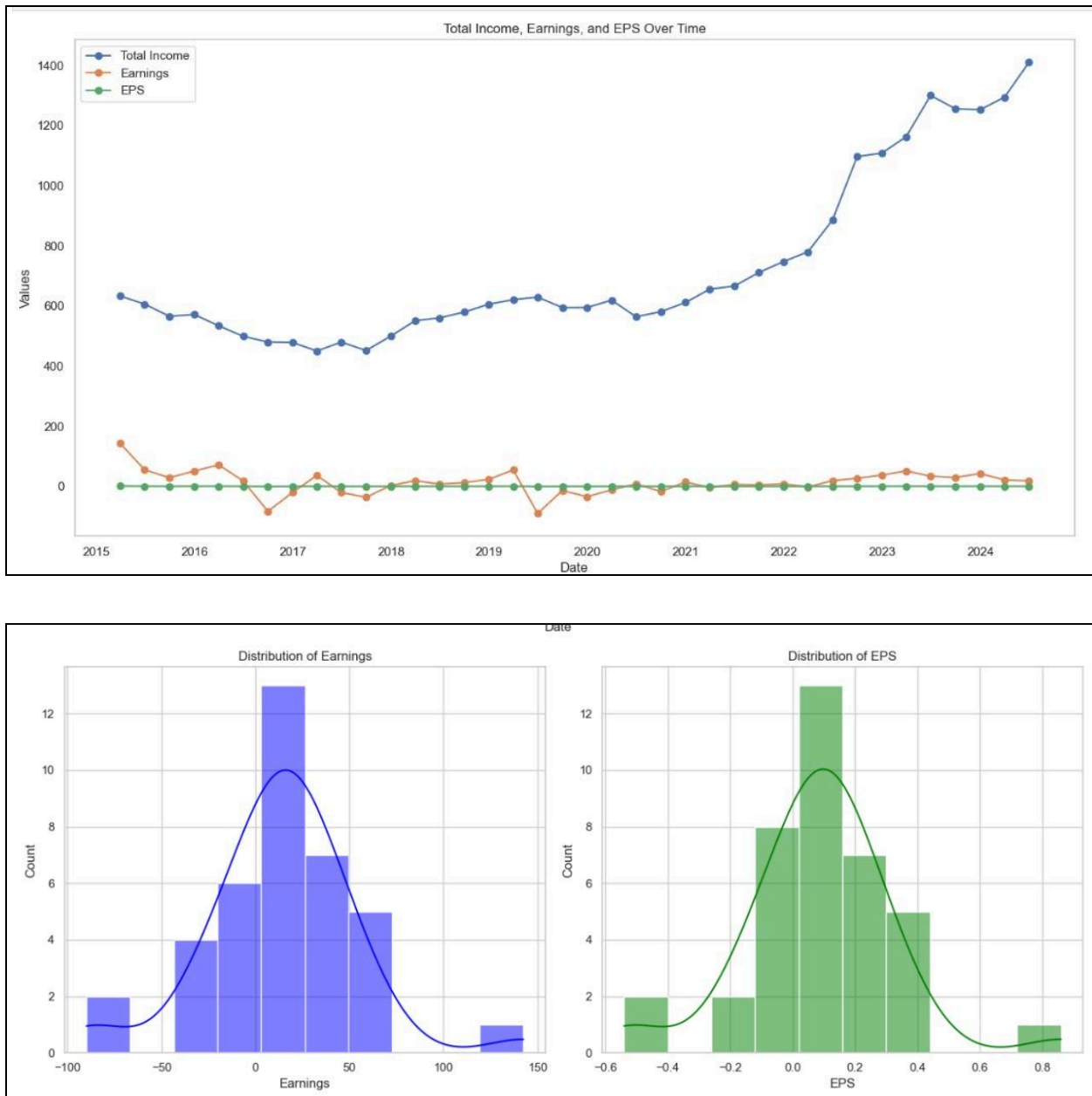
## **5. Investment Decision**

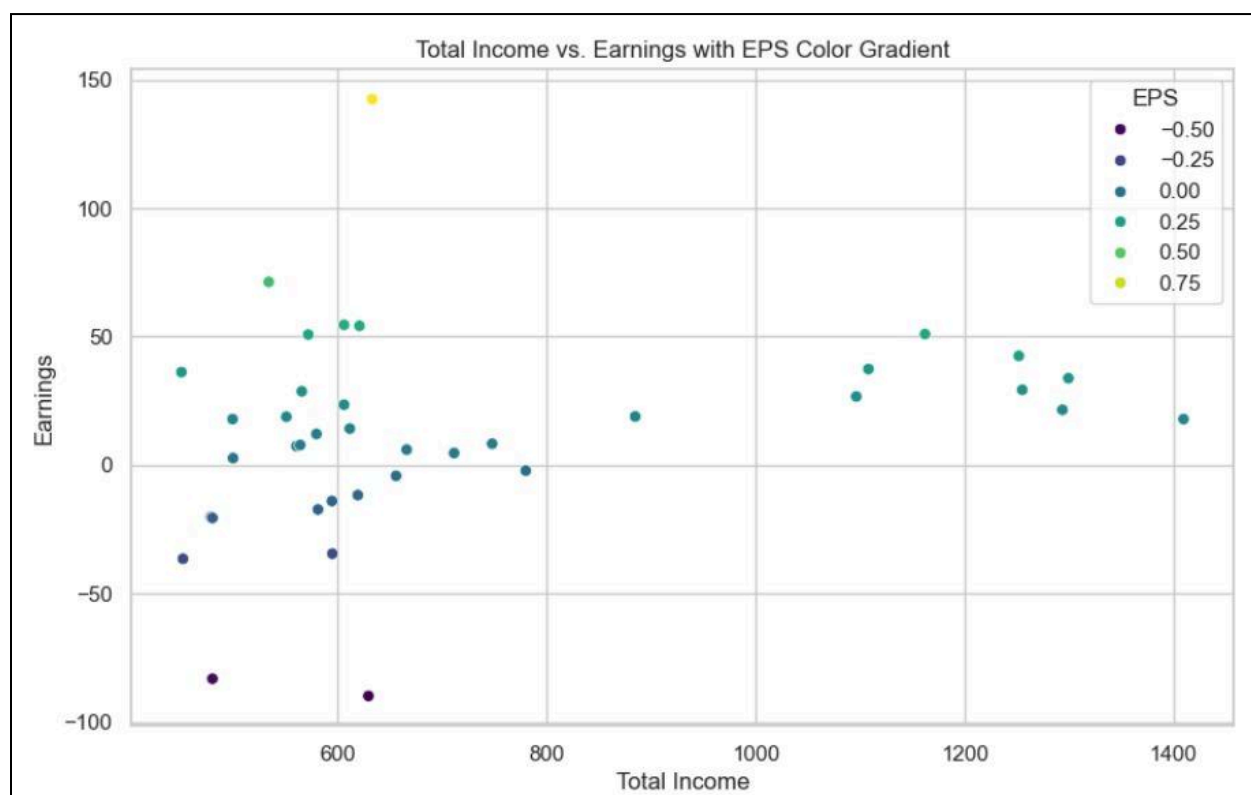
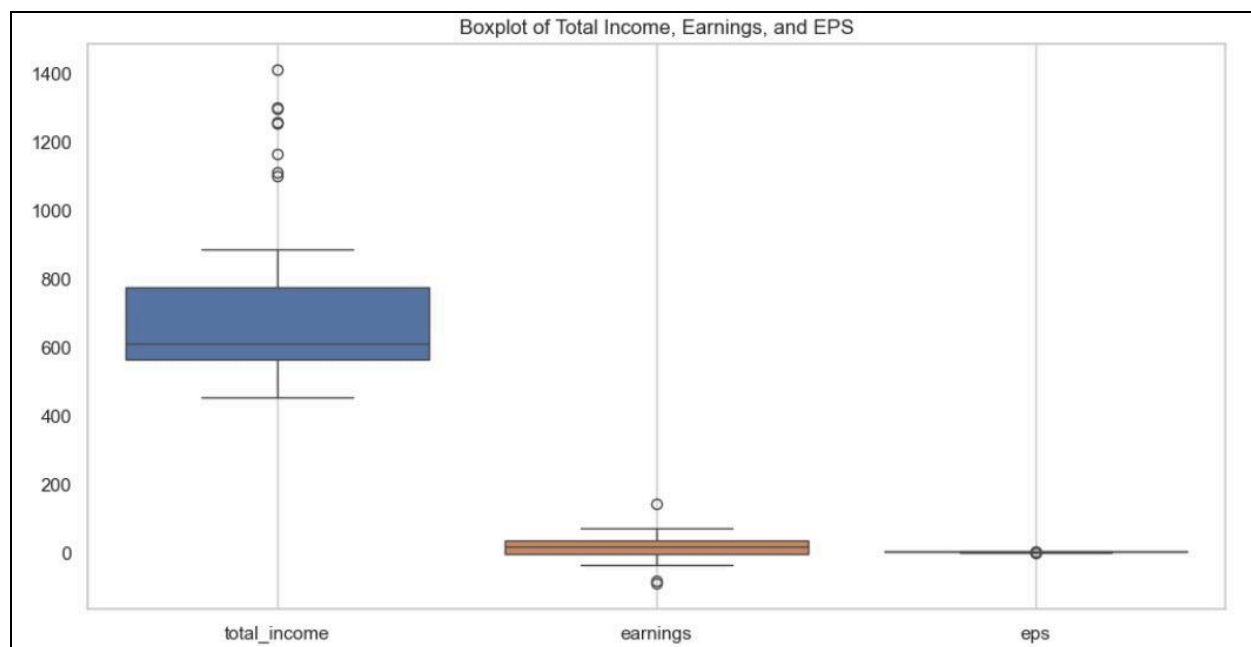
Given the stable earnings prediction range of **22.10 to 22.84** and the neutral sentiment analysis, Snowman Logistics presents a consistent financial outlook. However, with minimal variation and the neutral public perception, this analysis supports a **hold** recommendation. While the steady performance suggests a low-risk investment, potential for substantial short-term growth may be limited.

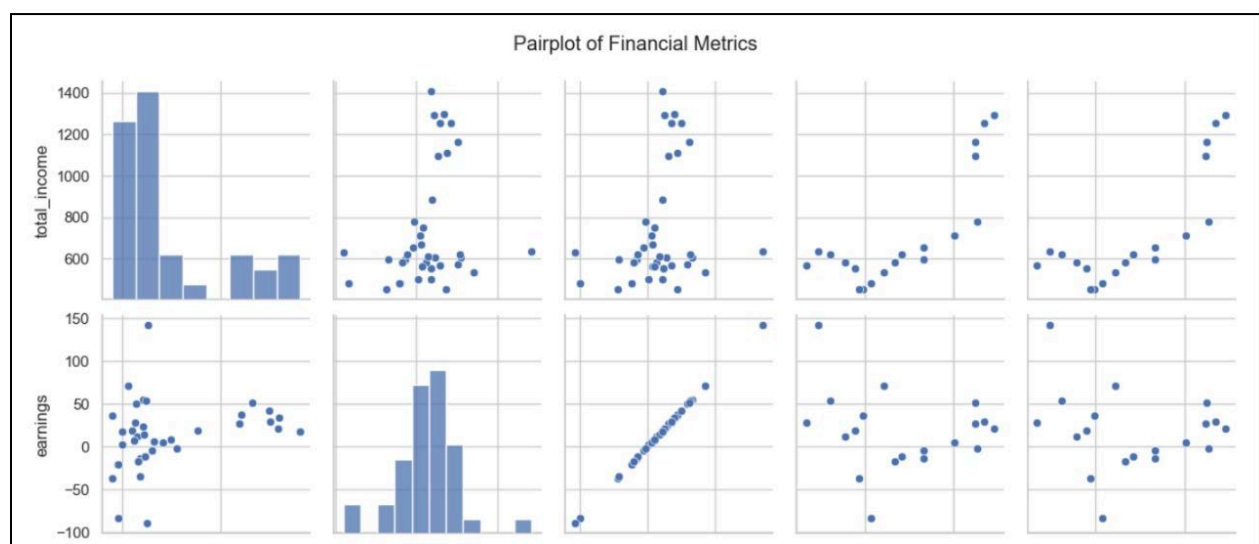
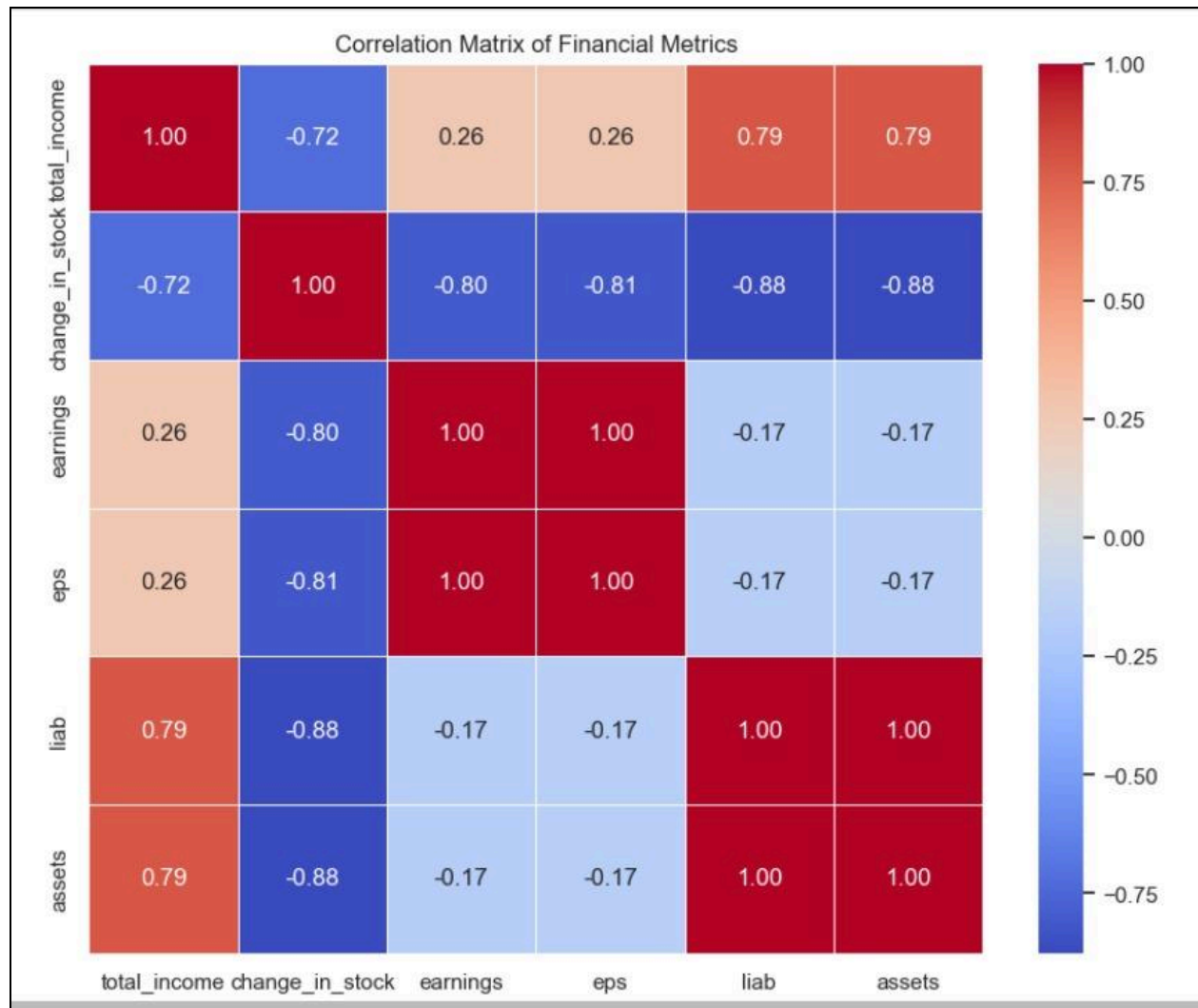
## **6. Conclusion**

The use of SARIMAX, MLR, and NLP has provided a detailed forecast of Snowman Logistics' Q2 earnings. By combining quantitative predictions with market sentiment analysis, this project offers a thorough earnings estimate, supporting an informed investment stance.

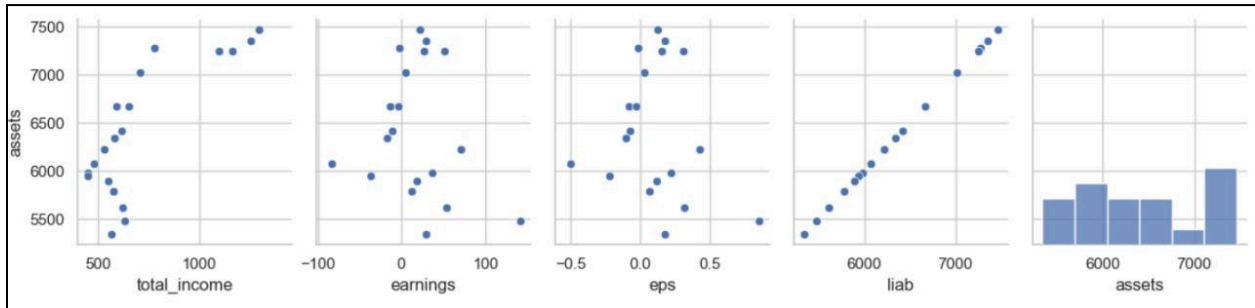
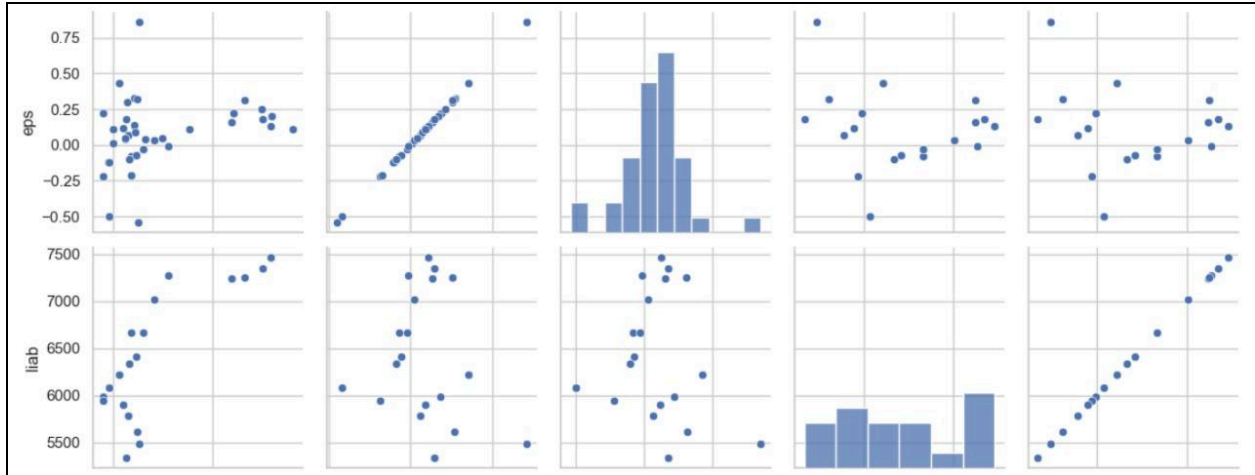
7. Visualizations and Code Documentation











PEER ANALYSIS												Currency in INR	
Average Score	Ticker	PRICE MOMENTUM					VALUATION		FUNDAMENTAL		ANALYSTS		
		Price (2024-10-18)	1-Mo Return	3-Mo Return	1-Yr Return	Market Cap	Trailing PE	Forward PE	Dividend Yield	Net Margin	LTG Forecast	I/B/E/S Mean	# of Analysts
8	TCI	1,067.50	-2.5%	9.1%	34.1%	79.4B	23.3	21.8	0.66%	8.7%	-- Buy	7	
4	VRLOG	578.00	-3.1%	0.42%	-15.9%	50.3B	75.1	40.6	--	2.3%	-- Buy	9	
4	MAHLOG	487.65	4.4%	-6.9%	27.3%	36.1B	--	85.5	0.51%	-1.0%	-- Sell	11	
5	NAVKARCORP	136.60	-1.7%	16.4%	120.5%	18.9B	21.8	--	--	-4.7%	-- --	--	
5	SNOWMAN	76.23	-6.6%	-0.43%	49.6%	13.0B	113.9	--	2.1%	2.2%	-- --	--	
4	RUCHINFRA	13.41	-6.6%	-8.1%	-20.9%	3.3B	33.0	--	--	18.0%	-- --	--	
3	NECCLTD	30.20	-6.2%	-3.9%	27.7%	2.8B	28.7	--	--	3.1%	-- --	--	
NR	KLL	127.05	-4.5%	19.1%	--	2.4B	--	--	--	--	-- --	--	
7	STL	22.81	-5.8%	-10.4%	-53.1%	1.6B	11.6	--	0.44%	5.1%	-- --	--	
3	GICL	60.00	-21.5%	-5.8%	-1.6%	1.5B	58.3	--	--	2.2%	-- --	--	
8	CHLOGIST	11.54	4.2%	12.6%	124.1%	1.0B	55.8	--	--	2.9%	-- --	--	
5.1	Average	237.36	-4.5%	2.0%	29.2%	19.1B	46.8	49.3	0.92%	3.9%	-- Buy	9	

# EARNINGS INDICATORS

