ANALYSIS OF BUSINESS PERFORMANCE MANAGEMENT OF SOFT DRINK ENTERPRISES

Through Coca - Cola

Presented by
GROUP 01

27 December, 2024

MEMBERS OF GROUP 01

NAME	STUDENT ID
PHẠM HỒNG NHUNG (Leader)	22080070
NGUYỄN THỊ THÙY DUNG	22080021
NGUYỄN HỮU MINH HOÀNG	22080034
TRẦN PHÚC HƯNG	22080037
NGUYỄN THƯƠNG HUYỀN	22080042
LÊ MAI LINH	22080047
TRẦN HƯƠNG THẢO	22080083
NGUYỄN DUY TÙNG	22080092

Introduction to Coca-Cola's Competitive Landscape

Consumer Priorities:
- Increasing focus on healthy and sustainable products

Competitive Dynamics:
- Coca-Cola (CCE):
  - Innovative promotions.
  - Portfolio of over 4,300 products.
- PepsiCo:
  - Generates billions annually with 22 major brands.

Research Objectives:
Evaluate business performance through financial indicators (focus on stock metrics)
Assess human resource factors influencing performance.

Methodology:
Qualitative Analysis: Identifies strategies impacting performance.
Quantitative Analysis: Utilizes algorithms (e.g., linear regression, KNN, logistic regression).

Coca-Cola's Business Performance Management

- **Market Leadership**
  - CCE leads the carbonated soft drink market with over 400 brands in 200 countries
- **Financial Performance**
  - 24% profit margin
  - 0.48 debt/equity ratio indicates financial stability.
- **Strategic Approaches**
  - "Glocal" Model: Tailors strategies to local markets.
  - Balanced Scorecard (BSC): Enhances efficiency and cost optimization.
- **Human Resource Strategies**
  - Group interviews and internal selection.
  - 70:20:10 training model boosts employee motivation and morale.
- **Global Success Factors**
  - $61 billion brand value and advanced supply chains.
  - Focus on healthier products and digital transformation.
  - Use of Decision Support Systems (DSS) for strategic decision-making.

Linear Regression Model for Coca-Cola Stock Prices

I. INTRODUCTION

Objective:

Implement a Linear Regression model to predict future closing prices of Coca-Cola stock.

Dataset Features:

Input Variable: Opening price

Target Variable: Closing price

1. Data Pre-Processing

Step 1: Check data types for cleaning.

```
df.info()
```

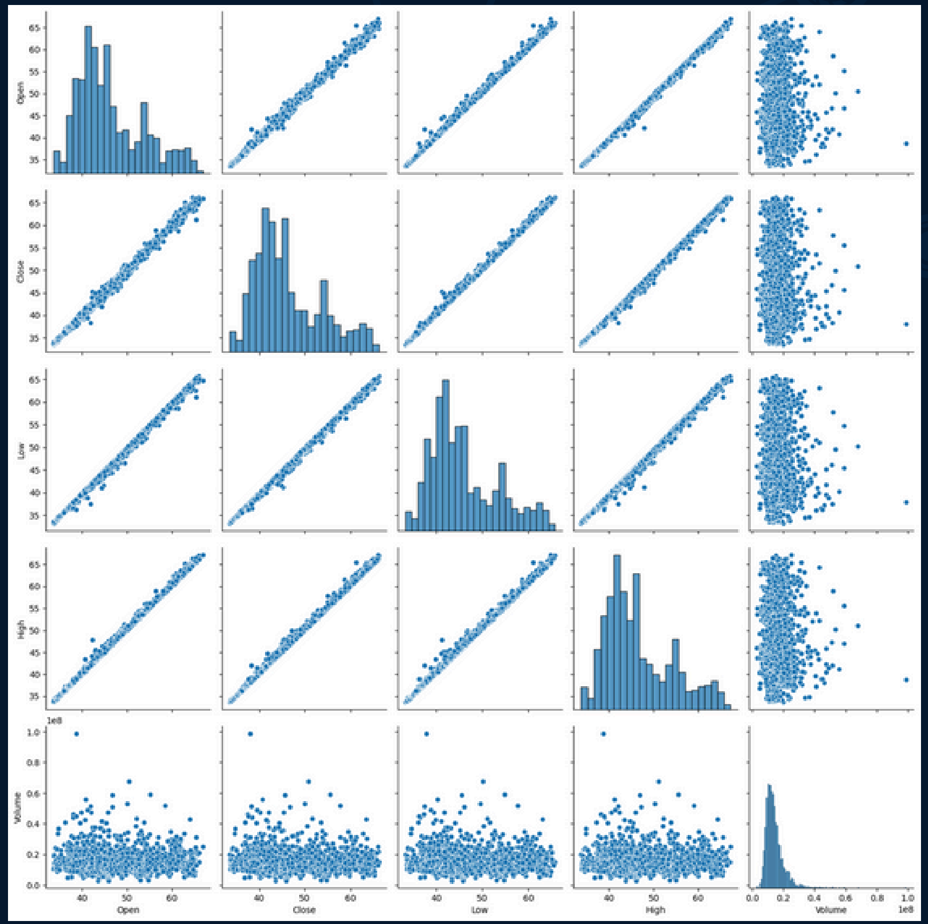
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2768 entries, 0 to 2767
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Date          2768 non-null   object
 1   Open          2768 non-null   float64
 2   High          2768 non-null   float64
 3   Low           2768 non-null   float64
 4   Close         2768 non-null   float64
 5   Volume        2768 non-null   int64
 6   Dividends     2768 non-null   float64
 7   Stock Splits  2768 non-null   int64
dtypes: float64(5), int64(2), object(1)
memory usage: 173.1+ KB
```

	Open	High	Low	Close	Volume	Dividends	Stock Splits
count	2768.000000	2768.000000	2768.000000	2768.000000	2.768000e+03	2768.000000	2768.000000
mean	46.240986	46.550983	45.925972	46.244491	1.439106e+07	0.005549	0.000723
std	7.337608	7.415198	7.255777	7.340632	6.255951e+06	0.044749	0.038014
min	33.650000	33.720000	33.280000	33.490000	2.996300e+06	0.000000	0.000000
25%	40.857500	41.110000	40.630000	40.880000	1.050740e+07	0.000000	0.000000
50%	44.495000	44.775000	44.150000	44.510000	1.307250e+07	0.000000	0.000000
75%	50.990000	51.365000	50.472500	50.872500	1.650562e+07	0.000000	0.000000
max	67.000000	67.200000	65.720000	66.210000	9.896750e+07	0.440000	2.000000

Step 2: Confirm no null or missing values.

Step 3: Remove unnecessary columns:

- Dividends and Stock Splits (constant values).



Summary: Data is clean; Date column is not required for modeling.

```
irrelevant_columns = ['Dividends', 'Stock Splits']
df = df.drop(columns=irrelevant_columns)
```

Data Analysis

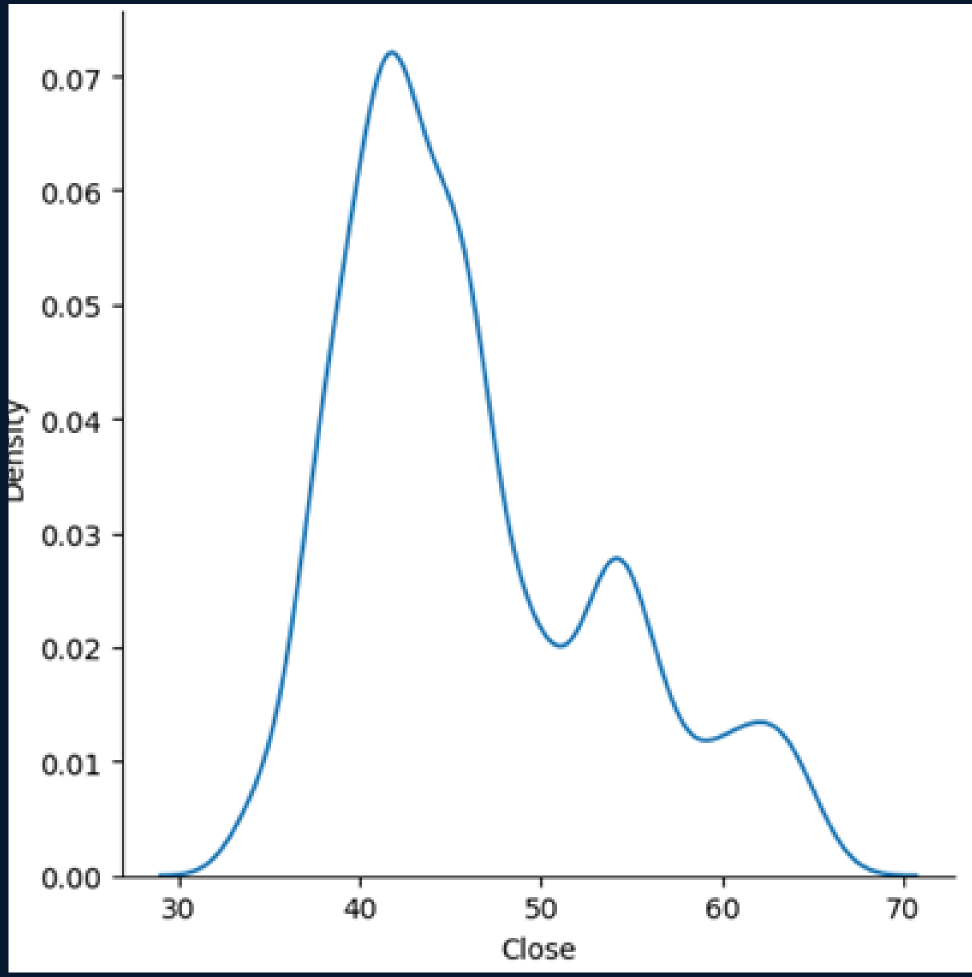
Correlation: Strong correlation (0.998) between Opening and Closing prices.

```
Correlation = df['Open'].corr(df['Close'])
print(f'The correlation coefficient between "Open" and "Close" is: {Correlation}')
```

```
The correlation coefficient between "Open" and "Close" is: 0.998191824395903
```

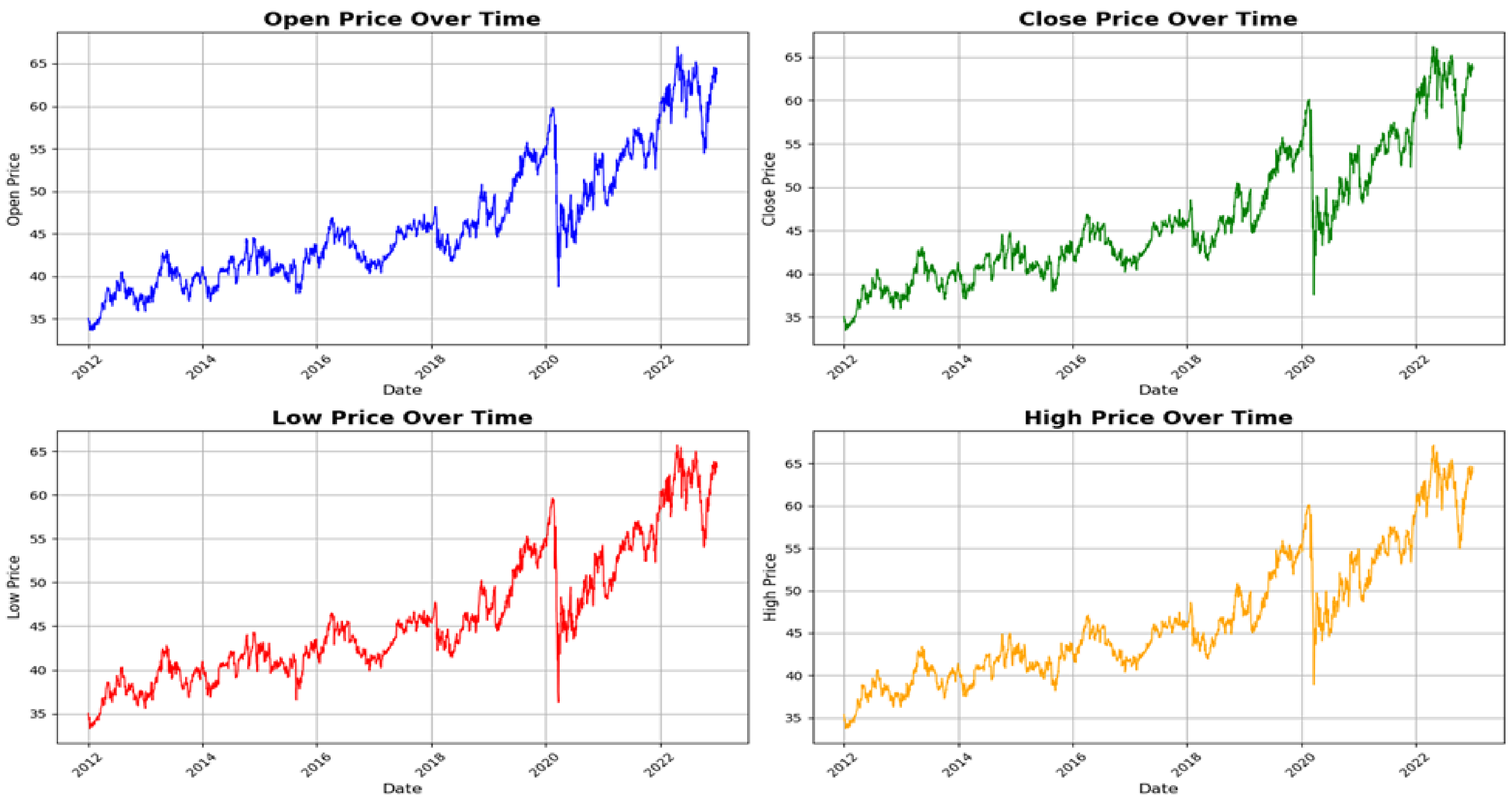
Visualizations

- **KDE Chart**: Distribution of Closing prices.



Visualizations

- **OHLC Chart**: Stock price fluctuations over time.



2. Building the model

Visualizing Relationships

- Chart A & B:Visual representation of two key variables:
  - Target Variable: Close
  - Input Variable: Open

Linear Regression Model

- Utilized Linear Regression to understand the relationship between variables.
- **Coefficients:**
  - Slope (a): 0.99860318
  - Intercept (b): 0.06809476
- The regression line is illustrated in the charts, indicating a strong correlation.

Model Training and Testing

- **Divided the test data into:**
  - X_train: Input features for training
  - y_train: Target variable for training
  - X_test: Input features for testing
  - y_test: Target variable for testing
- **Conducted training to produce predictions:**
  - y_train_pred: Predictions on training data
  - y_test_pred: Predictions on test data

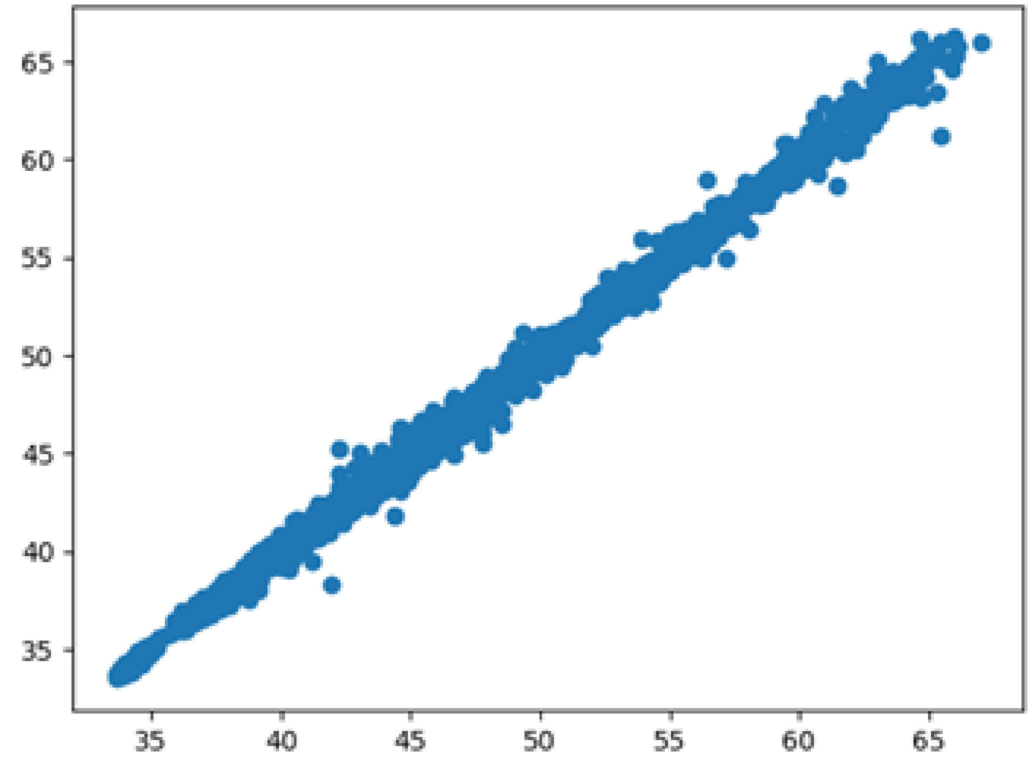


Chart A

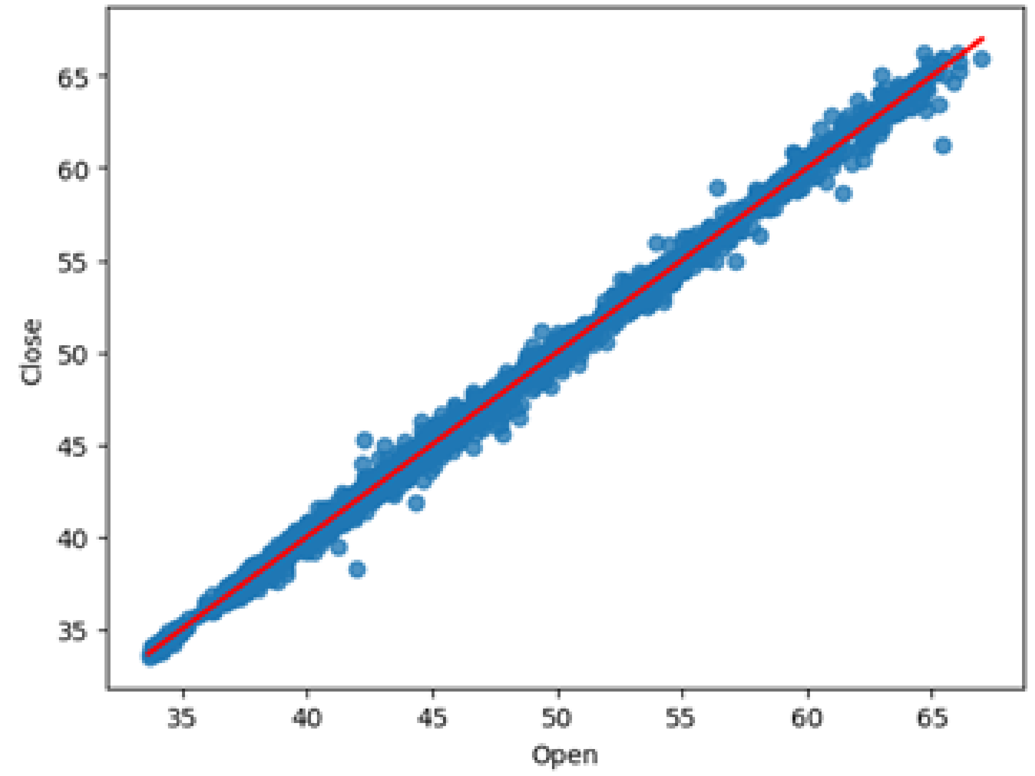


Chart B

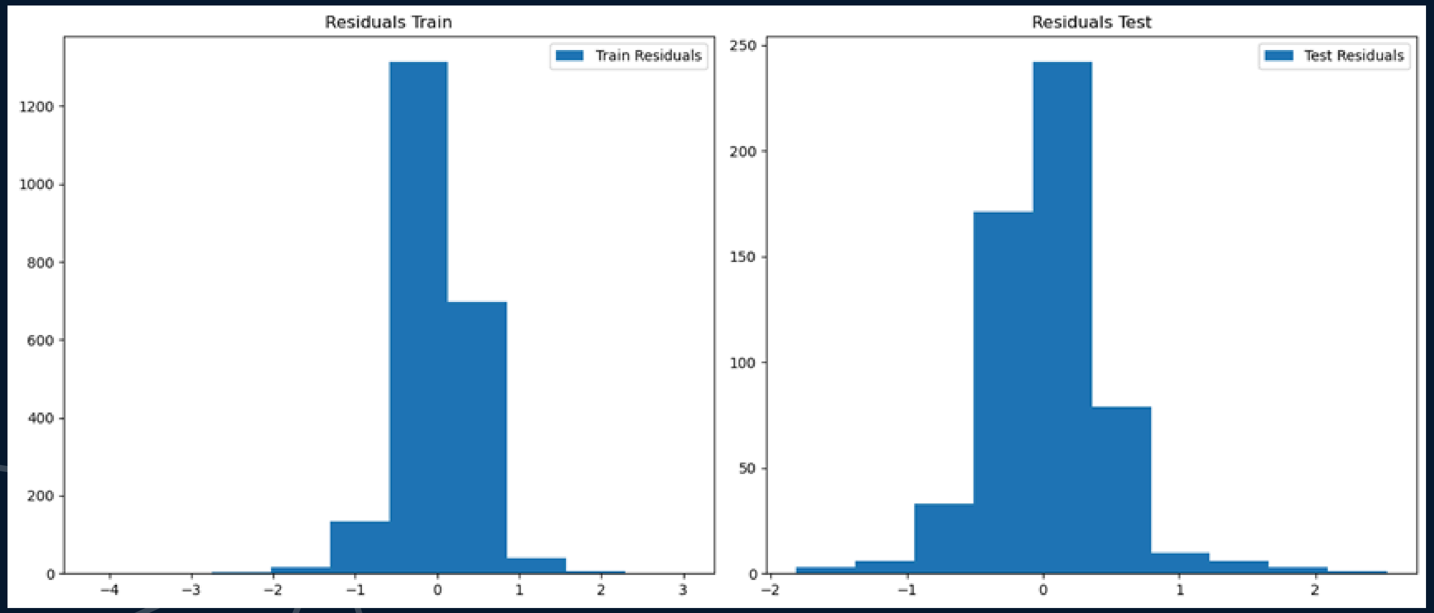
3. Make Prediction

R^2 Values:

- R^2_train: 0.9963
- R^2_test: 0.9967

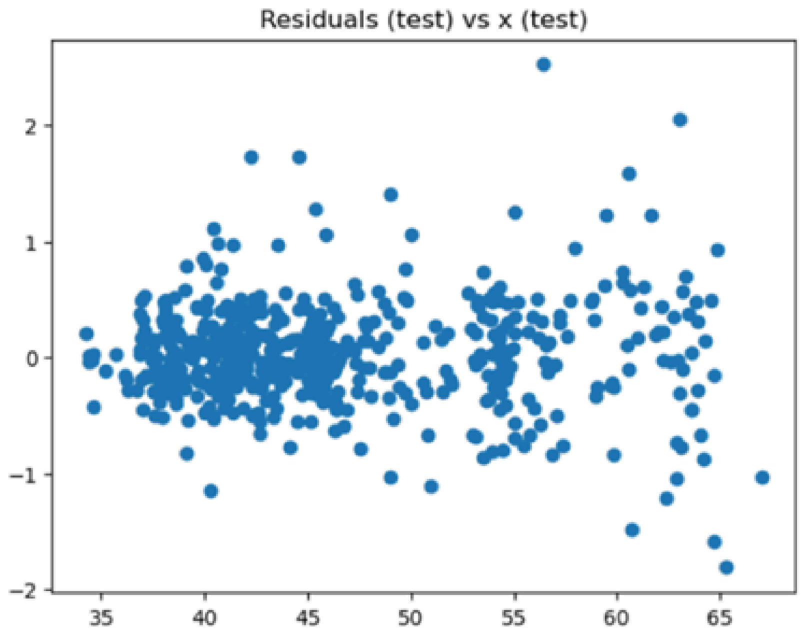
Mean Squared Error (MSE):

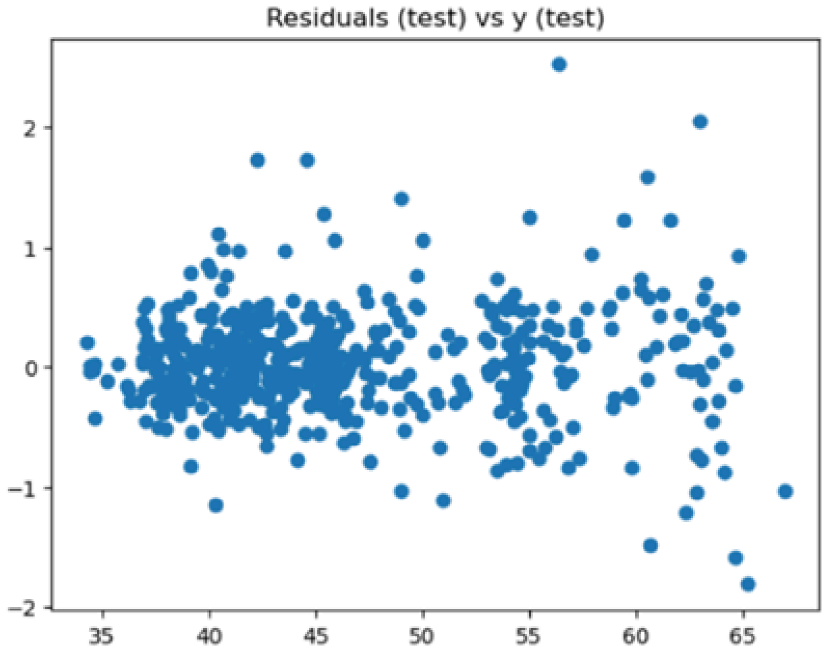
- MSE_train: 0.1949
- MSE_test: 0.1939



Residuals:

- Range of (-2, 2), indicating good model fit





Model Evaluation

Durbin-Watson Statistic: 1.91 (indicates no significant autocorrelation).

Significance:

- Open variable P-value < 0.01 shows statistical relevance.

Conclusion

The model effectively predicts Coca-Cola's closing prices, capturing 99.6% of market fluctuations with minimal error.

OLS Regression Results

Dep. Variable:	Close	**R-squared:**	0.996
Model:	OLS	**Adj. R-squared:**	0.996
Method:	Least Squares	**F-statistic:**	5.974e+05
Date:	Wed, 18 Dec 2024	**Prob (F-statistic):**	0.00
Time:	18:21:14	**Log-Likelihood:**	-1331.1
No. Observations:	2214	**AIC:**	2666.
Df Residuals:	2212	**BIC:**	2678.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0548	0.060	0.906	0.365	-0.064	0.173
Open	0.9987	0.001	772.938	0.000	0.996	1.001

Omnibus:	616.469	**Durbin-Watson:**	1.918
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	9101.514
Skew:	-0.900	**Prob(JB):**	0.00
Kurtosis:	12.768	**Cond. No.**	302.

Human Resource Dataset Analysis

I. INTRODUCTION

Objective:

Implement the classification algorithm: Logistic Regression.

Dataset Description:

Due to confidentiality in enterprises like Coca-Cola, a publicly accessible HR dataset is used to derive insights relevant to business performance.

1. Data Preprocessing

Step 1: Check data types for cleaning.

Step 2: Select "Attrition" as the target variable (predicting employee departures).

Step 3: Address binary classification issues.

Step 4: Encode Attrition to binary.

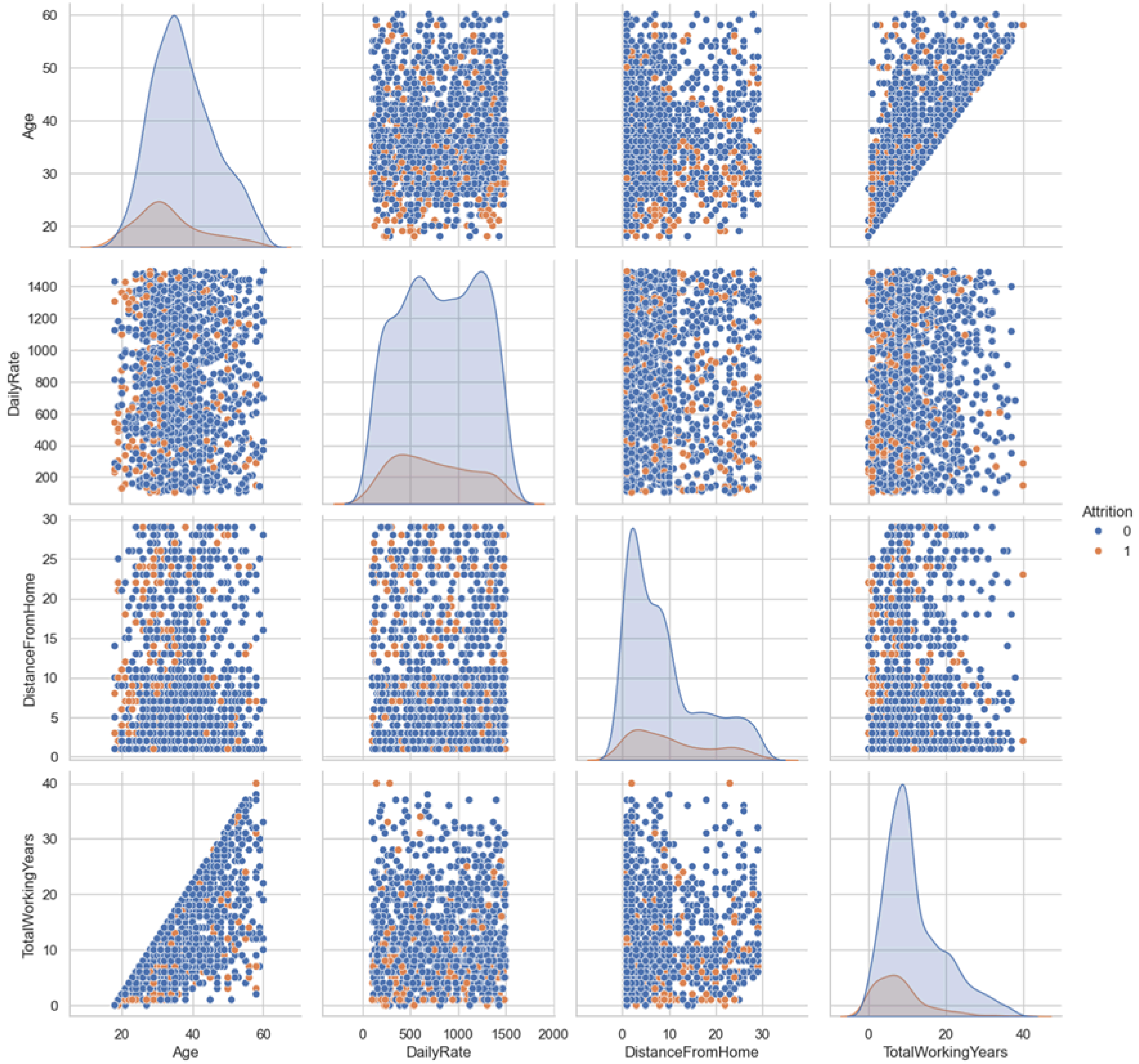
Step 5: Separate dataset into X (features) and y (target).

Step 6: Drop irrelevant or constant columns.

Step 7: Identify numerical and categorical features.

Step 8: Standardize numerical data and apply One-Hot Encoding for categorical variables.

Step 9: Visualize relationships using a pairplot.

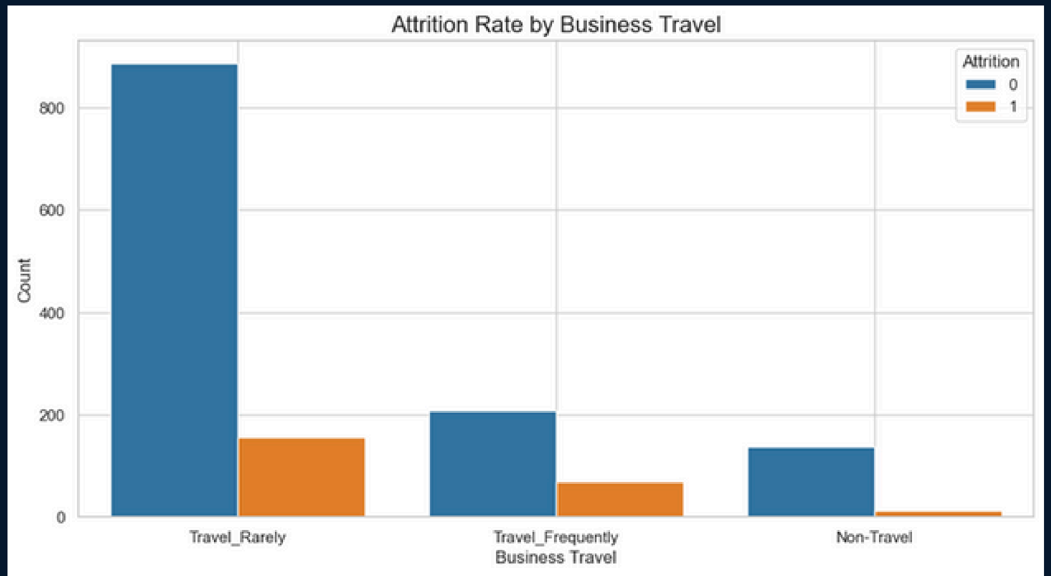


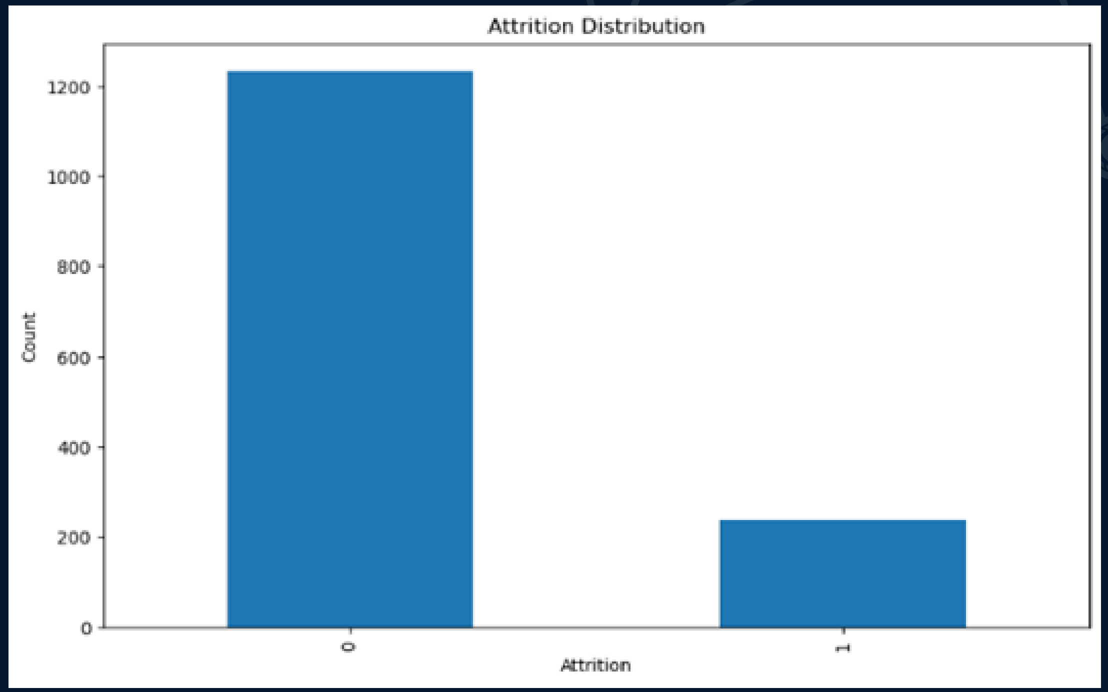
2. Building the Model

Step 1: Preprocess data and prepare to build the model.

```
attrition_counts = df['Attrition'].value_counts()
print(attrition_counts)
```

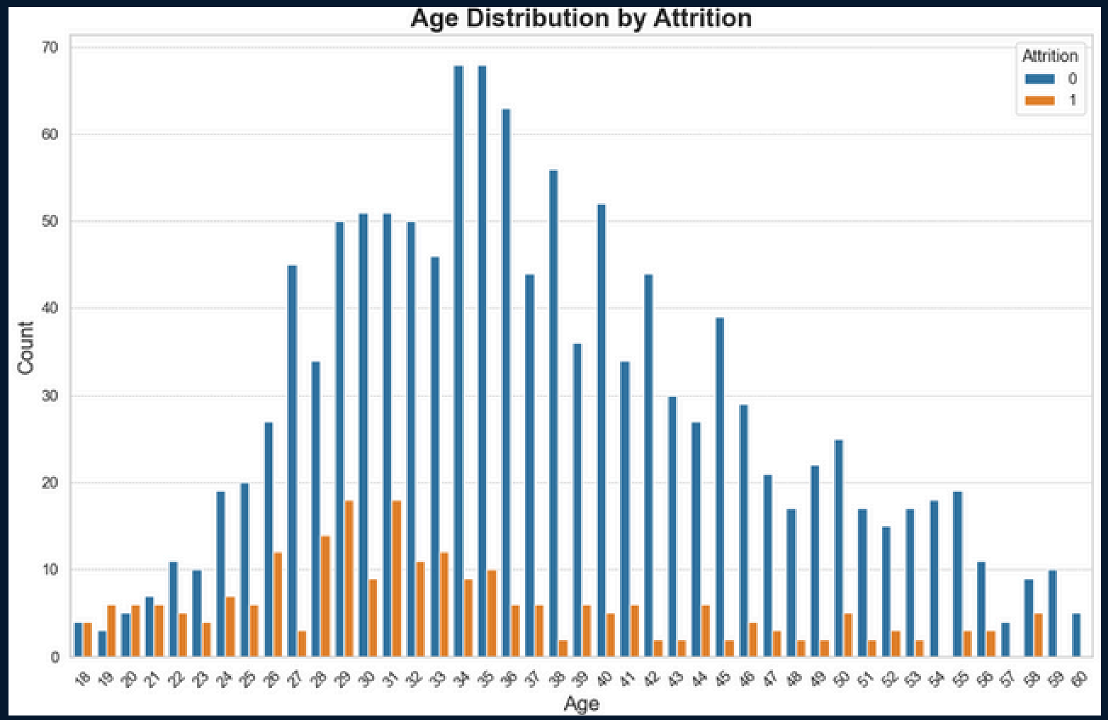
```
Attrition
0    1233
1     237
Name: count, dtype: int64
```





Step 2: Check class distribution in "Attrition":

- Imbalance detected; fewer employees left than stayed.
- Techniques to address imbalance:
  - Use stratify=y during train-test split to preserve class proportions.
  - Set class_weight='balanced' to adjust the importance of each class.



Step 3: Train-test split and build the logistic regression model.

3. Make Prediction

Step 1: Build the model and make predictions.

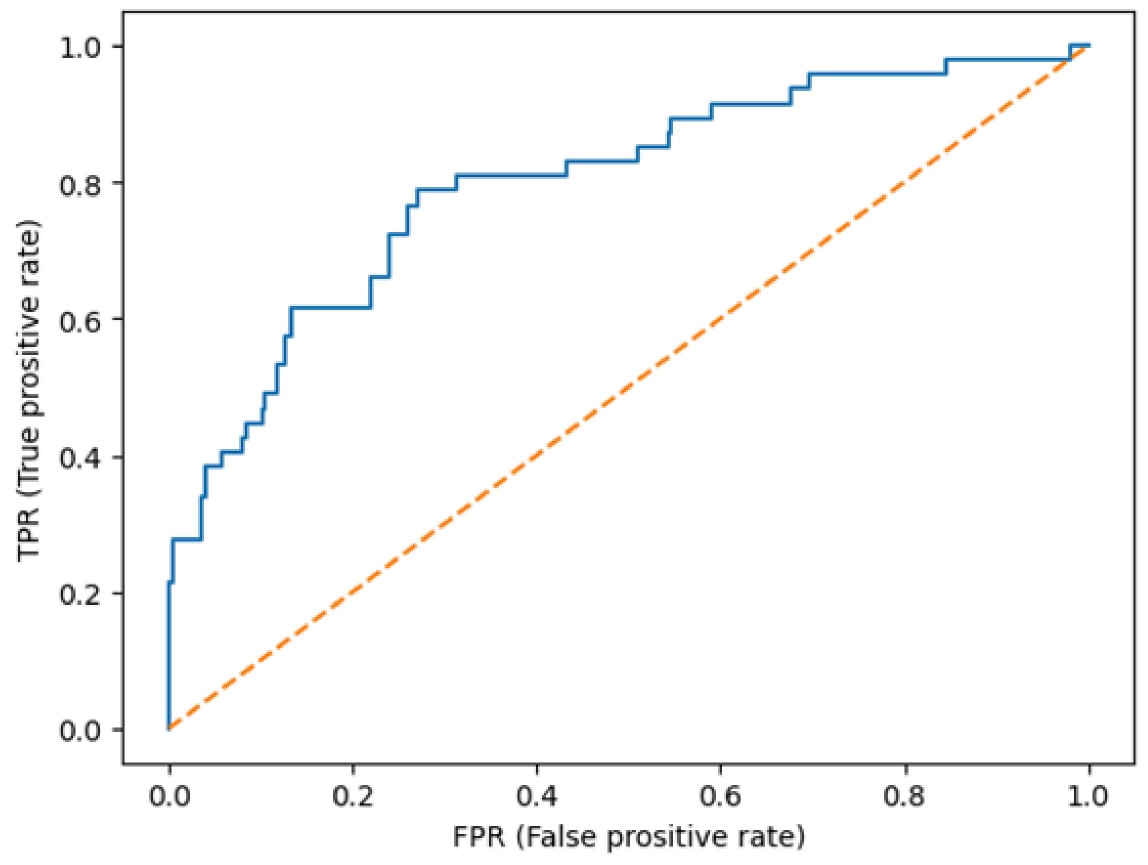
Step 2: Evaluate model accuracy:
Accuracy: 75.17%

- Confusion Matrix:
  - True Negatives (TN): 190
  - True Positives (TP): 31
  - False Positives (FP): 57
  - False Negatives (FN): 16

Classification Report:

	precision	recall	f1-score	support
0	0.92	0.77	0.84	247
1	0.35	0.66	0.46	47
accuracy			0.75	294
macro avg	0.64	0.71	0.65	294
weighted avg	0.83	0.75	0.78	294

- **AUC Value:** 0.7987 indicates good performance in distinguishing classes.
- **Insights for HR:** Utilize ROC curve analysis to enhance engagement strategies and reduce turnover.



SUMMARY OF KEY FINDINGS

HR Analysis:

- **Model Performance:** Logistic Regression achieved an accuracy of 75.17% and an AUC of 0.7987.
- **Significant Predictors:** Key factors influencing employee attrition include age and travel frequency.
- **Strategic Insights:** Highlights the importance of employee engagement and tailored retention strategies for sustaining business performance.

Stock Analysis:

- **Correlation:** Strong correlation of 0.9987 between opening and closing stock prices.
- **R-squared Value:** R-squared of 0.996, indicating that 99.6% of price fluctuations can be explained.
- **Financial Metrics:** Stock-related metrics effectively reflect Coca-Cola's market performance and inform strategic planning.

Overall Conclusion:

- **Data-Driven Strategies:** Emphasizes the role of financial ratios and human factors in guiding HR management and financial decision-making.
- **Sustained Competitiveness:** Underlines the necessity of leveraging data insights for continued success and competitiveness in the market.

THANKS FOR LISTENING