GRAPHICAL DATA REPRESENTATION IN BANKRUPTCY ANALYSIS

21BCE7192 Karthika
21BCE7727 Gyanada

Objectives:

- Introduction to bankruptcy and bankruptcy analysis.
- Importance of leveraging graphical data representation in bankruptcy analysis.
- Exploring various types of graphs, charts, and visualizations that can be employed to analyze financial data, identify trends, and predict potential bankruptcies.
- Visualization tools used for bankruptcy analysis.
- Case study and EDA.
- Benefits and challenges.

INTRODUCTION

BANKRUPTCY

- Company bankruptcy occurs when a company cannot pay its debts and obligations to creditors, resulting in the company's assets being liquidated to repay those debts.
- This can lead to the company ceasing operations and potentially going out of business.

Why it is important to study bankruptcy?

- Studying bankruptcy is important because it is a significant economic event that affects individuals, businesses, and society as a whole.
- Understanding bankruptcy can help make informed financial decisions and promote economic stability.

INTRODUCTION

BANKRUPTCY ANALYSIS

- Bankruptcy analysis involves evaluating the financial condition and performance of individuals, businesses, or organizations to determine their likelihood of facing insolvency or bankruptcy.
- This analysis typically includes assessing factors such as cash flow, debt levels, profitability, asset quality, and overall financial health.
- The goal is to identify early warning signs of financial distress and make informed decisions to mitigate risks or take appropriate actions.
- Bankruptcy analysis helps stakeholders, including creditors, investors, and regulators, understand the financial stability and viability of entities guiding their investment, lending, or regulatory decisions.

Importance of Graphical Data Representation in Bankruptcy analysis

Graphical data representation in bankruptcy analysis is crucial for :

- simplifying complex financial data
- enabling efficient data exploration
- identifying trends
- facilitating communication of insights.
- enhancing risk assessment.
- supporting predictive analytics in bankruptcy analysis

TYPES OF GRAPHS AND TOOLS USED:

- Python:
1. Pandas
2. NumPy
3. Matplotlib
4. Seaborn
5. Plotly
- Jupyter Notebook
- Excel

- Line chart
- Barchart
- Scatterplot
- Stacked barchart
- Correlation heatmap
- Histogram distribution curve.

EXPLORATORY DATA ANALYSIS:

- **Steps of EDA for Bankruptcy Analysis:**
  - Data Collection
  - Data Cleaning and Preprocessing
  - Summary Statistics
  - Time-Series Analysis
  - Comparative Analysis
  - Ratio Analysis
  - Visualization
  - Insights
  - Modelling

DATA COLLECTION AND PREPROCESSING

COLLECTION

Data Collected (Case Study):

- The Taiwan Economic Journal for the years 1999 to 2009 has listed the details of company bankruptcy based on the business regulations of the Taiwan Stock Exchange.
- It has over 900 listed companies. The data includes a majority of numerical attributes that help understand the possibility of bankruptcy.
- The collected dataset has 96 attributes, we take 54 relevant attributes for our analysis.

- We can demonstrate how to pick the best features and analyze how it affects the data model.

DATA COLLECTION AND PREPROCESSING

CLEANING AND PREPROCESSING

- Loading the Dataset.
- Dropping unnecessary columns in a DataFrame.
- Changing the index of a DataFrame.
- Using .str() method to clean columns.
- Dropping null values by dropna().
- Dropping duplicates by drop_duplicates().
- Renaming columns to a more recognizable set of labels.
- Skipping unnecessary rows in a CSV file.

```
data = pd.read_csv("C:\\Users\\gyanada\\Downloads\\bankruptcy\\data.csv")
```

```
data.columns = data.columns.str.lower()
data.columns = data.columns.str.replace(' ', '_')
data.columns = data.columns.str.replace('(', '').str.replace(')', '')
data.columns = data.columns.str.strip()
print(data.head())
```

```
data.dropna(inplace=True)
print(data.head())
data = data.drop_duplicates()
```

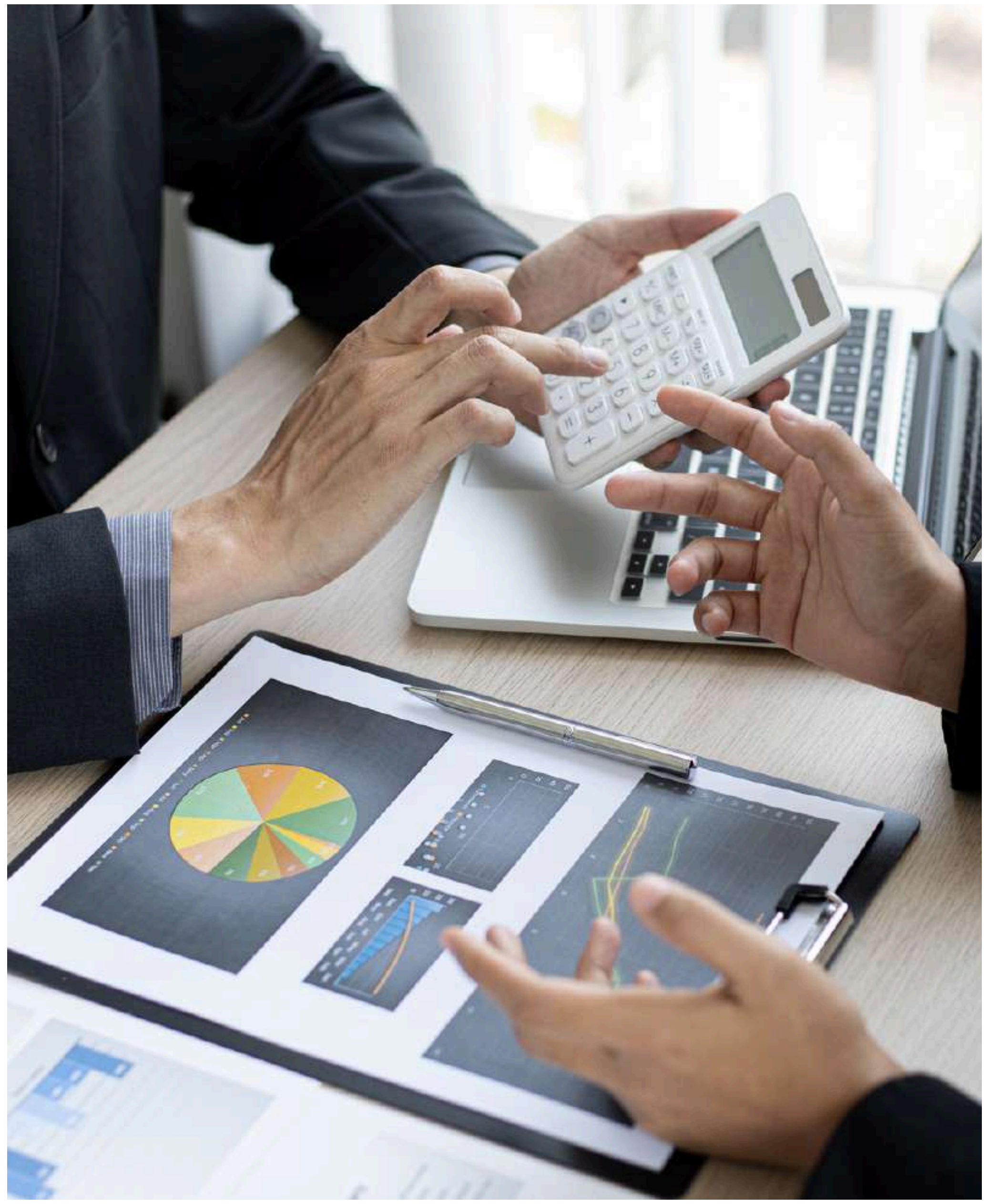
STATISTICS

SUMMARY STATISTICS

- Descriptive Statistics by df.describe()
- Correlation Analysis by df.corr()

```
In [29]: print(data.describe())

        bankrupt?  _roac_before_interest_and_depreciation_before_interest  \
count  6819.000000                                        6819.000000
mean      0.032263                                           0.505180
std       0.176710                                           0.060686
min       0.000000                                           0.000000
25%       0.000000                                           0.476527
50%       0.000000                                           0.502706
75%       0.000000                                           0.535563
max       1.000000                                           1.000000

       _roaa_before_interest_and_%_after_tax  \
count                            6819.000000
mean                                0.558625
std                                 0.065620
min                                 0.000000
25%                                 0.535543
50%                                 0.559802
75%                                 0.589157
max                                 1.000000
```

```
In [30]: print(data.corr())

                                                   bankrupt?  \
bankrupt?                                           1.000000
_roac_before_interest_and_depreciation_before_i...  -0.260807
_roaa_before_interest_and_%_after_tax               -0.282941
_roab_before_interest_and_depreciation_after_tax    -0.273051
_operating_gross_margin                             -0.100043
...                                                      ...
_liability_to_equity                                 0.166812
_degree_of_financial_leverage_dfl                    0.010508
_interest_coverage_ratio_interest_expense_to_ebit   -0.005509
_net_income_flag                                          NaN
_equity_to_liability                                -0.083048

                                                   _roac_before_interest_and_depreciation_before_interest  \
bankrupt?                                                                               -0.260807
_roac_before_interest_and_depreciation_before_i...                                       1.000000
_roaa_before_interest_and_%_after_tax                                                    0.940124
_roab_before_interest_and_depreciation_after_tax                                         0.986849
_operating_gross_margin                                                                  0.334719
```



ANALYSIS

TIME SERIES ANALYSIS

- Time series analysis is a statistical technique used to analyze patterns in time-ordered data to understand underlying patterns.
- In the context of bankruptcy prediction or financial analysis, time series analysis helps in identifying historical trends and forecasting future financial performance.

COMPARATIVE ANALYSIS

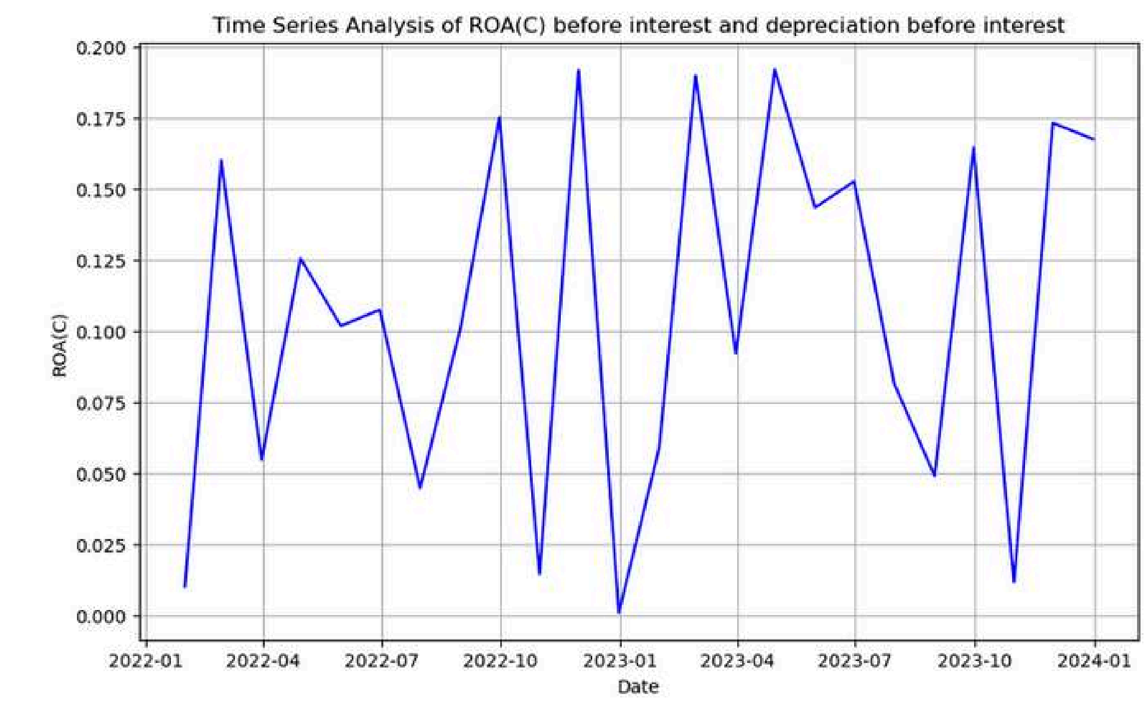
- Comparative analysis involves comparing the financial performance of a company with that of its peers, industry averages, or historical performance.
- In the context of bankruptcy by comparing key financial metrics such as profitability, liquidity, and leverage ratios, analysts can identify relative strengths and weaknesses.

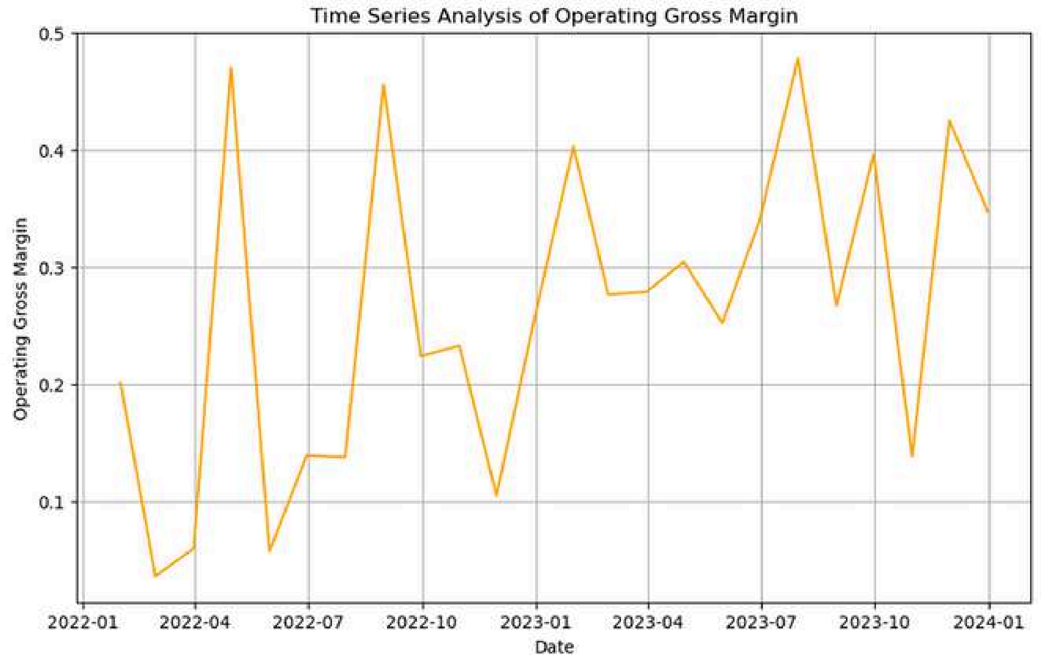
RATIO ANALYSIS

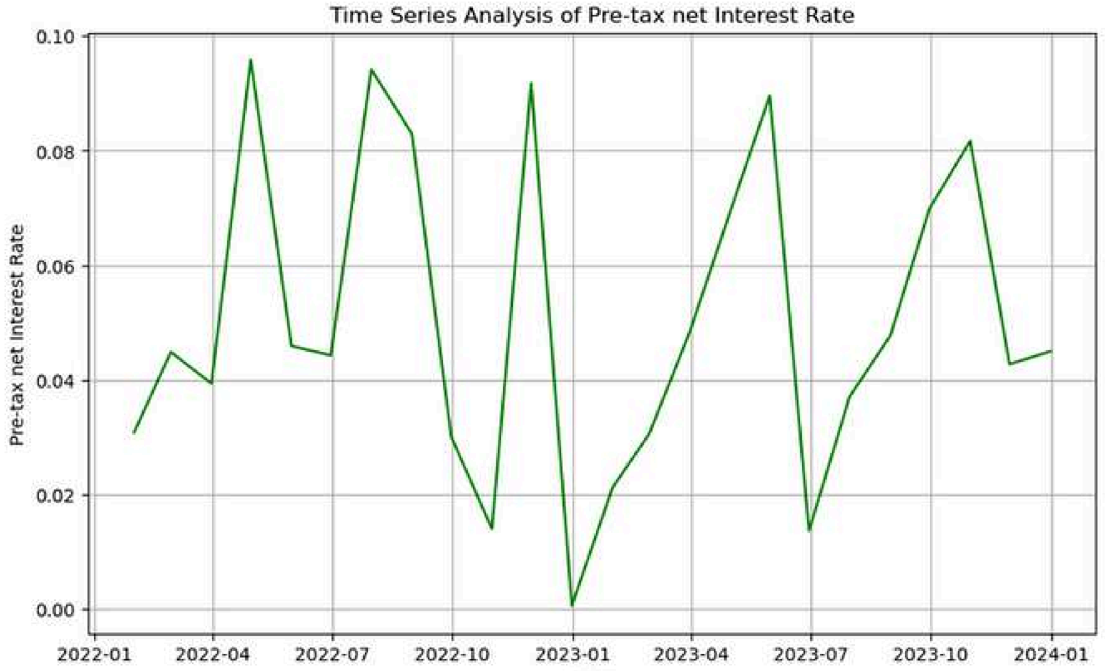
- Ratio analysis involves calculating and interpreting various financial ratios to assess a company's financial health and performance.
- By analyzing ratios such as liquidity ratios, profitability ratios and leverage ratios, analysts can evaluate the company's ability to meet its financial obligations, generate profits, and manage debt.

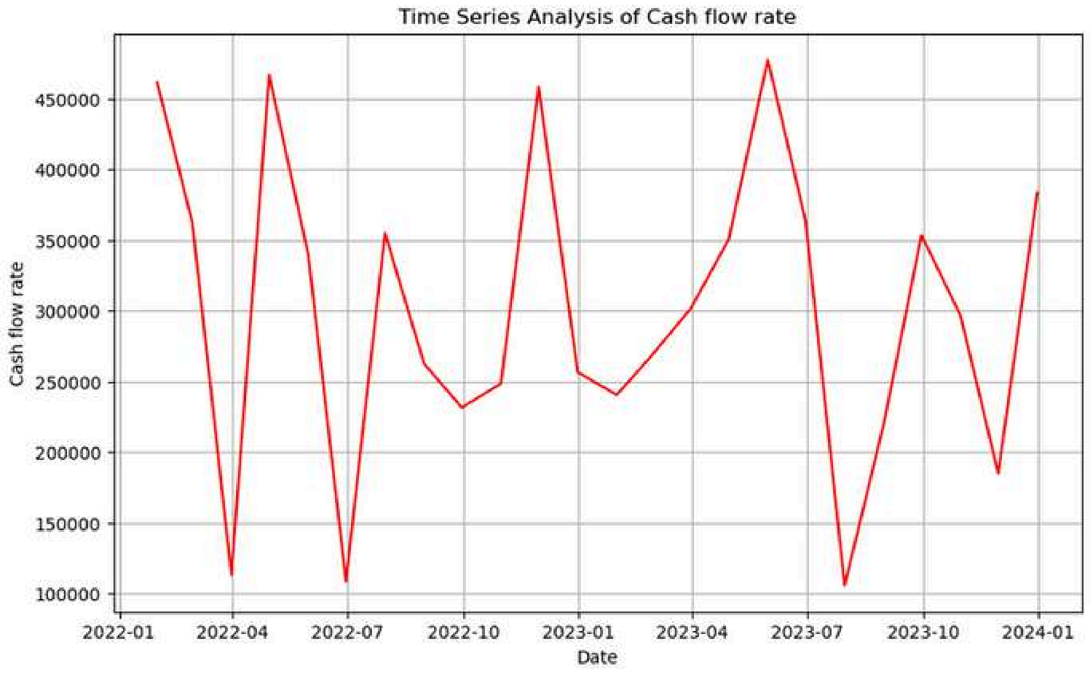
ANALYSIS

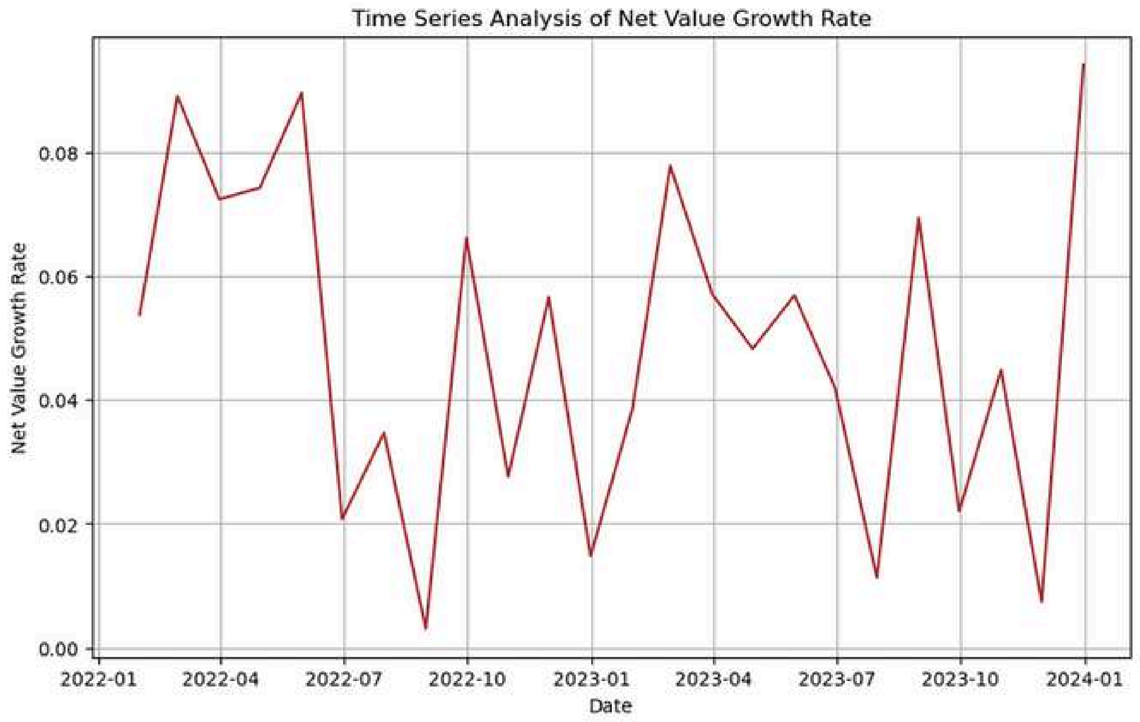
VISUALIZATIONS

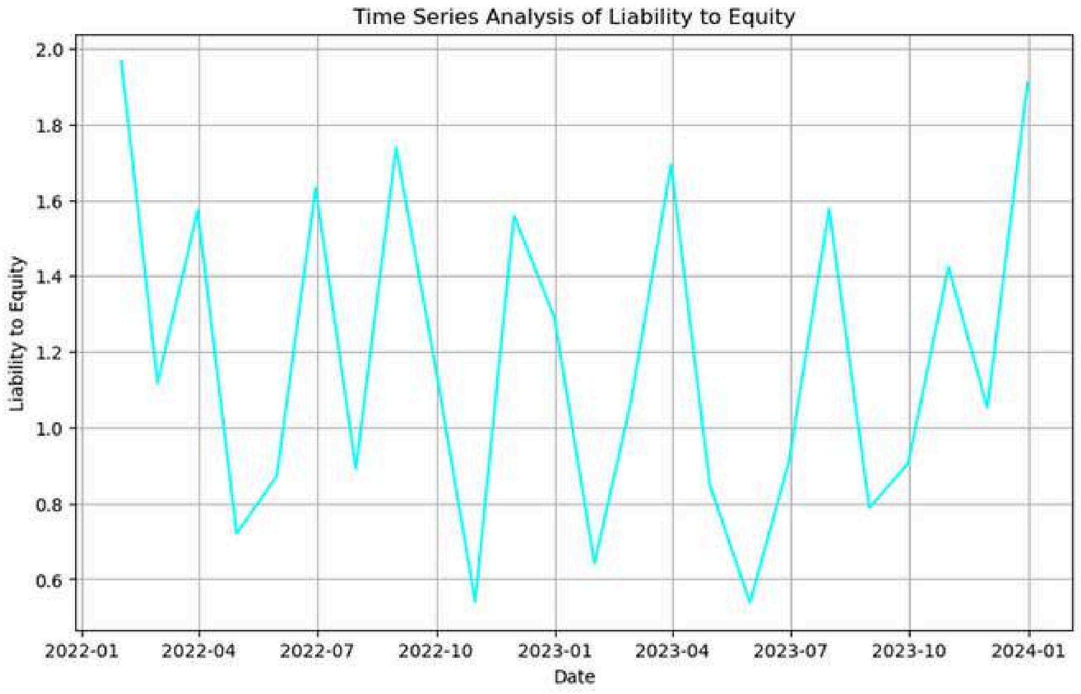




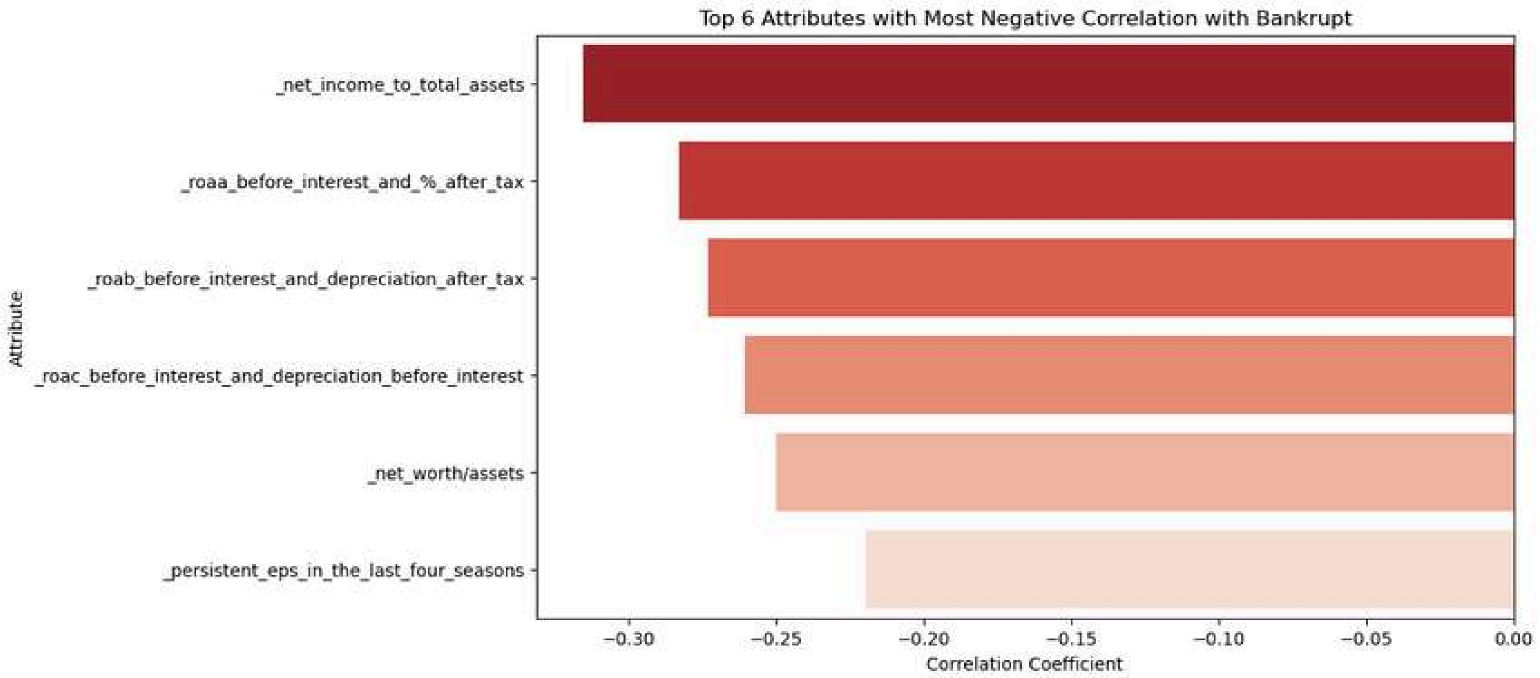


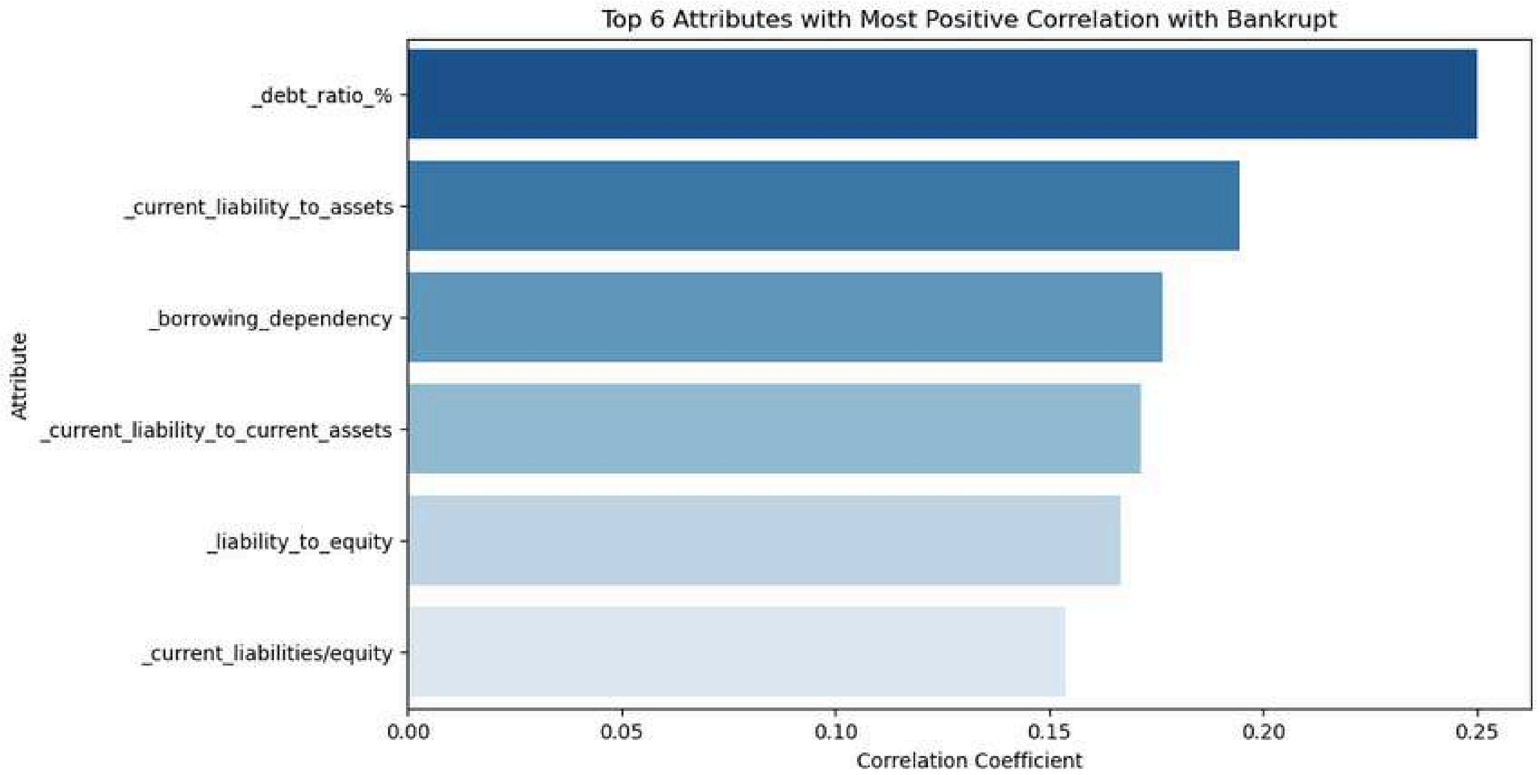




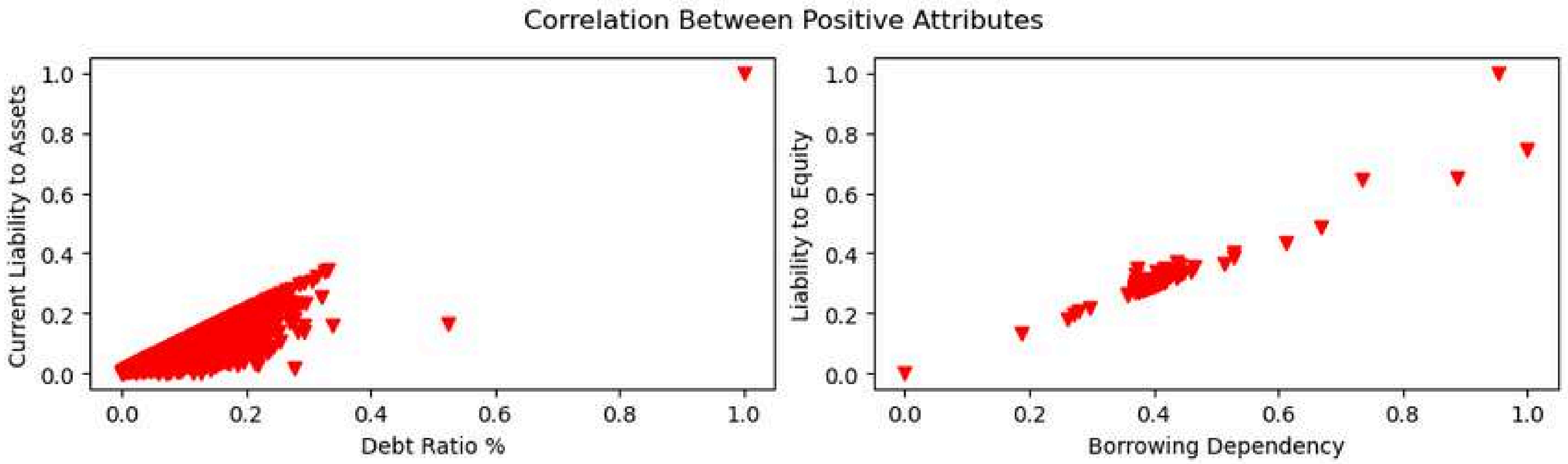


ANALYSIS

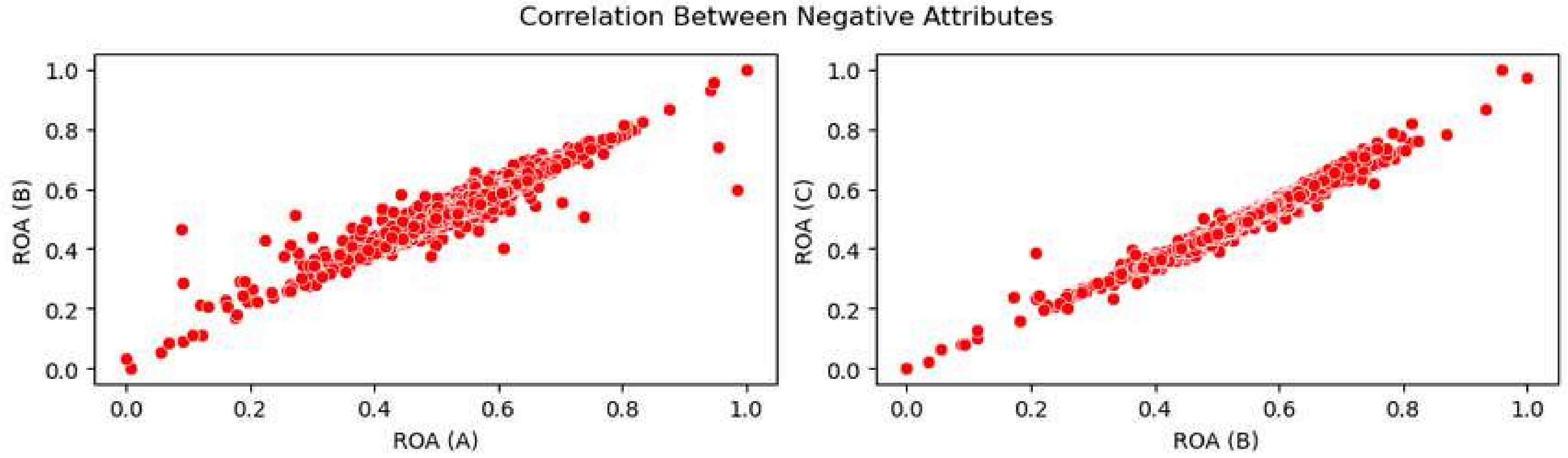




- We see that three attributes - "Debt Ratio %, Current Liability To Assets, Current Liability To Current Assets" are commonly high in bankrupt organizations.

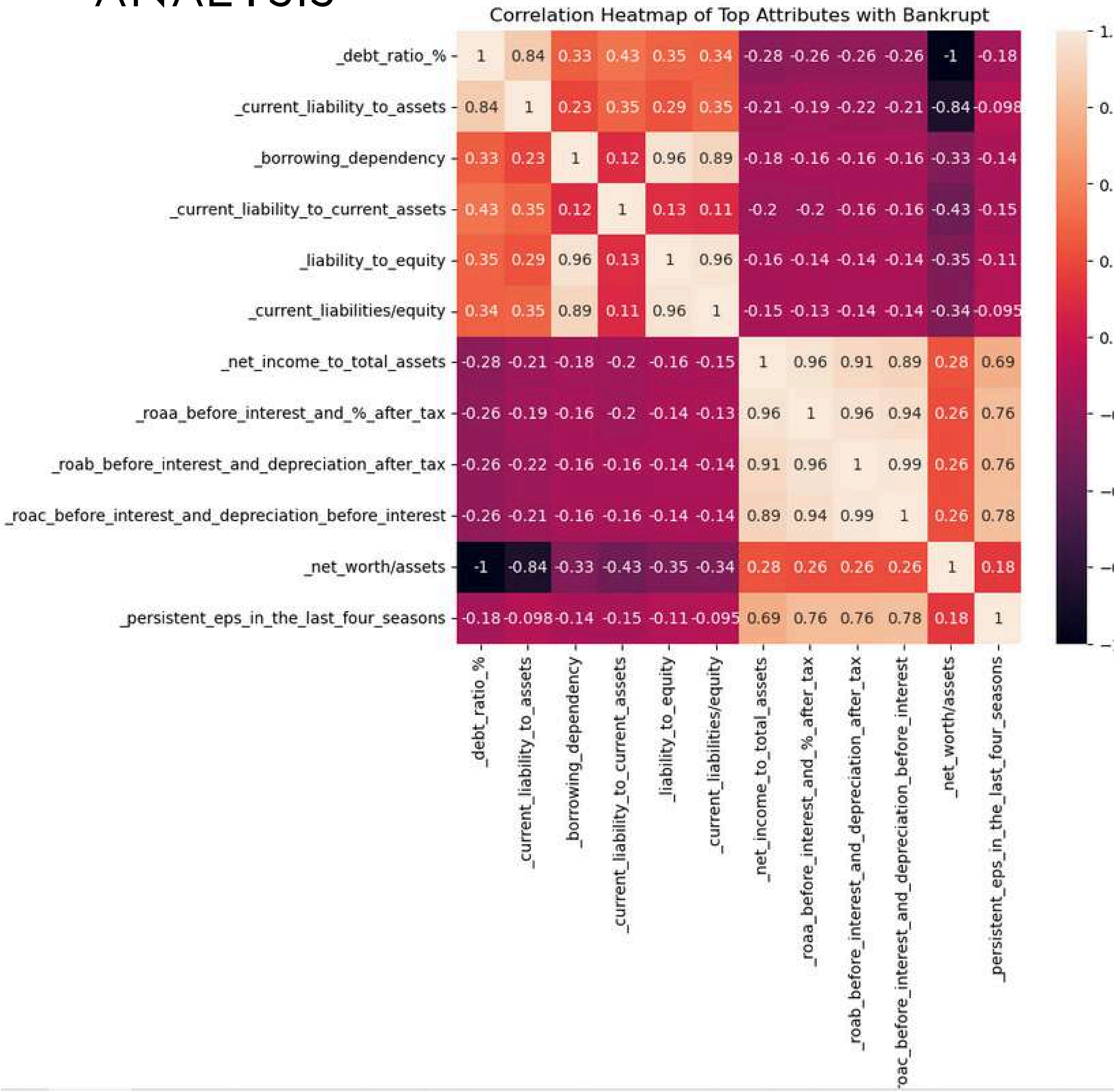


- There is a positive relation between attributes that have a high correlation with the target attribute.



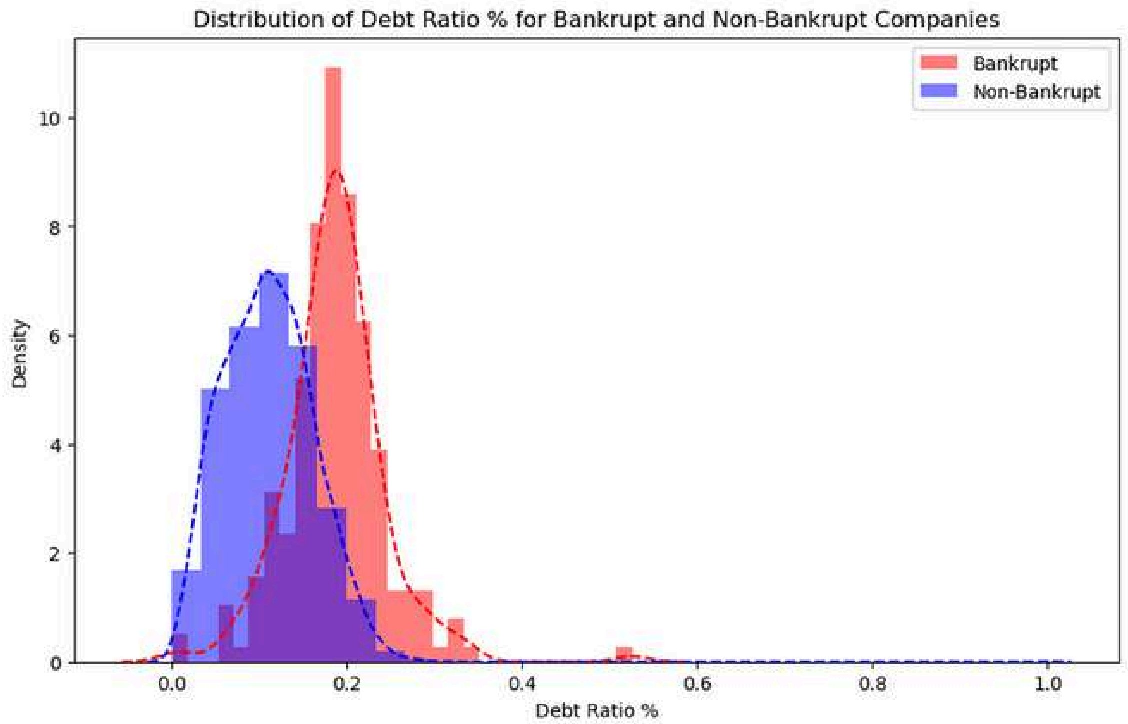
- There is a positive relation between attributes that have a low correlation with the target attribute.

ANALYSIS

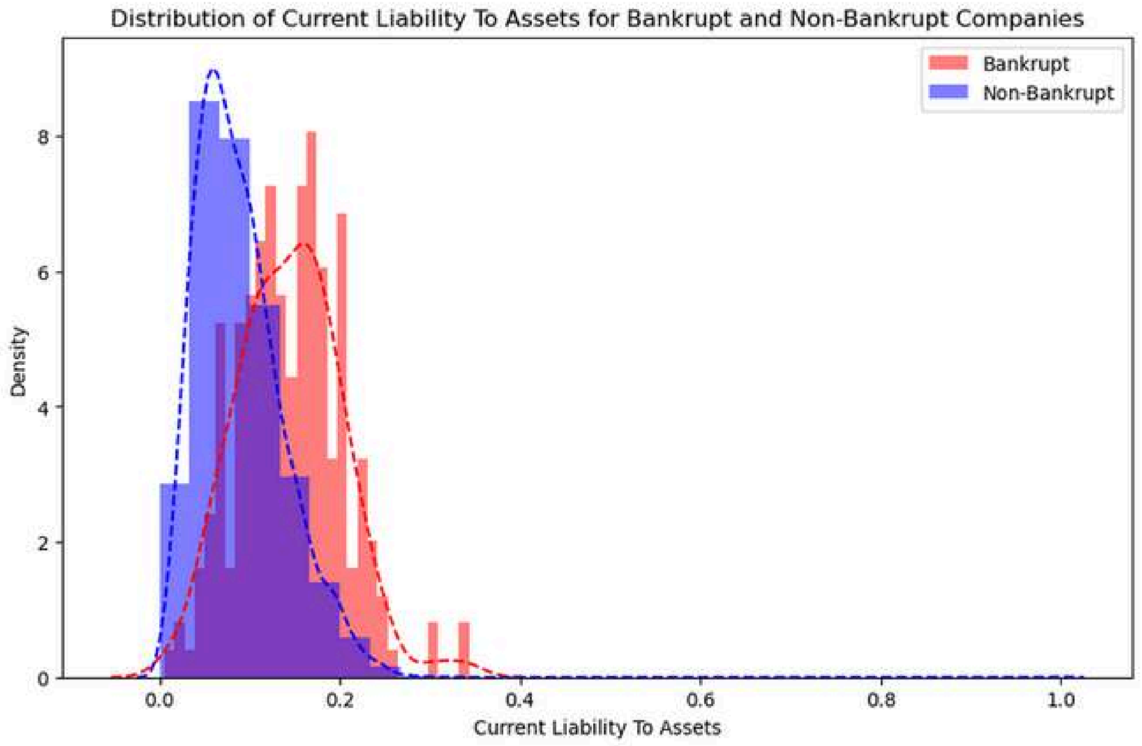


- We observed several correlations among the top 12 attributes.
- "Net Worth/Assets and Debt Ratio %" highly negatively correlated with one another.
- "Debt ratio and Current Liability Assests are highly positively correlated with one other.

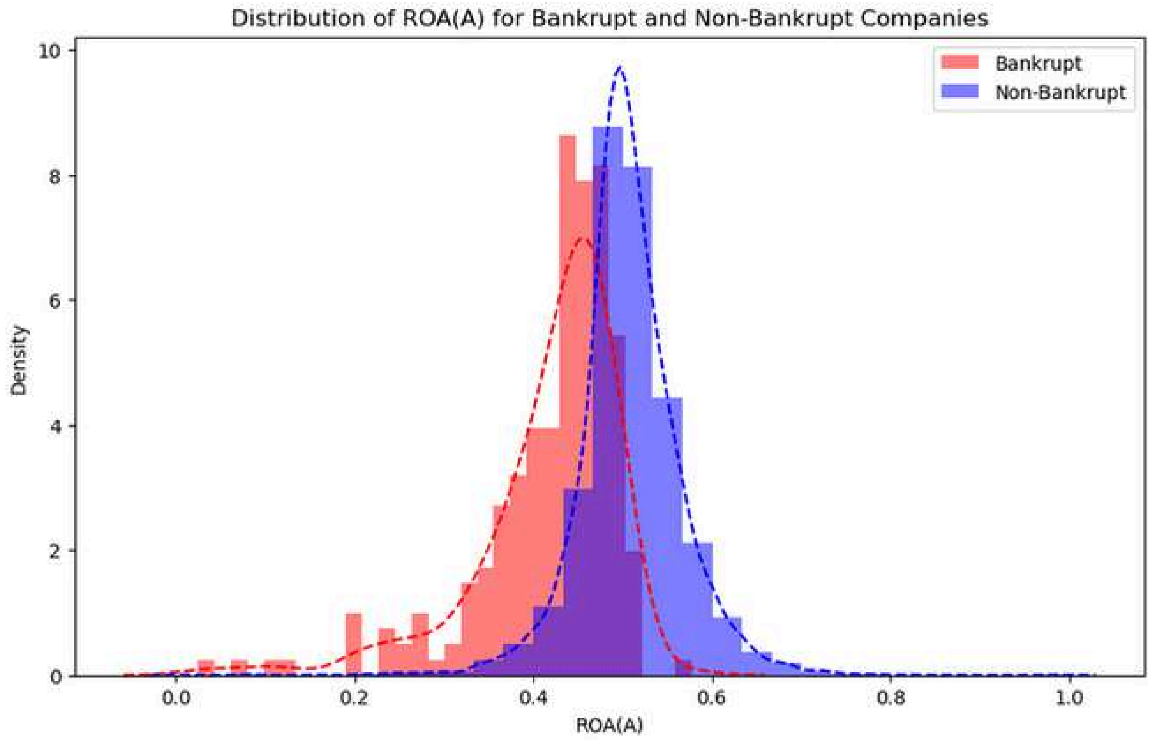
ANALYSIS



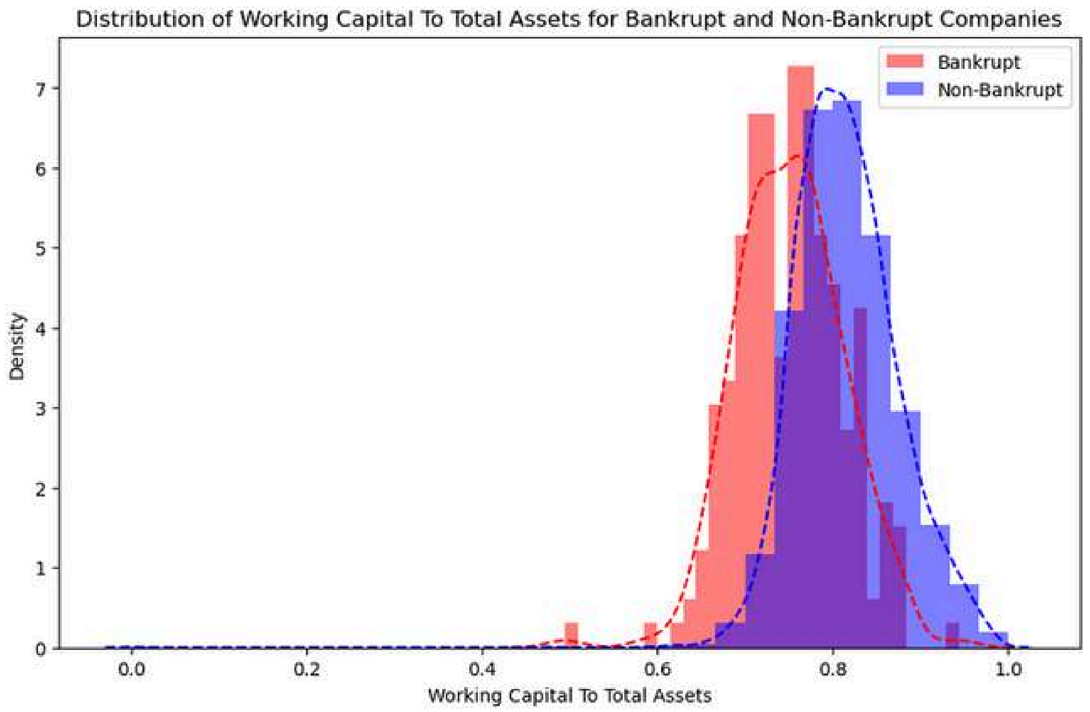
- Debt Ratio % is Liability/Total Assets and measures the extent of a company's leverage.
- Bankrupt companies tend to have higher and more right skewed debt ratios showing that they are typically more leveraged.



- Current Liability to Assets is comparable to Debt Ratio % but measures shorter tem liabilities compared to all liabilities.
- The plot again shows that bankrupt companies will typically have higher Current Liability to Assets relative to companies that didn't go bankrupt.

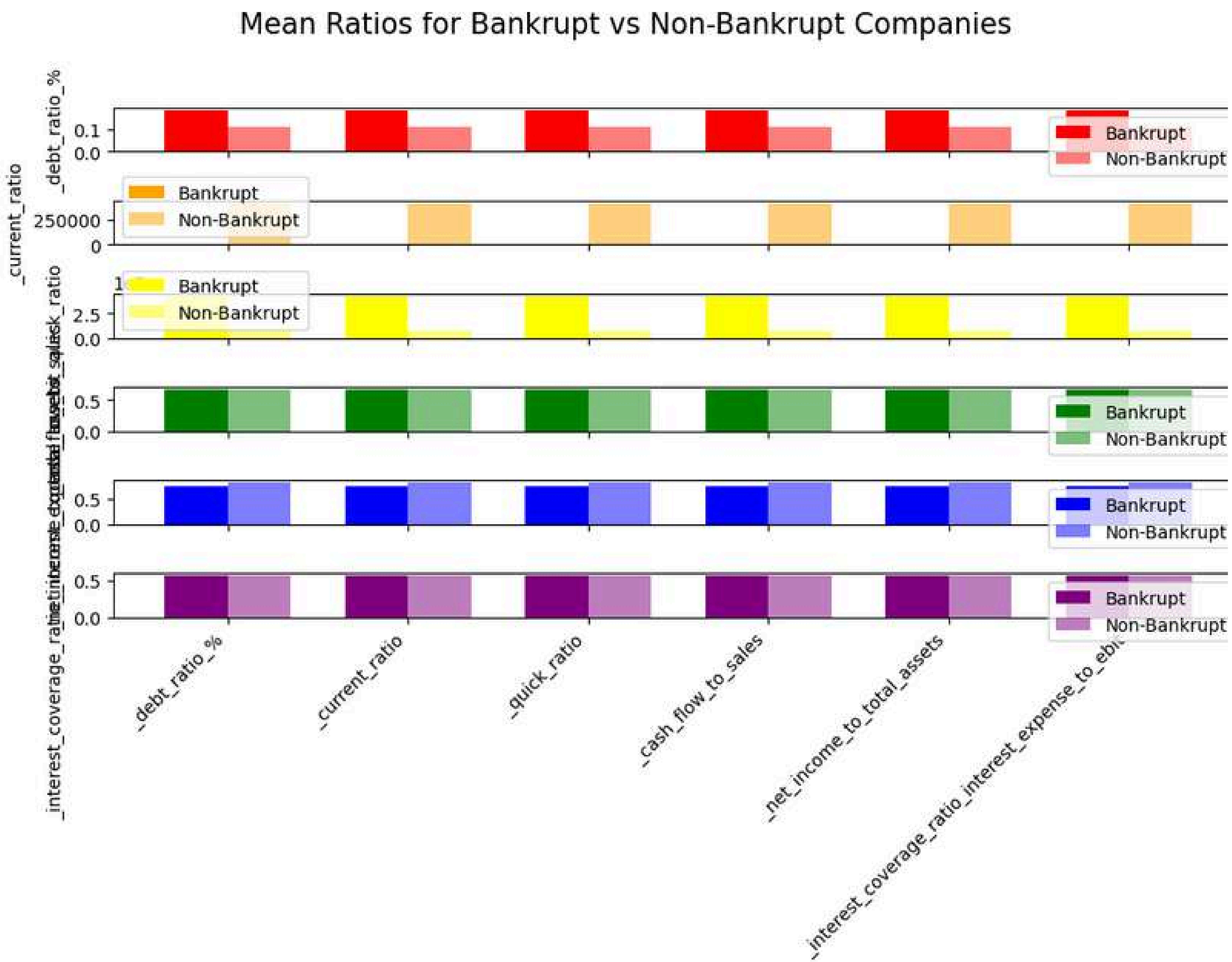


- ROA(A) is Net income before interest and % after tax/Total Assets. ROA(A) is an indicator of how profitable a company is relative to its total assets.
- We can clearly see differences in the distributions for bankrupt companies who have lower and more left skewed ROA(A) relative to not bankrupt companies.



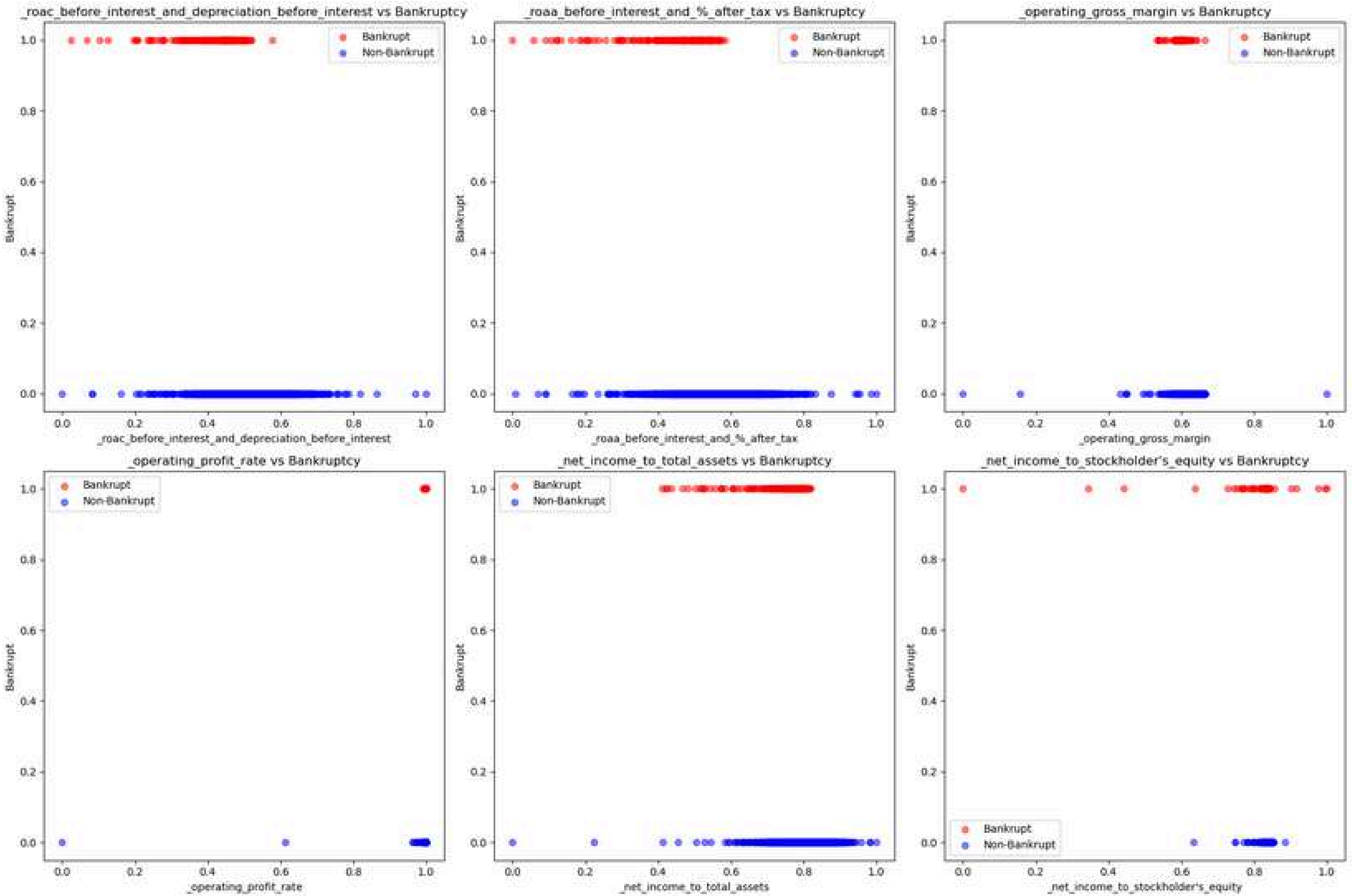
- Working Capital to Total Assets compares liquid assets to total assets.
- We can see from the distributions that bankrupt companies have smaller ratios of liquid assets to total assets.
- This may indicate that in time of financial distress, these companies would have more difficult time offloading assets to satisfy their liabilities and avoid bankruptcy.

ANALYSIS



- **Debt Ratio Analysis:**A higher debt ratio indicates that bankrupt companies have a higher proportion of debt relative to their total assets.
- **Liquidity Ratios (Current Ratio and Quick Ratio):**Bankrupt companies may have lower liquidity ratios, such as the current ratio and quick ratio, indicating difficulties in meeting short-term obligations.
- **Cash Flow to Sales Ratio:** A lower cash flow to sales ratio suggests that bankrupt companies may have insufficient cash flow relative to their sales revenue.
- **Profitability Ratios (Net Income to Total Assets Ratio):**Bankrupt companies may exhibit lower profitability ratios, such as net income to total assets ratio, indicating lower profitability relative to their total assets.
- **Interest Coverage Ratio:**A lower interest coverage ratio implies that bankrupt companies may struggle to cover their interest expenses with operating income.

ANALYSIS



_roac_before_interest_and_depreciation_before_interest vs Bankruptcy
_roaa_before_interest_and_%_after_tax vs Bankruptcy
_operating_gross_margin vs Bankruptcy
Bankrupt
Non-Bankrupt
Bankrupt
0.0
0.2
0.4
0.6
0.8
1.0
_roac_before_interest_and_depreciation_before_interest
_roaa_before_interest_and_%_after_tax
_operating_gross_margin
_operating_profit_rate vs Bankruptcy
_net_income_to_total_assets vs Bankruptcy
_net_income_to_stockholder's_equity vs Bankruptcy
_operating_profit_rate
_net_income_to_total_assets
_net_income_to_stockholder's_equity

ANALYSIS

- **ROAC (Return on Average Capital) Before Interest and Depreciation Before Interest:**
  - Bankrupt Companies: The scatter plot suggests that bankrupt companies tend to have lower ROAC values compared to non-bankrupt companies.
- **ROAA (Return on Average Assets) Before Interest and % After Tax:**
  - Bankrupt Companies: Bankrupt companies typically exhibit lower ROAA values, implying less efficient utilization of assets to generate profits.
- **Operating Gross Margin:**
  - Bankrupt Companies: The scatter plot shows that bankrupt companies generally have lower operating gross margins, indicating lower profitability on each unit of sales.
- **Operating Profit Rate:**
  - Bankrupt Companies: Bankrupt companies tend to have lower operating profit rates, indicating lower profitability from core business operations. This suggests that bankrupt companies may face challenges in generating profits from their primary business activities.
- **Net Income to Total Assets:**
  - Bankrupt Companies: The scatter plot suggests that bankrupt companies have lower ratios of net income to total assets, indicating lower profitability relative to their total assets. This implies that bankrupt companies may struggle to generate sufficient profits compared to their asset base.

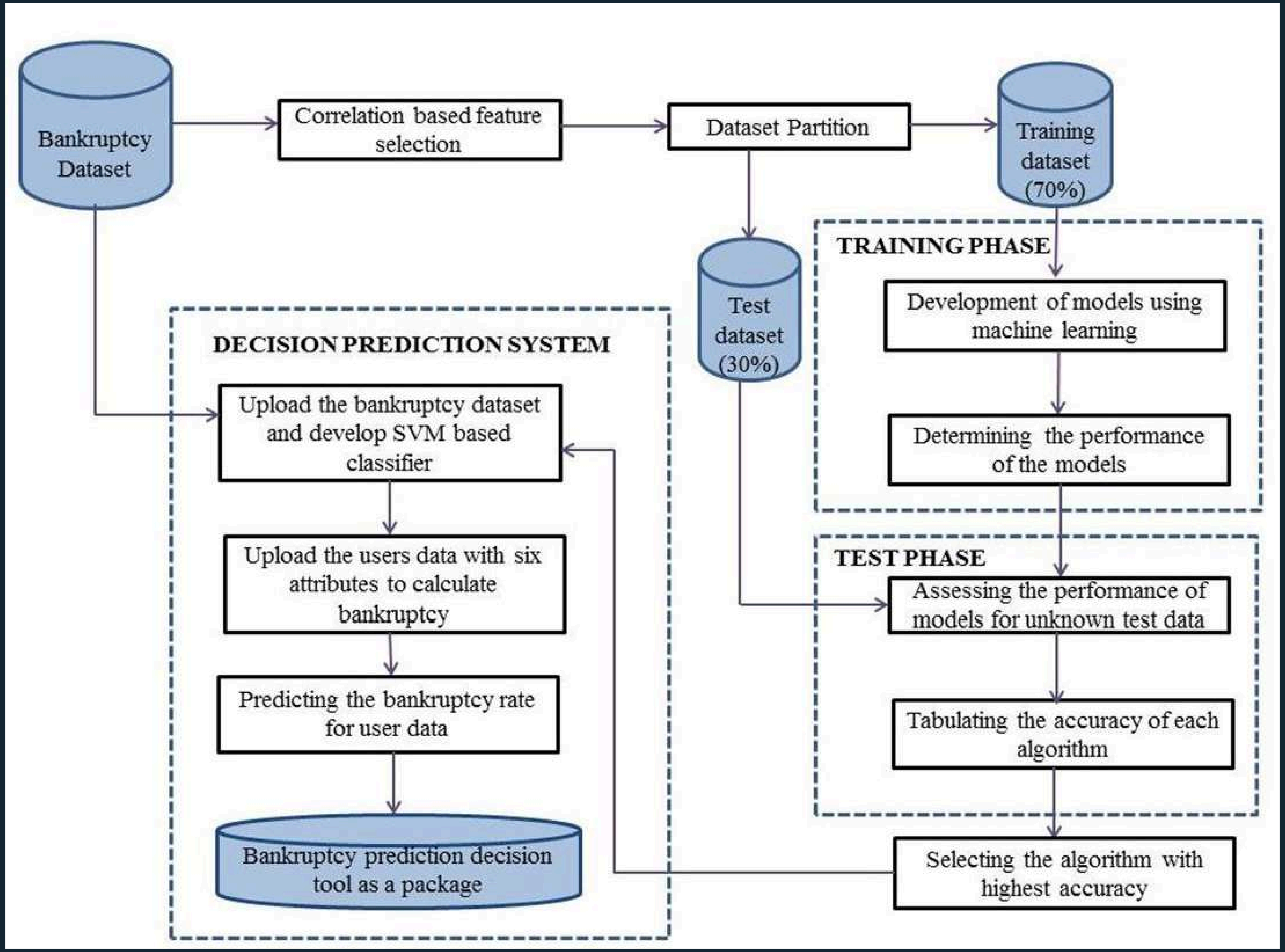
VALUABLE INSIGHTS:

- The number of organizations that have gone bankrupt in 10 years between 1999 – 2000 is few.
- Several companies possess many assets, which is always a good sign for an organization.
- An organization cannot guarantee not being bankrupt, although owning several assets.
- The organizations in the dataset are running into losses for the past two years as their net income poses to be negative.
- Very few of the organizations that have had negative income in the past two years suffer from bankruptcy.
- It is observed that "Debt Ratio %, Current Liability To Assets, Current Liability To Current Assets" attributes are a few of the attributes that have a high correlation with the target attribute.
- An increase in the values of the attributes "Debt Ratio %, Current Liability To Assets, Current Liability To Current Assets" causes an organization to suffer heavy losses, thus resulting in bankruptcy.
- An increase in the values of the attributes that have a negative correlation with the target attribute helps an organization avoid bankruptcy.
- There seems to be a relation between attributes that have a high correlation with the target attribute and a low correlation with the target attribute.
- We observed several correlations among the top 12 attributes, one of which being "Net Worth/Assets and Debt Ratio %" that is negatively correlated with one another.

MODELLING:

Creating a predictive model for brankruptcy involves several steps:

- Split the dataset into training, validation, and test sets to evaluate the performance of the model.
- Choose the appropriate algorithm(s) based on the problem type (e.g., classification, regression), data characteristics, and performance requirements.
- Train the selected model using the training data.
- Validate the final model on the test set.
- Deploy the trained model into production or use it to make predictions on new data.



```python
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, classification_report, confusion_matrix
import pandas as pd

# Define features (attributes) and target variable
features = ['_roac_before_interest_and_depreciation_before_interest',
            '_roaa_before_interest_and_%_after_tax',
            '_operating_gross_margin',
            '_operating_profit_rate',
            '_net_income_to_total_assets',
            "_net_income_to_stockholder's_equity",
            '_total_asset_turnover',
            '_accounts_receivable_turnover',
            '_inventory_turnover_rate_times',
            '_cash_turnover_rate',
            '_debt_ratio_%',
            '_current_liability_to_assets',
            '_current_liability_to_current_assets']

target = 'bankrupt?'

# Split data into training and testing sets
X = data[features]
y = data[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize different models
models = {
    'Logistic Regression': LogisticRegression(random_state=42),
    'Random Forest': RandomForestClassifier(random_state=42),
    'SVM': SVC(random_state=42),
    'Gradient Boosting': GradientBoostingClassifier(random_state=42),
    'kNN': KNeighborsClassifier(),
    'Decision Tree': DecisionTreeClassifier(random_state=42)
}

# Train and evaluate each model
results = {'Model': [], 'Accuracy': [], 'Precision (Bankrupt)': [], 'Precision (Non-Bankrupt)': []}

for name, model in models.items():
    print(f"Training {name}...")
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    # Evaluate the model
    accuracy = accuracy_score(y_test, y_pred)
    precision_bankrupt = precision_score(y_test, y_pred, pos_label=1)
    precision_non_bankrupt = precision_score(y_test, y_pred, pos_label=0)

    # Append results to the dictionary
    results['Model'].append(name)
    results['Accuracy'].append(accuracy)
    results['Precision (Bankrupt)'].append(precision_bankrupt)
    results['Precision (Non-Bankrupt)'].append(precision_non_bankrupt
```

```
--------------------------------------------------------------

Results:
                 Model  Accuracy  Precision (Bankrupt)  \
0  Logistic Regression  0.962610              0.000000
1        Random Forest  0.964076              0.571429
2                  SVM  0.962610              0.000000
3    Gradient Boosting  0.962610              0.500000
4                  kNN  0.961877              0.000000
5        Decision Tree  0.947214              0.266667

   Precision (Non-Bankrupt)
0                  0.962610
1                  0.968148
2                  0.962610
3                  0.968796
4                  0.962583
5                  0.970432
```

- Choosing random forest as best model.

- Choosing an input from the dataset(Brankrupt company) for Prediction Model:

In [114]:
```
import numpy as np

# Given values from the dataset
given_values = [0.456101009, 0.52071522, 0.510894588, 0.597630407, 0.597817783, 0.998937659, 0.797333017,
                0.80927091, 0.303491413, 0.781535588, 0.455702339, 0.173528296, 0.173528296]

# Selected features used in the model
selected_features = ['_roac_before_interest_and_depreciation_before_interest',
                     '_roaa_before_interest_and_%_after_tax',
                     '_operating_gross_margin',
                     '_operating_profit_rate',
                     '_net_income_to_total_assets',
                     "_net_income_to_stockholder's_equity",
                     '_total_asset_turnover',
                     '_accounts_receivable_turnover',
                     '_inventory_turnover_rate_times',
                     '_cash_turnover_rate',
                     '_debt_ratio_%',
                     '_current_liability_to_assets',
                     '_current_liability_to_current_assets']

# Create a dictionary with selected features and their corresponding values
input_data = {feature: value for feature, value in zip(selected_features, given_values)}

# Reshape the values into a 2D array
input_data = np.array([[input_data[feature] for feature in selected_features]])

# Make prediction
prediction = model.predict(input_data)

# Print prediction
if prediction[0] == 1:
    print("The company is predicted to go bankrupt.")
else:
    print("The company is predicted to stay solvent.")
```

```
The company is predicted to go bankrupt.
```

```
C:\Users\gyanada\anaconda3\Lib\site-packages\sklearn\base.py:464: UserWarning: X does not have valid feature names, but Ra
ndomForestClassifier was fitted with feature names
  warnings.warn(
```

BENEFITS

- **Early Warning System:** Bankruptcy analysis can serve as an early warning system for identifying companies at risk of financial distress.
- **Decision Making:** It helps stakeholders, including investors, creditors, and managers, make informed decisions.
- **Risk Management:** By understanding the factors contributing to bankruptcy risk, organizations can develop risk management strategies to safeguard their businesses.
- **Financial Health Assessment:** Bankruptcy analysis provides insights into a company's financial health.
- **Competitive Benchmarking:** Comparative analysis allows companies to benchmark their financial performance against competitors.

CHALLENGES

- **Data Quality:** Ensuring the accuracy, completeness, and reliability of financial data is a significant challenge.
- **Complexity**: Bankruptcy analysis involves analyzing multiple financial metrics and indicators which can be complex.
- **Data Availability:** Availability of timely and relevant financial data can be a challenge, especially for private companies.
- **Modeling Uncertainty:** Predictive models used for bankruptcy analysis may suffer from uncertainty and limitations, leading to potential inaccuracies.
- **Regulatory Changes:** Changes in accounting standards, regulations, and economic policies can impact bankruptcy analysis methodologies.

FUTURE SCOPE

- **Integration of Alternative Data Sources:** Using non-traditional data like satellite imagery or social media sentiment alongside financial data can provide new insights into a company's financial health.
- **Feature Engineering and Selection:** Developing new predictive features and refining existing ones to improve the accuracy of bankruptcy prediction models.
- **Ensemble and Hybrid Models:** Combining multiple predictive models to leverage their collective intelligence, or integrating machine learning with traditional methods for more robust predictions.
- **Explainable AI (XAI):** Making complex models more interpretable and transparent to understand the factors driving predictions.
- **Dynamic and Real-Time Monitoring:** Creating models that adapt to changing market conditions and incorporate real-time data for more timely risk assessments.
- **Application of Natural Language Processing (NLP):** Analyzing textual data to extract insights from financial reports, news articles, and regulatory filings.
- **Industry-Specific Models:** Tailoring bankruptcy prediction models to specific industries to better capture sector-specific dynamics and challenges.

CONCLUSIONS

- **Feature Importance:** Certain financial metrics such as Debt Ratio %, Current Liability To Assets, and Current Liability To Current Assets emerged as crucial indicators of bankruptcy risk.
- **Visualization Insights:** Various types of visualizations, including scatter plots, bar charts were employed to understand the relationships between financial metrics and bankruptcy risk. These visualizations provided insights into the distribution and trends of key variables.
- **Model Performance:** Different machine learning models, including Logistic Regression, Random Forest, SVM, Gradient Boosting, kNN, and Decision Trees, were evaluated for their effectiveness in predicting bankruptcy.
- **Future Directions:** Future research in bankruptcy analysis could focus on integrating alternative data sources, refining feature engineering techniques, and developing explainable AI models.

REFERENCES

https://www.sciencedirect.com/science/article/pii/S0957417419305123

https://www.kaggle.com/code/ginelledsouza/bankruptcy-analysis

THANK YOU!