# PHASE 3 PROJECT

BY

Simon Muema

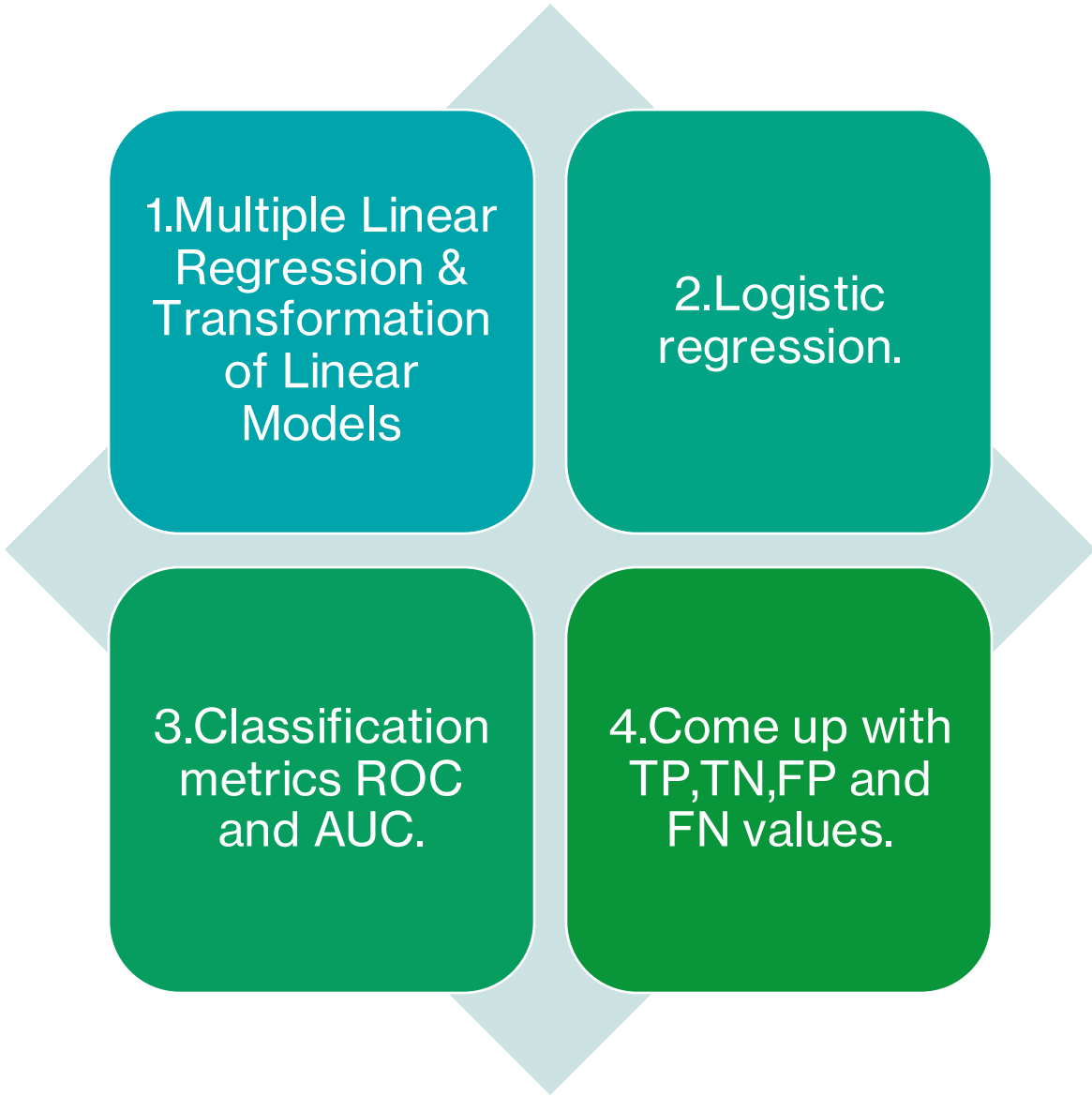
# PROBLEM STATEMENT.

- To provide actionable investment insights by analysing the most profitable companies arranging them by financial metrics, employee count, geographical location, profitability analysis, market value and employment analysis to offer specific investment recommendations and visual representations to guide investors in making informed decisions on which companies to choose when investing.

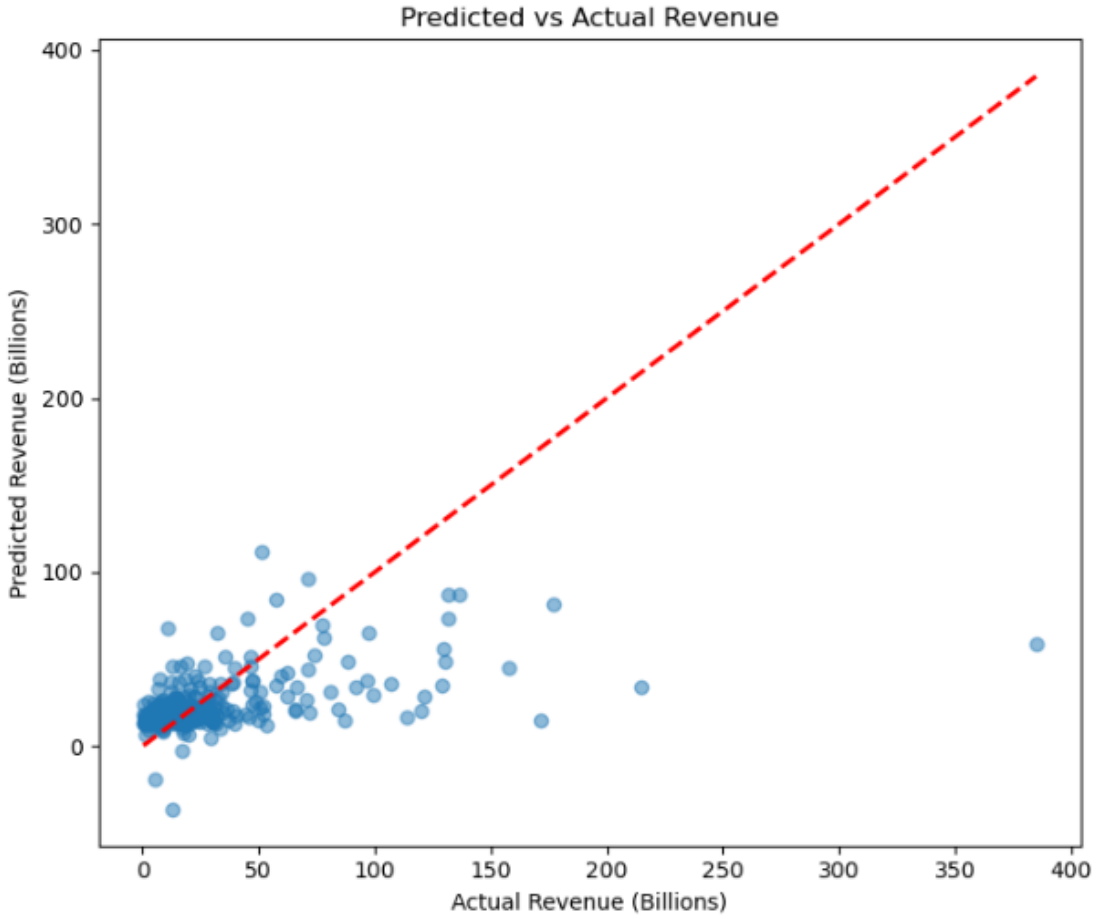
# MY PROJECT ENTAILS THE TOP FORBES COMPANIES IN THE WORLD WHEREBY WE LOOK FOR:

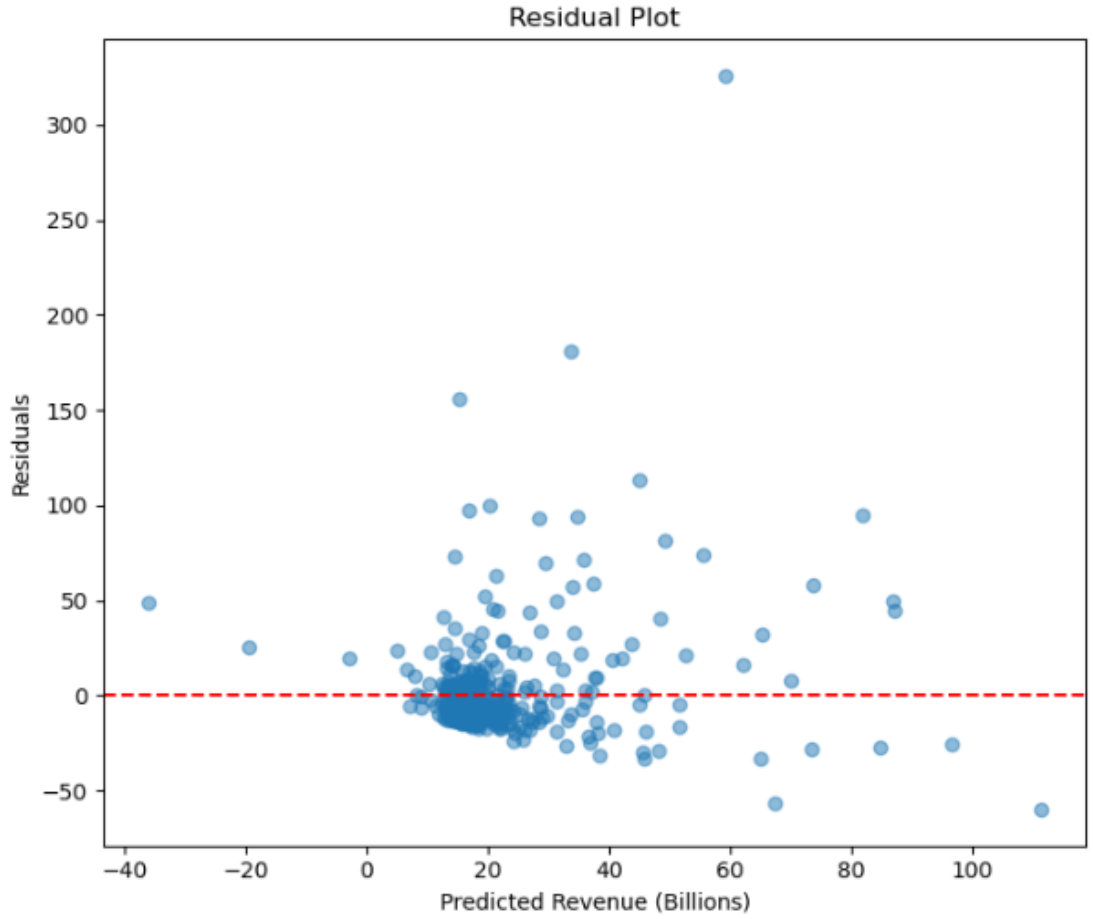
- **THE MOST PROFITABLE COMPANIES TO INVEST IN GLOBALLY.**
- **TYPE OF COMPANIES BASED ON REVENUE PROFITS AND ASSETS THAT ARE IN THE TOP AND MINOR LEAGUES.**
- **COMPANIES WITH THE MOST MARKET VALUES AND EMPLOYEES EMPLOYED IN DESCENDING ORDER.**
- **LOCATIONS THAT ARE THE BEST I.E COUNTRIES AND ORGANIZATION NAMES BASED ON RANKINGS.**
- **YEARS THAT THE TOP COMPANIES HAD THE BEST MARKET VALUES TO DETERMINE IF THEY ARE WORTH INVESTING IN.**
- **CONCLUSION TO INVESTORS WITH SPECIFIC RECOMENDATIONS BASED ON VISUAL REPRESENTATIONS ON THE BEST PLACES TO PUT THEIR INVESTMENT IN. THE DATA SET IS FROM KAGGLE FORBES SECTOR AND DEALS WITH THE GLOBAL ECONOMY.**

# Project Analysis Tools



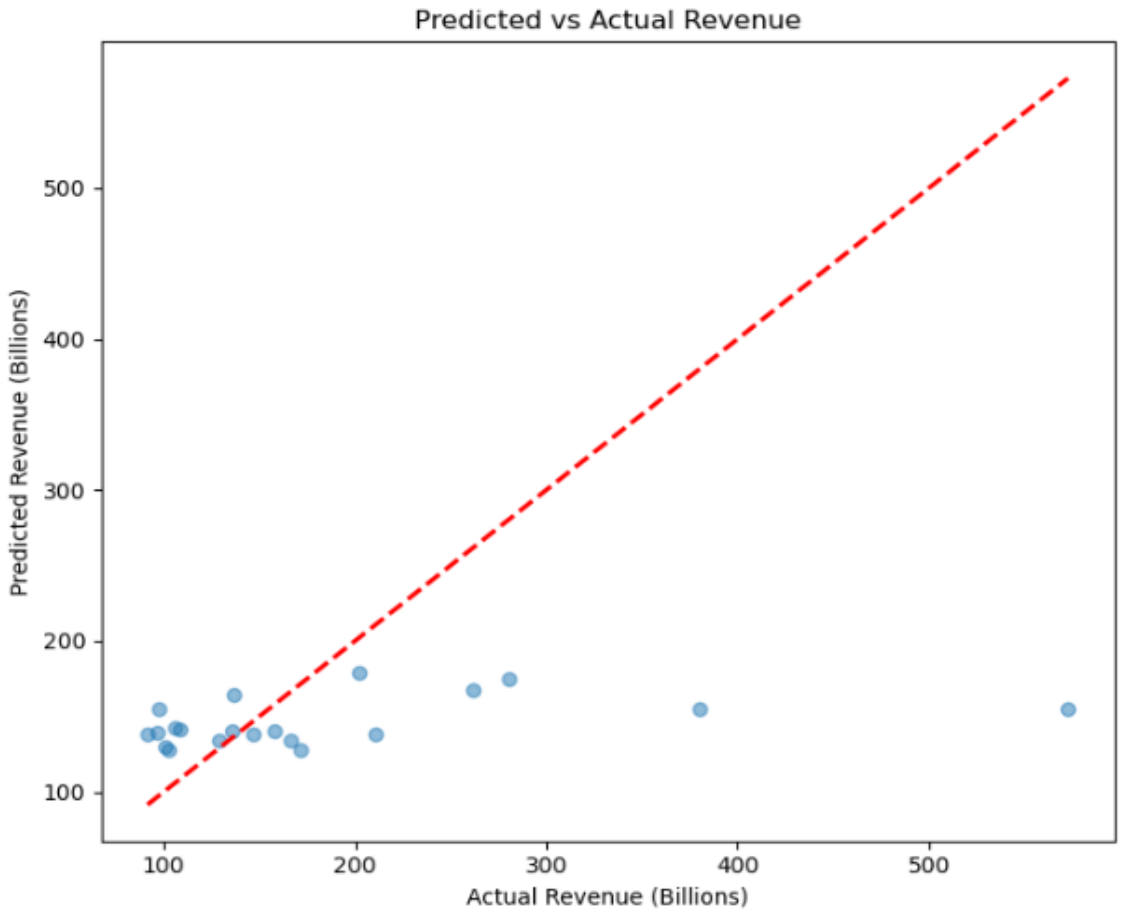
# Scatter plot for Profits_Billions vs Revenue_Billions

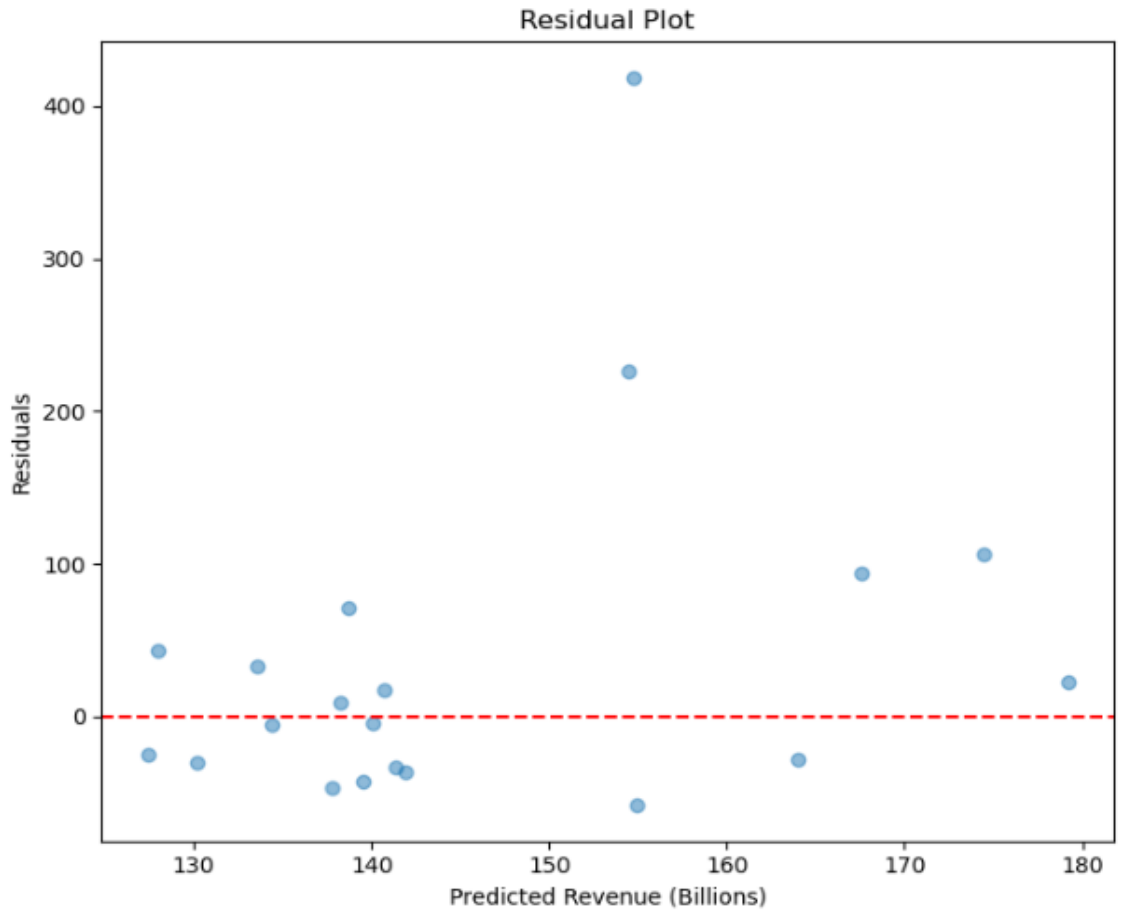


- Conclusion The multiple regression analysis aimed to understand the relationship between Revenue (Billions) and two independent variables: Total Employees and Market Value (Billions). The results of the analysis showed that:
- Model Fit: The model's R-squared value of 0.939 indicates that approximately 93.9% of the variance in revenue can be explained by the number of employees and market value. This suggests a strong fit of the model to the data.

# MULTIPLE REGRESSION FOR TOP 30.



# Multiple Regression Analysis in Python Using statsmodels

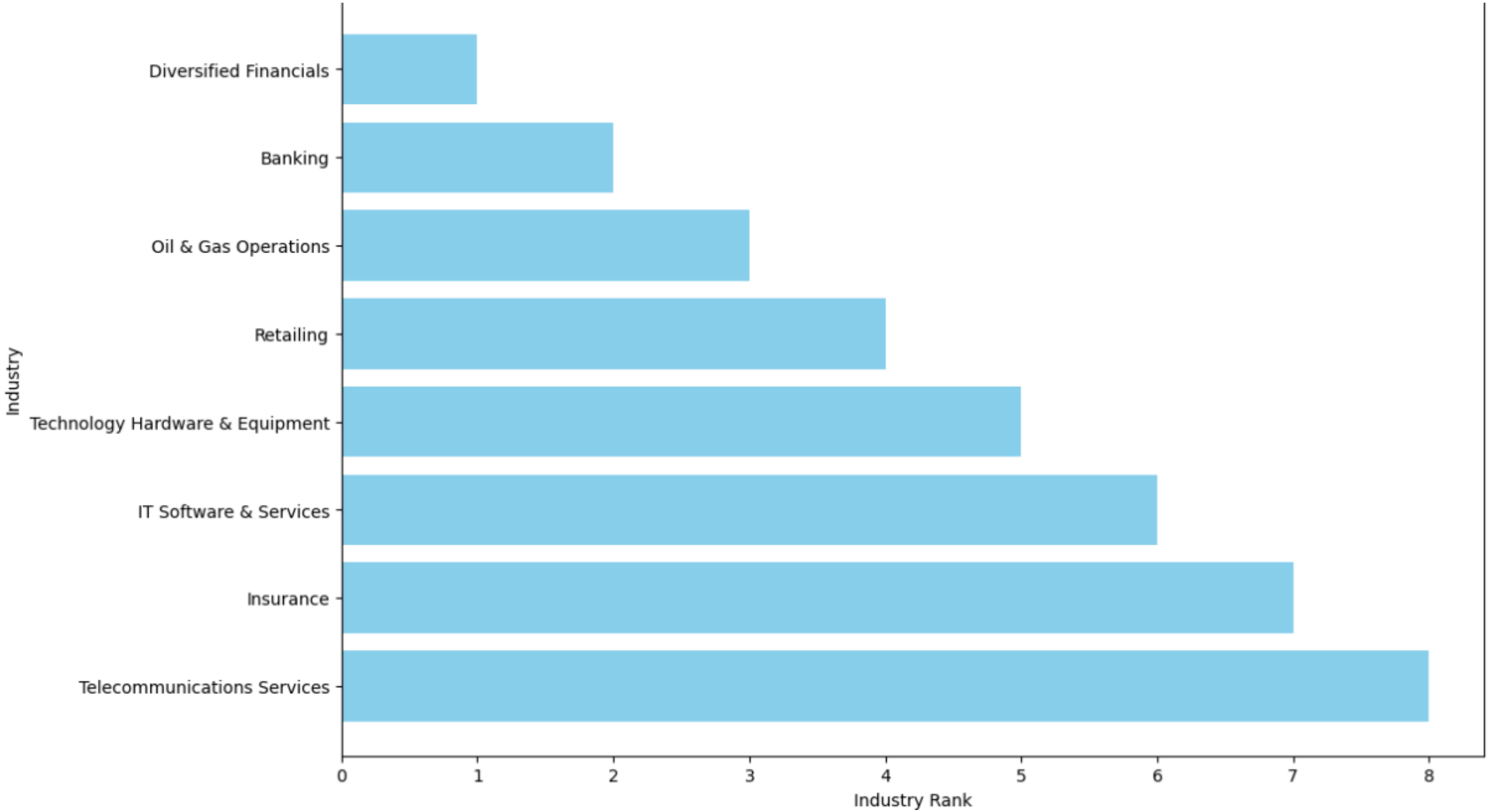
- Conclusion The multiple regression analysis aimed to understand the relationship between Revenue (Billions) and two independent variables: Total Employees and Market Value (Billions). The results of the analysis showed that:
- Model Fit: The model's R-squared value of 0.939 indicates that approximately 93.9% of the variance in revenue can be explained by the number of employees and market value. This suggests a strong fit of the model to the data.
- The p-value for Market Value (Billions) was close to the significance threshold, suggesting that it might be an important predictor of revenue, especially in larger datasets. The p-value for Total Employees was higher, indicating that the number of employees may not be as significant in predicting revenue within the sample used. Coefficients: The positive coefficients for both Total Employees and Market Value suggest that increases in either variable are associated with higher revenue, holding the other variable constant.

# I will rank the companies in descending order primarily by their Market Value and then by Total Employees as a secondary criterion.

- Here’s the ranking:
- 1.Saudi Arabian Oil Company (Saudi Aramco) Market Value: $2292.08 billion Total Employees: 68,493
- 2.Berkshire Hathaway Market Value: $741.48 billion Total Employees: 372,000
- 3.JPMorgan Chase Market Value: $374.45 billion Total Employees: 271,025
- 4.ICBC Market Value: $214.42 billion Total Employees: 449,296
- 5.China Construction Bank Market Value: $181.32 billion Total Employees: 352,621

# Inference

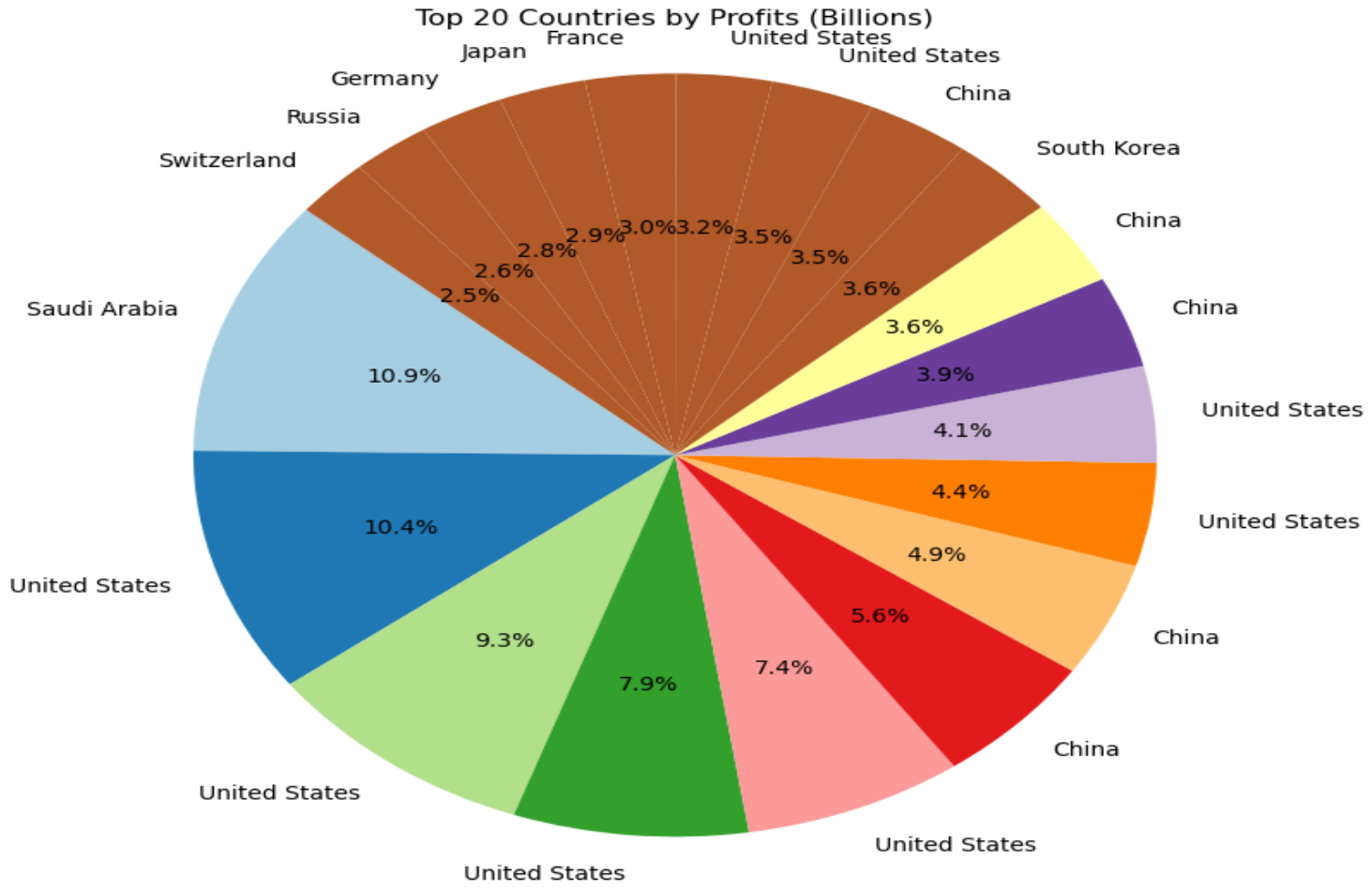
- Based on the results obtained from the regression analysis, Saudi Arabian Oil Company (Saudi Aramco) ranks as the top company due to its significantly higher market value. Berkshire Hathaway and JPMorgan Chase follow, with notable market values and a large number of employees. This ranking reflects the companies' overall financial power, with market value being the strongest determinant in the context of this analysis.


Diversified Financials
Banking
Oil & Gas Operations
Retailing
Technology Hardware & Equipment
IT Software & Services
Insurance
Telecommunications Services
Industry
0
1
2
3
4
5
6
7
8
Industry Rank

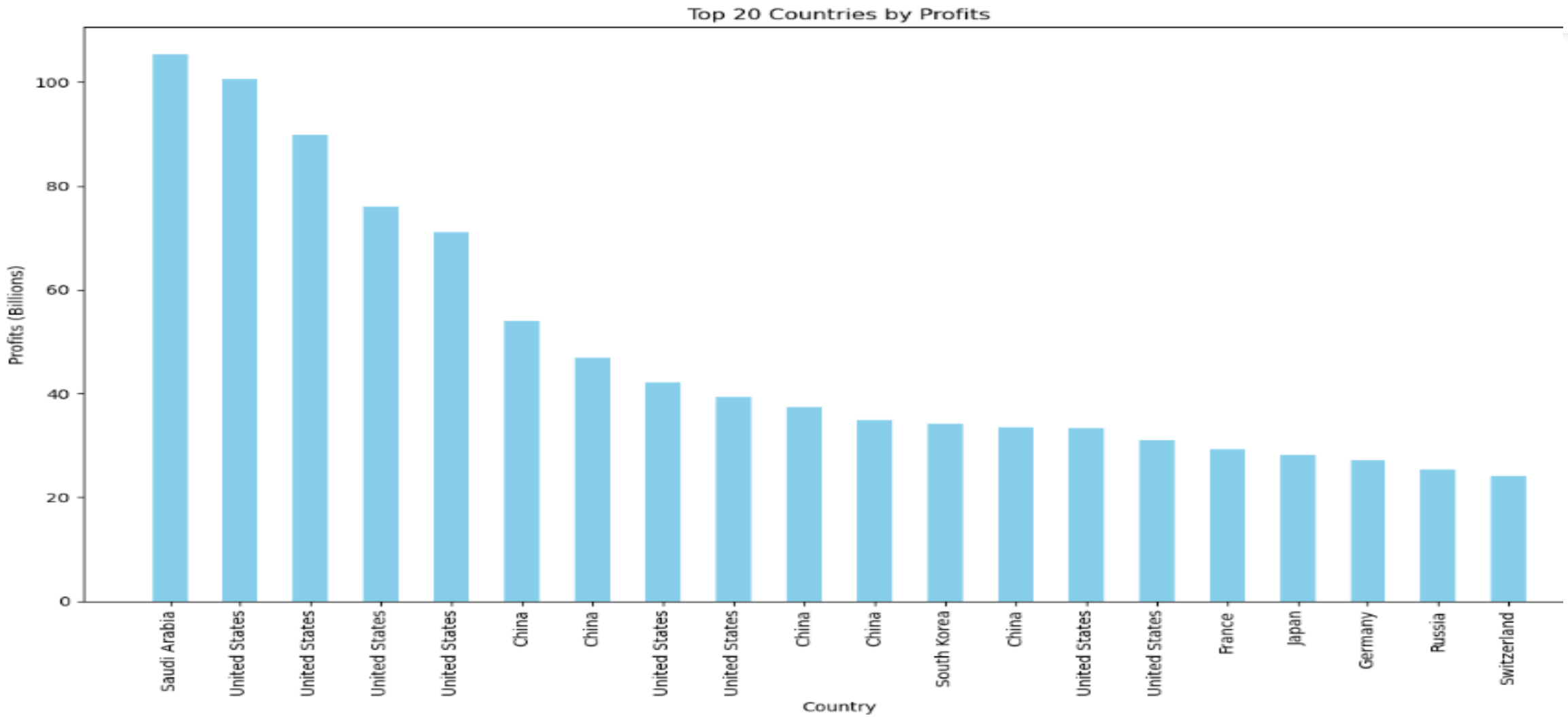
# Following industries and ranks:

- Banking will have the shortest bar, as it has the highest rank (2). IT Software & Services, Insurance and Telecomunication Services will have the longest bar, as it has the lowest ranks (6-8). This visualization helps in quickly understanding how different industries are ranked relative to each other, with higher-ranked industries appearing prominently at the top.

# USA THE BEST PLACE TO GET PROFITS.

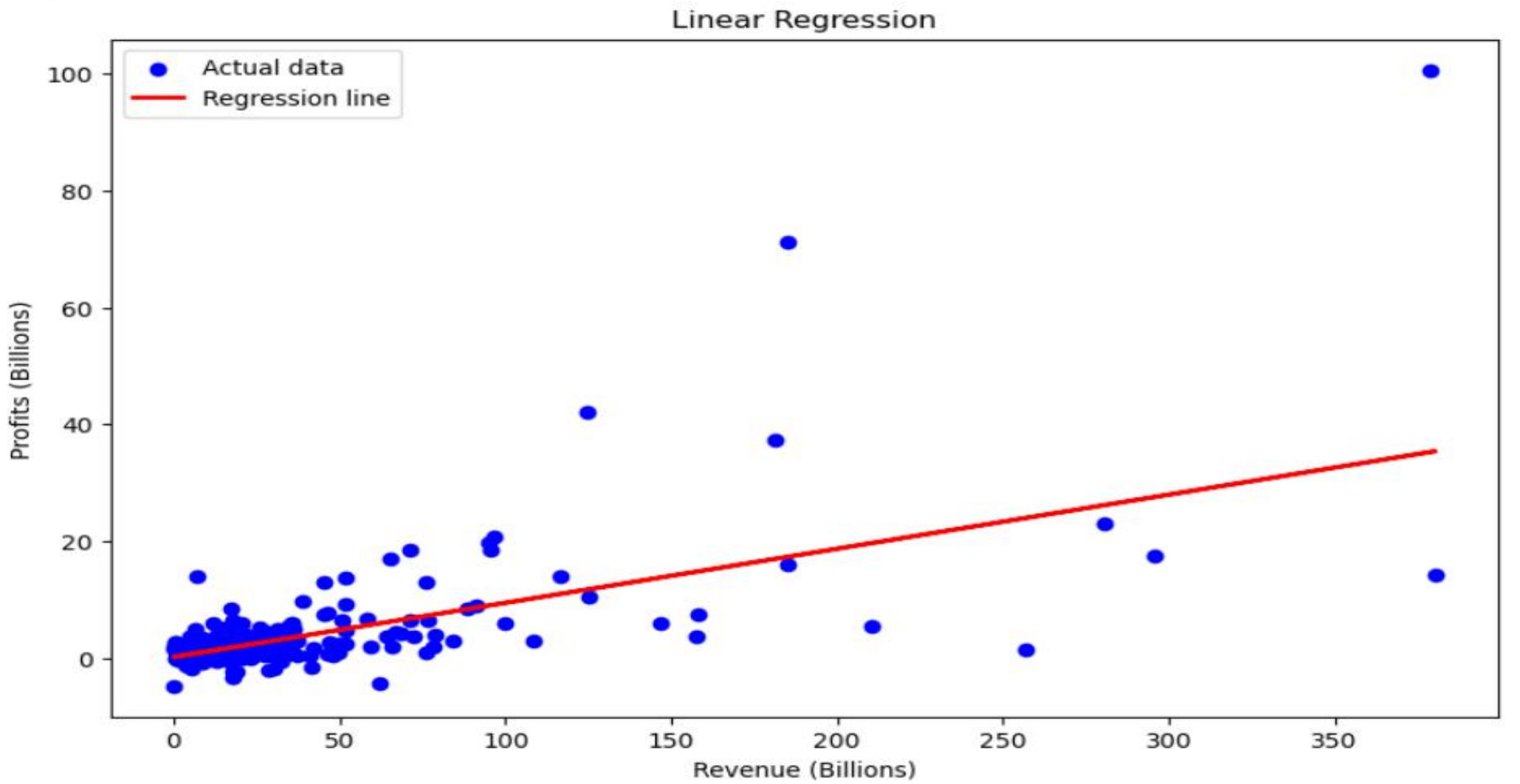


# Profits in Descending order.



- Saudi Arabia has the most profits and hence would recommend Investors to invest In this country.
- USA, CHINA AND SOUTH KOREA close to each other in profits hence I will recommend these countries also as significant profits are there.

# Multiple Linear Regression



- This interpretation is critical for understanding the financial landscape of the companies in your dataset, indicating that while the average revenue is around 23.89 billion dollars, there is a significant dispersion around this mean, reflecting diverse scales of operations among the companies.
- The average revenue for the companies is about 23.89 billion dollars. This gives us an idea of the central value around which the companies' revenues are distributed.

# logistic regression model

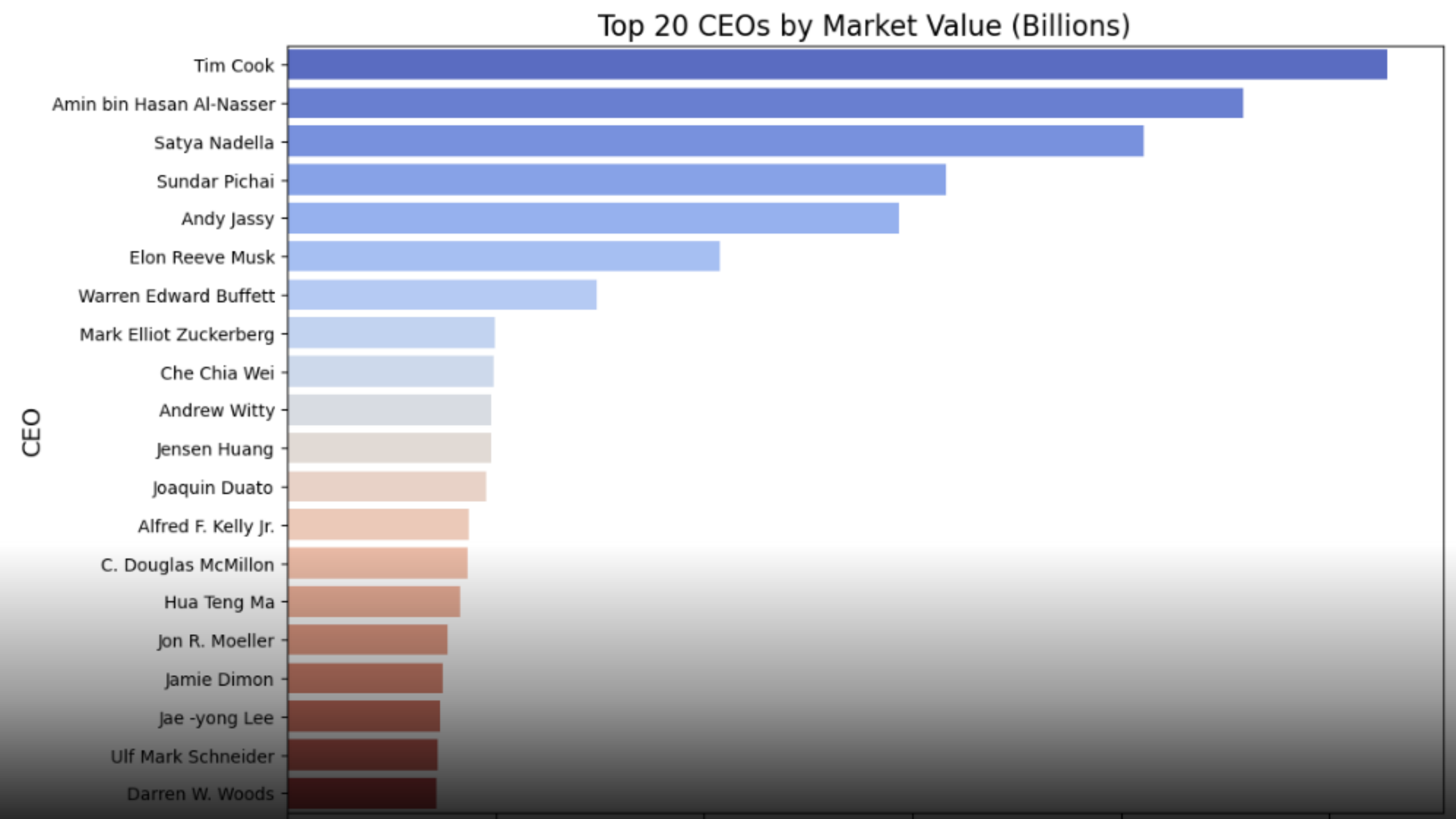
```
(0.9083333333333333,
 array([[523,   8],
        [ 47,  22]], dtype=int64),
 '              precision    recall  f1-score   support\n\n           0       0.92      0.98      0.95       531\n           1       0.73      0.32
0.44        69\n\n    accuracy                           0.91       600\n   macro avg       0.83      0.65      0.70       600\nweighted avg       0.90
0.91      0.89       600\n')
```

# Interpretation

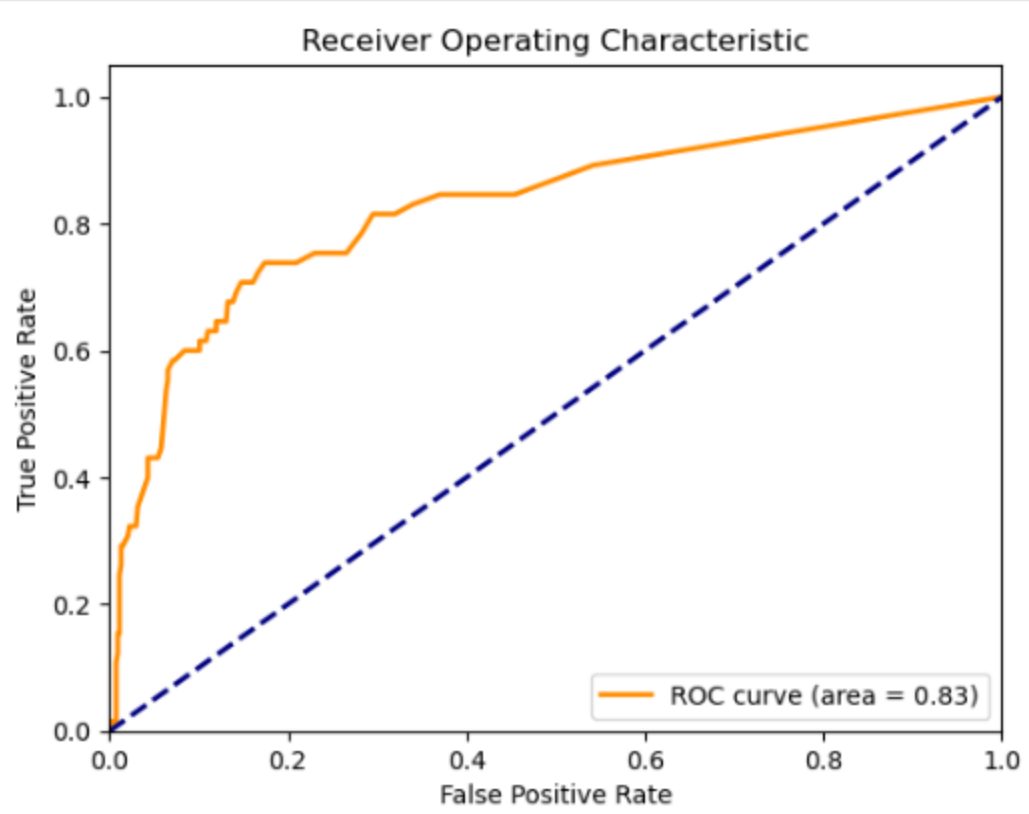
- The confusion matrix is a 2x2 array because we have a binary classification problem (0 and 1). True Negatives (TN) = 4: The model correctly predicted that 4 companies have 100,000 or fewer employees. False Positives (FP) = 0: The model did not incorrectly predict any companies with 100,000 or fewer employees as having more than 100,000 employees. False Negatives (FN) = 0: The model did not incorrectly predict any companies with more than 100,000 employees as having 100,000 or fewer employees. True Positives (TP) = 2: The model correctly predicted that 2 companies have more than 100,000 employees. This perfect confusion matrix (no false positives or false negatives) indicates that the model made no errors in classifying the companies.

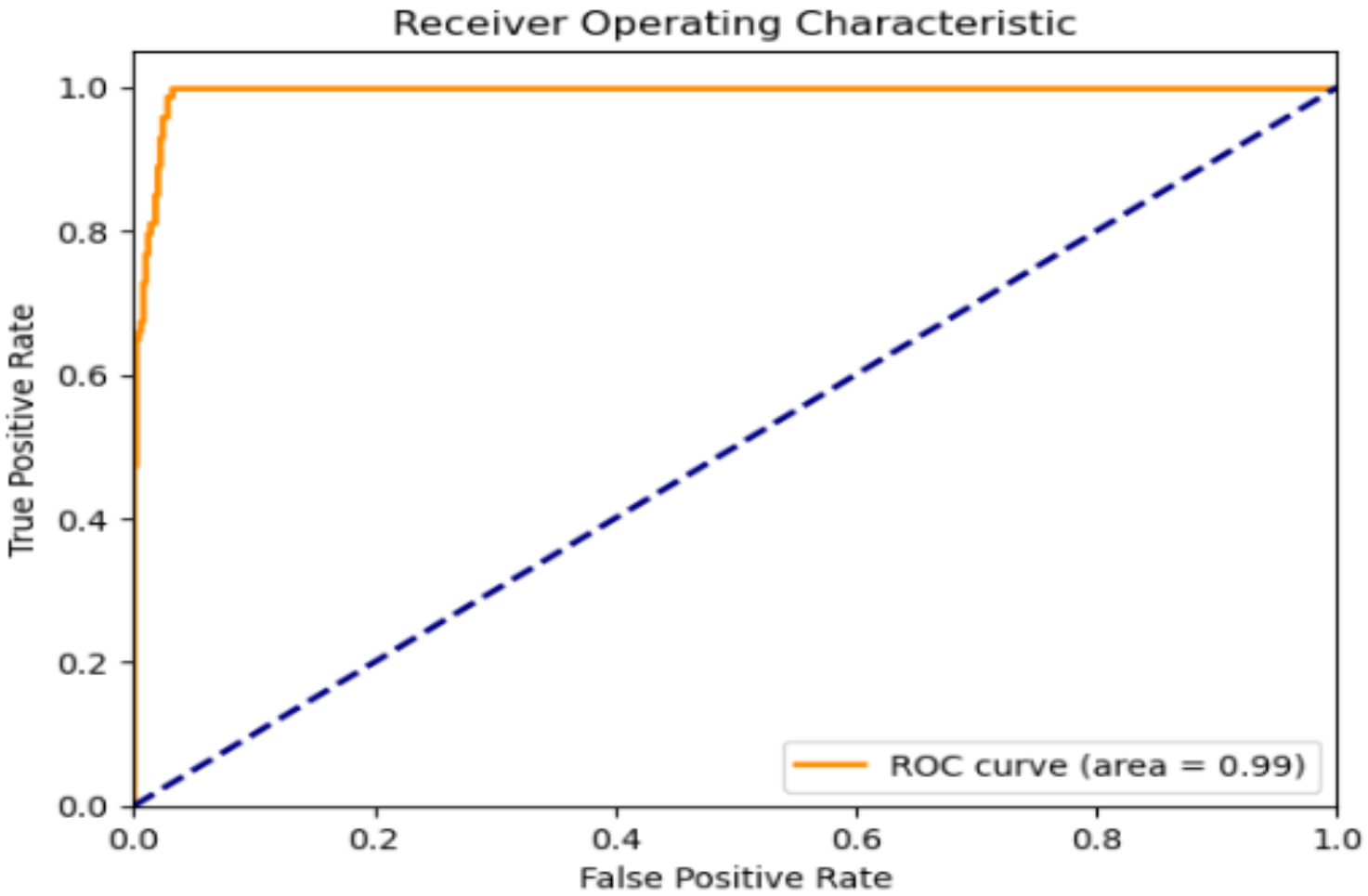
# Apple is The organization with largest market value.

| 2022 Ranking | CEO | Market_Value_Billions | Organization_Name |
|---|---|---|---|
| 7 | Tim Cook | 2640.32 | Apple |
| 3 | Amin bin Hasan Al-Nasser | 2292.08 | Saudi Arabian Oil Company (Saudi Aramco) |
| 12 | Satya Nadella | 2054.37 | Microsoft |
| 11 | Sundar Pichai | 1581.72 | Alphabet |
| 6 | Andy Jassy | 1468.40 | Amazon |
| 151 | Elon Reeve Musk | 1038.73 | tesla |
| 1 | Warren Edward Buffett | 741.48 | Berkshire Hathaway |
| 34 | Mark Elliot Zuckerberg | 499.86 | Meta Platforms |
| 58 | Che Chia Wei | 494.60 | Taiwan Semiconductor |
| 22 | Andrew Witty | 490.15 | UnitedHealth Group |
| 209 | Jensen Huang | 489.83 | NVIDIA |
| 40 | Joaquin Duato | 477.38 | Johnson & Johnson |
| 148 | Alfred F. Kelly Jr. | 436.49 | Visa |
| 23 | C. Douglas McMillon | 431.64 | Walmart |
| 28 | Hua Teng Ma | 414.28 | Tencent Holdings |
| 63 | Jon R. Moeller | 386.53 | Procter & Gamble |
| 4 | Jamie Dimon | 374.45 | JPMorgan Chase |


Top 20 CEOs by Market Value (Billions)
CEO
Tim Cook
Amin bin Hasan Al-Nasser
Satya Nadella
Sundar Pichai
Andy Jassy
Elon Reeve Musk
Warren Edward Buffett
Mark Elliot Zuckerberg
Che Chia Wei
Andrew Witty
Jensen Huang
Joaquin Duato
Alfred F. Kelly Jr.
C. Douglas McMillon
Hua Teng Ma
Jon R. Moeller
Jamie Dimon
Jae -yong Lee
Ulf Mark Schneider
Darren W. Woods

# ROC and AUC


Receiver Operating Characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.83)


Receiver Operating Characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.99)

# ROC and AUC explanation.

- When Target = 1: Positive Class: In this case, Target = 1 represents the positive class. For example, if Revenue_Billions > 50 was used to assign Target = 1, then organizations with revenue greater than 50 are considered positives. Model Performance: The ROC curve and AUC will show how well the model can correctly identify these high-revenue organizations (i.e., those where Target = 1) while minimizing false positives (organizations incorrectly classified as having high revenue).
- A curve closer to the top-left corner indicates a better model with higher sensitivity and lower false positive rate.

High AUC (close to 1): The model effectively separates the classes, indicating good performance.It means that your model is quite good at distinguishing between organizations with high and low revenue. It correctly identifies positive examples (high revenue) most of the time.

We have balanced against the risk of false positives.

# CONCLUSION

The data is Accurate since it is leading towards 1 as the ROC two different occasions that are 0.83 and 0.99 proving the data is accurate.

USA is the country with the largest return of invesment since it has the most organizations that have a significant share in the market value.

Forbes has the best companies that investors can invest in i.e based on country, industry, organization and market value.

# Recommendations.

Recommend USA,CHINA, SOUTH KOREA and SAUDI ARABIA as the best places to invest in.

Tim Cook, Amin Al Nassar, Satya Nadella and  Sundar Pichai as the CEO'S that yo can trust with you're investments.

Following industries and ranks; Diversified financials(Rank 1), Banking (Rank 2), Oil & Gas Operations (Rank 3), Retailing (Rank 4), Technology Hardware & Equipment (Rank 5) and IT Software & Services (Rank 6) recommended fields to invest in.