From figures to fortune: Comparative analysis of banking sectors using gamma regression model

Amrita Nath, Poushali Dutta, Sreyasi Dey
3rd year, St. Xavier's College (Autonomous), Kolkata

Introduction

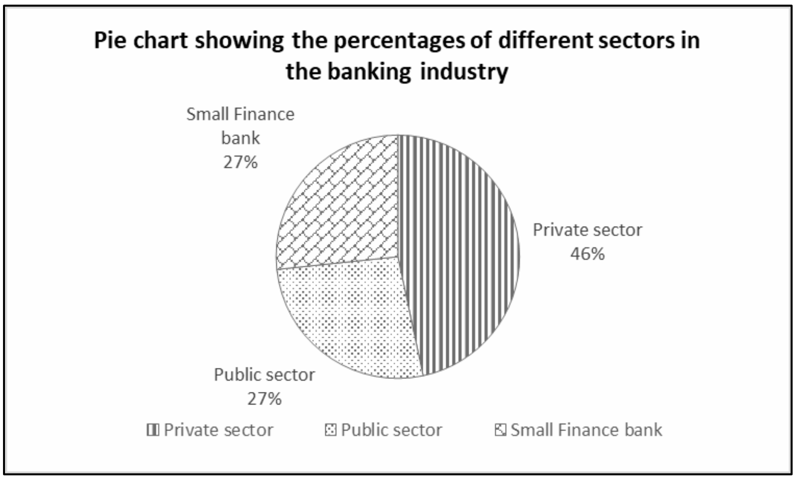
The financial sector plays a critical role in the global economy, serving as the backbone for economic growth and development. Banks are a vital part of this sector as they manage vast amounts of capital and provide essential services such as giving loans, investments, and risk management. This project aims to analyse financial data from Indian banks to uncover patterns, trends, and relationships that impact their profitability, stability, and growth. Through statistical analysis, we want to have a comprehensive understanding of the banking industry's current scenario and offer valuable insights for the improvement and betterment of the industry.

Data Description

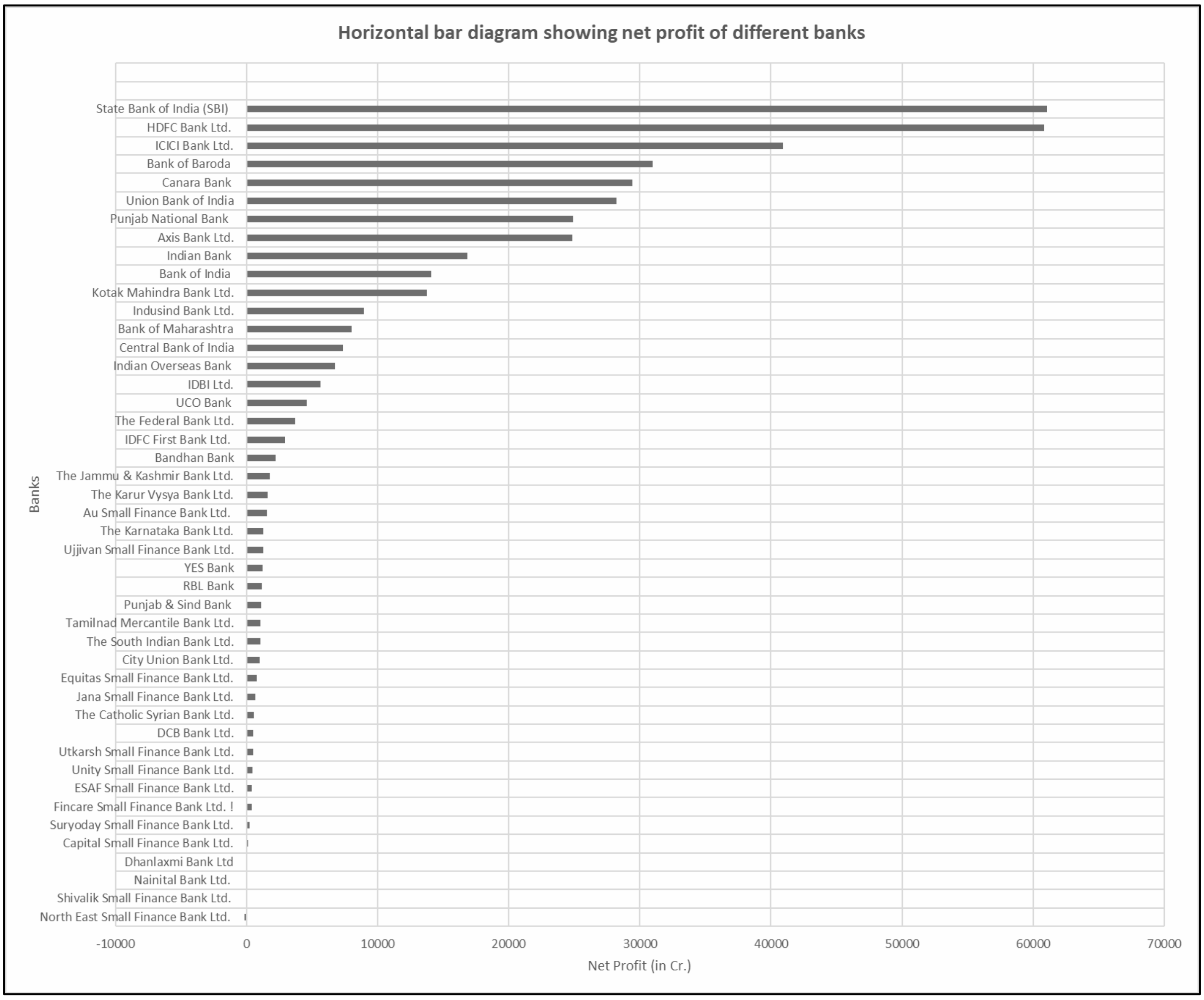
In this project, we will carry out the analysis on the bank performance data of the financial year 2023-2024. The data consists of information of 45 banks of India across three different sectors: Public Sector, Private sector and Small finance bank. We have extracted the information on the variables 'Deposits', 'Total assets', 'Net NPA', 'Total income', 'Total expenditure', 'Net profit', 'Credit deposit ratio', 'Investment deposit ratio', 'Return on assets', 'Capital adequacy ratio' and 'Net NPA as percentage to net advances' for the year 2024. Although, 'Operating expenses' is a part of Total expenditure, we have included it as well, to calculate efficiency ratio. All these variables are continuous in nature. So, we can measure them using ratio scale.

SOURCE: https://www.kaggle.com/datasets/lordpatil/indian-banks-key-business-statistics (October 2024 version)

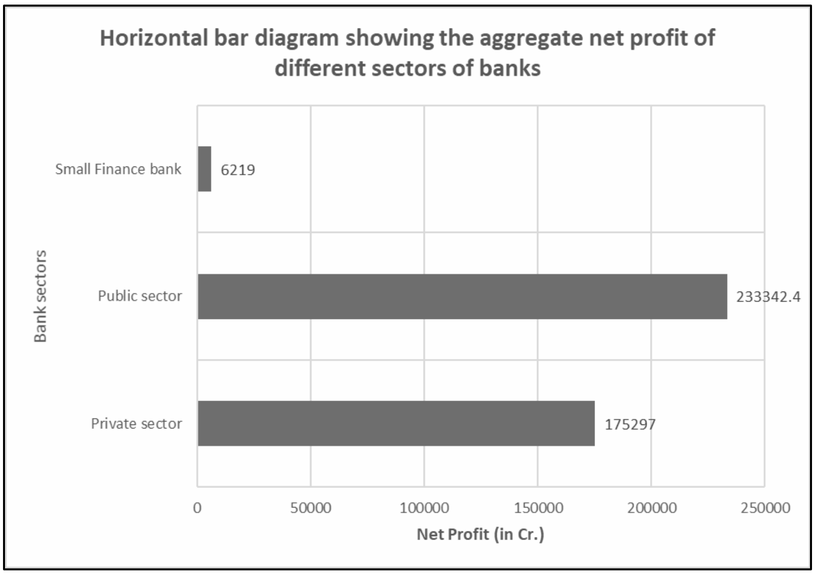
Graphical representations



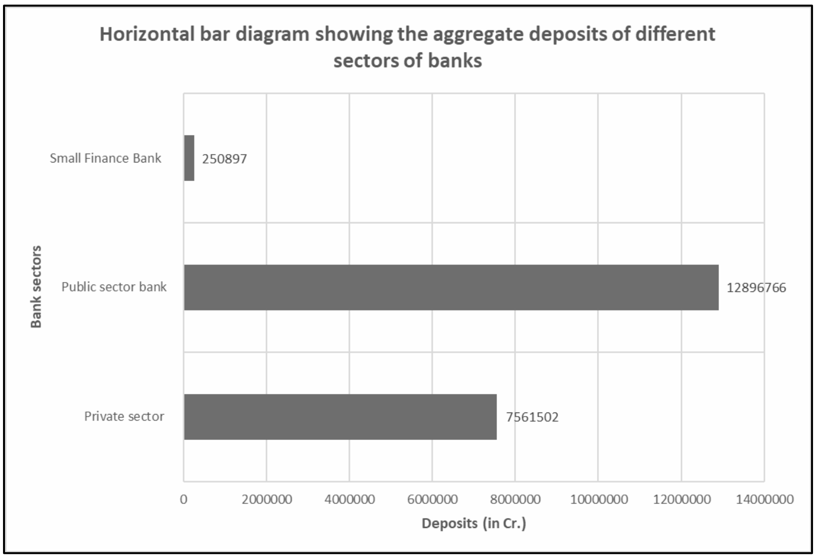
The pie chart shows that the percentage of private sector banks in the banking industry is higher than the other two sectors.



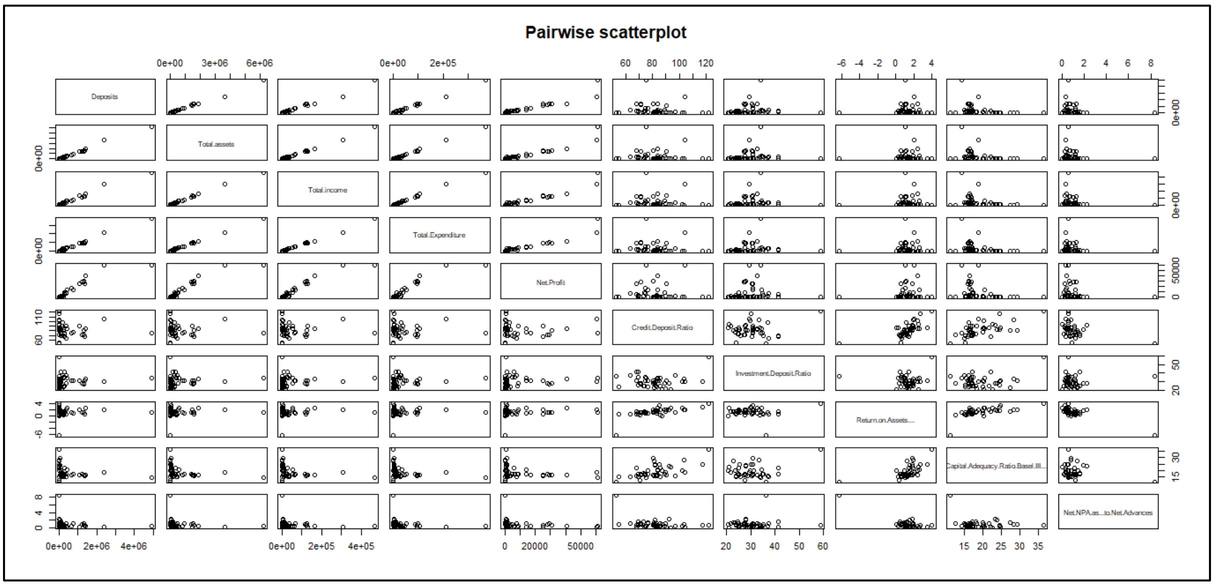
Here we have plotted a horizontal bar diagram showing net profit of 45 banks on the basis of data of 2023-24. From the plot we can say that according to net profit, the top 5 banks are State Bank of India (SBI), HDFC Bank, ICICI Bank, Bank of Baroda and Canara Bank.



We have drawn a horizontal bar diagram of net profits vs different sectors of banks. From the above graph, we can observe that the public sector banks make the most profit. Also, the small finance banks are making much lesser profit as compared to the public and private sector banks.

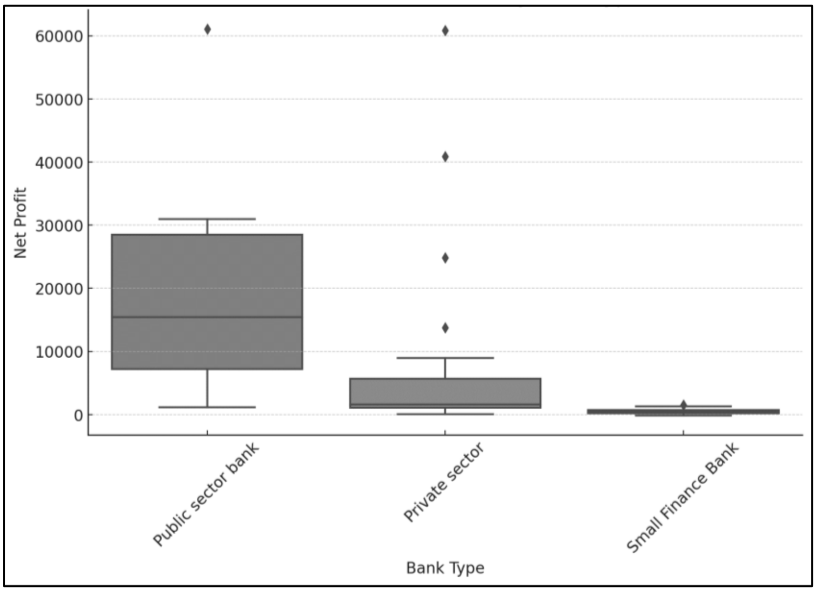


We have drawn a horizontal bar chart which shows that public sector banks have the most deposits. Also, the amount of deposits in public sector banks are very high as compared to private sector and small finance banks. This indicates that customers have more faith in public sector banks.



Next, we have created the scatterplot of pairs of different variables. From the scatterplot, we can understand the correlation between the pairs. We are interested in the factors which affect the Net profit of the banks. Here, we observe that Net profit is positively correlated (high) with Deposits, Total assets, Total income and Total expenditure.

Boxplot of Net Profit across different sectors of banks

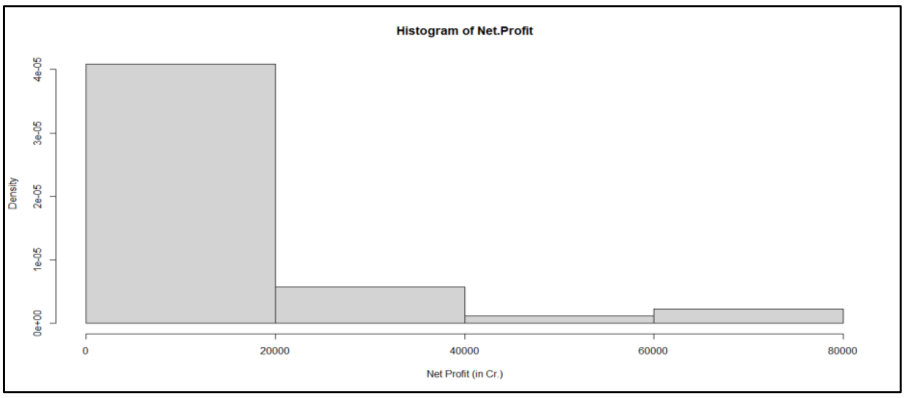


From the boxplot we observe that Net Profit varies significantly over the different sectors of banks. Median Net Profit is higher in Public Sector banks compared to other sectors. Outliers are present in private and public sector banks indicating some banks have exceptionally high Net Profit.

We are interested in studying the relationship between Net Profit and different sectors of banks. We denote Public Sector Bank by 1, Private Sector Bank by 2 and Small Finance Bank by 3.

GAMMA REGRESSION

In classical linear regression models, we assume that the response is normally distributed. To observe the distribution of Net profit, we plot its histogram.



From the histogram, we observe that the distribution of Net Profit is positively skewed. Hence, we use Generalized linear models, assuming that Net Profit follows a Gamma distribution.

Let Y denote the response variable “Net Profit” and x_1 and x_2 denote the predictors “Total assets” and “different sectors of banks” respectively.
Here, x_2 is a categorical predictor consisting of three categories: Public Sector, Private Sector and Small finance Bank. So, here we cannot incorporate sectors of banks using a single regression coefficient in the model. This is because the effect on the model when one moves from category 1 to category 2 of sectors might not be the same as when one moves from category 2 to category 3 of sectors. In such a case we define two new predictors x_{22} and x_{23} as follows,

$$x_{22} = \begin{cases} 1 \quad , \quad \textit{if the bank is a Private Setor Bank} \\ 0 \quad , \quad \quad \quad \quad \quad \quad \quad \textit{Otherwise} \end{cases}$$

$$x_{23} = \begin{cases} 1 \quad , \quad \textit{if the bank is a Small Finance Bank} \\ 0 \quad , \quad \quad \quad \quad \quad \quad \quad \textit{Otherwise} \end{cases}$$

Note that, here we have considered two predictors: Total assets and Sectors of the bank. To check if multicollinearity exists, we calculate the Generalized Variance Inflation Factor (GVIF). Generally, GVIF values less than 5 implies low multicollinearity among the predictors.

	GVIF	df	GVIF^(1/(2*df))
Total.assets	1.2158	1	1.1026
Bank.Type2	1.5931	1	1.2622
Bank.Type3	1.7352	1	1.3173

Since, here the values are less than 5, we can say that the effect of multicollinearity is not so pronounced.

We assume Y ~ Gamma ($\boldsymbol{\alpha}$,p) where p>0 is the shape parameter and $\boldsymbol{\alpha}$>0 is the scale parameter. The density of Y is given by,

$$f(y) = \frac{\alpha^p}{\Gamma(p)} e^{-\alpha y} y^{p-1} \quad , \alpha>0, p>0$$

We have observations on n=44 individual banks (We omit the observation corresponding to North East Small Finance Bank Ltd. because its Net Profit is < 0).

We generally assume the shape parameter to be constant which controls the skewness of the distribution. Therefore,
$Y_i \sim \text{Gamma}(\boldsymbol{\alpha}_i, p)$ for all i=1(1)44

Let $E(Y_i)=\mu_i$ and $\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_{22} x_{22i} + \beta_{23} x_{23i}$ which is a linear function of predictors.

In case of Gamma regression, the canonical link i.e. the inverse link does not preserve the range of η. So, the log link is generally used which is given by $g(\mu)=\ln(\mu)$ where ln(.) is the natural logarithm.

The Gamma regression model is given by,

$\ln(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_{22} x_{22i} + \beta_{23} x_{23i}$ $i=1(1)44$

Using the GLM function in R, we obtain the following summary table.

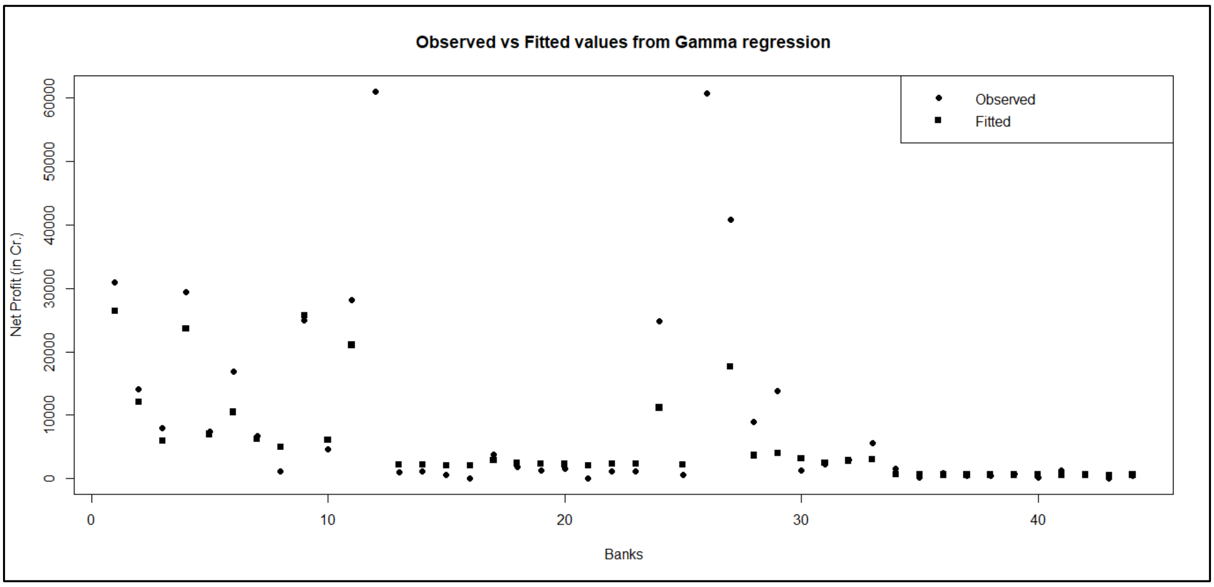
```
Call:
glm(formula = Net.Profit ~ Total.assets + Bank.Type2 + Bank.Type3,
    family = Gamma(link = "log"), data = data)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.327e+00  2.791e-01  29.835  < 2e-16 ***
Total.assets  1.170e-06  1.208e-07   9.682 4.86e-12 ***
Bank.Type21  -7.354e-01  3.054e-01  -2.408   0.0208 *
Bank.Type31  -2.023e+00  3.677e-01  -5.501 2.38e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.6428164)

    Null deviance: 131.657  on 43  degrees of freedom
Residual deviance:  45.312  on 40  degrees of freedom
AIC: 817.2

Number of Fisher Scoring iterations: 20
```

Interpretations:

- The intercept value (8.327) is on the logarithmic scale. In the model we are using a log link function. To interpret it in terms of net profit, we need to exponentiate the value:
  $e^{8.327} \approx 4124.74$
  This implies that the predicted average net profit for public sector banks, when total assets are zero, is approximately **₹4124.74 cr**. Economically, total assets are unlikely to be zero for a bank. This interpretation is absurd, possibly due to some errors or presence of outliers.
- Positive coefficient of total assets indicates that larger total assets are associated with higher net profits. For a unit increase in Total Assets, the expected change in the log of average Net Profit is 1.170e-06 when sector is fixed at category 1 . The coefficient is significant since p-value is very small (at 5% level of significance), which indicates that total asset is a significant predictor of Net Profit.
- For private sector banks (**β**22), the log of average Net Profit decreases approximately by 0.7354 as compared to public sector banks, when total assets are fixed at a particular level. This indicates that private sector banks have lower net profits compared to public sector banks, assuming the other variables are constant.
- For small finance banks (**β**23), the log of average Net Profit decreases approximately by 2.023 as compared to public sector banks, when total assets are fixed at a particular level. This indicates that small finance banks have significantly lower net profits compared to public sector banks, assuming the other variables are constant.

Goodness of fit

After fitting the model, we obtain the fitted values of the data using fitted( ) function in R. Now, we plot the observed and the fitted values on the same graph.



Here, we note that the observed and fitted values are more or less close to each other except a few extreme values. Moreover, from the fit, we observe that,

Dispersion parameter: 0.6428 which is the estimate of the dispersion (or variance) in the Gamma distribution, indicating the spread of the Net Profit data around the fitted model.

Null deviance: 131.657 (with df 43)

Residual deviance: 45.312 (on df 40)

As we include the variables “Total assets” and “Sectors” in the model, the deviance decreases indicating that the variables have significant role in predicting the response variable.

The reduction in deviance (from 131.657 to 45.312) indicates that the model explains a substantial portion of the variation in the dependent variable, indicating that our fit is moderately good.

EFFICIENCY ANALYSIS

The efficiency ratio shows how well a bank manages its cost. A low efficiency ratio is better as it indicates that the bank is spending less to generate income.

Efficiency=Operating Expense/Total Income

The top 5 most efficient banks are as follows:

Bank	Total income	Operating expense	Efficiency
Canara Bank	127654.4	26119.79	0.204613
Bank of Maharashtra	23492.56	4814.38	0.204932
HDFC Bank Ltd.	307581.6	63386.01	0.206079
Union Bank of India	115858.2	24439.96	0.210947
Bank of Baroda	127101.3	28251.68	0.222277

From the above table we can observe that banks which have high net profit like SBI, HDFC Bank, ICICI Bank, Bank of Baroda, Canara Bank may not always be efficient in managing their expenses. We observe that among the top five efficient banks, four of them belong to the public sector. Among the private sector banks, HDFC has the highest net profit and is the most efficient.

Conclusion

In this project, we compared the performance of different sectors in the banking industry and observed that public sector banks dominate the industry with the most deposits, highest efficiency and highest profit. We also analysed the effect of different sectors and total assets on the profitability of banks. The Gamma regression model provides a clear understanding of these relationships, offering valuable insights for improving decision-making and strategy in the financial sector. An efficiency indicator like efficiency ratio is also important for sustainable growth. The findings highlight how statistical data analysis can guide financial institutions to make better decisions, navigate challenges, and enhance their performance.

References

1. Indian Bank Association : https://www.iba.org.in/depart-res-stcs/key-bus-stcs.html
2. Pervez, A., Ali, I. Robust Regression Analysis in Analysing Financial Performance of Public Sector Banks: A Case Study of India. *Ann. Data. Sci.* **11**, 677–691 (2024). https://doi.org/10.1007/s40745-022-00427-3
3. Ford, C. 2022 "Getting Started with Gamma Regression” UVA Library StatLab https://library.virginia.edu/data/articles/getting-started-with-gamma-regression