FIT5147

Data exploration and visualisation
DEP: PART-2

COMPARATIVE STUDY OF BANKING INDUSTRY IN AUSTRALIA AND INDIA

Bhavna Balakrishnan
33954437
Applied Session 1
Tutors: Jeffery Liu, Niranjan Nanjunda

INTRODUCTION

The Banking Industry plays a crucial role in a country's economy. It is quite important to understand the workings and performance of banks since they play a major role in financing. In general, understanding the performance of a company will help give potential investors insights that aid them to make well informed decisions and formulate optimal investment strategies.

The goal of this project is to first and foremost understand the performance of the Top 5 Industry leaders of the banking Industries in Australia and India and then try to draw a comparison between them. Finally, we will also explore if the banks' performance is any way impacting or have impacted the country's interest rates. The analysis will be done during the time period of December 2018-December 2023. This 5-year scrutiny of the performances of major banks will help determine the insights and implications from the analysis to aid stakeholders in decision-making processes, such as investment strategies, policy formulation, and risk management. To put in short, below are the main objectives of this project-

- Understand and analyse the performance of the banks
- Compute the volatility of returns.
- Compare banks in Australia with those in India.
- Determine any relationship between Bank performance with Interest Rates.

MOTIVATION

The motivation I had to do this project is because I have a background in the field of Finance. I have done my Bachelors in Business Administration specialising in Finance and have worked in an American Bank called Wells Fargo for 1.5 years. Given my background, I am interested to see how the Banking Industries of both Australia and India compare to each other in terms of performance. I believe my knowledge of finance combined with the new learning of visualisation tools would make this a very interesting project to take and draw insights from.

DATA WRANGLING

Prior to the wrangling I have identified my primary data sources for historical stock data which were Yahoo Finance and BSE Stock Index. Once the data was collected it was wrangled as follows-

- Flipping Data into chronological order: The data for some banks came in the opposite order of most recent to past. In order to remain consistent all were made from past to most recent. Below is a screenshot as an example of what it was-

	A	B	C	D	E	F	G	H
1	Date	Open Price	High Price	Low Price	Close Price	WAP	No. of Trades	
2	29-Dec-23	646.45	649.5	639.55	641.95	642.419324	19339	
3	28-Dec-23	650	653.35	646.4	651.1	650.198611	8194	
4	27-Dec-23	641.4	649.6	639.05	648.35	644.733308	6340	
5	26-Dec-23	639.85	641.3	635.8	637.95	639.232103	8665	
6	22-Dec-23	645.4	649.4	635.25	636.65	641.243574	24444	
7	21-Dec-23	628.25	648.75	627.2	643.9	637.478499	9840	
8	20-Dec-23	659.4	660.4	633.7	636.15	646.116699	28522	

- Deleting unnecessary columns: The only columns that are required for this project are the dates, open, close, high, low, adjusted and volume columns. Other columns were deleted.

- Computing Estimated Returns: In order to understand how well a stock is performing, we can calculate the estimated returns of the stock by using the below formula-

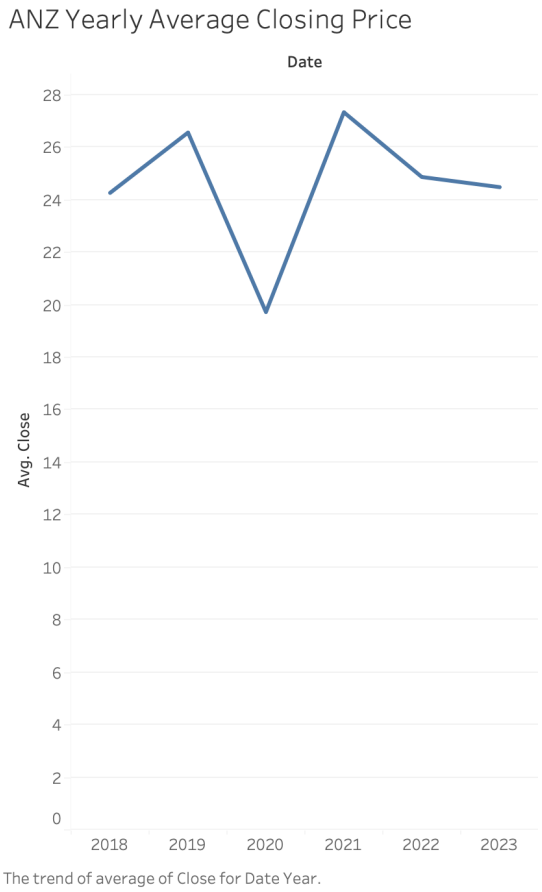
$$R_i = \frac{(P_1 - P_0) + D}{P_0} \times 100$$

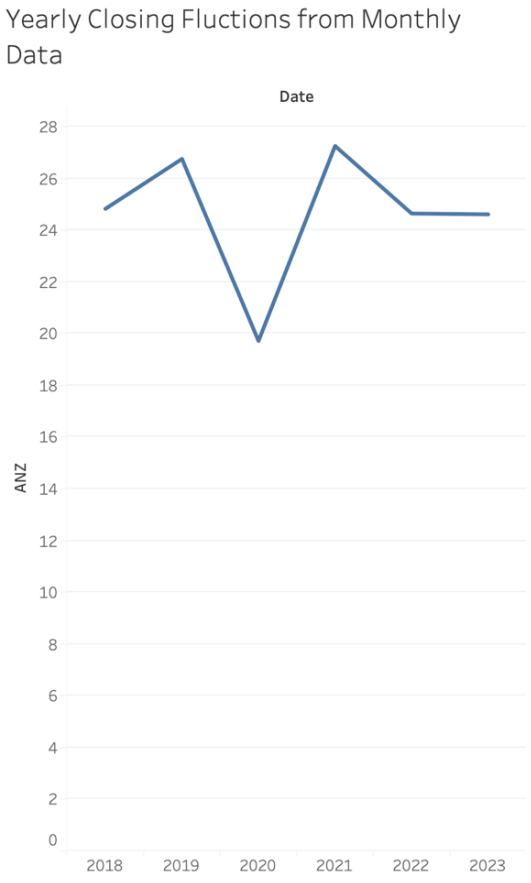
Here we have not included dividends since we are analysing and a continual basis and not a specified holding period.

- Identifying and merging relevant data from the data source: Two sheets from RBI's archives on Lending rates were taken and merged to get the required data within the timeframe of 2019-2023, unfortunately RBI only provided data from range 2020-2023 which will be used for the regression testing. The Weighted Average Interest Rate data across various typed of banks like Public, Private and International banks were averaged to get a more holistic number to run against the data from our selected banks.

DATA CHECKING

- All the data was taken for very reliable sources like Yahoo Finance that sourced it from the Australian Securities Exchange, RBA, RBI and BSE which is the Bombay Stock Exchange. After the data was sourced, it was roughly checked on excel for null and duplicate values.
- Random formula checks on the estimated returns were done to ensure accuracy of the formula used in Excel.
- Calculating Averages as necessary for merging datasets.
- Computing that the sum of estimated returns on a daily basis for a month was roughly similar to the monthly estimated return. Let us take a look at the same below-





What we can see is that there are similar and non-anomalous trends between the yearly closing prices. The diagram on the left was taken from daily stock data whereas the one on the right is from monthly data. The fact that both graphs are similar indicate consistency in the data.

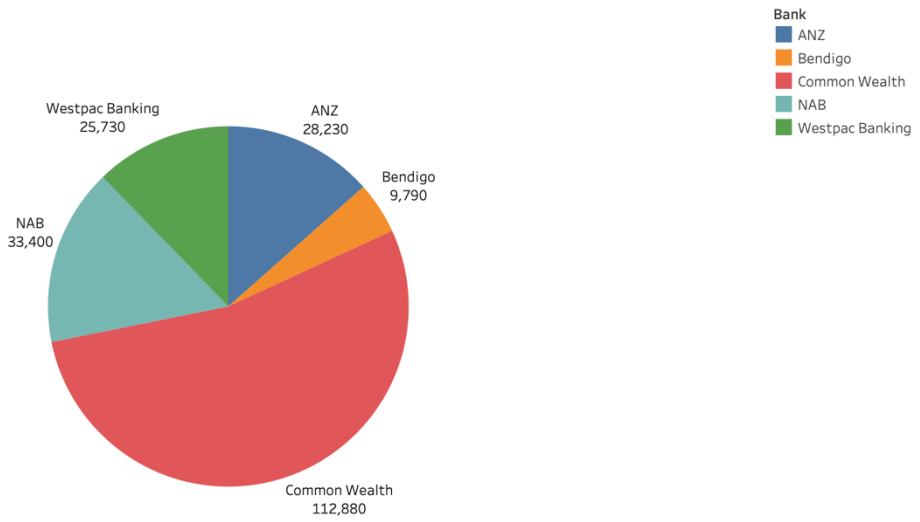
PERFORMANCE INDICATORS:

Market Capitalisation

We will begin by first identifying the top 5 banks of each country to analyse their financial data. The best way to understand the industry's performance is by reviewing the Industry Leaders' performances. I am choosing the top 5 on the basis of Market Capitalisation which is a good factor to consider when determining a company's worth.

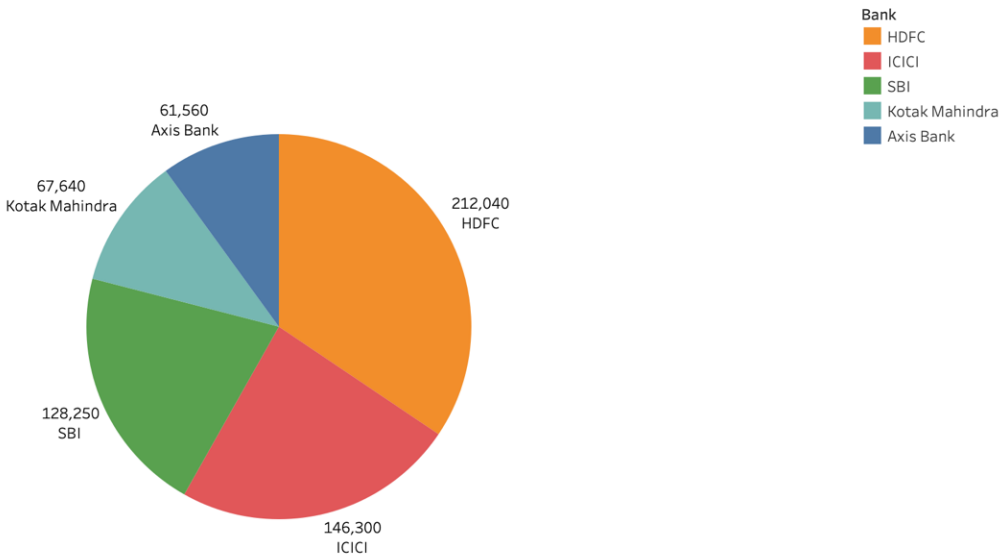
Australia

Below we can see a pie chart depicting the market capitalisation of the top 5 Industry Leaders in Australia per Million Dollars-



India

Similarly, we can see below the market leaders for India-

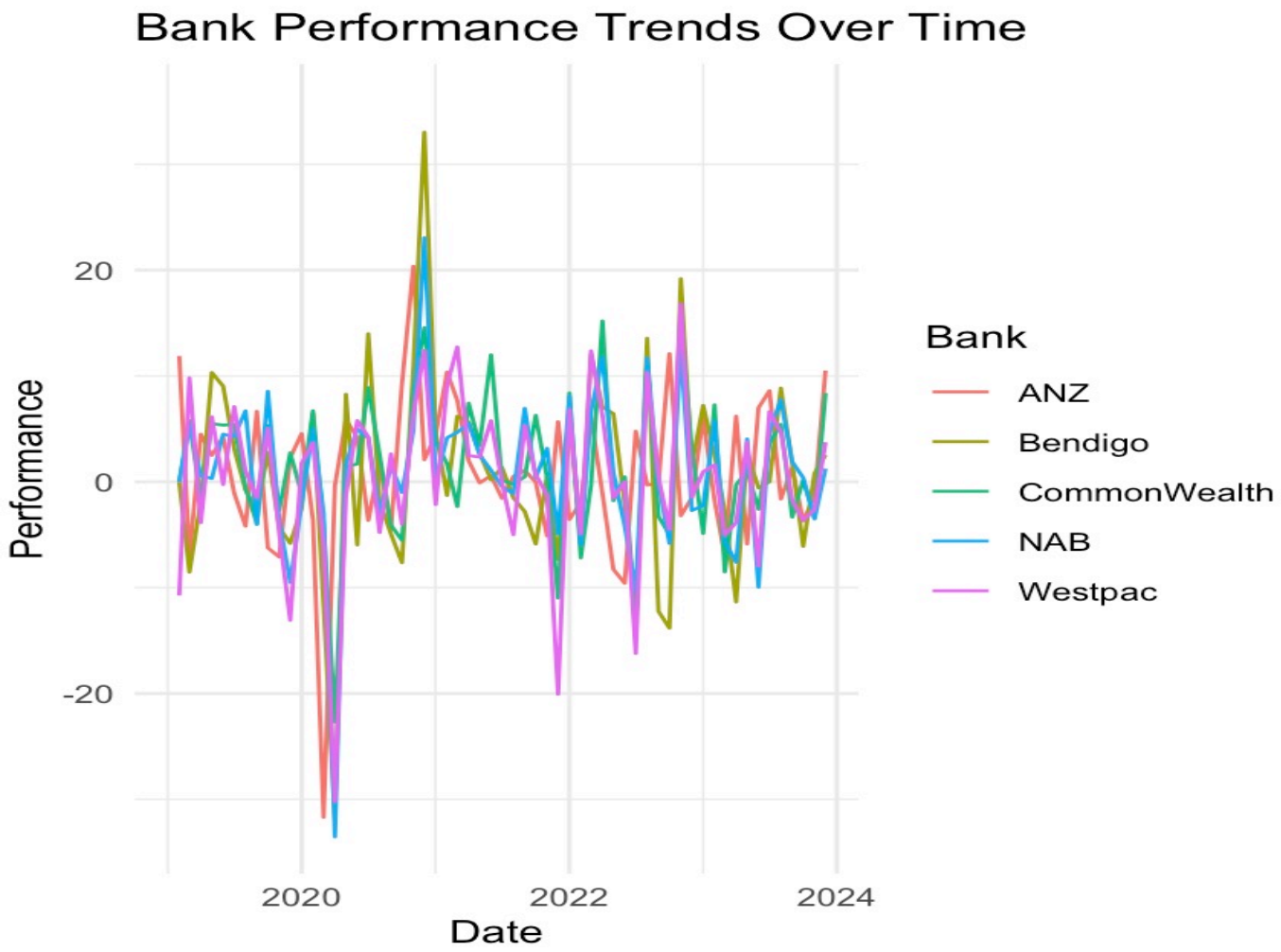


Trend Analysis

Now that we have identified the top banks, we must determine a suitable timeframe in which to examine their performance. For this report we will be scrutinizing performance between December 2018- December 2023 since the year 2024 is still progressing and the inclusion of December 2018 is required as the first input to calculate estimated returns from January 2019.

Let us start looking at the trends we can observe from the performances of these banks:

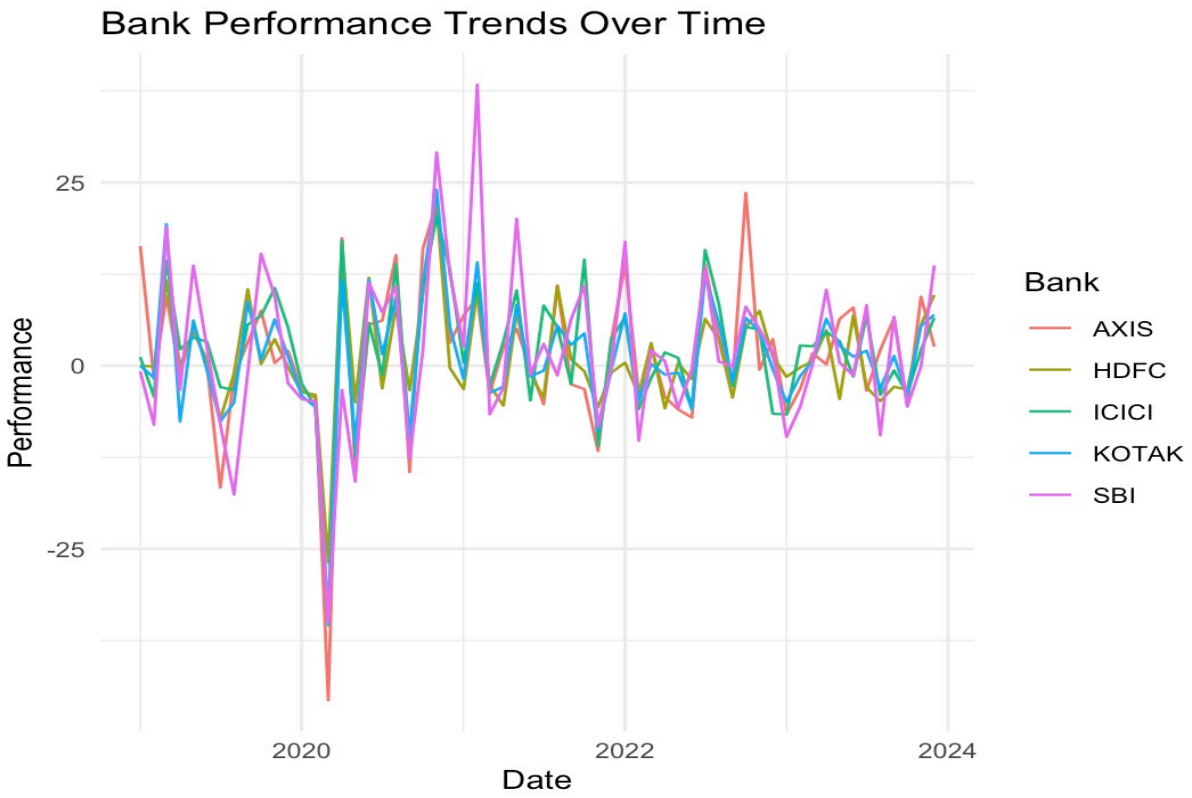
Australian Banks



From the above figure, we can identify that all the banks faced a severe drop in their returns during 2020, which was when the COVID-19 Pandemic hit. It also appears that they managed to recover around 2021when performance peaked, and have been fluctuating since. In 2022, NAB appears to take the lead with Westpac closely following. 2023 saw similar performances among banks and in 2024, ANZ performed the best while NAB had the least returns.

Indian Banks

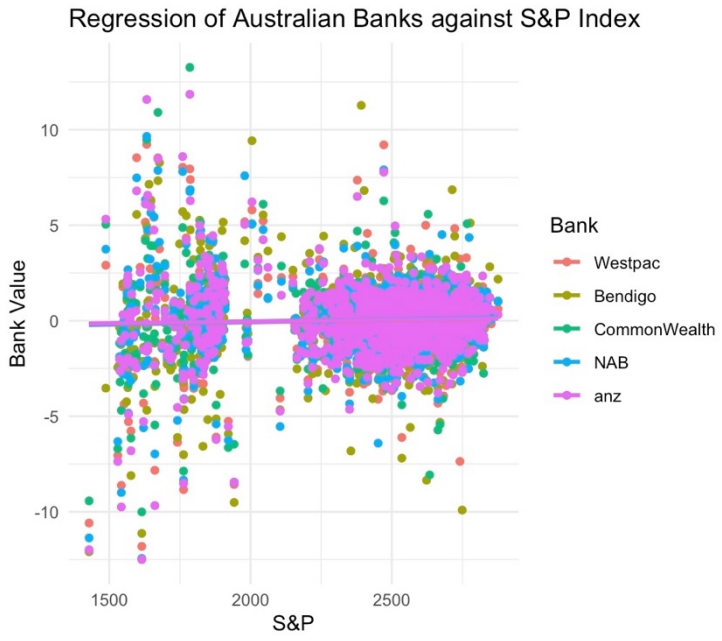
Now we will observe the trends in the banks in India over the years. Below we can see that similar to Australia, Indian Banks too, faced a hit in 2020 most likely due to the COVID-19 Pandemic and starting recovering closer to 2021 where there performance reached an all-time high with SBI outperforming the rest by a high margin.SBI remains the market leader for most part, except for near 2022 where ICICI beat them and in 2023 where Axis Bank overtook them.



Benchmarking against S&P Index

Conducting a regression test against the S&P Index will help us understand how the banks are performing in contrast to the broader market and how much the performance is impacted by the index.

<u>Australian Banks</u>



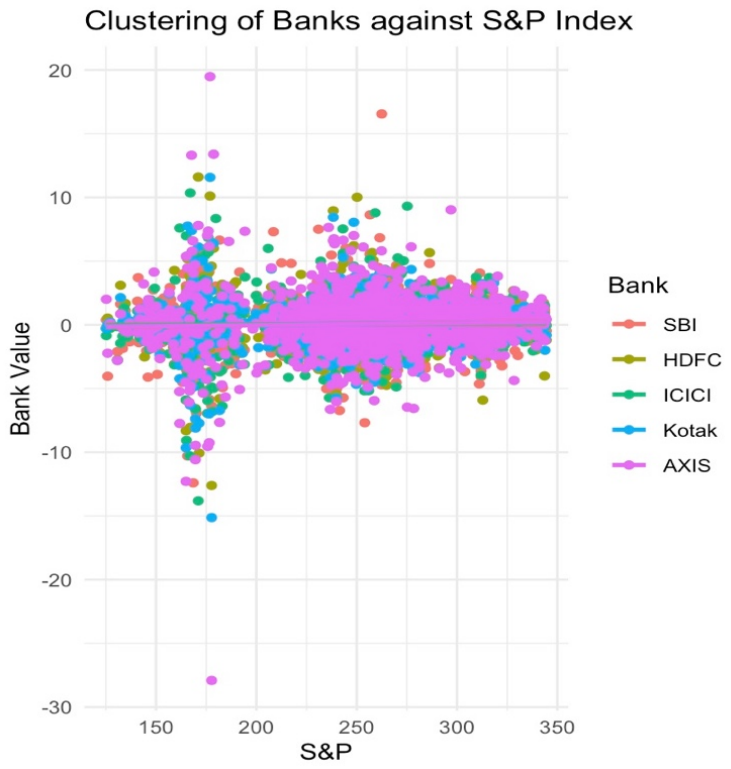
	Bank	Slope	R_squared
sp	Westpac	0.0002495887	0.002138530
sp1	Bendigo	0.0001879371	0.001025347
sp2	CommonWealth	0.0002154602	0.001898943
sp3	NAB	0.0002549528	0.002368677
sp4	anz	0.0001840229	0.001135948

From the above, there a few things we can conclude:

- There is a very weak relationship between the S&P index and performance of the Australian Banks. The low R^2 values indicate that the relationship is weak which just means that the performance of S&P Index barely affects the performance of the banks. This means they are potentially good options for a diversified portfolio.

- There is a positive relationship between the Index and the Banks as given by the positive slope. This means the directionality of returns are the same as that of the Index. The clustering around the middle is also indicative of the same.

Indian Banks



	Bank	Slope	R_squared
SnP	SBI	0.0019163499	0.0025151841
SnP1	HDFC	0.0004623845	0.0001768858
SnP2	ICICI	0.0010871701	0.0008515770
SnP3	Kotak	0.0013881940	0.0017594983
SnP4	AXIS	0.0016047350	0.0011307853

We can draw similar conclusions for India as well;
The above banks we see that all of them have positive slopes and a weak coefficient of determination (R^2). The positive slope indicates that there is a positive relation between the index and returns, we can also see that the data is fairly clustered around the middle which concludes the same. The positive low coefficient of determination (R^2) indicates that there is a weak relationship between the S&P Index with the estimated returns of the banks, i.e. While there is a positive relationship, the outcome of the S&P index does not wholly determine the result of the return.

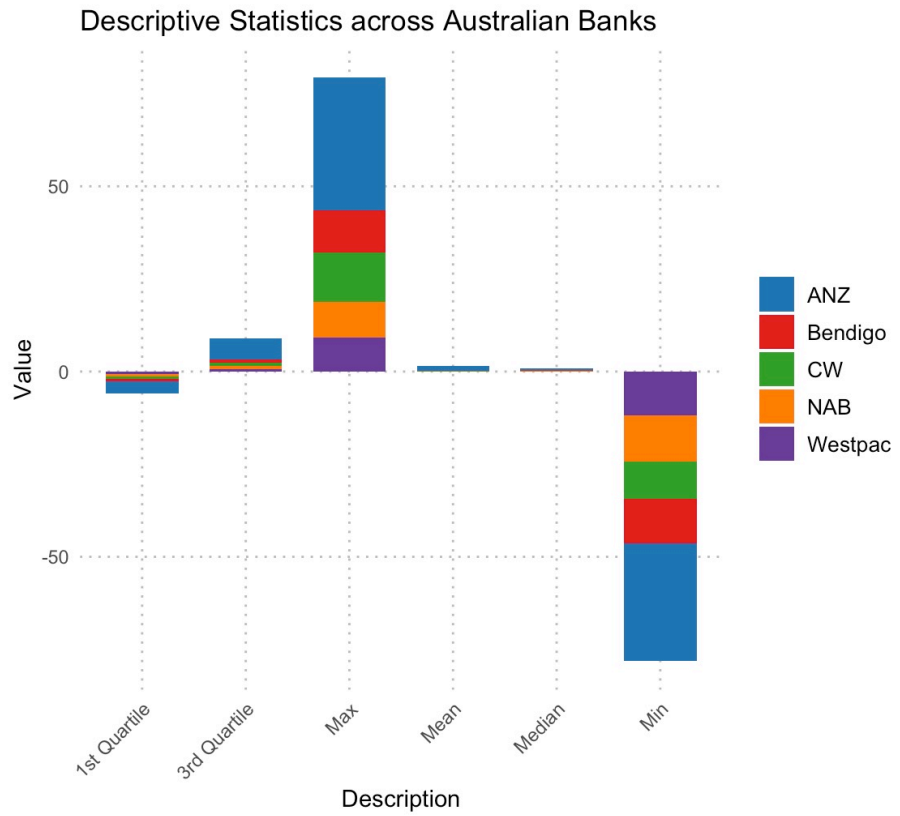
VOLATILITY

It is important to gauge the volatility of returns in order for an investor to make sense of what kind of risk they might be taking on and build a portfolio according to their risk tolerance. Since it is known that there exists a positive correlation between risk and reward there is a definite trade-off decision the investor will have to make. We can better comprehend the volatility by computing the variance and standard deviation of the datasets. We can also take a look at the descriptive statistics to understand the distribution of data better.

Australian Banks

Descriptive Statistics

The use of descriptive statistics will help us draw better comparisons between the banks and how they are performing relatively. Below is a graph showing us the descriptive statistics for Australian Banks using their estimated returns-



From the graph we can infer that ANZ's Estimated return have the widest range and most substantial return amongst the lot. It also has the highest average returns in the group making it a top choice to invest in.

Standard Deviation and Variance

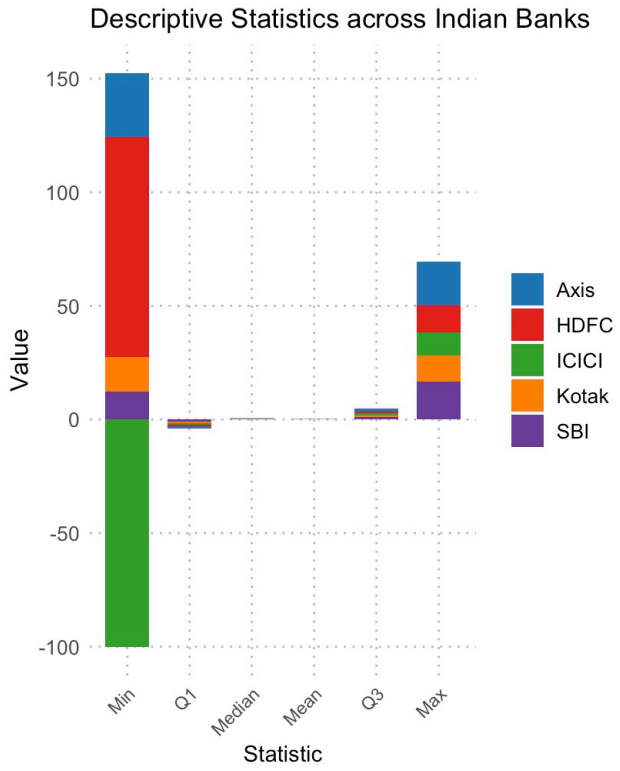
From the above, we observe that Bendigo has the highest variance and that Commonwealth bank has the least. This means that Bendigo's returns experience more fluctuations and can be recommended to risk takers whereas Commonwealth bank has a more consistent performance meant for more stable investors with low-risk appetite.

	Bank	Standard_Deviation	Variance
Westpac	Westpac	1.664177	2.769485
Bendigo	Bendigo	1.809711	3.275054
CommonWealth	CommonWealth	1.524555	2.324269
NAB	NAB	1.615247	2.609024
anz	anz	1.683546	2.834326

Indian Banks

Descriptive Statistics

From below, we can see that Axis bank looks like the best investment since it has both a high minimum and the highest maximum values. Other than this, all banks seem to be performing similarly as seen from their quartiles and mean values.



Standard Deviation and Variance

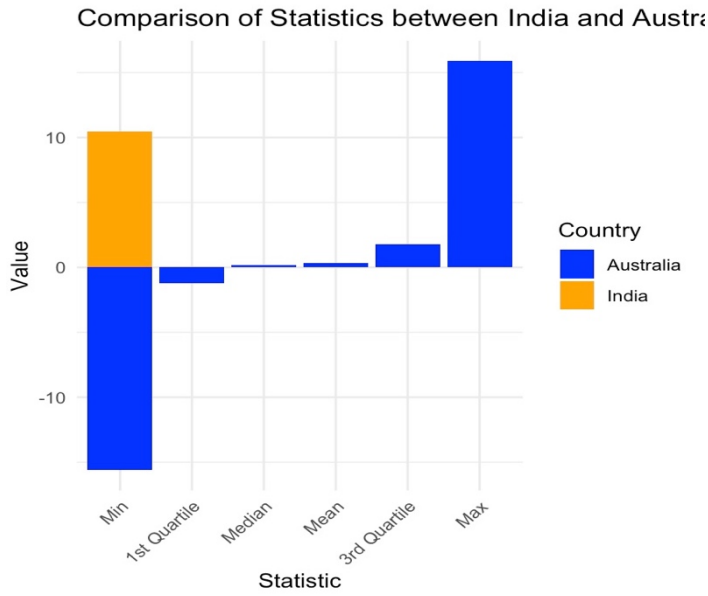
	Bank	Standard_Deviation	Variance
SBI	SBI	1.859734	3.458612
HDFC	HDFC	1.692068	2.863094
ICICI	ICICI	1.813203	3.287706
Kotak	Kotak	1.610708	2.594379
AXIS	AXIS	2.322599	5.394466

From the above, we can say that Axis Bank has the maximum variance and can therefore be recommended to a high risk-taker and Kotak Bank has the least variance and therefore a more stable performance preferred by a low-risk taker.

Comparing Indian and Australian Banks' performances

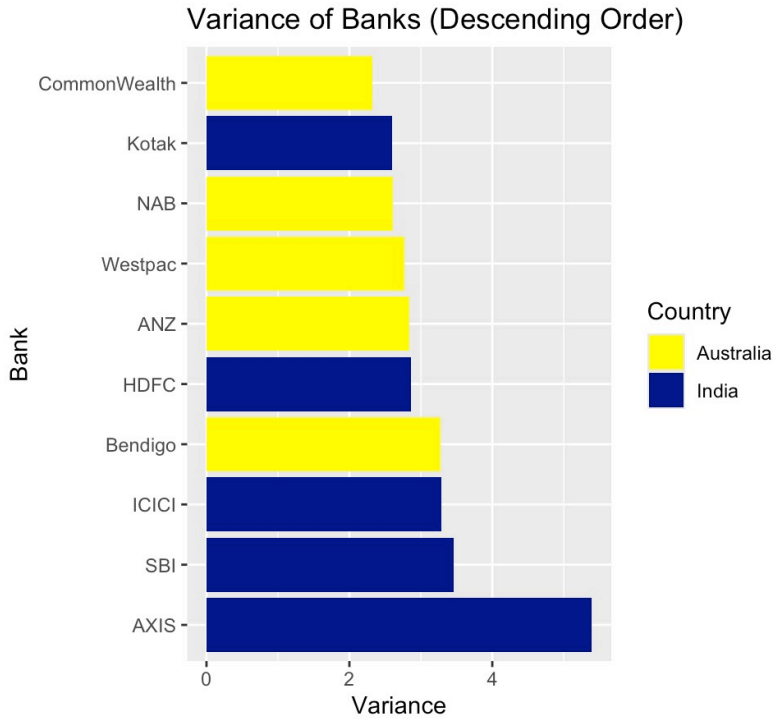
In terms of variance, we see that Indian Banks are having more volatility in comparison to Australian Banks, they also hold more market capital than Australian Banks. Banks of both countries have weak impact on the S&P Index making them a more diverse portfolio option. From the trend analysis, we can also conclude that Australian Banks seem to be at par with their competitors making all their performances more or less similar whereas in India there are clearly more differences between competing banks.

By performing an ANOVA Test, we can further determine and test the null hypothesis that both Country's Performances are related to each other to understand whether investing on them is a more diversified strategy. Also, we can continue to look at descriptive statistics to give us a clearer picture as seen below-




```
                Df Sum Sq Mean Sq F value   Pr(>F)
India            1   44.6   44.56   19.24 1.25e-05 ***
Residuals     1233 2855.3    2.32
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

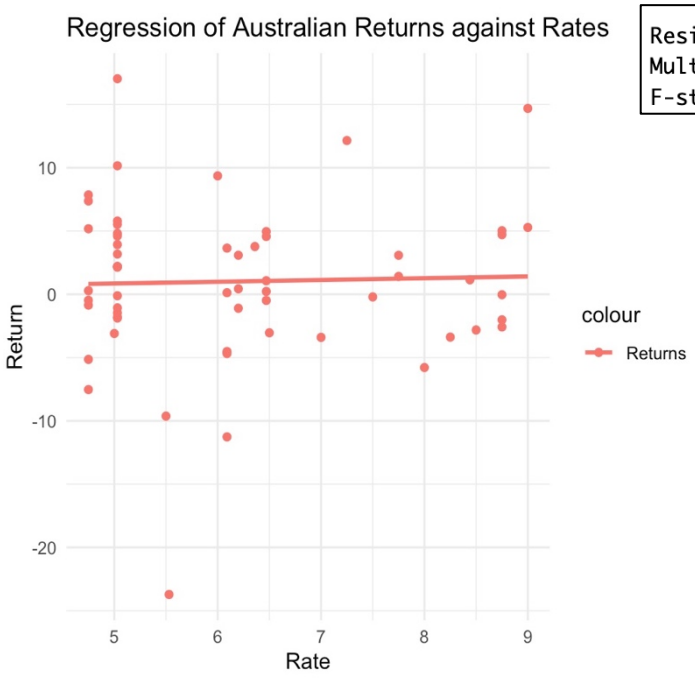
From the above we can say that the degree of freedom between groups India and Australia is 1 and residual is just the difference between the number of observations and number of groups (2). Next, we can see that the sum of squares for the independent variable India is 44.56 which represents the total variability in the data tells us the differences between Australia and India. Post this, the F-statistic is 19.24 which indicates that the differences between groups are statistically significant. The extremely small p value negates the null hypothesis that both groups are similar which basically means that there is a significant difference between them. This makes investing in both diversified. Furthermore, below is a snapshot of the variance amongst banks-



To conclude, while Australian Banks do happen to give us higher returns, Indian Banks are not far behind and therefore a mix of both stocks would make a good portfolio. Australian returns have a starker minimum return that could be offset by having a longer holding period to maximize the margins of return whereas Indian Banks would still give us lesser but substantial returns as well.

Impact of Interest Rates

Australian Banks

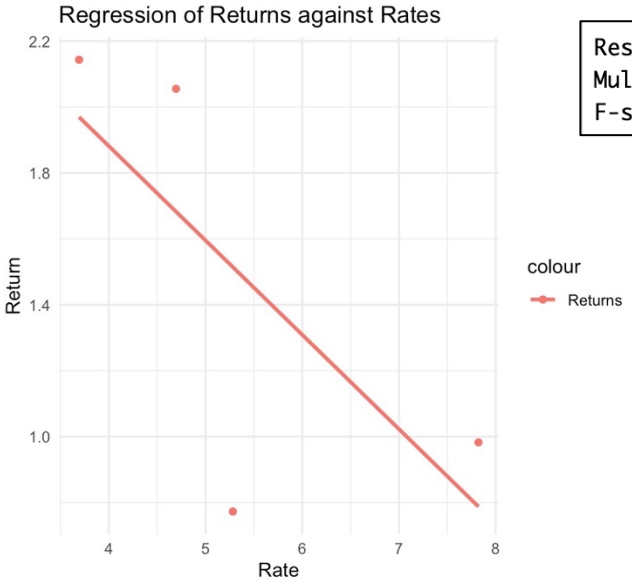



```
Residual standard error: 6.243 on 58 degrees of freedom
Multiple R-squared:  0.001004,  Adjusted R-squared:  -0.01622
F-statistic: 0.0583 on 1 and 58 DF,  p-value: 0.8101
```

```
> print(correlation)
[1] 0.03168791
```

Based on the regression analysis conducted, it appears that there is no significant relationship between the interest rate (Rate) and the stock returns (Return) in the provided dataset. The model's low R-squared values indicate that only a negligible proportion of the variation in returns can be explained by changes in the interest rate. Additionally, the high p-value for the F-statistic suggests that the regression model is not statistically significant, further supporting the conclusion that Rate does not have a significant effect on Return. Therefore, based on this analysis, it does not seem that changes in interest rates can reliably predict stock returns in this context. We also compute a small positive correlation meaning higher interest rates give us more returns.

Indian Banks




```
Residual standard error: 0.6154 on 2 degrees of freedom
Multiple R-squared:  0.5013,    Adjusted R-squared:  0.2519
F-statistic:  2.01 on 1 and 2 DF,  p-value: 0.292
```

The regression analysis suggests that there is a moderate relationship between Rate and Return for the Indian data, with Rate explaining approximately 50.13% of the variance in Return. However, the overall regression model is not statistically significant (F-statistic = 2.01, p-value = 0.292), indicating that the relationship may not be reliably predicted by Rate alone.

CONCLUSION

From the entire report we can conclude that the best portfolio for one to hold would be a mix of both Australian and Indian Bank stocks. Australian Banks have a more consistent performance across its own competitors and gives higher rewards than their Indian counterparts. Finally, both country's' banks do not have any significant relationship between their performances and the S&P Index or the Interest rates of their respective countries.

My final reflection of this project is the deep understanding of the complexity of data and how to visualise and interpret it. I believe this project enabled me to gain more clarity on the same.

REFERENCES

(n.d.). *Previous Next Top 5 Best Banks in Australia for 2024*. TaxLeopard. https://taxleopard.com.au/top-5-best-banks-in-australia/

Segal, T. (2022, October 29). *Is There a Positive Correlation Between Risk and Return?* Investopedia. https://www.investopedia.com/ask/answers/040715/there-positive-correlation-between-risk-and-return.asp#:~:text=key%20takeaways,of%20uncertainty%20with%20high%20returns.

(2024, April 23). *Largest Australian (ASX) Banks Stocks by Market Cap*. SimplyWallSt. https://simplywall.st/stocks/au/banks/market-cap-large

(2024, April 1). *Top 10 banks in India by market cap in 2024*. Forbes India. https://www.forbesindia.com/article/explainers/top-10-banks-india-by-market-cap/87913/1

(n.d.). *Rate of Return*. CFI. https://corporatefinanceinstitute.com/resources/valuation/rate-of-return-guide/

(n.d.). *S&P/ASX 200 Banks (Industry)*. S&P Dow Jones Indices. https://www.spglobal.com/spdji/en/indices/equity/sp-asx-200-banks-industry/#overview

(n.d.). *Statistical Tables*. Reserve Bank of Australia. https://www.rba.gov.au/statistics/tables/#interest-rates

(n.d.). *Database on Indian Economy*. Reserve Bank of India. https://cimsdbie.rbi.org.in/DBIE/#/

(n.d.). *Stock Prices*. Bombay Stock Exchange. https://www.bseindia.com/markets/equity/EQReports/StockPrcHistori.aspx?expandable=7&scripcode=532174&flag=sp&Submit=G

(n.d.). Yahoo Finance. https://au.finance.yahoo.com/lookup

APPENDIX

```r
# Load necessary packages

library(readxl)
library(ggplot2)

# Read data from Excel file
data <- read_excel("ind.xlsx", sheet = "Performance_Monthly")

# Reshape the data from wide to long format
library(reshape2)
data_long <- melt(data, id.vars = "Date", variable.name = "Bank", value.name = "Performance")

# Create line plot using ggplot2
ggplot(data_long, aes(x = Date, y = Performance, color = Bank)) +
  geom_line() +
  labs(title = "Indian Bank Performance Trends Over Time",
       x = "Date", y = "Performance") +
  theme_minimal()
```

Performance Line graph Code

```r
library(readxl)

# Read data from Excel file
data <- read_excel("aus.xlsx", sheet = "countrycompare")

# Perform ANOVA
anova_result <- aov(Australia ~ India, data = data)

# Print ANOVA summary
summary(anova_result)
```

ANOVA Code

```r
# Load necessary packages

library(readxl)
library(ggplot2)

# Read data from Excel file
data <- read_excel("ind.xlsx", sheet = "compare")


# Select data for India and Australia
india_data <- data[data$Country == "India", ]
aus_data <- data[data$Country == "Australia", ]

# Reshape the data from wide to long format
library(reshape2)
india_data_long <- melt(india_data, id.vars = "Country", variable.name = "Statistic", value.name = "Value")
aus_data_long <- melt(aus_data, id.vars = "Country", variable.name = "Statistic", value.name = "Value")

# Create stacked bar plot using ggplot2
ggplot() +
  geom_bar(data = india_data_long, aes(x = Statistic, y = Value, fill = Country), stat = "identity") +
  geom_bar(data = aus_data_long, aes(x = Statistic, y = Value, fill = Country), stat = "identity", position = "dodge") +
  labs(title = "Comparison of Statistics between India and Australia",
       x = "Statistic", y = "Value") +
  scale_fill_manual(values = c("India" = "orange", "Australia" = "blue")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

Stacked Bar Graph

```r
# Load necessary packages

library(readxl)
library(ggplot2)

# Read data from Excel file
data <- read_excel("aus.xlsx", sheet = "stats")

# Reshape the data from wide to long format
library(reshape2)
data_long <- melt(data, id.vars = "Description", variable.name = "Bank", value.name = "Value")

# Create stacked bar plot with beautifications
ggplot(data_long, aes(x = Description, y = Value, fill = Bank)) +
  geom_bar(stat = "identity", position = "stack", width = 0.7) +
  labs(title = "Descriptive Statistics across Australian Banks",
       x = "Description", y = "Value") +
  scale_fill_manual(values = c("ANZ" = "#1f78b4", "Bendigo" = "#e31a1c", "CW" = "#33a02c", "NAB" = "#ff7f00", "Westpac" = "#6a3d9d")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        panel.grid.major = element_line(color = "gray", linetype = "dotted"),
        panel.grid.minor = element_blank(),
        legend.position = "right", # Move legend to the right
        legend.title = element_blank(), # Remove legend title
        legend.text = element_text(size = 10)) # Adjust legend text size

```

Stacked Bar Graph Code

```r
# Load necessary packages

library(readxl)
library(ggplot2)

# Read data from Excel file
regression_data <- read_excel("ind.xlsx", sheet = "regression")

# Reshape the data from wide to long format
library(reshape2)
regression_data_long <- melt(regression_data, id.vars = "SnP", variable.name = "Bank", value.name = "Value")

# Create scatter plot for each bank against SnP
ggplot(regression_data_long, aes(x = SnP, y = Value, color = Bank)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Clustering of Banks against S&P Index",
       x = "S&P", y = "Bank Value") +
  theme_minimal() +
  theme(legend.position = "right")

```

Scatter Plot Code

```r
# Load necessary packages
library(readxl)

# Read data from Excel file
regression_data <- read_excel("ind.xlsx", sheet = "regression")

# Extract the columns for SnP and banks
SnP <- regression_data$SnP
banks <- subset(regression_data, select = -c(SnP))

# Create an empty dataframe to store results
results <- data.frame(Bank = character(), Slope = numeric(), R_squared = numeric(), stringsAsFactors = FALSE)

# Loop through each bank
for (col in colnames(banks)) {
  # Fit linear regression model
  model <- lm(banks[[col]] ~ SnP)

  # Extract coefficients
  slope <- coef(model)[2]

  # Calculate R-squared
  R_squared <- summary(model)$r.squared

  # Append results to dataframe
  results <- rbind(results, data.frame(Bank = col, Slope = slope, R_squared = R_squared))
}

# Print results
print(results)
```

Regression Code

```r
# Load necessary packages
library(readxl)

# Read data from Excel file
regression_data <- read_excel("ind.xlsx", sheet = "regression")

# Exclude SnP column
banks_data <- subset(regression_data, select = -c(SnP))

# Calculate standard deviation
std_dev <- apply(banks_data, 2, sd)

# Calculate variance
variance <- apply(banks_data, 2, var)

# Print results
result <- data.frame(Bank = colnames(banks_data), Standard_Deviation = std_dev, Variance = variance)
print(result)
```

Standard Deviation and Variance code

```r
# Install and load required packages

library(readxl)

# Load the data from the Excel file
data <- read_excel("aus.xlsx", sheet = "interest")

# Calculate the correlation coefficient between Rate and Return
correlation <- cor(data$Rate, data$Return)

# Print the correlation coefficient
print(correlation)
```

Correlation