Global Loan Dynamics: Analyzing Regional Disparities in IBRD Financial Commitments

Amirou Sanoussy

2024-08-07

1 Introduction

For this project, I wanted to explore a dataset that combined my experiences and interests. As a third-culture kid, I wanted a dataset that incorporated information on a global, regional, or national level while also allowing me to delve into my growing interest in economics.

While scouring the web for suitable data, I stumbled upon a dataset from the International Bank for Reconstruction and Development (IBRD), a department of the World Bank. What initially caught my attention was the connection to a previous project I worked on during my internship at the American Embassy in Nairobi. There, I assisted an Economic Officer with a report highlighting the importance and impact of international loans on developing nations.

This experience showed how crucial these loans are for nations, especially those of third-world status, and it motivated me to see what further insights I could uncover from this dataset. Through this analysis, I seek to explore the IBRD's financial activities and commitments, focusing on critical variables such as loan amounts, interest rates, and regional distributions.

Research Question

The primary research quesiton guiding this analysis is: **What are the differences in loan amounts and types across different regions? Which regions receive the most significant financial commitments, and what factors might explain these disparities?**

2 Description of the Data

The data set used for this project is sourced from the World Bank's The International Bank for Reconstruction and Development (IBRD) Statement of Loans - Latest Available Snapshot. It includes detailed information on loans issued by the IBRD. The Data corresponds to our scope of economics.

Key features of the data include:

Column_Name	Description	Data_Type	Example_Values
Country	The country receiving the loan.	String	"Kenya", "Brazil", "India"
Project ID	Unique identifier for each loan project.	String/Integer	"P123456", "P789012"
Loan Type	Type of loan issued.	String	"Investment", "Policy Loan"
Approval Date	Date the loan was approved.	Date	"2023-07-15", "2019-03-22"
Currency of Commitment	Currency in which the loan is issued.	String	"USD", "EUR", "JPY"
Total Commitment Amount	Total amount committed for the loan.	Numeric	50000000, 150000000
Loan Status	Current status of the loan (e.g., active, closed).	String	"Active", "Closed", "Disbursing"

This data helps in analyzing global financial assistance provided by the IBRD, understanding loan distribution, and identifying trends in international development finance.

3 Methods

In this section, we describe the methods used to analyze the differences in loan amounts and types across different regions and identify the areas that receive the most significant financial commitments. Our investigation aims to uncover patterns and insights that can explain these disparities.

My initial step when processing the dataset was cleaning and processing the data. This included converting date columns to appropriate date formats, handling missing values, and ensuring consistency across the dataset. Specifically, we assigned missing values for the "Interest Rate" column with the mean of existing values.

3.1 Exploratory Data Analysis and Method Justification

Following, we conducted an Exploratory Data Analysis (EDA) to understand the basic characteristics of the data. This involved generating summary statistics and visualizations to identify trends, outliers, and potential relationships between different variables.

In order to produce a comprehensive overview of the dataset, we calculated the aggregate statistics for key variables. These statistics help justify the choice of our methods and provide a basis for further analysis as shown in Table 1.

Table 2: Aggregated Statistics by Region

Region	Total_Loan_Amount	Avg_Loan_Amount	Median_Loan_Amount	Loan_Count
EAST ASIA AND PACIFIC	193663865849	97760659	47100000	1981
EASTERN AND SOUTHERN AFRICA	23641891341	57383231	12992500	412
EUROPE AND CENTRAL ASIA	191354548570	103044991	40410000	1857
LATIN AMERICA AND CARIBBEAN	278106992333	96800206	37000000	2873
MIDDLE EAST AND NORTH AFRICA	94126024000	86274999	36000000	1091
SOUTH ASIA	97527315633	171703020	108500000	568
WESTERN AND CENTRAL AFRICA	18072400000	44403931	16804719	407

By examining Table 1, we can derive specific understandings to guide the creation of bivariate plots. The total loan amount reveals that Latin America and the Caribbean received the highest total loan amount ($278,106,992,333), followed by East Asia and the Pacific ($193,663,865,849) and Europe and Central Asia ($191,354,548,570). These highest total loan amounts indicate that these regions have the most significant financial commitments from the IBRD. When considering the average loan amount, South Asia has the highest figure ($171,703,020), suggesting larger individual loans than other regions. The median loan amounts highlight the varying distribution of loans within regions; for instance, South Asia has a median loan amount of $108,500,000, demonstrating a high level of loan commitment consistency within the region. Additionally, the loan count shows that Latin America and the Caribbean have the highest number of loans (2873), indicating frequent financial engagements in this region. These disparities guided our decision to use specific visualizations to explore these patterns further.

The first visualization, a scatter plot of the original principal amount versus the amount repaid to IBRD by loan status, was chosen to understand repayment dynamics and potential issues in different loan categories. This was particularly relevant given the variation in loan amounts and repayment rates across regions. The second visualization, a box plot of interest rates by region, is a key tool in highlighting regional variations in borrowing costs, which could explain some of the differences in loan amounts and repayment behavior. Lastly, we created a correlation heat map of key financial variables to identify significant relationships within the data, providing a comprehensive view of how different financial metrics interact across regions.

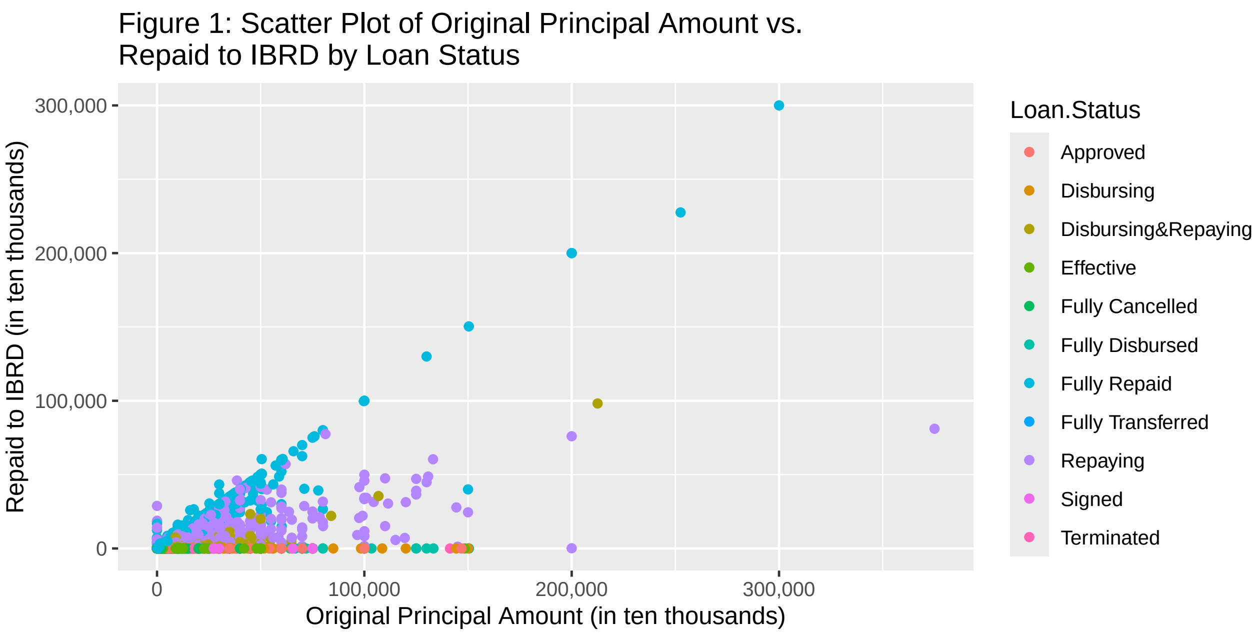
Additionally, we applied statistical learning methods such as K-means clustering to identify natural groupings in the data, leveraging the regional differences in loan amounts and counts. Principal Component Analysis (PCA) was used for dimensional reduction, simplifying the dataset while preserving its essential patterns.

These methods were selected for their ability to discover meaningful patterns and insights from the dataset, ultimately providing a deeper understanding of the differences in loan amounts and types across regions.

4 Results and Discussion

Scatter Plot of Original Principal Amount vs. Repaid to IBRD by Loan Status

We first create a Scatter Plot of Original Principa Amount vs. Repaid to IRBD by Loan Status which shows the relationship between the original principal amount and the amount repaid to IBRD, which is categorized by loan status as seen in Figure 1 below:



From Figure 1, the plot reveals distinct repayment patterns across different loan statuses. Fully repaid loans generally align along the diagonal, showing a direct relationship between the principal and repaid amounts. However, active loans and those with partial repayment show more variability, suggesting differences in repayment schedules and financial management practices. The variability in repayment patterns and loan statuses can provide insights into how regions manage their financial commitments. Regions with more fully repaid loans might indicate better financial management and repayment practices. In contrast, regions with more active and partially repaid loans could suggest challenges in sticking to repayment schedules, possibly due to economic instability or differing financial practices. By understanding these patterns, we can better explain the differences in loan amounts and types across regions and identify the factors contributing to these differences.

To explore these differences further, it is crucial to consider the borrowing costs associated with these loans. Interest rates play a significant role in the cost of borrowing and can impact a region's ability to repay loans. The following box plot compares the distribution of interest rates across different regions, highlighting regional variations in borrowing costs.

4.1 Box Plot of Interest Rate by Region

We then created a boxplot in Figure 2, that compares the distribution of interest rates across different regions, highlighting regional variations in borrowing costs.

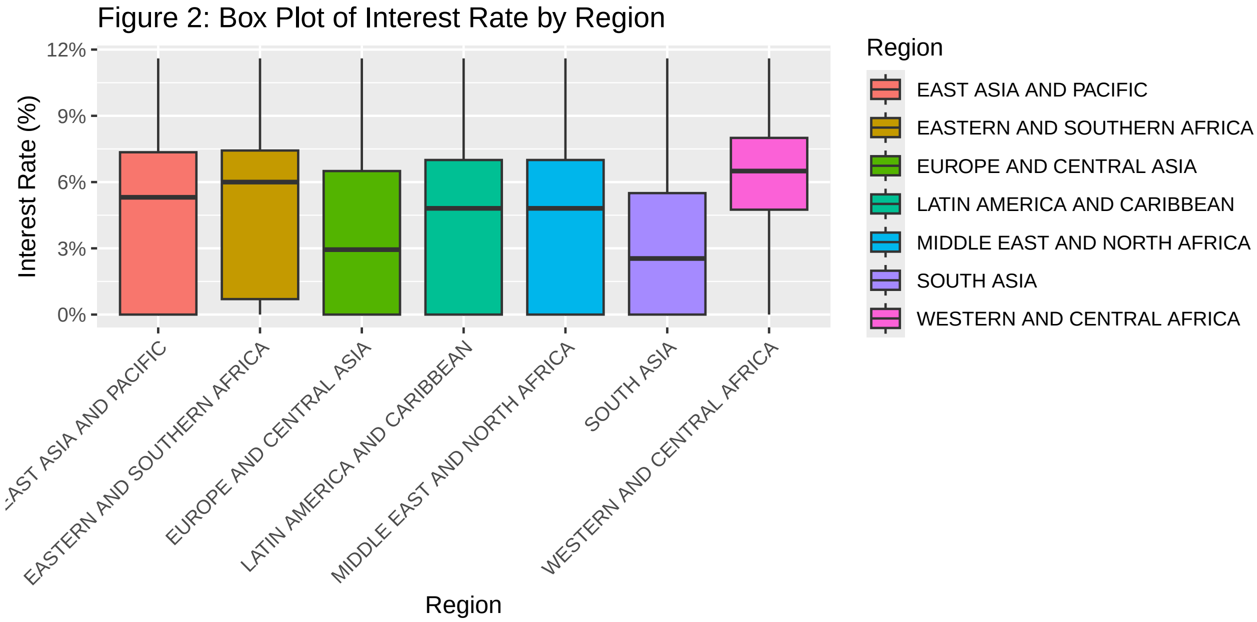


Figure 2 shows significant regional differences in interest rates. South Asia and Western and Central Africa regions have higher median interest rates than others, which makes sense in the grand scheme of things as they are region with many nations that are considered to be developing nations. This variation can influence the total cost of borrowing and repayment schedules. Higher interest rates can increase the financial burden on borrowers, potentially leading to difficulties in loan repayment and increased variability in repayment patterns, as observed in the scatter plot.

Understanding these differences is essential for evaluating the financial burden on borrowers and the sustainability of loan repayments in different regions. The Box plot suggests regions with higher interest rates may face more significant challenges in managing their financial commitments, which varies for per nation in those respective regions. The challenges of managing their financial obligations can further explain the disparities in loan amounts and types across different areas as they affect the credential reputation of nations' trustworthiness in paying back their debts.

To delve deeper into the underlying relationships between various financial variables, we can examine the correlations between key financial metrics. The following heatmap effectively visualizes these correlations and provides exciting insights into how different financial aspects interact.

4.2 Correlation Heatmap of Key Financial Variables

We then made a heatmap in Figure 3 to visualize the correlations between key financial variables, providing insights into significant relationships within the data.

Figure 3: Correlation Heatmap of Key Financial Variables

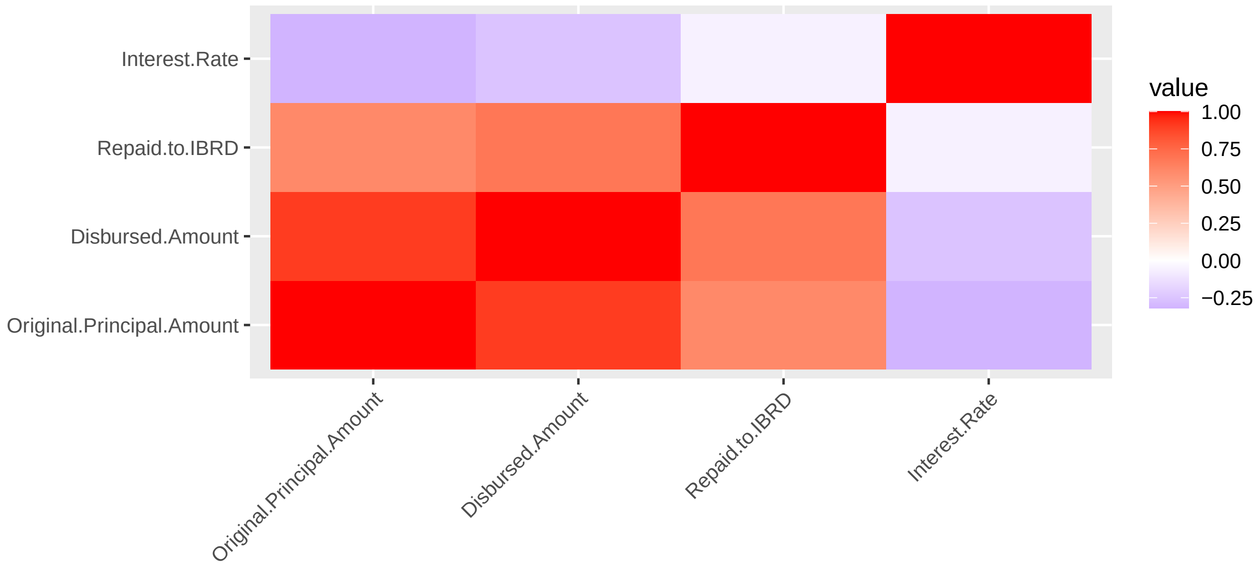


Figure 3 highlights strong positive correlations between the original principal amount and the disbursed amount and between the original principal amount and the repaid amount to IBRD. These relationships indicate that higher principal amounts generally lead to higher disbursements and repayments. This pattern suggests that larger loans result in more substantial financial commitments and repayments, consistent across the dataset.

However, the interest rate exhibits weaker correlations with other financial variables, implying that factors other than loan size significantly influence interest rates. These could include the borrowing country's economic stability, the borrower's creditworthiness, and macroeconomic conditions at the time of the loan agreement. This complexity underscores that interest rates are influenced by a diverse set of factors beyond just the loan amount.

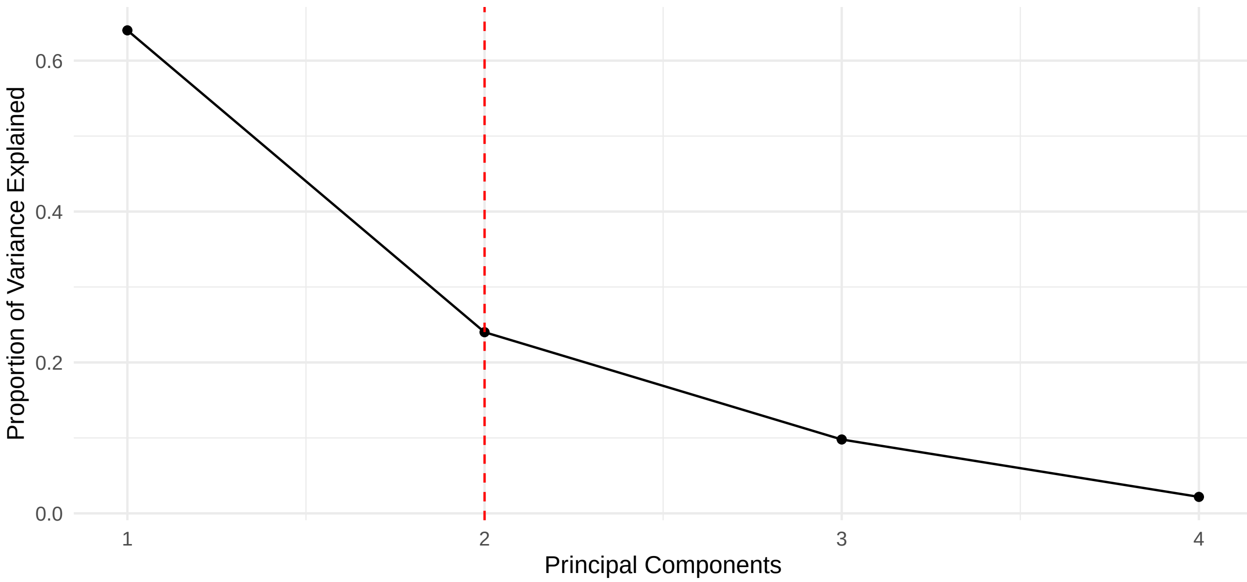
Relating this to the research question, the correlation analysis provides valuable understanding. Higher principal amounts lead to higher disbursements and repayments, indicating a direct relationship between the size of the loan and the financial commitments involved by region. The direct relationship helps to understand which regions receive significant financial support and potentially manage large-scale financial projects. The weaker correlations involving interest rates suggest that borrowing costs are influenced by multiple factors, meaning regions with similar loan sizes may face different borrowing costs. This complexity impacts a region's ability to manage and repay these loans, making it important to understand these influences to explain regional disparities in loan amounts and types.

The heatmap helps show the dataset's basic financial dynamics and interactions, providing a clearer picture of how different financial variables relate.

4.3 Principal Component Analysis (PCA)

We then use PCA to to reduce the dimensionality of the dataset and identify the key components contributing to the variance. In Figure 4 we show a scree plot below shows the proportion of variance explained by each principal component.

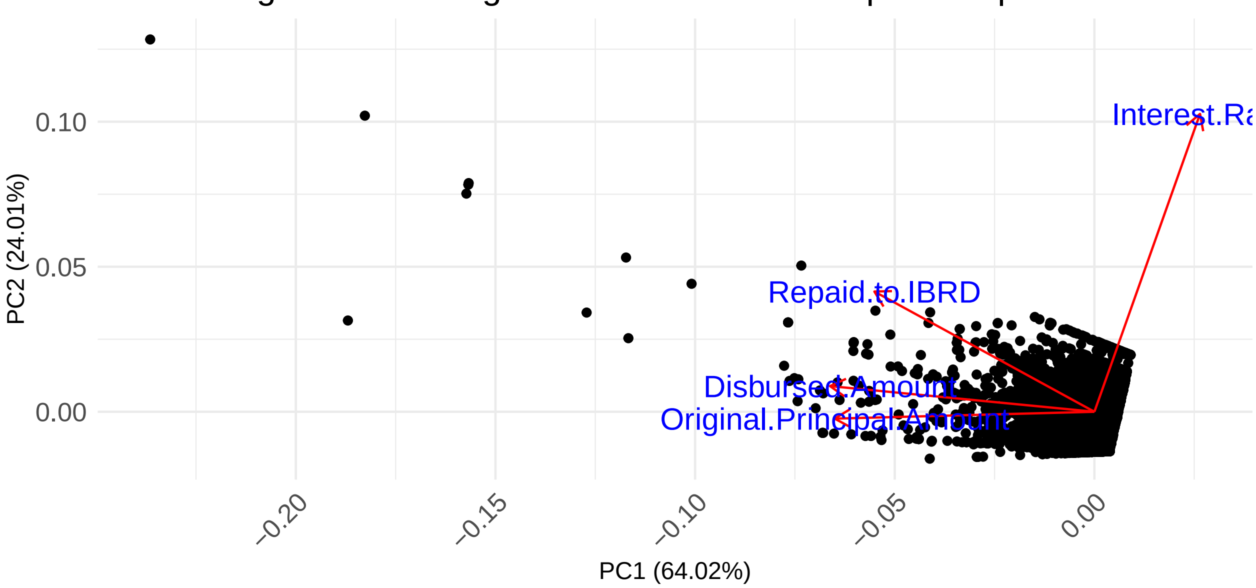
Figure 4: Scree Plot



The scree plot reveals that the first two principal components explain a significant portion of the variance in the IBRD dataset. The elbow point helps identify that the first two components are sufficient to capture most of the variance, simplifying the analysis by focusing on these principal components.

We then make a loading plot for the first two principal components as they provide insights into the variable that most contributes to these components as seen in Figure 5.

Figure 5: Loadings of the First Two Principal Components



As we can see from Figure 5, both components explain approximately 88.06% of the dataset's variability. The loading reveals that the primary contributor to PC1 is "Original.Principal.Amount" with a high negative loading. This suggests that PC1 primarily captures the variance related to the original loan amounts. Regions with higher values on PC1 are likely to have larger loan sizes. Also it reveals that the primary contributor to PC2 is "Interest. Rate" with a high positive loading. This indicates that PC2 captures the variance related to the loans' interest rates. Regions with higher values on PC2 are associated with higher borrowing costs.

Combining the insights from both the scree plot and the loadings plot, we can draw meaningful conclusions that relate to the research question: "What are the differences in loan amounts and types across different regions? Which regions receive the most significant financial commitments, and what factors might explain these disparities?"

The scree plot validates our focus on the first two principal components, simplifying the study while retaining the majority of the variance in the data. By examining the loadings plot, we identify that the original principal amount and interest rates are the key variables driving the differences across regions.

The plot analysis through PCA has highlighted two critical factors: Loan Size (PC1) and Borrowing Costs (PC2). Regions with higher original principal amounts manage larger financial projects, corresponding to significant financial support. This suggests that the distribution of economic resources is heavily influenced by the scale of the projects undertaken in different regions. Regions with larger loan sizes are likely to receive more significant financial commitments, providing crucial insights into understanding regional disparities in loan amounts. On the other hand, the interest rates significantly impact borrowing costs, which vary across regions. Higher interest rates are associated with increased financial burdens on borrowers, affecting their ability to manage and repay loans. This variation in borrowing costs can explain why some regions face more loan repayment and management challenges despite having similar loan sizes.

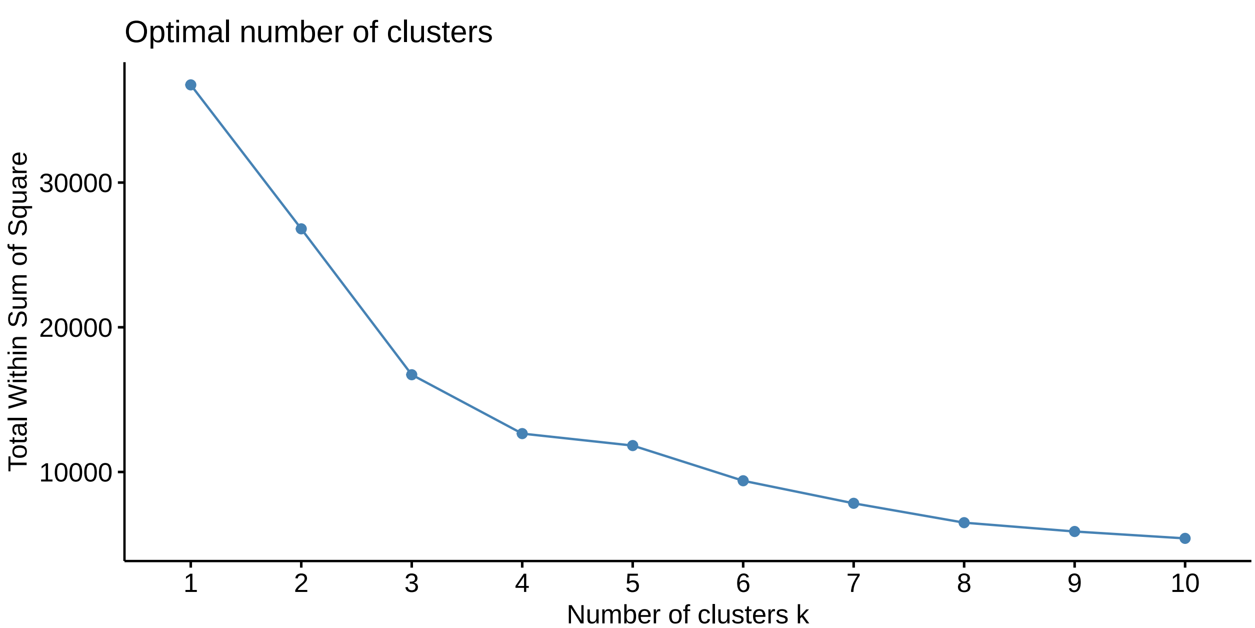
By focusing on these principal components, we target the key factors affecting loan distribution and repayment across regions. This understanding can guide policies to address financial disparities and ensure regions with higher financial burdens receive appropriate support.

4.4 K-Means Clustering

Having found the key variables that explain the most variance in our dataset, we can now transition to clustering analysis. The principal components derived from the PCA serve as an ideal basis for clustering, as they lower the complexity of the data while retaining the most essential information.

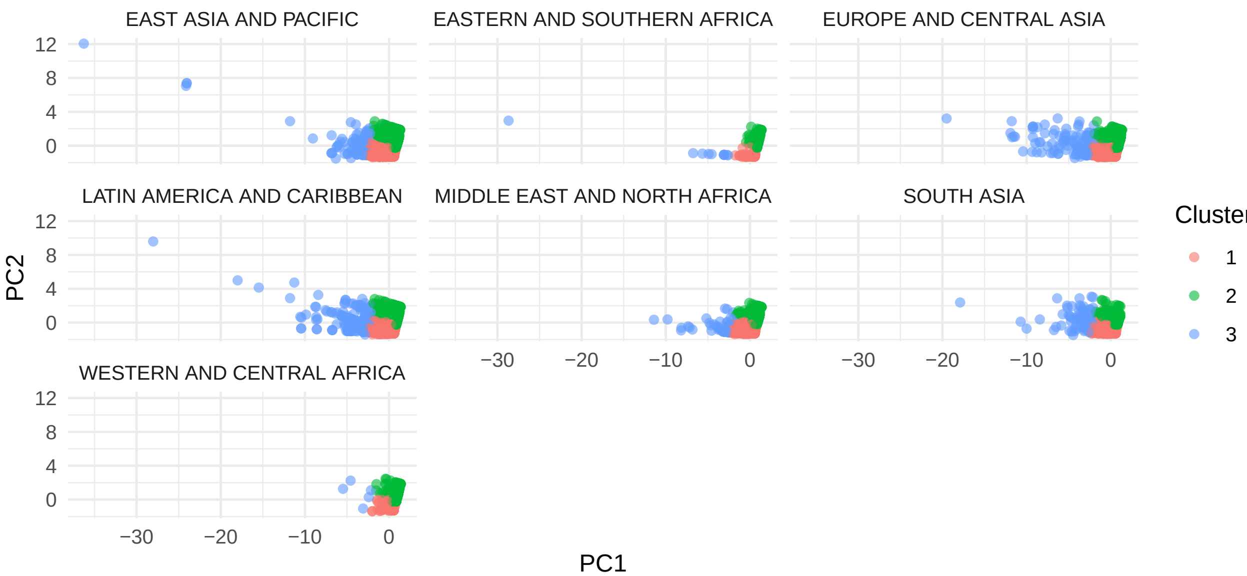
Clustering analysis, with the use of principal components, will allow us to group countries with similar loan characteristics. This approach ensures that the clusters are formed based on the most significant factors identified in our PCA, providing further insights into the patterns and disparities observed across regions.

To determine the optimal number of clusters, we will use the elbow method, which involves plotting the within-cluster sum of squares (WSS) against the number of clusters and looking for an "elbow" point where the rate of decrease sharply slows as seen in Figure 4.



The elbow point suggest that three clusters are optimal for grouping the countries based on their loan characteristics because it is at three where the plot shows a significant bend. We will proceed with applying K-means clustering of loan data by regions.

Figure 6: K−means Clustering of Loan Data by Region



The facet plot shows how loan data clusters across different regions based on the first two principal components (PC1 and PC2). Each area is displayed in its panel, with points colored by their assigned cluster. This layout helps us see patterns within each region and compare loan characteristics across different areas.

Cluster 1 (Red) is a significant cluster that includes loans with distinct features, such as higher original principal amounts or different repayment terms. The fact that Cluster 1 appears in most regions strongly indicates that these loan characteristics are not region-specific but common globally. Cluster 2 (Green) is spread widely across regions, representing loans with moderate traits that might be closer to average in terms of size, interest rates, or other financial variables. This cluster is especially prominent in regions like Eastern and Southern Africa, Latin America, and the Caribbean, indicating that these regions have more loans with balanced terms. Although less common, cluster 3 (Blue) includes loans with smaller principal amounts or more favorable terms, such as lower interest rates or shorter repayment periods. The presence of Cluster 3 in several regions suggests that some areas might have access to more favorable loan conditions.

When we look at regional patterns, the East Asia and Pacific region stand out with a diverse loan portfolio. This is clear from the strong presence of Cluster 1 and Cluster 2, indicating a mix of more significant and moderate loans. This diversity suggests a range of financial commitments, from substantial to more manageable loans. In Eastern and Southern Africa, the dominance of Cluster 2 indicates that loans here are generally moderate in size and terms, suggesting a balance between financial commitments and the region's economic capacity. Latin America and the Caribbean also show a significant presence of Cluster 1 and Cluster 2, pointing to a mix of large and moderate loans.

In Europe and Central Asia, the spread of clusters suggests a balanced loan portfolio with a mix of high, moderate, and low financial commitments. The Middle East and North Africa region show a wide range of loan sizes and terms, as reflected in the presence of all three clusters. South Asia mainly features loans in Cluster 2, indicating moderate financial commitments with balanced characteristics. Lastly, Western and Central Africa primarily handles moderate loans, as seen in the dominance of Cluster 2, with some loans falling into Cluster 1 and Cluster 3.

5 Conclusion

This analysis clearly explains the differences in loan amounts and types across various regions and the factors driving these differences. The results reveal significant regional differences in loan amounts. Regions like Latin America and the Caribbean, East Asia and the Pacific, Europe and Central Asia received the most important financial commitments, as shown by their high original principal amounts. These regions manage larger financial projects, attracting substantial loans from the IBRD. In contrast, while receiving fewer loans overall, areas such as South Asia and Western and Central Africa often deal with larger individual loans on average. The regions receiving lower loans suggest that these areas require targeted financial interventions for specific large-scale projects.

Two key factors explain these regional disparities: loan size (PC1) and borrowing costs (PC2). Regions with larger loans tend to manage more extensive financial projects, indicating that the scale of a region's projects significantly influences how economic resources are distributed. This emphasizes the need for financial support that aligns with the specific needs of regions undertaking large-scale development projects. Borrowing costs, reflected in interest rates (PC2), vary widely across regions and play a crucial role in shaping the financial burden on borrowers. Higher interest rates, especially in areas like South Asia and Western and Central Africa, increase the difficulty of managing and repaying loans. These variations in borrowing costs add to some regions' challenges, even when loan sizes are similar to those in other areas.

Based on these findings, the IBRD should adopt a more tailored loan distribution and management approach. Regions managing larger projects should continue to receive substantial financial support, with careful monitoring to ensure funds are used effectively. For areas with higher borrowing costs, the IBRD should consider more favorable loan terms, such as lower interest rates or alternative financing options, to alleviate financial strain and improve loan repayment capabilities.

The clustering analysis further emphasizes that regional loan profiles are unique, suggesting that region-specific strategies are needed. Regions with a high presence of large, distinct loans (Cluster 1) would benefit from targeted interventions to support these significant commitments. Regions dominated by moderate loan profiles (Cluster 2) might require policies that balance financial support with sustainable repayment terms. Finally, for areas with access to more favorable loan conditions (Cluster 3), the IBRD could explore more innovative financial products to enhance their development opportunities.

In conclusion, handling the disparities in loan amounts and borrowing costs across regions calls for tailored financial strategies that consider each region's unique needs and challenges. By focusing on the critical factors identified in this report, the IBRD can guarantee a more fair and effective distribution of financial resources, promoting sustainable development across all regions.

Citations

Source of IBRD Dataset: https://finances.worldbank.org/Loans-and-Credits/IBRD-Statement-of-Loans-Latest-Available-Snapshot/sfv5-tf7p/about_data

Source of Research: https://www.worldbank.org/en/news/feature/2012/07/26/getting_to_know_theworldbank#:~:text=The%20Bank%20lends%20only%20a,all%2C%20of%20the%20project%20itself.

Appendix

```r
# Data Preprocessing: Converting date columns to the appropriate date formats
sol$End.of.Period <- as.Date(sol$End.of.Period, format = "%m/%d/%Y %I:%M:%S %p")
sol$First.Repayment.Date <- as.Date(sol$First.Repayment.Date, format = "%m/%d/%Y %I:%M:%S %p")
sol$Last.Repayment.Date <- as.Date(sol$Last.Repayment.Date, format = "%m/%d/%Y %I:%M:%S %p")
sol$Agreement.Signing.Date <- as.Date(sol$Agreement.Signing.Date, format = "%m/%d/%Y %I:%M:%S %p")
sol$Board.Approval.Date <- as.Date(sol$Board.Approval.Date, format = "%m/%d/%Y %I:%M:%S %p")
sol$Effective.Date..Most.Recent. <- as.Date(sol$Effective.Date..Most.Recent., format = "%m/%d/%Y %I:%M:
sol$Closed.Date..Most.Recent. <- as.Date(sol$Closed.Date..Most.Recent., format = "%m/%d/%Y %I:%M:%S %p")
sol$Last.Disbursement.Date <- as.Date(sol$Last.Disbursement.Date, format = "%m/%d/%Y %I:%M:%S %p")

# Handling missing values: Replacing NA values in 'Country.Code' and 'Guarantor.Country.Code' with "Unk
sol$Country.Code[is.na(sol$Country.Code)] <- "Unknown"
sol$Guarantor.Country.Code[is.na(sol$Guarantor.Country.Code)] <- "Unknown"

# Handling missing values: Assigning the mean value to NA values in the 'Interest.Rate' column
sol$Interest.Rate[is.na(sol$Interest.Rate)] <- mean(sol$Interest.Rate, na.rm = TRUE)

# Aggregating statistics by region: Calculating total, average,
and median loan amounts, along with the loan count per region
region_stats <- sol %>%
  group_by(Region) %>%
  summarise(
    Total_Loan_Amount = sum(`Original.Principal.Amount`, na.rm = TRUE),
    Avg_Loan_Amount = mean(`Original.Principal.Amount`, na.rm = TRUE),
    Median_Loan_Amount = median(`Original.Principal.Amount`, na.rm = TRUE),
    Loan_Count = n()
  )

# Displaying the aggregated statistics in a table format
kable(region_stats, caption = "Table 1: Aggregated Statistics by Region")

# Creating a scatter plot: Visualizing the relationship between the original principal amount and the a
ggplot(sol, aes(x = Original.Principal.Amount, y = Repaid.to.IBRD, color = Loan.Status)) +
  geom_point() +
  ggtitle("Figure 1: Scatter Plot of Original Principal Amount vs. \nRepaid to IBRD by Loan Status") +
  xlab("Original Principal Amount (in ten thousands)") +
  ylab("Repaid to IBRD (in ten thousands)") +
  scale_x_continuous(labels = scales::comma_format(scale = 0.0001)) +
  scale_y_continuous(labels = scales::comma_format(scale = 0.0001))

# Creating a box plot: Comparing the distribution of interest rates across different regions to highlig
ggplot(sol, aes(x = Region, y = Interest.Rate, fill = Region)) +
  geom_boxplot() +
  ggtitle("Figure 2: Box Plot of Interest Rate by Region") +
```

```
  xlab("Region") +
  ylab("Interest Rate (%)") +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Selecting key financial variables for correlation analysis
financial_vars <- sol %>% select(Original.Principal.Amount, Disbursed.Amount, Repaid.to.IBRD, Interest.

# Calculating the correlation matrix
cor_matrix <- cor(financial_vars, use = "complete.obs")

# Reshaping the correlation matrix for visualization
cor_matrix_melt <- reshape2::melt(cor_matrix)

# Creating a heatmap: Visualizing the correlations between key financial variables to understand their
ggplot(cor_matrix_melt, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0) +
  ggtitle("Figure 3: Correlation Heatmap of Key Financial Variables") +
  xlab("") +
  ylab("") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Handling infinite values and missing data: Replacing infinite values with NA and then replacing NA wi
financial_vars <- financial_vars %>%
  mutate(across(everything(), ~ replace(., is.infinite(.), NA))) %>%
  mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))

# Standardizing the financial variables to prepare for PCA
financial_vars_standardized <- scale(financial_vars)

# Performing PCA: Identifying the principal components that explain the variance in the dataset
pca <- prcomp(financial_vars_standardized, center = TRUE, scale. = TRUE)

# Calculating the proportion of variance explained by each principal component
var_explained <- pca$sdev^2 / sum(pca$sdev^2)

# Identifying the elbow point: Finding the optimal number of components to retain
elbow_point <- which(diff(diff(var_explained)) == min(diff(diff(var_explained))))

# Creating a scree plot: Visualizing the proportion of variance explained by each principal component
ggplot(data.frame(PC = 1:length(var_explained), Variance = var_explained), aes(x = PC, y = Variance)) +
  geom_point() +
  geom_line() +
  geom_vline(xintercept = elbow_point, col = "red", lty = 2) +
  ggtitle("Figure 4: Scree Plot") +
  xlab("Principal Components") +
  ylab("Proportion of Variance Explained") +
  theme_minimal()

# Creating a loadings plot: Visualizing the contribution of each variable to the first two principal co
autoplot(pca, data = as.data.frame(financial_vars_standardized),
                          loadings = TRUE, loadings.label = TRUE, loadings.label.size = 5,
```

```
                          loadings.label.colour = "blue") +
  ggtitle("Figure 5: Loadings of the First Two Principal Components") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 12),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(hjust = 0.5, size = 16),
        legend.position = "bottom")
# Handling infinite values and missing data: Replacing infinite values with NA and then replacing NA wi
financial_vars <- financial_vars %>%
  mutate(across(everything(), ~ replace(., is.infinite(.), NA))) %>%
  mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))

# Standardizing the financial variables to prepare for clustering
financial_vars_standardized <- scale(financial_vars)

# Performing PCA again: Extracting the first two principal components for clustering
pca <- prcomp(financial_vars_standardized, center = TRUE, scale. = TRUE)
pca_scores <- as.data.frame(pca$x[, 1:2])
colnames(pca_scores) <- c("PC1", "PC2")

# Binding PCA scores to the original dataset
sol <- bind_cols(sol, pca_scores)

# Determining the optimal number of clusters: Using the elbow method to find the ideal number of cluste
fviz_nbclust(financial_vars_standardized, kmeans, method = "wss")

# Applying K-means clustering: Grouping the data into clusters based on the principal components
optimal_clusters <- 3
kmeans_result <- kmeans(pca_scores, centers = optimal_clusters)

# Adding the cluster assignments to the original dataset
sol$Cluster <- as.factor(kmeans_result$cluster)

# Creating a facet plot: Visualizing how loan data clusters across different regions based on the first
ggplot(sol, aes(x = `PC1`, y = `PC2`, color = Cluster)) +
  geom_point(alpha = 0.6) +
  facet_wrap(~ Region) +
  ggtitle("Figure 6: K-means Clustering of Loan Data by Region") +
  theme_minimal()
```