**Final Data Analysis Report**

**Introduction**

This project aims to analyse the closure of banks over time since October 1, 2000, in the US, focusing on the distribution of closures by state and acquiring institution. The primary goal is to understand the trends and patterns in bank closures, which can inform policy decisions and strategic planning in the banking sector. The original data source for this analysis is a collection of bank closure records, available at [link](https://www.kaggle.com/datasets/sahirmaharajj/failed-banks-dataset?resource=download). This project was selected due to the relevance of understanding the dynamics of bank closures, which can impact financial stability and consumer trust.

**Extract-Transform-Load**

**Extract**

1. **Data Extraction**: The extract class is responsible for reading data from CSV files. It uses the **csv.DictReader** to parse the CSV file and converts each row into a dictionary, storing all dictionaries in a list. This process extracts raw data from the source CSV file.

**Transform**

1. **Data Transformation**:
   * **Head and Tail**: The **transform** class provides methods to retrieve the top or bottom N records from the dataset, allowing for quick access to specific parts of the data.
   * **Rename Attribute**: Allows renaming a single column in the dataset.
   * **Remove Attribute**: Enables the removal of a single column from the dataset.
   * **Rename Attributes**: Supports renaming multiple columns at once.
   * **Remove Attributes**: Facilitates the removal of multiple columns simultaneously.
   * **Transform**: Combines the above operations to allow for complex transformations by specifying which columns to remove and/or rename.
   * **Add Full State Name**: Enhances the dataset by adding full state names using an external CSV file that maps state abbreviations to full state names.
2. **Data Aggregation**:
   * The script demonstrates several transformations to aggregate data in different ways, such as counting the number of closed banks per date, per state, and per acquiring bank. These transformations involve selecting specific columns to remove and possibly renaming others to focus on the desired aggregation criteria.

**Load**

1. **Data Loading**: The **load** class is responsible for writing the transformed dataset back into CSV files. It uses the **csv.DictWriter** to write each dictionary in the dataset as a row in the CSV file, with the keys of the dictionaries serving as field names.

**Visualizations and Interpretations**

1. **Number of Closed Banks Over Time**: This line plot shows the trend in the number of bank closures over the years.

A graph with a line graph

Description automatically generated

The data shows a general trend of bank closures in the US over the past two decades. There seems to be a significant increase in closures around the late 2000s, potentially coinciding with the 2008 financial crisis. Other periods with seemingly higher closure rates include late 2002, mid-2003, late 2005, mid-2006, and spread throughout 2013 and 2014.

The analysis of bank closures in the US over the years reveals a dynamic and evolving banking sector. The increasing frequency of closures, especially during economic downturns, highlights the sector's vulnerability to external shocks. However, the lack of a clear seasonal pattern and the potential impact of regulatory changes and technological advancements suggest a complex interplay of factors influencing the sector's stability. Further analysis, including geographic distribution and the impact of specific regulatory changes, would provide deeper insights into these trends.

1. **Distribution of Closed Banks by State**: This pie chart illustrates the proportion of bank closures by state. The visualization highlights the states with the highest number of closures, providing insights into regional disparities in bank health and stability.

A pie chart with different colors and numbers

Description automatically generated

**Illinois**, **Georgia**, and **Florida** also show a high frequency of bank closures, indicating these states may have unique challenges affecting the banking industry.

The distribution of bank closures is widespread across the country, covering both coastal and inland states. This suggests that the issue of bank closures is not limited to any particular region but affects banks nationwide.

1. **Top Acquiring Banks by Number of Closed Banks**: This bar chart displays the top 30 acquiring banks by the number of closed banks.

A graph of a financial report

Description automatically generated with medium confidence

A significant portion of the closed banks were acquired by other institutions. This suggests consolidation within the banking industry.

Several banks, like U.S. Bank, State Bank & Trust Company, and First Financial Bank, acquired multiple closed banks. This suggests they might be actively expanding their footprint.

A considerable number of closures don't have an acquiring bank listed. This could be due to various reasons, including:

* Failure Resolution: The Federal Deposit Insurance Corporation (FDIC) might have directly resolved the failure without involving another bank.
* Asset Sale: The closed bank's assets might have been sold off piecemeal instead of acquiring the entire institution.

**Conclusion**

The analysis of bank closures in the United States from October 1, 2000, to the present has provided valuable insights into the dynamics of the banking sector. The data reveals a complex interplay of economic, regulatory, and technological factors that influence the frequency and distribution of bank closures. The trend analysis indicates a significant increase in closures around the late 2000s, which could be attributed to the 2008 financial crisis. This period also saw notable spikes in closure rates in 2002, 2003, 2005, 2006, and 2013-2014, suggesting that external shocks and economic downturns play a crucial role in the banking sector's stability.

The geographical distribution of bank closures across the country underscores the widespread nature of the issue, affecting both coastal and inland states. States like Illinois, Georgia, and Florida stand out for their high frequency of closures, indicating unique regional challenges. This distribution suggests that the banking sector's stability is not uniform across the country, with some states facing more significant challenges than others.

The analysis of acquiring banks by the number of closed banks highlights the trend of consolidation within the banking industry. A significant portion of closed banks were acquired by other institutions, indicating a strategic response to economic downturns and regulatory changes. However, the presence of closures without an acquiring bank suggests that not all bank failures result in consolidation, possibly due to direct resolution by regulatory bodies or asset sales.

Overall, the analysis of bank closures in the US offers a nuanced understanding of the banking sector's resilience and vulnerabilities. It underscores the importance of regulatory oversight, economic stability, and technological innovation in maintaining the health and stability of the banking industry. Further research could explore the impact of specific regulatory changes, the role of technological advancements in banking, and the long-term effects of bank closures on local economies and consumer trust.

In conclusion, the analysis of bank closures in the US provides a comprehensive view of the banking sector's dynamics, highlighting the need for lasting monitoring and strategic planning to ensure financial stability and consumer protection. The insights gained from this analysis can inform policy decisions and strategic planning in the banking sector, contributing to a more stable financial ecosystem.

Author: Mikolaj Sobczyk C385

https://www.kaggle.com/datasets/sahirmaharajj/failed-banks-dataset?resource=download