Tiffany Tran, Christian Sierra Perez, Ziqi Polimeros

Professor Jefferson Bien-Aime

CIS 4400-S1DA

Term Project Report

**Project Title:** Data Warehouse Project of Financial Companies That Issued PPP Loan

**Authors:**

Project Manager: Christian Sierra Perez

Senior Data Analysis : Ziqi Polimeros

Data Analysis : Tiffany Tran

**Abstract:**

The purpose of this project is to evaluate the financial statements, stocks, and financial ratios of all the Financial Companies that were issued a Paycheck Protection Program (PPP) loan during COVID-19. Therefore, the data warehouse project displays skills in data warehouse design, data integration and management. It extracted data from diverse compatibility and transformations to ensure its compatibility and loaded it into the target system. The project offers insightful visualizations of line charts and bar charts, empowering users with a clear understanding of the data.

**Kanban Methodology:**

To begin, we utilized the Kanban methodology to have continuous improvement over the course of the project. It is an agile method that provides flexibility and a workflow that is enhanced and easily understood.

**The Kanban board indicates:**

1. The tasks that are currently being completed
2. The tasks that will be completed in the future.
3. The tasks that have been completed

**Data Source:**

* PPP Loan datasets were from the U.S. [Small Business Administration](https://data.sba.gov/dataset/ppp-foia) official website while stock price history, balance sheet and stock summary were from [Yahoo Finance](https://finance.yahoo.com/).
* Each stock price and balance dataset were about 30 k, while Each ppp loan dataset was about 400 MB. We combined all the ppp loan datasets and the final merged dataset was over five gigabytes.
* PPP Loan datasets were from the U.S. [Small Business Administration](https://data.sba.gov/dataset/ppp-foia) official website to get the Loan data dictionary from there and build a data dictionary for the stocks, balance sheet and ratios. To build the Dimensional Modeling on DB Schema. We created our physical and logical model with DB Schema.

**Data Warehouse Design:**

We were able to merge all the PPP Loan datasets from the U.S. [Small Business Administration](https://data.sba.gov/dataset/ppp-foia) official website to a CSV and use Tableau to connect to Snowflakes to do the Visualization. Also, to put the merge data to Data Warehouse.

**Data Transformation and Data Extraction:**

Using Python, we transformed the PPP loan dataset into dimensional and fact tables, enabling efficient access for users. The transformation process encompassed tasks like data mapping, merging, cleaning, filtering, and conducting comprehensive statistical analyses. We used Python web scraping the data from the websites above.

**Data Storage:**

With that, our ETL or extraction, transforming, and loading was fully created to be pushed into Microsoft Azure and built a container to source the data. PPP Loan datasets were from the U.S. Small Business Administration official website while stock price history, balance sheet and stock summary were from Yahoo Finance. This means that the data were pushed into Microsoft Azure and built a container to source the data.

**Data visualization:**

The stock price datasets were turned into line charts while balance sheet and ppp-loan were turned into bar charts, which was primed for insightful exploration and informed decision-making. Finally, with the utilization of Tableau, we created visualizations to present the data that we have transformed with filtering to demonstrate the different information that can be gathered from the visualizations.

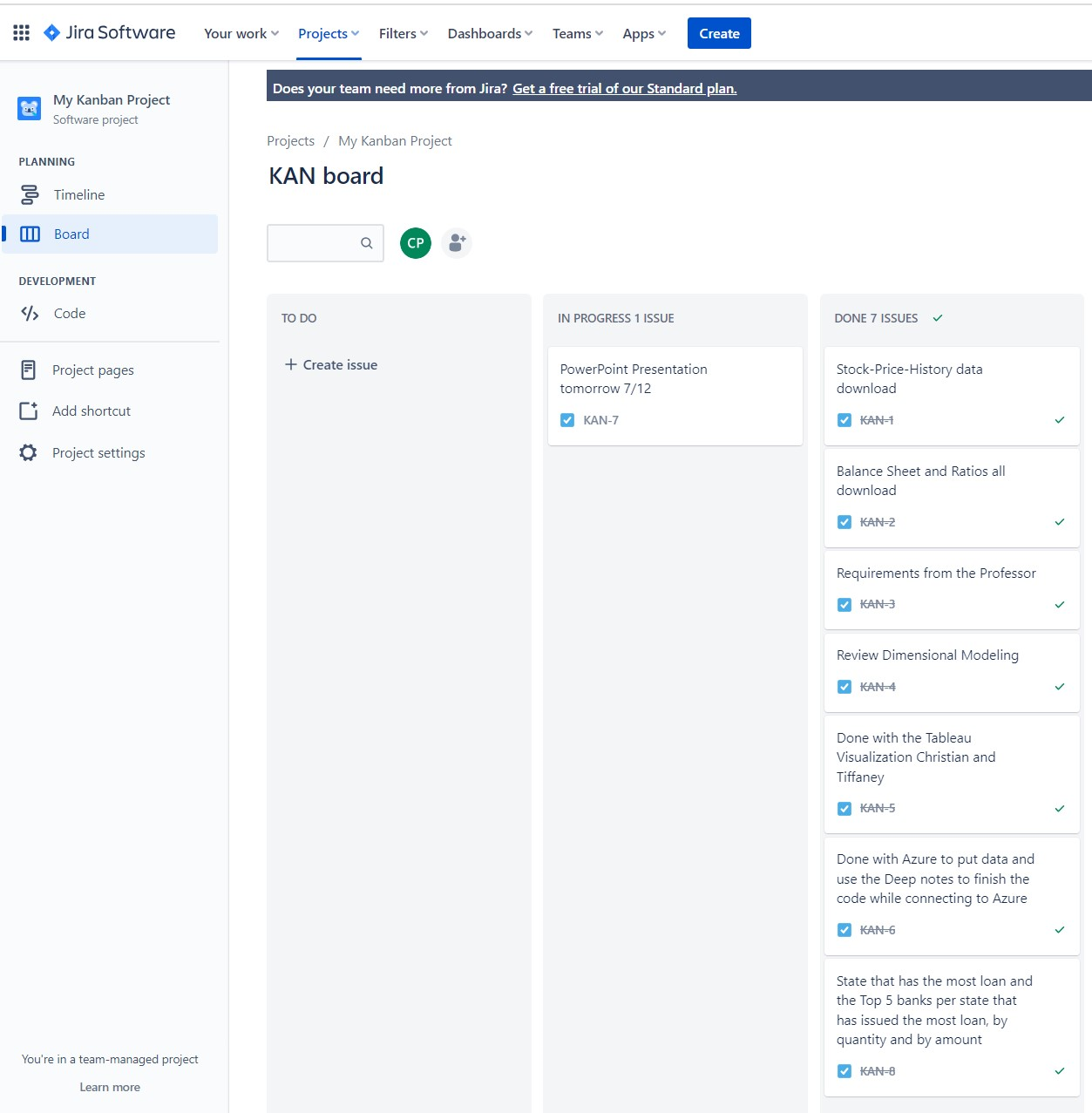
**Insight:**

* The stock price and total assets of the twenty companies went down during the pandemic. Subsequently, they displayed a steady upward trend.
* The forgiveness amount of loan was over 90% of the approval amount, but it did not bring a negative impact on the lender banks; instead, the stock price and total assets went up during and after the ppp loan program.
* The PPP Loan program not only provided vital support to numerous small businesses but also played a crucial role in safeguarding the financial industry.

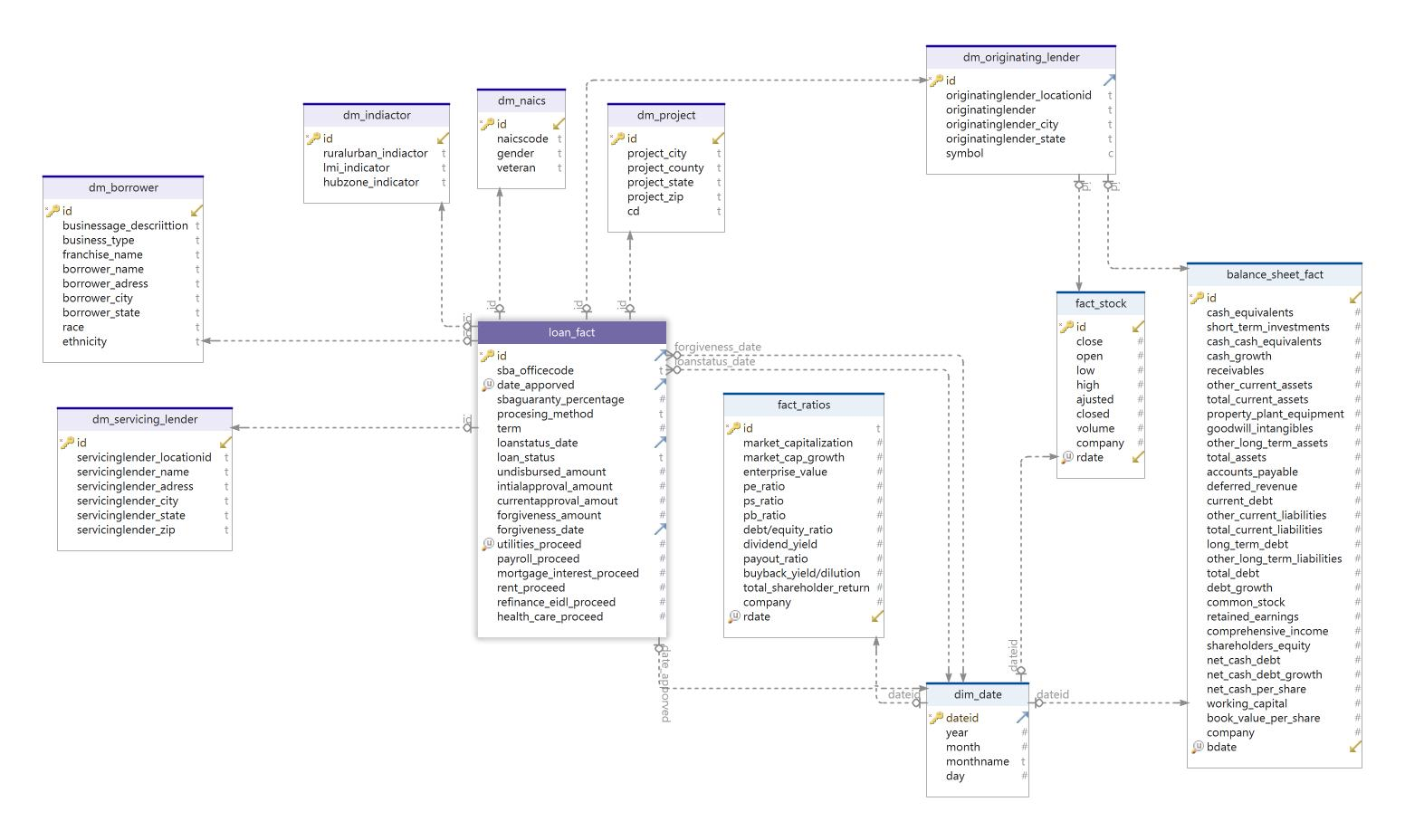
**Conclusion:**

* The Project was challenging at first, but my team members were able to complete the project on time in a brief period of time.
* We learned a lot of how to extract data from diverse sources, performed transformations to ensure its compatibility, and loaded it into the target system.
* We learned a lot of insightful visualizations of line charts and bar charts, empowering users with a clear understanding of the data.
* We had our difficulties on the project, but we were able to ask help and questions from the professor to help out with the project.

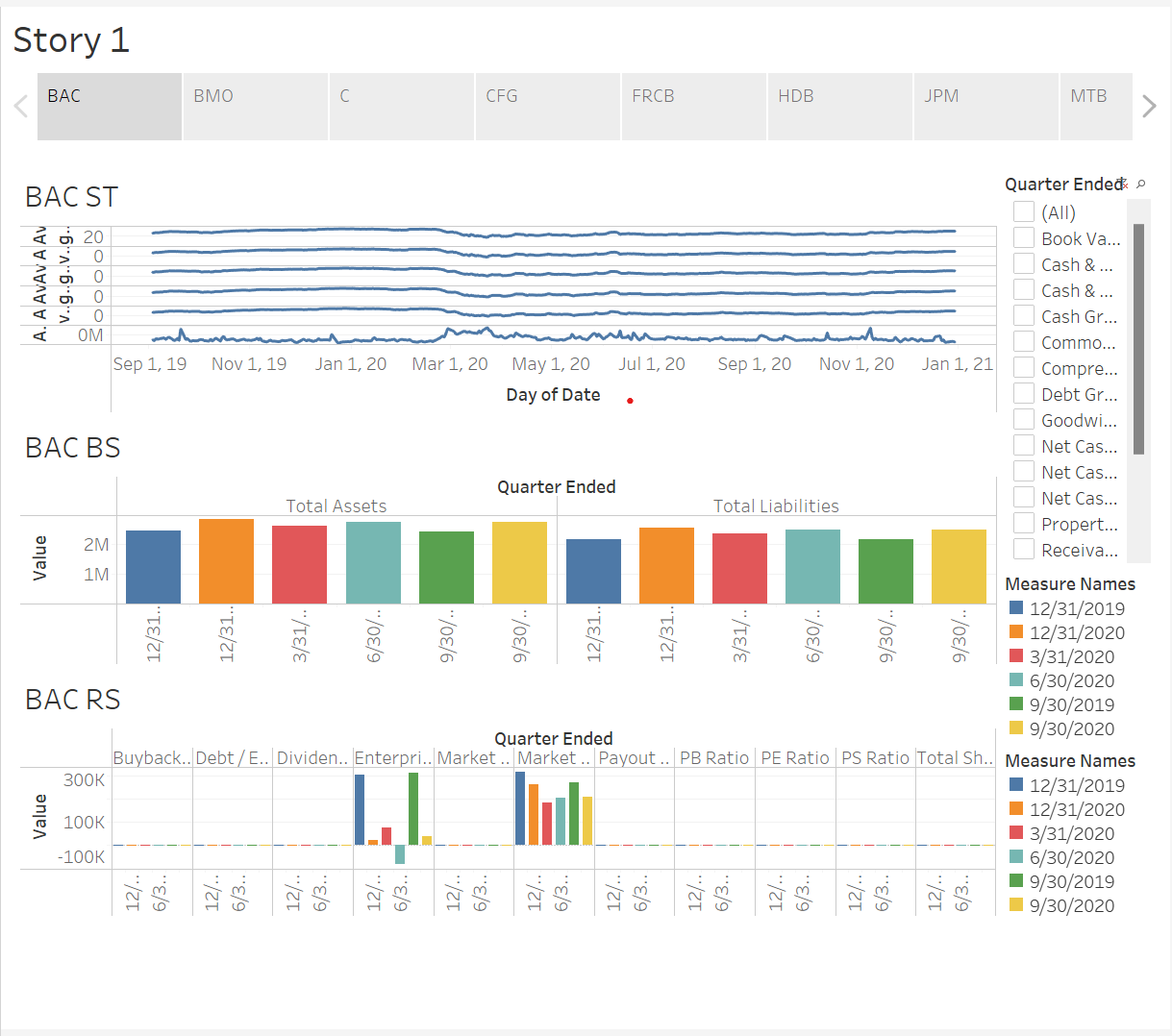
***Kanban Methodology***

******

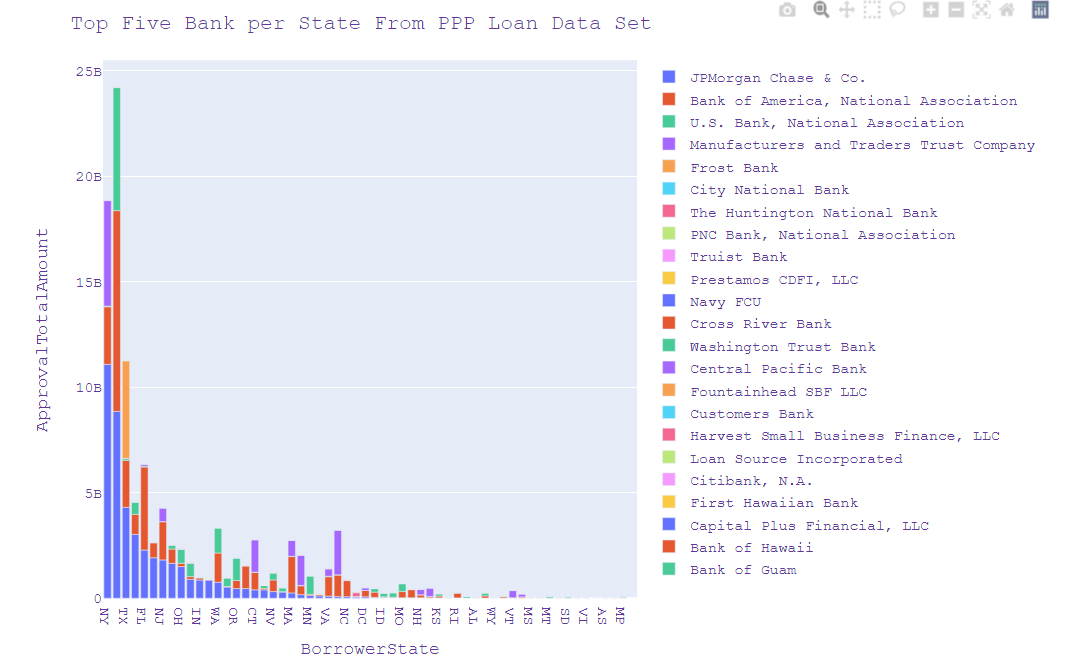
***DB Schema***

******

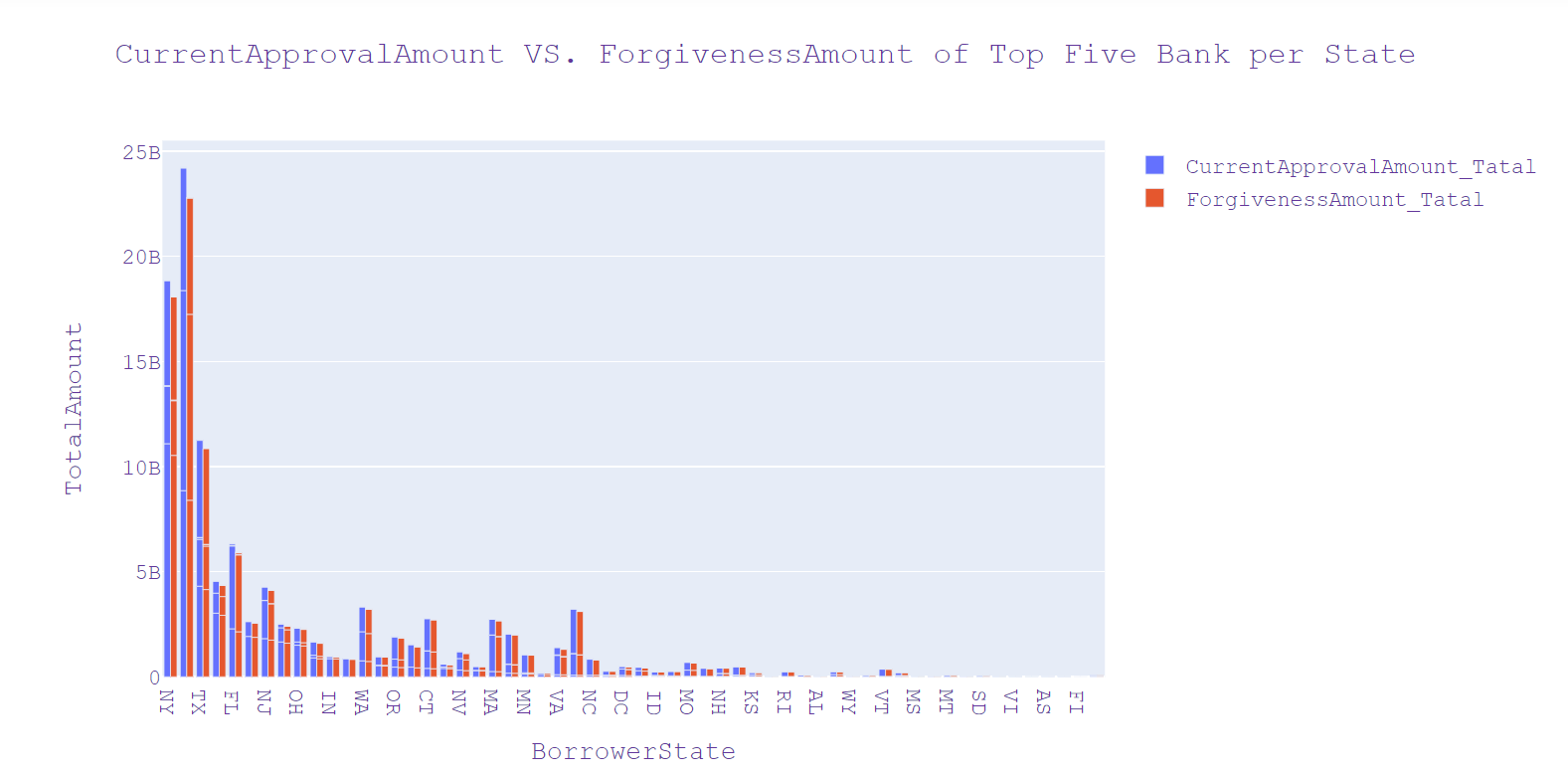
***Tableau Visualization***

******

***Data Visualization – PPP Loan***

******

***Insight***

******

***List of Links for Our Term Project***

***GitHub Link:***

<https://github.com/Christ244/CIS4400>

***\* Deep notes Link \*:***

<https://deepnote.com/workspace/cis4440-af30dd6c-7034-49d4-8ffd-f5f64dc5d870/project/KXW-ca436d51-26cb-4a7b-9af3-0ce82bc99b1b/notebook/christian_tiffany_ziqi-28eecb5ebd224e84830d08c034d50ea3>

***Project PowerPoint:***

<https://docs.google.com/presentation/d/1KXKwsI8TYJ0FEIYt7-qRqH4FOfVE9nB4tE4OsMCjMhQ/edit?usp=sharing>

**List of Links for Our Term Project**