Option 1: Capstone Project—Business Intelligence Solution for U.S. Organization

Nichole Kang

Colorado State University Global

MIS480: Capstone - Business Analytics and Information Systems

Dr. Mamdouh Babi

August 6, 2023

Option 1: Capstone Project—Business Intelligence Solution for U.S. Organization

The project covers the presentation of business intelligence solutions for the bank, Financial Services Bank (FSB). The selected organization for this project is a leading bank in the US offering an array of financial services products to its customers. This bank was selected because it has a large amount of data to work with and has several business opportunities that can benefit from business intelligence tools. This organization can enhance its operations to improve marketing strategies, as well as increase customer satisfaction. Shao et al. state that "business intelligence establishes a genuine business benefit for data properties and provides significant advancements in recognizing and utilizing consumer potential" (Shao et al., 2022, p. 13). Incorporating business intelligence can help to improve the decision making process by providing insightful information to leaders.

Business Problem

FSB seeks to better understand the acceptance rates of their credit card offers to customers. The organization wants to look at segmenting customers to send out higher quality credit cards offers. "For effective marketing, it is essential to identify a specific group of customers who share similar preferences and respond to a specific marketing signal" (Fan et al., 2015, p. 7). The segmentation process will consider financial factors such as credit scores, number of credit cards, home ownership status, and other factors. These factors will help to determine which credit card offers are best suited to the customer based on their financial profile. The objective is that having better quality offers will lead to a higher credit card acceptance rate.

Dataset

The dataset used for this project contains data for 10,000 customers who currently bank with FSB. Any personal identifying information is omitted for client confidentiality. There are 16 variables in this dataset, such as offer accepted, reward type, income level, credit score, and number of credit cards held. Additionally, the dataset includes data on the customers home ownership status and various deposit account data. The Offer_Accepted variable is binary, with 0 = “No” and 1 = “Yes”. The Credit Rating variable is binary, with 1 = “Poor” representing a credit score of 300-629, 2 = “Good” a credit score of 630-719, and 3 = “Excellent” credit score of 720-850. FSB stores this data in a cloud-based warehouse such as AWS, which also provides data security and is a highly reputable company. “The advantages of cloud storage include unlimited data storage space, convenient, safe and efficient file accessibility and offsite backup, and low cost of use” (Yang et al., 2020, p. 2). An API will be used for real-time access to the data.

Business Intelligence Tools

R Studio will be used and the primary business intelligence tool for this project. This is a powerful tool and performs particularly well with statistical analysis and visualization. One advantage to using R Studio is that it has a variety of packages that can be used for analysis and data manipulation. It will be useful in creating predictive models and segmenting customers, in order to produce insightful information and easy to interpret visualizations. “Predictive analytics can assist financial institutions for optimizing business processes, knowing customer behaviour, identifying unexpected opportunities, and preventing problems before its occurred” (Indriasari et al., 2019, p. 877).

Evaluating the Data, Visualization and Data Analytics Outcomes

Figure 1 shows the dataset uploaded into R Studio along with the output of the Summary Statistics for the credit score variable. It shows that the lowest credit score in the dataset is 429, and the highest credit score is 850. The mean credit score is 653, which falls into the “Good” credit rating category.

```
> current_time <- Sys.time()
> print(current_time)
[1] "2023-08-05 23:54:40 PDT"
> #Nichole Kang SID: 321150
>
> customer_data <- read_excel(path = "C:\\Users\\nkang\\Desktop\\MIS480\\creditcardsbank.xlsx")
>
>
> # Summary Statistics of Customer dataset
> summary_stats <- list(
+   Min = min(customer_data$credit_score),
+   Max = max(customer_data$credit_score),
+   Mean = mean(customer_data$credit_score),
+   Median = median(customer_data$credit_score),
+   SD = sd(customer_data$credit_score)
+ )
> print(summary_stats)
$Min
[1] 429

$Max
[1] 850

$Mean
[1] 653.6913

$Median
[1] 661

$SD
[1] 99.488
```

Figure 1

Screenshot of dataset uploaded into R Studio and Summary Statistics

Figure 2 displays the histogram for the credit scores of the customers in the dataset. This is a left skewed distribution. The majority of the customers appear to be in the 600 - 750 credit score range.

```
> current_time <- Sys.time()
> print(current_time)
[1] "2023-08-05 23:55:16 PDT"
>
> # Histogram of credit_score for Customer dataset
> options(scipen = 999999)
> hist(customer_data$credit_score,
+      main = "Histogram of Credit Score for Customer dataset",
+      xlab = "Credit Score",
+      ylab = "Frequency",
+      xlim = c(300, 900),
+      ylim = c(0, 4000),
+      labels = TRUE,
+      col = "blue",
+      border = "black",
+      breaks = 10)
>
```

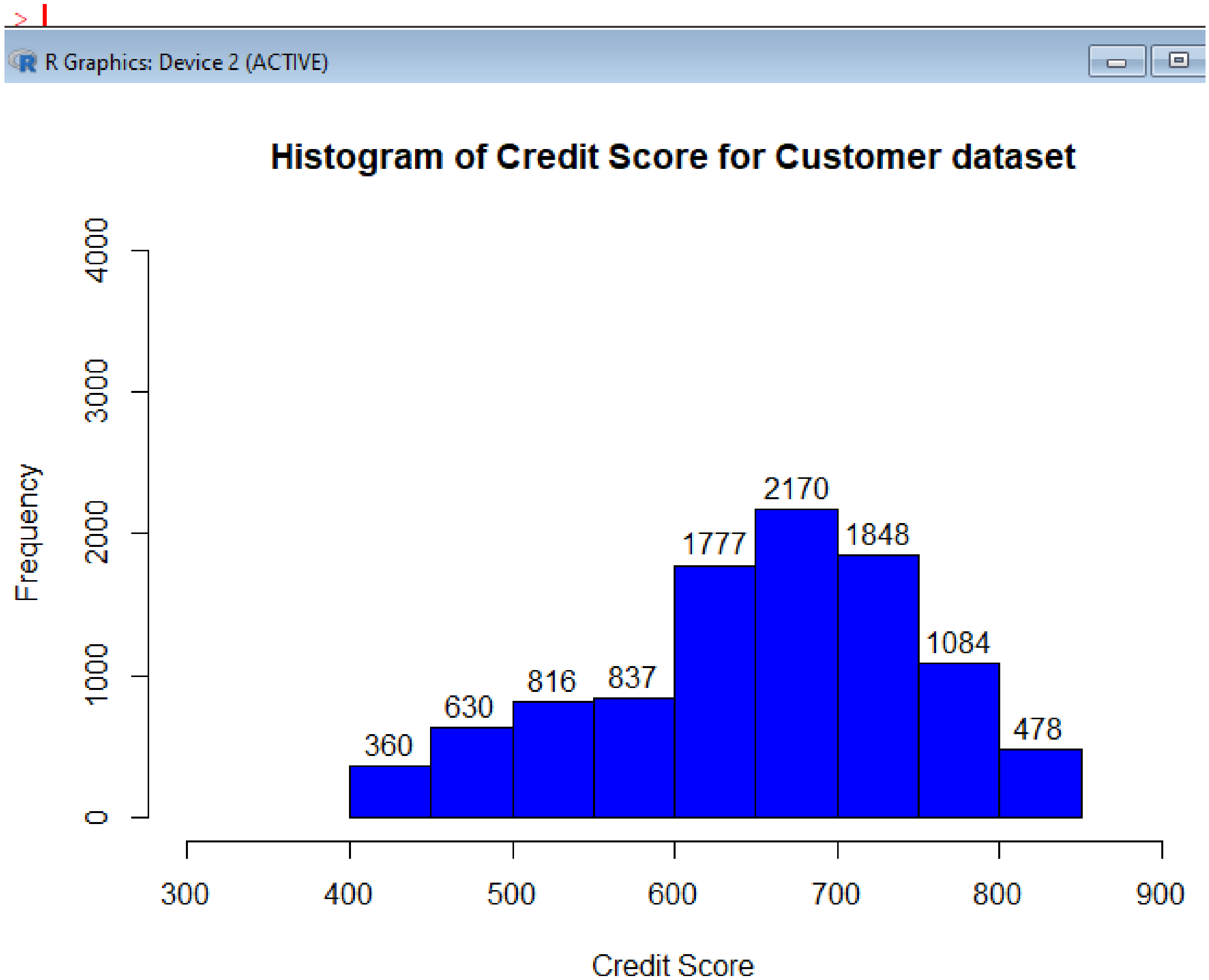

Figure 2

Screenshot of Histogram for Credit_Score variable

Figure 3 displays a pie chart showing the distribution of credit ratings for the 10,000 customers listed in the dataset. 30.6% have a “Poor” credit rating, these customers likely have faced some financial difficulties or bad financial behaviors, such as missed payments or defaulting on loans. This makes them a higher risk customer for the bank. 39% have a “Good” credit rating, these customers are the majority of the distribution. This type of customer has a generally consistent credit history, but may have some areas of opportunities in improving their financial situation, such as building more credit history or paying down revolving credit balances. 30.4% have an “Excellent” credit rating and typically have long standing credit history, low credit utilization, and on time payments. These are considered the more favorable customers as they tend to be the lowest risk.

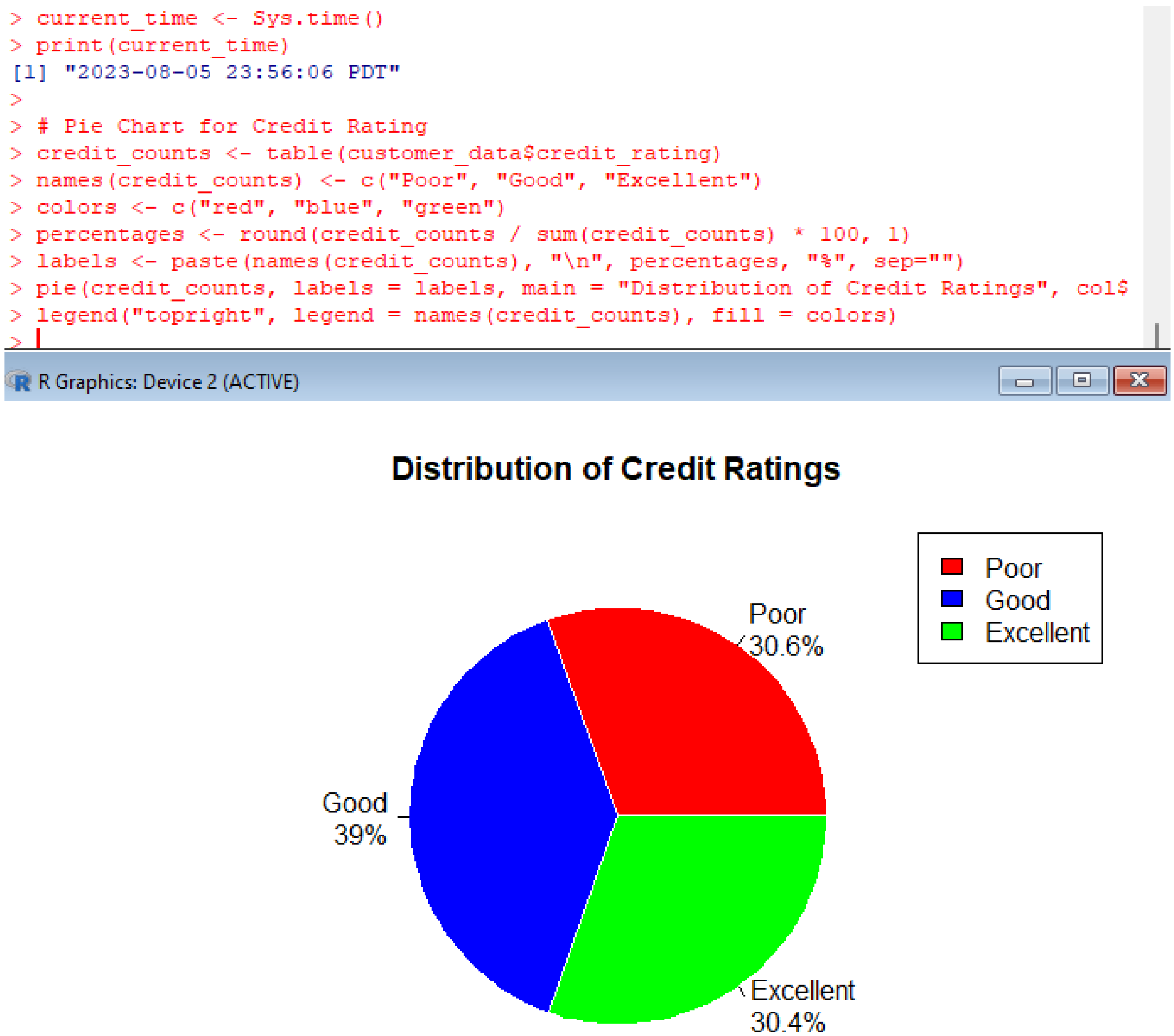

Figure 3

Screenshot of Pie Chart of Credit Ratings

Figure 4 shows a stacked bar chart of the offers accepted by the credit rating. The proportion of "Poor" credit rating shows that less than 1/8 of offers are accepted. For "Good" credit rating customers, the proportion shows a little more than 1/4 of offers are accepted. For "Excellent" credit ratings, it is only slightly above the 1/8 mark of credit offers accepted.

```
> current_time <- Sys.time()
> print(current_time)
[1] "2023-08-05 23:58:00 PDT"
>
> # Stacked Bar Plot of Credit Rating by Offer Accepted
> customer_data$credit_rating <- factor(customer_data$credit_rating,
+                                       levels = c(1, 2, 3),
+                                       labels = c("Poor", "Good", "Excellent"))
>
> customer_data$offer_accepted <- factor(customer_data$offer_accepted,
+                                        levels = c(0, 1),
+                                        labels = c("No", "Yes"))
>
> ggplot(customer_data, aes(x = credit_rating, fill = offer_accepted)) +
+   geom_bar(position = "fill") +
+   labs(title = "Proportion of Offers Accepted by Credit Rating",
+        x = "Credit Rating",
+        y = "Proportion") +
+   scale_fill_brewer(palette="Set1", name = "Offer Accepted") +
+   theme_minimal()
>
```

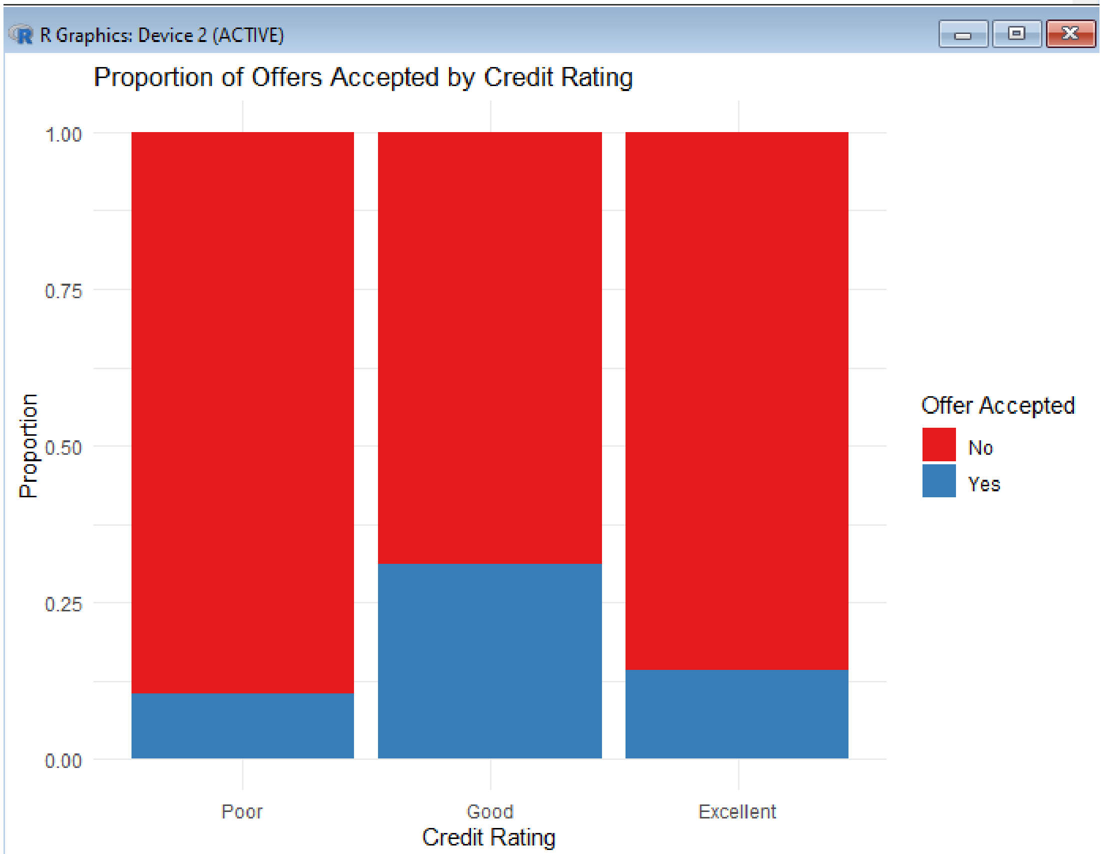

Figure 4

Screenshot of Stacked Bar Plot for Credit Rating by Offer Accepted

Figure 5 shows the executed Logistic Regression model for the offers accepted by credit rating and annual income. This shows that when a customer has a “Good” credit rating, they are 1.52 times more likely to accept a credit card offer from the bank than if they had a “Poor” credit rating. If the customer has an “Excellent” credit rating, they are only 0.31 times more likely to accept a credit card offer than those with a “Poor” credit rating. Additionally, it shows that for each unit the annual income increases, the likelihood of the customer accepting a credit card offer goes up by 0.000006139, which is an insignificant amount. These findings are shown to have strong statistical significance as displayed by the output < 0.0000000000000002.

```
> current_time <- Sys.time()
> print(current_time)
[1] "2023-08-05 23:49:32 PDT"
>
> # Logistic Regression
> customer_data$credit_rating <- as.factor(customer_data$credit_rating)
> logistic_model <- glm(offer_accepted ~ credit_rating + annual_income,
+                       data = customer_data, family = binomial)
> summary(logistic_model)

Call:
glm(formula = offer_accepted ~ credit_rating + annual_income,
    family = binomial, data = customer_data)

Coefficients:
                          Estimate   Std. Error z value            Pr(>|z|)
(Intercept)           -2.995263313  0.073978523  -40.49 < 0.0000000000000002 ***
credit_ratingGood      1.518078788  0.071587079   21.21 < 0.0000000000000002 ***
credit_ratingExcellent 0.309407359  0.080787377    3.83             0.000128 ***
annual_income          0.000006139  0.000000262   23.43 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9913.1  on 9999  degrees of freedom
Residual deviance: 8821.5  on 9996  degrees of freedom
AIC: 8829.5

Number of Fisher Scoring iterations: 5
```

Figure 5

Screenshot of Logistic Regression

Figure 6 shows the output of the predict() executed using the previous logistic regression. The output displays the confusion matrix where 0 represents “not accepted offers” and 1 represents “accepted offers”. The 7831 figure represents the correct figure for “offers not accepted”. 306 is the correct figure for “offers accepted”. 203 were not predicted accurately, and were predicted as “accepted offers”, but were actually “not accepted offers”. 1660 were not predicted accurately, and were predicted as “not accepted offers”, but were actually “ accepted offers”.

```
> current_time <- Sys.time()
> print(current_time)
[1] "2023-08-06 00:06:10 PDT"
>
> # Predict using Credit Rating and Annual Income
> predicted_probs <- predict(logistic_model, newdata = customer_data, type = "response")
> predicted_values <- ifelse(predicted_probs > 0.5, 1, 0)
> confusion_matrix <- table(customer_data$offer_accepted, predicted_values)
> print(confusion_matrix)
     predicted_values
         0    1
  No  7831  203
  Yes 1660  306
>
```

Figure 6

Screenshot of Prediction Function

Benefits of Business Intelligence

Given the multiple tasks performed in this project, it can be seen that there are many benefits of using business intelligence. Firstly, having the ability to make data-driven decisions is imperative in the business world. This brings confidence to stakeholders that decisions are being made with actual data rather than relying on one's intuition or assumptions on what they think would be best. In this scenario, the data shows how the business can utilize credit card offers to target marketing campaigns to customers who are more likely to accept the offers, one such customer segment is the “Good” credit rating customers. Secondly, using business intelligence provides the ability to create visualizations. These are immensely helpful in interpreting the data into useful information. This is especially true for persons who are not as skilled in interpreting data. In this project, bar charts, pie charts, and stack bar graphs were utilized to create visual representations of the data, in order to gain a better understanding of what the data is telling us. This information creates easy to interpret graphs from complex data. Thirdly, the insights obtained from business intelligence gives organizations a competitive advantage. The information from this project provides Financial Services Bank (FSB) the ability to effectively segment their customers for more impactful marketing campaigns. These insights were further supported by logistic regression and predictive analytics.

Programming Code and GitHub

Figure 7 shows the business intelligence solution uploaded to GitHub. GitHub is a great way to showcase projects and build a portfolio. It also provides a connection of like minded professionals to collaborate on projects and also network with other individuals in the tech community. Tsay et al. states that the social cues given by developers on GitHub are beneficial and "they use these inferences in practical ways, for instance to help manage their projects, discover user needs, and recruit developers" (Tsay et al., 2014, p. 358).

Github Link: https://github.com/nichlka/CapstoneProject

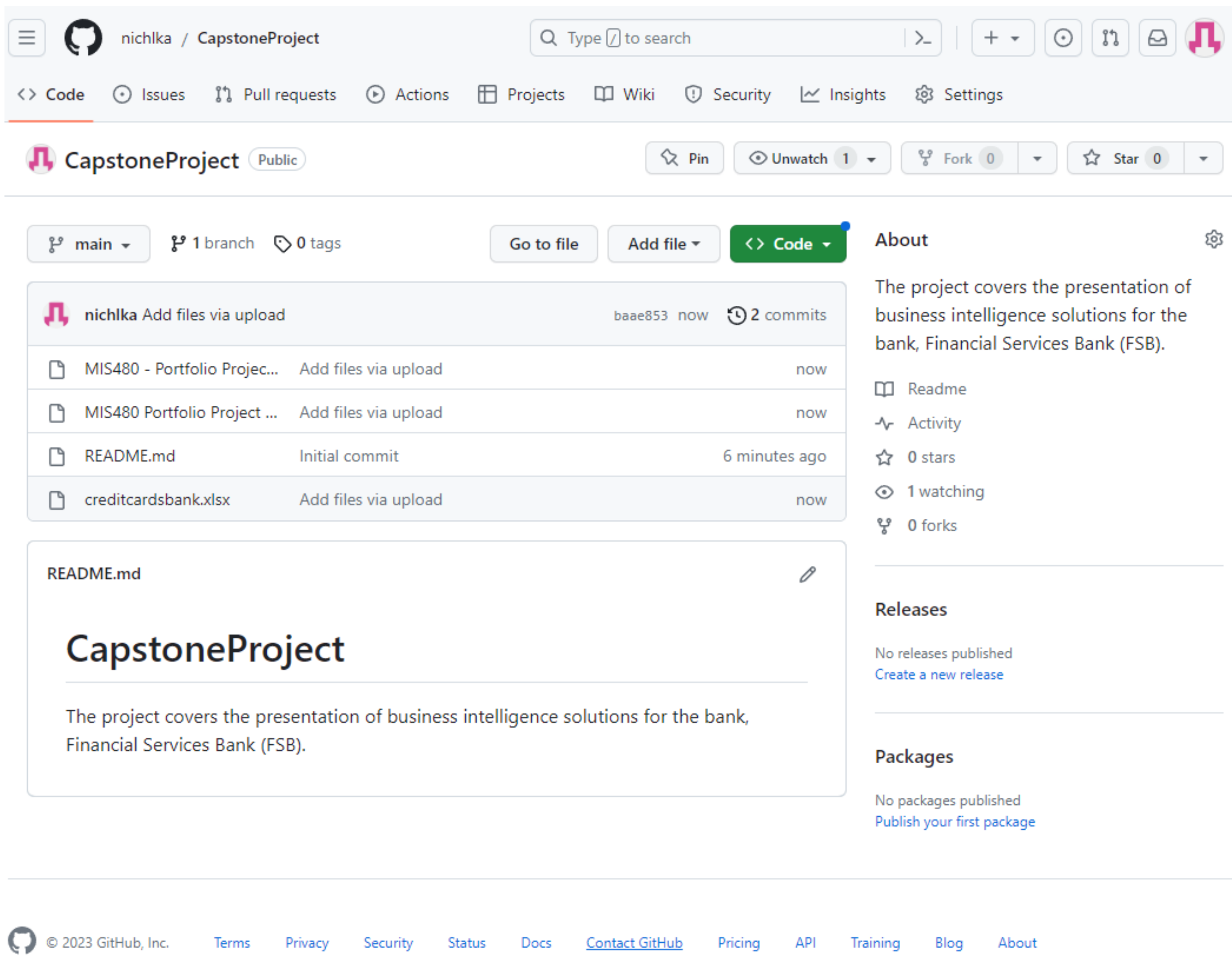

Figure 7

Screenshot of BI Solution uploaded to GitHub

Conclusion

This business intelligence solution for the Financial Service Bank (FSB) concluded with success in gaining more insight on the credit card offers accepted by customers. The organization will be able to make data-driven decisions in order to improve in the goal of understanding customer segments to increase the rate of credit card offer acceptance. The outcome concluded that the customer segment with a “Good” credit rating has a higher likelihood of accepting a credit card offer. The visualization provides easy to understand insights on the data, which allows leaders and stakeholders to make more informed decisions. By utilizing R Studio, this project was able to use statistical methods and execute predictive analytics. Using business intelligence to solve business needs, helps to improve efficiencies in business operations. Outcomes, such as this scenario, can lead to a better customer experience, for example, having more customized and applicable marketing. This gives businesses, like FSB, a competitive advantage by having a more strategic approach than other competitors.

References

Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, *2*(1), 28–32. https://doi.org/10.1016/j.bdr.2015.02.006

Indriasari, E., Soeparno, H., Gaol, F. L., & Matsuo, T. (2019). Application of Predictive Analytics at Financial Institutions: A Systematic Literature Review. *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 877–883. https://doi.org/10.1109/iiai-aai.2019.00178

Shao, C., Yang, Y., Juneja, S., & GSeetharam, T. (2022). IoT data visualization for business intelligence in corporate finance. *Information Processing & Management*, *59*(1), 1–14. https://doi.org/10.1016/j.ipm.2021.102736

Tsay, J., Dabbish, L., & Herbsleb, J. (2014). Influence of social and technical factors for evaluating contribution in GitHub. *Proceedings of the 36th International Conference on Software Engineering*, 356–366. https://doi.org/10.1145/2568225.2568315

Yang, P., Xiong, N., & Ren, J. (2020). Data Security and Privacy Protection for Cloud Storage: A Survey. *IEEE Access*, *4*, 1–18. https://doi.org/10.1109/access.2020.3009876