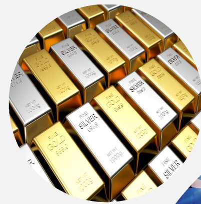
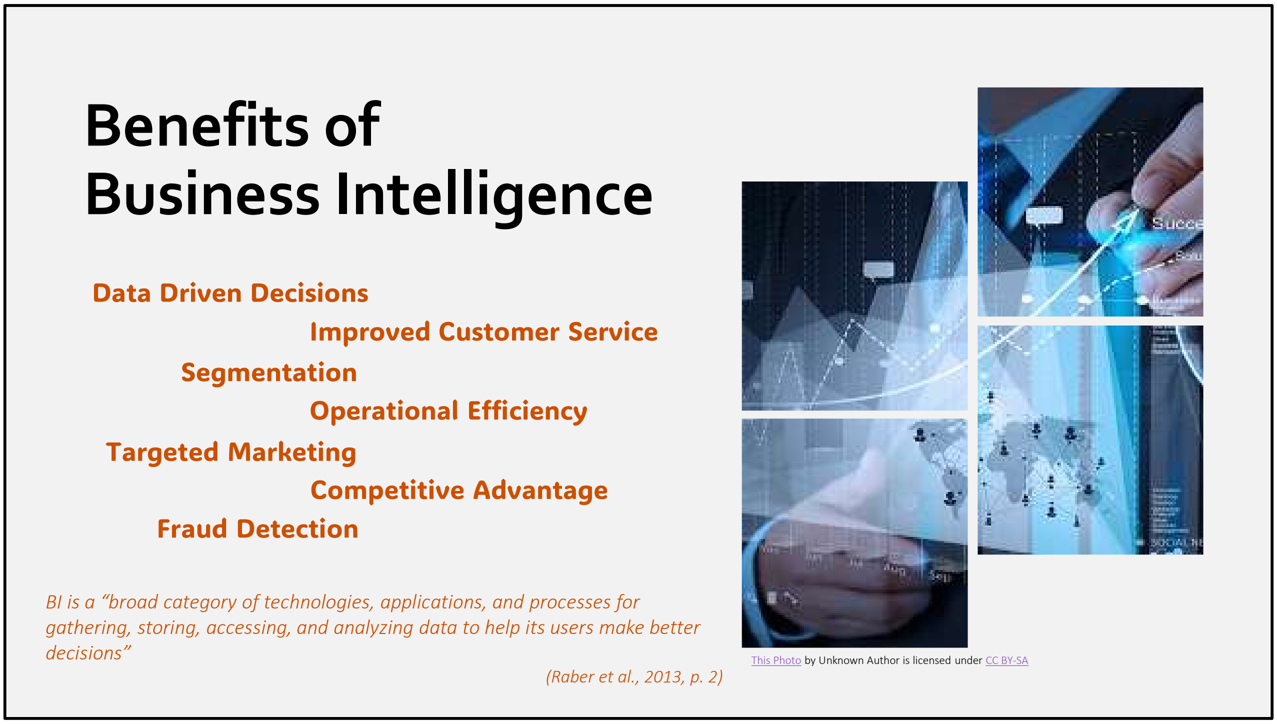
I will be introducing my capstone project, which highlights the importance of incorporating business intelligence solutions into business operations. This project will provide business intelligence solutions for a specific organization, Financial Services Bank (FSB), which is a leading bank in the US. The primary object will be to determine ways in which FSB can use the power of BI to enhance their business processes and develop in their growth and efficiencies.

Financial Services Bank

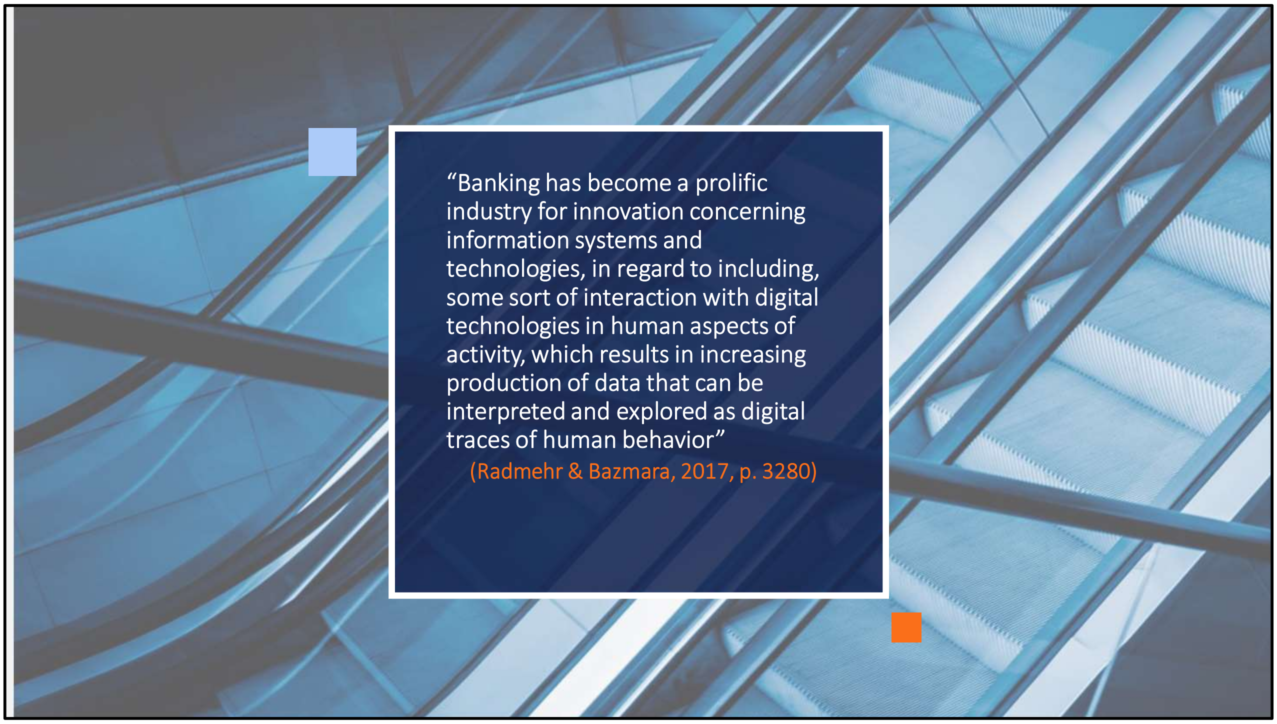


- Financial Services Bank
  - A leading bank in the US
  - Offers a wide variety of products and services
- Data Availability at FSB
  - Large amount of data from diverse sources
  - Data contains customer interactions, transaction records, and more
- The role of business intelligence at FSB
  - Supports informed decision making
  - Helps identify business opportunities
  - Enables improvement in marketing strategies and customer satisfaction

Financial Services Bank offers many financial services and has a large customer base. The reason it was selected for this project is that it has a large amount of diverse data available and opportunities to transform that data into powerful insights. The bank maintains a lot of quality data, such as customer interactions, transaction records, and financial information. This type of data can provide many insights to the bank to improve processes, leading to higher profits for the bank and a better experience for the customer.

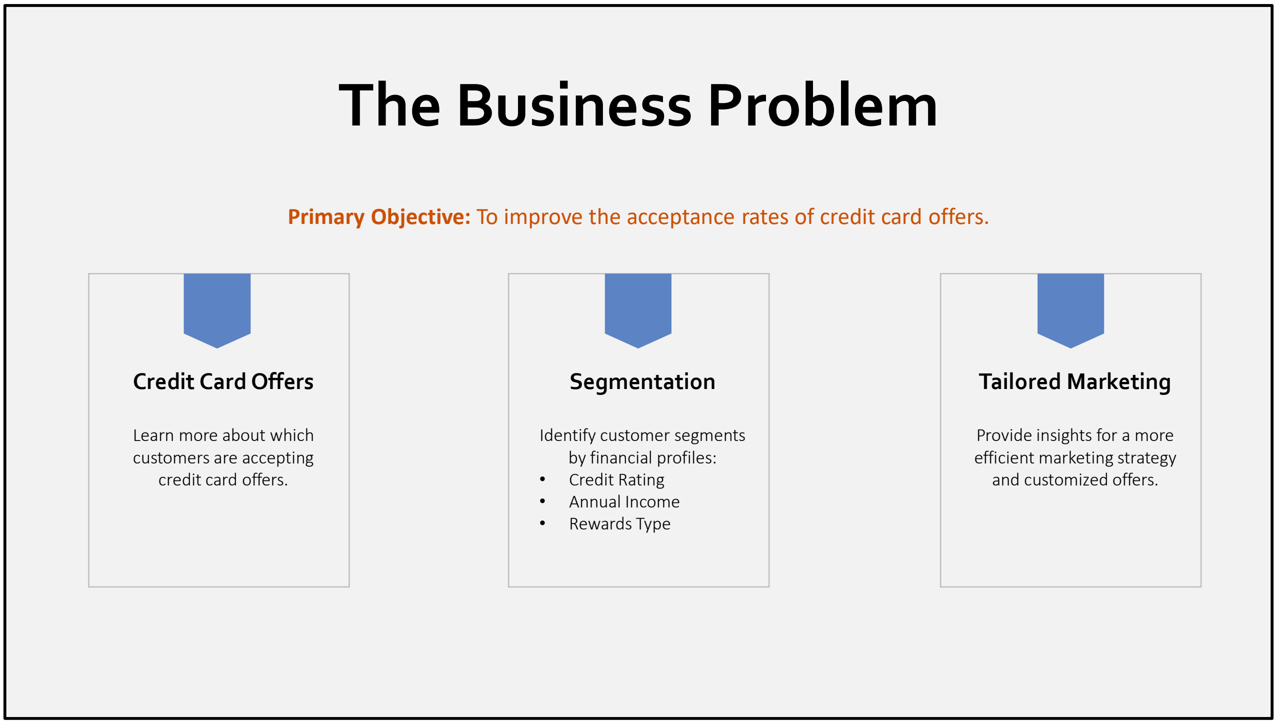


Business intelligence can provide a wide variety of benefits to organizations. Banking in particular can utilize business intelligence in many different segments of the business. For example, business intelligence is highly involved in fraud monitoring and protection. BI tools can help to analyze patterns in customers transaction behavior and detect suspicious activity. This helps to minimize losses for both the customer and the bank and saves the customer a lot of stress in preventing fraud from occurring.



I'd like to draw your attention to a quote from Radmehr and Bazmara.
"Banking has become a prolific industry for innovation concerning information systems and technologies, in regard to including, some sort of interaction with digital technologies in human aspects of activity, which results in increasing production of data that can be interpreted and explored as digital traces of human behavior" (Radmehr & Bazmara, 2017, p. 3280).
This idea captures the importance and value

that BI delivers to the banking industry.

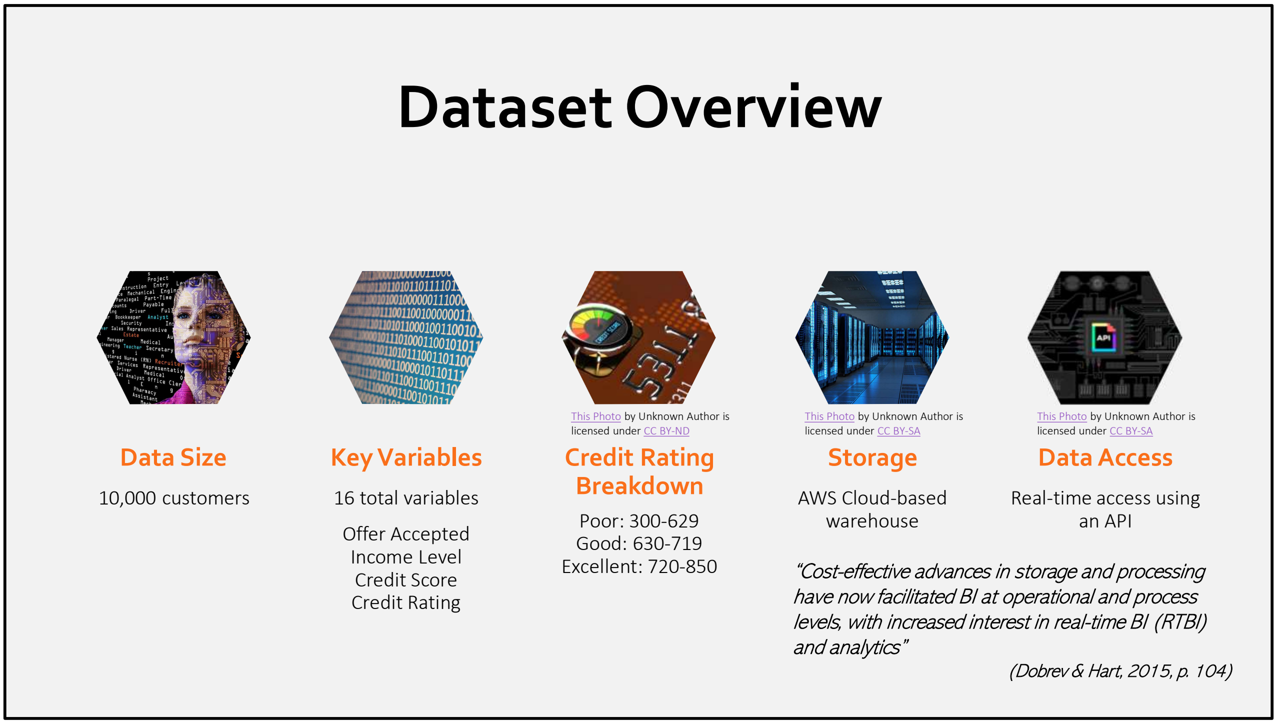


FSB wants to be able to improve their processes in order to increase the rate that customers are accepting credit card offers sent to them by the bank. They want to learn more about what factors are contributing to the current customers who are accepting credit cards. We will seek to have a deeper understanding of FSBs customers who receive credit cards offers and learn more about what factors impact offer acceptance. With this information FSB can better segment their customers and provide more tailored marketing strategies.

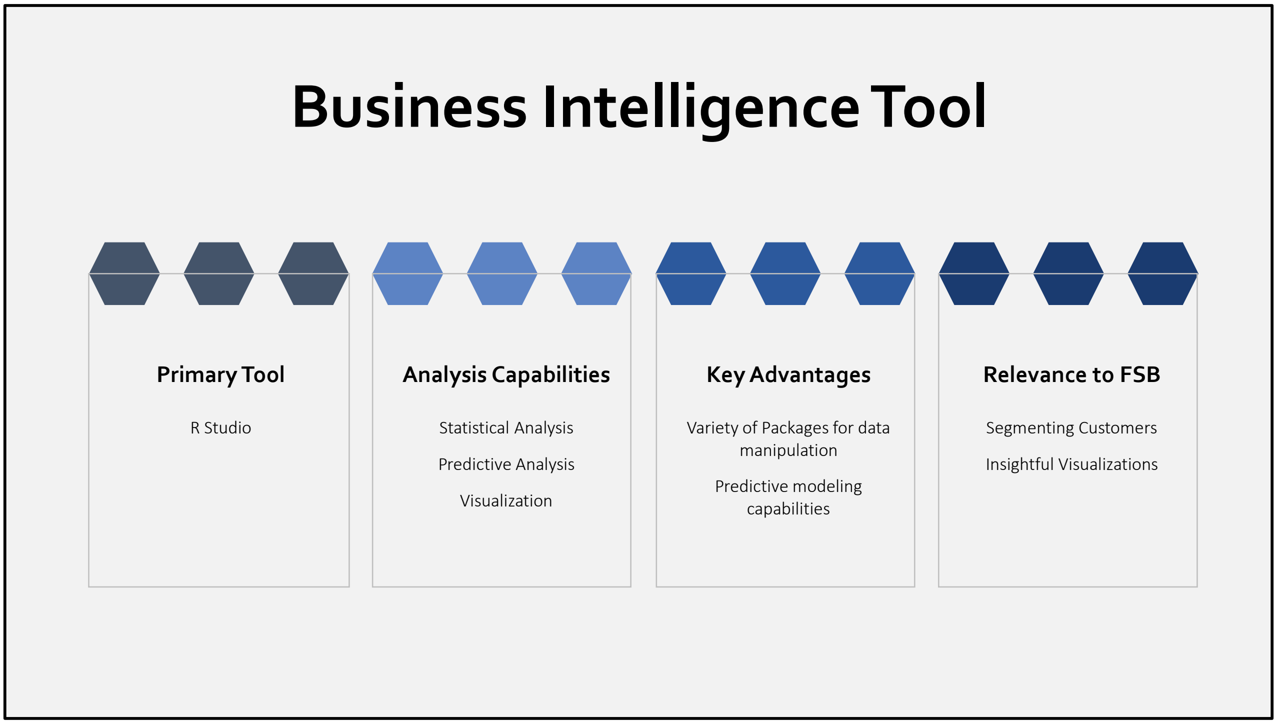
Segmentation Criteria

Credit Rating	Categories: Poor (1), Good (2), Excellent (3)
Credit Score	Categories: 300-629 (Poor), 630-719 (Good), 720-850 (Excellent)
Annual Income	Categories: < $48,500 (Low), $48,501 < x < $115,000 (Medium), > $115,001 (High)
Offer Acceptance	Categories: No (0), Yes (1)

Segmentations helps us to better understand the customers. By breaking down the customer base into manageable segments, the bank will be able to offer targeted offers that are more customized to the customers individuals needs. The segmentation for this project focused on four key variables: credit rating, credit score, annual income, and offer acceptance. Though there are many more variables that can and should be considered, these were the primary focuses on this project and what their influence was in the ultimate outcome of offer acceptance. The credit rating and credit score provide valuable insight to the customers financial credibility and risk levels. This are separated categories for the benefit of using various analysis processes to evaluate the customers. Annual income is helpful to determine the customers spending power and what offers would benefit them. The offer acceptance history was a key variable to determine how the customer financial profile can predict their willingness to accept a credit card offer.



It is important to understand the dataset before digging into the analysis of the project. The dataset gives us access to data for 10,000 customers, which is a substantial amount and will help in creating meaningful insights. The dataset provides 16 total variables that can be used, for this project the key variables are the offer acceptance, income levels, and the credit rating and score. The credit rating is broken down into three categories for analysis: poor, good, and excellent, and is will have binary representation for accuracy in statistical and predictive analysis. The dataset is securely store on AWS, which is a popular cloud platform that provides reliability and security. The dataset can be accessed real-time using API, which provides confidence that the analysis will always be up to date.



There are many options for business intelligence tools that can be used for this project. Some options include, Power BI, SAS, and Python. For this task, I chose to use R Studio for its statistical and predictive analysis capabilities. In addition to the numerical calculations that can be done in R Studio, it can also create visualizations that help to turn the data into something understandable. A major bonus with using R Studio that it has access to an extensive library of packages. An example of one such package is ‘ggplot2’ which is a visualization package that I used for this project. R Studio will help to provide insights to FSB on customer segmentation through its analysis and visualizations.

Evaluating the Data

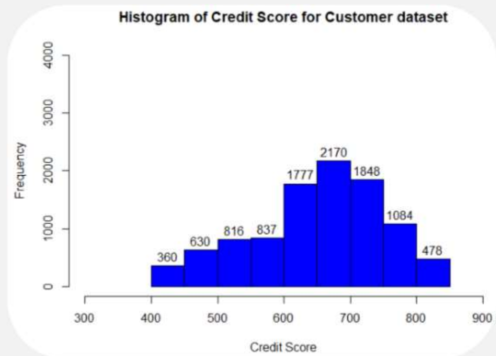


Figure 1 Screenshot of Histogram for Credit Score variable

Upload the Dataset

- Dataset is uploaded into R Studio

Key Metrics

- Summary Statistics for the credit score variable

Distribution Analysis

- Histogram for credit scores of customers

Key Findings

- Range of credit scores: 429 (lowest) to 850 highest
- Mean: 653 (Good credit rating)

Digging into evaluating the data starts with uploading the dataset into R Studio. As the credit score variable provides a good picture of the customers financial health, this is the first variable that was dug into. The summary statistics was run on the credit score variable in order to understand the distribution and characteristics. To better understand the spread, a histogram was created, and this provided clarity where most of the customer were with their credit score (See Figure 1). Based on this analysis, it was found that the minimum credit score in the dataset was 429, and the highest was 850. Also, the mean score was 653, which falls into the Good category for credit rating. The histogram has a left tail that shows that not as many customers are on the extreme low end of the credit score. This provides a good place to start in understanding the customers financial profiles.

Evaluating the Data

Rating Distribution

- Poor: 30.6%
- Good: 39%
- Excellent: 30.4%

Key Findings

- Majority of customers in “Good” credit rating
- “Poor” and “Excellent” have also equal distribution

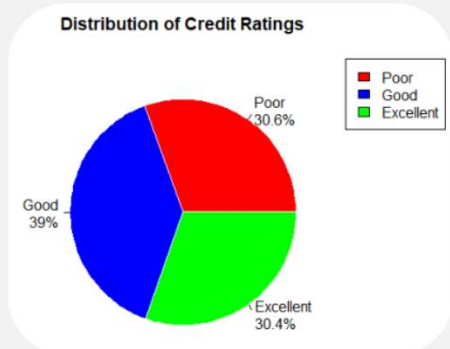


Figure 2 Screenshot of Pie Chart of Credit Ratings

To further analyze the data, a pie chart was created to visualize the distribution of credit ratings among the customers. In this distribution, it is seen that 30.6% of the customers fall into the “Poor” credit rating category, these customers may have had some past financial troubles. The most significant percentage of customers fall into the “Good” category, which represents a reasonably consistent financial history. The remaining 30.4% of customers are in the “Excellent” category, which indicate the more financially responsible customers. One take away is that the “Poor” and “Excellent” customers are almost equal in this distribution, which gives further insight on tailoring marketing strategies to customer segments.

Evaluating the Data

Segments on the Bar Chart

- Offers Accepted
- Offers Not Accepted
- Credit Rating

Key Findings

- The "Good" bar shows the most notable difference in offer acceptance

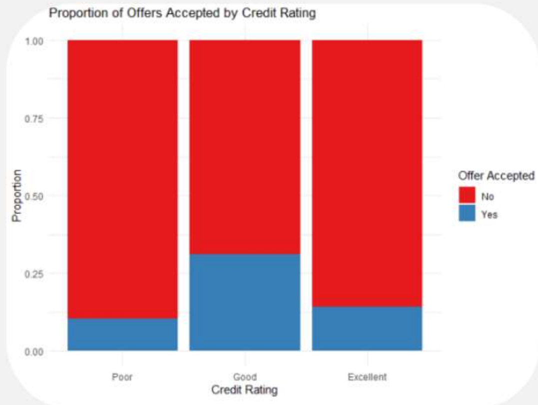


Figure 3 Screenshot of Stacked Bar Plot for Credit Rating by Offer Accepted

Continuing to evaluate the data, a stacked bar chart was selected as a visualization for the customers who accepted or declined a credit card offer. This chart includes three bars, representing each of the credit rating categories for poor, good, and excellent. By stacking the offers accepted and declined, it can be quickly seen which groups are responding to offers. The "Poor" and "Excellent" categories seem to have the lowest acceptance proportions compared to the "Good" category, which is a little over the 0.25 marker. It's helpful to have this visualization to show the gaps in the accepted and declined offers. The insights on this will help the marketing strategy to create more appealing offers to these segments showing lower offer acceptance.

Logistic Regression

Results

- Credit Rating: Good (1.52)
- Credit Rating: Excellent (0.31)
- Annual Income: 0.000006139

```
> # 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
```

"Logistic regression is used to obtain odds ratio in the presence of more than one explanatory variable"
(Sperandei, 2014, p. 12)

Using Logistic Regression is a model that can leverage two different predictors, and can model the probability of a binary outcome, such as offer accepted Yes or No. For this project, the credit rating and annual income variables were used as predictors to determine the likelihood of an offer being accepted by the customer. This model demonstrates that a customer's acceptance rate for credit card offers from the bank is 1.52 times higher when they have "Good" credit than when they have "Poor" credit. Only 0.31 times as many customers with "Excellent" credit ratings as those with "Poor" credit are more likely to accept a credit card offer. For every unit that the annual income increases, the probability that a consumer will accept a credit card offer is minimal. This model tells us that there is significant statistical significance of these findings.

Predictive Model

Results

- 7831 “offers not accepted” (True)
- 306 “offers accepted” (True)
- 203 “accepted offers” (False)
- 1660 “not accepted offers” (False)

```
> # 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
```

“Predictive models aim to specify a probabilistic model that provides a good fit to testing data that were not used to estimate the model’s parameters”
(Cranmer & Desmarais, 2017, p. 1)

Using the predict function, a predictive model was generated using the logistic regression model. 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”.

This project for FSB brought a wealth of insights that came from using BI tools and predictive modeling. This project delivers several insights the FSB incorporate into their business operations to see positive results. One such insight is to incorporate customized marketing strategies tailor campaigns to the customer specific needs and segment. Additionally, utilizing personalized offers well how to increase the acceptance rates and credit card offers. These strategies will help FSB in the short and long term by continuing to utilize business intelligence tools. Business intelligence well help to identify patterns and trends and deliver powerful data-driven insights so business leaders can make the most informed business decisions.

References

Cranmer, S. J., & Desmarais, B. A. (2017). What Can We Learn from Predictive Modeling? *Political Analysis*, *25*(2), 145–166. https://doi.org/10.1017/pan.2017.3

Dobrev, K., & Hart, M. (2015). Benefits, Justification and Implementation Planning of Real-Time Business Intelligence Systems. *The Electronic Journal Information Systems Evaluation*, *18*(2), 104–118.

Raber, D., Wortmann, F., & Winter, R. (2013). Towards The Measurement Of Business Intelligence Maturity. *21st European Conference on Information Systems*, 1–12.

Radmehr, E., & Bazmara, M. (2017). A Survey on Business Intelligence Solutions in Banking Industry and Big Data Applications. *International Journal of Mechatronics, Electrical and Computer Technology*, *7*(23), 3280–3298.

Sperandei, S. (2014). Understanding Logistic Regression Analysis. *Biochemia Medica*, *24*(1), 12–18. https://doi.org/10.11613/bm.2014.003