Analysis of JGA Bank Credit Card Customer Data and Churn Characteristics

Student's Name

Western Governors University

Table of Contents

A. Project Overview

A1. Research Question or Organizational Need

The purpose of this project is to create an automated process utilizing Python that will allow JGA Bank to gain insights from its existing credit card customer dataset. Of particular interest to JGA Bank, is understanding the characteristics of customers who've cancelled their credit cards and no longer use their credit services. JGA Bank is currently in the process of creating a new credit card program and believes that by having insights into the customers who have cancelled, or churned, the bank will be better able to build its new credit card product and reduce the risk that customer churn poses to its revenue stream. Currently, JGA Bank has no reporting solutions in place that allow it to utilize its existing dataset to gain these insights. This project will address this need and provide JGA with a tool to do so.

A2. Context and Background

JGA Bank (fictional), is a financial institution that provides banking and lending solutions to over 25,000 customers in its local community. JGA and its associates strive to provide their customers with a wide variety of products that both meet their customer's needs and provides additional revenue streams to fuel the banks growth. Of the services offered, JGA Bank launched its first credit card five years ago and has since seen much success with its initial card offering. Since its inception, JGA's credit card has been adopted by nearly half of its current customers. Due to this success, the bank is looking to increase its credit card portfolio with a new credit card offering that will allow JGA to expand its customer base and further grow revenue. Since launching its first credit card, the bank has built a database containing characteristics of each customer, purchase patterns and credit card usage. As JGA is looking to develop its new credit card program, the bank seeks to gain insights from its existing database of card holders to help build their new product. JGA Bank is particularly interested in leveraging its existing customer dataset to identify characteristics of customers that have churned since acquiring its credit card. While overall, JGA has been very successful in retaining its credit card customer base, it is particularly concerned about the potential impact churn may have on its credit card revenue as it seeks to expand its customer base with

the new credit card product. The bank believes that utilizing its current data to understand its existing and former customers, will better allow it to build its new product and reduce the risk that customer churn poses to its revenue.

A3. Summary of Published Works

Across many of today's industries and businesses, there is one metric that is continually increasing in its importance as an indicator of company success: churn. Churn is significant to all businesses that sell consumer goods but is especially important to those that offer a service as a product. These services can either be in recurring subscriptions or for the financial sector, in terms of lending through loans or credit cards. To understand the importance of churn, first one must understand what churn is. A simple explanation of churn is when a consumer decides to no longer purchase a product or to no longer subscribe to a service. When this happens, the business offering that good or service, loses a consistent source of revenue and in turn, reduced profits from that offering. In the case of banks, churn can be of significant cost. According to Deepak Dube in an article for Forbes.com in May 2020, "The annual attrition rate for banks in North America is about 11%. This high number forces banks to spend considerable marketing dollars to sign up new customers just to keep the customer count flat." This loss of revenue and additional expense of attracting new customers can be devastating to a company's bottom line and its future.

To combat this threat, it is important that businesses develop strategies to understand and address the underlying issues that drive churn. In an article for Investopedia.com, author Jake Frankenfield discusses how analyzing churn data can give companies insight into the quality of their business model and how it may indicate where a business needs to focus its attention to address potential shortcomings. According to Frankenfield:

> If a company sees that its churn rate is increasing from period to period, then it understands that a fundamental component of how it is running its business is flawed. The company may be

> providing a faulty product, it may have poor customer service, or its product may not be attractive to individuals who decided the cost is not worth the utility.
>
> The churn rate will indicate to a company that it needs to understand why its clients are leaving and where to fix its business. The cost of acquiring new customers is much higher than it is to retain current customers, so as you ensure that the customers you worked hard to attract remain as paying customers, it makes sense to understand the quality of your business. (2022)

As Frankenfield points out, churn can be indicative of a flawed business model, or product that doesn't meet customer expectations. To remedy this, it is imperative that a business maintains a thorough understanding of who its customers are, and if their products meet the customers' needs. To this end, it is beneficial for a business to understand the demographics of its clientele, such as age, education, and income. With datapoints such as these, a profile can then be built to allow for customer segmentation, and a closer analysis of how customers with similar backgrounds may interact with a product. In an article for Amplitude.com, Senior Product Marketing Manager, Anastasia Fullerton speaks to how using customer segmentation, or cohorting, can provide deeper insights when doing a churn analysis. In her article Ms. Fullerton states:

> When you're doing a churn analysis, remember to recognize that not all of your users will interact with your product in the same way. Different types of customers have different propensities to churn because they encounter different roadblocks or challenges when using your product. Running a churn analysis using cohorts will uncover issues specific to different types of users. Behavioral cohorting is a powerful way to identify causes of churn that would easily be missed if you only looked at your overall churn rate. (2020)

For JGA Bank, as it looks to grow its credit card offerings, understanding its current and former credit card customers is vital. As Ms. Fullerton points out in her article, it is important to recognize that not at

all customers use a product the same or have the same priorities. However, it is possible that customers with similar backgrounds may have similar expectations and behaviors. Bearing this in mind, the need for JGA Bank to leverage its existing data to understand its customers characteristics is further justified. Once the bank has completed this vital first step, the insights gained from this understanding will instruct the bank on how to better build its new credit card offering. Additionally, this insight into its customers will enable the bank to build upon this understanding to develop strategies to potentially address churn before it even occurs. In the afore mentioned article for Forbes, Mr. Dube speaks to how banks should tailor their churn strategies to encompass all organizations within the business, specifically those organizations that work directly with consumers such as customer care. Deepak speaks to how if a business has a solid understanding of their customers, they can tailor the customer experience, and leverage insights from customer data to build proactive solutions to ensure consumers are taken care of, and are happy with the service they receive, ultimately resulting in less churn. The initial steps that JGA Bank is undertaking in this project, will place the bank on a firm footing to enable JGA Bank to be successful in its future endeavors.

A4. Summary of Data Analytics Solution

To assist JGA Bank in its desire to better understand its credit card customer base, I will be creating an application using Python to automate the process of analyzing the banks credit card customer data set. The application will create customer segmentations based upon factors such as age, education and income and then will determine how much, if any, these customer attributes affect the likelihood that a customer will churn, based upon historical data. The application will require a user to input a recent extract of the credit card customer dataset, in the form of a .csv file, and will then run the analysis in a quick and efficient manner to give the Bank insights based upon the input. The output of the application will include customer segmentations, a summary of how many customers is included in each segmentation, along with relevant percentages that each customer segment represents in total observable churn rates. The application will also include visualizations in the form of bar charts and tables to depict the trends

observed from review of the data. Overall, this application will save JGA Bank both time and money, allowing members of the organization to create and enact actionable plans based on the insights provided from the analysis of customer data.

A5. Benefit to Organization and Decision-Making Process

Presently, JGA Bank has limited visibility into its customer dataset, particularly where churn is concerned, and needs a solution that can provide insights into the bank's customers in a timely, efficient, and accurate manner. The proposed solution will benefit the bank by providing a reporting application that will allow JGA to leverage its data to provide these insights, ultimately saving time and money. By utilizing the new reporting application, analysts will be able to quickly complete an analysis of customer data, eliminating the need for a time-consuming manual analytical process. The results provided from the application will provide JGA leadership insights previously unavailable to better tailor their credit card program to their customer's needs, and in turn reducing the likelihood a customer will churn, which erodes profit margins and increases operating costs.

B. Data Analytics Plan

B1. Goals, Objectives, and Deliverables

The goal of this project is to produce a reporting solution using Python that will automate the analysis of JGA Banks credit card customer dataset and provide insights into the number of customers with similar traits, and trends related to the level of churn among customer segments.

To accomplish this goal, there are three objectives:

1. Create customer segmentations to analyze churn behavior among customers.
    a. The deliverable for this objective is to create customer segmentations for the entire customer base by grouping customers upon factors including age, education, and income.
2. Create a customer segmentation summary for customers who have churned.

   a. The deliverable for this objective will be a summary detailing how many customers are in each customer segment, along with the percentage each customer segment represents of overall churn rate.

3. Create an active customer segmentation summary.
   a. The deliverable for this objective will be a summary detailing how many active customers are in each segment, highlighting those customers who fit the same profile as those identified in the previous summary with higher churn rates, as these customers may be at higher risk of churning, and should be further reviewed.

B2. Scope of Project

The scope of this project is to provide JGA Bank with a reporting solution via a Python application. The application will use data provided from the banks credit card customer database, input by a user via a .csv format. The application will analyze the data to provide an output detailing the number of customers in segments based upon similar traits, the percentage of churn each segment represents, as well as the number of active customers in each segment that may be at risk of churning. The insights provided from the analysis will act as a guide to point JGA Bank in the direction of customers it should investigate further to find the underlying reasons why these customers churned. This analysis does not intend to provide an explanation as to the cause of churn, or a definitive answer to which customers are ultimately responsible for churn, but merely a starting point for a broader analysis.

B3. Standard Methodology

To produce and complete this project, a Waterfall methodology will be used. This methodology uses a linear progression, following five phases that require each phase to be completed before proceeding to the next. Each phase is outlined below, along with an explanation of how each step will be executed during the completion of this project.

- **Requirements**: During this phase of the project, it will be important to work with the customer to gather all requirements, assumptions, dependencies, and determine costs. During this phase of the

project, I will coordinate with the customer to gather all required details to ensure the project can begin and to gain a full understanding of the customers' expectations.

- **Design**: After gathering all relevant details and completing the Requirements phase, I will then begin the Design phase. During the Design phase, I will review available data, and determine how to segment customers and best complete a churn analysis for each segment. During this phase, I will be able to determine which analytical method best fits the stated requirements and customer expectations, and then begin to craft a reporting solution that addresses the customer need.
- **Implementation**: After determining an appropriate design to satisfy project requirements, I will then begin to develop the Python script during the Implementation phase using a test dataset.
- **Verification & Testing**: The Verification phase can begin once coding has been completed based upon a test dataset during the Implementation phase. For this phase, I will begin introducing new data from the larger dataset to test and validate that the application is working as intended, and there are no errors or issues identified that may affect the validity of results or efficiency of the application.
- **Deployment & Maintenance**: The last phase of the project is the Deployment and Maintenance phase. During this phase, the application will be deployed to the organization and made available to users. During this phase it will be important that users are provided with documentation on how to use the application and are aware of how to maintain the application as needed. For this project once it is deployed, there will be no on-going maintenance provided to the organization, further emphasizing the need for user documentation.

B4. Timeline and Milestones

Milestone	Projected Start Date	Projected End Date	Duration (days/hours)
Gather project requirements	02/01/2023	02/01/2023	1 day / 8 hours
Review requirements and develop workflow	02/02/2023	02/02/2023	1 day / 8 hours
Gather data, create test dataset	02/03/2023	02/03/2023	1 day / 8 hours

Python script development	02/06/2023	02/08/2023	3 days / 24 hours
Test Python application	02/09/2023	02/09/2023	1 day / 8 hours
Develop User Training and Maintenance Materials	02/10/2023	02/10/2023	1 day / 8 hours

B5. Resources and Costs

Resource	Resource Type	Cost	Quantity	Duration	Total
Professional Service Hours	Personnel	$40 / hr	1	64 Hours	$2,560
Python	Technology	$0	1	NA	$0
Anaconda Navigator / Jupyter Notebook	Technology	$0	1	NA	$0
Laptop	Technology	$0	1	NA	$0
Infrastructure	Infrastructure	$0	0	NA	$0
5% Contingency	Contingency	5%	1	NA	$128
Expected Total	-	-	-	-	**$2,688**

All required resources to complete this project, with the exclusion of professional service hours are either already available within the organization (Infrastructure and Technology) or are free software downloads (Python). As a professional contractor, I am equipped with my own laptop, so no additional technology charges apply to this project. The expected duration of this project is a total of 64 billable hours at a rate of $40 per hour, totaling $2,560. A five percent contingency has been built into the estimate as a precaution to account for any unexpected costs, however unlikely, and will be deducted if no further expense is incurred.

B6. Criteria for Success

The goal of this project is to provide JGA Bank with insights into the traits of customers who have churned so that the bank may identify a strategy to understand what motivates customers with similar traits to no longer use its credit card services. To accomplish this, the project must produce accurate customer segmentations, that represent all customer types available in the dataset. Additionally, it is important to understand what percentage of overall churn each customer segment represents, as this is key to identifying which customer segments require additional research. Lastly, it is imperative that the bank

understands the risk each customer segment represents within its active customer base, both regarding its current and future credit card offerings.

To measure success for the first objective will require validation that all customers within the dataset are successfully assigned to a customer segment. This step will be completed within the Validation and Testing phase, or Phase 4 of the Waterfall project management methodology. Once the application has been coded, I will then begin to introduce the full data sample into the application and verify the number of customers represented in the source data ties to the total number of customers in the overall customer segment analysis. If both totals agree with one another, the criteria for this objective will be considered satisfied.

In evaluating the second objective for success, I will determine the total churn rate for the observed period and will review the percentage of churn that each customer segment represents. If the individual customer segments total the overall percentage of churn, the criteria for success will have been achieved for the second objective.

Lastly, to evaluate the third objective has been successfully achieved, I will validate that all customers marked as active have been correctly assigned to a customer segment and that each customer segment matches in both the active and inactive customer groups. If customers are segmented within the same type groupings based upon customer traits, regardless of status, this objective will be considered successful.

Criterion/Metric	**Required Data**	**Cut Score for Success**
Were customer correctly identified in segments?	Output of reporting solution	Considered successful if there are no outlier customers that have not been assigned to a segment
Was churn percentage correctly captured for each customer segment?	Output of reporting solution	Considered successful if churn percentage from each customer

		segment correctly adds up to the observed total churn percentage
Were all customers with an active status successfully assigned to a customer segment, and all customer segments are consistent based upon customer characteristics?	Output of reporting solution	Considered successful if all customers are correctly assigned to a customer segment, and segments are consistent based upon customer characteristics regardless of active status

C. Design of Data Analytics Solution

C1. Hypothesis

Creating customer segments, based upon grouping customers with similar traits, such as age, education, and income will identify that certain customer segments are more likely to churn than others. Specifically, I predict that customers of an older age group, higher income and educational backgrounds will represent a larger percentage of the overall churn rate than other customer segments.

C2. Analytical Method

I will use a one-way ANOVA quantitative analytical method. This descriptive method will allow me to measure the difference in mean churn among multiple customer segments. My categorical independent variable will be Customer Segment, and my quantitative dependent variable will be Customer Status (active or inactive).

C2a. Justification of Analytical Method

I chose to use the one-way ANOVA quantitative analytical method, as it is a fitting method to use when establishing whether two or more groups, or categories have a statistically significant difference from each other. The one-way ANOVA quantitative analytical method requires that the independent variable must have at least three different categories or groups, as well as one

quantitative dependent variable. These requirements align with my approach as I am seeking to measure the difference in mean churn among multiple customer segments that vary by customer attributes. This method will allow me to measure if there is a statistical significance between segments.

C3. Tools and Environments of Solution

To create my reporting solution, as well as review and manipulate data, I will use Python. Python is a flexible and easy to use language that is ideal to process large sets of data quickly and efficiently, as well as provide meaningful data outputs in a variety of formats.

C4. Methods and Metrics to Evaluate Statistical Significance

To evaluate statistical significance, I will implore a cohort analytical approach and use a descriptive one-way ANOVA analytical method to evaluate if there is a statistically significant difference in how churn rate is influenced when creating customer segments. To do this, I will create three customer segments to review. The first segment will include customers who are of an older age group, who have a higher income and advanced educational backgrounds. I will refer to this group as Segment A. Segment A will include customers who are 50 years old or older, have incomes of at least $80k per year, and possesses a college degree. The next segment will be Segment B and will contain individuals who are between the ages of 26 to 49 years old, have incomes ranging between $40k and $80k per year and possess a college degree. The last segment will be referred to as Segment C and will contain all other individuals who do not meet the criteria to be considered part of either Segment A or B. Customers who have unknown or unreported educational backgrounds or incomes will be excluded from this exercise and will not be included in a customer segment.

Once I have successfully created customer segments, I will test for statistical significance using the one-way ANOVA method to evaluate whether the mean of churn by customer segments is significantly different among varying groups and if any group has a significantly different mean churn than the overall churn rate. The quantitative dependent variable to calculate churn rate will be Customer Status, active or

inactive in conjunction with a count of unique Customer Number by each status. I will then propose a null hypothesis, stating that if there is no difference among segments, or no segment shows a significant difference to the overall mean churn, the hypothesis is null and invalid. Alternately, if one or more groups differs significantly from other segments or from the overall mean churn rate, this will indicate the null hypothesis is void and there is a statistical significance when reviewing churn by customer segmentation. In turn, if the null hypothesis is proved invalid and Segment A is the group that shows the most significant statistical difference, this will indicate my original theory is likely.

C4a. Justification of Methods and Metrics

This approach is fitting as it demonstrates a valid method to test whether a statistical difference is present among customers when using a segmentation approach to reviewing customers when grouping by traits and characteristics.

C5. Practical Significance

The practical significance of identifying customer segments with higher levels of churn will enable JGA Bank and its team to better understand which customers may be at most risk of churn, and better craft its credit card products and programs to the bank's customers' needs. With the insights provided from this analysis, the JGA team will be able to better target which customers the bank needs to investigate further to discover the underlying reasons the customer churned and make more informed decisions to safe-guard revenues derived from their credit card program.

C6. Visual Communication

For my analysis I will be using a combination of bar charts and tables to visualize the percentage of churn across customer segments. Bar charts are useful when viewing information in segments as well as showing trends across categories. Due to this, a bar chart will help communicate the impact each customer segmentation has versus others. Additionally, I will also utilize tables to illustrate more in-depth information such as the number of customers that comprise each segment as well as the attributes covered within the segment, and the total number of customers and percentage of churn for each. Tables are useful

when presenting detailed information that can easily cause other visualizations, such as bar charts, to become cluttered, and in-turn losing their effectiveness to convey information.

D. Description of Datasets

D1. Source of Data

For this project, I retrieved the Credit Card Customers dataset made publicly available for use on Kaggle.com by S. Goyal.

Goyal, S. (2020, November 19). Credit Card Customers. Kaggle. Retrieved January 6, 2023, from https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers

D2. Appropriateness of Dataset

The data set retrieved was appropriate for this projects goal as it contains data similar in nature to what a financial institution would have on record for their clients. The data contained within the source file, contains customer demographic attributes such as gender, age, income, education, dependents as well as financial attributes such as credit limit and credit level. The data provides a variety of attributes and measures that can be used to analyze churn as it is related to customer qualities and behaviors. Due to this, I felt the dataset was ideal for a project of this nature, as the intent of the project is to look for trends to guide the client in directions they may want to further investigate as they look to fully understand and address churn.

D3. Data Collection Methods

Data collection for this project was conducted via a download from Kaggle.com. Retrieving data from Kaggle was beneficial as datasets provided through the website are intended for data analytics projects. Kaggle has a good reputation in the analytics world and is recognized as a reliable source of training datasets. Generally, datasets receive a usability score from other users, in this case the dataset used for this project received a high score of 10, meaning the data contained within the file has high likely hood of being useful. Additionally, by retrieving a dataset from Kaggle, the data is publicly available and does not contain sensitive information that would require approvals, or risk compromise of classified information.

D4. Data Quality

Overall, the quality of data present within the source file is of a clean nature. All rows and columns within the dataset are populated with relevant data, and there are no nulls present. All data points appear to be populated with relevant values in each respective column. Due to the nature of this analysis and the need to group individual customer records into corresponding segments, there is a need to add additional columns to the data to help group and organize customer records. For my analysis I chose to focus on creating customer segments that are based on customer age, education level and income category, as well as customer status (active or inactive). Because of this, there will be many columns rendered unnecessary in the dataset, and will be discarded during the coding process, leaving only necessary inputs.

D5. Data Governance, Privacy and Security, Ethical, Legal, and Regulatory Compliance

Due to the source of data for this analysis, there was no confidential information available within the source data. However, as this analysis is aimed at the financial industry, specifically banking, customer and company confidential information should be of high consideration if this exercise were to be practiced in a real-world setting. If the analysis were to be performed for an actual bank, it is likely that Social Security Numbers, credit scores, addresses and other customer sensitive information could be contained within the data and would require the utmost security and concern when conducting an analysis. In addition to customer information, the data would also likely reveal a considerable amount of detail regarding the financial health of the institution conducting the analysis and would likely be highly regulated. With these considerations in mind, a project of this nature would require considerable security and planning of precautions to safeguard data.

D5a. Precautions

While the data used in this analysis is fictional, I excluded any customer attributes that could be specific to an individual customer and could be used to identify a person. This approach will allow for the reporting solution to function in a way that satisfies the required outputs, without compromising sensitive information. The output from this application will be sensitive in nature

to the organization conducting the analysis but will not contain information that if it were breached, would identify the institution or customers involved in the research. As such, the information contained in the output can be distributed amongst all parties involved in the project.

E. Sources

Goyal, S. (2020, November 19). Credit Card Customers. Kaggle. Retrieved January 6, 2023, from https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers

Frankenfield, J. (2022, May 18). Churn Rate. Retrieved January 6, 2023, from https://www.investopedia.com/terms/c/churnrate.asp

Fullterton, A. (2020, July 20). Why Churn Analysis is the key to improving Customer Retention. Retrieved January 6, 2023, from https://amplitude.com/blog/why-churn-analysis-is-the-key-to-improving-customer-retention

Dube, D. (2020, May 27). Why retaining customers for Banks is as important as winning new ones. Retrieved January 6, 2023, from https://www.forbes.com/sites/forbestechcouncil/2020/05/27/why-retaining-customers-for-banks-is-as-important-as-winning-new-ones/?sh=15c7fe883f98