Capstone Analytics Plan

Diana Paola Vega Feriz

Centennial College

Business Analytics Capstone

Bilal Hasanzadah

July 15, 2022

Analytics Startup Plan

Synopsis: ***This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

Project	*Prediction of Credit Card Approval*
Requestor	*Centennial College*
Date of Request	*July 13, 2022*
Target Quarter for Delivery	*August 19, 2022*
Epic Link(s)	*Not Applicable*
Business Impact	*Reduce the credit risk for the institution in credit card approval process by ensuring efficient approvals.*

1.0 Business Opportunity Brief

Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve for. An important step is to understand the nature of the business, system or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment.

Taking into account the high and manual steps involve in credit card approvals such as a wrong approval or missing a right candidate, can both have significant negative financial impact for the company.

The objective is to develop a right analytical tool to ensure efficient credit card approval process and therefore, decrease the financial loss by missing to select a good candidate and on the other hand increasing the credit risk by approving a bad candidate.

The specific ask:

To make more accurate the credit card approval assessment.

1.1 Supporting Insights

Define any supporting insights, trends and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of market, nationally and regionally?

This dataset is related to the credit card approvals which have been more requested over the years. These days, credit cards are not exclusively linked to banking sector but also there are many commercial companies such as airlines, supermarkets, among others that issue its own credit cards.

According to Shift Credit Card Processing (2021) as per August 2021 there were 2.8 billion credit cards in use worldwide of which 1.06 billion were in use in United States of America. In research from TheGlobalEconomy.com (2022) Canada accounted in 2017 the highest percentage of people aged 15+ who have a credit card with 82.58%.

Currently there are 4 major credit card companies, Visa, Mastercard, Discover, and American Express being the two first the biggest.

In this industry, the approval of a right candidate with healthy background and credit history is extremely important to avoid mid and long-term risks for the issuing companies.

1.2 Project Gains

Describe any revenue gains, quality improvements, cost and time savings (as applicable). What will you do differently and why would our customers care. What are the implications if we do nothing? *This section is particularly key for prioritization against company goals and KPI's.*

By conducting this analysis, I will help improve the credit card approval assessment since the issuing companies will be able to predict who is a good candidate and who is not.

Making an accurate assessment reduces the credit risk and losses caused by the customers' default and increases earnings by issuing credit cards to those who are good candidates.

Also, since this process still manual in many businesses, the assessment is error-prone and time-consuming, so this model will automate the approval process and therefore improving and achieving the KPI's of accuracy of the assessments and responses time.

Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority and data availability.

2.0 Analytics Objective

List the key questions, assumptions and define the hypotheses. Often the deliverable may not just be an analysis output, however a recommended operating model or blueprint for a pilot etc.

Note: Asking the right questions and truly understanding the problem will lead to the right data, right mathematics, and right techniques to be employed.

Since companies issue credit cards based on the applicant's personal and historic information, my objective is to predict if an applicant is a "good" or "bad" candidate thus if they can issue a credit card or not.

2.1 Other related questions and Assumptions:

List any assumptions that may affect the analysis

I am assuming that this dataset (cleaned version of the original) has not omitted any important and relevant from the original database.

2.2 Success measures/metrics

What does success look like? Define the key performance indicators (success definition/indicators, drivers and key metrics) against which the objectives will be analyzed. These should be drawn from the interlock meeting with key stakeholders and will inform the approach and methodology for the analysis.

1. To assess each model for its accuracy and select the best model.
2. To achieve the accuracy of my best model above 85%.
3. To apply minimum 3 different algorithms to the model.
4. To determine which variables have the greatest impact on the positive outcome.

2.3 Methodology and Approach

Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology.

Type of Analysis: I will implement a binary classification model. The initial algorithm I will use is decision tree to determine which demographics variables are most significant for the approval of the credit card. Also, I will use other techniques and algorithms to validate these findings.

Methodology: I will start running a descriptive analysis of my data to understand its content and clean my dataset. Then I will define the target variable to be 1 for an approved application and 0 for a rejected application. After that, I will transform my categorical variables into dummy variables. I will build a decision tree, logistic regression, and K-Nearest Neighbors algorithms. I will make confusion matrix to evaluate its accuracy.

Output: my output will be a set of insights that will help me to assess whether a potential customer should be issued a credit card or not.

3.0 Population, Variable Selection, considerations

Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today.

Audience/population selection: Not applicable

Observation window: Not applicable

Inclusions: I will include 20 variables out of the 21 in the dataset.

Exclusions: I will exclude Applicant_ID variable from my analysis.

Data Sources: I am using the dataset of Mario Caesar in Kaggle “Credit Card Approval Prediction (Cleaned Version) https://www.kaggle.com/datasets/caesarmario/application-data?select=Application_Data.csv.

This is a cleaned version of the original dataset from Seanny (@rikdifos) also in Kaggle who merged two different tables by the ID of the customers.

Audience Level: my stakeholders are the Risk Management Department of the credit card issuing company.

Variable Selection: Applicant_Gender, Owned_Car, Owned_Realty, Total_Children, Total_Income, Income_Type, Education_Type, Family_Status, Housing_Type, Owned_Mobile_Phone, Owned_Work_Phone, Owned_Phone, Owned_Email, Job_Title, Total_Family_Members, Applicant_Age, Years_of_Working, Total_Bad_Debt, Total_Good_Debt, Status (target variable).

Derived Variables: I will have derived variables from the dummy of the categorical variables Income_Type, Education_Type, Family_Status, Housing_Type, and Job_Title.

Assumptions and data limitations: My limitation is that the data I am using is not from a specific credit card issuing company, so I do not have enough background.

4.0 Dependencies and Risks

Identification of key factors that may influence the outcome of the project and likelihood of it happening:

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
This dataset is a cleaned version of the original one. If this cleaning was not done appropriately, it may create bias in the outcome.	Low	Not applicable	Minimum impact expected as I will run a data exploration analysis to mitigate this risk.

5.0 Deliverable Timelines

List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process.

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	Kick-off / Formal Request	Get the first consultation with my advisor to present my preliminary analysis and possible dataset.	Get the approval of the dataset.		July 15
2.	Analysis Plan	Make the analysis plan based on the template of Professor Mustafa.	Build an accurate and achievable plan.	3	July 15
3.	Data Exploration & Analysis	Make the data exploration to identify some issues that could arise and to ensure the data is cleaned.	Get the dataset ready to start modeling.	7	July 22
4.	Modeling	Run different models and algorithms. Also, check the accuracy of those models.	Define the best model which fits best to my objective.	14	August 5
5.	Documentation	Prepare the complete documentation of the analysis.	Include all the models and algorithms run.	7	August 12
6.	Peer Feedback	To discuss the overall project with the selected peers.	To have their feedback.	7	August 12
7.	Presentation	Prepare the power point of the presentation to stakeholders.	Convince the stakeholders the efficiency of the model.	10	August 22
8.	Portfolio	Update my portfolio with the capstone documentation and the github of the notebook.	Make the portfolio ready so possible employers can access to it.	4	August 26

References

Shift Credit Card Processing. (2021, August). *Credit Card Statistics [Updated August 2021] Shift Processing*. Retrieved July 14, 2022, from https://shiftprocessing.com/credit-card/

TheGlobalEconomy.com. (2022). *Percent people with credit cards by country, around the world*. Retrieved July 14, 2022, from https://www.theglobaleconomy.com/rankings/people_with_credit_cards/