**Analytics Plan**

**Sebastian Gonzalez – 301336125**

**Centennial College**

**Bilal Hassanzadah**

**Business Analytics Capstone**

**July 15, 2024**

Contents

[**1.0** **Business Opportunity Brief** 3](#_Toc171977368)

[**1.1 Project Gains** 3](#_Toc171977369)

[*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.* 3](#_Toc171977370)

[2.0 Analytics Objective 3](#_Toc171977371)

[2.1 Other related questions and Assumptions: 3](#_Toc171977372)

[2.2 Success measures/metrics 4](#_Toc171977373)

[2.3 Methodology and Approach 4](#_Toc171977374)

[3.0 Population, Variable Selection, considerations 4](#_Toc171977375)

[4.0 Dependencies and Risks 5](#_Toc171977376)

[5.0 Deliverable Timelines 5](#_Toc171977377)

## **1.0 Business Opportunity Brief**

This capstone aims to identify the most important variables determining credit card approval acceptance and using a database with different customer information. This project will create insights to create a strategy based on customer characteristics. Additionally, bank users sometimes do not understand why their application was rejected. This project will also help to identify the main variables for rejecting these applications. Data analytics practices will uncover these insights.

## **1.1 Project Gains**

This project will help banks to develop strategies based on their customer’s information. The insights created on this capstone will also help develop new types of financial products. The data analytics practices that will be used in the project will determine which one is the most accurate for this type of business case and the most aligned with the business strategy.

## **2.0 Analytics Objective**

To achieve a targeted analysis, it is important to determine which variables are the most important when customers apply for a credit card. Based on data analytics models will be determined which are the most important. Also, it is important to make assumptions and test some hypotheses to get more insights during the project.

## **2.2 Success measures/metrics**

Our main indicator or key metric to determine or qualify the results is the percentage of the accuracy of each model. Additionally, we will include statistical metrics which will give additional insights. For example, mean squared error.

## **2.3 Methodology and Approach**

**Type of analysis:** Logistic Regression, decision trees, neural networks.

These are the data analytics models which will be used to address the problem. It is important to determine which mode has more accuracy and which fits better in the business case.

**Methodology:** Data analytics models through Python.

Python would be the data analytics tool use for all the stages of the capstone. Starting with data exploration. It is important to understand the data, apply different methods to clean our database. After data exploration we will apply the data models in python to seek the insights for project.

**Output:** The output will be which is the best data model to be considered, the most important variables for the model and recommendations for future business strategies.

## 3.0 Population, Variable Selection, considerations

**Population selection:** Bank customers.

**Data Sources:** Data was taken from Kaggle.

Link: https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction?resource=download&select=application\_record.csv

**Audience Level:** Banks or people interested in banking

**Variable Selection:** The database includes around 30 variables for each customer. All variables will be considered for the analysis. Through the data analytics models, execution will be determined which of them are going to be included.

**Assumptions and data limitations:** There is a data limitation on the number of customers that were rejected when they applied for credit card. This number of customers on the database is lower than 20%. This could create issues for training the models.

## **4.0 Dependencies and Risks**

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Delay (based on historical data)** | **Impact** |
| Low rejection cases on the database | Low |  | We can experience some issues when training the models. Through data analytics tools we can manage this issue. |

## **5.0 Deliverable Timelines**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item** | **Major Events / Milestones** | **Description** | **Days** | **Date** |
| 1. | Analytics plan | Create an analytics plan. | 7 | *Jul 7 to jul 15* |
| 2. | Data Exploration | Explore dataset and use data analytics tools to have a proper database align with our purpose. | 15 | *Jul 7 to Jul 21* |
| 3. | Modeling | Use data analytics models to achieve our main objective. | 15 | *Jul 21 to Aug 4* |
| 4. | Conclusion and Insights | Analyze data analytics models and made final conclusions and insights. | 7 | *Aug 4 to Aug 12* |
| 5. | Presentation | Present the project. | 1 | *Aug 14* |