

Credit Card Approvals Final Project

Executive Summary

A dataset was chosen that provides individual credit card applicants' statuses, from total income and occupation type, to whether they own realty and have children. This document covers the goals in completing this project, the scope of the project, functional requirements, personnel requirements, and the delivery schedule that I will adhere to to deliver on my project successfully. Contingencies for risks to the course of the project are also outlined at the end of this document.

Business Objectives

This project will give me more of a real-world opportunity to be able to plan my project and follow through from beginning to end. I will use what I've learned in the Academy, as well as pick up some skills from others' code to wrangle, explore, and analyze the data.

First, I'll determine if an applicants' status as it pertains to owning a car, owning realty, length of employment, education level, and occupation, influences their annual income. In this way, I can visualize the data as is and attempt to draw correlations, if any. Then, I will use machine learning to create a model that will aid in approving credit card applicants who present low risk of defaulting on their credit card bills. While going through this process, I'll have to establish my own method of using different variables to either approve or deny an application.

By analyzing the data and developing a model to approve credit card applications, I can use the tools and skills I've gained during my time in the Academy to help clients in the future who want to improve upon current processes.

Background

Financial institutions need to vet applicants thoroughly before approving an applicant for a credit card. The institution is taking on risk by granting consumers credit cards. Many variables go into the application process, and we know that credit worthiness, or the credit score, weighs heavily on the institution's willingness to issue a credit card to an individual. However, are there other factors that can improve the chances that an individual can successfully manage their credit without defaulting? Machine learning may be able to help

find a better model to be used when determining which applicants are the highest risk to a credit issuing institution.

Scope

The project encompasses data analysis and machine learning using both R and Python. Tableau is also an option if some visuals cannot be captured best using R or Python. Kaggle user notebooks may be referenced for ideas as the project progresses.

Analyses to be used for this project include independent t-tests, MANOVA, linear regression, and multiple regression. If any analyses can be excluded from the list or included to improve the project, I will do my best to plan for them.

Outside research might be needed as it pertains to other ways credit card issuers make their judgements in approvals. Flexibility in availability to work and seek guidance on this project are of great importance.

Functional Requirements

- *GitHub Final Project repository to include Assignments, Code, Notebooks, Data, and Media
 - *Within GitHub, a README file will be maintained week by week with a to do list and retrospective**
- *Exeter LMS assignments completion*
- *R will be used as part of the data analysis portion of this project (simple linear regression, logistic regression, MANOVA)*
- *Python will be used in the data wrangling, independent t-test, correlations, and machine learning process.*
- *Data was collected from Kaggle.com and is open source.*
- *The variables of interest to this project are: gender, owns car, owns realty, total income, education level, years employed, occupation, number of children, and approval status.*

Personnel Requirements

This will be an individual project. I will be responsible for the completion of all curriculum assignments, wrangling, coding, analyses, GitHub management, research, and presentation. A minimum of 20 hours per week will be dedicated to the project.

Delivery Schedule

Week 1

Project planning – GitHub repository, meet with instructor, choose dataset, LMS Lesson 1 completion

Week 2

Data wrangling – subset, identify any missing values, rename columns, recode, data type conversions, data unit conversions, etc.

GitHub updates, Meet with instructor, LMS Lesson 2 completion

Week 3

Data exploration – histograms, pie charts, variable distributions, correlations.

GitHub updates, Meet with instructor, LMS Lesson 3 completion

Week 4

Data analysis – independent t- test, MANOVA, linear regression, multiple regression

GitHub updates, Meet with instructor, LMS Lesson 4 completion

Week 5

Data Visualization – histograms, pie charts, variable distributions, Tableau.

GitHub updates, Meet with instructor, LMS Lesson 5 completion

Week 6

Data Reporting and Presentation – put together slide decks

GitHub updates, Meet with instructor, LMS Lesson 6 completion

Other Requirements

None foreseen.

Assumptions

Programming languages are assumed to be working without bugs. Analyses are able to be run without errors or failure to pass assumptions. Mentor help is available when needed.

Limitations

Something may come up in my personal life to where I can't devote all the time necessary to complete a quality project. Potential changes at my full-time job can derail my progress due to having to work longer hours. I may come across something that I need outside help with to get past.

Risks

Technical issues may arise, such as loss of computer, internet, or programming language failure. I could get sick or hurt and not be able to complete the project.

In order to mitigate these risks and limitations, I plan to take care of myself both physically and mentally above all. If I lose an employee at work during this time and I have to work longer hours, I'll have to be as flexible as possible to continue working on the project. If I need outside help, I can't be afraid to ask for it from the instructor and or other mentors. A loss of computer could be avoided by taking care of my current one and backing up files to OneDrive. A loss of internet for extended periods of time or failures of R and Python are out of anyone's control. Any impediments to the progress of the project will be communicated to the instructor promptly.