CREDIT DEFAULT ANALYSIS

Project Process and Initial Data Analysis/Recommendations

Kevin Burr

April 2020

WHAT WE WILL COVER

- Background
- Objective/Goals
- Data Science Process
- Data Management
- Data Summary
- Known Data Issues
- Recommendations
- Questions

BACKGROUND

CONTEXT OF PROGRAM

Your Goal: Identify a reliable, statistically sound method of identifying credit accounts at risk of default before it happens to stem total potential losses.

We will analyze sample data collected to identify potential patterns in the data that may indicate an account owners propensity to default before default occurs.

OBJECTIVE/GOALS

SUMMARY OF REQUEST

For this project, we will:

- Analyze a very large set of credit information for existing accounts including account owner demographics, credit limit, and past payment activity.
- Provide an overview of the data analysis.
- Provide recommendations for identifying potential at risk accounts at CreditOne.

DATA SCIENCE PROCESS

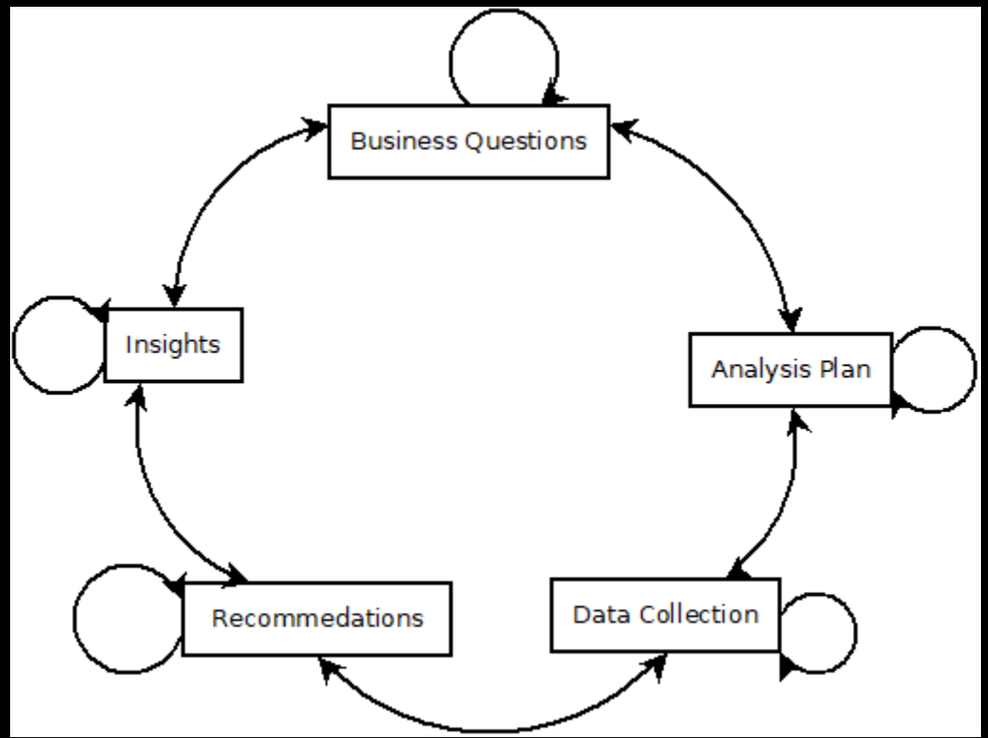
SUMMARY OF PROCESS

For this project, we will follow the BADIR process:

- B(usiness Questions) – ensure an understanding of the business question, why this is important, and any considerations that could impact the analysis
- A(nalysis Plan) – understand the goal of the analysis, the hypotheses, data required and timeline
- D(ata Collection) – from where is the data, what data cleansing is required for analysis
- I(nsights) – identify patterns in the anlysis, validation of hypotheses, measured confidence in analysis, and a ranked assessment of the outcomes of the analysis to business importance
- R(ecommendation) – any recommendations based on the analysis

PROCESS FLOW

DIAGRAM OF ANALYSIS PROCESS



Find the real and actionable Business question	Formulate a hypothesis-driven Analysis plan	Collect relevant Data based on the Analysis Plan	Derive Insights using machine learning & statistics	Drive KPI's with actionable Recommendation

DATA MANAGEMENT

DATA SECURITY DURING COURSE OF PROJECT

You customer data is protected by:

- Username and password protection on the data analyst's laptop, which is also protected through hard drive encryption software
- All customer data will be removed from the company hardware upon project analysis completion
- In adherence with any additional requirements based on your own internal data security and handling compliance requirements

DATA SOURCES

SUMMARY OF DATA

For this project, we will:

- Have access to an historical data set representative of CreditOne clients which includes demographic, balance limits, and payment activity over time.
- Data for project analysis will be provided by FTP, email, or other internally/externally protected/encrypted processes in compliance with all in internal procedures.
- Data for analysis is presumed to be primary contained in the file: “default of credit card clients.csv”

DATA SUMMARY

DESCRIPTION AND LOCATION OF PROJECT DATA

Data included for analysis:

- 30,000 client accounts with complete (no missing) attribute values for each record (April to September 2005)
- Each observation includes the balance limit and demographics about the borrower
  - Balance Limit – the amount that can be borrowed; currently approved limit to borrow for the account
  - Sex – gender of borrower
  - Education – education level of borrower
  - Marriage – marital status of borrower
  - Age – age of the borrower
- Payment details (April to September, 2005)
  - A history of past payment status – no consumption, paid in full, or number of months delayed payment
  - A history of monthly balances at statement time – amount in dollars each month at statement billing period end
  - A history of monthly payments – total amount paid each month by borrower toward balance
  - A value for default/not default – 1/0 to indicate whether the account is in default or not

KNOWN DATA ISSUES

ISSUES AND REMEDIATION PLAN

- April – September 2005 provided
  - No observed issues with the data based on summary review
- Note: data is 15 years old – it is possible more current data would provide a better representation of customer performance based on current market dynamics
  - Recommendation: provide a more current data set at some point for further analysis.

DATA SUMMARY

STATISTICAL SUMMARY OF ANALYZED DATA

Account Status Summary

Total Accounts: 30,000
Default Accounts: 6,636
Accounts in Good Standing: 23,364

Customer Demographics

Age: Min: 21, Max: 79, Mean: 35.49
Sex: M = 11,888, F = 18,112
Education:
- Grad School: 10,585
- University: 14,030
- High School: 4,917
- Other: 468

Marriage Status:
- Married: 13,659
- Single: 15,964
- Divorced: 323
- Other: 54

RECOMMENDATIONS

ENHANCEMENTS FOR MEASUREMENT DATA

- Updated data
  - Provide more current data for analysis that more accurately models current market/customer conditions
- Expand Demographics
  - Include general geographic identifiers for customers to more accurately model current market/customer conditions by geography – different geographies tend to perform differently (i.e. Detroit versus Austin)
- Include Profession
  - Different professions may have different payback profiles, possibly even by geography
- Include general market data
  - Performance of the market overall and by geography might indicate customer performance

QUESTIONS

-