Course Two Task One: Framework

***Framework Two -*BADIR***(***Jain and Sharma, *Behind Every Good Decision,*chapter 4):**

**B**usiness question

* What is the stated business question?

Credit One needs a better way to understand how to set a customer credit limit or, at the very least, if someone should be approved or not.

* What is the intent underlying the question (e.g., what is the context, what is the impacted segment, and what are stakeholders’ current thoughts about the underlying reasons?

We need to correct a trend of defaulting customers and determine if customer credit history can help predict future customer loan limits/approvals.

* What business considerations (e.g., stakeholders, timeline, and cost) are likely to impact the analysis?

Critical importance to the business requires determination as fast as possible.

**A**nalysis plan

* What is the analysis goal?

To build a linear regression predictive model to help set a credit limit for a customer, or a classification model to determine if they should be extended credit at all.

* What hypotheses are to be tested?

That customers with similar data points will behave in a similar way. That customers with a good credit history are good candidates for higher credit limits.

* What data is required/available to test the hypotheses?

We have data from roughly 3000 customers that describes the following:

-Amount of credit given

-Gender

-Marital Status

-Level of higher education obtained

-Age

-6 months (April-September) of payment info including:

-History of past payment status ( -2=No consumption, -1=Paid in full, 0=Use of revolving credit, No payment made for x number of months.)

-Amount due per month

-Amount of previous payment

-Status of default (Y/N)

* What methodology(-ies) will you employ?

I will use data analytics techniques to clean the data as well as select or engineer features from the data set to help build an accurately predictive linear regression model.

* What is the project plan (timeline and milestones, risks, phasing, prioritization, …)?

Framework and Project Outline due in One week

Data EDA due in three weeks

Build and evaluate models, due in 5 weeks

Present work, due in 5-6 weeks

**D**ata collection

* From where can the data be obtained?

Obtained from an SQL database owned by Credit One

* How must the data be cleansed and validated?

The data needs to be cleaned by correctly labeling/renaming columns, locating and deleting duplicate rows, imputing missing values, assigning coded numeric values in place of categorical values, and correcting dtype.

**I**nsights

* What patterns do you see in the data?

TBD

* Are each of the hypotheses proven or disproven?

TBD

* How much confidence should stakeholders place in the results?

TBD

* How do you rank your findings in terms of quantified impact on the business?

TBD

**R**ecommendation

* How can you most effectively present the results of your analysis to your stakeholders (in terms they can understand and in alignment with information they’ll value)?

TBD

* Note: A generic template for a recommendation presentation or report might include:

TBD

* Objective
  + Background (optional)

The concerning trend in customers defaulting on credit payments is harming the business model of our third party credit partner, therefore threatening our own profitability.

* + Scope (optional)

The solutions determined by the outcome of this project will affect the success of credit approvals for all future customers.

* + Approach (optional)

Data analysis of 6 months worth of previous credit history of existing customers, cleaned, explored and analyzed for the purposes of creating a regression model to better predict the best credit limit for new customers.

* + Recommendations
  + Key insights with impact
  + Next steps