CREDIT ONE – ANALYZING CUSTOMERS

BY OMAR CHAVEZ

THE PROBLEM TO BE SOLVED

- Credit One is seeing an increase in the number of customers that have defaulted on their loans

- How do we decide whether to approve someone for a loan or not?
- If they are approved, how much do we credit do we approve them for?

THE DATA SCIENCE PROCESS

- Business Question – How do we determine which customers to extend credit to?
- Analysis plan – Analyze the dataset to look for patterns that occur in customers that default, and in those that do not default. Then use decision tree algorithms to determine if a given customer is credit worthy or not
- Data Collection – Data is coming from Credit One's SQL database. The data will be removed of duplicates and sorted to prepare for analysis.
- Insights – What are the attributes of a customer that defaults on their loan? Given a new customer, how accurate is the decision tree algorithm in determining whether that customer will default or not?
- Recommendation – What are the key insights from the analysis? What kind of customer can we recommend that Credit One does or doesn't extend credit to?

THE DATA

- Data comes from the CreditOne SQL Database
- Features of the Data:
  - Amount of given credit
  - Male or Female
  - Education Level
  - Marital status
  - Age
  - Repayment status for 6 past payments
  - Amount of 6 past statements
  - Amount of 6 past payments

DATA MANAGEMENT

- Data will be securely managed by the Data Science team to ensure privacy of the customers is maintained
- No personal data will be include in the analysis
- Data will be analyzed and managed using Jupyter notebook and excel

ISSUES WITH THE DATA

- Data provided will need to be modified in following ways:
  - Data will need to be sorted
  - Headers will need to be modified (Currently data has two headers)
  - Duplicated will need to be removed
  - Any blank data will need to be given a value
  - The ‘Sex’ and ‘Education; fields are currently strings, so they will need to be assigned an integer value
    - Gender: 1 = male, 2 = female
    - Education: 1 = graduate school, 2 = university, 3 = high school, 0,4,5,6 = others

OVERVIEW OF THE DATA SCIENCE PROCESS