Determining Creditworthiness with Data Set

Proposed Data Analysis Approach

Tom Snyder May 2021

Goal:

Determine if an applicant should be approved for a line of credit

- Credit One desires to better evaluate those applying for a line of credit to reduce the number of defaults
- We will use the provided data set from Credit One
- **Plan**: Use machine learning to create a model that will classify applicants as PASS or FAIL

Proposed Framework: Zumel & Mount

- This framework was chosen because it includes:
  - Model-building
  - Model assessment and refinement
  - Explanation and presentation to stakeholders
  - Model deployment
  - Maintenance of model

Data Source:

SQL Database provided by Client

- Data will be imported from the SQL database into Jupyter Notebook to be manipulated via Python.
- Data is a set of applicant records with features including demographics and credit history.

Data Management:

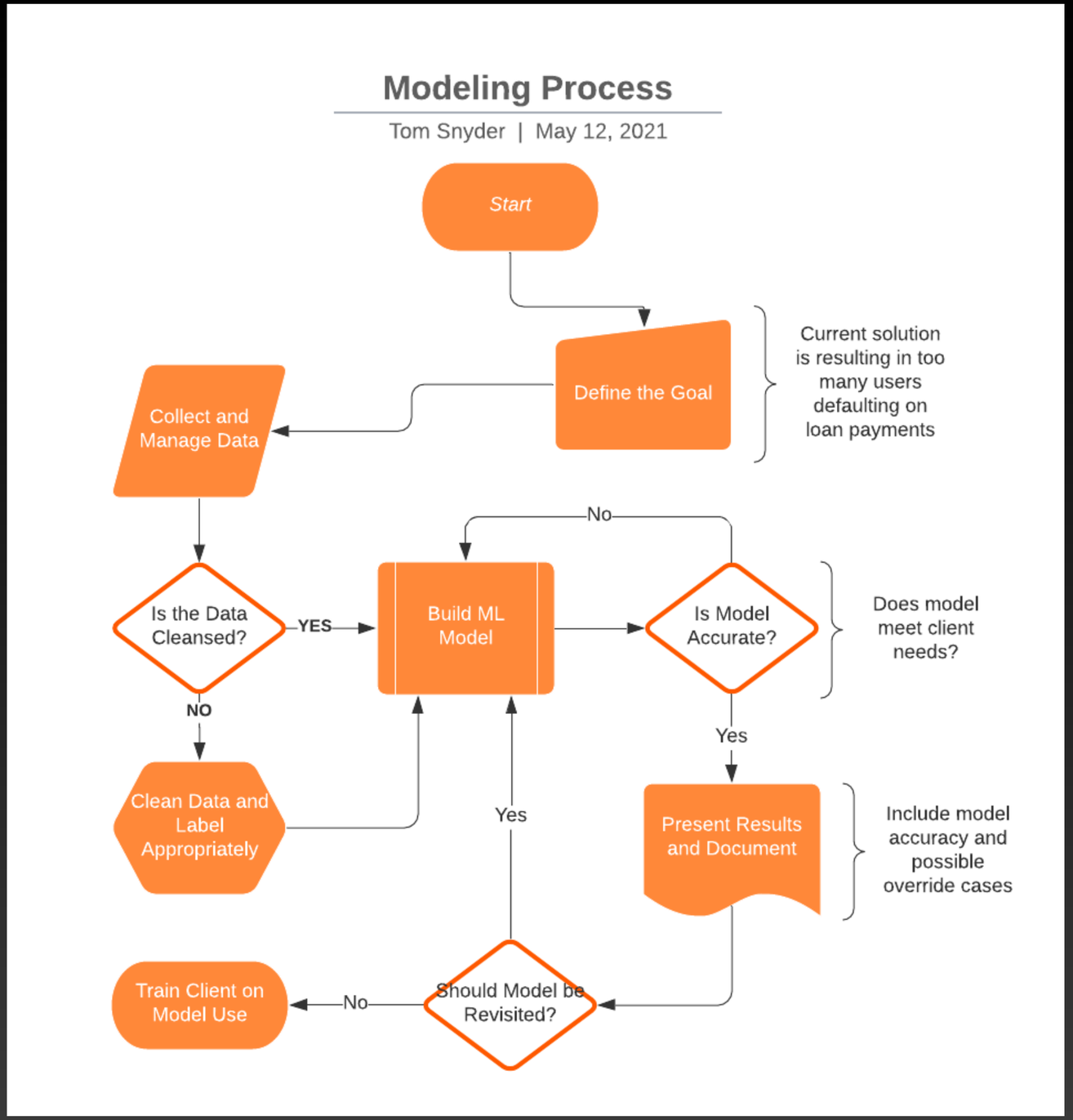
- Data will be munged and analyzed via Jupyter Notebook
- Data will be cleansed for duplicates, missing values, etc.
- **Visualizations**
  - Demographic ratios
  - Averages of relevant payment history
- **Models**
  - Explore how features affect creditworthiness
  - Multiple model types will be built to identify the most accurate for predictions

Known Issues

Initial exploration has revealed several issues

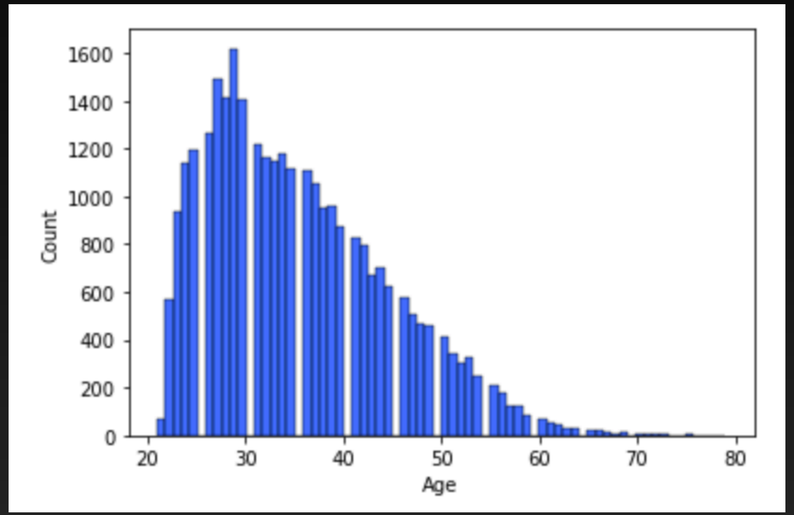
- The issues and proposed solutions include:
  - **Issue**: Data is in SQL database
    - **Solution**: Import into Jupyter Notebook to manipulate with Python
  - **Issue**: Data is improperly indexed
    - **Solution**: Reassign index headers to fit column names, remove duplicate header names
  - **Issue**: Data contains values as strings, rather than numbers
    - **Solution**: Identify numerical values and convert them

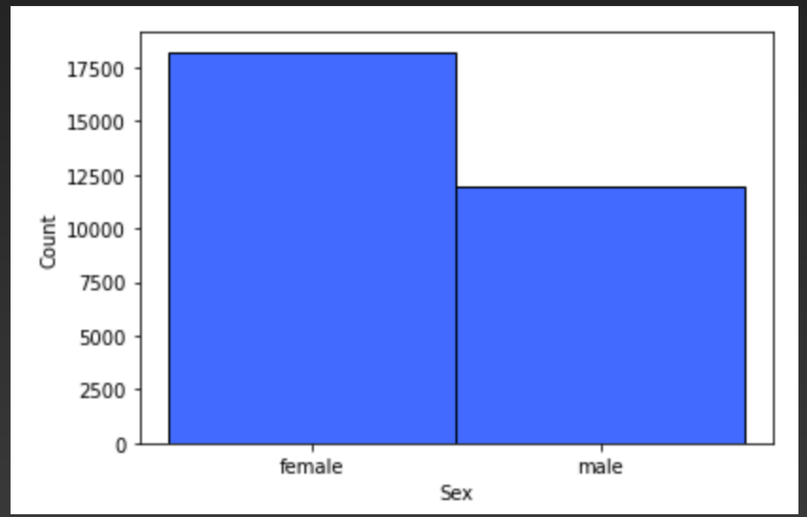
Modeling Process to be Followed

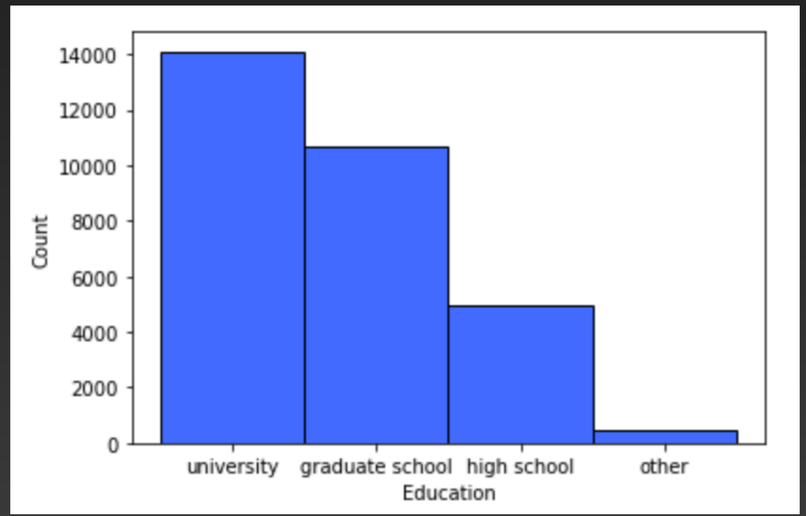


Initial Insights

Data skews young, female, and educated

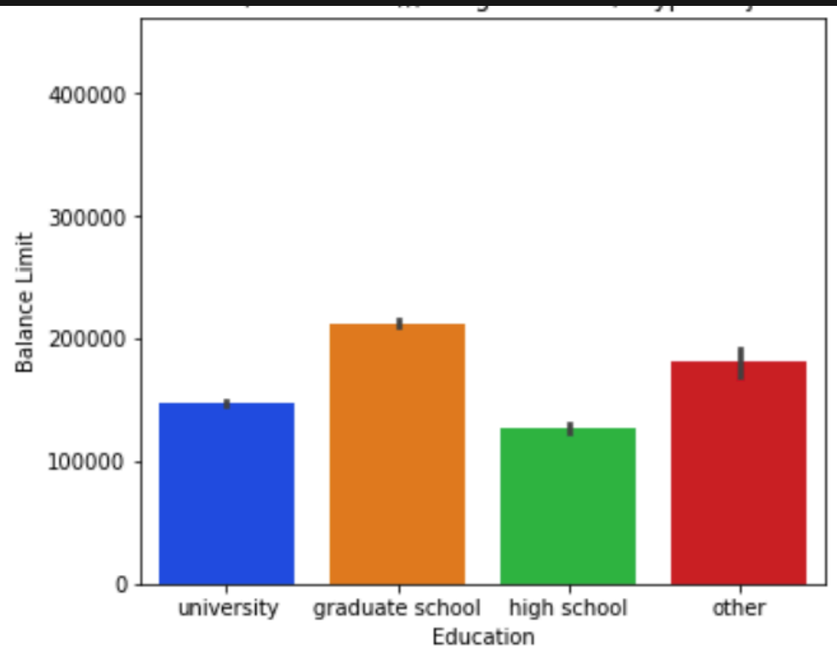


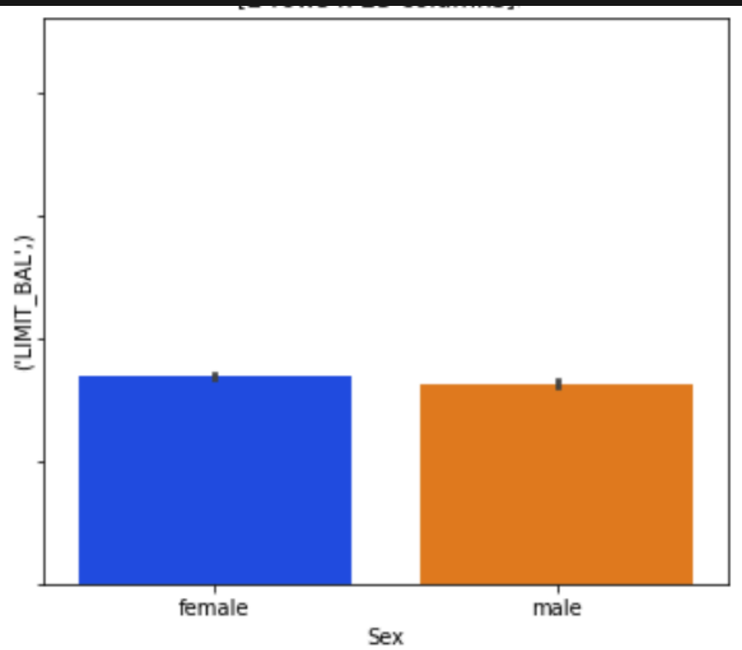


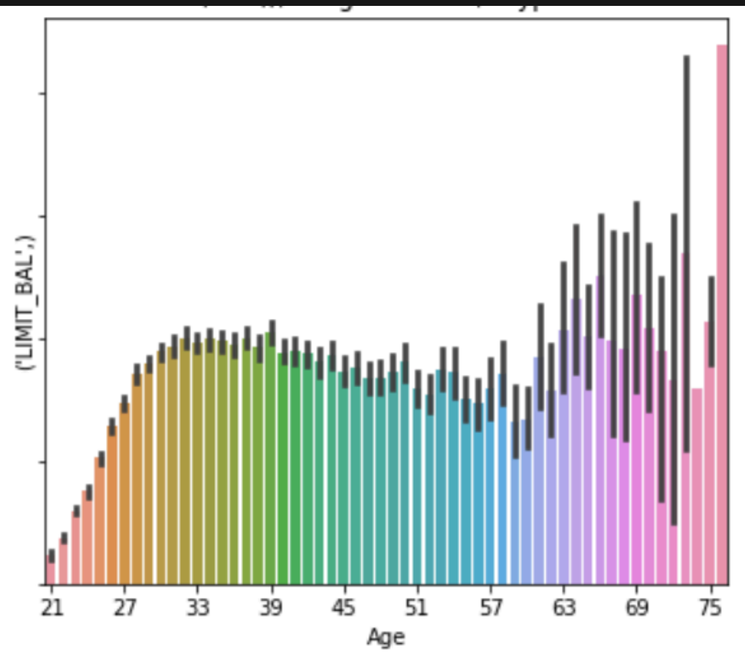


Initial Insights

Sex has negligible impact on credit limit balance







Initial Insights

Lower credit limit customers default more often