Standard Bank

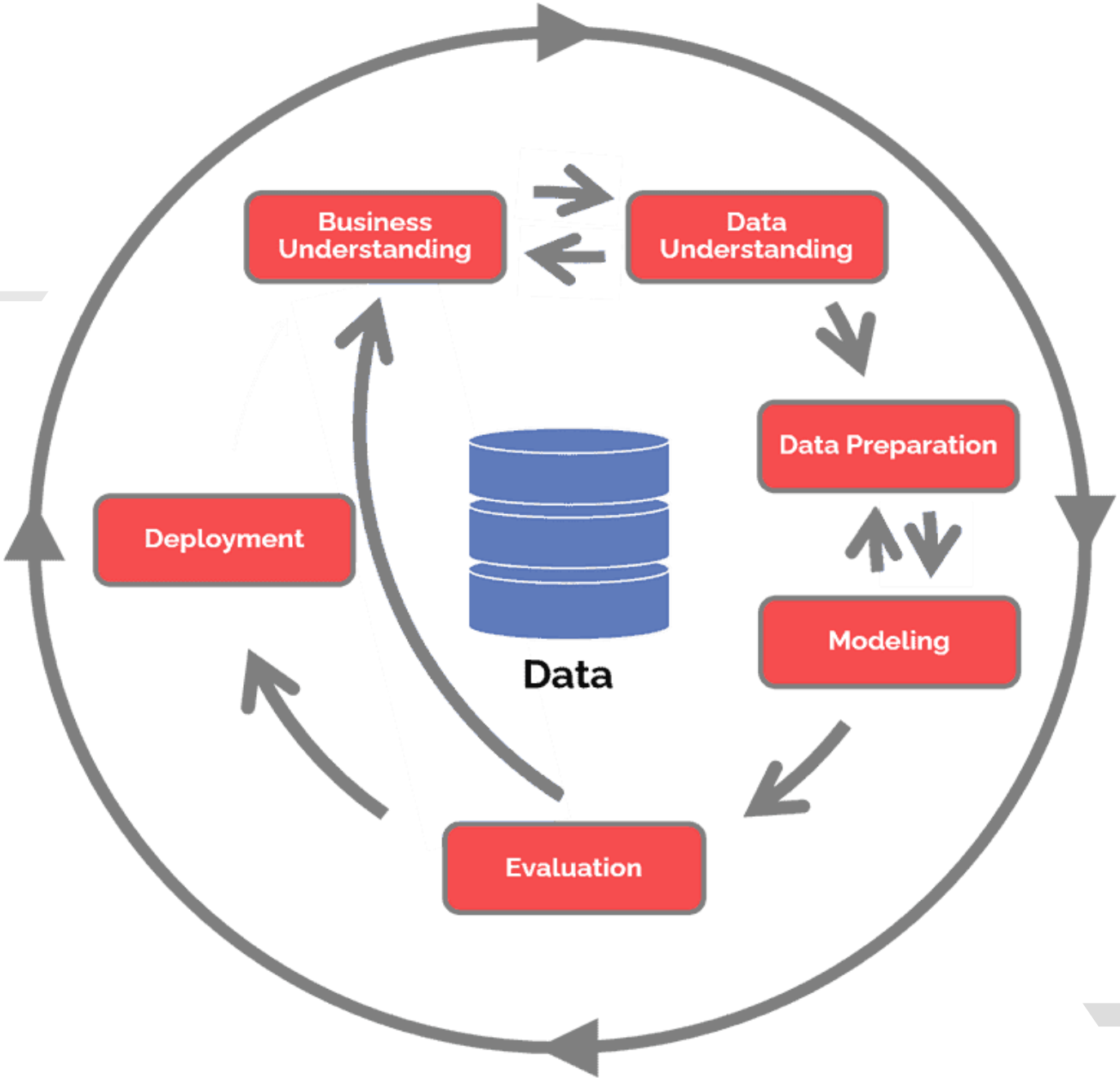
Home Loan Data Science Project

31-Oct-22

Agenda

- **Data Science Lifecycle**
- **Project Overview**
- **Data**
- **Analysis**
- **Modeling**
- **Model Evaluation**
- **Recommendations**

Data Science Lifecycle



Project Overview

♦ **Business Problem**:

At the moment, applying for a home loan is a laborious procedure. It takes 2 to 3 days, so the applicant won't learn the results of their application until after those 2 to 3 days.

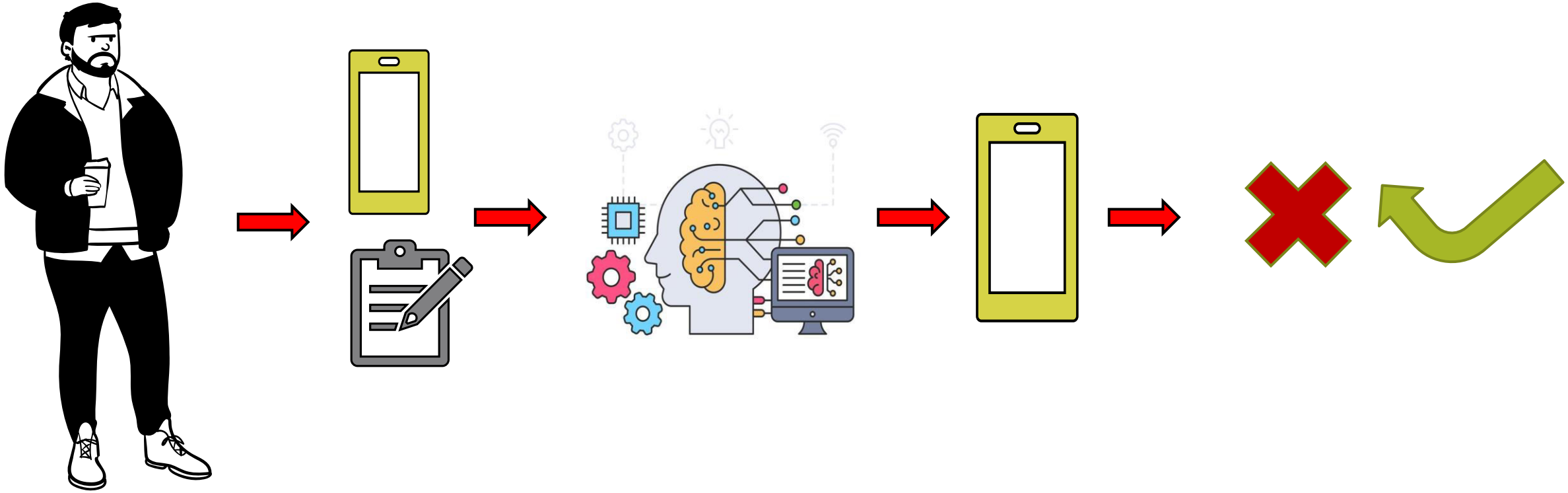
♦ **Business Objective**:

Help the user by getting information regarding the status of their loans in a matter of seconds.

♦ **Hypothesis**:

Machine learning may be used to forecast a future borrower's loan status based on historical data, greatly reducing the time it takes for them to discover their separate statuses.

Process Overview / Solution

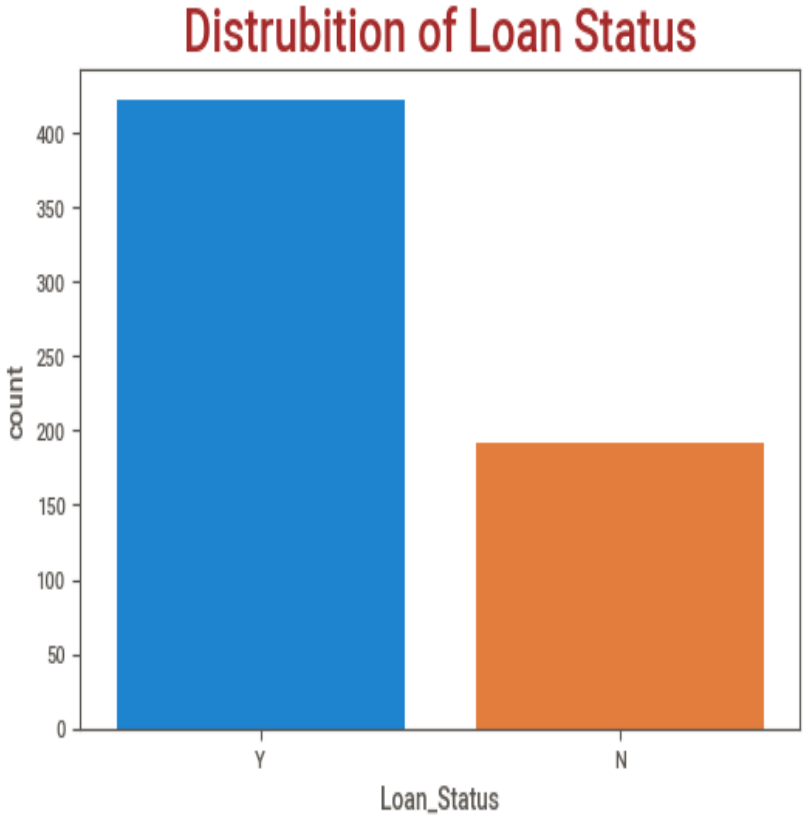


Data (Historical Data)

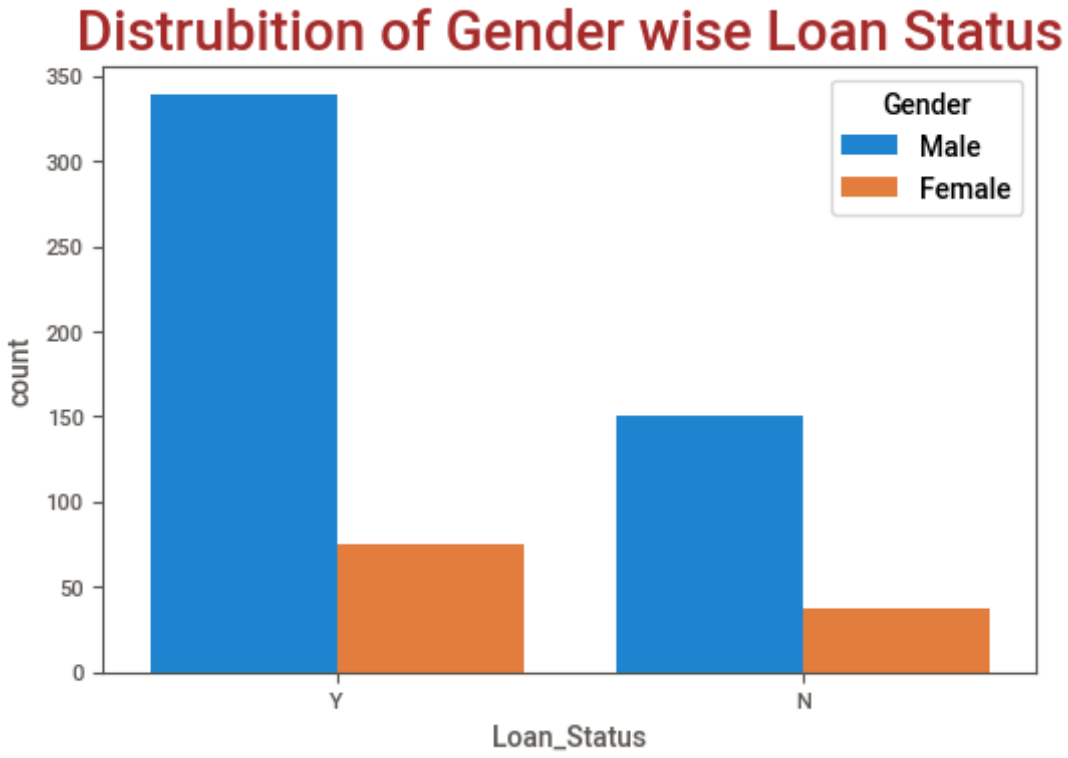
- Train Data contains 614 Rows and total 13 columns
- Out of 13 columns there are 5 numerical columns

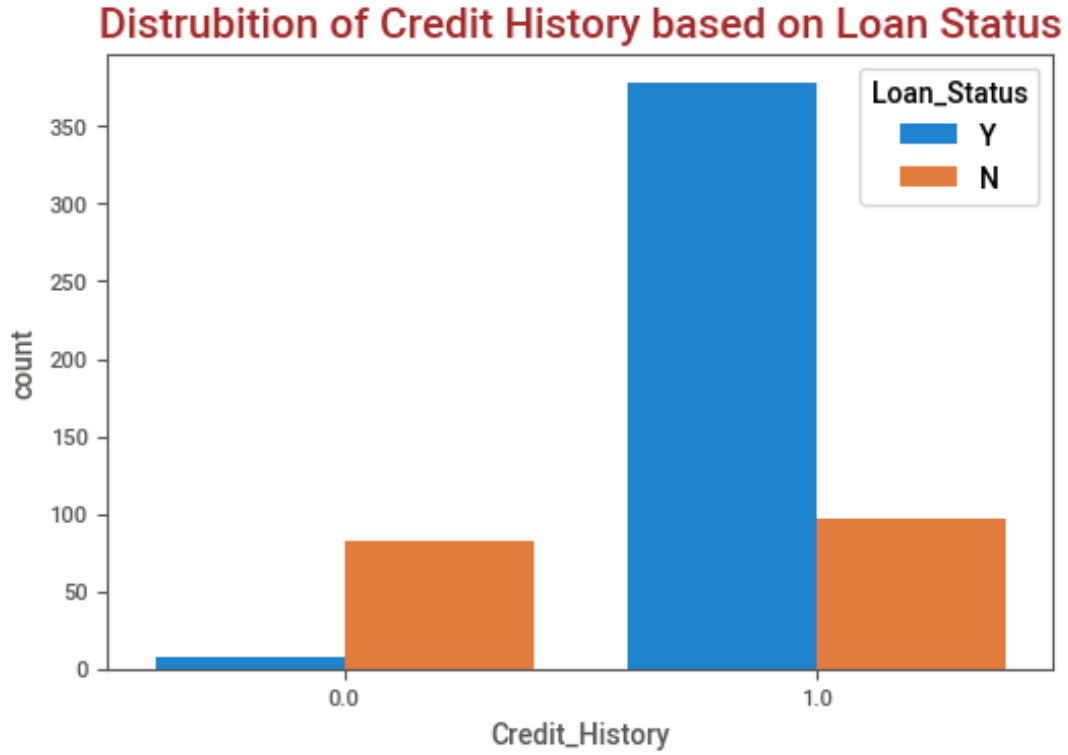
(4 float, 1 integer) and 8 object columns

- Target variable – Loan Status Y(422) and N (192)



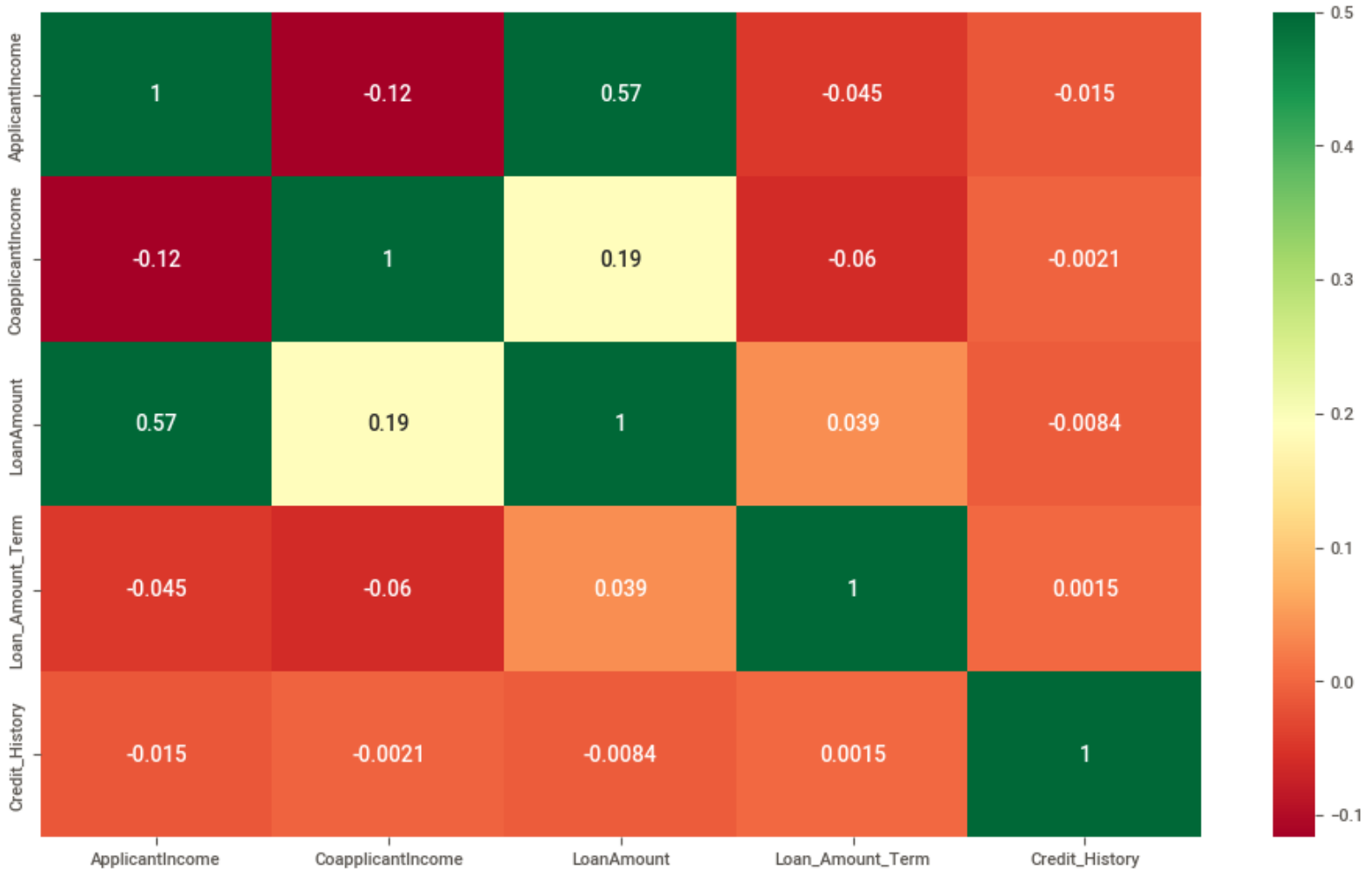
Analysis





Analysis

There is a correlation between the applicant's income and the loan amount they applied for.



Modeling

- The RandomForestClassifier machine learning model is one that has been trained.
- Along with traditional machine learning models, AutoML is also applied.
- Custom-made machine learning model required pre-processing
- AutoML did not required pre-processing
- Results from AutoML and conventional machine learning models are equivalent.

Model Evaluation

	AutoML	Custom-made ML
Accuracy	78%	77%

- **Where accuracy is the total of all the model's correctly predicted outcomes over all predicted outcomes.**

Recommendations

- Custom-made ML is better than AutoML
- We are fully aware of what was used, how it was used, and what algorithm was applied to accomplish the goal.
- If we can train and forecast in real time, it will benefit us and take less time to do so, although this use case may not allow for it.
- The ideal application for AutoML is as a foundational model.