Home Loans
Data Science Project

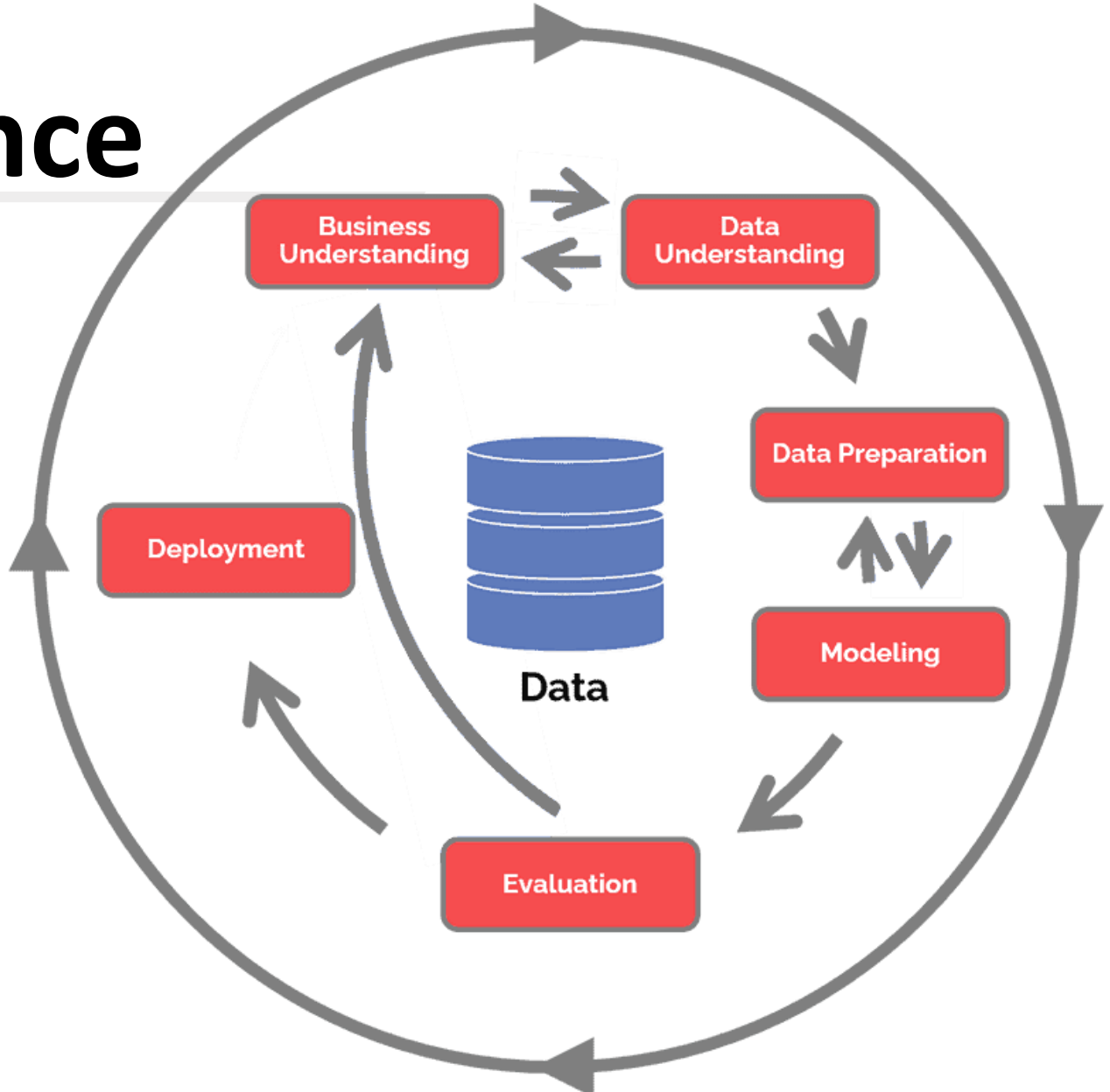
INTERNAL

Agenda

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

INTERNAL

Data Science Lifecycle



INTERNAL

Project Overview

Business Problem:

Currently the home loan application process is a manual one. It which takes 2-3 days, which mean that the applicant will only be notified after 2-3 days of the application outcome.

Business Objective:

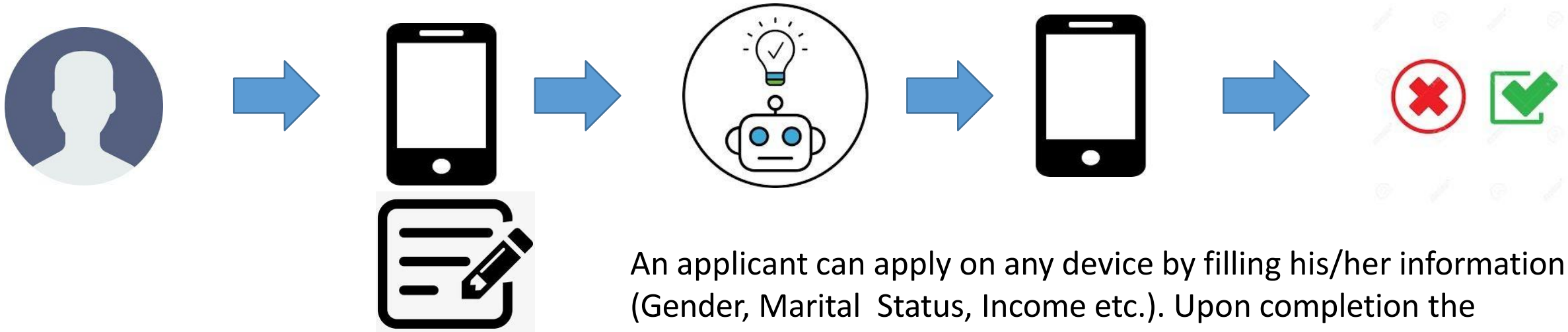
Reduce the amount of time it takes for applicants to be notified about their loan statuses (to a matter of seconds).

Hypothesis:

Based on historical data we can use machine learning to predict the loan status of a potential borrower such that the time taken for them to receive their respective statuses is reduced significantly.

INTERNAL

Process Overview Solution



An applicant can apply on any device by filling his/her information (Gender, Marital Status, Income etc.). Upon completion the machine learning model will be triggered to make a predict (based on historical data that it has been trained). The prediction will appear on the device as Accept or Decline on the same device in a matter of seconds

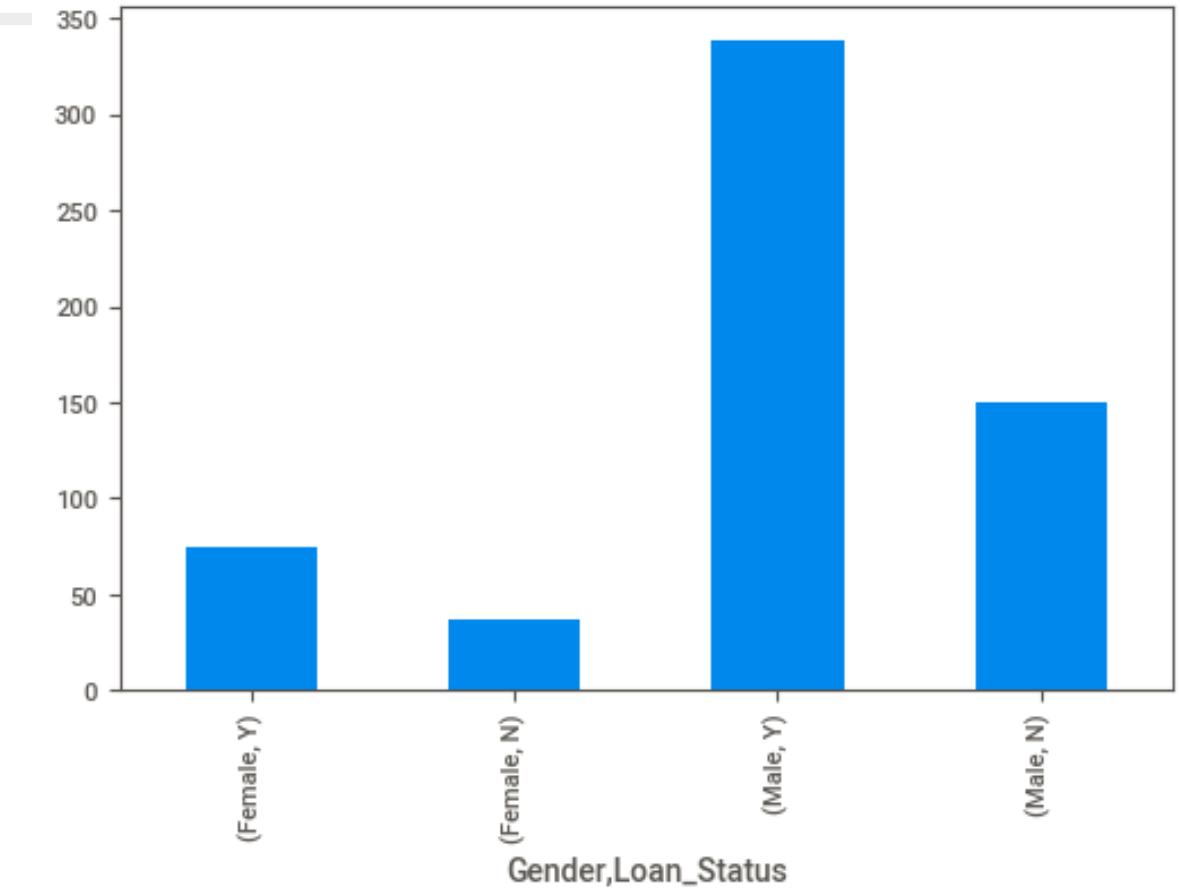
INTERNAL

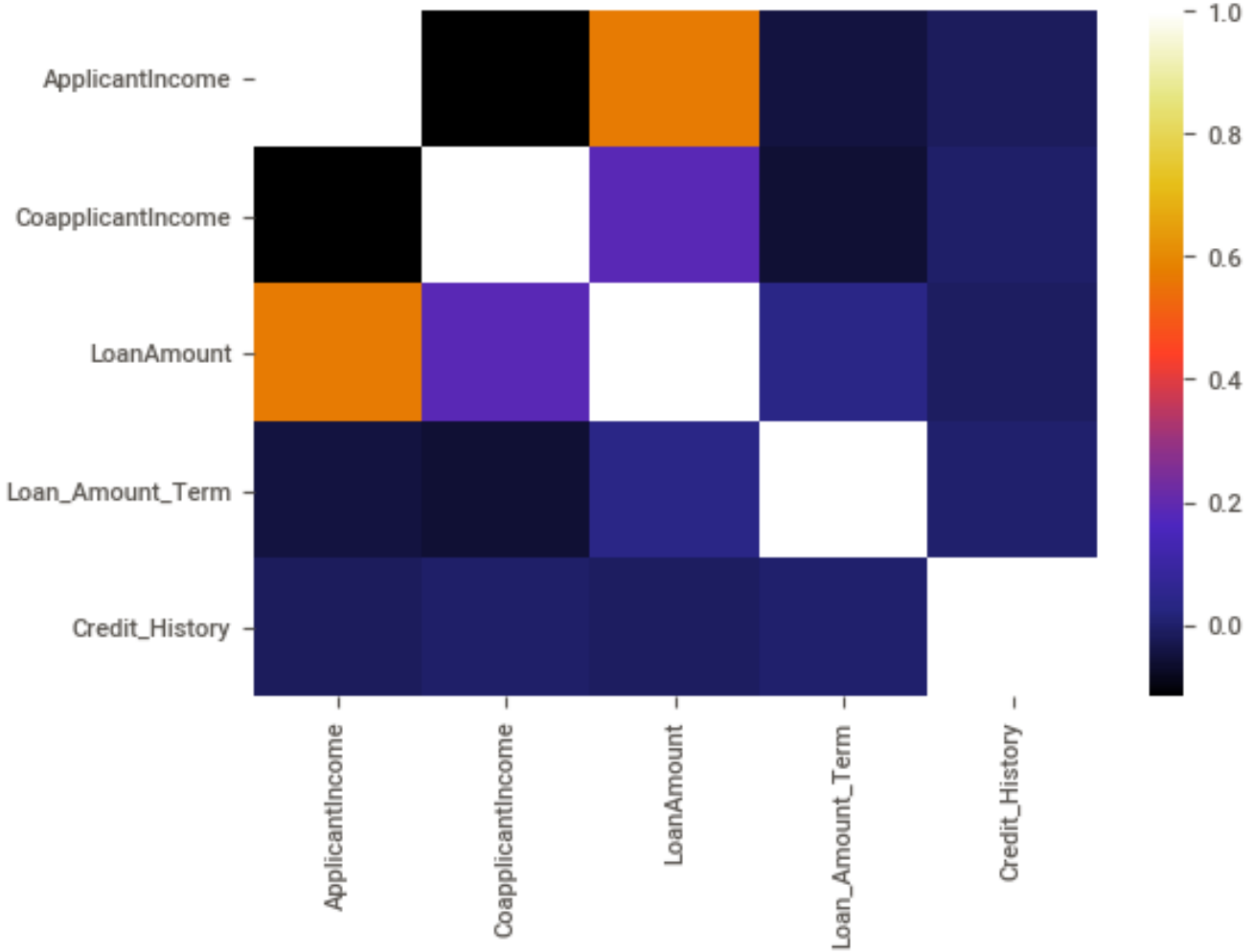
Data

- The number of records – 614
- The number of columns – 13
- The number of numerical columns – 5
- The number of categorical columns – 8
- Target/Loan Status – Y (422) vs N (192)

INTERNAL

Analysis





INTERNAL

Modeling

One machine learning model trained and AutoML used as well.

- Bespoke model required preprocessing
- AutoML did not
- Results fairly similar

INTERNAL

Model Evaluation

	AutoML	Bespoke ML
Accuracy	79,67%	78,05%

- where accuracy is the sum of all the correct predictions made by the model over all predictions made.

INTERNAL

Recommendations

- Bespoke ML > AutoML
- We understand/know exactly what went in, how it went in and what algorithm was used to achieve the objective
- Less time training (works in our favour if we train and predict in real time – maybe not applicable in this use case)
- AutoML is best used as a baseline model.

INTERNAL