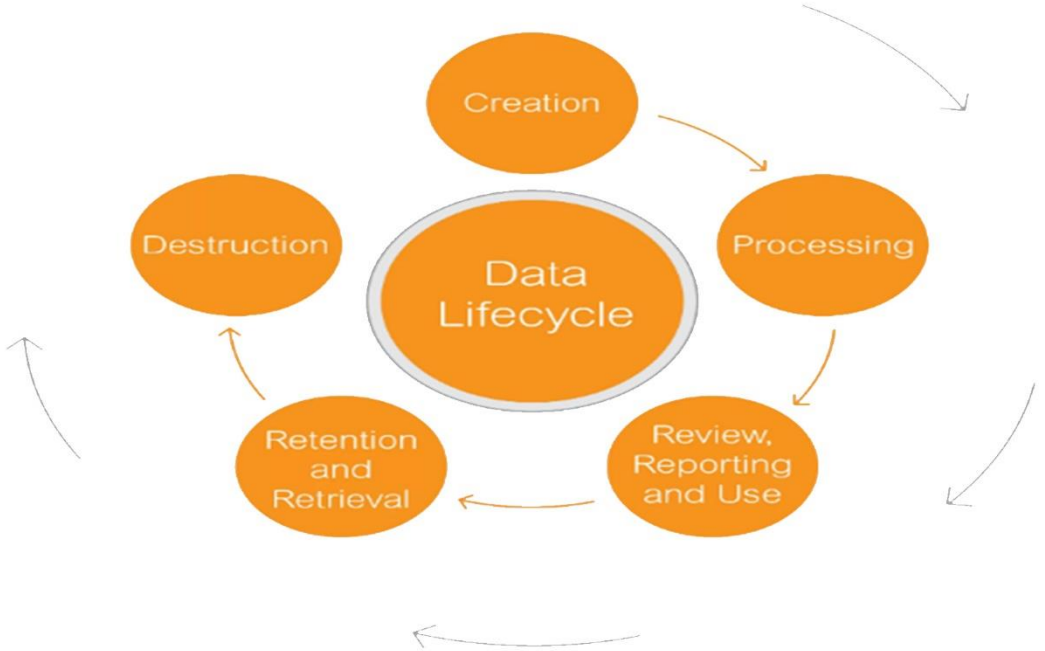
# Home Loans Data Science Project

# Agenda

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

# Data Science Lifecycle



# 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.

# Process Overview / Solution

Applicants fill out their information on any device (Gender, Marital Status, Income, etc.). The machine learning model will then make a prediction based on historical data. The prediction will appear on the device as Accept or Decline on the same device in a matter of seconds

# Data

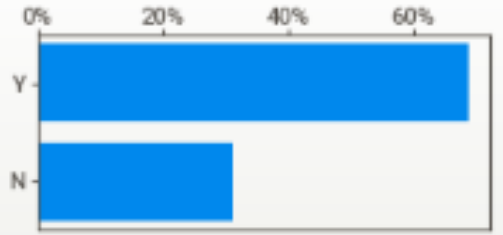
- The number of records – 614
- The number of columns – 13
- The number of numerical columns – 5
- The number of categorical columns – 8
- Target/Loan Status – Y (422) vs N (192)

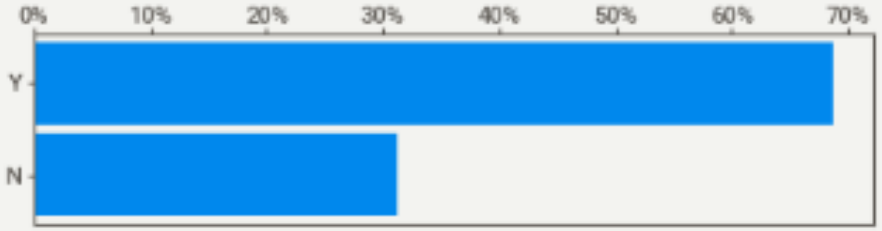
# Analysis

## Loan_Status

VALUES: 614 (100%)
MISSING: ---

DISTINCT: 2 (<1%)




### TOP CATEGORIES

| | | |
|---|---|---|
| Y | 422 | 69% |
| N | 192 | 31% |
| ALL | 614 | 100% |

CATEGORICAL ASSOCIATIONS
(UNCERTAINTY COEFFICIENT, 0 to 1)

### Loan_Status PROVIDES INFORMATION ON...

| | |
|---|---|
| Credit_History | 0.17 |
| Loan_Amount_Term | 0.02 |
| Property_Area | 0.01 |
| Married | 0.01 |
| Education | 0.01 |
| Dependents | 0.00 |
| Gender | 0.00 |
| Self_Employed | 0.00 |

### THESE FEATURES GIVE INFORMATION ON Loan_Status:

| | |
|---|---|
| Credit_History | 0.14 |
| Loan_Amount_Term | 0.02 |
| Property_Area | 0.02 |
| Married | 0.01 |
| Education | 0.01 |
| Dependents | 0.00 |
| Gender | 0.00 |
| Self_Employed | 0.00 |

# Modeling

A device getting to know version changed into skilled and AutoML was used as nicely.

- Bespoke models require preprocessing

- AutoML did now not

- outcomes are pretty similar

# Model Evaluation

| | AutoML | Bespoke ML |
|---|---|---|
| Accuracy | 79% | 77% |

- where accuracy is the total of all the right predictions produced by the model over all forecasts.

# Recommendations

- Bespoke ML trumps AutoML We are fully aware of what was used, how it was used, and what algorithm was employed to accomplish the goal.
- Less training time (works in our favour if we train and predict in real time – maybe not applicable in this use case)It is advisable to utilise AutoML as a foundational model.