



BrainStation Capstone Proposal Bank Loan Status

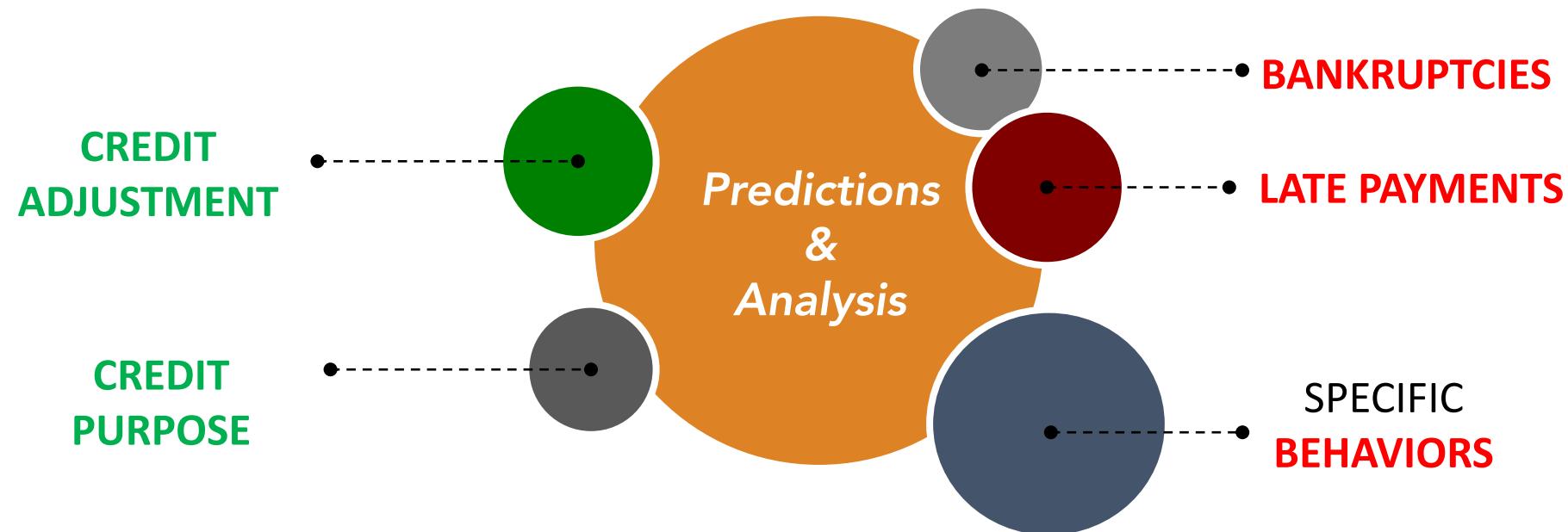
Mourad A. Abdel-Meseh



What?

What?

What I am Trying to Solve For



What?

The Tools

Data Acquisition

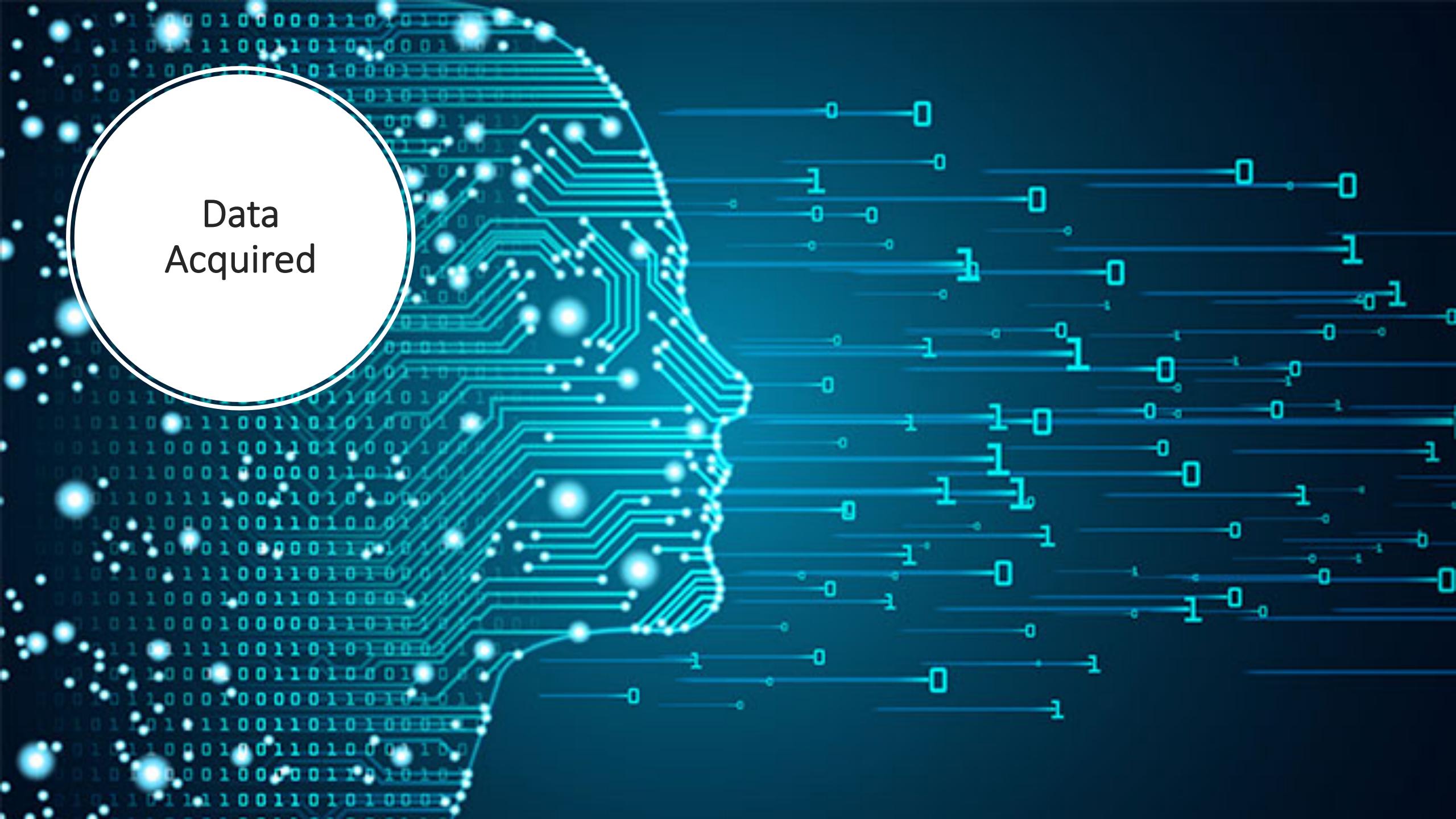


Preprocessing & Analysis



Modelling





Data
Acquired

Data Abstract



The Dataset we will be working with has been downloaded from Kaggle. As for the source of this set is currently unknown, I will try my best to get some sense and describe the features of this dataset.

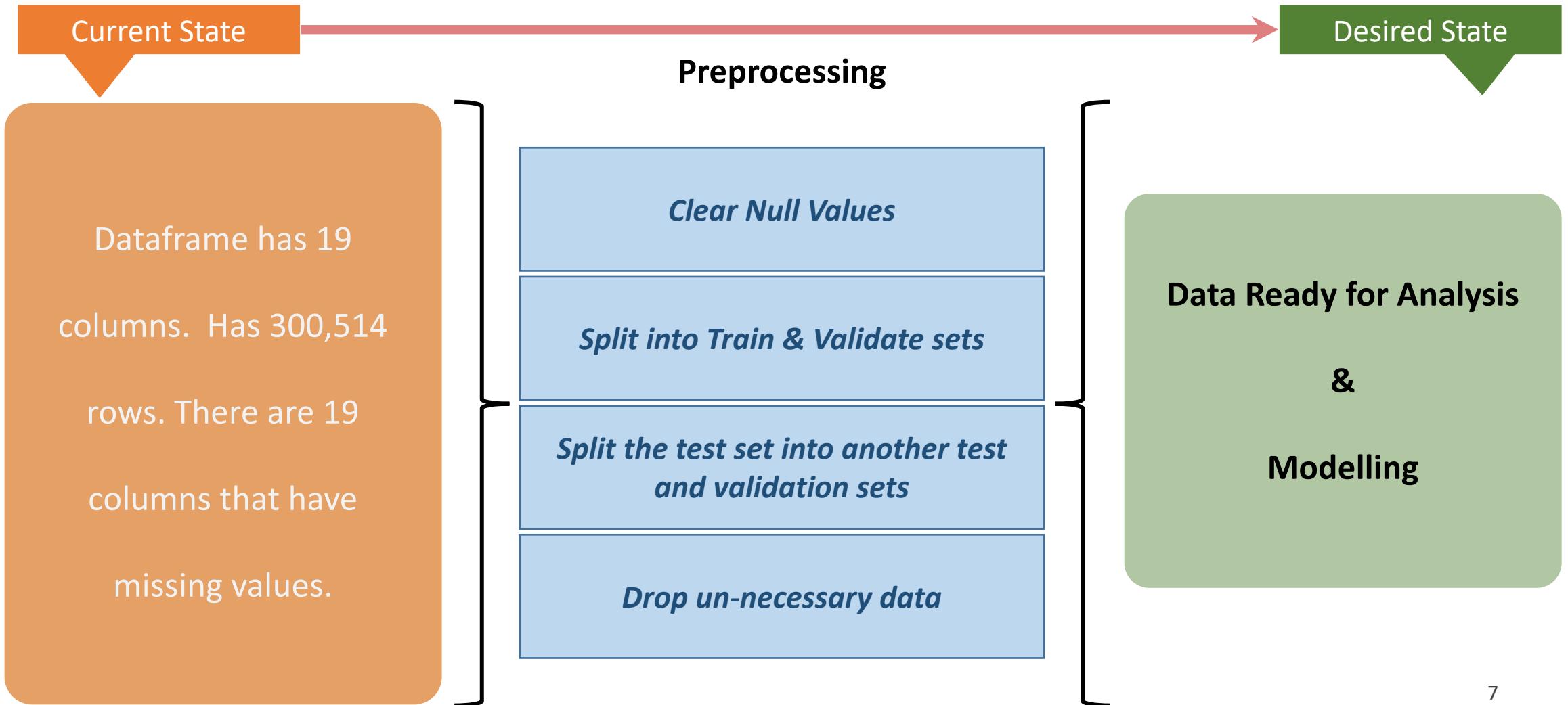


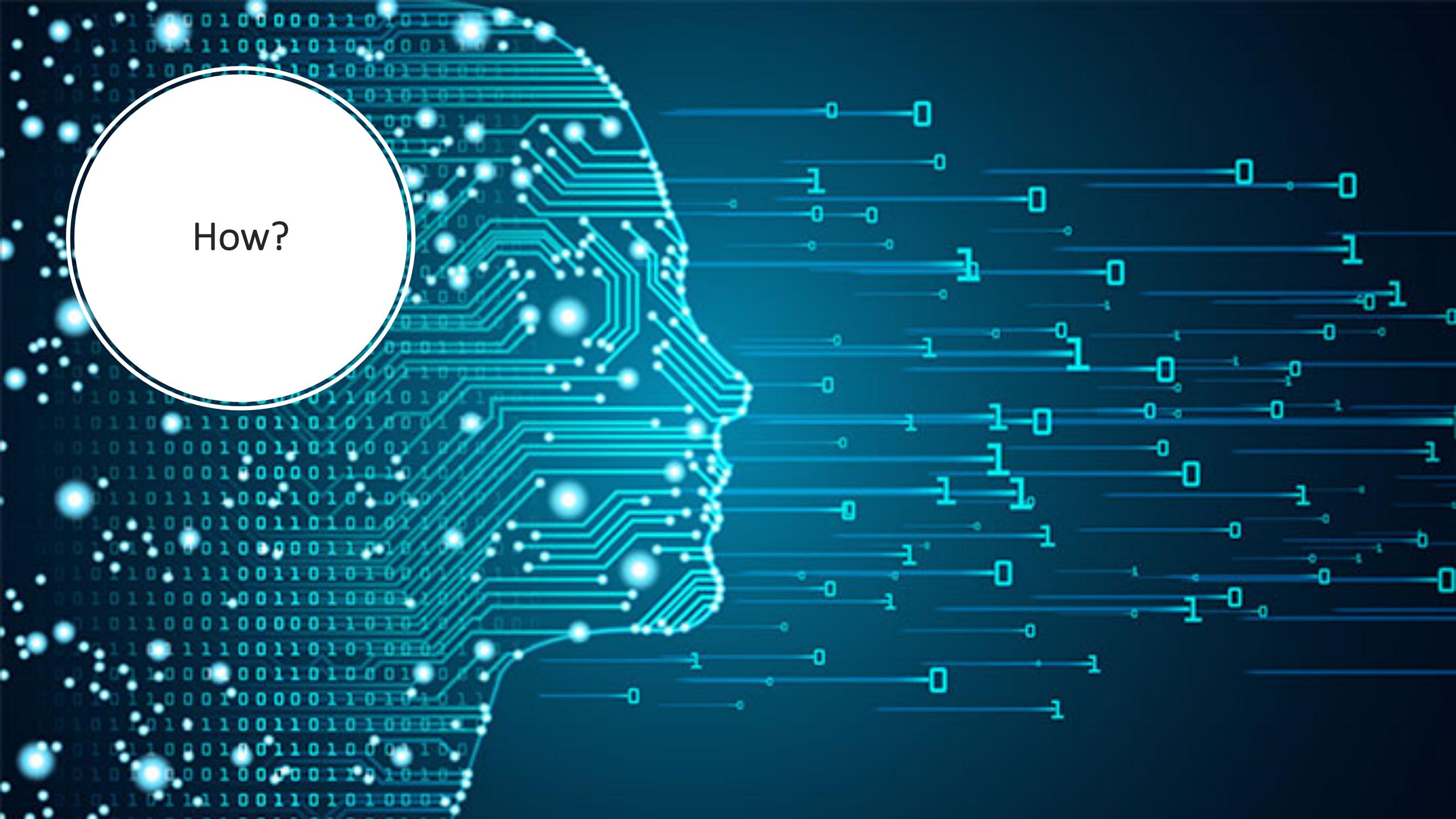
The dataset consists of 2 csv files already labelled into train & test sets.



From the initial look at the dataset we can deduce that it is for credit information per customer and consisting of 19 attributes of 300,514 customers.

Dataset Transformation

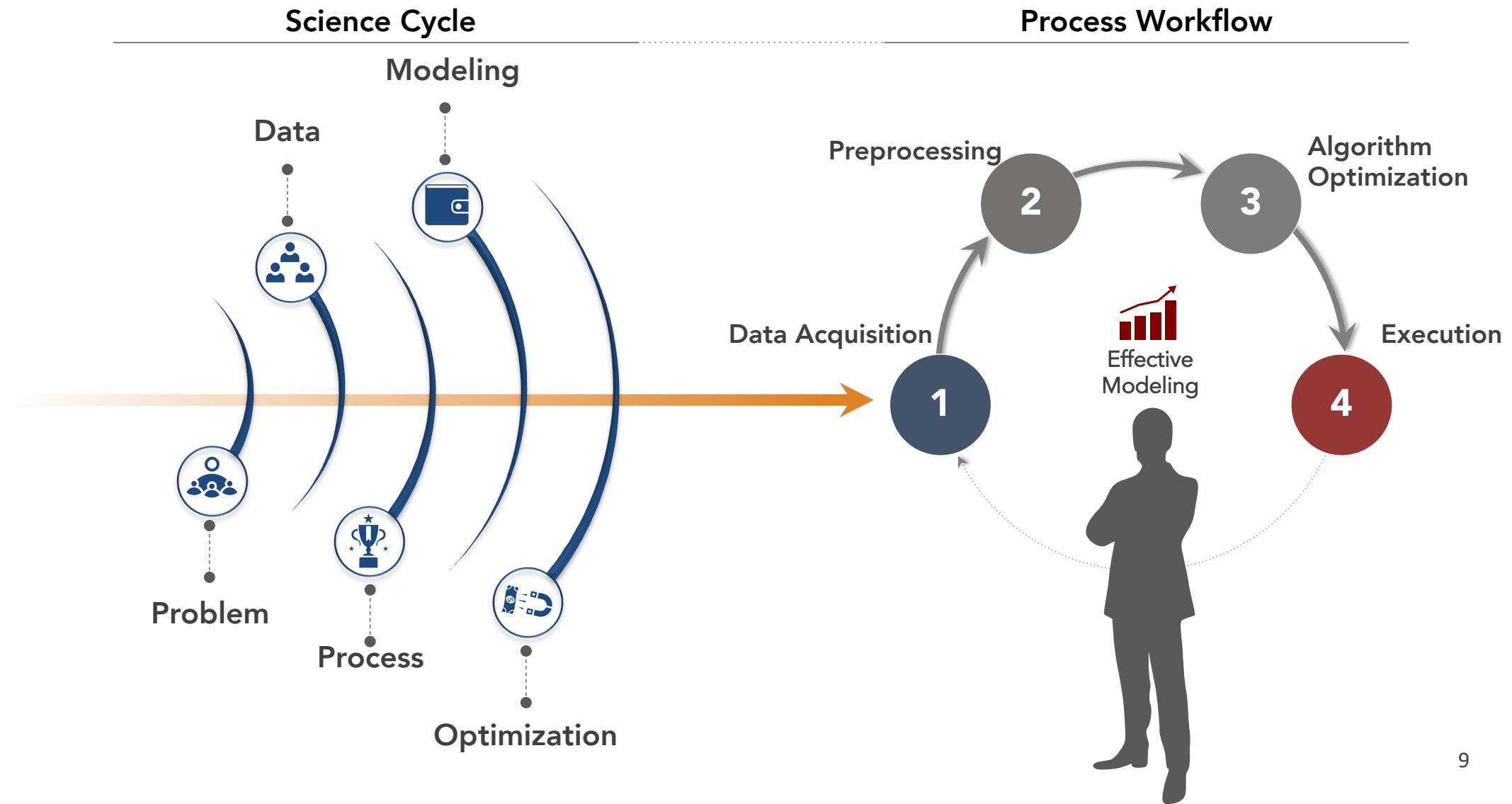




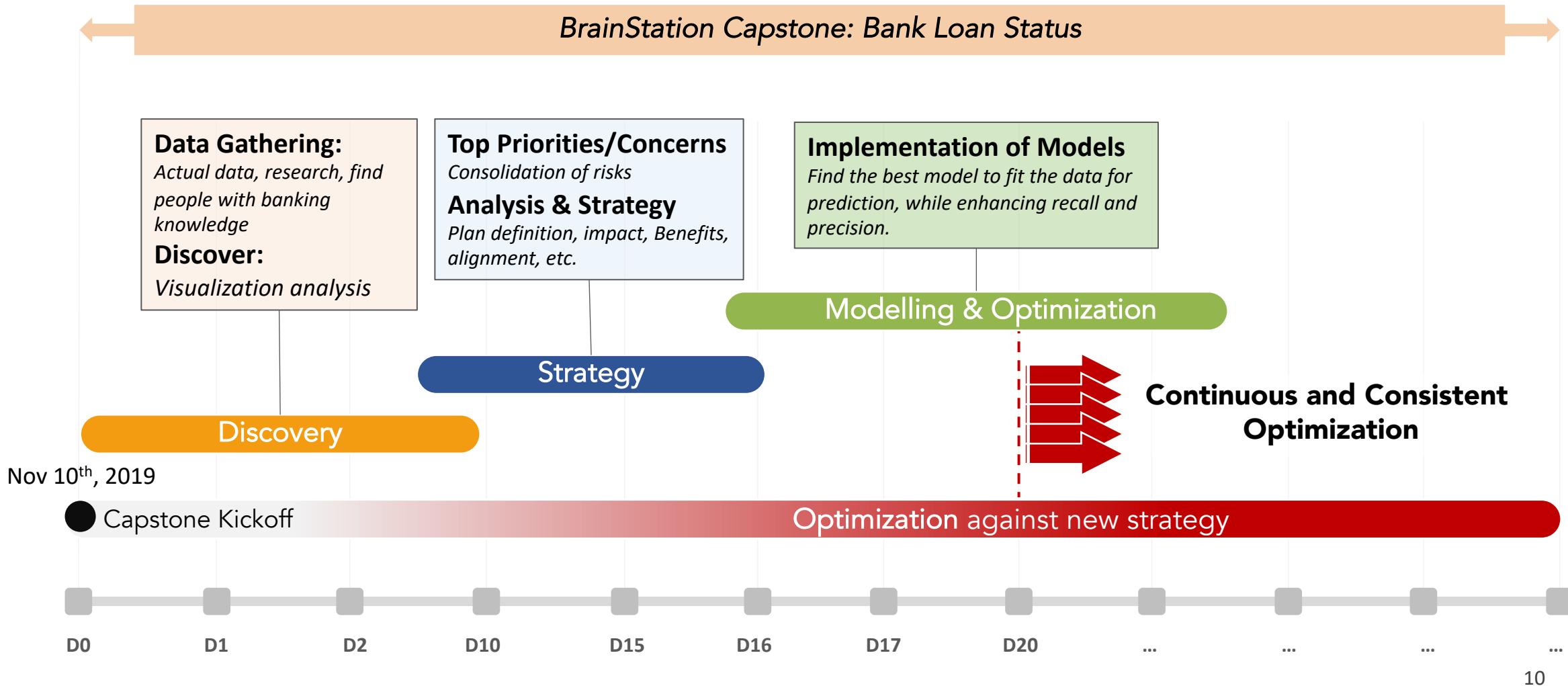
How?



Methodology Approach

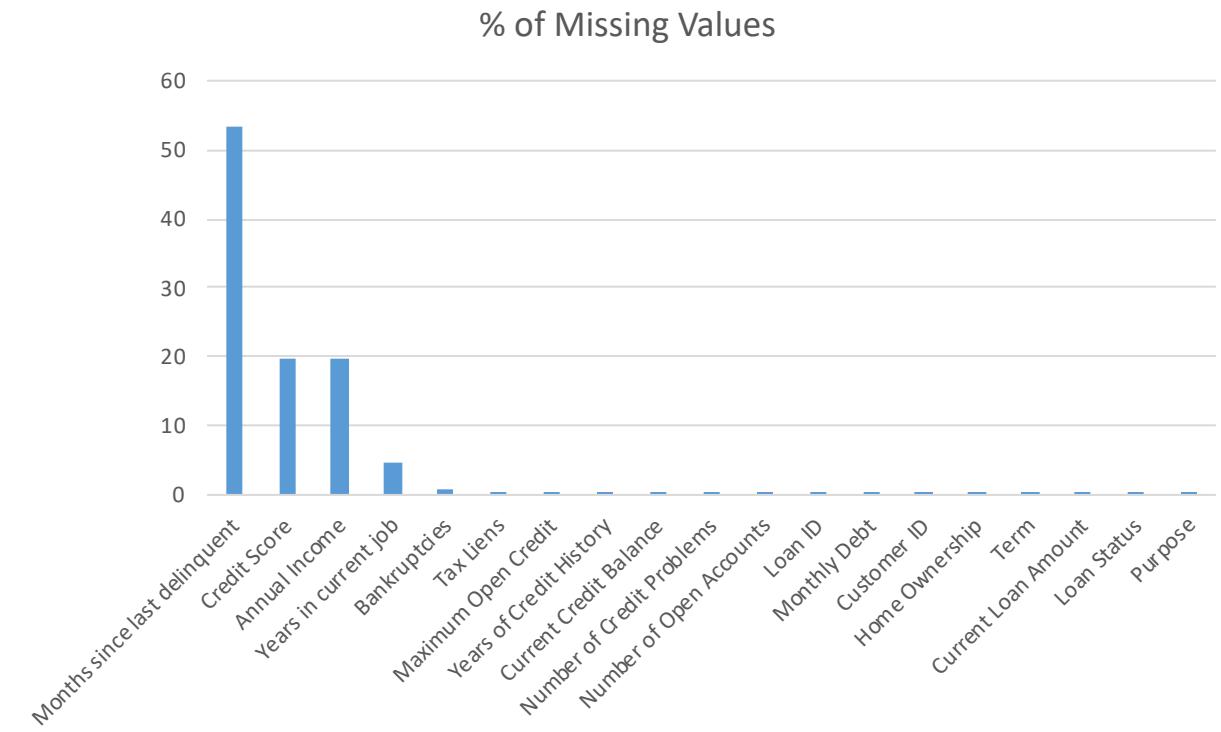


Capstone Timeline



Missing Values

Feature	Missing Values	% of Total Values
Months since last delinquent	53655	53.4
Credit Score	19668	19.6
Annual Income	19668	19.6
Years in current job	4736	4.7
Bankruptcies	718	0.7
Tax Liens	524	0.5
Maximum Open Credit	516	0.5
Years of Credit History	514	0.5
Current Credit Balance	514	0.5
Number of Credit Problems	514	0.5
Number of Open Accounts	514	0.5
Loan ID	514	0.5
Monthly Debt	514	0.5
Customer ID	514	0.5
Home Ownership	514	0.5
Term	514	0.5
Current Loan Amount	514	0.5
Loan Status	514	0.5
Purpose	514	0.5



The Risks

~54% Missing Value	Months Since Last Delinquent	In Both Training & Test Datasets
~19% Missing Value	Credit Score	In Both Training & Test Datasets
~19% Missing Value	Annual Income	In Both Training & Test Datasets

...

Process Summary

1 ► Drop rows with all missing values and impute for the rest.

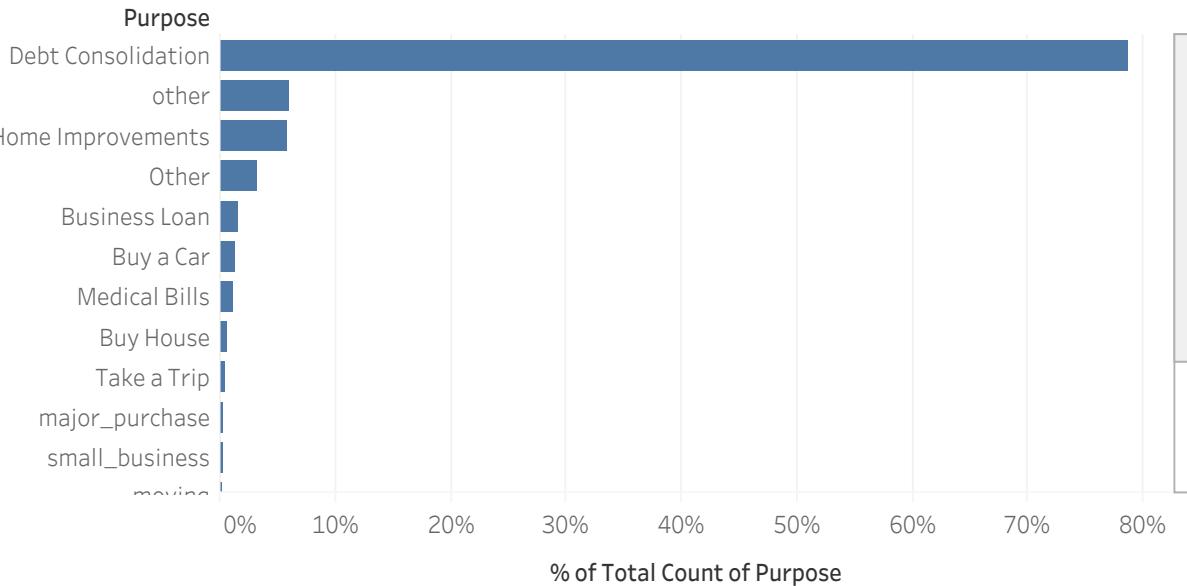
2 ► Concatenate all the data for analysis

3 ► Using Machine Learning find the best model to predict
Loan Status, Bankruptcy & Months since last
Delinquent

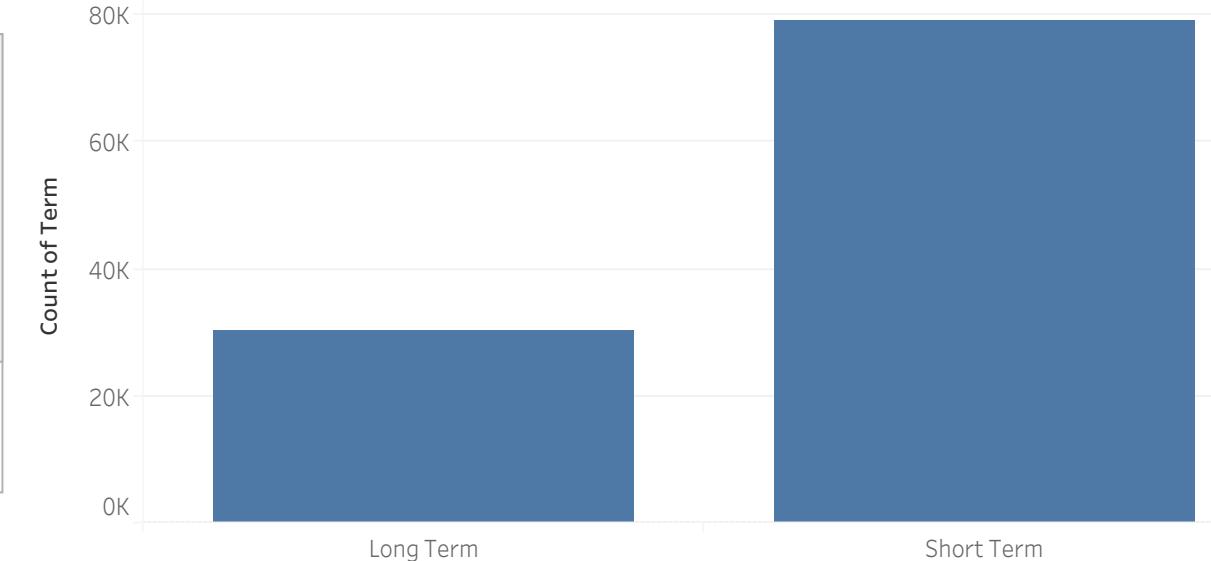
4 ► Compare all models

MODEL	ACCURACY	TEST SCORE	TRAIN SCORE	F1 SCORE	COMPUTATION TIME
KNN	95.94%	95.94%	96.43%	98%	18MIN 37S
XGBOOST	98.71%	98.71%	100%	99%	2MIN 50S
RANDOM FOREST	98.51%	98.51	99%	99%	1.99S

Purpose of Loan

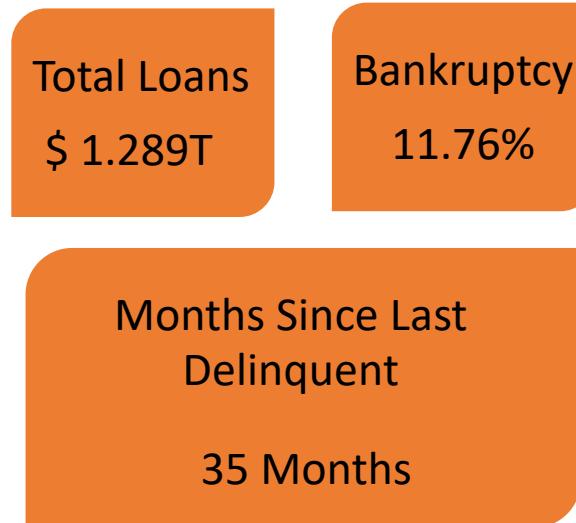


Loan Term

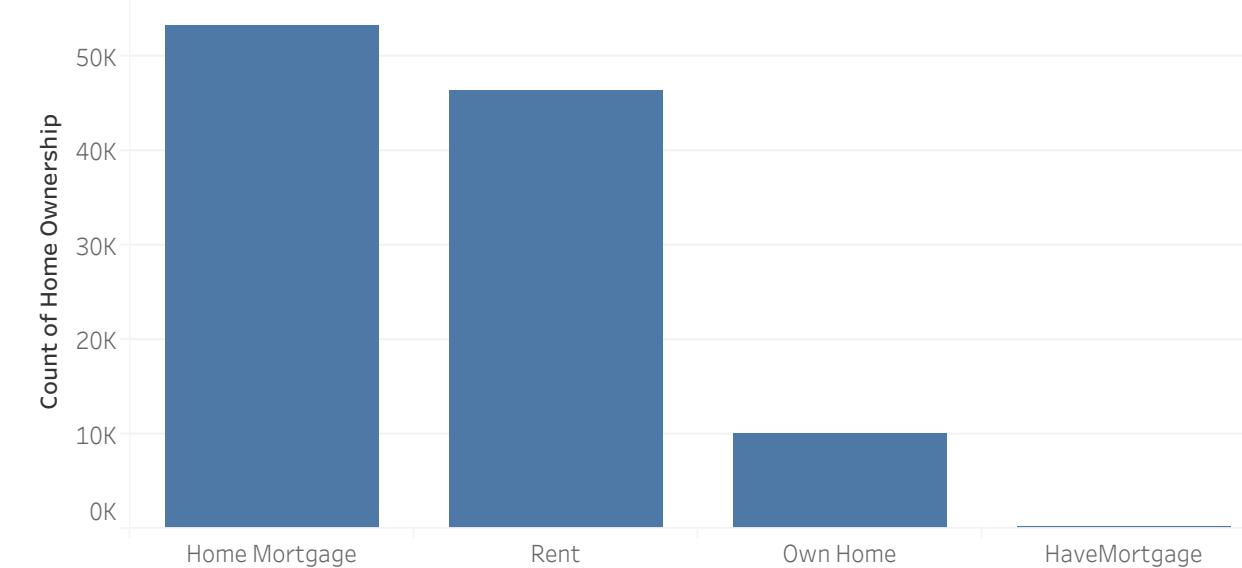


Analysis

Current Loan Amount	1,288,940,702,610
Maximum Open Credit	82,452,739,116
Current Credit Balance	32,308,700,879
Avg. Monthly Debt	18,483
Avg. Annual Income	1,377,565
Number of Open Accounts	1,221,418
Number of Records	109,772
Number of Credit Problems	18,478
Bankruptcies	12,911
Tax Liens	3,239
Median Credit Score	722
Median Months since last delinq..	34
Avg. Years of Credit History	18



Home Ownership





Q & A

Thank You