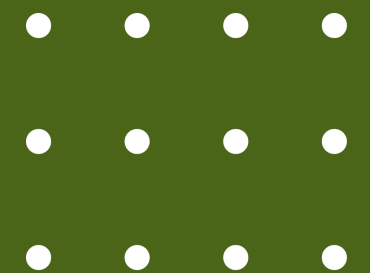
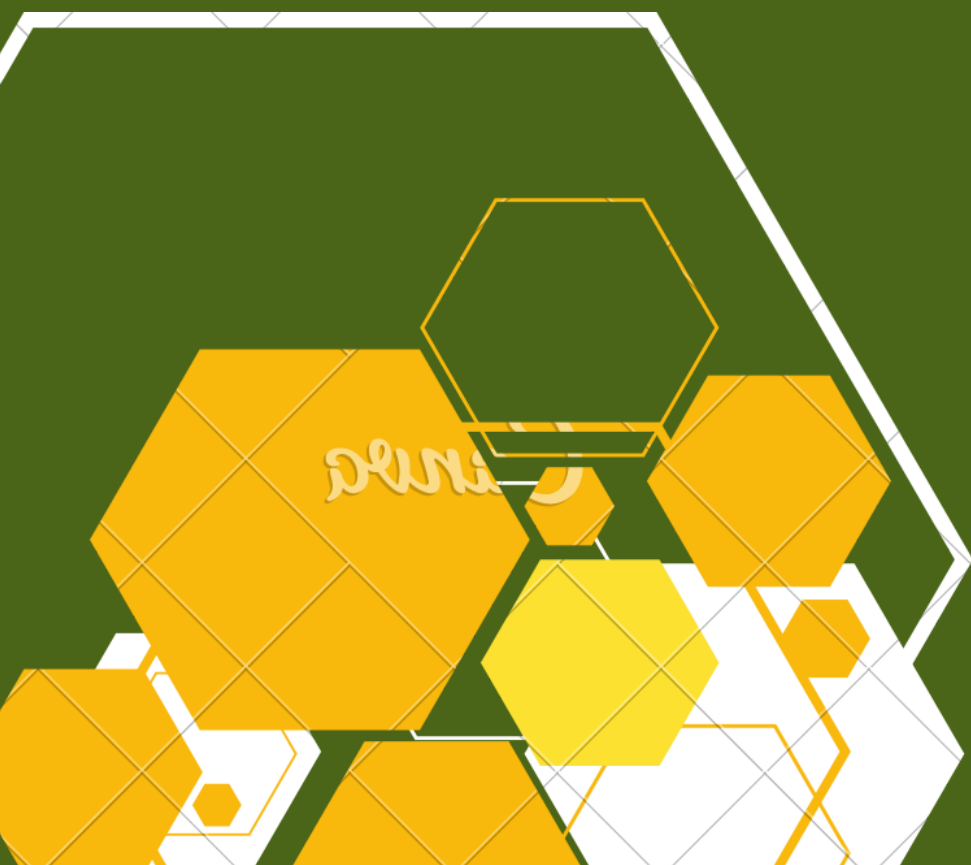


Fraud Detection **Medicare Claims**

Exploring ALTERYX



Anisha Joshi, IDS 506, 29th March 2022

Detailed Overview

Fraud Detection using Medicare Claims dataset

- CMS dataset(M): Medicare Provider Utilization & Payment Data
- LEIE Dataset(L) - List of fraudulent providers
- Goal - With the two datasets - identify the patterns in fraudulent claims and develop a classifier model.

01

Ingestion

Load all the data
Filter only the valid NPI data from L
Take Union of all the M datasets
Inner Join with NPI: Left join on L and Right join on M
Define Fraud Label with Y/N
Keep only non-null unique values & save as csv

02

Curation

Use the ingested data
Only Unique NPIs & HCPCS code
Convert to respective datatypes
Define & add some custom metrics
Summarize the values with group by, count & average
Save the output data as csv file

03

ML Model

Performed t-test (Num var.)
Contingency Table (Cat var.)
Tried Logistic Regression & Boosted Model

Statistics Data Report

2 records displayed, 2 fields, 1,985 bytes

Report Profile

1 of 1 Fields ✓ Records 1 to 2

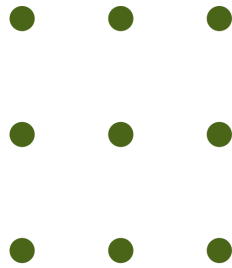
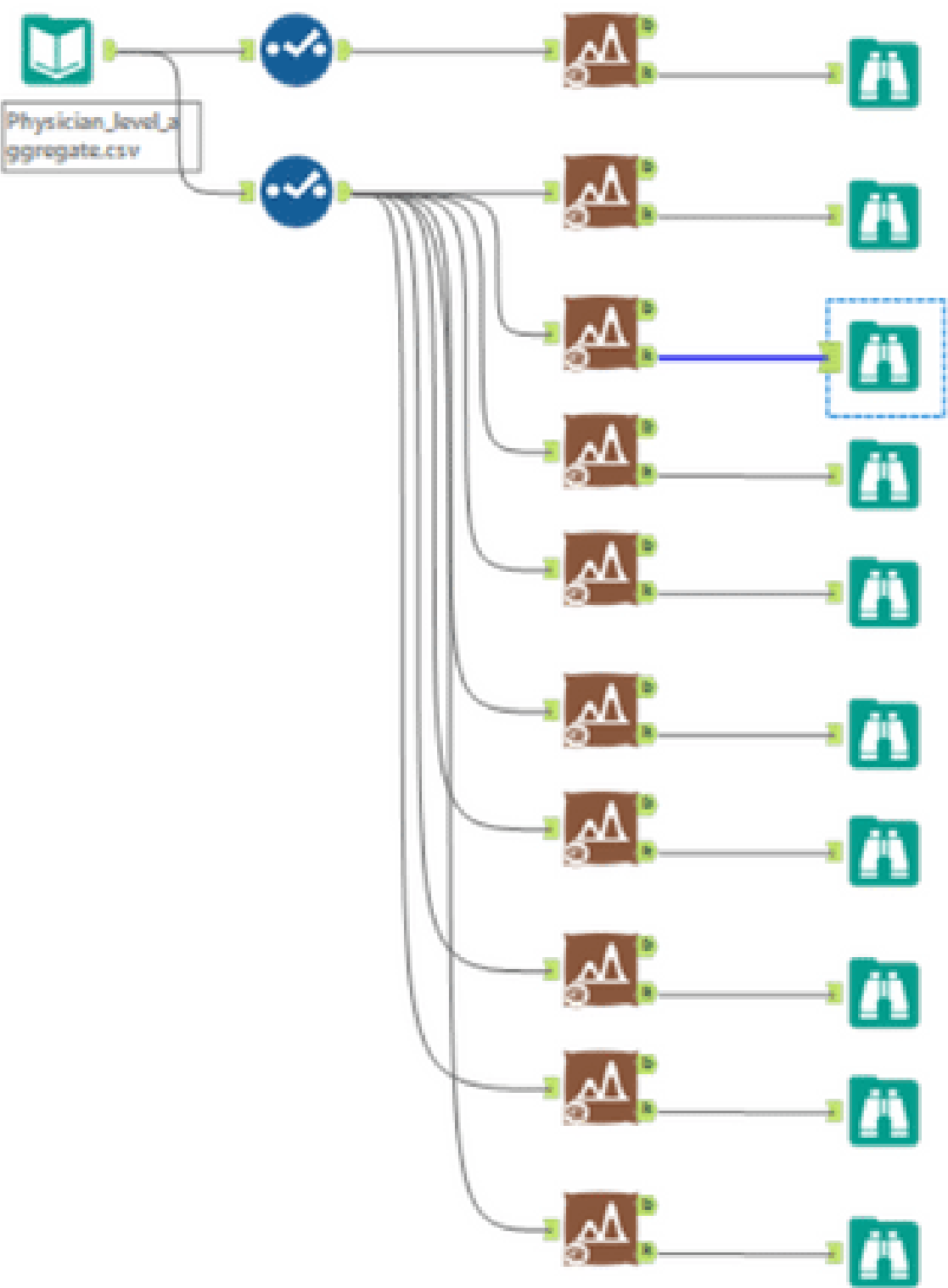
Record Report

1 **Welch's Two Sample t-test(s) of Avg_Final_amount_recieved by Fraud_Label**

2

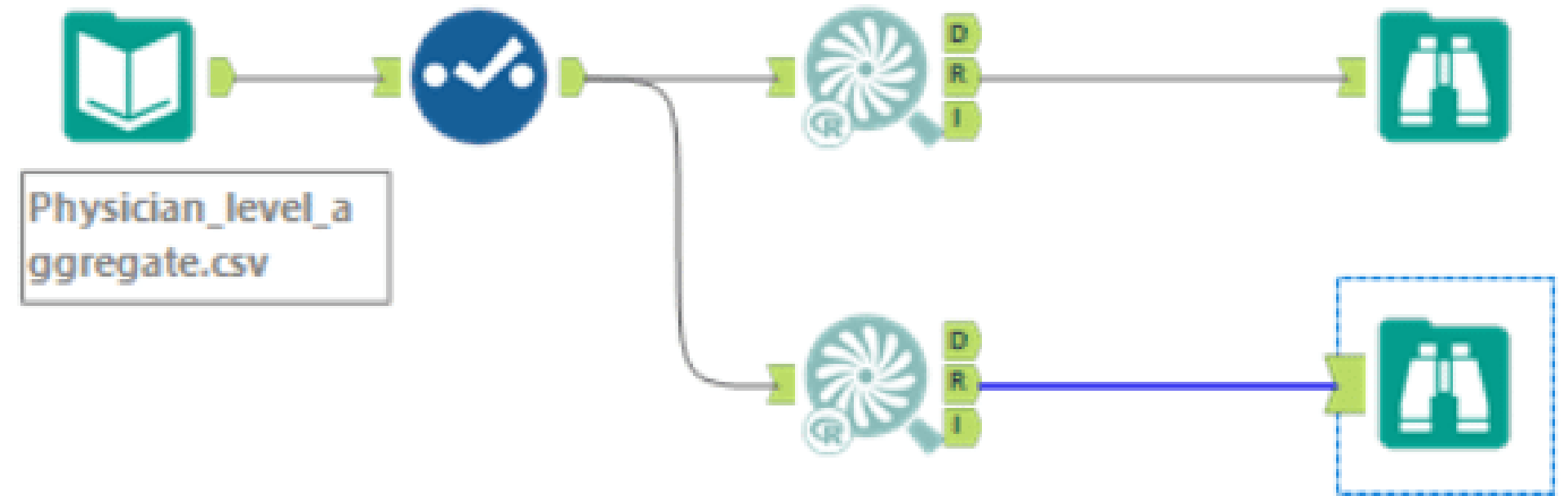
Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-0.432438	434.34	0.66564

Numerical Variables:
take only those where p-
values < 0.05



Select the Categorical Variables

Used contingency table
Refer to p values



Provider Type & Target Value(Fraud_Label)

Chi-squared = 184.5167, df = 62, p-value < 0.0000

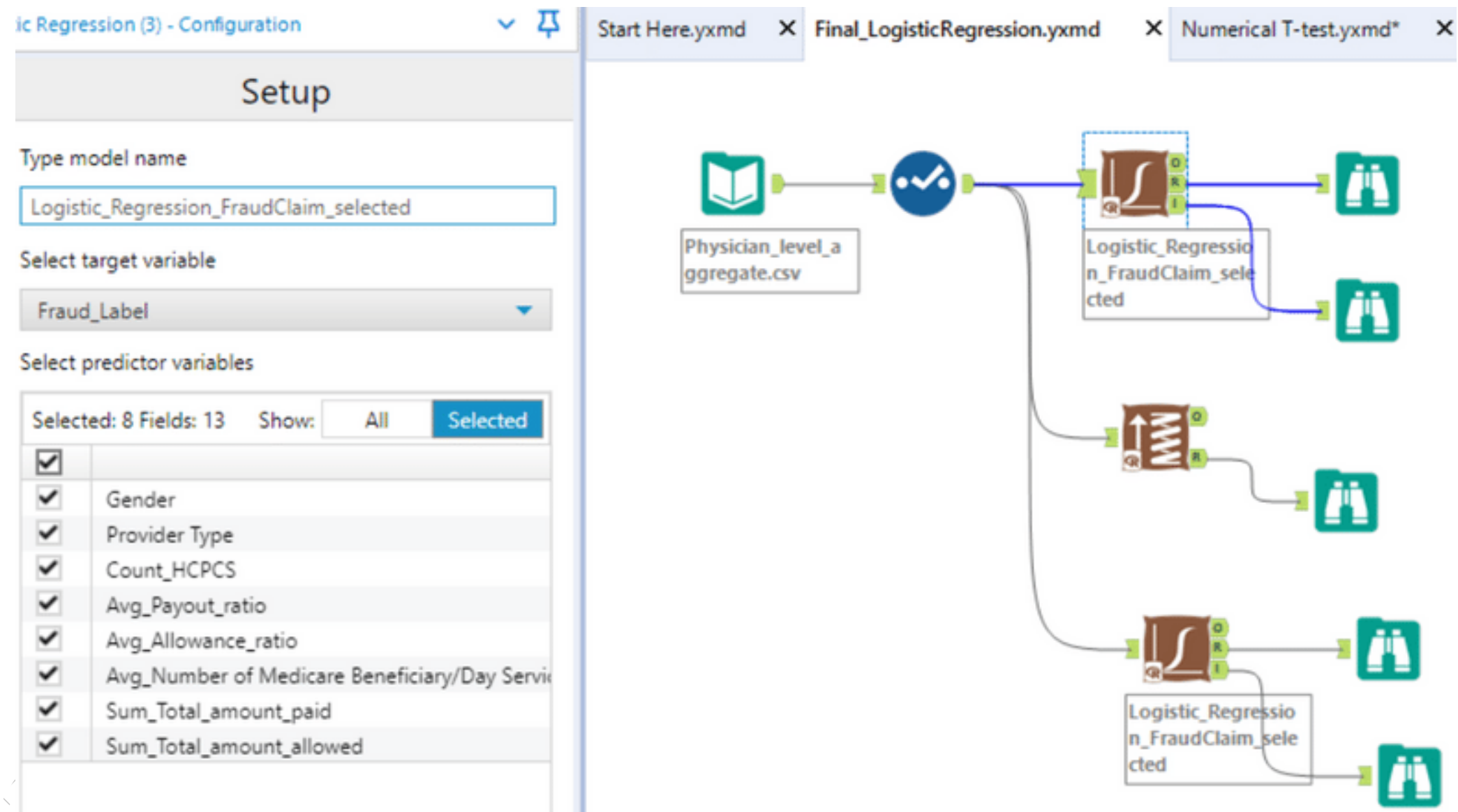
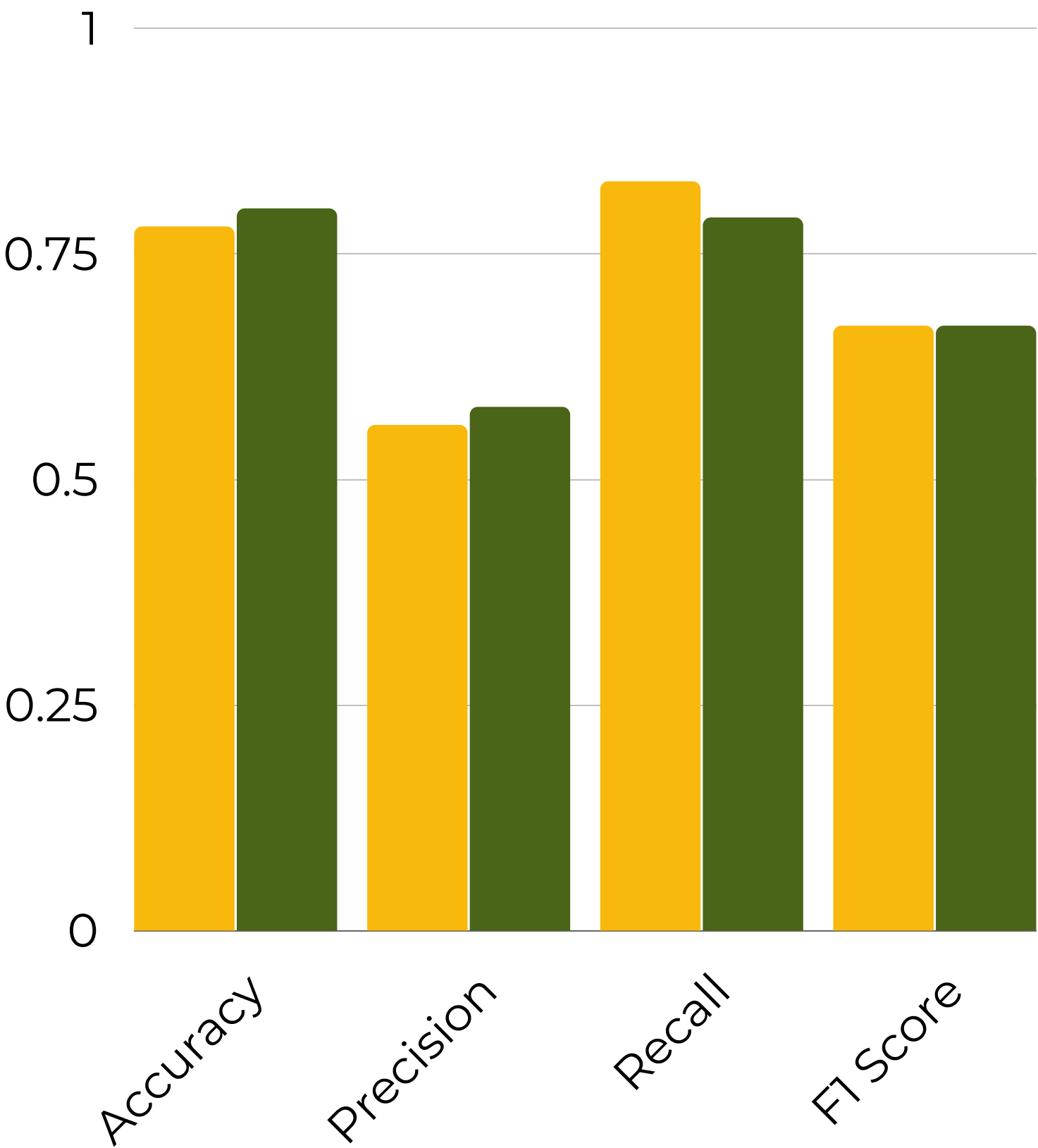
Gender & Target Value(Fraud_Label)

Chi-squared = 101.1603, df = 2, p-value < 0.0000

Compare the Classifier

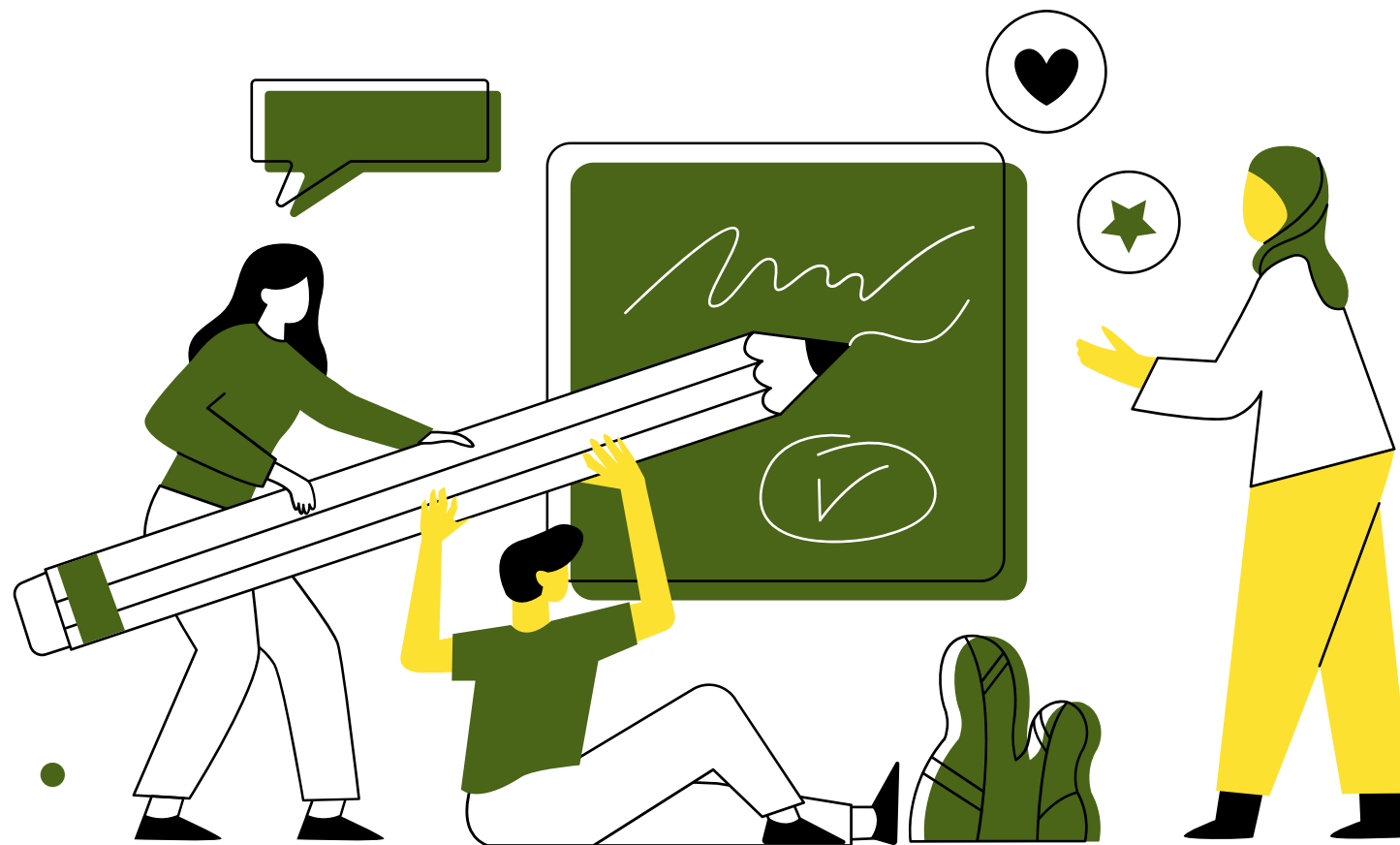
FP is more important - Precision

■ All Features ■ Selected Features



Feature Importance

From Boosted Model
(Same Table & Features as previous)



Provider.Type

Gender

Sum_Total_amount_allowed

Avg_Number.of.Medicare.Beneficiary.Day.Services

Avg_Allowance_ratio

Sum_Total_amount_paid

Count_HCPCS



Variable Importance Plot provides information about the relative importance



Further Steps

Lasso/Ridge regression
Have the Model with only Imp
Features

