US HEALTH INSURANCE MARKETPLACE

Yağmur Bali

Istanbul Data Science Academy

Health care in US can be very expensive

Health insurance offers a way to reduce such costs to more reasonable, affordable amounts.

About Data

US Health Insurance Marketplace Dataset from Kaggle

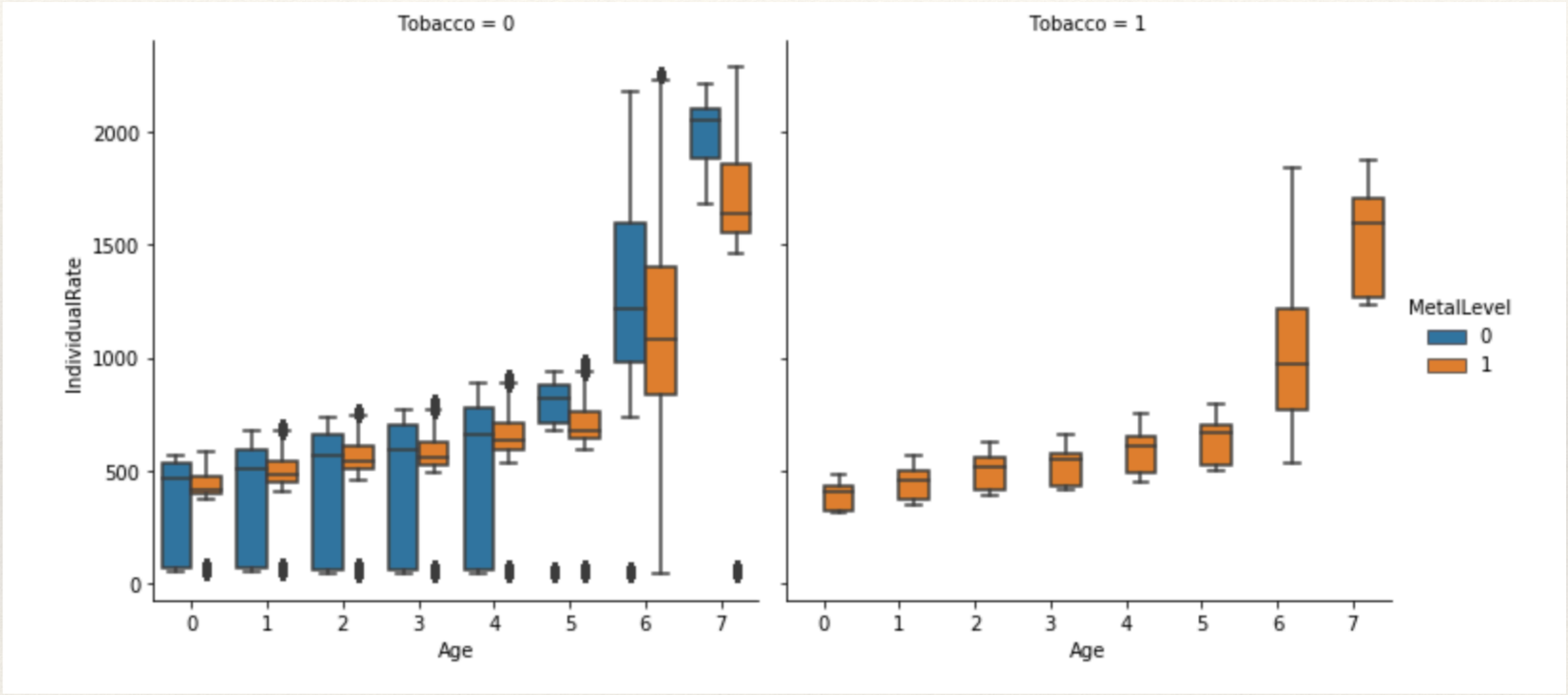
- Age
- Tobacco Usage
- Benefit Name
- Coverage
- In Network Payments
- Out of Network Payments
- Is EHB (Essential Health Benefits)
- Individual Rate
- Market Coverage: Small Group / Individual
- Metal Level: Dental / Medical

Methodology

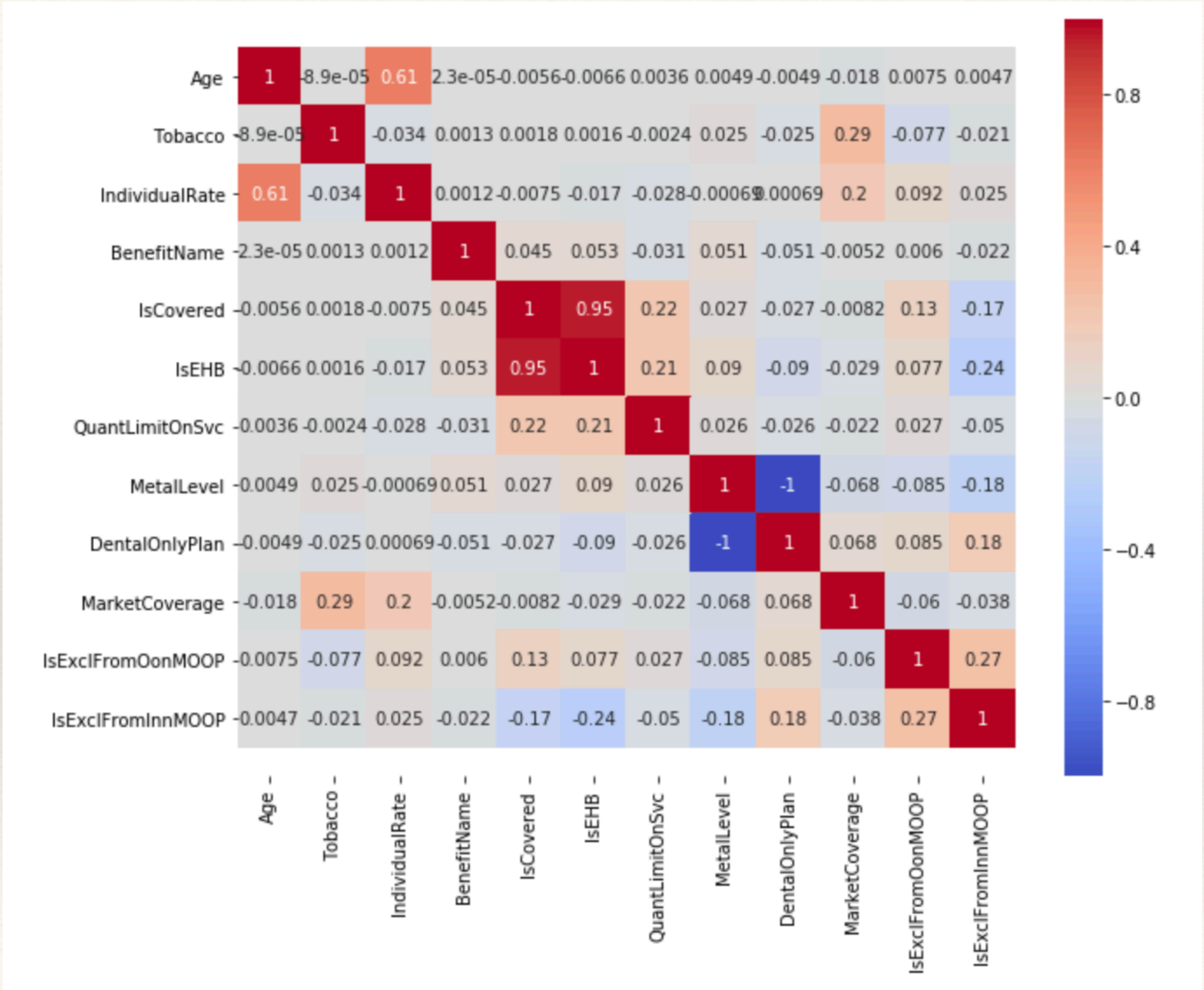
Which insurance plan is preferred?

The columns below can affect the metal level:
* Age
* Tobacco Usage
* Benefit Name
* Is EHB
* In Network Payments
* Out of Network Payments
* Coverage

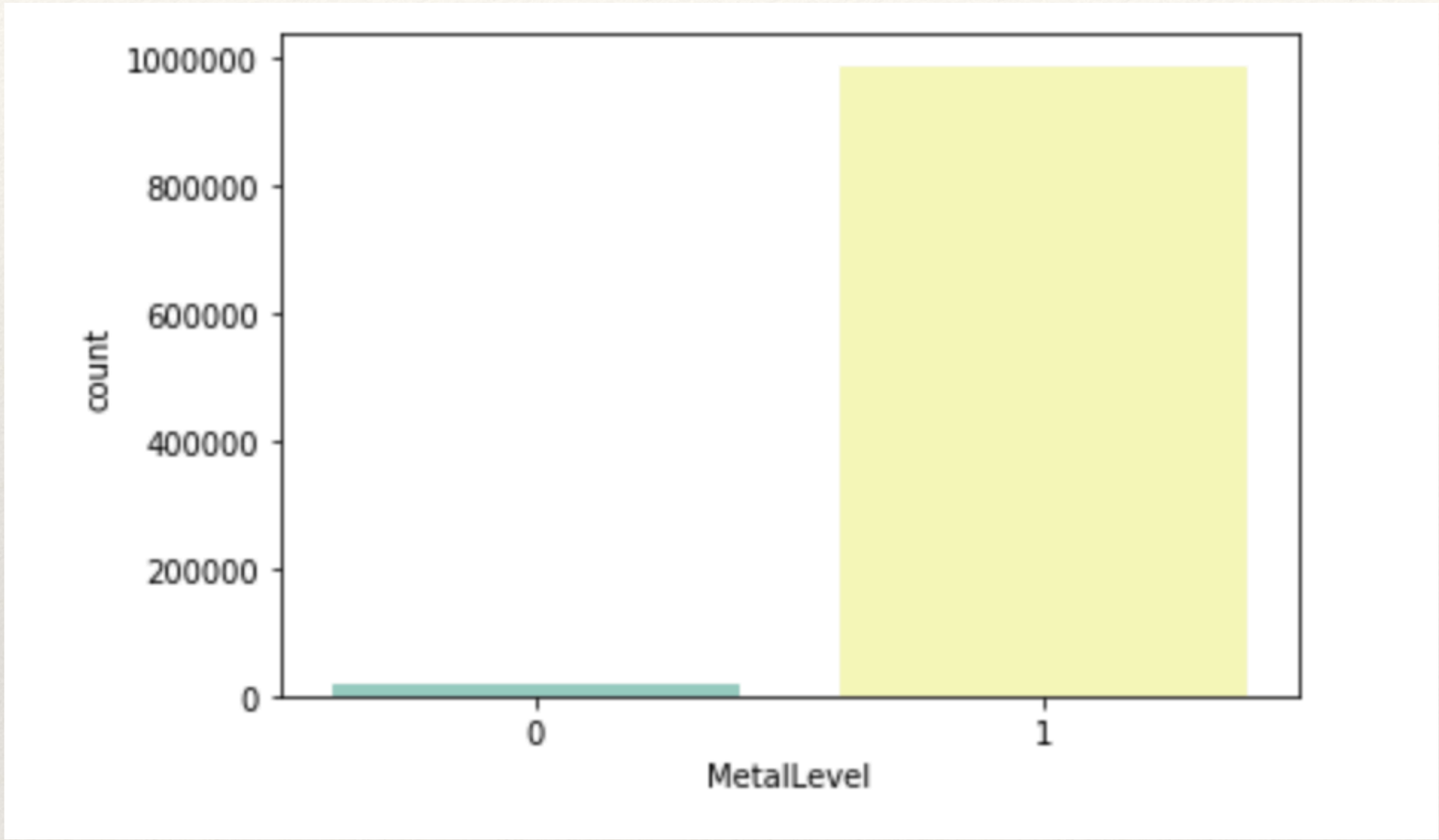
Tobacco Usage by Individual Rate and Age



CORRELATION HEATMAP



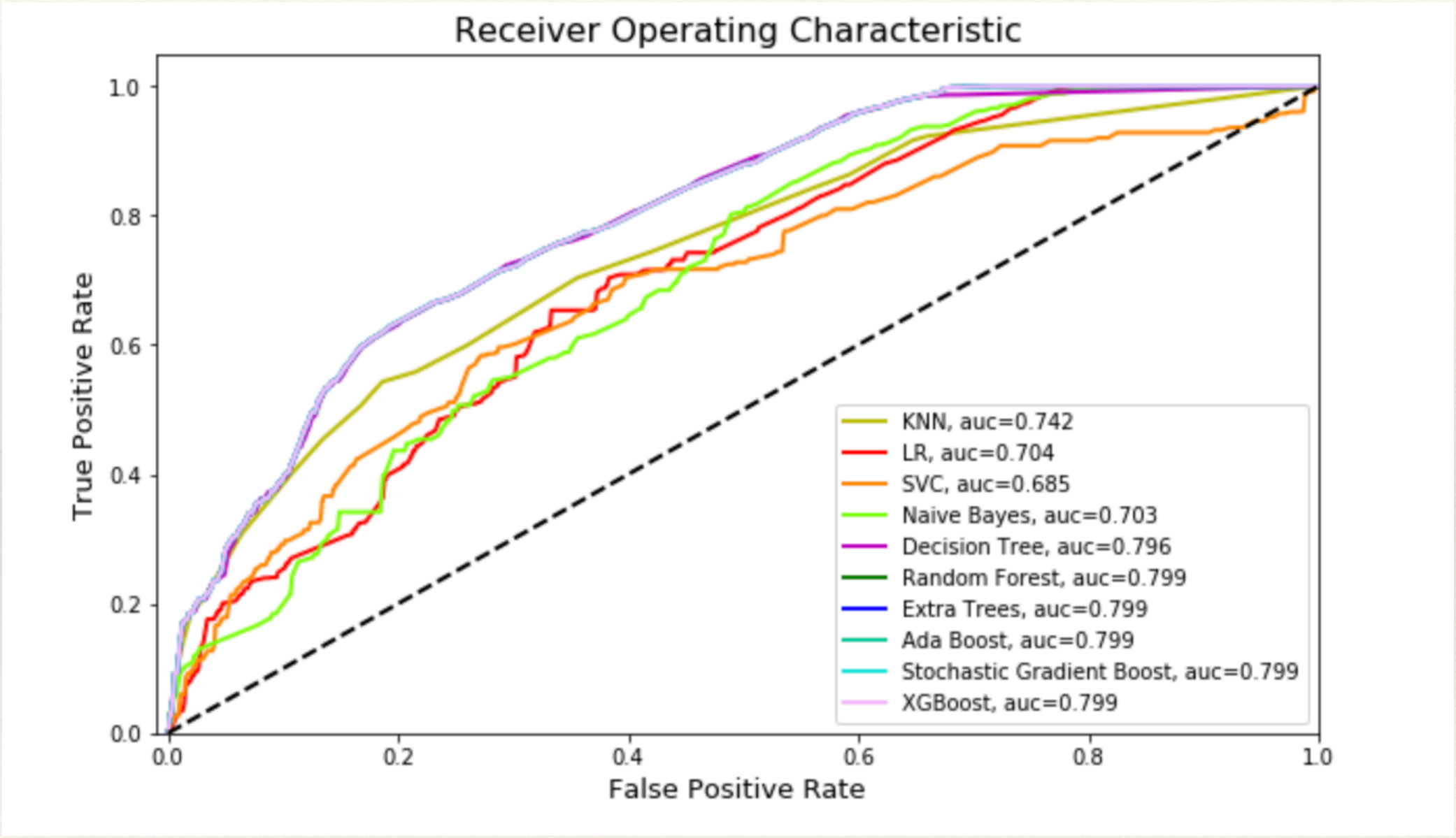
Imbalance Data



Dental: 2%
Medical: 98%

Undersampling- NearMiss Algorithm

Model Selection



Which Models are the Best?

	Model	Train Accuracy	Test Accuracy	Precision	Recall	F1
0	KNN	0.75	0.680000	0.69	0.68	0.67
1	Logistic Regression	0.69	0.650000	0.65	0.65	0.64
2	Naive Bayes	0.59	0.590000	0.77	0.59	0.51
3	SVC	0.66	0.630000	0.64	0.63	0.63
4	Decision Tree	0.77	0.774384	0.73	0.72	0.71
5	Random Forest	0.77	0.720000	0.73	0.72	0.71
6	Extra Trees	0.76	0.720000	0.73	0.72	0.71
7	Ada Boost	0.69	0.720000	0.73	0.72	0.71
8	Stochastic Gradient Boost	0.66	0.720000	0.73	0.72	0.71
9	XGBoost	0.75	0.720000	0.73	0.72	0.71
10	Random Forest GS	0.77	0.720000	0.73	0.72	0.71
11	Random Forest RS	0.77	0.710000	0.71	0.71	0.71
12	Decision Tree GS	0.77	0.710000	0.72	0.71	0.71

1- Decision Tree
2- Random Forest
3- Extra Trees Classifier

FUTURE WORK

Looking up other rows in network and plan attributes

Focus on just medical plans and separate them as Bronze, Silver, Gold and Platinum

Focus on just dental plans and separate them as High and Low

THANK YOU