Health Insurance Marketplace

Vishal Kumar

Springboard - Data Science Career Track

Capstone project

Introduction

- A service that helps people shop for and enroll in affordable health insurance.
- The federal government operates the Marketplace, available at HealthCare.gov, for most states. Some states run their own Marketplaces.
- The Health Insurance Marketplace (also known as the “Marketplace” or “exchange”) provides health plan shopping and enrollment services through websites, call centers, and in-person help.
- Small businesses can use the Small Business Health Options Program (SHOP) Marketplace to provide health insurance for their employees.
- ACA(Affordable Care Act) or Obamacare, health exchanges were fully certified and operational by January 1, 2014, under federal law.
- Since the inception of the exchange, in year 2014, ACA has been the hot topic of discussion across the U.S.

Potential Clients

- Companies selling health insurance in state based marketplace or federal run exchange or private plan outside the marketplace.
- Few examples: Freedom Life Insurance Company of America, CIGNA Health and Life Insurance Company, Celtic Insurance Company, Blue Cross and Blue Shield, Aetna Life Insurance Company.

Problem Definition

- Project is divided into 2 parts: EDA(Explanatory Data Analysis and Machine Leaning.
- **EDA**:
  - Data exploration to answer some analytical questions related to dataset.
  - Like, how do plan rates and benefits vary across states? how do plan benefits relate to plan rates? how do plan rates vary by age? how do plans vary across insurance network providers?
  - Any other related analysis on rate and benefits. We found some trends during the data exploration

- **Machine Learning:**
  - Build different models and using regression to predict individual and individual tobacco monthly premium.
  - Comparison and improvement in the models.
- **Dataset:**
  - The dataset has been downloaded from the kaggle and from the National Bureau of Economic Research.
  - CMS' Center for Consumer Information & Insurance Oversight produces the Health Insurance
  - Marketplace Public Use Files (Marketplace PUF) to increase transparency in the Health Insurance Marketplace and to support benefit and rate analysis.

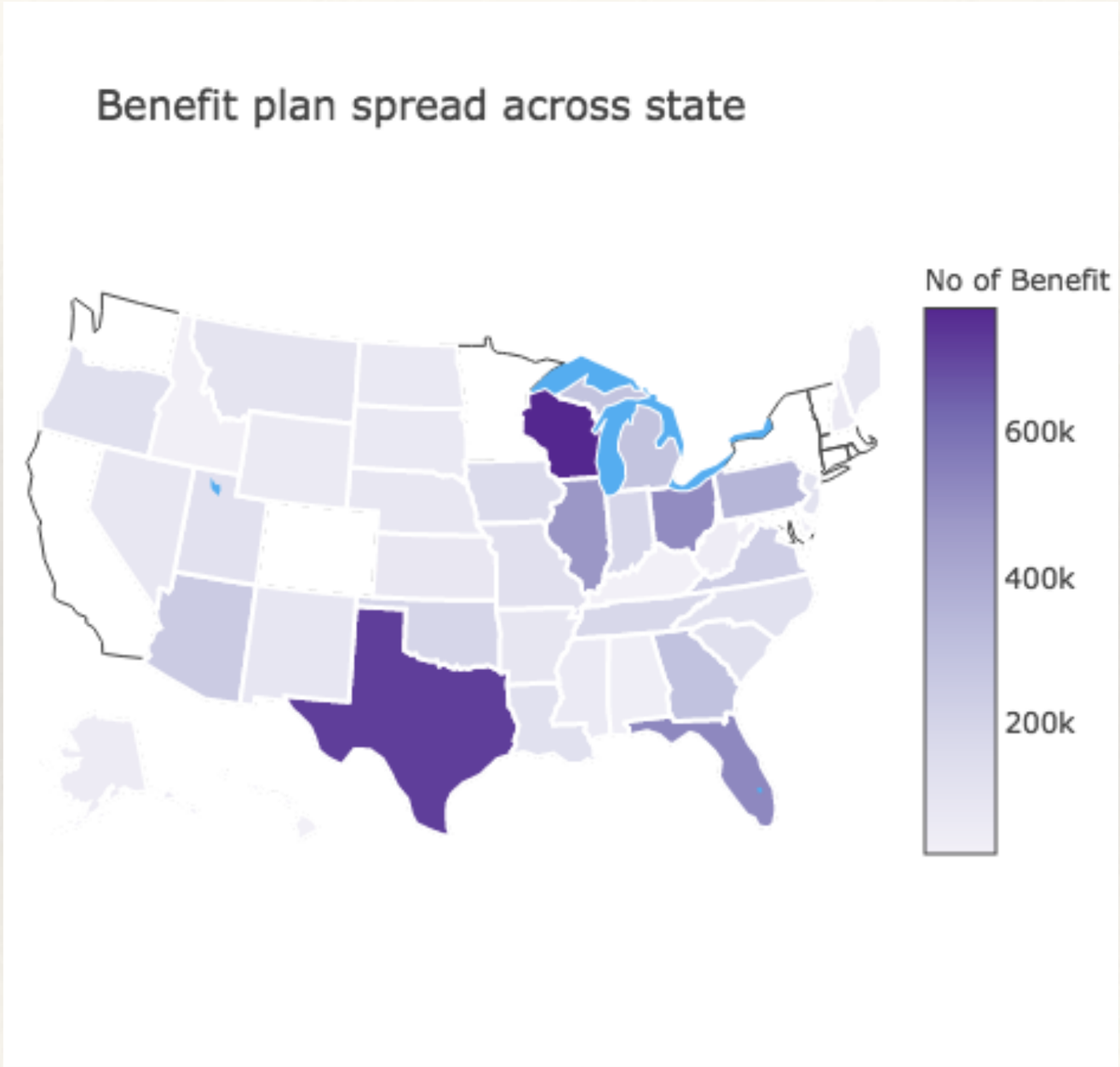
Data Wrangling

- Bringing data to the desire format:
  - Merger of the dataset of year 2017 and 2018.
  - Merger of the dataset of the year 2014-2016 to above merged dataset.
  - Removed columns which were not required for this project.
- We had following data problems:
  - Missing or NaN values
  - Outliers
  - Duplicate rows

- **Data Cleaning:**
  - Columns with “$” sign: to remove the prefix “$”, used replace function and pandas to_numeric function.
  - Age and RatingAreaId columns changed from object data type to numeric for prediction model.
  - MetalLevel column changed to numeric data type using OneHotEncoder.

EDA - Benefits Plan

- Since the inception of exchange in year 2014, there were 40 states have participated in the program. Currently there are 39 states which are using Federal run exchange.
- In 5 years total of 7.2 Million plans offered to the American people.
- Out of those 7.2 Million plans 964 were unique.
- Wisconsin, Texas and Florida top 3 states to offer highest number of plans.



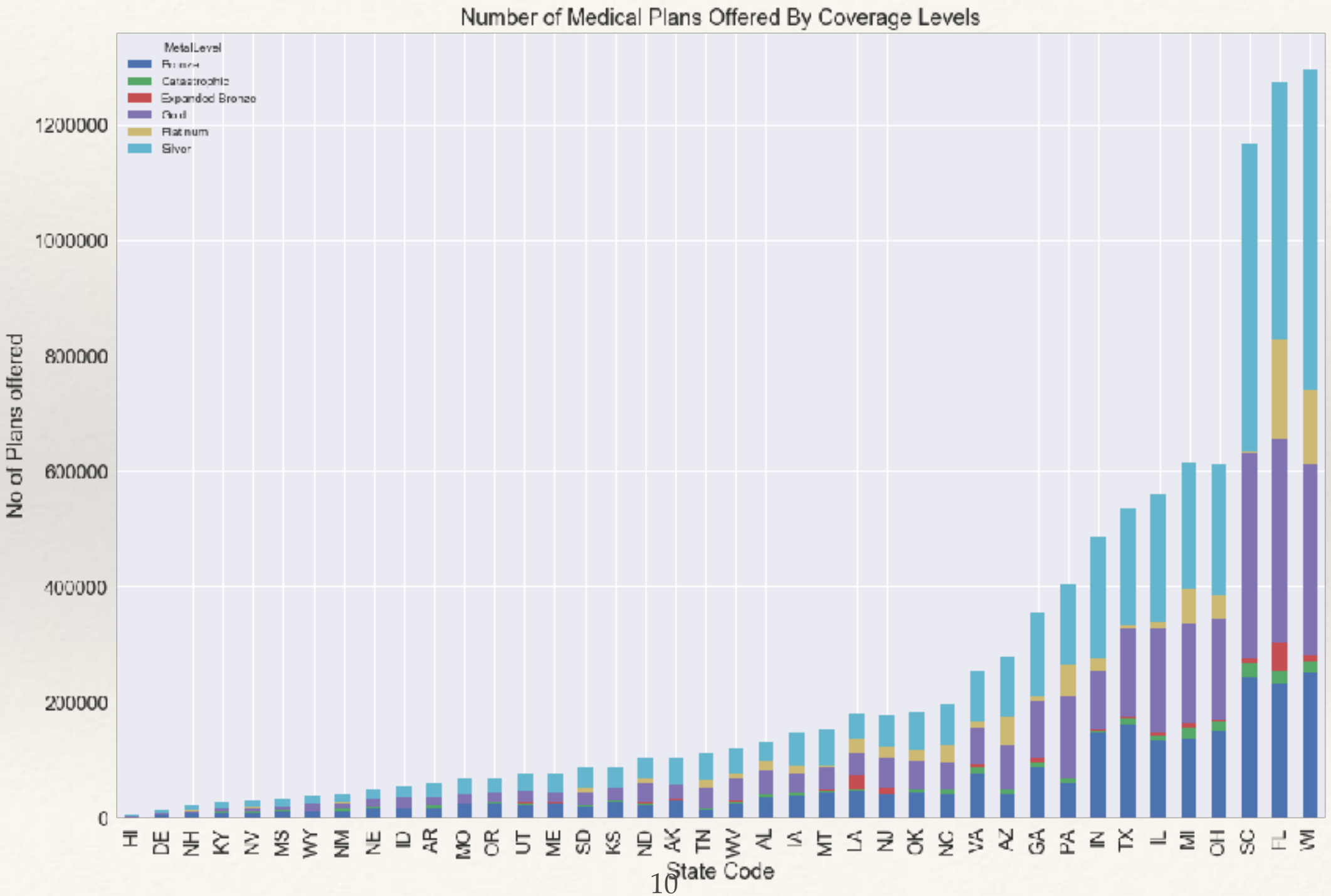
EDA - Individual Monthly Rate

Monthly Individual Rate

5 Cheapest states		5 Costliest states	
State Code	**Average Rate**	**State Code**	**Average Rate**
MO	$167.83	AK	$684.63
MI	$210.09	WI	$490.37
TX	$215.49	IL	$426.92
AR	$230.65	NJ	$415.19
MS	$247.00	SD	$388.62

- Montana is the cheapest states in terms of healthcare individual monthly rate and Alaska is the costliest.
- People of Alaska pay 400% more for same or similar plan than in Montana.
- The premiums for Alaskans on the exchanges for 2017 are at the top(approx. $790) among the states.

EDA - Coverage by: Metal Level



- Silver plan is the most offered and popular plan compared to other metal plan.
- Three states(South Carolina, Florida, and Wisconsin) offered more than 1 million plans to its population, which is almost **14%** of the total plans offered in the United States.

EDA - Notable Findings

- Health Insurance rates go up as a policyholder gets older, with the largest increases coming after age 50.
- This reflects the higher expected share of health care costs that older Americans are expected to utilize.
- Consumers, 64 and older have their premiums capped at 3 times the premiums of the 21 year old base rate.
- When the number of issuer(companies which provides health insurance), is less then premium is high and where the number of issuer is high the premium is low. This way it provides more competitive rates to the consumers, else there will be monopoly of fewer issuers and the consumer will have to pay more monthly cost for healthcare insurance.

Predictive Model

- Predict Individual Rates and Individual Tobacco Rates using regression models.
- Factors that decides the rate
- Under the health care law(ACA), insurance companies can account for only 5 things when setting premiums.
  - 1.Age.
  - 2.Location.
  - 3.Tobacco use
  - 4.Individual vs. family enrollment.
  - 5.Plan category.
- We are predicting Individual rate and Individual Tobacco rate.

- **OLS(Ordinary Least Squares Method):** First we had implemented the OLS method to predict the individual monthly rate. The result for individuals and individual tobacco users are 64% and 62.24% respectively on training dataset. When compared with the true value of the test dataset the results were very disappointing at about 25% and 27% respectively.

- **LinearRegression()** in scikit learn package inPython: This model also gave almost same result as of OLS model.

- **Decision Tree:** Even the Decision tree results were not very satisfactory.
- **Decision Tree With Cross validation:** Once decision tree was implemented with the cross validation the performance of the model increased from **25%** to **73%**.
- 5 - Fold cross validation on the training dataset to train the model in much better way than simply train the model.

Conclusion

- Not all 50 states are participant of health insurance marketplace.
- Monthly insurance premium increases with the increase in age.
- Alaska is the costliest state in terms of health care insurance, where as Montana is the cheapest.
- The number of issuers decreasing in the health insurance marketplace.

- Decision tree with 5 fold cross validation resulted in the model that performs best with the unseen test data with around 73% accuracy.
- The buzz around the exchange: Human decisions are the biggest factor in this model that is performing only 73%.

Recommendations

- Issuers can target states where the number of plans are less, to expand the business and can provided competitive monthly premium options.
- The model can be scaled to predict the monthly premiums of family groups, like couple with one dependent and couple with two dependent and so on.
- This helps in proper decision making for the consumers before buying heath insurance with their budget and kind of coverage they need.

Practical Considerations and Suggestions

- Availability of data about the enrollment in the health insurance market place will help in better analysis, by comparing the actual enrollment with the available options.
- Scale out the model to predict many other premium rates with family group option.
- Dental plan considerations which was not part of this project.

References

- **Dataset:**
  - https://www.kaggle.com/hhs/health-insurance-marketplace
  - http://www.nber.org/data/cms-marketplace.html
- **Health Insurance marketplace:**
  - https://www.healthcare.gov
- **Help:**
  - https://en.wikipedia.org/wiki/Health_insurance_marketplace
  - http://akcommonground.org/high-health-care-costs-in-alaska-facts-causesconsequences-and-remedies/

Thank You!

I would like to thank Springboard and my Mentor Srdjan Santic, for his valuable feedback through out this project and keep me motivated to complete the project.

Questions?