Capstone Project 1 - Milestone Report

Prepared By: Vishal Kumar

Prepared On: February 27, 2018

Healthcare Insurance Marketplace

Milestone Report

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.

Problem Definition

Building a model which will predict the monthly insurance premiums for an individual and individuals tobacco users. Explore the available data to answer some analytical questions 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?

There can be many other analysis based on the findings of the specified questions. Like how the numbers of companies provides insurance has decreased in last two years after the change in federal government. Similarly there can be many more interesting facts can be explored using this kind of analysis.

Potential Clients

This model can be used by anyone who wants to enroll in on the exchange and wants to know what would be his monthly premium based on coverage so that he/she can decide on the plans which would be most suitable for them. Any company which sells healthcare insurance and wants to know whether they can invest more in exchange or they should come out of the exchange based on the kind if environment and smog around this healthcare exchange program. For example the data shows that the last two year the number of companies in the exchange has

dropped significantly around 45%. If this trend continues the marketplace would not survive and any company would not want to invest in the exchange.

Also this model can be used by health insurance providers to improve their sales by comparing the similar plans with their competitors in particular state and specific areas in that states.

About Data

The dataset has been download from the kaggle.com (link is below) and from the National Bureau of Economic Research the link is below.

https://www.kaggle.com/hhs/health-insurance-marketplace
http://www.nber.org/data/cms-marketplace.html

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.

The Marketplace, also known as the Affordable Insurance Exchange, is a key component of the Affordable Care Act (ACA).

CMS describes the Marketplace PUF data in the following way:

The Marketplace PUF includes plan and issuer level information for certified Qualified Health Plans (QHPs) and stand-alone dental plans (SADPs) offered to individuals and small businesses through the Health Insurance Marketplace. The Marketplace PUF includes data from states participating in the Federally Facilitated Marketplaces (FFM), which include State Partnership Marketplaces (SPMs), and states whose State-based Marketplaces rely on the federal information technology platform for QHP eligibility and enrollment functionality. The Marketplace PUF does not contain any data on plans offered in states that established and operate their own Marketplace (State-based Marketplace) and do not rely on the federal platform for QHP eligibility and enrollment functionality, nor does it contain enrollment or claims data. Due to the limitations that this dataset does not contain any data about enrollment we cannot establish the relationship between the total number of insured and uninsured population state wise and in the entire USA. It would have been very interesting to verify the claims that after the establishment of marketplace or exchange the number of uninsured population has been decreased. We could have measured the number and percent and how the trend is over the years.

Data Cleaning/Wrangling Techniques

Generally we do not get the data in the formate best suited for our problems. We always need to do some data cleaning and wrangling to bring the data to the desired format. Following are some common data problems we may come across when dealing with datasets.

- Inconsistent column names
- Missing data (NaN values)
- Outliers
- Duplicate rows

In my dataset, fortunately I have to deal with missing values(NaN values). And hope to see some outliers as well.

To check for missing data points I have used the following method.

1) Read the csv files into pandas dataframe.
2) use ***dataframe.info()*** method to find if there is any missing data. In some cases the dataframe.info() method doesn't show the desire output for these kind of dataframes, I have used ***isnull()*** method to find if there is any missing values in the columns.

We have used following options to deal with the missing values:

1. ***fillna()*** method: There are lot of options available with the fillna() method to deal with the missing values. For example:

   fillna(0) will replace all the NaN values with 0(zero)

2. ***dropna()*** method: to drop the entire data point if there are missing values in all the columns.

There are columns in the data frames which has floating numbers representing the amount with "$" sign as prefix. For our analysis we had to change the column as numeric so that we can perform mathematical calculation on the columns. To convert these kind of values to numerical values we had to get rid of the "$" sign. To do this we have used ***replace*** function on data frame column name and then pandas ***to_numeric*** function to convert the object datatype into numeric datatype. Similar kind of conversion has been done on Age column of the rate data frame, which has the values as object datatype and with the text values as "0-20" where we have to get rid of "-" and covert it to numerical values. Also RatingAreaId column has the similar values, like "Rating Area 10", in these type of values we had to get rid of space and then convert these area into numbers. We needed to perform this data cleaning on these columns as these were part of prediction model using regression.

Initial Findings

Some of the initial findings are:

- Not all the states are part of this marketplace or exchange. Out of 50 states only 40 states were or are part of the federal healthcare insurance marketplace. There are some states which joined this program later and not at the starting and there are also states which have moved out after the inaugural year i.e. year 2014. These states which are not part of the Federal government marketplace program have their own state based marketplace also called SBM.
- People of Wisconsin has the highest number of benefits available to choose from. Texas and Florida are on 2nd and 3rd position in this category. So we would expect the number of companies selling health insurance in Wisconsin would be highest but that is not the case. Texas has the highest number of companies selling health insurance followed by Michigan, Ohio, Pennsylvania and then stands Wisconsin. This leads to the further investigation of what is the ratio of number of companies and the number of benefits offered in states, if the ratio is less then we can expect those states to be more expansive in terms of monthly premium rates.
- The number of Issuers(companies that sells health insurance in the marketplace) has been decreased significantly in last two years. After the inception of exchange in the year 2014, the number of issuers in each participating states were increasing or were same for the year 2015 and 2016. In some states these numbers were decrease by very small numbers. But it has started to fall down in year 2017 and by the year 2018 these numbers were significantly dropped about 45%. This can be explained by the fact that Federal Government has changed late 2016 and we all know that ACA was in jeopardy. Hence issuers would have thought of getting themselves as distant as possible.
- Except for the individual tobacco rate, all other category has range of approximately $5 - $25 fluctuation since the inception of ACA(Affordable Care Act) in year 2014. The monthly premium rates has not been increased or decreased significantly for all the category except for the individual tobacco rate. It has increased approximately around $200 since the inception of marketplace or exchange.