**Who is buying dental insurance from the health exchanges?**

**By Justin Vena**

**Background:**

The ACA (Patient Protection and Affordable Care Act) was passed in 2010, and changed the healthcare industry. ACA introduced the Health Insurance Marketplace (or called the health exchanges) to the United States. Health and Dental Insurance products became available for sale in 2013 and came into effect for the 2014 plan year.

Insurance Products on the Health Exchange: Before the Health Exchange was established; most people received their insurance from their employer or buying directly from the insurer. This lead to the question: what types of consumers will buy from the exchanges? The exchanges make it easy for a consumer to shop around between insurance products, knowing the offerings of price and plans available.

**Project Question**:

What characteristics can we successfully use to classify which ZIP codes are more likely to buy an ACA dental insurance product?

**Data Collection:**

In order to solve this problem, I needed to gather insurance, Medicare, and Census information. The data came from two sources, Kaggle.com and Census.gov.

Kaggle.gov is a data science competition website. It has several public data sets that are available to look at. I used the Health Insurance Marketplace dataset. The dataset provides us with insurance and medicare data from 2014 to 2016. Has data on 35 states and the health and dental insurance products sold to each state. Several different datasets are included, but I will primarily use the Service Area Data, which provides information about the exchange plan geographic information.

Census.gov has public datasets the US government publishes. I used the 2014 ZIP Code/Income dataset to pull geographic, and income data.

**Features used for Analysis:**

Kaggle Data Features:

* Geographic: State (35), County, ZIP
* Plan Standard: Low or High
* Child/Adult Plan: If the plan is for a child and/or an adult
* Year: 2014-2016

Census.gov Data Features:

* ZIP
* Income

Kaggle Data Clean-Up and Issues:

* Missing ZIP Codes is currently the biggest issue
* How to manipulate data to improve model
  + Exclude the missing ZIP codes from the model
  + Replace the missing ZIP with average income of all the ZIP codes offered per state
* Run using data from all years separately and see how it impacts the results.
* Run with and without the Child/Adult Flag. Child only plans are usually Medicaid only, which would be interesting to see if they are sold in certain area codes over others.
* Run initial program with just Texas since has largest number of plans

Census Data Clean-Up and Issues:

* ZIP/Income information still needs to be cleaned.
* Will do a check of average using bracket counts versus the average given

**Selecting and Running the Model:**

We want to solve our problem by segmenting the different types of plans into clusters that exhibit similar characteristics.

First attempt will be to use K-Means Algorithm to cluster the groups into categories. We test using a cluster of 3, but will also check using greater. Will validate using silhouette score.

After testing if time, check using DBSCAN clustering as well for comparison.

**Code:**

See Github

**Business Applications and Conclusions:**

The study is good for looking to see what types of plans should be sold in different area codes. If we knew the profitability of types of plans, we could focus on how to sell individual and segment plans around the United States. If there is no true difference, start including more variables to more effectively segment types of plans.

Since I am still working on the program, I have not come to any conclusions about if the model is valid or not yet.