Who is buying dental insurance from the health exchanges?

Justin Vena
DS5

Adapting to the changes in the Healthcare Marketplace

ACA (Patient Protection and Affordable Care Act)

- Established in 2010, changed the healthcare industry in many different ways.
- Introduced the Health Insurance Marketplace (or called the health exchanges)
- Health and Dental Products became available for sale in 2013 and came into effect for 2014

Adapting to the changes in the Healthcare marketplace

Insurance Products on the Health Exchange:

- Before the Health Exchange was established, most people received their insurance from (1) their employer or (2) buying directly from the insurer.
- What types of consumers will buy from the exchanges?
- The exchanges make it easy for a consumer to shop around between insurance products.

How will we approach this?

Project Question:

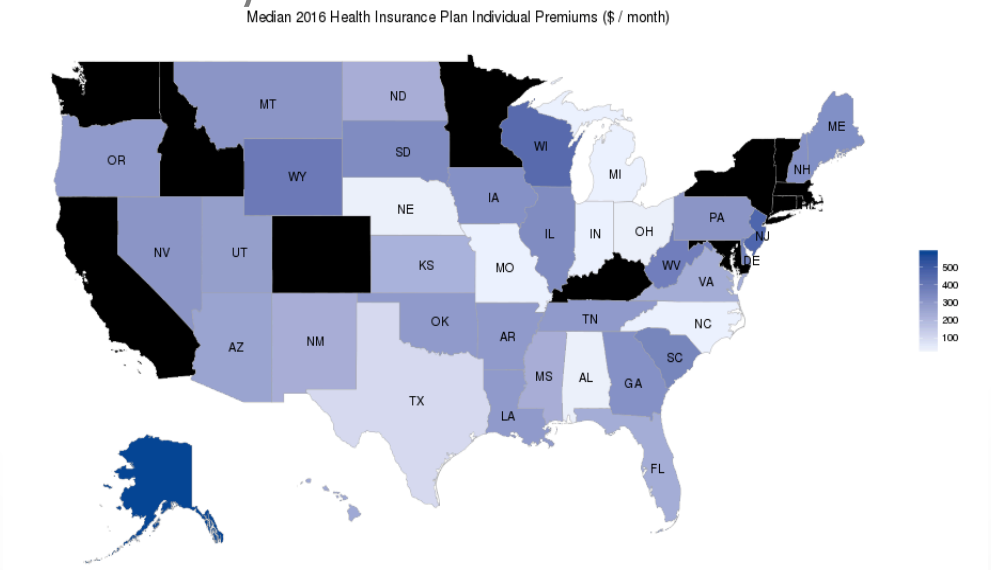
What characteristics can we successfully use to classify which ZIP codes are more likely to buy an ACA dental insurance product?

Let's gather insurance, Medicare, and Census information to solve this.

Where's the Data?

Kaggle.com Health Insurance Marketplace Dataset:

- Providing us with the Insurance and Medicare Data
- Contains data from 2014 to 2016 from the CMS (Center for Medicare Services)on 35 States (No Washington State)



- Several different datasets are included on Kaggle,
- but will primarily use the Service Area Data

Where's the Data?

Census.gov's public data:

- 2014 ZIP Code/Income data set
- Geographical: State, ZIP
- Income: Counts by Income Bracket, Average Income

Data Exploration and Investigation

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

State by Plan Level:

	State	High	Low	Grand Total
1	TX	4,254	5,778	10,032
2	GA	1,749	1,908	3,657
3	MI	1,680	1,752	3,432
4	VA	1,607	1,654	3,261
5	NE	1,419	1,419	2,838
…	…	…	…	…
31	NJ	84	126	210
32	AK	87	116	203
33	ME	64	64	128
34	NH	40	40	80
35	DE	12	24	36
	Grand Total	23,465	27,411	50,876

Data Exploration and Investigation

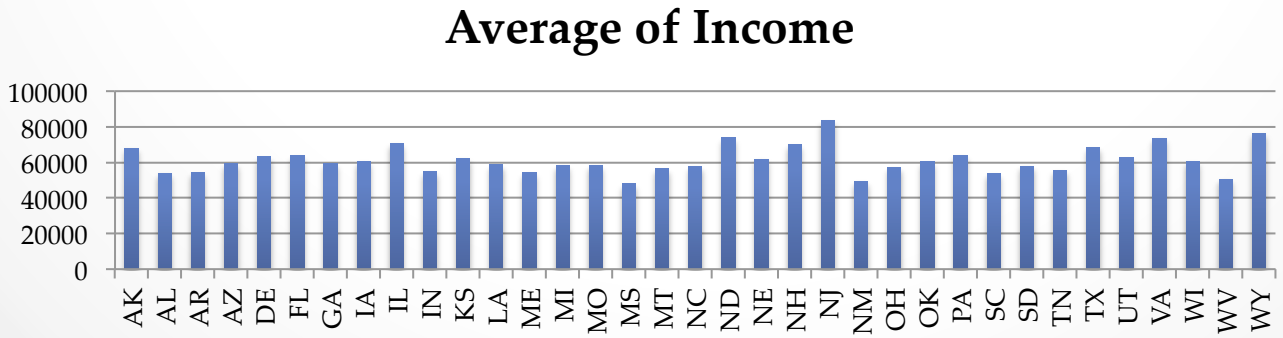
Kaggle Data Clean-Up and Issues:

- Missing ZIP Codes is currently the biggest issue
- Manipulated the data to improve model in two ways:
  - Exclude the missing ZIP codes from the model
  - Replace the missing ZIP with average income of all the ZIP codes offered per state
- Replaced the missing ZIP with average income per state
- Ran only using 2014 data since its when Census and Kaggle data best aligned

Data Exploration and Investigation

Census Data Clean-Up and Issues:

- ZIP/Income information needed to be cleaned and organized
- Did reasonable checks of averages using bracket counts versus the average given.
- Used Total Family Income Unadjusted per ZIP code. Used average of all ZIP codes per state to be used for "0" or unreported ZIP codes.



Next Steps: Running the Model

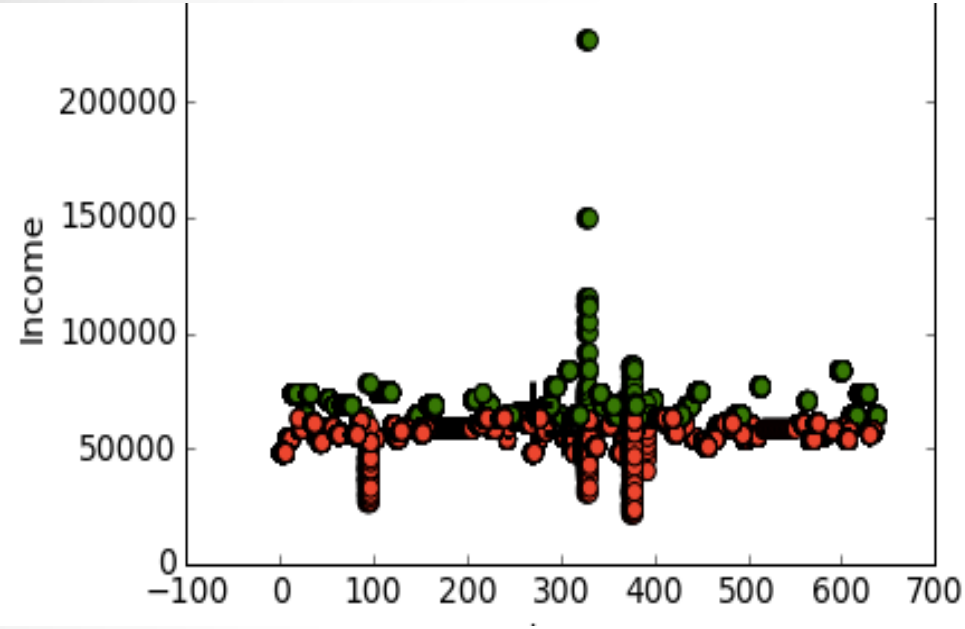
Before running and preparing the data:

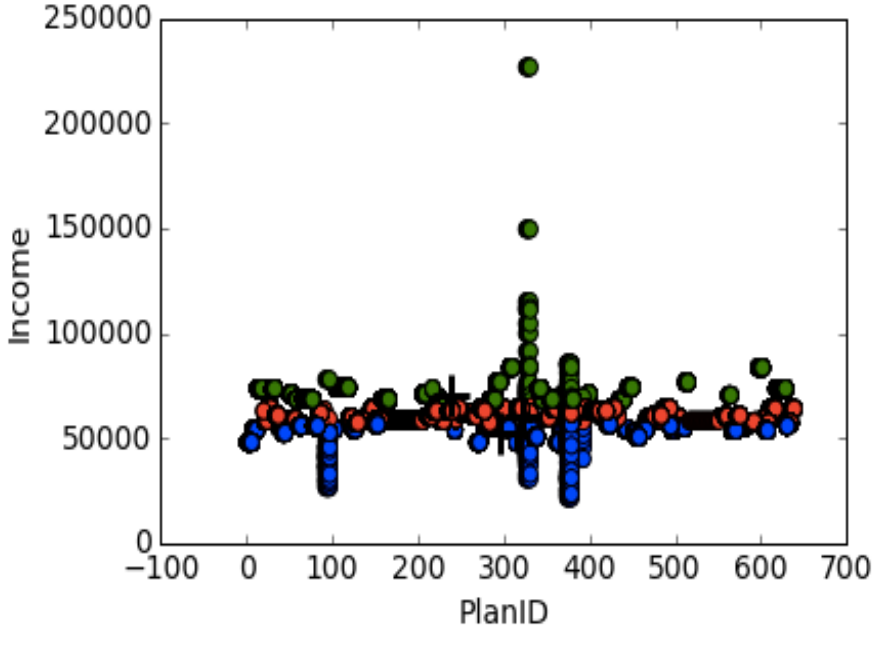
- Mapped categories so would run in Python
- Filled in missing values with state average incomes

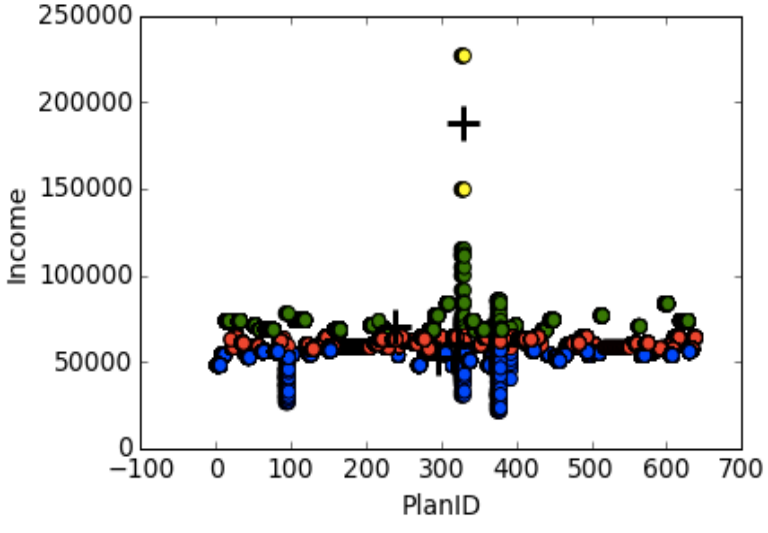
Model selection:

- We want to segment the dental plans into groups, so will use clustering
- Use KMeans Clustering and use silhouette scores to evaluate the model.

Model Results







Model Results

Clusters and Silhouette Scores:

Clusters	Silhouette Scores:
2	0.60891
3	0.61235
4	0.61455
5	0.62639

Average Income Per Cluster:

2 Clusters	3 Clusters	4 Clusters	5 Clusters
57,163	52,460	52,460	47,884
69,045	60,499	60,499	57,350
	70,536	70,452	62,487
		188,042	70,459
			188,042

Model Results

Minimum Income Per Cluster:

2 Clusters	3 Clusters	4 Clusters	5 Clusters
21,521	21,521	21,521	21,521
63,258	56,710	56,710	52,644
	65,878	65,878	60,440
		114,958	66,698
			114,958

Maximum Income Per Cluster:

2 Clusters	3 Clusters	4 Clusters	5 Clusters
62,927	56,452	56,452	52,426
226,625	65,415	65,415	59,754
	226,625	114,912	65,988
		226,625	114,912
			226,625

Business Applications and Conclusions

- The purpose of study is useful to see what types of plans should be sold in different ZIP codes.
- Corresponding how "rich" the benefits are for a particular plan to a certain income level, we could a better focus on how to sell individual and segment plans around the United States.

Business Applications and Conclusions

- With the average income known about the plans sold to which ZIP codes, insurers can focus advertising on, or creating plans and networks that better suit these customers.
- Looking at the range the clusters place the incomes with the corresponding richness of the plans would be a great marketing tactic.

Business Applications and Conclusions

Extra Time with Data and Model

- I would add more plan features and use dimensionality reduction to see how I could improve the model.
- More elements could be added to improve and make the model more robust such as Maximums, Deductibles, Individual v Family Rules, Multi-State Plans