
Health Insurance Marketplace Using Census and Rent Cost

DurianCandy

Sunny Kwong Charles Siu
Sean Tey Andrew Young

Agenda

- **Intro**
- **Datasets**
- **Pre Processing**
- **Visualizations**



Most Flexible



Most Affordable

	PPO Preferred Provider Organization	EPO Exclusive Provider Organization	POS Point-of- Service	HMO Health Maintenance Organization
Primary Care Physician (PCP) required	✓	S	✓	✓
Referral required to see a specialist	✗	✗	S	✓
"In-network" benefits	✓	✓	✓	✓
Non-emergency "out-of-network" benefits	✓	✗	✓	✗
Emergency coverage	✓	✓	✓	✓
TIP: Find the most affordable network with your doctors				

S = Sometimes

How much do **insurance costs** varies between **plan types**?

How much do **insurance costs** varies between **states**?

Do **insurance costs** share patterns with **rental prices** and **census data**?

Datasets

Title: Health Insurance Marketplace

Source: Kaggle (Centers for Medicare & Medicaid Services)

Shape:

- **Rate:** 24 Cols x 12.7M Obs
- **PlanAttr:** 176 Cols x 77.4K Obs
 - Each observation represents an insurance plan
 - Each column represents an attribute of the insurance plan

Title: Zillow Rental Values

Source: Zillow Research

Shape: 112 Cols x 934 Obs

- Each observation represents an a region
- Each column represents monthly *Zillow Rental Index* (A smoothed measure of the median estimated market rate rent across a given region and housing type.)

Title: US Census Demographic Data

Source: Kaggle (US Census Bureau)

Shape:

2015 County: 37 Cols x 3220 Obs

- Each observation represents an a county in US
- Each column represents a measurement in census

Pre-Processing Steps

- Joining rent data and insurance data
 - Average Zillow Rent Index by Year and State
 - Average Insurance Rate by Plan Type, Year, and State
- Joining census data and insurance data
 - Average Insurance Rate by State
 - Average of Census Data by State

EMR Configuration Goals

- Do you get a significant speed up everytime you double the number of nodes?
- Is one beefy machine better than 16 or 32 smaller ones?

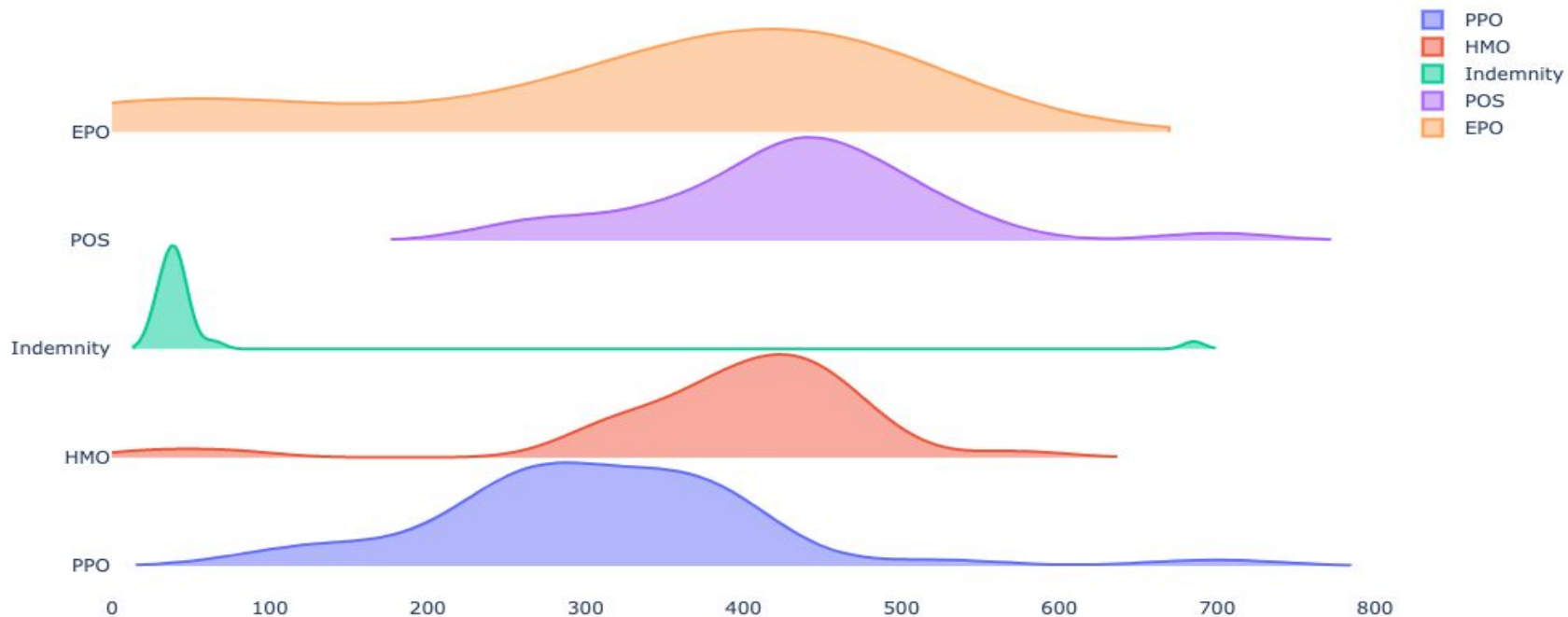
EMR Configurations Summary

Configurations we will test:

Instance Type	Node Count	Total vCPU	RAM (GiB)	Total price per hour (\$)	Running Time (min)
C5.xlarge	4	16	32	0.68	6.52
C5.xlarge	8	32	64	1.36	7.27
C5.xlarge	16	64	128	2.72	7.05
C5.xlarge	32	128	256	5.44	6.85
C5.18xlarge	1	72	144	3.06	6.3

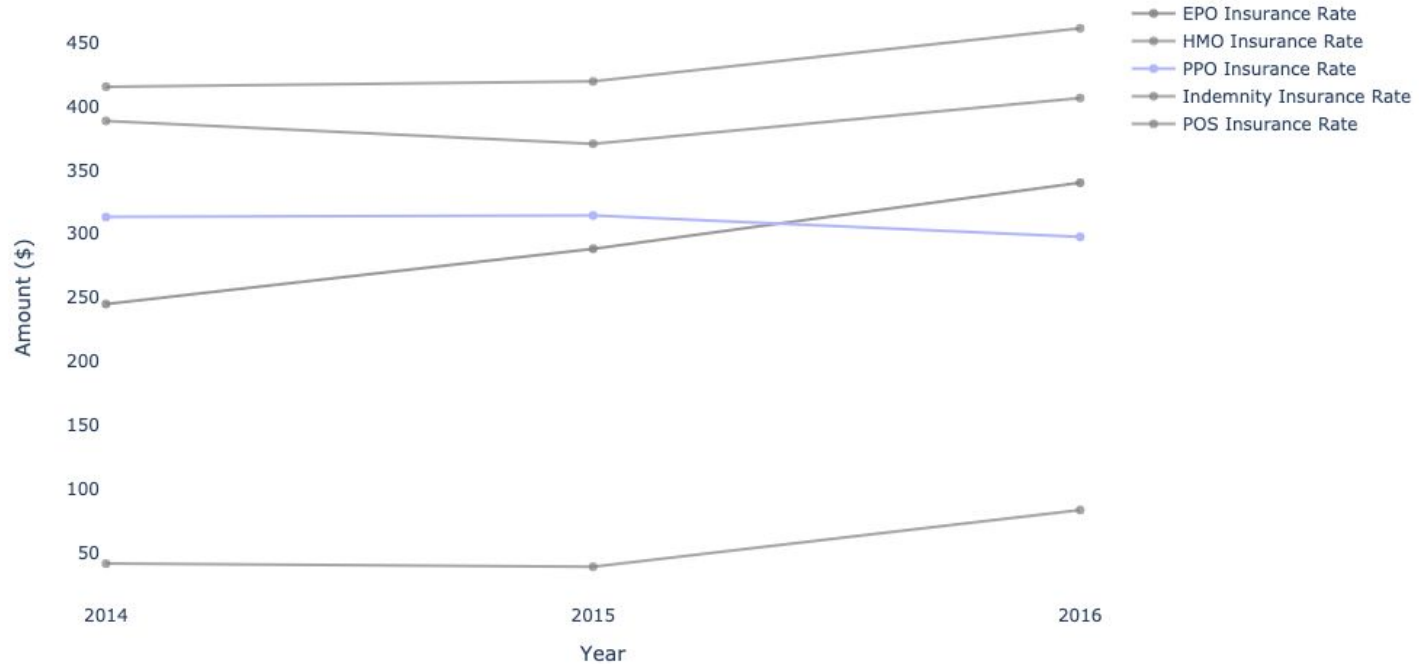
Visualization #1

Distribution of Rate by Plan Type



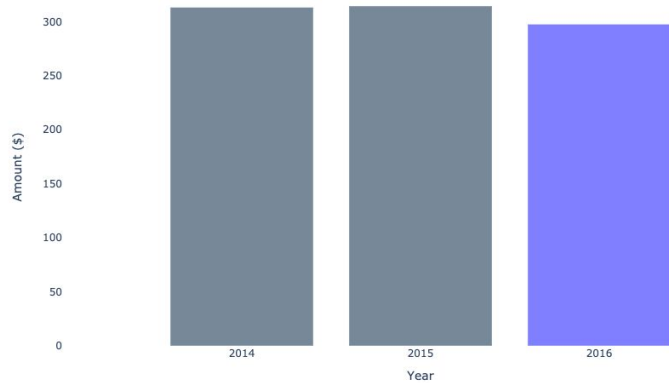
Visualization #2

Trend of Insurance Rate vs Zillow Rental Index

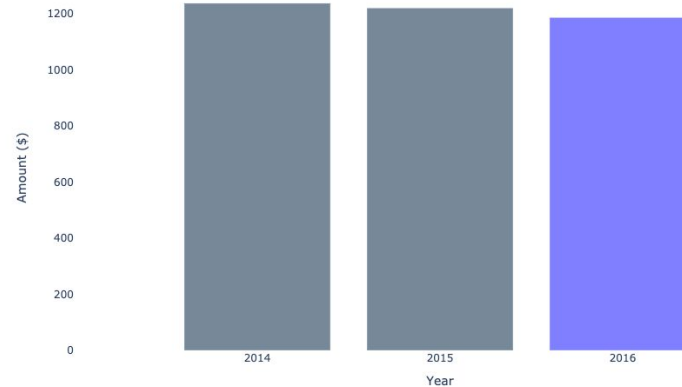


Visualization #3

PPO Rate by Year

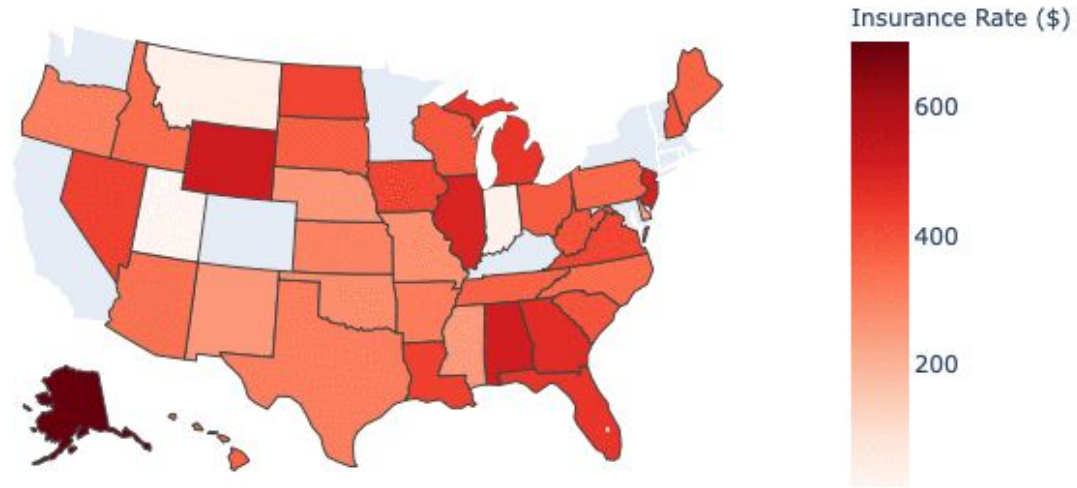


ZRI Rate by Year



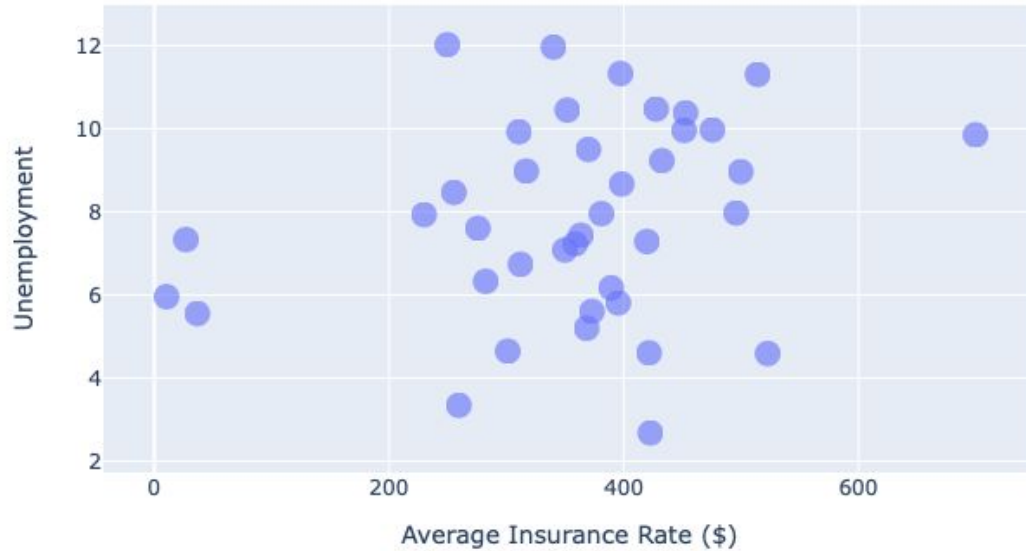
Visualization #4

Insurance Rate by State



Visualization #5

Average Insurance Rate vs. Unemployment



Thank you