How Prudential Can Use Mortality Rate Data

By Ramkrishan Sahu

Problem Statement:

We have to find out how Prudential can use the Mortality rate data set of the states in USA.

About The Data Set:

Provider	Institute for Health Metrics and Evaluation (IHME)
Geography	United States (USA)
Coverage type	Country
Time period covered	01/1980 - 12/2014
Summary: This dataset provides the estimates of mortality rates of 21 mutually exclusive causes at the county-level for each state in the United States between 1980-2014 (quinquennial.	

Understanding The Data Set:

US county-level trends in mortality rates for 21 major causes of death, 1980–2014

Location	FIPS	Mortality Rate, 1980*	Mortality Rate, 1985*	Mortality Rate, 1990*	Mortality Rate, 1995*	Mortality Rate, 2000*	Mortality Rate, 2005*	Mortalit
United States		1.52 (1.44, 1.61)	3.16 (3.11, 3.22)	11.45 (11.34, 11.56)	16.61 (16.48, 16.74)	5.97 (5.92, 6.02)	4.87 (4.83, 4.91)	3.20 (3.1
Alabama	1	1.46 (1.33, 1.59)	2.15 (2.03, 2.27)	8.03 (7.79, 8.25)	14.10 (13.76, 14.43)	5.47 (5.30, 5.65)	4.67 (4.51, 4.83)	3.35 (3.2
Autauga County, Alabama	1001	0.95 (0.68, 1.29)	1.44 (1.12, 1.86)	6.57 (5.49, 7.84)	10.75 (9.18, 12.50)	3.62 (2.94, 4.37)	3.32 (2.66, 4.14)	2.37 (1.8
Baldwin County, Alabama	1003	0.84 (0.63, 1.12)	1.44 (1.18, 1.75)	6.51 (5.71, 7.40)	11.18 (9.89, 12.59)	3.82 (3.28, 4.44)	3.18 (2.71, 3.72)	2.27 (1.8
Barbour County, Alabama	1005	1.11 (0.82, 1.48)	1.81 (1.43, 2.27)	7.86 (6.66, 9.25)	15.65 (13.49, 18.12)	6.33 (5.23, 7.50)	6.19 (5.08, 7.50)	4.67 (3.6
Bibb County, Alabama	1007	0.88 (0.59, 1.26)	1.21 (0.88, 1.60)	5.13 (4.06, 6.25)	9.69 (7.92, 11.75)	2.89 (2.24, 3.73)	2.65 (1.96, 3.52)	1.74 (1.2
Blount County, Alabama	1009	1.08 (0.76, 1.53)	1.24 (0.94, 1.64)	4.66 (3.78, 5.64)	8.50 (7.13, 10.05)	2.71 (2.16, 3.38)	2.33 (1.77, 2.96)	1.48 (1.0
Bullock County, Alabama	1011	1.48 (1.04, 2.01)	2.53 (1.90, 3.24)	11.13 (9.19, 13.41)	20.31 (16.99, 24.05)	7.63 (6.01, 9.44)	7.05 (5.39, 8.97)	5.10 (3.7
Butler County, Alabama	1013	1.30 (0.91, 1.77)	2.04 (1.52, 2.60)	8.75 (7.26, 10.47)	15.09 (12.75, 17.75)	5.89 (4.75, 7.26)	5.60 (4.33, 7.01)	4.09 (3.0
Calhoun County, Alabama	1015	1.98 (1.46, 2.67)	2.38 (1.89, 2.99)	8.09 (6.94, 9.25)	14.09 (12.57, 15.85)	4.66 (3.96, 5.48)	3.98 (3.27, 4.83)	2.55 (2.0
Chambers County, Alabama	1017	1.40 (1.02, 1.91)	2.44 (1.96, 2.99)	10.03 (8.53, 11.67)	17.45 (15.20, 19.79)	6.82 (5.69, 7.99)	6.03 (4.89, 7.40)	4.41 (3.4
Cherokee County, Alabama	1019	1.39 (0.95, 1.99)	1.50 (1.09, 2.06)	5.18 (4.21, 6.38)	9.36 (7.70, 11.40)	2.88 (2.14, 3.77)	2.29 (1.67, 3.04)	1.43 (0.9
Chilton County, Alabama	1021	0.65 (0.45, 0.91)	0.94 (0.70, 1.24)	4.68 (3.85, 5.66)	9.16 (7.64, 11.08)	2.68 (2.13, 3.27)	2.27 (1.72, 2.88)	1.49 (1.0
Choctaw County, Alabama	1023	1.02 (0.71, 1.40)	1.67 (1.26, 2.17)	6.94 (5.64, 8.36)	12.27 (10.15, 14.57)	4.65 (3.68, 5.77)	4.02 (3.09, 5.13)	2.84 (2.0
Clarke County, Alabama	1025	1.12 (0.80, 1.56)	1.74 (1.37, 2.25)	7.08 (5.88, 8.42)	12.41 (10.33, 14.55)	4.48 (3.66, 5.44)	3.89 (3.04, 4.88)	2.70 (2.0
Clay County, Alabama	1027	1.08 (0.72, 1.57)	1.37 (0.99, 1.89)	5.37 (4.11, 6.86)	9.92 (7.82, 12.22)	3.18 (2.25, 4.33)	2.71 (1.83, 3.94)	1.65 (1.0
Cleburne County, Alabama	1029	0.84 (0.60, 1.17)	1.10 (0.82, 1.46)	4.76 (3.83, 5.80)	8.93 (7.23, 10.92)	2.69 (2.08, 3.44)	2.25 (1.65, 2.91)	1.39 (0.9
Coffee County, Alabama	1031	0.71 (0.48, 1.01)	1.10 (0.82, 1.41)	4.98 (4.11, 5.97)	9.54 (8.20, 11.06)	3.35 (2.73, 4.06)	3.06 (2.41, 3.88)	2.16 (1.6
Colbert County, Alabama	1033	1.25 (0.85, 1.75)	1.51 (1.10, 1.98)	5.60 (4.66, 6.78)	9.57 (8.01, 11.24)	3.13 (2.45, 3.92)	2.57 (1.97, 3.29)	1.74 (1.2
Conecuh County, Alabama	1035	1.64 (1.12, 2.38)	2.48 (1.86, 3.28)	9.31 (7.61, 11.25)	16.02 (13.32, 18.86)	6.52 (5.16, 8.16)	6.14 (4.75, 7.80)	4.57 (3.3
Coosa County, Alabama	1037	1.06 (0.73, 1.49)	1.49 (1.09, 1.96)	6.23 (4.95, 7.61)	11.41 (9.29, 13.66)	3.80 (2.85, 4.82)	3.27 (2.34, 4.30)	2.16 (1.4

HIV-AIDS & tuberculosis | Diarrhea, lower respiratory | Neglected tropical diseases | Maternal disc ...

Understanding The Data Set:

Each sheet is shares the Data on the Mortality rates caused by one out of 21 different reasons.

Columns:

- Location: 56 States and sub their locations
- FIPS code (i.e. Location codes)
- Mortality rates (quinquennial i.e. 5 years of data) with the Average, Minimum and Maximum Mortality rates

We have 21 such sheets for each of the 21-major cause of death

Pre-processing & Cleaning

Pre-processing steps:

- We merged the 21 sheets and separated the Average, Minimum and Maximum rated.
- We are keeping:
  - The mortality rates for each state as a whole
  - And the average mortality rates for each state

Data Sheet After Pre-processing

Columns:

- Location: 56 States
- FIPS code (i.e. Location codes)
- Category: 21-major cause of death
- Average mortality rates (quinquennial i.e. 5 years of data)
- Also created 3 new empty columns for prediction:
  - To predict the mortality rate for 2020 and 2025
  - Trend in Premium: Predicts if we should Increase or Decrease the premium in the future for these causes

Theory Of Process

1. Try to predict the trend in mortality rate for the year 2015-2020 & 2020-2025
2. If mortality rate is Increasing (or positive) increase the premium to reduce the risk.
3. If mortality rate is Decreasing (or negative) decreases the premium to a more competitive premium to acquire more customers.

Tools & Algorithm Used:

Jupyter Notebook

Language: Python

Libraries: Pandas, Numpy, Matplotlib,

Algorithm: Linear model from sklearn

Code Logic

1. Finding the Line of Best Fit or Linear Regression Line
2. Predicting mortality rate for year 2015-2020 and 2020-2025:
    1. Predicting the mortality rate from the coefficient and constant found from the Linear Regression and inserting the column 2020, 2025
3. Predicting trend in Premium
    1. If coefficient > 0 => "Increase the premium"
    2. Else => " Decrease the premium"

Output Data Set

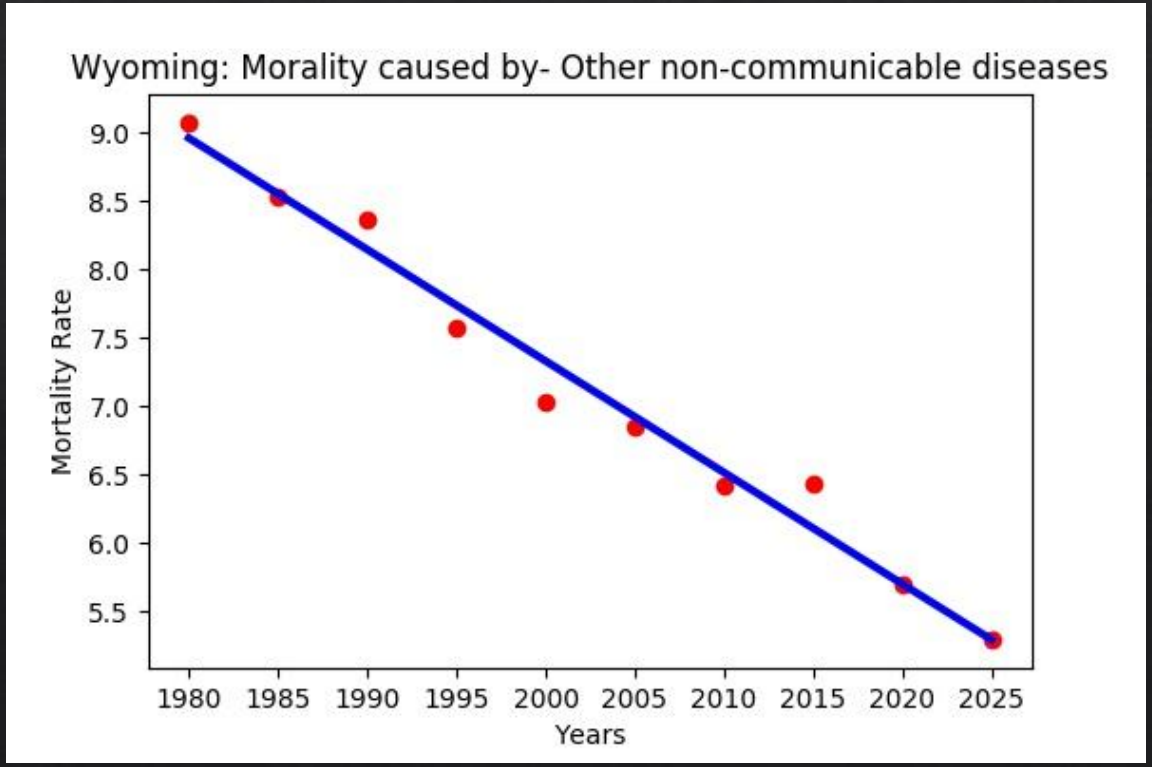
Apart from the initial data it includes:

Mortality Rate for year 2015-2020 in column 2020

Mortality Rate for year 2020-2025 in column 2025

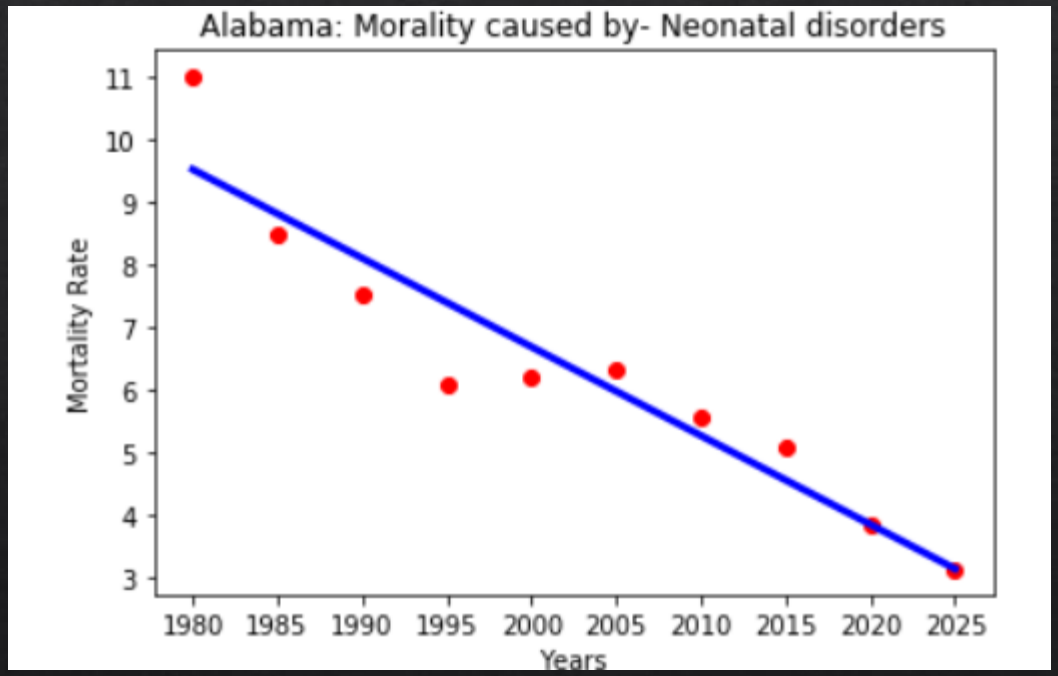
Graphs For The Particular State Due To Cause:

Creating plot for mortality rate from 1980 – 2025 in Wyoming caused by Other non-communicable diseases



Graphs For The Particular State Due To Cause:

Creating plot for mortality rate from 1980 – 2025 in Alabama caused by Neonatal disorders



Further Work:

- Can use better algorithms like ensemble of multi algorithms, xgboost.
- If current Premium data is provided we can even get the most optimal premium trends for the future.

Thank you

Ramkrishan Sahu

✆: +91 9029527880

✉: ram04190@gmail.com

in : Linkedin.com/in/ramkrishansahu/