Diabetes Prevalence and Unexplored Relations

INDS 4997

Capstone Project

Kayley Reith

05-04-2025

Literature Review

There are 38.4 million Americans diagnosed with diabetes, accounting for over 11.5% of the population (CDC 2024). This disease is prevalent as 1 in 10 adults (CDC 2024) will be diagnosed with this chronic metabolic disorder affecting blood glucose levels (NIH, n.d.). Diabetes is the 8th leading cause of death and has a morbidity rate of 16.2% (ABA 2023). Given its prevalence and health impact, approximately 413 billion dollars have been allocated toward expenses of diabetes research, intervention, and treatments (ABA 2024). While different forms of diabetes exist (Prediabetes, Type 1, Type 2, Unspecified) 90-95% of all diagnosed diabetes take the form of Type 2 (CDC 2024). Recent projections and forecasts suggest that without any form of intervention, diabetes prevalence will rise to "643 million Americans by 2030 and 783 million Americans by 2045", ultimately normalizing the livelihood of the disease (Cleveland Clinic 2023).

Gap in Knowledge

Despite previous and existing research efforts and public health initiatives, diabetes continues to be a growing crisis and health epidemic. The stagnation in progress and efforts call out a need for data-driven approaches to extract new insight. This research explores lesser-unstudied variables at the state level to uncover significance and correlation between diabetes prevalence. In order to identify overlooked contributors to diabetes prevalence, this study conducts a thorough analysis at a detailed level, which will allow health, medical, and policy professionals the information and opportunity to implement tailored strategies and intervention methods to reduce diabetes risk.

Data Collection and Management

Data was obtained by online scraping publically accessible state-level indicator data sources. Numerous sources were combined, including the Bureau of Labor Statistics (BLS), Kaiser Family Foundation (KFF) News, The Sentencing Project, the Centers for Disease Control (CDC), and the United States Census Bureau. Because there was insufficient information on the prevalence of diabetes in all states, Kentucky and Pennsylvania were excluded from the data management process. Due to the small number of missing values 10 out of 3550 observations, the values were imputed using variable averages in order to address missingness and noise reduction. Standardization of variables was also achieved by eliminating percentages, ratios, and commas in order to establish numerical unison.

Candidate Techniques and Methodology

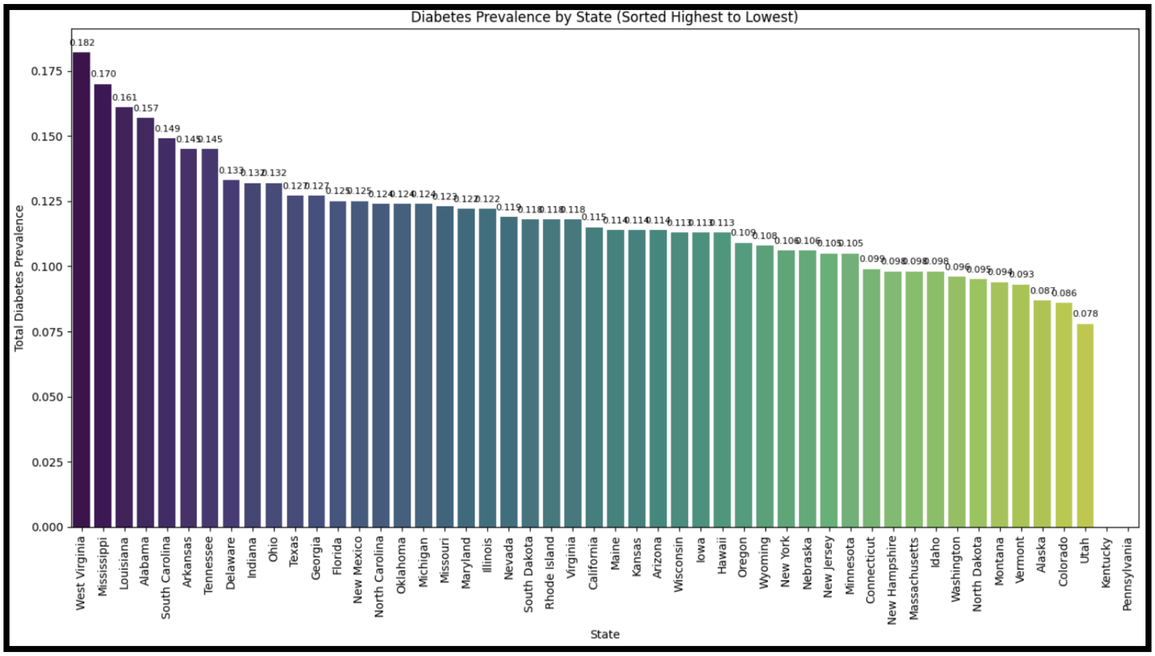
Several algorithmic models were constructed and tested to ensure optimal performance in order to thoroughly explore whether aspects are significant and potential associations between diabetes risk and prevalence using state-level indicators. A combination of supervised and unsupervised methods were examined in order to fully comprehend the data's patterns.

The ensemble learning method, Random Forest Regressor handles complex relationships between features, creating a robust way to identify feature importances without overfitting, to the factors influencing diabetes prevalence. The Simulation Deterministic Sensitivity Analysis (DSA) is a method used to assess how changes in specific input parameters of a model affect the model's output. This model simulated future impacts based on incremental differences on variables increasing and decreasing in volume in relation to diabetes state prevalence (+/- 5%).

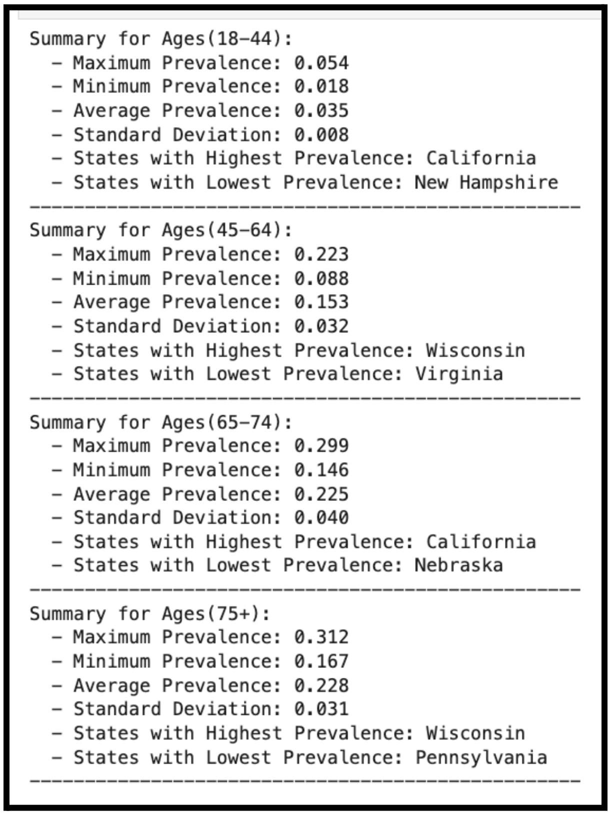
This method gave new information for public health planning by enabling the evaluation of the potential effects of changes and fluctuations in individual characteristics on estimated diabetes prevalence rates. The K-means clustering algorithm was applied to group similar data points together and discovers underlying patterns and relationships between state diabetes prevalence levels and understudied random variables to highlight influences on state health. This unsupervised learning approach helped uncover potential health disparities and regional trends by analyzing how states naturally cluster based on the input data.

Results and Findings

During the data exploration phase, the prevalence of diabetes was investigated at the state level by examining the total number of reported and diagnosed cases. Additionally, prevalence was reported individually for four age groups: 18–44 years, 45–64 years, 65–74 years, and 75 years and older. Visualising the frequency of cases (Figure 1.1) showcases how West Virginia has the highest state prevalence of 18.2% and Utah has the lowest with 7.8%. Kentucky and Pennsylvania were excluded from the total prevalence analysis but included in the age group clusters. Amidst the age group diabetes prevalence breakdown (Appendix A), six states (California, New Hampshire, Wisconsin, Virginia, Nebraska, and Pennsylvania) were identified as having the highest and lowest risk levels (Figure 1.2). These states were selected for further exploration.



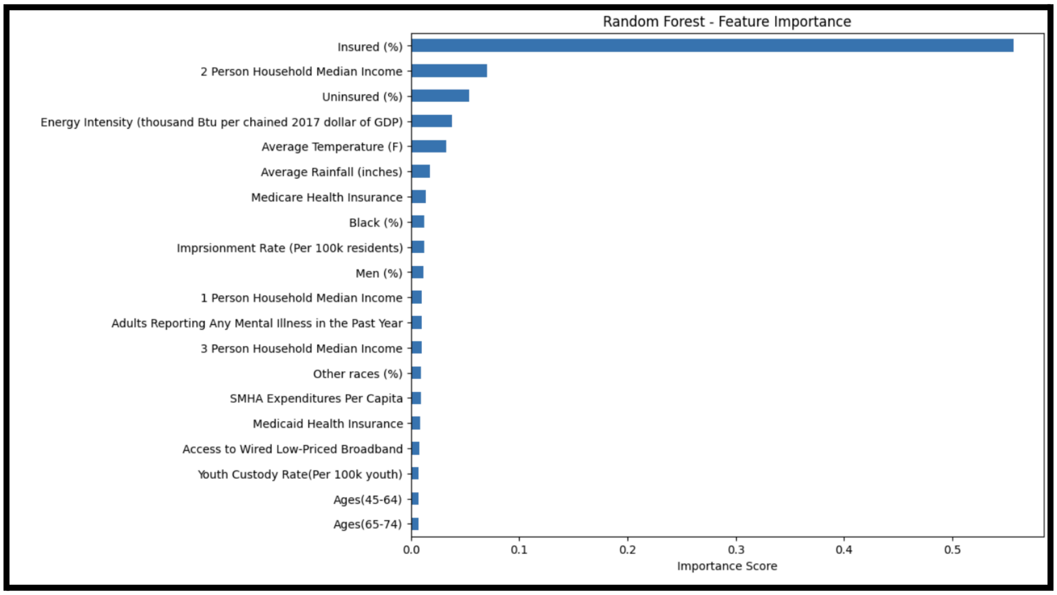
(Figure 1.1)



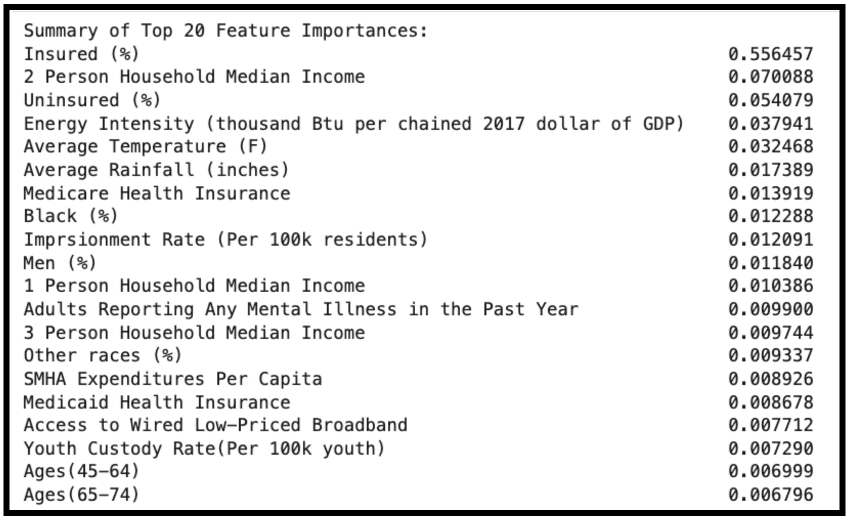
```
Summary for Ages(18-44):
  - Maximum Prevalence: 0.054
  - Minimum Prevalence: 0.018
  - Average Prevalence: 0.035
  - Standard Deviation: 0.008
  - States with Highest Prevalence: California
  - States with Lowest Prevalence: New Hampshire
---------------------------------------------------
Summary for Ages(45-64):
  - Maximum Prevalence: 0.223
  - Minimum Prevalence: 0.088
  - Average Prevalence: 0.153
  - Standard Deviation: 0.032
  - States with Highest Prevalence: Wisconsin
  - States with Lowest Prevalence: Virginia
---------------------------------------------------
Summary for Ages(65-74):
  - Maximum Prevalence: 0.299
  - Minimum Prevalence: 0.146
  - Average Prevalence: 0.225
  - Standard Deviation: 0.040
  - States with Highest Prevalence: California
  - States with Lowest Prevalence: Nebraska
---------------------------------------------------
Summary for Ages(75+):
  - Maximum Prevalence: 0.312
  - Minimum Prevalence: 0.167
  - Average Prevalence: 0.228
  - Standard Deviation: 0.031
  - States with Highest Prevalence: Wisconsin
  - States with Lowest Prevalence: Pennsylvania
---------------------------------------------------
```

(Figure 1.2)

Continuing investigation all 50 states were included in the generalization of the Random Forest Regressor results including all variables in the dataset (Figure 1.3). The most significant characteristic was determined to be Insured (%), underscoring the vital role that insurance coverage plays in forecasting the frequency of diabetes at the state level. Features like SMHA (State Mental Health Agencies) Expenditures Per Capita and Electric Vehicle Count, were significantly less important, suggesting they hold little bearing on the model's projections (Figure 1.4).



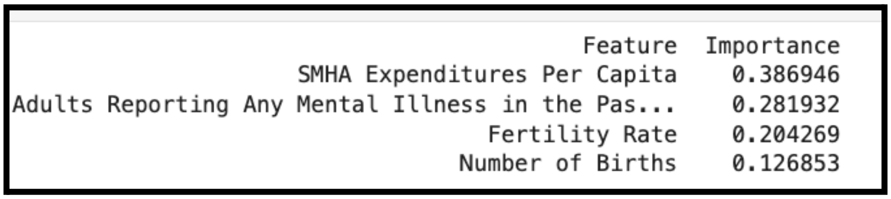
(Figure 1.3)



```
Summary of Top 20 Feature Importances:
Insured (%)                                                     0.556457
2 Person Household Median Income                                0.070088
Uninsured (%)                                                   0.054079
Energy Intensity (thousand Btu per chained 2017 dollar of GDP)  0.037941
Average Temperature (F)                                         0.032468
Average Rainfall (inches)                                       0.017389
Medicare Health Insurance                                       0.013919
Black (%)                                                       0.012288
Imprsionment Rate (Per 100k residents)                          0.012091
Men (%)                                                         0.011840
1 Person Household Median Income                                0.010386
Adults Reporting Any Mental Illness in the Past Year            0.009900
3 Person Household Median Income                                0.009744
Other races (%)                                                 0.009337
SMHA Expenditures Per Capita                                    0.008926
Medicaid Health Insurance                                       0.008678
Access to Wired Low-Priced Broadband                            0.007712
Youth Custody Rate(Per 100k youth)                              0.007290
Ages(45-64)                                                     0.006999
Ages(65-74)                                                     0.006796
```

(Figure 1.4)

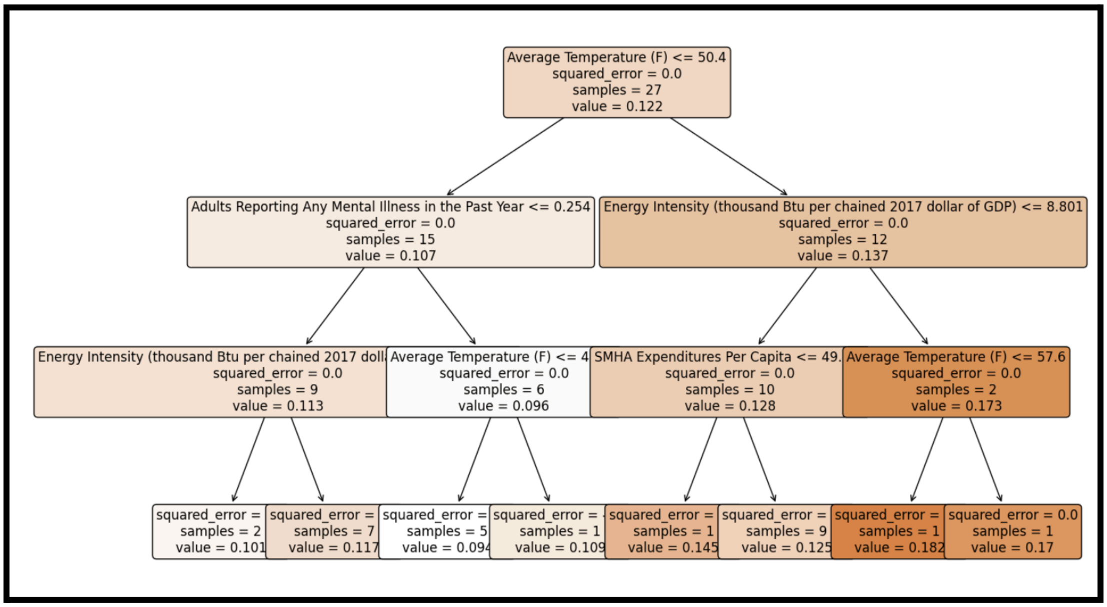
The healthcare model utilized health-related data variables (Figure 1.5) in the Random Forest Regressor model, with the dataset split into 80% training and 20% testing. The model demonstrated that heath related data generalized at the state level had strong precision with a Mean Squared Error (MSE) of 0.000236 and an average prediction error of about 1.2% as indicated by the Mean Absolute Error (MAE). While the health data model explains approximately 31% (R² value) of the variance in state-level diabetes rates, a significant portion of the variance remains unexplained, highlighting potential areas for further exploration. Among the healthcare variables, SMHA Expenditures Per Capita emerged as the most influential predictor (38% significance), followed by Adults Reporting Any Mental Illness in the Past Year (28%) suggesting that factors related to mental health may play a crucial role in shaping diabetes prevalence across states.



```
                                          Feature  Importance
                     SMHA Expenditures Per Capita    0.386946
Adults Reporting Any Mental Illness in the Pas...    0.281932
                                   Fertility Rate    0.204269
                                 Number of Births    0.126853
```

(Figure 1.5)

Since additional variables were needed to gain a better understanding of the model and diabetes prevalence, energy (carbon emissions and energy intensity by dollar GDP value) and environmental-related (average rainfall and temperature) variables were integrated alongside the health data into the model. This model achieved high results with a strong performance with a Mean Squared Error (MSE) of 0.000068 and Mean Absolute Error (MAE) of 0.006633. The MAE indicates that, on average, the new model's predictions were off by approximately 0.6633%, reflecting a high degree of accuracy. The R² value (0.8034) means that 80.34% of the variation in state-level diabetes prevalence was explained by the selected features, highlighting the model's ability to capture a significant portion of the underlying patterns in the data with energy and environmental information. This was further illustrated in a decision tree (Figure 1.6) that highlights how Average Temperature plays the most important role in diabetes prevalence as the root node. It emphasizes that when Average Temperature is above 50.4°F, higher energy intensity levels (>8.80) and greater SMHA Expenditures Per Capita lead to predicted values of 18%. This model exhibits a relationship and association between energy use, mental health, and healthcare spending with diabetes prevalence.

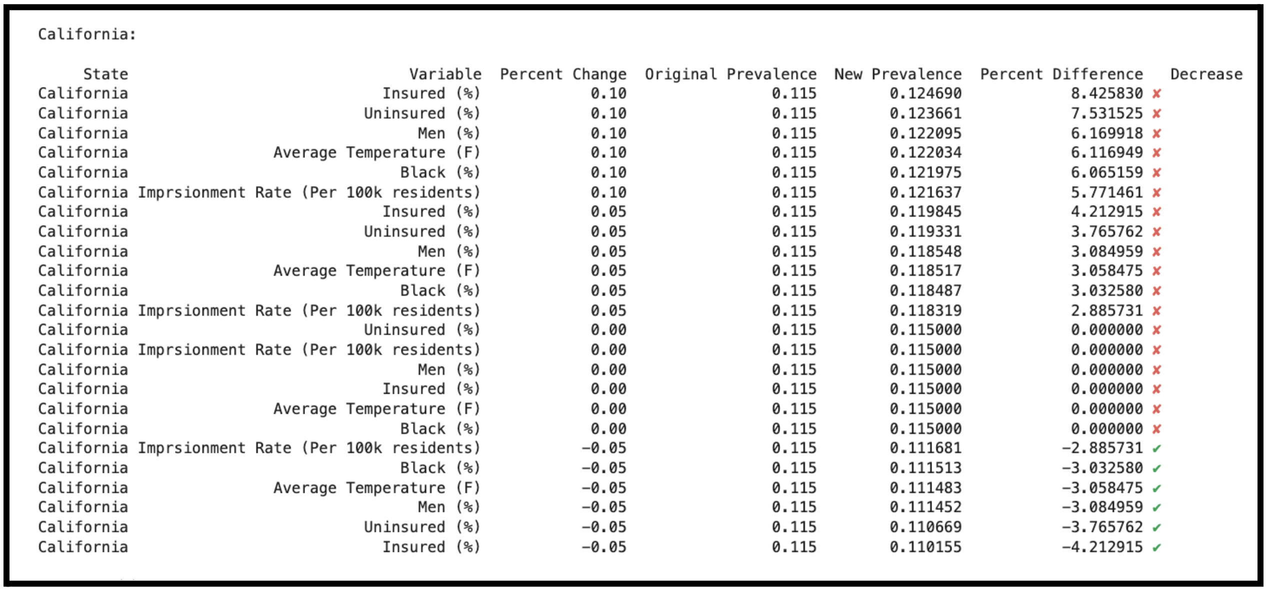


(Figure 1.6)

The subset of significant states was used by the Simulation of Deterministic Sensitivity Analysis (DSA) to model hypothetical changes with the strongest and most significant correlation factors. The percentage changes used fall within a narrow range of -5% to 10% in order to reflect the reality of minute changes in diabetes risk, prevalence, and intervention that happen gradually over time.

The DSA reveals that in states like California, a 10% increase in insured individuals corresponded to an approximate 8.4% rise in the high-risk diabetes population. This finding contradicts the usual expectation that higher insurance coverage would lead to lower risk and prevalence of diabetes. Instead this could indicate a flaw in the model, a delay in the impact of insurance increases, or that more insured individuals are getting tested which leads to an increase in the total diagnoses count at young stages of the disease. This doesn't equate to a worsening of health but the prevalence is reported quicker. This information trends across all states in the subset.

While this model unveils new insights, its evaluation deems it inadequate. The R^2 value indicates approximately 10% of the variation in diabetes prevalence is explained by the model, suggesting weak explanatory power and uncertainty. The Root Mean Squared Error (RMSE) shows how predictions are off by about 8.94%, indicating moderate accuracy but room for growth. The Mean Absolute Error (MAE) shows how the model's predictions deviate by 7.91% from the actual values, showcasing a low error magnitude but still yielding inaccuracies. Ultimately this model alone can not capture the complex associations of diabetes prevalence.



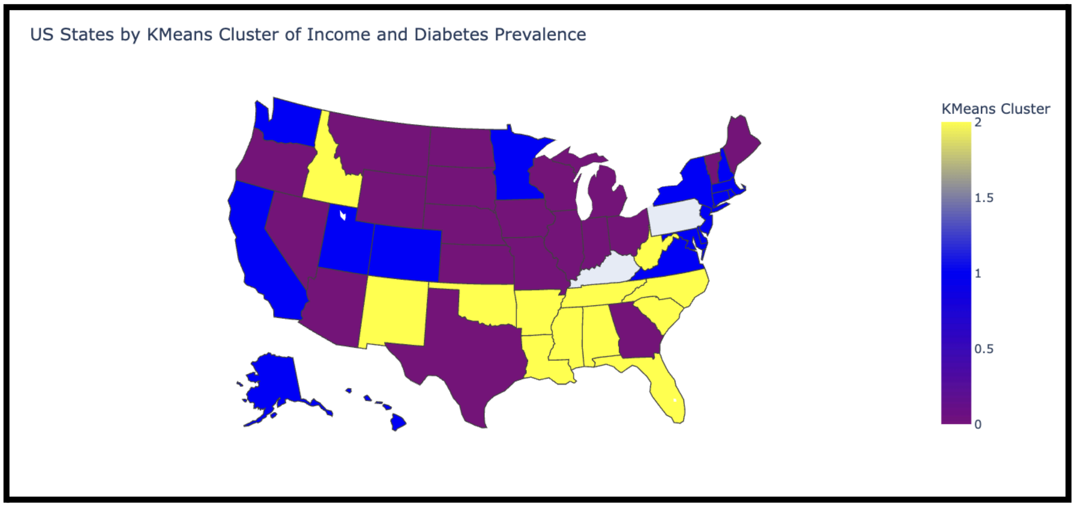
```
California:

     State                            Variable  Percent Change  Original Prevalence  New Prevalence  Percent Difference  Decrease
California                        Insured (%)            0.10                0.115        0.124690            8.425830 ✘
California                      Uninsured (%)            0.10                0.115        0.123661            7.531525 ✘
California                            Men (%)            0.10                0.115        0.122095            6.169918 ✘
California            Average Temperature (F)            0.10                0.115        0.122034            6.116949 ✘
California                          Black (%)            0.10                0.115        0.121975            6.065159 ✘
California Imprsionment Rate (Per 100k residents)        0.10                0.115        0.121637            5.771461 ✘
California                        Insured (%)            0.05                0.115        0.119845            4.212915 ✘
California                      Uninsured (%)            0.05                0.115        0.119331            3.765762 ✘
California                            Men (%)            0.05                0.115        0.118548            3.084959 ✘
California            Average Temperature (F)            0.05                0.115        0.118517            3.058475 ✘
California                          Black (%)            0.05                0.115        0.118487            3.032580 ✘
California Imprsionment Rate (Per 100k residents)        0.05                0.115        0.118319            2.885731 ✘
California                      Uninsured (%)            0.00                0.115        0.115000            0.000000 ✘
California Imprsionment Rate (Per 100k residents)        0.00                0.115        0.115000            0.000000 ✘
California                            Men (%)            0.00                0.115        0.115000            0.000000 ✘
California                        Insured (%)            0.00                0.115        0.115000            0.000000 ✘
California            Average Temperature (F)            0.00                0.115        0.115000            0.000000 ✘
California                          Black (%)            0.00                0.115        0.115000            0.000000 ✘
California Imprsionment Rate (Per 100k residents)       -0.05                0.115        0.111681           -2.885731 ✔
California                          Black (%)           -0.05                0.115        0.111513           -3.032580 ✔
California            Average Temperature (F)           -0.05                0.115        0.111483           -3.058475 ✔
California                            Men (%)           -0.05                0.115        0.111452           -3.084959 ✔
California                      Uninsured (%)           -0.05                0.115        0.110669           -3.765762 ✔
California                        Insured (%)           -0.05                0.115        0.110155           -4.212915 ✔
```

(Figure 1.7)

The K-Means Clustering algorithm was deployed several times with a focus on socioeconomic factors like income and employments extrapolating across the four age groups diabetes prevalence and a model with all states' prevalence centred around technology factors like number of internet providers and BroadBand (internet connection high speed wide-bandwidth) access.

The optimal K-cluster group was three found through a looping mechanism and shown on the elbow plot (Appendix B). The subset of socioeconomic factors were different levels of median house income by the number of people, employment percent changes, and total gross income across the four age groups (Appendix C). The results of the three clusters revealed distinct patterns in diabetes prevalence across states. The states that funneled into cluster 0 had lower socioeconomic factors with lower income, lower employment and often lead towards lower diabetes prevalence like Illinois and Indiana. Meanwhile cluster 1 had states with higher income and moderate employment which led to the lowest diabetes prevalence overall in states like Utah and Colorado. Lastly, cluster 3 showcases states with lower income and employment, and higher diabetes prevalence, especially among older populations (75+) the prevalence was highest in West Virginia, Tennessee, and Oklahoma. This was visualized in a heatmap by clusters excluding Kentucky and Pennsylvania (Figure 1.8).


US States by KMeans Cluster of Income and Diabetes Prevalence
KMeans Cluster
2
1.5
1
0.5
0

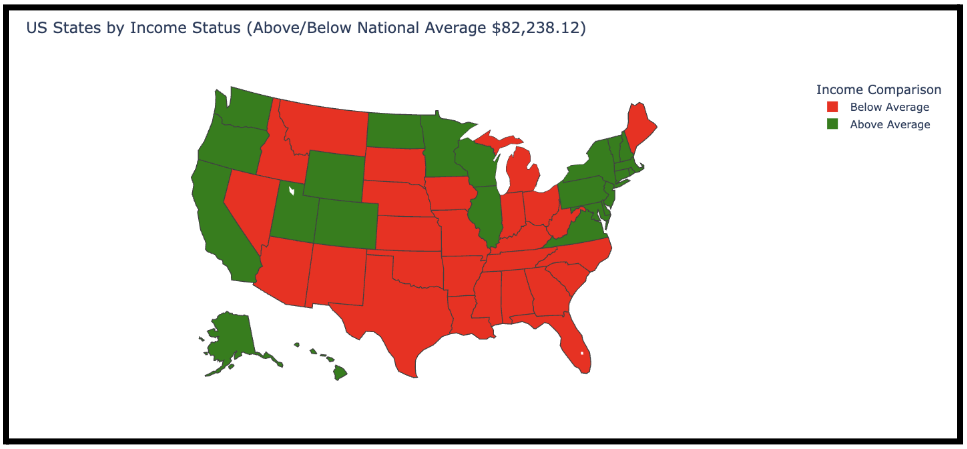
(Figure 1.8)

The geographical display highlights that much of the Midwest region (purple states), along with Texas and several scattered states on each coast, falls into Cluster 0, characterized by lower socioeconomic status and moderate levels of diabetes prevalence. The blue colored states scattered throughout the United States have higher income levels and lower state employment rates and lend themselves towards the lowest levels of diabetes prevalence in Cluster 1. Meanwhile the southeastern region (yellow states) of the United States belong to Cluster 2, where the highest risk of diabetes is apparent, additionally West Virginia and Idaho are high risk as low income states with moderate employment.

Verification of income and employment status was cross-referenced by establishing the baseline threshold of the national average income across all 50 states being $82,238.12 (Figure 1.9). There are 22 states that are above the average income level and 16 states in Cluster 1 identified as low risk for diabetes prevalence. Out of the 22 states with an income above the national average, 16 were classified into the low-risk diabetes cluster (Cluster 1), while the remaining 6 were funneled into high-risk clusters (Cluster 0 or 2). This results in a conditional probability of the chances of low diabetes prevalence given the levels of above-average income.

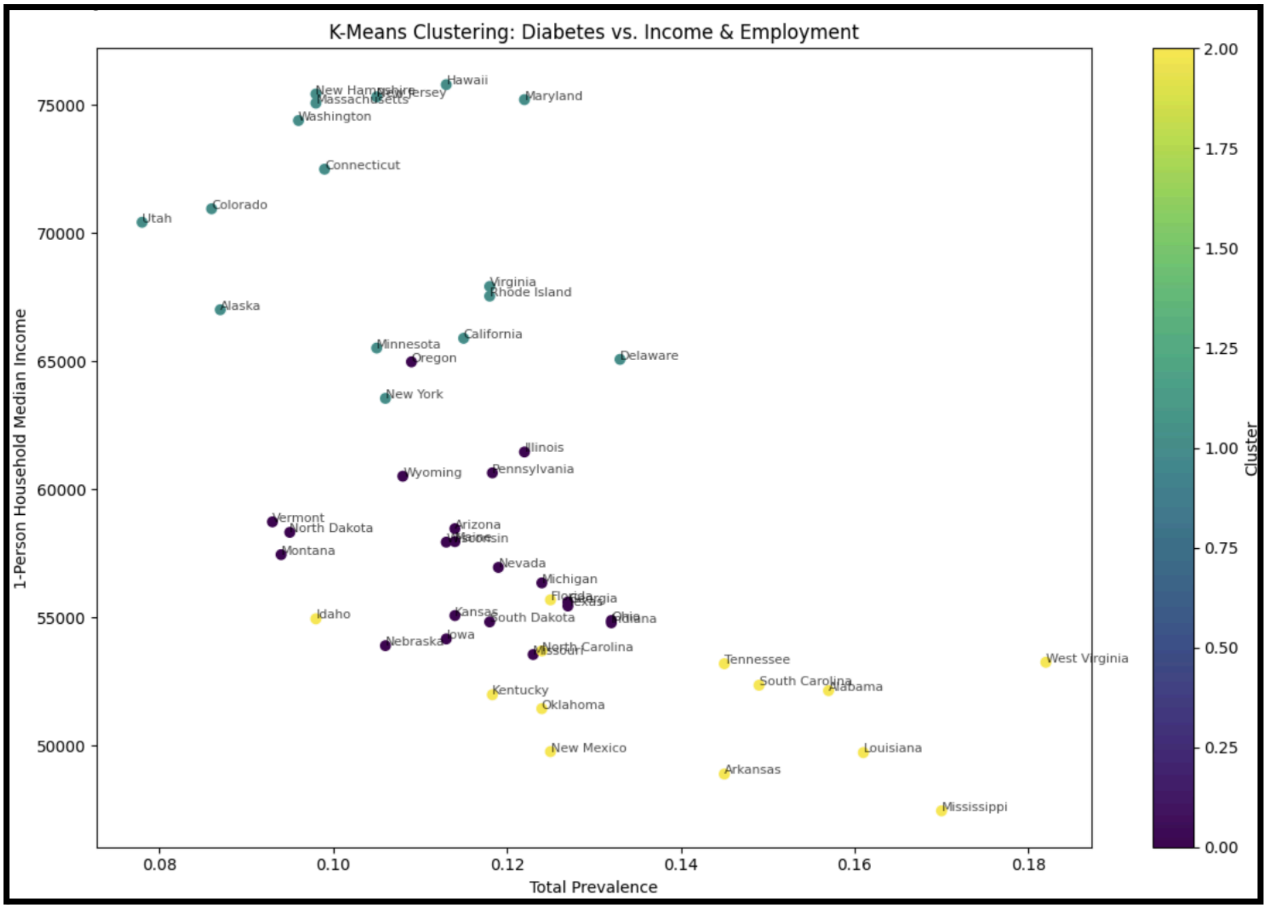
$$P(Low\ Diabetes \mid High\ Income) \;=\; \frac{16}{22} = 0.727$$

Therefore there is a 72.7% chance that a state with above-average income also has low diabetes prevalence. This finding highlights a strong correlation of income levels and employment status linked to healthier outcomes and less risk for diabetes.


US States by Income Status (Above/Below National Average $82,238.12)
Income Comparison
🟥 Below Average
🟩 Above Average

(Figure 1.9)

The other K-Means model is clustering states' diabetes prevalence centered on technology-related factors of internet access, providers, and speed (Figure 2.1). The three clusters were further dissected based on the national average of internet providers in each state (137.36) and categorized into above or below the threshold. Cluster 0 (purple states) had 10 of 35 states above the average of internet providers, while Cluster 1 (blue states) had 3 of 3 states above the average and Cluster 2 (yellow states) had 3 of 12 states above the average.



(Figure 2.1)

The model's results yielded that Cluster 0 had the highest average prevalence of approximately 0.1235 and Cluster 2 had the lowest average prevalence of approximately 0.1023. Cluster 0 has 10 states with above average access to internet providers and has the highest average diabetes prevalence. This correlation could indicate a possible relationship between the number of internet providers linking towards a higher state diabetes prevalence (not implying causation). Whereas clusters that had states with less access to the number of internet providers had a lower average of diabetes prevalence by 2.11%.

The magnitude of this difference is shown in a state like Wisconsin where a 2.11% difference could affect 119,200 people. Or a denser populated state like Florida, where the 2.11% difference could affect over 464,000 more people, reinforcing how even small shifts in prevalence can significantly impact public health. It was reported that "on average people with diabetes incur annual medical expenditures of $19,736", thus this shift in prevalence could potentially affect the spending of diabetes medical expenses by $1.97 billion to $9.1 billion (Parker, et.al 2024). Which allows more resources and funding to be allocated for the redesign of future diabetic health initiatives.

Limitations and Constraints

Several limitations and constraints impacted the scope and accuracy of this analysis. The stagnancy in public health initiatives with social determinants of health and diabetes risk have made none to little progress. Therefore establishing a baseline of unexplored variables was difficult to navigate at the state level. The lack of publicly available diabetes and health related data are not consistently accessible across all states, limiting the granularity and inclusiveness of the study. This also calls into question the gross oversimplification of analyzing health results at the state level. Multicollinearity was a restriction of the variables obtained; for instance the problem arose when examining the associations between insured and uninsured populations, as well as between employment by industry and unemployment rates. More significant results could be gained through localized data from city, regional, or individual health groups statistics. The expansion and acquisition of more variables would provide a better opportunity for the models to perform adequately and uncover more relationships that could impact diabetic health.

Future Directions

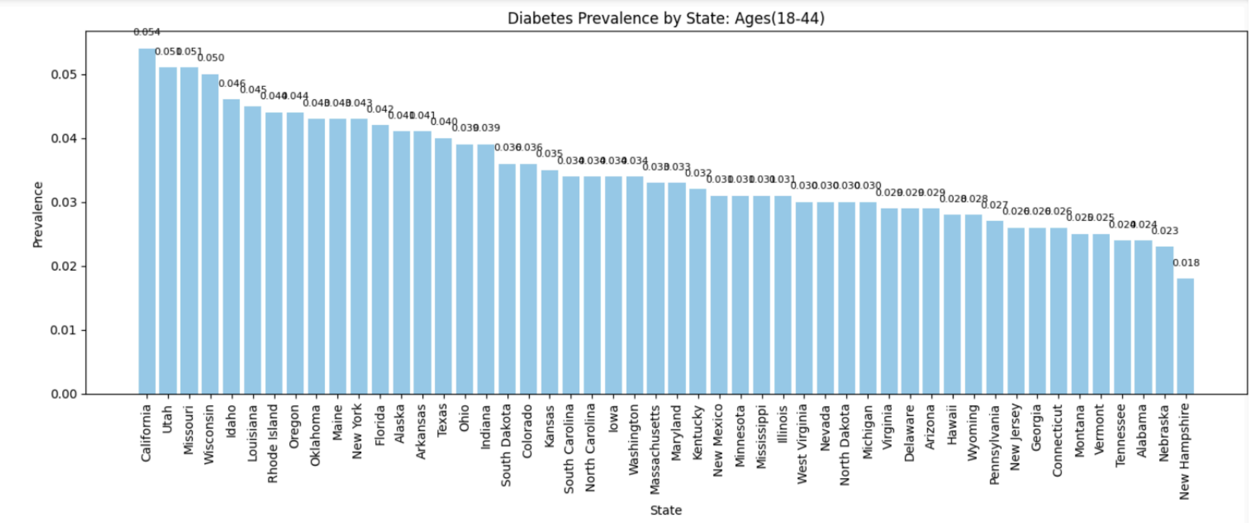
This analysis laid the foundation for more comprehensive exploration of diabetes across the United States of America with future efforts aimed at scaling the project both in scope and depth. Expanding the dataset towards a larger variety of unexplored variables like medical and biological factors consisting of family history, genetics, diet, physical activity levels, pet ownership rates, and other technology-related variables can yield more adequate models that uncover meaningful information about diabetes risk and prevalence. Magnifying the variables' scope from state-level to more local instances with increased precision can improve findings and offer better strategies to target diabetes and public health interventions.

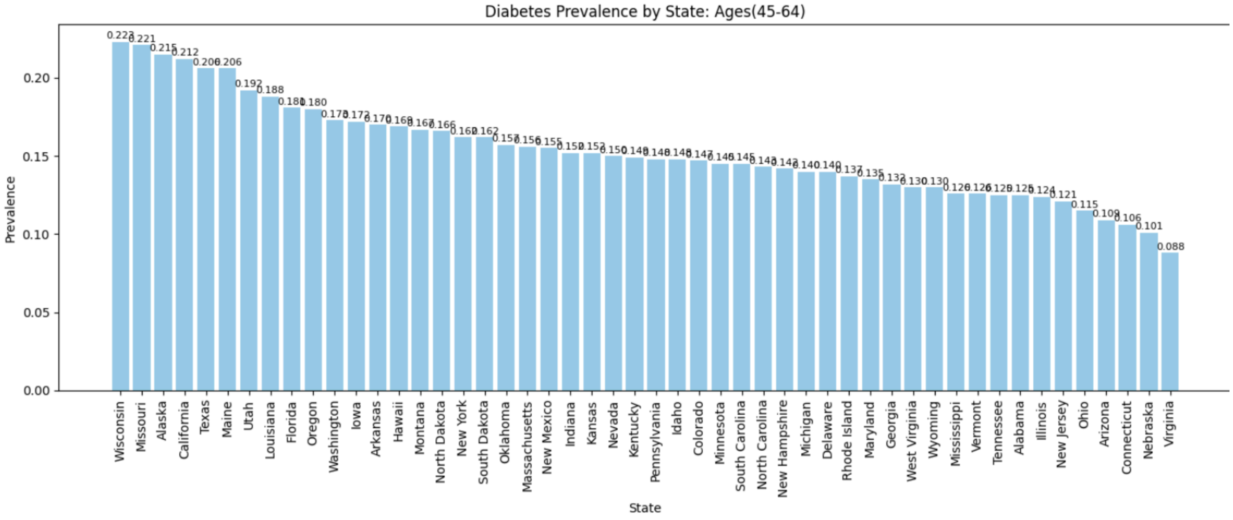
Conclusion

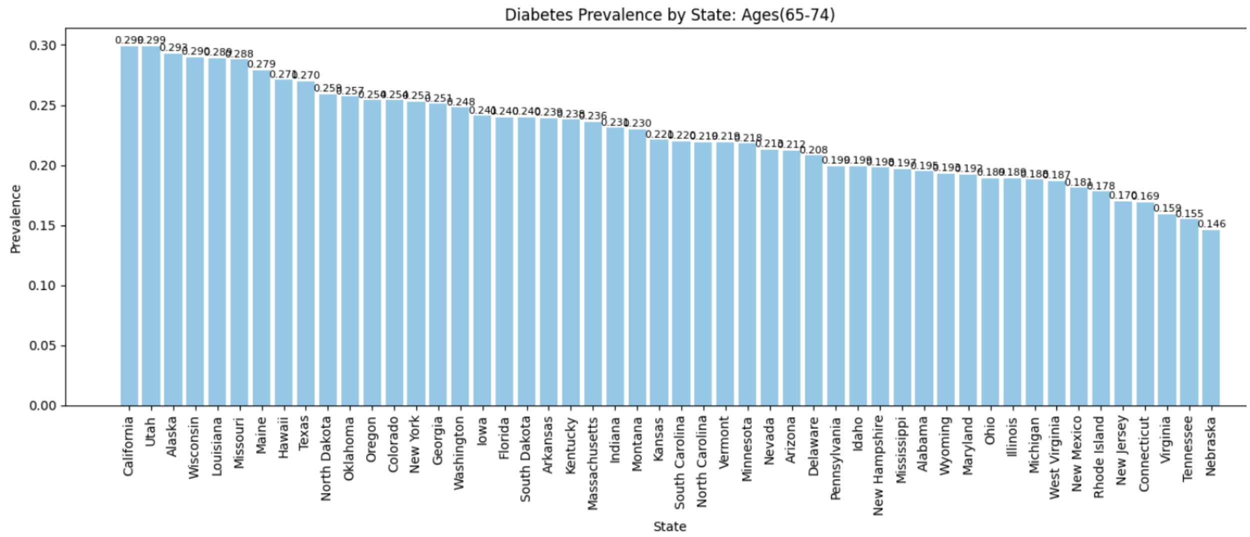
This study concludes that diabetes prevalence in the United States needs further exploration of more understudied variables to identify relations. The factors of mental health, socioeconomic employment status, internet access, and environmental contributors have some critical role and effect on diabetes prevalence. Despite its severity we have showcased how small shifts in prevalence can affect hundreds of thousands of Americans, making positive impacts in diabetic health and providing new information for public health planning for diabetes treatment and costs.

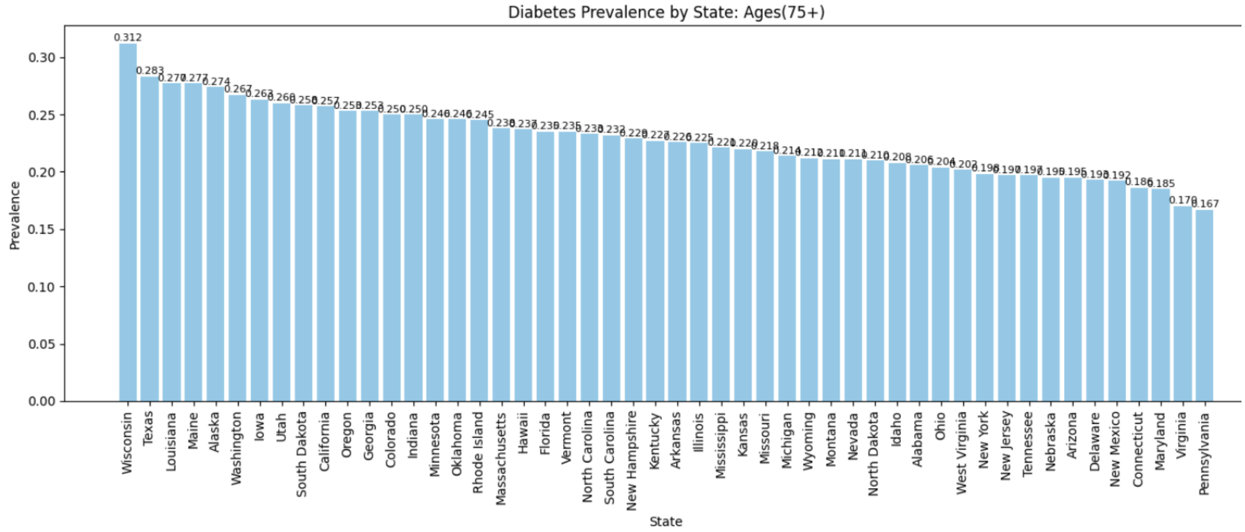
Appendix A

Bar plot visualizations of state-level diabetes prevalence, broken down by four age group clusters, revealed a subset of six states with the most significant highest and lowest prevalence rates for further exploration: California, New Hampshire, Wisconsin, Virginia, Nebraska, and Pennsylvania. The addition of West Virginia and Utah were added to the state subset since they had the most significant prevalence generalized across all states and age groups.



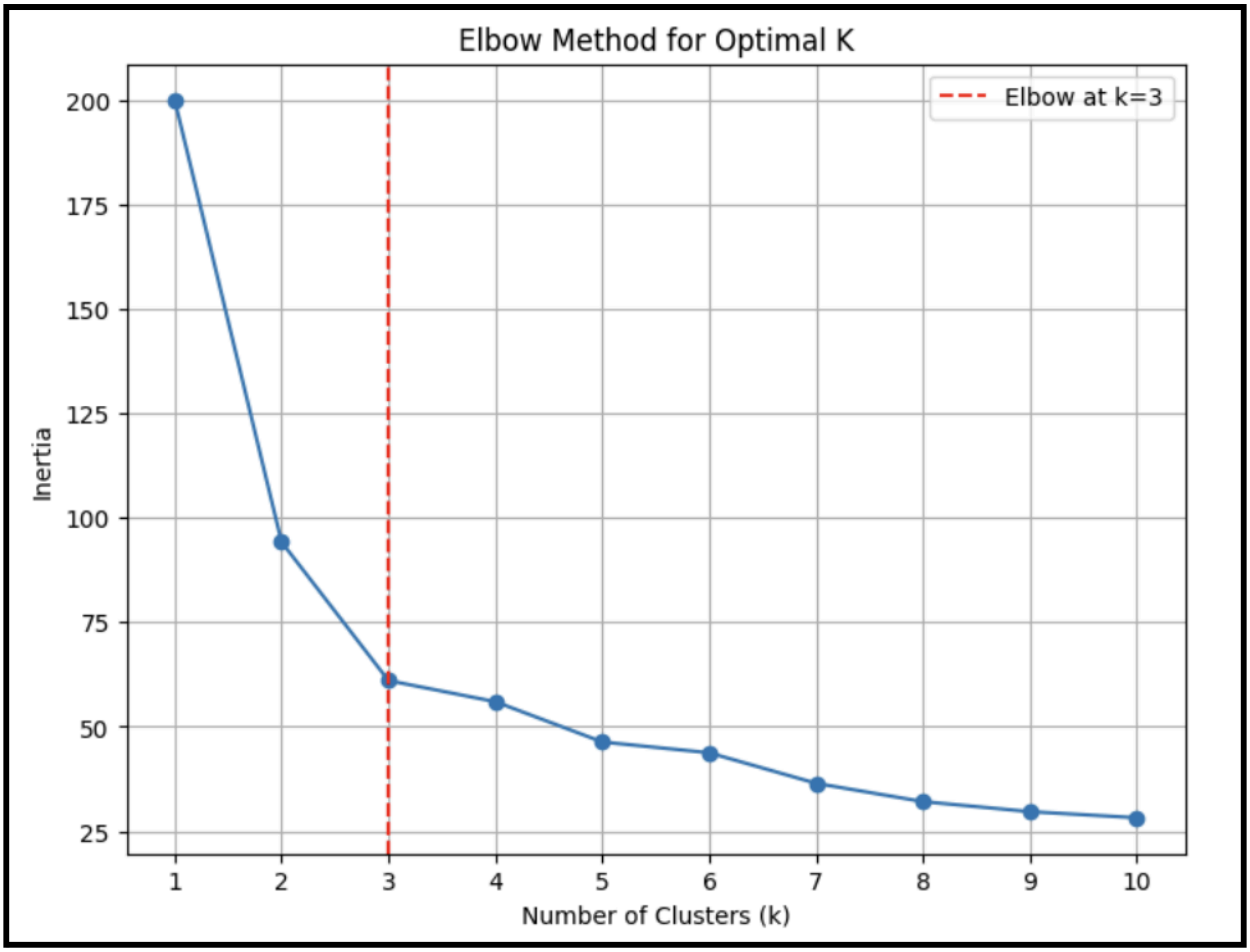



Diabetes Prevalence by State: Ages(65-74)
Prevalence
0.00
0.05
0.10
0.15
0.20
0.25
0.30
State
0.299
0.299
0.293
0.290
0.289
0.288
0.279
0.271
0.270
0.259
0.257
0.254
0.254
0.253
0.251
0.248
0.241
0.240
0.240
0.239
0.238
0.236
0.231
0.230
0.221
0.220
0.219
0.219
0.218
0.213
0.212
0.208
0.199
0.199
0.198
0.197
0.195
0.193
0.192
0.189
0.189
0.188
0.187
0.181
0.178
0.170
0.169
0.159
0.155
0.146
California
Utah
Alaska
Wisconsin
Louisiana
Missouri
Maine
Hawaii
Texas
North Dakota
Oklahoma
Oregon
Colorado
New York
Georgia
Washington
Iowa
Florida
South Dakota
Arkansas
Kentucky
Massachusetts
Indiana
Montana
Kansas
South Carolina
North Carolina
Vermont
Minnesota
Nevada
Arizona
Delaware
Pennsylvania
Idaho
New Hampshire
Mississippi
Alabama
Wyoming
Maryland
Ohio
Illinois
Michigan
West Virginia
New Mexico
Rhode Island
New Jersey
Connecticut
Virginia
Tennessee
Nebraska


Diabetes Prevalence by State: Ages(75+)
Prevalence
0.00
0.05
0.10
0.15
0.20
0.25
0.30
State
0.312
0.283
0.278
0.277
0.274
0.267
0.263
0.260
0.258
0.257
0.253
0.253
0.250
0.250
0.246
0.246
0.245
0.238
0.237
0.235
0.235
0.233
0.232
0.229
0.227
0.226
0.225
0.221
0.220
0.218
0.214
0.212
0.211
0.211
0.210
0.208
0.206
0.204
0.202
0.198
0.197
0.197
0.195
0.195
0.193
0.192
0.186
0.185
0.170
0.167
Wisconsin
Texas
Louisiana
Maine
Alaska
Washington
Iowa
Utah
South Dakota
California
Oregon
Georgia
Colorado
Indiana
Minnesota
Oklahoma
Rhode Island
Massachusetts
Hawaii
Florida
Vermont
North Carolina
South Carolina
New Hampshire
Kentucky
Arkansas
Illinois
Mississippi
Kansas
Missouri
Michigan
Wyoming
Montana
Nevada
North Dakota
Idaho
Alabama
Ohio
West Virginia
New York
New Jersey
Tennessee
Nebraska
Arizona
Delaware
New Mexico
Connecticut
Maryland
Virginia
Pennsylvania

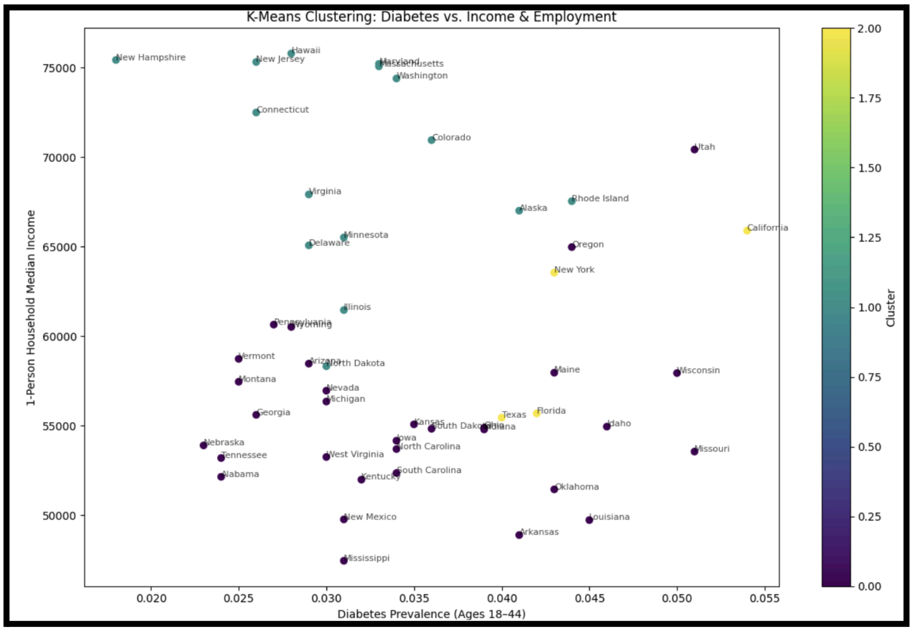
Appendix B

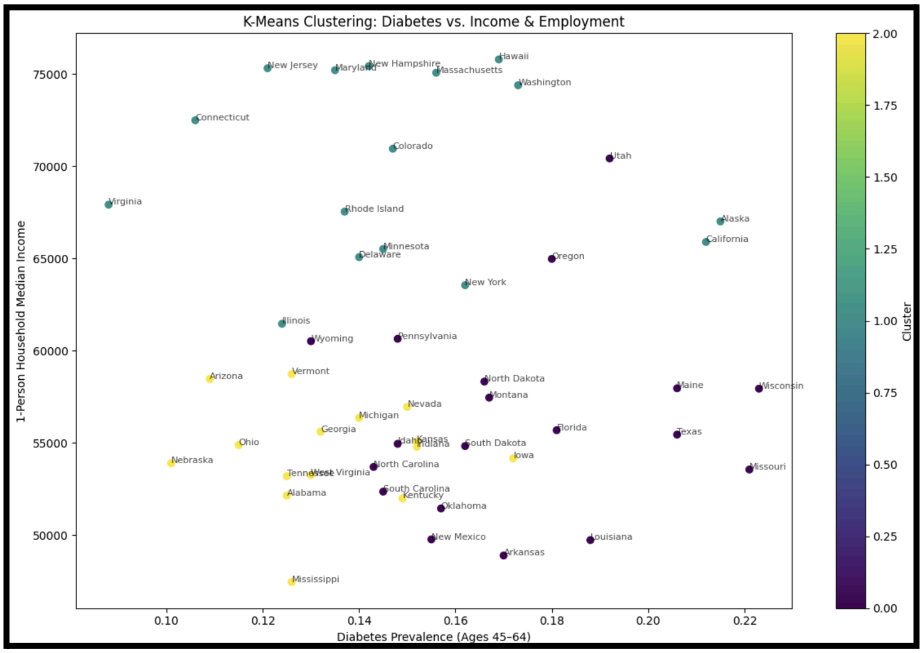
Showcasing the optimal K value for K-means Clustering model based upon socioeconomic and employment factors across all states. The elbow plot reveals that K=3 is the optimal number of clusters for the model.

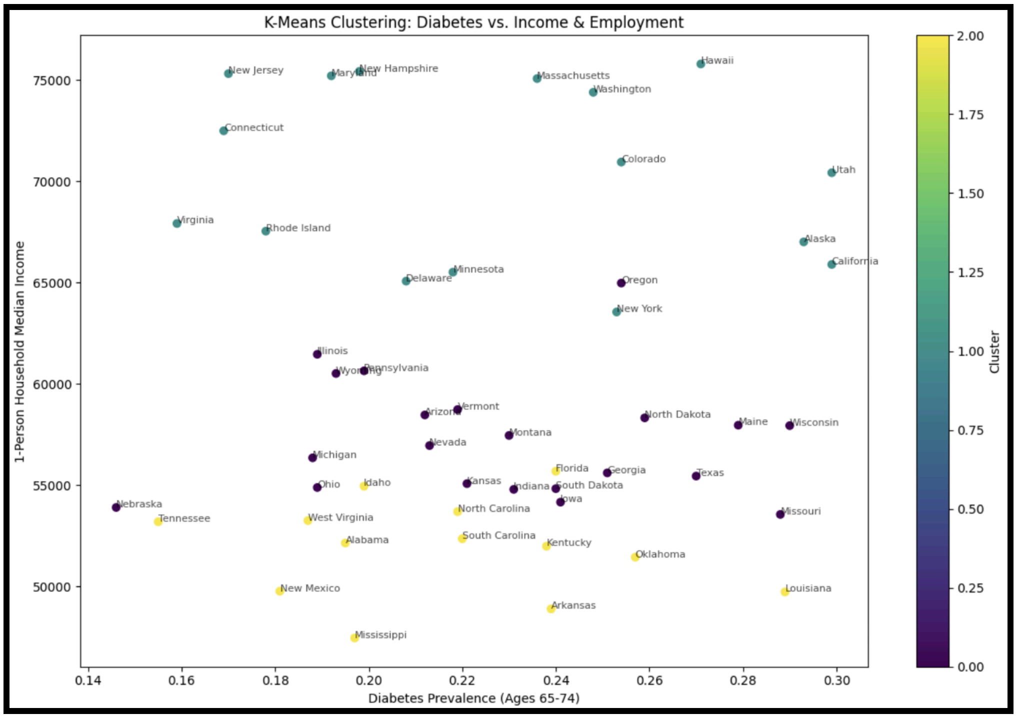


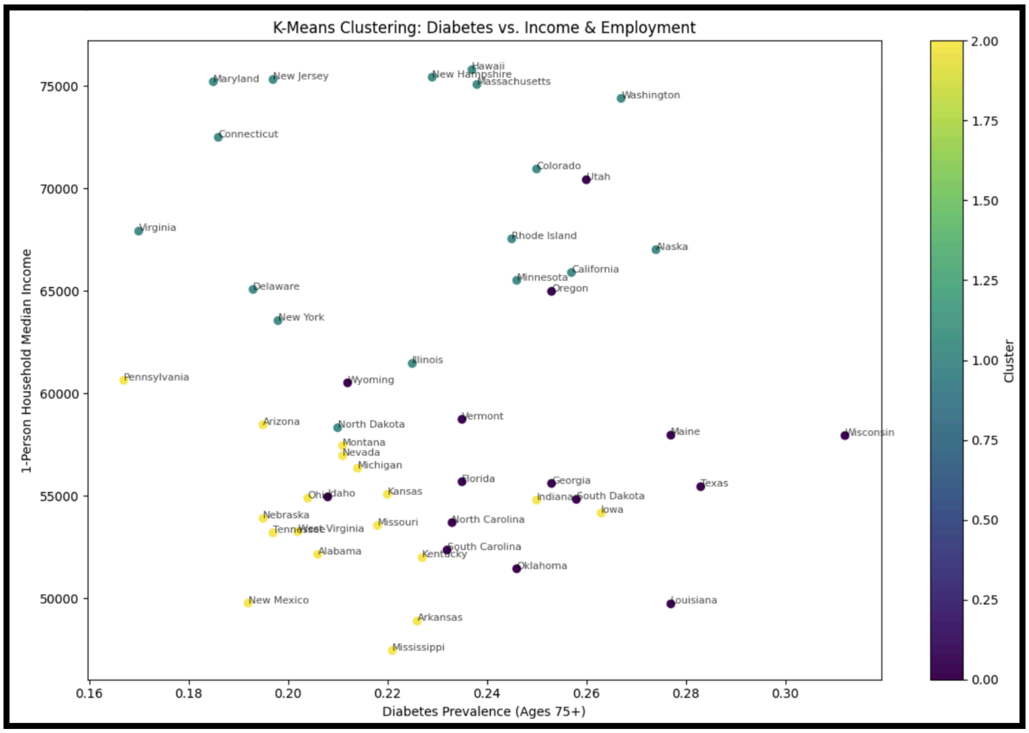
Appendix C

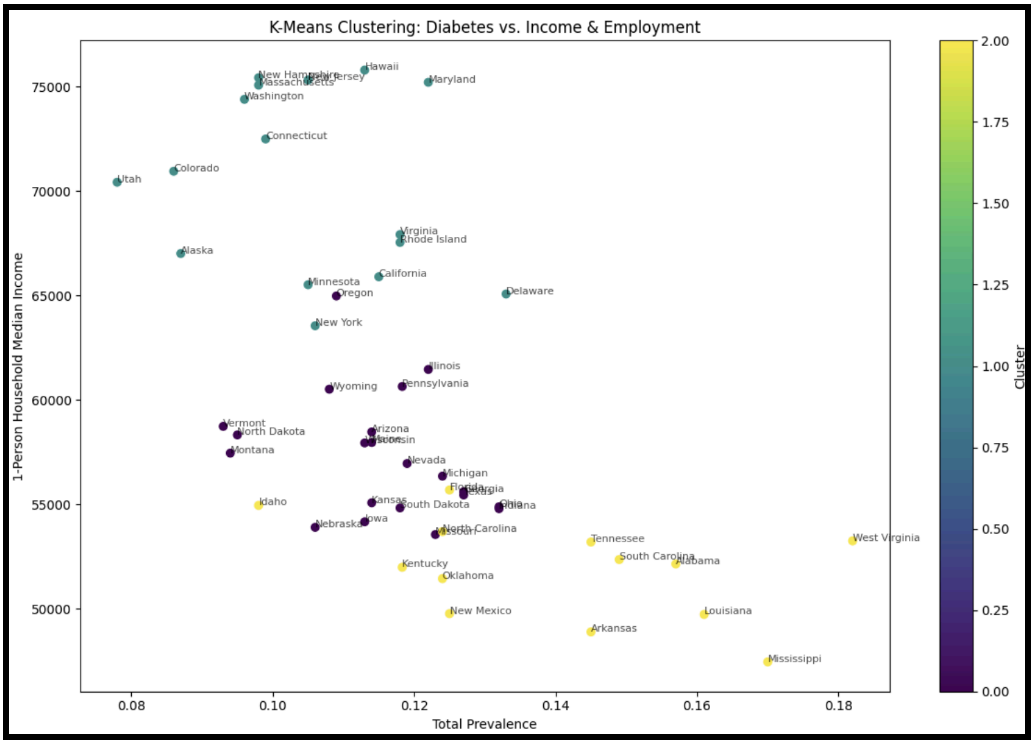
These visualizations showcase the K-Means clusters of states grouped by ages and their diabetes prevalence by income-related factors. The graphs below are ordered chronologically by increasing age groups and the last group is total prevalence including all ages.






K-Means Clustering: Diabetes vs. Income & Employment
1-Person Household Median Income
Diabetes Prevalence (Ages 65-74)
Cluster
75000
70000
65000
60000
55000
50000
0.14
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
New Jersey
Maryland
New Hampshire
Massachusetts
Hawaii
Washington
Connecticut
Colorado
Utah
Virginia
Rhode Island
Alaska
California
Delaware
Minnesota
Oregon
New York
Illinois
Wyoming
Pennsylvania
Arizona
Vermont
North Dakota
Maine
Wisconsin
Montana
Nevada
Michigan
Florida
Georgia
Texas
Ohio
Idaho
Kansas
Indiana
South Dakota
Iowa
Nebraska
Tennessee
West Virginia
North Carolina
Missouri
Alabama
South Carolina
Kentucky
Oklahoma
New Mexico
Louisiana
Arkansas
Mississippi


K-Means Clustering: Diabetes vs. Income & Employment
1-Person Household Median Income
Diabetes Prevalence (Ages 75+)
Cluster
75000
70000
65000
60000
55000
50000
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Maryland
New Jersey
Hawaii
New Hampshire
Massachusetts
Washington
Connecticut
Colorado
Utah
Virginia
Rhode Island
Alaska
California
Minnesota
Oregon
Delaware
New York
Illinois
Pennsylvania
Wyoming
Vermont
Arizona
North Dakota
Maine
Wisconsin
Montana
Nevada
Michigan
Florida
Georgia
Texas
Ohio
Idaho
Kansas
Indiana
South Dakota
Iowa
Nebraska
North Carolina
Tennessee
West Virginia
Missouri
South Carolina
Alabama
Kentucky
Oklahoma
New Mexico
Louisiana
Arkansas
Mississippi


K-Means Clustering: Diabetes vs. Income & Employment
1-Person Household Median Income
Total Prevalence
Cluster
75000
70000
65000
60000
55000
50000
0.08
0.10
0.12
0.14
0.16
0.18
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Hawaii
New Hampshire
New Jersey
Massachusetts
Maryland
Washington
Connecticut
Colorado
Utah
Virginia
Rhode Island
Alaska
California
Minnesota
Oregon
Delaware
New York
Illinois
Wyoming
Pennsylvania
Vermont
North Dakota
Arizona
Maine
Wisconsin
Montana
Nevada
Michigan
Florida
Georgia
Texas
Idaho
Kansas
South Dakota
Ohio
Indiana
Iowa
Nebraska
North Carolina
Missouri
Tennessee
West Virginia
South Carolina
Alabama
Kentucky
Oklahoma
New Mexico
Louisiana
Arkansas
Mississippi

References

American Diabetes Association. (2023, November 2). Statistics about diabetes. https://diabetes.org/about-diabetes/statistics/about-diabetes

American Diabetes Association. (2024, September). The staggering costs of diabetes [Infographic]. https://diabetes.org/sites/default/files/2024-09/ADA_2024_StaggeringCostsOfDiabetes.pdf

Centers for Disease Control and Prevention. (2024, May 15). National Diabetes Statistics Report. U.S. Department of Health & Human Services. https://www.cdc.gov/diabetes/php/data-research/index.html

Cleveland Clinic. (2023, February 17). Diabetes. https://my.clevelandclinic.org/health/diseases/7104-diabetes

National Institute of Health of Diabetes and Digestive and Kidney Diseases. (n.d.). What is diabetes? U.S. Department of Health & Human Services. https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes

Parker ED, Lin J, Mahoney T, Ume N, Yang G, Gabbay RA, ElSayed NA, Bannuru RR. Economic Costs of Diabetes in the U.S. in 2022. Diabetes Care. 2024 Jan 1;47(1):26-43. doi: 10.2337/dci23-0085. PMID: 37909353.