Project Report

Identify the questions being answered, and includes answering the starting questions in the scenario.

The aim of this project was to learn more about the quality of life in the United States. The two main questions we needed to answer were what are the factors that impact the quality of life and could the quality of life factors vary from area to area. My main focus in this project was to analyze the given datasets to see if those would give me my answers. The behavioral dataset provided questions people were asked from surveys as well as the result to those answers. Some of those questions include "Mean days of activity limitation" or "Mean mentally unhealthy days". All of the questions are within the topic of activity limitation, mental health, general health, and physical health which are all aspects of quality of life. This dataset also provides age ranges, race, and gender to see the differences in those fields. It also provides other factors such as state and year. Another dataset we used was the census data set. This provided all the different zip codes within the states, gender, population, min and max ages, city, as well as longitude and latitude. The first dataset provided only had zip codes and very few information that I could tie into my main census table which is why I used the other zip codes dataset from Kaggle with the states included. Lastly, after doing some research I found my last dataset which is state annual personal income from a government site. The dataset contained personal income per capita, states, population, and years dating back all the way to the early 1900s. This dataset was useful to see the differences in income especially in the early 2000s to see what other outside factors contribute to quality of life.

I mostly focused only two of those questions to answer the main question of what factors affect quality of life. The two I chose was the mean mentally unhealthy days and mean physically unhealthy days. I chose these because mental and physical health are key indicators to quality of life. My hypothesis was if there are more mentally/ physically unhealthy days the quality of life would not be sufficient. This could be analyzed by looking at the data value and if the number is higher that means the mentally/physically unhealthy days are more within 30 days which means people had more unhealthy days. Another component to answering the question was does the quality of life vary from area to area and it does. I was able to answer this question when doing my pivot tables and creating my graphs to see the top 10 and bottom 10 states and their mean average days mentally and physically. I was able to see more Midwest states having better mental and physical health rather than further south states. I also wanted to see if other factors like environmental factors affected quality of life. I was able to see that it does with my income dataset because income from 2008 to 2009 changed drastically and this made me realize that this could be because of the great recession in 2008. I also wanted to see if income was related to quality of life, and I compared my top 10 states from my income table with my risk factors table with the same states to see if higher income would result to better mental and physical health. I saw a slight correlation between California having worst unhealthy days and making less in my income graph. There is a limitation to my analysis because those two tables were separate and the years may vary since my income in 2000, 2007, 2008, and 2010 while my other table goes back to 1990s.

Introduce the datasets and cite them accordingly.

Behavioral risk factor HRQOL - dataset by CDC. data.world. (2017, February 2). Retrieved November 28th, 2022, from https://data.world/cdc/behavioral-risk-factor-hrqol

This dataset was taken from the CDC and measures well-being from different states of the United States.

State personal income: Revised estimates for 2010. BEA. (n.d.). Retrieved December 2nd, 2022, from https://apps.bea.gov/regional/histdata/releases/0911spi/index.cfm

This dataset measures the state income per capita in the United States from the early 1920s to 2010.

US Census Bureau . (n.d). US Population by Zipcode, Version 2. Retrieved November 28th, 2022 from https://www.kaggle.com/datasets/census/us-population-by-zip-code

This dataset contains Census data from 2000 and 2010.

Describe what was discovered in the research.

There was a lot that was discovered in the research. There are many factors that could directly affect quality of life. Factors such as mental health and physical health are important factors in quality of life but are far from being the only important ones. I also discovered and predicted that race would also play a major role in quality of life especially with the United States history. In one of my graphs, I was able to see the average mean days with breakout and filtered it out to be race/ethnicity. Native Americans reported to have the worst of mean unhealthy days compared to other races and this makes a lot of sense. When looking at outside factors this finding was not surprising. Another factor I discovered was age. When doing the same graph as the one for race but filter out ages instead and percentage instead of mean days you can see 75 + have a higher percentage of unhealthy days physically.

Support your descriptions with visualizations (Half should be created in Excel; Half in Power BI).

- **Two PowerBI and 5 Excel**
-

Explain what your visualizations represent and why they are relevant to the proposed questions.

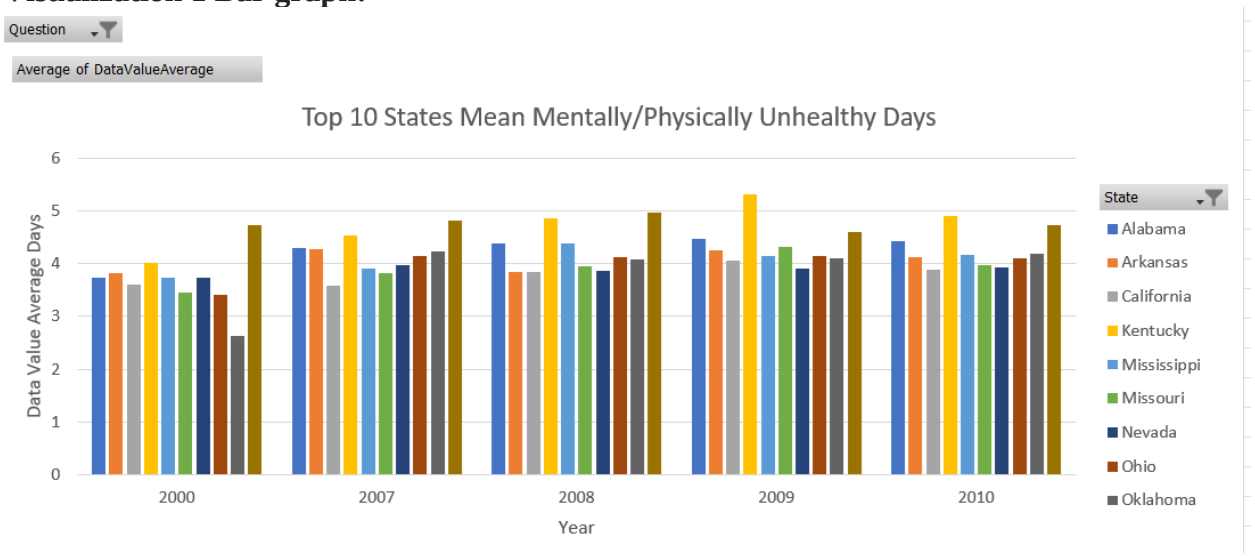
PowerBI:

Visualization 1: I was able to see the average mean days with breakout and filtered it out to be race/ethnicity. Native Americans reported to have the worst of mean unhealthy days compared to other races and this makes a lot of sense. When looking at outside factors this finding was not surprising. This is relevant to the proposed question because it provides other factors that contribute to quality of life in the United States.

Visualization 2: I did data value percentage and breakout and filtered it out by age. When analyzed it can be observed that 75 + have a higher percentage of unhealthy days physically. This makes sense because as we age our physical abilities deteriorate over the years and this answered the question because age is also another factor to quality of life.

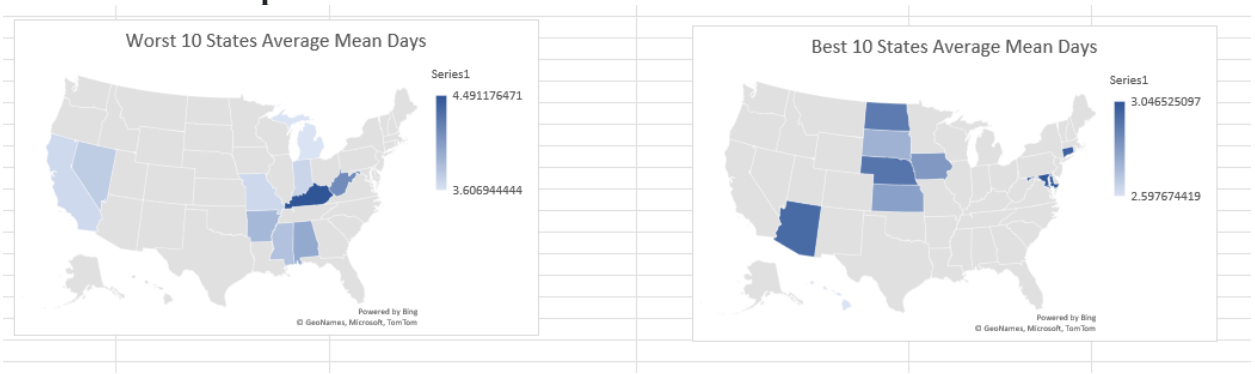
Excel:

Visualization 1 Bar graph:



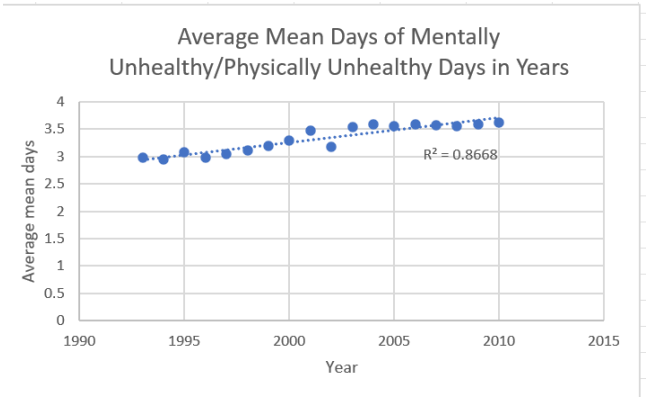
I looked at average mean unhealthy days physically and mentally in 2000, 2007, 2008, 2009, and 2010. I then looked at top 10 mean unhealthy days for those two questions to see which states were worse in these categories and how they changed over these years. This graph helped answer both the questions of what factors and do they vary from area to area and they do.

Visualization 2 Maps:



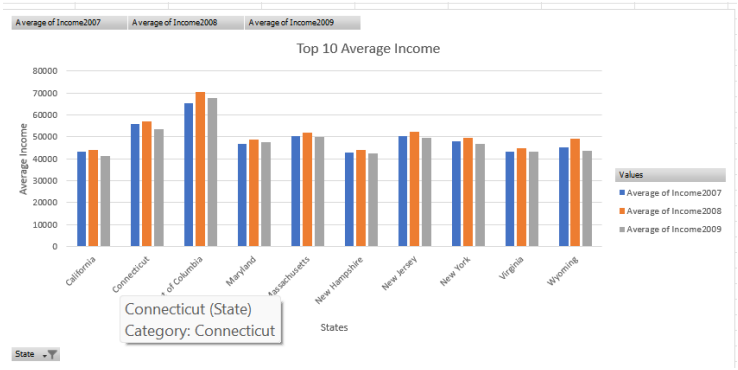
I was able to answer this question when doing my pivot tables and creating my graphs to see the top 10 and bottom 10 states and their mean average days mentally and physically. I was able to see more Midwest states having better mental and physical health rather than further south states.

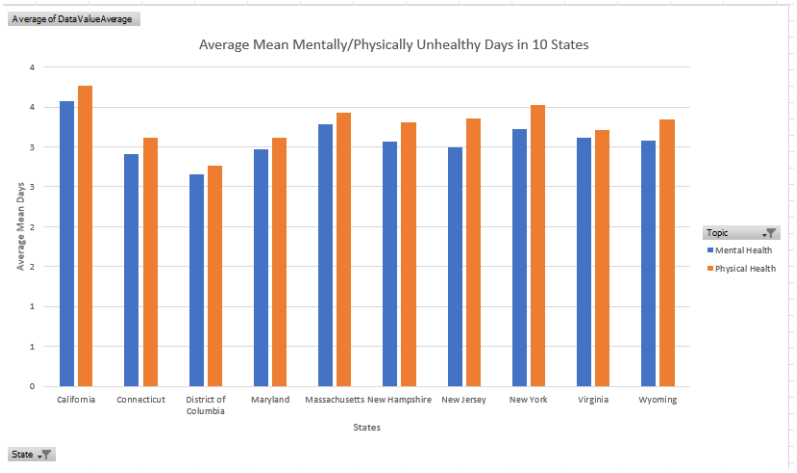
Visualization 3 Scatter Plot:



In this graph I wanted to see if there was a correlation between the mean unhealthy days mentally and physically in years. This graph answers the question because it shows a positive correlation and that could be related to environmental factors. There have been major changes in technology and major life events such as 9/11 and 2008 recession which can explain why average mean days gets worse throughout the years.

Visualization 4 & 5 Bar graphs:





For my first income graph I took average incomes in 2007, 2008, 2009 in the top 10 states**.** I also wanted to see if income was related to quality of life, and I compared my top 10 states from my income table with my risk factors table with the same states to see if higher income would result to better mental and physical health which was my second bar graph. I saw a slight correlation between California having worst unhealthy days and making less in my income graph. There is a limitation to my analysis because those two tables were separate and the years may vary since my income in 2000, 2007, 2008, and 2010 while my other table goes back to 1990s.

Include a correlation matrix or some other table that shows the Analysis Toolpak from Excel.

- **I made a descriptive statistic with my two data value columns**

Address any decisions that can be made from this data.

After analyzing the datasets, there can be some decisions that can be made. One of the more important decisions that can be made is improvements towards mental health and in certain areas and certain groups of people as well. Mental health has only continued to decline well over the years analyzed in these datasets. Major events and other environmental factors affect mental health. Another could be advances toward physical health because that too affects the quality of life someone has. Although these are only some suggestions there could be all types of decisions made with more research and data to make more educated and impactful developments.