**Mapping the Workforce: Exploring US Census Bureau Data**

Name: Vrund Patel

FA-550 (Data Visualization Application)

**Research Questions:**

1. How does unemployment vary across different states and in the United States, and are there any seasonal or cyclical trends?
2. What is the current unemployment rate in the United States, and how does it compare to historical data?
3. How does the unemployment rate vary across different states in the United States?
4. Are there any seasonal or cyclical patterns in unemployment rates, such as higher rates during certain months or specific economic cycles?
5. What are the major industries in terms of employment in the United States, and how has the employment distribution across industries changed over time? Are there any notable differences in employment patterns between different states?
6. What are the top industries in terms of employment in the United States, and what percentage of the workforce do they represent?
7. How has the employment distribution across industries changed over time period?
8. Are there any industries that have experienced significant employment variations?
9. Are there any notable differences in employment patterns between different states? Which industries are more prominent in certain areas compared to others?
10. What are the most common occupations in the United States, how has the distribution of occupations evolved over the years, and are there any notable gender or racial disparities within specific occupations?
11. What are the top occupations in terms of employment in the United States, and what percentage of the workforce do they represent?
12. How has the distribution of occupations changed over specific time period?
13. Are there any notable gender or racial disparities within specific occupations in terms of representation, pay, or advancement opportunities?
14. How has employment in specific industries, such as healthcare, technology, or manufacturing, changed over time?
15. How has employment in the healthcare, technology, and manufacturing industries changed over the past decade or specific time period of interest?
16. What is the average wage or salary for different occupations within these industries, and how does it compare to the overall average wage in the United States?
17. Are there any significant differences in wages or salaries between states within these industries?

**Tool Choice:**

For this project, the chosen visualization tools are R for data cleaning and Tableau for visualization.

R is a widely-used programming language for statistical computing and data analysis. It offers a comprehensive set of libraries and packages specifically designed for data cleaning, manipulation, and transformation. By utilizing R for data cleaning, you can ensure the accuracy and quality of the dataset before proceeding with visualization. Additionally, R provides a robust environment for conducting advanced statistical analyses if needed for further insights.

Tableau is a powerful and user-friendly data visualization tool that allows for the creation of interactive and visually appealing visualizations. Its intuitive drag-and-drop interface enables the creation of varied graph types, making it suitable for visualizing different aspects of the employment data in this project. Tableau also offers features for handling time series data, which will be beneficial for analyzing trends and patterns over time. Furthermore, Tableau provides options for interactive dashboards and storytelling, which can help present the findings in a compelling and informative manner.

Overall, the combination of R for data cleaning and Tableau for visualization provides a robust and efficient workflow. R's data cleaning capabilities ensure the dataset's integrity, while Tableau's visualization features enable the creation of dynamic and informative visualizations. This combination allows for a comprehensive analysis of the employment data and effective communication of the findings.

Some Charts that might be used to support visualization (subject to change):

1. Line charts: Line charts are useful for visualizing trends over time. I might use line charts to show the changes in the unemployment rate over the years or to track the employment distribution across different industries over time.
2. Bar charts: Bar charts are effective for comparing different categories or groups. I might use bar charts to compare the unemployment rates across different states or metropolitan areas, showcasing any regional variations.
3. Stacked area charts: Stacked area charts are helpful for showcasing the composition of a whole over time. It can be used stacked area charts to illustrate the distribution of employment across different industries over the years, highlighting the relative contributions of each industry.
4. Choropleth maps: Choropleth maps are ideal for displaying geographic variations.
5. Scatter plots: Scatter plots are effective for examining relationships or correlations between variables. Scatter plots can be used to explore the relationship between educational attainment and earnings, or to analyze the relationship between employment rates and factors like population density or educational attainment at the state or metropolitan level.
6. Heatmaps: Heatmaps are useful for visualizing dense data or patterns. Heatmaps might be used to display the average wages or salaries for different occupations or industries, highlighting any variations across categories.
7. Pie charts: Pie charts can be useful for visualizing proportions or percentages within a whole. Pie charts can be applied to show things like distribution of Employment by Industry.

**Timeline:**   
  
11/15: Data Collection and Dataset Identification

Identify and gather relevant datasets related to employment, occupational trends, industry-specific analysis, and wage/salary information.

Verify the reliability and quality of the selected datasets.

11/22: Data Cleaning and Preparation

Utilize R language for data cleaning and preprocessing tasks.

Remove any inconsistencies, outliers, or missing values in the datasets.

Prepare the datasets for further analysis and visualization.

11/29 and 12/6: Exploratory Data Analysis

Conduct exploratory data analysis to gain insights into the datasets.

Perform descriptive statistics, identify patterns, and visualize preliminary findings.

Evaluate the suitability of different graph types for visualization.

Develop line charts, bar charts, stacked area charts, choropleth maps, scatter plots, heatmaps, geographical charts, and pie charts as appropriate.

Utilize Tableau for creating visualizations-based stories on the research questions.

12/13-12/14: Visualization Development Review

Review and refine the code for data processing, cleaning, and visualization.

Ensure that everything is well-documented, optimized, and follows best practices.

12/15: Final Data Visualization Stories Submission

Review and polish the visualizations and narratives.

Create a comprehensive and cohesive set of data visualization stories that address the research questions.

Conduct a thorough recheck on the code, ensuring its correctness and effectiveness.

Submit the finalized code and data visualization stories for evaluation.

**Data:**

For my research project, I have chosen to utilize datasets from the US Census Bureau ([Census Bureau Tables](https://data.census.gov/table)) as my primary data source. The US Census Bureau provides a wealth of data related to employment, occupational trends, industry-specific analysis, and wage/salary information.

The data on the US Census Bureau site is distributed across various filters and components, allowing for flexibility in extracting relevant information for visualization and analysis.

This process may involves several key steps:

Identify Relevant Variables: I will carefully examine the available variables within the datasets to determine which ones align with my research questions and objectives. These variables may include employment levels, occupational categories, industry classifications, geographic locations, and time periods.

Filter and Subset Data: Based on the identified variables, I will apply filters to extract specific subsets of data that are most relevant to my research. For example, I may filter the data by industry sector, occupation category, or geographic region to focus on specific areas of interest.

Transform and Aggregate Data: In some cases, it may be necessary to transform or aggregate the data to derive meaningful insights. This could involve calculating proportions, percentages, or averages, or grouping data into meaningful categories. These transformations will enable me to create visualizations that effectively represent the patterns and trends present in the data.

Organize Data for Visualization: Once the relevant components have been extracted and transformed, I will organize the data in a structured format that is suitable for visualization. This may involve creating datasets or tables that contain the necessary variables and associated data points for each visualization.