**Design Document: US Census Data Analysis Project**

**1. Project Overview**

**Project Name:**

* US Census Data Analysis: Male-to-Female Population Ratio and Trends (2013-2022)

**Objective:**

* The project aims to retrieve US Census data for the past 10 years (2013-2022), calculate and analyze the male-to-female population ratio, and visualize these trends in a meaningful way. The goal is to display both the male-to-female ratio as a percentage and the overall male and female population in millions on a graph.

**2. System Architecture**

**2.1 High-Level Architecture:**

1. **Data Retrieval**:
   * Census data is pulled from the US Census Bureau’s API.
2. **Data Processing**:
   * Male and female population values are extracted from the response.
   * The male-to-female ratio is calculated for each year.
3. **Data Visualization**:
   * A line graph is generated that displays:
     + Left y-axis: male-to-female percentage.
     + Right y-axis: male and female population in millions.
4. **Graph Customization**:
   * Y-axis labels, legends, and visual markers are customized to show trends and offer better readability.

**2.2 Components:**

* **API**: US Census Bureau API.
* **Programming Language**: Python.
* **Libraries**:
  + **Requests**: To make API calls.
  + **Pandas**: To structure and process the retrieved data.
  + **Matplotlib**: To visualize the data through graphs.

**3. Functional Requirements**

**3.1 Data Retrieval:**

* Retrieve male and female population data from the US Census Bureau API for each year between 2013 and 2022.
* Handle cases where data might be missing or incomplete.

**3.2 Data Processing:**

* Extract the population values from the JSON response.
* Calculate the male-to-female ratio using the formula: Male-to-Female Ratio=Male PopulationFemale Population\text{Male-to-Female Ratio} = \frac{\text{Male Population}}{\text{Female Population}}Male-to-Female Ratio=Female PopulationMale Population​
* Convert the male and female population values to millions for ease of representation.

**3.3 Data Visualization:**

* Plot the male-to-female ratio as a percentage on the left y-axis.
* Plot the male and female population values in millions on the right y-axis.
* Customize the x-axis to show the years (2013-2022).
* Include an indicator at the 100% ratio point to show equal male-to-female representation.
* Legends should be placed outside the graph to avoid clutter.

**4. Non-Functional Requirements**

**4.1 Performance:**

* The system should fetch and process data within seconds.
* The visualization should render efficiently even when working with large datasets.

**4.2 Scalability:**

* The system should be designed to allow easy extension to include other variables from the US Census API or to visualize data for individual states or other demographic factors.

**4.3 Usability:**

* The script should be user-friendly and require minimal setup.
* The final graph should be clear and easy to interpret, with labeled axes, clear legends, and a balanced layout.

**5. Technical Specifications**

**5.1 Data Sources:**

* **US Census Bureau API**:
  + **Endpoint**:
    - https://api.census.gov/data/{year}/acs/acs5
  + **Key Data Fields**:
    - B01001\_002E: Male population.
    - B01001\_026E: Female population.

**5.2 Tools and Technologies:**

* **Programming Language**: Python 3.x.
* **Packages**:
  + requests: For API calls.
  + pandas: For data manipulation.
  + matplotlib: For data visualization.

**5.3 Code Structure:**

1. **Data Retrieval**:
   * Fetch Census data using the API for each year in the range 2013-2022.
   * Extract male and female population data.
2. **Data Processing**:
   * Check if the data is valid (numeric) and calculate the male-to-female ratio.
   * Convert population numbers to millions for easier visualization.
3. **Visualization**:
   * Plot the male-to-female percentage on the left y-axis.
   * Plot the population values in millions on the right y-axis.
   * Add custom ticks, legends, and title.

**6. User Interface**

This project is a command-line interface (CLI) Python script that runs the following steps:

1. Pulls data from the US Census Bureau API.
2. Processes the data to compute the male-to-female ratios.
3. Displays a graph visualizing the trends with the following features:
   * Left y-axis for the male-to-female percentage.
   * Right y-axis for population (in millions).
   * Legends for both the ratio and the population data placed outside the graph.
   * A clear marker at the 100% male-to-female ratio for reference.

**7. Assumptions and Constraints**

* The US Census API will provide valid data for the years requested.
* There may be occasional missing or invalid data for some years, which is handled by assigning zero values.
* The project focuses only on national-level data ("for": "us:\*") and does not break down by state or other demographics.

**8. Testing Plan**

1. **Unit Testing**:
   * Ensure that the API call retrieves the correct data format.
   * Validate that the male and female population values are extracted properly.
2. **Integration Testing**:
   * Verify that the data processing works correctly when integrated with the API response.
   * Ensure that the ratio and population values are plotted correctly on the graph.
3. **Visual Testing**:
   * Check that the graph displays correctly with legends, axes labels, and visual markers as expected.

**9. Future Enhancements**

* Add more demographic breakdowns (by state, income range, or age group).
* Include other variables from the Census API (e.g., income data, race, and ethnicity).
* Integrate additional time periods and perform long-term trend analysis.
* Create an interactive dashboard for user input to analyze data for specific years or states.

**10. Conclusion**

* This project effectively demonstrates how to retrieve and process population data from the US Census Bureau API. It also highlights key gender-related population trends over a 10-year period using clear and customizable visualizations. The code is modular, allowing easy extension for additional data and further customization.