**PS 23 - Tableau Visualization Guide for Adult Census Dataset**

This document provides detailed, beginner-friendly instructions to perform data visualization tasks on the Adult Census dataset using Tableau. The guide assumes you have no prior experience with Tableau and covers everything from installation to creating the required visualizations. Each task (a to g) is explained step-by-step, with explanations of why each step is performed and how to interpret the results.

**Prerequisites**

* **Computer**: Windows or Mac with internet access.
* **Dataset**: The Adult Census dataset (adult.csv), which you’ve provided.
* **Tableau Public**: A free version of Tableau for creating visualizations. (Note: Tableau Public saves workbooks to the cloud, so ensure you’re comfortable with this. Alternatively, use Tableau Desktop if you have a license.)
* **Microsoft Word**: To save or edit this guide.

**Step 1: Install Tableau Public**

1. **Download Tableau Public**:
   * Visit Tableau Public’s website.
   * Click “Download Tableau Public” and select the version for your operating system (Windows or Mac).
   * Explanation: Tableau Public is free and suitable for beginners to create and share visualizations. It’s ideal for learning and performing the tasks in this guide.
2. **Install Tableau**:
   * Run the downloaded installer and follow the on-screen instructions (e.g., accept terms, choose installation location).
   * Explanation: This sets up Tableau on your computer, allowing you to open and analyze the dataset.
3. **Launch Tableau Public**:
   * Open Tableau Public from your applications or desktop.
   * You’ll see a welcome screen with options to connect to data sources.
   * Explanation: The welcome screen is your starting point for loading data and creating visualizations.

**Step 2: Load the Adult Census Dataset**

1. **Connect to the Dataset**:
   * On the Tableau welcome screen, under “Connect” (left pane), click “Text File”.
   * Browse to the location where you saved adult.csv and select it.
   * Click “Open”.
   * Explanation: Tableau supports CSV files, and this step loads the Adult Census dataset into Tableau for analysis. The dataset contains columns like age, education, workclass, native-country, sex, and Target (income class).
2. **Preview the Data**:
   * Tableau will display a data preview window showing the dataset’s columns and sample rows.
   * Ensure all columns (e.g., age, education, Target) are visible and correctly formatted (e.g., age as a number, education as a string).
   * If any column has incorrect data types (e.g., age as a string), click the data type icon above the column (e.g., “Abc” for string, “#” for number) and select the correct type (e.g., Number (Whole) for age).
   * Explanation: This step ensures Tableau interprets the data correctly, which is crucial for accurate visualizations.
3. **Go to Worksheet**:
   * Click the “Sheet 1” tab at the bottom of the screen to start creating visualizations.
   * Explanation: A worksheet is where you build charts and graphs by dragging and dropping fields.

**Step 3: Understanding Tableau’s Interface**

Before creating visualizations, let’s familiarize ourselves with Tableau’s interface:

* **Data Pane** (left):
  + Lists all columns (fields) from the dataset, divided into **Dimensions** (categorical fields like education, native-country) and **Measures** (numerical fields like age, capital-gain).
  + Explanation: Dimensions are used for grouping or filtering, while Measures are used for calculations or aggregations.
* **Shelves** (top and middle):
  + **Columns** and **Rows**: Drag fields here to define the axes of your chart.
  + **Marks**: Drag fields to control color, size, or labels.
  + **Filters**: Drag fields to filter data (e.g., show only Master’s degree holders).
  + Explanation: Shelves determine how data is visualized (e.g., as bars, lines, or maps).
* **Show Me** (top-right):
  + Suggests chart types based on selected fields.
  + Explanation: This is helpful for beginners to choose appropriate visualizations.
* **Worksheet Area** (center):
  + Where your chart appears as you build it.
  + Explanation: This is your canvas for visualizing data.

**Step 4: Perform Visualization Tasks**

Below are detailed instructions for each task (a to g). Each task includes steps to create the visualization, explanations of why each step is performed, and how to interpret the result. Save your work frequently by clicking **File > Save to Tableau Public** and giving your workbook a name (e.g., “Adult\_Census\_Visualizations”).

**a. Find and Plot Income Class of People Whose Education is Master’s and Doctorate**

**Goal**: Show the distribution of income classes (<=50K vs. >50K) for individuals with Master’s or Doctorate degrees.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button at the bottom (a tab with a “+” icon).
   * Name the worksheet “Income by Master’s & Doctorate” (double-click the tab to rename).
   * Explanation: Each visualization gets its own worksheet for clarity.
2. **Filter for Master’s and Doctorate**:
   * Drag the education field from the Data Pane to the **Filters** shelf.
   * In the filter dialog, uncheck all options except “Masters” and “Doctorate”.
   * Click “OK”.
   * Explanation: This filters the dataset to include only rows where education is Master’s or Doctorate, focusing our analysis on these groups.
3. **Build a Bar Chart**:
   * Drag Target (income class) to the **Columns** shelf.
   * Drag Number of Records from the Measures section to the **Rows** shelf.
   * Explanation: Target creates bars for <=50K and >50K. Number of Records counts the number of individuals in each income class. Tableau automatically aggregates it as a sum.
4. **Enhance the Visualization**:
   * In the **Marks** card, ensure the mark type is “Bar” (use the dropdown).
   * Drag Target to the **Color** mark to color-code the bars (e.g., blue for <=50K, orange for >50K).
   * Explanation: Colors make it easier to distinguish income classes.
5. **Add Labels**:
   * Drag Number of Records to the **Label** mark.
   * Explanation: This displays the count of individuals on each bar for clarity.
6. **Format the Chart**:
   * Click the “T” button in the toolbar to show labels.
   * Right-click the y-axis, select “Format”, and set the axis title to “Count of Individuals”.
   * Right-click the chart title (double-click “Sheet 1” at the top) and edit it to “Income Class for Master’s and Doctorate Holders”.
   * Explanation: Formatting improves readability and makes the chart professional.
7. **Interpret the Result**:
   * The bar chart shows two bars: one for <=50K and one for >50K, representing individuals with Master’s or Doctorate degrees.
   * Compare the heights to see which income class is more common. For example, if the >50K bar is taller, it suggests that people with these degrees are more likely to earn over $50K.
   * Explanation: This visualization answers the question by showing the income distribution for highly educated individuals.

**b. Find and Plot Income Class of People Who Have Private Jobs**

**Goal**: Show the income class distribution for individuals with workclass as “Private”.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button and name it “Income by Private Jobs”.
   * Explanation: A new worksheet keeps this visualization separate.
2. **Filter for Private Workclass**:
   * Drag workclass to the **Filters** shelf.
   * In the filter dialog, check only “Private” and click “OK”.
   * Explanation: This restricts the data to individuals in the private sector.
3. **Build a Bar Chart**:
   * Drag Target to the **Columns** shelf.
   * Drag Number of Records to the **Rows** shelf.
   * Explanation: This creates bars for each income class, with heights showing the count of individuals.
4. **Enhance the Visualization**:
   * Set the mark type to “Bar” in the **Marks** card.
   * Drag Target to the **Color** mark to differentiate income classes.
   * Drag Number of Records to the **Label** mark to show counts.
   * Explanation: These steps make the chart visually clear and informative.
5. **Format the Chart**:
   * Show labels using the “T” button.
   * Format the y-axis title to “Count of Individuals”.
   * Edit the chart title to “Income Class for Private Sector Workers”.
   * Explanation: Proper formatting ensures the chart is easy to understand.
6. **Interpret the Result**:
   * The chart shows the number of private sector workers in each income class.
   * A taller <=50K bar suggests most private sector workers earn less than $50K, while a significant >50K bar indicates higher earners exist.
   * Explanation: This answers the question by focusing on the private sector’s income distribution.

**c. Find and Plot Yearly Sales Comparison**

**Note**: The Adult Census dataset does not contain explicit “sales” or “year” columns. I’ll assume this task refers to comparing income classes (Target) across different workclass categories (e.g., Private, Self-emp-inc, Federal-gov) as a proxy for “sales” (interpreting higher income as related to sales-oriented roles or sectors). If you meant a different dataset or specific sales data, please clarify, and I can adjust the instructions.

**Goal**: Show income class distribution across workclasses to compare “sales” (income) performance.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button and name it “Income by Workclass”.
   * Explanation: This keeps the visualization organized.
2. **Build a Stacked Bar Chart**:
   * Drag workclass to the **Columns** shelf.
   * Drag Number of Records to the **Rows** shelf.
   * Drag Target to the **Color** mark.
   * Explanation: Each workclass gets a bar, with segments colored by income class (<=50K vs. >50K). The height represents the total count, and colors show the proportion of each income class.
3. **Enhance the Visualization**:
   * Set the mark type to “Bar”.
   * Drag Number of Records to the **Label** mark and choose “Show mark labels” to display counts.
   * Explanation: Labels show the exact counts for each segment, aiding interpretation.
4. **Format the Chart**:
   * Format the y-axis title to “Count of Individuals”.
   * Edit the chart title to “Income Class Comparison Across Workclasses”.
   * Rotate x-axis labels (right-click x-axis, select “Rotate Labels”) for readability if needed.
   * Explanation: Formatting ensures clarity, especially with multiple workclasses.
5. **Interpret the Result**:
   * Each bar represents a workclass (e.g., Private, Federal-gov).
   * The bar is split into two colors: one for <=50K and one for >50K.
   * Compare the proportions to see which workclasses have more high earners. For example, “Self-emp-inc” might have a larger >50K segment, suggesting better “sales” or income performance.
   * Explanation: This visualization serves as a proxy for comparing income (as “sales”) across work sectors.

**d. Find and Plot Country-Wise Statistics on Geospatial Graph**

**Goal**: Create a map showing the count of individuals by native-country and their income class.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button and name it “Country-Wise Statistics”.
   * Explanation: A new worksheet is used for the map.
2. **Set Up the Map**:
   * Drag native-country to the center of the worksheet (or to the **Detail** mark).
   * Tableau will recognize native-country as geographic data and suggest a map. If not, click “Show Me” and select the “Map” chart type.
   * Explanation: Tableau’s geocoding maps country names to geographic coordinates.
3. **Add Metrics**:
   * Drag Number of Records to the **Size** mark to control the size of map markers (larger markers for countries with more individuals).
   * Drag Target to the **Color** mark to color-code by income class.
   * Explanation: Size shows the volume of individuals, and color shows the income class distribution.
4. **Enhance the Visualization**:
   * In the **Marks** card, ensure the mark type is “Map” or “Circle”.
   * Drag Number of Records to the **Label** mark to show counts on the map.
   * Explanation: Labels provide exact counts, and circles make the map clear.
5. **Format the Map**:
   * Edit the chart title to “Country-Wise Income Class Distribution”.
   * Adjust the map style (click **Map > Map Layers** and choose a style like “Normal” or “Dark”) for better visibility.
   * Explanation: A clear map style ensures the data stands out.
6. **Handle Missing Countries**:
   * If some countries (e.g., “?” in the dataset) don’t map, filter them out:
     + Drag native-country to the **Filters** shelf.
     + Exclude “?” or any invalid entries.
   * Explanation: This ensures the map only shows valid geographic data.
7. **Interpret the Result**:
   * The map shows circles for each country, sized by the number of individuals and colored by income class.
   * A large circle for the United States with a mix of <=50K and >50K colors indicates many individuals with varied incomes.
   * Smaller circles for other countries (e.g., Mexico, India) show fewer individuals, with colors indicating their income distribution.
   * Explanation: This answers the question by visualizing country-wise statistics geographically.

**e. Plot Age-Wise Education vs Salary Statistics**

**Goal**: Show how education and income class vary across age groups.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button and name it “Age-Education vs Salary”.
   * Explanation: A new worksheet organizes this complex visualization.
2. **Build a Heatmap**:
   * Drag age to the **Columns** shelf.
   * Drag education to the **Rows** shelf.
   * Drag Number of Records to the **Color** mark.
   * Drag Target to the **Detail** mark to split the heatmap by income class.
   * Explanation: The heatmap shows a grid where each cell represents a combination of age and education, colored by the count of individuals. Target adds income class detail.
3. **Enhance the Visualization**:
   * Set the mark type to “Square” in the **Marks** card.
   * Drag Number of Records to the **Label** mark to show counts in each cell.
   * Explanation: Squares and labels make the heatmap easier to interpret.
4. **Group Ages (Optional)**:
   * If there are too many age values, create age bins:
     + Right-click age in the Data Pane, select “Create > Bins”.
     + Set the bin size to 10 (for age groups like 20-29, 30-39).
     + Drag the new age (bin) field to **Columns** instead of age.
   * Explanation: Binning reduces clutter and groups ages into meaningful ranges.
5. **Format the Chart**:
   * Format the color scale (click the **Color** mark, choose a diverging palette like Blue-Orange).
   * Edit the chart title to “Age-Wise Education vs Salary Statistics”.
   * Explanation: A clear color scale and title improve readability.
6. **Interpret the Result**:
   * The heatmap shows how education levels (e.g., HS-grad, Bachelors) and income classes distribute across ages.
   * Darker cells indicate more individuals. For example, a dark cell for age 30-39, Bachelors, >50K suggests many individuals in that age group with a Bachelor’s degree earn over $50K.
   * Explanation: This visualization reveals patterns in how education and income correlate with age.

**f. Plot Country-Wise Male-Female Ratio**

**Goal**: Show the ratio of males to females by native-country.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button and name it “Country Male-Female Ratio”.
   * Explanation: A new worksheet is used for this visualization.
2. **Build a Stacked Bar Chart**:
   * Drag native-country to the **Columns** shelf.
   * Drag Number of Records to the **Rows** shelf.
   * Drag sex to the **Color** mark.
   * Explanation: Each country gets a bar, split into segments for Male and Female, with heights showing counts.
3. **Calculate the Ratio (Optional)**:
   * To show the ratio explicitly, create a calculated field:
     + Click **Analysis > Create Calculated Field**.
     + Name it “Male-Female Ratio”.
     + Enter the formula: SUM(IF [sex] = "Male" THEN 1 ELSE 0 END) / SUM(IF [sex] = "Female" THEN 1 ELSE 0 END).
     + Click “OK”.
   * Drag the “Male-Female Ratio” field to the **Label** mark.
   * Explanation: This calculates the male-to-female ratio for each country and displays it on the chart.
4. **Enhance the Visualization**:
   * Set the mark type to “Bar”.
   * Drag Number of Records to the **Label** mark to show counts.
   * Explanation: Labels provide exact counts and ratios.
5. **Format the Chart**:
   * Format the y-axis title to “Count of Individuals”.
   * Edit the chart title to “Country-Wise Male-Female Ratio”.
   * Rotate x-axis labels if needed.
   * Explanation: Formatting ensures the chart is clear and professional.
6. **Interpret the Result**:
   * Each bar shows the number of males and females for a country.
   * The ratio (if calculated) indicates the male-to-female balance (e.g., 1.5 means 1.5 males per female).
   * A country with a larger blue segment (Male) than orange (Female) has more males.
   * Explanation: This answers the question by visualizing gender ratios across countries.

**g. Plot Income Class Based on Workclass (Government and Other)**

**Goal**: Show income class distribution for government workclasses (Federal-gov, State-gov, Local-gov) vs. other workclasses.

1. **Create a New Worksheet**:
   * Click the “New Worksheet” button and name it “Income by Workclass Type”.
   * Explanation: A new worksheet organizes this visualization.
2. **Create a Workclass Group**:
   * Right-click workclass in the Data Pane, select “Create > Group”.
   * Name the group “Workclass Type”.
   * Select “Federal-gov”, “State-gov”, and “Local-gov”, then click “Group” and name it “Government”.
   * Select all other valid workclasses (e.g., Private, Self-emp-inc, Self-emp-not-inc), click “Group”, and name it “Other”.
   * Exclude invalid entries like “?” or “Never-worked”.
   * Click “OK”.
   * Explanation: Grouping simplifies the analysis by categorizing workclasses into Government and Other.
3. **Build a Stacked Bar Chart**:
   * Drag the new “Workclass Type” field to the **Columns** shelf.
   * Drag Number of Records to the **Rows** shelf.
   * Drag Target to the **Color** mark.
   * Explanation: This creates bars for Government and Other, split by income class.
4. **Enhance the Visualization**:
   * Set the mark type to “Bar”.
   * Drag Number of Records to the **Label** mark to show counts.
   * Explanation: Labels make the counts clear.
5. **Format the Chart**:
   * Format the y-axis title to “Count of Individuals”.
   * Edit the chart title to “Income Class by Workclass (Government vs Other)”.
   * Explanation: Formatting improves clarity.
6. **Interpret the Result**:
   * The chart shows two bars: one for Government and one for Other workclasses.
   * Each bar is split by income class (<=50K vs. >50K).
   * Compare the proportions to see if Government workers have a higher share of >50K than Other workers.
   * Explanation: This answers the question by comparing income distributions between government and non-government sectors.

**Step 5: Create a Dashboard (Optional)**

To present all visualizations together:

1. **Create a Dashboard**:
   * Click the “New Dashboard” button at the bottom.
   * Name it “Adult Census Dashboard”.
   * Explanation: A dashboard combines multiple visualizations into one view.
2. **Add Worksheets**:
   * Drag each worksheet (a to g) from the left pane onto the dashboard canvas.
   * Arrange them (e.g., in a grid or tiled layout) by dragging their edges.
   * Explanation: This organizes the visualizations for easy comparison.
3. **Format the Dashboard**:
   * Add a title: Click **Dashboard > New Title** and enter “Adult Census Data Analysis”.
   * Adjust sizes and layouts for clarity.
   * Explanation: A well-formatted dashboard is professional and user-friendly.
4. **Save the Dashboard**:
   * Click **File > Save to Tableau Public** and name it “Adult\_Census\_Dashboard”.
   * Explanation: Saving to Tableau Public allows you to share or access your work later.

**Step 6: Save and Share Your Work**

1. **Save the Workbook**:
   * Click **File > Save to Tableau Public As**.
   * Enter a name (e.g., “Adult\_Census\_Visualizations”) and click “Save”.
   * Explanation: This saves your entire workbook, including all worksheets and dashboards, to Tableau Public’s cloud.
2. **Export for Sharing (Optional)**:
   * To share as a PDF: Click **File > Print to PDF**, select the dashboard or worksheets, and save.
   * To share online: Publish to Tableau Public by clicking **File > Publish to Tableau Public** and following the prompts.
   * Explanation: These options let you share your visualizations with others.
3. **Copy to Word**:
   * Copy the visualizations into a Word document:
     + For each worksheet, click **Worksheet > Copy > Image**, paste into Word, and add captions.
     + Alternatively, include this guide in Word and reference the Tableau Public link.
   * Explanation: This fulfills the requirement for a Word document.

**Troubleshooting Tips**

* **Data Not Showing Correctly**:
  + Check data types in the data preview (Step 2). Ensure age is a number, Target is a string, etc.
  + Remove invalid entries (e.g., “?”) using filters.
* **Map Not Displaying**:
  + Ensure native-country is recognized as a geographic field (look for a globe icon next to it). If not, right-click native-country, select **Geographic Role > Country/Region**.
* **Too Many Data Points**:
  + Use filters or bins (as in task e) to reduce clutter.
* **Tableau Public Login**:
  + You may need to create a free Tableau Public account to save or publish workbooks.

**Conclusion**

This guide walked you through installing Tableau Public, loading the Adult Census dataset, and creating seven visualizations to answer specific questions. Each task was broken down into detailed steps with explanations to help you understand Tableau’s interface and the logic behind each visualization. The resulting visualizations provide insights into income distributions, education, workclass, gender ratios, and geographic patterns in the dataset.

To use this in a Word document:

1. Copy the text from this document.
2. Paste it into Microsoft Word.
3. Format as needed (e.g., add headings, adjust fonts).
4. Include screenshots or images of your Tableau visualizations by copying them from Tableau (Worksheet > Copy > Image).

If you encounter issues or need clarification on any step, feel free to ask for further assistance. Happy visualizing!