**Web Analytics and Business Intelligence**

**Data Cleaning**

1. **Load Data Source**

**Steps:**

* Ensure the dataset is properly loaded into Tableau and visible in the Data Source tab.
* Open Tableau and select "Connect" from the start page.
* Choose the appropriate data source type (Excel, CSV, or database).
* Locate the data file or use valid credentials to connect to a database.

**2. Explore and Understand the Data**

Identify potential issues such as null values, inconsistent data formats, or incorrect classifications.

**Steps:**

* Examine data in the Data Source tab.
* Verify field types and correct mismatches:
* Branch ID → Integer
* Duration → Float/Decimal
* Time → Date/Time
* Worker ID → Integer

**3. Perform Basic Cleaning**

Resolve common data quality issues for better analysis.

**Steps:**

* Remove ‘o’ from Branch ID using a calculated field (Branch ID 2) and hide the original Branch ID field.
* Rename columns, e.g., renaming F1 to Record ID.
* Address null values:
* Replace nulls in numerical fields with “0” and in text fields with appropriate default text.
* Clean textual data using calculated fields and regex:
* Query 7: Extract alphabetic text into a new column.
* Query 9: Replace incorrect entries with the text "Will you recommend this factory to a friend or family member?"

**Visualizations and Insights**

**1. Bar Chart: Distribution of Responses**

To identify the most frequent responses for a specific survey question.

**Steps:**

* Drag Query 1.0 to Columns and Response 1.0 to Rows.
* Use COUNT for aggregation and sort bars in descending order.
* Add labels for clarity.

**2. Line Chart: Temporal Trends in Survey Duration**

To show how survey durations evolve over time, segmented by branch.

**Steps:**

* Drag Time to Columns (convert to continuous if needed) and Duration to Rows.
* Add Branch ID 2 to Color for branch-wise differentiation.
* Adjust the Time axis to display weekly granularity.

**3. Scatter Plot: Correlation Between Response Categories**

To highlight relationships between responses.

**Steps:**

* Place Response 2.0 on Columns and Response 3.0 on Rows.
* Add Worker ID to Color for heatmap-style visualization.
* Adjust mark type to Square and customize the color gradient.
* Add labels for Record ID and Worker ID to show counts clearly.

**4. Histogram: Distribution of Survey Durations**

To examine the frequency of various survey durations to identify outliers or common ranges.

**Steps:**

* Drag Duration to Columns and select the histogram option from the Show Me menu.
* Adjust bin sizes for meaningful granularity.

**5. Box-and-Whisker Plot: Variability in Responses Across Branches**

Visualize the spread of survey durations across branches, detect outliers, and analyze response variability.

**Steps:**

* Drag Branch ID 2 and Response 5.0 to Rows, and Duration to Columns.
* Use the Analytics Pane to add a Box Plot with interquartile range (IQR).
* Customize with colors and labels for better clarity.

**Dashboard: Survey Insights**

This dashboard consolidates the analyses to offer a complete view of survey results. It combines bar charts, line charts, scatter plots, and box-and-whisker plots to reveal trends, correlations, and branch-level insights. Decision-makers can interact with filters and visual elements to pinpoint operational inefficiencies and improve survey strategies.









