Sevenette data analytics Project 2 questionnaire

These questions are for all groups



Data Cleaning and Quality

1. Data Quality and Cleaning:
   * Identify and handle any inconsistencies or anomalies in the dataset.
   * Ensure that all columns are in the correct data format (e.g., date columns as datetime objects).
   * Visualization: Create a heatmap to identify any missing values or anomalies in the dataset.
2. Data Integrity:
   * Verify that all trips with the same start and stop locations have reasonable distances.
   * Ensure that the start and end times of trips do not overlap.
   * Visualization: Use a scatter plot to visualize the relationship between start and stop locations and the distances traveled.

Data Processing

1. Missing Data Imputation:
   * Identify trips with missing mileage and estimate an average mileage for these trips based on the available data.
   * Propose a method to handle the missing mileage for the trip on 1/2/2016 1:25 to 1/2/2016 1:37.
   * Visualization: Create a histogram to visualize the distribution of trip distances, including imputed values.
2. Date and Time:
   * Extract the month and day from the START\_DATE and END\_DATE columns.
   * Create a new column that indicates whether the trip started and ended on the same day.
   * Visualization: Plot the frequency of trips by month and day.

Data Manipulation

1. Distance Calculations:
   * Create a new column categorizing trips as either short (<= 5 miles), medium (> 5 and <= 15 miles), or long (> 15 miles).
   * Visualization: Use a bar chart to show the distribution of trips by distance category.

Descriptive Statistics

1. Basic Summary:
   * How many trips were made in total?
   * What is the total number of miles traveled?
   * Visualization: Create a summary table and a pie chart to show the proportion of total miles traveled for different purposes.
2. Trip Durations:
   * Calculate the average duration of a trip in minutes.
   * Identify the longest and shortest trip durations.
   * Visualization: Use a box plot to visualize the distribution of trip durations.
3. Purpose Analysis:
   * What is the most common purpose for the trips?
   * Calculate the total distance traveled for each purpose category.
   * Visualization: Create a bar chart showing the total distance traveled for each purpose.

Advanced Analysis

1. Location Insights:
   * Identify the most frequently visited start and stop locations.
   * Calculate the average distance traveled between major locations (e.g., Fort Pierce to West Palm Beach).
   * Visualization: Use a heatmap to show the frequency of trips between start and stop locations.
2. Customer Visits:
   * Calculate the average distance and duration for customer visits.
   * Compare these statistics with those of other trip purposes.
   * Visualization: Create a comparative bar chart to visualize the average distance and duration for customer visits versus other purposes.
3. Time of Day Analysis:
   * Analyze the distribution of trips across different times of the day (e.g., morning, afternoon, evening, night).
   * Identify any patterns or trends in trip purposes based on the time of day.
   * Visualization: Use a line chart to show the distribution of trips throughout the day and a stacked bar chart to show trip purposes by time of day.

Predictive Analysis

1. Predictive Modeling:
   * Develop a simple predictive model to estimate the distance of a trip based on the start and stop locations, and trip duration.
   * Suggest possible features that could be used to improve the model's accuracy.
   * Visualization: Visualize the model's predictions versus actual distances using a scatter plot with a line of best fit.

Visualization

1. Trend Analysis:
   * Plot the number of trips taken each day.
   * Visualize the distribution of trip distances.
   * Visualization: Use a time series plot to show the number of trips taken each day and a histogram for trip distances.
2. Purpose Distribution:
   * Create a pie chart or bar graph showing the distribution of trip purposes.
   * Visualize the relationship between trip purposes and distances traveled.
   * Visualization: Use a pie chart for the distribution of trip purposes and a scatter plot to show the relationship between trip purposes and distances.

Top of Form

Bottom of Form