Contents

[Requirement 1: Total trip count per month. 1](#_Toc201739303)

[Requirement 2: Total passenger count per month. 2](#_Toc201739304)

[Requirement 3: Average daily income. 2](#_Toc201739305)

[Requirement 4: Tips distribution by passenger count. 3](#_Toc201739306)

[Requirement 5: Average fare per mile. 4](#_Toc201739307)

As a seasoned data engineer, here's a step-by-step guide to achieving each of the data visualization requirements in Apache Superset, assuming you have the necessary data loaded and accessible in Superset as a dataset (e.g., a trips dataset containing trip information with columns like trip\_id, passenger\_count, fare\_amount, tip\_amount, trip\_distance, and a timestamp column like pickup\_datetime).

For each visualization, the general flow will be:

1. **Create/Select Dataset:** Ensure you have the relevant dataset ready.
2. **Create New Chart:** Start building a new chart.
3. **Configure Chart:** Select chart type, define metrics, dimensions, and other properties.
4. **Save Chart:** Give it a meaningful name.
5. **Add to Dashboard:** Place the chart on your desired dashboard.

Let's break down each requirement:

## Requirement 1: Total trip count per month.

**Goal:** Show the number of trips grouped by month.

**Step-by-step Guidelines:**

1. **Navigate to Charts:**
   * From the Superset home page, click the **"+"** sign in the top right corner and select **"Chart."**
2. **Choose Dataset:**
   * In the "Choose a dataset" dialog, select your relevant dataset (e.g., trips). Click **"Create new chart."**
3. **Configure Chart:**
   * **Visualization Type:** Select **"Bar Chart"** or **"Line Chart"** (Bar Chart is often clearer for counts per period).
   * **Time Column:**
     + Locate the time column (e.g., pickup\_datetime).
     + For "Time Grain," select **"month"**.
   * **Metrics:**
     + The default "COUNT(\*)" or "row\_count" metric should suffice. If not, add a new metric:
       - Click **"Add Metric."**
       - Set "Label" to Total Trips.
       - Set "Aggregate" to COUNT.
       - Set "Column" to trip\_id (or any non-null column that represents a single trip).
       - Click **"Save."**
   * **X-axis:** The "Time Column" configured with "month" time grain will automatically serve as your X-axis.
   * **Y-axis:** The "Total Trips" metric will be on the Y-axis.
   * **Run Query:** Click **"Run Query"** to see the chart preview.
4. **Save Chart:**
   * Click **"Save."**
   * **Chart Name:** Enter Total Trip Count per Month.
   * **Add to Dashboard:** Select an existing dashboard or choose **"Create new dashboard"** and provide a name (e.g., Trip Analysis Dashboard).
   * Click **"Save & Go to Dashboard"** (or "Save & Add to Dashboard" if creating new).

## Requirement 2: Total passenger count per month.

**Goal:** Show the sum of passengers across all trips, grouped by month.

**Step-by-step Guidelines:**

1. **Navigate to Charts:**
   * Click the **"+"** sign and select **"Chart."**
2. **Choose Dataset:**
   * Select the trips dataset. Click **"Create new chart."**
3. **Configure Chart:**
   * **Visualization Type:** Select **"Bar Chart"** or **"Line Chart."**
   * **Time Column:**
     + Locate the time column (e.g., pickup\_datetime).
     + For "Time Grain," select **"month"**.
   * **Metrics:**
     + Click **"Add Metric."**
     + Set "Label" to Total Passengers.
     + Set "Aggregate" to SUM.
     + Set "Column" to passenger\_count.
     + Click **"Save."**
   * **X-axis:** The "Time Column" (monthly grain) will be your X-axis.
   * **Y-axis:** The "Total Passengers" metric will be on the Y-axis.
   * **Run Query:** Click **"Run Query."**
4. **Save Chart:**
   * Click **"Save."**
   * **Chart Name:** Enter Total Passenger Count per Month.
   * **Add to Dashboard:** Select your Trip Analysis Dashboard (or the one you created earlier).
   * Click **"Save & Go to Dashboard"** (or "Save & Add to Dashboard").

## Requirement 3: Average daily income.

**Goal:** Calculate the average daily total income (e.g., fare\_amount + tip\_amount + other charges if applicable), aggregated daily.

**Step-by-step Guidelines:**

1. **Navigate to Charts:**
   * Click the **"+"** sign and select **"Chart."**
2. **Choose Dataset:**
   * Select the trips dataset. Click **"Create new chart."**
3. **Configure Chart:**
   * **Visualization Type:** Select **"Line Chart"** (good for trends over time) or **"Big Number"** (if you want just the grand average). For daily trend, Line Chart is better.
   * **Time Column:**
     + Locate the time column (e.g., pickup\_datetime).
     + For "Time Grain," select **"day"**.
   * **Metrics:**
     + First, you might need a calculated column for Total Income if it doesn't exist. Go to your dataset in "Datasets" tab, click "Edit", then "Calculated Columns" and add a column like fare\_amount + tip\_amount + tolls\_amount (adjust as per your dataset's income components). Let's assume you've named it total\_income.
     + Now back in the chart, add a new metric:
       - Click **"Add Metric."**
       - Set "Label" to Average Daily Income.
       - Set "Aggregate" to AVG.
       - Set "Column" to total\_income (or the expression fare\_amount + tip\_amount if not pre-calculated).
       - Click **"Save."**
   * **X-axis:** The "Time Column" (daily grain) will be your X-axis.
   * **Y-axis:** The "Average Daily Income" metric will be on the Y-axis.
   * **Run Query:** Click **"Run Query."**
4. **Save Chart:**
   * Click **"Save."**
   * **Chart Name:** Enter Average Daily Income.
   * **Add to Dashboard:** Select your Trip Analysis Dashboard.
   * Click **"Save & Go to Dashboard."**

## Requirement 4: Tips distribution by passenger count.

**Goal:** Understand how tip amounts vary based on the number of passengers in a trip.

**Step-by-step Guidelines:**

1. **Navigate to Charts:**
   * Click the **"+"** sign and select **"Chart."**
2. **Choose Dataset:**
   * Select the trips dataset. Click **"Create new chart."**
3. **Configure Chart:**
   * **Visualization Type:** A **"Bar Chart"** (showing average tips per passenger count) or a **"Box Plot"** (showing distribution) would be effective. Let's use Bar Chart for simplicity.
   * **Metrics:**
     + Click **"Add Metric."**
     + Set "Label" to Average Tip Amount.
     + Set "Aggregate" to AVG.
     + Set "Column" to tip\_amount.
     + Click **"Save."**
   * **Group By:**
     + For "Group By," select passenger\_count. This will group the average tip by each distinct passenger count.
   * **Sort By (Optional):** You might want to sort by passenger\_count or Average Tip Amount for better readability.
   * **Run Query:** Click **"Run Query."**
4. **Save Chart:**
   * Click **"Save."**
   * **Chart Name:** Enter Tips Distribution by Passenger Count.
   * **Add to Dashboard:** Select your Trip Analysis Dashboard.
   * Click **"Save & Go to Dashboard."**

## Requirement 5: Average fare per mile.

**Goal:** Calculate the average fare amount divided by the trip distance, to understand the fare efficiency.

**Step-by-step Guidelines:**

1. **Navigate to Charts:**
   * Click the **"+"** sign and select **"Chart."**
2. **Choose Dataset:**
   * Select the trips dataset. Click **"Create new chart."**
3. **Configure Chart:**
   * **Visualization Type:** A **"Big Number"** chart if you want a single average value across all trips, or a **"Bar Chart"** if you want to see this average grouped by another dimension (e.g., payment type, or even just the overall average over time with "Line Chart"). Let's go with "Big Number" for the overall average.
   * **Metrics:**
     + Click **"Add Metric."**
     + Set "Label" to Average Fare per Mile.
     + Set "Aggregate" to AVG.
     + Set "SQL Expression": fare\_amount / trip\_distance
       - **Important Note:** Handle potential trip\_distance being zero to avoid division by zero errors. A more robust expression would be CASE WHEN trip\_distance > 0 THEN fare\_amount / trip\_distance ELSE 0 END or AVG(fare\_amount) / AVG(trip\_distance) depending on your desired aggregation. For simplicity, we'll use fare\_amount / trip\_distance but be aware of data quality.
     + Click **"Save."**
   * **Time Column (Optional for Big Number):** For a single "Big Number," you might not need a time column. If you want to see the trend of "Average fare per mile" over time, you would select pickup\_datetime and a relevant "Time Grain" (e.g., day or month) and use a "Line Chart."
   * **Run Query:** Click **"Run Query."**
4. **Save Chart:**
   * Click **"Save."**
   * **Chart Name:** Enter Average Fare per Mile.
   * **Add to Dashboard:** Select your Trip Analysis Dashboard.
   * Click **"Save & Go to Dashboard."**

After creating all charts, you'll be directed to your Trip Analysis Dashboard. From there, you can **"Edit Dashboard"** to:

* Drag and drop charts to arrange them.
* Resize charts for optimal viewing.
* Add layout elements like rows, columns, and tabs for better organization.
* Set interactive filters if needed (e.g., filter by payment\_type or passenger\_count).
* Finally, **"Save"** your dashboard and **"Publish"** it for broader access.