**Descriptive Statistics Insights**

**Summarizing Core Trip Characteristics:**

* **Passenger Count:**
  + Average passengers per trip: ~49.
  + Range: 1 to 99, with a standard deviation of ±26.
  + *Insight:* The wide variability suggests diverse demand patterns. This presents opportunities for better load balancing and dynamic capacity adjustments.
* **Fare Amount:**
  + Average fare: £25.36.
  + Distribution is skewed towards the lower end.
  + *Insight:* The concentration of lower fares may reflect a prevalence of short-distance trips or the impact of subsidized pricing. This warrants a deeper look into fare structures.
* **Trip Duration:**
  + Mean: 95 minutes; Median: ~97.5 minutes.
  + Range: 5 to 179 minutes.
  + *Insight:* The broad range indicates varying trip complexities. Categorizing durations (short, medium, long) could aid in route planning and demand segmentation for improved service delivery.

**Understanding Relationships Between Key Metrics:**

* **Fare vs. Duration:** There is a **moderate positive correlation** between Fare\_Amount and Trip\_Duration\_Minutes.
  + *Insight:* This is an expected finding, suggesting that longer trips generally incur higher fares. This validates the existing fare model's foundational logic.
* **Fare vs. Passenger Count:** A **weak correlation** was observed between Fare\_Amount and Passenger\_Count.
  + *Insight:* This indicates that the current fare structure is primarily time-based or distance-based, rather than being significantly influenced by the number of passengers on a given trip. This suggests potential for dynamic pricing models based on demand.
* **Passenger Load vs. Time/Fare:** Passenger load weakly correlates with both time and fare.
  + *Insight:* This reinforces that current pricing is not heavily influenced by occupancy. For Ferry and Tram services, higher passenger counts sometimes correlate with lower per-passenger fares, suggesting existing bulk pricing or subsidies that could be leveraged or re-evaluated.
* **Passenger Count Distribution (Histogram/Density Plot):**
  + *Expected Outcome:* Likely shows peaks around certain passenger counts, possibly indicating common vehicle capacities or typical demand levels.
  + *Insight:* Helps MetroMove understand typical passenger loads. Identifying frequent low-occupancy trips (e.g., many trips with <20 passengers) highlights opportunities for cost savings through smaller vehicles or reduced frequency during specific times/routes. Conversely, consistent high-occupancy trips indicate routes needing increased capacity.
* **Fare Amount Distribution (Histogram/Box Plot):**
  + *Expected Outcome:* Confirms the skew towards lower fares. Outliers might be visible.
  + *Insight:* Visualizes the revenue landscape. The skew confirms the prevalence of lower-priced trips. Outliers (e.g., highest fare recorded at £49.86, mostly in Tram and Bus services) require investigation to understand if they represent premium services, special events, or data anomalies.
* **Trip Duration Distribution (Histogram/Box Plot):**
  + *Expected Outcome:* Shows the spread of trip times, potentially with multiple peaks for short, medium, and long trips.
  + *Insight:* Reveals common trip lengths and helps identify unusually long trips. Trips exceeding 120 minutes (22% of total) are significant and directly impact passenger satisfaction and operational efficiency, signaling potential bottlenecks or inefficiencies.

**Exploring Relationships Between Two Variables:**

* **Average Trip Duration by Transport Mode (Bar Plot):**
  + *Insight:* Trains show the longest average duration, aligning with their role in longer-distance commutes. Ferries have the shortest average duration, indicating efficient water-based transit for specific routes. Trams, despite sometimes having high individual fares, show short average durations, suggesting efficiency for quick city hops. Buses show wide occupancy variance and longer durations, likely due to urban congestion.
  + *Actionable:* Focus optimization efforts on high-duration modes (Trains, Buses) to improve overall network speed.
* **Stacked Average Fare by Route (Departure vs. Arrival Stations - Bar Plot):**
  + *Insight:* This visual highlights the most profitable routes and the fare contribution of different arrival stations for a given departure station.
  + *Actionable:* MetroMove can identify high-revenue corridors (e.g., Airport-Central Station) to prioritize service enhancements or marketing. Conversely, low-fare routes might indicate services needing re-evaluation or subsidy analysis.
* **Passenger Count vs. Day of Week:**
  + *Insight:* Weekends (especially Sunday) show traffic peaks, likely driven by leisure travel. Weekdays, particularly Wednesday, Thursday, and Friday, show a mid-week dip.
  + *Actionable:* This directly supports dynamic scheduling, allowing MetroMove to increase capacity during peak weekend leisure travel and potentially reduce trips during mid-week off-peak windows to optimize resource allocation.

**Uncovering Complex Patterns with Multiple Variables:**

* **Route-Specific Performance (e.g., Departure\_Station to Arrival\_Station with Passenger\_Count and Trip\_Duration\_Minutes):**
  + *Insight:* This analysis identifies specific routes that are either highly efficient/popular or significantly underperforming. For example, some long-duration bus trips with consistently low passenger counts (< 20 passengers) are likely unprofitable operations.
  + *Actionable:* Flag underperforming routes for re-evaluation (e.g., reduced frequency, rerouting, or even suspension). Investigate long-duration, low-fare anomalies for potential manual errors or niche services that need targeted adjustments.
* **Station-Level Traffic Patterns:**
  + *Insight:* Central Station processes 20.5% of total arrivals, indicating a significant bottleneck risk. The top 10 departure/arrival stations are critical hubs.
  + *Actionable:* Prioritize infrastructure investment and operational improvements at high-traffic stations. Implement strategies like priority boarding at Central Station and enhance staff deployment and security at these key hubs to manage flow and improve passenger experience.

**Creating New Perspectives from Existing Data:**

* **Temporal Features (Hour of Day, Day of Week, Month, Year):**
  + *Insight:* Extracting these features allowed for granular temporal analysis, revealing peak commuting hours (6:00-9:00 AM) and weekend leisure travel patterns. This transforms raw timestamps into actionable time-based demand signals.
* **Route Features (e.g., Route\_ID from Departure\_Station and Arrival\_Station):**
  + *Insight:* Combining departure and arrival stations into a unique route identifier enabled direct analysis of route performance, fare patterns, and efficiency, which was not possible with individual station data alone. This is crucial for identifying high-demand corridors like "Airport–Central."
* **Duration Categories (Short, Medium, Long Trips):**
  + *Insight:* Categorizing Trip\_Duration\_Minutes provides a simplified view for high-level planning and identifying problematic long trips.
* **Distribution of Key Variables:**
  + *Insight:* The continued skewness in distributions of new features (e.g., Trip\_Duration\_Hours) indicates persistent inefficiencies or unmet demand patterns that require strategic intervention.

**Recommendations**

**1. Resource Reallocation:**

* **Reduce low-utilization bus routes:** Specifically, those with consistently less than 15 passengers, especially during off-peak hours, to optimize operational costs.
* **Increase train capacity:** Focus on the 7:00–9:00 AM window to meet high morning commute demand.
* **Eliminate circular same-station trips:** These are likely operational dead-heads or anomalies; addressing them could save approximately 120 hours/week in operational time.

**2. Fare & Pricing:**

* **Introduce off-peak discounts:** Especially for trams, to stimulate ridership during quieter periods and maximize asset utilization.
* **Standardize fares:** Implement a cap on 20% deviation for routes of similar distance to ensure fairness and predictability for passengers.

**3. Route Optimization:**

* **Focus on high-demand corridors:** Prioritize improvements and capacity on routes like the "Airport–Central" corridor, which show high average fares and passenger counts.
* **Redeploy buses:** Shift resources from underused zones to peak corridors to better match supply with demand.

**Slide 14: Strategic Business Recommendations (2/2)**

**4. Service Quality:**

* **Tackle long trips:** Address the 22% of trips exceeding 120 minutes by implementing dedicated transit lanes for buses and improving ferry scheduling to reduce delays.

**5. Demand-Driven Scheduling:**

* **Utilize forecasting:** Implement a system for 7-day projections of passenger demand.
* **Dynamically adjust vehicle allocation:** Use demand forecasts to adapt vehicle deployment by time of day and day of week, ensuring optimal fleet utilization.

**6. Fare Restructuring:**

* **Distance-based pricing for trains:** Align fares more closely with the distance traveled, promoting fairness and potentially increasing revenue on longer routes.
* **Flat fares for short-duration bus rides:** Introduce flat fares for bus trips under 30 minutes to encourage short-distance travel and simplify pricing.

**7. Infrastructure Investments:**

* **Expand Central Station capacity:** Given its role as a major bottleneck (20.5% of total arrivals), invest in expanding its capacity to improve passenger flow and reduce congestion.
* **Add ferry docks at South Point:** This would reduce wait times and improve the efficiency and appeal of ferry services in that area.