

Project Task for for the Jersey City Citi Bike Dataset

Group: 11

GitHub Repository: <https://github.com/tsaichen10/jc-citibike-vis>

To connect our analysis to real-world problems, we have created five user stories from the perspective of different stakeholders. These stories illustrate the tasks of typical users for the Jersey City Citi Bike dataset. The overarching theme is **"Understanding and optimizing Citi Bike usage patterns and operational efficiency in Jersey City."**

1. Marketing Analysis

Persona: Sarah, a marketing analyst for Citi Bike.

Sarah is planning a new campaign to convert "casual" riders into "member" riders. To effectively target her campaign—for instance, deciding whether to offer discounts on weekends or during weekday commute times—she first needs to understand: **"How do member and casual users differ in their riding activity across hours of the day and days of the week?"**

Abstraction:

- **High level:** Discover (The user is asking "how" things differ, which is exploratory).
- **Mid/Search:** Lookup (She knows the two groups she wants to compare: 'member' and 'casual').
- **Low/Query:** Compare (The keyword "differ" maps directly to this action).
- **Target:** Distributions (by group).
- **Final:** Compare distributions.

2. Urban Planner

Persona: BardiC, an urban planner for the Jersey City Department of Transportation.

BardiC is tasked with identifying high-traffic corridors to prioritize the expansion of protected bike lanes. She needs to understand the city's primary arteries of bike traffic. Her core question is: **"Which station-to-station routes are the most heavily trafficked across the city, both overall and during peak commuter hours (7-9 AM & 5-7 PM)?"**

Abstraction:

- **High level:** Discover (Asking "which" are the "most" is an exploratory task).
- **Mid/Search:** Explore (She is exploring the entire set to find the top-ranked items).
- **Low/Query:** Summarize (Rank) (The goal is to aggregate all trips and rank the resulting routes by volume).

- **Target:** Paths (The routes between stations).
- **Final:** Rank paths.

3. Commuter Route Planning

Persona: Ms. Marvel, a Jersey City resident.

Ms. Marvel lives near Hamilton Park and works near the Exchange Place PATH station. She wants to start commuting by bike but is concerned about traffic and bike availability. She needs to explore her options by asking: **"What are the main commuter flows from the Hamilton Park area towards the waterfront business district during the morning rush hour (7-9 AM)? Which specific station-to-station routes are most popular, and are there any less crowded but still efficient alternatives?"**

Abstraction:

- **High level:** Discover (She is exploring unknown "patterns").
- **Mid/Search:** Explore (She doesn't know the specific routes and is exploring all possibilities within a corridor).
- **Low/Query:** Summarize (She wants to understand the "main" flows, which is a summarization task).
- **Target:** Network Paths (The trips between stations).
- **Final:** Summarize flows.

4. Fleet Maintenance

Persona: David, a fleet maintenance supervisor for Citi Bike.

David manages the maintenance schedules for different bike models. He knows that electric bikes (`rideable_type`) have different usage patterns and can be parked anywhere, leading to missing end station data. He needs to know: **"How does the usage of electric bikes differ from classic bikes in terms of average trip duration and common parking zones (using `end_lat` , `end_lng`), especially for trips that do not end at a designated station?"**

Abstraction: This is a compound task.

1. Task 1: Compare Trip Durations

- **High level:** Discover.
- **Mid/Search:** Lookup (Compare known groups: 'electric' vs. 'classic').
- **Low/Query:** Compare.
- **Target:** Distribution (The distribution of trip durations for each type).
- **Final:** Compare distributions.

2. Task 2: Identify & Compare Parking Zones

- **High level:** Discover.
- **Mid/Search:** Explore (Find unknown spatial patterns).
- **Low/Query:** Summarize (Cluster) (Group lat/lng points into dense zones).

- **Target:** Spatial Clusters / Hotspots.
- **Final:** Identify clusters and compare their locations/densities.

5. Operations Management

Persona: Alex, a station rebalancer for Citi Bike.

Alex's job is to move bikes between stations to ensure availability, optimizing his truck routes and schedules. His daily question is: **"Which stations are the most popular start and end points, and what is the net inflow/outflow for the top 20 busiest stations during morning (7-10 AM) and evening (5-7 PM) rush hours?"** This helps him predict when stations will likely become full or empty. (Uses `start_station_id` and `end_station_id`).

Abstraction: This is a compound task.

1. Task 1: Find Busiest Stations

- **High level:** Discover.
- **Mid/Search:** Explore (Find the top N items).
- **Low/Query:** Summarize (Rank) (Rank stations by total ride count).
- **Target:** Nodes (Stations).
- **Final:** Rank stations.

2. Task 2: Analyze Inflow/Outflow Trends for Top Stations

- **High level:** Discover.
- **Mid/Search:** Lookup (Look up known stations from Task 1).
- **Low/Query:** Identify (Identify the time-series pattern for each station).
- **Target:** Trend (The net bike flow over time).
- **Final:** Identify trend.