Analysis of Citi Bike Usage Patterns: Visual Analytics for Urban Mobility Optimization

Author: Harsha Saketh Konjeti
Github: https://github.com/hkonjeti3/Visualisation_project-hkonjeti

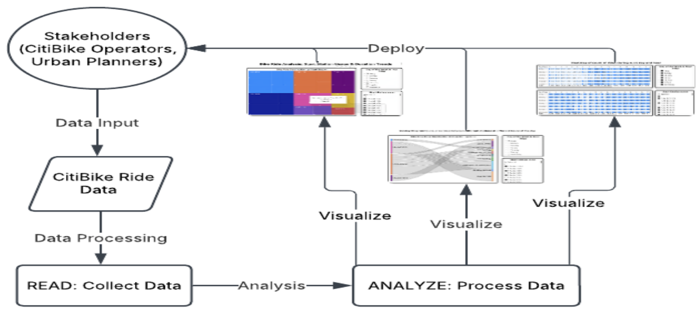

Fig. 1 Visual Abstract

1. INTRODUCTION AND PRIOR WORK

Bike-sharing systems provide a sustainable and cost-effective transit alternative, reducing congestion and promoting healthier lifestyles. However, efficient bike allocation remains a challenge, requiring insights into usage patterns and peak demand periods. While prior studies have explored predictive modeling for demand forecasting [1] and spatial analyses of station activity [2], they often overlook categorical and linguistic data that enhance user behavior analysis. This project integrates structured ride data with treemaps, Sankey diagrams, and heatmaps to optimize CitiBike operations, improve resource allocation, and provide actionable insights for CitiBike operators, urban planners, and commuters.

1.1 Stakeholder Groups

The project caters to three primary stakeholders. CitiBike operators require insights on high-demand stations, peak usage times, and ride duration trends for effective bike rebalancing. Urban planners need ride flow data to optimize bike lane placements and improve urban cycling infrastructure. Commuters and riders can benefit from understanding bike availability patterns across stations and times, allowing for better trip planning.

1.2 Stakeholder Needs

CitiBike operators need demand forecasting, planners require spatial flow insights, and riders seek peak availability trends. The project's visualizations address these needs by offering filtering, interactive analytics, and ride pattern analysis.

2. DATA ACQUISITION

2.1 Data Sources

The dataset was sourced from CitiBike's open data portal, containing structured records on ride timestamps, start/end stations, and trip duration. The primary dataset includes attributes such as ride timestamps (start and end times of trips), station names (categorical labels of bike stations), ride duration (trip time in minutes), and day of the week (extracted from timestamps).

2.2 Data Description, Quality and Coverage

The dataset consists of 1,000,000 ride records, capturing CitiBike usage trends across various stations and time periods. It includes ride timestamps, station names, geospatial coordinates, ride duration, and user types (member or casual rider). While the dataset is mostly complete, missing values exist for start and end station names (3,349 total) and latitude/longitude data (713 total), which may impact geospatial analysis. Data quality measures include timestamp standardization, categorical cleaning, and filtering out incomplete records to enhance accuracy. The dataset provides broad temporal and spatial coverage, allowing analysis of ride frequencies, peak usage hours, and station demand trends. Despite minor inconsistencies, the data remains highly valuable for optimizing bike redistribution, urban planning, and commuter decision-making.

3. DATA ANALYSIS

The dataset was processed using Python libraries such as Pandas and NumPy to clean and structure the data efficiently. Ride timestamps were converted into day-of-week and hour-of-day formats to enable time-based analysis, and missing values in station names and geospatial data were handled through filtering and interpolation. Data aggregation techniques were applied to compute key metrics such as total rides per station, average ride duration, and high-traffic routes. Statistical methods, including frequency distributions and correlation analysis, were used to identify patterns in ride demand and peak usage periods. Visualization software, primarily Power BI, was employed to create treemaps, Sankey diagrams, and heatmaps, providing an interactive and intuitive understanding of station utilization and ride flows. These methods facilitated accurate analysis, aligning with research objectives by offering insights for CitiBike operators, urban planners, and commuters to enhance bike allocation and urban mobility planning.

4. VISUALIZATIONS

This project employed multiple visualization techniques to represent CitiBike ride data effectively, improving stakeholder understanding and facilitating decision-making. The selected visualizations include treemaps, Sankey diagrams, and heatmaps, each serving distinct analytical purposes.

4.1 Treemap - Bike Ride Distribution by Start Station

The treemap visualizes ride distribution by start station, allowing CitiBike operators to identify high-demand locations. Larger blocks represent stations with higher ride counts, enabling effective bike redistribution planning. **Example Conclusion:** The treemap indicates that high-volume stations like **1 Ave & E 6 St** require more frequent bike rebalancing compared to lower-demand stations. See Fig 2.

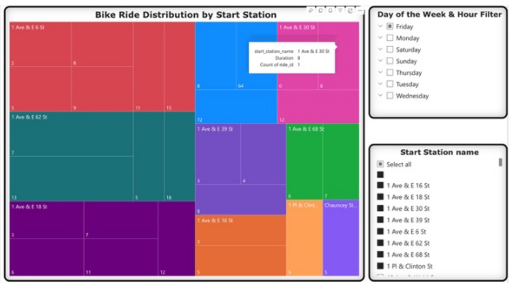

Fig 2 Tree Visualization

4.2 Sankey Diagram - Ride Flow from Start to End Stations

The Sankey diagram illustrates ride flows between start and end stations, highlighting commonly traveled routes and helping urban planners optimize bike lane placements. Thicker lines indicate a higher frequency of trips between connected stations. **Example Conclusion:** The Sankey diagram shows that rides between **Grand Army Plaza & Central Park S** and **2 Ave & E 72 St** are among the most common, suggesting an opportunity to improve connectivity between these locations. see **Fig. 3**.

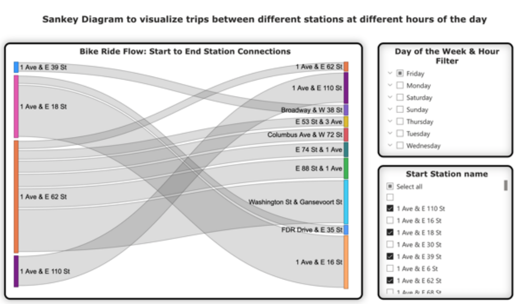

Fig 3 Sankey Diagram

4.3 Heatmap - Peak Bike Usage by Hour & Day

The heatmap identifies peak bike usage hours across different days, providing insights into ride demand trends and ensuring bike availability during high-traffic periods. Darker shades represent higher usage intensities. **Example Conclusion:** The heatmap confirms that **morning (8-10 AM) and evening (4-7 PM) rush hours** see the highest demand, emphasizing the need for increased bike availability during these times. See Fig 4.

These visualizations were developed using Power BI for interactivity and ease of analysis. Each figure is linked to larger-format or interactive versions to enhance exploration. By employing these visualization techniques, this project offers actionable insights into CitiBike usage patterns, supporting data-driven urban mobility strategies.

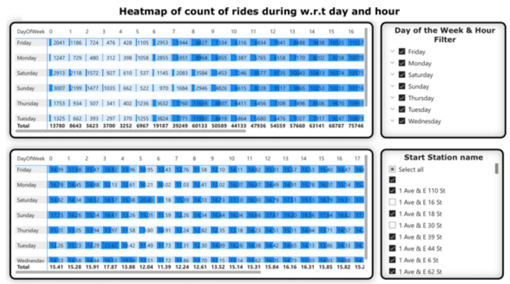

Fig 4 Matrix chart

5. USAGE AND CRITIQUE OF AI TOOLS

AI tools were instrumental in data analysis, visualization recommendations, and report structuring. ChatGPT (February 2025) suggested suitable visualization techniques, summarized data trends, and refined workflows. AI-assisted pattern recognition streamlined large dataset exploration, enhancing efficiency.

6. INTERPRETATION OF RESULTS

The analysis revealed peak ride demand during morning (8-10 AM) and evening (4-7 PM) hours, with business districts and major transit hubs experiencing the highest usage. High-demand stations like 1 Ave & E 6 St require frequent bike rebalancing, while common travel routes, such as Grand Army Plaza & Central Park S to 2 Ave & E 72 St, indicate key connectivity needs. Weekend ride patterns suggest increased recreational use, contrasting with weekday trends. Unexpectedly, certain stations showed imbalanced bike inflow vs. outflow, highlighting the need for more efficient bike redistribution strategies. The findings emphasize opportunities for urban planners to improve bike lane placements and CitiBike operators to optimize station rebalancing efforts. Future research could explore predictive analytics for demand forecasting and real-time ride data integration for dynamic bike allocation strategies.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the college for providing academic resources and guidance throughout this project. Special thanks to CitiBike and its open data initiative for making real-world datasets available for analysis. I also appreciate the insights and feedback from faculty members and peers, whose constructive discussions helped shape the direction of this research. Finally, I acknowledge the role of AI-assisted tools like ChatGPT, which aided in refining data processing workflows and enhancing the report structure.

REFERENCES

1. [1] E. Fishman, "Bikeshare: A Review of Recent Literature," *Transport Reviews*, vol. 36, no. 1, pp. 92-113, 2016.
2. [2] S. A. Shaheen, A. P. Cohen, and E. W. Martin, "Public Bikesharing in North America: Early Operator and User Understanding," *Mineta Transportation Institute Report*, 2013.