Pedalling to Profit

KEY FINDINGS FROM EDA

Trip Duration Insights

- Median trip duration: 14 min
- 75% of rides under: 22 min → Commuters & casual riders dominate
- Only 1.62% of trips under 2 min
  - Just 0.60% had same start & end → likely false starts or docking errors

Interpretation & Recommendation:

- Mostly genuine usage, low misuse → a reliable customer base
- Consider a 2-min grace period or waived unlock fee to reduce friction
- Flag stations with frequent short loops for inspection (bike/station faults)

Outliers & System Load

Trips Exceeding 42 Minutes

- 6.5% of trips last longer than 42 minutes
  - Significantly above the typical range (median: 14 min)
  - Often reflects tourist or leisure use, inefficient returns, or user negligence
- These trips reduce bike turnover, limiting availability for other users
  - Contributes to redistribution burden and operational drag

Recommendations

- Tiered Pricing: Encourage timely returns with higher fees for prolonged use
- Idle-Time Alerts: Use real time GPS tracking + alerts for bikes inactive > x hours

Extremely Long Trips (> 12h)

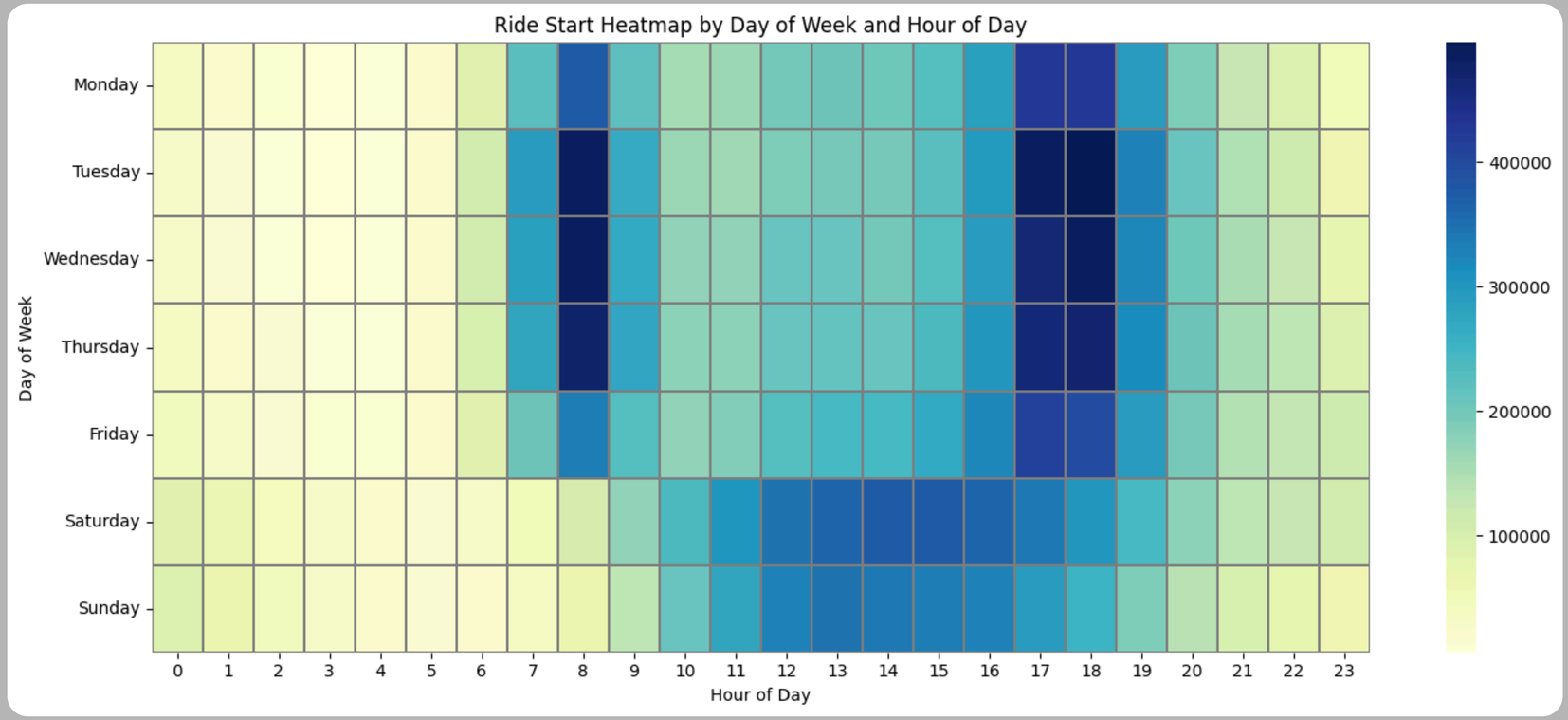
Extremely Long Trips (> 12 hours)

- 0.11% of all trips (~32,970) lasted over 12 hours
- While rare, they pose serious operational risks and may indicate:
  - Forgotten or abandoned bikes
  - Docking station malfunctions
  - Potential theft or misuse

Interpretation:

- Operational Cost & Supply Chain Risk:
  - Long trips, especially extreme outliers, limit inventory rotation, increase wear, and strain rebalancing resources
  - Leads to customer dissatisfaction at under-stocked stations

TEMPORAL USAGE TRENDS



Analysis on when people ride most frequently by hour of day and day of week:

- *Weekday mornings (7–9 am) and evenings (5–7 pm) show commuting peaks.*
- *Weekend afternoons (10 am–6 pm) show high leisure activity.*

These insights help the business:

- *Understand peak demand windows*
- *Schedule rebalancing and maintenance*
- *Target marketing or pricing strategies*

Operational Insights

Stations with high positive net flow (sinks) may overflow with bikes during peak periods and need regular redistribution.

- Hop Exchange, The borough
- St. James's Square, St. James's
- Queen street, 1 Bank
- Brushfield street, Liverpool street

Stations with high negative net flow (sources) can run out of bikes frustrating customers during high-demand periods.

- Waterloo station, 2 Waterloo
- Lancaster Gate, Bayswater
- Cloudesley road, Angel
- Boston place, Marylebone

These insights can guide:

- Strategic station placement
- Bike redistribution planning
- Demand forecasting models (predict station-level inventory
- Incentive programs to encourage returns to underused stations

DATA SCIENCE USE CASES

1

Demand Forecasting

Predict usage by hour/day/station to optimize bike distribution using Time-series models

2

Customer Segmentation

Use clustering to identify commuter vs tourist vs leisure profiles using clustering techniques

3

Bike Availability Prediction

Predict when a station is likely to run out of bikes or become completely full using regression or classification models

4

Station Rebalancing Optimization

Recommend how many bikes to move between stations to maintain service levels using Reinforcement Learning

5

Trip Duration Prediction

Estimate how long a user will ride based on start time, station pair, and bike model.

Thanks