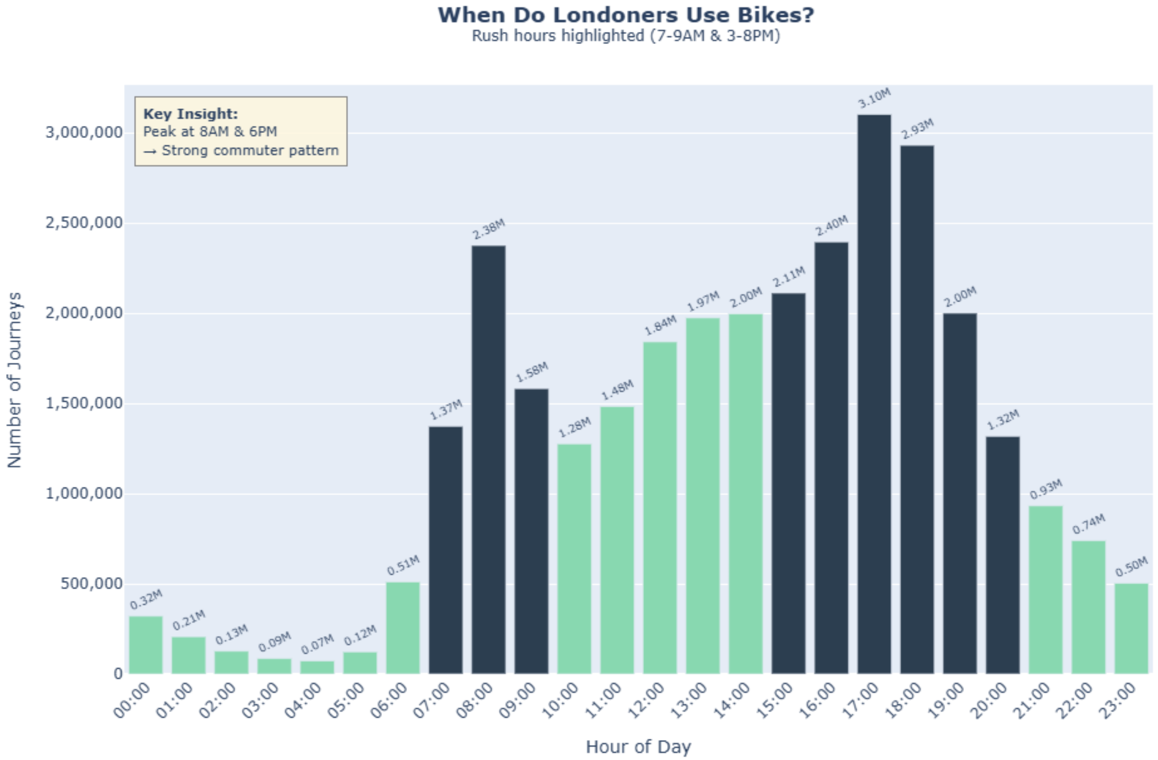
Bike Demand Forecasting

Machine learning model for predicting bike-sharing demand using time series analysis and seasonal decomposition

Nitya Devaraj

Data overview (2019-2021)

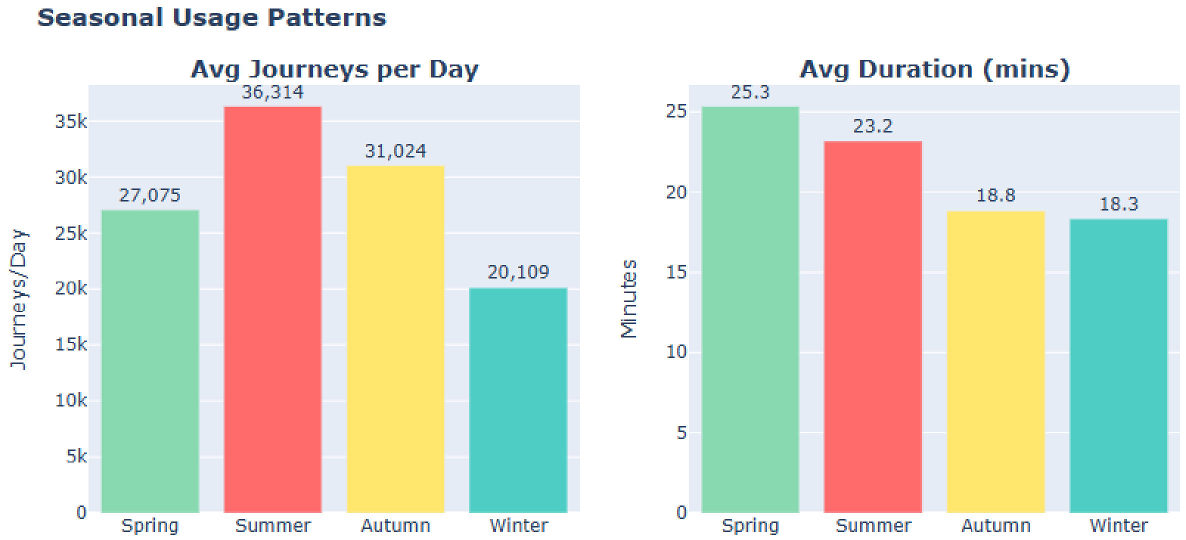
- EDA Outcomes
  - **Infrastructure**: 17,766 unique bikes
  - **Scale**: ~10 million journeys per year
  - **Consistent Demand**: Only 11% variation between busiest and quietest days
  - **Short-Term Usage**: 87% of journeys ≤ 30 minutes
  - **Post-COVID Growth**: 5.5% increase in journeys per day
  - **Commuter Pattern**: Peaks at 7–9am and 4–8pm
  - **Operational Insight**: Bikes need redistribution before 7am; low overnight usage = maintenance window

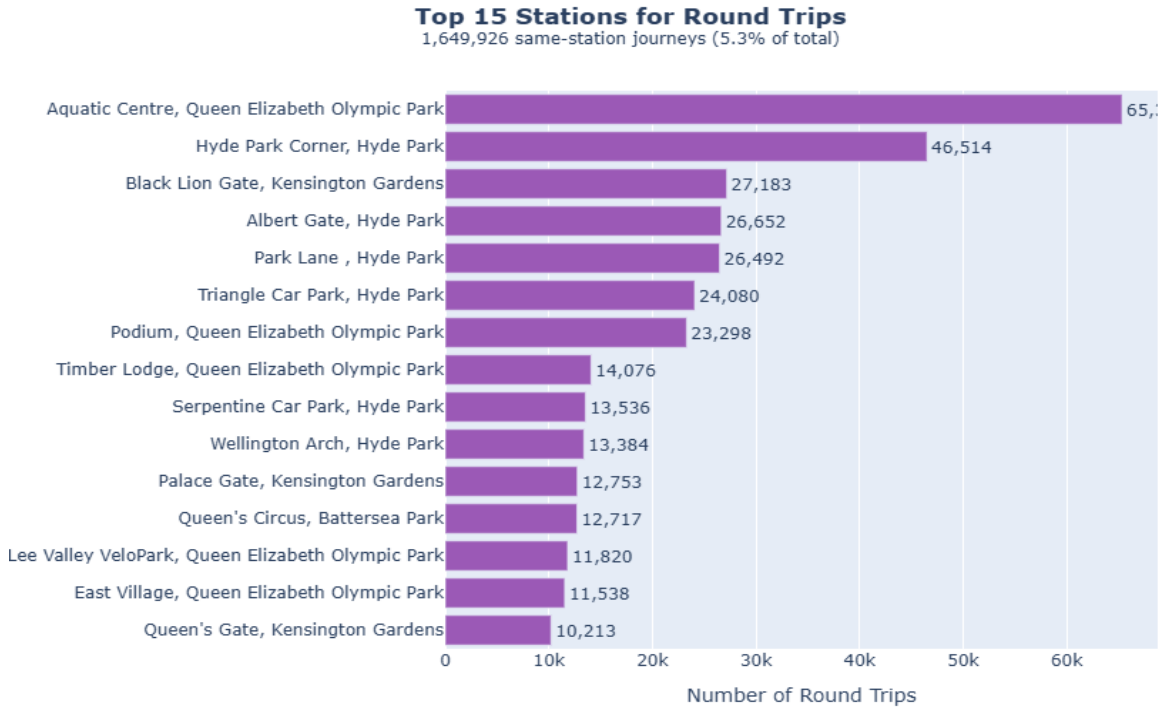


- Data science use cases
  - **Demand Forecasting**: Predict hourly/daily demand per station for fleet sizing, pricing, etc.
  - **Bike Redistribution Optimization**: Predict where bikes accumulate/deplete to improve availability and reduce costs
  - **Customer Segmentation**: Identify user types (commuter, leisure, tourist) for targeted marketing and products

Usage pattern- customer behaviour and seasonality

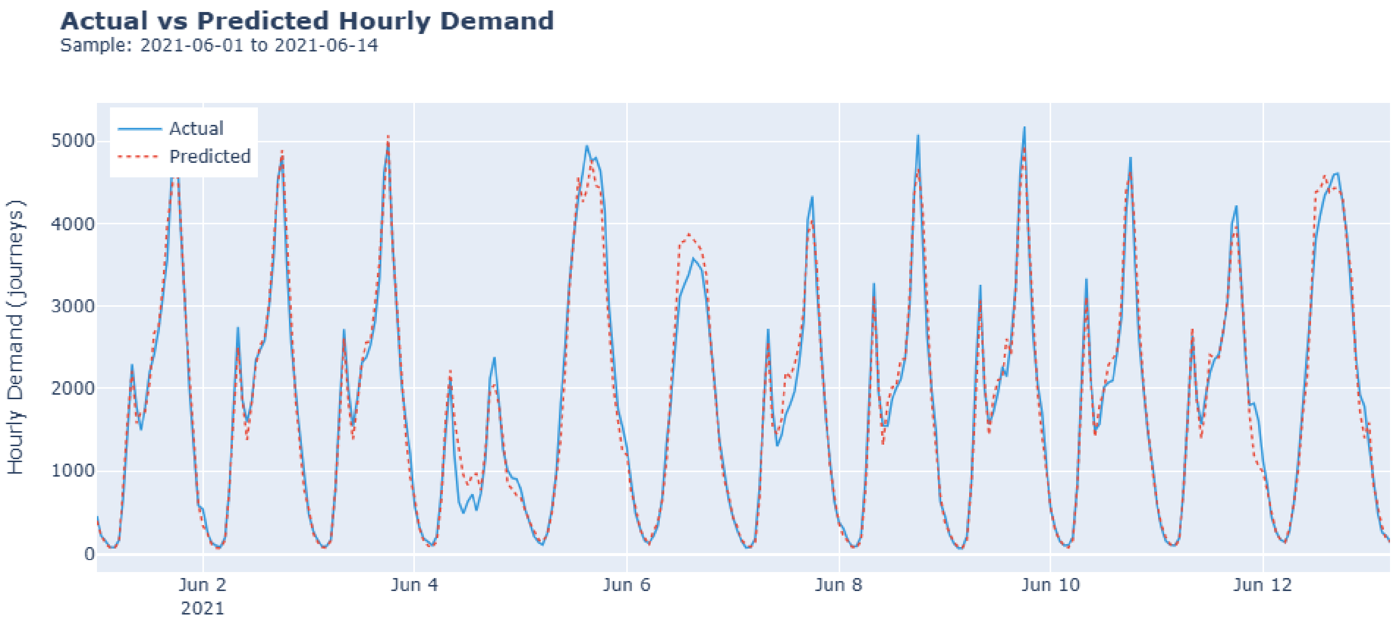
- Seasonal demand patterns
  - Summer: highest demand (scale fleet +20%)
  - Spring: longest ride duration (leisure behaviour)
  - Autumn/Winter: shorter commuter dominated rides
- Customer segments
  - Two distinct user types with different needs
  - Commuters
  - Leisure users: Round trip journeys at parks/tourist areas
  - 5.3% of all journeys are same station round trips
  - Hotspots: Hyde Park, Kensington Gardens, Olympic Park

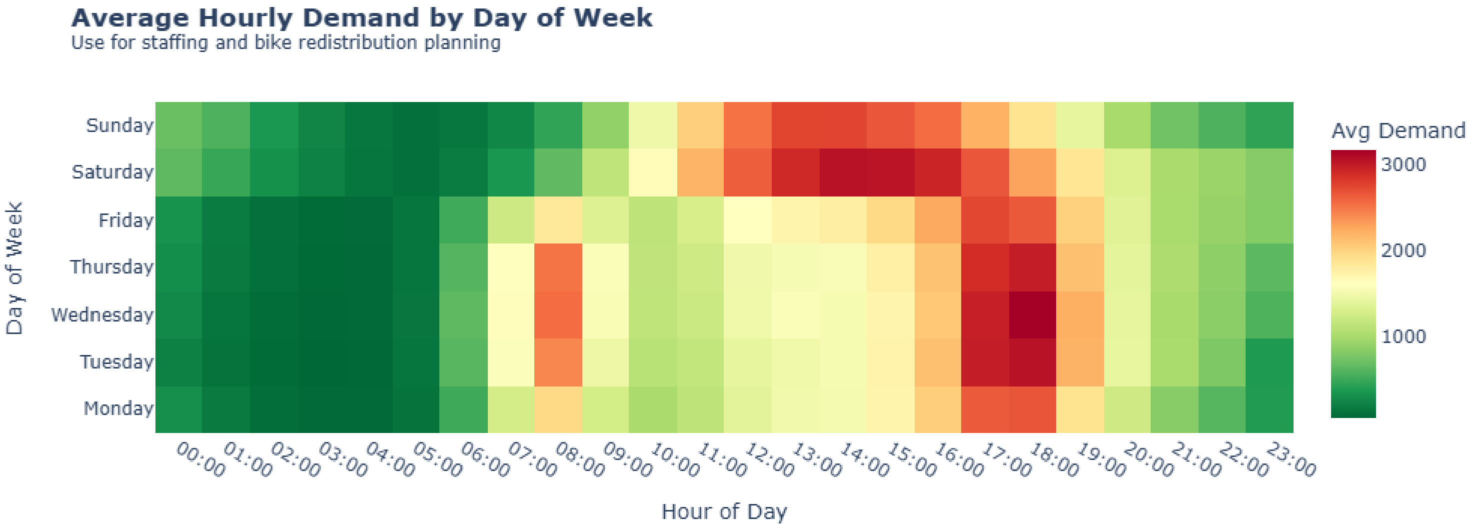




Demand Forecasting Model

- Model performance
  - XGBoost model predicts hourly demand with 96% accuracy (R^2= 0.966)
  - Visual 1- actual vs predicted line chart (2-week sample)
- Demand patterns for operations
  - Visual 2- predicted hourly demand heatmap by day of week
  - Weekday commuter peaks and weekend midday patterns
  - Identifies optimal redistribution windows
- Business application
  - **Fleet Sizing**: Predict peak demand to determine required inventory + buffer
  - **Staff Scheduling**: Deploy redistribution crews during demand valleys (10PM-6AM)
  - **Dynamic Pricing**: Increase prices during predicted high-demand periods
  - **Bike Redistribution**: Pre-position bikes before morning/evening rushes
- Next Steps
  - Integrate weather/events data for improved accuracy
  - Deploy as real-time API for live operational decisions





Recommendations

- 1. **Optimise bike redistribution**
  - **Strategy**: Combine station imbalance data + hourly demand forecasts
  - **Actions**:
  - Pre-position bikes at **source stations** before 7AM rush
  - Deploy collection crews at **sink stations** overnight 10PM-6AM Prioritise top 10 imbalanced stations (80% of redistribution need)
  - **Visual**: Station imbalance bar chart (sources vs sinks)
  - **Impact**: ↓ operational costs, ↑ bike availability, ↑ customer satisfaction

- 2. **Customer-Segmented Pricing Strategy**
  - **Strategy**: Two products for two distinct user types
  - **Actions**:
  - **Commuter Pass**: Flat monthly fee for unlimited 30-min rides (87% of journeys qualify)
    - Target: Weekday 7-9am, 4-8pm users
  - **Leisure Premium**: Higher per-ride pricing, round-trip/park stations
    - Target: Weekend users, tourists, round-trip journeys (longer duration = higher revenue, see slide 3)
  - **Dynamic Pricing**: Surge pricing during predicted peak hours (6-8PM weekdays)
  - **Impact**: ↑ revenue per journey, ↑ customer retention (locked-in commuters), optimised capacity