Toronto Bike Share Ridership Analysis (Jan–Sep 2024)

Objective: Analyze Toronto Bike Share ridership data to identify trends in daily usage, peak hours, and popular stations, and to compare weekday vs weekend usage. This project demonstrates data analysis, visualization, and A/B-style experimentation skills.

---

1. Data Acquisition

- **Source:** Bike Share Toronto Ridership Data
- **Files:** 9 CSV files covering Jan–Sep 2024, each containing individual ride records.
- **Columns:** Trip ID, trip duration, start/end station ID and name, start/end time, bike ID, user type, model.

---

2. Tools & Libraries

- Python 3.11
- Libraries:
    - pandas for data cleaning and manipulation
    - numpy for numerical operations
    - matplotlib and seaborn for visualization

    - glob for handling multiple CSV files
    - scipy.stats for statistical analysis (t-test)
- IDE: VS Code

---

3. Data Preparation & Cleaning

Step 1: Combine CSVs

Step 2: Inspect and clean data

- Initial shape: (5,342,388 rows, 12 columns)
- Extra BOM column 'ï»¿Trip Id' removed
- Column names standardized to lowercase and underscores

Step 3: Handle missing values and datetime conversion

- Converted start_time and end_time to datetime format
- Dropped rows with invalid datetime entries

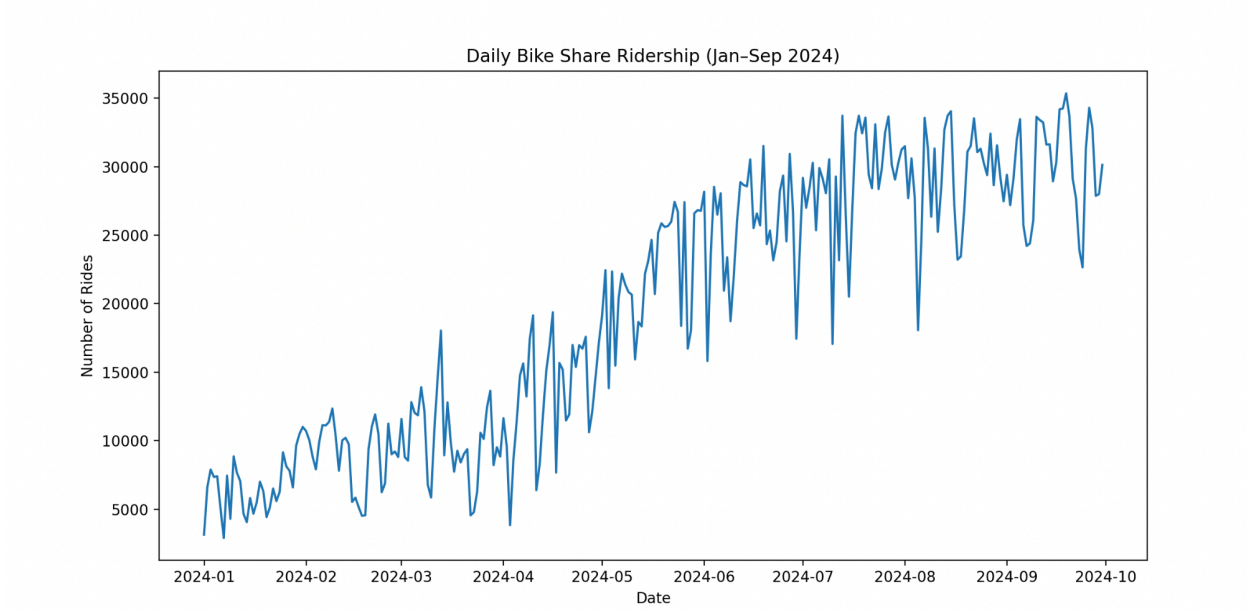
Step 4: Check dataset after cleaning

- Final shape: (5,342,388 rows, 11 columns)
- Some missing values in start_station_name, end_station_name, and model were retained (not critical for overall trends).

---

4. Exploratory Data Analysis

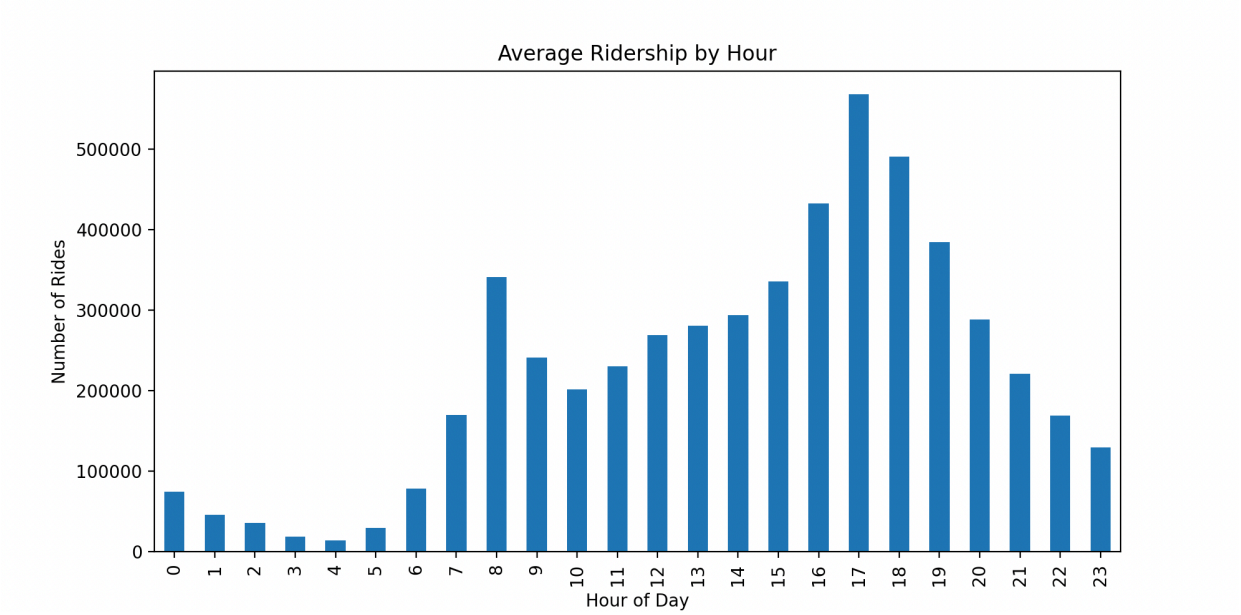
4.1 Daily Ridership



Findings:

- Ridership shows a **general upward trend from January through October**, reflecting growing usage over the year.
- Peaks in **July and September–October** suggest that both summer months and early fall are the busiest periods, likely influenced by favorable weather, seasonal commuting patterns, and recreational use.
- This trend highlights **seasonal effects** and can help prioritize resource allocation during high-demand months.

4.2 Hourly Ridership



Findings:

- Daily ridership exhibits **clear commuting patterns**, with a **morning peak around 8 AM** and the **highest usage in the evening around 5 PM**.
- Midday ridership gradually increases from 10 AM to 5 PM, then declines into the late evening, while very early morning hours have minimal usage.
- These patterns suggest that bikes should be **strategically deployed during peak commuting hours**, and maintenance or redistribution can be scheduled during low-demand periods.

---

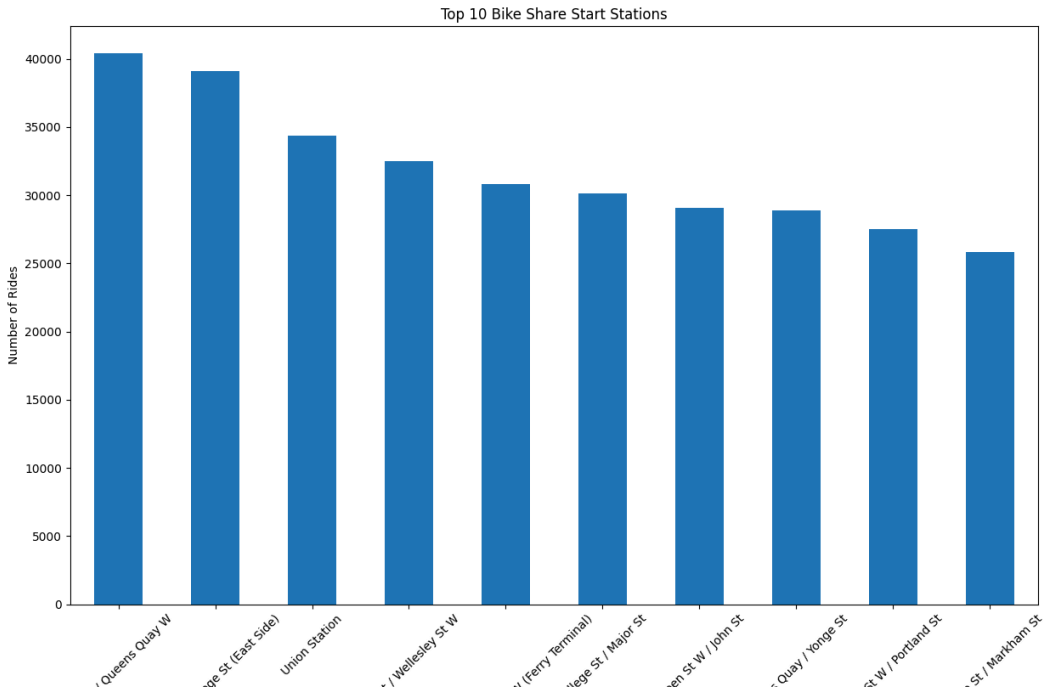
5. Weekday vs Weekend Analysis (A/B-style)

Results:

- **Total rides:** Weekdays = 3,914,865; Weekends = 1,427,523
- **T-test:** T-statistic = 1.273, p-value = 0.205
- **Interpretation:**
    - Difference in daily ridership is not statistically significant at 5% level, though visual trends clearly show higher weekday usage.
    - However, the practical difference is meaningful: weekdays consistently show higher ridership, reflecting commuter behavior and weekday activity patterns.
    - This suggests that while variability in daily counts reduces statistical significance, the trend itself is operationally important for planning bike availability, especially during workdays

---

6. Top Start Stations



Findings:

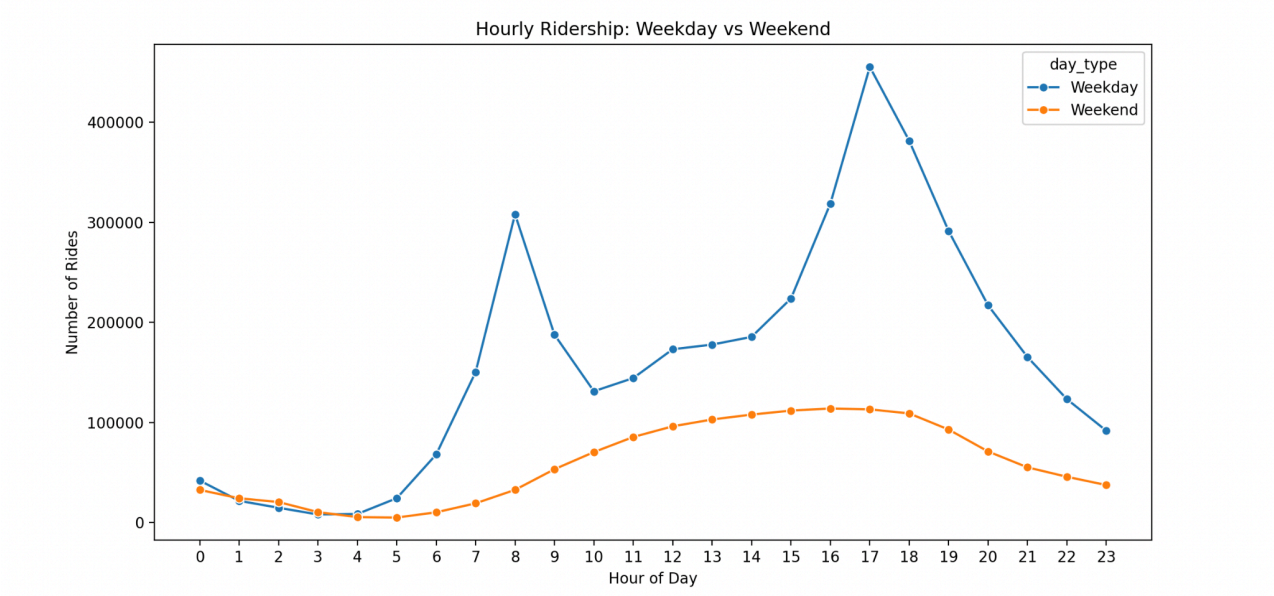
Top 10 Start Stations:

York St / Queens Quay W	40377
Bay St / College St (East Side)	39100
Union Station	34350
Bay St / Wellesley St W	32475
Bay St / Queens Quay W (Ferry Terminal)	30798
College St / Major St	30155
Queen St W / John St	29060
Queens Quay / Yonge St	28875
King St W / Portland St	27542
College St / Markham St	25848

- The busiest stations are **York St / Queens Quay W, Bay St / College St (East Side), and Union Station**, indicating concentrated demand in downtown and high-traffic areas.
- These stations collectively account for a **large share of total rides**, suggesting that both commuter and recreational users heavily rely on these locations.
- The data highlights **key hubs for operational focus**, where increasing bike availability, expanding docking capacity, or scheduling maintenance during off-peak times could maximize efficiency and user satisfaction.
- Understanding these high-demand nodes is critical for **resource allocation and optimizing service coverage** across the network.

---

7. Hourly Ridership by Day Type



Findings:

- Ridership is much higher overall on weekdays compared to weekends.
- Weekdays show sharp peaks around 8 AM and 5 PM.
- Weekends have a smoother curve, peaking in the afternoon.
- Weekday ridership is strongly tied to commuting, while weekend ridership reflects leisure and social activity.
- **Interpretation:**
    - Weekday ridership is dominated by commuting, with sharp peaks at 8 AM and 5 PM.
    - The evening peak is higher, reflecting both work trips and after-work activities.
    - Weekend ridership is lower, smoother, and peaks mid-afternoon due to leisure and social travel.
    - Overall, weekdays show structured travel patterns, while weekends reflect flexible, varied activity.

---

8. Conclusions & Recommendations

- **Key Points:**
    - Ridership peaks in July and September–October, and weekdays see much higher usage than weekends.
    - Morning (8 AM) and evening (5 PM) peaks dominate weekday patterns, while weekends show smoother, mid-afternoon activity.
    - Certain stations (York St / Queens Quay W, Bay St / College St, Union Station) handle a large share of rides.

- **Recommendations:**
    - Increase bike availability and docking capacity at top stations during weekday commuting hours.
    - Adjust deployment for midday and early afternoon weekend demand.
    - Schedule maintenance or redistribution during low-demand hours to minimize service disruption.

Project Highlights

- **Skills demonstrated:** Data cleaning, handling large datasets (~5.3M rows), Python data analysis, visualization, A/B-style testing, interpreting statistical results.
- **Tools used:** Python, Pandas, Matplotlib, Seaborn, SciPy.
- **Impact:** Clear insights on ridership trends, peak hours, and station demand — exactly the type of quantitative analysis a transportation-focused internship seeks.