Exploring Rider Behavior and Station Expansion Opportunities in POGOH Bike Share

Introduction

This report examines ridership patterns and system usage for Pittsburgh's POGOH bike share program, focusing on trip data collected between April 2022 and July 2025. POGOH has become a central part of the city's sustainable transportation network, offering both pedal and e-assist bikes to students, faculty, community members, and tourists. The organization's unique funding model, which provides free unlimited access to university affiliates, has shaped ridership dynamics and concentrated usage around campus areas. At the same time, tourists and casual riders represent a growing market, particularly during summer months and weekends.

The purpose of this analysis is to better understand these distinct user groups, evaluate data quality and consistency, and identify insights that can inform future station expansion, promotional strategies, and system planning. The report follows the CRISP-DM framework: beginning with business understanding, then moving through data exploration and data quality assessment, and concluding with hypotheses that will guide modeling and decision-making in subsequent phases.

1.1Business Understanding: Background

POGOH is the only public bike share in Pittsburgh. The first bike share program was called Healthy Ride, and it started on May 31, 2015. In 2022, Healthy Ride changed its name and became POGOH, with all new stations and bikes. The new system started on May 6, 2022 with 38 stations and about 350 bikes, and it included both pedal and e-assist (electric) bikes. As beginning 2025, POGOH has 60 stations and 600 bikes.

POGOH gives people another way to get around the city without using a car. Riders can choose pedal bikes or e-bikes, and they unlock them with the Lyft phone app also single ride can use cards(Lyft Urban Solutions, n.d.). The program is nonprofit and supported by local partners. It also has low-cost options for students and people with limited income. Besides transportation, POGOH helps people stay active, reduces traffic, and supports small businesses near the stations.

There are several positive impacts of the program to the city:

- The bike share system has created several benefits for Pittsburgh. By replacing short car trips, it lowers CO_2 emissions and supports a healthier city (Lyft Urban Solutions, n.d.).
- Riders get exercise and more affordable travel choices, while "community rides" and adaptive bikes make the program open to more people.
- POGOH also helps local businesses by bringing customers near stations. Through partnerships with schools and employers, members can ride pedal or e-bikes at a low cost or sometimes free under 30 minutes, which builds stronger community connections

Pittsburgh is one of the top 15 U.S. cities for bike commuting, and the city already has more than 117 miles of marked bike routes. This shows that Pittsburgh is a strong place for bike share and still has room for growth. POGOH has been expanding quickly. In 2024, riders completed 464,383 trips, which was more than double the year before (POGOH, 2025). Some stations are much busier than others, especially near university areas, which shows where demand is highest (University of Pittsburgh, n.d.). Riding patterns also change by season and by time of day, with fall and afternoon rush hours being the busiest.

The new e-bikes have made it easier for more people to ride, even on Pittsburgh's hills. Membership and price changes also affect how often people use the system.

Looking ahead, POGOH plans to add more than 50 new stations by end of 2025 which including new locations in Wilkinsburg and other neighborhoods that do not yet have service (POGOH, n.d.). These expansion efforts, along with partnerships with employers and the addition of new rider features, show how POGOH continues to grow and strengthen its role in Pittsburgh's transportation.

Why use data science analysis the growth and create future explaining strategy for the program? Because looking at the data helps explain how and why people ride. With data science, we can see patterns like which stations are busiest, when trips happen, and how new changes affect use. This kind of data makes planning more accurate, so the system can grow in ways that match what the city needs. Data analysis also helps the organization plan its next phase, such as where to expand stations and how to manage areas that are already crowded or underused. By studying trip duration, station usage rates, and times when bikes are in highest demand, POGOH can see limitations (for example, stations that fill up too fast have no spot to park).

Pricing is another important factor. Data science can show how membership costs or pay-per-ride fees influence usage — such as whether lower-cost memberships increase short commuter trips or attract more casual riders. Together, these insights guide decisions that help the company growth.

Major stakeholders of the program:

POGOH has several groups of people who are connected to the system. The riders are the most important because they use the bikes every day. They care most about cost, easy access, and having bikes available when they need them (POGOH, 2025). Local businesses are also affected since stations bring more traffic and increase sales (Lyft Urban Solutions, n.d.). The City of Pittsburgh is another key stakeholder, especially the Department of Mobility and Infrastructure, because the city sets the rules and helps connect bike share to other forms of transportation (City of Pittsburgh, n.d.). The management team also play important role: they are responsible for making decisions that keep operations running smoothly reduce churn rate, planning expansion beyond the city core to reach more residents, and making sure existing bikes and stations are well maintained.

1.2 Project Objectives and Success Criteria

The main goal of this project is to use data to better understand how people use POGOH and to turn those findings into steps that can actually improve the system. The goal is to provide insights that support decisions on expand new stations, pricing strategies, operation improvements, and what opportunities exist for continued growth.

Supporting Goals

In order to reach the main objective, there are three supporting goals to serve as stepping stones:

1. Analyze where rides start and end
    - *measure*: Track trip counts by station and identify the top and bottom 10 stations which will help enhance maintains and expand new stations. Learning which stations often empty or full at what time during the week. Plan successful reducing the number of problem stations by 90% or above
2. Analysis the peak time find weekly and seasonal patten
    - measure: Compare ridership at peak hours, days of the week, and across seasons. Success will be shown in a clear report that explains usage patterns and helps

make the service more efficient. At the same time, the goal is to reduce rider complaints about not finding a nearby bike or an open return spot by at least 80%.

3. Study how prices and e-bikes effect usage
   - measure: Track changes in trip numbers after pricing updates or promotional offers. This analysis is challenging because there are no interviews included in the study. It is also difficult to isolate the effects of each factor—such as upgrades, system expansion, and different promotions—since multiple changes may happen at the same time. That makes it harder to identify the corresponding components that drive usage.

Overall project success will be measured in two ways.

1. Quantitative success. Ridership has already more than doubled in each of the last two years, and the program now operates 60 stations and around 600 bikes. With plans to add more than 50 additional stations, especially near downtown and the university area, it is reasonable to expect ridership to grow by another 100% by the summer of 2026 compared to 2024 data.

   Another measurable target is an 80% reduction in empty or full station incidents, which can be achieved by building more stations and using live app data to guide bike redistribution.

   Finally, a long-term goal is to increase new user sign-ups. This means expanding beyond the existing downtown core and bringing in riders from more neighborhoods across the city.

2. Qualitative success. Beyond the numbers, success will also be seen in stronger community satisfaction include both riders and local businesses, positive feedback from the City of Pittsburgh and other partners, and evidence that the analysis is actually being

used for decision-making. If riders feel the system is more reliable and accessible, and management team adopt the recommendations to guide future growth, then the project will have achieved meaningful impact.

The impact of this project can be looked at under operations and community.

On the operations side, the study should help keep stations more balanced, make bike moves smoother, and give better ideas on where new stations will setup, what's the ratio of the bikes to install.

On the community side, success means more people will see POGOH as an easy and affordable way to travel, even for people who have not used it before. Shops and restaurants close to stations may also get more customers, and the city can move closer to its goals of less traffic and fewer car emissions. POGOH plans to add more than 50 new stations by the end of 2025.

Because Pittsburgh has long winters that slow down bike use, the real improvements will probably be best measured in the September of 2026 since 70% of existing usage is close to the university area.

1.3 Assessment of Responsibility

Any project that uses live ride data, such as POGOH's bike share system, has responsibilities at several levels.

Legal and Regulatory: Transportation data includes trip start and end times, GPS locations, and user accounts. Since this information can sometimes be linked back to individuals, the organization must follow federal and state privacy rules.

Data security is another major issue. POGOH collects geographic information that could show where people live, work, or study. If such details were leaked, it might put users at risk or expose personal routines. In order to prevent this, the organization should use encryption, access controls, and clear data storage rules. It is also important to only collect the information that is truly necessary for system operations. Collecting unnecessary data increases risks and makes it harder to guarantee privacy

Another key responsibility is consent and transparency. Riders should be clearly informed that their trip data is being collected and how it will be used. This means providing written consent when signing up, and also giving regular updates so riders understand what data is collected, why it matters, and how it is protected.

Besides that, user agreements should include waivers where riders accept responsibility for safe use, they understand they are under city traffic laws.

There are also ethical responsibilities. Fairness means making sure that pricing, memberships, and station locations do not leave out low-income or underserved riders. POGOH already offers discounted memberships for riders who qualify for assistance programs, and many universities and colleges in Pittsburgh have joined the employer program. Through this program, students, faculty, and staff can take unlimited trips of up to 30 minutes for free (POGOH, n.d.). This is especially important because students are often considered a low-income group, and the benefit makes the system more affordable and accessible for them.

Stakeholders share in this responsibility. Schools can help by setting up safe bike routes through campus, protecting students at busy intersections, and reporting missing bikes. Local businesses benefit from bike stations, so they also play a role. They can provide safe parking near their shops, offer promotions to riders, take part in awareness campaigns.

To reduce risks, both technical and community strategies should be used. On the technical side, anonymization can remove personal information from public datasets. Data can also be shown in ranges rather than exact numbers, lowering the chance of identifying individuals. Regular security updates and audits add another layer of protection. On the community side, outreach and education can build trust. Schools and employers could share short safety videos and reminders to teach safe riding and traffic awareness.

In conclusion, responsibility in this project is more than just following laws. It is about protecting people and communities. With strong data protection, informed consent, collecting only what is needed, and active community involvement, the bike share program can grow in a way that is responsible, safe, and beneficial for everyone.

1.4 Data Science Goals & Success Criteria

The organizational need is to understand ridership patterns, station demand, and the effect of pricing and e-bike adoption on system usage. This can be framed as a combination of regression, time series analysis, and clustering.

- Regression models will help quantify relationships between pricing, membership type, and ridership growth.
- Time series analysis will capture seasonal and hourly patterns of trips.
- Clustering methods can be used to group stations or riders with similar usage patterns, which can guide targeted strategies.

Together, these approaches directly connect to POGOH's objectives of expansion, equity, and operational efficiency.

The primary data science objective is to develop models that can explain and predict ridership demand at different stations and times. This will show which stations are busiest, when they are used most, what's the average length of the trips.

A secondary goal is to evaluate e-bike adoption, including identifying what share of trips involves e-bikes and whether they help reach new rider groups. Another objective is to explore station clustering, so stations with similar demand patterns can be managed in a similar way. These outcomes provide direct value by informing where to add new stations, how to balance bikes, and how to design fair membership programs.

Success will be measured in two ways: numbers and usefulness. For the prediction models, success means the results should be close enough to real ridership that they can guide decisions. For example, if the model says a station will have 200 rides on a summer weekday, we want it to be within about 20–25 rides of the actual number. For time-based forecasts, success means being able to predict general patterns such as which months are busiest and what times of day see the most demand, even if the exact numbers are not perfect. For clustering, success means being able to group stations into 3–5 categories (such as “busy shopping area,” “university area,” or “low-use residential”) that clearly show different usage styles.

On the qualitative side, success means the analysis produces insights that city officials, businesses, and POGOH staff can easily understand. The results should lead directly to actions like knowing where to place new stations, how to balance bikes between start and end trips, and whether membership or pricing changes are helping more people ride.

The analysis will begin with exploratory data analysis to confirm patterns. For demand prediction, multiple regression and different forest regression will be tested and compared. Time series methods such as ARIMA or Prophet will be used to forecast ridership by season and day of week. For clustering, K-means and hierarchical clustering will be applied to group stations and riders. Model evaluation will use cross-validation and performance metrics relevant to the organizational goals. Visualizations will also play a key role in communicating findings to POGOH leadership and community partners.

Objective–Technical Mapping

- Organizational goal: Identify busiest stations and where to expand.
    - *Technical deliverable:* Regression and clustering models that show demand and group stations by usage patterns.
- Organizational goal: Reduce empty/full station problems.
    - *Technical deliverable:* Time series forecasts to predict peak hours and guide bike redistribution.
- Organizational goal: Understand effect of pricing and e-bike rollout.
    - *Technical deliverable:* Regression analysis on trip data with membership type, trip length, and bike type variables.

Constraints and Assumptions

1. The analysis assumes that ridership patterns will stay mostly consistent over the study period, so past data can be used to predict future demand.
2. Data quality may be affected by missing records, incorrect trip times, or GPS errors, which could lower the accuracy of the results.
3. The project is limited by the type of data available. There is not enough detailed citywide geographic or resident activity data to fully predict where new stations might be needed.
4. It is difficult to measure how much e-bikes have changed usage because both pedal bikes and e-bikes are free for member trips under 30 minutes. As a result, there is no clear data to prove that the increase in ridership is directly caused by the addition of e-bikes.

1.5 Project Plan

This project will follow the CRISP-DM framework, focusing on four phases: Data Preparation, Modeling, Evaluation, and Deployment. Each step is designed to answer POGOH's main questions about ridership, station use, and future expand.

Phase-Specific Planning

Data Preparation**:** The first step is to clean the trip data by fixing errors, removing duplicates, and handling missing records.

Modeling: regression will be used to study how pricing and membership affect ridership. For example, the model can test if discounted memberships increase short commuter trips. Time series models will look at seasonal and hourly demand, such as predicting higher ridership near Pitt and CMU when students return in September. Clustering will group stations with similar usage patterns, such as busy downtown hubs versus school neighborhood stations.

Evaluation: Models will be checked for accuracy and usefulness. The goal is not perfect predictions but clear patterns that help decision-making. For example, the results should confirm if new school year will bring double the rides compared to summer school break. Final results will be shown with charts to make sure they are easy for shareholder who have no professional background to understand.

Deployment: The final report will include deployment success measurement, recommendations, and documentation. In the future, POGOH and the city could use live forecasting to guide station expansion and bike redistribution, especially during peak times at university or downtown shopping area.

Timeline and Milestones

- Weeks 1–2: Clean data test correlations and create features.
- Weeks 3–6: Build regression, time series, and clustering models.
- Weeks 7–8: Refine models, draft results, and make an implementation plan.
- Weeks 9–10: Write final report includes evaluation, deployment and present recommendations.
- Deployment 2026: Apply changes March–Jun; measure impact August-October.

Risks and Likelihoods

Risks include missing or inaccurate data. Backup plans include using simpler features, testing different models, and focusing on accomplished the main goal first.

2.1 Data inventory

The project uses a combination of primary trip data and supporting contextual datasets. The table below summarizes the available files, formats, and sources

File Name / Dataset	Format	Source	Description
2022c; 2023c;2024c;2025c;	CSV	POGOH internal database	Detailed trip-level data from 2022 annual to 2025 (to July 31). Includes user type, trip duration, start/end stations, and timestamps.

changed_capacities	CSV	POGOH internal database	Historical updates to bike station capacities.
old_stations	CSV	POGOH internal database	Records of retired stations.
POGOH Annual Reports 2020–2023(additional data sources)	PDF	POGOH website (public reports)	Summary data on user demographics (age, gender, income), promotions, event highlights, and financial breakdown.
station_export	CSV	POGOH internal database	list of 60 active stations with attributes such as name, group, capacity and location as 2025 before phase 3 expanding
2022 & 2025 Pricing	PDF	POGOH internal database	List of different membership and pricing. In order to understand how e bike impacted the user.

Data Collection Methodology:

The primary trip database contains trip information automatically recorded through the application software. When users unlock and return bikes, data are captured and transmitted to the central database. These records include station location, membership type, bike type, trip duration, and timestamps. Occasional gaps or errors may occur due to equipment malfunctions or recording inaccuracies.

The ‘changed-capacities’ table are extracted from POGOH’s internal management system. The ‘changed-capacities’ file documents modifications to station capacity over time, while the ‘old - stations’ file contains records of stations that were retired. The ‘station_export’ file represents the current list of active stations. These datasets were compiled and maintained by the POGOH management team.

The Annual Reports (2020–2023) were obtained directly from the official POGOH website. These reports are published annually by the organization and summarize program outcomes, survey results, and financial highlights. Demographic information (such as age, gender, and income) is based on user surveys, which may carry sampling bias and do not perfectly align with the trip data. Promotion and event details are selective and highlight only major activities, rather than providing a full account of all initiatives.

2.2 Data Description:

Variable Documentation

Column Name	Data Type	Example Values	Units / Format	Domain Meaning
Id	Integer	1080806, 1080805	Unique numeric ID	Unique identifier for each trip record.
Closed Status	String	NORMAL, GRACE_PERIOD	Text	Status of the trip record (valid,free).
Duration	Integer	124, 185, 703	Seconds	Length of the trip measured in seconds.
Start Station Id	Integer	13, 41, 6	Numeric	Unique identifier for the station where the trip began.
Start Date	String	7/31/2025 23:57	Datetime (MM/DD/YYYY HH:MM)	Timestamp when the trip started.
Start Station Name	String	S Bouquet Ave & Sennott St	Text	Descriptive name of the trip’s starting station.

End Date	String	8/1/2025 0:01	Datetime (MM/DD/YYYY HH:MM)	Timestamp when the trip ended.
End Station Id	Float	50.0, 8.0, 39.0	Numeric (nullable)	Unique identifier for the destination station (null if missing).
End Station Name	String	Glasshouse, Atwood St & Bates	Text	Descriptive name of the destination station.
acct ID	Integer	9723, 17380	Numeric	Unique rider account identifier.
Rider Type	String	MEMBER, CASUAL	Text (categorical)	Classification of rider type.
Bike Model	String	EFIT, PBSC	Text (categorical)	Model of the bike used for the trip.
Product Name	String	Annual, Monthly, Mobility Justice Membership	Text	Type of subscription or pricing product associated with the trip.

Dataset Structure and Relationships

- Primary Key: Id uniquely identifies each trip.
- Station Relationship: Start Station Id and End Station Id link trips to station reference tables (e.g., station_export).
- User Relationship: acct ID identifies the rider; this can be linked with aggregated demographic information if available.
- Subscription Relationship: Product Name ties trips to membership products and pricing plans.

Time Period, Volume, and Scale

- Coverage: January 2022 April – July, 2025
- Records: per data file,

Year	Trips
2022 data started April 29	72,347
2023	211,251
2024	463,384
2025 up to July 31	236,198

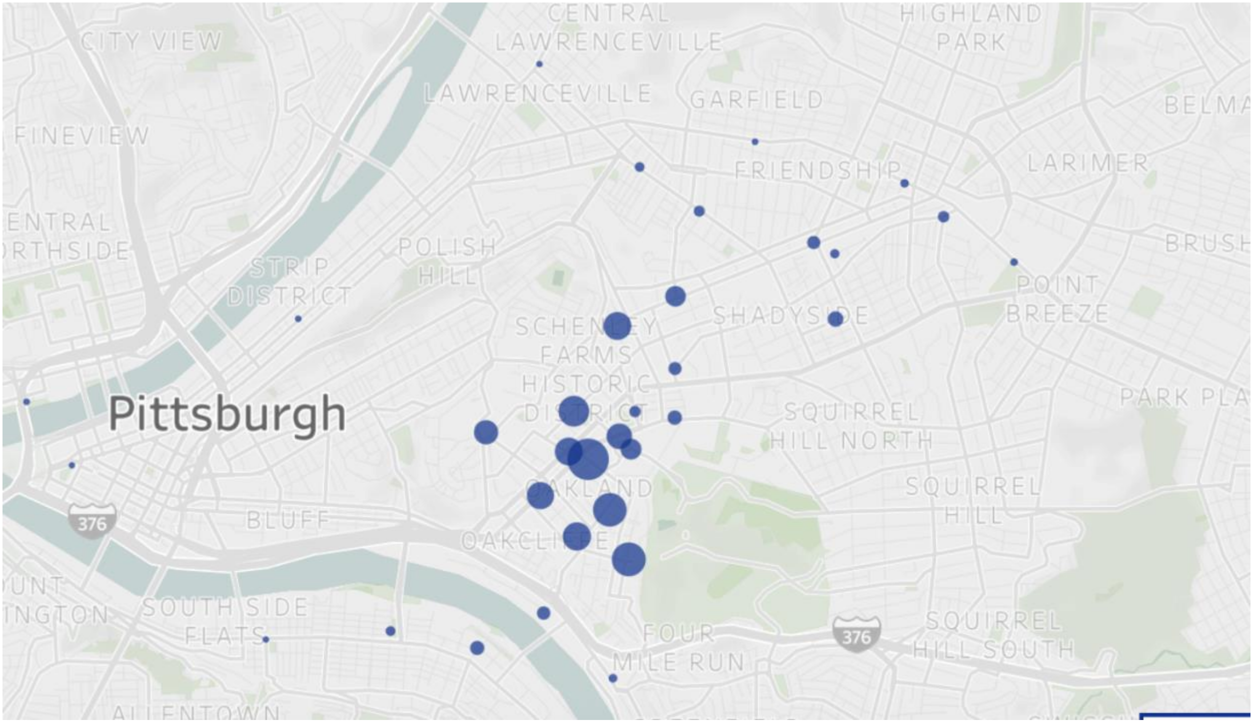
- Scale: Large dataset with granular trip-level details, suitable for time-series, ridership forecasting, and station-level analyses.

Categorical Variable Encoding

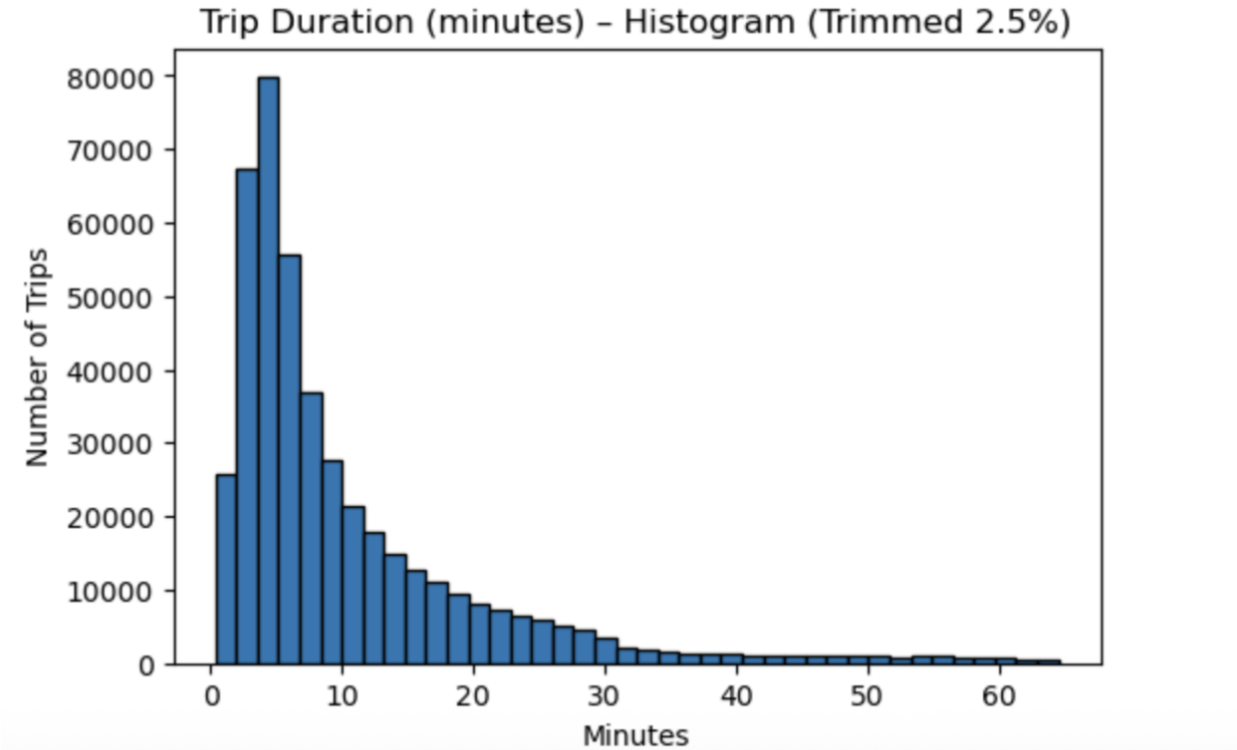
- Closed Status: Categorical — NORMAL, GRACE_PERIOD (no charge).
- Rider Type: Categorical — MEMBER, CASUAL (CASUAL = single ride).
- Bike Model: Encoded as EFIT = e-bike; FIT = pedal bike.
- Product Name: Categorical subscription types — e.g., Annual, Monthly, Single Ride, Corporate Membership, Mobility Justice Membership.

2.3 data exploration Report

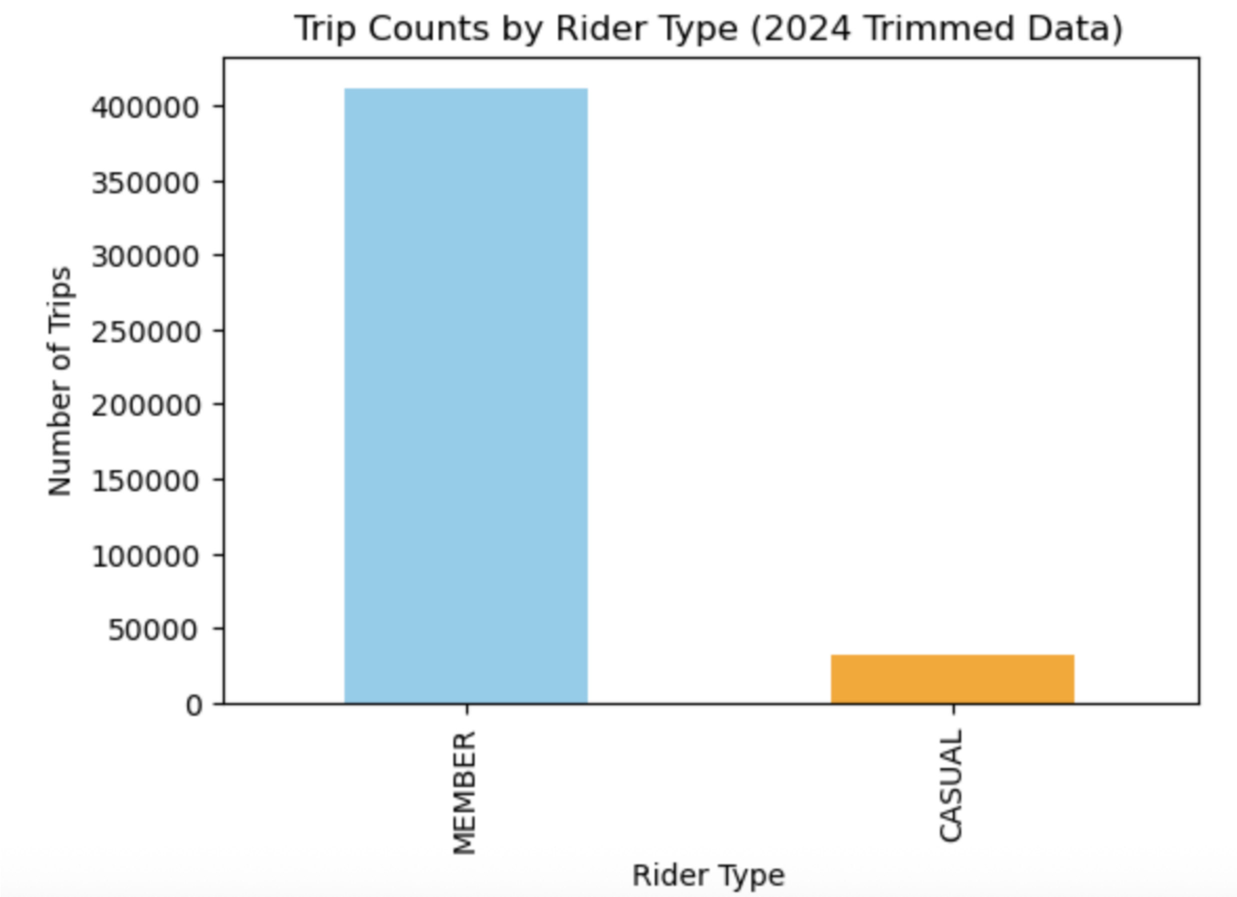
In the study of trip data, the first seven months of 2025 recorded a total of 236,198 trips, which represents an increase of nearly 15% compared to the same period in 2024. Since the 2024 dataset covers the full year, I used the first seven months of 2024 for a fair comparison with 2025. The descriptive statistics for both years are very similar, so I focused on the 2024 dataset for detailed analysis.



The map above is from Pittsburgh university website which show ride usage data from July 2024 to July 2025, limited to stations with a minimum of 1,000 trips per year. The busiest station is S Bouquet Ave & Sennott St**,** which recorded over 46,000 rides. Notably, all of the top 10 high-traffic stations are located within the immediate radius of this central hub, highlighting a strong concentration of demand in that area.



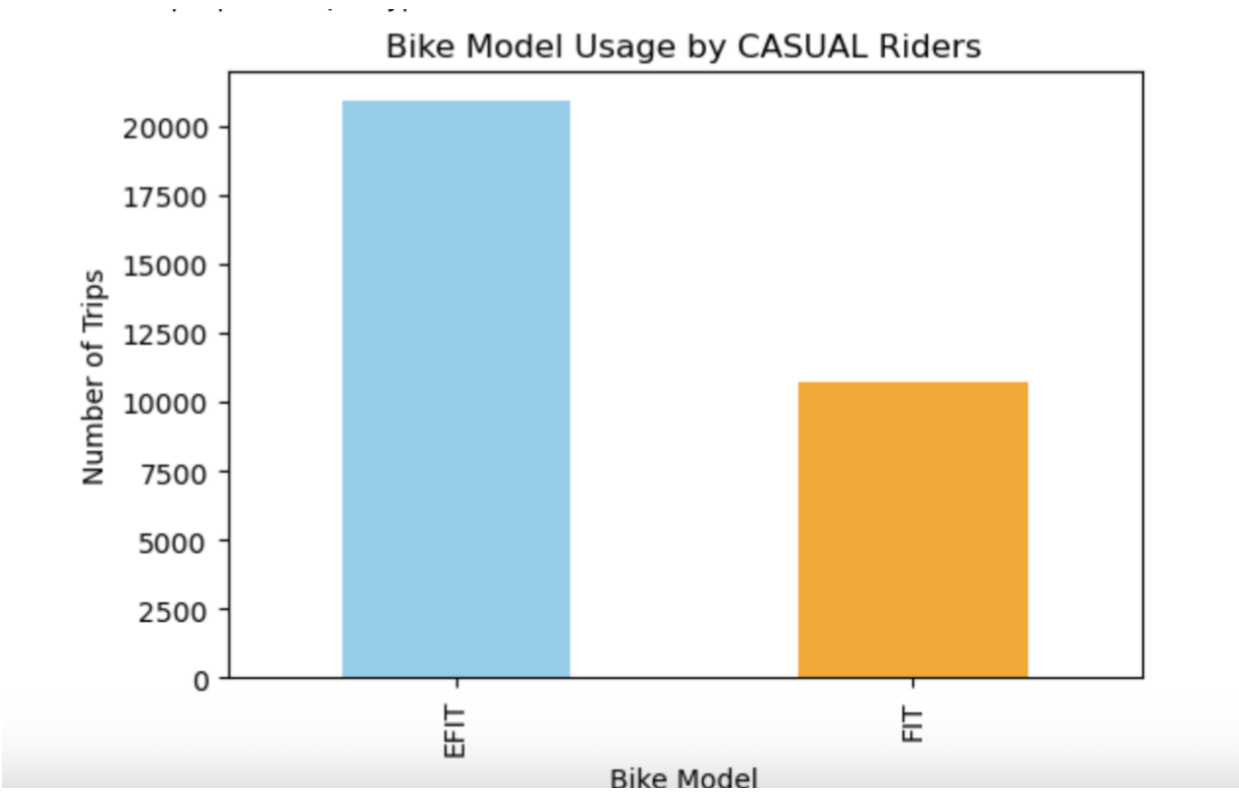
After reviewing the 2024 trip data, I found 199 trips with missing end station information, which indicates bikes that were never properly returned to a station. These records would distort the average trip duration, so I removed them. To further minimize the effect of extreme outliers, I also removed the top 2.5% of the longest trips and the bottom 2.5% of the shortest trips (the latter may reflect broken bikes that ended immediately after being unlocked). After trimming, the minimum trip duration was about 0.38 minutes, and the maximum was about 64 minutes. The mean duration is approximately 10.5 minutes, with 75% of trips lasting 13 minutes or less. This indicates that the majority of rides comfortably fit within the free 30-minute limit provided by membership plans.



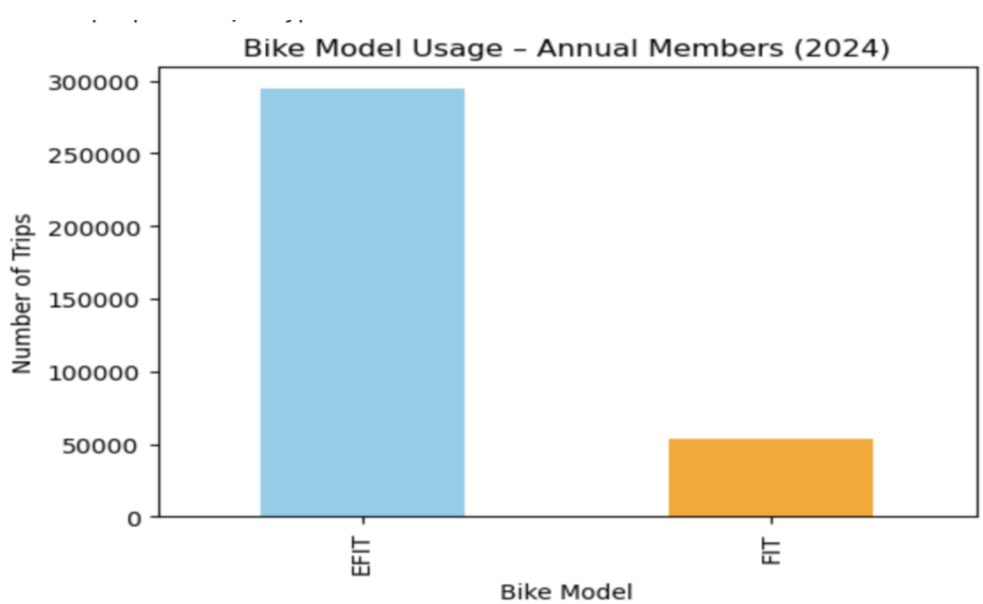
Compare Rider Type of 2024 data, member did 92.87% trip, Casual did 7.13% trip.

- CASUAL: trips taken without a membership, usually Single Ride or Day Pass purchases.
- MEMBER: trips taken by people with an Annual, Monthly, hours or special discounted membership.

Since most members receive unlimited 30-minute rides that include e-bikes**,** the analysis will focus more closely on e-bike usage among Single Ride riders**.** For these casual riders, e-bike trips incur an extra $1.5 per 30 minutes rental**,** making their usage patterns and willingness to pay particularly important to understand.

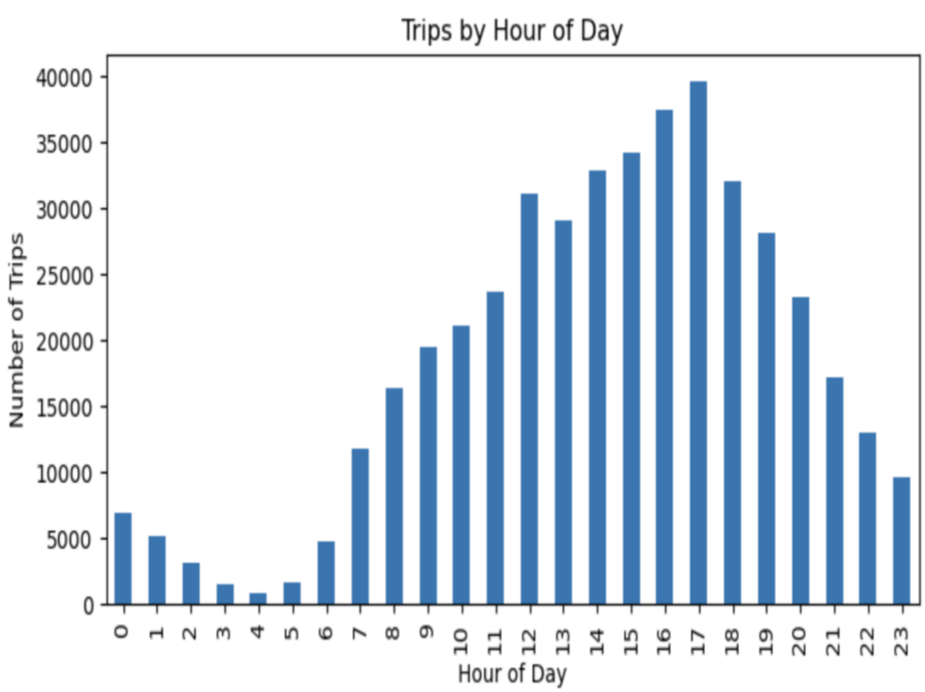


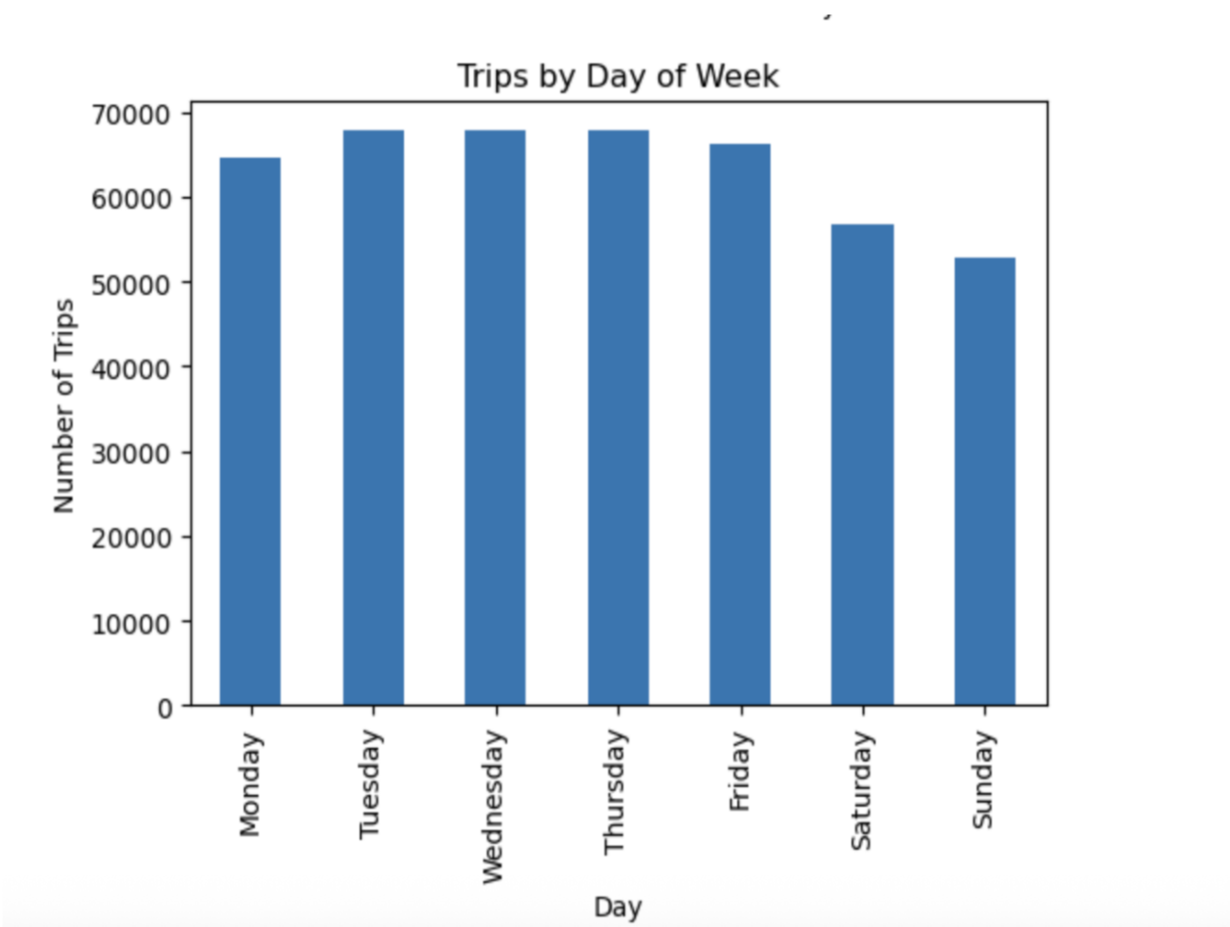
The study shows that approximately 66% of casual riders chose e-bikes, even though each trip incurred an additional $1 surcharge. In contrast, about 34% of casual riders selected standard pedal bikes.

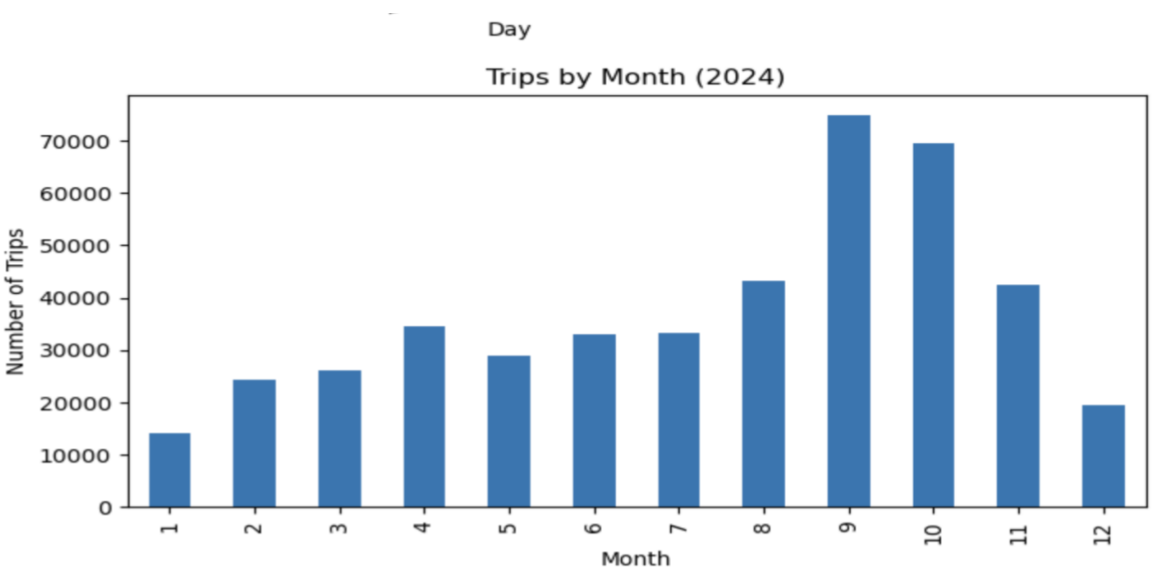


When examining the 2024 data for Annual Membership riders, where both pedal and e-bikes are included at no additional cost for trips under 30 minutes, the share of e-bike usage rose significantly to 85.5%, while pedal bike usage declined correspondingly. This suggests that when the price barrier is removed, riders show a strong preference for e-assist bikes.

This doesn't mean install 85.5% e-bike is the simple answer. According to the program website, the replacement cost of a pedal bike is approximately 1,000 dollars, while an e-bike costs around 2,500 dollars. In addition, the maintenance costs of e-bikes are higher than those of standard pedal bikes. These cost considerations play an important role in decisions about expanding or upgrading the fleet, as the financial burden of acquiring and maintaining e-bikes is substantially greater than that of regular bikes.

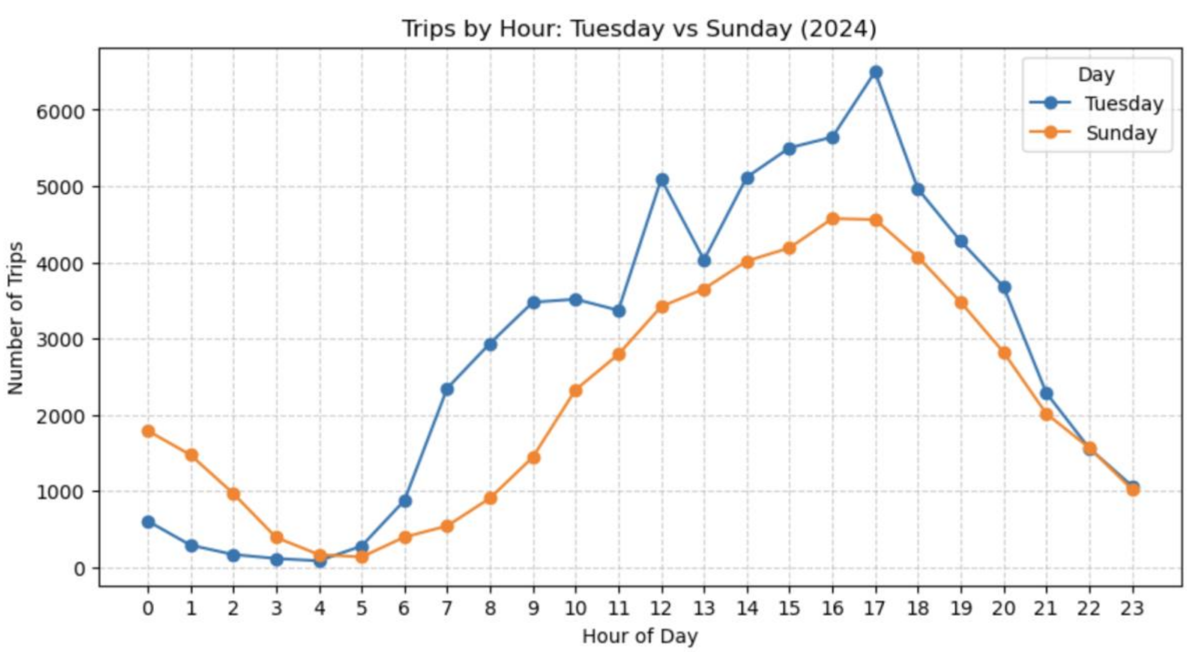






By examined usage by hour, day of the week, and seasonal patterns. The data show that September and October are the peak months, with ridership volumes nearly double those of average months. During the winter, usage declines significantly as colder temperatures discourage bike riding. Trips begin to increase again in the spring, but because many riders are students and university employees, demand is lower during the summer break. Ridership then rises sharply at the end of August as academic activities resume, leading into the high levels observed in September and October.

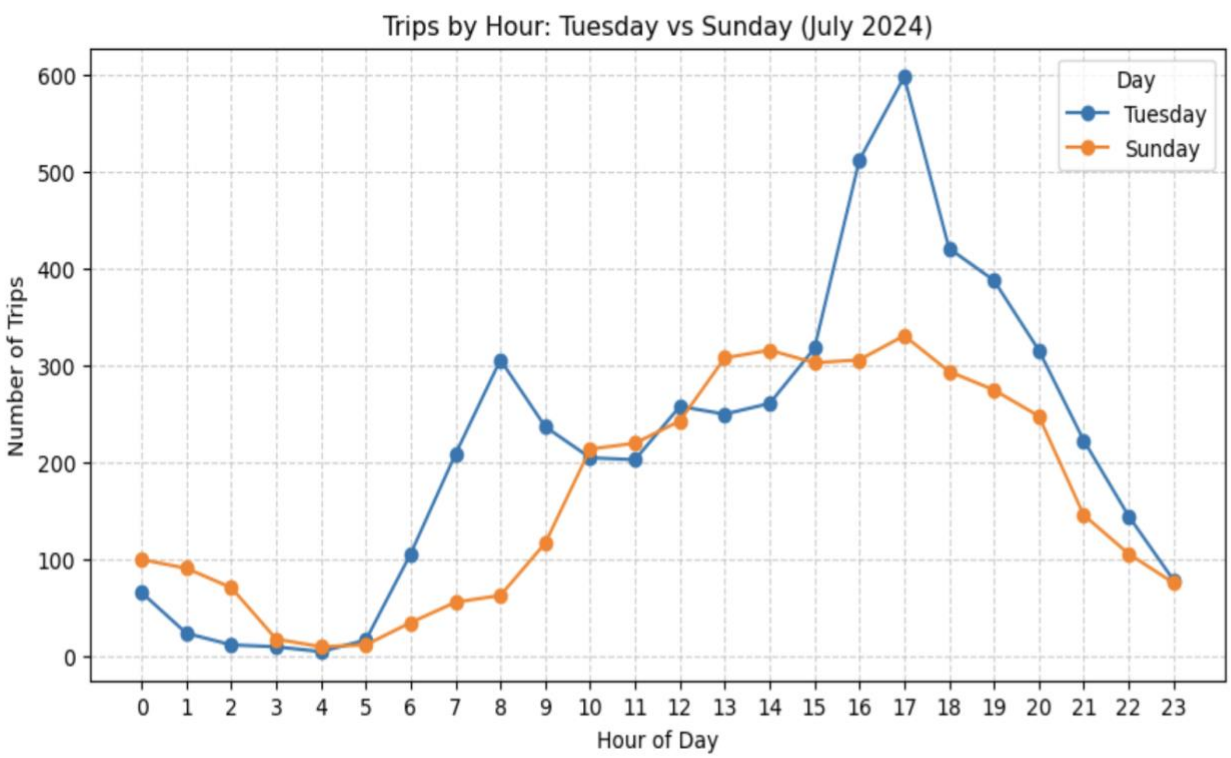
In terms of time of day and day of the week, the patterns mostly mirror university schedules. Unlike systems in larger commuter-heavy cities, there is no pronounced morning and afternoon rush hour. Instead, ridership peaks steadily between noon and 7 p.m. on weekdays, with the single busiest hour around 5 p.m. This reflects the lifestyle of the primary user group: students and faculty often have flexible schedules, afternoon classes, or campus events, which create a higher concentration of trips later in the day. The 5 p.m. peak likely corresponds to the end of classes and the transition to evening activities, rather than to a traditional office-commute rush.



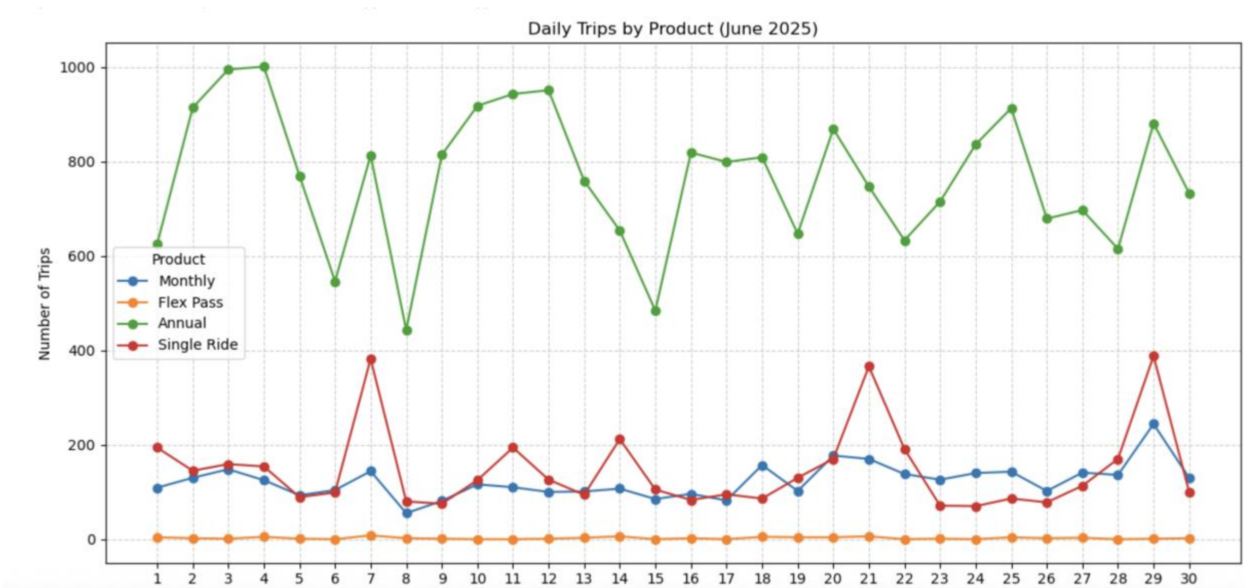
Weekday vs Weekend: By comparing Tuesday and Sunday to observe the differences between weekdays and weekends, distinct patterns emerge. On weekends, particularly Sunday, there are fewer peaks during the afternoon but more rides occurring after midnight.

Finding: Sunday trips typically start later in the day and follow a more recreational or leisure-oriented pattern.

Implication: Weekend usage aligns more closely with tourism and leisure activities, which could justify targeted promotions such as day passes for visitors or discounts on casual rides during weekends.

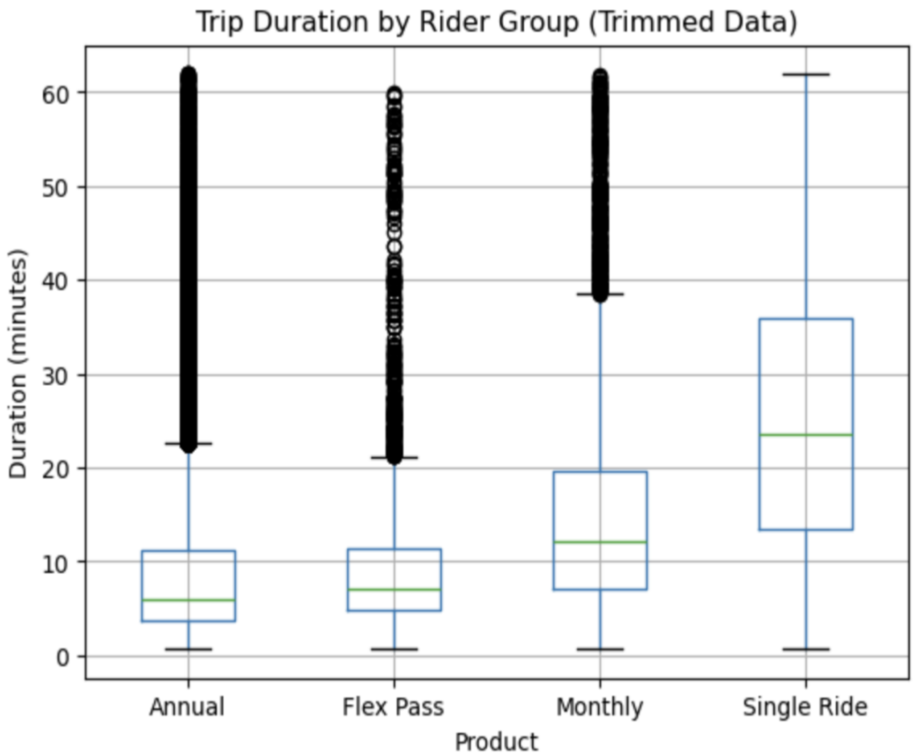


Analysis summer season: When analyzing July 2024 weekday usage specifically, the patterns differ. Since much less students attend summer school, weekday ridership clearly reflects morning and afternoon rush-hour traffic. However, because school activities are reduced in July, the number of trips between rush hours declined, showing fewer rides during the day. In contrast, weekend casual riders maintained the same pattern without being significantly impacted by the decrease in school-related activity.



Using June 2025 as the most recent dataset during the school lower activities month, we analyzed differences between community riders and tourists in a summer month. The number of tourist users, represented by Single Ride riders, nearly doubled on most Saturdays, although trips were noticeably lower on the second Saturday due to thunderstorms. In contrast, Monthly riders**,** who represent local community users not covered by student or corporate plans, showed much lower Saturday peaks. This contrast highlights an opportunity to further promote shared bike use among tourists during the summer months, when leisure and recreational demand is naturally higher.

```
=== Trip Duration Stats by Product (trimmed data) ===
               count   mean  median   min    max
Product Name
Annual        169475   8.58    5.92  0.65  61.93
Flex Pass       2559   9.97    7.15  0.67  59.95
Monthly        13304  14.41   12.22  0.65  61.85
Single Ride    16325  25.93   23.57  0.65  61.93
<Figure size 800x500 with 0 Axes>
```



From the boxplot analysis, we can see that Monthly riders (community users) and Single Ride users (tourists) record higher average trip durations compared to other groups. The mean trip duration for Single Ride users is close to 26 minutes, with the 75th of trips lasting 37 minutes. This means many tourist trips are long enough to be billed at the equivalent of two 30-minute blocks. In contrast, Annual members average only about 8.5 minutes per trip, reflecting shorter, point-to-point rides. This difference suggests that tourists typically ride for longer distances and durations, e-bike will much helpful for them while annual members primarily use the service for quick, utilitarian trips.

```
=== Top 10 Start Stations (Single Ride, July 2025 Weekends) ===
Start Station Name
North Shore Trail & Fort Duquesne Bridge    201
21st St & Penn Ave                          137
Penn Ave & 7th St                           122
10th St & Penn Ave                          111
S 27th St & Sidney St. (Southside Works)    106
Liberty Ave & Stanwix St                     97
42nd St & Butler St                          90
17th St & Penn Ave                           85
52nd St & Butler St                          72
Forbes Ave & Market Square                   70
Name: count, dtype: int64

=== Top 10 End Stations (Single Ride, July 2025 Weekends) ===
End Station Name
North Shore Trail & Fort Duquesne Bridge    200
Penn Ave & 7th St                           139
21st St & Penn Ave                          135
Liberty Ave & Stanwix St                    102
10th St & Penn Ave                          101
S 27th St & Sidney St. (Southside Works)    100
42nd St & Butler St                          78
17th St & Penn Ave                           77
52nd St & Butler St                          69
Burns White Center at 3 Crossings            69
Name: count, dtype: int64
```

In a further analysis of tourist behavior, we examined Single Ride trips during weekends in July 2025. The results show that tourist activity is concentrated at stations located near river trails and the downtown core. This pattern is distinct from Annual members, whose usage is heavily clustered around the university area.

Based on the exploratory analysis of POGOH trip data, several hypotheses emerge that could be tested through further modeling and analysis:

1. Tourist Behavior Hypothesis
    - *Observation:* Single Ride trips are concentrated near river trails and downtown, with longer average durations compared to members.
    - *Hypothesis:* Tourists primarily use the system for leisure and recreational purposes, leading to longer trip times and clustering around recreational or visitor-oriented stations.
    - *Organizational Action:* Create more stations connect downtown attractions, partner with surrounding hotels, and cultural venues to encourage tourist ridership.
2. Community Rider Hypothesis
    - *Observation:* Monthly and Annual members, particularly those associated with the university, have much shorter average trip times and frequent rides around campus areas.
    - *Hypothesis:* Local community members and students use the system primarily for short, utilitarian trips
    - *Organizational Action:* Maintain and expand station density around educational and employment hubs to support reliable commuting use.
3. Seasonality Hypothesis
    - *Observation:* Ridership peaks in September and October and declines in winter, with student activity strongly influencing seasonal patterns.
    - *Hypothesis:* Seasonal variation in ridership is driven both by weather and the academic calendar.

    - *Organizational Action:* Allocate more bikes and increase maintenance capacity during the fall peak, storage the bikes according to the usage during off-season months reduce the cost of operation.

4. Bike Type Preference Hypothesis
    - *Observation:* Even with an additional $1 fee, 66% of riders prefer e-assist bikes; usage rises to over 85% when the additional cost is removed for Annual members.
    - *Hypothesis:* When price barriers are lowered, riders strongly prefer e-assist bikes, especially for tourist longer trips.
    - *Organizational Action:* Prioritize investment in e-assist bikes for top tourists' stations.

All of these hypotheses can be translated into actionable strategies that support stakeholder goals. For the organization, they provide guidance on where to expand new station locations to meet growing demand. For local businesses, they open opportunities to partner with POGOH in promoting tourism and downtown activity. For the community, they encourage healthier lifestyles and more sustainable transportation choices by making shared bikes a more attractive option.

2.4 Data Quality Assessment

The majority of this case study relies on trip data automatically recorded by the application. Within the dataset of over one million entries, a few hundred records have no return station, which may be related to damaged or missing bikes, or equipment errors. In addition, approximately 2.5% of trips lasted fewer than 25 seconds, which fall within the grace period and likely reflect bike malfunctions or aborted rides. After trimming the top and bottom 2.5% of trip lengths to remove such anomalies, no other missing or suspicious data were identified.

Beyond the trip data itself, program documentation also plays an important role in quality assessment. For example, I noted several updates on the official website compared to the file labeled "2025 pricing." For this analysis, I used the most recent 2025 pricing information and compared it with the 2022 pricing structure. The website also provided additional details about eligibility rules, such as how students and faculty qualify for free 30-minute unlimited rides. These details helped clarify funding sources and revenue flows discussed in the annual report. However, the 2024 annual report has not yet been released, limiting the ability to reference the latest promotional initiatives and strategic development plans. Pricing and funding data are especially critical because they directly influence rider behavior, system revenue, and future decisions about station expansion.

It is also important to recognize that ridership patterns are shaped by the program's funding model. Free unlimited access for students and faculty has led to heavy usage around university areas, while broader community ridership is less consistently represented. This is not a flaw in the dataset itself but rather a structural characteristic of the system.

Finally, while the trip dataset is relatively simple and internally consistent, additional contextual data would make the analysis more actionable. Neighborhood-level information, such as population density or downtown residential statistics, would be especially valuable for guiding station expansion decisions. These data are likely available through city planning or census sources. Incorporating them in future work would strengthen the relevance of the analysis for both organizational planning and community engagement.

Conclusion

The analysis of POGOH trip data from April 2022 to July 2025 highlights several critical findings. First, the dataset is generally consistent and reliable, though a small number of anomalies—such as 2.5% of trips under 25 seconds long or hundreds rides with no return station—required trimming. Pricing updates and funding details from the program's website provided important context, revealing how free access for students and faculty strongly influences community ridership patterns. In contrast, tourists using single rides show longer trip durations and a concentration at downtown and riverfront stations, underscoring different usage motivations.

Seasonal and temporal analyses revealed clear patterns: ridership peaks in September and October, declines in winter, and surges again during summer weekends, particularly among casual riders. These insights suggest opportunities for targeted promotions to attract tourists during peak leisure months and to balance station distribution beyond university areas.

These findings give a clear starting point for the next phase of modeling. They show where demand is highest, how different groups of riders use the system, and what strategies—like adding new stations, adjusting prices, or running promotions—could help POGOH reach its goals. By being honest about both the strengths and limits of the data, this analysis makes sure that the next steps are based on realistic assumptions and remain focused on what matters to stakeholders.

References

Lyft Urban Solutions. (n.d.). *Pittsburgh bike share*. Retrieved from https://lyfturbansolutions.com/cities/pittsburgh-bike-share

University of Pittsburgh. (n.d.). *Draft POGOH dashboard*. Retrieved from https://analytics.pitt.edu/t/PUB/views/DraftPOGOHDashboard/TripsOverview

POGOH. (2025, January 22). *2024 ride report*. POGOH Pittsburgh bike share. Retrieved from https://pogoh.com/blog/2024-ride-report/