IE1171 Final Report: POGOH EXPANSION – Justin Do, Xinkai Li, Yuru Zhang, Haoji Wang.

Background/Introduction

POGOH is Pittsburgh’s bikeshare system that is meant for residents, students and visitors that want to take a short trip around Pittsburgh. POGOH is a complete system replacement and rebrand of Healthy rider which was the previous bikeshare system in Pittsburgh. Bike Share Pittsburgh owned and operated the bikeshare system using new equipment and software from PBSC, a new leader in the industry. Sixty stations with over 600 bikes span the Pittsburgh area allowing for easy and quick transportation at a relatively low cost. Our group objectives is to understand & analyze trends in ridership using available data, evaluate the impact of new bike lanes on ridership, analyze the effect of bike infrastructure changes on bus ridership and study the effect of bike share network expansion on transportation choices.

Starting off, the first task that we encountered when tackling our objectives is data manipulation & comprehension. Healthy Rider and POGOH data set allows us to the duration ride time, start/end station ID, date, and name. What we can comprehend from this is looking at how a covariant such as seasonality affects average duration for Healthy Rider and POGOH from 2017-2024 and when the stations are being added based on year and month. The monthly average ridership dataset allows us to see the route name, ID and route code, as well as the starting date/end with average daily riders. From this dataset we can see the average number of daily riders for any given bus route on any given day from 2017 to 2024. Monthly bus stoppage data set allows us to see the average on and off boarding for any given bus route on any given day from the time of 2019-2021. Something to note is that the data within this dataset is only for 2019 quarter three, 2020 quarters one and three, and 2021 quarter two. The transit stop usage data set shows us all bus stops active from July 2018 to June 2019; this also has average boarding for any given bus route for any given stop. Our last dataset is bike pavement, which shows us the bike lanes that are added from 1980-2015 with information of year added, street names/intersections, project ID, and street distance in miles. Furthermore, a suggestion from David Walker from our zoom meeting was to look at the data from inside and outside of Oakland. From this we can make four large hypotheses to see the interaction of bike/bus data as well as how seasonality and how inside/outside Oakland differs.

Hypothesis

1. Is there a difference in overall quantity/duration/boarding of ridership of POGOH/Healthy rider and Bus riders in and outside of Oakland
   1. HO\_a: there is a difference in average duration for POGOH outside and inside Oakland.
   2. H1\_a: there is not a difference in average duration for POGOH outside and inside Oakland.
      1. HO\_a1: There is a difference in average duration for POGOH outside and inside Oakland based on season.
      2. HO\_a2: There is not a difference in average duration for POGOH outside and inside Oakland based on season.
   3. HO\_b: there is a difference in average duration for Healthy Rider outside and inside Oakland.
   4. H1\_b: there is not a difference in average duration for Healthy Rider outside and inside Oakland.
      1. HO\_b1: There is a difference in average duration for Healthy Rider outside and inside Oakland based on season.
      2. H1\_b2: There is not a difference in average duration for Healthy Rider outside and inside Oakland based on season.
   5. HO\_c: There is a difference in average daily riders for any given bus route for any given day outside and inside Oakland.
   6. H1\_c: There is not a difference in average daily riders for any given bus route for any given day outside and inside Oakland.
      1. HO\_c1: There is a difference in average daily riders for any given bus route for any given day outside and inside Oakland based on season.
      2. H1\_c2: There is not a difference in average daily riders for any given bus route for any given day outside and inside Oakland based on season.
   7. HO\_d: There is a difference in average on/off boarding for any given bus stop for any given day outside and inside Oakland
   8. H1\_d: There is not a difference in average on/off boarding for any given bus stop for any given day outside and inside Oakland
      1. HO\_d1: There is a difference in average # of riders boarding on/off for any given bus route for any given stop outside and inside Oakland based on season.
      2. H1\_d2There is not a difference in average # of riders boarding on/off for any given bus route for any given stop outside and inside Oakland based on season.
2. Introduction of new bike lanes & stations to bus ridership
   1. HO\_e: The POGOH bike infrastructure does affect bus ridership in Pittsburgh
   2. H1\_e: The POGOH bike infrastructure leads to a reduction in bus ridership in Pittsburgh
3. Introduction of new bike lanes near existing stations increases POGOH ridership
   1. HO\_g: There is no significant effect in ridership before and after the addition of bike lanes
   2. H1\_g: There is a significant effect in ridership with the addition of new bike lanes near existing stations
4. Introduction of new bike stations increases POGOH ridership at nearby stations
   1. HO\_h: New bike stations does not have a significant effect on POGOH ridership at nearby stations
   2. H1\_h: New bike stations does have a significant effect on POGOH ridership at nearby stations

Methods

* Data manipulation
  + Healthy Rider
  + POGOH
    - Looking at yearly/quarterly data for these datasets inside and outside Oakland
  + Monthly Bus riders/stoppage
    - Looking at yearly/quarterly data for these datasets inside and outside Oakland
  + Bike lane additions
    - Looking at the addition of bike lanes by year and location
* Linear Regression
  + DID
    - The Difference-in-Differences (DiD) method was used to assess the impact of the intervention (e.g., installation of new bike lanes) on the number of cyclists. By comparing changes in ridership before and after the intervention, between the intervention and control groups.
  + GML
    - The data is count-based (ridership), making Poisson regression a suitable choice for modeling discrete data. GLM allows us to account for both the intervention effect and group differences, providing clear insights into whether the intervention impacted ridership.

Results (what is found by data manipulation)

POGOH average ridership duration from 2022-2024 per session

* Seasonality overall in and outside of Oakland
  + Q1(Jan, Feb, March) = 13.85 minutes
  + Q2(April, May, June) = 26.57 minutes
  + Q3(July, Aug, Sep) =18.72 minutes
  + Q4(Oct, Nov, Dec) = 12.75 minutes
* Seasonality overall inside of Oakland
  + Q1(Jan, Feb, March) = 10.08 minutes
  + Q2(April, May, June) = 11.75 minutes
  + Q3(July, Aug, Sep) =9.45 minutes
  + Q4(Oct, Nov, Dec) = 8.27 minutes
* Seasonality overall outside of Oakland
  + Q1(Jan, Feb, March) = 17.56 minutes
  + Q2(April, May, June) = 29.43 minutes
  + Q3(July, Aug, Sep) =22.02 minutes
  + Q4(Oct, Nov, Dec) = 17.07 minutes

Healthy ride average ridership duration decreases from 2017-2022 per session

* Seasonality overall in and outside of Oakland
  + Q1(Jan, Feb, March) = 61.7 minutes
  + Q2(April, May, June) = 87.91minutes
  + Q3(July, Aug, Sep) = 83.94 minutes
  + Q4(Oct, Nov, Dec) = 56.96 minutes
* Seasonality overall inside of Oakland
  + Q1(Jan, Feb, March) = 47 minutes
  + Q2(April, May, June) = 61.39 minutes
  + Q3(July, Aug, Sep) = 58.04 minutes
  + Q4(Oct, Nov, Dec) = 36.36 minutes
* Seasonality overall outside of Oakland
  + Q1(Jan, Feb, March) = 63.144 minutes
  + Q2(April, May, June) = 89.45 minutes
  + Q3(July, Aug, Sep) = 85.79 minutes
  + Q4(Oct, Nov, Dec) = 59.36 minutes

Bus average daily number of riders from 2017-2024

* Seasonality
  + Q1(Jan, Feb, March) = 1163 riders
  + Q2(April, May, June) = 1155 riders
  + Q3(July, Aug, Sep) =1233 riders
  + Q4(Oct, Nov, Dec) = 1209 riders
* Seasonality of riders inside of Oakland
  + Q1(Jan, Feb, March) = 2664 riders
  + Q2(April, May, June) = 2348 riders
  + Q3(July, Aug, Sep) =2510 riders
  + Q4(Oct, Nov, Dec) = 2516 riders
* Seasonality of riders outside of Oakland
  + Q1(Jan, Feb, March) = 1083 riders
  + Q2(April, May, June) = 1097 riders
  + Q3(July, Aug, Sep) =1171 riders
  + Q4(Oct, Nov, Dec) = 1136 riders

Daily boarding on and off riders for any given bus route

* Daily avg riders boarding on in and outside of Oakland
  + 2019 Q3: 9.42 riders
  + 2020 Q1: 7.96 riders
  + 2020 Q3: 3.68 riders
  + 2021 Q2: 3.77 riders
* Daily avg riders boarding off in and outside of Oakland
  + 2019 Q3: 9.37 riders
  + 2020 Q1: 7.92 riders
  + 2020 Q3: 3.68 riders
  + 2021 Q2: 3.76 riders
* Daily avg riders boarding on in of Oakland
  + 2019 Q3: 76.89 riders
  + 2020 Q1: 19.24 riders
  + 2020 Q3: 19.24 riders
  + 2021 Q2: 17.16 riders
* Daily avg riders boarding off in of Oakland
  + 2019 Q3: 76.49 riders
  + 2020 Q1: 19.24 riders
  + 2020 Q3: 19.24 riders
  + 2021 Q2: 17.14 riders
* Daily avg riders boarding on not in Oakland
  + 2019 Q3: 9.31 riders
  + 2020 Q1: 3.65 riders
  + 2020 Q3: 3.65 riders
  + 2021 Q2: 3.75 riders
* Daily avg riders boarding off not in Oakland
  + 2019 Q3: 9.26 riders
  + 2020 Q1: 3.657 riders
  + 2020 Q3: 3.657 riders
  + 2021 Q2: 3.73 riders

Bus average amount affected by new related bike stations from 2018 July to 2019 June

- 7 new bike stations

* + 7/1/2018 to 9/1/2018 = 48 riders
  + 9/2/2018 to 11/17/2018 = 50 riders
  + 11/18/2018 to 3/16/2019 =45 riders
  + 3/17/2019 to 6/15/2019 = 47 riders
  + 6/16/2019 to 6/30/2019 = 43 riders

Increase of bike stations from 2018-2019 affects the bus ridership

Conclusion

In conclusion, the results from the first hypothesis gives us insight of the relationship between bike share system, bus ridership, and seasonal factors in the Pittsburgh’s transit system. For the first set of sub hypothesis, we are able to fail to reject the null for both HO\_a & HO\_a1 and conclude that there is a difference in average duration for Healthy Rider based on inside/outside and seasonal data. For the second sub hypothesis HO\_b, we fail to reject the null as there is a difference in average duration for POGOH inside/outside of Oakland but we accept the alternative for HO\_b2. We cannot conclude that there is a difference in average duration for POGOH inside/outside Oakland based on season. For the data of average daily riders, HO\_c, we fail to reject the null here and see that there is a difference in average daily riders for any given bus route on any given day outside/inside Oakland. Although, we accept the alternative for HO\_c2, as we can not conclude that there is a difference for average daily riders for any given bus route for any given day inside/outside Oakland based on season. For the last sub hypothesis within our first hypothesis, HO\_d, we fail to reject the null and can see that there is a difference in average number of riders boarding on/off for any given bus stop for any given route on any given day inside/outside Oakland. Due to the small quantity of data within this dataset, we are not able to conclude that there is a seasonality difference in average boarding on/off in Oakland.

Furthermore, the results from the second hypothesis allows us to view the relationship between new additions of bike infrastructure such as lanes and stations and how it affects the overall bus ridership in Pittsburgh. First, it is noted that most of the bike stations are mainly concentrated in the center of Pittsburgh and nearby urban/school areas. From 2019-2021, there are several new stations along the river that have been added alongside the river from the original number in 2019/2020. From here, we can look for nearby bus stops based on the new location. Due to the lack of data for specific bus station ridership from 2022-2024, we can not accurately see the impact of new bike stations and how it affects bus ridership within this time. From what we analyzed of new bike stations from 2022-2024, we see that new bike stations in the past three years have a greater relationship with the bus stations in urban areas, Pitt area, and densely populated areas. What we are able to conclude is, we accept the alternative for H1\_e, and determine that the POGOH bike infrastructure leads to a reduction in bus ridership in Pittsburgh.

In addition, analyzing the results of the third hypothesis we have found mixed findings of the introduction of bike lanes near existing pogo stations on bus ridership. From our DiD analysis, we observe that there is a weak impact of intervention on ridership with a low R^2 value of .0278. Although, the post-intervention period demonstrated a significant effect with a p-value of 0, suggesting that the time after the intervention has an impact on ridership. This suggests that there is no significant impact on treatment groups and the combined effect of the treatment and post-intervention period. While looking at our generalized linear model, it reveals a significant increase in the number of bikers after interventions. This suggests that the intervention had a positive impact on the number of bikers. However, there was no significant difference between the experimental and control groups, implying that the changes om the experimental group itself were not sufficient to independently explain the increase of cycling. A Poisson Regression model shows a significant effect of the intervention, the overall fit was low suggesting that we need to consider other additional factors.

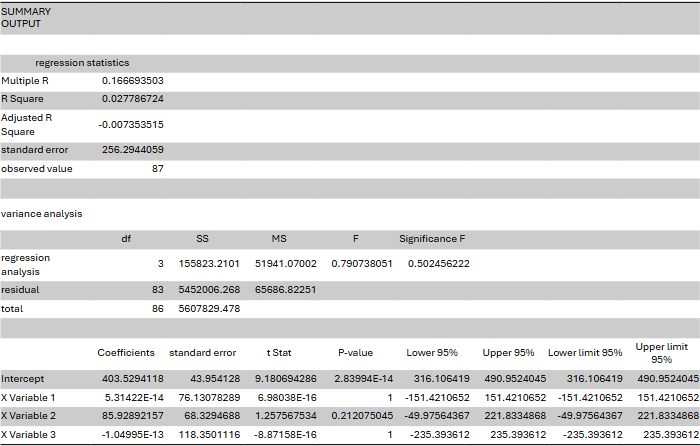
Last but not least, to verify the fourth hypothesis - Introduction of new bike stations increases POGOH ridership at nearby stations, we firstly generate a graph showing all the POGOH ridership data, which can roughly show that the overall ridership data for each station has increased after 4 stations having been added in June 2023 and 18 stations having been added in July same year. Then in order to validate our hypothesis more seriously, we draw all the stations on one graph according to their coordinates, by which we could figure out which stations specifically are the stations affected by nearby newly added stations. After we conclude a group of stations which are affected, we utilized a DiD analysis. The analysis results did show an increasing trend of station usage data, but not significantly (with a p value = 0.23). Then we chose a smaller group of stations, which is in the Oakland around University of Pittsburgh. For these specific stations, new stations had a significant positive effect on the data (with a p value nearly equal to 0). So our conclusion for hypothesis 4 would be: Generally speaking, new biking stations have some effect on nearby stations. Some of the specific stations (Oakland Pitt) are strongly affected – the usage increased significantly.

Sources

* <https://data.wprdc.org/dataset/pogoh-trip-data>
* (<https://data.wprdc.org/dataset/station-locations>)
* <https://data.wprdc.org/dataset/prt-monthly-average-ridership-by-route>
* <https://pogoh.com/blog/coming-soon-22-new-pogoh-stations/>
* <https://data.wprdc.org/dataset/prt-transit-stop-usage>
* <https://data.wprdc.org/organization/pogoh>
* <https://data.wprdc.org/organization/healthy-ride>
* <https://data.wprdc.org/dataset/shape-files-for-bikepgh-s-pittsburgh-bike-map>
* <https://data.wprdc.org/dataset/data-protected-bike-lanes-oct-2019>
* <https://data.wprdc.org/dataset/on-road-bicycle-pavement-markings>
* <https://pittsburghpa.gov/domi/map-hub>
* <https://pittsburghpa.maps.arcgis.com/apps/instant/interactivelegend/index.html?appid=b56a2a11f5a74f64948c3f214b207e50>
* <https://pittsburghpa.maps.arcgis.com/apps/instant/interactivelegend/index.html?appid=6ea6d6a573bc4994ae14276834e9cc1d>
* <https://data.wprdc.org/dataset/shape-files-for-bikepgh-s-pittsburgh-bike-map>
* [Pennsylvania's first protected intersections a feature of the new Allegheny Circle two-way conversion - BikePGH : BikePGH](https://bikepgh.org/2022/07/22/pennsylvanias-first-protected-intersections-a-feature-of-the-new-allegheny-circle-two-way-conversion/)
* [Pennsylvania's first protected intersections a feature of the new Allegheny Circle two-way conversion - BikePGH : BikePGH](https://bikepgh.org/2022/07/22/pennsylvanias-first-protected-intersections-a-feature-of-the-new-allegheny-circle-two-way-conversion/)
* [Downtown Penn Ave Upgrades | Complete Streets | Engage Pittsburgh](https://engage.pittsburghpa.gov/complete-streets/downtown-penn-ave-upgrades)
* [Pittsburgh bike infrastructure - Google My Maps](https://www.google.com/maps/d/u/0/viewer?mid=1UrHsHsb1UR7PD1AQPSKKbA-JH7s&ll=40.44378280687967%2C-79.99683046398778&z=17)

Appendix

For DID

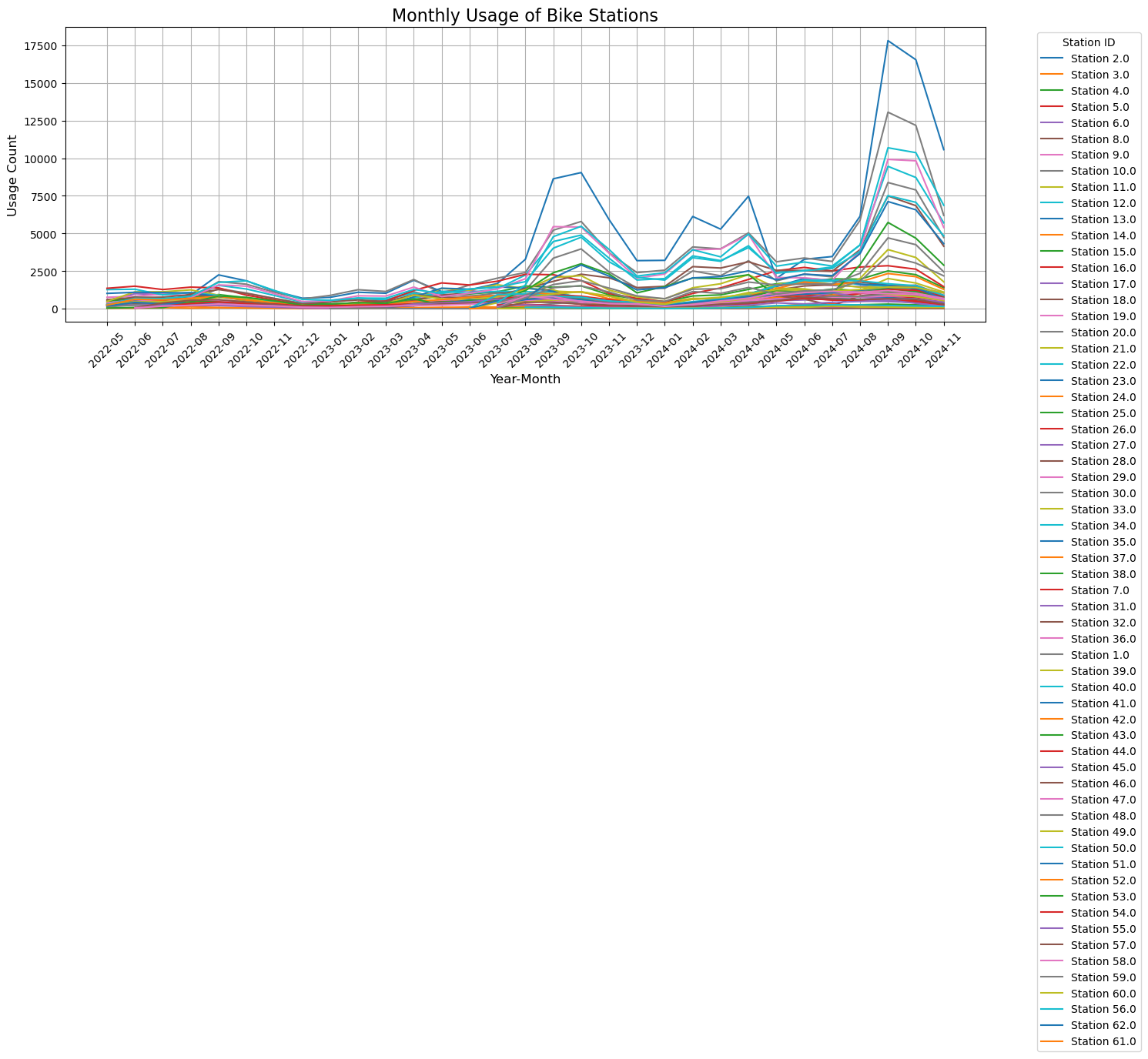
For GML

A screenshot of a computer

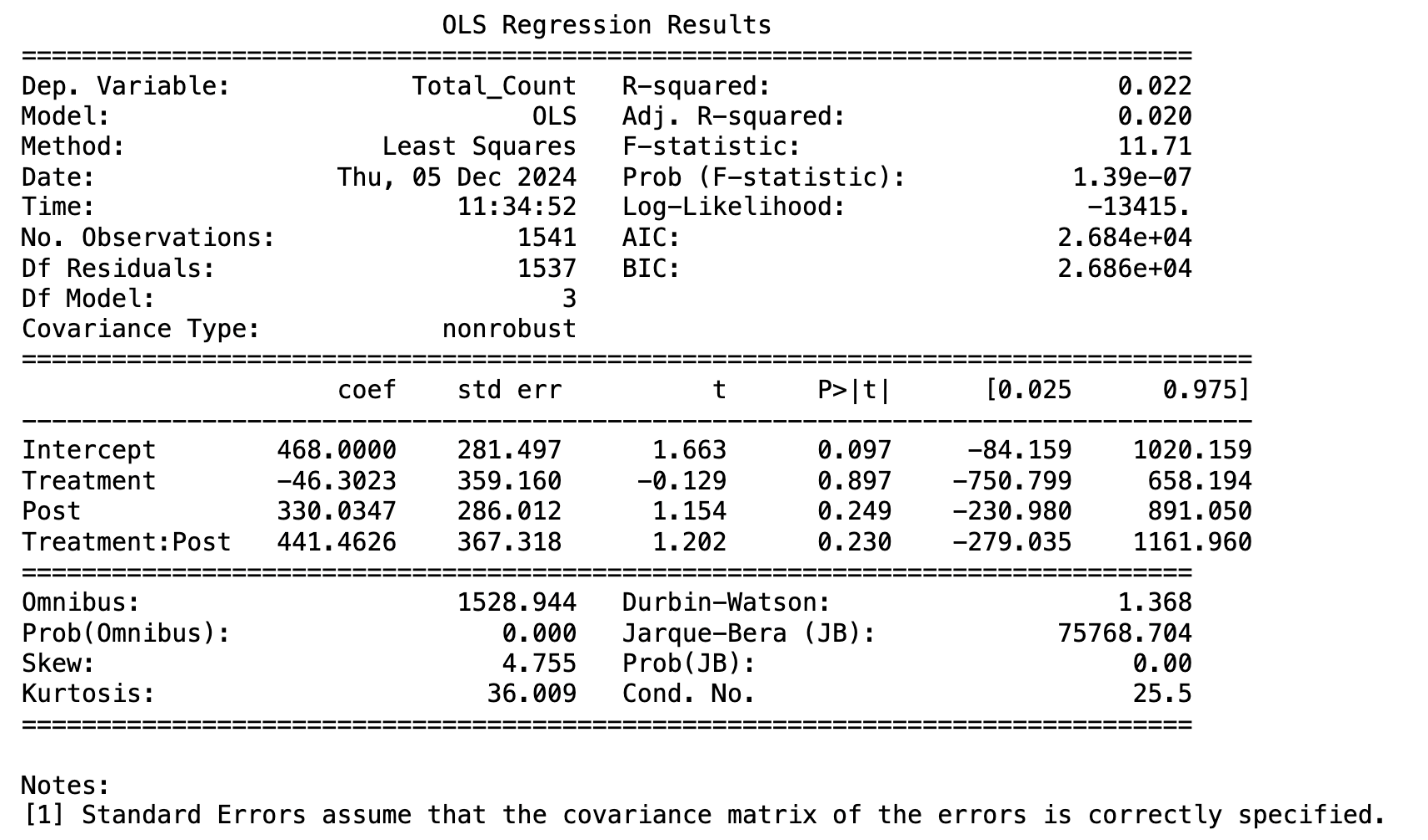
Description automatically generated

Hypothesis 4:

Overall monthly usage of bike stations:



The result for all the stations who has a new station nearby:



The result for the specific stations which are in Oakland near Pitt:

