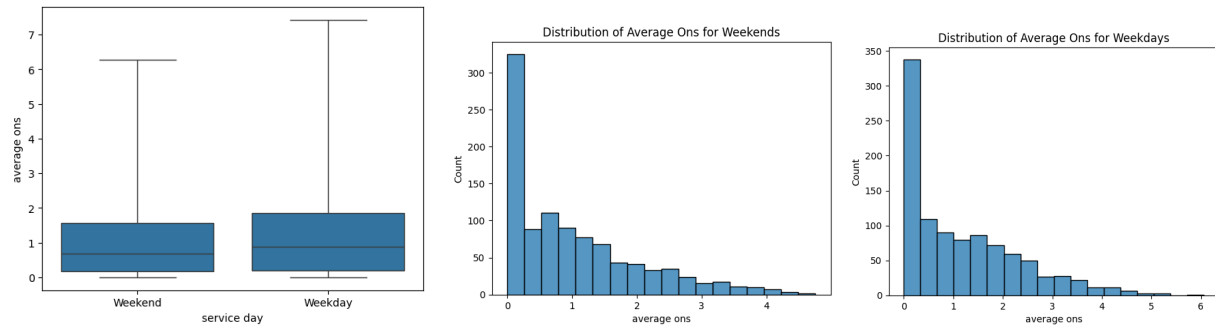


# Introduction

Public transportation usage changes throughout the week as people have different routines on weekdays and weekends. Factors such as work schedules, school, social activities, and daily responsibilities all influence when individuals decide to use public transportation. Understanding these patterns are important because it helps explain the demand for public transportation during different times. For this project, we are interested in specifically examining the bus usage in Pittsburgh on weekdays compared to weekends. Our research question is, do average bus boardings per stop differ significantly between weekdays and weekends in Pittsburgh, and is weekend service associated with significantly lower ridership?

## Data Description

To answer this question, we used a dataset from the Western Pennsylvania Regional Data Center which looked at monthly bus stop usage. Our analysis focused on the average number of boardings per stop during weekdays and weekends. No data cleaning or preparation was necessary as there were no missing or invalid values.

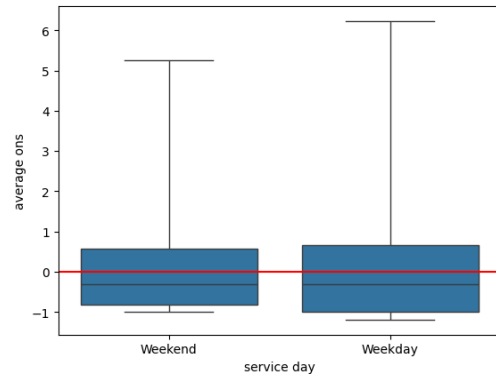


The graphs compare average boardings for weekday stops and weekend stops. The box plot shows that there is a higher usage of buses on weekdays compared to weekends. The histograms show how ridership is distributed on weekends compared to weekdays. Both graphs are heavily skewed to the right meaning most stops have very low boardings and only a small number of stops are heavily used.

## Methodology

To answer our research question we are using a regression model to predict if weekend service is associated with significantly lower ridership. The variable `avg_ons` represents the average number of boardings per stop.  $\beta_0$  captures the mean weekday ridership, and measures the change in ridership associated with weekends. The variable `weekend` is coded as 0 for weekdays and 1 for weekends.

$$\text{avg\_ons} = \beta_0 + \beta_1(\text{weekend}) + \epsilon$$



The box plot also shows the greater amount of onboardings due to its higher 4th quartile, but the residual line shows us that there is not a high amount of variance in the means of the two categories, weekend and weekdays.

## Results

The linear regression produced the following output.

	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.1928	0.005	233.775	0.000	1.183	1.203
service_day[T.Weekend]	-0.1891	0.007	-28.084	0.000	-0.202	-0.176

This model shows that the intercept coefficient ( $\beta_0$ ) is 1.1928 and the weekend coefficient ( $\beta_1$ ) is -0.1891. The intercept coefficient tells us that the log average number of weekday boardings is 1.1928. The weekend coefficient tells us that weekend service is associated with a decrease of 0.1891 log average boardings per stop compared to weekdays. Both p-values are 0.000, meaning the results are statistically significant.

## Conclusion

As we can see from the data and graphs, there is a greater number of onboarding riders on the weekdays than the weekends. While we can't pinpoint which factors this can be attributed to such as weather or whether there is a Pittsburgh game day or not within the dataset provided, we can still conclude that weekdays are busier than weekends for onboarding.

## References

<https://data.wprdc.org/dataset/prt-transit-stop-usage/resource/3f40b94b-4ac4-48f1-8c61-8439d2d2f420>