The Impact of Public Transit on Rideshare Demand in Austin, Texas

MGT 6203 Team 16

Jeffrey Finucane, Justin Heinzekehr, Jingya Ye, Ting Sit, Warren Ehrenfried

1.Overview of Problem Statement

The study is to analyze if there is an existing relationship between the number of rideshares versus public transportation services. Do rideshare services serve as a complement to public transportation or cannibalize public transit utilization? What motivates a passenger to switch mode of transportation?

The analysis leverages the dataset of rideshares in the Austin, Texas area and the ride volume from CapMetro, the main bus service agency in the region. The outcome of the analysis will be used to provide recommendations to CapMetro to improve bus services, which could help attract more passengers to take public transit instead of using rideshares.

The main benefits of the study are:

- Reduce greenhouse gas emission by shifting mode of transportation to public transit, which has a 69% lower emission (Anair et al., 2020) than rideshares services
- Mitigate traffic congestion as a result of reducing number of cars on roads from rideshare
- Achieve 50/50 mode of transportation goal for Austin

2.Literature review

Existing studies on the relationship among rideshare and public transportation is inconclusive. While the fear of rideshare service may replace public transit usage sounds logical, the substitution effect may not be completely valid (Hoffmann Pham et al., 2019). In fact, a complementary effect has been suggested, which provides opportunities of integration among the public transit and ridesharing systems (Zhang & Zhang, 2018). Cats et al. (2022) argues that both substitution and complementary effects could co-exist depending on the circumstance, including the personal monetary evaluation of travel time, service accessibility, and costs, that transition to a different mode of transportation may not be voluntary. This argument agrees with observations from Hoffmann Pham et al. (2019) study that a rise in Uber and Lyft services leads to a decline in Yellow and Green Taxi ride volume, but no obvious decline in subway usage.

Nonetheless, some aspects of rideshare service show strong substitution effect, although the effect is uneven across various demographic factors (Pan & Qiu, 2022). Quantifying those effect factors could become the blueprint for public transit agencies to improve service quality to make public transit options more desirable for passengers. Increasing the service quality of public transit services often leads to an increase in ridership (Erhardt et al., 2022), and we assume natural migration of trips from rideshare as public services improve. Our study focuses on identifying such factors for Austin residents to provide CapMetro ideas on what investment area they can prioritize. Examples of factors as suggested by research study includes spatial disparities, availability (Cats et al., 2022), poverty and unemployment rates, minority (Pan & Qiu, 2022), cost, travel time, and trip purpose (Hoffmann Pham et al., 2019).

One area that studies agree on is the environmental benefits of public transit, which serve as the common motivation for conducting the research. Anair et al., (2020) paper that looks at CO2 emissions confirmed that public transit does indeed provide a more environmentally friendly approach to transit compared to ridesharing.

3. Methodologies and approach

3.1 Exploratory analysis on rideshare dataset

Performing exploratory analysis on rideshare data to identify potential independent variables which may explain the passenger behaviors of using rideshare. Variables may include hour of day, day of week, weather condition, distance traveled, duration, location, and costs. For each variable, perform a statistical test to determine if the change in distribution is significant on each variable. Also, perform tests to confirm the independence among the selected variables.

3.2 Distance variable

Euclidean measurement is used to find the average distance of the n-nearest bus stops for each pick up and drop off location. Calculating the distance of the n-nearest bus stops to each rideshare pick-up and drop-off point to test if we can account for some variance in our regression model through this factor, although our rideshare coordinate data is coarse and may not be accurate enough to act as a key indicator on choice of mode of transportation. In order to achieve the goal, the distance measured in data is standardized to km; Any rideshare trip that is greater than 20 km is removed from the analysis by assuming such distance would be feasible to be traveled by bus to narrow down the scope. The data which have start or end lat/lon information missing is removed.

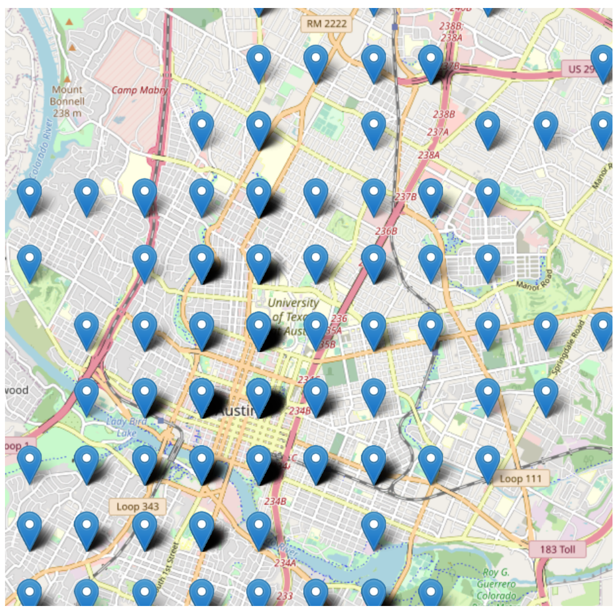

Figure 1 Rideshare coordinate precision is coarse

The rideshare dataset is only precise to 2 decimal places resulting in a grainy set of geospatial data that will decrease the quality of our analysis and distance calculations for nearest bus stops (Figure 3.2.1). At this level of coordinate precision, each rideshare pick-up location is up to a kilometer away from its actual, accurate location.

Also, the use of euclidean measurements are not ideal for these distance calculations as when traveling in a city, walking to a bus stop is rarely linear from beginning to destination.To calculate this distance measurement much more accurately, millions of walking routes would need to be calculated through the use of a navigation software which is out of the scope of the current project.

3.3 Exploratory analysis on bus ride dataset

Performing similar exploratory analysis on bus ride volume and coming up with similar distributions as rideshare on variables identified before. Also run the same set of statistical test on variable independence and significance.

3.4 Explore correlation between rideshare and bus rides volume

Identify if the distribution among rideshare rides and bus rides are similar or contrast each other, which will help inform whether potential relationships exist. Perform statistical tests to assess if there are significant correlations among the two ride volumes information.
If all the statistical tests are passed, a linear regression model can be applied, and the equation could be like this:

Rideshares ~ Bus Ride + independent variables identified in steps 1, 2, & 3

Once the model is established, a recommendation will be made on bus services that will factor in the independent variable and lead to an anticipated increase in bus rides. Rideshares volume will be reduced should the correlation be negative.

3.5 CO2 emission model for justification

The reduced rideshare volume will be converted to a CO2 calculation model to estimate the greenhouse emission reduction once compared to original CO2 consumption.

4. Experiments

4.1 Baseline set up

To set up a baseline of the impact of ride shares on the Austin area, CO2 emissions from the rides were analyzed. To get this data two data sets were combined. These data sets were the Austin rideshare data and the Vehicle Fuel Economy dataset from the US department of energy. The keys that were used to combine the data were year, make, and model of the ride vehicle

There were two issues that were encountered when merging the sources. The first was that the vehicle data contained multiple keys with the same year make and model. This was due to some cars having hybrid, electric, and gas models. To counter this error, a random sample was taken from the year, make, and model combo resulting in a random assignment for a specific ride. Although not perfect, this helped get an estimate of the actual emissions of the rideshare ride. The second issue encountered resulted from the rideshare ride's car not being present in the vehicle database. Since the data set was very large with 911,057 rides, excluding the NA values was the route chosen. This resulted in a new sample of 400,349 rides to estimate the data. The third issue was that hybrid modeled cars had two different values for CO2 emitted based on fuel type. Due to this if there was a value for two different fuel sources an average was taken. However, a majority of the rides had just a single value so that was used for calculation.

After cleansing the data, a total amount of CO2 emission was able to be calculated for the sample. This resulted in a value of 229,406 kilograms of CO2 emission. After volume adjusting back to our original sample size the rideshare data used in the regression below accounted for 408,058 kilograms of emission. Comparing that to the total CO2 emissions in 2017 ("City of Austin", 2021), our rideshare data counted for 1% of all emissions that year. Therefore a reduction through recommendations to consolidate rides would have a significant impact on the carbon footprint of Austin, Texas

4.2 Exploratory Analysis

4.2.1 Initial Hypothesis

1. Control on day of week, hour of day, and weather condition, **higher** bus ride volume lead to **lower** rideshare volume (negatively correlated).
2. For the same to and from location, the **longer** the difference in traveled time among rideshare and public transposition, the **higher** the rideshare volume.
3. The shorter the distance between a pickup/dropoff location from a bus stop, the **lower** the share ride volume.
4. The **fewer** rides on the road through rideshare the **less** fuel consumed and **less** CO2 emissions (positive correlation).

4.2.2 Initial Exploratory analysis

4.2.2.1 Rideshare dataset:

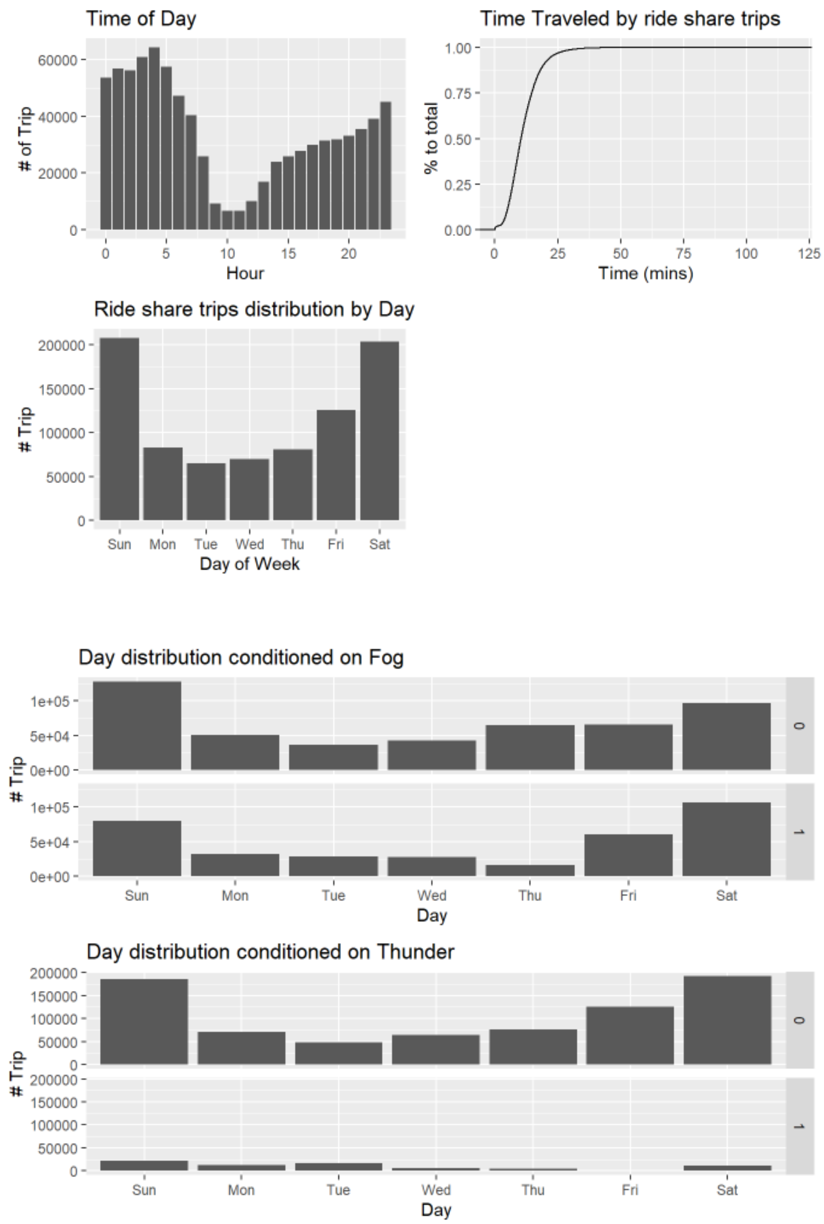

Figure 2 Exploratory analysis of potential significant variables on rideshare dataset

1. Rideshares tend to be more populated around early morning time and late evening
   - Hypothesis to Test: public transportation service hours impact passenger choice
2. Almost 90% of rideshare complete in < 25 minutes
   - Hypothesis to Test: travel time in public transportation impact passenger choice
3. Rideshare happen more frequently over weekend
   - Hypothesis to Test: public transportation frequency during weekend impact passenger choice

- Hypothesis to Test: rideshare over weekend is more effective in terms of time/cost relative to public transportation

4. There is no significant distribution with/without weather conditions from observations
   - Hypothesis to Test: Passenger preference to use rideshare is not impacted by weather conditions

4.2.2.2 Bus ride dataset

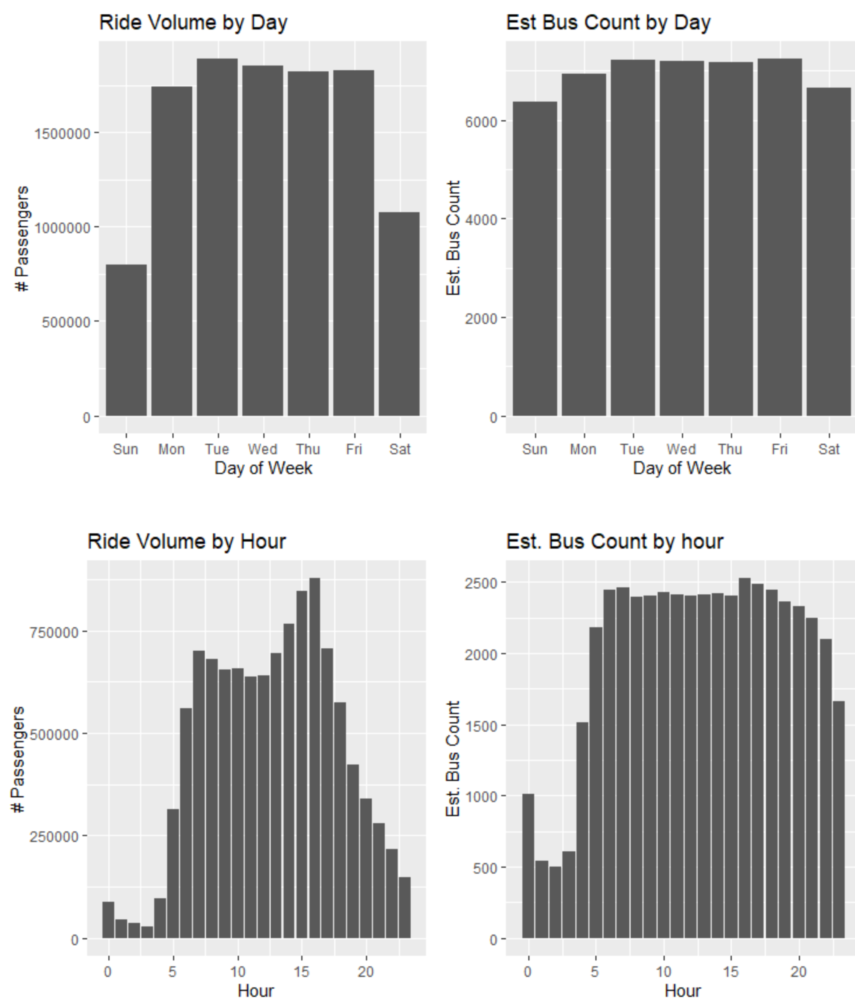

Figure 3 Exploratory analysis of potential significant variables on bus rides

1. Bus rides occur more often during weekdays vs weekend, even though the estimated number of bus count are not significant different among day of week.

2. Bus rides occur mostly from 7am to 4pm. Ridership drop drastically till mid-night even though bus frequency stay relatively the same throughout the day.
3. Minimal bus rides from midnight to 5am, probably driven by reduction in bus frequency.

5. Analysis and Findings

5.1 Data Cleansing and Transformation

Based on the dataset summary, rideshare trips that travel more than 11km will be removed due to assuming that those won't be replaced by public transit due to undesirable long distance. Also, remove trips that take longer 120 minutes, assuming such time duration or longer is not desirable for using public transit. All distance measures in this analysis will be standardized into kilometers (km).

Two new feature columns will be created. One is *wkd_en* to represent trips happening on a weekday or weekend. Weekend here is defined as Saturday and Sunday, the rest are Weekday. Another one is hour of the day (*hr_cat*), and will be grouped into 4 categories: hour 22 pm to 6am as 'darkout', 6am to 10am as 'rush hour', 10am to 17pm as 'work_hour', and 17pm to 22pm as 'evening'.

5.2 Statistical Testing

This section is to test on whether certain variables would have explanatory power to the target variable, which is *ride_vol.*

<u>Test weather condition such as Fog and Thunder would impact *ride_vol.*</u>
Kolmogorov-Smirnov (K-S) test is applied. High p-value suggest that fog and thunder <u>do not impact</u> *ride_vol*

Table 1 K-S test on weather condition Fog and Thunder

Fog	Thunder
data: no_fog$n and fog$n D = 0.28571, p-value = 0.9627 alternative hypothesis: two-sided	data: no_thunder$n and thunder$n D = 0.2381, p-value = 0.9627 alternative hypothesis: two-sided

Test whether *wkd_end* and *hr_cat* would impact *ride_vol.* Running a linear regression across those variables and check if coefficients are significant and R2 to be above 0.5.

```
Call:
lm(formula = ride_vol ~ wkd_end + hr_cat, data = rideshare_gp)

Residuals:
    Min      1Q  Median      3Q     Max 
-7768.4 -1765.2  -266.3  1928.4 13363.2 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       8899.7      614.8  14.475  < 2e-16 ***
wkd_endweekend    6610.6      648.8  10.188  < 2e-16 ***
hr_catevening    -4468.9      829.0  -5.390 2.43e-07 ***
hr_catrush_hour  -6689.0      829.0  -8.068 1.47e-13 ***
hr_catwork_hour  -6957.1      829.0  -8.392 2.19e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3799 on 163 degrees of freedom
Multiple R-squared:  0.5437,    Adjusted R-squared:  0.5325 
F-statistic: 48.56 on 4 and 163 DF,  p-value: < 2.2e-16
```

Both *wkd_end* and *hr_cat* variables are significant and R2 reached above 0.5, which will be included in the linear regression model.

Test whether bus volume (*bus_vol*), when is control on day of week, and hour of day, has significant relationship with *ride_vol.* Perform similar grouping on the CapMetro ride volume dataset to be the same as rideshare dataset by creating the *wkd_end* and *hr_cat* features with the same classification requirements.

Run Pearson test to check correlation among bus_vol & ride_vol

```
        Pearson's product-moment correlation

data:  final_df$bus_vol and final_df$ride_vol
t = -6.3726, df = 166, p-value = 1.762e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5573409 -0.3129403
sample estimates:
       cor 
-0.4433435 
```

bus_vol is negatively correlated with *ride_vol* by -0.44 and is significant

Add *bus_vol* to the linear regression model

```
Call:
lm(formula = ride_vol ~ bus_vol + wkd_end + hr_cat + start_grid_cell,
    data = final_df_1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3002.7  -369.2    58.3   267.9  8721.8 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         6.769e+02  9.292e+01   7.284 5.02e-13 ***
bus_vol            -4.329e-02  3.418e-03 -12.666  < 2e-16 ***
wkd_endweekend      6.190e+02  5.522e+01  11.210  < 2e-16 ***
hr_catevening      -3.676e+02  7.509e+01  -4.896 1.08e-06 ***
hr_catrush_hour    -3.898e+02  7.991e+01  -4.878 1.18e-06 ***
hr_catwork_hour    -3.459e+02  8.079e+01  -4.282 1.96e-05 ***
start_grid_cell2-B -2.046e+02  1.116e+02  -1.833 0.067053 .  
start_grid_cell2-C -3.946e+02  1.188e+02  -3.323 0.000912 ***
start_grid_cell2-D -6.452e+02  2.947e+02  -2.190 0.028677 *  
start_grid_cell3-A  1.772e+03  1.116e+02  15.880  < 2e-16 ***
start_grid_cell3-B  3.822e+03  1.434e+02  26.657  < 2e-16 ***
start_grid_cell3-C -4.412e+02  1.117e+02  -3.950 8.14e-05 ***
start_grid_cell3-D -5.326e+02  1.840e+02  -2.894 0.003852 ** 
start_grid_cell4-A -4.301e+02  1.156e+02  -3.720 0.000206 ***
start_grid_cell4-B -7.122e+01  1.111e+02  -0.641 0.521743    
start_grid_cell4-C -4.739e+02  1.130e+02  -4.193 2.91e-05 ***
start_grid_cell4-D -6.739e+02  1.020e+03  -0.660 0.509078    
start_grid_cell5-A -5.619e+02  1.213e+02  -4.632 3.91e-06 ***
start_grid_cell5-B -3.132e+02  3.939e+02  -0.795 0.426611    
start_grid_cell5-C -6.149e+02  7.240e+02  -0.849 0.395836    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1016 on 1623 degrees of freedom
Multiple R-squared:  0.5511,	Adjusted R-squared:  0.5458 
F-statistic: 104.9 on 19 and 1623 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = ride_vol ~ bus_vol + wkd_end + hr_cat, data = final_df)

Residuals:
    Min      1Q  Median      3Q     Max 
-7700.1 -1794.7  -298.4  2041.2 13080.3 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      8.981e+03  6.440e+02  13.947  < 2e-16 ***
bus_vol         -4.980e-03  1.137e-02  -0.438    0.662    
wkd_endweekend   6.576e+03  6.552e+02  10.036  < 2e-16 ***
hr_catevening   -4.306e+03  9.104e+02  -4.730 4.86e-06 ***
hr_catrush_hour -6.300e+03  1.216e+03  -5.182 6.45e-07 ***
hr_catwork_hour -6.492e+03  1.348e+03  -4.815 3.35e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3809 on 162 degrees of freedom
Multiple R-squared:  0.5443,	Adjusted R-squared:  0.5302 
F-statistic: 38.69 on 5 and 162 DF,  p-value: < 2.2e-16
```

The negative correlation persists, but the explanatory coefficient is not significant. It is NOT statistically significant that increase in *bus_vol* would lead to lower *ride_vol*.

5.3 Spatial binning by location for further investigation

In an attempt to find possible regions where there might be some statistical significance that an increase in bus volume would lead to lower rideshare volume, the city of Austin was divided into a grid. The city was split into 25-equal sized cells in a 5-by-5 grid layout and each subset of data (region) was iterated upon using a similar regression model but with each region acting as a new categorical variable.

One other benefit of this method relates to the coarseness of our coordinate accuracy for rideshare pick-up locations. Each cell contains locations specific to a certain region of Austin and the data should be more closely related or likely to show correlation with data points near it. In an area like downtown Austin, the density of bus stops are higher resulting in an average walking distance to the n-nearest bus stops to be more inaccurate than in a less dense area. This method should make up for some of those discrepancies in coordinate precision by using a unique coefficient for each cell.

By filtering out rides that fell well outside the main population of the Austin metropolitan area, the dataset contained more relevant data points of only rides that began within one of the 25

5-by-5 cells. The result shows that bus volume (bus_vol), has a negative relationship with rideshare volume (ride_vol) and is somewhat significant with an R-squared value remaining > 0.5.

5.4 Identifying possible areas for expansion

The regression equation above suggests that expanding public transportation use would make a small reduction in rideshare use, after controlling for area of the city, time of day and a weekend/weekday factor. The next step is to make a practical recommendation for expansion of Austin's public transportation services.

With the understanding that different areas of the city show distinct patterns of rideshare and public transportation usage, we gathered data from the 2020 US Census to help us make predictions about the feasibility of expansion in specific neighborhoods.

The Census' American Community Survey (ACS) provides us with the median household income and population density for each census tract (about the size of a neighborhood). We merged these variables into our ridership data by using the Census' geocoder (Census Geocoder,n.d) to connect latitude and longitude pairs to its census tract number. Both household income and population density turned out to be significant predictors of demand for public transportation. Another significant predictive factor was the variability of demand for transportation. In other words, for a given census tract, if ridesharing occurred at many different times of day, days of week, and starting locations, public transportation use tended to be higher than if ridesharing was used only during a few times, even after controlling for the total volume.

After combining rideshare volume (passgrs), population density (ppsm, or people per square mile), median household income (median_hincome) and the variety of transportation demand (n_var), we arrived at a model for predicting the number of public transportation rides that could be gained by expanding into specific census tracts. The model has good explanatory power, with an R2 value of 0.525.

```
Call:
lm(formula = rides_on ~ passgrs + ppsm + median_hincome + n_var,
    data = master_grp)

Residuals:
    Min      1Q  Median      3Q     Max
-254352  -31992   -7874   21262  491868

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     5249.4874 20410.2518   0.257  0.79730
passgrs            4.5259     0.4132  10.952  < 2e-16 ***
ppsm               6.4172     1.7971   3.571  0.00045 ***
median_hincome    -0.5114     0.1832  -2.792  0.00577 **
n_var            127.2927    28.6557   4.442  1.5e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 74510 on 192 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.5347,    Adjusted R-squared:  0.525
F-statistic: 55.15 on 4 and 192 DF,  p-value: < 2.2e-16
```

By applying this model to census tracts in which there is currently no public transportation access (as of January 2017), we generated a list of potential census tracts for expansion. This list shows all census tracts within a half mile of the current public transportation network, since it is likely not feasible to integrate locations further than a half mile away from the current infrastructure.

Table 2 23 census tracts that in the model that can make for expansion

	tract	ppsm	median_hincome	start_dist_from_bus	passgrs	rides_on	passgrs_per_ppsm	n_var	pred
1	215.14	10037.8	68284	2407.5052	25.2	0	0.002510510	17	37022.867
2	23.13	4041.5	48891	2645.1747	1400.0	0	0.346406037	158	32631.186
3	205.12	3858.3	49778	2052.4605	442.4	0	0.114661898	180	29468.400
4	207.04	6690.3	59617	931.4502	81.2	0	0.012136974	44	23663.830
5	205.11	7509.8	84201	1211.2101	109.2	0	0.014541000	57	18132.487
6	215.02	5778.7	55388	926.7219	18.2	0	0.003149497	13	15745.345
7	207.10	3181.5	71367	2075.4541	350.0	0	0.110011001	129	7174.957
8	215.15	8532.8	105022	2524.6908	5.6	0	0.000656291	4	6834.434
9	207.09	3849.1	78058	1821.6191	225.4	0	0.058559144	113	5436.816
10	22.15	674.9	50527	22101.8581	33.6	0	0.049785153	12	0.000
11	24.34	86.9	65395	27111.0174	7.0	0	0.080552359	5	0.000
12	109.20	294.6	71250	45831.0218	1.4	0	0.004752206	1	0.000
13	205.03	969.6	121281	1804.4589	257.6	0	0.265676568	113	0.000
14	207.07	5022.3	84027	2529.7110	43.4	0	0.008641459	21	0.000
15	208.10	2320.1	96528	31212.3249	2.8	0	0.001206845	2	0.000
16	215.09	2619.4	107069	1620.2491	26.6	0	0.010154997	17	0.000
17	215.16	641.3	73006	2315.3816	16.8	0	0.026196788	8	0.000
18	215.17	4262.4	86750	1232.6009	184.8	0	0.043355856	71	0.000
19	316.00	1632.9	118155	2421.6833	340.2	0	0.208340988	98	0.000
20	435.00	2127.5	75112	2304.4043	70.0	0	0.032902468	41	0.000
21	9508.03	458.6	83779	28807.6934	1.4	0	0.003052769	1	0.000
22	9508.04	134.4	95046	30735.8952	1.4	0	0.010416667	1	0.000
23	9508.05	280.1	73636	46217.5089	7.0	0	0.024991075	5	0.000

According to our model, 9 out of these 23 census tracts are candidates for expansion and would generate at least some new public transportation usage. Eight of those nine are located on the north boundary of the current service area (i.e., all the census tracts above 200), so this area represents the clearest opportunity to expand public transportation use. The location of these census tracts is such that all eight could likely be connected by one or two routes (see map below).

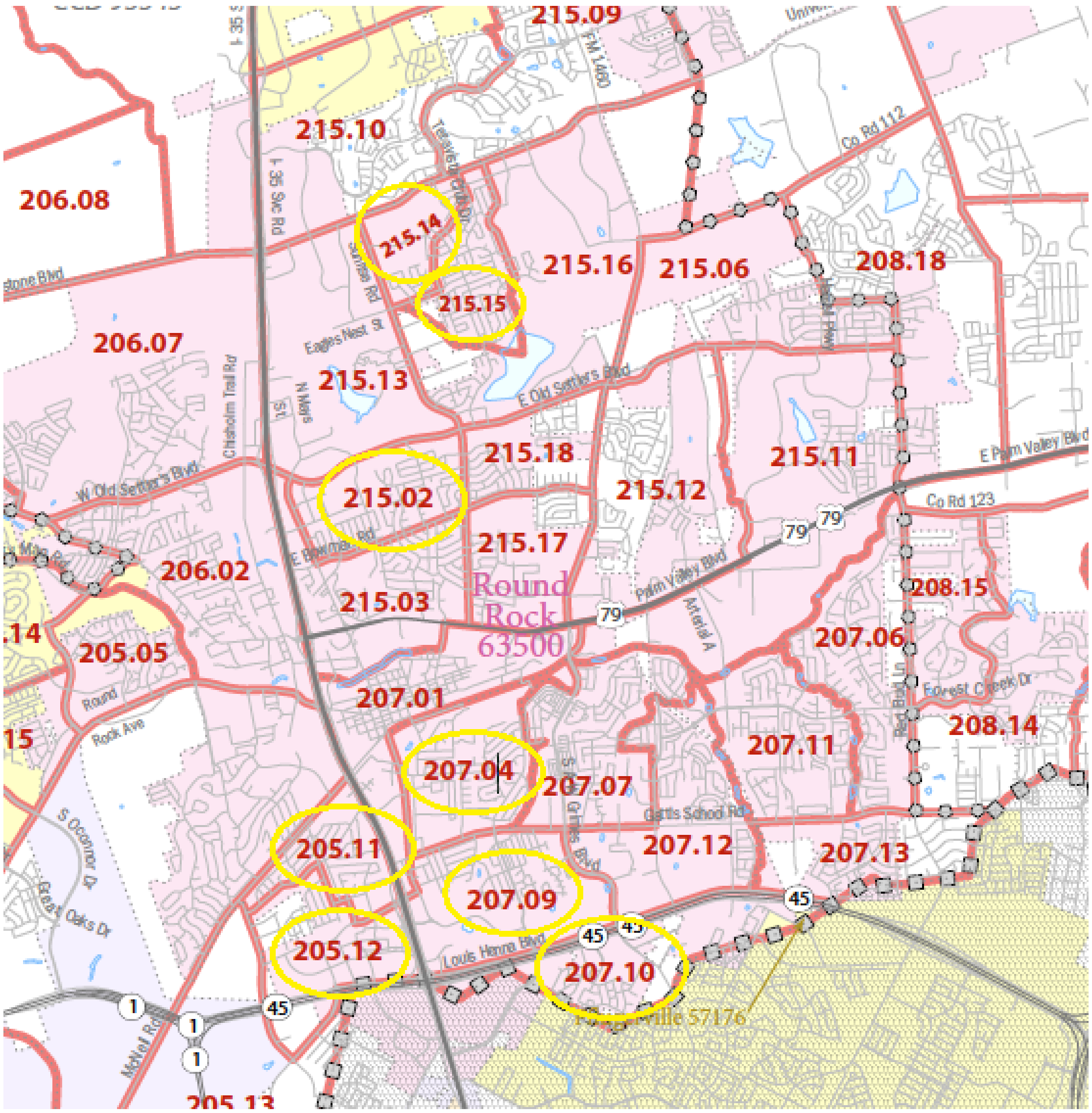

Figure 4 Nine tracts are the potential expansion

If CapMetro added one or more bus routes that served all eight of the census tracts identified, our model predicts that CapMetro would gain a total of 143,479 rides over a six-month period, or 286,958 rides per year. Assuming that we need at least one bus stop per census tract, and that the cost of constructing a new bus stop and associated amenities is between $35,000 and $50,000 (Go Durham, 2023), the cost of integrating these eight census tracts into the current public transportation system would be between $280,000 and $400,000.

To confirm that our recommendation aligns with common sense, we checked this against the map of Austin. The area we are recommending for expansion is a neighborhood called Round Rock, which includes several higher education institutions - Austin Community College, the Round Rock branch of Texas State University, and a health science center of Texas A&M University. This may explain why this area is particularly well suited for public transportation - students would represent a relatively dense population with low median income who would be likely to use public transportation if it was available.

Interestingly, we found further practical confirmation of our analysis after discovering that CapMetro did in fact expand in the Round Rock area in the year immediately following the time period of our dataset. CapMetro added four new routes across the Round Rock neighborhood as part of a five-year contract beginning in August 2017 (CBS Austin, 2017).

The majority of these public transportation rides will likely replace private car rides rather than rideshare use. According to the model developed in step 5.3, we can expect that an increase in public transportation use will decrease rideshare usage by about 4% of the public transportation rides - in this case, about 12,400 rideshare trips. The remaining 96% of the new public transportation use could replace private car use, replace walking or cycling, or represent new trips because of the convenience of the new public transportation options. The correlation between public transportation and private car use was outside the scope of our research project, but if we conservatively estimate that half of the new public transportation trips would otherwise have been taken by private car, then the total replaced rides would be 143,479 private car rides + 12,400 rideshare rides, a grand total of about 156,000 car trips replaced per year.

This proposal makes good business sense for CapMetro in terms of expanding market share in Austin. However, it is probably not a cost-effective way to reduce the overall carbon footprint of the city. We know from step 4.1 of our analysis that the average rideshare in Austin generates 0.57 kg of carbon (229,406 kg over a sample of 400,349 rides), or 0.0006 tons. Replacing 156,000 vehicle trips per year would therefore reduce the city's carbon emissions by 98 tons per year. The official current estimate of the social cost of carbon emissions is $51 per ton, although the EPA is proposing to raise this estimate to $190 per ton (Hersher et al., 2023). Even if we use that higher estimate of $190, the social value of the reduction in carbon emissions based on our proposal would be about $18,620 per year (98 tons of carbon x $190). At that rate, it would take 15 years of operation (with no additional maintenance costs) for the new bus routes to reduce carbon emissions enough that the social value would equal the cost of bus stop construction.

6. Conclusion

Regression analysis in this study suggests substitution effect of public transit versus rideshare usage of about 4% within Austin, Texas. Factoring demographic difference among districts within the region, 9 tracts are identified as good candidates for service expansion in terms of frequency and coverage, with 8 tracks concentrate on the northern region. Such expansion could lead to replacement of 156,000 car trips, which is roughly estimate to be ~$18,620. With expansion cost estimated to be $280,000 and $400,000, the investment breakeven period is 15 years.

Several challenges were identified throughout the analysis. Firstly, the imprecision of the coordinate data for the rideshare dataset are not sufficient to extract meaningful distances from bus stops. After calculating mean distance from n-nearest bus stops for each rideshare pick-up location, the coefficients found for these distances proved to be insignificant. Secondly, it didn't get to test due to limitation of data for demographic and consuming behavior, customer preference because of time constraints. Lastly, Austin is a growing metro city with increasing population and business activities which drive the user pattern of both public transit and rideshare services, but those impacts are not included in this study since data only include 1 year of history.

Suggest further studies to look at bus ride and rideshare relationship at a more micro level, with more precise coordinate information. Data related to customer preference on mode of transportation will be valuable to estimate the sensitivity between public transit and rideshare usage. As Austin is a growing city, the future study ride volume both on bus and rides datasets can be expanded from monthly throughout one year to a continuous annual studies to reflect the effect of growing population.

Data Sources

- Austin Ride Volume: This data set summarizes the total volume of rideshare on each day from June 4th, 2016 to February 7th, 2017. Number of rideshares will be the response to the model to be trained for. (https://data.world/andytryba/rideaustin)
- Rideshare Austin Data: This data set contains key information which includes geographic start and end locations of the rideshare, detailed vehicle list and weather condition for each registered rideshare. (https://data.world/andytryba/rideaustin)
- Vehicle Fuel Economy: This will help to get fuel consumption in both city and highway with CO2 emission on each specific vehicle. (https://www.fueleconomy.gov/feg/download.shtml)
- CapMetro ridership data: Dataset contains the number and starting location of public transportation rides from June 2016 through January 2017.
- CapMetro Shapefiles: Dataset contains the geographic locations of current public transit stops in the Austin CapMetro public transit city. (https://www.capmetro.org/destinations)(https://data.texas.gov/Transportation/CapMetro-Shapefiles-AUGUST-2017/5d4c-snum)

References

Anair, D., Martin, J., Pinto de Moura, M. C., & Goldman, J. (2020, February 25). *Ride-Hailing Climate Risks | Union of Concerned Scientists*. https://www.ucsusa.org/resources/ride-hailing-climate-risks(Anair et al., 2020)

Austin Motion. (2017). *2016 Mobility Bond program Overview And Implementation Plan* [Capital Planning]. Austin City Government. https://www.austintexas.gov/sites/default/files/files/Capital_Planning/2016_Bond/FINAL_2016Bond_ProgramOverview_ImplementationUpdate_backup_reducedsize.pdf

CBS Austin. (2017). "Round Rock to get CapMetro bus service in August." https://cbsaustin.com/news/local/round-rock-to-get-capmetro-bus-service-in-august

City of Austin Carbon Footprint (Number of Metric Tons of Carbon Dioxide Emissions from Our Government Activity)." *Data.Austintexas.Gov*, Sept. 2021, data.austintexas.gov/stories/s/HE-E-3-City-of-Austin-carbon-footprint-number-of-m/k8ja-qxaz/.

Erhardt, G. D., Hoque, J. M., Goyal, V., Berrebi, S., Brakewood, C., & Watkins, K. E. (2022). Why has public transit ridership declined in the United States? *Transportation Research Part A: Policy and Practice*, *161*, 68–87. https://doi.org/10.1016/j.tra.2022.04.006

Go Durham. (2023, March 16). *Story Map Shortlist*. https://triangletransit.maps.arcgis.com/apps/Shortlist/index.html?appid=5c5787e5a105433c904bd45fcb422332

Hersher, R., Ramirez, R., Scott, A., & Cirino, M. (2023, February 9). EPA's proposal to raise the cost of carbon is a powerful tool and ethics nightmare. *NPR*. https://www.npr.org/2023/02/08/1155544378/epas-proposal-to-raise-the-cost-of-carbon-is-a-powerful-tool-and-ethics-nightmar

Hoffmann Pham, K., Ipeirotis, P. G., & Sundararajan, A. (2019). Ridesharing and the Use of Public Transportation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4099122

Pan, Y., & Qiu, L. (2022). How Ride-Sharing Is Shaping Public Transit System: A Counterfactual Estimator Approach. *Production And Operations Management*, *31*(2), 906–927. https://doi.org/10.1111/poms.13582

Zhang, Y., & Zhang, Y. (2018). Exploring the Relationship between Ridesharing and Public Transit Use in the United States. *International Journal of Environmental Research and Public Health*, *15*(8), 1763. https://doi.org/10.3390/ijerph15081763