2025

How do Annual Members and Casual Riders use Cyclistic Bikes Differently?

A DATA ANALYSIS CASE STUDY

ANDRE MANNS

How do Annual Members and Casual Riders use Cyclistic Bikes Differently?

Executive Summary

Despite successful integration of its bike-share scheme in Chicago, it is believed Cyclistic profits could be improved by converting Casual Members (CM's) into Annual Members (AM's). CM's can be regarded as riders who purchase short-term passes expiring within 24 hours, while AM passes provide users with access for 12 consecutive months [1]. Before a new marketing campaign targeting CM conversion can begin, a greater understanding of CM and AM bike usage differences is required. The purpose of this report is to identify these differences and offer recommendations that can be used by Lily Moreno (Head of Marketing) and the Marketing Department to develop the new marketing campaign.

Cyclistic GPS data is used to compare member group mean ride durations, as well as ride share partitioned by duration. Next, preferred time of riding is compared, with investigations focusing on seasonal and weekday vs weekend member ride rates. Additionally, US governmental geo-spatial records are joined with company data to compare spatial and spatiotemporal member trends across Chicago's 77 communities. The data is cleaned and analysed using Google Big Query; statistical significance testing is performed in Microsoft Excel and visualisations are created in Tableau.

Significant differences between both memberships are observed for mean ride duration, seasonal ride rates, weekday vs weekend ride rates, community preference and community preference over time. From these findings, the following actions are recommended:

1. Concentrate annual membership promotions at bike stations in shoreline communities between Lake View and Loop over the weekend.
2. Increase CM ride rates in shoreline communities between Lake View and Loop over the weekend.
3. Increase CM ride rates for the summer months.
4. Offer discounted annual membership in the summer months.
5. Increase CM rates for rides greater than or equal to 20 minutes.

Difficulties encountered include detection of erroneous data caused by faulty GPS tracking and user-based error. These issues are amended by applying a scenario-led error mode filter. Additionally, the absence of user ID means individuals with 'extreme' riding habits cannot be detected. Fortunately, the impact of such outliers on the study findings is likely to be negligible due to large sample size.

Following evaluation of bike-share themed literature, user 'ride purpose' is identified as a potentially significant factor causing CM and AM bike usage distinctions. Further analysis is recommended to investigate the topic and assess its influence on Cyclistic user behaviour in Chicago. This could have benefits for future strategic decision making.

Table of Contents

Abbreviations

Abbreviation	Meaning
AM	Annual Member
CM	Casual Member
EUB	Extreme Upper Boundary
UQ	Upper Quartile
IQR	Inter Quartile Range
UTC	Coordinated Universal Time
SD	Standard Deviation
RD	Ride Duration
DRP	Daily Ride Percentage
CRP	Community Ride Percentage

Symbols and Notation

Symbol	Meaning
$\hat{Y}$	Estimated Value
df	Degrees of Freedom
X^2	Chi Square Value
V	Cramer's V
N	Sample Size
$P < 0.01$	Probability of result occurring by random chance is less than 1%
Δ	Change in Value
Σ	Sum of Value
X	Distance

Assumptions

Assumption	Explanation
Weekday	Monday to Thursday
Weekend	Friday to Sunday
Ride Location	Ride Start Coordinates
Ride Date and Time	Ride Start Timestamp
Time Zone	Central Time Zone

Section 1.0: Introduction

Since its inauguration in 2016, Cyclistic bikes has offered a comprehensive bike share scheme in Chicago; at the time of writing, over 580 stations and 5800 bikes are accessible across the city [1]. Additionally, flexible payment options mean users can choose to become Annual Members (AMs) and pay a fixed sum or enrol as Casual Members (CMs) and pay either a daily fee or at a minute-based rate. Together, this extensive metropolitan network and variable pricing structure, has enabled successful bike-share integration and a large customer base with nearly 5.5 million rides recorded in the past year (see Figure B. 1). Despite this success, new evidence suggests that AMs are more profitable than CMs, and in the year preceding April 2025, it was discovered that casual user rides still accounted for more than a third of all Cyclistic rides in Chicago (see Figure B. 1).

This has led to a strategic shift and Lily Moreno, head of marketing, has authorised the development of a new marketing campaign which aims to convert CMs to AMs. Before this can begin, an understanding of how these groups use Cyclistic bikes differently is required. The purpose of this report is to identify these differences, by using Cyclistic bikes GPS tracking information, as well as US Census Bureau and City of Chicago geographical data. Recommendations from this report will be used by Cyclistic Executives and the Cyclistic Marketing department to develop a new marketing strategy.

Before differences in AM and CM bike usage can be investigated, it is important to first define ways in which the two groups may diverge. In this report the dependent variables have been identified as ride duration, ride time and ride location. The study alternate hypotheses are as follows:

$H_1$1: Ride duration is different between AMs and CMs.

$H_1$2: Seasonal riding preference between AMs and CMs is different.

$H_1$3: Weekend and weekday riding popularity is different for AMs and CMs.

$H_1$4: AMs and CMs prefer to ride in different communities.

The report has been structured in a chronological format. In Section 2.0, a complete methodology including definition of the study parameters and use of resources is provided. Results are found in section 3.0. Section 4.0 discusses the significance of these findings with regards to the original hypotheses and identifies limitations of the study. Conclusions and recommended actions are stated in Section 5.0. Finally, References and Appendices can be found in Sections 6.0 and 7.0 respectively.

Section 2.0: Methodology

2.1: Study Framework & Data Sourcing

Determination of the study parameters was the start point for this report. These are shown in Table 1.

Table 1. Lists the study parameters. Note that electric scooters were omitted from the analysis in line with the scope of work.

Classification	Definition
Date	01/05/2024 – 30/04/2025
Location	Chicago (as defined by US Census Bureau Geographical Data)
Study Groups	Casual and Annual Members of Cyclistic bikes
Transport Modes	'Classic' and Electric Bikes

As its primary data source, the study used Cyclistic bikes GPS tracking data. Access to the data source was facilitated by its storage in an open access online data lake comprised of monthly data 'buckets' [2]. These buckets were transferred to google clouds big query and aggregated, in accordance with the study date parameters, to form an initial table: '0_0_1_raw_data'. Microsoft Excel was considered for the analysis, but its 1,048,576-row limit [3] made it unsuitable for the investigation.

Evaluation of the schema (see Table A. 1) revealed one limitation which was the lack of geographical context. Latitude and longitude coordinates, used in isolation, were not sufficient for determining if rides had started or ended within Chicago's city limits. Grouping rides, by distinct regions within the city was another difficulty. Resolution of the first issue was achieved by joining US Census Bureau geometric polygons for Chicago, which defined an outer city perimeter [4]. Regional distinctions were facilitated by integrating the 'City of Chicago Community Areas' dataset, which divides Chicago into 77 governmentally recognised communities [5].

2.2: Data Cleaning

Following transfer and storage in Big Query, raw data was passed through the cleaning pipeline detailed in Table 2. Definition of the ride duration 'extreme upper boundary' (EUB), used for anomalous data detection in cleaning step I-J, was achieved by substituting UQ and IQR values from table '1_5_1_outliers_lwr_removed' (see Table 3) into (1) to give (2):

$$EUB = IQR + (3 \times UQ) \tag{1}$$
$$3137 = 699 + (3 \times 1040) \tag{2}$$

Table 2. Chronologically details the data cleaning steps performed in Google Big Query (for the corresponding SQL code see Code Link D. 1). A cleaning step refers to the data cleaning process linking two consecutive table revisions: in data cleaning step 'A-B' for example, the input table is '0_0_1_raw data'. 221 rows were removed to produce '1_1_1_duplicates_removed'. Note that new tables were created following each cleaning step, even if no data had been removed. This was for two reasons: first to securely store backup copies; second, to clearly trace and communicate the data cleaning actions that had been performed.

Table Revision	Table Name	Row Count	Data Cleaning Method
A	0_0_1_raw_data	5,735,884	N/A
B	1_1_1_duplicates_removed	5,735,673	Duplicated results were identified by grouping ride id's. 221 rows from 31/05/2024 were removed.
C	1_2_1_nulls_lat_lng_removed	5,729,314	6359 rows were found to contain null values in at least two of the four coordinate fields (see Table A. 1). Additionally, ride durations inconsistent with the rest of the dataset (greater than 1000 minutes), were observed in 6353 of these same rows. Bike computer errors, network issues or user misuse were found to be probable explanations [6]. After concluding that the data was unreliable and could not be replaced, all 6359 rows were removed from the dataset. The impact of this is thought to be minimal since the omitted data accounted for 0.11% of the total row count.
D	1_2_2_nulls_name_id_removed	5,729,314	Null values in the station name or station id columns accounted for nearly 20% of the total row count. This can be attributed to users starting and/or ending their rides away from official docking points. All null values in these columns were replaced with 'unregistered_location' to reflect the start/end location and validity of the ride.
E	1_3_1_filtered	5,584,440	US Census Bureau geometric polygons were joined to the dataset (using ride start and end coordinates) to remove rides occurring outside city limits. A 1km 'buffer zone' was added to ensure that valid rides were retained. Additionally, UTC timestamps were converted to local time, and a filter was applied to ensure all rides started and ended within the study date parameters. Finally, 144337 'electric scooter' rides were removed to meet the parameters outlined in the Scope of Work document. Curiously, all electric scooter rides were recorded between 31/08/2024 and 30/09/2024 (see Figure B. 2).
F	1_4_1_trimmed	5,584,440	Misplaced starting or ending spaces were removed from all string data types using the 'trim' function.
G	1_4_2_double_spaces_removed	5,584,440	Double spaces in string data types were identified and replaced with single spaces, using the REGEXP_CONTAINS and REGEXP_REPLACE functions.

H	1_4_3_typos_sp_characters_removed	5,584,440	The 20 most repeated words from start and end station fields were added to a REGEXP_CONTAINS function to ensure correct spelling. Asterisks appearing at the end of station names were removed. Their inclusion resulted in 11 duplicate station entries, with these same stations also appearing without an asterisk. The purpose of their inclusion remains unclear. Lastly, for consistency, ‘/’ were replaced with ‘&’ symbols.
I	1_5_1_outliers_lwr_removed	5,498,195	86,245 rows with ride durations less 30 seconds were removed from the dataset. It was assumed that these had been caused by incorrect docking or bike computer error.
J	1_5_2_outliers_upr_removed	5,492,966	5 error modes were considered for rides at the upper end of the ride duration distribution. Rides that satisfied the criteria listed below were removed from the dataset. The results are illustrated in figure 1. 1. *‘Timed out’ or automated rides:* First, rides were filtered by duration to leave only those greater than 10799 seconds. Durations were then divided by 3600 (1 hour) to 4dp and the integer was removed. Rides falling within 0.015 of an exact hour, were excluded from the dataset. 2. *Unsuccessful docking or sessions left active after use:* First, rides were filtered to leave only durations greater than 21599 seconds. A logical argument was implemented to exclude rides with a combination of unexpected start times, end times and duration. Rides were retained if they started after 05:59:59, finished before 00:59:59 the following day and had a ride duration less than or equal to 27799 seconds. Rides with durations greater than 27799 were retained if they started after 05:59:59 local time and finished the same day. 3. *Rides activated by overnight maintenance:* Exclusion criteria required that rides started and ended on different days, and that ride duration exceeded 10799. A count was applied whereby rows were grouped by start date and start location. Groups containing 2 or more rows, were removed. 4. *Rides affected by network issues at the end of a trip:*

			Rides classed as 'extreme upper end outliers', see (1) and (2), were grouped by end time and end location. Groups containing 2 or more rows were removed. 5. *Rides affected by network issues at the beginning of a trip:* The same approach to that used in the previous error mode, but here rides were grouped by start location and start time.
K	1_5_3_outliers_speed_removed	5,485,233	Average ride speed was calculated using timestamp and coordinate data. Rides with average speeds greater than 35 Km/hr were removed.
L	1_6_1_quality_check	5,485,233	It was found that the 'Griffin Museum of Science and Industry' was also entered as the 'Museum of Science and Industry'. Further reading revealed a name change in May 2024 [7]. All entries were updated to the 'Griffin Museum of Science and Industry'.
M	2_1_1_format	5,485,233	Column headings were amended and rearranged into a more user-friendly format (see Table A. 2).
N	3_1_1_analysis	5,485,233	The format was checked and approved.

2.3: Data Validation

Validation of the data cleaning process was achieved by measuring ride duration dispersion and central tendency for each step of the data cleaning pipeline (see Table 3). A 67.52% reduction in Standard Deviation (SD) between tables 0_0_1_raw_data (SD = 3513) and 3_1_1_analysis (SD = 1141) aligned with pre-cleaning expectations that removal of erroneous data would result in increased data consistency. A right skew in ride duration distribution was observed by comparing mean (mean = 896) and median (median = 590) ride duration in the final dataset. This compared favourably with the duration distribution of other bike-share schemes in Boston and Shanghai [8], [9].

Table 3. A statistical summary of ride duration for all users at each stage of the data cleaning pipeline (see Table 2). Note that a prerequisite for statistical analysis was row removal, hence some tables from the data cleaning pipeline are not listed. For the corresponding SQL code, see Code Link D.2.

Table Name	Mean [s]	Median [s]	IQR [s]	SD [s]	Min [s]	Max [s]
0_0_1_raw_data	1013	577	694	3513	-164899	93595
1_1_1_duplicates_removed	1013	577	694	3504	-164899	93595
1_2_1_nulls_lat_lng_removed	914	576	692	1872	-164899	89997
1_3_1_filtered	921	580	697	1892	-164899	89997
1_5_1_outliers_lwr_removed	936	590	699	1902	30	89997
1_5_2_outliers_upr_removed	895	589	697	1140	30	61979
1_5_3_outliers_speed_removed	896	590	697	1141	30	61979
3_1_1_analysis	896	590	697	1141	30	61979

Additionally, cleaning steps were validated through graphical comparison. The histogram in figure 1 compares ride count, categorised by minute-long duration intervals, before and after cleaning step I-J. Figure 2, shows the impact of cleaning steps I-N on average station durations.

Figure 1a. 1_5_1_lwr_outliers_removed

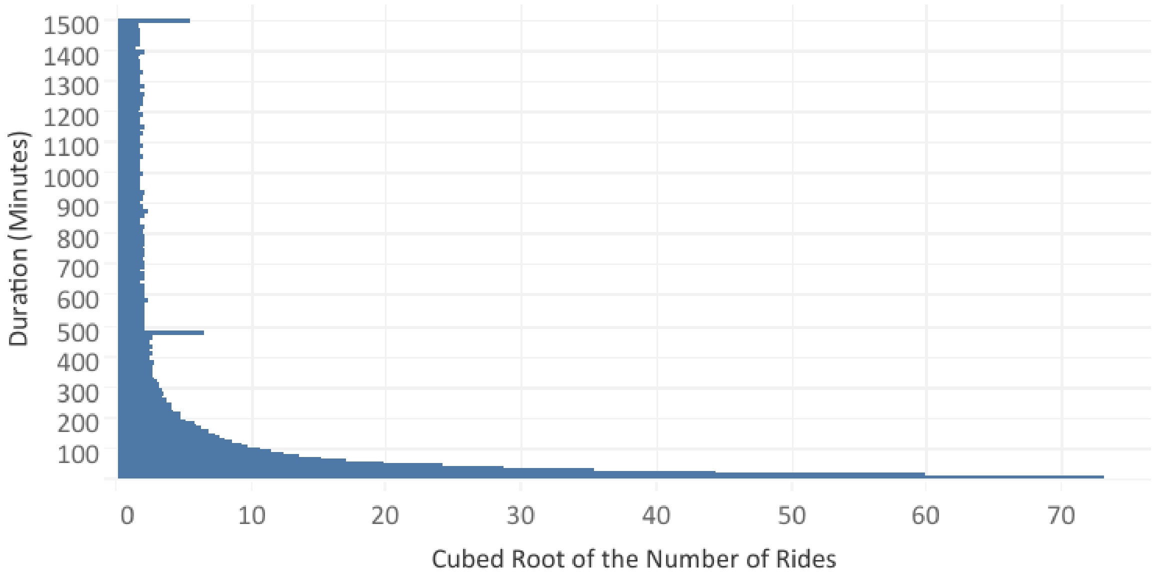

Figure 1b. 1_5_2_upr_outliers_removed

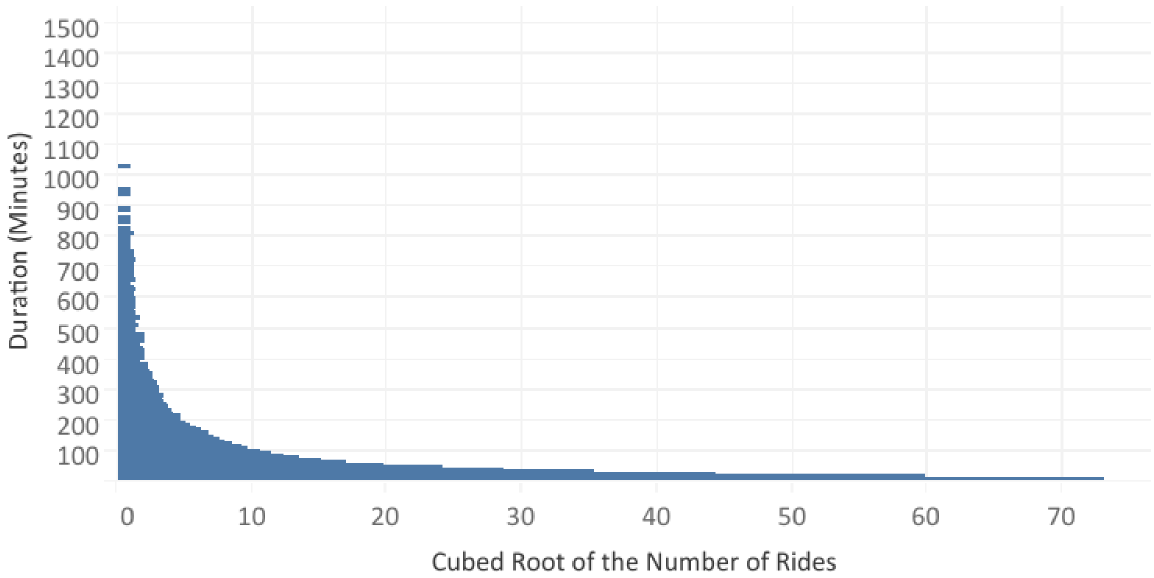

Figure 1. Illustrates the relationship between ride duration and ride count before and after cleaning step I-J (see Code Link D. 3 for corresponding SQL code). Rides were categorised by minute, and a cubed root function for ride count was used to compensate for the long-tailed distribution. Inconsistent ride counts were observed for 480 and 1500-minute-long rides in figure 1a ($Ride\ Count_{480} = 272$*,* $Ride\ Count_{1500} = 161$*), which when combined accounted for more than 10% of all rides greater than 479 minutes. Figure 1b shows the resulting ride count distribution following cleaning step I-J. 1500 minute and 480-minute ride groups were omitted completely, with maximum ride duration falling by 31% (*$Ride\ Duration_{MAX} = 1033\ [m]$*). No obvious irregularities in the distribution were observed. In total, 5229 rides were omitted, representing a 0.095% reduction in sample size.*

Figure 2a. 1_5_1_lwr_outliers_removed

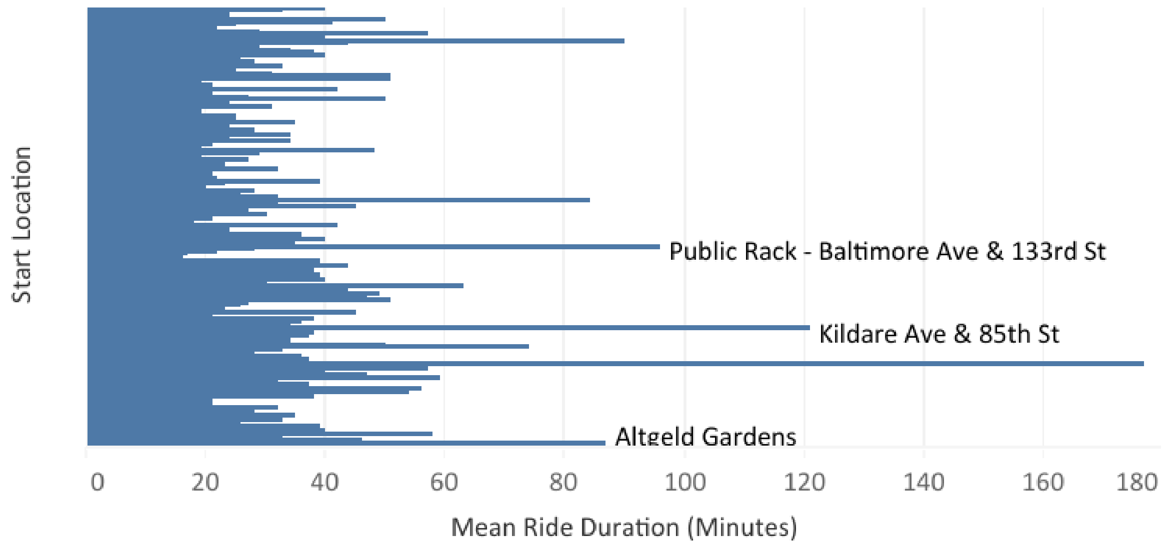

Figure 2b. 3_1_1_analysis

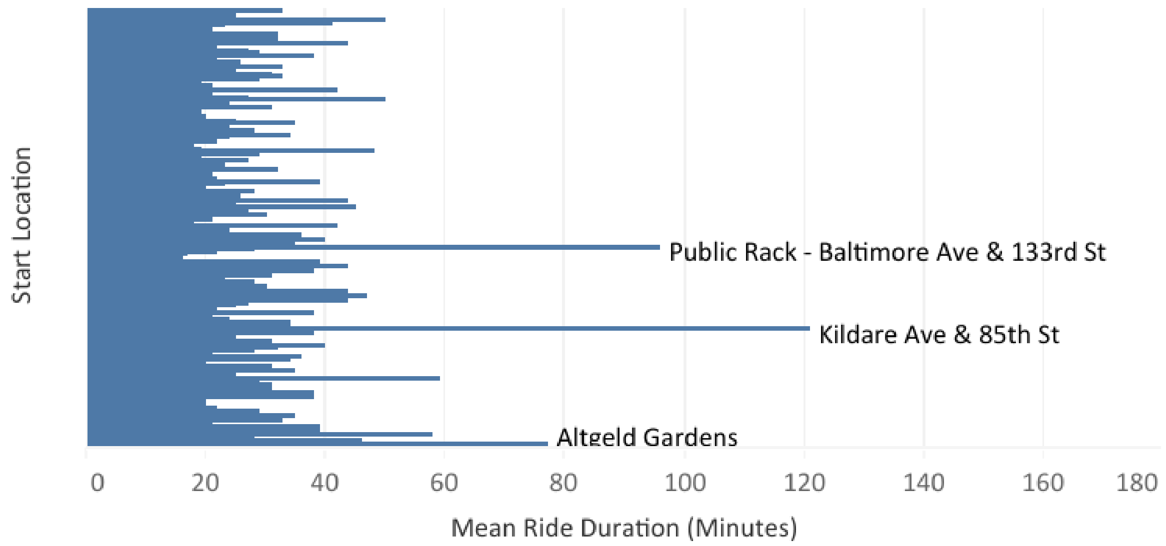

Figure 2. Shows the effect of cleaning steps I-N on mean ride durations associated with each start station (see Code Link D. 4 for corresponding SQL code). The number of stations averaging durations greater than 40 minutes fell by 42%, while a 67% reduction was observed for stations averaging over 50 minutes. Unusually high averages at 'Kildare Ave & 85th St' and 'Public Rack – Baltimore Ave & 133rd St' were justified by their only being 1 and 2 rides leaving from these stations respectively. 40 trips were made from 'Altgeld Gardens' (see Figure 2b). These trips were reasonably spread through the year, with only November, December and January registering no rides. Therefore, ride activity from this station was deemed to be genuine, and no data was removed.

2.4: Analysis Methods

All metrics and data required for investigating the report hypotheses were obtained by querying table 3_1_1_analysis in Big Query and transferring the output to Excel for statistical analysis. The significance level for all tests was set at $p < 0.01$. The statistical method for assessing each hypothesis is outlined below:

$H_1$1: Rides were partitioned by membership type, and difference in mean ride duration was evaluated using a t-test (see Code Link D. 5). Although not typically used for skewed distributions, a t-test was deemed suitable in this case due to the very large sample size [10]. For corresponding t-test formulae see (C. 1), (C. 2); confidence interval calculations are shown in (C. 3), (C. 4), (C. 5) and (C. 6). Later, rides were categorised by duration into discreet intervals. A chi-square test was used to determine if member ride patterns differed significantly. The Chi-square formula is shown by (C. 7). Ride counts and accompanying chi-square values are found in Table A. 3.

$H_1$2: CM and AM annual ride trends were visually compared by plotting daily ride count percentage against ride start date on a line graph. In a second analysis, ride counts were grouped according to membership type and season. A chi-square test was used to determine if seasonal disparities were significant and combined with a Cramer's V assessment to measure the effect size of any such relationship. Ride counts and accompanying chi-square values are found in Table A. 4. The Cramer's V equation is shown in (C. 8).

$H_1$3: Member ride count percentages were categorised by day of the week and compared in a grouped bar chart. Next, ride counts were summated for each member group and divided into weekday and weekend categories. A Chi-square assessment was used to determine if member weekday and weekend ride counts differed significantly and paired with a Cramer's V analysis. To compensate for sampling bias (caused by an unequal share of days), weekdays were normalised relative to weekend days. Associated ride counts and statistical values are shown in (Table A. 5)

$H_1$4: Following a 'negligible' Cramer's V score when comparing start location and end location ride count by community (see Table A. 6), all geo-spatial comparisons were performed using start location only. Rides were grouped by community and membership. A geospatial heatmap was used to visually identify dense regions of cycling activity and compare membership group user behaviours. Additionally, a ranked bar chart displaying the 20 most ridden communities, enabled community ride count scores to be directly compared. A Chi-square test was used to determine if community ride count distributions were significantly different between member groups and accompanied by a Cramer's V assessment. Community ride counts and chi-square values are shown by in Table A. 7.

Section 3.0: Results

3.1: Ride Duration

The analysis set out to determine how ride duration differed between casual and annual members. Mean ride duration was found to be higher for CMs (mean = 1211, SD = 1552, $t(5485233) = 417$, $p < 0.01$) compared to AMs (mean = 719, SD = 770) suggesting that CM ride duration was, on average, significantly longer than AM ride duration (99% CI 489 to 495). To explain this finding, ride duration was divided into discreet time intervals, with ride counts summed and proportioned for each membership. The results are displayed in Figure 3.

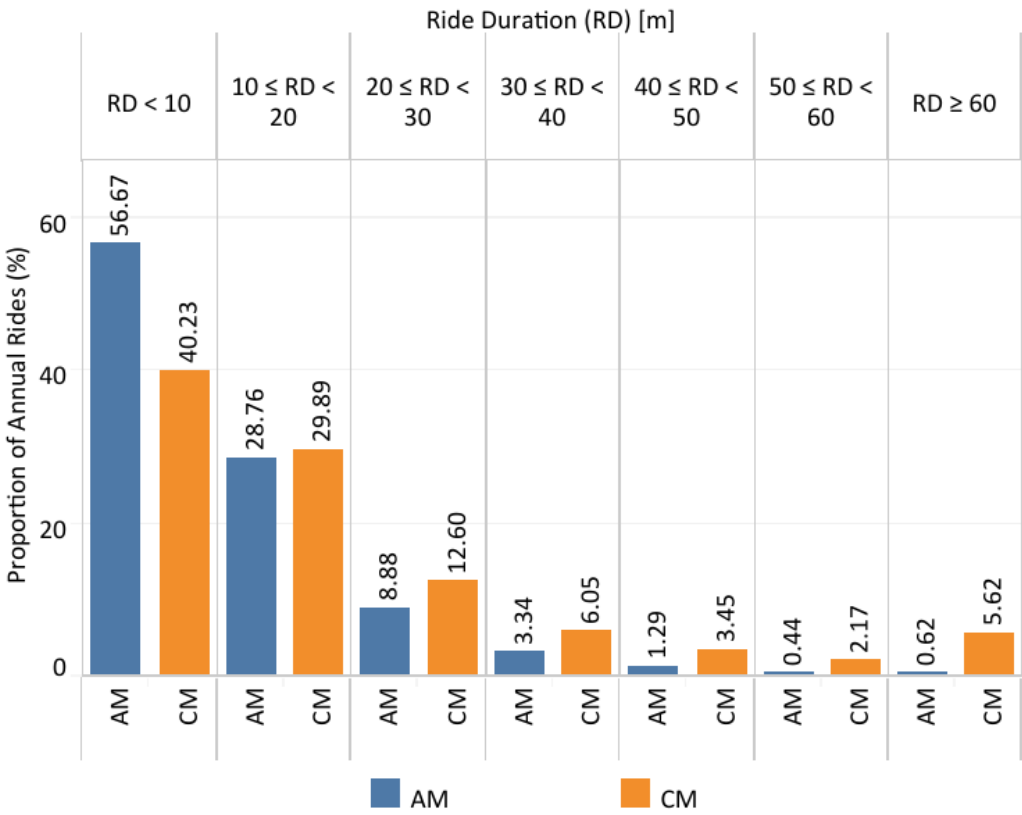

Figure 3. Illustrates grouped ride duration percentages for AMs and CMs. Percentages were calculated by dividing grouped counts by AM or CM population size. See Code Link D. 6 for the corresponding SQL code.

A right-skew was observed across both memberships, with proportional ride rates peaking for rides lasting less than 10 minutes. Skewness was more severe for AMs, with more than half of all rides (56.67%) lasting less than 10 minutes. This was comparatively higher than CM ride share for the same duration category (40.23%). Ride duration variability was greater for AMs (mean = 100/7 = 14.29%, SD = 21.22pp, range = 56.23pp) than it was for CMs (mean = 100/7 = 14.29%, SD = 14.87pp, range = 38.06pp). Similar ride rates were seen for rides between 10 and 20 minutes long (casual = 29.89%, member = 28.76%), but rides lasting longer than 20 minutes were considerably more popular amongst CMs (15.31pp) supporting earlier t-test results, that casual members preferred longer rides. A chi-square test of independence found there to be a significant difference between both member groups (X^2 (df=6, N = 5485233) = 300917.901, $p < 0.01$)) validating earlier t-test results, that ride duration differed significantly.

T-test and chi-square findings allowed for rejection of the null hypothesis ($H_0 1$): that trip duration between CMs and AMs is not significantly different.

3.2: Ride Time

The aim was to determine if CM and AM riding tendencies differed over the course of the year. Figure 4 displays Daily Ride Percentages (DRP's) for each group over this period. Evaluation of the line of best fit revealed a similar sinusoidal relationship between percentage ride occurrence and time of year for both member groups. Greater annual variation in DRP was observed amongst CMs with a higher peak in July ($\hat{y}_{CM}$= 0.53%, $\hat{y}_{AM}$ = 0.43%) and lower general level in February ($\hat{y}_{CM}$ = 0.02%, $\hat{y}_{AM}$ = 0.12%). Additionally, CM DRP was generally higher than that of AMs between May and late October, but lower during the other months. Daily fluctuations were greater for CMs between March and October (see Table 4).

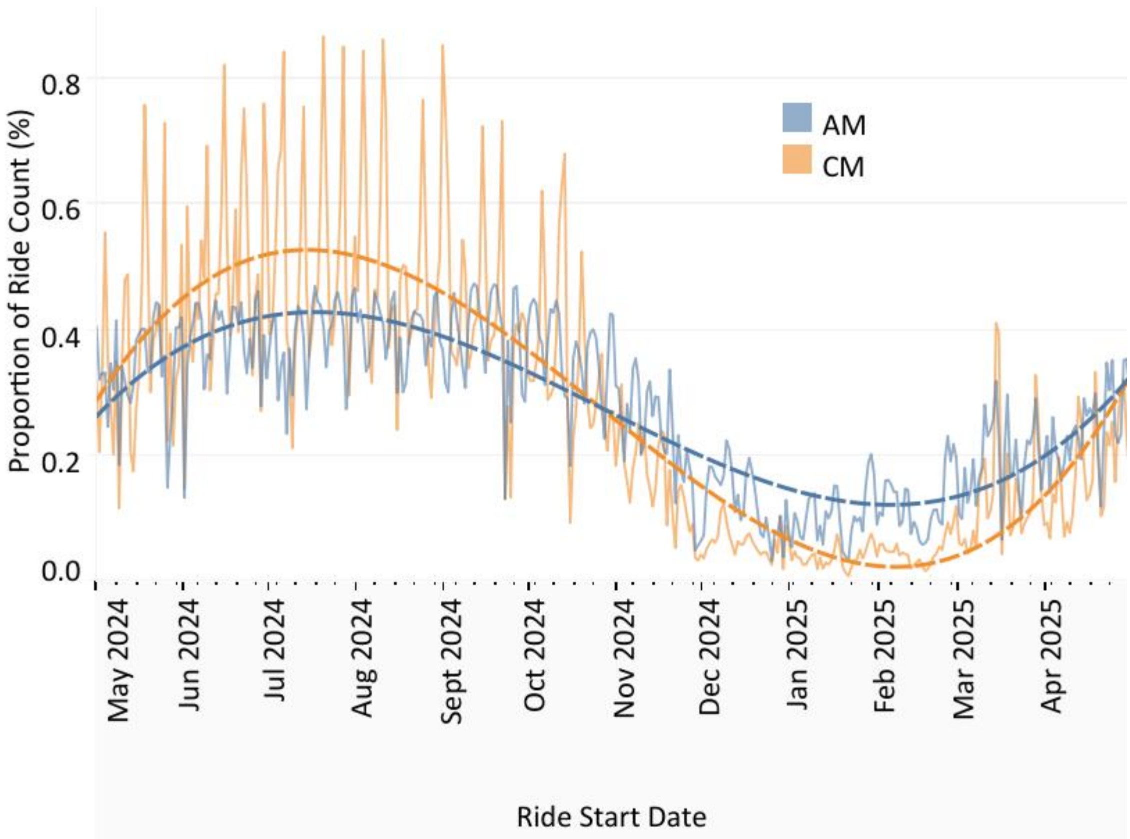

Figure 4. Shows DRPs for AMs and CMs between May 2024 and April 2025. Note the 'cross-over' point in late October, after which DRP is generally higher for AMs. See Code Link D. 7 for the corresponding SQL code.

Table 4. Shows the SD for CM and AM ride count percentages. Note that greater CM variability is evident between March and October. This is also evidenced visually, by the larger peaks in Figure 4.

SD of Monthly Percentage Ride Count (PP)												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CM	0.016	0.024	0.095	0.078	0.151	0.155	0.166	0.166	0.145	0.138	0.071	0.024
AM	0.043	0.049	0.063	0.061	0.072	0.069	0.063	0.047	0.081	0.067	0.083	0.051

From these findings, it was proposed that seasonality and day of the week affected CM DRP more than AMs. To investigate, rides were partitioned by membership type and season (Figure 5), as well as membership type and day of the week (Figure 6).

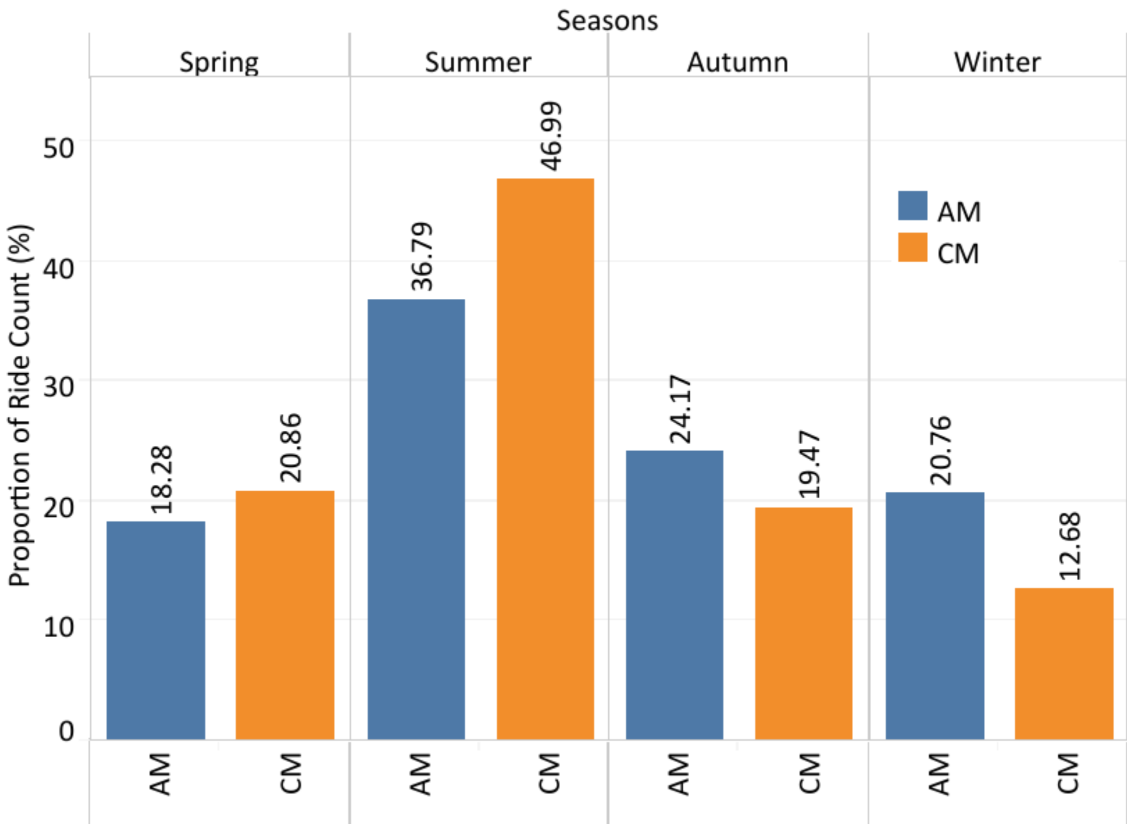

Figure 5. Shows seasonal ride count percentages by member type. See Code Link D. 8 for the corresponding SQL code.

The proportion of CMs riding in Summer exceeded that of AMs (10.20pp), marking the biggest seasonal percentage difference. Spring had the most comparable ride rates (2.58pp), while AM ride percentage was higher for the remainder of the year. Seasonal dispersion was higher for CMs (mean = 100/4 = 25%, SD = 15.09pp, range = 34.31pp) compared to AMs (mean = 100/4 = 25%, SD = 8.22pp, range = 18.51pp) aligning with the annual fluctuations observed in Figure 4. A chi-square assessment revealed that member seasonal riding habits were significantly different $X^2(df=3, N = 5485233) = 95491.328$, $p < 0.01$), although a 'small' Cramer's V effect size (V = 0.13) indicated that factors, other than seasonality, were likely implicated [11]. The findings allowed for rejection of the null hypothesis ($H_0 2$): that seasonal riding tendencies between AMs and CMs is the same.

To compare member weekend and weekday riding habits, annual ride counts were partitioned by day of the week and converted into percentages (see Figure 6). Average daily ride share was higher for CMs over the weekend ($mean_{CM}$ = 17.61%, $mean_{AM}$ = 12.90%) with casual ride percentage peaking on Saturday (20.80%). Conversely, AMs preferred to ride during the working week ($mean_{AM}$ = 15.33%, $mean_{CM}$ = 11.80%), with ride share peaking on Wednesday (16.41%). Greater ride variability was seen in CMs over the course of the week (mean = 100/7 = 14.29%, SD = 3.56pp, range = 10.18pp), with AMs demonstrating more consistent riding habits (mean = 100/7 = 14.29%, SD = 1.89pp, range = 5.65pp).

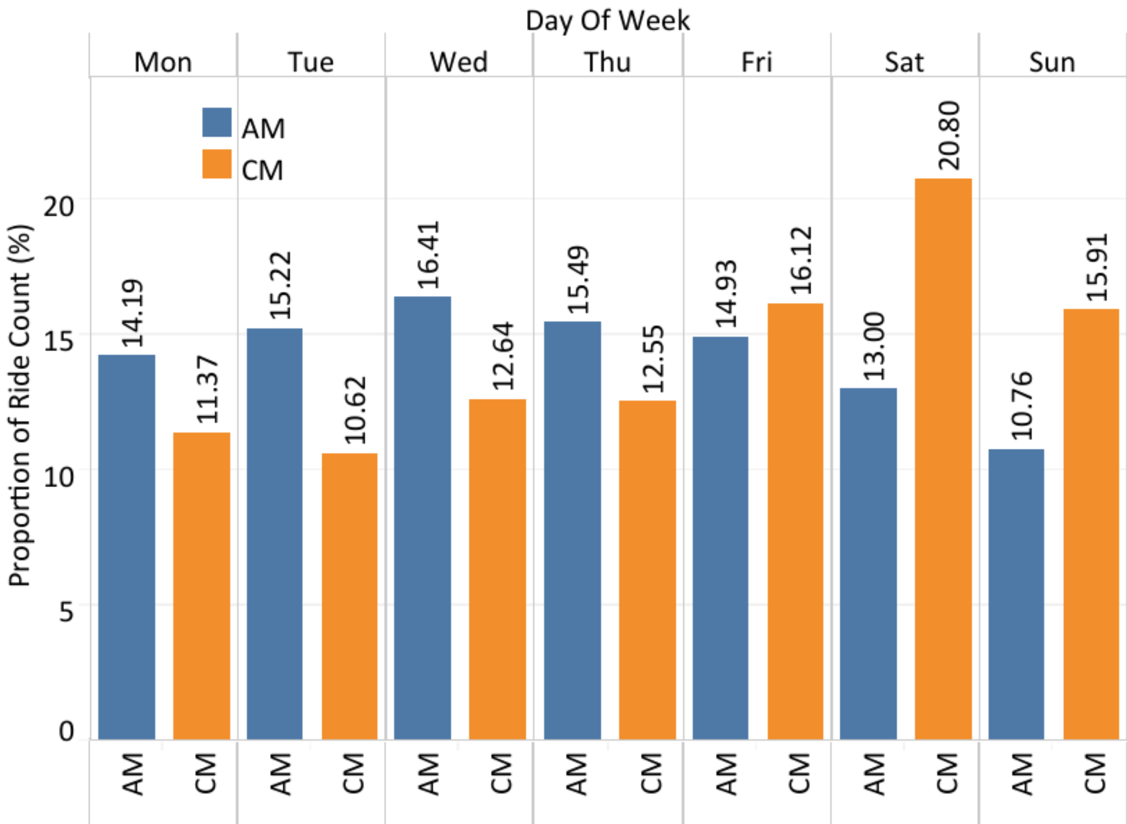

Figure 6. Shows AM and CM percentage ride counts partitioned by day of the week. Note that AM ride share is higher on all weekdays, but CM ride share is higher on all weekend days. See Code Link D. 9 for the corresponding SQL code.

Following analysis of Figure 6, both memberships were divided into 'weekday' and 'weekend' groupings. A chi-square test of independence, found there to be significance between the two member groups $X^2(df=1, N = 5485233) = 88254.267, p < 0.01$). The subsequent Cramer's V score (V = 0.14) suggested there was a small influence of weekday or weekend timing on differing ride rates [11].

These findings allowed for rejection of the null hypothesis ($H_0 3$): that weekend and weekday riding popularity is the same for CMs and AMs.

3.3: Ride location

The objective was to determine if CMs and AMs preferred to ride in different communities across Chicago. A heat map displaying community ride percentage (CRP), derived from annual figures (see Table A. 8), is shown in Figure 7. Locational distributions were similar, with most rides densely clustered in North-Easterly regions between Uptown, Near West Side, Logan Square and Near South Side. CRP outside of this region was considerably lower; the exception being Hyde Park which ranked as the 8th most popular community overall (see Table A. 8). Rides were disproportionately shared among the 77 communities, with the 20 most popular communities accounting for nearly all annual rides (96.04%). Similar dispersion was observed in both member groups (mean = 100/77 = 1.30%, SD_{CM} = 3.52pp, SD_{AM} = 3.44pp). The heat map in Figure 8 focuses on the top 20 communities for each member group.

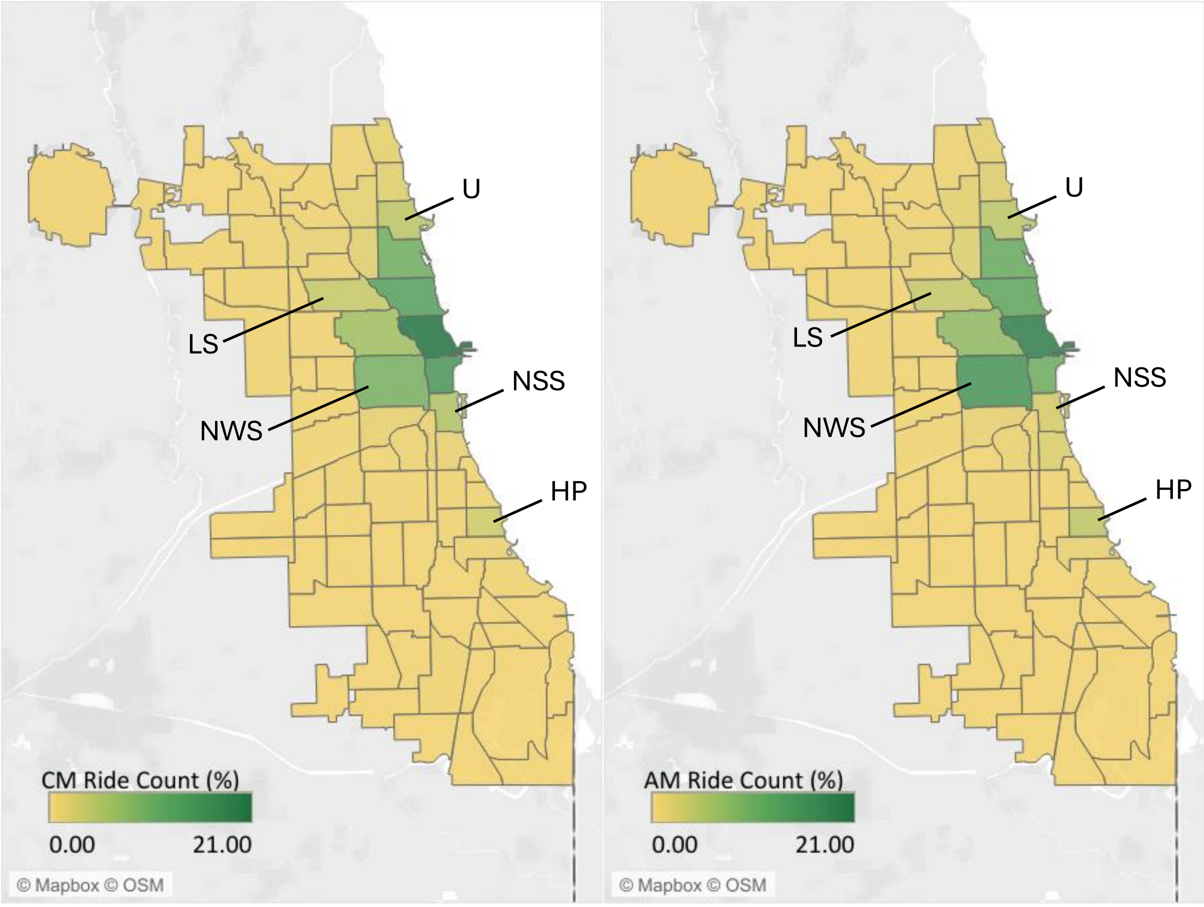

Figure 7. Displays ride count percentage between May 2024 and April 2025 by community. Starting furthest North and heading clockwise: Uptown (U); Near South Side (NSS); Hyde Park (HP); Near West Side (NWS); Logan Square (LS). To identify unlabelled communities, see Figure B. 3. For the corresponding SQL code see Code Link D. 10 and Code Link D. 11.

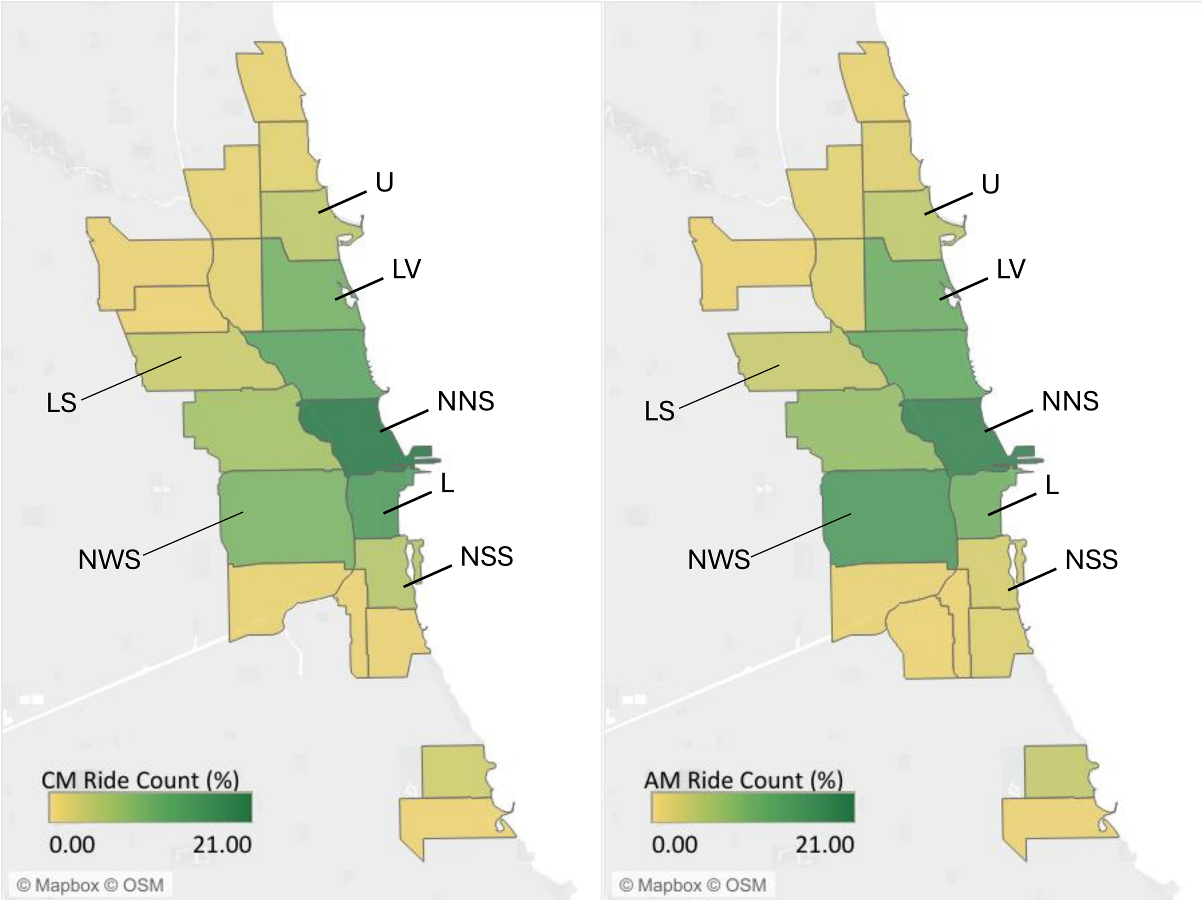

Figure 8. Shows the top 20 communities for CMs and AMs by ride count percentage. Note that all 77 communities were included in calculations of CRP (see Table A. 8). Starting furthest North and heading clockwise: Uptown (U); Lake View (LV); Near North Side (NNS); Loop (L); Near South Side (NSS); Near West Side (NWS); Logan Square (LS).

As shown by Figure 8, CMs preferred the shoreline communities between Lake View and Loop compared to AMs (8.94pp), while AMs were more likely to ride inland between Near West Side and Logan Square (6.89pp). Near North Side was the most popular community for CMs (20.61%) and AMs (18.82%). The largest differences were observed in Near West Side (5.50pp) and Loop (4.31pp) with AMs preferring the former, and CMs the latter. Considering only the top 20 communities for each member group, ride rate dispersion was similar (mean = 100/20 = 5%, SD_{CM} = 5.69pp, SD_{AM} = 5.44pp).

CM and AM CRP scores are displayed again in Figure 9. Commonalities between AMs and CMs included the same 6 communities with highest CRP, and the same 10 communities with lowest CRP. The largest rank differences were found at Loop, Near West Side and Near South Side (ΔRank = 3). The mean rank difference across all 20 communities was 1.3.

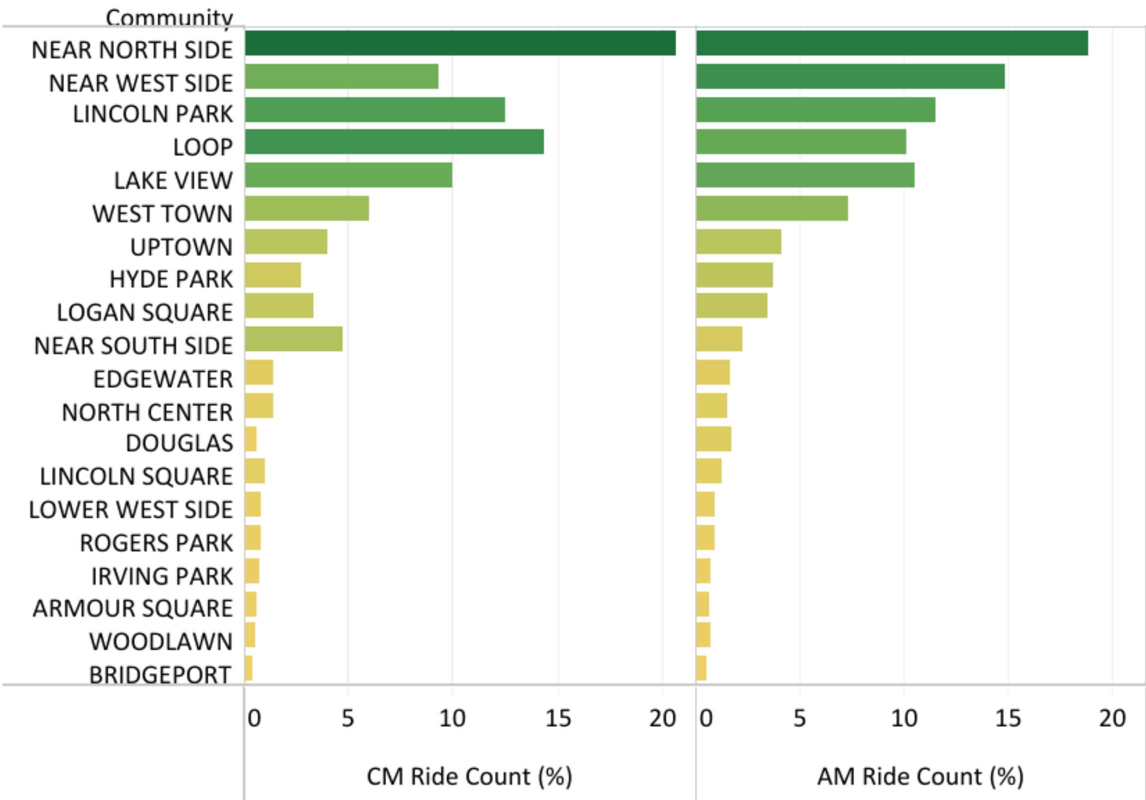

Figure 9. Displays CM and AM CRP. Communities are sorted in descending order according to total annual ride count. The top 20 communities are displayed. For CRP scores see Table A. 8.

A chi-square test of independence for all 77 communities revealed a significant distinction between both member groups ($X^2(df=76, N=5401873) = 110046.409, p < 0.01$). A subsequent Cramer's V assessment yielded a 'small' effect size ($V = 0.14$). From these findings, it was possible to reject the null hypothesis ($H_0 4$): that AMs and CMs prefer to ride in the same communities.

3.4: Weekday vs Weekend Ride Location

Following the rejection of null hypotheses $H_0 3$ and $H_0 4$, and the determination that both AM and CM community-level ride preferences changed significantly between the working week and weekend (see Table A. 9 and Table A. 10), an additional comparison was made to determine if these fluctuations were significantly different. The alternate hypothesis is stated below:

$H_1 5$ – Community-level ride tendencies change disproportionately between CMs and AMs from the working week to the weekend.

A ranked comparison of CRP fluctuations for the 20 most popular communities is shown in Figure 11. Note that all CRP fluctuations discussed in this section, are relative to working week CRPs. Common trends were detected in both memberships. The greatest CRP increases were observed at Lake View and Lincoln Park for both CMs (Lake View = 1.74pp, Lincoln Park = 1.96pp) and AMs (Lake View = 3.01pp, Lincoln Park = 2.71pp). Additionally, CRP decreases were greatest at Near West Side for both memberships (CM = -3.59pp, AM = -5.28pp). The three largest CRP differences between both memberships were observed at Loop (Δpp = 3.77), Near West Side (Δpp = 1.69) and Near North Side (Δpp = 1.61). Variation in CRP fluctuations between the working week and weekend, was higher for AMs (SD = 1.75pp) compared to CMs (SD = 1.07pp) across the 20 most popular communities.

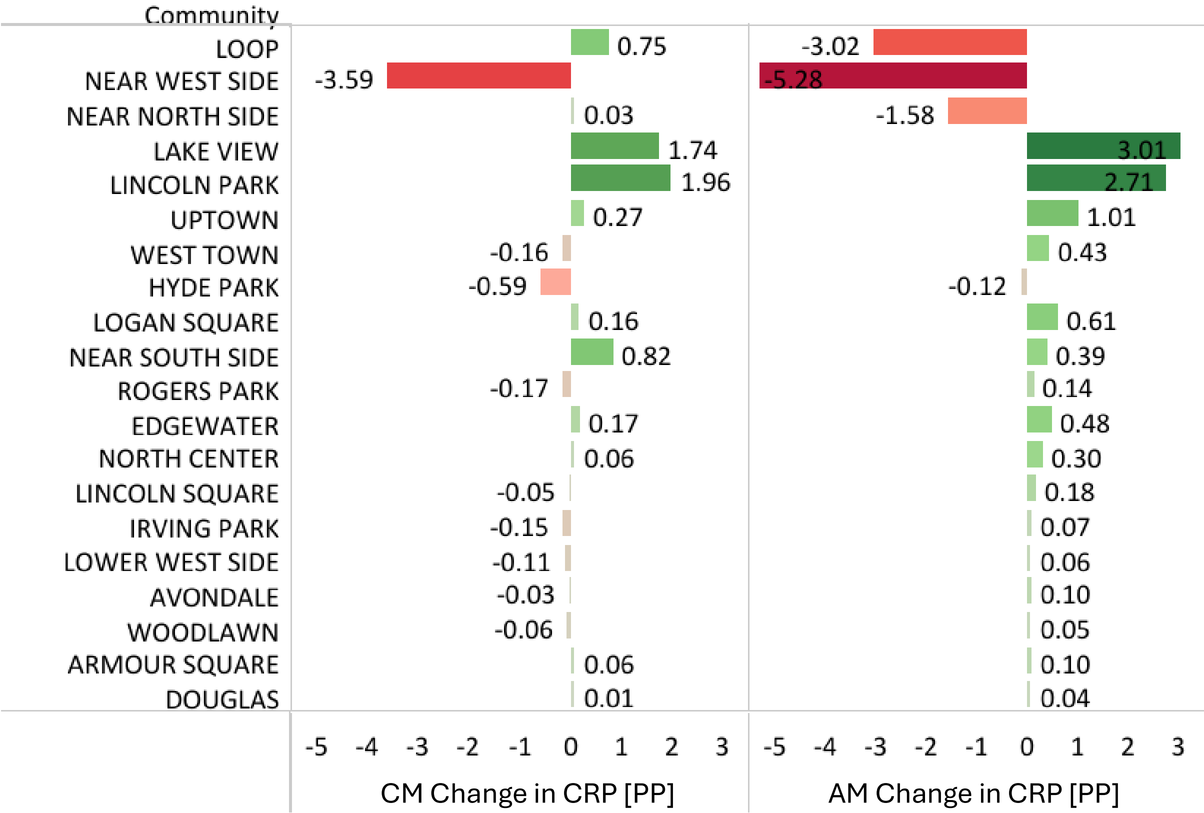

Figure 10. Shows AM and CM CRP fluctuations between weekdays and the weekend. Note that communities are sorted in order of AM and CM CRP differences.

Analysis of all 77 communities revealed a net CRP shift of CMs towards shoreline communities ($\sum pp, shoreline = +3.19pp$) and away from inland communities ($\sum pp, inland = -3.19pp$). For AMs this was also noted to a smaller magnitude ($\sum pp, shoreline = +0.69pp, \sum pp, inland = -0.69pp$). As shown by Figure 11, CRP increases were most prevalent for CMs between the shoreline communities of Edgewater and Douglas ($\sum pp = +5.75pp$); gains for AMs in these same communities were comparatively less ($\sum pp = +3.03pp$) due to negative CRP shifts at Loop (-3.02pp) and Near North Side (-1.58pp). The findings indicate that, over the course of the year, community preferences generally changed disproportionately between member groups heading into the weekend.

Looking at ride coordinates, average start locations remained in Near North Side for both member groups, although AMs preferred to ride further North at the weekend (ΔX_{AM} = 0.56 km, bearing = 351˚), while CM ride location shifted North-East (ΔX_{CM} = 0.28 km, bearing = 037˚). A greater distance separating member group average ride locations ($X_{Weekday}$ = 0.28 km, $X_{Weekend}$ = 0.43 km) further evidenced divergent group trends.

Based on variable CRP shifts, and increased distance between mean group ride locations, it was possible to reject the null hypothesis ($H_0 5$): that community-level ride tendencies change proportionately between CMs and AMs from the working week to the weekend.

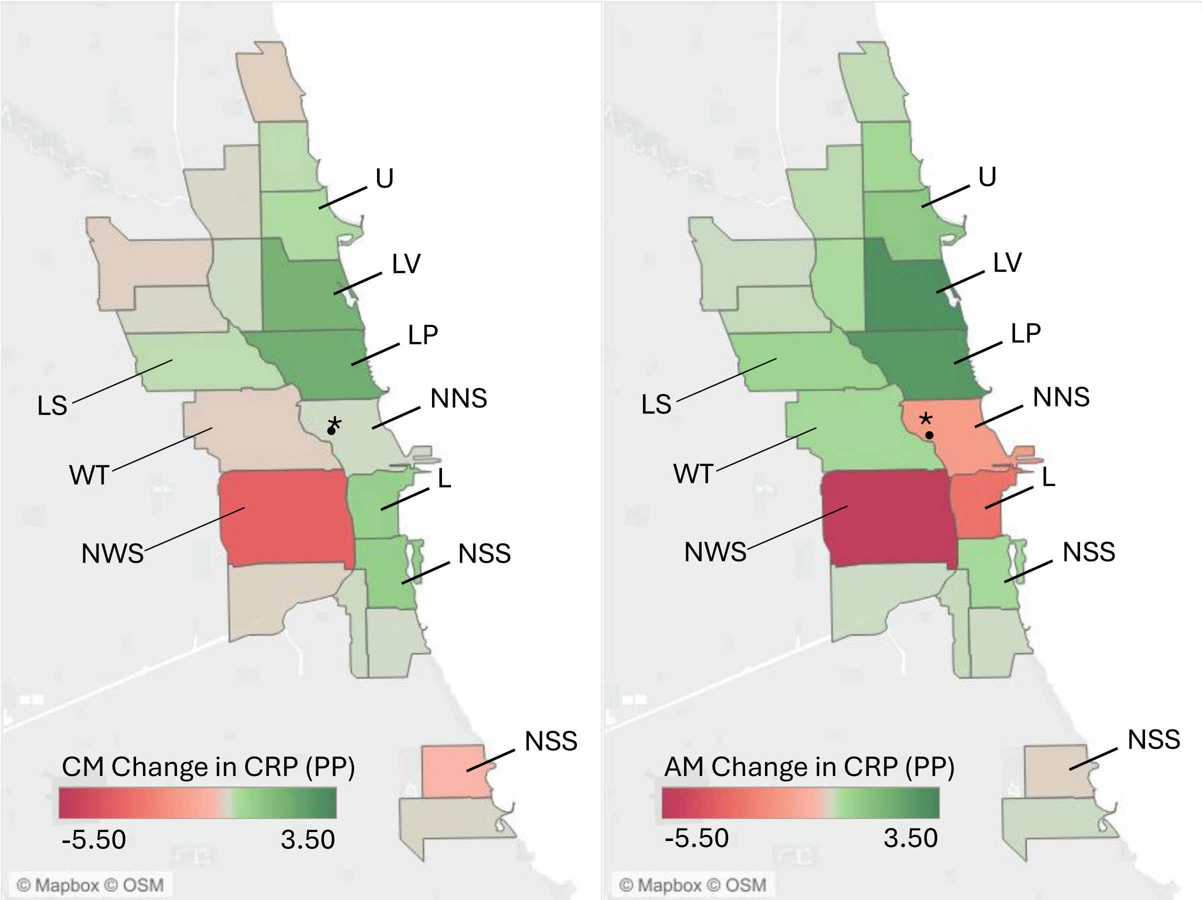

Figure 11. Shows CM and AM CRP fluctuations between the working week and weekend for rides between May 2025 and April 2024. Average weekday starting coordinates are shown by '•', while average weekend starting coordinates are shown by '' (SQL code used to generate figure coordinates is available through Code Link D. 12).*

Section 4.0: Discussion

Evaluation of Cyclistic trip data revealed a significant difference in member trip duration, with CM's averaging between 489 and 495 seconds longer per trip ($p<0.01$). These findings align with observations by Yang et al. (2024) who found mean ride duration and distance to be higher amongst single day users compared to AMs, for a bike share scheme in Richmond [12]. With current data, it is difficult to state why CM's ride for longer than AM's, although Mohamed et al. (2024) found a significant difference in the distances travelled by commuters and those cycling recreationally [13], suggesting differing ride durations could be linked to travel purpose. Additionally, employment, or living circumstances may contribute to longer trip durations. More research is required to explain the relationship between ride duration and member type.

Another significant distinction was seasonality, with CMs 10.20pp more likely to ride during the summer months than AMs. Meteorological records show that Chicago is both warmer and wetter during the summer months [14]. Therefore, considering seasonal ride trends, it would be expected that CMs prefer to ride in these conditions. However, Jaber & Csonka (2023) contradict this, observing that 'occasional user' ride rates in Budapest are negatively impacted by precipitation [15]. Kim (2022) on the other hand, observes a positive association between temperature and 'one-day' user ride rates [16]. It could be that anomalous weather patterns affected user behaviour during the summer months. Another explanation could be improved cycling infrastructure in Chicago, meaning users feel safe in wet conditions. More research is required to determine the exact cause of differential seasonal ride count.

Normalised working week to weekend ride tallies were significantly different between member groups. Mean daily ride share for CMs was 4.71pp higher than AMs for any given weekend day, suggesting that CMs prefer to ride over the weekend, while AMs give more priority to weekday rides. The findings compare well with other bike share systems: positive associations between short-term bike passes and weekend ride rates, have been reported in Budapest and Seoul [15], [16]. A possible reason for this trend could be rider profile, with short-term visitors or tourists more likely to visit on weekends; conversely most commuters would be expected to travel during the working week. Research on member demographic is required before drawing such conclusions, but the author of this report believes it could be advantageous for streamlining future strategic shifts.

In addition to time and duration, locational differences were observed, with CMs preferring to ride in significantly different communities across Chicago. Near West Side was one of the most polarising communities, ranking 2nd for AMs but only 5th for CMs. According to Chicago Metropolitan Agency for Planning (CMAP) data, Near West Side is the 10th most populous community with employment rates 9.1pp higher than average city levels [17]. Additionally, nearly 1 in 4 Near West Side inhabitants regularly cycle or walk to work [17]. Cyclistic riders in the community may, therefore, be disproportionately comprised of commuters, and could explain higher AM ride rates. This would complement week to weekend trends that suggest AM are more likely to ride on weekdays.

An analysis of member CRP shifts over the week and weekend, revealed a greater CM CRP shift towards shoreline communities ($\sum pp, shoreline = +3.19pp$) compared to AMs ($\sum pp, shoreline = +0.69pp$). A possible reason for this could be Chicago's Lakefront trail (see Figure 12). The trail is recognised for being one of the most popular recreational areas in the city, encompassing 18 miles of pathways intended specifically for cyclists and pedestrians [18]. If CMs prefer to ride recreationally, it makes sense that they would 'migrate' to this area in their free time. Conversely, a lower CRP increase for AMs in the same communities, suggests that recreational riding could be a lower priority to its constituents. In terms of supporting evidence, Buck et al. (2013) were able to show that short-term members in Washington DC prefer recreational riding [19]. More data is required to reach similar conclusions for Cyclistic members; the author of this report recommends user surveys as a suitable information source for such research.

Figure 12. Shows a scale map of the 'Lakefront Trail' [20]. Note that the trail runs between the communities of Uptown and South Shore, nearly the entire length of the Chicago's shoreline.

4.1: Limitations

Cyclistic bike tracking-computers, were the primary data source used in the analysis. As detailed by Lißner and Huber (2021), 'inaccurate data' is common with these systems and often caused by GPS or user errors [21]. Distinguishing between inaccurate and accurate data, was further complicated by the absence of a user or vehicle ID. Inclusion of such data would make it easier to filter problematic accounts, or faulty GPS systems. In the end, a work-around was achieved by recreating documented error modes. These were then used to filter suspicious data. Similar cleaning methodologies were used by DongDong et al. (2021) who anticipated erroneous data, by using scenario-led error mode prediction in a successful analysis of London's bikeshare system [22].

An absence of user ID information also made it impossible to identify 'extreme' users. Winters et al. (2019) identified 'super users' for a bike share system in Vancouver, with ride habits significantly dissimilar to other users [23]. It is theoretically possible that this type of rider could skew results, and therefore the assumptions made for an entire membership group. Considering the sample size of the study, this outcome is considered unlikely, but future studies would benefit from inclusion of the data.

Section 5.0: Conclusion

The study used company tracking data to determine how CMs and AMs use Cyclistic bikes differently. The original goal was to compare ride duration, ride-time and ride-location, however spatiotemporal trends were also investigated following completion of the first three analyses. CMs are shown to ride for longer on average, with group constituents displaying a greater preference for ride intervals that last longer than 20 minutes (15.31pp). Seasonal ride rates are significantly different, with CMs riding more frequently in the summer, and less in the winter months. Results show significant distinctions between weekend and weekday ride preferences, with CM mean daily ride share increasing at the weekend (5.81pp); AM mean daily ride share is higher during the week (2.43pp). Analysis of geo-spatial data reveals that CMs prefer to ride in significantly different communities, favouring shoreline regions between Lake View and Near South Side over inland communities like Near West Side. Finally, there is evidence to suggest that member community preference changes disproportionately over the weekend, with CMs displaying a greater CRP increase in shoreline communities over this period. Following identification of these member bike-usage disparities, the following actions are recommended for converting CMs to AMs:

1. Concentrate annual membership promotions at bike stations in shoreline communities between Lake View and Loop over the weekend.
2. Increase CM ride rates in shoreline communities between Lake View and Loop over the weekend.
3. Increase CM prices for the summer months.
4. Offer discounted annual membership in the summer months.
5. Increase CM prices for rides greater than or equal to 20 minutes.

The reader is reminded that implementation of these proposals is no guarantee of increased CM conversion, and further analysis is advised to monitor overall success.

Section 6.0: References

[1] Google Data Analytics Professional Certificate, “Case Study 1: How does a bike-share navigate speedy success?,” [Online]. [Accessed 10 05 2025].

[2] “Index of bucket "divvy-tripdata",” Divvy, [Online]. Available: https://divvy-tripdata.s3.amazonaws.com/index.html. [Accessed 17 05 2025].

[3] N. End, “The Excel Row Limit is 1,048,576 Rows - Top 5 Solutions,” row zero, 03 01 2025. [Online]. Available: https://rowzero.io/blog/excel-row-limit. [Accessed 17 05 2025].

[4] Boundaries, Census Bureau US, “Urban Area National Shapefile (2010 Census),” TIGER/Line® Shapefiles, 2010.

[5] City of Chicago , “Chicago Data Portal, Boundaries - Community Areas - Map,” Chicago, 2025.

[6] Divvy Bikes, “I’m done riding, but my trip is still open,” Divvy Bikes, [Online]. Available: https://help.divvybikes.com/hc/en-us/articles/360033485491-I-m-done-riding-but-my-trip-is-still-open. [Accessed 25 05 2025].

[7] Museum of Science and Industry, “INTRODUCING THE KENNETH C. GRIFFIN MUSEUM OF SCIENCE AND INDUSTRY,” Kenneth C. Griffin Museum of Science and Industry, 2024 05 14. [Online]. Available: https://www.msichicago.org/press/press-releases/introducing-the-kenneth-c-griffin-museum-of-science-and-industry. [Accessed 2025 05 27].

[8] F. É. d. S. H. D. F. S. P. R. C. Kon F, “Abstracting mobility flows from bike-sharing systems,” *Public Transp,* vol. 14, no. 3, p. 559, 2022.

[9] A. &. Z. P. &. H. Y. &. G. K. &. A. K. Li, “An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China,” *Journal of Transport Geography,* vol. 88, p. 5, 2020.

[10] M. Tushev, “Does my sample have to be normally distributed for a t-test?,” Medium, 26 10 2021. [Online]. Available: https://miroslavtushev.medium.com/does-my-sample-have-to-be-normally-distributed-for-a-t-test-7ee91aaaca2a. [Accessed 25 05 2025].

[11] S. Lee, “The Ultimate Guide to Cramer's V Analysis,” Number Analytics, 15 May 2025. [Online]. Available: https://www.numberanalytics.com/blog/ultimate-cramers-v-guide. [Accessed 10 June 2025].

[12] E. S. N. K. C. C. P. P. V. F.-M. P. Yifan Yang, “Comparing E-bike and Conventional Bicycle Use Patterns in a Public Bike Share System: A Case Study of Richmond, VA,” *Journal of Cycling and Micromobility Research,* vol. Volume 2, p. 5, 2024.

[13] F. Mohamed, M. Chandrabose, A. Mohammad Forkan, N. Owen and T. Sugiyama, “Variations in Cycling Distances by Trip Purpose and Socio-Demographic Attributes: Implications for Spatial Scales to Assess Environmental Correlates of Cycling.,” *International Journal of Environmental Research and Public Health,* vol. 21, no. 12, p. 1, 2024.

[14] "Climate and Average Weather Year Round in Chicago Illinois, United States," Weather Spark, [Online]. Available: https://weatherspark.com/y/14091/Average-Weather-in-Chicago-Illinois-United-States-Year-Round. [Accessed 17 07 2025].

[15] B. C. Ahmed Jaber, "Investigating the temporal differences among bike-sharing users through comparative analysis based on count, time series, and data mining models," *Alexandria Engineering Journal,* vol. 77, p. 6, 2023.

[16] K. K, "Discovering spatiotemporal usage patterns of a bike-sharing system by type of pass: a case study from Seoul," *Transportation (Amst),* vol. 21, p. 13, 2023.

[17] C. M. A. f. Planning, "Near West Side, Community Data Snapshot, Chicago Community Area Series, July 2025 Release," American Community Survey , Chicago, 2025.

[18] "Chicago's Lakefront Trail," Choose Chicago, [Online]. Available: https://www.choosechicago.com/listing/chicagos-lakefront-trail/. [Accessed 18 07 2025].

[19] D. B. R. H. P. R. B. C. P. &. B. N. Buck, "Are Bikeshare Users Different from Regular Cyclists? A First Look at Short-Term Users, Annual Members, and Area Cyclists in the Washington, D.C., Region," *Transportation Research Record,* vol. 2387, no. 1, p. 112, 2013.

[20] J. &. T. Z. &. S.-M. D. &. B. J. &. J. M. &. C. e. a. Wolch, "Proximity and Perceived Safety as Determinants of Urban Trail Use: Findings from a Three-City Study," *Environment and Planning A,* vol. 42, p. 65, 2010.

[21] S. H. S. Lißner, "Facing the needs for clean bicycle data – a bicycle-specific approach of GPS data processing," *European Transport Research Review volume,* vol. 13, no. 8, p. 1, 2021.

[22] D. L. C. a. M. D. Feng, "Exploring the Impact of Dockless Bikeshare on Docked Bikeshare—A Case Study in London," *Sustainability,* vol. 12, no. 15, p. 5, 2020.

[23] K. H. S. J. Meghan Winters, "Who are the 'super-users' of public bike share? An analysis of public bike share members in Vancouver, BC,," *Preventive Medicine Reports,* vol. 15, p. 1, 2019.

Section 7.0: Appendix

Appendix A

Table A. 1. Lists the schema for data sourced directly from the Divvy public data lake [2].

Table Name	Field Name	
0_0_1_raw_data	ride_id	String
	rideable_type	String
	started_at	Timestamp
	ended_at	Timestamp
	start_station_name	String
	start_station_id	String
	end_station_name	String
	end_station_id	String
	start_lat	Float
	start_lng	Float
	end_lat	Float
	end_lng	Float
	member_casual	String

Table A. 2. Lists the schema for the dataset following data cleaning step L-M.

Table Name	Field Name	Data Type
3_1_1_analysis	ride_id	String
	started_local	Datetime
	ended_local	Datetime
	start_location	String
	end_location	String
	ride_duration_seconds	Float
	member_type	String
	rideable_type	String
	start_lng	Float
	start_lat	Float
	end_lng	Float
	end_lat	Float
	start_location_id	String
	end_location_id	String

Table A. 3. Shows CM and AM observed ride counts, expected ride counts and chi-square values for 7 different duration intervals.

Member Group	RD < 10 [m]	10 ≤ RD < 20 [m]	20 ≤ RD < 30 [m]	30 ≤ RD < 40 [m]	40 ≤ RD < 50 [m]	50 ≤ RD < 60 [m]	RD ≤ 60 [m]
CM Count	792650	588952	248198	119206	67954	42677	110837
AM Count	1991885	1010715	312081	117276	45429	15570	21803
CM Exp	1000295	201270	574652	40730	47648	84952	20924
AM Exp	1784239	359008	1025014	72652	84991	151529	37322
Chi Term	43103.91	10941.48	355.83	18195.03	83796.24	13811.74	22614.2
	24165.28	6134.11	199.49	10200.66	46978.56	7743.25	12678.1
X^2 = 300917.901							

Table A. 4. Shows CM and AM observed ride counts, expected ride counts and chi-square values for each season within the study date parameters.

Member Group	Spring	Summer	Autumn	Winter
CM Count	411114	925893	383553	249914
AM Count	642437	1292905	849683	729734
CM Exp	378469.8	797064	443018	351921
AM Exp	675081.2	1421734	790218	627727
Chi Term	2815.67	20822.4	7981.9	29568
	1578.545	11673.6	4474.9	16577
X^2 = 95491.328, V = 0.13				

Table A. 5. Shows CM and AM observed and normalised ride counts, expected ride counts and chi-square values for weekdays and weekends within the study date parameters. Note that normalised counts were calculated by multiplying weekday counts by 0.75 (2dp).

Member Group	Weekday	Weekend
CM Count	929588	1040886
AM Count	2154998	1359761
CM Norm Count	693855	1040886
AM Norm Count	1608515	1359761
CM Exp	849246	885496
AM Exp	1453125	1515151
Chi Term	28432.50727	27268.55
	16616.72849	15936.48
X^2 = 88254.267, V = 0.14		

Table A. 6. Shows observed ride counts and chi-square values for the start and end locations of rides categorised by community. The Chi-square value and Cramer's V are shown in the final row. A 0.01 Cramer's V value indicated that the significance between both categories was of little-to no importance [11].

Community	Ride Start Count	Ride End Count
Near North Side	1051056	1031867
Near West Side	693282	686231
Lincoln Park	638006	653723
Loop	626175	613529
Lake View	554357	564334
West Town	369772	373199
Uptown	216888	217302
Hyde Park	179846	179034
Logan Square	179269	182020
Near South Side	166189	157587
Edgewater	81069	83865
North Center	77862	80572
Douglas	69718	69927
Lincoln Square	60144	60781
Lower West Side	47633	49643

Rogers Park	45635	45821
Irving Park	37524	38080
Armour Square	33985	35568
Woodlawn	32954	33053
Bridgeport	26392	26720
Avondale	26067	26272
West Ridge	18528	19261
Kenwood	14285	13634
Humboldt Park	12695	13224
Albany Park	11478	11867
North Park	9465	9453
Portage Park	8744	8994
Belmont Cragin	7491	7763
Grand Boulevard	7421	7890
South Lawndale	6476	6611
Austin	5866	5943
Hermosa	5579	5831
South Shore	5409	5524
North Lawndale	5100	5073
Oakland	5003	4800
East Garfield Park	4758	4666
Washington Park	4374	4409
New City	4215	4314
Brighton Park	3637	3625
Jefferson Park	3573	3517
Mckinley Park	3511	3456
Dunning	2740	2950
South Chicago	2125	2178
Greater Grand Crossing	2047	2051
Chicago Lawn	1996	2150
Archer Heights	1985	1832
Roseland	1902	1776
Gage Park	1869	1876
Englewood	1783	1874
Forest Glen	1708	1767
Norwood Park	1682	1722
Chatham	1438	1497
Garfield Ridge	1389	1368
Beverly	1383	1418
West Lawn	1381	1610
West Englewood	1358	1456
Ashburn	1217	1264
Fuller Park	1190	1208
East Side	1182	1187
Auburn Gresham	1112	1118
Morgan Park	952	952
South Deering	943	1002
West Garfield Park	806	837

Washington Heights	793	801
West Elsdon	687	644
Hegewisch	618	618
Pullman	557	537
West Pullman	556	578
Ohare	516	560
Montclare	497	511
Mount Greenwood	443	408
Clearing	433	463
Avalon Park	381	365
Edison Park	253	268
Calumet Heights	242	267
Riverdale	185	160
Burnside	93	95
X^2 = 1222, V = 0.01		

Table A. 7. Shows observed ride counts, Cramer's V and the chi-square value for CM and AM ride start locations by community.

Community	CM Count	AM Count
Near North Side	397315	653741
Near West Side	179289	513993
Lincoln Park	239480	398526
Loop	276849	349326
Lake View	191141	363216
West Town	115805	253967
Uptown	75829	141059
Hyde Park	51329	128517
Logan Square	62857	116412
Near South Side	89636	76553
Edgewater	27033	54036
North Center	25976	51886
Douglas	12099	57619
Lincoln Square	19649	40495
Lower West Side	16072	31561
Rogers Park	15364	30271
Irving Park	12980	24544
Armour Square	11784	22201
Woodlawn	9793	23161
Bridgeport	7315	19077
Avondale	9749	16318
West Ridge	5976	12552
Kenwood	4465	9820
Humboldt Park	4529	8166
Albany Park	3807	7671
North Park	3432	6033
Portage Park	4215	4529
Belmont Cragin	3316	4175

Grand Boulevard	3099	4322
South Lawndale	2578	3898
Austin	2899	2967
Hermosa	2024	3555
South Shore	2207	3202
North Lawndale	2429	2671
Oakland	1696	3307
East Garfield Park	2131	2627
Washington Park	1809	2565
New City	1855	2360
Brighton Park	1982	1655
Jefferson Park	1437	2136
Mckinley Park	1260	2251
Dunning	1761	979
South Chicago	865	1260
Greater Grand Crossing	1177	870
Chicago Lawn	1058	938
Archer Heights	1189	796
Roseland	1321	581
Gage Park	1049	820
Englewood	1044	739
Forest Glen	628	1080
Norwood Park	759	923
Chatham	713	725
Garfield Ridge	974	415
Beverly	766	617
West Lawn	910	471
West Englewood	725	633
Ashburn	943	274
Fuller Park	446	744
East Side	785	397
Auburn Gresham	618	494
Morgan Park	380	572
South Deering	587	356
West Garfield Park	385	421
Washington Heights	551	242
West Elsdon	445	242
Hegewisch	502	116
Pullman	380	177
West Pullman	316	240
Ohare	287	229
Montclare	252	245
Mount Greenwood	355	88
Clearing	396	37
Avalon Park	263	118
Edison Park	106	147
Calumet Heights	188	54
Riverdale	157	28

Burnside	80	13
$X^2 = 110046.409$, V = 0.14		

Table A. 8. Lists community initials, ride count by start location, and percentage ride count by start location.

Community Name	Initials	Total Count	Percent CM	Percent AM
Near North Side	NNS	1051056	20.609	18.818
Near West Side	NWS	693282	9.300	14.795
Lincoln Park	LP	638006	12.422	11.472
Loop	L	626175	14.360	10.055
Lake View	LV	554357	9.915	10.455
West Town	WT	369772	6.007	7.310
Uptown	U	216888	3.933	4.060
Hyde Park	HP	179846	2.662	3.699
Logan Square	LS	179269	3.260	3.351
Near South Side	NSS	166189	4.650	2.204
Edgewater	E	81069	1.402	1.555
North Center	NC	77862	1.347	1.494
Douglas	D	69718	0.628	1.659
Lincoln Square	LSQ	60144	1.019	1.166
Lower West Side	LWS	47633	0.834	0.908
Rogers Park	RP	45635	0.797	0.871
Irving Park	IP	37524	0.673	0.707
Armour Square	AS	33985	0.611	0.639
Woodlawn	W	32954	0.508	0.667
Bridgeport	B	26392	0.379	0.549
Avondale	A	26067	0.506	0.470
West Ridge	WR	18528	0.310	0.361
Kenwood	K	14285	0.232	0.283
Humboldt Park	HPK	12695	0.235	0.235
Albany Park	AP	11478	0.197	0.221
North Park	NP	9465	0.178	0.174
Portage Park	PP	8744	0.219	0.130
Belmont Cragin	BC	7491	0.172	0.120
Grand Boulevard	GB	7421	0.161	0.124
South Lawndale	SL	6476	0.134	0.112
Austin	AUS	5866	0.150	0.085
Hermosa	HE	5579	0.105	0.102
South Shore	SS	5409	0.114	0.092
North Lawndale	NL	5100	0.126	0.077
Oakland	O	5003	0.088	0.095
East Garfield Park	EGP	4758	0.111	0.076
Washington Park	WP	4374	0.094	0.074
New City	NC	4215	0.096	0.068
Brighton Park	BPK	3637	0.103	0.048
Jefferson Park	JP	3573	0.075	0.061

Mckinley Park	MP	3511	0.065	0.065
Dunning	DU	2740	0.091	0.028
South Chicago	SC	2125	0.045	0.036
Greater Grand Crossing	GGC	2047	0.061	0.025
Chicago Lawn	CL	1996	0.055	0.027
Archer Heights	AH	1985	0.062	0.023
Roseland	RSD	1902	0.069	0.017
Gage Park	GPK	1869	0.054	0.024
Englewood	EGD	1783	0.054	0.021
Forest Glen	FGN	1708	0.033	0.031
Norwood Park	NPK	1682	0.039	0.027
Chatham	C	1438	0.037	0.021
Garfield Ridge	GR	1389	0.051	0.012
Beverly	BEV	1383	0.040	0.018
West Lawn	WLN	1381	0.047	0.014
West Englewood	WEGD	1358	0.038	0.018
Ashburn	ASH	1217	0.049	0.008
Fuller Park	FPK	1190	0.023	0.021
East Side	ESD	1182	0.041	0.011
Auburn Gresham	AG	1112	0.032	0.014
Morgan Park	MPK	952	0.020	0.016
South Deering	SDRG	943	0.030	0.010
West Garfield Park	WGP	806	0.020	0.012
Washington Heights	WH	793	0.029	0.007
West Elsdon	WEN	687	0.023	0.007
Hegewisch	HEG	618	0.026	0.003
Pullman	PN	557	0.020	0.005
West Pullman	WPN	556	0.016	0.007
Ohare	OHR	516	0.015	0.007
Montclare	MCL	497	0.013	0.007
Mount Greenwood	MGR	443	0.018	0.003
Clearing	CLR	433	0.021	0.001
Avalon Park	APK	381	0.014	0.003
Edison Park	EPK	253	0.005	0.004
Calumet Heights	CHS	242	0.010	0.002
Riverdale	RDA	185	0.008	0.001
Burnside	BUR	93	0.004	0.000

Table A. 9. Shows CM weekday and weekend ride counts by community. The Chi-square value is shown on the final row.

Community	Weekday Count	Weekday Count (Normalised)	Weekend Count
Near North Side	187083	139640.9	210232
Loop	126881	94705.44	149968
Lincoln Park	103425	77197.61	136055
Lake View	81708	60987.79	109433
Near West Side	101741	75940.65	77548
West Town	55369	41328.06	60436
Near South Side	38324	28605.47	51312
Uptown	34461	25722.09	41368
Logan Square	28863	21543.67	33994
Hyde Park	27037	20180.73	24292
Edgewater	11935	8908.421	15098
North Center	11945	8915.885	14031
Lincoln Square	9506	7095.388	10143
Lower West Side	8106	6050.411	7966
Rogers Park	8074	6026.526	7290
Irving Park	6843	5107.694	6137
Douglas	5670	4232.153	6429
Armour Square	5258	3924.632	6526
Woodlawn	4927	3677.569	4866
Avondale	4749	3544.708	5000
Bridgeport	3740	2791.579	3575
West Ridge	3260	2433.301	2716
Humboldt Park	2318	1730.182	2211
Kenwood	2219	1656.287	2246
Portage Park	2298	1715.254	1917
Albany Park	1948	1454.01	1859
North Park	1874	1398.775	1558
Belmont Cragin	1858	1386.833	1458
Grand Boulevard	1522	1136.038	1577
Austin	1636	1221.129	1263
South Lawndale	1376	1027.062	1202
North Lawndale	1133	845.6842	1296
South Shore	1129	842.6986	1078
East Garfield Park	1042	777.7608	1089
Hermosa	1108	827.0239	916
Brighton Park	1125	839.7129	857
New City	1053	785.9713	802
Washington Park	1007	751.6364	802
Dunning	980	731.4833	781
Oakland	742	553.8373	954
Jefferson Park	830	619.5215	607
Roseland	745	556.0766	576

Mckinley Park	662	494.1244	598
Archer Heights	698	520.9952	491
Greater Grand Crossing	663	494.8708	514
Chicago Lawn	543	405.3014	515
Gage Park	584	435.9043	465
Englewood	625	466.5072	419
Garfield Ridge	549	409.7799	425
Ashburn	526	392.6124	417
West Lawn	500	373.2057	410
South Chicago	453	338.1244	412
East Side	406	303.0431	379
Beverly	419	312.7464	347
Norwood Park	394	294.0861	365
West Englewood	426	317.9713	299
Chatham	380	283.6364	333
Forest Glen	346	258.2584	282
Auburn Gresham	335	250.0478	283
South Deering	304	226.9091	283
Washington Heights	327	244.0766	224
Hegewisch	296	220.9378	206
Fuller Park	213	158.9856	233
West Elsdon	254	189.5885	191
Clearing	183	136.5933	213
West Garfield Park	191	142.5646	194
Pullman	234	174.6603	146
Morgan Park	192	143.311	188
Mount Greenwood	165	123.1579	190
West Pullman	172	128.3828	144
Ohare	145	108.2297	142
Avalon Park	154	114.9474	109
Montclare	124	92.55502	128
Calumet Heights	104	77.62679	84
Riverdale	81	60.45933	76
Edison Park	46	34.33493	60
Burnside	31	23.13876	49
X^2 = 11674.764			

Table A. 10. Shows AM weekday and weekend ride counts by community. The Chi-square value is shown on the final row.

Community Name	Weekday Count	Weekday Count (Normalised)	Weekend Count
Near North Side	414170	309141.2	210202
Loop	239242	178573	149927
Lincoln Park	222186	165842.2	136040
Lake View	198114	147874.6	109412
Near West Side	358892	267881.1	77535

West Town	152296	113675.5	60431
Near South Side	43797	32690.58	51304
Uptown	78258	58412.67	41357
Logan Square	66395	49557.99	33988
Hyde Park	79842	59594.99	24287
Edgewater	29222	21811.64	15093
North Center	29359	21913.89	14029
Lincoln Square	23387	17456.33	10141
Lower West Side	18891	14100.46	7964
Rogers Park	17405	12991.29	7289
Irving Park	14499	10822.22	6137
Douglas	35038	26152.77	6425
Armour Square	12779	9538.392	6524
Woodlawn	13807	10305.7	4866
Avondale	9195	6863.254	5000
Bridgeport	11157	8327.713	3574
West Ridge	7395	5519.713	2713
Humboldt Park	4950	3694.737	2211
Kenwood	5738	4282.909	2245
Portage Park	2775	2071.292	1917
Albany Park	4396	3281.225	1859
North Park	3577	2669.914	1557
Belmont Cragin	2392	1785.416	1458
Grand Boulevard	2588	1931.713	1577
Austin	1709	1275.617	1262
South Lawndale	2520	1880.957	1201
North Lawndale	1582	1180.823	1295
South Shore	1755	1309.952	1076
East Garfield Park	1505	1123.349	1088
Hermosa	2267	1692.115	916
Brighton Park	978	729.9904	857
New City	1438	1073.34	802
Washington Park	1645	1227.847	802
Dunning	602	449.3397	781
Oakland	1877	1401.014	954
Jefferson Park	1313	980.0383	607
Roseland	387	288.8612	575
Mckinley Park	1335	996.4593	598
Archer Heights	463	345.5885	491
Greater Grand Crossing	482	359.7703	513
Chicago Lawn	540	403.0622	515
Gage Park	455	339.6172	465
Englewood	425	317.2249	419
Garfield Ridge	258	192.5742	425
Ashburn	129	96.28708	417
West Lawn	249	185.8565	410
South Chicago	685	511.2919	412
East Side	233	173.9139	379

Beverly	381	284.3828	347
Norwood Park	604	450.8325	365
West Englewood	387	288.8612	299
Chatham	372	277.6651	333
Forest Glen	668	498.6029	282
Auburn Gresham	296	220.9378	283
South Deering	177	132.1148	283
Washington Heights	162	120.9187	224
Hegewisch	75	55.98086	204
Fuller Park	429	320.2105	233
West Elsdon	143	106.7368	191
Clearing	17	12.689	213
West Garfield Park	260	194.067	194
Pullman	127	94.79426	146
Morgan Park	332	247.8086	188
Mount Greenwood	50	37.32057	190
West Pullman	146	108.9761	144
Ohare	142	105.9904	141
Avalon Park	63	47.02392	109
Montclare	177	132.1148	128
Calumet Heights	33	24.63158	84
Riverdale	10	7.464115	74
Edison Park	76	56.72727	60
Burnside	7	5.22488	49
X^2 = 85972.964			

Appendix B

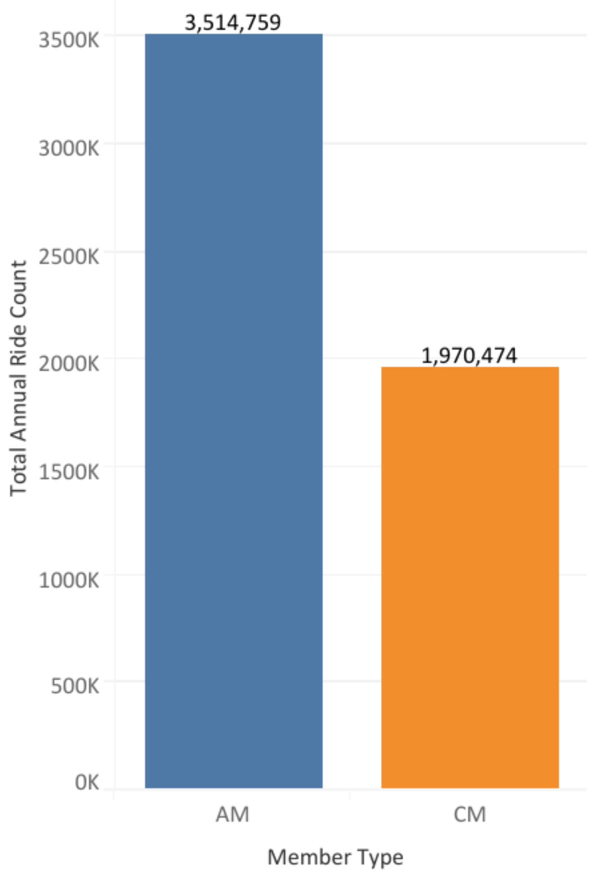

Figure B. 1. Illustrates AM and CM total ride counts between May 2024 and April 2025.

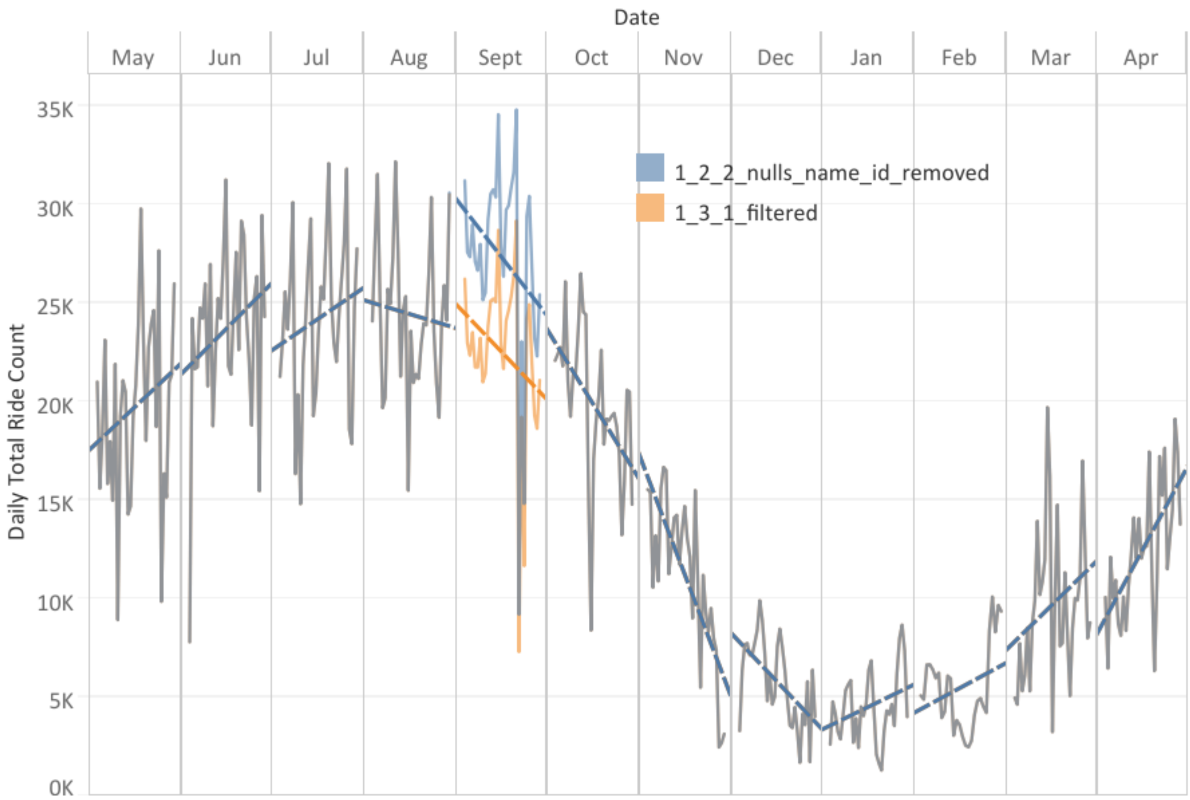

Figure B. 2. Shows total daily ride count between May 2024 and April 2025, before and after cleaning step D-E. A line of best fit for CM and AM illustrates monthly trends.

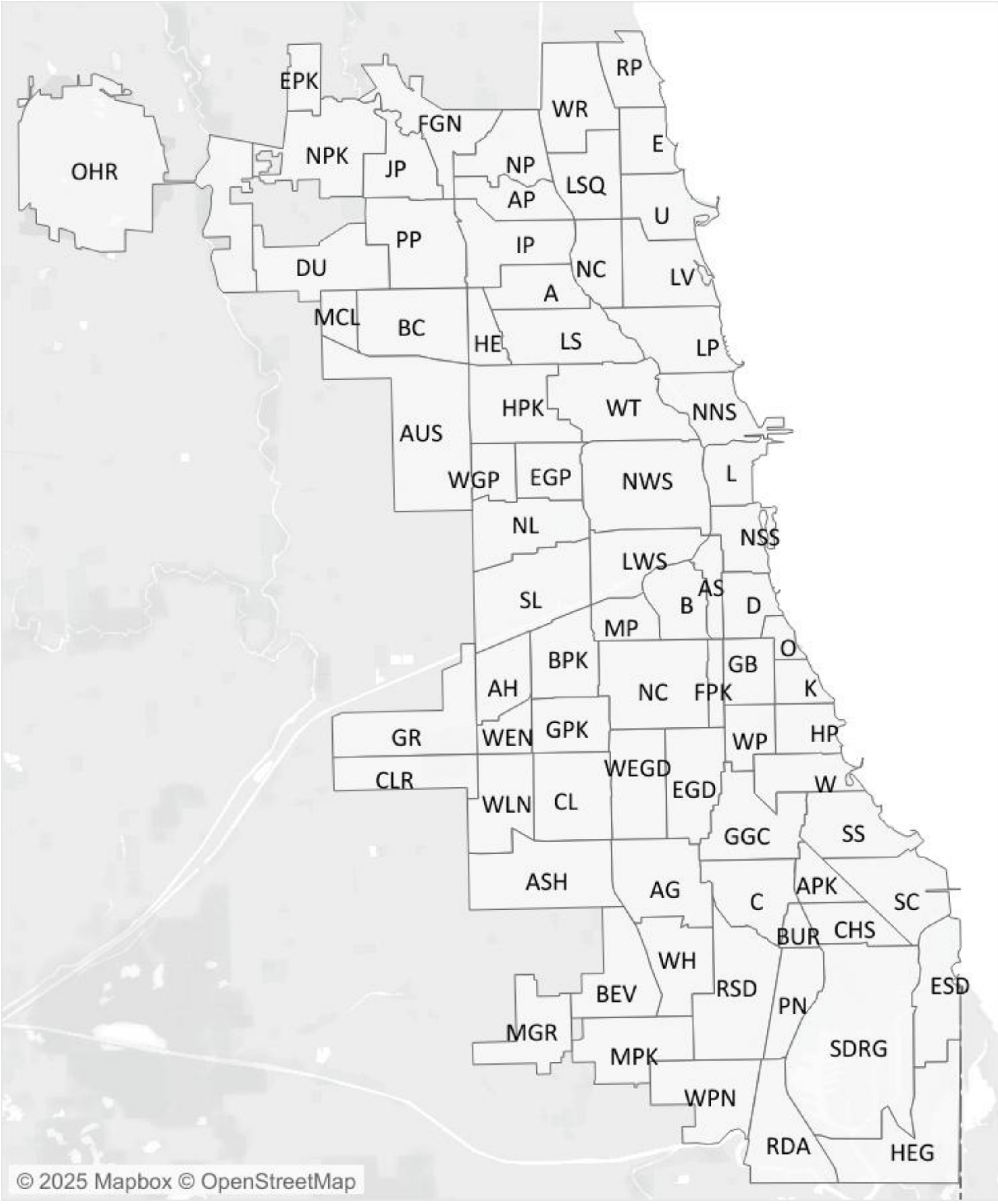

Figure B. 3. Identifies all communities in Chicago by their initials. For full names see Table A. 8.

Appendix C

$$t = \frac{\bar{x}_{casual} - \bar{x}_{annual}}{\sqrt{\frac{(s_{casual})^2}{n_{casual}} + \frac{(s_{annual})^2}{n_{annual}}}} \tag{C. 1}$$

Where:

$\bar{x}$ = mean ride duration

s = group variance

n = group sample size

t = t-value

Substituting CM and AM values into (C. 1) gives (C. 2):

$$416.821 = \frac{1211.095 - 719.373}{\sqrt{\frac{(2410007.297)^2}{1970474} + \frac{(592681.231)^2}{3514759}}} \tag{C. 2}$$

$$Lower\ Bound = (\bar{x}_{casual} - \bar{x}_{annual}) - t_{\frac{\propto}{2}} \times \sqrt{\frac{(s_{casual})^2}{n_{casual}} + \frac{(s_{annual})^2}{n_{annual}}} \tag{C. 3}$$

Where:

$\bar{x}$ = mean ride duration

s = group variance

n = group sample size

$t_{\frac{\propto}{2}}$ = critical value

Substituting CM and AM values into (C. 3) gives (C. 4):

$$488.683 = (1211.095 - 719.373) - 2.576 \times \sqrt{\frac{(2410007.297)^2}{1970474} + \frac{(592681.231)^2}{3514759}} \tag{C. 4}$$

$$Upper\ Bound = (\bar{x}_{casual} - \bar{x}_{annual}) + t_{\frac{\propto}{2}} \times \sqrt{\frac{(s_{casual})^2}{n_{casual}} + \frac{(s_{annual})^2}{n_{annual}}} \tag{C. 5}$$

Where:

$\bar{x}$ = mean ride duration

s = group variance

n = group sample size

$t_{\frac{\alpha}{2}}$ = critical value

Substituting CM and AM values into (C. 5) gives (C. 6):

$$494.761 = (1211.095 - 719.373) + 2.576 \times \sqrt{\frac{(2410007.297)^2}{1970474} + \frac{(592681.231)^2}{3514759}} \quad \textit{(C. 6)}$$

$$x^2 = \sum \frac{(O - E)^2}{E} \quad \textit{(C. 7)}$$

Where:

x^2 = Chi-square value

O = observed ride count

E = Expected ride count

$$V = \sqrt{\frac{X^2}{N(k-1)}} \quad \textit{(C. 8)}$$

Where:

V = Cramer's V

X^2 = Chi-square value

N = Sample size

k = min of rows or columns

Appendix D

Code Link D. 1. Lists SQL code from the data cleaning pipeline (see Table 2).

Table Revision	Table Name	Row Count	Code Link
A	0_0_1_raw_data	5,735,884	cyclistic_trips/data_cleaning/0_0_1_raw_data.sql at main · andremanns/cyclistic_trips
B	1_1_1_duplicates_removed	5,735,673	cyclistic_trips/data_cleaning/1_1_1_duplicates_removed.sql at main · andremanns/cyclistic_trips
C	1_2_1_nulls_lat_lng_removed	5,729,314	cyclistic_trips/data_cleaning/1_2_1_nulls_lat_lng_removed.sql at main · andremanns/cyclistic_trips
D	1_2_2_nulls_name_id_removed	5,729,314	cyclistic_trips/data_cleaning/1_2_2_nulls_name_id_removed.sql at main · andremanns/cyclistic_trips
E	1_3_1_filtered	5,584,440	cyclistic_trips/data_cleaning/1_3_1_filtered.sql at main · andremanns/cyclistic_trips
F	1_4_1_trimmed	5,584,440	cyclistic_trips/data_cleaning/1_4_1_trimmed.sql at main · andremanns/cyclistic_trips
G	1_4_2_double_spaces_removed	5,584,440	cyclistic_trips/data_cleaning/1_4_2_double_spaces_removed.sql at main · andremanns/cyclistic_trips
H	1_4_3_typos_sp_characters_removed	5,584,440	cyclistic_trips/data_cleaning/1_4_3_typos_sp_characters_removed.sql at main · andremanns/cyclistic_trips
I	1_5_1_outliers_lwr_removed	5,498,195	cyclistic_trips/data_cleaning/1_5_1_outliers_lwr_removed.sql at main · andremanns/cyclistic_trips
J	1_5_2_outliers_upr_removed	5,492,966	cyclistic_trips/data_cleaning/1_5_2_outliers_upr_removed.sql at main · andremanns/cyclistic_trips
K	1_5_3_outliers_speed_removed	5,485,233	cyclistic_trips/data_cleaning/1_5_3_speed_outliers_removed.sql at main · andremanns/cyclistic_trips
L	1_6_1_quality_check	5,485,233	cyclistic_trips/data_cleaning/1_6_1_quality_check.sql at main · andremanns/cyclistic_trips
M	2_1_1_format	5,485,233	cyclistic_trips/data_cleaning/2_1_1_format.sql at main · andremanns/cyclistic_trips
N	3_1_1_analysis	5,485,233	cyclistic_trips/data_cleaning/3_1_1_analysis.sql at main · andremanns/cyclistic_trips

Code Link D. 2. A statistical summary for ride duration, Table 3:

cyclistic_trips/data_validation/statistical_summary.sql at main · andremanns/cyclistic_trips

Code Link D. 3. Ride duration vs ride count, Figure 1:

cyclistic_trips/data_validation/ride_duration_vs_ride_count.sql at main · andremanns/cyclistic_trips

Code Link D. 4. Start station vs average ride duration, Figure 2:

cyclistic_trips/data_validation/start_location_vs_avg_ride_duration.sql at main · andremanns/cyclistic_trips

Code Link D. 5. T-test input values for ride duration, Section 3.1:

cyclistic_trips/results/ride_duration/t-test_input_values.sql at main · andremanns/cyclistic_trips

Code Link D. 6. Partitioned ride duration, Figure 3:

cyclistic_trips/results/ride_duration/partitioned_ride_duration_by_member_type.sql at main · andremanns/cyclistic_trips

Code Link D. 7. DRP through the year, Figure 4:

cyclistic_trips/results/ride_time/ride_percentages_through_the_year.sql at main · andremanns/cyclistic_trips

Code Link D. 8. Ride count partitioned by season, Figure 5:

cyclistic_trips/results/ride_time/ride_percentage_by_season.sql at main · andremanns/cyclistic_trips

Code Link D. 9. Ride count partitioned by day of the week, Figure 6:

cyclistic_trips/results/ride_time/partitioned_days_of_week.sql at main · andremanns/cyclistic_trips

Code Link D. 10. Ride count partitioned by community during the working week. Used in Figure 7, Figure 8, Figure 9, Figure 10 and Figure 11:

cyclistic_trips/results/ride_location/ride_count_by_community_weekday.sql at main · andremanns/cyclistic_trips

Code Link D. 11. Ride count partitioned by community over the weekend. Used in Figure 7, Figure 8, Figure 9, Figure 10 and Figure 11:

cyclistic_trips/results/ride_location/ride_count_by_community_weekend.sql at main · andremanns/cyclistic_trips

Code Link D. 12. Average starting coordinates by member type and week portion, Figure 11:

cyclistic_trips/results/ride_spatiotemporal/coord_by_member_and_week_portion.sql at main · andremanns/cyclistic_trips