Capstone

Project

on

Cyclistic Bikes

for the Google/Coursera

Certificate in Data Analytics

Presented

by

Joachim Kolb

from Ottawa

on

2022/07/24

Contact information:
joachim.kolb@gmail.com
www.reinbekconsulting.ca

Executive Summary:

Cyclitis's ride data allow for significant conclusions on the profiles of member and casual users of the Cyclistic bike share. With a median duration of just over eleven minutes, most rides were found to be shorter than twenty minutes, and almost all rides are shorter than three hours. In the overall data, there are two different patterns for weekdays and weekends, with the number of rides in the former having two peaks, in the morning and in the evening, and the latter reaching a single peak in mid-afternoon.

On the basis of the rides they take, members account for 56.3% of all rides and 40.8% of all bike use minutes. Members use the services more steadily across the year, and there is evidence that at least some of them use Cyclistic bikes for their morning commute. With a mean duration of 12.4 minutes and a median duration of under ten minutes, members' rides tend to be shorter, and they use their bikes more heavily during the week and comparatively less during the weekend than casual riders.

Accounting for nearly 60 percent of all minutes of use, casual riders take longer mean and median rides, are over-represented among weekend riders, and are more likely to make use of e-bikes. They also for almost all cases where a ride was logged with a docked bike, meaning a bike not actually available for riding at the time. This may reflect the limited experience of a subset of the casual riders with the Cyclistic infrastructure or indeed with life in Chicago.

The evidence is compatible with casual riders having a more experiential, leisure-oriented attitude to cycling, while members see it more as a means to an end, here the satisfaction of their transportation needs.

The most glaring limitation from the data is the lack of user-related data, which makes it impossible to trace actual individual usage patterns, not to mention a cluster analysis that might yield sub-groups among the users to be addressable by particular marketing techniques. Specifically, cases where members occasionally buy specific casual services are likely to occur in practice, but cannot be picked up from these data. It would hence be useful to consider the use of anonymized individual account data, addressing privacy concerns.

Regarding an expansion of Cyclistic member services to areas more distant from the city center, it would be useful to make non-electric bikes also available to suburbia, given the relative aversion to electric bikes on the part of member users.

Finally, while the Cyclistic data did not come with pricing information, a look at how existing pricing information from Divvy bikes plays out in the light of these data casts doubt on the financial desirability of increasing the number of member users specifically. At the present time, casuals appear to be far more lucrative.

Table of Contents

Executive Summary:.....2
1. Introduction.....4
2. The Business Task.....4
3. Data sources used, Data Cleaning and Data Manipulation.....5
3.1 Time.....6
3.2 Locations.....8
3.3 Ride Types.....8
3.4 Distance.....9
4. Analysis.....9
4.1 Framing the Role of Cycling in the US and the Windy City.....9
4.2 Hypotheses of the Analysis.....11
4.3 Evidence from Time-Related Data.....11
4.4 Evidence related to bike type.....14
4.5 Evidence from Location-Related Data.....15
5 Evidence on the Hypotheses.....19
6 Recommendations to the Marketing Team.....19
Bibliography:.....20

1. Introduction

This report analyzes the twelve months of ride data from the Cyclistic bike share in Chicago ranging from 1 June 2021 to 30 May 2022, with a view to providing recommendations on how to increase the number of Cyclistic members using the company's bike ride services on a subscription basis.

After clearly outlining the business task in Section 2, all data sources used will be described in Section 3, which includes an embedding in the

Section 3 documents how the data were cleaned and manipulated. Section 4 contains a summary of the analysis, with supporting visualizations illustrating key findings. Finally, we present three main recommendations based on this analysis.

2. The Business Task

The question posed to us by the Cyclistic Head of Marketing, Ms. Lily Moreno, concerns the way in which casual users and annual members use bikes differently. It is formulated with a view to increasing member numbers, given that this constituency is apparently considered more lucrative than casual riders by Cyclistic management.

The main focus of this assignment is to provide data on how casual and member riders differ in their usage patterns. In order to do so, usage patterns of member and casual users will be investigated across their variation over months, weeks, and the times of the day within this one-year period. Geographical information will also be taken into account. Attributes to be included are: type of bicycle, start point, start time, end point, end time, duration and distance. A more detailed characterization of these categories is found in Section 3.

Given the rides database on the basis of which this question is to be answered, it will not be possible to produce profiles of individual riders or rider types, as the customer data are not available, except as anonymized users of individual rides. We also do not know whether these rides were taken using a day pass or a ride pass, nor do we know about Cyclistic-specific pricing schemes for such passes. It will, however, be occasionally enlightening to relate this consideration to the data of Divvy Bikes, the company whose data we actually analyze, where such information is available. In principle, it is possible to coordinate this information with information about day-time patterns (working days vs weekend days and public holidays), temperature data (how do the users compare between winter and summer rides, for example).

In addition, this report will provide three recommendations to management

3. Data sources used, Data Cleaning and Data Manipulation

The following table provides some information on these categories in the data.

Category	Description
Type of bicycle	The three categories given comprise two different types of bikes, classic bikes and electric bikes. The third category, "docked bikes", appears to refer to bikes not currently available for riding.
Start point	The GPS coordinates of the point of departure
Start time	The time of departure in datetime format
End point	The GPS coordinates of the point of arrival
End time	The time at which the rental ended, in datetime format
duration	The difference between the above two points in time.
distance	A measure of the difference between the above two points in space.

Beyond the data received from Cyclistic, a search for existing bicycle rental businesses in Chicago led to Divvy, the site of which, divvybikes.com, provides ample information on one such business, which appears to have a dominant or monopoly position in this space in Chicago. Divvy is all the more relevant as its managing company, Motivate International Inc., actually provided the data for this research project, which was originally presented by Kevin Hartmann (undated) as a project about Divvy. In other words, to the extent that they reflect authentic rides, the observations that this research project is based on reflects Divvy rides.

While it may not be legitimate to equate Divvy and Cyclistic *tout court*, the information about bike rides, stations, user categories and pricing on this site put the data into perspective.

In addition, select publications from the City of Chicago and specifically the Chicago Department of Transport (2021) were viewed for further background information about the Chicago urban transport ecosystem as a whole and the place of bicycles within it and the plans of those who run it.

An old adage in data analysis projects proclaims that most of the time is usually not spent on the actual analysis of the data, but on data preparation and data cleaning. Further hampered by a relative lack of experience with the main analysis software, RStudio, I certainly found this to be the case with this project. The main modifications of the data that were required are as follows:

3.1 Time

Although they looked complete at first sight, there was considerable cleaning work to be done on the data related to date at time. For February and March 2022, the data format of the two columns *start_at* and *end_at* had to be changed to the *date_time* format using the mdy_hm() function, as they were in a different sequence and would not otherwise have been recognized. For all other months, this transformation could be done using the using the ymd_hms() function. Finding, understanding and addressing the problem was a laborious, but educational cleaning process that cost me over a week before finally ending on 22/07/15 with the creation of a unified dataset named *cyclistic5*.[1]

Subtracting the departure time from the arrival time yields the duration of the ride. It was found that some of the rides had a negative duration, which could not be interpreted. Others had an extremely long duration, with a maximum of over five weeks. While this could conceivably be interpreted, for example in the case of a convict on the run, it is not rational under the Divvy/Cyclistic business model to keep a bike for such a long period. Both rides with a negative duration and rides with a duration of more than 24 hours were hence eliminated, resulting in the loss of 4405 observations, of which 139 had a negative duration. This constituted a loss of less than 0.1% of the original sample and yielded the distribution visualized in Image 1.

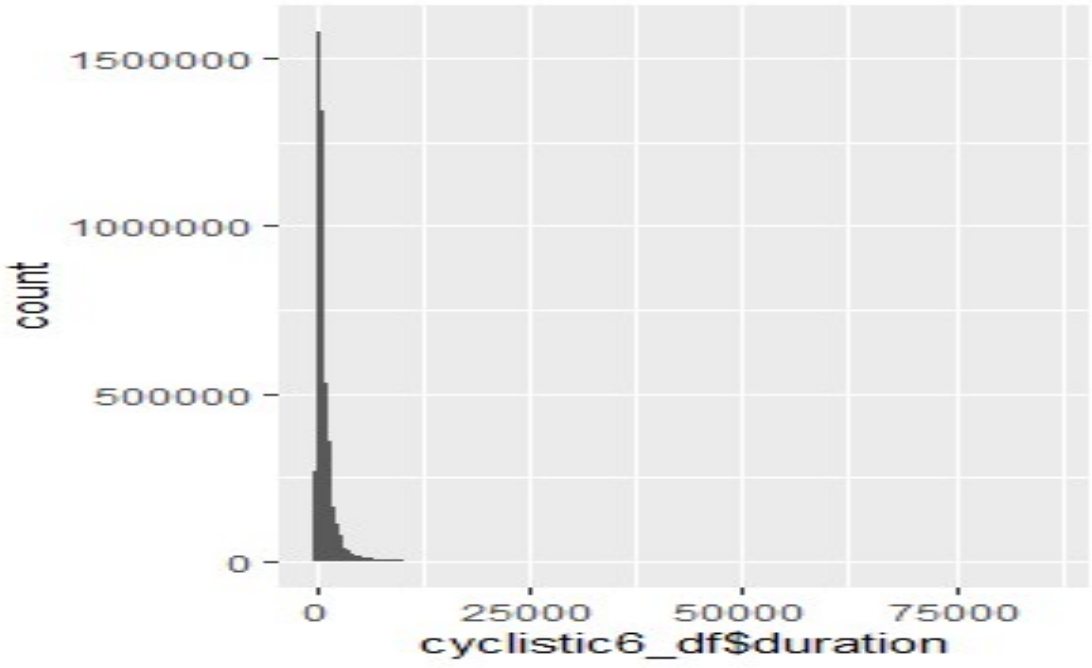


Figure 1 The distribution of ride length with durations of up to one calendar day.

[11] See the R script 220715wholeCycle4.R for details.

It was then found through plotting that the overwhelming majority of rides actually took less than 12500 seconds (under four hours), and the maximum duration was set to that value, entailing a loss of 19555 observations or 0.33% of the total.[2] The resulting distribution is shown in Figure 2 below.

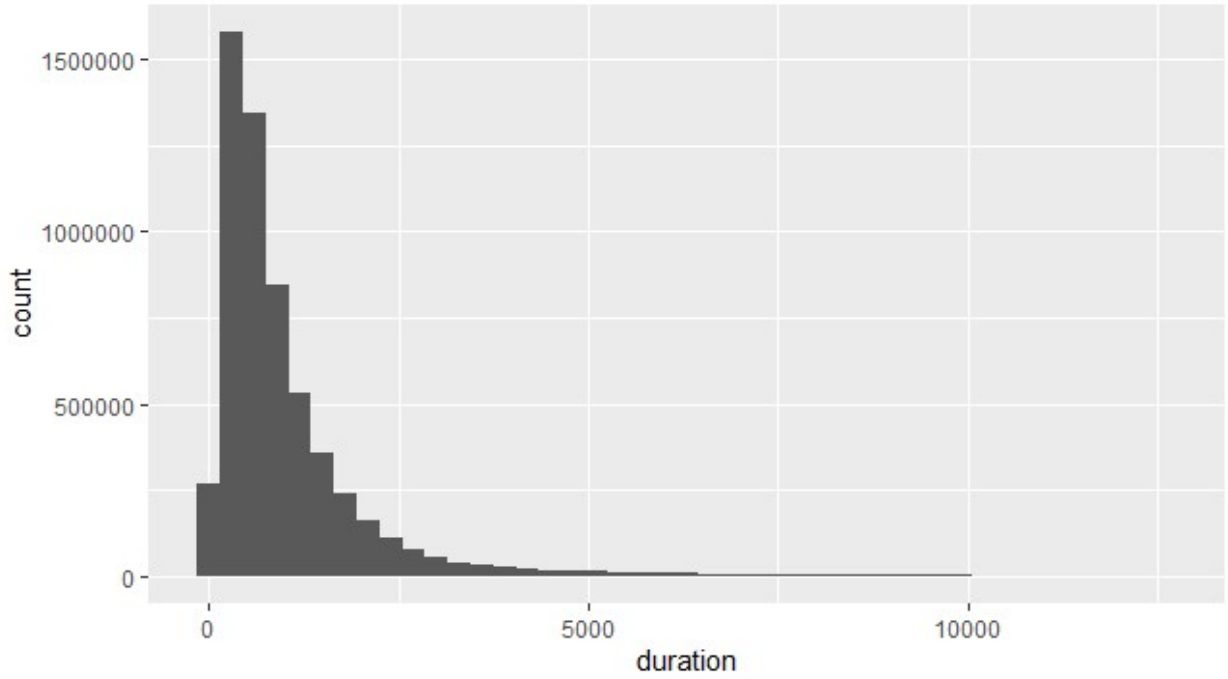


Figure 2: The duration of rides of up to 3.5 hours (12500 seconds)

Within this total, member data are found within a considerably lower range than non-member data. In other words, members prefer shorter rides.

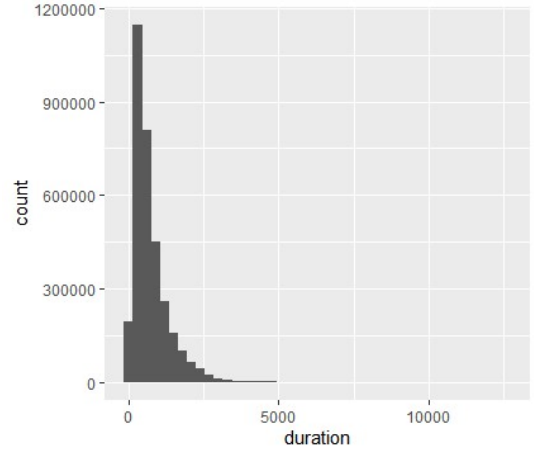


Figure 3: The duration distribution of member rides

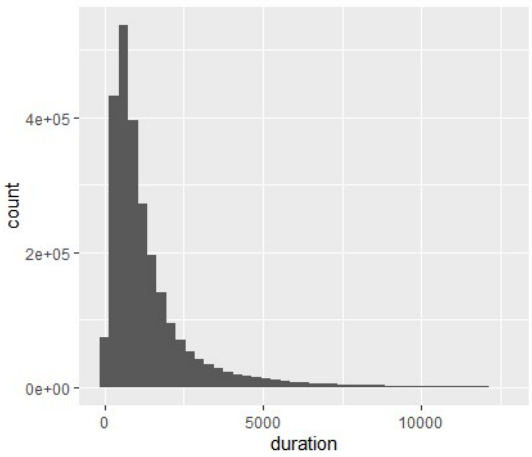
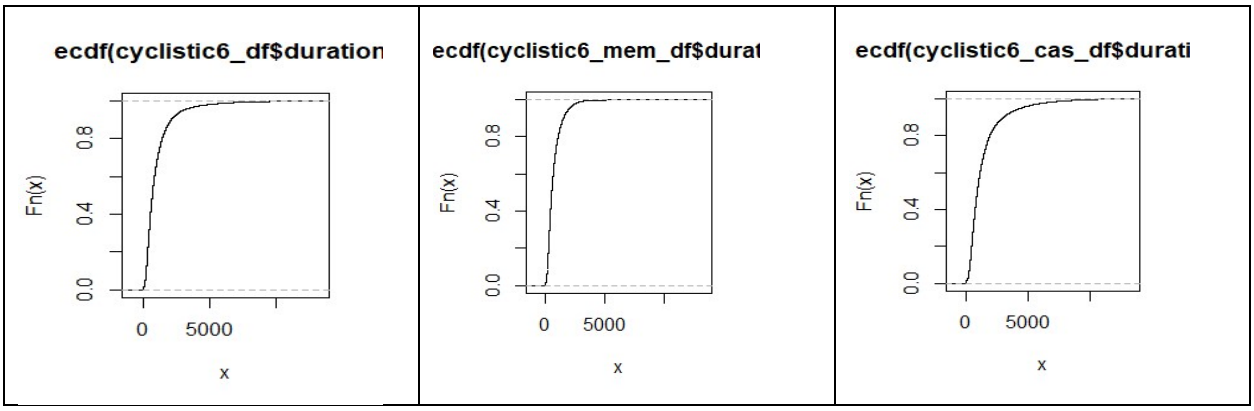


Figure 4 The duration distribution of casual rides

[2] The resulting dataset was stored in cyclistic6.

Correspondingly, the cumulative distribution functions for total, member and casual duration all rise quickly, but significantly more so for the member distribution than for the casual distribution, as the following plots in Table 1 and the numbers in Table 2 show.

Table 1: The cumulative distribution function for total, member, and casual duration (x-axis: time in seconds)



	CDF of Total duration	CDF of Member Duration	CDF of Casual Duration
5 minutes	17.35%	23.37%	9.541%
10 minutes	44.51%	54.8%	31.11%
20 minutes	74.23%	83.32%	62.45%
45 minutes	93.89%	98.38%	88.07%
1 hour	96.4%	99.38%	92.58%
2 hours	99.3%	99.9%	98.5%
3 hours	99.9%	99.99%	99.79%
4 hours	1	1	1

3.2 Locations

The *start_station* and *end_station* names and IDs were often unavailable, and there were some NA values for *end_lat* and *end_lng*. However, the corresponding columns were retained regardless, as the GPS coordinates were present in the respective start columns and some information might still be gleaned from the frequency distribution of the cases where this information was present. In the future, new information from the principals should permit a better understanding of what exactly the reason for their absence might be, and if this absence is information that can be used in a systematic way.

3.3 Ride Types

The three ride types, *classic_bike*, *electric_bike*, and *docked_bike*, represent two distinct models of bikes, "classic" mechanical ones and electric bicycles, and one status, apparently that of not being eligible for riding, *docked_bike*.

This *docked_bike* category comprises roughly 4.6% of all rides or 267,454 of a total of 5,841,028. Of these, 267,454 rides, 100% concern casual customers. For members, *docked_rides* do not appear to exist.

The mean duration of these rides is longer than that of the other categories, reaching 40 minutes, but the start and the end point tend to be identical, raising the question of whether locomotion occurred at all.

As there is not enough clarity regarding the status of these bikes, it would be premature to delete them, but it should be noted that, if they are indeed non-rides, then this would reduce the average duration of casual rides.

3.4 Distance

A measure of geographic distance is provided by the two parameters start_point and end_point, from the distance between which the degree of displacement achieved by the ride may be deduced.

This does not necessarily provide information on the length of the ride: It will be informative in the case of a commute, where a rider moved from A to B by the shortest route, for example in the course of a commute; it will be less informative in the case of a weekend ride on Lake Michigan, during which a rider spent three hours riding up and down the coast and had a coffee by the lake during this period, before leaving the bike at same bus-stop where he first picked it up. The 25-mile trajectory will go unrecorded, and all that can be found in the data is a distance of zero.

Due to this problematic nature of distance data, the information has not been used heavily for this analysis. A far richer context, incorporating fee waiver zones and the status of different neighborhoods, as well as fee structures and their change over time, would be required to make meaningful use of these data. As this effort goes beyond the scope of this exploratory study, distance data have not been used in the below analysis.

Meanwhile, it is possible to use distance data to calculate a speed measure, achieved by dividing the distance by the duration of the ride. It is expected that this measure should be higher for the weekend commuting period and lower for the leisure-heavy weekend.

4. Analysis

Section 4.1 embeds the analysis of Cyclistic rider categories into information on the role of cycling in the US in general, and specifically in Chicago, Illinois. Section 4.2 then formulates some hypotheses of the analysis. Section 4.3 provides presents data on the temporal distribution of rides, while Section 4.4 formulates some requirements of an eventual analysis of spatial data that is, however, beyond the scope of this analysis. Section 4.5 summarizes the findings.

4.1 Framing the Role of Cycling in the US and the Windy City

In the North American context, bicycles generally play a marginal part in the transportation ecosystem, although there are efforts to create a more prominent role for them in urban and particularly metropolitan locations. Both the inherent health benefits of self-propelled transportation and ecological considerations (lowering the carbon footprint) may induce riders to take up cycling, and public institutions to support such tendencies.

That said, for many Americans, bicycles are tools of leisure, rather than primarily a means of transportation, something to be driven by car to scenic locations, where the owners of suitably equipped mountain bikes then engage in conspicuous consumption by means of experientially motivated locomotion, without any connection to their daily transportation needs.

The City of Chicago has published dedicated bike policies in five-year intervals since 2005 as so-called *Chicago Streets for Cycling* plans (Chicago Department of Transportation (CDOT), undated). It awarded a contract for the operation of a public bike sharing system to a company named Alta Bicycle Share in 2012, which has since been acquired and renamed *Motivate,* and is currently owned by *Lyft, Inc.* In March 2019, Lyft was granted a nine-year operating contract for Divvy, which gave it "*the exclusive right to operate the city-owned system. and receive a portion of the subsequent advertisement revenue. The deal required Lyft to invest US$50 million to add 175 stations and 10,500 bikes to the system, expand to all 50 city wards by 2021, and add electric pedal bikes which could lock to both Divvy stations and conventional bike racks. Lyft would additionally be required to make annual payments to the city starting at US$6 million and increasing by 4 percent each year; the city would share in at least US$1.5 million in advertisement revenue each year*" (Buckley 2019).[3] While this association with Lyft funded a rapid expansion of the Divvy system, changes of the pricing structure and a declared policy of phasing out classic bikes and replacing them with the more expensive electric ones were met with criticism of the operator pricing economically vulnerable segments of the population out of bike share market (Rosas & Greenfield (2022)). Going forward, this may lead to modifications in previously observed rider behavior, as several members announced in response to Rosas and Greenfield's blog post. One example may illustrate the tenor of the discussion:

> *"The blue bikes are not free, they are what you are paying for by having a membership. If those bikes disappear, I have no reason to keep my membership as the e-assist Divvies make no sense for those of us with CTA monthly passes that often just need the last mile solution and/or like the exercise that the blue bikes provide."*
> Reader comment on Rosas & Greenfield (2022)

If Divvy were a competitor of Cyclistic, any strategic recommendations for the company would have to take this towering competition into account. However, this is not the case. As evidenced by Hartmann

[3] In its 2020 annual report, Lyft (2021:111) characterizes its obligations as follows: "*In May 2019, the Company entered into a noncancelable arrangement with the City of Chicago, with respect to the Divvy bike share program, under which the Company has an obligation to pay approximately $7.5 million per year to the City of Chicago through January 2028 and to spend a minimum of $50 million on capital equipment for the bike share program through January 2023. The Company has made payments totaling $15.0 million and investments totaling $19.7 million as of December 31, 2020.*"

(undated), who presents an earlier version of the case study without using the Cyclistic name, Cyclistic is Divvy, and the data analyzed in this case study, made available by Motivate, Inc., all reflect Divvy rides. This is probably as it has to be, as the need for a comprehensive network of docking stations may not technically turn ride service providers into natural monopolies, but does come close to doing so, particularly in the light of extreme outcomes from unregulated competition such as those seen in the so-called Chinese bicycle graveyards (Ngo 2021).

4.2 Hypotheses of the Analysis

With respect to the characterization of differences in cycling behavior between Cyclistic members and casual riders, it is hence of interest to investigate to what extent these two prevalent conceptions of cycling, as a leisure activity and as a means of transportation, can be mapped to these two user categories.

The expectation was that members, who pay a monthly fee and have signed up to an annual contract, are more committed to cycling and take it more seriously.

This gives rise to a number of hypotheses:

1. Members are more likely to commute regularly to work or school than non-members.
2. Members are less likely to use the bike primarily or exclusively for leisure purposes. As a result, weekend usage, during which the leisure component is expected to loom large, should be less relevant for members than for casual riders.
   Late-night use may also tend to involve more casual riders for similar reasons.
3. Members are also more likely than casual riders to ride their bike throughout the year, braving cool winter weather to get their money's worth or pursue their dream of a carbon-free existence even when this reduces comfort.
4. Members are expected to be found where the benefits of *Cyclistic* use are great thanks to a well-developed infrastructure. They are thus more likely to be active in quarters with numerous dedicated bike lanes and docking stations than in parts of the city where these are not found.
5. Members are more aware of pricing and less likely to engage in costly rides than casual users. In particular, they are likely to stay well below the limit beyond which they will have to pay for additional minutes.

About other aspects of interest, it will be impossible to say much on the basis of rides data alone. For example, the declared and hidden motivations of casual and member riders cannot be exhaustively analyzed, although some pointers may be found in the data. To assess these more thoroughly, qualitative and interactive analysis tools such as surveys and focus groups should be more expedient.

4.3 Evidence from Time-Related Data

The time data available show clear differences between both rider groups at the annual, weekly and hourly levels.

Over the year, variations in the proclivity to cycle are a natural outcome of seasonally varying weather conditions and of the holiday season. The expected differences are, however, far more pronounced for casual riders than for members.

In terms of individual rides, members maintain a ratio of 4.3 between the number of rides in the busiest month, September (392,257 rides), and the slowest month, January (85,250 rides). For casual riders, the same ratio amounts to a whopping 23.9 between the busiest month, July (442,056 rides), and the slowest month, January (18520 rides). The following graphic illustrates the changes in ride frequency over the year.

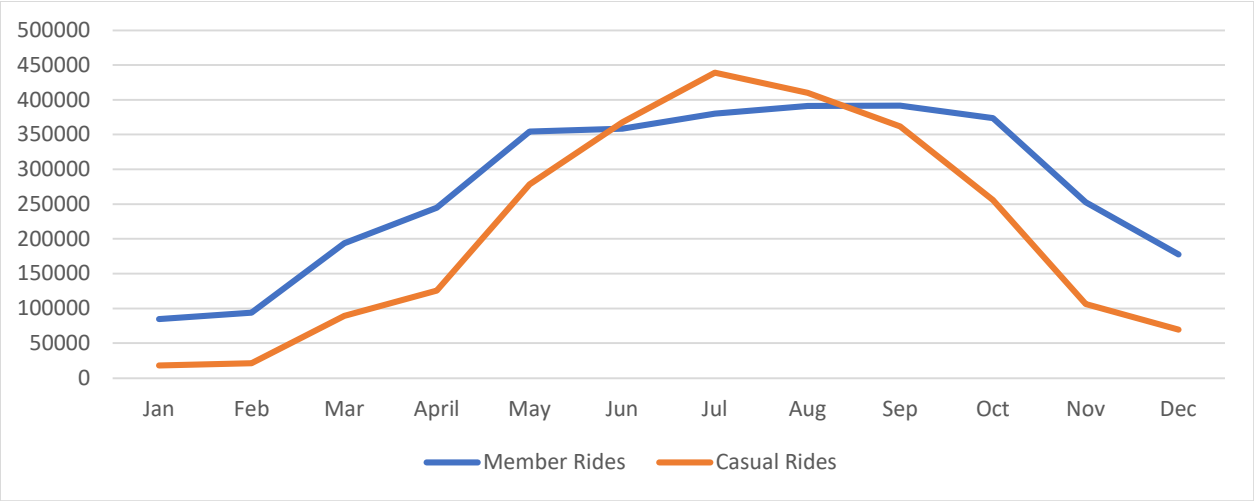


Figure 5: Member and Casual Rides by month

At the weekly level, a marked difference is found between the weekend distribution of rides and the weekday distribution of rides.

On weekdays, the aggregated data display two daily peaks, with a lower peak around 8 am in the morning and another, far higher one in the early evening between 4 pm and 8 pm. The existence of these peaks is compatible with the assumption that the commuter population contributes significantly to daily rides over this period, while the difference in their level reflects the fact that the non-commuting population is more active later in the day.

By contrast, aggregated weekend data only show one daily peak, reached early in the afternoon, with hourly averages rising monotonously before then and falling afterwards. The following graphics illustrate the contrast in the ride distribution between weekdays and weekends.

Figure 6: The twin-peaked total weekday pattern

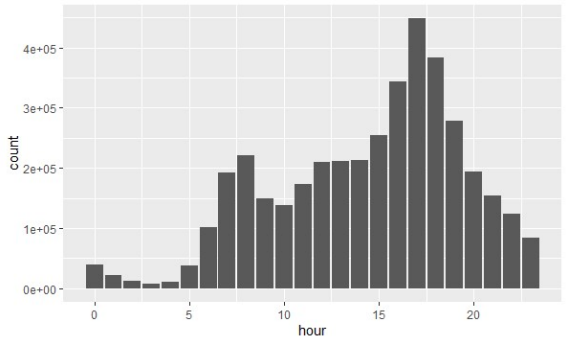
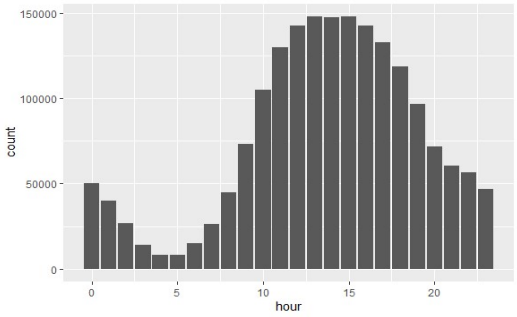


Figure 7 The single-peaked total weekend pattern



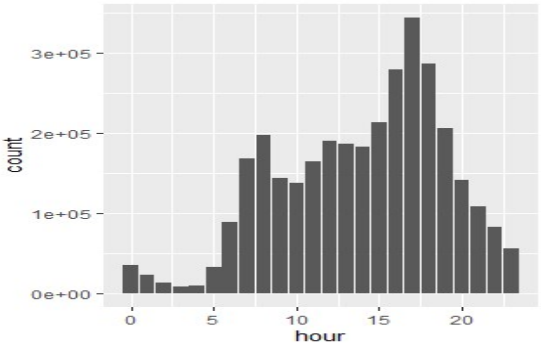


Figure 6: Overall member distribution

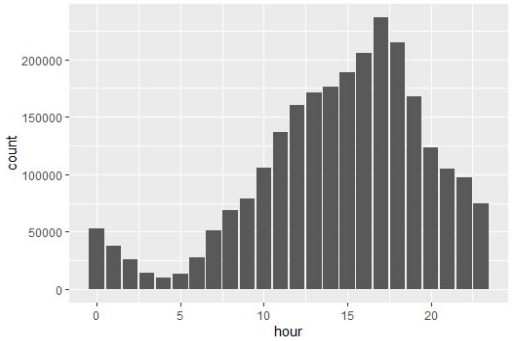


Figure 7: Overall casual distribution

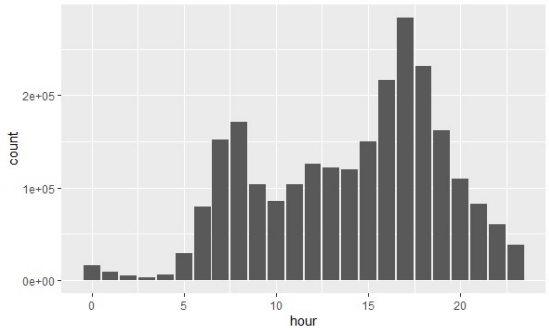


Figure 8: Member weekday distribution

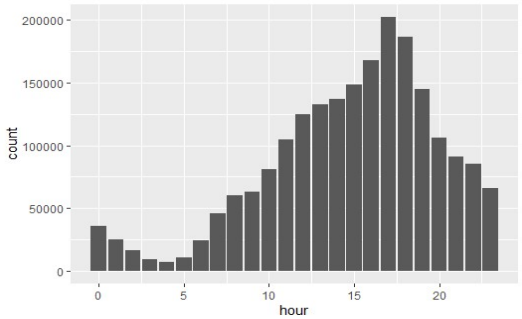


Figure 9 Casual weekday distribution

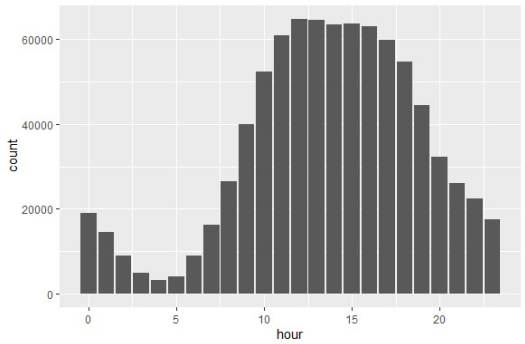


Figure 10: Member weekend distribution

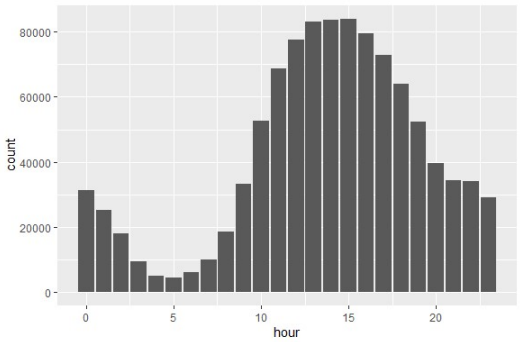


Figure 11: Casual weekend contribution

A look at the separate rider groups reveals significant differences, however. For members, a more accentuated twin-peak structure is found in the weekday rides, while for casual riders, there is only a single, evening peak, illustrating that commuting is largely a member affair. By contrast, all casual days have roughly the same structure, but the peak occurs earlier in the afternoon on weekends and around five pm on all other days.

By contrast, both groups have similar weekend patterns, although the weekend peak forms an earlier and more extended plateau in the case of members. Figures 8 to 13 illustrate these trends.

4.4 Evidence related to bike type

The data show a greater propensity for electric bike use among casual than among member riders, although the latter pay considerably less per unit of an electric bike ride than (USD 0.16 per kilometer) than casual riders (USD 0.39 per kilometer) in the only pricing plan available, that of the Divvy system.

Table 2: Some summary statistics on bike types

	Total	Members	Casual Riders
Number of electric bike rides	2,366,398	1,317,714	1,047,684
Percentage of electric bike rides	39.11%	37.90%	40.13%
Percentage of classic bike rides	60.88%	62.1%	59.87%
Median classic duration	11.52 minutes	9.57 minutes	15.6 minutes
Median electric duration	10.03 minutes	8.42 minutes	12.66 minutes

The following graphs illustrate the distribution of electric vs. classic ride durations:

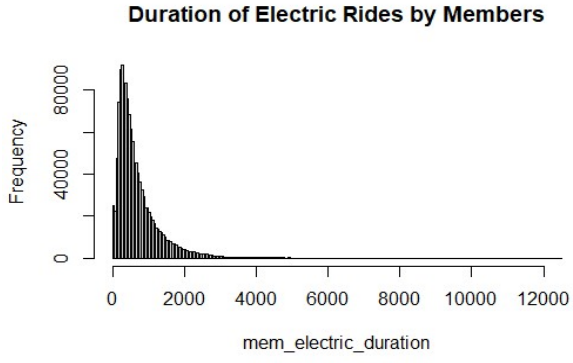


Figure 12 Electric ride Duration by Members

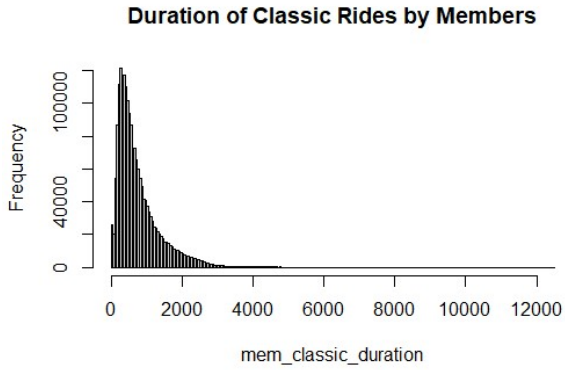


Figure 13: Duration of Classic Rides by Members

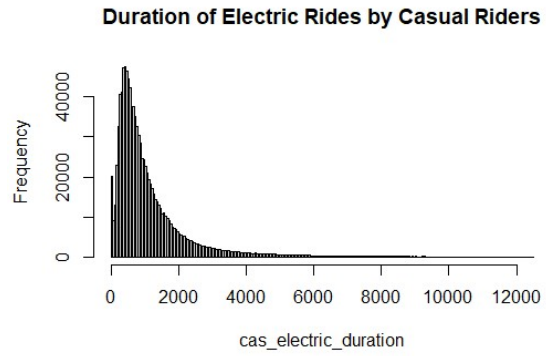


Figure 14: Duration of Electric Rides by Casual Riders

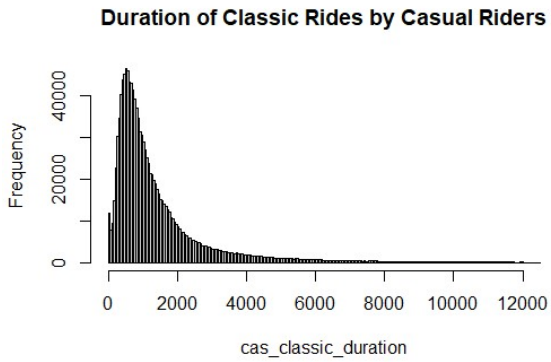


Figure 15 Duration of Classic Rides by Casual Riders

4.5 Evidence from Location-Related Data

The centrality of downtown and particularly the lake shore for bike users emerges clearly from the data. Out of the ten most popular departure locations in the network, six are within spitting distance of Lake Michigan and close to parks or cultural institutions such as the *Theatre on the Lake*, documenting the importance of leisure use for overall Divvy usage.

Most popular stations overall in the started at column (entire dataset, 822176 missing station names)			**Most popular departure locations by GPS coordinates (entire dataset)**	
1	Streeter Dr & Grand Ave	83768	Cloud Gate	41.892278, -87.612043
2	Wells St & Concord Ln	43973	Chicago Architecture River Cruise	41.79, -87.6
3	Michigan Ave & Oak St	43779	cloud Gate 2	41.911722, -87.626804
4	Millennium Park	40892	Chicago Yacht Club	41.880958, -87.616743
5	Clark St & Elm St	40770	N Wells St & W Concord Ln	41.912133, -87.634656
6	Wells St & Elm St	37464	Theater on the Lake, Chicago	41.926277, -87.630834

7	Kingsbury St & Kinzie St	36042	Clark St & Elm St, Chicago, IL 60610, United States	41.902973, -87.63128
8	Theater on the Lake	35807	Near North Side, Chicago, IL 60611, USA	41.90096039, -87.62377664
9	Clark St & Armitage Ave	32946	5225 S Blackstone Ave, Chicago, IL 60615, USA	41.8, -87.59
10	Wabash Ave & Grand Ave	32793	1143 N Wells St, Chicago, IL 60610, USA	41.903222, -87.634324
0	NA	822176	Over 822,000 rides are not associated with a station name.	

Figure 16: Most Popular Stations Overall

Time does not allow for an exhaustive analysis of all potentially relevant location data, but the most likely candidate for a different pattern shall be investigated: the weekday morning commute, taking place Monday to Friday, between 7 am and 9 am.

After filtering for this time frame and the weekday requirement, 221,121 observations remain. If there were to be a group of dedicated Divvy commuters riding to work on every one of the 260 weekdays in a year, then this group could hardly exceed 850 individuals, even if all other Chicago non-commuters abstained from riding Divvy bikes during this period. Even if Divvy commuters only commuted by bike every other workday, their number would not exceed 1,700 individuals. Ride numbers of morning commuters are thus far lower than for the most popular rides at across all times of the day. The following table contains information on this group:

Table 3: Facts about the morning commute slot

	Total	Member	Casual
Observations	221,121	171,077	50,044
Mean ride duration	12 min 12 s	11 min 4s	16 min 5s
Median ride duration	9 min	8 min 55s	10 min 36
Electric commutes	91,633 (41.4%)	66,381 (38.80%)	25,252 (50.45%)

What stands out here is the dominance of member rides, which outnumber casual rides by more than 3:1 in this timeslot. Also, as illustrated by the table below and the map in Figure 18, ride departure locations tend to be closer to major train stations (Ogilvie Transportation Center, Chicago Union Station, Clinton Street Metro Station), which points to the relevance of these bike rides for the last mile in a more complex commute involving other forms public transportation.

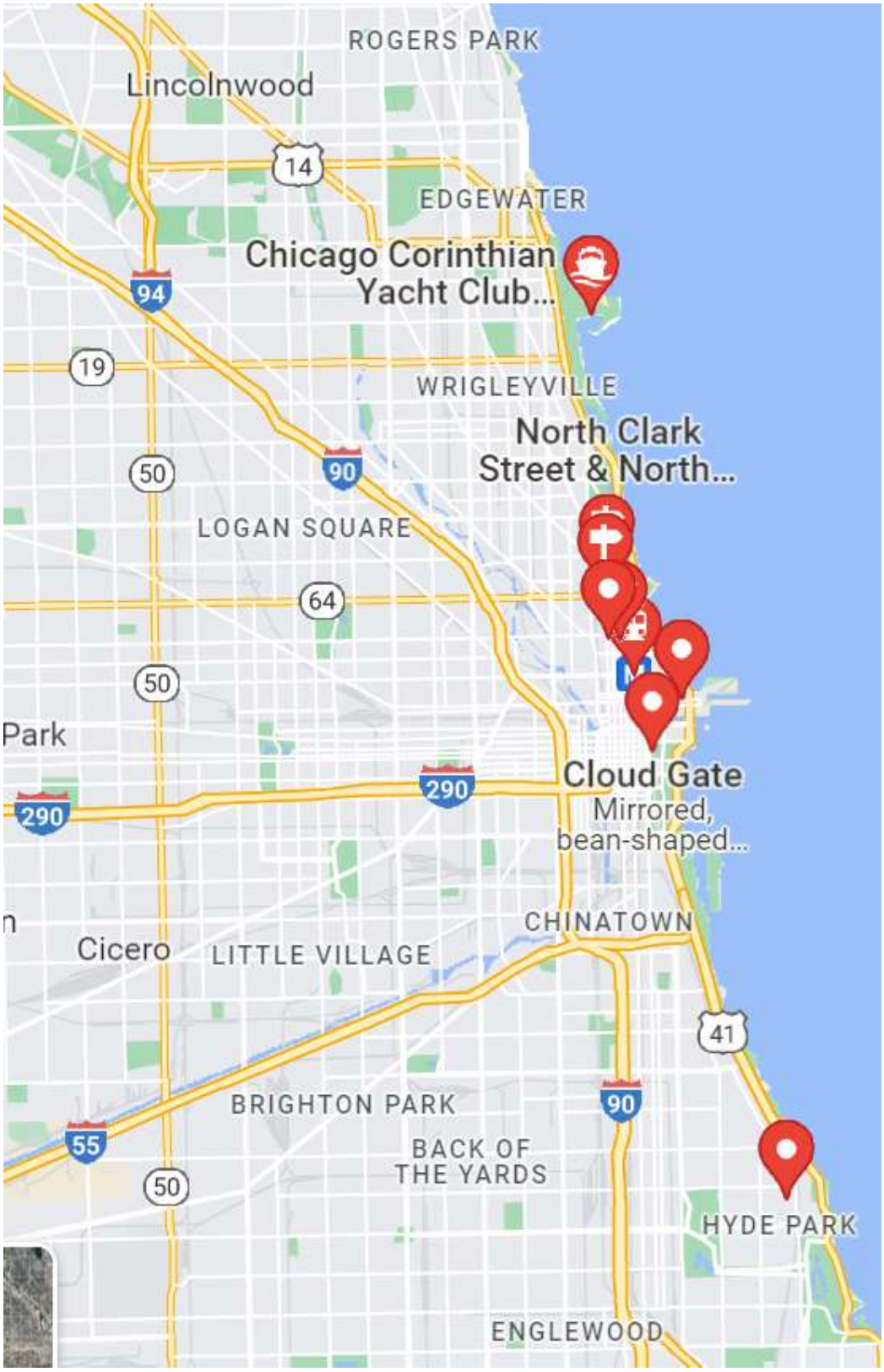


Figure 17: Overall, the most popular rides start from central lakeshore locations.
Source: Google Maps, using Divvy ride departure coordinates

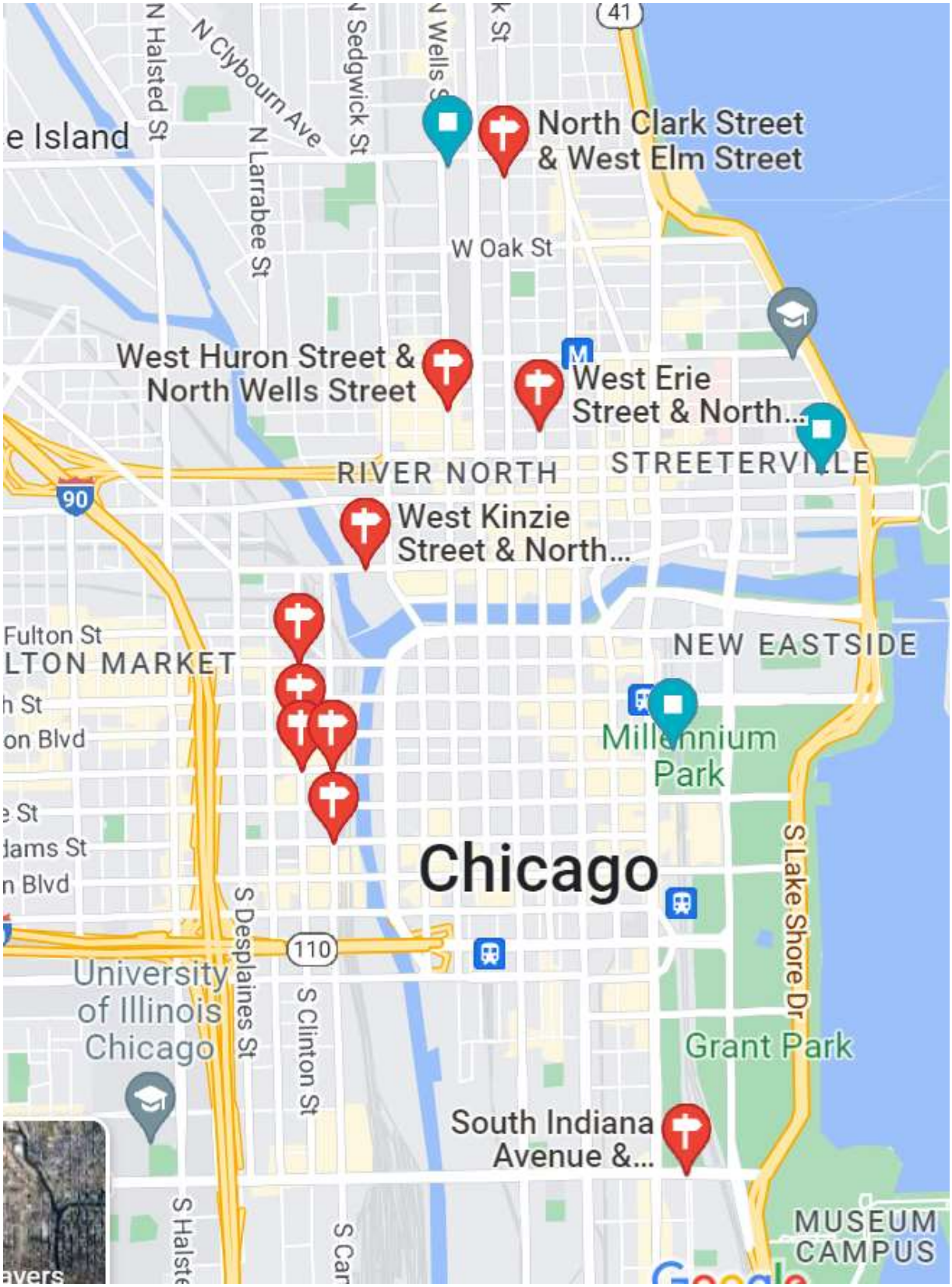


Figure 18: Commuter rides often start close to major train stations west of the Chicago river and stay clear of the lake shore. Popular departure points are marked by white signpost symbols on a red background.

Table 4: The ten most popular morning departure points

Rank by Riders	Starting Point	Annual Riders	Prominent Neighborhood Feature
1	Clinton St & Washington Blvd	2645	Ogilvie station
2	Clinton St & West Maddison Street	2431	Ogilvie Station
3	Kingsbury & Kinzie St	2287	Downtown
4	Canal St & Adams St	2018	Chicago Union Station
5	Wells St & Huron Street	1833	Downtown
6	Canal St & Madison Street	1812	Ogilvie Transportation Center
7	Clark ST & Elm St	1774	Downtown
8	Dearborn & Erie St	1654	Downtown
9	Clinton St & Lake St	1630	Ogilvie Stations, Clinton Metro Station
10	Indiana Ave & Roosevelt Rd	1614	Downtown, lake proximity, cultural

Due to time constraints, no extended research has been done on distance-related to date, given the reasons outlined in Section 3.4.

5 Evidence on the Hypotheses

Regarding the hypotheses formulated in Section 4.2, the following evidence has been found:

1. **Members are indeed more likely to commute regularly to work or school than non-members**. The prominence of the morning peak in member weekday data supports this hypothesis, and a closer look at the weekdays 7 to 9 am time slot shows that member rides outnumber casual rides by more than 3:1 in this time slot. However, the volume of morning commutes is relatively limited, with about 220,000 overall or an average of 850 per working day.
2. **Members are indeed less likely than non-members to use their subscription primarily or exclusively for leisure purposes**. The weekend ride incidence has been found to be less relevant for member rides than for casual rides.
   Late-night use has not been investigated yet at this point.
3. **Members are indeed more likely than casual riders to ride their bike throughout the year, braving cool winter weather**. What motivates them exactly has not been pinpointed.
4. The hypothesis on greater member prevalence in the vicinity of greater infrastructure has not been assessed.
5. **Member distributions indeed provide evidence on the price sensitivity of members**, with 98.3% of rides not exceeding the Divvy 45-minute free period, a significantly higher rate than in the case of casual riders (88.07%).

6 Recommendations to the Marketing Team

From the above analysis, the following recommendations to the marketing team emerge:

1. **Refrain from equating evidence from rides with evidence about riders**. Although this analysis has yielded some pointers, more pertinent information about the rider population can only be derived from a cluster analysis of anonymized rider data. It would be useful to consider the

viability of using anonymized rider information to be able to develop relevant information about the Cyclistic user population without compromising their privacy.
As it is, our data fail to capture cases where an individual purchases services from both categories. For example, a member may buy a day pass to go for an extended joyride, because the expected cost of the ride within his subscription would be superior to that of the more generous three-hour slots allowed for by the Divvy day pass. Our data currently do not capture these rides as belonging to the same individual, conveying a distorted picture of user dynamics by putting such weekend activities squarely in the casual camp.
It would also be desirable to be able to segment member and non-member activities by sub-cluster.
Thus, members may include both well-of individuals living in the city center and using the bikes largely for joy rides, as well as price-sensitive commuters seeking an economical alternative to other kinds of public transportation.
Another group may be individuals who support cycling for value-related reasons, with their economic circumstances having no role to play in their decision to cycle.
Among casual riders, the ability to capture the share of out-of-state tourists vs. residents would be important.

2. **Do not overestimate members' willingness to pay prime rates for premium services.** Data show a limited uptake of electric rides among member riders, who generally appear to be quite cost-sensitive, tending to take shorter rides than casual users, with more than 98% of all rides staying under the 45-minute grace period beyond which moderate per-minute fees become payable. Efforts to shift riders to greater e-bike use, for example by limiting the availability of 'classic' mechanical bikes, are likely to be met with ill-will by these riders (Rosas & Greenberg (2022)).
3. **Overall, casual riders appear to be more willing to pay for longer rides and more costly electric rides**. It is doubtful that shifting them to the member category will lead to greater overall revenue. Members do show a more sustained and regular use throughout the year, but as their rides are mostly short, this does not translate into additional revenue.

Bibliography:

Anonymous (undated), *Divvy Exercise R Script.* Referenced in the Coursera/Google study materials at https://docs.google.com/document/d/1TTj5KNKf4BWvEORGm10oNbpwTRk1hamsWJGj6qRWpuI/edit

Ngo, Hope (2021), "*Why some bikeshares work and some don't.*" BBC Future Planet feature, retrieved at https://www.bbc.com/future/article/20210112-the-vast-bicycle-graveyards-of-china on 2022/07/21

Buckley, Madeline (2019), "*Divvy to get $50 million upgrade from Lyft investment in exchange for ride revenue under contract proposal*." Published in the Chicago Tribune on 12 March, 2019, retrieved at https://www.chicagotribune.com/news/breaking/ct-met-divvy-expansion-deal-20190312-story.html on 2022/07/23.

Chicago Department of Transportation (CDOT) (undated, after 2015), *Chicago Streets for Cycling Plan 2020.* Chicago, Illinois.

Chicago Department of Transportation (CDOT) (June 2021), *Strategic Plan for Transportation.* Chicago, Illinois. Available online at https://www.chicago.gov/city/en/depts/cdot/supp_info/cdotstrategicplan.html

Florida, Richard (2011), *The Rise of the Creative Class Revisited*. Basic Books, New York

Hartman, Kevin (undated) "*'Sophisticated, Clear, and Polished': Divvy and Data Visualization*", available at https://artscience.blog/home/divvy-dataviz-case-study, referenced in the materials made available for this project at https://docs.google.com/document/d/1TTj5KNKf4BWvEORGm10oNbpwTRk1hamsWJGj6qRWpuI/edit

Illinois Department of Transportation (2022), *IDOT Annual Report 2021.*

Lyft (2021), *Annual Report 2020.* Brooklyn, New York. Available online at https://s27.q4cdn.com/263799617/files/doc_financials/2020/ar/Annual-Report-2020.pdf

Ma, Winston (2020), *"Here are 4 major bike-sharing trends from China after lockdown."* World Economic Forum, published online on July 22, 2020, at https://www.weforum.org/agenda/2020/07/4-big-bike-sharing-trends-from-china-that-could-outlast-covid-19/, retrieved on 22/07/21

Matloff, Norman (2012), The *Art of R Programming. A Tour of Statistical Software Design*. No starch press, San Francisco.

Matloff, Norman (2019) *FasteR: A Guide to Learning R*. Available from Github at https://github.com/matloff/fasteR

Matloff, Norman (2022) *Teaching R in a Kinder, Gentler, More Effective Manner: Use Base-R, not the Tidyverse*. Available from Github at https://github.com/matloff/TidyverseSkeptic

Rosas, Ruth & Greenfield, George (2022), "*Divvy's new pricing is unaffordable to many residents, and therefore inequitable.*" Post of 14 June, 2022, retrieved on 20 July, 2022 at https://chi.streetsblog.org/2022/06/14/divvys-new-pricing-is-unaffordable-to-many-residents-and-therefore-inequitable/.

Wickham, Hadley & Grolemund, Garrett (2017) *R for Data Science*. O'Reilly, a free electronic version is available at https://r4ds.had.co.nz/introduction.html#running-r-code