Cyclistic Case Study:

How Does a Bike-Share Navigate Speedy Success?

This case study represents Cyclistic bike-sharing data for the 2021 period to identify differences in rider habits to potentially further grow revenue and margin in the city of Chicago

About the company

- Launched in 2016
- Fleet of 5,824 geotracked bicycles
- Network of 692 stations across Chicago

Offer

Customer types	Pricing plans		
Casual	Single -ride	Full-day	
Members			Annual pass

Excluded information

- Product pricing
- Personal identifiable details of riders

Stakeholders and their intentions

Marketing analytics team
Team responsible (specifically S. Roberts) for collecting, analyzing, and reporting data within this presentation to solve the assigned business task, and subsequently help guide future marketing strategy.

Lily Moreno (Director of Marketing)
Development of campaigns and initiatives to promote the bike-share program. Lily believes the future success of the company lies in maximizing number of annual memberships.

Finance analytics team
Team responsible for analyzing and reporting financial metrics, which recently determined annual pass members are much more profitable than casual riders.

Executive team
Team who will decide whether to pursue the insights and recommendations from this presentation.

Case study data analysis roadmap

1. **Ask**
2. **Prepare**
3. **Process**
4. **Analyze**
5. **Share**
6. **Act**

Based on Google
6 phases of data analytics

Statement of business tasks

Primary
How do annual members and casual riders use Cyclistic bikes differently?

Secondary
How can Cyclistic use digital media to influence casual riders to become members?

As the analysis is for a 12 month period, data has been limited to 2021 only, and is sourced via a licensing agreement provided by Divvy to Google

Data hosting

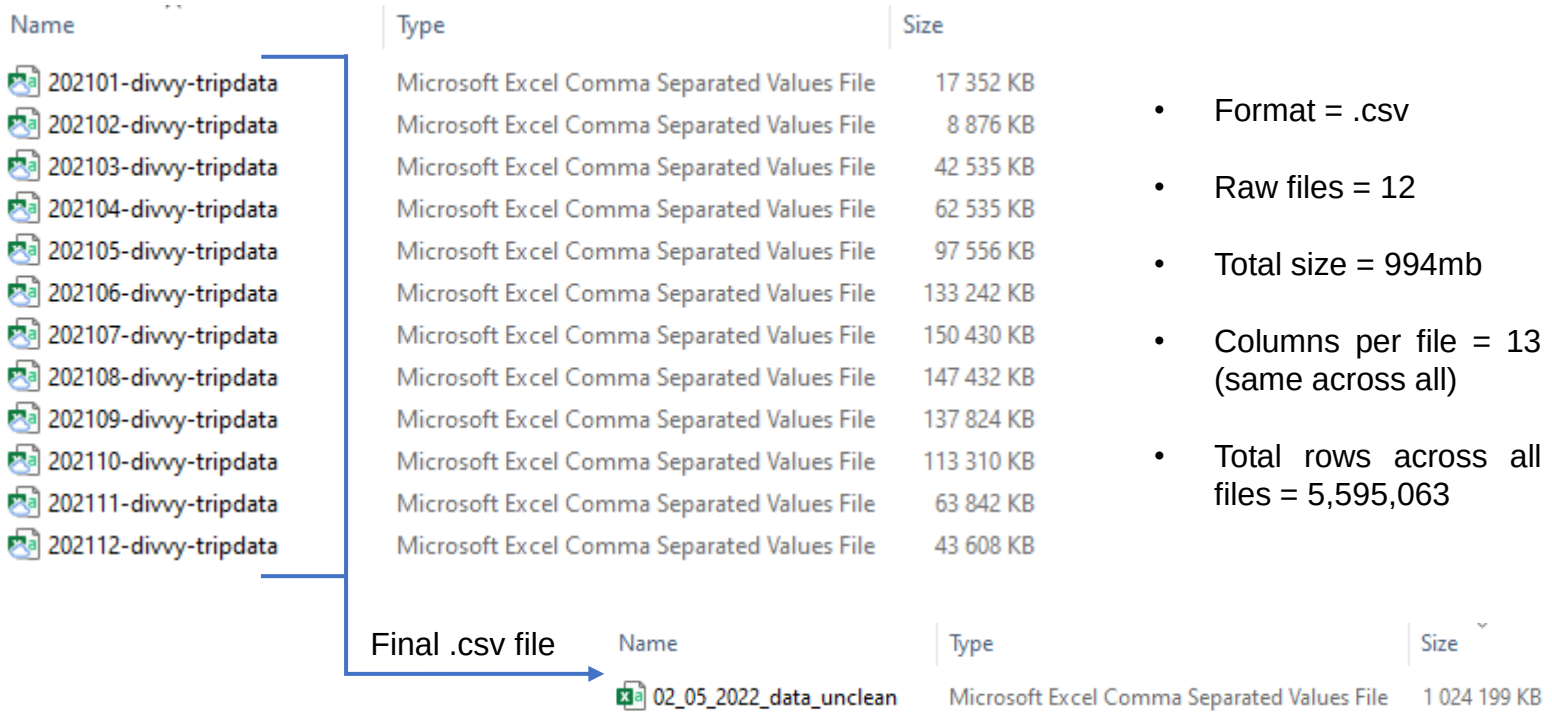
- Maintained by the Grow with Google team.
- Hosted by AWS.
- Link to source.

Data licensing

- Lyft Bikes and Scooters, LLC (“Bikeshare”) operates the City of Chicago’s (“City”) Divvy bicycle sharing service.

- Cyclistic is a fictional name given to the company representing the data.
- Link to data license agreement.

Data source files



- Format = .csv
- Raw files = 12
- Total size = 994mb
- Columns per file = 13 (same across all)
- Total rows across all files = 5,595,063

Sample of data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
2	89E7AA6C29227EFF	classic_bike	2021-02-12 16:14:56	2021-02-12 16:21:43	Glenwood Ave & Tou	525	Sheridan Rd & Colum	660	42.0127	-87.666	42.0046	-87.661	member
3	0FEFDE2603568365	classic_bike	2021-02-14 17:52:38	2021-02-14 18:12:09	Glenwood Ave & Tou	525	Bosworth Ave & How	16806	42.0127	-87.666	42.0195	-87.67	casual
4	E6159D746B2DBB91	electric_bike	2021-02-09 19:10:18	2021-02-09 19:19:10	Clark St & Lake St	KA1503000012	State St & Randolph S	TA1305000029	41.8858	-87.631	41.8849	-87.627	member
5	B32D3199F1C2E75B	classic_bike	2021-02-02 17:49:41	2021-02-02 17:54:06	Wood St & Chicago A	637	Honore St & Division	TA1305000034	41.8956	-87.672	41.9031	-87.674	member
6	83E463F23575F4BF	electric_bike	2021-02-23 15:07:23	2021-02-23 15:22:37	State St & 33rd St	13216	Emerald Ave & 31st S	TA1309000055	41.8347	-87.626	41.8382	-87.645	member
7	BDAA7E3494E8D545	electric_bike	2021-02-24 15:43:33	2021-02-24 15:49:05	Fairbanks St & Superi	18003	LaSalle Dr & Huron St	KP1705001026	41.8958	-87.62	41.8949	-87.632	casual
8	A772742351171257	classic_bike	2021-02-01 17:47:42	2021-02-01 17:48:33	LaSalle Dr & Huron St	KP1705001026	LaSalle Dr & Huron St	KP1705001026	41.8949	-87.632	41.8949	-87.632	casual
9	295476889D9B79F8	classic_bike	2021-02-11 18:33:53	2021-02-11 18:35:09	Fairbanks St & Superi	18003	Fairbanks St & Superi	18003	41.8957	-87.62	41.8957	-87.62	member
10	362087194BA4CC9A	classic_bike	2021-02-27 15:13:39	2021-02-27 15:36:36	LaSalle Dr & Huron St	KP1705001026	LaSalle Dr & Huron St	KP1705001026	41.8949	-87.632	41.8949	-87.632	member

R Studio Desktop proved to be the most efficient and capable method of both cleaning and visualizing the data through the use of the R Programming language

Tool selection

Data cleaning

- Due to the large size of the consolidated .csv file, **R Studio Desktop** was selected.
- Alternatives such as Excel and BigQuery (SQL) Free Cloud Edition could not efficiently handle the large .csv file sizes.

Data analysis

- For the key data points listed in the next few slides, smaller **Excel** .csv files with only relevant data variables were drawn via R Studio (relevant to the specific visualization) to conduct specific calculations in Excel.
- For the insights, these were drawn from the actual visualizations.

Data visualization

- For the data map visualization, **Tableau Public** was utilized as it is the most efficient tool which visualizing data in a map format.
- For the remaining visualizations, **R Studio Desktop** was selected as it could best display heat maps, bar charts & scatter plots.

Data cleaning process

1. Checking data types, formats, and identifying errors such as duplication
2. Removing error variables
   - 2.1. NA values
   - 2.2. Unnecessary character strings
3. Transforming data
   - 3.1. Creating columns for date specific variables (i.e. day of the week, months etc)
   - 3.2. Converting variables to factors, and subsequently ordering them
   - 3.3 Creating columns with calculated variables (i.e. ride length, ride start times etc)

Data observations

- Reduction of 5,595,063 observations (rows) to 5,590,0146
- Increase of 13 variables (columns) to 23

Snapshot of cleaning

```
data_clean$yyyy_mm_dd <- format(as.Date(data_clean$date), "%Y-%

data_clean$day_of_week <- as.factor(data_clean$day_of_week) # c
data_clean$day_of_week <- ordered(data_clean$day_of_week,
                                  levels = c("Sunday", "Monday"

data_clean$ride_length_seconds <- difftime(data_clean$ended_at,

data_clean$started_at_time <- format(as.POSIXct(
  data_clean$started_at), format = "%H:%M:%S") # create column
```

R Packages used

RStudio

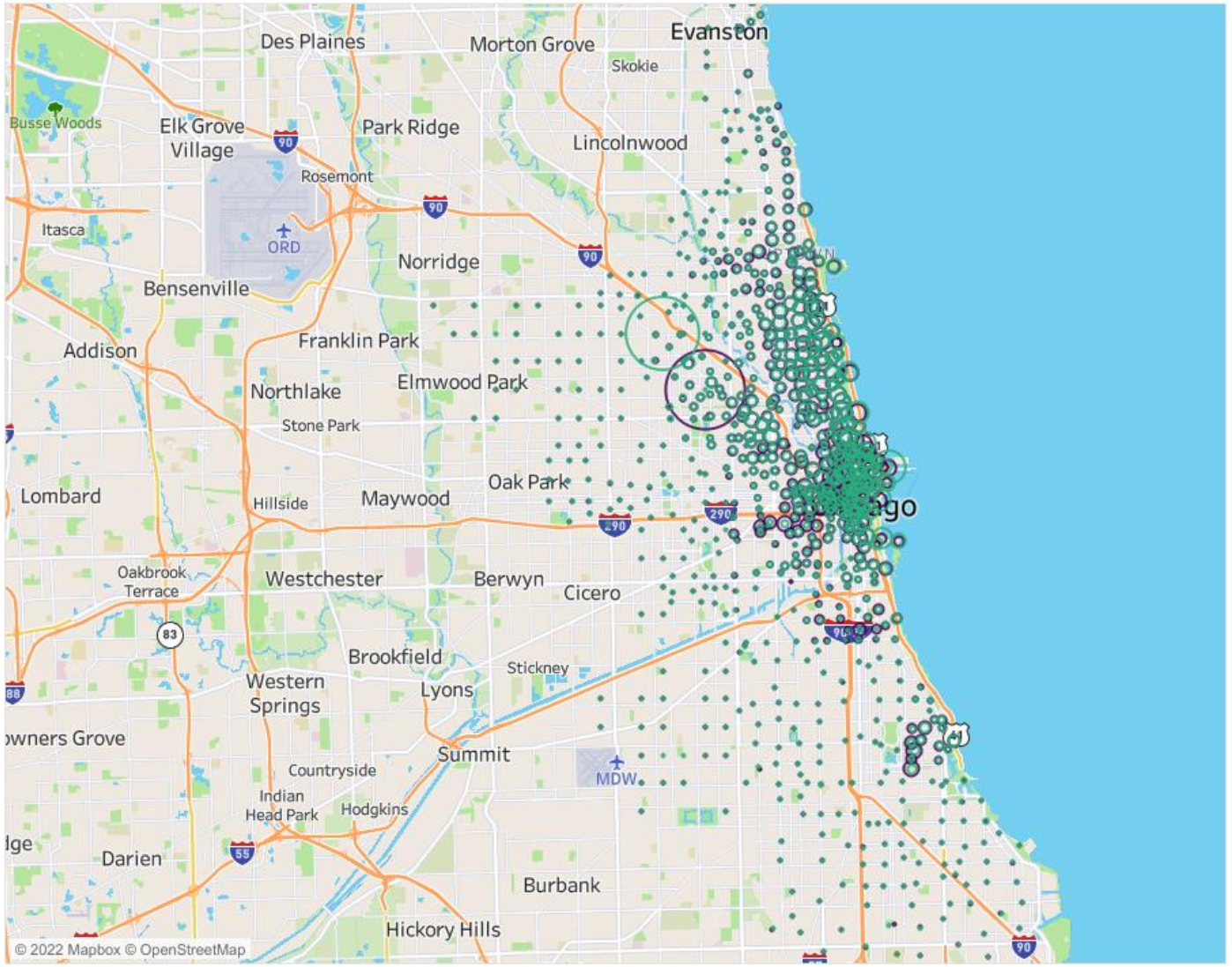
```
library("ggplot2")
library("plyr")
library("dplyr")
library("readr")
library("tidyr")
library("data.table")
library("stringr")
library("lubridate")
library("viridis")
library("scales")
```

R Markdown

Link to source.

While bike-sharing station options are available across the greater Chicago area, rides are heavily concentrated amongst the Central and Northern corridors, predominately lead by member riders

Total Annual Trips per Starting Station



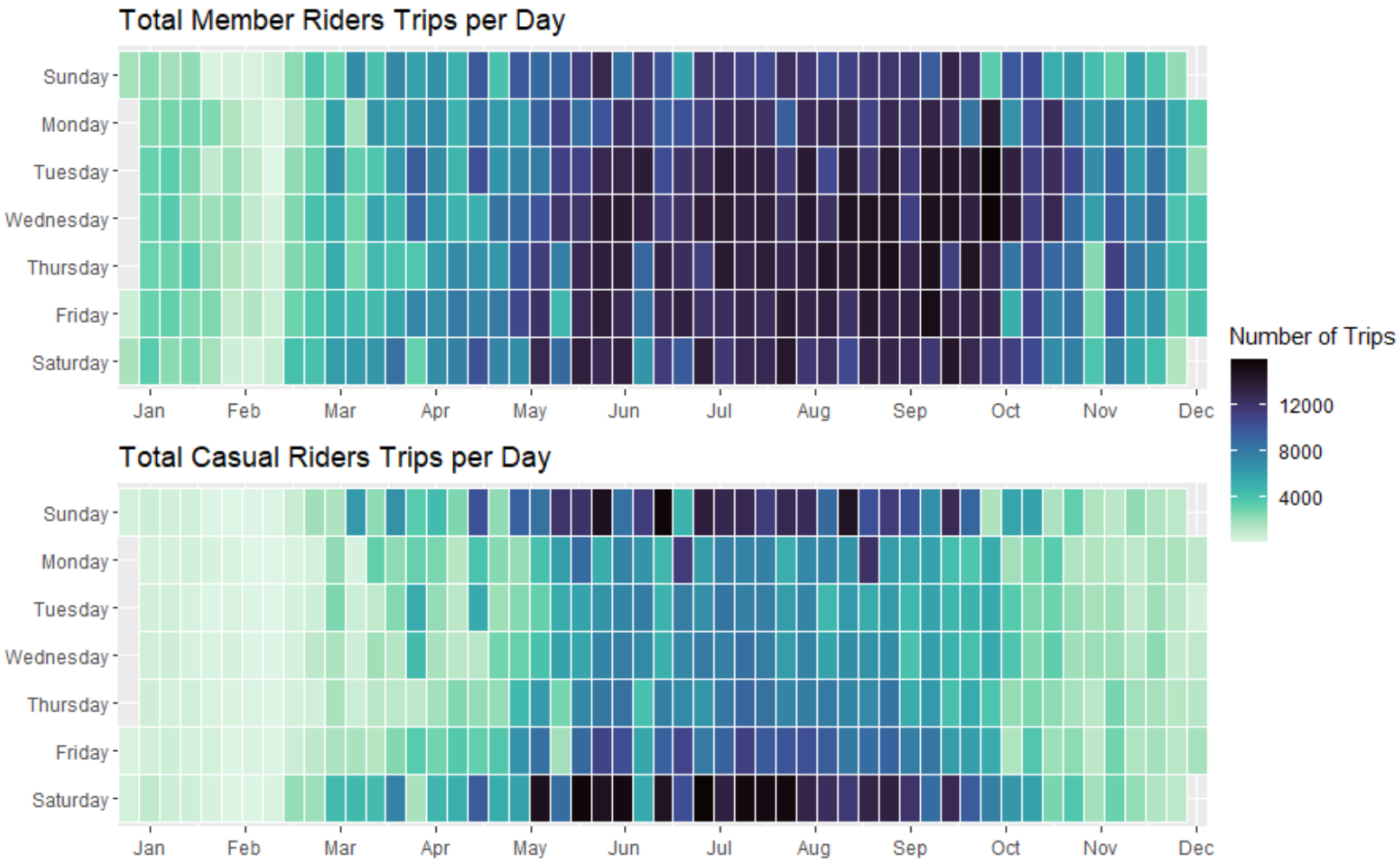
Key data points

- 849 unique stations
- Most annual trips at a station = 373,110 (6.7% of annual)
- Ave. trips per station = 6,584
- Distance from most Northern to Southern stations = 37.7 miles
- Distance from most Eastern to Western stations = 30.8 miles
- Members represent 55% of trips, while Casuals are 45%

Insights

- Member riders heavily concentrate their trips amongst Chicago's Central Business District (CBD), as well as the more trendy northern districts known for their restaurants and entertainment facilities
- Member riders utilize stations more frequently on a consistent basis than Casuals (22%+)
- Casual riders are much more geographically dispersed across the entire Chicago area.

While member riders see more usage over their bikes over the course of a whole week, casual riders, who outnumber members, see much more heavy usage over the weekend



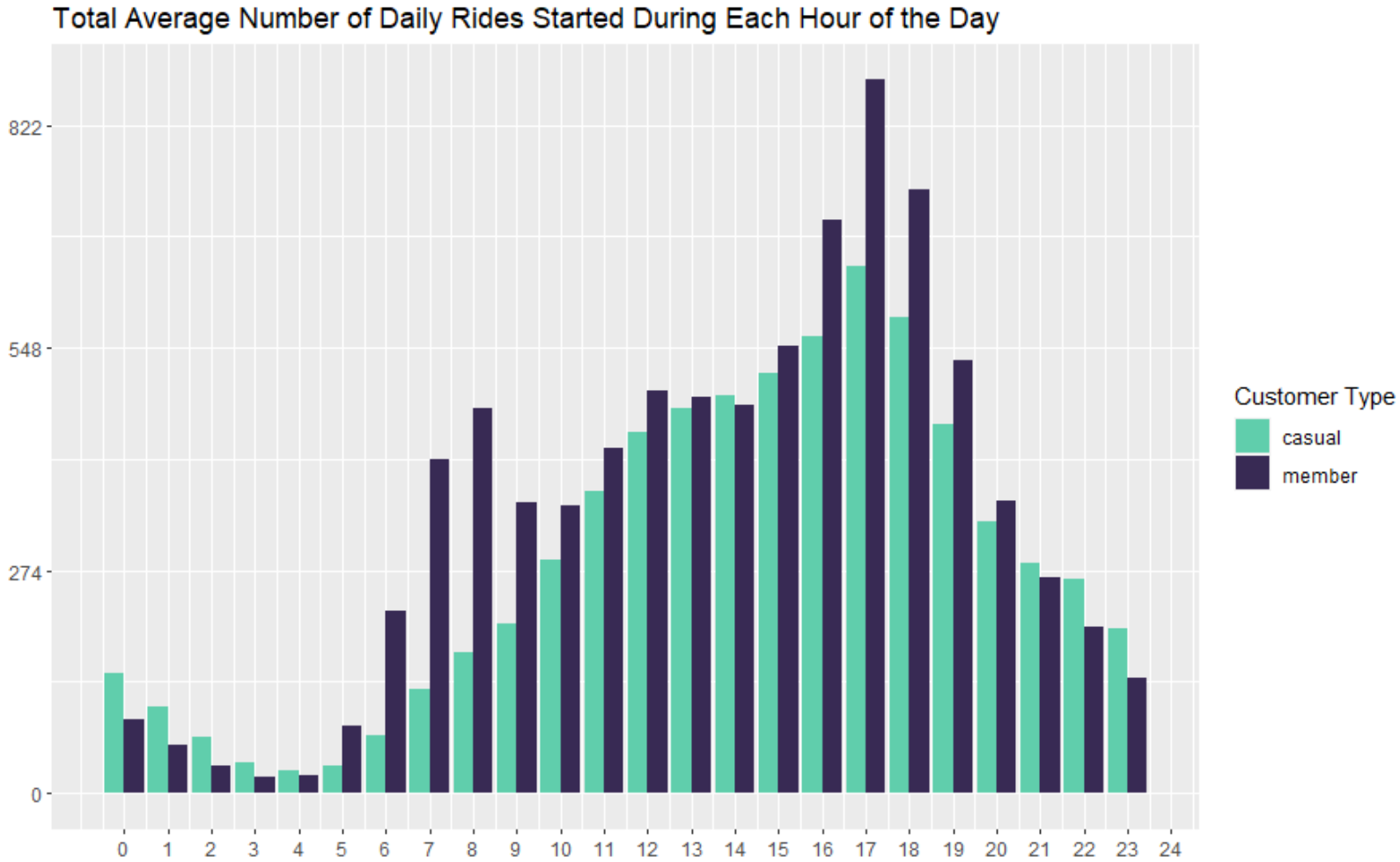
Key data points

- Total annual trips = 5,590,146
- Avg. trips per day = 7,658
- Winter months average trips = 131,178
- Summer months average trips = 784,719 (5.96x increase)
- Weekend avg. trips (member vs casual) = 7,777 vs 9,976
- Weekday avg. trips (member vs casual) = 8,643 vs 5,701

Insights

- Riders are much more likely to utilize the service in the warmer Summer months. This may also be due to more dangerous Winter conditions (e.g. ice, snow, darkness) present in Chicago during that period.
- Member riders trips are more concentrated during weekdays, while casual riders trips are more likely during weekends. This may reinforce the premise that members are workers commuting from work, to restaurants / entertainment, and back home.
- Casual riders are assumed to be tourists, as well as residents who may have the option between alternative transport modes available to them (e.g. rail, taxi's etc.)

Daily riders of both types tend to jump on their bikes more as the day progresses, seeing peak usage around the early hours of the evening



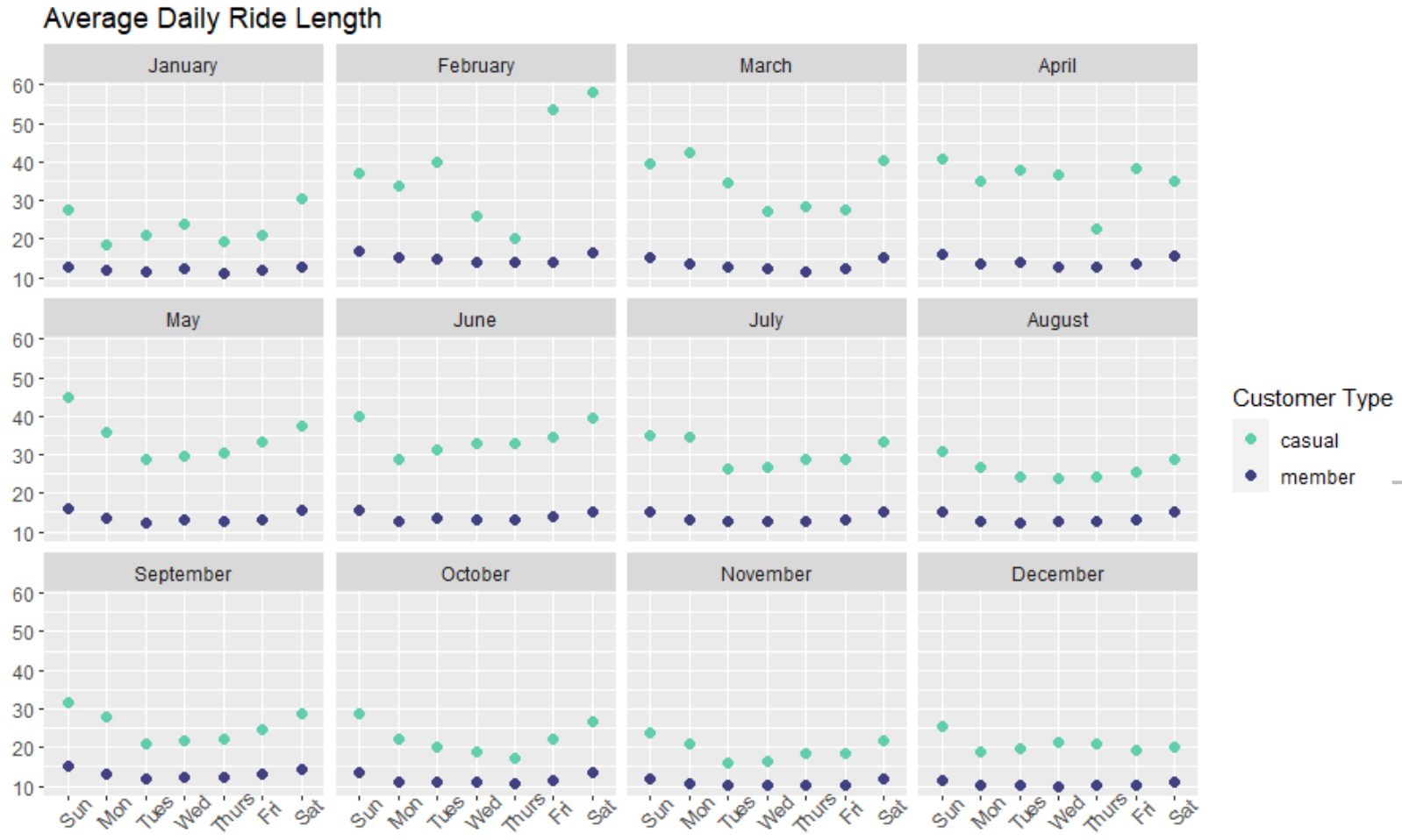
Key data points

- Daily total avg. trips (member vs casual) = 8,396 vs 6,919
- Daily low & peak hours (member & casual) = 04h00 & 18h00
- 25% of daily member rides are completed by midday, while only 19% of casual rides
- Between 07h00-08h00 daily, member riders account for 69% more rides
- 50% of casual rides are completed between 12h00-19h00 daily

Insights

- Member riders tend to peak during the early morning work commute (07h00-09h00), their lunchtime break (12h00-14h00), and the evening commute home (16h00-19h00)
- Casual riders seem to start the day later, with riders being more consistent from 11h00 onwards, while also finishing up much later in the early hours of the morning
- We can assume member riders have fixed schedules (i.e. workers), while casual riders are both tourists with no morning commitments and the ability to stay out later, but also parts of the rider community who may utilize alternative methods of mobility depending on conditions such as weather

Winter, as well as the festive period, seem to influence a decline in usage over the cooler months of the year, while there is a considerable uptick as the warmer weather moves in through Spring



Key data points

- Daily total avg. ride length (member vs casual) = 90 vs 201 minutes (123% increase)
- Lowest avg. months = November (casual) & December (member)
- Peak seasonal months (member vs casual) = Summer (June-August) vs Spring (March-May)
- Member & casual riders avg. daily ride length increases by 17-25% over weekends

Insights

- Member riders spend less than 50% of their time on their bikes than casual riders
- The colder Winter months see a slow down in usage, possibly due to lower tourist (i.e. casual) levels prior (Nov) to the Christmas period, while seeing workers (i.e. member) going on their annual holiday (Dec)
- All riders significantly increase usage over the weekends, showing the benefits of having a member pass that can allow for usage for both work and pleasure requirements

Opportunities, driven by further customer segmentation, are available to convert casual riders over to annual members if targeted digital advertising is conducted during specific periods of the year

Issues to be solved

Primary
How do annual members and casual riders use Cyclistic bikes differently?

Secondary
How can Cyclistic use digital media to influence casual riders to become members?

Final thoughts and recommendations

1. Casual riders are predominantly workers who have multiple means of transport, as well as tourists who require a **mobility solution** for them to see and traverse the city of Chicago, while member riders are those who see most of their usage through **weekday commuting**, while still seeing the benefits of using the bike-share program over the weekends. Further categorization should be conducted via:

    1.1 **Spend analysis** to see the frequency of repeat purchases by casual riders (opportunity to conduct upselling through promotions).

    1.2 **Segmenting** casual riders into both those who are actual tourists, and those who are locals (i.e. workers) in the Chicago area (targeting offers depending on where you are from).

    1.3 **Location analysis** to determine whether there is a need for the outlying (lower) usage stations (which see usage mostly by casual riders), or if there more benefit from promoting the higher concentration areas (like the CBD and the Northern corridor).

1. **In-app notifications** to push lower pricing to casual riders if they utilize the service in the traditionally lower peak hours between 06h00-09h00 – this may encourage them to see the benefits of using the bikes for their daily commute.

2. **Email promotions** in the month leading up to the seasonal period to promote lower annual membership prices if they sign up within the weaker Autumn and Winter months.

3. **Partnering** with local influencers and brands to show the benefits of bike sharing, whether it may be through the environmentally friendly nature of the transport, the benefits of not being confined in small and closed spaces with the covid-19 pandemic, or the intrinsic benefits of being healthy and active that come from utilizing a means of mobility that is human powered.