



Google Data Analytics Capstone

Cyclistic Case Study

;

Cyclistic: How can a bike-sharing company
maximize their annual memberships?

A case-study/project authored by

Waqar Shaozab

15 September, 2023

Introduction

Greetings! This case study comes from the **Google Data Analytics Professional Certificate** capstone project. It offers an in-depth exploration of the six critical stages of the data analysis process: **Ask, Prepare, Process, Analyze, Share, and Act.**

Allow me to introduce Cyclistic, a fictitious bike-sharing company based in Chicago, Illinois. In the year 2016, Cyclistic launched an innovative bike-share program, boasting an impressive fleet of **over 5,800 bicycles** and an extensive network of **600** docking stations. Setting itself apart from other bike-sharing companies, Cyclistic proudly provides a large selection of different bike options, including reclining bikes, hand tricycles, and cargo bikes. We want our bikes to be as inclusive and accessible for individuals with disabilities. Despite these options, only **8 percent** of riders use of these assistive features.



Bikes stationed in Chicago - Divvy

Furthermore, a substantial **30 percent** of Cyclistic's riders use the service for their daily work commutes, indicating that the majority of users primarily use our bikes leisurely. Cyclistic's current marketing strategy centers around creating a broad brand awareness and appealing to a wide consumer base. This strategy is dependant on the implementation of flexible pricing plans, encompassing single-ride passes, full-day passes, and annual passes. Customers who opt for single-ride and full-day passes fall into the category of

casual riders, while those who choose the annual passes are considered *Cyclistic members*.

Despite the efficacy of these pricing plans in attracting a diverse consumer base, it's important to note that **Lily Moreno, the Director of Marketing**, firmly believes that focusing on converting casual riders into annual members will provide substantial growth Cyclistic. Finance analysts at Cyclistic have also concluded that annual members contribute to a higher profit margin compared to casual riders. As a result, instead of initiating a marketing campaign targeting new customers, Moreno is confident that we can successfully transition casual riders into Cyclistic members.

Table of Contents

1	ASK	1
1.1	Key Tasks	1
1.2	Deliverable	1
1.3	Business Task	2
1.4	Business Objectives	2
1.5	Case-Study Deliverables	2
1.6	Key Stakeholders	2
2	PREPARE	3
2.1	Key Tasks	3
2.2	Deliverable	3
2.3	Information on Data Source	4
2.4	Limitations of Data	4
2.5	Is The Data ROCCC?	4
3	PROCESS	6
3.1	Key Tasks	6
3.2	Deliverable	6
3.3	Removed Duplicates	7
3.4	Checked Spelling	7
3.5	Searched for Null Values	7
3.6	Ensured <i>started_at</i> and <i>ended_at</i> Columns Had Consistent Timestamp Format	7
3.7	Used <code>=TRIM()</code> Function to Eliminate Extra Spaces	8
3.8	Used <code>=TRIM()</code> and <code>=PROPER()</code> Functions	8

3.9	Created New Column called <i>ride_length</i>	8
3.10	Created New Column Called <i>day_of_week</i>	8
4	ANALYZE	10
4.1	Key Tasks	10
4.2	Deliverable	10
4.3	Process and Analyze In R	11
4.4	Analyze in Excel	17
4.5	Create Pivot Table 1	18
4.6	Create Pivot Table 2	19
4.7	Summary of Analysis	21
5	SHARE	24
5.1	Key Tasks	24
5.2	Deliverable	24
5.3	Visualization in R	25
5.4	Visualization in Excel	26
6	ACT	30
6.1	Key Tasks	30
6.2	Deliverable	30
6.3	Three Recommendations	31
	References	32

Chapter 1

ASK



Asking the right questions are essential to the first phases of data analysis. If your questions are **S**pecific, **M**easurable, **A**ction-Oriented, **R**elevant, **T**ime-Bound, or **S.M.A.R.T.**, then you know your answers will be as thorough and accurate as possible. This will then lead to a complete and actionable analysis.

1.1 Key Tasks

- Identify the business task
- Consider key stakeholders

1.2 Deliverable

- ☐ Clear statement of the business task

1.3 Business Task

Design marketing strategies aimed at converting casual riders into annual members.

1.4 Business Objectives

- How do annual members and casual riders use Cyclistic bikes differently?
- Why would casual riders buy Cyclistic annual memberships?
- How can Cyclistic use digital media to influence casual riders to become members?

1.5 Case-Study Deliverables

1. A clear statement of the business task.
2. A description of all data sources used.
3. Documentation of any cleaning or manipulation of data.
4. A summary of analysis.
5. Supporting visualizations and key findings.
6. Three recommendations based on the analysis.

1.6 Key Stakeholders

- Lily Moreno Director of Marketing and Manager: Responsible for developing and implementing initiatives and campaigns for the bike-share program.
- Cyclistic marketing analytics team: A team of data analysts guiding Cyclistics marketing strategy.
- Cyclistic executive team: Any new marketing programs are subject to approval by the highly-detailed executive team.

Chapter 2

PREPARE



2.1 Key Tasks

- Download data and store it appropriately
- Identify how it's organized
- Sort and filter data

2.2 Deliverable

- Description of all data sources used.

2.3 Information on Data Source

- The data is stored **here** via 57 datasets saved as **.CSV (comma-separated values)** files ranging from years **2013-2023**.
- The datasets are publicly available for this case-study by Motivate International Inc. under this **license**. <https://www.overleaf.com/project/64a57b0b8f766fab10728056>
- The original attributes that were captured include: *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* with a total of 13 columns.
- Downloaded and unzipped all 12 files that corresponded to the year 2022.
- Housed the original files in a separate folder.
- Created sub-folders for .CSV files and .XLSX files with appropriate naming conventions.

2.4 Limitations of Data

- Data-privacy issues forbids us from accessing riders' personally identifiable information. That means I am unable to connect pass purchases to credit card numbers to determine some indicators like rider home location to Cyclistic services or if they purchased multiple single passes.
- Of the historical bike trip data, we are only using datasets that correspond to the last **12 months**.

2.5 Is The Data ROCCC?

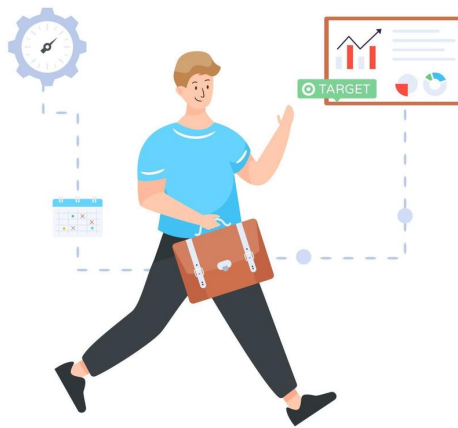
A credible data source is **ROCCC** which stands for **R**eliable, **O**riginal, **C**omprehensive, **C**urrent, and **C**ited.

1. **Reliable** - YES - Large sample size; representative of population.
2. **Original** - YES - Primary source of data comes internally from Cyclistic.

3. **Comprehensive** - MED - Data can be easily deciphered; does contain null values but not enough to lead to an incomplete analysis.
4. **Current** - YES - The most relevant data is available; entirety of the 2022 calendar year.
5. **Cited** - YES - Sources comes directly from internal databases; no third-party was involved.

Chapter 3

PROCESS



Work in Process

3.1 Key Tasks

- Check the data for errors.
- Choose your tools.
- Transform the data so you can work with it effectively.
- Document the cleaning process.

3.2 Deliverable

- ☐ Documentation of any cleaning and manipulation of data.

The process phase is an essential part for high-level decision making. This means cleaning and transforming data from dirty to clean in order to prepare for analysis. But the cleaner your data is, the better and more effective your analysis will be. Dirty data will lead to unreliable insights.

A number of tools can be used to process data like spreadsheets, SQL, and R. For this case-study, we'll start by cleaning the data using **Microsoft Excel** and then move over to coding in **R** due its ability to process large datasets.

Below will outline data cleaning process in Excel:

3.3 Removed Duplicates

1. *Data → Data Tools → Remove Duplicates → Select Columns → Ok*

3.4 Checked Spelling

1. *Select Column → Review → Proofing → Spelling*

3.5 Searched for Null Values

1. *Home → SelectSheet → Editing → Find and Select → GoTo Special → Blanks → Highlight Blanks Yellow*

Note: Null values fall under columns: *start_station_name start_station_id end_station_name end_station_id end_lat*, and *end_lng*. This could be due to customers not retrieving and docking their bikes at an appropriate station.

3.6 Ensured *started_at* and *ended_at* Columns Had Consistent Timestamp Format

1. *Home → Number → FormatCells → Time/Date → dd - mm - yyyyhh : mm : ss → Ok*

3.7 Used **=TRIM()** Function to Eliminate Extra Spaces

1. *Insert Column* → **=TRIM(A2, B3 etc...)** → *Drag Values Down* → *Copy New Column* → *Use Paste Special on Old Column* → *Values* → *Ok* → *Delete New Column*

3.8 Used **=TRIM()** and **=PROPER()** Functions

1. To be done on columns *start_station_id* and *end_station_id*
2. This is to eliminate spaces and format correctly
3. *Insert Column* → **=TRIM(PROPER(F2))** → *Drag Values Down* → *Copy New Column* → *Use Paste Special on Old Column* → *Values* → *Ok* → *Delete New Column*

3.9 Created New Column called ***ride_length***

1. Calculated Length of Each Ride by Subtracting *started_at* From the *ended_at* Columns
2. **=(D2-C2)**
3. Formatted Time as HH:MM:SS (37:30:55)
4. *Home* → *Number* → *Format Cells* → *Time* → *HH : MM : SS*(37 : 30 : 55)

3.10 Created New Column Called ***day_of_week***

1. Calculated Day of Week Each Ride Began Using the **=WEEKDAY** Command.
2. **=WEEKDAY(C2,1)**
3. Formatted as Number With No Decimals
4. *Home* → *Number* → *Format Cells* → *Number* → *Decimal Places* : 01 = *Sunday*, 7 = *Saturday*

These steps are to be done for each .CSV file.

Below will outline the data cleaning/manipulation process in R.

Chapter 4

ANALYZE



4.1 Key Tasks

1. Aggregate your data so it's useful and accessible.
2. Organize and format your data.
3. Perform calculations.
4. Identify trends and relationships.

4.2 Deliverable

- ☐ Summary of analysis

4.3 Process and Analyze In R

```
# Install and load all required packages
```{r}
install.packages("tidyverse")
install.packages("lubridate")
install.packages("ggplot2")
install.packages("dplyr")

library(tidyverse)
library(lubridate)
library(ggplot2)
library(dplyr)
library(hms)

```
```

```
## Collect Data
# Upload Cyclistic datasets (csv files) here
```{r}
JAN_2022.csv <- read_csv("JAN_2022")
FEB_2022.csv <- read_csv("FEB_2022")
MAR_2022.csv <- read_csv("MAR_2022")
APR_2022.csv <- read_csv("APR_2022")
MAY_2022.csv <- read_csv("MAY_2022")
JUN_2022.csv <- read_csv("JUN_2022")
JUL_2022.csv <- read_csv("JUL_2022")
AUG_2022.csv <- read_csv("AUG_2022")
SEP_2022.csv <- read_csv("SEP_2022")
OCT_2022.csv <- read_csv("OCT_2022")
NOV_2022.csv <- read_csv("NOV_2022")
DEC_2022.csv <- read_csv("DEC_2022")

```
```

```
## Wrangle Data and Combine into a Single File
# Compare column names for each of the files
```{r}
colnames(JAN_2022)
colnames(FEB_2022)
colnames(MAR_2022)
colnames(APR_2022)
colnames(MAY_2022)
colnames(JUN_2022)
colnames(JUL_2022)
colnames(AUG_2022)
colnames(SEP_2022)
colnames(OCT_2022)
colnames(NOV_2022)
colnames(DEC_2022)

```
```


[illegible]

```

# Inspect the dataframes and look for incongruencies
```{r}
str(JAN_2022)
str(FEB_2022)
str(MAR_2022)
str(APR_2022)
str(MAY_2022)
str(JUN_2022)
str(JUL_2022)
str(AUG_2022)
str(SEP_2022)
str(OCT_2022)
str(NOV_2022)
str(DEC_2022)
```

```

A link for the str() tibble can be found here <file:///Users/waqarshaozab/Downloads/str.tibble.html>

```

# Convert ride_id and rideable_type to character so that they can stack
correctly
```{r}
JAN_2022 <- mutate(JAN_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
FEB_2022 <- mutate(FEB_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
MAR_2022 <- mutate(MAR_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
APR_2022 <- mutate(APR_2022,ride_id =
as.character(ride_id),rideable_type = as.character(rideable_type))
MAY_2022 <- mutate(MAY_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
JUN_2022 <- mutate(JUN_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
JUL_2022 <- mutate(JUL_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
AUG_2022 <- mutate(AUG_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
SEP_2022 <- mutate(SEP_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
OCT_2022 <- mutate(JAN_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
NOV_2022 <- mutate(JAN_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
DEC_2022 <- mutate(DEC_2022,ride_id = as.character(ride_id),rideable_type
= as.character(rideable_type))
```

# Stack individual month's data frames into one big data frame
```{r}
all_trips <- bind_rows(JAN_2022,FEB_2022,MAR_2022,APR_2022,MAY_2022,JUN_2
022,JUL_2022,AUG_2022,SEP_2022,OCT_2022,NOV_2022,DEC_2022)
```

##Clean Up Data to Prepare for Analysis
# Inspect the new table that has been created
```{r}
colnames(all_trips)
nrow(all_trips)
dim(all_trips)
head(all_trips)
tail(all_trips)
str(all_trips)
summary(all_trips)
```

```

```

R Console
[1] "tbl_df"      "tbl" data frame
[2] "start_at"     "start_station_name"
[3] "start_lng"    "start_lat"
[4] "end_at"       "end_station_name"
[5] "end_lng"      "end_lat"
[6] "rideable_type"
[7] "started_at"   "ended_at"
[8] "start_station_id"
[9] "end_station_id"
[10] "ride_id"
[11] "rideable_type"
[12] "started_at"
[13] "ended_at"
[14] "start_station_name"
[15] "end_station_name"
[16] "start_station_id"
[17] "end_station_id"
[18] "rideable_type"
[19] "started_at"
[20] "ended_at"
[21] "start_station_name"
[22] "end_station_name"
[23] "start_station_id"
[24] "end_station_id"
[25] "rideable_type"
[26] "started_at"
[27] "ended_at"
[28] "start_station_name"
[29] "end_station_name"
[30] "start_station_id"
[31] "end_station_id"
[32] "rideable_type"
[33] "started_at"
[34] "ended_at"
[35] "start_station_name"
[36] "end_station_name"
[37] "start_station_id"
[38] "end_station_id"
[39] "rideable_type"
[40] "started_at"
[41] "ended_at"
[42] "start_station_name"
[43] "end_station_name"
[44] "start_station_id"
[45] "end_station_id"
[46] "rideable_type"
[47] "started_at"
[48] "ended_at"
[49] "start_station_name"
[50] "end_station_name"
[51] "start_station_id"
[52] "end_station_id"
[53] "rideable_type"
[54] "started_at"
[55] "ended_at"
[56] "start_station_name"
[57] "end_station_name"
[58] "start_station_id"
[59] "end_station_id"
[60] "rideable_type"
[61] "started_at"
[62] "ended_at"
[63] "start_station_name"
[64] "end_station_name"
[65] "start_station_id"
[66] "end_station_id"
[67] "rideable_type"
[68] "started_at"
[69] "ended_at"
[70] "start_station_name"
[71] "end_station_name"
[72] "start_station_id"
[73] "end_station_id"
[74] "rideable_type"
[75] "started_at"
[76] "ended_at"
[77] "start_station_name"
[78] "end_station_name"
[79] "start_station_id"
[80] "end_station_id"
[81] "rideable_type"
[82] "started_at"
[83] "ended_at"
[84] "start_station_name"
[85] "end_station_name"
[86] "start_station_id"
[87] "end_station_id"
[88] "rideable_type"
[89] "started_at"
[90] "ended_at"
[91] "start_station_name"
[92] "end_station_name"
[93] "start_station_id"
[94] "end_station_id"
[95] "rideable_type"
[96] "started_at"
[97] "ended_at"
[98] "start_station_name"
[99] "end_station_name"
[100] "start_station_id"
[101] "end_station_id"
[102] "rideable_type"
[103] "started_at"
[104] "ended_at"
[105] "start_station_name"
[106] "end_station_name"
[107] "start_station_id"
[108] "end_station_id"
[109] "rideable_type"
[110] "started_at"
[111] "ended_at"
[112] "start_station_name"
[113] "end_station_name"
[114] "start_station_id"
[115] "end_station_id"
[116] "rideable_type"
[117] "started_at"
[118] "ended_at"
[119] "start_station_name"
[120] "end_station_name"
[121] "start_station_id"
[122] "end_station_id"
[123] "rideable_type"
[124] "started_at"
[125] "ended_at"
[126] "start_station_name"
[127] "end_station_name"
[128] "start_station_id"
[129] "end_station_id"
[130] "rideable_type"
[131] "started_at"
[132] "ended_at"
[133] "start_station_name"
[134] "end_station_name"
[135] "start_station_id"
[136] "end_station_id"
[137] "rideable_type"
[138] "started_at"
[139] "ended_at"
[140] "start_station_name"
[141] "end_station_name"
[142] "start_station_id"
[143] "end_station_id"
[144] "rideable_type"
[145] "started_at"
[146] "ended_at"
[147] "start_station_name"
[148] "end_station_name"
[149] "start_station_id"
[150] "end_station_id"
[151] "rideable_type"
[152] "started_at"
[153] "ended_at"
[154] "start_station_name"
[155] "end_station_name"
[156] "start_station_id"
[157] "end_station_id"
[158] "rideable_type"
[159] "started_at"
[160] "ended_at"
[161] "start_station_name"
[162] "end_station_name"
[163] "start_station_id"
[164] "end_station_id"
[165] "rideable_type"
[166] "started_at"
[167] "ended_at"
[168] "start_station_name"
[169] "end_station_name"
[170] "start_station_id"
[171] "end_station_id"
[172] "rideable_type"
[173] "started_at"
[174] "ended_at"
[175] "start_station_name"
[176] "end_station_name"
[177] "start_station_id"
[178] "end_station_id"
[179] "rideable_type"
[180] "started_at"
[181] "ended_at"
[182] "start_station_name"
[183] "end_station_name"
[184] "start_station_id"
[185] "end_station_id"
[186] "rideable_type"
[187] "started_at"
[188] "ended_at"
[189] "start_station_name"
[190] "end_station_name"
[191] "start_station_id"
[192] "end_station_id"
[193] "rideable_type"
[194] "started_at"
[195] "ended_at"
[196] "start_station_name"
[197] "end_station_name"
[198] "start_station_id"
[199] "end_station_id"
[200] "rideable_type"
[201] "started_at"
[202] "ended_at"
[203] "start_station_name"
[204] "end_station_name"
[205] "start_station_id"
[206] "end_station_id"
[207] "rideable_type"
[208] "started_at"
[209] "ended_at"
[210] "start_station_name"
[211] "end_station_name"
[212] "start_station_id"
[213] "end_station_id"
[214] "rideable_type"
[215] "started_at"
[216] "ended_at"
[217] "start_station_name"
[218] "end_station_name"
[219] "start_station_id"
[220] "end_station_id"
[221] "rideable_type"
[222] "started_at"
[223] "ended_at"
[224] "start_station_name"
[225] "end_station_name"
[226] "start_station_id"
[227] "end_station_id"
[228] "rideable_type"
[229] "started_at"
[230] "ended_at"
[231] "start_station_name"
[232] "end_station_name"
[233] "start_station_id"
[234] "end_station_id"
[235] "rideable_type"
[236] "started_at"
[237] "ended_at"
[238] "start_station_name"
[239] "end_station_name"
[240] "start_station_id"
[241] "end_station_id"
[242] "rideable_type"
[243] "started_at"
[244] "ended_at"
[245] "start_station_name"
[246] "end_station_name"
[247] "start_station_id"
[248] "end_station_id"
[249] "rideable_type"
[250] "started_at"
[251] "ended_at"
[252] "start_station_name"
[253] "end_station_name"
[254] "start_station_id"
[255] "end_station_id"
[256] "rideable_type"
[257] "started_at"
[258] "ended_at"
[259] "start_station_name"
[260] "end_station_name"
[261] "start_station_id"
[262] "end_station_id"
[263] "rideable_type"
[264] "started_at"
[265] "ended_at"
[266] "start_station_name"
[267] "end_station_name"
[268] "start_station_id"
[269] "end_station_id"
[270] "rideable_type"
[271] "started_at"
[272] "ended_at"
[273] "start_station_name"
[274] "end_station_name"
[275] "start_station_id"
[276] "end_station_id"
[277] "rideable_type"
[278] "started_at"
[279] "ended_at"
[280] "start_station_name"
[281] "end_station_name"
[282] "start_station_id"
[283] "end_station_id"
[284] "rideable_type"
[285] "started_at"
[286] "ended_at"
[287] "start_station_name"
[288] "end_station_name"
[289] "start_station_id"
[290] "end_station_id"
[291] "rideable_type"
[292] "started_at"
[293] "ended_at"
[294] "start_station_name"
[295] "end_station_name"
[296] "start_station_id"
[297] "end_station_id"
[298] "rideable_type"
[299] "started_at"
[300] "ended_at"
[301] "start_station_name"
[302] "end_station_name"
[303] "start_station_id"
[304] "end_station_id"
[305] "rideable_type"
[306] "started_at"
[307] "ended_at"
[308] "start_station_name"
[309] "end_station_name"
[310] "start_station_id"
[311] "end_station_id"
[312] "rideable_type"
[313] "started_at"
[314] "ended_at"
[315] "start_station_name"
[316] "end_station_name"
[317] "start_station_id"
[318] "end_station_id"
[319] "rideable_type"
[320] "started_at"
[321] "ended_at"
[322] "start_station_name"
[323] "end_station_name"
[324] "start_station_id"
[325] "end_station_id"
[326] "rideable_type"
[327] "started_at"
[328] "ended_at"
[329] "start_station_name"
[330] "end_station_name"
[331] "start_station_id"
[332] "end_station_id"
[333] "rideable_type"
[334] "started_at"
[335] "ended_at"
[336] "start_station_name"
[337] "end_station_name"
[338] "start_station_id"
[339] "end_station_id"
[340] "rideable_type"
[341] "started_at"
[342] "ended_at"
[343] "start_station_name"
[344] "end_station_name"
[345] "start_station_id"
[346] "end_station_id"
[347] "rideable_type"
[348] "started_at"
[349] "ended_at"
[350] "start_station_name"
[351] "end_station_name"
[352] "start_station_id"
[353] "end_station_id"
[354] "rideable_type"
[355] "started_at"
[356] "ended_at"
[357] "start_station_name"
[358] "end_station_name"
[359] "start_station_id"
[360] "end_station_id"
[361] "rideable_type"
[362] "started_at"
[363] "ended_at"
[364] "start_station_name"
[365] "end_station_name"
[366] "start_station_id"
[367] "end_station_id"
[368] "rideable_type"
[369] "started_at"
[370] "ended_at"
[371] "start_station_name"
[372] "end_station_name"
[373] "start_station_id"
[374] "end_station_id"
[375] "rideable_type"
[376] "started_at"
[377] "ended_at"
[378] "start_station_name"
[379] "end_station_name"
[380] "start_station_id"
[381] "end_station_id"
[382] "rideable_type"
[383] "started_at"
[384] "ended_at"
[385] "start_station_name"
[386] "end_station_name"
[387] "start_station_id"
[388] "end_station_id"
[389] "rideable_type"
[390] "started_at"
[391] "ended_at"
[392] "start_station_name"
[393] "end_station_name"
[394] "start_station_id"
[395] "end_station_id"
[396] "rideable_type"
[397] "started_at"
[398] "ended_at"
[399] "start_station_name"
[400] "end_station_name"
[401] "start_station_id"
[402] "end_station_id"
[403] "rideable_type"
[404] "started_at"
[405] "ended_at"
[406] "start_station_name"
[407] "end_station_name"
[408] "start_station_id"
[409] "end_station_id"
[410] "rideable_type"
[411] "started_at"
[412] "ended_at"
[413] "start_station_name"
[414] "end_station_name"
[415] "start_station_id"
[416] "end_station_id"
[417] "rideable_type"
[418] "started_at"
[419] "ended_at"
[420] "start_station_name"
[421] "end_station_name"
[422] "start_station_id"
[423] "end_station_id"
[424] "rideable_type"
[425] "started_at"
[426] "ended_at"
[427] "start_station_name"
[428] "end_station_name"
[429] "start_station_id"
[430] "end_station_id"
[431] "rideable_type"
[432] "started_at"
[433] "ended_at"
[434] "start_station_name"
[435] "end_station_name"
[436] "start_station_id"
[437] "end_station_id"
[438] "rideable_type"
[439] "started_at"
[440] "ended_at"
[441] "start_station_name"
[442] "end_station_name"
[443] "start_station_id"
[444] "end_station_id"
[445] "rideable_type"
[446] "started_at"
[447] "ended_at"
[448] "start_station_name"
[449] "end_station_name"
[450] "start_station_id"
[451] "end_station_id"
[452] "rideable_type"
[453] "started_at"
[454] "ended_at"
[455] "start_station_name"
[456] "end_station_name"
[457] "start_station_id"
[458] "end_station_id"
[459] "rideable_type"
[460] "started_at"
[461] "ended_at"
[462] "start_station_name"
[463] "end_station_name"
[464] "start_station_id"
[465] "end_station_id"
[466] "rideable_type"
[467] "started_at"
[468] "ended_at"
[469] "start_station_name"
[470] "end_station_name"
[471] "start_station_id"
[472] "end_station_id"
[473] "rideable_type"
[474] "started_at"
[475] "ended_at"
[476] "start_station_name"
[477] "end_station_name"
[478] "start_station_id"
[479] "end_station_id"
[480] "rideable_type"
[481] "started_at"
[482] "ended_at"
[483] "start_station_name"
[484] "end_station_name"
[485] "start_station_id"
[486] "end_station_id"
[487] "rideable_type"
[488] "started_at"
[489] "ended_at"
[490] "start_station_name"
[491] "end_station_name"
[492] "start_station_id"
[493] "end_station_id"
[494] "rideable_type"
[495] "started_at"
[496] "ended_at"
[497] "start_station_name"
[498] "end_station_name"
[499] "start_station_id"
[500] "end_station_id"
[501] "rideable_type"
[502] "started_at"
[503] "ended_at"
[504] "start_station_name"
[505] "end_station_name"
[506] "start_station_id"
[507] "end_station_id"
[508] "rideable_type"
[509] "started_at"
[510] "ended_at"
[511] "start_station_name"
[512] "end_station_name"
[513] "start_station_id"
[514] "end_station_id"
[515] "rideable_type"
[516] "started_at"
[517] "ended_at"
[518] "start_station_name"
[519] "end_station_name"
[520] "start_station_id"
[521] "end_station_id"
[522] "rideable_type"
[523] "started_at"
[524] "ended_at"
[525] "start_station_name"
[526] "end_station_name"
[527] "start_station_id"
[528] "end_station_id"
[529] "rideable_type"
[530] "started_at"
[531] "ended_at"
[532] "start_station_name"
[533] "end_station_name"
[534] "start_station_id"
[535] "end_station_id"
[536] "rideable_type"
[537] "started_at"
[538] "ended_at"
[539] "start_station_name"
[540] "end_station_name"
[541] "start_station_id"
[542] "end_station_id"
[543] "rideable_type"
[544] "started_at"
[545] "ended_at"
[546] "start_station_name"
[547] "end_station_name"
[548] "start_station_id"
[549] "end_station_id"
[550] "rideable_type"
[551] "started_at"
[552] "ended_at"
[553] "start_station_name"
[554] "end_station_name"
[555] "start_station_id"
[556] "end_station_id"
[557] "rideable_type"
[558] "started_at"
[559] "ended_at"
[560] "start_station_name"
[561] "end_station_name"
[562] "start_station_id"
[563] "end_station_id"
[564] "rideable_type"
[565] "started_at"
[566] "ended_at"
[567] "start_station_name"
[568] "end_station_name"
[569] "start_station_id"
[570] "end_station_id"
[571] "rideable_type"
[572] "started_at"
[573] "ended_at"
[574] "start_station_name"
[575] "end_station_name"
[576] "start_station_id"
[577] "end_station_id"
[578] "rideable_type"
[579] "started_at"
[580] "ended_at"
[581] "start_station_name"
[582] "end_station_name"
[583] "start_station_id"
[584] "end_station_id"
[585] "rideable_type"
[586] "started_at"
[587] "ended_at"
[588] "start_station_name"
[589] "end_station_name"
[590] "start_station_id"
[591] "end_station_id"
[592] "rideable_type"
[593] "started_at"
[594] "ended_at"
[595] "start_station_name"
[596] "end_station_name"
[597] "start_station_id"
[598] "end_station_id"
[599] "rideable_type"
[600] "started_at"
[601] "ended_at"
[602] "start_station_name"
[603] "end_station_name"
[604] "start_station_id"
[605] "end_station_id"
[606] "rideable_type"
[607] "started_at"
[608] "ended_at"
[609] "start_station_name"
[610] "end_station_name"
[611] "start_station_id"
[612] "end_station_id"
[613] "rideable_type"
[614] "started_at"
[615] "ended_at"
[616] "start_station_name"
[617] "end_station_name"
[618] "start_station_id"
[619] "end_station_id"
[620] "rideable_type"
[621] "started_at"
[622] "ended_at"
[623] "start_station_name"
[624] "end_station_name"
[625] "start_station_id"
[626] "end_station_id"
[627] "rideable_type"
[628] "started_at"
[629] "ended_at"
[630] "start_station_name"
[631] "end_station_name"
[632] "start_station_id"
[633] "end_station_id"
[634] "rideable_type"
[635] "started_at"
[636] "ended_at"
[637] "start_station_name"
[638] "end_station_name"
[639] "start_station_id"
[640] "end_station_id"
[641] "rideable_type"
[642] "started_at"
[643] "ended_at"
[644] "start_station_name"
[645] "end_station_name"
[646] "start_station_id"
[647] "end_station_id"
[648] "rideable_type"
[649] "started_at"
[650] "ended_at"
[651] "start_station_name"
[652] "end_station_name"
[653] "start_station_id"
[654] "end_station_id"
[655] "rideable_type"
[656] "started_at"
[657] "ended_at"
[658] "start_station_name"
[659] "end_station_name"
[660] "start_station_id"
[661] "end_station_id"
[662] "rideable_type"
[663] "started_at"
[664] "ended_at"
[665] "start_station_name"
[666] "end_station_name"
[667] "start_station_id"
[668] "end_station_id"
[669] "rideable_type"
[670] "started_at"
[671] "ended_at"
[672] "start_station_name"
[673] "end_station_name"
[674] "start_station_id"
[675] "end_station_id"
[676] "rideable_type"
[677] "started_at"
[678] "ended_at"
[679] "start_station_name"
[680] "end_station_name"
[681] "start_station_id"
[682] "end_station_id"
[683] "rideable_type"
[684] "started_at"
[685] "ended_at"
[686] "start_station_name"
[687] "end_station_name"
[688] "start_station_id"
[689] "end_station_id"
[690] "rideable_type"
[691] "started_at"
[692] "ended_at"
[693] "start_station_name"
[694] "end_station_name"
[695] "start_station_id"
[696] "end_station_id"
[697] "rideable_type"
[698] "started_at"
[699] "ended_at"
[700] "start_station_name"
[701] "end_station_name"
[702] "start_station_id"
[703] "end_station_id"
[704] "rideable_type"
[705] "started_at"
[706] "ended_at"
[707] "start_station_name"
[708] "end_station_name"
[709] "start_station_id"
[710] "end_station_id"
[711] "rideable_type"
[712] "started_at"
[713] "ended_at"
[714] "start_station_name"
[715] "end_station_name"
[716] "start_station_id"
[717] "end_station_id"
[718] "rideable_type"
[719] "started_at"
[720] "ended_at"
[721] "start_station_name"
[722] "end_station_name"
[723] "start_station_id"
[724] "end_station_id"
[725] "rideable_type"
[726] "started_at"
[727] "ended_at"
[728] "start_station_name"
[729] "end_station_name"
[730] "start_station_id"
[731] "end_station_id"
[732] "rideable_type"
[733] "started_at"
[734] "ended_at"
[735] "start_station_name"
[736] "end_station_name"
[737] "start_station_id"
[738] "end_station_id"
[739] "rideable_type"
[740] "started_at"
[741] "ended_at"
[742] "start_station_name"
[743] "end_station_name"
[744] "start_station_id"
[745] "end_station_id"
[746] "rideable_type"
[747] "started_at"
[748] "ended_at"
[749] "start_station_name"
[750] "end_station_name"
[751] "start_station_id"
[752] "end_station_id"
[753] "rideable_type"
[754] "started_at"
[755] "ended_at"
[756] "start_station_name"
[757] "end_station_name"
[758] "start_station_id"
[759] "end_station_id"
[760] "rideable_type"
[761] "started_at"
[762] "ended_at"
[763] "start_station_name"
[764] "end_station_name"
[765] "start_station_id"
[766] "end_station_id"
[767] "rideable_type"
[768] "started_at"
[769] "ended_at"
[770] "start_station_name"
[771] "end_station_name"
[772] "start_station_id"
[773] "end_station_id"
[774] "rideable_type"
[775] "started_at"
[776] "ended_at"
[777] "start_station_name"
[778] "end_station_name"
[779] "start_station_id"
[780] "end_station_id"
[781] "rideable_type"
[782] "started_at"
[783] "ended_at"
[784] "start_station_name"
[785] "end_station_name"
[786] "start_station_id"
[787] "end_station_id"
[788] "rideable_type"
[789] "started_at"
[790] "ended_at"
[791] "start_station_name"
[792] "end_station_name"
[793] "start_station_id"
[794] "end_station_id"
[795] "rideable_type"
[796] "started_at"
[797] "ended_at"
[798] "start_station_name"
[799] "end_station_name"
[799] "start_station_id"
[800] "end_station_id"

```

| | |
|--------|--------|
| tbl_df | 6 x 19 |
|--------|--------|

| | |
|--------|--------|
| tbl_df | 6 x 19 |
|--------|--------|

3rd Qu.: 0.000
Max.: 7.000

```

end_lat...11      end_lat...12      start_ride_station
Min.   :42         Min.   :~-88      Length:4978837
1st Qu.:42         1st Qu.:~-88      Class :character
Median :42         Median :~-88      Mode  :character
Mean   :42         Mean   :~-88
3rd Qu.:42         3rd Qu.:~-88
Max.   :42         Max.   :~-88
NA's   :4210688    NA's   :4210688
...8
Length:4978837
Class :character
Mode  :character

```

| ride_id
<chr> | rideable_type
<chr> | started_at
<chr> |
|------------------|------------------------|---------------------|
| C2F7DD78E82EC875 | electric_bike | 1/13/2022 11:59 |
| A6CF8980A652D272 | electric_bike | 1/10/2022 8:41 |
| BD0F91DFF741C66D | classic_bike | 1/25/2022 4:53 |
| CBB80ED419105406 | classic_bike | 1/4/2022 0:18 |
| DDC963BFDDA51EEA | classic_bike | 1/20/2022 1:31 |
| A39C6F6CC0586C0B | classic_bike | 1/11/2022 18:48 |

| ended_at
<chr> | start_station_name
<chr> |
|-------------------|-------------------------------|
| 1/13/2022 12:02 | Glenwood Ave & Touhy Ave |
| 1/10/2022 8:46 | Glenwood Ave & Touhy Ave |
| 1/25/2022 4:58 | Sheffield Ave & Fullerton Ave |
| 1/4/2022 0:33 | Clark St & Bryn Mawr Ave |
| 1/20/2022 1:37 | Michigan Ave & Jackson Blvd |
| 1/11/2022 18:51 | Wood St & Chicago Ave |

| start_station_id
<chr> | end_station_name
<chr> |
|---------------------------|-------------------------------|
| 525 | Clark St & Touhy Ave |
| 525 | Clark St & Touhy Ave |
| Ta1306000016 | Greenview Ave & Fullerton Ave |
| Ka1504000151 | Paulina St & Montrose Ave |
| Ta1309000002 | State St & Randolph St |
| 637 | Honore St & Division St |

| end_station_id
<chr> | start_lat
<dbl> | start_lng
<dbl> | end_lat
<dbl> |
|-------------------------|--------------------|--------------------|------------------|
| Rp-007 | 42.01280 | -87.66591 | 42.01256 |
| Rp-007 | 42.01276 | -87.66597 | 42.01256 |
| Ta1307000001 | 41.92560 | -87.65371 | 41.92533 |
| Ta1309000021 | 41.98359 | -87.66915 | 41.96151 |
| Ta1305000029 | 41.87785 | -87.62408 | 41.88462 |
| Ta1305000034 | 41.89563 | -87.67207 | 41.90312 |

| member_casual
<chr> | ride_length
<S3: hms> | day_of_week
<dbl> | end_lat...11
<dbl> |
|------------------------|--------------------------|----------------------|-----------------------|
| casual | 00:03:00 | 5 | NA |
| casual | 00:05:00 | 2 | NA |
| member | 00:05:00 | 3 | NA |
| casual | 00:15:00 | 3 | NA |
| member | 00:06:00 | 5 | NA |
| member | 00:03:00 | 3 | NA |

| ride_id
<chr> | rideable_type
<chr> | started_at
<chr> |
|------------------|------------------------|---------------------|
| 7BDEDE9860418B53 | classic_bike | 12/7/2022 6:52 |
| 43ABEE85B6E15DCA | classic_bike | 12/5/2022 6:51 |
| F041C89A3D1F0270 | electric_bike | 12/14/2022 17:06 |
| A2BECB88430BE156 | classic_bike | 12/8/2022 16:27 |
| 37B392960E566F58 | classic_bike | 12/28/2022 9:37 |
| 2DD1587210BA45AE | classic_bike | 12/9/2022 0:27 |

| ended_at
<chr> | start_station_name
<chr> |
|-------------------|-------------------------------|
| 12/7/2022 6:56 | Sangamon St & Washington Blvd |
| 12/5/2022 6:54 | Sangamon St & Washington Blvd |
| 12/14/2022 17:19 | Bernard St & Elston Ave |
| 12/8/2022 16:32 | Wacker Dr & Washington St |
| 12/28/2022 9:41 | Sangamon St & Washington Blvd |
| 12/9/2022 0:35 | Southport Ave & Waveland Ave |

| start_station_id
<chr> | end_station_name
<chr> | end_station_id
<chr> |
|---------------------------|---------------------------|-------------------------|
| 13409 | Peoria St & Jackson Blvd | 13158 |
| 13409 | Peoria St & Jackson Blvd | 13158 |
| 18016 | Seeley Ave & Roscoe St | 13144 |
| Ka1503000072 | Green St & Madison St | Ta1307000120 |
| 13409 | Peoria St & Jackson Blvd | 13158 |
| 13235 | Seeley Ave & Roscoe St | 13144 |

| start_lat
<dbl> | start_lng
<dbl> | end_lat
<dbl> | end_lng
<dbl> | member_casual
<chr> |
|--------------------|--------------------|------------------|------------------|------------------------|
| 41.88316 | -87.65110 | 41.87764 | -87.64962 | member |
| 41.88316 | -87.65110 | 41.87764 | -87.64962 | member |
| 41.94998 | -87.71402 | 41.94340 | -87.67962 | member |
| 41.88314 | -87.63724 | 41.88186 | -87.64926 | member |
| 41.88316 | -87.65110 | 41.87764 | -87.64962 | member |
| 41.94815 | -87.66394 | 41.94340 | -87.67962 | casual |

| ride_length
<S3: hms> | day_of_week
<dbl> | end_lat...11
<dbl> | end_lat...12
<dbl> |
|--------------------------|----------------------|-----------------------|-----------------------|
| 00:04:00 | 4 | NA | NA |
| 00:03:00 | 2 | NA | NA |
| 00:13:00 | 4 | NA | NA |
| 00:05:00 | 5 | NA | NA |
| 00:04:00 | 4 | NA | NA |
| 00:08:00 | 6 | NA | NA |

```
# Convert "ride_length" from Factor to numeric so we can run calculations
on the data
```

```
```{r}
is.factor(all_trips$ride_length)
all_trips$ride_length <- as.numeric(all_trips$ride_length)
is.numeric(all_trips$ride_length)
```
```

```
[1] FALSE
[1] TRUE
```

```
## Conduct descriptive analysis
# Descriptive analysis on ride_length
```

```
```{r}
summary(all_trips$ride_length)
```
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
      0      360      660    1088    1140   359640    711
```

```
# Compare members and casual users
```

```
```{r}
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = mean)
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = median)
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = max)
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = min)
```
```

| all_trips\$member_casual
<chr> | all_trips\$ride_length
<dbl> |
|-----------------------------------|---------------------------------|
| casual | 1536.8251 |
| member | 773.3823 |

| all_trips\$member_casual
<chr> | all_trips\$ride_length
<dbl> |
|-----------------------------------|---------------------------------|
| casual | 780 |
| member | 540 |

| all_trips\$member_casual
<chr> | all_trips\$ride_length
<dbl> |
|-----------------------------------|---------------------------------|
| casual | 359640 |
| member | 93600 |

| all_trips\$member_casual
<chr> | all_trips\$ride_length
<dbl> |
|-----------------------------------|---------------------------------|
| casual | 0 |
| member | 0 |

```
# See the average ride time by each day for members vs casual users
```

```
```{r}
aggregate(all_trips$ride_length ~ all_trips$member_casual +
all_trips$day_of_week, FUN = mean)
```
```

| all_trips\$member_casual
<chr> | all_trips\$day_of_week
<dbl> |
|--|--|
| casual | 1 |
| member | 1 |
| casual | 2 |
| member | 2 |
| casual | 3 |
| member | 3 |
| casual | 4 |
| member | 4 |
| casual | 5 |
| member | 5 |

| all_trips\$member_casual
<chr> | all_trips\$day_of_week
<dbl> |
|--|--|
| casual | 6 |
| member | 6 |
| casual | 7 |
| member | 7 |

| all_trips\$day_of_week
<dbl> | all_trips\$ride_length
<dbl> |
|--|--|
| 1 | 1723.1598 |
| 1 | 841.8464 |
| 2 | 1526.6171 |
| 2 | 736.1577 |
| 3 | 1345.3279 |
| 3 | 727.7927 |
| 4 | 1283.9093 |
| 4 | 726.3266 |
| 5 | 1338.5988 |
| 5 | 737.5799 |

| all_trips\$day_of_week
<dbl> | all_trips\$ride_length
<dbl> |
|--|--|
| 6 | 1414.0916 |
| 6 | 751.9322 |
| 7 | 1657.4400 |
| 7 | 848.3786 |

```
#Finally lets create a .csv file that we will visualize in Tableau
```{r}
write.csv(all_trips,"C:/Users/waqarshaozab/Desktop/csv_files/all_trips.csv",
row.names =FALSE)
```
```

4.4 Analyze in Excel

Given my proficiency in data manipulation using Excel, I've chosen to export the R data frame to a .csv file. Once this is completed, I'm poised to conclude the analysis and transition into the data visualization phase. Below, I'll provide a detailed account of the steps executed in Excel.

4.5 Create Pivot Table 1

1. Select Column *member_casual* and *ride_length* → Insert → Pivot Table → New Worksheet → Drag *member_casual* to Rows Section → Drag *ride_length* to Value Section
2. Formatted Time as HH:MM:SS (37:30:55)
3. Home → Number → Format Cells → Time → HH : MM : SS(37 : 30 : 55)
4. Right click each table and set the measure to 'average' and repeat for each month

| | | | | | |
|-------------|------------------------|--|-------------|------------------------|--|
| JAN | | | FEB | | |
| Row Labels | Average of ride_length | | Row Labels | Average of ride_length | |
| casual | 0:30:22 | | casual | 0:26:43 | |
| member | 0:11:59 | | member | 0:11:24 | |
| (blank) | | | (blank) | | |
| Grand Total | 0:15:16 | | Grand Total | 0:14:14 | |
| MAR | | | APR | | |
| Row Labels | Average of ride_length | | Row Labels | Average of ride_length | |
| casual | 0:32:37 | | casual | 0:29:32 | |
| member | 0:11:58 | | member | 0:11:29 | |
| (blank) | | | (blank) | | |
| Grand Total | 0:18:30 | | Grand Total | 0:17:38 | |
| MAY | | | JUN | | |
| Row Labels | Average of ride_length | | Row Labels | Average of ride_length | |
| casual | 0:30:52 | | casual | 0:32:06 | |
| member | 0:13:22 | | member | 0:14:00 | |
| (blank) | | | (blank) | | |
| Grand Total | 0:21:06 | | Grand Total | 0:22:41 | |
| JUL | | | AUG | | |
| Row Labels | Average of ride_length | | Row Labels | Average of ride_length | |
| casual | 0:29:17 | | casual | 0:29:19 | |
| member | 0:13:43 | | member | 0:13:23 | |
| (blank) | | | (blank) | | |
| Grand Total | 0:21:23 | | Grand Total | 0:20:39 | |
| SEP | | | OCT | | |
| Row Labels | Average of ride_length | | Row Labels | Average of ride_length | |
| casual | 0:27:59 | | casual | 0:26:23 | |
| member | 0:12:59 | | member | 0:11:57 | |
| (blank) | | | (blank) | | |
| Grand Total | 0:19:20 | | Grand Total | 0:17:21 | |
| NOV | | | DEC | | |
| Row Labels | Average of ride_length | | Row Labels | Average of ride_length | |
| casual | 0:21:18 | | casual | 0:22:18 | |
| member | 0:11:08 | | member | 0:10:37 | |
| (blank) | | | (blank) | | |
| Grand Total | 0:14:10 | | Grand Total | 0:13:30 | |

| JAN | | FEB | |
|------------|------------------------|------------|------------------------|
| Row Lat▼ | Average of ride_length | Row Lat▼ | Average of ride_length |
| casual | 0:30:22 | casual | 0:26:43 |
| member | 0:11:59 | member | 0:11:24 |
| (blank) | | (blank) | |
| Grand Tota | 0:15:16 | Grand Tota | 0:14:14 |
| | | | |
| MAR | | APR | |
| Row Lat▼ | Average of ride_length | Row Lat▼ | Average of ride_length |
| casual | 0:32:37 | casual | 0:29:32 |
| member | 0:11:58 | member | 0:11:29 |
| (blank) | | (blank) | |
| Grand Tota | 0:18:30 | Grand Tota | 0:17:38 |
| | | | |
| MAY | | JUN | |
| Row Lat▼ | Average of ride_length | Row Lat▼ | Average of ride_length |
| casual | 0:30:52 | casual | 0:32:06 |
| member | 0:13:22 | member | 0:14:00 |
| (blank) | | (blank) | |
| Grand Tota | 0:21:06 | Grand Tota | 0:22:41 |
| | | | |
| JUL | | AUG | |
| Row Lat▼ | Average of ride_length | Row Lat▼ | Average of ride_length |
| casual | 0:29:17 | casual | 0:29:19 |
| member | 0:13:43 | member | 0:13:23 |
| (blank) | | (blank) | |
| Grand Tota | 0:21:23 | Grand Tota | 0:20:39 |
| | | | |
| SEP | | OCT | |
| Row Lat▼ | Average of ride_length | Row Lat▼ | Average of ride_length |
| casual | 0:27:59 | casual | 0:26:23 |
| member | 0:12:59 | member | 0:11:57 |
| (blank) | | (blank) | |
| Grand Tota | 0:19:20 | Grand Tota | 0:17:21 |
| | | | |
| NOV | | DEC | |
| Row Lat▼ | Average of ride_length | Row Lat▼ | Average of ride_length |
| casual | 0:21:18 | casual | 0:22:18 |
| member | 0:11:08 | member | 0:10:37 |
| (blank) | | (blank) | |
| Grand Tota | 0:14:10 | Grand Tota | 0:13:30 |

4.6 Create Pivot Table 2

1. *Select Column member_casual, ride_length, and day_of_week → Insert → Pivot Table → New Worksheet → Drag member_casual to Rows Section → Drag ride_length to Value Section, and Drag day_of_week to Column Section*
2. Formatted Time as HH:MM:SS (37:30:55)
3. *Home → Number → Format Cells → Time → HH : MM : SS(37 : 30 : 55)*

4. Right click each table and set the measure to 'average' and repeat for each month

| | | | | | | | | | |
|-------------------|----------|----------|----------|----------|-----------|-----------|-------------------|-------------|--|
| JAN | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 1 | 2 | 3 | 4 | 5 | 6 | 7 (blank) | Grand Total | |
| casual | 0:26:34 | 0:28:08 | 0:19:25 | 0:36:12 | 0:35:27 | 0:24:31 | 0:37:59 | 0:30:22 | |
| member | 0:13:08 | 0:11:28 | 0:12:08 | 0:11:38 | 0:11:37 | 0:11:56 | 0:12:24 | 0:11:59 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:16:04 | 0:14:02 | 0:13:13 | 0:15:30 | 0:15:17 | 0:14:11 | 0:18:58 | 0:15:16 | |
| FEB | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 1 | 2 | 3 | 4 | 5 | 6 | 7 (blank) | Grand Total | |
| casual | 0:32:59 | 0:24:54 | 0:26:54 | 0:23:09 | 0:27:28 | 0:22:07 | 0:27:10 | 0:26:43 | |
| member | 0:12:15 | 0:11:22 | 0:11:16 | 0:10:50 | 0:11:04 | 0:11:41 | 0:11:38 | 0:11:24 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:17:44 | 0:13:59 | 0:13:33 | 0:12:42 | 0:13:20 | 0:13:36 | 0:15:08 | 0:14:14 | |
| MAR | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 1 | 2 | 3 | 4 | 5 | 6 | 7 (blank) | Grand Total | |
| casual | 0:38:48 | 0:35:24 | 0:25:02 | 0:30:04 | 0:29:53 | 0:25:39 | 0:36:16 | 0:32:37 | |
| member | 0:13:29 | 0:12:37 | 0:10:59 | 0:11:58 | 0:10:51 | 0:11:04 | 0:13:41 | 0:11:58 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:24:21 | 0:20:07 | 0:14:11 | 0:17:10 | 0:16:02 | 0:14:51 | 0:23:27 | 0:18:30 | |
| APR | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:32:58 | 0:29:07 | 0:26:19 | 0:21:21 | 0:25:52 | 0:26:08 | 0:34:45 | 0:29:32 | |
| member | 0:12:25 | 0:10:48 | 0:10:50 | 0:10:32 | 0:11:24 | 0:10:58 | 0:13:35 | 0:11:29 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:21:18 | 0:15:37 | 0:14:56 | 0:13:11 | 0:15:47 | 0:15:48 | 0:23:55 | 0:17:38 | |
| MAY | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:33:43 | 0:32:27 | 0:26:45 | 0:25:48 | 0:29:16 | 0:29:28 | 0:33:23 | 0:30:52 | |
| member | 0:14:29 | 0:13:24 | 0:12:47 | 0:12:19 | 0:13:10 | 0:12:49 | 0:14:42 | 0:13:22 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:24:42 | 0:21:39 | 0:17:57 | 0:17:01 | 0:19:29 | 0:20:01 | 0:24:48 | 0:21:06 | |

| | | | | | | | | | |
|-------------------|----------|----------|----------|----------|-----------|-----------|-------------------|-------------|--|
| JUN | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:36:08 | 0:31:29 | 0:31:29 | 0:29:04 | 0:30:12 | 0:32:38 | 0:32:14 | 0:32:06 | |
| member | 0:15:41 | 0:13:25 | 0:13:39 | 0:13:06 | 0:13:42 | 0:13:48 | 0:15:10 | 0:14:00 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:27:24 | 0:21:24 | 0:21:02 | 0:19:42 | 0:20:59 | 0:23:06 | 0:24:51 | 0:22:41 | |
| JUL | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:33:35 | 0:31:06 | 0:26:23 | 0:24:00 | 0:24:21 | 0:26:05 | 0:32:54 | 0:29:17 | |
| member | 0:15:12 | 0:13:23 | 0:12:56 | 0:12:55 | 0:13:01 | 0:13:01 | 0:15:18 | 0:13:43 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:25:42 | 0:21:41 | 0:18:34 | 0:17:33 | 0:17:59 | 0:19:16 | 0:25:31 | 0:21:23 | |
| AUG | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:33:45 | 0:28:53 | 0:28:11 | 0:25:39 | 0:25:11 | 0:30:13 | 0:31:57 | 0:29:19 | |
| member | 0:14:38 | 0:12:28 | 0:13:04 | 0:13:00 | 0:13:03 | 0:13:33 | 0:14:41 | 0:13:23 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:24:44 | 0:19:06 | 0:19:08 | 0:18:05 | 0:18:11 | 0:21:46 | 0:24:22 | 0:20:39 | |
| SEP | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:34:40 | 0:28:37 | 0:21:10 | 0:22:48 | 0:21:11 | 0:27:08 | 0:35:23 | 0:27:59 | |
| member | 0:14:08 | 0:12:33 | 0:12:14 | 0:12:24 | 0:12:26 | 0:13:00 | 0:14:47 | 0:12:59 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:24:28 | 0:18:55 | 0:15:17 | 0:16:04 | 0:15:42 | 0:19:11 | 0:25:58 | 0:19:20 | |
| OCT | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:32:23 | 0:22:14 | 0:23:20 | 0:20:08 | 0:19:38 | 0:25:21 | 0:30:15 | 0:26:23 | |
| member | 0:13:20 | 0:11:25 | 0:11:11 | 0:11:16 | 0:10:57 | 0:11:42 | 0:13:29 | 0:11:57 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:22:17 | 0:14:52 | 0:14:37 | 0:13:53 | 0:13:41 | 0:16:42 | 0:21:25 | 0:17:21 | |
| NOV | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:34:06 | 0:17:43 | 0:16:42 | 0:16:24 | 0:24:10 | 0:19:16 | 0:22:34 | 0:21:18 | |
| member | 0:11:52 | 0:10:34 | 0:10:47 | 0:10:55 | 0:11:41 | 0:11:21 | 0:11:11 | 0:11:08 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:20:07 | 0:12:18 | 0:12:18 | 0:12:25 | 0:15:37 | 0:13:55 | 0:15:22 | 0:14:10 | |
| DEC | | | | | | | | | |
| Average of Column | | | | | | | | | |
| Row Lat | 24:00:00 | 48:00:00 | 72:00:00 | 96:00:00 | 120:00:00 | 144:00:00 | 168:00:00 (blank) | Grand Total | |
| casual | 0:25:06 | 0:18:18 | 0:15:20 | 0:24:13 | 0:19:42 | 0:29:04 | 0:23:18 | 0:22:18 | |
| member | 0:10:51 | 0:10:29 | 0:10:20 | 0:10:23 | 0:10:24 | 0:10:53 | 0:11:17 | 0:10:37 | |
| (blank) | | | | | | | | | |
| Grand Total | 0:15:13 | 0:12:06 | 0:11:23 | 0:13:26 | 0:12:33 | 0:15:44 | 0:15:03 | 0:13:30 | |

4.7 Summary of Analysis

After analyzing the data using tools like R and Excel, we've uncovered valuable insights that can help us address the business task at hand: creating marketing strategies to turn casual riders into annual members. These insights directly address the three business objectives set by Lily Moreno, our Director of Marketing: 1. How do annual members and casual riders use Cyclistic differently? 2. Why would casual riders buy Cyclistic annual memberships? 3. How can Cyclistic use digital media to influence casual riders to become members?

Addressing the first objective, “How do annual members and casual riders use Cyclistic differently?” We should be able to differentiate the two groups by underscoring these three insights:

1. **Understanding Riding Behavior:** The data reveals that annual members and casual riders use Cyclistic bikes differently. In 2022, members took more rides overall than casual riders. However, when we look at daily usage patterns, both groups ride similarly on weekends. The key distinction arises on weekdays, where members significantly outpace casual riders in bike usage. This suggests that members likely use Cyclistic bikes for their daily commutes.
2. **Ride Duration:** When examining ride duration by month, casual riders spend considerably more time on bikes than members. The average ride duration for casual riders is approximately 28 minutes, while for members, it’s around 12 minutes. Additionally, both casual riders and members tend to spend more time riding on weekends compared to weekdays. Notably, members reduce their weekday riding time.
3. **Bike Type Preferences:** Our analysis also revealed differences in bike type usage. Members primarily opt for classic bikes, while casual riders predominantly choose electric and docked bikes. Notably, docked bikes are exclusively used by casual riders, and these bikes require docking at a station to stop the meter.

Addressing the second objective, “Why would casual riders buy Cyclistic annual memberships?” We can effectively persuade casual riders to become members by highlighting two critical insights:

1. **Frequency vs. Duration:** Members ride more frequently than casual riders, particularly on weekdays. However, they spend less time per ride. This highlights Cyclistic’s ability to offer an affordable and efficient means of commuting to work. Our ad campaign can emphasize this by showcasing how members save time and money while staying active.
2. **Health Benefits:** We can promote the benefits of exercise for physical, mental, and spiritual well-being. Linking exercise to improved productivity at work is a

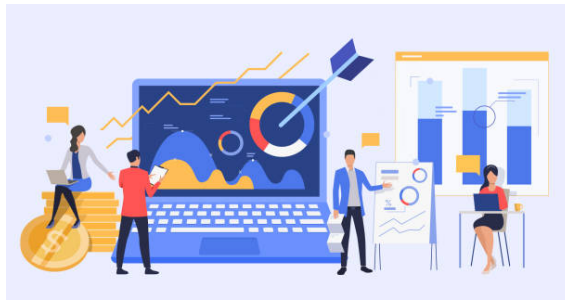
persuasive angle. Given that many city dwellers prefer public transit, we should target this audience with a campaign designed for a general audience, using a third-grade reading level and vibrant graphics to engage Chicago residents. Additionally, we should highlight the bike options available, making cycling more accessible, even for people with disabilities.

Addressing the third objective, “How can Cyclistic use digital media to influence casual riders to become members?” I believe that the benefits of digital media can generate this impactful insight to fruition:

1. **Online Presence:** We can use digital media platforms like Facebook, Twitter, and Instagram to provide efficient tools for reaching a wider audience. Leveraging recommendation algorithms, we can ensure our advertisements are seen by the right people. By harnessing digital media effectively, we can confidently promote our product and extend our reach to potential customers.

Chapter 5

SHARE



5.1 Key Tasks

1. Determine the best way to share your findings
2. Create effective data visualizations
3. Present your findings
4. Ensure your work is accessible

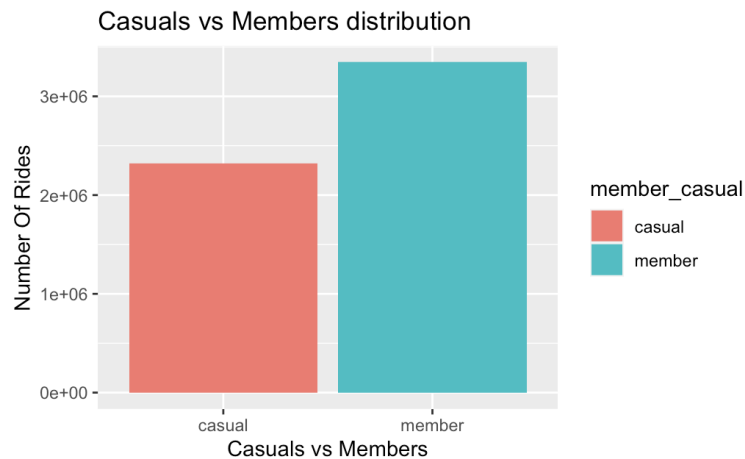
5.2 Deliverable

- ☐ Supporting visualizations and key findings

5.3 Visualization in R

Let's visualize the number of rides by rider type

```
```{r}
ggplot(all_trips, aes(x = member_casual, fill=member_casual)) +
 geom_bar() +
 labs(x="Casuals vs Members", y="Number Of Rides", title= "Casuals vs
Members distribution")
```
```



As you can see, members take more rides than casuals. So based on the cyclic data,

Comparison between Members Casual riders depending on ride length
(mean, median, minimum, maximum)

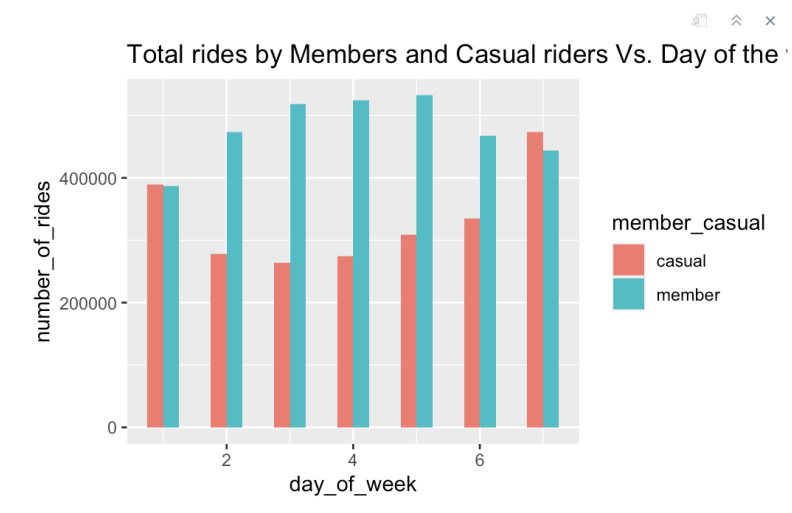
```
```{r}
all_trips %>%
 group_by(member_casual) %>%
 summarise(average_ride_length = mean(ride_length, na.rm = TRUE),
 median_length = median(ride_length, na.rm = TRUE),
 max_ride_length = max(ride_length, na.rm = TRUE),
 min_ride_length = min(ride_length, na.rm = TRUE))
```
```

| member_casual
<chr> | average_ride_length
<dbl> | median_length
<dbl> |
|------------------------|------------------------------|------------------------|
| casual | 1495.6630 | 780 |
| member | 762.8505 | 540 |

| median_length
<dbl> | max_ride_length
<dbl> | min_ride_length
<dbl> |
|------------------------|--------------------------|--------------------------|
| 780 | 359640 | 0 |
| 540 | 93600 | 0 |

Lets visualize total rides data by type and day of week

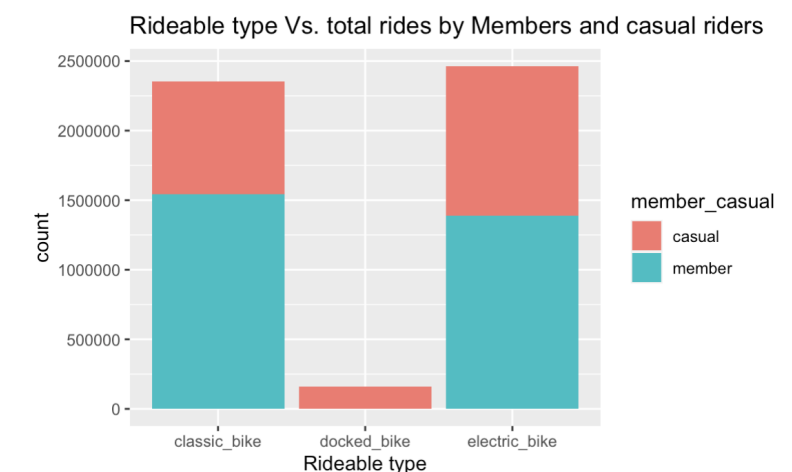
```
```{r}
all_trips %>%
 group_by(member_casual, day_of_week) %>%
 summarise(number_of_rides = n(), .groups="drop") %>%
 arrange(member_casual, day_of_week) %>%
 ggplot(aes(x = day_of_week, y = number_of_rides, fill = member_casual))
+
 labs(title = "Total rides by Members and Casual riders Vs. Day of the
week") +
 geom_col(width=0.5, position = position_dodge(width=0.5)) +
 scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```
```



```
#Rideable type Vs. total rides by Members and casual riders
{r}
all_trips %>%
  group_by(rideable_type) %>%
  summarise(count = length(ride_id))

ggplot(all_trips, aes(x=rideable_type, fill=member_casual)) +
  labs(x="Rideable type", title="Rideable type Vs. total rides by Members and
casual riders") +
  geom_bar()
{r}
```

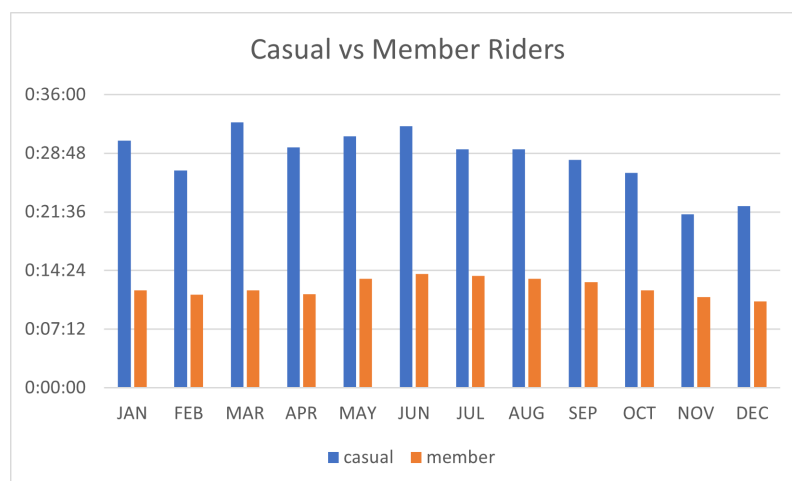
| rideable_type
<chr> | count
<int> |
|------------------------|----------------|
| classic_bike | 2353187 |
| docked_bike | 160896 |
| electric_bike | 2464754 |



5.4 Visualization in Excel

This table was created from **Pivot Table 1** in order to create the first Excel visual

| casual | member |
|----------------|----------------|
| 0:30:22 | 0:11:59 |
| 0:26:43 | 0:11:24 |
| 0:32:37 | 0:11:58 |
| 0:29:32 | 0:11:29 |
| 0:30:52 | 0:13:22 |
| 0:32:06 | 0:14:00 |
| 0:29:17 | 0:13:43 |
| 0:29:19 | 0:13:23 |
| 0:27:59 | 0:12:59 |
| 0:26:23 | 0:11:57 |
| 0:21:18 | 0:11:08 |
| 0:22:18 | 0:10:37 |
| 0:28:14 | 0:12:20 |



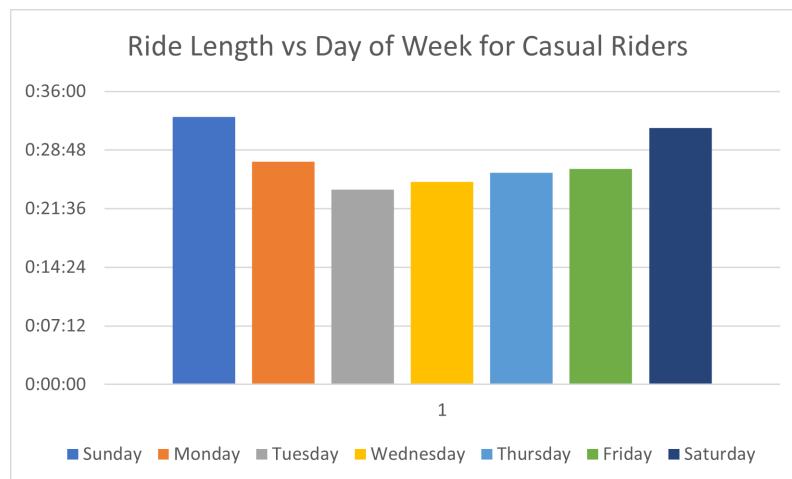
These tables were created from **Pivot Table 2** in order to create the next Excel visuals

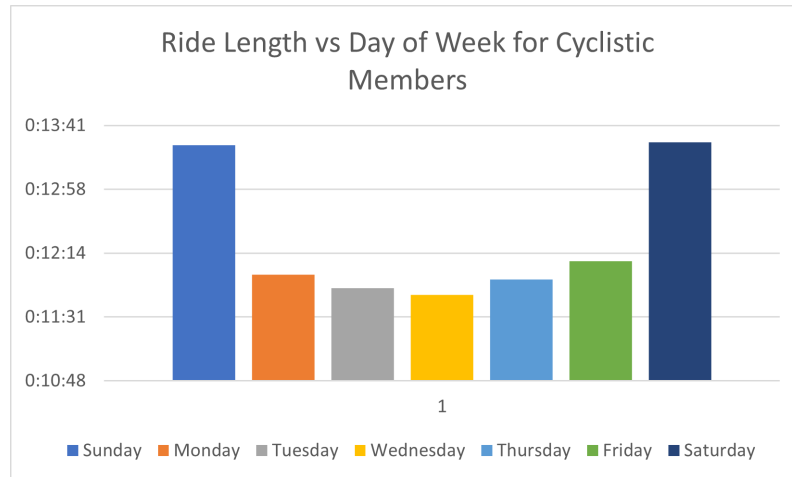
| 1_casual | 2_casual | 3_casual | 4_casual | 5_casual | 6_casual | 7_casual |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0:26:34 | 0:28:08 | 0:19:25 | 0:36:12 | 0:35:27 | 0:24:31 | 0:37:59 |
| 0:32:59 | 0:24:54 | 0:26:54 | 0:23:09 | 0:27:28 | 0:22:07 | 0:27:10 |
| 0:38:48 | 0:35:24 | 0:25:02 | 0:30:04 | 0:29:53 | 0:25:39 | 0:36:16 |
| 0:32:58 | 0:29:07 | 0:26:19 | 0:21:21 | 0:25:52 | 0:26:08 | 0:34:45 |
| 0:33:43 | 0:32:27 | 0:26:45 | 0:25:48 | 0:29:16 | 0:29:28 | 0:33:23 |
| 0:36:08 | 0:31:29 | 0:31:29 | 0:29:04 | 0:30:12 | 0:32:38 | 0:32:14 |
| 0:33:35 | 0:31:06 | 0:26:23 | 0:24:00 | 0:24:21 | 0:26:05 | 0:32:54 |
| 0:33:45 | 0:28:53 | 0:28:11 | 0:25:39 | 0:25:11 | 0:30:13 | 0:31:57 |
| 0:34:40 | 0:28:37 | 0:21:10 | 0:22:48 | 0:21:11 | 0:27:08 | 0:35:23 |
| 0:32:23 | 0:22:14 | 0:23:20 | 0:20:08 | 0:19:38 | 0:25:21 | 0:30:15 |
| 0:34:06 | 0:17:43 | 0:16:42 | 0:16:24 | 0:24:10 | 0:19:16 | 0:22:34 |
| 0:25:06 | 0:18:18 | 0:15:20 | 0:24:13 | 0:19:42 | 0:29:04 | 0:23:18 |
| 0:32:54 | 0:27:22 | 0:23:55 | 0:24:54 | 0:26:02 | 0:26:28 | 0:31:31 |

| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0:32:54 | 0:27:22 | 0:23:55 | 0:24:54 | 0:26:02 | 0:26:28 | 0:31:31 |

| 1_membe | 2_membe | 3_membe | 4_membe | 5_membe | 6_membe | 7_membe |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0:13:08 | 0:11:28 | 0:12:08 | 0:11:38 | 0:11:37 | 0:11:56 | 0:12:24 |
| 0:12:15 | 0:11:22 | 0:11:16 | 0:10:50 | 0:11:04 | 0:11:41 | 0:11:38 |
| 0:13:29 | 0:12:37 | 0:10:59 | 0:11:58 | 0:10:51 | 0:11:04 | 0:13:41 |
| 0:12:25 | 0:10:48 | 0:10:50 | 0:10:32 | 0:11:24 | 0:10:58 | 0:13:35 |
| 0:14:29 | 0:13:24 | 0:12:47 | 0:12:19 | 0:13:10 | 0:12:49 | 0:14:42 |
| 0:15:41 | 0:13:25 | 0:13:39 | 0:13:06 | 0:13:42 | 0:13:48 | 0:15:10 |
| 0:15:12 | 0:13:23 | 0:12:56 | 0:12:55 | 0:13:01 | 0:13:01 | 0:15:18 |
| 0:14:38 | 0:12:28 | 0:13:04 | 0:13:00 | 0:13:03 | 0:13:33 | 0:14:41 |
| 0:14:08 | 0:12:33 | 0:12:14 | 0:12:24 | 0:12:26 | 0:13:00 | 0:14:47 |
| 0:13:20 | 0:11:25 | 0:11:11 | 0:11:16 | 0:10:57 | 0:11:42 | 0:13:29 |
| 0:11:52 | 0:10:34 | 0:10:47 | 0:10:55 | 0:11:41 | 0:11:21 | 0:11:11 |
| 0:10:51 | 0:10:29 | 0:10:20 | 0:10:23 | 0:10:24 | 0:10:53 | 0:11:17 |
| 0:13:27 | 0:12:00 | 0:11:51 | 0:11:46 | 0:11:56 | 0:12:09 | 0:13:30 |

| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0:13:27 | 0:12:00 | 0:11:51 | 0:11:46 | 0:11:56 | 0:12:09 | 0:13:30 |





Chapter 6

ACT



6.1 Key Tasks

1. Create recommendations for stakeholders
2. Add case study to portfolio

6.2 Deliverable

- ☐ Top three recommendations based on analysis

Once we've wrapped everything up, I'd like to share three suggestions with our stakeholders on how we can turn casual riders into Cyclistic members. These ideas come from our data

analysis and insights. I truly think that by implementing these three steps, we can increase Cyclistic's ability to increase member ridership.

6.3 Three Recommendations

1. Engage in a proactive effort to advertise the benefits becoming a member
2. Use the power of digital media
- 3.

Conclusion

In concluding this data analyst project centered around Cyclistic, I want to emphasize the significance of our findings and recommendations. Through a thorough analysis of the data, we've gained valuable insights that can greatly benefit our stakeholders as they attempt to convert casual riders into Cyclistic members.

Our analysis has given us insight on key trends, patterns, and user behaviors within the biking community. It has allowed us to identify actionable strategies that can enhance Cyclistic's appeal and encourage more riders to become members.

As we move forward, I am confident that the implementation of our recommendations will lead to a significant increase in membership. These suggestions are not educated guesses; they stem from real data and observations, making them practical and achievable.

By adopting these strategies, Cyclistic can effectively engage with its target audience, improve the overall rider experience, and ultimately boost membership numbers. We have a unique opportunity to inspire Chicago residents to choose Cyclistic as their preferred bike-sharing service.

In conclusion, our analysis and recommendations provide a roadmap for Cyclistic's growth in memberships. I'm enthusiastic about the impact these insights can have on our company, and I believe that they will allow our stakeholders to make informed decisions that will drive Cyclistic's success in converting casual riders into members.