OUTLINE

1. Introduction

Located in the metropolitan Los Angeles, Metro Bike Share provides a convenient, flexible and affordable transportation method for Angelenos and visitors to explore the beauty of the city. Metro Bike Share partners with Metro and the City of Los Angeles, setting over 200 bike stations throughout Downtown LA, Central LA, North Hollywood and Westside in 2019. Hence, Riders may easily access to bicycle and complete their first/last mile from their distention by renting Electric Metro Bike, Smart Metro Bike or Classic Metro Bike from Metro Bike Share. In addition, Metro Bike Share offers multiple price plans to fit various requirements of their customers, such as daily pass, monthly pass, annual pass, and other promotion passes for the elder, students, and groups. Similarly, Metro Bike Share offers its business partners more budget plans to their employees.

Since 2016 July, Metro Bike Share has already sold 74,584 passes to its customers, and the total trips reach 889407 till now (Metro Bike Share, 2019). But, to lure more customers use sharing bikes, Metro Bike Share considers promoting its service through several marketing campaigns. As a business consultant to Metro Bike Share, we will picture the current business operation status, customer profiles and several factors that might influence customer behavior. suggest who would be the target audience, when and where should our client launch marketing campaigns and what price plan should be in the marketing campaign via bike data in this report.

2. Data Preparation and Data Cleaning

In order to analyze the business questions, we established the database in bike-sharing past data, weather data, metro bike station data, and holiday data from 2016Q3 to 2019Q2, processing the dataset under three main steps.

- **Data collecting**: Basically, bike-sharing data, bike station location data were downloaded from the Metro Bike Share official website, while past weather data were manually downloaded from Wunderground official website. However, we faced a data missing issue when downloading the weather data. Specifically, the past weather data in the Santa Monica region and Los Angeles region are completely the same, hence we had no other choice but to give up Santa Monica's weather data. Moreover, there are missing values such as pollutants, weather quality, which are the key factors that might influence customer's behavior, in the history of weather data. To solve this problem, we downloaded the missing value from the United States Environmental Protection Agency's official website.

- **Data cleaning**: In this process, we faced several challenges in an inconsistent data column format among different quarterly bike-sharing data, as well as, the date format came with date and time that cause difficulties in selecting a certain period. The inconsistent column format lay in *bike-id* column among bike-sharing data. Some of the datasets demonstrated the numeric format, while some were character. Similarly, the duration column was recorded as second in some dataset, while recorded as minute in other datasets. Hence, we manually checked the column format in each file and unified the data format.

Apart from data format issues, we also faced a human issue in data cleaning which results in the column inconsistency problem, due to different people processed different time period datasets. For example, some might rename the column as *start_station_id* or *end_station_id*, while others kept the original column name as *start_station*. Hence, as the previous solution, we manually checked the column name in each file and unified it. Overall, both data format issues and human issues caused troubles when we combined quarterly bike-sharing data into a whole dataset. But by manually modifying the issues, the problem solved quickly.

- **Import data into SAS**: Lastly, we imported the data into SAS and created new data tables for further analysis. The attributes and time period of all database are shown in Table 1.

Name of Database	Time Period	Columns
Bike Sharing Data	2016Q3-2019Q2	Trip_id, duration, start_date, start_time, end_date, end_time, start_lat, start_lon,end_station, end_lat, end_lon, bike_id, Plan_duration, trip_rout_category, passholder_type, bike_type
Bike Station	-----	Station_ID, Station_Name, Go_Live_Date, Region, Status
Weather Data	2016.07.01-2019.06.30	Date, Avg_Temp, Avg_Humidity, Avg_Wind, Precipitation, Pollutant, Quality, Location,
Holiday Date	2016-2019	Holoday_name, Date

Table1 Data Columns

3. Data Analysis

An efficient and effective marketing campaign relies on a clear business profile, an accurate target audience, and an appropriate time and location. In this section, our first step would be depicting a general company profile in terms of business operation, best-sell product and the peak month. Next, we would like to picture who would be the target audience and their profile in the following analysis.

3.1 Overview of Business Operation

We categorize the research questions into three categories, namely, bike usage, best-selling pass type, and the popular stations.

- **Bike Usage**
  In the beginning, we wonder what the monthly bike usage is every year. Is this change related to the amount of bike supply by Metro Bike Share? Is there a seasonal pattern on bike usage?
  To analyze these questions, we selected year, month, a summation of *trip_id* and summation of *bike_id* from bike sharing dataset. The query can be found in Appendix A1.1 on P.12. Because the datasets of 2016 and 2019 are not a complete year, we used an average amount to examine the annual change in total trips and the amount of bike

supply. As figure 1 shows, the amount of bike supply and trips are positively correlated, both increasing by approximately 136% per year from 2016 to 2018. However, even though the supply of bike was similar in 2018 and 2019, the average bike usage in the first 6 months of 2019 less than that of 2018. As figure 2 shown, bike usage is relatively low but slightly increases from January to Jun, and fast raises and peaks in July and August. Finally, bike usage slightly decreases until December.

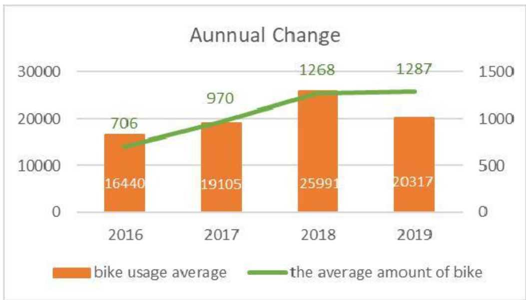


Figure 1 Annual Change

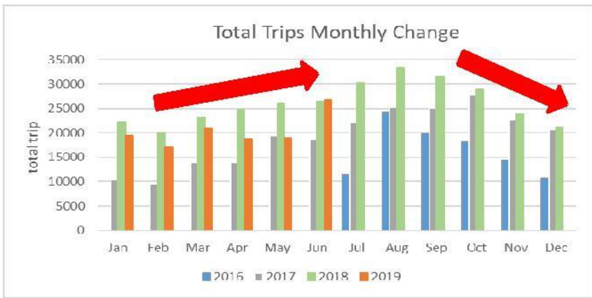


Figure 2 Total Trips Monthly Change

■ **Best-Sell Pass Type**

After understanding the monthly bike usage, our next question would be what's the best-selling pass to riders? What's the average duration among different pass types? Does the popular route change to a different type of pass holder?

To analyze this question we selected year, pass holder type, route category and the summation of *trip_id* in route category. The query can be found in Appendix A1.2 on P.12. Figure 3 indicates that the Monthly Pass is the most popular pass type among customers, accounts for 62% in 2016 but decrease to 53% in 2019, due to more pass options for customers. Monthly pass holders average ride 16 minutes per ride. Following, the walk-up pass is the second most popular pass. The usage of Walk-Up Pass slightly increased from 31% of total usage in 2016 to 36% of that in 2019, and the users usually ride over 1 hour.

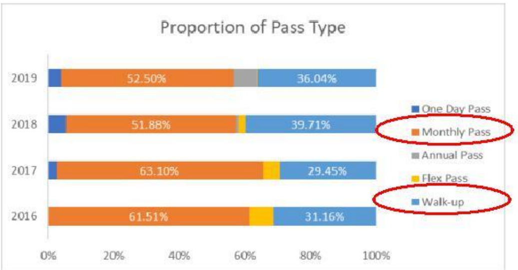


Figure 3 Proportion of Pass Type

Figure 4 Average Trip Duration (min)

The long riding duration of walk-up pass holder could be explained by the route that they chose. We found that 69% of round-trip riders hold Walk-Up Pass. On the contrary, 62% of one-way trips are used by the Monthly Passholders. Overall, to Metro Bike Share, the monthly pass is the top one product certainly, but users tend to use it for short time commute, while walk-up pass is a potential product for the company to expand userbase.

- **Popular Station**

In this section, we wondered where is the popular bike station? We chose 2019 trip data to analyze this research question. The query can be found in Appendix A1.3 on P. 12. The reasons for selecting only 2019 data are that the top 10 stations among different years were stable and the marketing campaign would be effective in the current hot spot. Similarly, we chose the start station under the consideration that most of the bike-sharing users use it for a one-way trip, so the start station converges at certain regions and would be related to the tourist attraction or financial center in downtown LA.

Generally, the bike-share stations located mainly in four regions, DTLA, Westside of LA, Pasadena, and Port of LA. We decided to explore the top 10 popular bike stations in download LA in 2019, due to that 84% of trips happened in Downtown LA (figure 5). Without doubt, the top 10 popular bike stations, which are marked in red in figure 6, located in the financial center, Jewelry District, such as 7^{th}& Flower (station_id: 3005) and Figueroa & 8^{th} (station_id: 3035) Similarly, Railway transportation center, Union station as well as the LA City Hall were the popular bike stations in 2019. Besides, tourist attractions such as union plaza, Walt Disney Concert Hall and art district were also popular among bike-sharing users, which are marked yellow in figure 7.

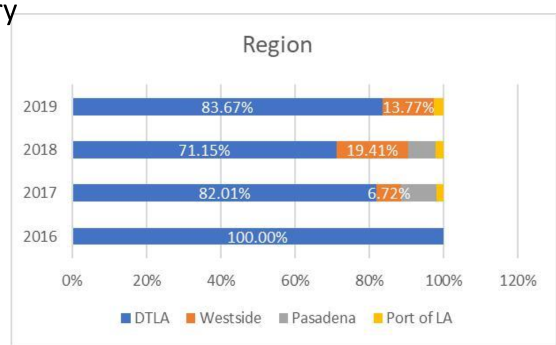


Figure 5 Region

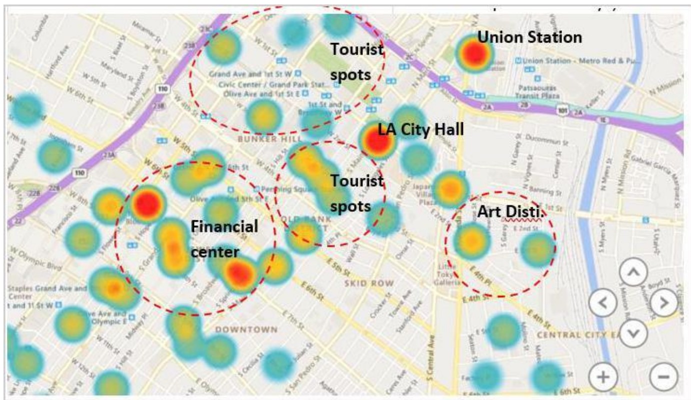


Figure 6 Popular bike-sharing station

3.2 Setting Up Customer Profile

Setting clear customer profiles helps to formulate effective marketing campaigns. Based on analysis of the current dataset, we find out three main customer groups: tourists, commuters and LA residents who don't usually ride. Those groups have very different characteristics, which will be introduced in the following parts.

„ **Methodology**

In order to picture customer behavior, we analyzed the total bike usage distribution in hour and separated the figure into weekdays and weekend. After we got the result, we analyzed the bike stations as well as the pass holder type during the peak hour. Next, we match the characteristic of result into a certain group of people. The querier of trip analysis, station analysis, pass type analysis in weekday could be found in Appendix 2.1 on P.13, while those analysis of weekend analysis are in Appendix 2.2 on P.16.

„ **Finding Four Main Customer Groups**

Commuters

From weekday trip analysis, we found three obvious peak hours, which were around 8am, 13pm, and 18pm as shown in figure 7. Bike usage increase significantly at those three peak hours and then decrease quickly. The ride duration is relatively low, around 20 minutes per trip. Moreover, we cross analyzed the pass type and bike station and discovered that most bike stations during weekday peak hour were located in the city center or are next to other transportation stations. And we also find that most of the passholder types at those stations and at those moments are monthly pass.

Based on our finding, we match the characteristic to commuters. Basically, these group of people go and off work around 8am (80.4% of trips from monthly pass) and 18pm (65.5% of trips from monthly pass) and might rent a

bicycle for their lunch break around 13pm. (57.3% of trips from monthly pass). As the majority of bike station are near to metro or public transportation, commuters rent the shared bike to finish their last mile to work. Moreover, monthly pass is economic choose for commuters to rent a bike every weekday.

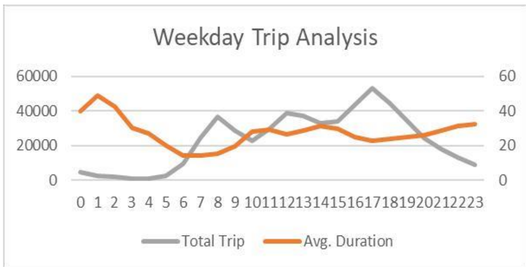


Figure 7 Weekday Trip Analysis

Tourists

In weekend trip analysis, we concluded into a totally different trip usage pattern from weekday. Figure 8 presents one gradual trip rise, which starts from around 6 and begins to fall at 16. The riding duration was quite constant in different hours and relatively longer than that in weekday. In addition, the peak hour on weekends are from 14 to 16. After cross analyzing the bike station and pass type during the peak hour, the most popular stations are around the oceanfront walkway, which is a good place for people to take a walk or take some pictures. And the most passholders found at those stations are walk-up, which is a bulk offer for group riders (Metro Bike Share, 2019).

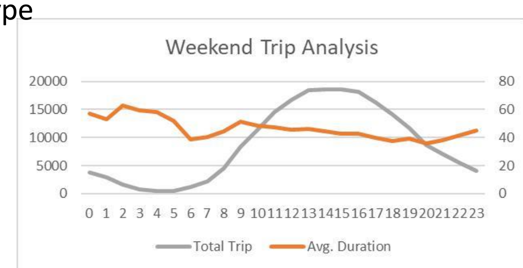


Figure 8 Weekend Trip Analysis

To sum up, according to our finding, we defined these group of users as tourists. They start to go out at 6 and stay outdoors until 16 when they begin to go back home, preferring to purchase walk-up or one-day pass and take a long trip along the oceanfront to enjoy the beauty of Los Angeles.

LA Residents Who Don t Usually Ride

Apart from the commuters and tourists, we discovered a group of people who might be LA residents but don't usually ride. One evidence is from the analysis of the total trips of different passholder types in different years. The queries can be found in Appendix 2.3 on P.18. As figure 9 and figure 10 show, there are some people choosing annual pass but the total trips from annual pass are much smaller than other passholder types like monthly pass and walk-up. We assumed that this might connect to that the annual pass launched from 2018.

People who choose annual pass can't be the tourists or people who only stay in LA for a short period. So, they are most likely to be long-time LA residents. But the problem is that they don't usually ride, which only contributes to a small number of total trips.

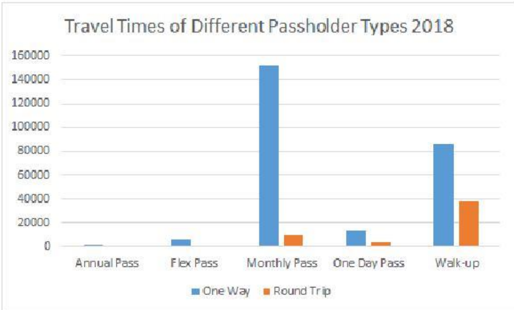


Figure 9 Travel Times of Different Passholder Types 2018

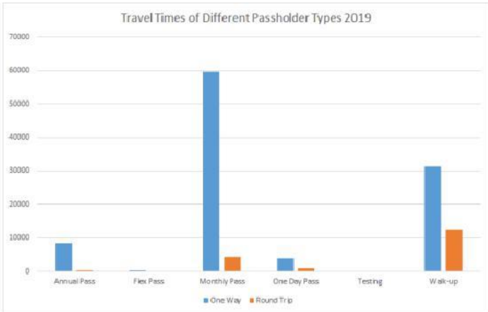


Figure 10 Travel Times of Different Passholder Types 2019

Late-Night Riders

From the trip analysis, we found an interesting fact that 7.9% of total trips occurred during 22:00 to 06:00 in 2019. Both weekday and weekend data all indicate this situation. It's a small amount of customer to Metro Bike Share, but as figure 11 indicates that the average riding duration is average over 50 minutes that is larger than morning period in a day. Also, from figure 12, the pass types were basically monthly pass (48.63%) and Walk-Up pass (41.34%). The queries of late-night rider can be found in Appendix 2.5 on P.19. Moreover, the bike stations which were frequently used during late night for monthly passholder located around the union train station, while that station located around CBD such as Main & 1st (station_id: 3030) and 7th & Flower (station_id: 3005) for walk-up passholders. Consequently, we assume that this category of users composed by the work-overtime commuters, LA residents or tourists.

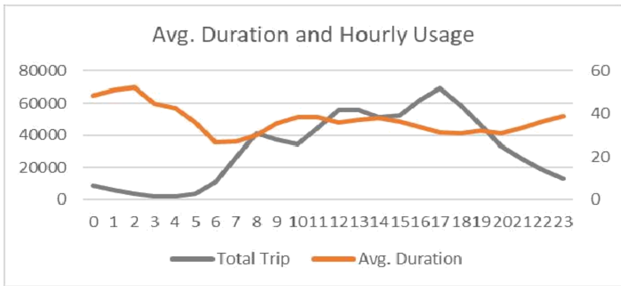


Figure 11 Avg. Duration and Hourly Usage

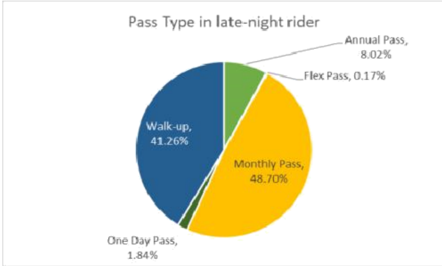


Figure 12 Pass Type in Late-night Rider

3.3 Other Variables in Customer Behavior

■ **The Public Holidays**

We collected the information of American national holidays and calculated the daily usage. The queries of the analysis can be found in Appendix 3.1 P.20. From figure 13, we can tell that during the holidays, the usage of bicycles decreased, especially on New Year's Day and Christmas Day, which is much lower than the average daily usage rate of that year. This result is exactly the opposite of our assumption at the beginning. We believe that this phenomenon could be reasonably explained by that people prefer staying at home more to have a rest instead of hanging out on holidays.

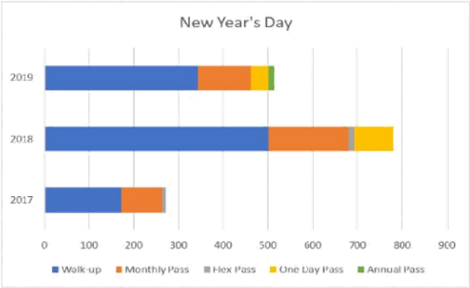
Holiday Usage	2016	2017	2018	2019
Average	536	628	855	675
New Year's Day		262	765	515
Martin Luther King Jr. Day		304	795	556
Washington's Birthday		263	449	624
Cesar Chavez Day		528	775	633
Memorial Day		633	917	638
Independence Day		503	1209	
Labor Day	650	618	1144	
California Admission Day	757	880	874	
Native American Day	687	963	1040	
Veterans Day	549	781	845	
Thanksgiving	251	575	399	
Black Friday	321	693	587	
Christmas Day	230	469	490	

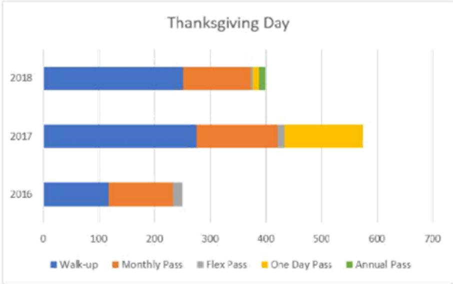
Figure 13 Holiday Usage

Average Walk-up times in JAN	
2017	106
2018	233
2019	242
Average Walk-up times in NOV	
2016	135
2017	232
2018	341
Average Walk-up times in DEC	
2016	98
2017	220
2018	298

Figure 14 Average Walk-up times in Jan

Next, we take New Year's Day, Thanksgiving Day and Christmas Day as examples to analyze the specific situation because of their remarkable data. The results could be finds from figure 13 to figure 17. First of all, we found that at these special holidays, while the total usage was lower than average usage, the bike usage of Walk-up was relatively higher than other types. Especially at New Year's and Christmas Day, Walk-up users can be very high even twice trips than the average happened on that day. Consequently, we conclude that national holiday affect user's riding incentives. Tourists who usually purchase walk-up pass prefer to ride a bicycle during vacations. On the contrary, commuters who accounts for the majority of userbase don't rent a share bike during holidays.





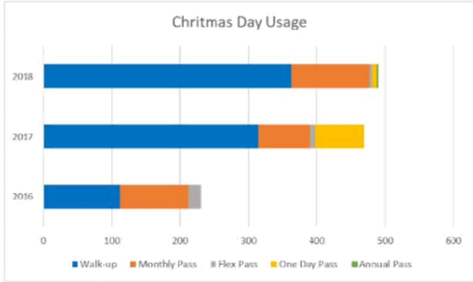


Figure 15 New Year Day

Figure 16 Thanksgiving Day

Figure 17 Christmas Day Usage

- **Weather Analysis**

Weather is another vital factor that influences a rider's behavior. The query could be found in Appendix 3.2 P. 21. Here, we will demonstrate the relationship between weather indicators and trips. We selected average usage by month, monthly average temperature, humidity, total precipitation, and the weather quality from weather dataset. However, the weather indicators of Los Angelos were stable in terms of humidity, temperature, wind, and quality. We found that the temperature and usage amount share a weak positive correlation. People tend to ride a bicycle when it is warm. Figure 18 indicates that the usage amount increases when the temperature becomes higher. However, the correlation between weather and bike usage is only 0.28. It is a weak relationship.

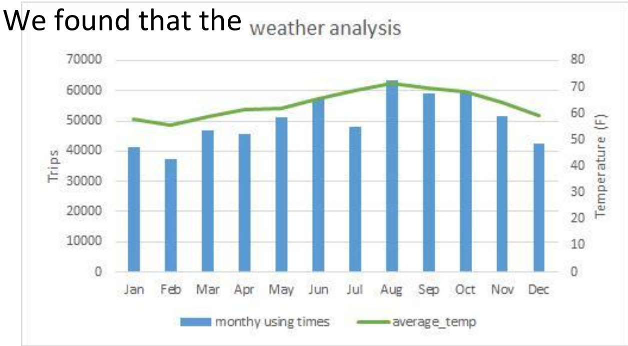


Figure 18 weather analysis

4.Business Insights and Marketing Recommendations

4.1 Business Insights

„ **Company Profile**

Metro Bike Share is a growing business, with a stable increase in bike usage every year. However, we found that even the bike supply remained the same in 2019, the monthly bike usage was relatively low compared with that of 2018 from January to May. Besides, according to the usage data, the monthly usage in the first 6 months is relatively low but slightly increases, after peaking in August then slightly decreases until December. We found one evidence to support this phenomenon which is the temperature. Even though the temperature and bike usage share a weak correlation, these two factors are positively correlated. Lastly, after doing the location analysis, we discovered that the top 10 popular bike stations located in the financial center in LA, railway transportation center and multiple tourist attractions.

„ **Customer Profile**

Customers are divided into four groups: commuters, tourists, LA residents who don't usually ride and late-night rider. Commuters, who are the main customer to Metro Bike Share, are people with monthly pass and usually take trips on daily travel to work and launch time. They have a low average trip duration and take trips in the city center or near the transportation stations. Tourists are people with walk-up pass and usually take trips for group rides on weekends near the tourist attractions. They spend a longer time on trips than commuters. The third group is LA residents who don't usually ride. They have annual pass but only contribute to a small number of total trips. Similarly, the late -night rider account for a small portion of users, they rent the bicycle from 22:00 to 06:00 the next day, holding monthly pass or walk-in pass and usually renting the bike in a long period.

4.2 Marketing Strategies:

„ **Commuters**

Based on the previous analysis, a lot of commuter ride bicycles during the peak hours during weekdays. Even though they're the main and regular customer to our client, we can't ensure that each commuter can stick on shared bike service when there are other competitors in LA. Therefore, we recommend that Metro Bike Share run a series of marketing campaign like "ride more and earn more" to strengthen stickiness among regular customers. Users can earn points when they ride bikes every day and share it to their colleges or friends. This form of campaigns makes daily ride more than a routine to commuters but create a motivation and fun for users. Moreover, the active interaction between users and Metro Bike Share would lead a positive influence on brand image and awareness among users, strengthening the linkage between our client and commuters.

„ **Tourists or Residents Who Prefer Not Ride.**

As analyzed in previous paragraph, we found that tourists and those who do not frequently use bicycles are the main consumers of bike sharing during the holidays. However, this group of people is not the main customer group. Therefore, we recommend that the company should strengthen its brand awareness among this group. The main focus is increasing their usage rate. For example, we can educate the public about the positive effects of riding bicycles and also provide flexible holiday pass for tourists and residents to attract more users.

„ **Late Night Riders**

Based up this finding, we suggest that Metro Bike Share should add more security check measurement to ensure user's safety. For example, the security check would be a real-time location share to friends or family member on the app. Or when some dangerous situation happen, Metro Bike Share could share the GPS of that bike to the police, ambulance or local authority. When customers feel more safety in riding bike at night, they might use it more often. Or it could lure people who don't want to ride a bike at midnight period to try bike share service.

Reference

Metro Bike Share, 2019, Data, Metro Bike Share, Accessed 13. Nov. 2019. <https://bikeshare.metro.net/about/data/>.

Metro Bike Share 2019, *Buy in Bulk*, Metro Bike Share, accessed 10 November 2019, <https://student.unsw.edu.au/how-do-i-cite-electronic-sources>.

United States Environmental Protection Agency, Outdoor Air Quality Data, accessed 10. Nov. 2019, < https://www.epa.gov/outdoor-air-quality-data/download-daily-data>

Appendix

Appendix 1.1 Overview of Business Operation- bike usage analysis

Query

```
proc sql;
select year(start_date) as year, month(start_date) as month,
count(trip_id) as total_number, count(distinct bike_id) as total_number_of_bike
from orion.bike_sharing_all
group by year, month;
quit;
```

Attributes: Year, month, total_number, total_number_bike

year	month	total_number	total_number_of_bike
2016	7	11420	692
2016	8	24153	753
2016	9	19866	742
2016	10	18159	699
2016	11	14427	680
2016	12	10612	671
2017	1	10347	706
2017	2	9533	693
2017	3	13906	739
2017	4	13859	728

Appendix 1.2 Overview of Business Operation- Best-Sell Pass Type

Query

```
proc sql;
select distinct year(start_date) as year, passholder_type, trip_route_category,
count(trip_id) as sum
from orion.bike_sharing_all
group by year, passholder_type, trip_route_category;
quit;
```

Attributes: year, passholder_type, trip_route_category, sum

year	passholder_type	trip_route_category	sum
2016	Flex Pass	One Way	6858
2016	Flex Pass	Round Trip	367
2016	Monthly Pass	One Way	57694
2016	Monthly Pass	Round Trip	2981
2016	Walk-up	One Way	24446
2016	Walk-up	Round Trip	6291
2017	Flex Pass	One Way	10574
2017	Flex Pass	Round Trip	560
2017	Monthly Pass	One Way	135816
2017	Monthly Pass	Round Trip	8852

Appendix 1.3 Overview of Business Operation- Popular station

Query

```
proc sql;
select distinct year(a.start_date) as year, a.start_station, count(a.trip_id) as sum,
```

```
    count(distinct a.bike_id) as bike, a.start_lat, a.start_lon,
      b.station_name, b.Region
from orion.bike_sharing_all as a
inner join orion.metro_bike_share_stations as
b on (a.start_station = b.station_ID) group by
Calculated year,a.start_station
having Calculated year = 2016
order by calculated sum desc;
quit;
```

Attribute: year, start_station, sum, bike, start_lat, start_lon, station_name, region

year	start_station	sum	bike	start_lat	start_lon	Station_Name	Region
2016	3030	3816	739	34.0519409	-118.24353	Main & 1st	DTLA
2016	3030	3816	739	34.051941	-118.24353	Main & 1st	DTLA
2016	3069	3757	742	34.05088	-118.248253	Broadway & 3rd	DTLA
2016	3069	3757	742	34.0508804	-118.24825	Broadway & 3rd	DTLA
2016	3005	3633	737	34.0485	-118.258537	7th & Flower	DTLA
2016	3005	3633	737	34.0485497	-118.25905	7th & Flower	DTLA
2016	3014	3520	737	34.05661	-118.237213	Union Station West Portal	DTLA
2016	3014	3520	737	34.0566101	-118.23721	Union Station West Portal	DTLA
2016	3031	3474	733	34.0447006	-118.25244	7th & Spring	DTLA
2016	3031	3474	733	34.044701	-118.252441	7th & Spring	DTLA
2016	3064	3419	728	34.04681	-118.256981	Grand & 7th	DTLA
2016	3064	3419	728	34.0468102	-118.25698	Grand & 7th	DTLA
2016	3042	2949	726	34.049301	-118.238808	1st & Central	DTLA
2016	3042	2949	726	34.0493011	-118.23881	1st & Central	DTLA

Appendix 2.1 Setting up Customer Profile- Weekday Trip, station, pass type Analysis

„ Total Trips and Average Duration in Different Hours Weekday

Query:

```
proc sql;
title "Total Trips in Different Hours Weekday";
select hour(start_time) as daytime "Hour", count(trip_id) "Total Trips"
from orion.data_of_all
where weekday(start_date) ^= 1
and weekday(start_date) ^= 7
group by daytime;
quit;
title;
```

Attribute: hour, trips

Total Trips in Different Hours Weekday

Hour	Total Trips
0	4554
1	2779
2	1852
3	1111
4	1310
5	2843
6	9651
7	24412
8	36711
9	28967
10	23100
11	29802
12	39065
13	37369
14	32902
15	33967
16	43914
17	53073
18	44737
19	34672
20	24508
21	18170
22	13154
23	8849

„ Average Duration in Different Hours Weekday

Query:

```
proc sql;
title "Average Duration in Different Hours Weekday";
select hour(start_time) as daytime "Hour", sum(duration)/count(trip_id) "Average Duration" format=4.1
from orion.data_of_all
where weekday(start_date) ^= 1
and weekday(start_date) ^= 7
and duration ^= 1440
group by daytime;
quit;
title;
```

Attribute: hour, average duration

Average Duration in Different Hours Weekday

Hour	Average Duration
0	40.1
1	49.1
2	42.6
3	30.3
4	26.9
5	20.2
6	14.5
7	14.2
8	15.3
9	19.9
10	28.3
11	29.4
12	26.6
13	28.5
14	31.5
15	29.9
16	25.2
17	22.9
18	24.0
19	25.2
20	25.9
21	28.6
22	31.3
23	32.3

„ Top 10 Stations with Most Trips During Peak Hours at Weekdays

Query:

```
proc sql;
title "Top 10 Stations at 9 Weekday";
select start_station "Station", count(trip_id) as trips "Trips"
from orion.data_of_all
where weekday(start_date) ^= 1
and weekday(start_date) ^= 7
and hour(start_time) = 9
group by start_station
order by trips desc;
quit;
title;
```

Attribute: station, trips

Top 10 Stations at 9 Weekday

Station	Trips
3014	1733
3042	1255
3034	1245
3031	1140
3005	870
3082	830
3064	789
3030	776
3038	714
3049	693

„ Passholder Types of Top 3 Stations at 9/13/18 Weekday

Query:

```
proc sql;
title "Passholder Types of Top 10 Stations at 9 Weekday";
select start_station "Station", passholder_type "Passholder Type", count(trip_id) as trips "Trips"
from orion.data_of_all
where weekday(start_date) ^= 1
and weekday(start_date) ^= 7
and hour(start_time) = 9
and start_station = 3034
group by start_station, passholder_type;
quit;
title;
```

Attribute: station, passholder type, trips

Passholder Types of Top 10 Stations at 9 Weekday

Station	Passholder Type	Trips
3034	Annual Pass	108
3034	Flex Pass	79
3034	Monthly Pass	997
3034	One Day Pass	2
3034	Walk-up	59

Appendix 2.2 Setting up Customer Profile- Weekend Trip, station, pass type Analysis

„ Total Trips and Average Duration in Different Hours Weekends

Query

```
proc sql;
title "Total Trips in Different Hours Weekend";
select hour(start_time) as daytime "Hour", count(trip_id) "Total Trips"
from orion.data_of_all
where weekday(start_date) = 1
or weekday(start_date) = 7
group by daytime;
quit;
title;
```

Attribute: hour, total trips

Total Trips in Different Hours Weekend

Hour	Total Trips
0	3763
1	2913
2	1708
3	798
4	462
5	432
6	1160
7	2271
8	4502
9	8327
10	11421
11	14636
12	16699
13	18496
14	18596
15	18565
16	18095
17	16249
18	14124
19	11690
20	8734
21	7031
22	5454
23	4091

„ Average Duration in Different Hours Weekend

Query

```
proc sql;
title "Average Duration in Different Hours Weekend";
select hour(start_time) as daytime "Hour", sum(duration)/count(trip_id) "Average Duration" format=4.1
from orion.data_of_all
where weekday(start_date) = 1
or weekday(start_date) = 7
and duration ^= 1440
group by daytime;
quit;
title;
```

Attribute: hour, average duration

Average Duration in Different Hours Weekend

Hour	Average Duration
0	57.4
1	53.2
2	62.6
3	59.5
4	58.1
5	52.1
6	38.8
7	40.3
8	44.7
9	51.2
10	48.5
11	47.5
12	45.0
13	46.0
14	44.6
15	42.7
16	42.5
17	40.1
18	37.7
19	39.5
20	35.9
21	37.9
22	41.7
23	45.2

„ Top 10 Stations with Most Trips During Peak Hours on Weekends

Query

```
proc sql;
title "Top 10 Stations at 14 Weekend";
select start_station "Station", count(trip_id) as trips "Trips"
from orion.data_of_all
where weekday(start_date) ^= 2
and weekday(start_date) ^= 3
and weekday(start_date) ^= 4
and weekday(start_date) ^= 5
and weekday(start_date) ^= 6
and hour(start_time) = 14
group by start_station
order by trips desc;
quit;
title;
```

Attribute: station, trips

Top 10 Stations at 14 Weekend

Station	Trips
4214	985
4210	632
3069	630
3082	537
3005	501
4215	473
3031	461
3064	372
3067	370
3038	368

„ Passholder Types of Top 3 Stations at 14/15/16 Weekend

Query

```
title "Passholder Types of Top 3 Stations at 14 Weekend";
select start_station "Station", passholder_type "Passholder Type", count(trip_id) as trips "Trips"
from orion.data_of_all
where weekday(start_date) ^= 2
and weekday(start_date) ^= 3
and weekday(start_date) ^= 4
and weekday(start_date) ^= 5
and weekday(start_date) ^= 6
and hour(start_time) = 14
and start_station = 3069
group by start_station, passholder_type;
quit;
title;
```

Attribute: station, trips

Top 10 Stations at 14 Weekend

Station	Trips
4214	985
4210	632
3069	630
3082	537
3005	501
4215	473
3031	461
3064	372
3067	370
3038	368

Appendix 2.3 Setting up Customer Profile- Travel Times Analysis

Travel Times of Different Passholder Types 2018/2019

Query

```
proc sql;
title "Travel Times of Different Passholder Types 2018";
select passholder_type "Passholder Type", count(trip_route_category) "One Way"
from orion.data_of_all
where trip_route_category = "One Way"
and year(start_date) = 2018
group by passholder_type;
quit;
title;
```

Attribute: passholder type, one-way

Travel Times of Different Passholder Types 2018

Passholder Type	One Way
Annual Pass	1997
Flex Pass	6463
Monthly Pass	151452
One Day Pass	13600
Walk-up	85564

Appendix 2.4 Setting up Customer Profile: Passholder Types at 8/13/18 at Weekdays

```
proc sql;
title "Passholder Types at 8 Weekday";
select passholder_type "passholder_type", count(trip_id) as trips "Trips"
from orion.data_of_all
where weekday(start_date) ^= 1
and weekday(start_date) ^= 7
and hour(start_time) = 8
group by passholder_type;
quit;
title;
```

Attribute: passholder_type, trips

Passholder Types at 8 Weekday

passholder_type	Trips
Annual Pass	920
Flex Pass	1159
Monthly Pass	29533
One Day Pass	309
Testing	1
Walk-up	4789

Appendix 2.5 Setting up Customer Profile- Late-night rider analysis

„ Select data from the database to analyze

Query:

```
proc sql;
select trip_id, year(start_date) as year, weekday(start_date) as weekday, month(start_date) as month,
start_date, end_date, hour(start_time) as hour, duration, start_station, start_lat, start_lon,
passholder_type
from orion.bike_sharing_all
where Calculated year = 2019
and duration <1440
and 1< Calculated weekday < 7;
quit;
```

Attribute: trip_id, year, weekday, month, start_date, end_date, hour, duration, start_station, start_lat, start_lon, passholder_type

trip_id	year	weekday	month	start_date	end_date	hour	duration	start_station	start_lat	start_lon	passholder_type
112536773	2019	3	1	01JAN2019	01JAN2019	0	7	3046	34.052872	-118.24749	Walk-up
112536772	2019	3	1	01JAN2019	01JAN2019	0	6	3046	34.052872	-118.24749	Walk-up
112538689	2019	3	1	01JAN2019	01JAN2019	0	32	3030	34.051941	-118.24353	Walk-up
112538688	2019	3	1	01JAN2019	01JAN2019	0	30	3030	34.051941	-118.24353	Walk-up
112538687	2019	3	1	01JAN2019	01JAN2019	0	28	3030	34.051941	-118.24353	Walk-up
112538686	2019	3	1	01JAN2019	01JAN2019	0	28	3030	34.051941	-118.24353	Walk-up
112538685	2019	3	1	01JAN2019	01JAN2019	0	27	3030	34.051941	-118.24353	Walk-up

„ The hourly bike usage during late night

Query:

```
proc sql;
select distinct a.hour,count(trip_id) as sum, avg(a.duration) as avg_duration
from
(select trip_id, year(start_date) as year, weekday(start_date) as weekday, month(start_date) as
month, start_date, end_date, hour(start_time) as hour, duration,
start_station, start_lat, start_lon, passholder_type
from orion.bike_sharing_all where Calculated
year = 2019
and duration <1440
and 1< Calculated weekday < 7) as a
group by a.hour;
quit;
```

Attribute: hour, sum, avg_duration

hour	sum	avg_duration
0	668	34.86527
1	326	33.86503
2	218	28.27523
3	140	26.22143
4	196	22.5102

Appendix 3.1 Other Variables in Customer Behavior -Holiday Analysis

In this analysis, we search the US's national holidays on the internet and summarized the date into the table as below:

Holiday Information				
	2016	2017	2018	2019
New Year's Day		01JAN2017	01JAN2018	01JAN2019
Martin Luther King Jr. Day		16JAN2017	15JAN2018	21JAN2019
Washington's Birthday		20FEB2017	19FEB218	18FEB2019
Cesar Chavez Day		31MAR2017	31MAR2018	01APR2019
Memorial Day		29MAY2017	28MAY2018	27MAY2019
Independence Day	04JUN2016	04JUN2017	04JUN2018	
Labor Day	05SEP2016	04SEP2017	03SEP2018	
California Admission Day	09SEP2016	09SEP2017	09SEP2018	
Native American Day	23SEP2016	22SEP2017	28SEP2018	
Veterans Day	11NOV2016	10NOV2017	12NOV2018	
Thanksgiving	24NOV2016	23NOV2017	22NOV2018	
Black Friday	25NOV2016	24NOV2017	23NOV2018	
Christmas Day	26DEC2016	25DEC2017	25DEC2018	

„ The daily usage in a specific date

Query

```
proc sql;
title 'Daliy Usage';
select start_date, count(trip_id) as times
from orion.bike_sharing_all
where start_date = '27MAY2019'd
```

```
group by start_date;
quit;
title;
```

Daliy Usage

start_date	times
27MAY2019	638

„ The usage of different passholder types in the specific holidays

Query

```
proc sql;
title 'Christmas Day Usage';
select start_date 'Christmas Day', passholder_type 'Passholder Type', count(start_date) as
Trips from orion.bike_sharing_all
where start_date = '26DEC2016'd and start_station<>3000 or
        start_date = '25DEC2017'd and start_station<>3000 or
        start_date = '25DEC2018'd and start_station<>3000
group by start_date, passholder_type
order by start_date, Trips desc;
quit;
```

Attribute: name of holiday, passholder_type, Trips

Christmas Day Usage

Christmas Day	Passholder Type	Trips
26DEC2016	Walk-up	111
26DEC2016	Monthly Pass	101
26DEC2016	Flex Pass	18
25DEC2017	Walk-up	314
25DEC2017	Monthly Pass	75
25DEC2017	One Day Pass	71
25DEC2017	Flex Pass	9
25DEC2018	Walk-up	362
25DEC2018	Monthly Pass	115
25DEC2018	One Day Pass	7
25DEC2018	Flex Pass	4
25DEC2018	Annual Pass	2

Page Break

Appendix 3.2 Other Variables in Customer Behavior -Weather Analysis

„ Combine the bike data and weather data

Query

```
proc sql;
create table orion.bike_weather_dataset as
select distinct a.start_date as date, count(a.trip_id) as total_num, c.Region as station_name,
    sum(geodist(a.start_lat, a.start_lon, a.end_lat, a.end_lon)) as daily_distance format=6.2,
    b.avg_temp, b.avg_humidity, b.avg_wind, b.pollutant, b.quality, b.Precipation
from orion.bike_sharing_all as a
```

```
inner join
    orion.weather_data as b
            on (a.start_date = b.date)
inner join
   orion.metro_bike_share_stations as c
          on (a.start_station = c.Station_ID)
where c.Region = "DTLA"
group by a.start_date, c.Region
order by a.start_date asc;
quit;
```

Attribute: date, total_num, station_name. daily_distance, avg_temp, avg_humidity, avg_wind, pollutant, quality, precipation,

date	total_num	station_name	daily_distance	avg_temp	avg_humidity	avg_wind	pollutant	quality	Precipation
07JUL2016	190	DTLA	211.40	58.2	78.9	15.8	10.5	Excellent	0
08JUL2016	341	DTLA	363.74	61.1	78	13.9	10.7	Excellent	0
09JUL2016	416	DTLA	462.35	63.1	77.5	15.6	10.9	Excellent	0
10JUL2016	331	DTLA	375.32	62.8	66.2	19.6	9.4	Excellent	0
11JUL2016	347	DTLA	379.61	61.3	69.3	16	12	Excellent	0
12JUL2016	531	DTLA	25688	60.4	72.4	15.5	15.1	Good	0

„ Select monthly average data

Query

```
proc sql;
select month(date) as month,
sum(total_num) as sum, sum(daily_distance) as distance, avg(avg_temp) as average_temp,
avg(avg_humidity) as average_humidity, avg(avg_wind) as average_wind, sum(Precipation) as
Precipation
from orion.bike_weather_dataset
group by month;
quit;
```

Attribute: month, sum, distance, average_temp, average_humidity, average_wind, precepation.

month	sum	distance	average_temp	average_humidity	average_wind	Precipation
1	41263	44944.37	57.95376	67.84516	6.517204	13.19
2	28062	30516.2	55.5807	63.15088	8.126316	5.3
3	81556	87614.27	58.56083	72.69583	7.845833	8.31
4	45582	49868.04	61.64444	69.03556	8.688889	0.45
5	51394	55851.99	61.85914	73.83333	8.617204	0.95
6	57630	63236.93	65.65556	77.17667	7.984444	0
7	48107	151786.1	68.7908	75.47931	9.864368	0
8	63643	205613.9	71.22903	76.47097	8.017204	0
9	59042	176803.9	69.80111	73.00778	7.608889	0.09
10	59697	65624.21	68.32903	67.26882	6.807527	0.94
11	51407	55127.09	64.25165	62.54835	6.13956	3.3
12	42466	44627.11	58.9871	59.25161	6.160215	4.17

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.028953
R Square	0.000838
Adjusted R	-0.0002
Standard Er	254.8783
Observatior	961