DSO-545 Final Report

LA Metro Bike Share Analysis

Haoyi Su

Jialu Lu

Shentan Zhang

Yuan Li

Yiqing Song

Yin Wang

Instructor: Dr. Abbass Al Sharif

04/20/2019

Abstract

Students from the University of Southern California start a project to clean and analyze the past two years' data from Metro Bike, a bike share company whose business is in Greater Los Angeles area. The objective of this study is to find out the company's current problems and opportunities in general operation, bike allocation, client behavior, pricing strategy and new bike introduction by implementing exploratory data analysis with statistical computing and data visualization methods. Finally, based on the findings of this study, detailed recommendations are proposed and the plan for future work is made.

1. Introduction

Bike share is a popular service used by people at all ages. It's common to see people using bike share on campus, at bus and subway stations, in residential area and business districts, almost everywhere. Sharing bicycle is a new kind of environmental protection sharing economy. As a time-sharing rental model, bike share has great potential and development space.

Metro Bike was launched in 2016 and mainly operates in the Greater Los Angeles area. It aims to provide convenience for citizens for short trips. With smart lock integrated with GPS and communication module on each bike, users can easily locate and use the nearby Metro Bikes anytime and anywhere through the smartphone app. After arriving the destination, users need to park the bike on a fixed parking pile. To enhance customer loyalty, besides pay-per-ride, the Metro Bike offers three bike-sharing pass options that allow long-term customers to utilize the bikes at reduced rates.

This project focuses on analyzing and visualizing the behavioral data of Metro Bike share in the last two years. The data analysis would provide the company with insights in all key aspects for future system upgrade including bike allocation, marketing strategies, pricing policy adjustment, etc.

2. Data Description

To analyze trips data of metro bikes in Los Angeles, we use datasets from LA Metro Bike. The link of raw datasets is: https://bikeshare.metro.net/about/data/.

- These files are separated by quarters. During the process of data cleaning, they are combined into a new dataset which includes data from 2017 to 2018. The combined dataset contains 541,149 records of metro bikes' trips and the following 15 fields.
- The fields of the combined dataset are:

- **trip_id**: Locally unique integer that identifies the trip
- **duration**: Length of trip in minutes
- **start_time**: The date/time when the trip began, presented in ISO 8601 format in local time
- **end_time**: The date/time when the trip ended, presented in ISO 8601 format in local time
- **start_station**: The station ID where the trip originated (for station name and more information on each station see the Station Table)
- **start_lat**: The latitude of the station where the trip originated
- **start_lon**: The longitude of the station where the trip originated
- **end_station**: The station ID where the trip terminated (for station name and more information on each station see the Station Table)
- **end_lat**: The latitude of the station where the trip terminated
- **end_lon**: The longitude of the station where the trip terminated
- **bike_id**: Locally unique integer that identifies the bike
- **plan_duration**: The number of days that the plan the passholder is using entitles them to ride; 0 is used for a single ride plan (Walk-up)
- **trip_route_category**: "Round Trip" for trips starting and ending at the same station or "One Way" for all other trips
- **passholder_type**: The name of the passholder's plan
- **bike_type**: The kind of bike used on the trip, including standard and electric assist bikes.

3. Data Cleaning

- **Unify column names:** Make the column names of all the data frames aligned to combine all the quarterly trip data
- **Data combination:** Combine all the data frames through ‘rbind’ command
- **Missing values handling:** Bike Station 3039 and 4108 lack latitude and longitude info in 2017Q4 quarterly report but do not miss that info in the others. Thus, missing location info is added through referring to the other intact information in other reports.
- **Mutate new variables:** Some new variables concerning time, such as quarter, month, year, weekday, hour and minute, and variables concerning location, including start region and end region, are created.
- **Modify passholder type:** According to the “Pricing” page on the official website, Metro Bike has launch three passes, including Flex Pass (Annual Pass), Monthly Pass, and One-day Pass. But in the raw tables, the company used both Flex Pass and Annual Pass to indicate the Yearly Pass. Thus, Annual Pass is renamed as Flex Pass.
- **Align the unit of duration**: In 2017Q1 report, the unit of duration is second but in the others the unit is minute. Thus, re-calculation is implemented on variable Duration in 2017Q1 to represent the length of trips in minutes.

4. Exploratory Data Analysis

4.1 General Operating Situation

The analysis on Metro Bike begins with an overview of general trend by month. Figure 1 indicates the same trend in 2017 and 2018. In total, the number of trips is obviously increasing from 2017 to 2018, which means Metro Bike's business has been in a period of steady growth and expansion during 2018. However, in the last quarter of 2018, the difference between the two years' trip count is very small, which might be a sign of saturation. Therefore, after a rapid growing period, Metro Bike is expected to enter a plateau in 2019.

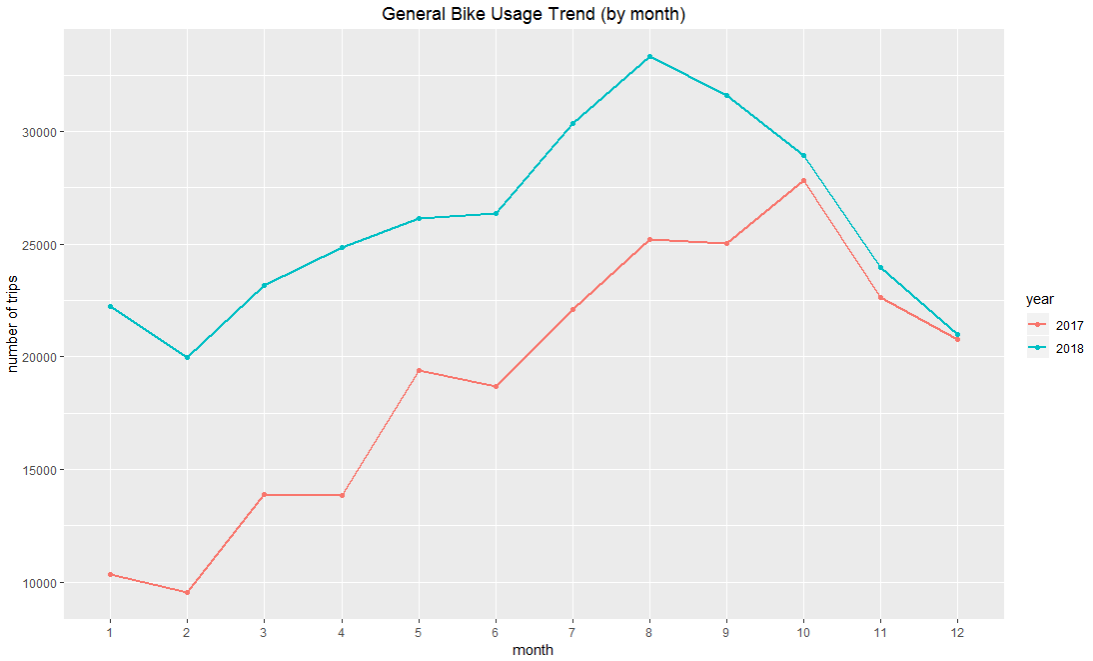

Figure 1. General Bike Usage Trend

Within each year, an upward trend from the first quarter to approximately the third quarter is followed by a drop in the last quarter. Therefore, it indicates that bike renting has seasonality that people prefer to rent bikes at warmer weather than a cold one. The third quarter of every year, which is the peak of bike share usage, requires more attention on bike allocation to deal with the problems of bike shortage.

4.2 Geographical Distribution

4.2.1 General Station Usage Analysis

Visualizing the station usage onto LA city map, Figure 2 shows the overall station usage in the past two years. There are mainly four operation regions: DTLA, Santa Monica,

Port of LA and Pasadena. By comparing the station usage in different regions by the time variable quarter, Metro Bike can understand the changes occurring in the overall LA market.

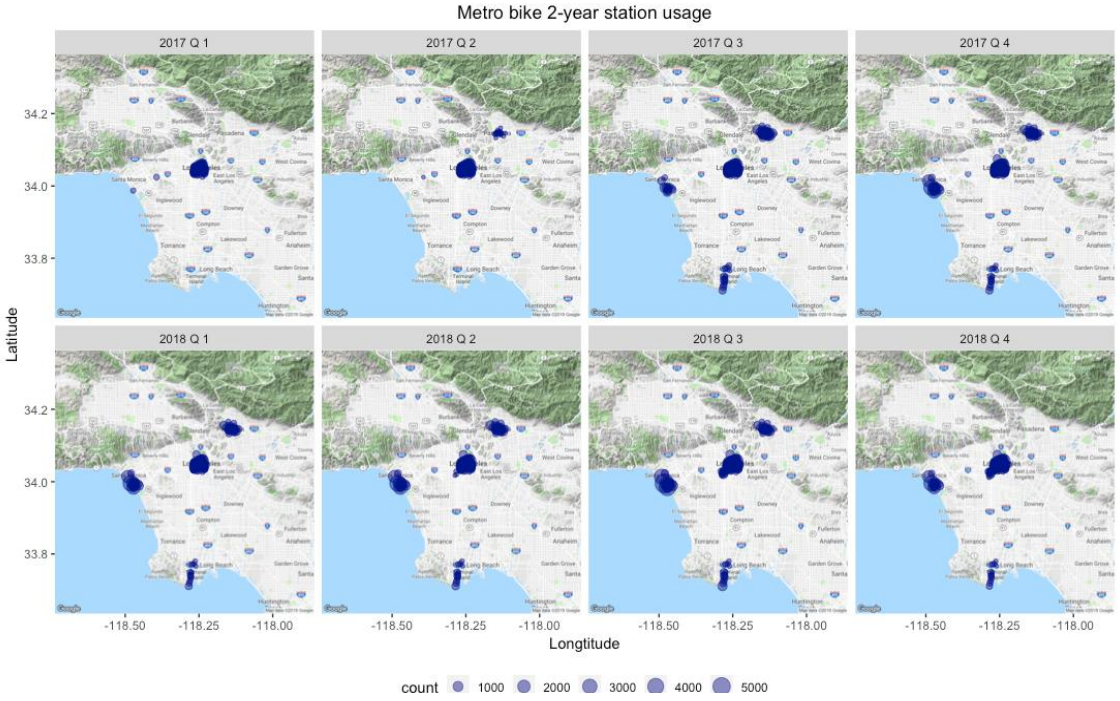

Figure 2. General Station Usage Maps

According to the maps, it is obvious that Metro Bike start with mainly DTLA region, and then expanded its business scope to Pasadena, and then came to Port of LA (Long beach area) and Santa Monica. By the end of 2017, Metro Bike had already built its market in four major regions. However, in 2018Q4, Metro Bike stopped its operation in Pasadena so that only three regions were left. Referring to the chart in Figure 3 concerning the station usage changes by region,

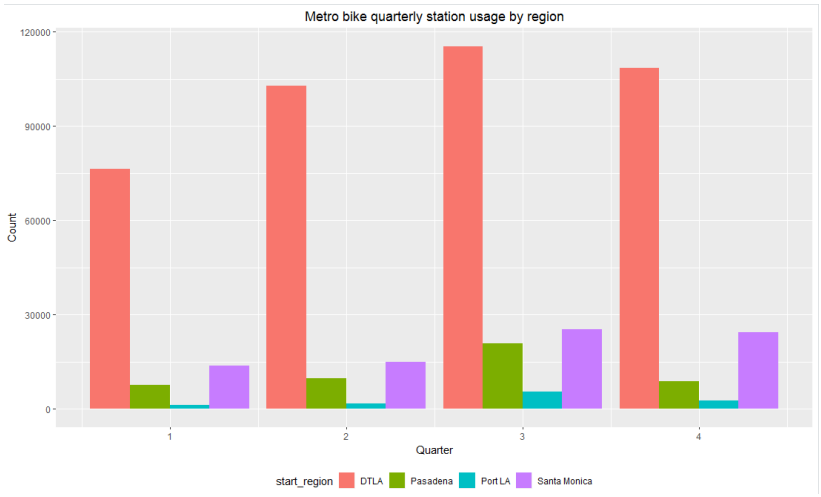

Figure 3. Quarterly Station Usage by Region

the reason why the Pasadena operation was suspended may be the station usage of Pasadena declined continually by quarter from 2017Q3. For the future operation, DTLA and Santa Monica will still be the two largest contributors.

4.2.2 Regional Station Usage Analysis

For each of the four regions, analysis is based on its map.

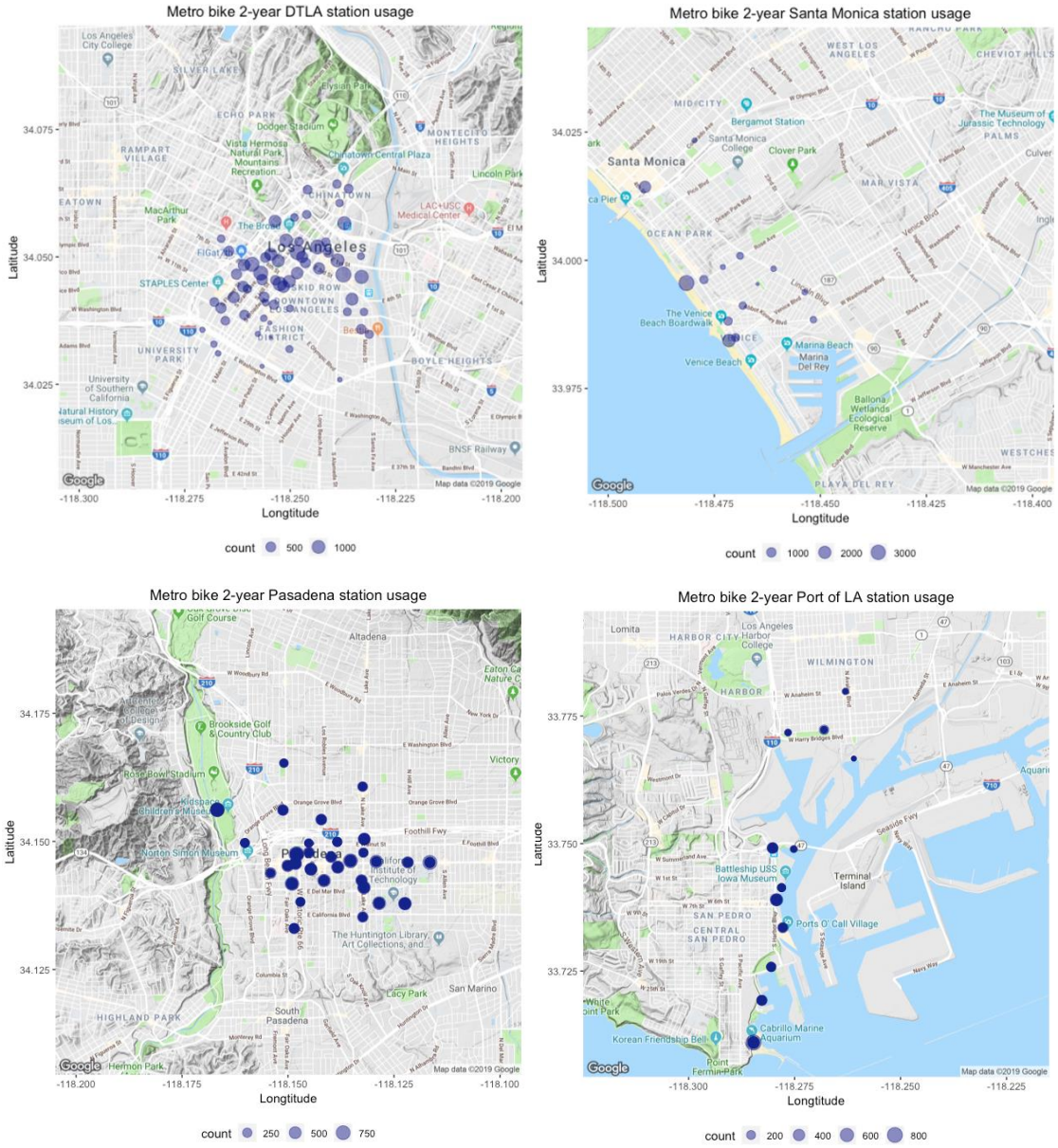

Figure 4. Station Usage Map of Four Regions

- **DTLA** (the largest market):

In the DTLA area, the Metro Bike stations are mainly distributed from Staples Center all the way down to the old Chinatown, plus the USC campus. Also, this area has a

significantly heavy bike share usage. It may be because that the parking spaces are so expensive and limited. Office workers may prefer to use the Metro Bike to finish the last mile of commuting. Surprisingly, the metro bike around the USC campus is not quite frequently used. The reason may be students choose some more convenient ways for school, such as by skateboards, their own bikes, shuttles and cars.

- **Santa Monica**

Santa Monica came into the market in 2017Q3, and it rapidly surpassed Pasadena and grew to be the second largest market. The Metro Bike system is especially welcomed in the beach areas, where users can simply tap to rent a bike and ride it alongside the coastline.

- **Pasadena**

The Metro Bike system in the Pasadena region has a smaller usage than the one in DTLA.

The maximum count appears in the center of Pasadena, around some shopper places, such as Old Pasadena, and California Institute of Technology.

- **Port of LA**

The smallest market in the LA mainly operated around sightseeing places.

4.2.3 Bike Supply Situation

For bike share companies, the allocation of bike can be a big challenge. Due to the imbalance of bike flux between starting and ending regions, some regions are lacking in bikes during rush hour while others may have a plethora of bikes. This could be a waste of resource and would have a bad influence on user experience. Therefore, it's of great interest to find the regions always short of bikes and those with superfluous bikes all the time.

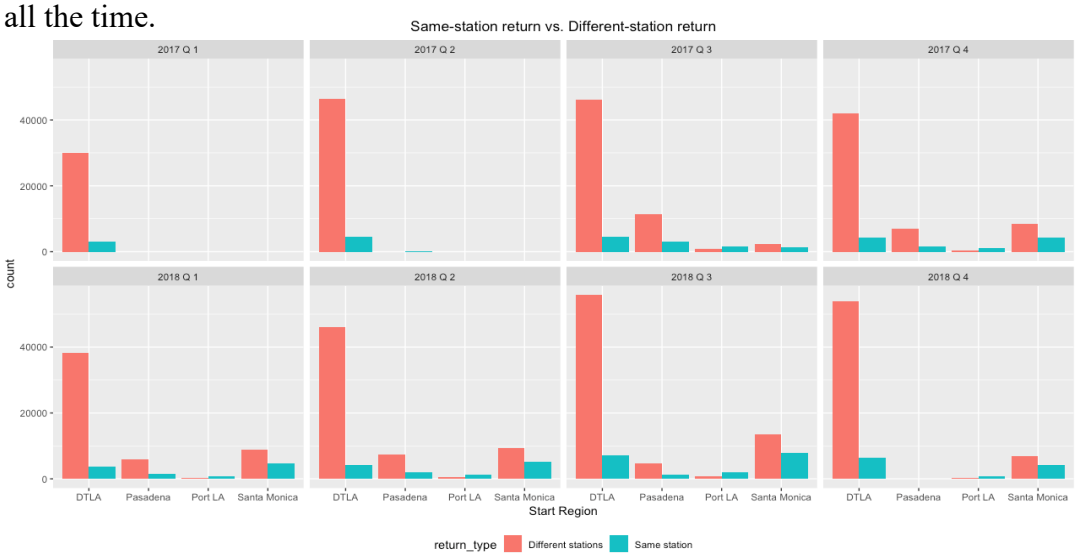

Figure 5. Same and Different Station Return Comparison

Firstly, the bike return behavior is studied to provide the company some insight for bike allocation management. Figure 5 shows how people return bikes they used. It seems that the return behaviors vary across regions but does not change a lot across regions. Obviously, people in DTLA and Pasadena tend to return their bikes in different stations while in Santa Monica and Port of LA people are more likely to return to the same station. It might be because that people in DTLA and Pasadena used the bike system for short-trip commuting, such as office workers ride bikes for a meal at break time or students go to and from campus. But in Santa Monica and Port of LA, the users are more probably to rent a bike for sightseeing, they may park cars and then take bikes nearby to get around, and after trips, they will return their bikes to the same station and walk to their cars. For now, this is just a conjecture. More evidence will be analyzed in the following sections to verify it.

Secondly, noticing that different-station return may cause unbalanced bike allocation, it is necessary to look at where the bikes travel to. By regions, for all the different-station returns, Figure 6 indicates that Santa Monica has the highest ratio that the bikes are returned to other regions, while in other regions bikes are almost move within the regions. A surmise is made that people who start their trips at Santa Monica are more likely to do sightseeing more than just one area, they may put the bikes into their car and use them at the next destinations.

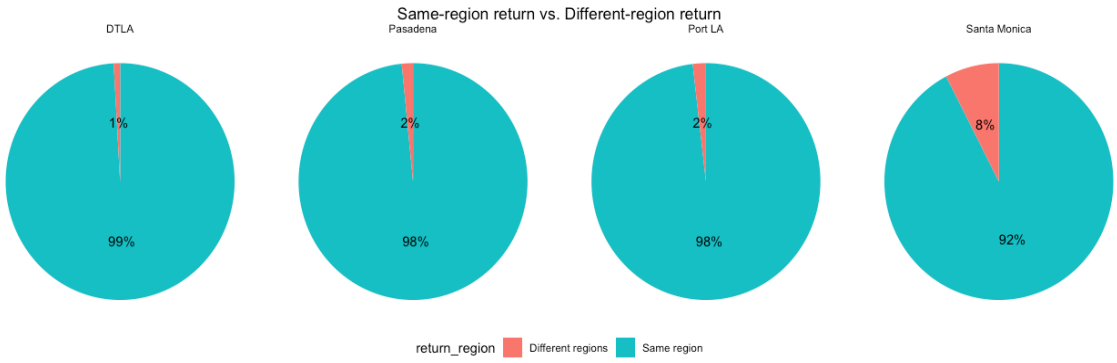

Figure 6. Same and Different Station Return Proportion

However, due to the market size, DTLA, as the start region, still has the largest number of bikes that are returned to other regions. According to the map in Figure 7 showing the trip start stations (red points), end stations (pink points) and the routes, the Metro Bikes mainly flow among DTLA, Pasadena and Santa Monica. The busiest route of inter-regional ride trip is between Santa Monica and DTLA.

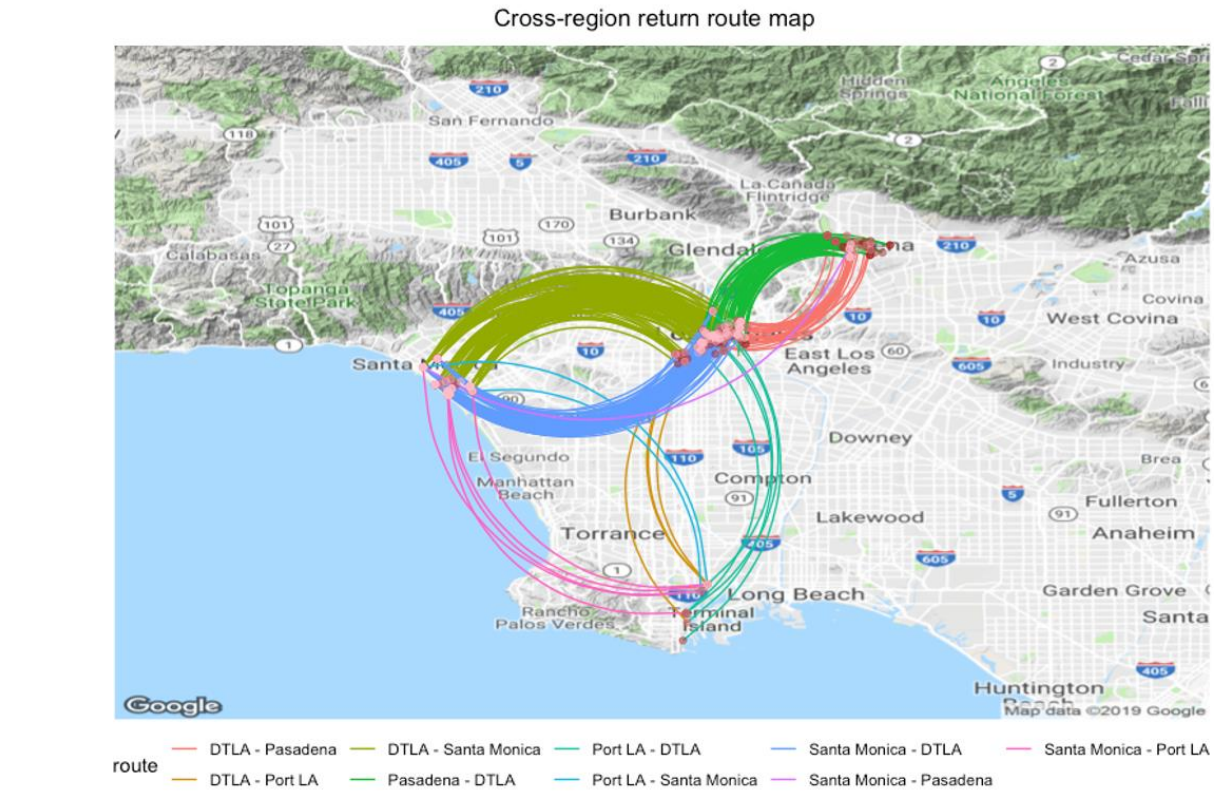

Figure 7. Cross-Region Return Route Map

Then, to figure out exactly which region is short of bike supply, regional bike shortages are calculated by the difference of bike gain from other regions and bike loss to other regions. Figure 8 shows the regional shortage situation by year. It seems that the storage region varies with year. In 2018, DTLA turned to the largest bike shortage region from a former bike surplus region.

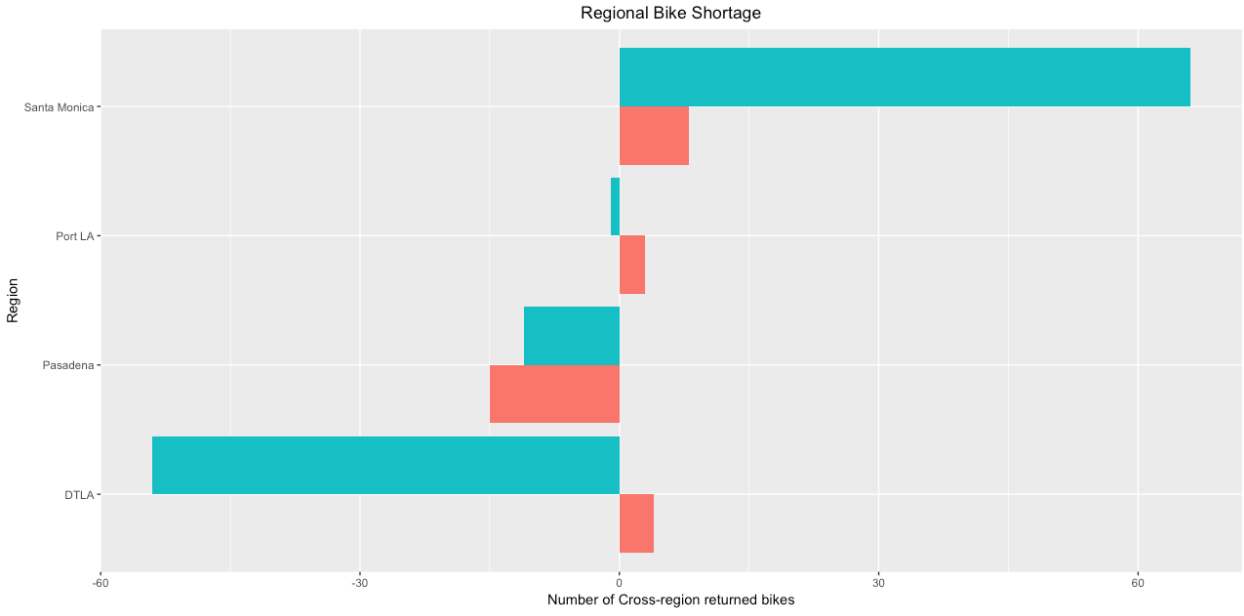

Figure 8. Bike Flux - Region

Finally, to dig even deeper into the question, the flux between stations are analyzed to find out the exact stations with plethora and shortage of bikes. Figure 9 shows the net bike flux of each station in year 2017 and 2018. Net bike flux is calculated as the difference between bikes inflow and outflow. Stations with net bike flux below 100 are defined as “balanced stations”. These stations are not included in the charts since the purpose is to find the extreme cases.

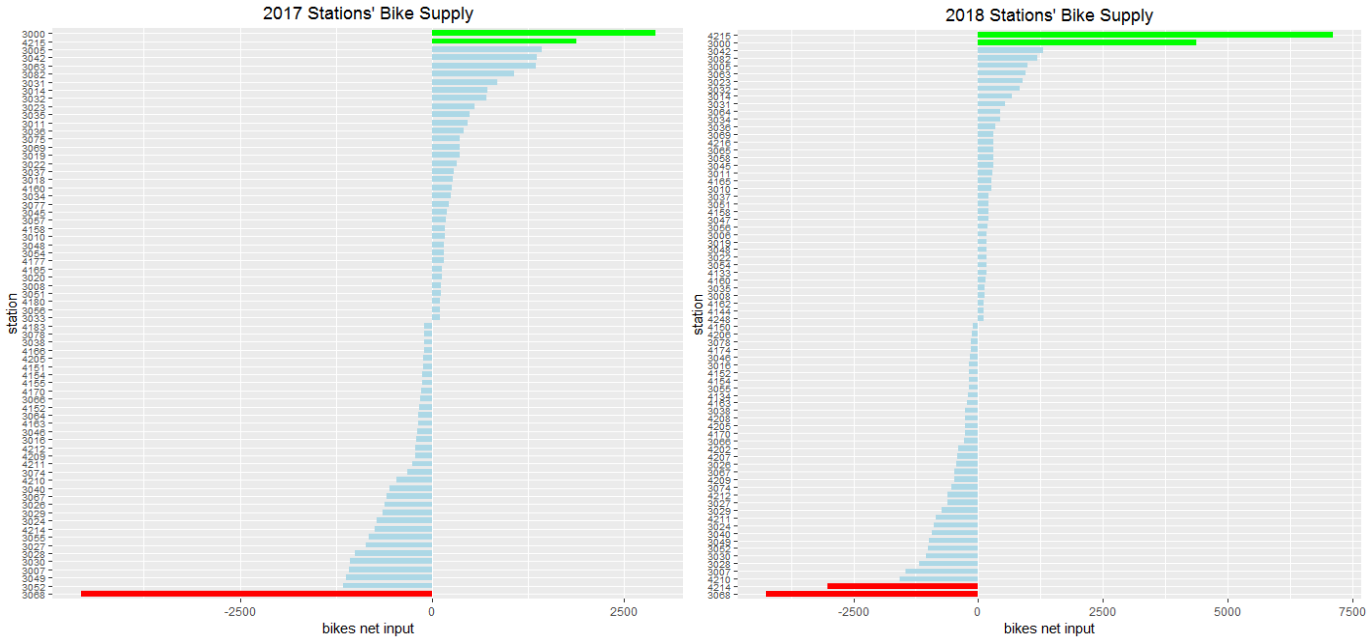

Figure 9. Bike Flux - Station

From the charts above, in both years, station 3068 is the one most lacking in bikes. Apart from station 3000 which is a virtual station and not taken into consideration, station 4215 is the one with most redundant bikes. Station 4215 is in Santa Monica and 3068 is in downtown. Therefore, to supplement bikes to station 3068 and avoid wasting resources at station 4215, a more advanced real-time monitoring system can be developed. If the development of this system is a long-term plan, a special allocation operation which is in charge of transporting bikes from 4215 to 3068 every day should be introduced as soon as possible.

4.3 Rush Hour Analysis

4.3.1 Weekly Rush Hour Pattern

Analysis of bike usage changing within a week time is very important for more reasonable bike allocation strategies. Finding out weekly rush hour pattern is a perfect

way to implement this analysis.

This study firstly takes a brief look at general rush hour which contains information of all trips in 2017 and 2018 from all regions. The heatmap in Figure 10 clearly shows that during workdays (Monday to Friday), the rush hours of using bike are 8:30 to 9:00 and 17:00 to 18:00. These times are the peak of going to and leaving from office. Therefore, it's highly possible that for most people in general, the main application of bike share during workdays in LA is to cover the distance between home and office/school.

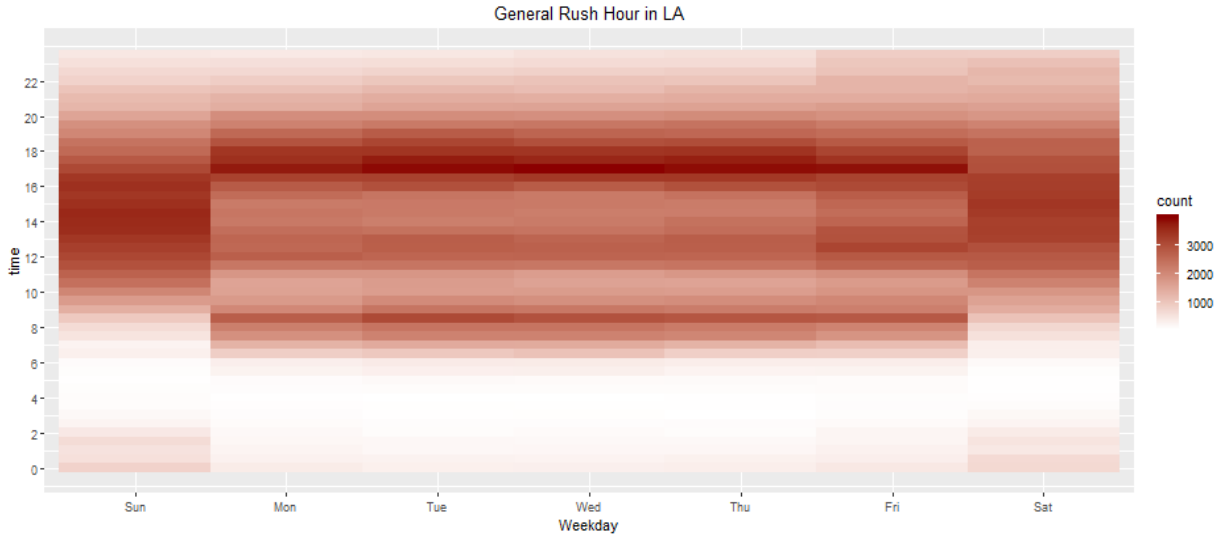

Figure 10. General Rush Hour

However, story is quite different at weekends. On both Saturday and Sunday, the rush hours are typically from 12:00 to 17:00, with very few people using bikes during regular morning peak time. This situation is well-founded since after five days' hard work, people tend to take a rest at weekends and get up late. But why do they use bikes in the afternoon? To answer this question, it's necessary to compare the trip duration at weekends to that at workdays.

According to Figure 11 the durations of workdays' bike use are more centered at a short time around 8 minutes while at weekends the distribution is relatively more evenly and the average trip duration is longer. The reason behind the scene could be that at workdays, people are in such a great hurry to their destinations that they tend to try their best to shorten the duration of the trip. At weekends, however, there is no need to rush against the clock. After lunch, they may hang around on a Metro Bike just for relaxation, exercise and fresh air without specific destination. The durations are thus longer at weekends.

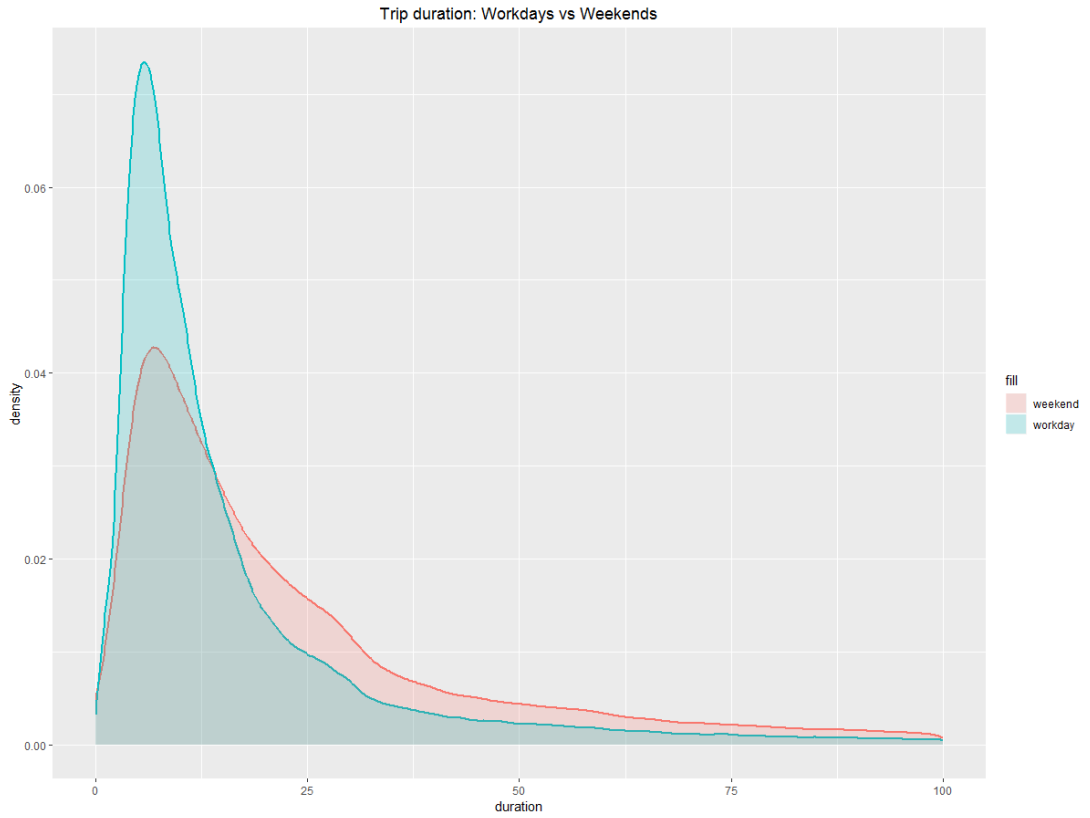

Figure 11. Trip Duration Comparison: Workdays vs Weekends

This pattern is very interesting. But is this still the case now? From the comparison of rush hour and trip duration between 2017 and 2018 in Figure 12, although the absolute count of trips is increasing, patterns are quite uniform, which means people in LA have been used to Metro Bike and they have developed a stable habit of using Metro Bikes. This is uplifting because this pattern can be used for the years to follow.

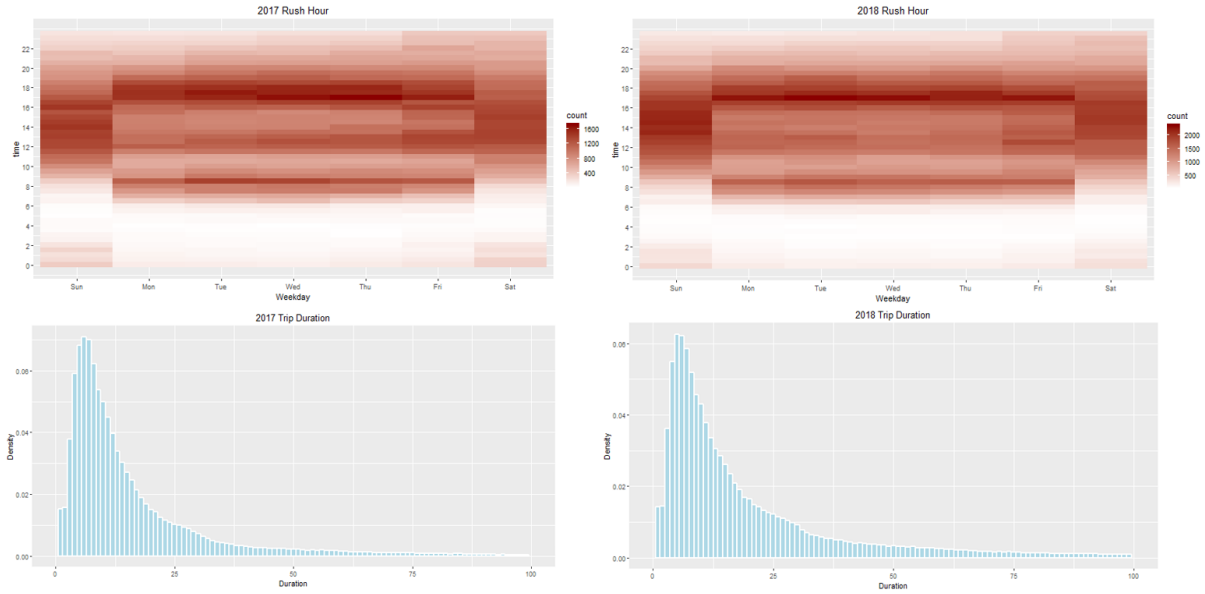

Figure 12. Rush Hour/Trip Duration Comparison: 2017 vs 2018

4.3.2 Regional Pattern

As mentioned before, in this study, greater LA area is divided into four regions based on Metro Bike use: DTLA, Pasadena, Port LA and Santa Monica. It's of great interest to see whether people in different regions have different habits of using Metro Bikes. This analysis is very important for optimizing geographical allocation of bike resources. Based on Figure 13, the majority of bike use happen in downtown which is a cluster of apartments and offices. Therefore, the pattern is nearly identical to general case. Pasadena is also similar except for the population base. However, when it comes to Santa Monica and Port LA, the pattern is quite different. Very few people have time to

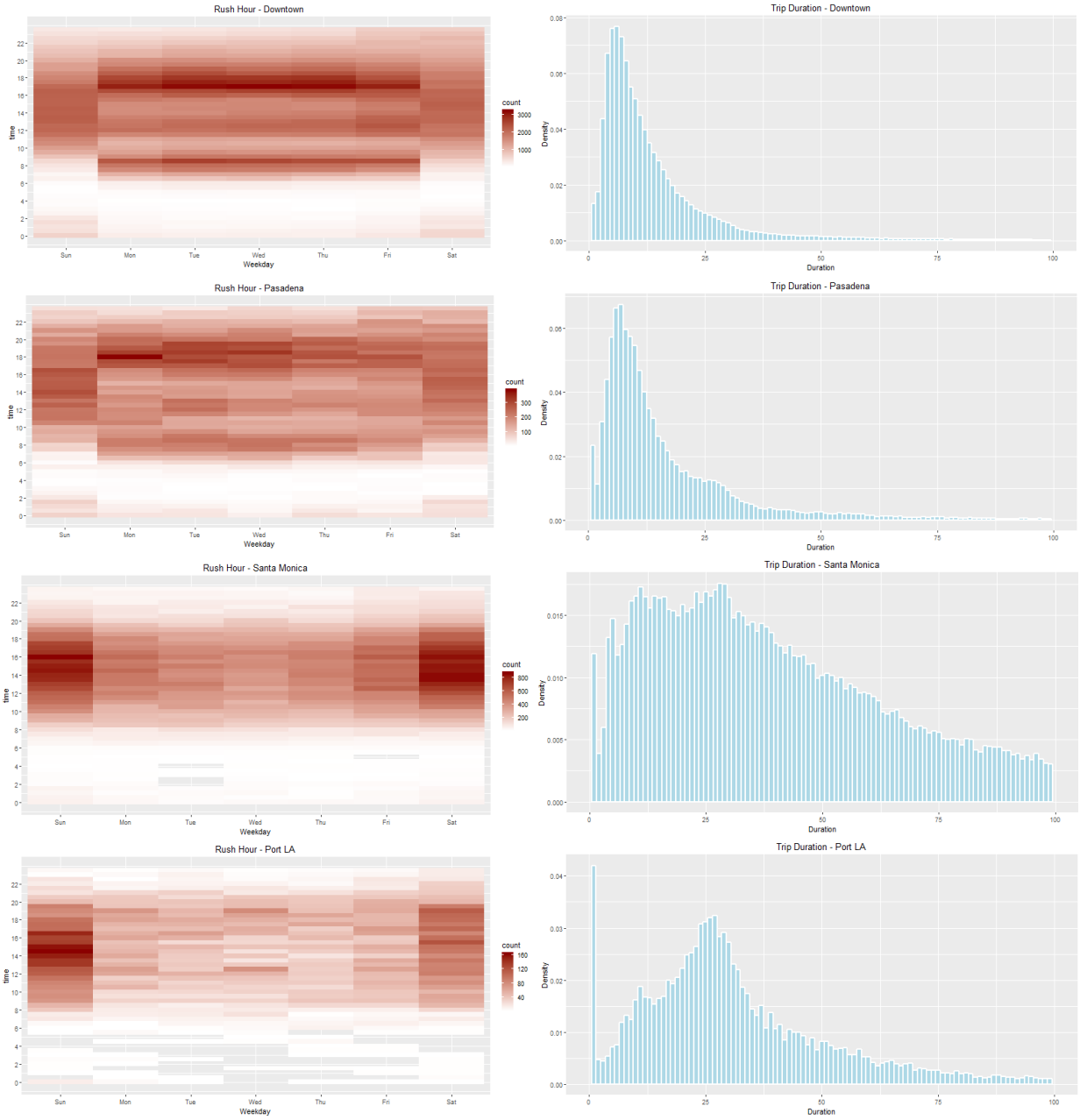

Figure 13. Rush Hour/Trip Duration Comparison among Regions

take a bike trip by the sea at workdays, because most of them don't work or live here. But at weekends, they love to do this, for exercise and relaxation. So, as tourist spots, these two regions become hot spots only at weekends. Compared to Port LA, Santa Monica is more popular among the riders since it's more famous and closer to downtown. The average duration of each trip at Santa Monica is much longer than any other regions. Many people use the Metro Bike for over an hour, which is rare in downtown and Pasadena. This could be a strong evidence of the fact that at weekends, riders aren't in a hurry and they may not even have a destination. Thus, it's likely that they just drive to the seaside, park the car and take a bike ride. If this conjecture is true, then these bunch of people must take a round trip, since they need to go back to their car. To verify this conjecture, Figure 14 is made to compare the proportion of round trips between seaside (Santa Monica and Port) and city (Downtown and Pasadena).

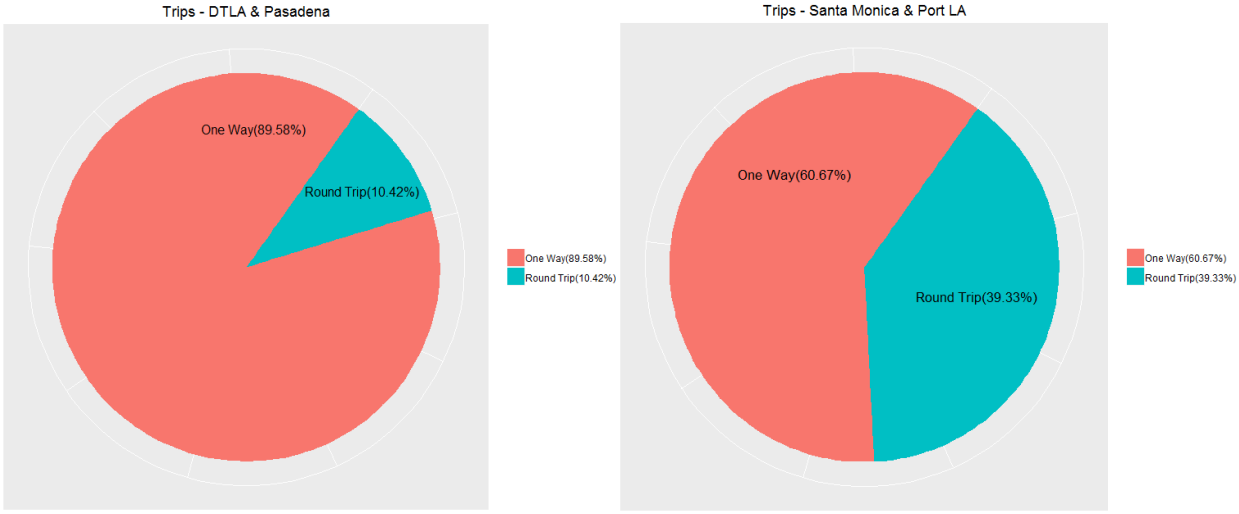

Figure 14. Trip Type Proportion Comparison: City vs Seaside

The result shows a greater proportion of round-trip takers at Santa Monica and Port LA, and therefore corroborates the conjecture.

4.4 Pass Type Analysis

Metro Bike now has the 4 types of passes:

Figure 15. Pass Types of Metro Bike

From 2017 to 2018, with the rapid development of Metro Bike, the proportion of pass type used has changed a lot. In order to capture the trend and invest more effort in the most promising pass type, the analysis on pass type is imperative.

4.4.1 Pass Type Temporal Change

Firstly, what's stable and what has changed from 2017 to 2018? According to Figure 16, Monthly Pass has always been the most popular one, since it's the most economical choice. In addition, with the total number of passholder increasing in 2018, Walk-up Pass has a significant growth. But the number of Flex Pass decreased, which means people began to find it not economical to use it and the riding frequency is not high enough. Therefore, it's very clear that Flex Pass is becoming less and less appealing to users. The price of Flex Pass should be adjusted, or we could even cancel this Pass.

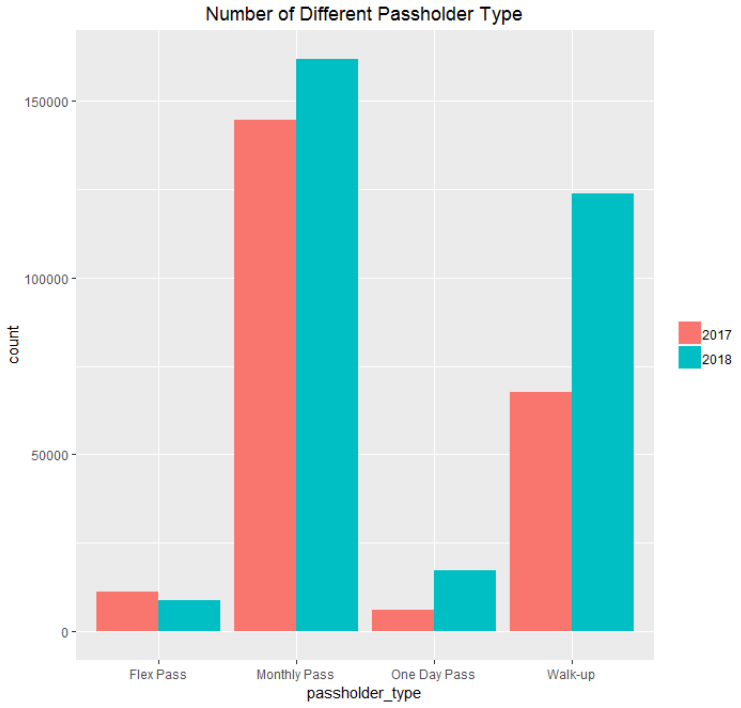

Figure 16. Number of different Pass Type

4.4.2 Pass Type Regional Discrepancy

Based on the difference of pattern, our regions are classified into two groups: city(DTLA and Pasadena) and seaside(Santa Monica and Port LA). This part focuses on finding the discrepancy of pass type distribution between city and seaside. Tog begin with, DTLA is the biggest market for Metro Bike. Here, office workers mainly take short trips(less than 30 minutes per trip) to commute from home and office, so

calculations can be made for the total bike cost for a year to decide which pass is the most economical one for these people. Assume they take three short-rides every day, each of which are less than 30 minutes. Five days a week and four weeks a month, the number of working day is approximately 240 days. Here are costs for each pass:

Pass Type	Annual Costs
Pay-Per-Ride	$1260
One Day Pass	$1200
Monthly Pass	$204
Flex Pass	$150

Table 1. Pass Types and Annual Costs

There is no doubt that for these people, Flex and Monthly pass are better choices. But what is the current situation? Here, Figure 17 shows the pass type distribution of short trips(<=30 minutes) in DTLA as well as all trips at seaside.

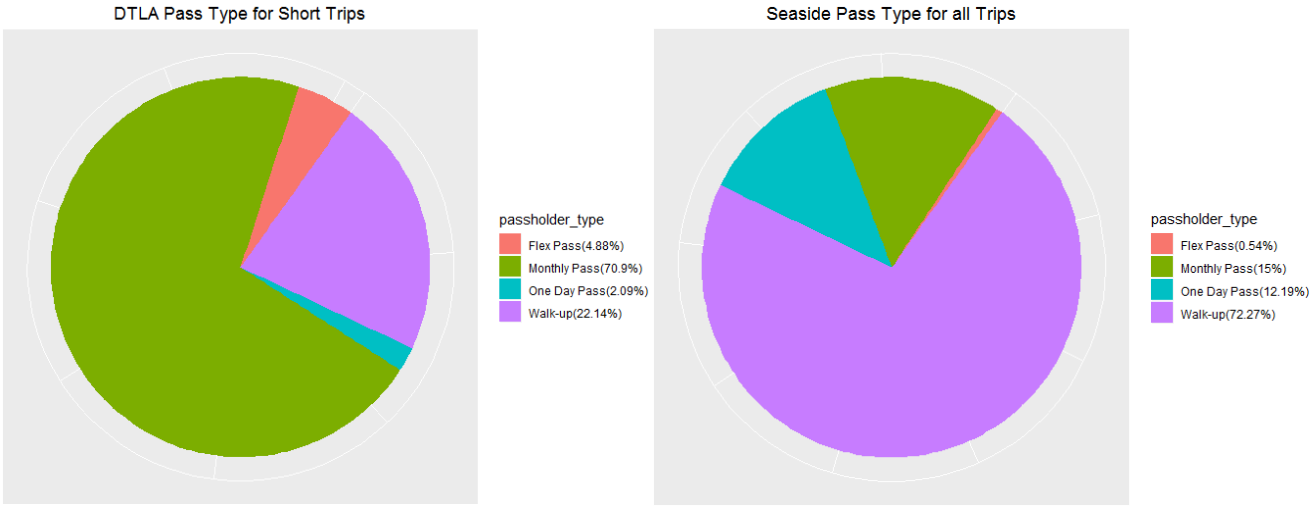

Figure 17. Pass Type Proportion Comparison: City vs Seaside

In DTLA, over 70% people use Monthly Pass. Most of these people are office workers and they know exactly how to pinch pennies. Therefore, Monthly Pass should still be the main source of revenue for years to come. To take advantage of the current situation, two aspects of actions can be taken. Firstly, Metro Bike should spare no efforts to keep the loyalty of this group. Flex Pass can be merged into Monthly Pass and the new pass type can be named “Business Pass” so that more people would notice the advantage of the pass. Since 22.14% people still choose to pay by ride, they may not be aware that they themselves are the target customer of this pass type. For these people, perhaps they

don’t belong to the group of office workers, or they don’t know how to save bike expense and thus need more guidance. For this new “Business Pass”, advertisement investment should be made to both keep old users and attract new users. Also, bonuses like breakfast coupons would be included in the new pass to cater to the need of early birds.

For riders at seaside, they choose to pay per ride. As mentioned before, these riders are relaxing here at weekends and it makes sense that they don’t care about long-term pass, or they don’t have a good choice now. Therefore, it’s the responsibility of Metro Bike to do a survey about a new pass type for exercisers. Then based on the result, new pass type “Seaside Pass” could be introduced next year with discount on weekend users and round trip riders. Also, smarter bikes equipped with fitness device(like heart monitor) should be used.

4.5 Bike Type Comparison

Since 2018Q4, electric bikes have been put into market. As the trail run of new bikes, the number of electric bikes was much fewer than standard bikes. During the trial period, great potential as well as challenges of the new type of bike, are shown.

4.5.1 General Situation of Electric Bikes

In 2018Q4, 96.5% trips are still covered by traditional standard bikes. Though the number of usages varied significantly, the average riding times of standard bikes was continuously decreasing during 2018Q4 while electric bikes are being used more and

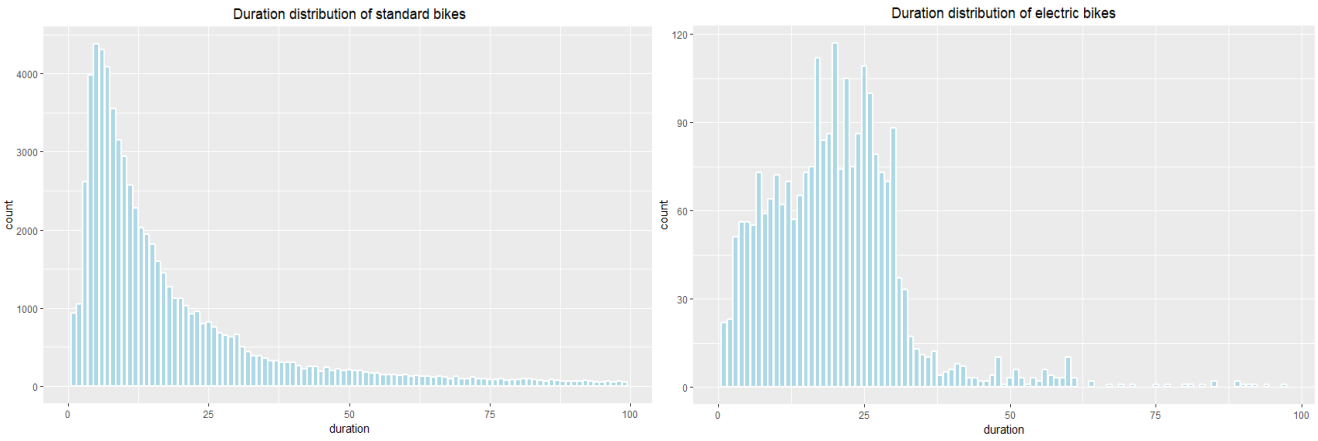

Figure 18. Duration Comparison: Standard vs Electric

more frequently. The average riding counts of an electric bike was 314 times versus 56 times for standard bikes. Figure 18 shows the duration distribution of riding trips with each bike type. Most people ride standard bikes around 7 minutes each trip. The duration is much longer when it comes to electric bikes. Obviously, people prefer to ride a relatively long trip with an electric bike. Therefore, there is no doubt that the company should arrange more electric bikes into the market.

4.5.2 Electric Bikes' Station Supply

But which stations should electric bikes be put into? At this stage, making a deliberate plan about bike allocation is important for its future development. Figure 19 shows the quantity of electric bikes borrowed from and returned to each station.

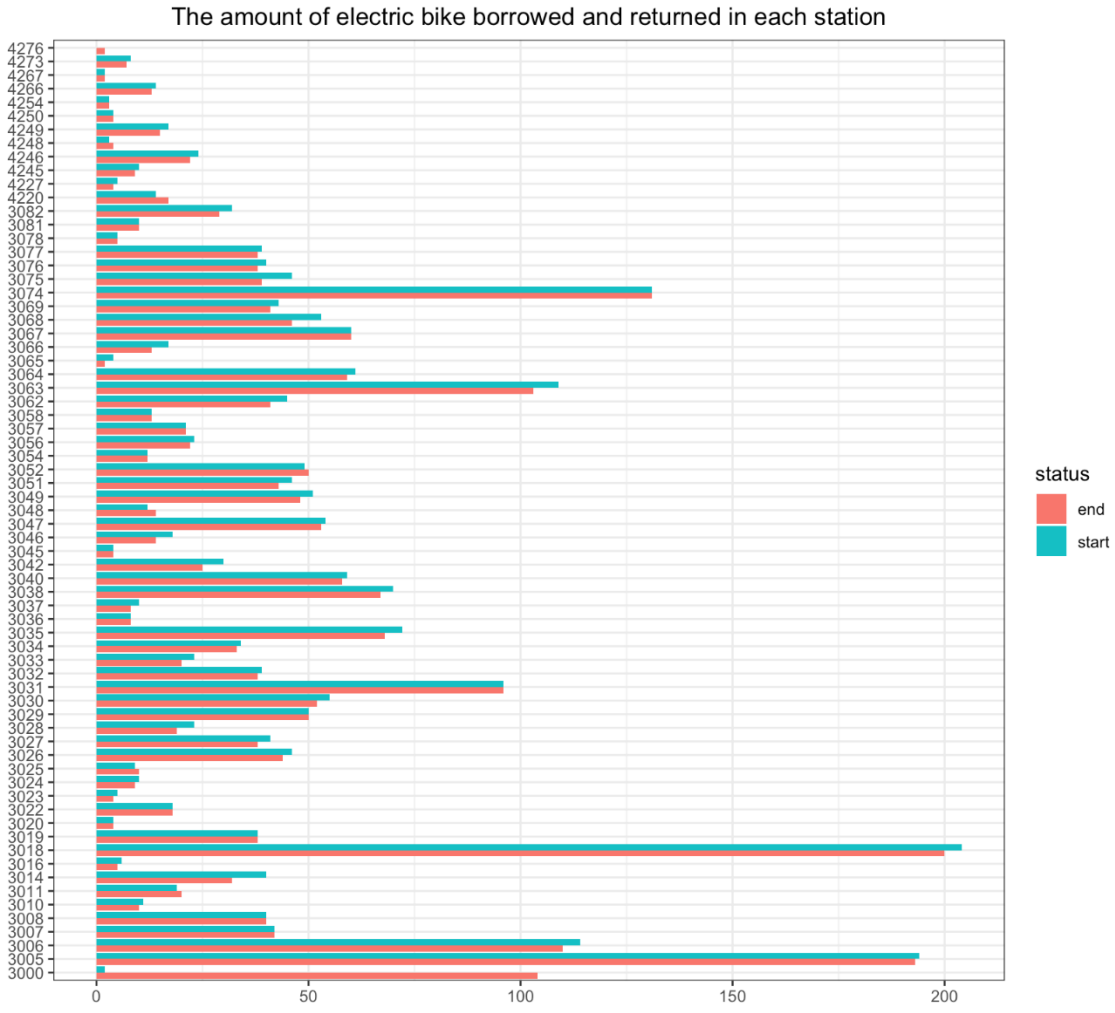

Figure 19. The Flux of Stations with Electric Bikes

Although there are some gaps between the amount of borrowing and returning within each station, these gaps are small and totally acceptable (station 3000 is the virtual station and is not included here). However, the difference between stations are significant. In general, the larger quantity of bike usage one station has, the more popular it is. The top 3 stations with highest borrowed and returned quantity are station 3018, 3005, and 3074. Since their quantity are about 2 or 3 times bigger than normal stations, it's urgent to put more electric bikes in these stations to deal with the large quantity of flow to prevent the potential of shortage of electric bikes. Since currently there are only 8 electric bikes, this number should be increased to 16 in the first quarter next year. Most of the newly added electric bikes should be allocated to station 3018, 3005, and 3074.

Also, the current price for using electric bikes is the same as using the standard ones. This is only for trial period, since the maintenance fee for electric bikes is much higher. Thus, the next step is to make a new pricing strategy for electric bikes.

5. Shiny Dashboard

The dashboard integrated an interactive map from package leaflet. It enables users to zoom in and out as much as they want. It is an intuitive way to show how the station usage varies among regions during different time period.

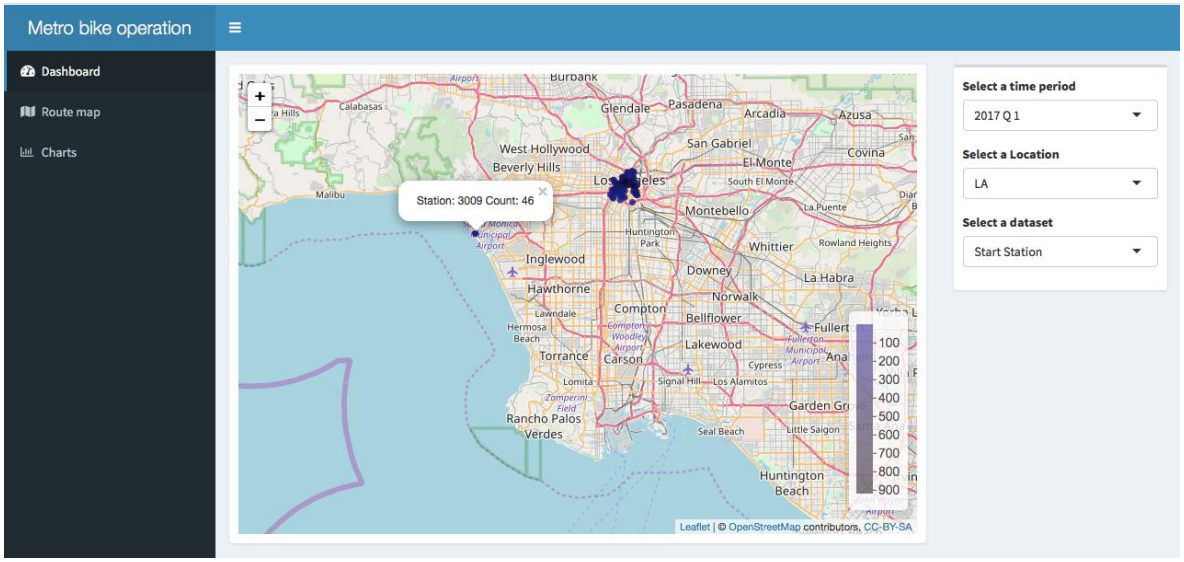

Figure 20. Shiny Dashboard Interface

Main Functions of the dashboard:

- **Station usage info integrated with Interactive map**
  The map shows each station's usage by the time period user selected, both yearly or quarterly. The darker the point is, the heavier station usage. Further, users can select a specific region to see the detailed station usage.
- **Popup message**
  When users click on a point on the map, it will show a popup message indicating the precise Station ID and the count of usage. It's quite convenient to locate a busy station and check the precise information.
- **Location shift for details**
  Users can select a specific region to see the detailed station usage situation.
- **Start/End station selection**
  Users can select datasets to display the information of start or end station.

6. Conclusions and Recommendations

6.1 Conclusions

- In general, Metro Bike's business, after a period of steady growth and expansion in 2018, is expected to enter a plateau in 2019.
- The Metro bike system usage shows strong seasonality: the bike usage reaches the highest point at warmer seasons (Q2 and Q3) and then drops dramatically in Q4.
- People in Downtown LA and Santa Monica are the top 2 groups of contributors to Metro Bike's business. Different from Downtown LA's stability, Santa Monica has a great momentum of growth.
- From 2017 to 2018, Downtown LA's shortage of bikes is becoming much more severe while Santa Monica is having more redundant bikes. Station 3068 in Downtown and 4215 at Santa Monica are microcosm of this situation.
- Downtown LA has a stable large group of office workers who will remain the main source of revenues for the following years due to the nine-to-five timetable during workdays. Though most of them choose to buy Monthly Pass, a lot of people still have little idea about the advantage of long-term pass.
- During weekends, people tend to leave the city and gather at seaside like Santa Monica and Port LA in the afternoon, riding a bike for exercise and relaxation. Many of them don't care about time, take a round trip and pay per ride, which is not the most economical pass type.

- Electric bike is so popular among the users that in Downtown LA stations it's out of supply. Also, the duration of electric bike trip is much longer than traditional bike, which indicates that people enjoy riding electric bikes to cover long distance.

6.2 Recommendations

- To get rid of the imbalance of bike allocations between Downtown LA and Santa Monica, specifically, to supplement bikes to stations like 3068 and avoid wasting resources at stations like 4215, a more advanced real-time monitoring system should be developed. If the development of this system is a long-term plan, a special allocation operation which is in charge of transporting bikes from Santa Monica to Downtown on workdays should be introduced as soon as possible.

- For old users of Monthly Pass in Downtown, Metro bike should spare no efforts to keep the loyalty of this group. At the same time, for people working in Downtown but still have little idea about how to choose the most economical pass type, they need more guidance. Therefore, Flex Pass can be merged into Monthly Pass and the new pass type can be named "Business Pass" so that more people would notice the advantage of the pass. For this new "Business Pass", advertisement investment should be made to both keep old users and attract new users. Also, bonuses like breakfast coupons would be included in the new pass to cater to the need of early birds.

- For riders at seaside, Metro bike should do a survey about a new pass type for exercisers. Then based on the result, new pass type "Seaside Pass" could be introduced next year with discount on weekend users and round trip riders. Also, smarter bikes equipped with fitness device(like heart monitor) should be used.

- Metro Bike should put more electric bikes into Downtown area, especially station 3018, 3005, and 3074. Since currently there are only 8 electric bikes, this number should be increased to 16 in the first quarter next year. Also, due to the high maintenance fee for electric bikes, the next step is to make a new pricing strategy for electric bikes.

7. Future Work

The main problem of this study is the limitation on data. Therefore, the future work focuses on collect more detailed and oriented data to bring the analysis to a higher level:

- **Do surveys to collect user's personal information such as age, income, etc.**
  With these data, it would be convenient to cluster the users by means of unsupervised learning. Then, different strategies will be applied to specific group of users to adjust to their interest.

- **Make records on users' service situation such as regular routes.**
  With these data, the company could optimize the station position and improve the user experience by setting supply stations on the busiest routes.