Final Project Report

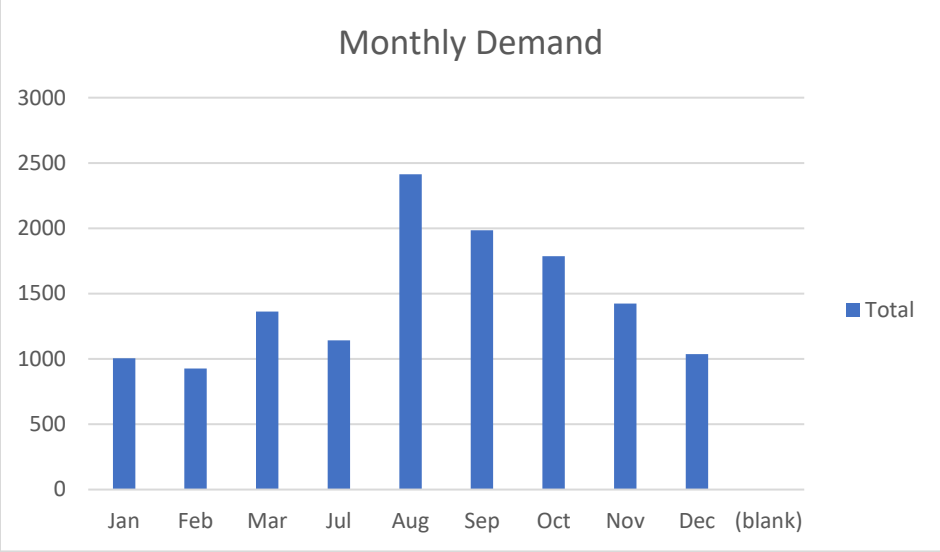
MSCI 719 - Operations Analytics

Yuchen Wang

*Note: this dataset is randomly generated, so some results may not provide practical conclusions, but the logic is correct.

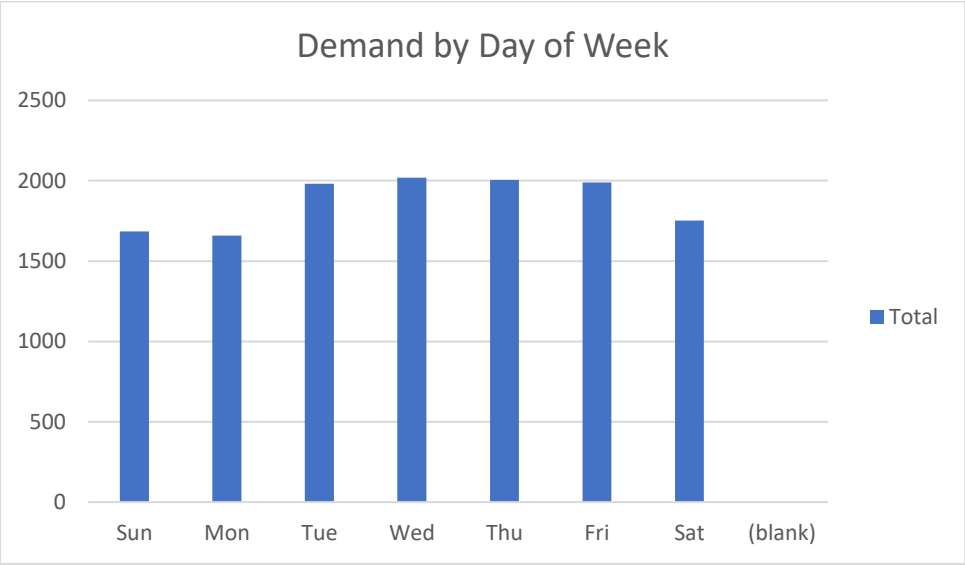
Section A

A1.1 Relationship between demand for renting bikes and time



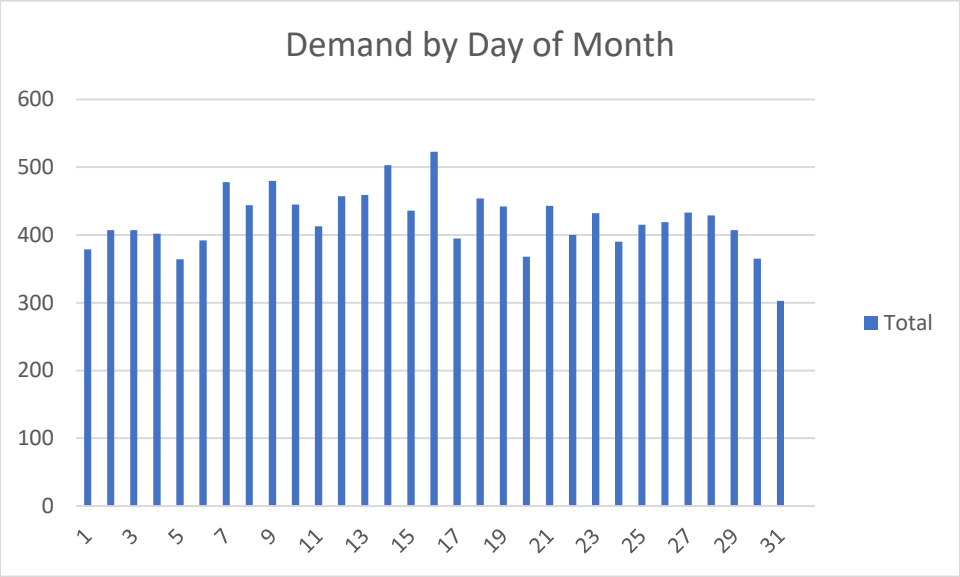
For monthly demand, we can conclude from this graph that there is a pattern of demand among different months within a year. And we can see that demand in August to November are relatively higher than other months.

This is probably because rainy days start around December to March in LA.



For demand by different days of a week, we can conclude from this graph that there is no significant relationship between DoW and demand. They are almost the same.

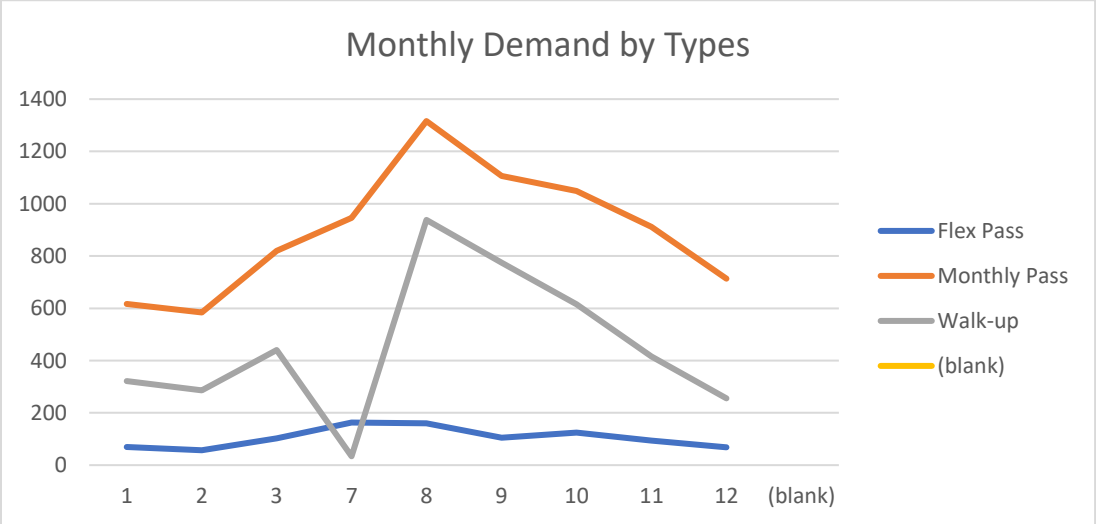
This probably is because weather conditions on different days within a week are very similar. And people rent bikes to work on weekdays and rent bikes for fun on weekends.



For demand by different days of a month, we can conclude from this graph that there is no significant relationship between day of month and demand. They are almost the same.

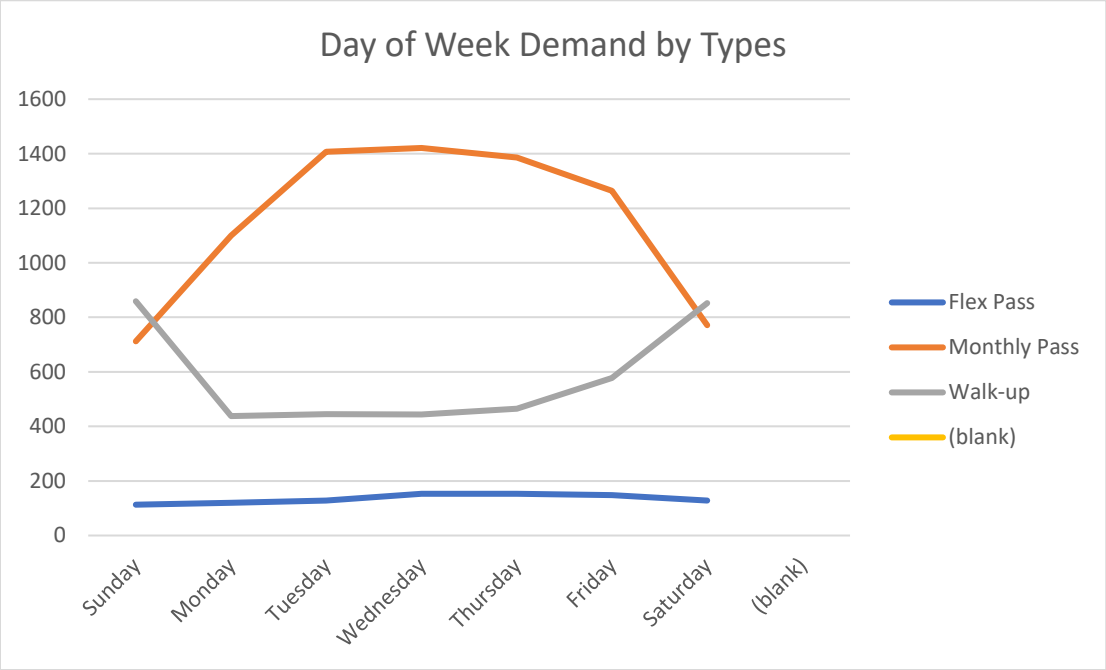
This probably is because weather conditions on different days within a month are very similar.

A1.2 Relationship between demand and time divided by different types



For monthly demand, we can see that there is a significant difference between different months for monthly pass and walk-up, while there is no significant difference between different months for flex pass.

This is probably because when people register a flex-pass (one-year pass), no matter it's rainy or not, people want to go biking to maximize the value of their membership.

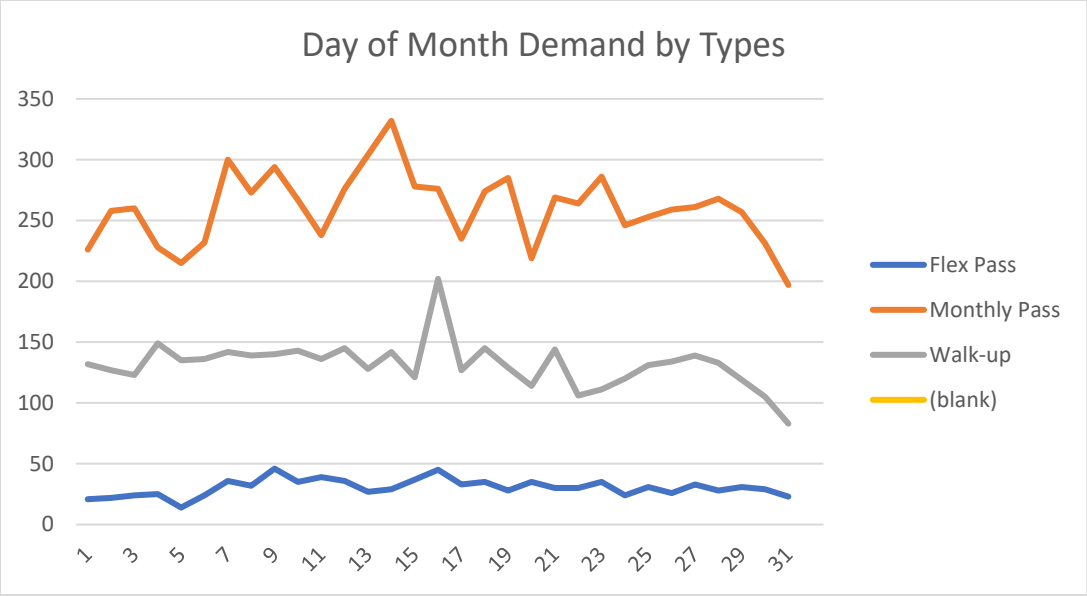


For day of week demand, we can see that there is a significant difference between different days of a week for monthly pass and walk-up, while there is no significant difference between different days for flex pass.

This is probably because when people register a flex-pass (one-year pass), no matter it's weekends or not, people want to go biking to maximize the value of their membership.

Besides, monthly pass demand of weekends are less than weekdays, while walk-up demand of weekends are more than weekdays.

This is probably because when weekends coming, people just feel like going for a biking and they buy a one-time ticket. And people ride to work buy monthly-pass since this is more valuable for them, so they often rent on weekdays.



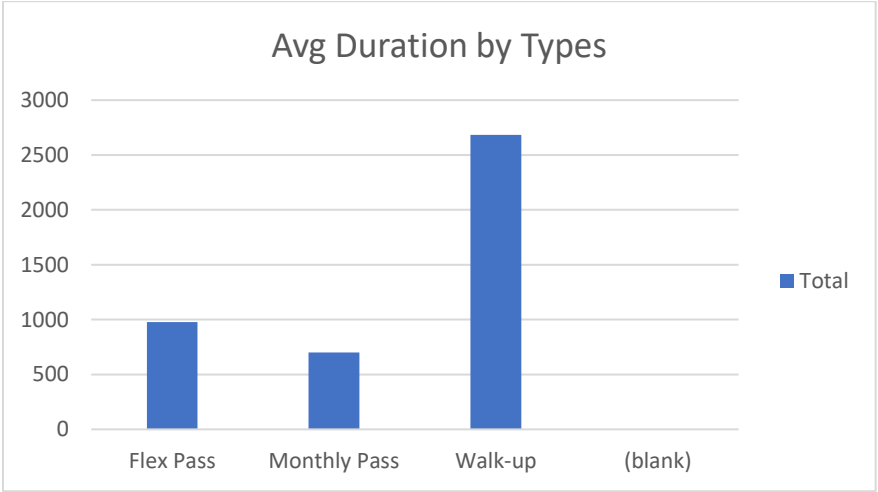
For day of month demand, all the three types show some relationship between demand and days of a month, but the difference of demand between different days is not huge.

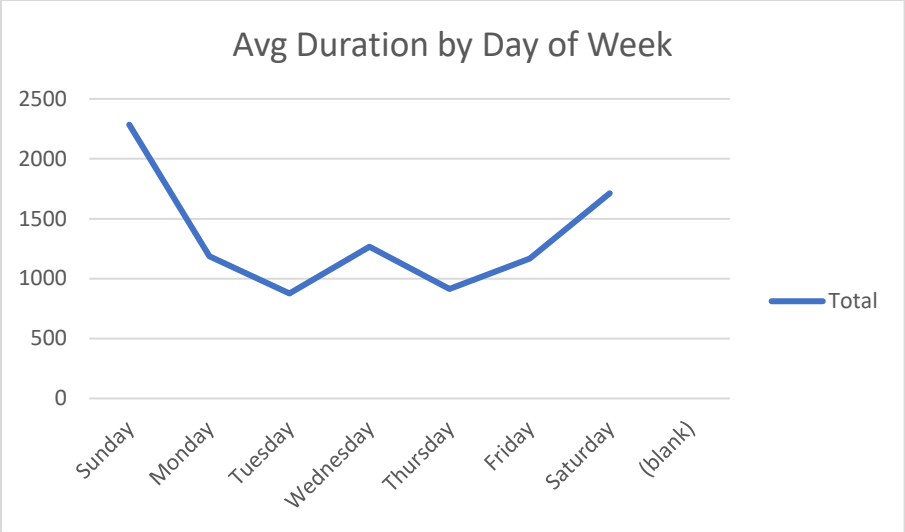
A2 Predict the trip duration of customers

Since rainy days start around December to March in LA, I did two regression for Dec-Mar and Jul-Nov separately.

[Dec-Mar]

I compared average duration for different types of tickets and for different days of week. There is a significant relationship between duration and types, also between duration and day-of-week.



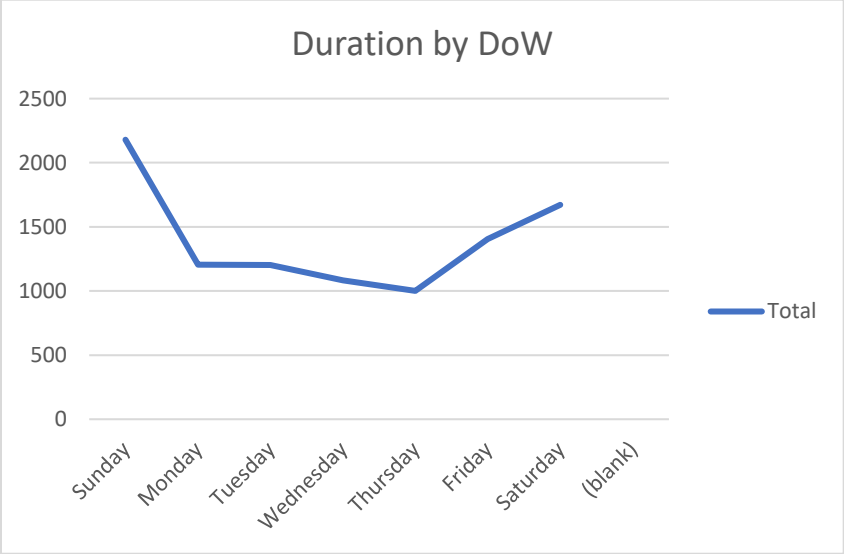


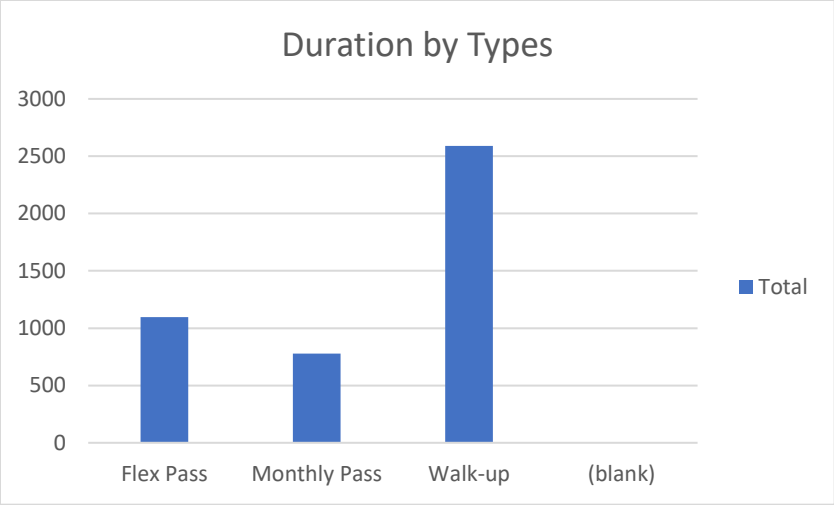
I did the regression by using weekends or not, types of tickets as dummy variables, and age. They are all significant (p-value<0.05).

	Coefficients	*P-value*
Intercept	4945.46	3.93E-45
Age	-73.88	1.57E-14
weekends?	352.22	0.02858
Flex pass	-1409.68	1.45E-06
Monthly pass	-1630.05	4.98E-24

[Jul-Nov]

Same as I did in [Dec-Mar].





The predictor variables are all significant (p-value<0.05).

	Coefficients	*P-value*
Intercept	5064.10	5.83E-96
Age	76.68	1.27E-32
weekends?	284.85	0.015482
Flex pass	1332.59	1.14E-10
Monthly pass	1543.91	4.48E-41

A3 Predict the daily demand of walk-up customers

Shown in A1, for walk-up consumers, demand in rain seasons and demand in dry seasons are significantly different. So I separated Dec-Mar and Jul-Nov here and did regression model for them respectively.

Walk-consumers' demand is different between weekdays and weekends, so I made weekdays or not as a dummy variable and an interaction variable.

Since demand on day T-7 would be very similar with demand on day T, I used T-7 demand as a predictor variable.

[Dec-Mar. Dummy]

	Coefficients
Intercept	7.04
Weekends?	6.33
T-7 Demand	0.20

[Dec-Mar. Interaction]

	Coefficients
Intercept	9.45
Weekends Demand	0.40
Weekdays Demand	0.049

[Dec-Mar. Combine]

	Coefficients
Intercept	7.76
T-7 Demand	0.12
Weekends?	4.95
Weekends Demand	0.13

The R square of this model if 0.26, slightly improved from dummy model and interaction model.

[Jul-Nov. Combine]

	Coefficients
Intercept	103.66
Weekends?	-96.58
Weekend Demand	2.42
T-7 demand	-1.85

As I saw in the output of this model, the R square of this model is very small (only 0.01), this is probably because in dry season, July to November, people go for biking no matter it's weekends or weekdays or because in dry season, people go for biking very randomly and intuitively, we can't use demand on T-7 to predict demand on day T.

Another reason for the "bad" output of this model is this dataset is not real, it's randomly generated.

Section B

1. Data Cleaning

I filtered data on March. 2017 from "Yuchen Wang. xlsx" and I calculated the total travelled distance for each trip by using the provided formula.

I used some aggregation, filtering and sorting then I got "List A" and "Top 10 bike with highest distance which is less than 18000". Bike IDs of them are 5960, 6549, 5810, 6604, 6679, 5755, 6227, 5923, 6176, 6239.

2. Regression from historical data:

I used historical data to conduct a linear regression with 2 predictor variables: "March travelled distance" and "Age of Bike" and 1 outcome variable: # of days before next maintenance.

Before doing it, I filtered out the bikes with a distance of less than 18000, since I want to make it paralleled with data of 2017.

I got the output of regression:

	Coefficients	*P-value*
Intercept	32.29865	2.1E-123
March travelled distance	-0.00032	1.2E-08
Age of Bike (months)	-0.15364	3.32E-07

All predictor variables are significant (p-value<0.05) and the R square is 0.69 showing that predictor variables explaining 47.6% variation of outcome variable.

3. Regular Maintenance

I predicted the # of days before next maintenance for top 10 bikes by using age of these bikes and the regression coefficients I calculated before.

Then I calculated "Maintenance efficiency" and "cost" for each of them.

BikeID	Age	Predicted #days	Maintain efficiency	Cost
5960	23	23	76.59%	$ 17.83
6549	15	24	80.92%	$ 17.62
5810	33	22	72.14%	$ 17.21
6604	12	25	83.04%	$ 17.08
6679	9	26	85.02%	$ 16.66
5755	45	20	66.98%	$ 16.30
6227	30	23	75.01%	$ 15.97
5923	23	24	79.59%	$ 15.05
6176	23	24	79.62%	$ 15.02
6239	29	23	76.75%	$ 14.84

4. Full Maintenance

All maintenance efficiency and cost for each of top 10 bikes are same: 100% and $28 respectively.

BikeID	maintain efficiency	Cost
5960	100%	28
6549	100%	28
5810	100%	28
6604	100%	28
6679	100%	28
5755	100%	28
6227	100%	28
5923	100%	28
6176	100%	28
6239	100%	28

5. Compare and Solver-LP

Given a budget of $230, we want to maximize the total average maintenance efficiency while total cost is less than $230.

I did an integer programming by using Solver and got the results below.

Decision Variables		
	regular	full
5960	0	1
6549	0	1
5810	0	1
6604	1	0
6679	1	0
5755	0	1
6227	0	1
5923	1	0
6176	1	0
6239	0.138237389	0.86176261

As shown in the table, for 6239, its value is not integer, given a budget limitation, I decided to make it a regular maintenance.

Results	Regular	Full	Total Avg Maintain Eff	Total cost
5960	0	1	90.40%	$ 218.66
6549	0	1		
5810	0	1		

6604	1	0		
6679	1	0		
5755	0	1		
6227	0	1		
5923	1	0		
6176	1	0		
6239	1	0		

By doing integer programming, I got the results as shown in this table, and the total average maintenance efficiency is 90.40% with a cost of $ 218.66 which is within the given budget.