



# BABD

INTERNATIONAL MASTER IN BUSINESS ANALYTICS AND BIG DATA

## Bike Sharing Analysis – San Francisco Bay Area

Federico Raimondi Cominesi



POLITECNICO DI MILANO  
GRADUATE SCHOOL  
OF BUSINESS



POLITECNICO DI MILANO

In partnership with



Executive Education  
Ranking 2017



European Business Schools  
Ranking 2017



# Agenda

1

ANALYSIS CONTEXT & SCOPE

2

THE DATA SCIENCE PROCESS

3

HOW DOES IT WORK?

4

BUSINESS QUESTIONS

5

THE DATASETS

6

DATA CLEANING & DATA EXPLORATION

7

IN DEPTH ANALYSIS & MODEL BUILDING

8

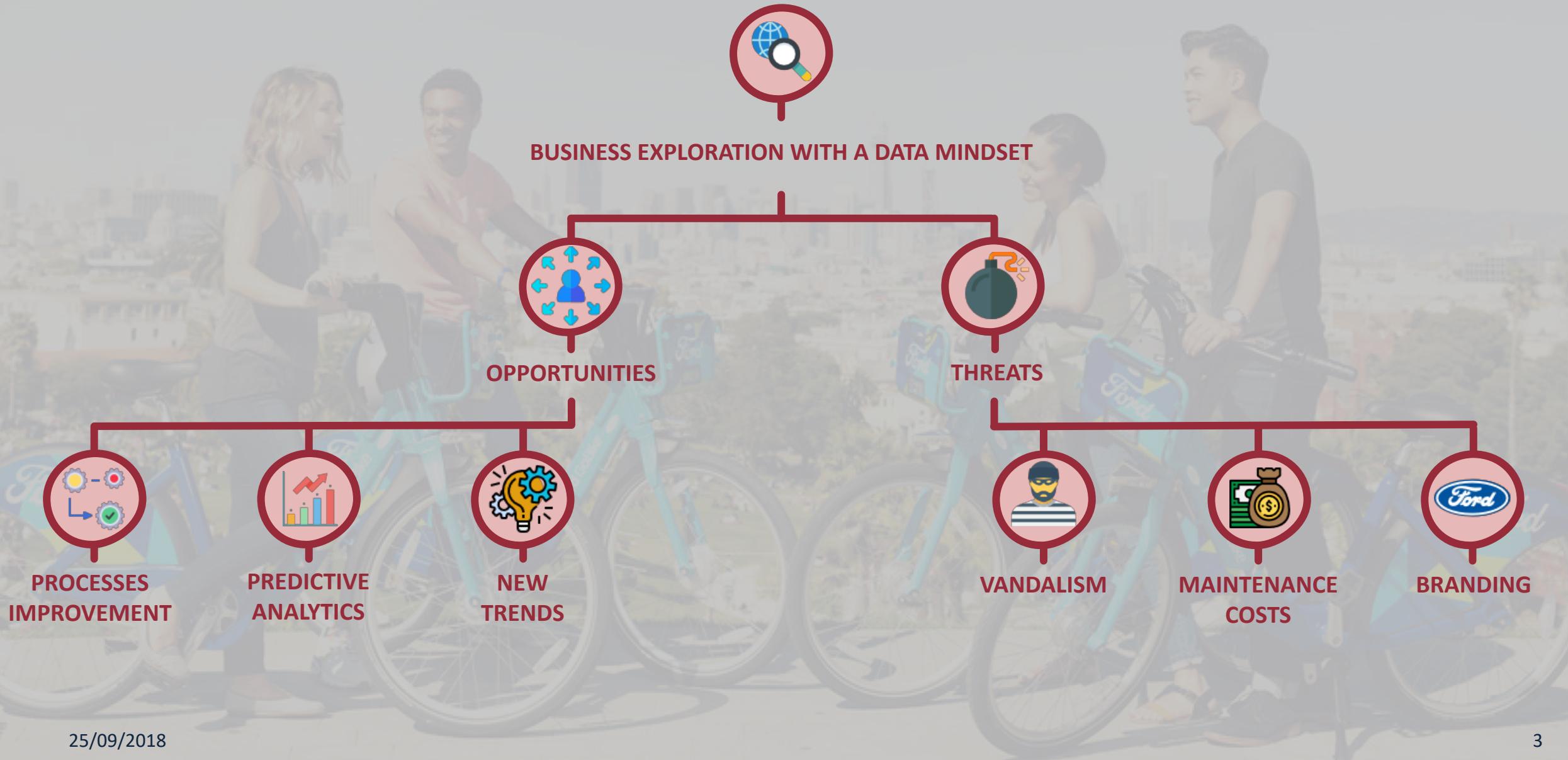
BUSINESS ACTIONS

9

NEW TRENDS AND OPPORTUNITIES

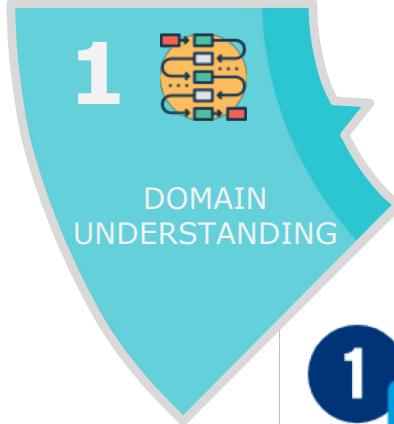


# Analysis Context & Scope



# The Data Science Process





## How does it work?



Join



Unlock



Ride



Return

### USERS:



*Subscriber:* Frequent user with a subscription plan



*Customer:* Occasional user who buys single rides

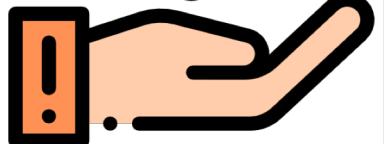
# Business Questions



CAN WE SEGMENT OUR CUSTOMERS  
TO OFFER BETTER PROMOTIONS?



HOW CAN WE FACE THE PROBLEM  
OF VANDALISM AND STOLEN BIKES?



CAN WE PREDICT HOW MANY TRIPS WILL  
BE DONE IN A DAY IN DIFFERENT AREAS?



IS OUR BUSINESS STILL GROWING?  
ARE THERE NEW OPPORTUNITIES THAT WE COULD CATCH?





### DATA COLLECTION

# 3

# The Datasets



[Ford GoBike Bike-Share Data](#)



[Bay Area Bike-Share Data](#)

#	id	# duration	# start_date	A start_station_name	# start_station_id	# end_date	A end_station_name	# end_station_id	# bike_id
1	4576	63	8/29/2013 14:13	South Van Ness at Market	66	8/29/2013 14:14	South Van Ness at Market	66	520
2	4607	70	8/29/2013 14:42	San Jose City Hall	10	8/29/2013 14:43	San Jose City Hall	10	661
3	4130	71	8/29/2013 10:16	Mountain View City Hall	27	8/29/2013 10:17	Mountain View City Hall	27	48
4	4251	77	8/29/2013 11:29	San Jose City Hall	10	8/29/2013 11:30	San Jose City Hall	10	26
5	4299	83	8/29/2013 12:02	South Van Ness at Market	66	8/29/2013 12:04	Market at 10th	67	319
6	4927	103	8/29/2013 18:54	Golden Gate at Polk	59	8/29/2013 18:56	Golden Gate at Polk	59	527
7	4500	109	8/29/2013 13:25	Santa Clara at Almaden	4	8/29/2013 13:27	Adobe on Almaden	5	679

4

DATA  
CLEANING &  
EXPLORATION

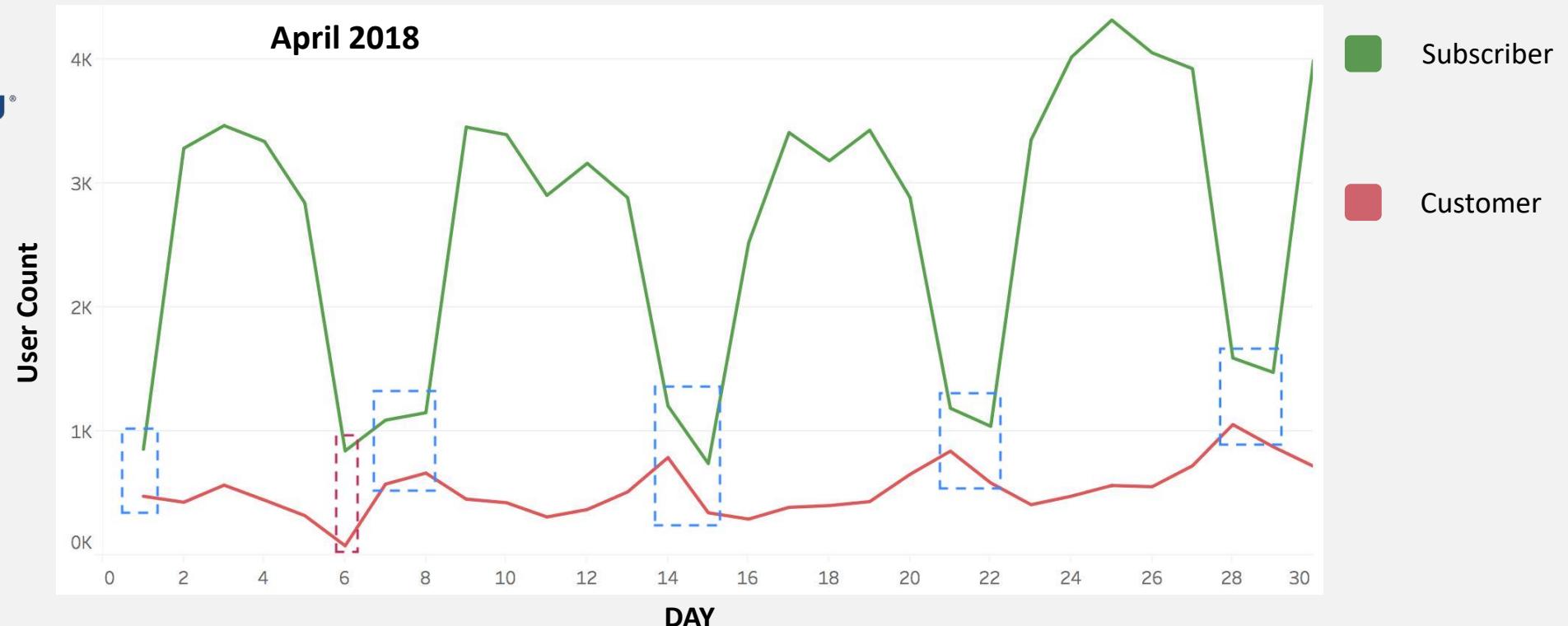
# Data Cleaning



Fixing date format in all files

Weekend  
Outlier

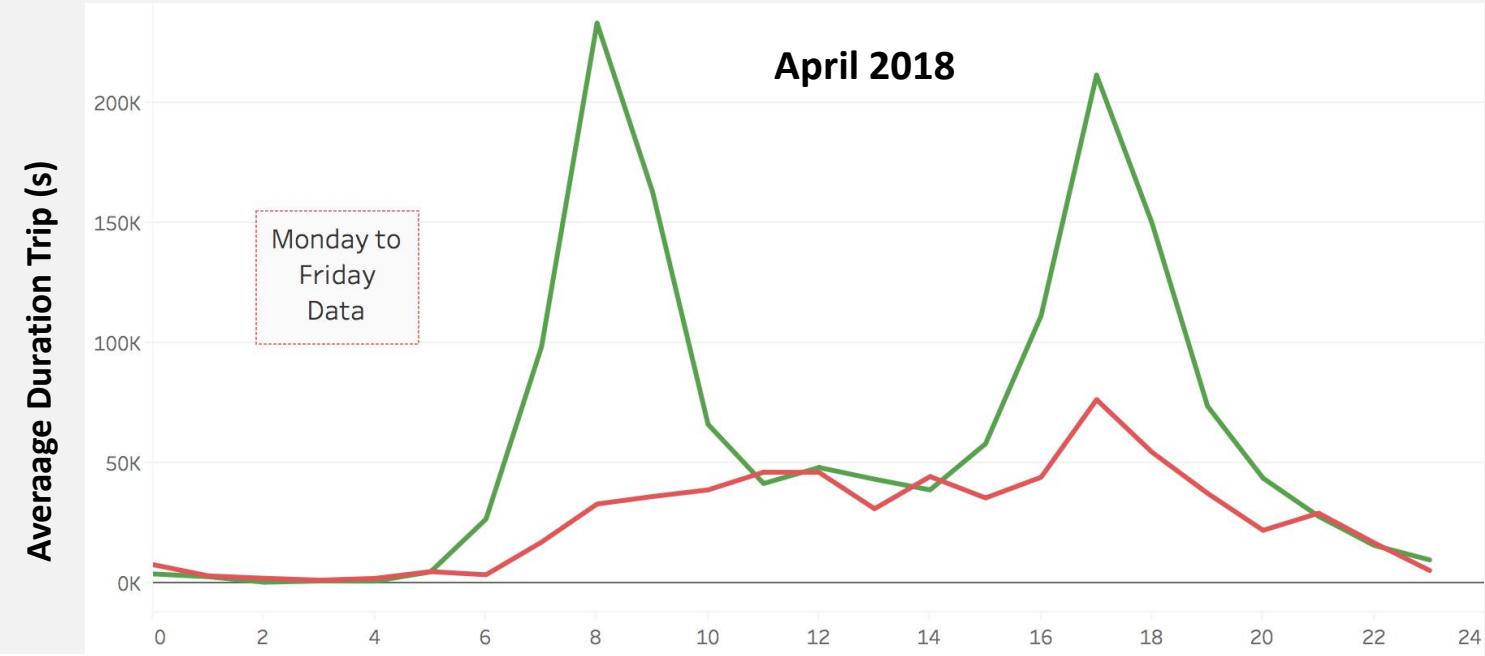
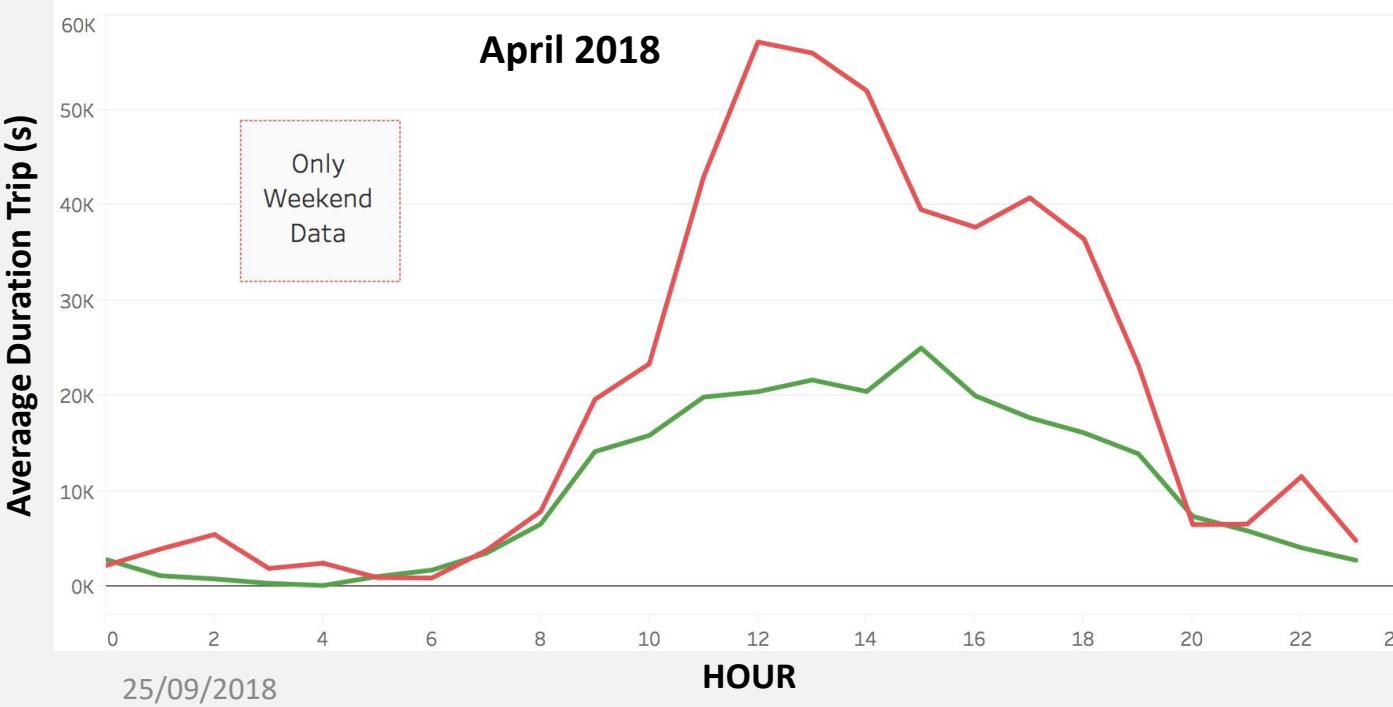
## Data Exploration



## 4

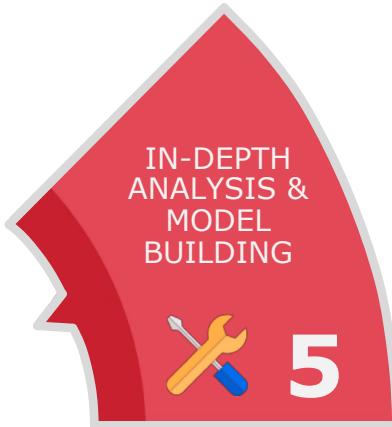
DATA  
CLEANING &  
EXPLORATION

# Data Exploration



User Type	Median Duration Trip (s)
Subscriber	536
Customer	1005.5

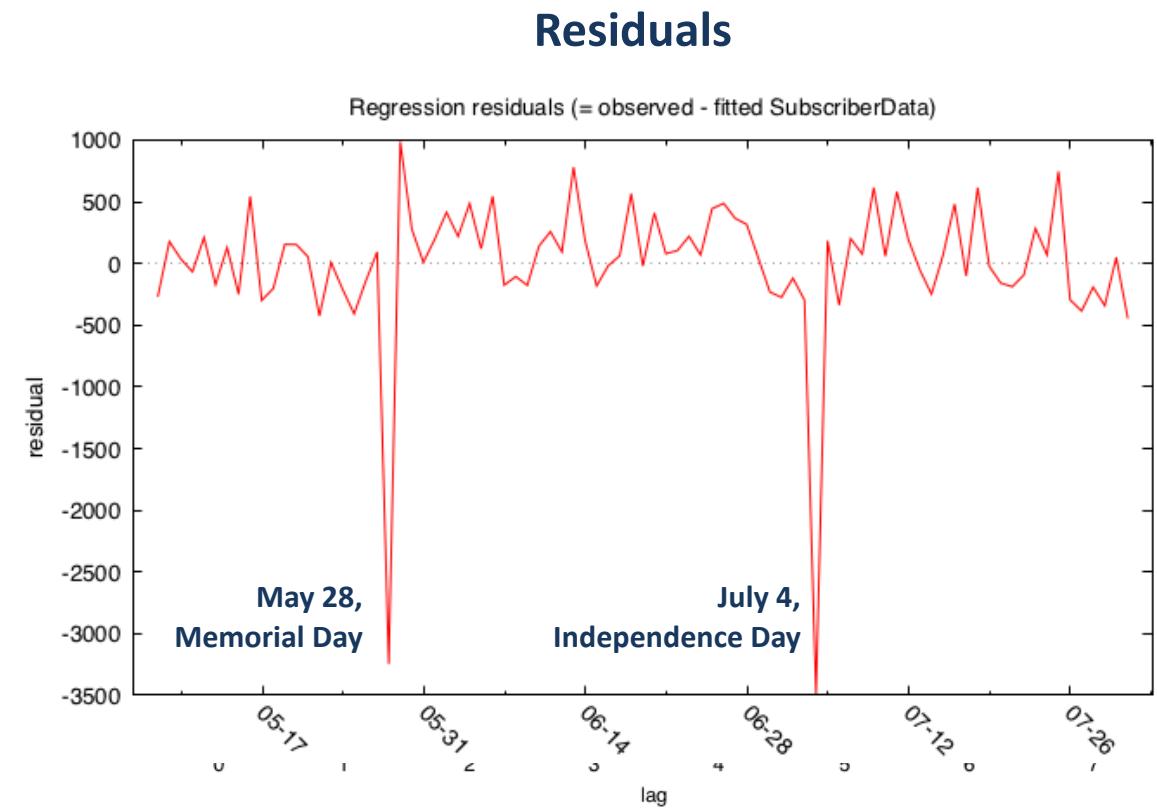
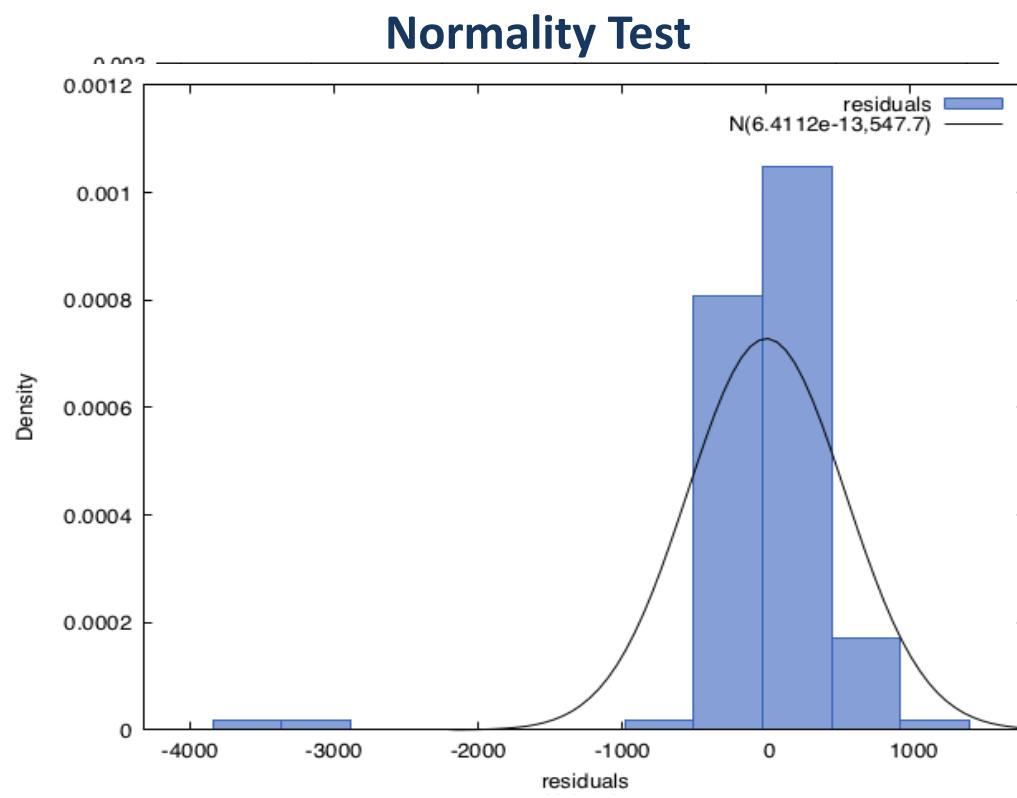
# In Depth Analysis & Model Building



GRETl SCRIPT



ARMA (1,0)





# In Depth Analysis & Model Building

Metric	Formula	Subscriber	Customer
<i>Mean Error</i>	$\frac{\sum_{i=1}^n (x_i - \hat{x}_i)}{n}$	-69.99	-119.24
<i>Mean Absolute Error</i>	$\frac{\sum_{i=1}^n  x_i - \hat{x}_i }{n}$	112.79	120.73
<i>Root Mean Squared Error</i>	$\sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}}$	134.21	154.06
<i>Theil's U</i>	$\frac{RMSE}{RMSE_0}$	0.035	0.628

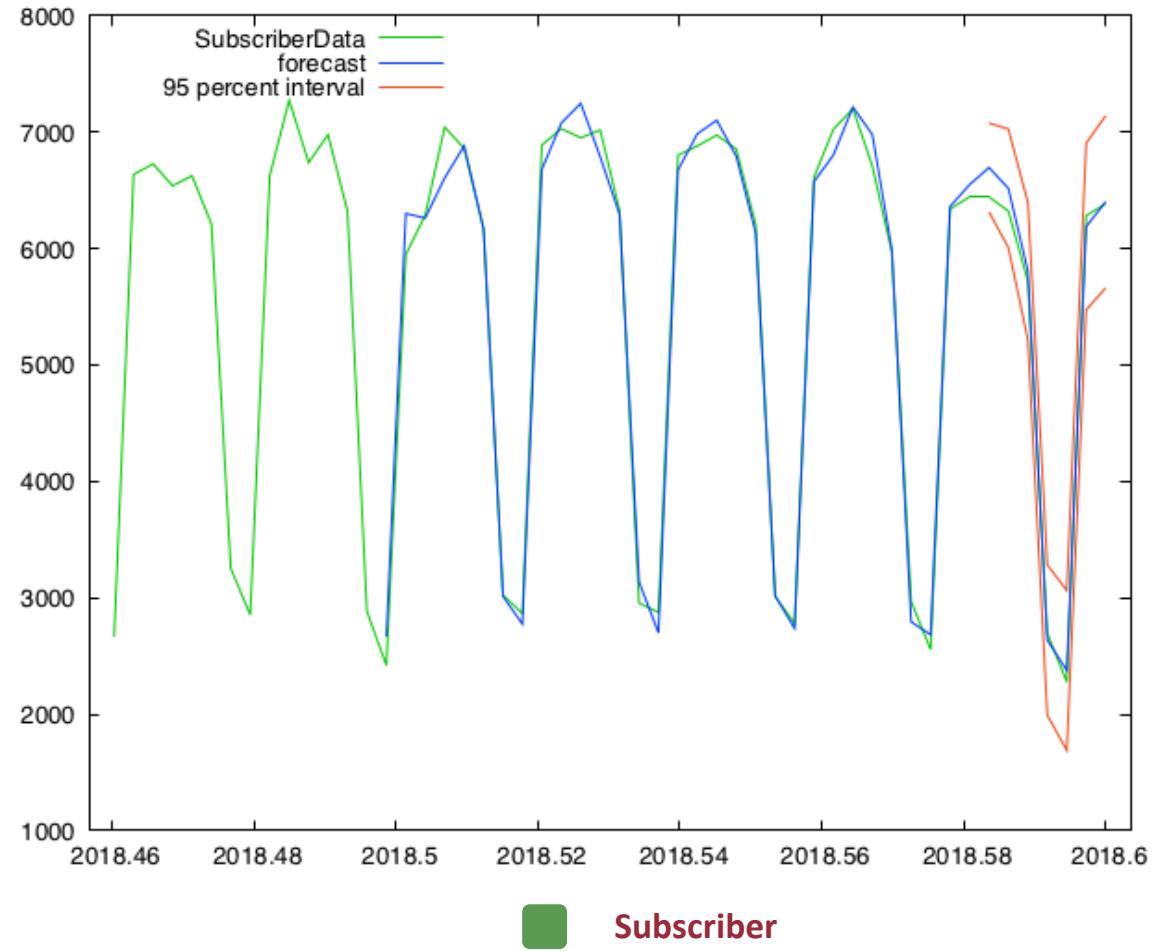
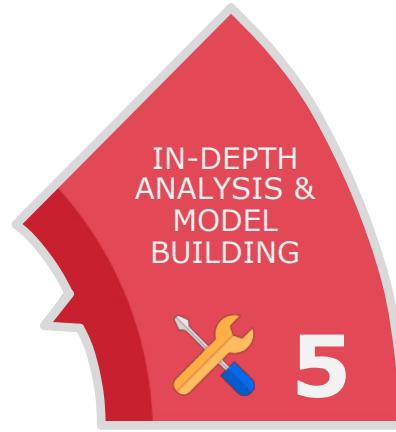
Where:

$\hat{x}_i$  : predicted value

$n$  : number of observations

$RMSE_0$ : this is the  $RMSE$  of a naïve model

# Predictions





6

BUSINESS  
ACTIONS

## Business Actions

**Rebalancing Schedule:**  
10 a.m. to 3 p.m.



**Maintenance Schedule:** 5-6 a.m. right before the pick

**Dashboard:** Real-time monitoring

**Customized Promotions:**  
Possible integration with a churn analysis



# Dashboard Demo

## Ford GoBike

San Francisco Stations

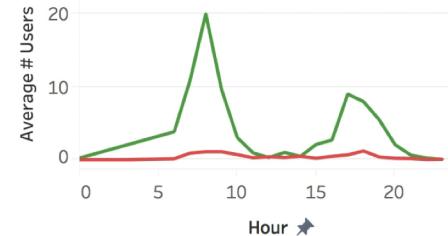
30 < > ○ ◀ ▶ □ △ ▨ ▩ Show history

User Type

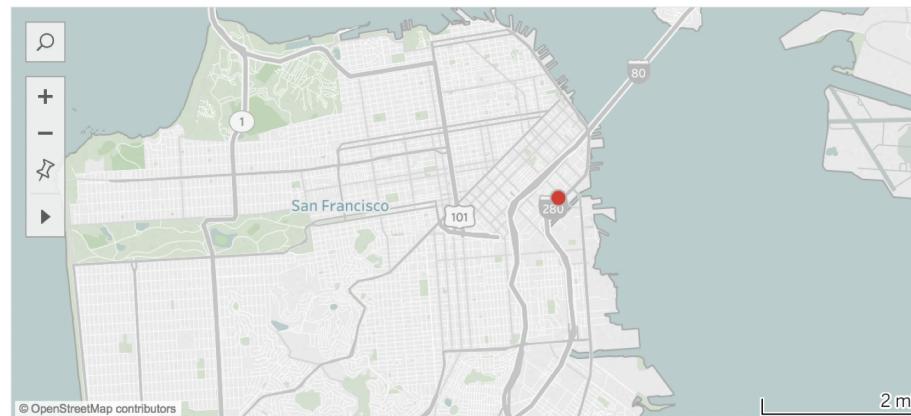
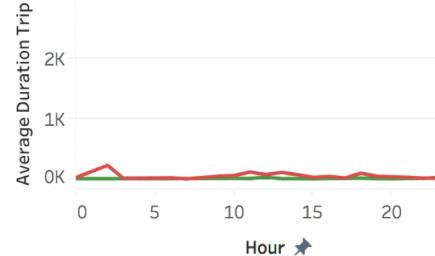
Customer

Subscriber

Average # Users

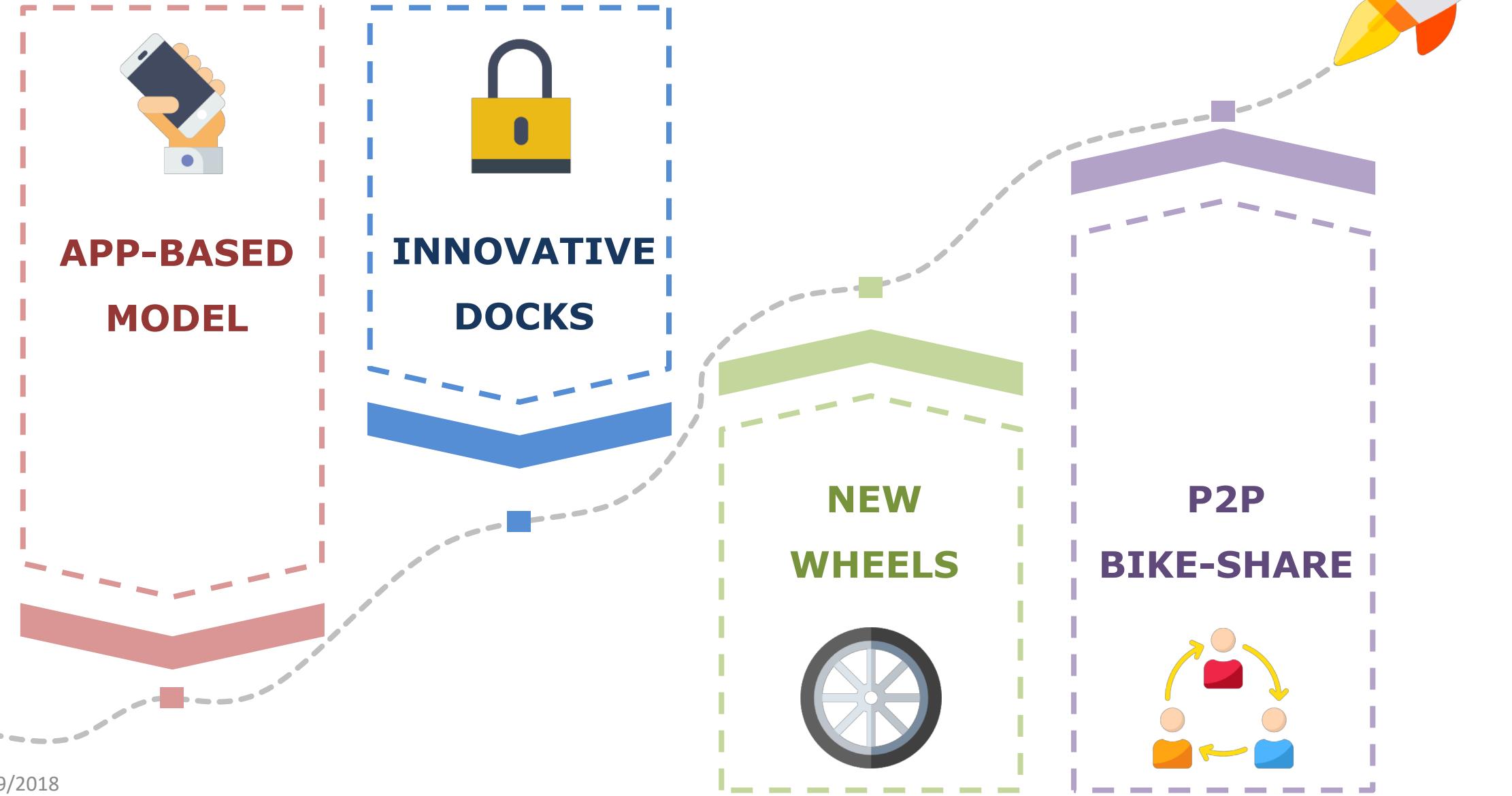


Adjusted Trip Duration



<https://www.youtube.com/watch?v=k90AnR1w1R4>

# New Trends and Opportunities



THANKS FOR YOUR ATTENTION

Q & A