

# Data Driven Decision Making

## Citi Bike Trip Histories

Submitted By  
Iftekhar Hossain

### **List of the variables and their data types:**

The dataset has fifteen variables:

1. Tripduration (number/numeric: in seconds)
2. Starttime (number/ Date and Time, date of number type)
3. Stoptime (number/ Date and Time, date of number type)
4. start station id (number/numeric)
5. start station name (text/string)
6. start station latitude (number/numeric)
7. start station longitude (number/numeric)
8. end station id (number/numeric)
9. end station name (text/string)
10. end station latitude (number/numeric)
11. end station longitude (number/numeric)
12. bikeid (number/numeric)
13. usertype (Customer = 24 hours pass or 03 days pass, Subscriber = Annual Membership)
14. birth year (number/numeric: in year)
15. gender (0=unknown, 1=male, 2=female)

### **Number of unique values in those fifteen variables:**

1. Trip Duration: 12305
2. Start Time: 37651
3. Stop Time: 37770
4. Start Station ID: 337
5. Start Station Name: 337
6. Start Station Latitude: 337
7. Start Station Longitude: 337
8. End Station ID: 337
9. End Station Name: 337
10. End Station Latitude: 337

11. End Station Longitude: 337
12. Bike ID : 5794
13. User Type: 02 (Subscriber = Annual Membership, Customer = 24 hours pass or 03 days pass)
14. Year of Birth: 76
15. Gender: 03 (0=unknown, 1=male, 2=female)

### Assignment 01 :

File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Paste

Clipboard

Calibri

11

A<sup>+</sup>

A<sup>-</sup>

B

I

U

Font

Alignment

Wrap Text

Merge & Center

Number

General

Conditional Formatting

Format as Table

Cell Styles

A2

695

	A	B	C	D	E	F	G	H	I	J
1	Trip Duration	Start Time	StartTime_Date	StartTime_Hour	startTime_DOW	Stop Time	StopTime_Date	StopTime_Hour	StopTime_DOW	Start Station ID
2	695	01/06/2013 0:00	01/06/2013	12:00:01 AM	7	01/06/2013 0:11	01/06/2013	12:11:36 AM	7	444
3	693	01/06/2013 0:00	01/06/2013	12:00:08 AM	7	01/06/2013 0:11	01/06/2013	12:11:41 AM	7	444
4	2059	01/06/2013 0:00	01/06/2013	12:00:44 AM	7	01/06/2013 0:35	01/06/2013	12:35:03 AM	7	406
5	123	01/06/2013 0:01	01/06/2013	12:01:04 AM	7	01/06/2013 0:03	01/06/2013	12:03:07 AM	7	475
6	1521	01/06/2013 0:01	01/06/2013	12:01:22 AM	7	01/06/2013 0:26	01/06/2013	12:26:43 AM	7	2008
7	2028	01/06/2013 0:01	01/06/2013	12:01:47 AM	7	01/06/2013 0:35	01/06/2013	12:35:35 AM	7	485
8	2057	01/06/2013 0:02	01/06/2013	12:02:33 AM	7	01/06/2013 0:36	01/06/2013	12:36:50 AM	7	285
9	369	01/06/2013 0:03	01/06/2013	12:03:29 AM	7	01/06/2013 0:09	01/06/2013	12:09:38 AM	7	509
10	1829	01/06/2013 0:03	01/06/2013	12:03:47 AM	7	01/06/2013 0:34	01/06/2013	12:34:16 AM	7	265
11	829	01/06/2013 0:04	01/06/2013	12:04:22 AM	7	01/06/2013 0:18	01/06/2013	12:18:11 AM	7	404
12	1316	01/06/2013 0:04	01/06/2013	12:04:28 AM	7	01/06/2013 0:26	01/06/2013	12:26:24 AM	7	423
13	1456	01/06/2013 0:04	01/06/2013	12:04:41 AM	7	01/06/2013 0:28	01/06/2013	12:28:57 AM	7	502

File

Home

Insert

Draw

Page Layout

Formulas

Data

Review

View

Help

Acrobat

Paste

Clipboard

Calibri

11

A<sup>+</sup>

A<sup>-</sup>

B

I

U

Font

Alignment

Date

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

Editing

Analyze Data

Sensitivity

D1

Day of Month

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Trip Duration	Start Time	StartTime_Date	Day of Month	StartHours_Day	StartTime_Hour	StartTime_DOW	Stop Time	StopTime_Date	Day of Month	StopTime_Hour	StopTime_DOW	Start Station ID
2	695	01/06/2013 0:00	01/06/2013	1	0	12:00:01 AM	7	01/06/2013 0:11	01/06/2013	1	12:11:36 AM	7	444
3	693	01/06/2013 0:00	01/06/2013	1	0	12:00:08 AM	7	01/06/2013 0:11	01/06/2013	1	12:11:41 AM	7	444
4	2059	01/06/2013 0:00	01/06/2013	1	0	12:00:44 AM	7	01/06/2013 0:35	01/06/2013	1	12:35:03 AM	7	406
5	123	01/06/2013 0:01	01/06/2013	1	0	12:01:04 AM	7	01/06/2013 0:03	01/06/2013	1	12:03:07 AM	7	475
6	1521	01/06/2013 0:01	01/06/2013	1	0	12:01:22 AM	7	01/06/2013 0:26	01/06/2013	1	12:26:43 AM	7	2008
7	2028	01/06/2013 0:01	01/06/2013	1	0	12:01:47 AM	7	01/06/2013 0:35	01/06/2013	1	12:35:35 AM	7	485
8	2057	01/06/2013 0:02	01/06/2013	1	0	12:02:33 AM	7	01/06/2013 0:36	01/06/2013	1	12:36:50 AM	7	285
9	369	01/06/2013 0:03	01/06/2013	1	0	12:03:29 AM	7	01/06/2013 0:09	01/06/2013	1	12:09:38 AM	7	509
10	1829	01/06/2013 0:03	01/06/2013	1	0	12:03:47 AM	7	01/06/2013 0:34	01/06/2013	1	12:34:16 AM	7	265
11	829	01/06/2013 0:04	01/06/2013	1	0	12:04:22 AM	7	01/06/2013 0:18	01/06/2013	1	12:18:11 AM	7	404
12	1316	01/06/2013 0:04	01/06/2013	1	0	12:04:28 AM	7	01/06/2013 0:26	01/06/2013	1	12:26:24 AM	7	423
13	1456	01/06/2013 0:04	01/06/2013	1	0	12:04:41 AM	7	01/06/2013 0:28	01/06/2013	1	12:28:57 AM	7	502
14	386	01/06/2013 0:05	01/06/2013	1	0	12:05:13 AM	7	01/06/2013 0:11	01/06/2013	1	12:11:39 AM	7	241
15	924	01/06/2013 0:05	01/06/2013	1	0	12:05:21 AM	7	01/06/2013 0:20	01/06/2013	1	12:20:45 AM	7	486
16	1233	01/06/2013 0:06	01/06/2013	1	0	12:06:44 AM	7	01/06/2013 0:27	01/06/2013	1	12:27:17 AM	7	527
17	512	01/06/2013 0:03	01/06/2013	1	0	12:03:36 AM	7	01/06/2013 0:12	01/06/2013	1	12:12:08 AM	7	309
18	505	01/06/2013 0:03	01/06/2013	1	0	12:03:45 AM	7	01/06/2013 0:12	01/06/2013	1	12:12:10 AM	7	309
19	833	01/06/2013 0:07	01/06/2013	1	0	12:07:29 AM	7	01/06/2013 0:21	01/06/2013	1	12:21:22 AM	7	503
20	1818	01/06/2013 0:08	01/06/2013	1	0	12:08:10 AM	7	01/06/2013 0:38	01/06/2013	1	12:38:28 AM	7	257
21	682	01/06/2013 0:08	01/06/2013	1	0	12:08:53 AM	7	01/06/2013 0:20	01/06/2013	1	12:20:15 AM	7	486

**Number of Missing data in the variables:**

1. Trip Duration: 0
2. Start Time: 0
3. Stop Time: 0
4. Start Station ID: 0
5. Start Station Name:0
6. Start Station Latitude: 0
7. Start Station Longitude: 0
8. End Station ID: 18059
9. End Station Name: 18059
10. End Station Latitude: 18059
11. End Station Longitude: 18059
12. Bike ID : 0
13. User Type: 0
14. Year of Birth: 240321
15. Gender: 0

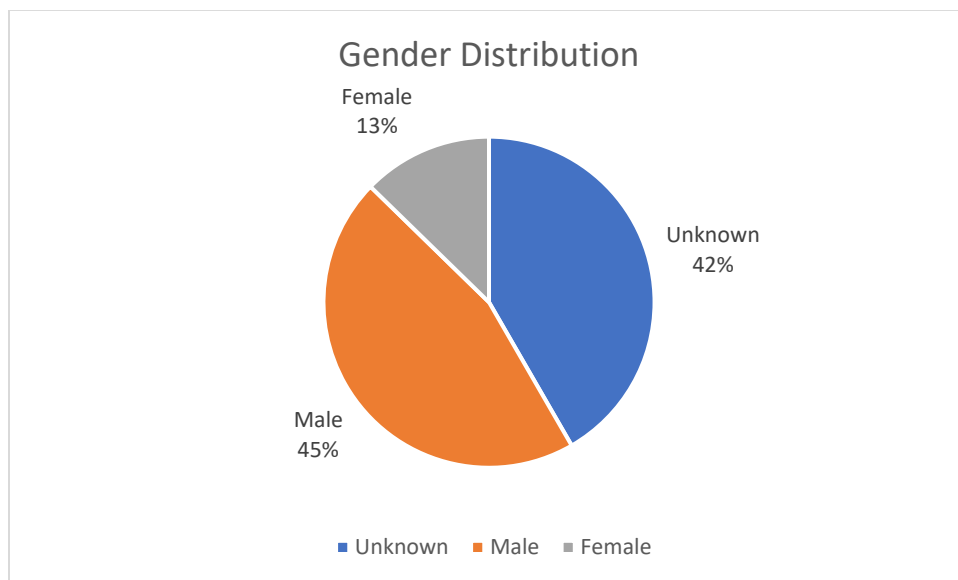
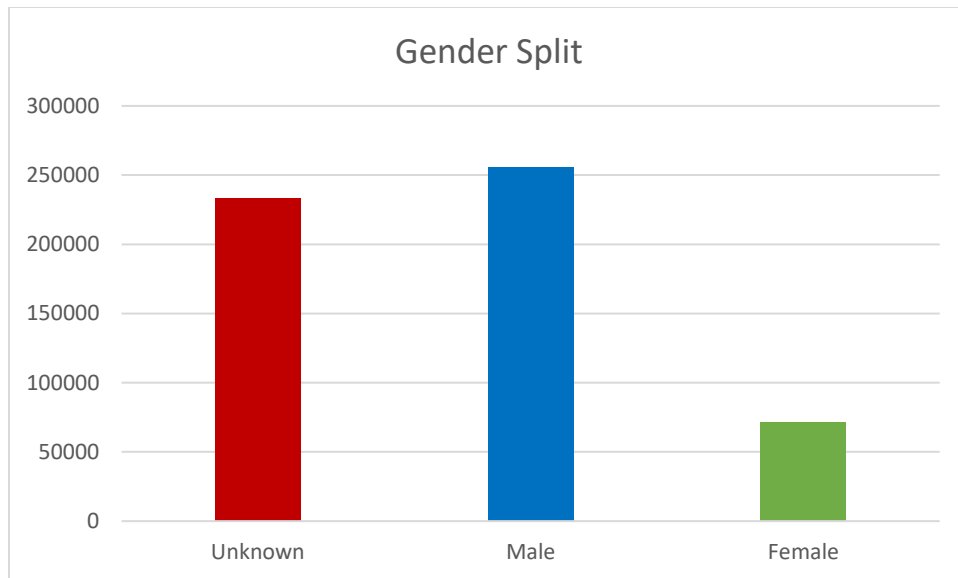
### **Dealing with Missing Values:**

The missing values in End Station ID, End Station Name, End Station Latitude, End Station Longitude columns observations (18059) are only 3.1% of the total (577703) observations(rows). We can drop the rows with missing values. Now we have 559644 observations to work with. We can ignore the missing values of Year of Birth column ( 240321 which is 41.6% of the total observations) and can delete the entire column. Since we are not using this column (Year of Birth) values in this analyzing.

=FILTER(A2:T577704,(N2:N577704<>"NULL")+(O2:O577704<>"NULL")+(P2:P577704<>"NULL")+(Q2:Q577704<>"NULL"))

### **Gender Split:**

Unknown	233215
Male	255339
Female	71090

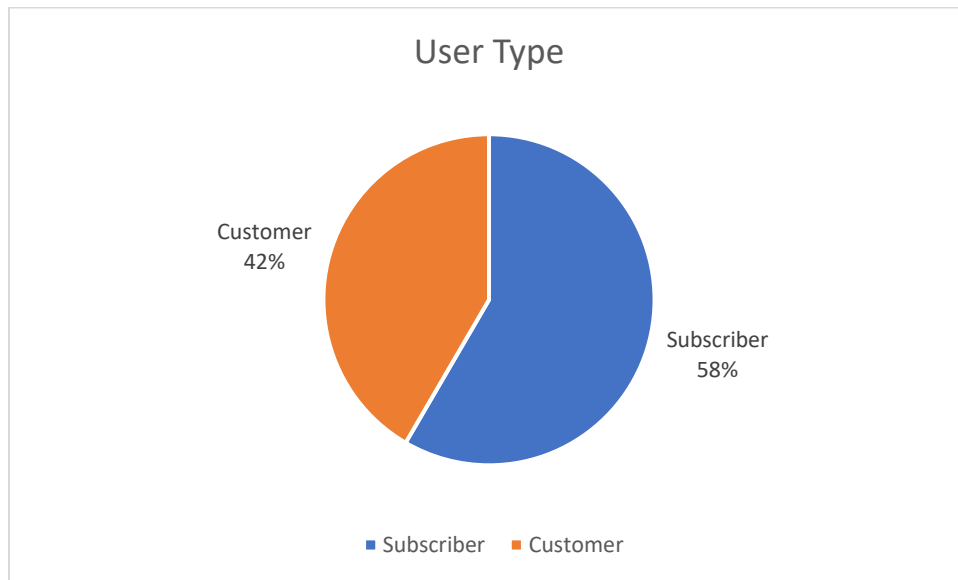


### **User Type Distribution:**

There are 02 types of user:

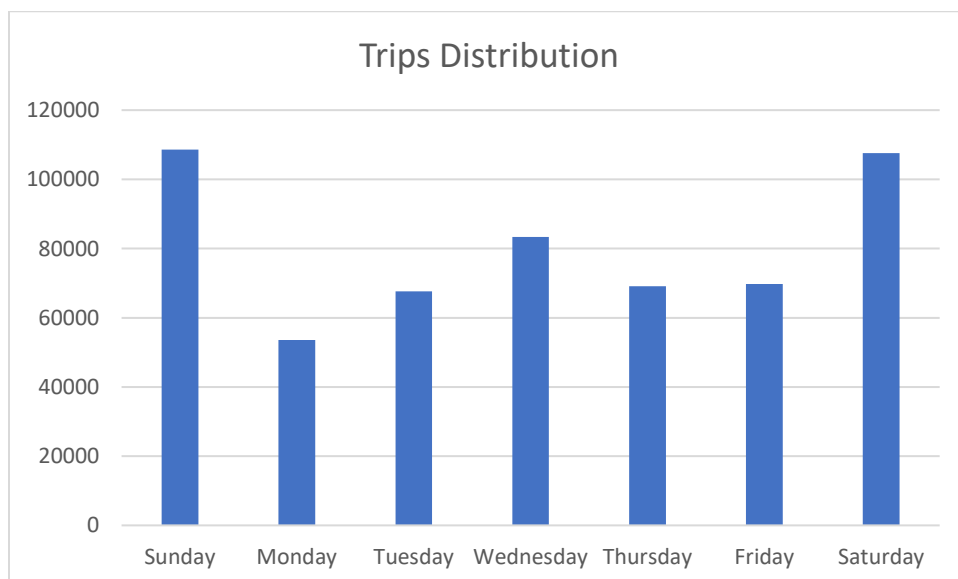
1. Subscriber = Annual Membership
2. Customer = 24 hours pass or 03 days pass

User Type	
Subscriber	326843
Customer	232801



### **Assignment 2:**

Days of the week	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Total Rides	108576	53624	67659	83359	69142	69720	107564



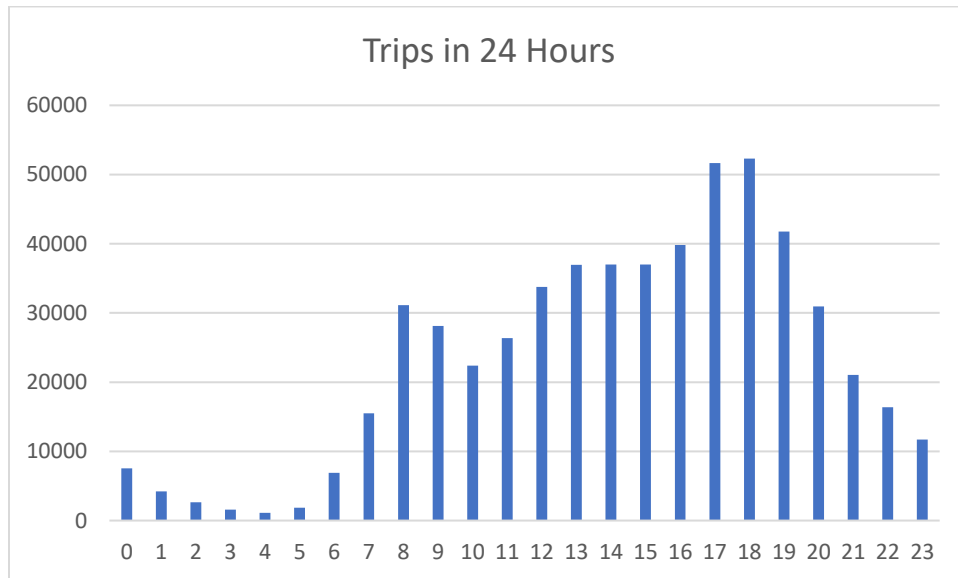
- On which days of the week bike trips/rentals are the most popular?  
 Answer: Sunday is the most popular, then the second most popular is Saturday and Wednesday is the thirds most popular for bike rental.

- Weekends are more popular for bike riding than working days.

2. On which days of the week Citi bike rentals are least popular?

Answer: Monday is the least popular. After that Tuesday, Thursday and Friday.

- In the working days of the week bike rentals are least popular.



Hour	0	1	2	3	4	5	6	7	8	9	10	11
Trips	7528	4235	2664	1563	1133	1875	6884	15505	31105	28122	22379	26367

12	13	14	15	16	17	18	19	20	21	22	23
33753	36935	36980	36992	39829	51669	52300	41753	30951	21035	16383	11704

3. During which hours of the day Citi bike rentals are most popular?

Answer: Between 16 and 19 Hours ( 04:00 PM to 07:00 PM) the bike rentals are most popular.

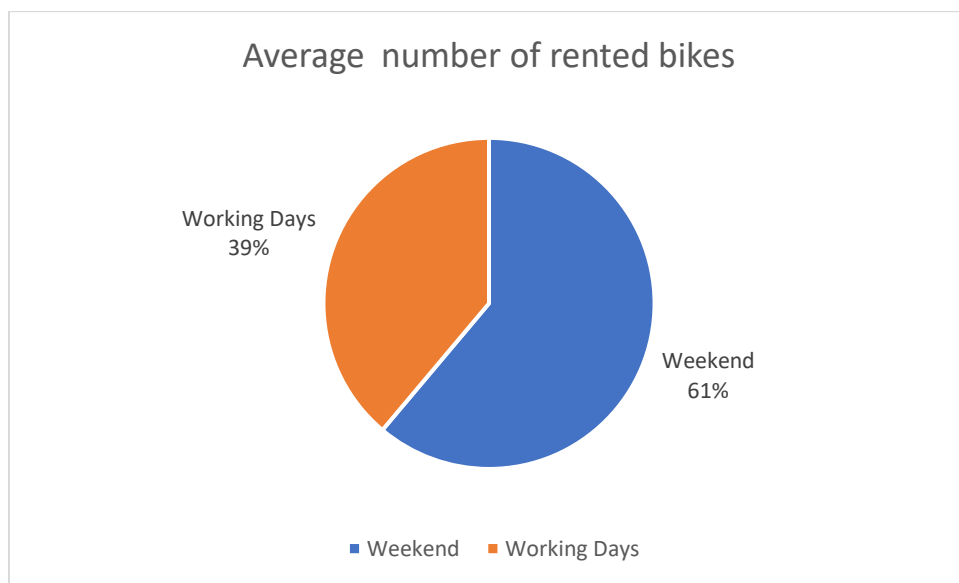
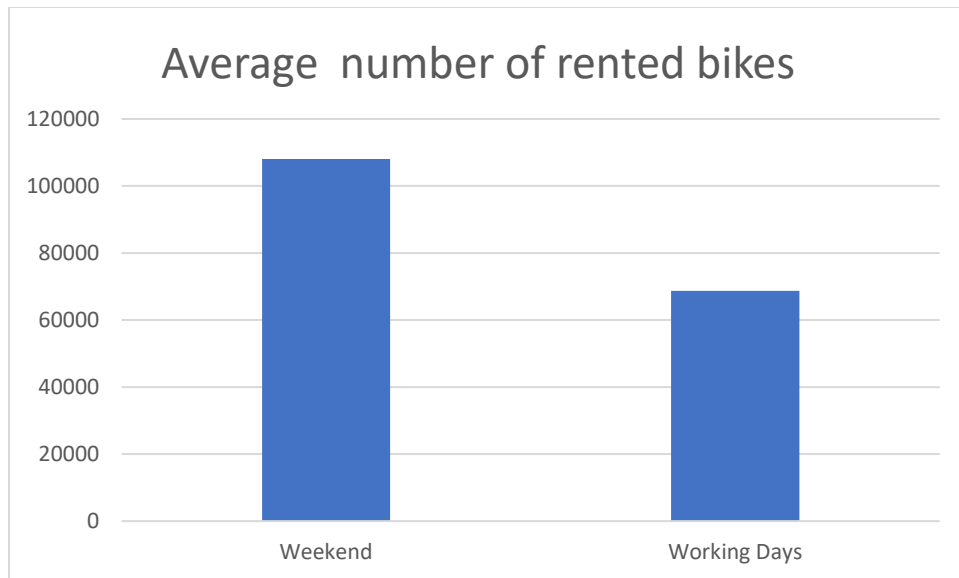
4. During which hours of the day Citi bike rentals are least popular?

Answer: 0 to 6 Hours (Between 00.00 hrs to 06:00 AM) bike rentals are least popular.

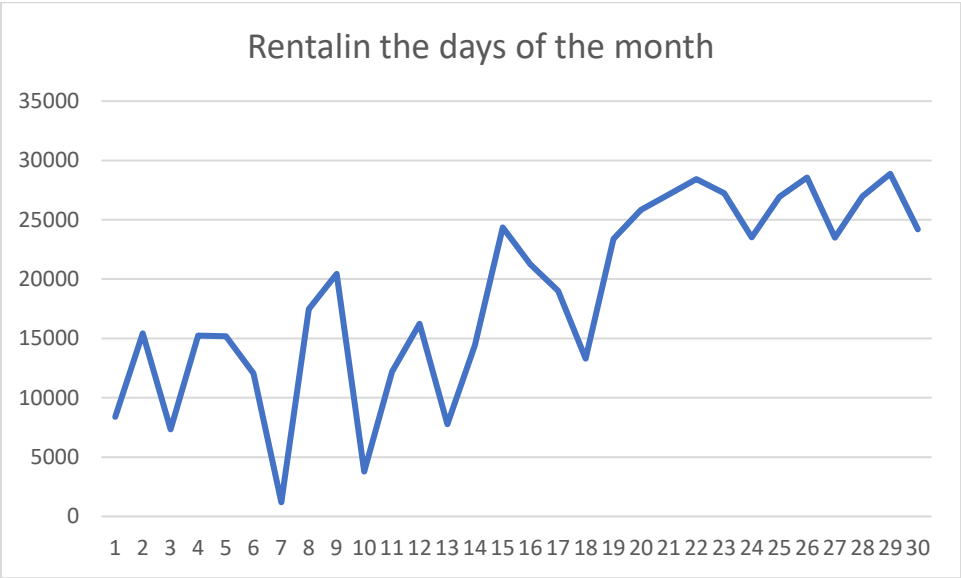
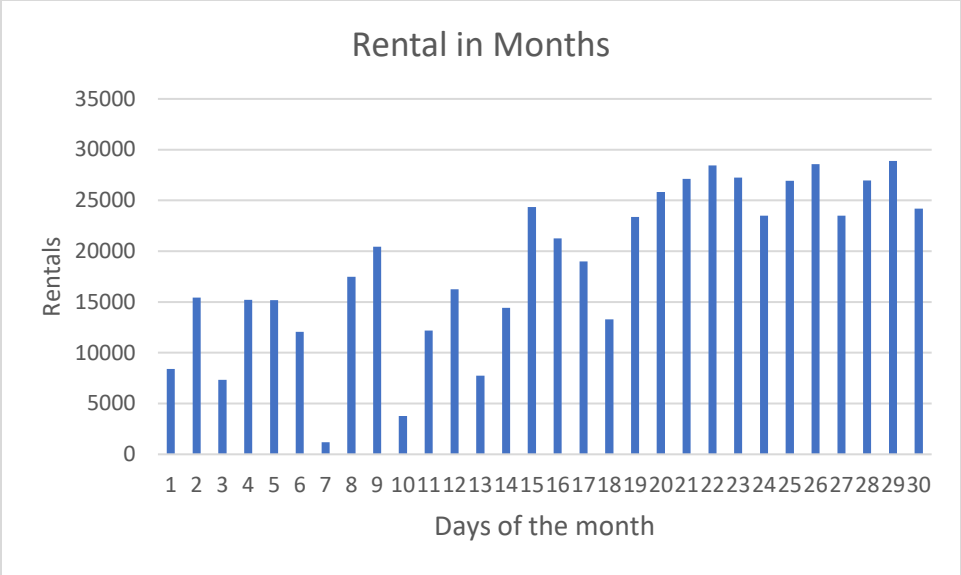
5. A comparison between the average number of rented Citi bikes on the working days and the average number of rented Citi bikes during the weekend of the week.

6. Answer:

Average number of rented bikes on Weekend	108070
Average number of rented bikes on Working days	68701

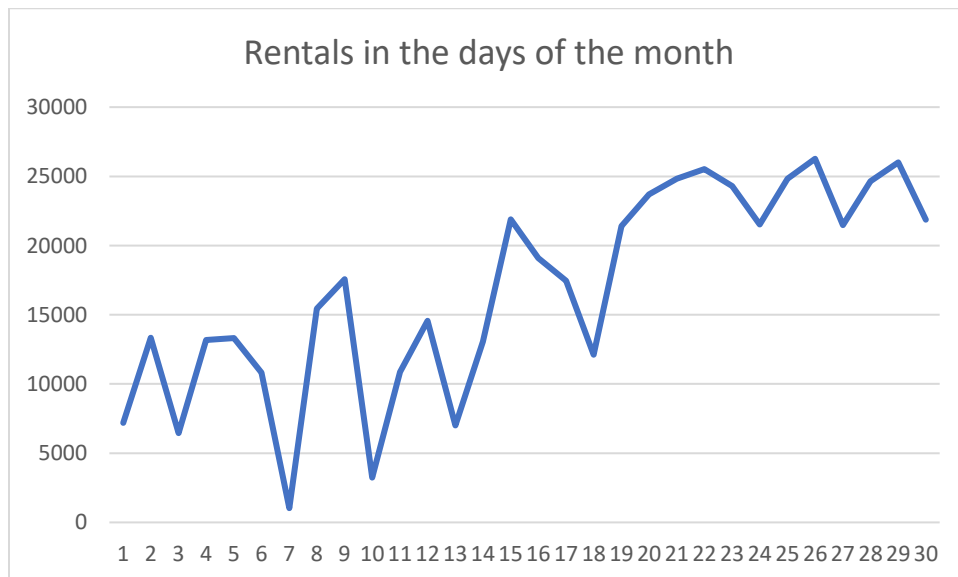
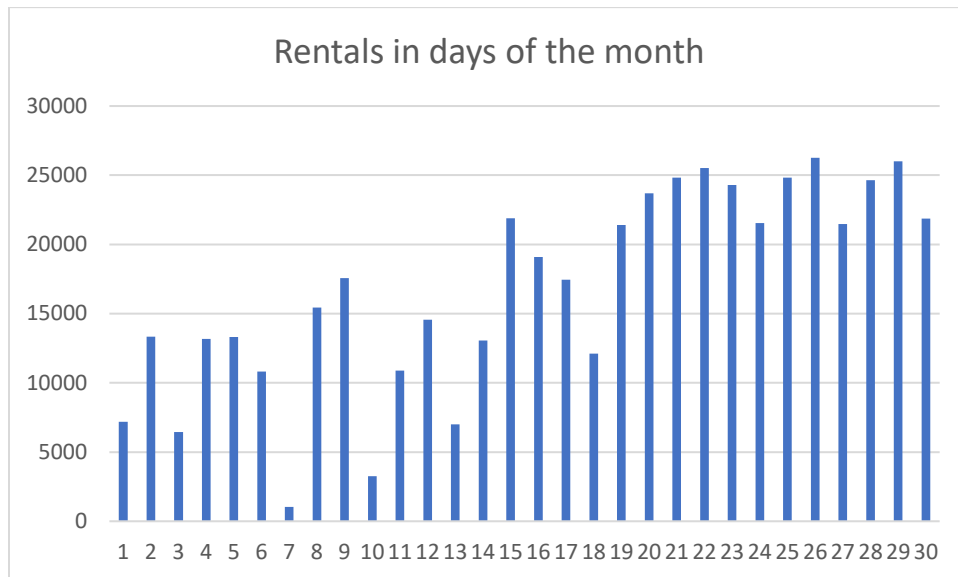


**Before removing Outliers:**





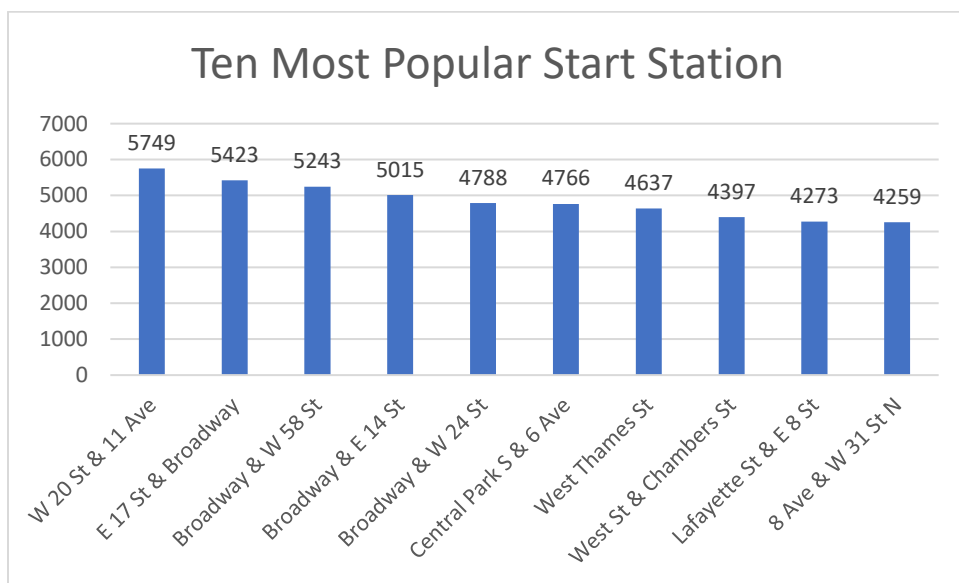
**After removing the outliers By 5<sup>th</sup> percentile and 95<sup>th</sup> percentile values:**



**1. The Ten most popular start stations for Citi bike rentals:**

Answer:

Name of the start stations	Total Rental
W 20 St & 11 Ave	5749
E 17 St & Broadway	5423
Broadway & W 58 St	5243
Broadway & E 14 St	5015
Broadway & W 24 St	4788
Central Park S & 6 Ave	4766
West Thames St	4637
West St & Chambers St	4397
Lafayette St & E 8 St	4273
8 Ave & W 31 St N	4259



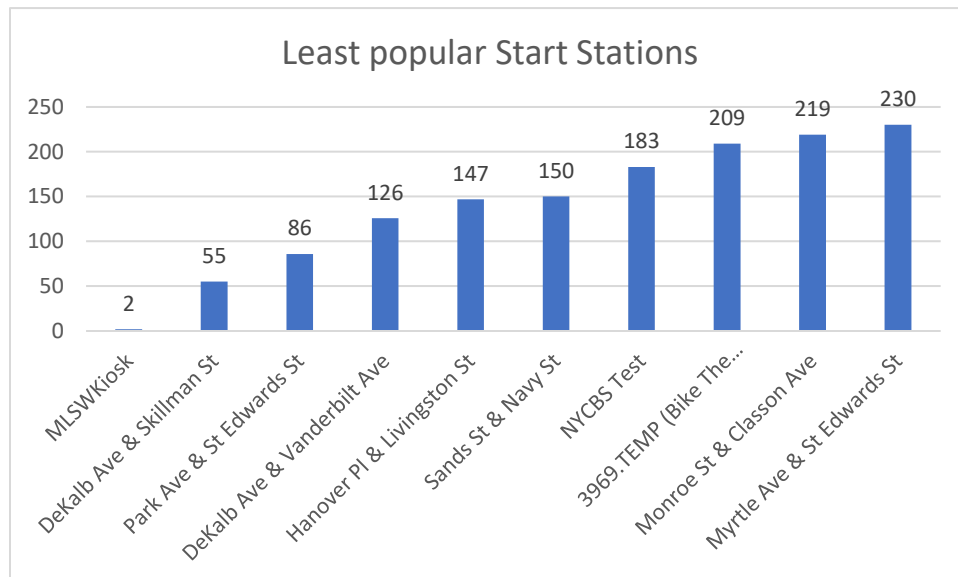
**The Ten least popular start stations for Citi bike rentals:**

Answer:

Name of the Start Station	Total number of rides
MLSWKiosk	2
DeKalb Ave & Skillman St	55
Park Ave & St Edwards St	86
DeKalb Ave & Vanderbilt Ave	126
Hanover Pl & Livingston St	147
Sands St & Navy St	150
NYCBS Test	183
3969.TEMP (Bike The Branches - Central Branch)	209
Monroe St & Classon Ave	219

Myrtle Ave & St Edwards St

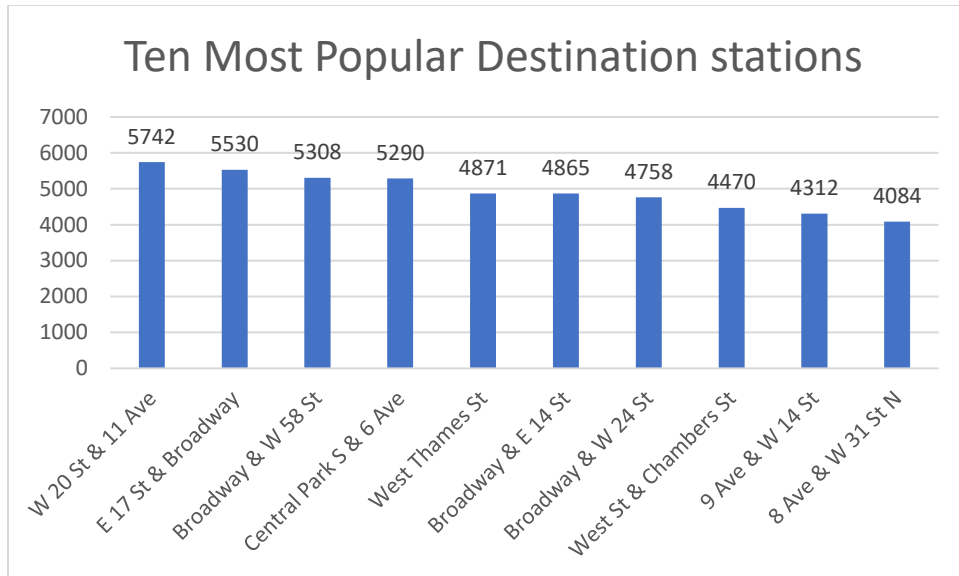
230



**1. Ten most popular destination stations for Citi bike rentals**

Answer:

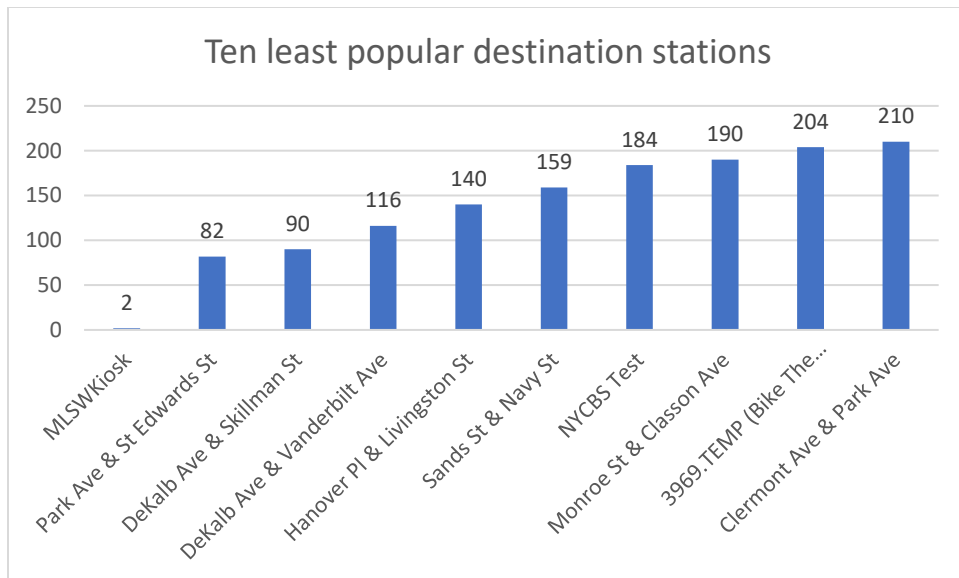
Name of the End Station	Total number of rides
W 20 St & 11 Ave	5742
E 17 St & Broadway	5530
Broadway & W 58 St	5308
Central Park S & 6 Ave	5290
West Thames St	4871
Broadway & E 14 St	4865
Broadway & W 24 St	4758
West St & Chambers St	4470
9 Ave & W 14 St	4312
8 Ave & W 31 St N	4084



## 2. Ten least popular destination stations for Citi bike rentals

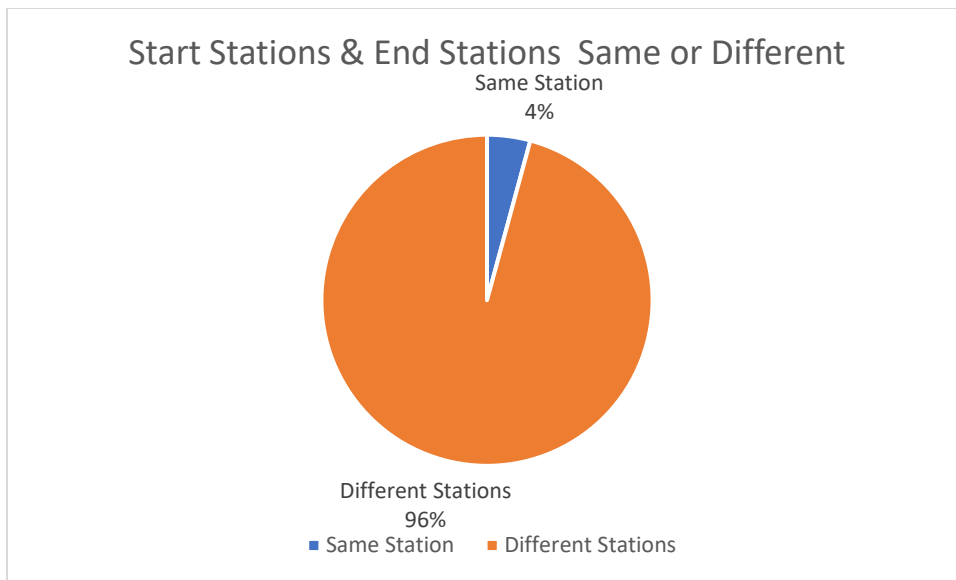
Answer:

Name of the Destination station	Number of rides
MLSWKiosk	2
Park Ave & St Edwards St	82
DeKalb Ave & Skillman St	90
DeKalb Ave & Vanderbilt Ave	116
Hanover Pl & Livingston St	140
Sands St & Navy St	159
NYCBS Test	184
Monroe St & Classon Ave	190
3969.TEMP (Bike The Branches - Central Branch)	204
Clermont Ave & Park Ave	210



3. Are the start stations and the destination stations usually being the same or different?  
Answer:

Same Station	23624
Different Stations	536020



### **Assignment 3:**

#### **Before removing Outliers:**

Before removing Outliers:	Trip Duration
Count	559644

Min	61
Max	3876479
Range	3876418
Mean	1371.34829
Median	874
Mode	547
stdev( $\sigma$ )	9118.59798
Q1	514
Q3	1398

1. What are the most frequent ride durations?

Answer: Between 400 to 650 Seconds

2. What are the mean value and median value of the distribution of Citi bike trip durations?

Answer:

Mean = 1371 (around) and Median = 874

But, After removing Outliers by Z-score ( $>3$  and  $<-3$ ), Mean (around)= 1147, Median=871

- The mean is greater than the median.
- The Trip Duration distribution is right-skewed.

3. What Citi bike trip durations fall within 1 standard deviation from the mean value?

Answer:

Values fall within one standard deviation from the mean value = From (Mean- Standard Deviation) to (Mean + Standard Deviation)

Before removing outliers,

Mean-SD	-7747.25
Mean+SD	10489.95

Our minimum Trip Durations is 61 Seconds. So, 61 to Mean+SD = 10489 seconds fall within one standard deviation from the mean value.

After removing outliers:

Mean -SD	-267
Mean +SD	2560

But, our minimum Trip Durations is 61 Seconds. So, 61 to 2560 seconds fall within one standard deviation from the mean value.

After removing outliers by 5<sup>th</sup> percentile and 95<sup>th</sup> percentile:

Mean	980
Standard Deviation	528
Mean-SD	452
Mean+SD	1433

From 452 to 1433 seconds fall within the 1 standard deviation from the mean value.

4. What are the percentiles of the trip durations at the levels of 5%, 25%, 50%, 75% and 95%?

Before removing the Outliers Trip Durations:

Trip Duration	
5th percentile	237
25th percentile	514
50th percentile	874
75th percentile	1398
95th percentile	2545

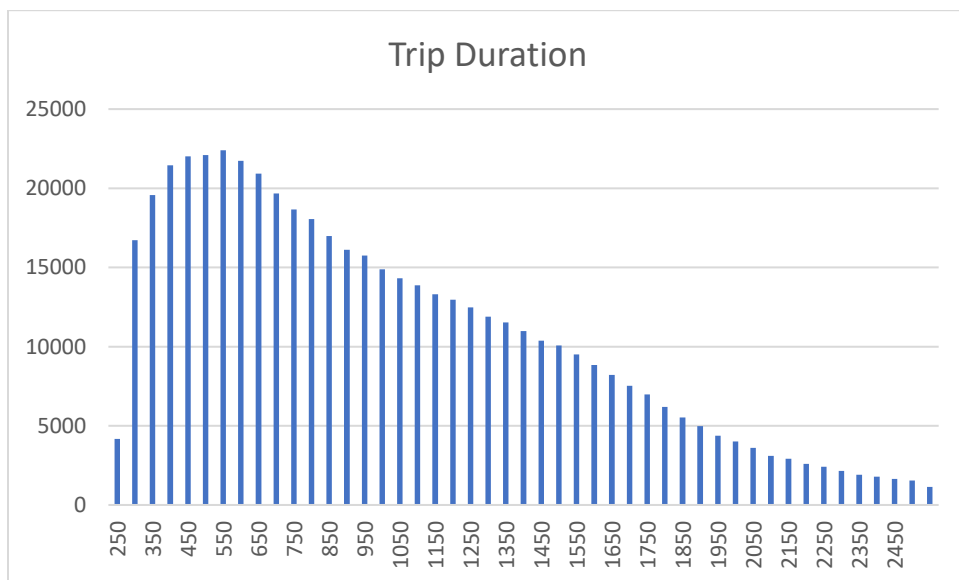
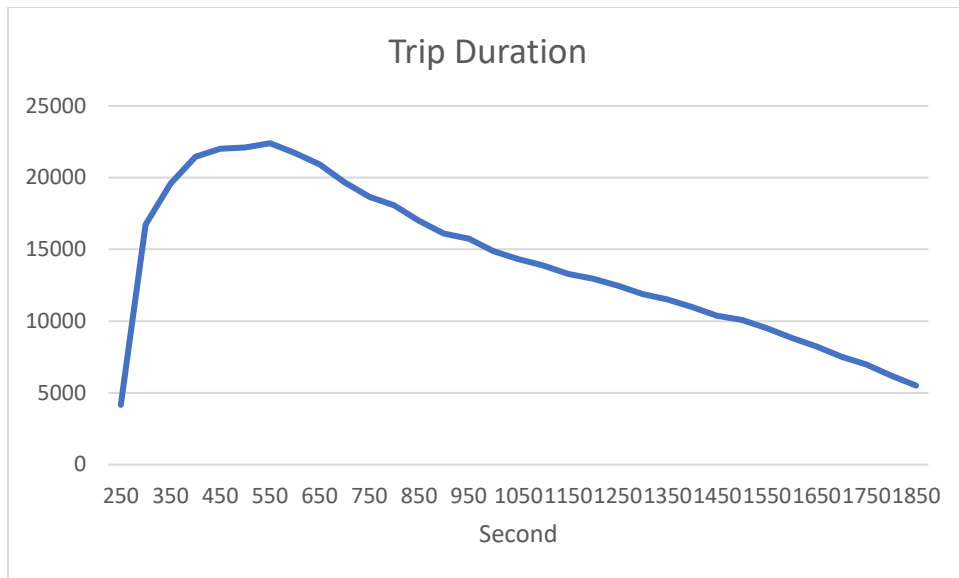
After removing the Outliers Trip Durations:

5th percentile	237
25th percentile	513
50th percentile	871
75th percentile	1393
95th percentile	2487

5. What is the shape of the distribution of Citi bike trip duration, i.e. is it symmetric or skewed?

Answer: It is not symmetric.

- The Trip Duration distribution is right-skewed.
- In Trip Duration variable, the mean is greater than the median.



6. What is the skewness coefficient of the distribution of Citi bike trip duration?

Answer: 0.71 (around)

SKEW.P_Duration
0.711250943

- The positive value of skewness coefficient means right (or positive) skewness.

7. What is the relation between the mean and the median of the distribution of Citi bike ride duration? How does it relate to the assessment of the skewness of the distribution?

Answer:



In Trip Duration distribution, the is Mean greater than Median value (Mean>Median).

- In positive skewness (Right Skewness): Mean>Median.
- In negative skewness (Left Skewness): Mean<Median

8. Which statistic, mean or median, better represents the central tendency of the distribution of the ride duration?

Answer:

Median. Because The Trip Duration distribution is Skewed.

Mean is better for mostly symmetrical distributions and if they have no outliers.

9. Does the distribution of ride duration look like a normal distribution?

Answer: No, the Trip Duration distribution is right-skewed.

Find the outliers of the ride duration variable using:

**The z-score method:**

If the z-score value is  $> 3$  or  $< -3$ , I consider the value as an outliers, There are 1616 Trips Duration values out of 559645 where the Z-score is  $> 3$ . But no Trip duration Z-score is  $< -3$ .

Number of Trip Duration where Z-score $<-3$	0
Number of Trip Duration where Z-score $>3$	1616

Number of outliers observations by using Z-score are 1616.

After removing Outliers	Trip Duration
Count	558028
Min	61
Max	28707
Range	28646
Mean	1146.575439
Median	871
Mode	547
Stdev ( $\sigma$ )	1413.104036
Q1	513
Q3	1393
5th percentile	237
25th percentile	513
50th percentile	871
75th percentile	1393
95th percentile	2487

After Removing of the outliers observations by 05<sup>th</sup> percentile and 95<sup>th</sup> percentile:

Min	237
Max	2545
Range	2308
Mean	980
Median	873
Mode	547
STDEV:P	528
Mean-SD	452
Mean+SD	1433

#### Assignment 4:

Euclidean\_Distance =SQRT((O2-S2)^2+(P2-T2)^2)  
and so on =SQRT((O503963-S503963)^2+(P503963-T503963)^2)

Manhattan\_Distance: =ABS(O2-S2)+ABS(P2-T2)  
And so on =ABS(O503963-S503963)+ABS(P503963-T503963)

File Home Insert Draw Page Layout Formulas Data Review View Help Acrobat										
<div> <div> <div>Undo</div> <div>Clipboard</div> </div> <div> <div>Font</div> <div>Alignment</div> <div>Number</div> </div> <div> <div>General</div> <div>Conditional Formatting</div> <div>Format as Table</div> <div>Cell Styles</div> </div> <div> <div>Insert</div> <div>Delete</div> <div>Format</div> </div> <div> <div>Σ</div> <div>Find</div> </div> <div> <div>Analyze Data</div> <div>Sensitivity</div> </div> <div> <div>Send to MindManager</div> </div> </div>										
Y2 =SQRT((O2-S2)^2+(P2-T2)^2)										
	R	S	T	U	V	W	X	Y	Z	
1	D End Station Name	End Station Latitude	End Station Longitude	Bike ID	User Type	Gender	S_E_Same	Euclidean_Distance	Manhattan_Distance	S
2	2 St James Pl & Pearl St	40.71117416	-74.00016545	17954	Subscriber	1	0	0.00778374	0.01011885	
3	5 W 21 St & 6 Ave	40.74173969	-73.99415556	18826	Subscriber	2	0	0.003173686	0.00375497	S
4	1 E 52 St & 2 Ave	40.756014	-73.967416	15748	Subscriber	1	0	0.002081071	0.00294293	
5	3 Lafayette St & E 8 St	40.73020661	-73.99102628	18746	Customer	0	1	0	0	
6	1 E 14 St & Avenue B	40.72938685	-73.97772429	18506	Subscriber	1	0	0.027298393	0.038298284	
7	9 Ave & W 22 St	40.7454973	-74.00197139	17548	Subscriber	1	0	0.008672187	0.01157344	
8	5 E 15 St & Irving Pl	40.73524276	-73.98758561	15035	Subscriber	1	0	0.006112674	0.00850949	
9	5 Mercer St & Bleecker St S	40.72679454	-73.99695094	17381	Subscriber	1	0	0.00960012	0.01357147	
10	8 Front St & Gold St	40.70224	-73.982578	17126	Subscriber	2	1	0	0	
11	7 Macon St & Nostrand Ave	40.68098339	-73.95004798	17889	Subscriber	2	0	0.087865212	0.114618729	
12	1 E 20 St & 2 Ave	40.73587678	-73.98205027	17878	Subscriber	1	0	0.010616952	0.014646187	
13	7 E 17 St & Broadway	40.73704984	-73.99009296	19416	Subscriber	2	0	0.014575894	0.02032589	
14	6 Bialystoker Pl & Delancey St	40.71622644	-73.98261206	18460	Subscriber	1	0	0.00550062	0.00640974	
15	7 E 33 St & 2 Ave	40.744023	-73.976056	18126	Customer	0	0	0.007792421	0.00955972	
16	7 E 6 St & Avenue B	40.72453734	-73.98185424	19026	Subscriber	2	0	0.023902563	0.02590956	
17	6 Bank St & Hudson St	40.73652889	-74.00618026	18001	Customer	0	0	0.017806721	0.02157686	

### **Assignment 5:**

After removing Outliers by using 5<sup>th</sup> percentile and 95<sup>th</sup> Percentile.

	Euclidean_Distance		Manhattan_Distance
Count	503962	Count	503962
Min	0	Min	0
Max	0.127477218	Max	0.169626
Range	0.127477218	Range	0.169626
Mean	0.022120532	Mean	0.028249373
Median	0.018657859	Median	0.02365919
Mode	0	Mode	0
stdev( $\sigma$ )	0.015258803	stdev( $\sigma$ )	0.019539297
Q1	0.010897601	Q1	0.01379603
Q3	0.029904661	Q3	0.038350148
5th percentile	0.004169926	5th percentile	0.00511489
25th percentile	0.010897601	25th percentile	0.01379603
50th percentile	0.018657859	50th percentile	0.02365919
75th percentile	0.029904661	75th percentile	0.038350148
95th percentile	0.052588522	95th percentile	0.06727859

- **Manhattan Distance is more practical to measure ride distances is New York Citi bike ride.**

1. What are the most frequent ride distances?

Answer: From 0.01 to 0.02 (Manhattan Distance)

2. What are the mean and median values of the distribution of ride distances?

Answer: Mean=0.028249373, Median=0.02365919 (Manhattan Distance)

3. What ride distances fall within the one standard deviation from the mean value?

Answer:

Values fall within one standard deviation from the mean value = From (Mean-Standard Deviation) to (Mean + Standard Deviation). From 0.00871 to 0.047789

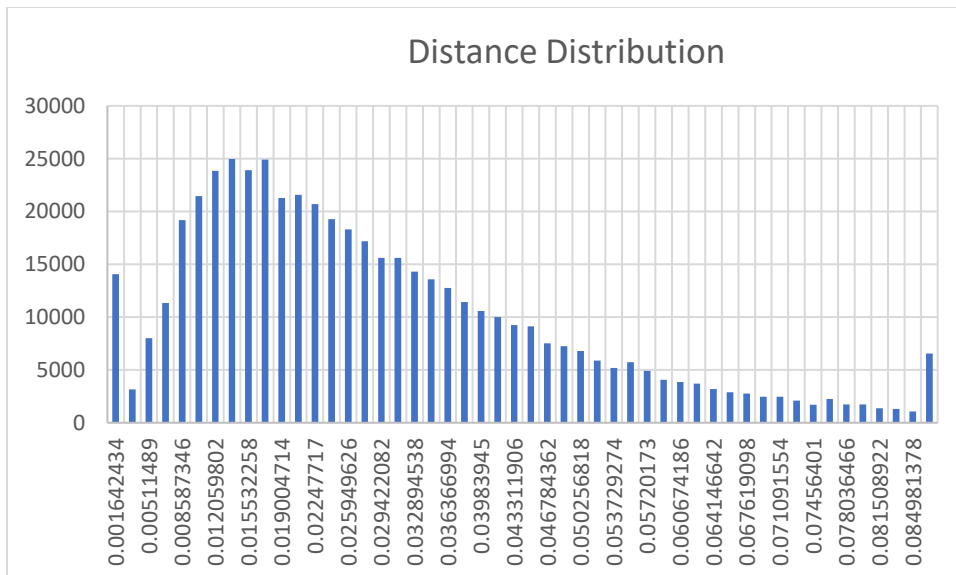
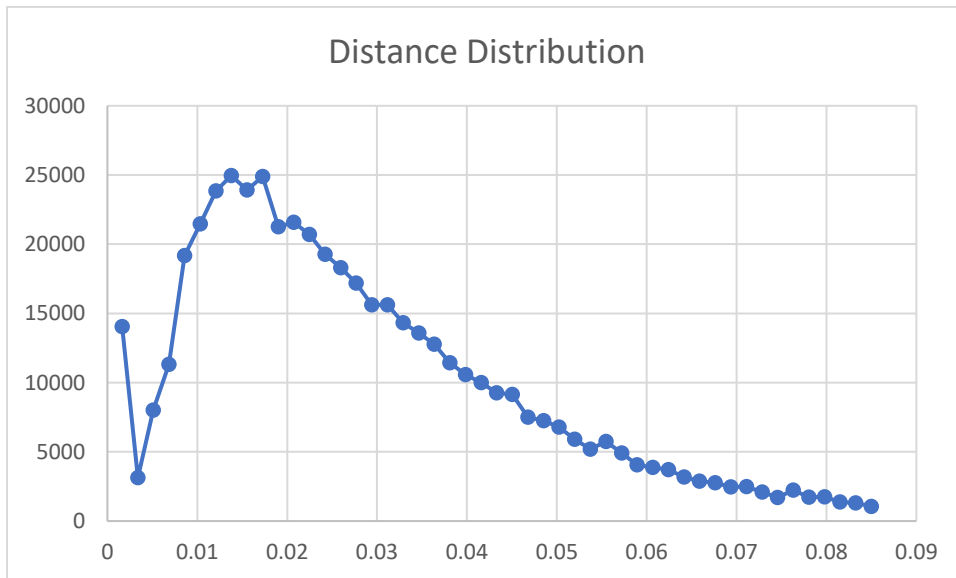
Mean-SD	0.00871
Mean+SD	0.047789

4. What are the percentiles of the ride distance at the level of 5%, 25%, 50%, 75% and 95%?

Answer: (For Manhattan Distance)

5th percentile	0.00511489
25th percentile	0.01379603
50th percentile	0.02365919
75th percentile	0.038350148
95th percentile	0.06727859

5. What is the shape of the distribution of the ride distance, i.e. is it symmetric or skewed?  
 Answer: Not symmetric. It is right (Positive) skewed.



4. What is the skewness coefficient of the distribution of the ride distance?  
 Answer: 1.19780466

7. What is the relation of the mean and median of the distribution of ride distance? How does it relate to the assessment of the skewness of the distribution?

Answer: In Citi bike ride distance distribution, the Mean is greater than Median value (Mean>Median). Mean=0.028249373, Median=0.02365919

- In positive skewness (Right Skewness): Mean>Median.
- In negative skewness (Left Skewness): Mean<Median

8. Which statistic – mean or median – better represents the central tendency of the distribution of distance?

Answer:

Median. Because The distribution of distance is Skewed.

Mean is better for mostly symmetrical distributions and if they have no outliers.

9. Does the distribution of ride distance look like a normal distribution?

Answer:

No, the distribution of distance is right-skewed.

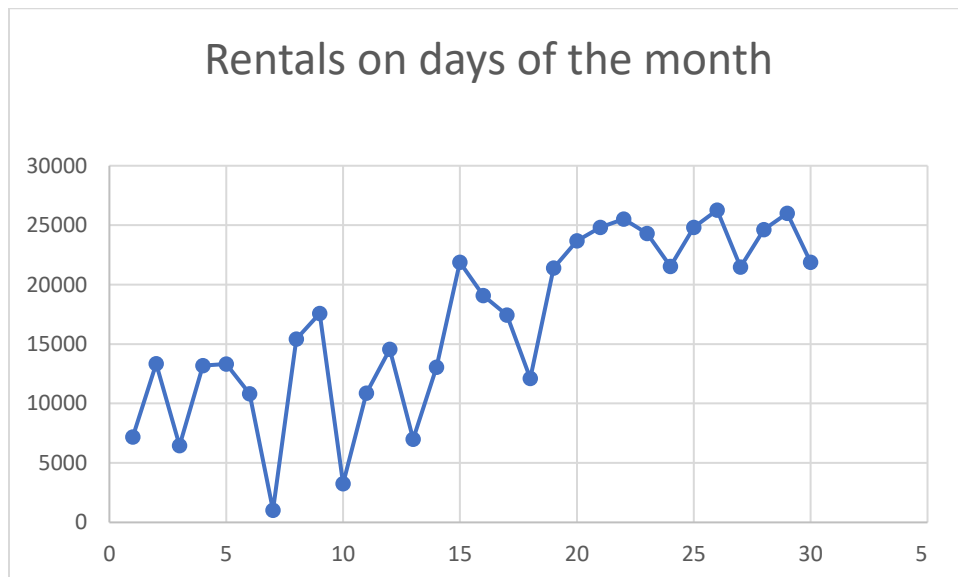
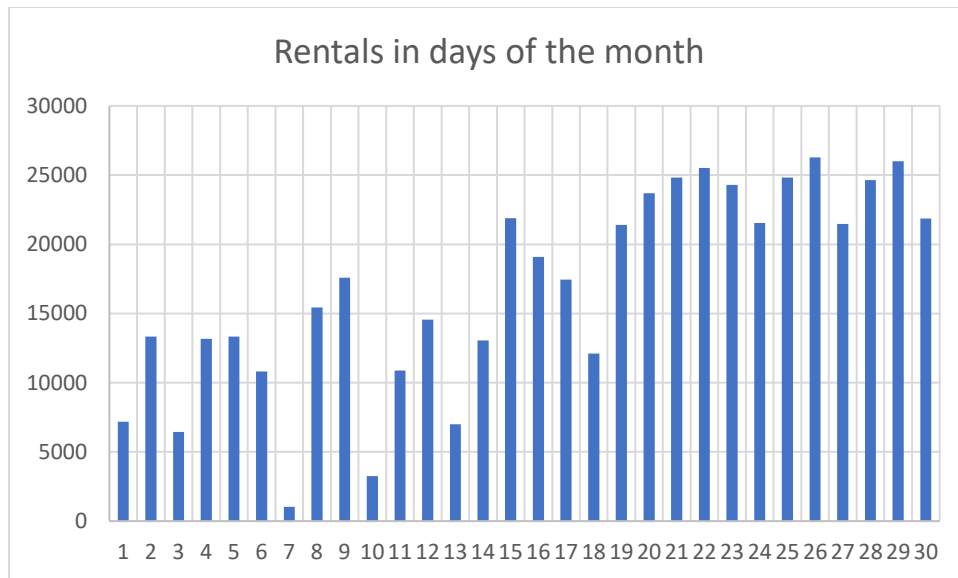
#### **A Comparison between the Euclidean and Manhattan distance metrics:**

	Euclidean Distance		Manhattan Distance
Count	503962	Count	503962
Min	0	Min	0
Max	0.127477218	Max	0.169626
Range	0.127477218	Range	0.169626
Mean	0.022120532	Mean	0.028249373
Median	0.018657859	Median	0.02365919
Mode	0	Mode	0
stdev( $\sigma$ )	0.015258803	stdev( $\sigma$ )	0.019539297
Q1	0.010897601	Q1	0.01379603
Q3	0.029904661	Q3	0.038350148
5th percentile	0.004169926	5th percentile	0.00511489
25th percentile	0.010897601	25th percentile	0.01379603
50th percentile	0.018657859	50th percentile	0.02365919
75th percentile	0.029904661	75th percentile	0.038350148
95th percentile	0.052588522	95th percentile	0.06727859

#### **Assignment 6:**

**Predicting the number of Trips for the first 07 days of the July 2013:**

Day	Rentals
1	7187
2	13344
3	6446
4	13176
5	13322
6	10813
7	1028
8	15433
9	17578
10	3247
11	10875
12	14563
13	6998
14	13058
15	21882
16	19088
17	17447
18	12109
19	21411
20	23697
21	24826
22	25522
23	24295
24	21534
25	24829
26	26271
27	21472
28	24636
29	26004
30	21871

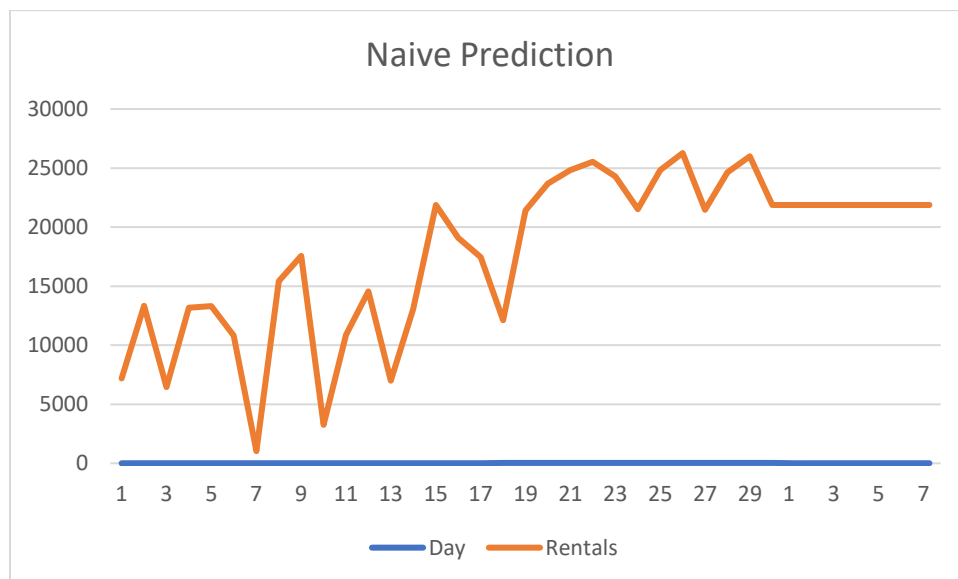


#### **A naïve prediction:**

In naïve prediction method, in a series of data, the last known observation is consider as the predictive value till the end of the future value. Here the last known value is 30 June 2013= 21871. So for the Forecasting number of Trips for the first seven days of the July 2013 will be 21871.

Date	Rentals
------	---------

26 June 2013	26271
27 June 2013	21472
28 June 2013	24636
29 June 2013	26004
30 June 2013	21871
01 July 2013	21871
02 July 2013	21871
03 July 2013	21871
04 July 2013	21871
05 July 2013	21871
06 July 2013	21871
07 July 2013	21871



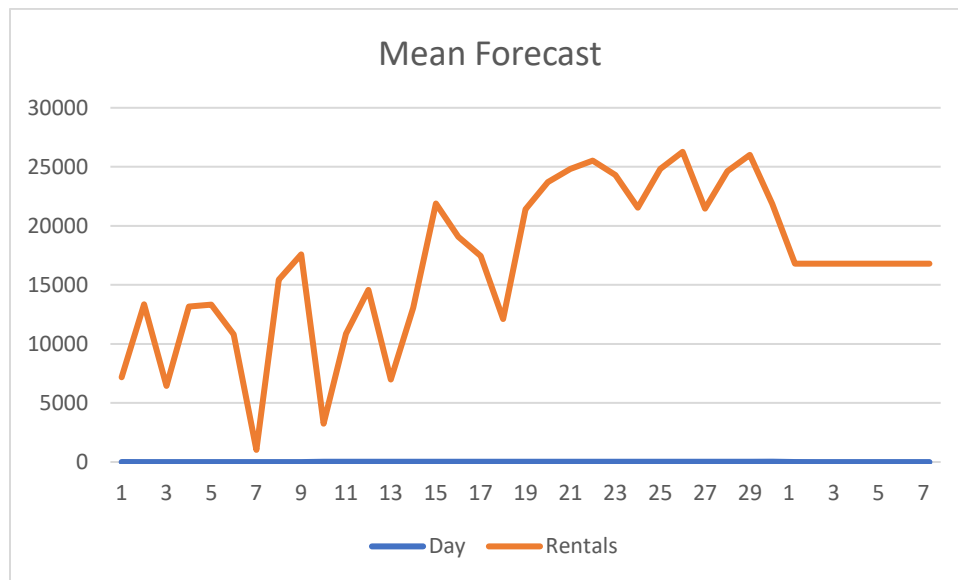
### **Mean forecast:**

In Mean or Average forecast method, in the series of data, the average value of the known observations (historical data) is considered as the predictive value till the end of the future values. Therefore, for the forecasting number of Trips for the first seven days of the July 2013 will be 16799 (around).

Date	Rentals
26 June 2013	26271
27 June 2013	21472
28 June 2013	24636
29 June 2013	26004
30 June 2013	21871
01 July 2013	16799



02 July 2013	16799
03 July 2013	16799
04 July 2013	16799
05 July 2013	16799
06 July 2013	16799
07 July 2013	16799

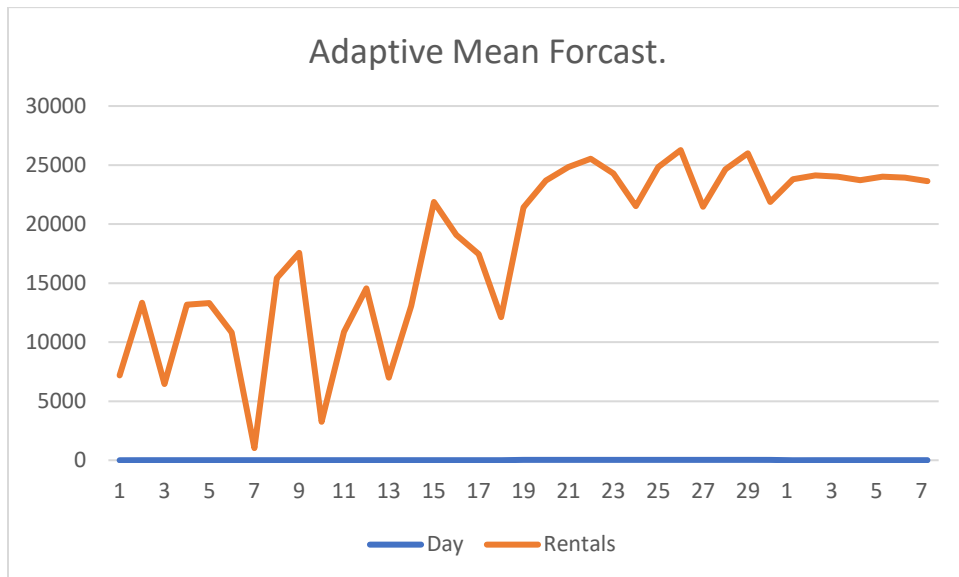


### **Adaptive mean forecast:**

An adaptive mean forecast or rolling average forecast, takes the average value of the last k number of values to predict the future value.

By taking the average of the last 07 observations.

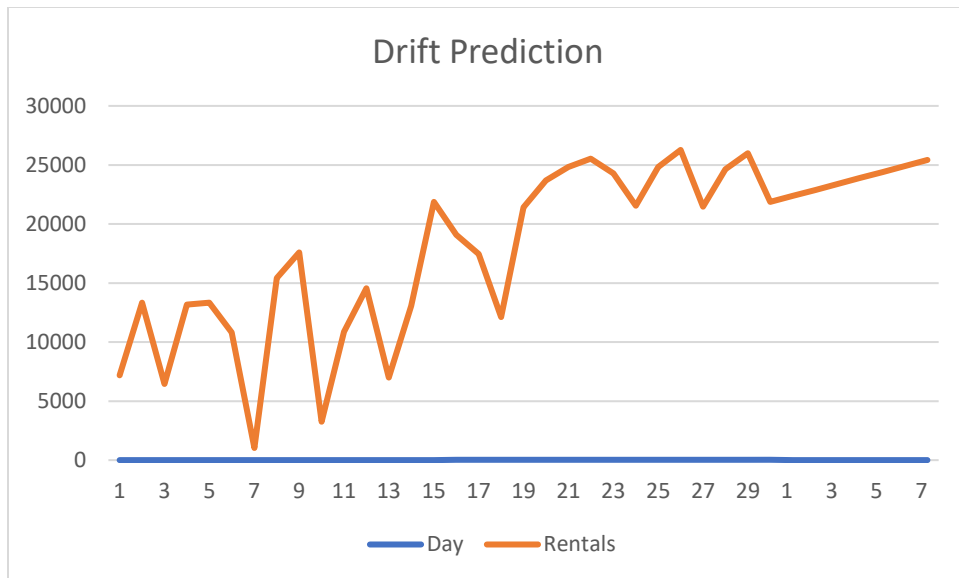
Date	Rentals
26 June 2013	26271
27 June 2013	21472
28 June 2013	24636
29 June 2013	26004
30 June 2013	21871
01 July 2013	23802
02 July 2013	24126
03 July 2013	24026
04 July 2013	23705
05 July 2013	24024
06 July 2013	23937
07 July 2013	23642



**A drift prediction:**

Answer: for the 01 July 2013 the formula was  $=\$B\$60+(\$B\$60-\$B\$31)/(30-1)$  and of the 07 July 2013 the formula was  $=\$B\$60+7*(\$B\$60-\$B\$31)/(30-1)$

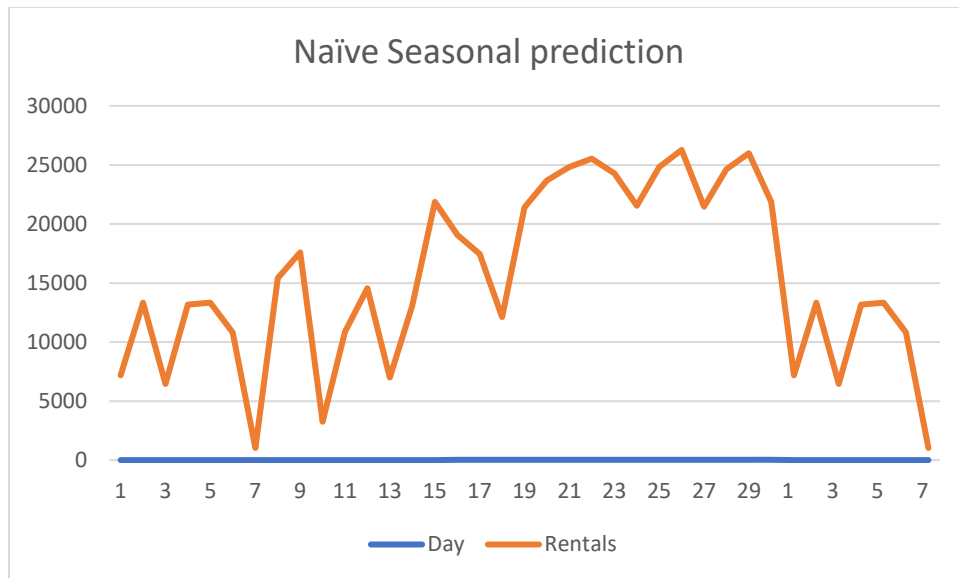
Date	Rentals
26 June 2013	26271
27 June 2013	21472
28 June 2013	24636
29 June 2013	26004
30 June 2013	21871
01 July 2013	22377
02 July 2013	22884
03 July 2013	23390
04 July 2013	23896
05 July 2013	24403
06 July 2013	24909
07 July 2013	25415



### **Naïve seasonal prediction:**

Answer: In data, where there is a visible seasonal pattern, it forecast by taking the data from the history of the same point (past) of the season.

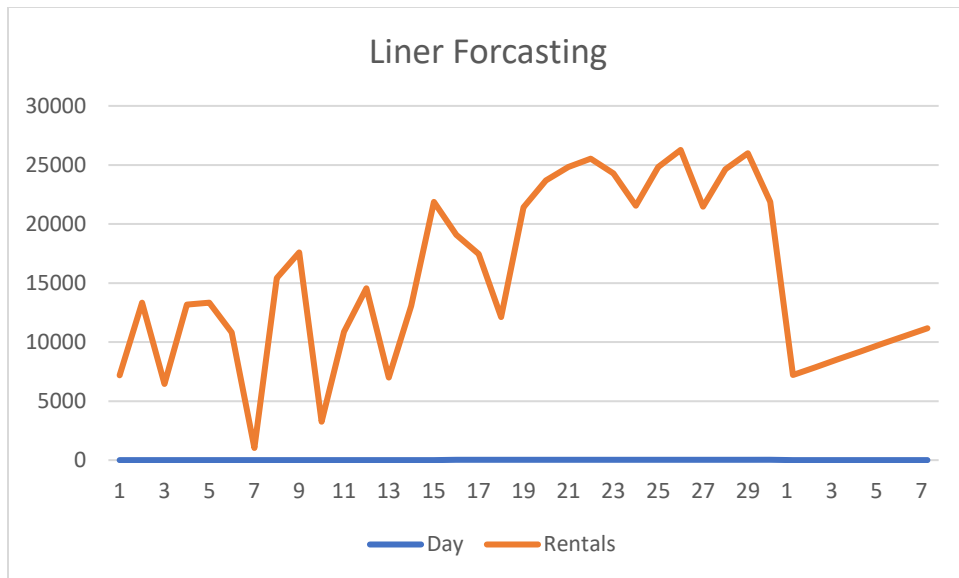
Day	Known Rentals	Day	Forecasting Rentals
01 June 2013	7187	01 July 2013	7187
02 June 2013	13344	02 July 2013	13344
03 June 2013	6446	03 July 2013	6446
04 June 2013	13176	04 July 2013	13176
05 June 2013	13322	05 July 2013	13322
06 June 2013	10813	06 July 2013	10813
07 June 2013	1028	07 July 2013	1028



### **Assignment 7:**

**A linear trend:** By using FORECAST.LINEAR(x, known\_y's, known\_x's) function.

Date	Rentals
26 June 2013	26271
27 June 2013	21472
28 June 2013	24636
29 June 2013	26004
30 June 2013	21871
01 July 2013	7209
02 July 2013	7870
03 July 2013	8532
04 July 2013	9193
05 July 2013	9854
06 July 2013	10516
07 July 2013	11177



An exponential trend:

1	7187
2	13344
3	6446
4	13176
5	13322
6	10813
7	1028
8	15433
9	17578
10	3247
11	10875
12	14563
13	6998
14	13058
15	21882
16	19088
17	17447
18	12109
19	21411
20	23697
21	24826
22	25522
23	24295
24	21534
25	24829

26	26271			
27	21472			
28	24636			
29	26004			
30	21871	21871	21871.00	21871.00
31		27031.45547	18502.75	35560.16
32		27693.27442	19164.53	36222.02
33		28355.09337	19826.28	36883.91
34		29016.91232	20487.99	37545.83
35		29678.73126	21149.66	38207.81
36		30340.55021	21811.27	38869.83
37		31002.36916	22472.81	39531.93
38		31664.1881	23134.29	40194.09

