

FINAL PROJECT

Bike Sharing Usage & Weather in Washington, D.C.



DATA OVERVIEW

Visualization using two datasets of shared bicycles in Washington, D.C:

Data set in hour contains 17379 observations and 17 variables.

Data set in day contains 731 observations and 16 variables.

PURPOSE OF RESEARCH

1

Observing the trend
of usage of bike
sharing in timeline in
Washington, D.C.

2

Defining the affects
of weather and
environmental
factors on bicycle
use in Washington,
D.C.

USED TOOLS FOR DATA ANALYSIS

Microsoft Excel

RStudio

Tableau

STATISTICAL VIEW

```

> str(hour)
'data.frame': 17379 obs. of 17 variables:
$ instant : int 1 2 3 4 5 6 7 8 9 10 ...
$ dteday : Factor w/ 731 levels "2011-01-01","2011-01-02",... 1 1 1 1 1 1 1 1 1 ...
$ season : int 1 1 1 1 1 1 1 1 1 ...
$ yr : int 0 0 0 0 0 0 0 0 0 ...
$ mnth : int 1 1 1 1 1 1 1 1 1 ...
$ hr : int 0 1 2 3 4 5 6 7 8 9 ...
$ holiday : int 0 0 0 0 0 0 0 0 0 ...
$ weekday : int 6 6 6 6 6 6 6 6 6 ...
$ workingday: int 0 0 0 0 0 0 0 0 0 ...
$ weathersit: int 1 1 1 1 2 1 1 1 1 ...
$ temp : num 0.24 0.22 0.22 0.24 0.24 0.24 0.22 0.2 0.24 0.32 ...
$ atemp : num 0.288 0.273 0.273 0.288 0.288 ...
$ hum : num 0.81 0.8 0.8 0.75 0.75 0.75 0.8 0.86 0.75 0.76 ...
$ windspeed : num 0 0 0 0 0 0.0896 0 0 0 ...
$ casual : int 3 8 5 3 0 0 2 1 1 8 ...
$ registered: int 13 32 27 10 1 1 0 2 7 6 ...
$ cnt : int 16 40 32 13 1 1 2 3 8 14 ...

> str(day)
'data.frame': 731 obs. of 16 variables:
$ instant : int 1 2 3 4 5 6 7 8 9 10 ...
$ dteday : Factor w/ 731 levels "1/1/11","1/1/12",... 1 23 45 51 53 55 57 59 61 3 ...
$ season : int 1 1 1 1 1 1 1 1 1 ...
$ yr : int 0 0 0 0 0 0 0 0 0 ...
$ mnth : int 1 1 1 1 1 1 1 1 1 ...
$ holiday : int 0 0 0 0 0 0 0 0 0 ...
$ weekday : int 6 0 1 2 3 4 5 6 0 1 ...
$ workingday: int 0 0 1 1 1 1 1 0 0 1 ...
$ weathersit: int 2 2 1 1 1 1 2 2 1 1 ...
$ temp : num 0.344 0.363 0.196 0.2 0.227 ...
$ atemp : num 0.364 0.354 0.189 0.212 0.229 ...
$ hum : num 0.806 0.696 0.437 0.59 0.437 ...
$ windspeed : num 0.16 0.249 0.248 0.16 0.187 ...
$ casual : int 331 131 120 108 82 88 148 68 54 41 ...
$ registered: int 654 670 1229 1454 1518 1518 1362 891 768 1280 ...
$ cnt : int 985 801 1349 1562 1600 1606 1510 959 822 1321 ...

```

```

> summary(hour)
   instant      dteday      season       yr      mnth      hr      holiday
Min.    : 1 2011-01-01: 24 Min.  :1.000  Min.  :0.0000  Min.  : 1.000  Min.  : 0.00  Min.  :0.00000
1st Qu.: 4346 2011-01-08: 24 1st Qu.:2.000  1st Qu.:0.0000  1st Qu.: 4.000  1st Qu.: 6.00  1st Qu.:0.00000
Median : 8690 2011-01-09: 24 Median :3.000  Median :1.0000  Median : 7.000  Median :12.00  Median :0.00000
Mean   : 8690 2011-01-10: 24 Mean   :2.502  Mean   :0.5026  Mean   : 6.538  Mean   :11.55  Mean   :0.02877
3rd Qu.:13034 2011-01-13: 24 3rd Qu.:3.000  3rd Qu.:1.0000  3rd Qu.:10.000 3rd Qu.:18.00  3rd Qu.:0.00000
Max.   :17379 2011-01-15: 24 Max.   :4.000  Max.   :1.0000  Max.   :12.000  Max.   :23.00  Max.   :1.00000
(Other) :17235

   weekday      workingday      weathersit      temp      atemp      hum      windspeed
Min.  :0.000  Min.  :0.0000  Min.  :1.000  Min.  :0.020  Min.  :0.0000  Min.  :0.0000  Min.  :0.0000
1st Qu.:1.000 1st Qu.:0.0000  1st Qu.:1.000  1st Qu.:0.340  1st Qu.:0.3333  1st Qu.:0.4800  1st Qu.:0.1045
Median :3.000  Median :1.0000  Median :1.000  Median :0.500  Median :0.4848  Median :0.6300  Median :0.1940
Mean   :3.004  Mean   :0.6827  Mean   :1.425  Mean   :0.497  Mean   :0.4758  Mean   :0.6272  Mean   :0.1901
3rd Qu.:5.000 3rd Qu.:1.0000  3rd Qu.:2.000  3rd Qu.:0.660  3rd Qu.:0.6212  3rd Qu.:0.7800  3rd Qu.:0.2537
Max.   :6.000  Max.   :1.0000  Max.   :4.000  Max.   :1.000  Max.   :1.0000  Max.   :1.0000  Max.   :0.8507

   casual      registered      cnt
Min.  : 0.00  Min.  : 0.0  Min.  : 1.0
1st Qu.: 4.00  1st Qu.: 34.0  1st Qu.: 40.0
Median :17.00  Median :115.0  Median :142.0
Mean   :35.68  Mean   :153.8  Mean   :189.5
3rd Qu.:48.00 3rd Qu.:220.0  3rd Qu.:281.0
Max.   :367.00  Max.   :886.0  Max.   :977.0

> summary(day)
   instant      dteday      season       yr      mnth      holiday      weekday      workingday
Min.  : 1.0  1/1/11 : 1  Min.  :1.000  Min.  :0.0000  Min.  : 1.00  Min.  :0.00000  Min.  :0.000  Min.  :0.000
1st Qu.:183.5 1/1/12 : 1  1st Qu.:2.000  1st Qu.:0.0000  1st Qu.: 4.00  1st Qu.:0.00000  1st Qu.:1.000  1st Qu.:0.000
Median :366.0 1/10/11: 1  Median :3.000  Median :1.0000  Median : 7.00  Median :0.00000  Median :3.000  Median :1.000
Mean   :366.0 1/10/12: 1  Mean   :2.497  Mean   :0.5007  Mean   : 6.52  Mean   :0.02873  Mean   :2.997  Mean   :0.684
3rd Qu.:548.5 1/11/11: 1  3rd Qu.:3.000  3rd Qu.:1.0000  3rd Qu.:10.00  3rd Qu.:0.00000  3rd Qu.:5.000  3rd Qu.:1.000
Max.   :731.0 1/11/12: 1  Max.   :4.000  Max.   :1.0000  Max.   :12.00  Max.   :1.00000  Max.   :6.000  Max.   :1.000
(Other):725

   weathersit      temp      atemp      hum      windspeed      casual      registered
Min.  :1.000  Min.  :0.05913  Min.  :0.07907  Min.  :0.0000  Min.  :0.02239  Min.  : 2.0  Min.  : 20
1st Qu.:1.000 1st Qu.:0.33708  1st Qu.:0.33784  1st Qu.:0.5200  1st Qu.:0.13495  1st Qu.: 315.5  1st Qu.:2497
Median :1.000  Median :0.49833  Median :0.48673  Median :0.6267  Median :0.18097  Median : 713.0  Median :3662
Mean   :1.395  Mean   :0.49538  Mean   :0.47435  Mean   :0.6279  Mean   :0.19049  Mean   : 848.2  Mean   :3656
3rd Qu.:2.000 3rd Qu.:0.65542  3rd Qu.:0.60860  3rd Qu.:0.7302  3rd Qu.:0.23321  3rd Qu.:1096.0  3rd Qu.:4776
Max.   :3.000  Max.   :0.86167  Max.   :0.84090  Max.   :0.9725  Max.   :0.50746  Max.   :3410.0  Max.   :6946

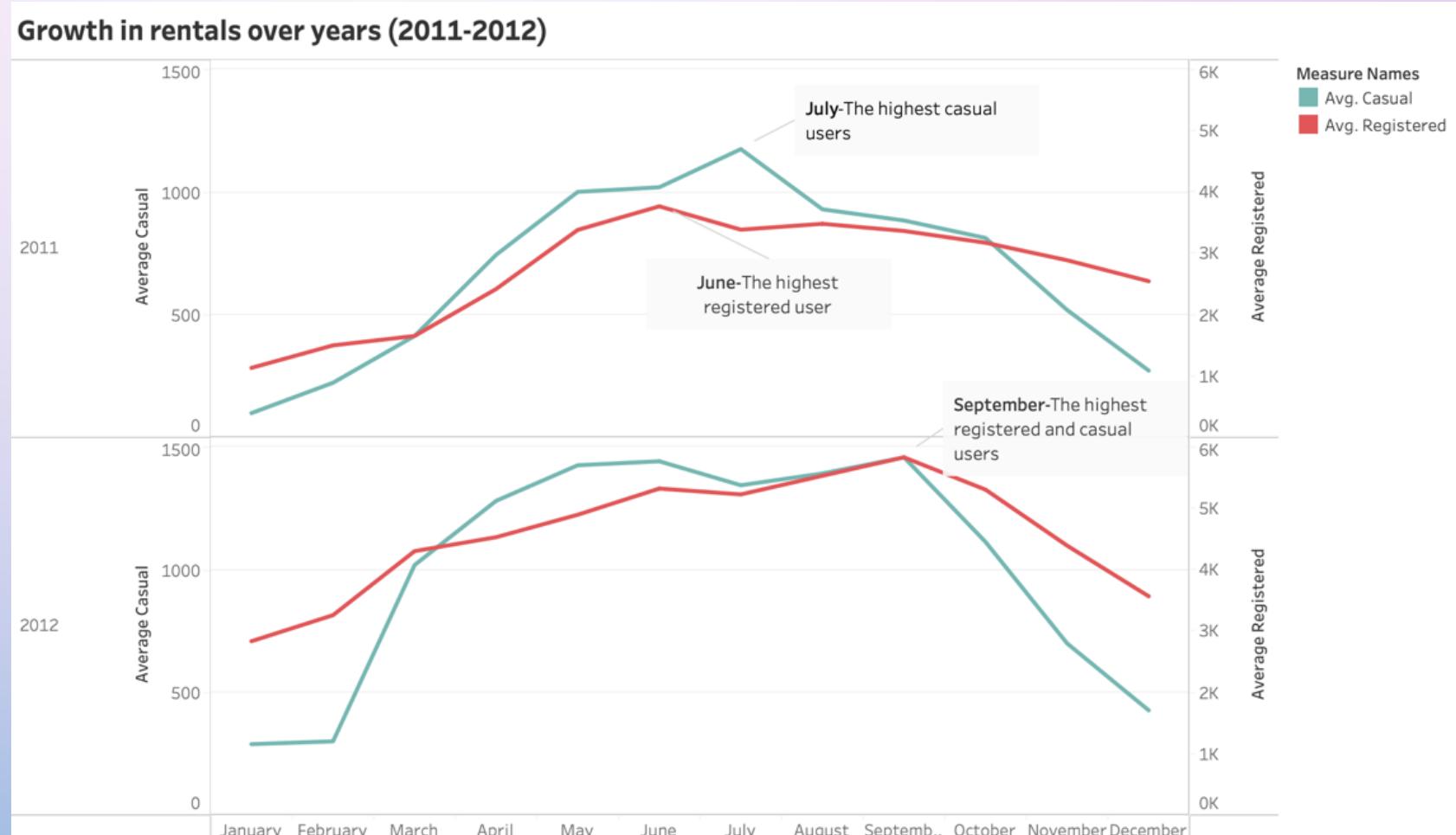
   cnt
Min.  : 22
1st Qu.:3152
Median :4548
Mean   :4504
3rd Qu.:5956
Max.   :8714

```

Outcome

Display Average casual and registered rentals over 2011-2012 years by months.

- The graph shows average casual and registered rentals for 2011 and 2012 by months.
- The highest casual user is in July 2011.
- The highest registered user is in June 2011.
- The highest casual and registered users are September 2012.

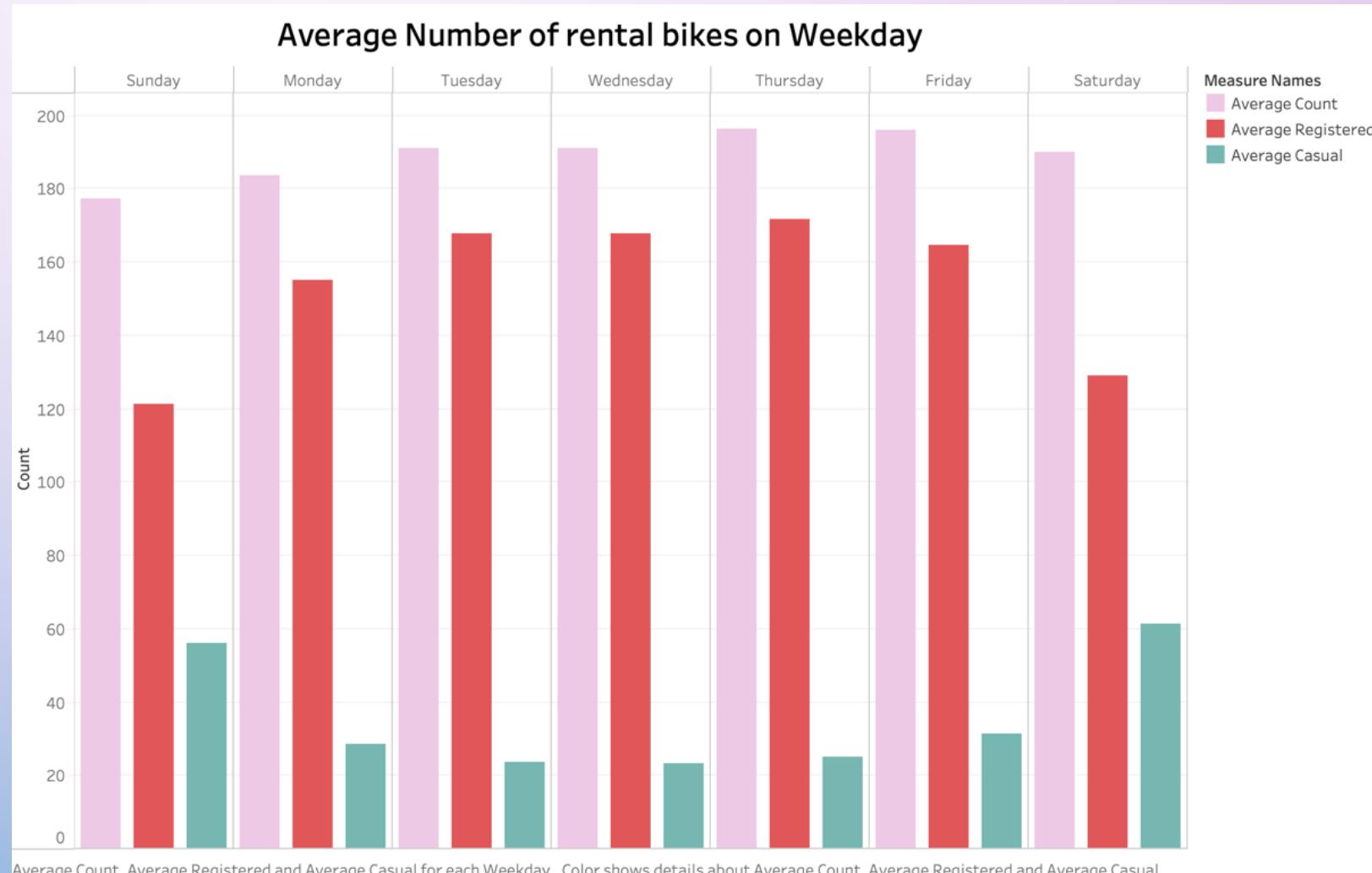


The trends of Avg. Casual and Avg. Registered for Dated Month broken down by Yr. Color shows details about Avg. Casual and Avg. Registered. The data is filtered on Mnth, which ranges from 1 to 12. The view is filtered on Yr, which keeps 2011 and 2012.

Outcome

Show comparison of average rentals by total, registered and casual on weekday.

- The graph shows Average bicycle rentals by total, registered and casual throughout weekday.
- Average total and registered demand are the highest level on Thursday while average casual demand doesn't show high demand.
- The high demand for casual users on average is observed for weekend at 60 counts.
- Average casual bicycle demand are low for between Monday and Friday and show the lowest demand on Wednesday at about 23 count.

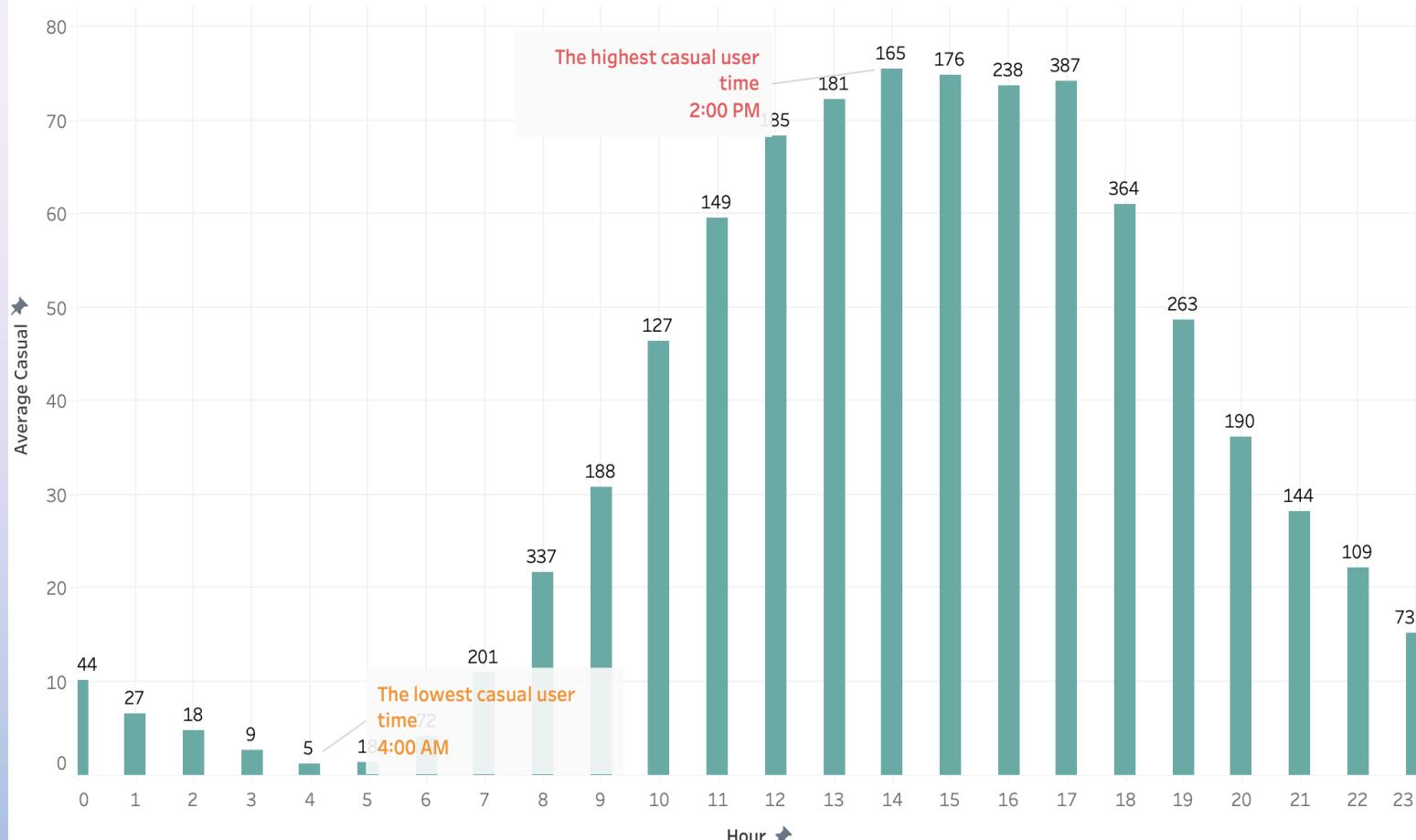


Outcome

Define distribution of Average registered and casual amount of bike rental over 24 hour.

- The graph shows distribution of average registered and casual bike rentals over 24 hour by help of bar chart.
- Lowest casual user is below 10 in 4:00 AM when average registered bike rentals also show the lowest level within 5 counts. Almost from 6:00 AM, average casual rentals started to increase and reach the highest level at 2:00 pm.
- Registered for bike rentals start from 6:00 am and reach the high registration level at 8:00 am; at 5:00 pm it shows actual amount with casual rentals in which it reach the highest registered numbers for rental bikes.

Average Registered and Casual users throughout 24 hour

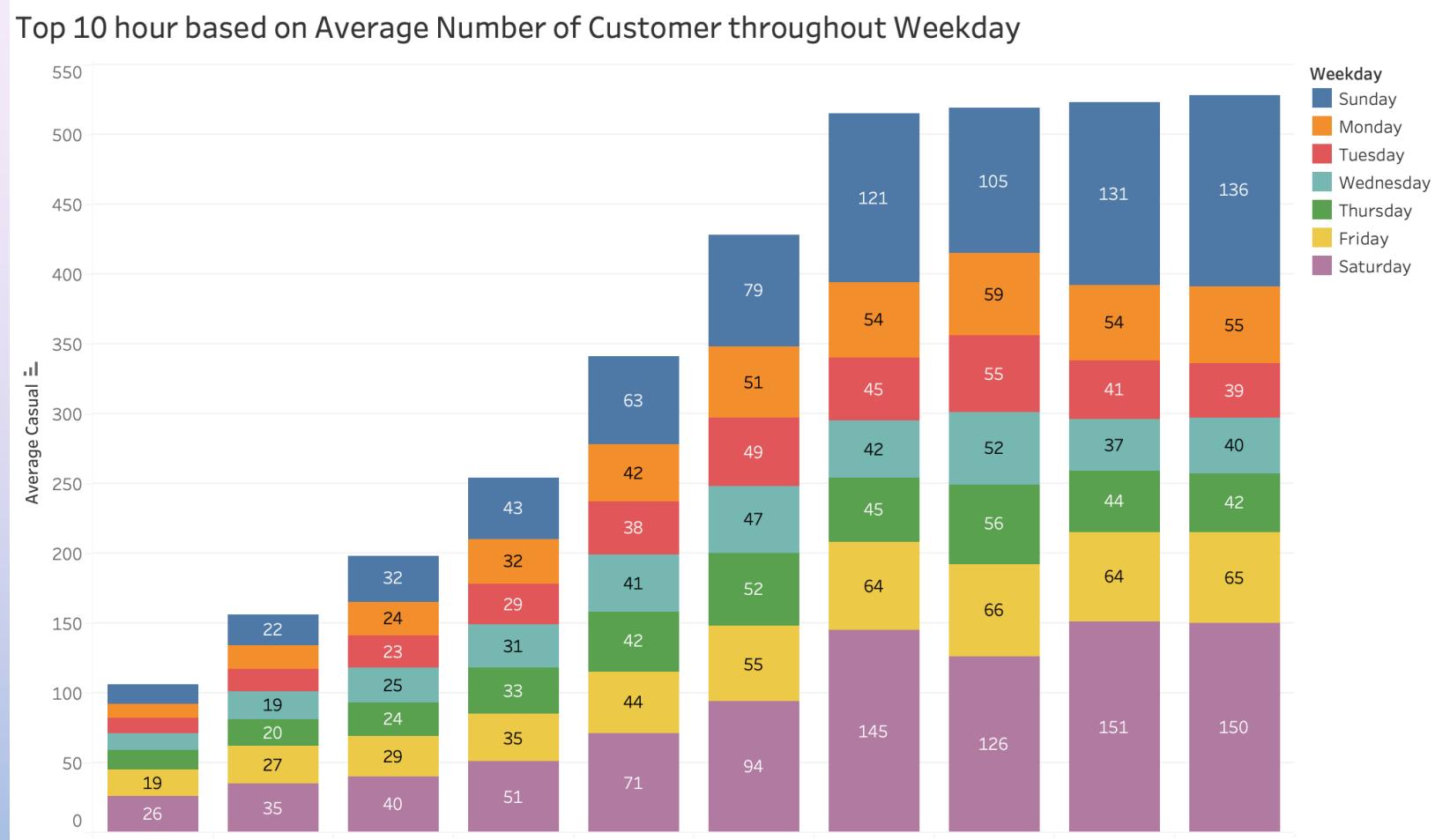


The plot of average of Casual (Hour(1).Csv) for Hr. The marks are labeled by average of Registered (Hour(1).Csv). The view is filtered on Hr and average of Casual (Hour(1).Csv). The Hr filter keeps non-Null values only. The average of Casual (Hour(1).Csv) filter keeps non-Null values only.

Outcome

Define top 10 hour based on Average casual number of bicycle rentals throughout Weekday.

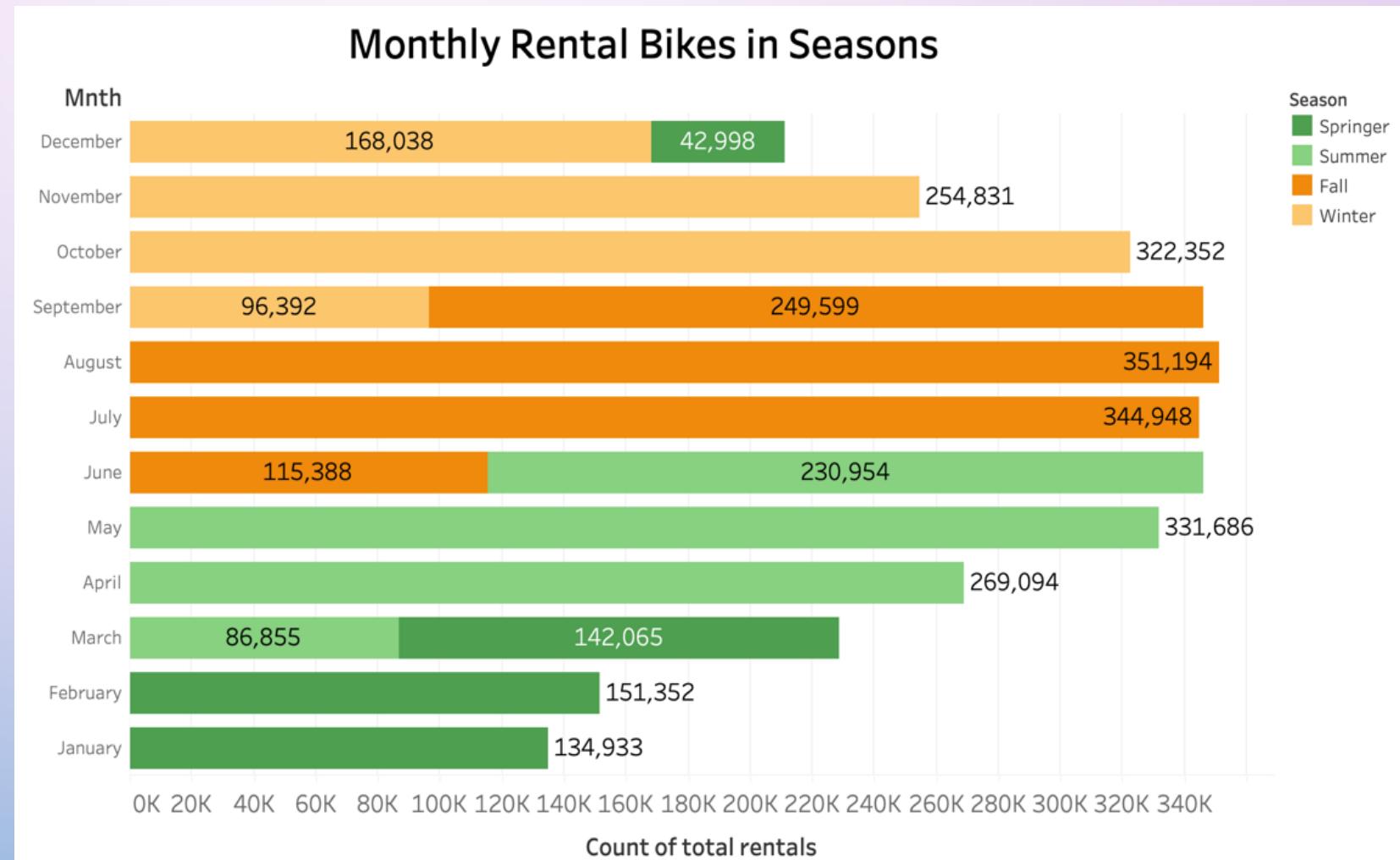
- The graph shows top 10 hours based on Average casual number of rental bike customer throughout Weekday.
- From bar chart, we can see the highest demanded time is 2:00 PM in which Saturday and Sunday are the highest demanded day within 150 and 136 counts while the lowest demand time by top hour is 11:00 PM in which Saturday and Friday users are 26 and 19 counts, respectively.
- On 03:00 PM Saturday and Sunday demand are almost same within 151 and 131 counts, respectively. From 02:00 PM until 05:00 PM, Friday demand doesn't change much while from 02:00 PM until 04:00 PM, Thursday demand are same count within 42-45.



Outcome

Predict rental bikes based on summary of total bike rentals according to season.

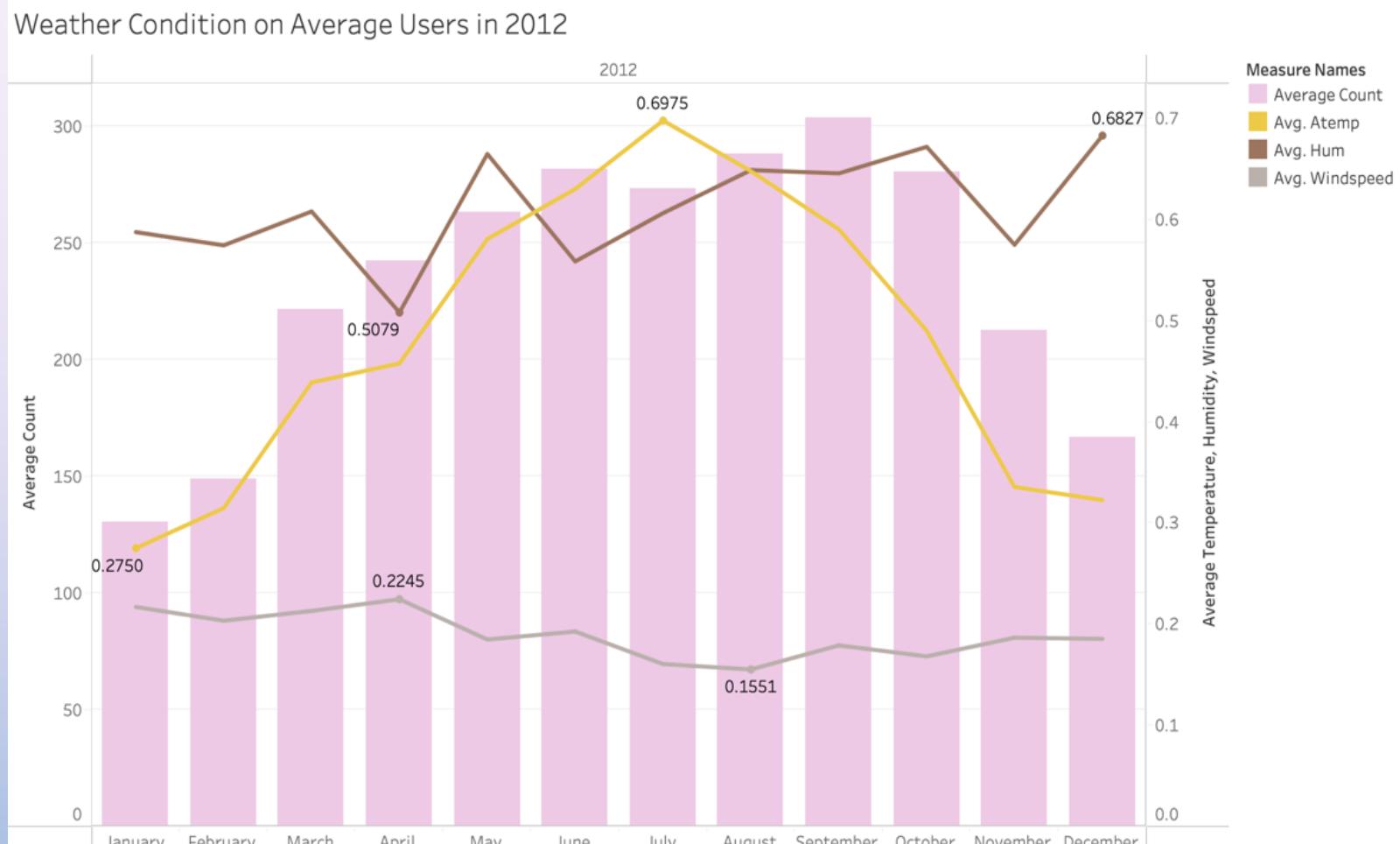
- The graph shows Monthly rental bikes according to seasons based on total summary of rentals.
- We can observe the highest total rentals in August within 351,194 counts in Fall season. Second high rentals happens on July in same season with 344,948 counts. First two months show the lowest total bicycle rentals which happen on Springer season.
- December and March are not appropriate months for rentals with lower demand in effect of Winter and early Summer period. Although that October is already considered winter season, it keeps high demand for bike rentals.



Outcome

Observe Average Count of bike rentals by effects of average actual temperature, humidity and windspeed in 2012 year.

- The graph shows average users for 2012 year by affection of weather condition such as average actual temperature, average humidity and average windspeed by months.
- The highest average count in September by reaching maximum level at 300.
- Maximum points for average actual temperature at 0.6975, average humidity at 0.6827 and average windspeed at 0.2245; minimum points like 0.2750, 0.5079, and 0.1551, respectively. When average count reach highest level in September, average temperature is almost 0.6, humidity is 0.65 and windspeed is about 0.18 which they create ideal weather condition for bike riders.
- January is seen for the lowest demanded month with high humidity and windspeed, low temperature which is not preferred by bike riders for rentals.



The trends of Average Count, Average Count, Avg. Atemp, Avg. Hum and Avg. Windspeed for Mnth broken down by Dteday Year. Color shows details about Average Count, Avg. Atemp, Avg. Hum and Avg. Windspeed. For pane Measure Values: The marks are labeled by Average Count, Avg. Atemp, Avg. Hum and Avg. Windspeed. The view is filtered on Dteday Year, which keeps 2012.

Outcome

Give prediction of Average Number of total Customer on four weather sit throughout months.

At first glance, we can think that heavy snow, rains and thunderstorms are the worst condition for bike rentals, however in January with light snow and light rain condition show the lowest average bike demand within 50 counts.



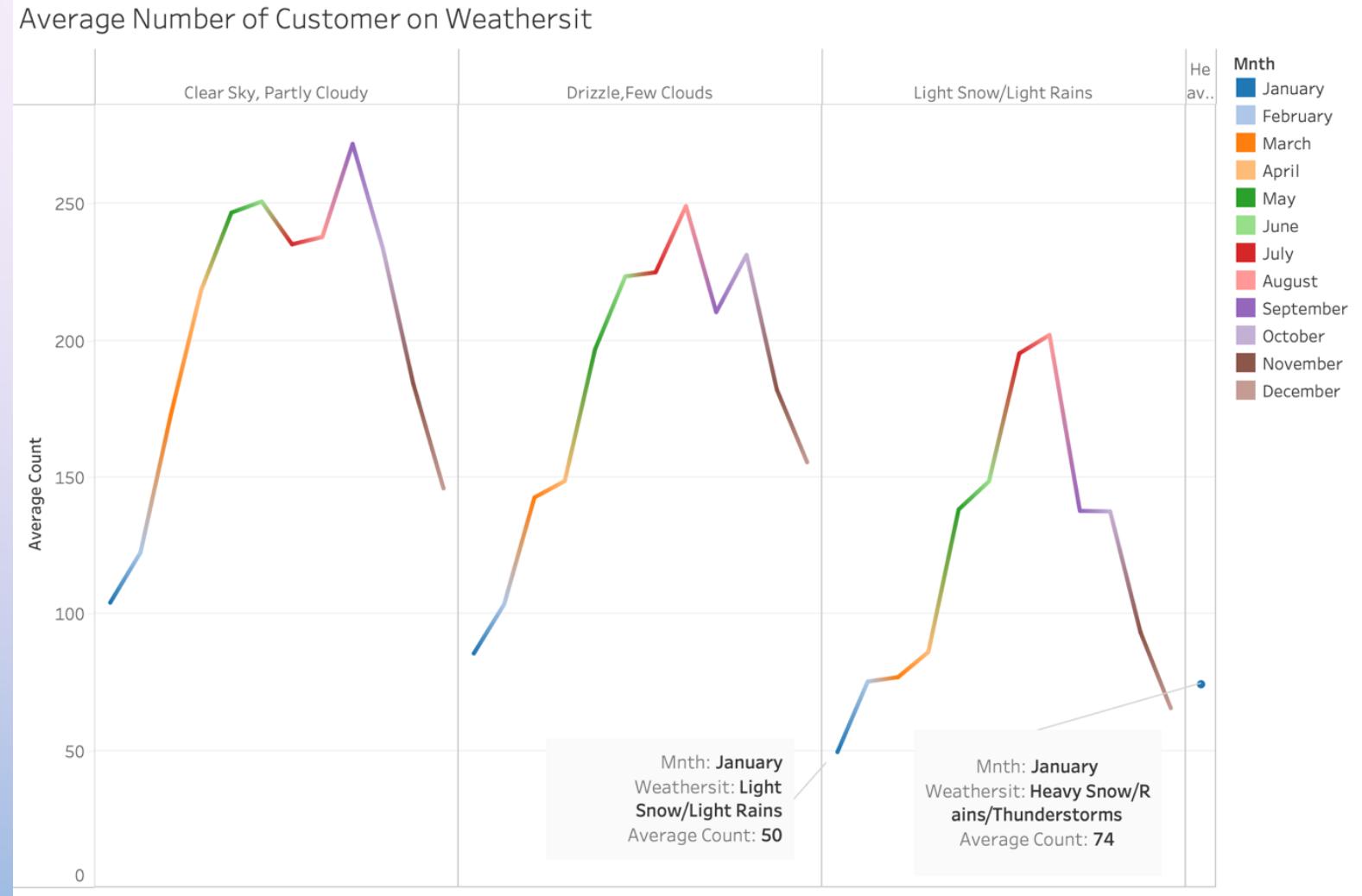
The graph shows average number of total bike rentals on four weathersit by months.



Drizzle and few cloudy weather condition are also suitable for bike rental in which August is considered as the highest demanded month.



High demand happens on clear sky and partly cloudy weather condition in which the highest demand are observed in September.



Summary

In a summary, we achieved our research purpose by defining solution for business questions by help of tools such as Microsoft Excel, RStudio, and Tableau. We observed difference between casual and registered bike rentals, affectionate of weather condition like humidity, temperature and windspeed; how to increase and decrease bike rentals in different seasons and weather sets; we observed hourly, weekly, monthly and yearly distribution of bicycle rentals and highlighted special points. At the end of each question, we defined which condition has been suitable for bike rentals on Washington D.C. location over 2011-2012.

REFERENCE

- Fanaee-T, H., & Gama, J. (2013, December 20). UCI Machine Learning Repository: Bike Sharing Dataset Data Set. UCI Machine Learning Repository. Retrieved March 19, 2022, from <https://archive.ics.uci.edu/ml/datasets/bike%2Bsharing%2Bdataset>
- Data sets are provided by instructors.

Thank you for
your
attention!

