Influence of differentfactors on the number of rented bikes in Washington DC, USA

Alexander Krotov (2727809)

Shuaiyujiao Tan (2728413)

Abhishek Das (2643742)

Anoosid Kiatkamolvong (2640909)

ABSTRACT

In this project, we analyzed bike sharing usage in Washington D.C. The data was collected from 2011 to 2012 by the bike sharing service called "Capital Bikeshare". The main purpose of this study is to find relationships between the number of rented bikes and various other factors which influence bike usage. All of this data, after carefully analysis can bring a lot of valuable insights which can be used to improve traffic systems and encourage people to use bicycles instead of motorized transportation.

INTRODUCTION

Bike sharing service is an automated process of renting and returning a bike via a network of bike stations throughout the city. There are a number of such services throughout the World. They allow a user to hire a bike from one location and return it to a bike station that is nearby their destination. These systems are becoming more and more popular around the world, mainly because it is a cost-effective and healthy way to get around the city. Moreover, in contrast with traditional transportation it is the most eco-friendly, especially in such a crowded places as modern megalopolises. Presently there are about 500 bike-sharing schemes running around the world consisting of over 500 thousand bicycles. In this era of datafication, the data generated from such systems provide a range of valuable information such as travel time, routes and the most popular places for cycling.

The following are the 5 questions that were answered in this analysis:

- What is the most popular season for bike users?
- Howweather conditions influence the number of rented bikes?
- What time is peak demand for bikes ?
- Are weekdays more popular than weekends?
- How does the temperature influence on the number of bikes being rented?

By understanding how temperature, part of the day, day of the week and the weather affect the number of rented bikes, bike sharing companies can take more actionable decisions in order to improve current services and maybe extend them. As a result, it could relieve the congestion in places where it will be taken into consideration.

DATASET

For the analysis we used two datasets in CSV format relating to the two-year historical log for years 2011 and 2012 from Capital Bikeshare system, Washington D.C., USA, available at http://capitalbikeshare.com/system-data. The data was aggregated hourly and daily and then added the corresponding weather and seasonal information. Weather information was extracted from http://www.freemeteo.com.

DATA PROCESSING AND MANIPULATION

The main tools that used to conduct the analysis were Python with some additional libraries such as Pandas and NumPy.

Our initial data consist of two files in CSV format. The data was checked for missing values using function Dropna but no missing values were found.
The temperature values in the original dataset were normalized. Consequently, in order to get appropriate outputs we had to convert our values into normal ones in Celsius using formulas which we were given with the data description.
Almost for all of our calculation used the Groupby method, also part of the Pandas library. This provided the opportunity to group our data and perform calculations for categories of the data.
In order to visualize our outputs, Matplotlib was used for plotting the data.

DISCUSSION

Most popular seasons for bike users

The first question we addressed was how seasons influence bike usage. It may seem obvious but our goal is to find it out quantitatively.
The column season contains the following quantitative designation: 1: spring, 2: summer, 3: fall, 4: winter.
First of all, we use cross tables for all our outputs in order to get the exact values. For the deep data understanding we use both overall and mean values for each season.
Table 1: Seasons and overall and mean values of shared bikes.

seasons	overall	mean
spring	4242	111
summer	4409	208
fall	4496	236
winter	4232	198

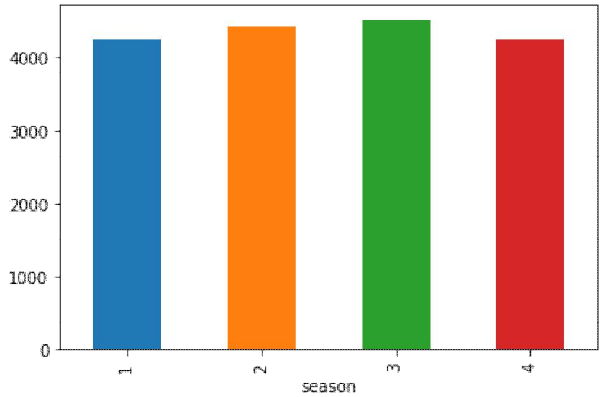

Figure 1: Seasons and overall amount of hired bikes.

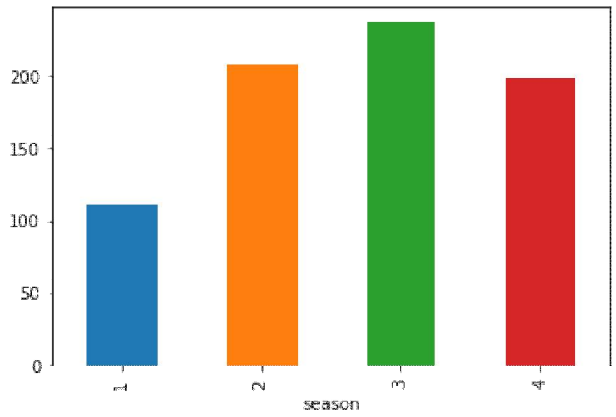

Figure 2: Seasons and mean value of hired bikes.

Despite the fact summer and winter are the most and least popular season for cyclists, our data provides the opposite information.
The crosstable and both histograms provide evidence that fall is the most popular season among those who are interested in hiring bikes. Whilst spring is, on the contrast, the least popular, according to mean values.

The impact of weather conditions on the no. of bikes being rented

As we saw on the previous part of the analysis, sometimes our longstanding concepts might go contrary with data. In the current part of the analysis, we would like to know, how weather affects the bike sharing.
Our data contains 4 weather patterns, which were extracted from the Freemeteosource :
1: Clear, Few clouds, Partly cloudy
2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

Table 2: Weathersit patterns and overall values of shared bikes.

Weathersit	Number of rented bikes
1	11413
2	4544
3	1419
4	3

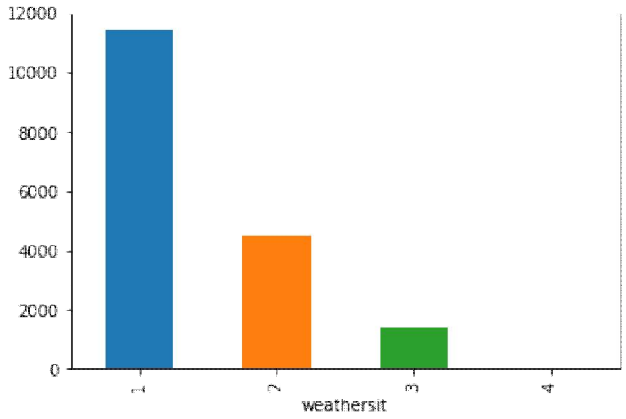

Figure 3: Weather and overall values of hired bikes.

Here we can see the strong dependence of the amount of rented bikes with weather.
Almost nobody uses bikes when it is rainy or snowy, and most people use shared service during warm and sunny days. In this case, it makes sense to calculate the correlation coefficient for 'weathersit' and 'cnt' variables, which shows us how strong the relationship between two variables.

Table 3: The correlation between weather and the number of rented bikes

	Weather	No. of rented bikes
Weather	1.000	-0.297
No. of rented bikes	-0.297	1.000

The R squared coefficient is -0.297391, so the relationship is negative. The negative signbefore the coefficient shows that our quantitative marking of weather patterns could be changed in order to change the relationship to a positive one.

What time of the day the demand of bikes peaks

Before we analyzed the relationship between weather and seasons, it seemed relatively obvious what was more popular and why. However, when it comes to recognizing which part of the day or particular period of time is higher in demand, the relationship was not so apparent and therefore the data helped us.

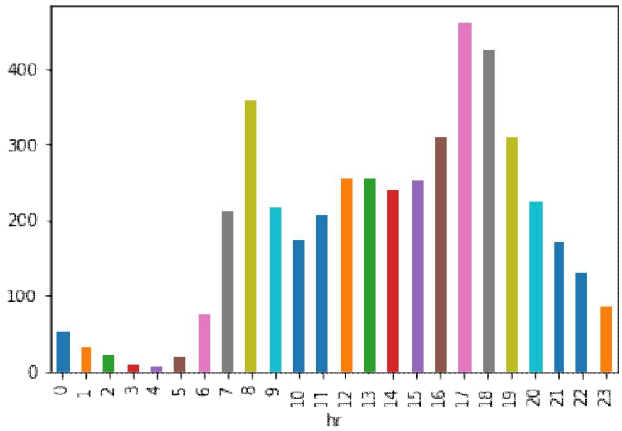

Figure 4: Hours and overall values of hired bikes.

Looking at the histogram, we can recognize two obvious peaks which mean rush hours. Consequently, the values of the number of bikes rented during that time is maximum. To be more precise, from seven to nine in the morning and from five to seven in the evening the bike sharing service is in higher demand.

On which days of the week the most bikes are rented

After analyzing the data, we saw the difference between the number of bikes rented was negligible irrespective of the

day of the week. The bike users remained almost the same, independent of the day of the week. The table below shows the days of the week and the average number of bike users on that day.

Table 4: The usage of bikes depends on the day of the week

Day of the week	Average number of users
0-Sunday	4334
1-Monday	4359
2-Tuesday	4576
3-Wednesday	4642
4-Thursday	4721
5-Friday	4601
6-Saturday	4521

In addition to the table above, the following bar graph shows the same results.

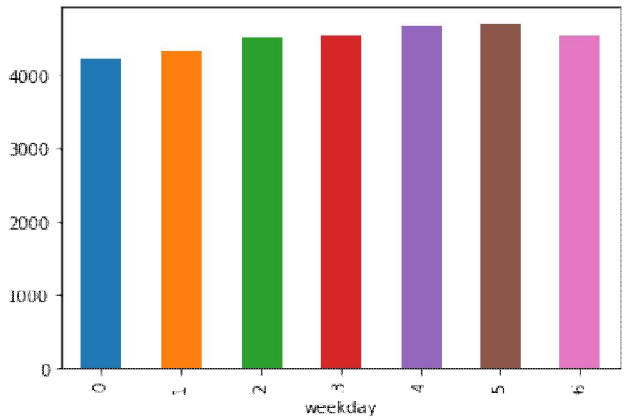

Figure 5: The overall number of rented bikes for each day of the week

From the graph above, it is evident that the bike users remained almost the same, demonstrating that the days of the week did not had any strong impact on the number of bikes rented.

<u>Relationship between the temperature and number of bikes rented</u>

In this part of our analysis we compared the temperature with the number of bikes being rented to see if there is any relation between them. After analysis, we can see that the number of bikes hired is positively correlated with temperature. The table below shows different temperature ranges and the corresponding number of rented bikes.

Table 5: The overall number of rented bikes for different temperature range

Normalized Temperature in Celsius	Number of Hired bikes
-10 - 0	433
1 - 10	1283
11 -20	1942
21 - 30	3057
31-40	2963

The table above shows that the most suitable temperature for bike renting is between 21-30 degrees Celsius. We can further see the relationship between these two variables in the following bar graph.

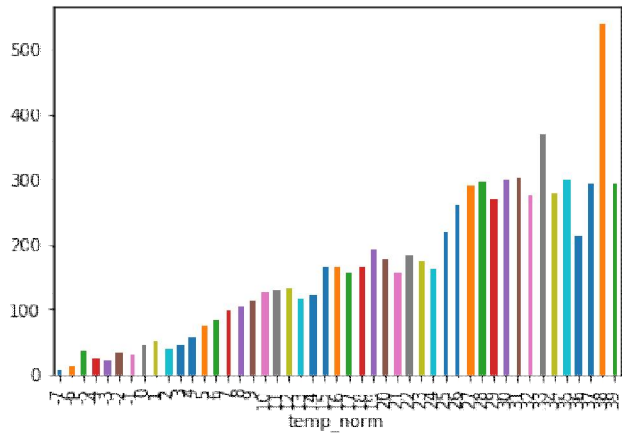

Figure 6: The overall number of rented bikes for temperature scale

The above graph clearly shows the strong positive correlation between the attributes, thus we can say that temperature has a significant impact on bike usage.

CONCLUSION

We have presented how various factors like weather conditions, temperature, hour of the day affect bike usage in Washington D.C. In contrast to our perception summer would be the most preferable season among cyclist we have discovered that more people ride bikes in fall than they do in summer. After analyzing how weather influences bike usage it became fairly obvious that more people tend to use bikes when it is clear than when it is rainy or snowy. When it came to analyzing the hours we found two peaks, rush hours, one of them in the morning and another one in the evening. Obviously, during that period of time people simply commute to and from work. The outputs of temperature analysis were more or less apparent. People are more likely to use bikes when it is warm rather than it is cold. In addition, we have analyzed weekdays and weekends in order to understand which part of the week is most popular among cyclist and which less. However, we have not managed to find anyrelationships between day of the week and the rent of the bikes.

While the data for our study is based on bike sharing service in Washington DC, the results could be fairly applicable to the bikesharing in general. Moreover, these results can be implemented to improve transportation system in the modern cities. Also, it can be considered as answers for many business questions in problem, related to already existing bike sharing services as well as to relieve running new bike sharing sturt ups.

REFERENCES:

1. Cited: Fanaee-T, Hadi, and Gama, Joao, "Event labeling combining ensemble detectors and background knowledge", Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg, DOI:10.1007/s13748-013-0040-3.
2. “The Weather.” The Weather. Weather Forecasts. Current Weather. Severe Weather Warnings | i-Weather.com, www.freemeteo.com./.
3. “Understanding Bike-Sharing Systems Using Data Mining: Exploring Activity Patterns.” NeuroImage, Academic Press, 6 Sept. 2011, www.sciencedirect.com/science/article/pii/S1877042811014388.
4. Transportation Research Record Journal Online, trrjournalonline.trb.org/doi/abs/10.3141/2143-20.
5. Raviv, Tal, et al. “Static Repositioning in a Bike-Sharing System: Models and Solution Approaches.” SpringerLink, Springer, 8 Jan. 2013, link.springer.com/article/10.1007/s13676-012-0017-6.
6. Motivate International, Inc. “System Data.” Capital Bikeshare, capitalbikeshare.com/system-data.
7. Kaghazgarian, Mark. “Bike Sharing in Washington D.C. Dataset.” RSNA Pneumonia Detection Challenge | Kaggle, 28 May 2018, www.kaggle.com/marklvl/bike-sharing-dataset.

APPENDIX

PARTICIPANTS

Alexander Krotov, 2727809

Shuaiyujiao Tan, 2728413

Anoosid Kiatkamolvong, 2640909

Abhishek Das, 2643742

PERSONAL CONTRIBUTIONS

Alexander Krotov (2727086)

I worked various segments of the code. This included the data grouping, as well as the data manipulation. This allowed me to try different options to see what worked best. I also focused on some specific parts of the data visualization, grouping, transforming and computing the average and overall values which formed a crucial part of our analysis. Individually was written the following report sections: Discussion of Results, first 3 analysis and Data processing and manipulation.

Shuaiyujiao Tan (2728413)

I worked on the code vizualization collaboratively. This allowed me to try different options to see how different libraries of Python work. I also Shared Contribution to the final report includes sections: Abstract and Introduction. Before that I learn a lot of related information about the project topic.

Abhishek Das (2643742)

I worked on the examining data on the missing values together with other group members. As well as this, I examined the outputs in order to find relationship between them. According to the outputs, I wrote individually Discussion (4-th and 5-th parts) and took part in Conclusion writing parts of the paper.

Anoosid Kiatkamolvong (2640909)

I have check the data for inappropriate values. Took part in visualization process and transformation of data. Also, wrote the initial version of conclusion which was improved within the team.

We are lucky to claim that all of our work was done altogether. By operating with data we were able to share our experience with programming. None of us were an expert in Python, working in the team allowed us the opportunity to learn how to operate with data more effectively and quicker. The writing of the report initially was divided into chunks but finally was worked on and improved by the team. Consequently, this project helped us not only to write code on Python but also develop team working skills, which is inevitable skill for any data scientist.