SEOUL BIKE SHARING DEMAND: EXPLORATORY DATA ANALYSIS

Introduction:

In a span of few decades, the sharing of bicycle system has seen enamours growth. This system is a recently developed transportation system which provides people with bicycle for common use. Bicycle system provides user to rent a bike from one docking station, where user can ride and then return in another docking station. Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Many countries have bike sharing system, such as Ddareungi is a bike sharing system in South Korea, which started in the year 2015, known as Seoul bike in English. It was started to overcome issues like greater oil prices, congestion in traffic and pollution in the environment and to develop a healthy environment for citizen of Seoul to live. With the advent of smart technology and convenience, the use of Rental bike is increasing every day. So, there is a need to manage the bike rental demand and manage the continuous and convenient service for the users. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes. This study proposes a data mining-based approach including weather data to predict whole city public bike demand.

Information about the dataset:

The 'Seoul Bike Sharing Demand' dataset has been obtained from the UCI Archives. The dataset contains count of public bikes rented at each hour in Seoul Bike sharing System daily from December 2017 to November 2018 with the corresponding Weather data and information about Holidays. The dataset consists of 8760 data points with 14 attributes, out of which 10 are numeric attributes and the rest are categorical features.

The dataset contains information on the following attributes:

Date : year-month-day
Rented Bike count - Count of bikes rented at each hour
Hour - Hour of the day
Temperature-Temperature in Celsius
Humidity - %
Windspeed - m/s
Visibility - 10m
Dew point temperature - Celsius
Solar radiation - MJ/m2
Rainfall - mm
Snowfall - cm
Seasons - Winter, Spring, Summer, Autumn
Holiday - Holiday/No holiday
Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

Any form of transportation mainly depends upon the Climatic conditions, hence we study the pattern of rentals for the corresponding weather information such as Temperature, Humidity, Wind-speed, Visibility, Dew point temperature, rainfall, and snowfall. We also study the effect of Holidays (official workday and holiday) and Functional days (Functional or Non-functional) of rental bike system. For the purpose of visual analysis, we have employed Kernel Density Estimate (KDE) plots, relplots and catplots.

Patterns of Bike Rentals:

1. The KDE plot across different seasons represents that the use of rental bikes is the least during winter season and highest during the summer season. This shows the effect of weather conditions in rental bike use.
2. The relplot between 'temperature' and count of rental bikes is indicative of the fact that there is a positive correlation between temperature and count of rentals. In fact, when the temperature drops below -10 degrees, less than 500 bikes are rented. Most bikes are rented when the temperature is between 20-30 degrees.
3. We observe that humidity rainfall as well as snowfall plays a role in the pattern of rentals. Maximum bikes are rented when the humidity was between 40-60% and an increase in rainfall causes a marked decrease in the number of rentals. In similar lines, increase in snowfall also causes the rentals to decrease. We also observe that when there is less visibility, bikes are rented less frequently. However, the effect of wind-speed and dew point temperature is not much on the frequency of rentals.
4. The catplot across the hour of the day represents the rental bike usage is high during hour 18 and less during hour 4. In fact, the rentals are highest at 7-8 am the morning and the rentals again peak at 6-7pm in the evening. This suggests that people use these rental bikes to commute to and from their places of work. The rental starts gradually increasing from 10am to 6pm and thereafter it starts decreasing. On an average, less than 500 bikes are rented between 4-5am and more than 2500 bikes are rented on an average at 8 hours and between 17 and 21 hours. As seen from the catplots the rental bike use has a strong time component associated with it.
5. The rental bike use is very high during the working day or no holiday compared to holidays, which implies the working days influence on the rental bike usage. This may be because people use rental bikes in a regular way for their work commute. The rentals are negligible during the winter holidays. Rentals are more or less similar on the holidays as far as hour of the day is considered. Comparatively, greater number of rentals is observed during 8am to 6pm, even on the holidays. These relations across different time components are used to predict the rental bike count at each hour more effectively.