**Bike Sharing Prediction**

The data is about the bike-sharing program and was hosted in Kaggle.

The data is hourly rental data spanning two years. The training set contains of the first 19 days of each month, while the test set is the 20th to the end of the month. One must predict the total count of bikes rented during each hour covered by the test set, using only information available prior to the rental period.

**About Dataset**

Overview Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. Currently, there are over 500 bike-sharing programs around the world.

**This notebook explains how we can go about explore and prepare data for model building. The notebook is structured in the following way**

* About Dataset
* Data Summary
* Feature Engineering
* Missing Value Analysis
* Outlier Analysis
* Correlation Analysis
* Visualizing Distribution Of Data
* Visualizing Count Vs (Month,Season,Hour,Weekday,Usertype)
* Filling 0's In Windspeed Using Random Forest
* Linear Regression Model
* Regularization Models
* Ensemble Models

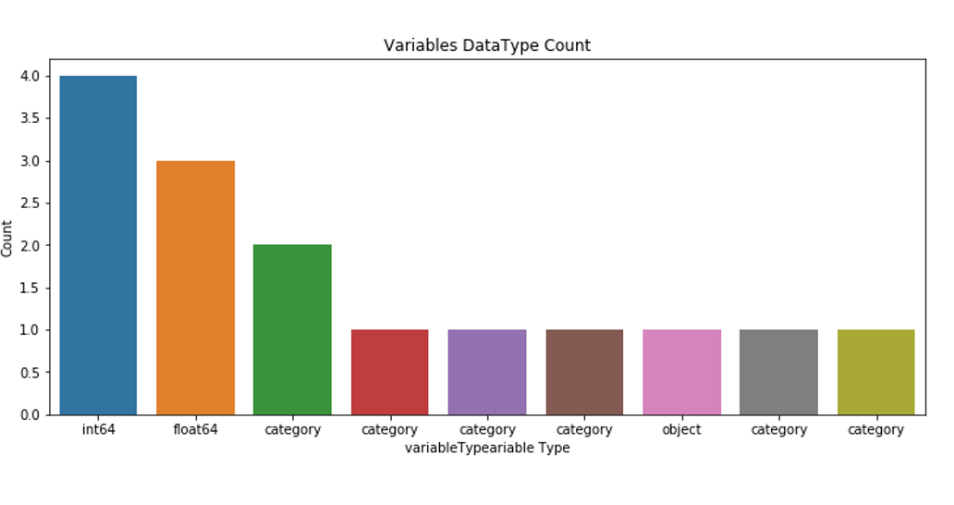
**Attribute Information:**

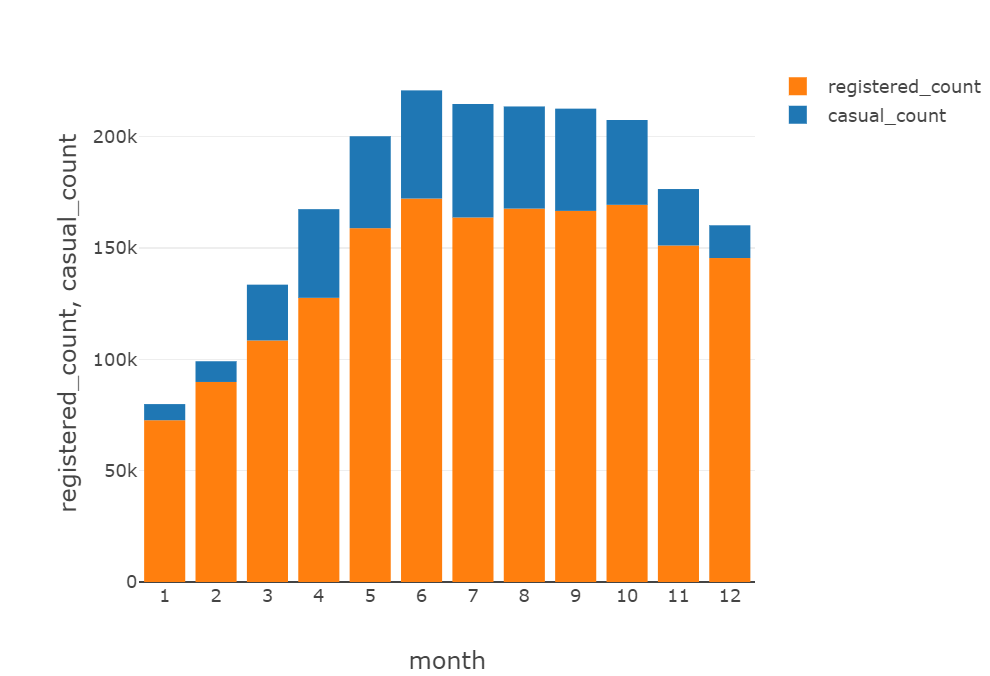
1. datetime - hourly date + timestamp
2. season - 1 = spring, 2 = summer, 3 = fall, 4 = winter
3. holiday - whether the day is considered a holiday
4. workingday - whether the day is neither a weekend nor holiday
5. weather - 1: Clear, Few clouds, Partly cloudy, Partly cloudy 2: Mist + Cloudy, Mist + 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
6. temp - temperature in Celsius
7. atemp - "feels like" temperature in Celsius
8. humidity - relative humidity
9. windspeed - wind speed
10. casual - number of non-registered user rentals initiated
11. registered - number of registered user rentals initiated
12. count - number of total rentals

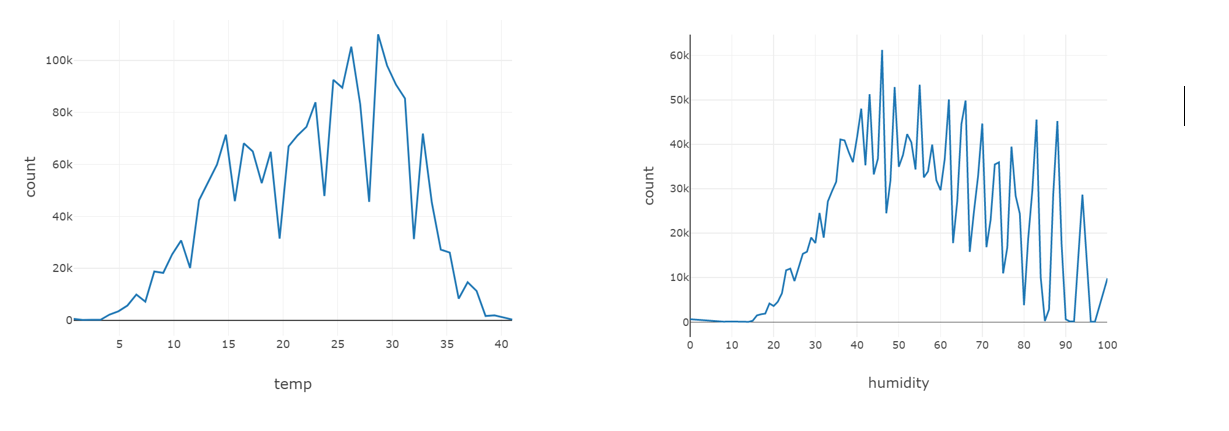
#### **Feature Engineering:**

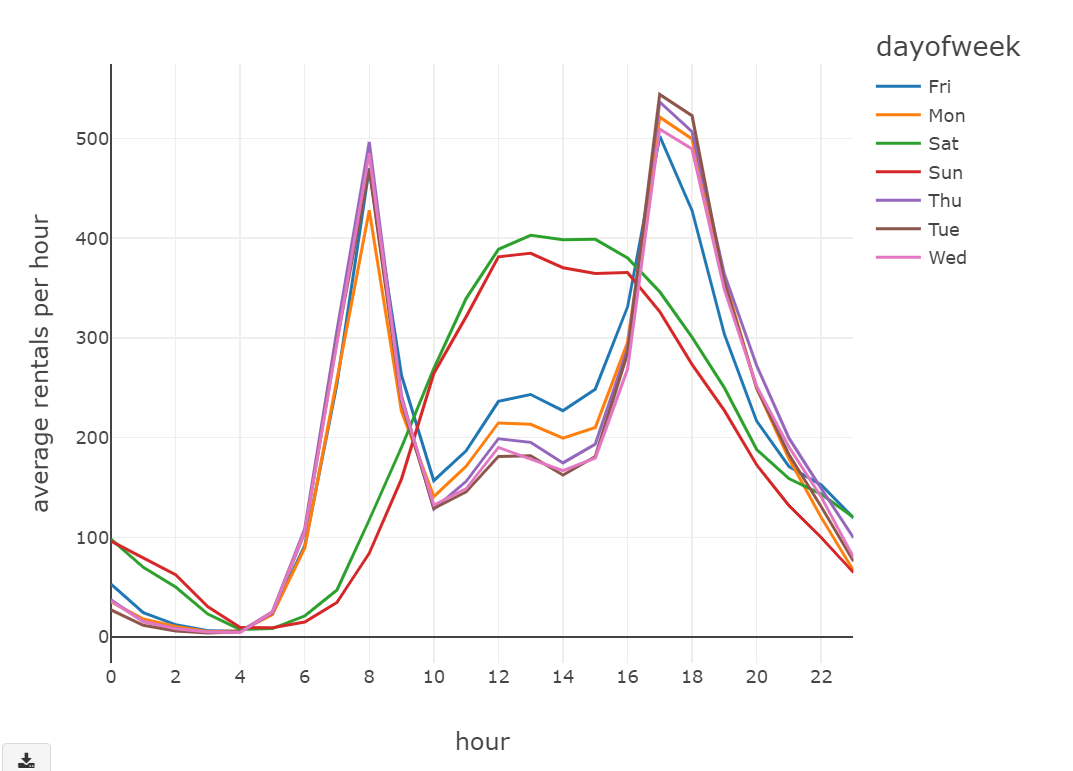
As we see from the above results, the columns "season", "holiday", "workingday" and "weather" should be of "categorical" data type.But the current data type is "int" for those columns. Let us transform the dataset in the following ways so that we can get started up with our EDA

Create new columns "date,"hour","weekDay","month" from "datetime" column. Coerce the datatype of "season","holiday","workingday" and weather to category. Drop the datetime column as we already extracted useful features from it. Creating New Columns From "Datetime" Column





* From the above plot we can see that for all the 12 months, number of registered bike rentals are very high compared to casual bike rentals. We have the least bike rentals in the month of January "79884" and maximum in the month of June "220733". Overall, bike rentals increased gradually until June and then there is a slight decrease in the further months
* This plot shows the variation of bike rentals at different temperatures. The rentals are very less at the extreme temperatures (very high and very low). The rentals are high when the temperature is between 25 - 30 degree celsius.
* This plot shows the variation of bike rentals at different humidity conditions.The rental count is very high at an humidity value of '46' and it keeps decreasing with slight fluctuations in between with increase in humidity value.



* The below plot shows the average rentals per hour for each day. Almost all the days have high rentals at 8:00 AM in the morning and during 5:00 PM in the evening. The rentals are low for all weekdays during the noon hours excluding the weekends, Saturday and Sunday have rental count average around 400 during noon. Apart from noon hours, saturday and sunday have less rentals during the morning and evening hours.
* The above plot shows the average rentals per hour at different seasons. The Fall season has more rentals followed by summer, winter and spring.
* 31% of the rentals are when the climate is clear and partly cloudy. Least rentals are when there is snow, rain and thunderstroms. 25% rentals are during heavy rain and thunderstorms.

