**Objective : predict Demand in a bike sharing program**

**Step-1**: First imported important libraries

* Numpy for calculation
* Pandas for reading and manipulating data
* Matplotlib and seaborn as visualization

**Step-2:**

Read the data check head and info of the data.

Dropped some unnecessary columns.

Check null values

**Step-3:**

Check histogram of every variable and visualize data

For continuous variable eg. temperature, atemp, humidity, windspeed used scatter plot and For categorical features eg. season, holiday, working day , month , hour used bar plot.

plt.subplot(3,3,1)

plt.title('Average Demand across season')

cat\_list = bikes\_prep['season'].unique()

cat\_average = bikes\_prep.groupby('season').mean()['demand']

c = ['g','m','b','r']

plt.bar(cat\_list,cat\_average,color = c)

Some of findings were

* Demand is not normally distributed Temperature
* Demand appears to have direct correlation with the plot for temp, temp and atemp appear almost identical.
* Humidity and windspeed affect demand but need more statistical analysis we can drop them
* Variation in demand based on **Season,Month,Holiday,Hour,weather.**
* No significant change due to weekday or working day year wise growth pattern not considered due to limited number of years

**Step-4:**

After all this check outliers.

Check multicollinearity using correlation matrix

Now check modified data.

Checked autocorrelation solving autocorrelation by preparing lag features.

Made demand normally distributed applying log transformation

Create dummy variable using pd.get\_dummies

**Step-5:**

Split train and test data

And fit our data using linear regression.

And check the prediction accuracy using rmse and rmsle