**Bike Sharing Sales Review Increasing Plan**

Project Proposal: Increase review for Bike Sharing business

1. **What is the Business about?**

A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short-term basis for a price or free.

1. **What kind of systems are there for bike sharing?**

A US bike-sharing provider BoomBikes has recently suffered considerable dips in their revenues due to the ongoing Corona pandemic. The company is finding it very difficult to sustain in the current market scenario. So, it has decided to come up with a mindful business plan to be able to accelerate its revenue as soon as the ongoing lockdown comes to an end, and the economy restores to a healthy state.

In such an attempt, BoomBikes aspires to understand the demand for shared bikes among the people after this ongoing quarantine situation ends across the nation due to Covid-19. They have planned this to prepare themselves to cater to the people's needs once the situation gets better all around and stand out from other service providers and make huge profits.

1. **Problem Statement and Business Use case:**

A US bike-sharing provider BoomBikes has recently suffered considerable dips in their revenues. They have contracted a consulting company to understand the factors on which the demand for these shared bikes depends. Specifically, they want to understand the factors affecting the demand for these shared bikes in the American market. The company wants to know:

* Which variables are significant in predicting the demand for shared bikes?
* How well those variables describe the bike demands

Required to model the demand for shared bikes with the available independent variables. It will be used by the management to understand how exactly the demands vary with different features. They can accordingly manipulate the business strategy to meet the demand levels and meet the customer's expectations. Further, the model will be a good way for management to understand the demand dynamics of a new market.

1. **Data Science Workflow:**
2. We want to measure the count of bike shared on a given day based on weather conditions and other factors. The data provided has factors like:

RangeIndex: 730 entries, 0 to 729

Data columns (total 16 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 instant 730 non-null int64

1 dteday 730 non-null object

2 season 730 non-null int64

3 yr 730 non-null int64

4 mnth 730 non-null int64

5 holiday 730 non-null int64

6 weekday 730 non-null int64

7 workingday 730 non-null int64

8 weathersit 730 non-null int64

9 temp 730 non-null float64

10 atemp 730 non-null float64

11 hum 730 non-null float64

12 windspeed 730 non-null float64

13 casual 730 non-null int64

14 registered 730 non-null int64

15 cnt 730 non-null int64

dtypes: float64(4), int64(11), object(1)

1. Data shared with us is collected over a span of two years across different seasons. Want to build Single linear regression model to predict demand. We want to analyse the different factor and understand which factor would be impact bike demands.
2. First solution is to identify which factors are important in describing bike demands. Second solution want to deliver is to build a model to predict demand in future based on these attributes.

* Bike Rentals are more during the Fall season and then in summer
* Bike Rentals are more in partly cloudy weather
* Bike rentals are more in the year 2019 compared to 2018
* Bike Rentals are observed at higher temperatures
* Bike rentals more at high humidity
* Bike Rentals are more during the Fall season and then in summer
* Bike Rentals are more in the year 2019 compared to 2018
* Bike Rentals are more in partly cloudy weather
* Bike Rentals are more on Saturday, Wednesday and Thursday

1. List down some EDA questions

* Season - We can notice a positive trend in the number of customers in 2 - Summer, 3 - Fall and 4 - Winter seasons
* Year - The overall business shows a increasing trend in their user base year on year
* Month - Similar to the season trend, there is a postive trend in the months of summer, fall and winter.
* Holiday: On holidays, the users show a wider spread in the counts. On normal days, the users are more than holidays
* Weekday: Weekdays or weekends do not show any specific trend here.
* Weather sit: Clearer weathers show a postive trend in the number of bike users - 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

1. Analytics Models:

* Descriptive Analytics
* Descriptive Insight
* Predictive analytics
* Prescriptive analytics
* Bike Rentals are more during the Fall season and then in summer
* Bike Rentals are more in partly cloudy weather
* Bike rentals are more in the year 2019 compared to 2018
* Bike Rentals are observed at higher temperatures
* Bike rentals more at high humidity

1. This will help to increase revenue provide better services to customer and reduce and sales loss.
2. Customers and Revenue stream will be impacted

Data Collection:

* dataset
* schema what kind of fields it has and the variable.
* Size of data
* file is csv
* data dictionary

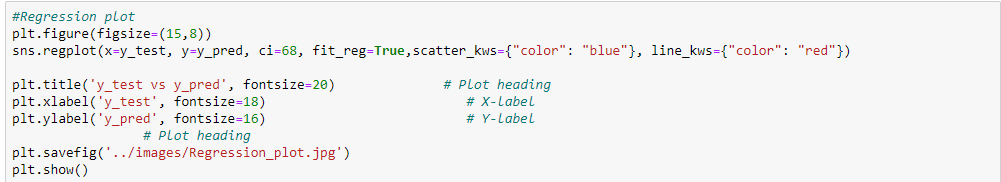
Please refer the Jupyter note book for Data related details.

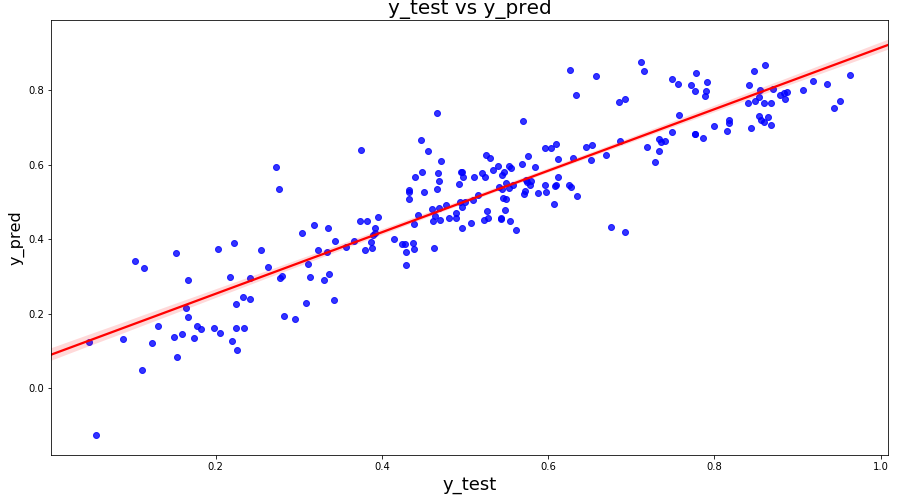
Data Pre-processing, Preparation and Feature Engineering:

* Identifying data quality
  + Check for data inconsistency, all data types are correct and no encoding issues.
* Handle Missing Values
* Treating categorical Values
* Treating outlier
* Do scaling
* Feature engineering

Machine Learning and Model selection:

* Multiple linear regression model
* ridge regression, lasso regression, stochastic gradient regressor.





We can see that the equation of our best fitted line is:

**count=0.4914×temp+0.0916×September+0.0645×Saturday+0.0527×summer+0.0970×winter+0.2334×Year +0.0566×workingday−0.03041×lightsnow−0.0786×mistcloudy−0.065×spring**

**Final Result Comparison between Train model and Test:**

* **Train R^2 : 0.826**
* **Train Adjusted R^2 : 0.82**
* **Test R^2: 0.8115**
* **Test Adjusted R^2: 0.790564**
* **Difference in R^2 between train and test: 1.5%**
* **Difference in adjusted R^2 between Train and test: 3.15% which is less than 5%**

**Yes! Its a best model**

**Interpretation:**

* We arrived at a very decent model for the the demand for shared bikes with the significant variables
* We can see that temperature variable is having the highest coefficient 0.4914, which means if the temperature increases by one unit the number of bike rentals increases by 0.4914 units.

Similary we can see coefficients of other variables in the equation for best fitted line.

We also see there are some variables with negative coefficients, A negative coefficient suggests that as the independent variable increases, the dependent variable tends to decrease. We have spring, mist cloudy , light snow variables with negative coefficient. The coefficient value signifies how much the mean of the dependent variable changes given a one-unit shift in the independent variable while holding other variables in the model constant.

Model persistence:

* Building an API to expose the model to predict for future inputs.
* Deployment on local machine(Ngrok and Flask)
* Cloud AWS , GCP