**Subjective Questions**

**Assignment-based Subjective Questions**

**Q1.** **From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?**

**Answer1.**

**Categorial Variables we have are: Year, Month, Day, Season**

**I inferred in 2019 Demand of Bike sharing was more as compared to 2018**

1. **As per Season analysis, I inferred usually there is high Demand in**

**“Fall” season, so we should be prepared ourselves with Bikes availability.**

**To maintain or increase current demand**

1. **We should provide attractive offers in other seasons**

**This will increase revenue and size of our Target Audience:**

**Suggestions For Offers:**

1. **Goodies (Bags, T-shirts with Boom Bikes printed)**

**: This will attract Teenagers and kids and help us in promotion of Boom Bikes**

1. **Vouchers: OTT coupon codes, shopping vouchers will attract Females and youngsters**
2. **Boom Bikes Membership: Helps to establish Boom Bikes as a brand in market**

**Q2: Why is it important to use drop\_first=True during dummy variable creation?**

**Answer2. drop\_first=True is important to remove unnecessary variables e.g. I have date:01-02-2018, after creating dummy variables like year, month, day, I can safely drop original variable using drop\_first= True, without any risk of Data Loss. drop\_first=true can also be used in case of duplicated data where u want to drop first occurrence of duplicated value**

**Q3:**  **Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?**

**Answer: Holiday, if holiday is 0 count is more. It indicates people prefer shared bikes,**

**for travelling to work.**

**Q4. How did you validate the assumptions of Linear Regression after building the model**

**on the training set?**

**Q5. Based on the final model, which are the top 3 features contributing significantly towards**

**explaining the demand of the shared bikes?**