5.1 Please discuss potential shortcomings of the methods of your analysis, including:

1. **Dataset:**

The given dataset contains the NYC subway information for the month of May 2011 only. Due the limitation of the dataset, the predication model won't be accurate/appropriate to predict the usage of NYC subway during other month of year apart from May.

If the basic and advance dataset are compared with each other then we can say that the very important information is missing in the basic dataset such as columns like 'day_week' and 'weekend'. This information could be very important for the analysis. Which concludes that it's difficult to commute the accurate/efficient prediction model for the basic dataset.

2. **Analysis, such as the linear regression model or statistical test.**

Linear Regression model:

The basic drawback of linear regression model is that it assumes the linear relationship between the features/dependent variable and output/independent variable. It's not always the case because it's possible that the relation between the features/dependent variable and output/independent variable could be curved (non-linear).

Apart from that the linear regression model is very sensitive to outlier. The presence of outlier affects the position of regression line in the model.

Examples: