4.1 From your analysis and interpretation of the data, do more people ride the NYC subway when it is raining or when it is not raining?

From my analysis and interpretation of the data, more people ride the NYC subway when it is raining compare to when it is not raining.

4.2 What analyses lead you to this conclusion? You should use results from both your statistical tests and your linear regression to support your analysis.

Statistical Tests:
Mann Whitney U test states that the distribution of the two samples, one in which people ride the NYC subway when it is raining and another in which people ride the NYC subway when it is not raining, are not equal. They are statistically different from each another.

Along with that Mann Whitney U test, mean and median value of NYC subway data, when it is raining, is higher than mean and median value of NYC subway date, when it is not raining.

The Mann-Whitney U test returns the following values.

1. The Mann-Whitney statistical value: **1924409167.0**
2. One-sided p-value assuming an asymptotic normal distribution: **0.024999912793489721**

The calculated p-value is less than the critical p value (0.05), which rejects the null hypothesis and stats that the distribution of both the samples (number of entries on rainy day and number of entries on non rainy day) are not equal.

Sample 1 (Numbers of entries on rainy days) mean: **1105.44637675**
Sample 2 (Numbers of entries on non rainy days) mean: **1090.27878015**

Sample 1 (Numbers of entries on rainy days) median: **282.00**
Sample 2 (Numbers of entries on non rainy days) median: **278.00**

Linear regression:

Though my current linear regression model doesn't include rain as feature but if rain is consider as an only feature for the linear regression model then it has positive co-efficient (182.6566) in the linear regression model.

The positive co-efficient of feature 'rain' means the present of rain (value = 1) will increase the number of entries per hour by 182.6566. That states that the number of entries per hour is higher when it is raining compare to when it is not raining.