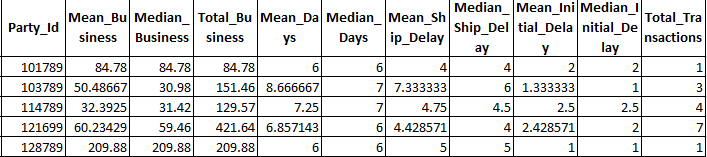
# Do customers that receive their product sooner purchase more than customers with longer delivery times?

To figure out the total delivery time we take difference between delivery confirmation date and order date.

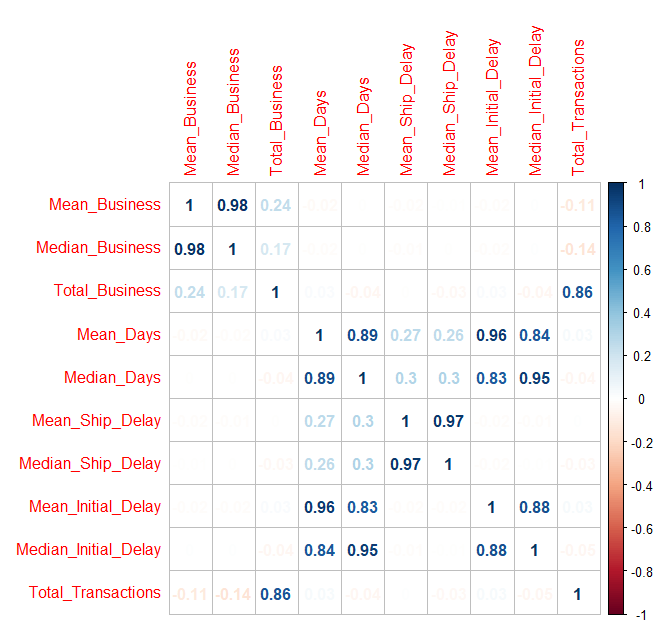
We also consider initial delay i.e. difference between the days of order date and shipping date, and delay in shipping i.e. the difference in between shipping date and delivery date.

We start by calculating median, mean of different types of delays and business earned from each customer.



In order to establish a relationship between business obtained from a given customer, number of purchases made by that customer and delivery time, we chose to build a linear regression model.

Firstly, we built a multi-collinearity matrix amongst above columns.



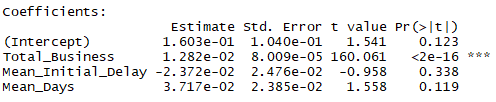
We eliminated all the columns that has high collinearity.

Finally we chose, *'Total\_Business' ,'Mean\_Days', 'Mean\_Initial\_Delay'.* We have further removed shipping delay, because upon running AIC, we found it not be significant.

-Total Business is revenue earned from customer.

- Mean Initial Delay is difference between shipping date and delivery date

- Mean Days is difference between order date and delivery date



Furthermore, upon observing P values, we see that P Value for Total Business is less than 0.05. So, we fail to reject null hypothesis and hence there is significant linear relation between Total Business and Total Transactions.

This model has 77% accuracy in predicting values