# In-depth Analysis (Machine Learning)

Key questions to be analysed using the machine learning:

1. Does the mode of payment impact the purchasing power of the customer?

For understanding this relationship, as payment type was a categorical variable, it was initially converted into dummy variables after which a OLS Regression was done to see if there was any relation between Payment Value of customer and Payment method.

**Code Used:**

from statsmodels.formula.api import ols

fit = ols('payment\_value ~ C(payment\_type\_boleto) + C(payment\_type\_credit\_card) + C(payment\_type\_debit\_card)+C(payment\_type\_voucher)', data=df\_f).fit()

fit.summary()

**Results:**



As seen, there is no relation between payment method and purchasing behaviour. Also, as noted previously 75% of the purchases are done on credit and it also accounts to 75% of the sales as well.

1. Is Delivery time affected by:
   1. Distance between customer and seller
   2. Volume of order
   3. Weight of order
   4. Freight value

To check this, I initially used OLS regression but saw a very small R squared value of 0.2 and also did Lasso regression to pick up the important features for the regression which also showed that none of the features are important. From this we can conclude, none of these factors affect the delivery time and hence delivery time is something that could be standardised and SLAs could be set.

**OLS Regression Code Used:**

import statsmodels.api as sm

X = sm.add\_constant(X)

est = sm.OLS(Y, X).fit()

est.summary()

**Result:**



**Lasso Regression Code:**

from sklearn.linear\_model import Lasso

names=df\_g.drop(['order\_id'],axis=1).columns

len(names)

# Instantiate a lasso regressor: lasso

lasso = Lasso(alpha=0.4,normalize=True)

# Fit the regressor to the data

lasso\_coef=lasso.fit(X,Y).coef\_

# Compute and print the coefficients

lasso\_coef = lasso.fit(X,Y).coef\_

print(lasso\_coef)

# Plot the coefficients

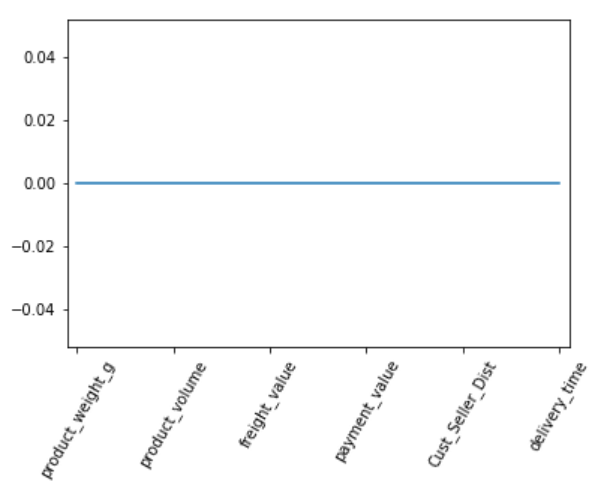
plt.plot(range(len(names)), lasso\_coef)

plt.xticks(range(len(names)), names.values, rotation=60)

plt.margins(0.02)

plt.show()

**Lasso Regression Results:**



As you can see, none of the features are significant to affect delivery time.

1. What factors do the customer rating depend upon?
   1. The delivery time
   2. Difference between estimated and actual delivery time

**Code Used:**

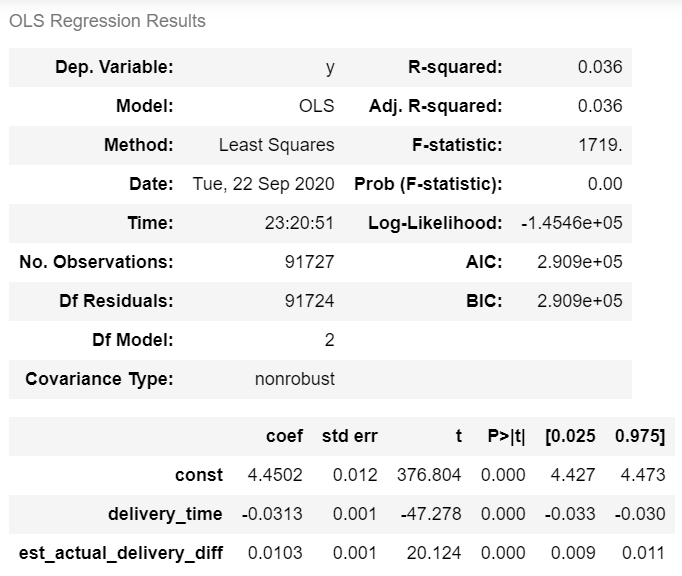
import statsmodels.api as sm

X1 = sm.add\_constant(X1)

est = sm.OLS(Y1, X1).fit()

est.summary()

**Result:**



As seen , there is no significant relationship between the delivery time or even anticipated delivery time- actual delivery time and the ratings of the customer.