Carlos He He

Francis Troy Kirinhakone

CPTS 575

Progress Report

We decided to do a linear regression using the total products value, timestamp, order city, and products category. Since the data runs from 2016 to 2018 we divided them into quarters. So looking at Q1 of 2016, we split a portion of the data to be our training data and built a model around it. We utilized that same model to see how it worked in predicting Q1 of 2017 and 2018. Once we got a good fit then we did the same for Q2, Q3, and Q4. We predicted the sales of 2019 and broke it into quarters. We are also predicting which categories will be mostly purchased for each month or quarter of 2019.

We’ve been brainstorming and decided that the following ideas or discoveries would be interesting to tackle:

Ranking the most frequent customers by city.

Ranking the most frequent bought products by category.

Analysis – seeing if the business should focus more on frequent buyers to optimize sales.

Check if transactions are processed/shipped/delivered on weekends. Are weekend transactions different from weekdays?

* Possible validation would be looking at customer reviews

Do payment types or sequential payments change delivery speed.

Are sequential payments slower? How much slower? Is there loss of profit? Is it worth doing sequential payments or upgrading to single payment?

After finding the above discoveries, we want to see if we can think of new practices to optimize the business. Since the last Olympics in 2016 was held in Rio de Janeiro, we would also like to see if this major world-wide event had an impact on business sales.