Module 20 – Final Project – Self assessment:

Mine was a solo project, so I played all the roles.

It was quite challenging put together all we learnt in this course. But it’s helpful and learnt lot more things during the project.

Most challenging part was exploring and finding a better suitable machine learning model for the dataset. Because it’s a fictious dataset. There was no real co-relation between sales and demographic information available in the dataset. So, at the end I had to settle with sales prediction and customer

calcification. Initial plan was to develop a model to predict the sales at new store based on the store location demographic information.

Tried with multiple regressor model for timeseries data. Finally end with Linear regression and Random Forest regression models. Although for customer classification I used

Unsupervised model(k-Means). Both model fits for this dataset.

During the course, we learnt how to interact with AWS S3 using pyspark. But in this project, I tried with python library boto3 by referring the documentation site.

# Overall Summary:

Product Sales Analysis:

Setting up an automated end-to-end data pipeline for processing and storing sales data using python, AWS and PostgreSQL.

Built the interactive dashboard.

Built the machine learning model to predict the future sales (linear regressor and Random Forest regressor)

For customer classification used the Unsupervised machine learning model (k-means).