

Self-assessment

I played different roles from circle to triangle. My role also in the project included also the Business User point of view. Even that I did not suggest the topic for the project I know about that industry, so it was not just about coding and reporting, I look after to create a sense of business on the analysis, challenging processes that we could identify from the dataset and make disruptive advisements with the results.

I was advising about what could make sense in the real world once we got the code and start reviewing the outputs. All the collaboration took place over whatsapp, slack and zoom, asking for more information and explaining some business processes related to the topic selected.

My greatest challenge (personal) was in the database creation and connection, so I asked for help to connect the AWS database, the team helped me.

At the very end I was in charge of creating the dashboard and draft some hypothesis around the code.

Team-assessment

We interacted three times a week over zoom, and exchanging information on a frequent basis over whatsapp and slack. Our team performed well organized, following the rules advised by the bootcamp lecture, but also based on specific roles, such as business users, DBA's, data scientists and data analysts, being these parallel roles one of our strengths. That created a better ambience and provoke the collaboration among all the teammates. We could also explode the experience of each other in different knowledge fields and make a cohesive project.

That talented teammates made the difference in the last mile of this bootcamp. I learnt a lot from my teammates and I just add more questions to the analysis with the aim of going further.

Brief-summary-of-the-project

The project is focused on the Data and Machine Learning applied to the day a day needs in the corporate world, over a Superstore dataset, which can be easily replicated or extrapolated to other industries due to the areas it covers. We decided to scout over the operations of a retailer with the aim of understanding its performance, create hypothesis to improve the ways the company is performing, and create a predictive model to anticipate potential undesired situations.

The dataset provided in a CSV file includes 48 sales information with variables such as discounts, shipping cost, shipping method, returns, category, subcategory, territory manager, market, etc. The dataset was loaded in a PostgreSQL database using Amazon Web Services, and per internal practices the database was synchronized with PostgreSQL locally for future analysis if the AWS is no longer required.

Once gathered all the information the most important part was to identify the relevant questions to be answered after the analysis. Due to the amount of orders and the money in exchange the "returns" and "profitability" were the main concerns (as an assumption).

The analysis includes three machine learning models, a) Supervised machine learning to classify and predict the profitable orders, b) Supervised machine learning to classify and predict if an order is going to be returned, and c) Neural Network to predict the weekly sales. Something new, the use of LSTM (long short-term memory) for the sales forecast.

Once we got the results of the models the information was exposed through a Tableau dashboard and as part of the Business Analysis the process was mapped and we pointed to the most critical activities that the business need to control or constraint in order to reduce the loses by returns or low, null, or negative profit.