Generalized retail pricing solution - Public project summary

The aim of this project is to create a generalized code framework which contains different parameterizable algorithms and functions that can serve as the basis for a pricing solution for different retail companies with different types of data sets. The codes should be easily parameterized and adjusted case by case. The desired outcome is not a model that provides the most accurate solutions but a useful framework, that could be applied in many different situations.

The client company serves as a data driven solutions consultancy and due to this they solve similar problems and issues for different third-party clients time to time. This framework will help them to create a pricing solution mainly for retail companies faster and in a more efficient manner as they will not need to start from scratch every time, but can build on this basic framework. The main problem I am going to solve is that currently every project is done one by one starting from zero, and the company realized a need to create a general solution that they can apply in many cases.

The codes could be easily understandable and structured in a way that the analysts find it easy to use and reproduce in the future. Beside this it should be applicable to different companies and thus different data sets as well. This is achieved by formulation functions that take different inputs and do the required transformations and calculations on it, rather than hardcoding these for the current data set.

For the project a transactions data set is used, which contain more than 17 million transactions for an unknown retailer in the past two years and 2 months. All the transactions happened before the Covid-19 pandemic so the effect of lockdowns and shopping craze does not affect the data.

The project builds on several divided steps and the code framework is written accordingly. The process starts with basis cleaning of the data in a generalized way, then some feature engineering is done and new variables are created that are needed in further steps and mainly in the modeling part. The next part is about filtering the data set and dropping some observations based on the preferences of the user. It is followed by dealing with extreme values. This part is also customizable and adaptable to different needs. For this project one of the most important parts is the next one which is the aggregation of the data. The process ends with the modelling and creating the product pricing. These are the main steps of the framework.

The methodology used for pricing is based on the price elasticity of demand. For this we are using the different unit prices the products were sold for and the quantity that customers demanded at a given price. This approach is based on the theory of the law of demand which states that quantity purchased varies inversely with price. So, if the price of a product increases the demand for it gets smaller and smaller. We are looking at three different models. In the first two we are modelling the demand curve for each product (how quantity demanded change for different values of price). In the first model we assume a linear relationship between them and calculate a simple linear regression, so in case of a unit change in price level. This model can be easily used to predict an optimal price that maximizes the revenue earned from the sales of the given product. In the second we model the relationship between demand and unit price with a log-log regression. Here we get how elastic our products are this can help with the pricing decisions. The last model takes into consideration that the demand curve may have any forms and to avoid this problem a loess model is built for price prediction.

The final output of the codes and the model is a table that contains all products sold by the retailer and beside some useful statistics, the elasticity value of each and two different price prediction.

Considering it from the view of the client the outcome is a code framework that they can use in the future. For this I am providing a Project_codes.R file which is a clean and straight forward process in which the users can set several parameters adapt some part of it and get the output they require. To make this file nice and easy to understand I moved complex calculations and codes to a Functions.R file. It serves as a supporting file which has the functions that are then used in the project codes. With this approach the main file contains sometime only a couple of lines of codes instead of dozens of lines that I put to the functions file. Beside the code framework, I create a Shiny Application as well. This is built upon the code framework. It has almost the same functionalities and options as the codes themselves but the user interface provides a much better and easier way to update the required fields and values. The application is very easy to use and due to this it cannot only serve as a base for

the analysts at the client company, but after some customizations it could be provided for the third-party users as well.

The most difficult part of the project was to find the best ways to deal with such a huge data set (which has more than 17 million observations). I needed to try different methods for the same purpose as some of them turned out to be too resource heavy. The other difficulty I needed to overcome is to make the codes applicable in different cases and very easily parameterizable. I solved this with generalized functions, and in the end with very little adjustments in the beginning the codes can be applied to different data inputs as well.

This project provides a basic framework for a pricing solution for retail companies and an application that is built on the framework. I only included three very basic models using elasticity modelling to find the optimal prices. In the future the project could be enhanced by including several additional pricing models to it and also by further generalizing the codes. Overall the final product is easily useable by anyone and can be parameterized really easily.

This was the first bigger project in my career where I have the opportunity to create a tool which actually solves a real-life problem. The problem how to price products is faced by almost each and every company. This project helped me understand pricing a little better and I also got an insight on what factors and how does affect demand on different products. Through the project I had the opportunity to think in more detail about retail prices and not just from a customer point of view.

I was building a standardized framework, so I had the opportunity to learn nice and efficient code structuring and function building in a way that other data scientists can easily understand and use it. I also needed to come over some issues due to the size of the data and learned that the first way of solution is nearly never the best one when you create codes. I learnt how to generalize my code further and further and also realized that even that I made quite a lot of work to generalize the codes there are a lot of opportunities left that could make it even more applicable to any situation. I created an application and could dig a bit deeper in the tiny whiny details of it as well as practices some data visualization.

Last but not least had the opportunity to try myself out as a consultant for an actual data driven solutions company. I worked with experts in this field and advanced by project planning and communication skills.