BUSINESS CASES WITH DATA SCIENCE

MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS – MAJOR IN BUSINESS ANALYTICS

Case 4 - ManyGiftUK Recommender System

Group AA

Ahmadov, Emil (m20201004)

Macean, Doris (m20200609)

Shin, Doyun (m20200565)

Tagiltseva, Anastasiia (m20200041)

May, 2021

INDEX

1. INTRODUCTION

With the introduction of online shopping, product preference data has become available for existing, as well as potential customers. Nonetheless, it is still not easy for general consumers to make purchasing decisions in that:

1. Forgetfulness being a basic human trait, when we are not in the immediate need of certain product, we often forget about our needs. Shopping, then, can be seen as acting upon the need we remember, as opposed to acting upon all the needs.
2. Even if the customer has general ideas of products to purchase, there are dozens if not hundreds of alternatives which are nearly impossible for him/her to compare.
3. Rapid innovation driven by increased competition have led vendors to offer highly diverse range of products. There is an ocean of products that the public does not even know exists, even though those products may be highly relevant to certain people.

All three points are related to our limited emotional and cognitive capacity. As such, scholars such as Iyengar and Lepper (2000) observed people tend to purchase more when choices are easy. Naturally, recommender systems have become highly popular to help customers make choices more easily. With regards to point 1, a simple reminder by the recommender system could trigger a purchase, providing a benefit to both the user and the company. For point 2, a handful of options could be recommended as opposed to hundreds of products. Point 3 would require the system to present the information of those potentially relevant products to customers. These solutions are especially important for companies like ManyGiftsUK, where customers can easily get lost due to the sheer number of offerings.

2. BUSINESS UNDERSTANDING

2.1. BACKGROUND

ManyGiftsUK is a UK-based online retailer for a wide range of products primarily for gifting purposes. Within the given data, there were around 4000 different types of items purchased by customers located in 38 different countries. Most of the products are sold directly from their website, with some products also available on Amazon.

2.2. BUSINESS OBJECTIVES

The main business objective that influenced ManyGiftsUK to explore the implementation of a recommender system is the increased revenue obtained by subtly prompting customers to purchase more products. The nudging here would also provide higher shopping convenience for clients. Expanding customer base is not the scope of this project, however, the increased shopping convenience of the website would help to attract and retain customers.

The firm has explicitly expressed the need to improve the shopping experience for both current and new customers. For the former, the company wishes to identify products that the customers may be interested in and make suggestions accordingly. For the latter, the company seeks to make recommendations based on the product the new customer is currently considering.

2.3. BUSINESS SUCCESS CRITERIA

The solution for the above business objectives would be considered successful if each customer after the deployment tends to spend more in monetary terms. Hence, improving customer satisfaction and increasing revenue.

2.4. SITUATION ASSESSMENT

GroupAA consists of four data analysts. We were provided with purchase history of registered customers and unregistered customers from ManyGiftsUK. Our analysis and model development has been done primarily through python, with ALS analysis and LightFM library. Further details on the tools we used are available in the enclosed notebook.

The available hardware was mid-end PCs with octacore 3.1Ghz CPU, 16GB RAM, and Intel UHD Graphics 620 (discrete) GPU, or equivalent, hence the computational power was somewhat limited, but not as to hinder our analysis significantly.

2.5. DETERMINE DATA MINING GOALS

As per our client's request, GroupAA aimed to develop two models. Both are types of recommender systems. One model was developed for the existing customers based on their past purchase history, and that of other similar users. The second model was developed to tackle the cold start problem. Here, a reliable recommendation system was created that is based on the product the visitor is currently checking. We would adopt both the ALS analysis and lightFM library to develop these systems using the collaborative filtering method. The success level can only be measured by performance values (based on prediction against test data) at the initial stage. Only during the survey and the first stage of full deployment (will be discussed in the deployment section in this report) we will be able to observe whether the nudging through recommendation would, in fact, change the buying behavior of customers in general. During the full deployment we will be able to measure the average/median spending per customer, as well as the change in GBP term spending of the existing customers.

3. PREDICTIVE ANALYTICS PROCESS

3.1. DATA UNDERSTANDING

We started with the exploratory data analysis of the given dataset for the purpose of gaining general information about the data. There are 541909 rows which contain 25900 unique transactions, 4070 products and 4372 unique customers, over 38 countries. We identified that about 25% of the transactions belong to the first-time customers.

During our analysis we found that variable "InvoiceDate" and "CustomerID" were wrongly formatted. We also found that the "Description" variable had 1454 rows of missing values and 5268 of our rows in the dataset were duplicate rows. Our team dug further in the data and found out some inconsistencies in the variables. To gain overall insights about the numerical variables we plotted boxplots.
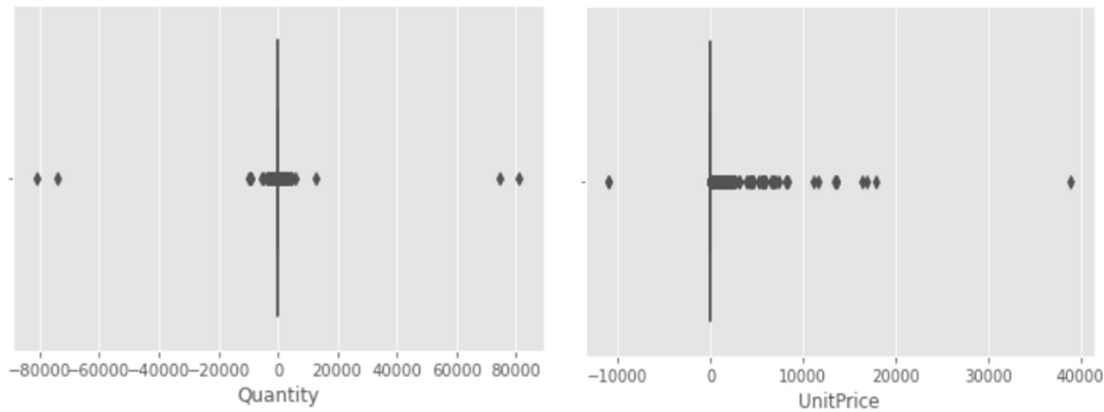
Figure 1 Boxplot for Quantity and UnitPrice

We identified some problems such as negative quantities and negative unit prices. To have a closer look to these values, we checked other variables related to the same records. Through our analysis, we found that some of the descriptions for negative quantities contained some notes for company's internal use such as "damaged", "smashed" etc.

3.2. DATA PREPARATION

We began our data preparation with deleting the duplicated rows which we considered as errors since they contained an identical invoice date. After that we transformed invoice data and customer id to datetime format and object, respectfully.

In order to eliminate the inconsistencies in the variables quantity and unit price, we analyzed these rows with negative values further and identified problematic descriptions related to them. We deleted all the rows which had negative unit prices. We also found values greater than 3000 that appeared to not be related to sales of specific products. As these did not provide information on the sales, they were removed as well.

While analyzing negative quantities we found out that almost all of them were cancelled orders because their stock code started with letter "C". We proceeded to exclude the cancelled orders from our analysis.

As mentioned above, we came across some problematic descriptions during our analysis and decided to take a closer look at the descriptions. We deleted all the rows that had "AMAZON FEE" or "POSTAGE" in their descriptions. While further analyzing, it was identified that most of the real product names were written in upper case letters. As such, we found that lower case descriptions were not accurate product names. We dropped all the descriptions that were written in lower case. In addition, we analyzed cases where the description was written in mixed cases. Some of these descriptions belonged to real products. we identified the ones that were not product names and dropped those rows from the dataset. Please refer to our enclosed notebook for a list of descriptions that were kept or dropped.

Additionally, descriptions with missing values but valid stock codes were filled with the most common description for that specific stock code. In cases where a stock code did not have any description, we ignored them from our analysis.

For further analysis, we engineered the feature revenue as the product of quantity and unit price.

Finally, we created two subsets of the data to be used to develop our two models. One dataset containing existing customers and dataset containing new users. We split the data into train and test set in order to be able to train our models and measure performance. As can be seen below, we made sure that the data within the test set took place after the data in the training set.
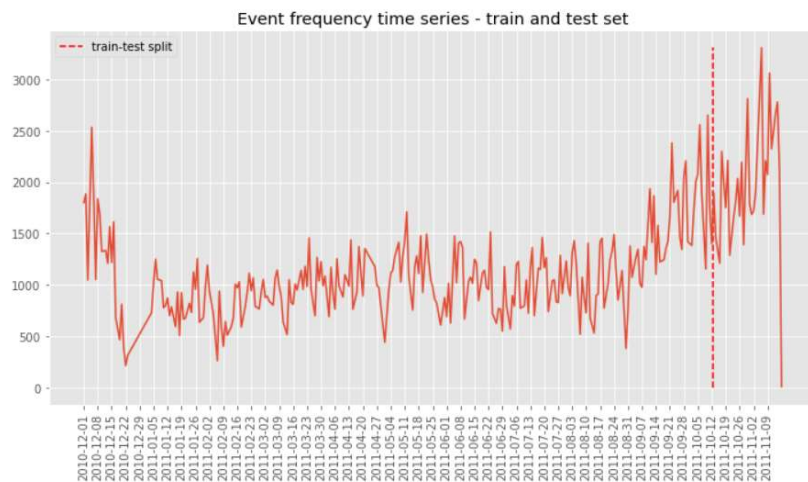

Figure 2: Train-test split of data set

3.3. EXPLORATORY DATA ANALYSIS

Our team conducted an analysis to find the most sold products based on sale amounts (on left) and frequency (on right). Most sold product in terms of quantity is 22197 - which is a popcorn holder, whereas in terms of frequency, white hanging heart t-light holder is the top seller.
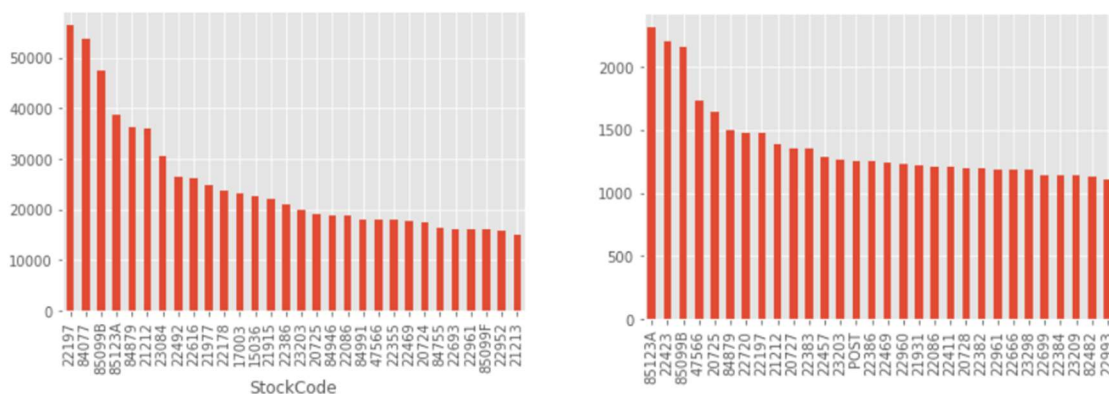

Figure 2 Quantity and frequency for top-30 products

Furthermore, we analysed customers of the company from the perspectives of frequency and amount of revenue they brought to the company. ABC and XYZ analyses were done for this purpose. ABC is based on Pareto principle (the 80 – 20 rule) of the "vital few and trivial many". It helps to identify the importance of the customers to the company based on their revenue contribution. XYZ analysis focuses on how frequent customers buy products from the company based on the number of months in a year. [2]
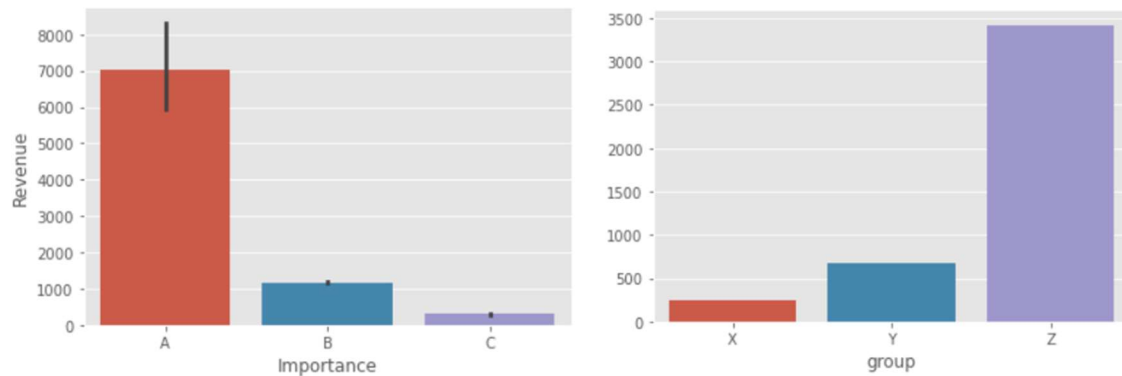
Figure 2 ABC and XYZ analysis

We found out that 2170 of the customers belonged to group C, 1302 of them belonged to group B and 867 of them belonged to group A. Group A customers contributed to almost 75 percent of the revenue, while the others contributed to 17.6 and 8 percentage of the revenue respectively. It was also found that 213 of the customers were loyal customers and made transactions in every month in a year.

3.4. MODELING

We have been presented with the task of developing a recommendation system for the company's clients. Without information about a customer's product ratings, we focused on models that use implicit data. This requires inferring item interest based on purchases made, Content based filtering methods focus on similarities between items and between users. In this case, as do not have information about our users, we focused on similarities between items. Collaborative filtering focuses on interactions between users and items.

3.4.1. Existing customers

We began with an analysis of recommendation models for existing customers. Initially, we applied a baseline model that recommends the most popular items to every user. We subsequently applied various user and item-based content based and collaborative filtering models. The goal is to find similar users based on similarity of purchases and recommending items accordingly. In contrast, item-item filtering will take an item, find users who liked that item and find other items that those users or similar users also liked. We applied the following models – Alternating Least Squares, Logistic Matrix Factorization, Bayesian Personalized Ranking and LightFM – to try to achieve an improvement on the performance of the baseline model.

Alternating Least Squares represents an approach to optimize the loss function. The key insight is that you can turn the non-convex optimization problem into an "easy" quadratic problem if you fix either p_u or q_i. ALS fixes each one of those alternatively. When one is fixed, the other one is computed, and vice versa [3]

Bayesian Personalized Ranking - Although, ALS approach by Yifan Hu reduces the impact of missing data using confidence and preference metrics, it does not directly optimize its model parameters for ranking. Instead, it optimizes to predict if an item is selected by a user or not. Bayesian Personalized Ranking optimization criterion involves pairs of items (the user-specific order of two items) to come up with more personalized rankings for each user [4]

Logistic Matrix Factorization - logistic loss for implicit matrix factorization [5]

LightFM is a hybrid model that uses both content based filtering and collaborative filtering in order to provide recommendations to existing users based on the known purchases of a customer. LightFM learns embeddings for the relationships between users and items. [6]

3.4.2. Cold start problem

Finally, while preparing recommendation systems for existing customers is essential, addressing the cold start problem is very important for capturing the potential sales to new customers. We decided to address the cold start problem by implementing a combination of methods according to the user's online experience. Before any selection is made, our baseline model will be used to simply recommend the most popular items.

We then developed an item segmentation by clustering the products. As a user begins to search for items, the item clusters that we developed will be applied. This clustering makes recommendations of products based on the current keyword entered by the user. This algorithm tries to predict what the user is looking for by suggesting relevant and related items. For example, if a user enters the word necklace, they will see products that are necklaces, as well as other jewelry.

Finally, when a user has a specific item selected, the LightFM model will be applied. The resulting recommendation using this model is based on items that are frequently bought together. Here you can see an example of the application with regards to a specific item. In this way, we can dynamically provide recommendations based on real time information regarding a customer's interests.

4. RESULTS EVALUATION

We used the precision at K as the model selection criteria as it measures the proportion of positive items among the highest-ranked items. In our case, we will only be recommending a subset of items to our customers making precision the most relevant metric. As can be seen in the table below, we found that the model with the best performance was LightFM. This model is more precise and provides an enhancement on traditional collaborative filtering models.

	Pop_model	ALS_model	LMF_model	BPR_model	LightFM
Precision	0.086880	0.046214	0.037283	0.055619	0.1095
AUC	0.513039	0.507264	0.504235	0.508511	0.6539

Table 1. Results for each model used.

5. DEPLOYMENT AND MAINTENANCE PLANS

5.1. PRE-DEPLOYMENT

We recommend starting the pre-deployment stage with a survey. Without having to actually implementing the recommender in the system, a replica of the webshop can be built, however only for a number of selected products. We recommend selecting products from different revenue levels

and purposes. Then, the items suggested by our recommendation models can be presented on different location of the product page. The survey should measure the customer's purchase intentions, focused around questions like "How attracted were you to check out the suggested items from scale 0 to 10?" From the surveyed group, a number of people can be selected for focus group analysis, to further design the optimal placement and timing of the recommendations. For these purposes, GroupAA analysts will work with customer relations/marketing team of ManyGiftsUK.

5.2. FULL-DEPLOYMENT

Full deployment will follow in 3 stages, targeting a randomly selected visitors initially, then expanding the size of the experimental group. We believe 10% of the total visitors for stage 1, and 30% for stage 2 will be acceptable number of the experimental groups. This will allow us to compare the behavior patterns between the group exposed to recommendation system against others, then evaluate our model further in large scale for further optimization. This gradual adaptation strategy would also minimize and control unforeseen effects of the recommender models to the customers. At stage 3, all visitors would be targeted. Note, depending on the company's situation and the result from the previous full-deployment stage, more gradual implementation with more stages will be possible.

5.3. POST-DEPLOYMENT

After the system integration for all visitors of the website, periodical assessment has to be done. Beside the continuous monitoring, we believe a quarterly in-depth evaluation of the model and analysis of customer behavior changes will be necessary to maintain the effectiveness of the recommendation system. The maintenance method from this point will depend on the company's situation. GroupAA will be happy to train ManyGiftUK's personnel to handle the day-to-day monitoring with one of our analysts stationed onsite. For deeper level analysis and optimization, we recommend to involve our company for consultation.

6. CONCLUSIONS

In this work, our team has conducted Exploratory Data Analysis to better understanding and gain interesting insights from the data. We have built 4 different recommendation engines based on different ideas and algorithms. Our models are simple to implement, highly scalable, and have shown good performance using a dataset composed of user purchasing behavior from ManyGiftsUK.

We have found LightFM as the optimal model for recommending items to existing users with a precision of 0.11. A mixture of popular recommender, item clusterizationn and LightFM has be suggested as a solution to the cold start problem. While the results are interesting, there is some room for improvement. The models can be expanded with some additional data on the users. The use of content-based filtering could also be enhanced.

7. REFERENCES

[1] Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? Journal of Personality and Social Psychology, 79(6), 995–1006. https://doi.org/10.1037/0022-3514.79.6.995

[2] Greene, S. (2020, October). *Industry Optimization using ABC and XYZ Analysis.* researchgate.net. https://www.researchgate.net/publication/345351238_Industry_Optimization_using_ABC_and_XYZ_Analysis.

[3] Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. 2008 Eighth IEEE International Conference on Data Mining, 263–272.

https://www.researchgate.net/publication/220765111_Collaborative_Filtering_for_Implicit_Feedback_Datasets

[4] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner and Lars Schmidt-Thieme (2009). BPR: Bayesian Personalized Ranking from Implicit Feedback. 2009 UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009

https://www.researchgate.net/publication/221404543_BPR_Bayesian_Personalized_Ranking_from_Implicit_Feedback

[5] Christopher C. Johnson (2014) Logistic Matrix Factorization for Implicit Feedback Data. NIPS Conference, 2014.

https://web.stanford.edu/~rezab/nips2014workshop/submits/logmat.pdf

[6] Kula, M. (2015, July). Metadata Embeddings for User and Item Cold-start Recommendations. https://arxiv.org/pdf/1507.08439.pdf

[7] Kordik, P (2018, June) Machine Learning for Recommender systems — Part 1 (algorithms, evaluation and cold start) https://medium.com/recombee-blog/machine-learning-for-recommender-systems-part-1-algorithms-evaluation-and-cold-start-6f696683d0ed