BUSINESS CASES WITH DATA SCIENCE

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

Recommender System - ManyGiftsUK

Ana Marta Silva: M20200971

Natalia Cristina Castañeda: M20200575

María Luisa Noguera: M20201005

Gustavo Tourinho: M20180846

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

1. BUSINESS UNDERSTANDING

1.1 Introduction and Background

Word of mouth or a “buzz” were the traditional recommendation systems for audiences to choose and purchase any item. However, with the development of technology in the digital era, more options to push products are now available to sellers. Amongst these options they now have the recommendation systems based on large databases generated by historical behavior of their customers on a digital platform.

Examples of such recommendation systems based on algorithms are the advertisements in social networks, video suggestions in Youtube, and of course similar, complementary or related products on Amazon. All products pushed by the recommendation systems are relevant to each customer precisely because its base on their previous purchases or views on the platform.

Although this recommendation algorithms have been available for some time, not all companies have transitioned to their use.

Such is the case for ManyGiftsUK, who, despite having a track record of over two decades, only two years prior has moved their business from analog to digital and are now generating enough data for them to start generating recommendations to their customers.

1.2 Business Objectives

a. Build a collaborative recommendation system for ManyGiftsUK’s customers based on the provided dataset.
b. Offer relevant items to new customers.
c. Improve customers satisfaction.

1.3 Business Success criteria

a. Top 10 recommendations showing on each user’s landing page
b. Increase in related item sales as a result of recommendations in each user’s page.

1.4 Situation Assessment

a. General comments and Resources

ManyGiftsUK provided us with a transaction data set that contained 541909 entries that translated into 25900 transactions with 4070 unique items they sell through their store and that was built with the business of 4372 customers.

The dataset also contains 8 variables for each entity which are:

InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

Since no rating or scoring was provided, we focus on developing a collaborative recommendation system for this client.

b. Terminology

Concept	**Definition**
ALS Model	An Alternating Least Squares model (algorithm) matrix type collaborative recommendation filter
Implicit data	Information deduced from historical data gathered from customer behavior
Explicit data	Information based on a rating system (or scoring) done by an audience

1.5 Recommendation System Goals

a. Build a collaborative recommendation system and determine which model is more effective (considering ALS, Bayesian, LMF, AALS, NMSLALS)

1.6 Machine Learning Success Criteria

a. Mean Average Precision
b. Normalized Discounted Cumulative Gain
c. Precision
d. AUC

Since several models are being evaluated, the final choice will be based in the overall scoring of this four metrics of the best executed one.

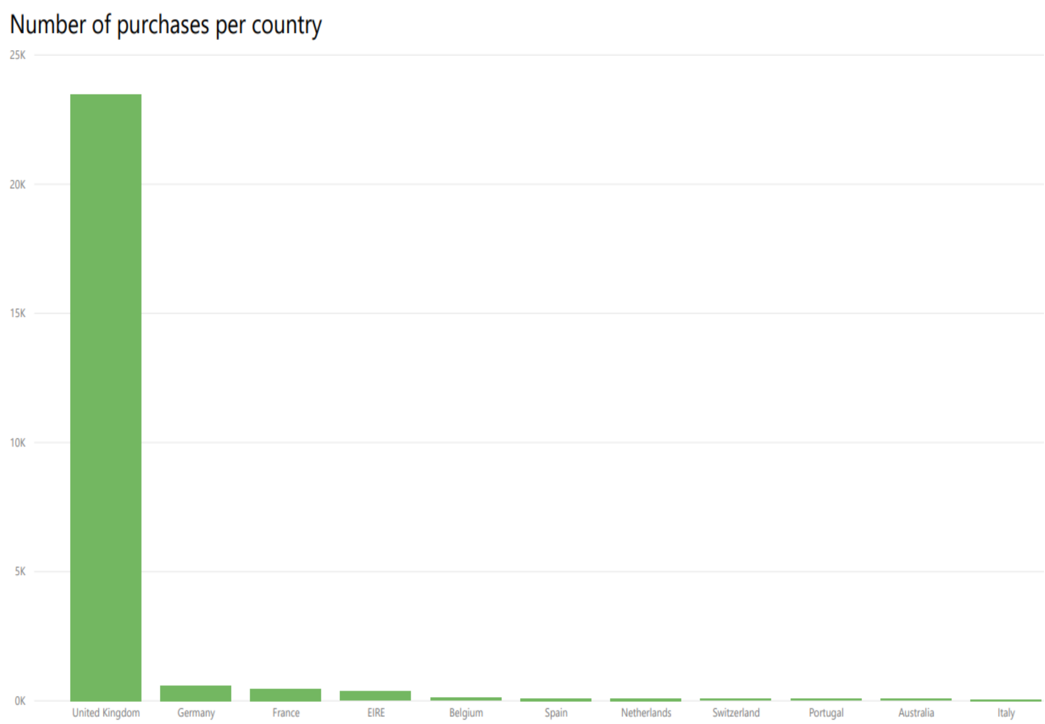
2. DATA UNDERSTANDING

Because of the lack of detail regarding the products and combined with the simple fact that no ratings are included in the dataset, it was evident that the models to build had to be collaborative.

Once we explored the dataset further, we found that cancellations represented a duplicity since it used the same customer ID and created a separate transaction (did not replaced the purchase), therefore these were later eliminated.

Other details such as average of items purchased by customer (3), average cost of 4,6 GBP, and most popular item is White Hanging Heart T-Light Holder, were taken into consideration for the rest of the analysis.

Besides, in terms of distribution of purchases per country, ManyGiftsUK sells for 38 distinct countries, being the UK its strongest market, followed by Germany and France, in a much smaller scale.



3. DATA PREPARATION

3.1 Data Quality

After a first analysis of the data, we found that the missing values corresponded to new customers (135080, 25%), meaning customers without a purchasing history. Since they do not provide the needed customer-item interactions appropriate for the considered strategy, they were filtered out. Additionally, we found that the dataset included 10624 cancelations (1.9%) which according to what was initially identified, they could not be used to infer preference.

3.2 Feature Engineering

a. The CustomerID values were transformed into integer numbers.
b. The instances with UnitPrice equal to zero were eliminated, as we assumed that they correspond to offers proposed to the customers and don't represent their individual preferences.
c. We grouped by Customer_ID and StockCode, and aggregated by Quantity in order to obtain their unique combinations.
d. Considering the existing cancellations which corresponded to a negative quantity values and the fact that these coexisted in the dataset with the original invoices, we decided to keep only the quantities higher than zero.

3.3 Feature Exclusion

a. The features InvoiceNo, Description, InvoiceDate, UnitPrice and Country were eliminated as they were not relevant for the models we aimed to build.

3.4 Format data

Before applying the recommender systems algorithms, we created a sparse matrix with a shape (consisting of unique customers as rows (4325) and unique StockCode as columns (3649), and the values in it correspond to the aggregated quantities previously obtained. The resulting matrix had a sparsity of 98.3%.

4. MODELING

4.1 Modeling Techniques, Assumptions, and Modeling.

Following, we proceeded to divide this matrix into train and test, having each a proportion of 80% and 20% respectively.
Next, in order to answer the business objectives and considering the type of information available, we decided to aim at converting user behavior into user preferences by creating collaborative based recommendation system models. These models recognize similarities between users or items based on explicit or implicit indications and make automatic predictions of user interests by learning their preferences considering their interaction with a set of items and other user's interactions. In other words, are based on the idea that similar items are liked by similar users and vice versa.

Likewise, we decided to apply various matrix factorization models, such as: Alternating Least Squares (ALS), Bayesian Personalized Ranking, Logistic Matrix Factorization and Approximate Alternating Least Squares which consider different methods of speeding up the ALS recommendations system (Non-Metric Space Library and Approximate Nearest Neighbours). These models differ because of different loss functions.

For all these models we defined the following parameters: alpha=15, which correspond to the level of confidence on the dataset for making recommendations, factors=20 which corresponds to the number of latent factors to compute, regularization of 0.1, and iterations=50 which correspond to the number of iterations used to fit the model in the data.

Besides, we later decided to tune the hyperparameters of the Alternating Least Squares by testing an alpha of 50 and 80 and in a third trial we opted for a regularization of 0.01.

On the other hand, in order to solve the cold start problem, resulting from the existence of new customers in the database who consequently don't have any buying history, we decided to create a model that provides recommendations to them based on the most popular items in the dataset.

4.2 Assess the Model

In order to assess the performance of our recommender system, we aim to analyze if the recommendations provided match the products that the customer ended up buying, thus if the recommendations are relevant and correctly ranked.

Consequently, we decided to compute the precision, the AUC and two famous metrics applied to ranking systems: the mean average precision (MAP) and the NDCG (Normalized Discounted Cumulative Gain). For this we applied the assessment method from the Implicit library.

We obtained the following results:

	pop_model	als_model	als_model2	als_model3	als_model4	bpr_model	lmf_model	annoy_model	nms_model
precision	0.181158	0.035754	0.016046	0.019580	0.037314	0.077080	0.011016	0.036072	0.037824
map	0.085186	0.009857	0.004456	0.005828	0.010125	0.041737	0.002809	0.009935	0.010411
ndcg	0.171152	0.030286	0.013990	0.017573	0.031248	0.090393	0.009172	0.030559	0.031863
auc	0.515691	0.502633	0.501553	0.501920	0.502757	0.504972	0.499246	0.502749	0.502641

However, we were not satisfied with the final scores obtained for these models for each metric and being informed of the limitations of the ranking_metrics_at_k method from the library implicit, we decided to compute the metric AUC for all models using a function created by us.

The AUC corresponds to the area under the ROC curve (Receiver Operating Characteristic). Greater areas represent recommendations where the products recommended are in fact bought. We will use the test set to calculate this metric.

With this in mind, we wrote a function to calculate the AUC for each user present in the test set and as a benchmark we calculated the mean AUC for the case when recommendations are based only on popular items. We obtained the following results:

	AUC
als_model	0.872
als_model2	0.860
als_model3	0.865
als_model4	0.873
bpr_model	0.675
lmf_model	0.649
annoy_model	0.872
nms_model	0.872

Consequently, that the model with highest AUC is the ALS4 which was obtained after tuning the regularization factor to 0.01instead of the 0.1 value used in the other models. This ALS model has a AUC of 87,3% which means that the model is very good at proposing relevant products as 87,3% of the times the recommendations were accepted.

In comparison, the AUC obtained by the popular model was of 81.3%, which means that the model we created has better performance, meaning it suggests more relevant recommendations compared with popular products.

5. EVALUATION

5.1 Evaluation results

After selecting the best recommender system algorithm based on the AUC we are now able to assess their recommendations for a particular costumer. First, we start by retrieving the purchased items of a specific customer and then proceed to obtain the **top 10** items recommended by the algorithm. For example, for a customer who bought a Pink Honeycomb Paper Fan, the ALS recommends the following 10 products:

	StockCode	Description
0	84789	ENCHANTED BIRD PLANT CAGE
1	85188B	PINK METAL SWINGING BUNNY
2	84535A	ENGLISH ROSE NOTEBOOK A6 SIZE
3	85214	TUB 24 PINK FLOWER PEGS
4	35004G	SET OF 3 GOLD FLYING DUCKS
5	90039B	FIRE POLISHED GLASS BRACELET MONTAN
6	85136A	YELLOW SHARK HELICOPTER
7	90155	RESIN NECKLACE W PASTEL BEADS
8	85086A	CANDY SPOT HEART DECORATION
9	84199	GLOW IN DARK DOLPHINS

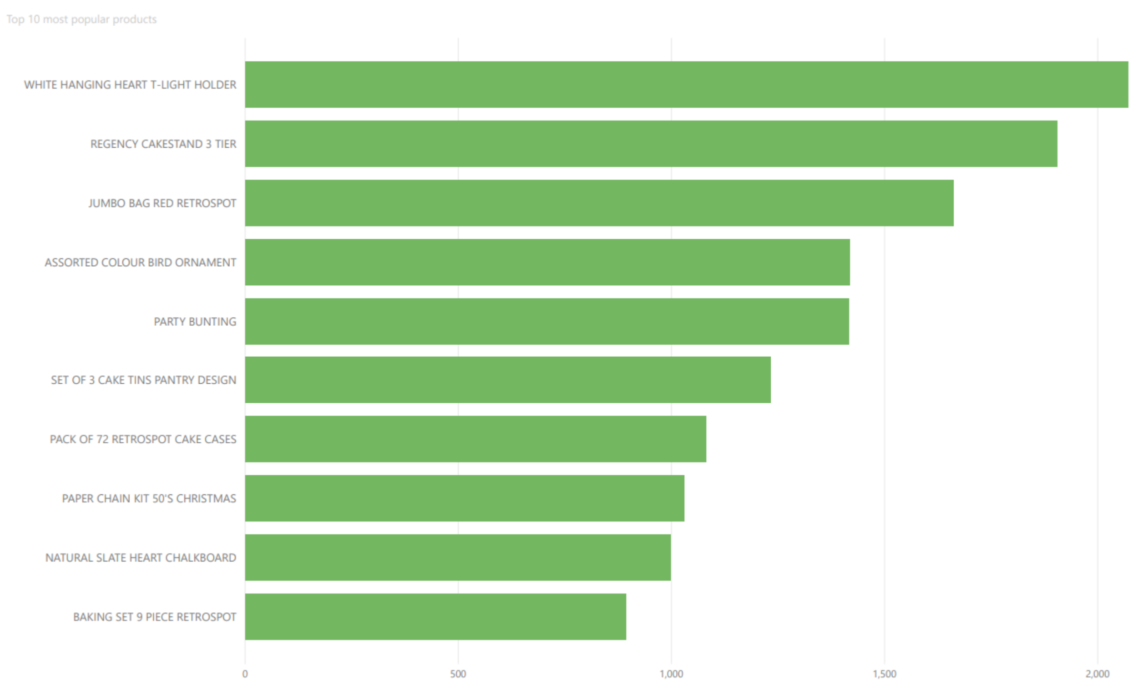
Regarding these recommendations, it is evident that the algorithm infers that the customer is interested in the pink color and motif decorations. These seems a reasonable result from the model, which proves the good quality of our recommendation system.

Another example is a customer who selects 3 products: a Pink Yellow Patch Cushion Cover, a Blue White Plastic Rings Lampshade and a Pack Of 12 Vintage Leaf Tissues, The recommendations provided by the system will be:

	StockCode	Description
0	44092C	PURPLE/COPPER HANGING LAMPSHADE
1	20902	VINTAGE KEEPSAKE BOX PARIS DAYS
2	23296	SET OF 6 TEA TIME BAKING CASES
3	84678	CLASSICAL ROSE SMALL VASE
4	84006	MAGIC TREE -PAPER FLOWERS
5	23293	SET OF 12 FAIRY CAKE BAKING CASES
6	22393	PAPERWEIGHT VINTAGE COLLAGE
7	79302M	ART LIGHTS,FUNK MONKEY
8	23439	HAND WARMER RED LOVE HEART
9	84568	GIRLS ALPHABET IRON ON PATCHES

Again, the system inferred that the customer would be interested in home decoration items which seems relevant suggestions considering the items the customer have previously bought.

For the new customers, the recommendation system will propose the top 10 most popular products. This model of recommendation has an AUC of 81.3% which although lower than the one obtained by applying the ALS model, it is still a very good score and makes us confident on its recommendations and adequate usage. The top 10 products considering the data available are:



After the customer makes a purchase, ALS recommendations will be proposed.

5.2 Review process

We believe that more information about the characteristics of the products, for example categories, would allow to make a content based recommendation system, by matching products' attributes to customer's preferences, thus allowing to create a hybrid system and improve the recommendations.

Besides, having more demographical data about the customers would advance the recommendations obtained.

6. DEPLOYMENT

Recommender systems allow to increase the number of items sold, their diversity and to increase user satisfaction and user fidelity.

In order to ManyGiftsUK to be able to provide relevant recommendations of products to their customers, therefore providing a better user experience on its website and facilitating the user's choices, the company should implement our ALS recommendation system for all their registered users (users with account in their website).

For the new users of their website, they should apply our popular recommendation system, which will suggest the top 10 most popular products and then proceed by suggesting the recommendations based on the ALS model when purchasing history is available.

Considering the constant changes in user preferences, the model should be updated on a real time basis to improve and update the recommendations. Also, more explicit data should be obtained by for example product ratings.

Moreover, we suggest that contextual information indicators besides the country of origin of the customer, date or the description of the product should be collected as they could allow to create a hybrid recommender system of content, context and collaborative recommendations, thus improving their relevance. For example, the items added to a *wishlist,* the content searched/clicked (browsing/search history) and the device type are information that could be collected.

Furthermore, if more information would be collected a clustering of customers could be performed thus suggesting their preferences to new users. More personal data about the customers would allow to build more detailed buying profiles.

In terms of challenges that can arise from applying a recommendation system, we can mention the following:

- Trust issues when the recommendations fail to meet the customer's interests and they don't understand what indicators affect it.
- Privacy concerns if the user doesn't trust the system, thus being more reluctant to disclose personal information.
- Visitors to the website who delete, or block cookies will be always identified as new visitors so the website should be able to provide specific identifiers to these users so in a next visit, the recommendation system can provide recommendations based on the previous search behavior.
- Malicious users that enter the website and start rating certain items to increase or decrease their popularity this decreasing the quality of the recommendations. These attacks can be detected by hit ratio for example.
- The recommender system may face a latency problem when products are added to the platform very frequently and, because they don't have purchases history they are not recommended to customers. To solve this a content based recommendation system should be built.
- Scalability is also a concern when the number of user and products in the platform increase significantly. Consequently, the recommender system finds difficulties for providing recommendations when needing to process such a large-scale data, for that reason the ALS model can be applied in small clusters instead of in the entire dataset.

Visual representations of like-minded users' purchases can generate more transparency and help customers identify matching interests. Besides, describing why products are being suggested, for example by stating that the products are proposed based on the customer previous or recent purchasing history can also help increasing trust.

Another technique to improve the recommendations would be to increase the involvement of the customer in the recommendation system by allowing to customize preference indicators (for example: category, price, ratings).

Finally, if a user of the website finds the recommendations interesting and relevant this will improve the user's opinion of the system and consequently of the company, thus increasing the usage of the system and the likelihood of the recommendations being accepted.

7. References

Khusro S., Ali Z., Ullah I. (2016) Recommender Systems: Issues, Challenges, and Research Opportunities. In: Kim K., Joukov N. (eds) Information Science and Applications (ICISA) 2016. Lecture Notes in Electrical Engineering, vol 376. Springer, Singapore. https://doi.org/10.1007/978-981-10-0557-2_112

He, C., Parra, D., & Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Syst. Appl., 56*, 9-27.