Influencing Buyers Decision with Product Recommendations

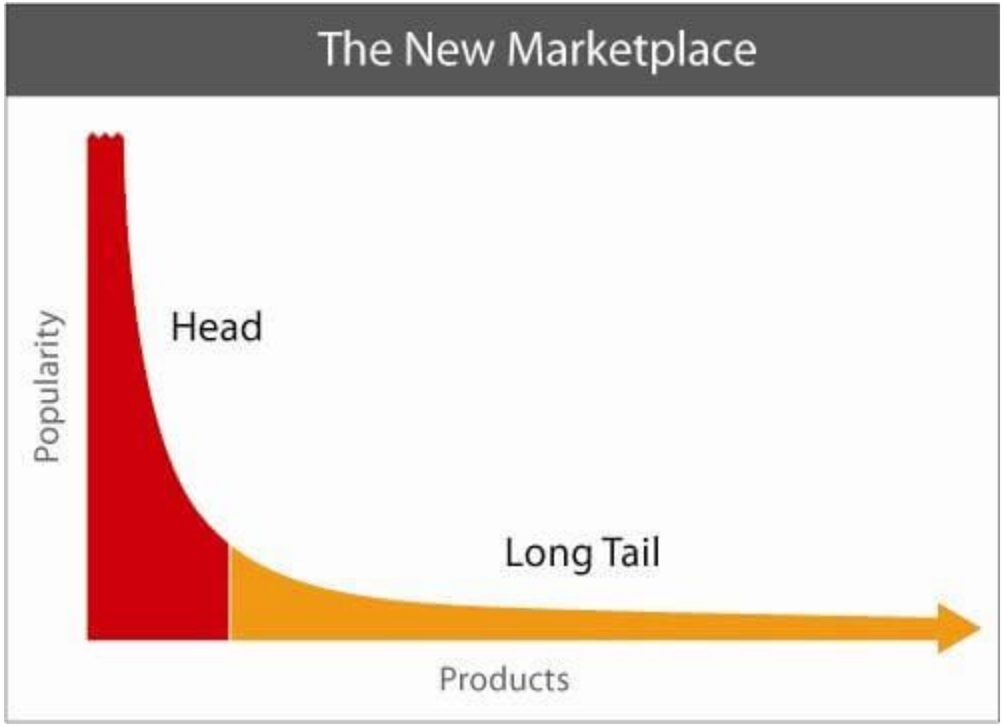
Final Capstone

Thapani Sawaengsri

CONTEXT

- According to Statista, over 4.33 billion people access the internet globally.
- Many service providers and retailers have moved online and the popularity of ecommerce platforms continue to rise .
- Unlike physical stores, online retailers are not restricted to limited shelf space. Retailers are able to sell a wide variety of products online because the costs of logistics is lower.

THE LONG TAIL

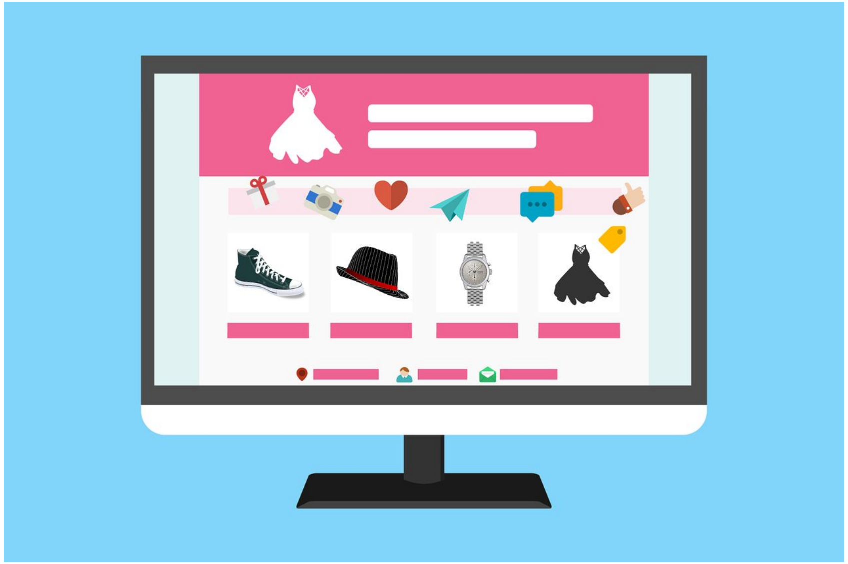


TASTING BOOTH EXPERIMENT

vs.

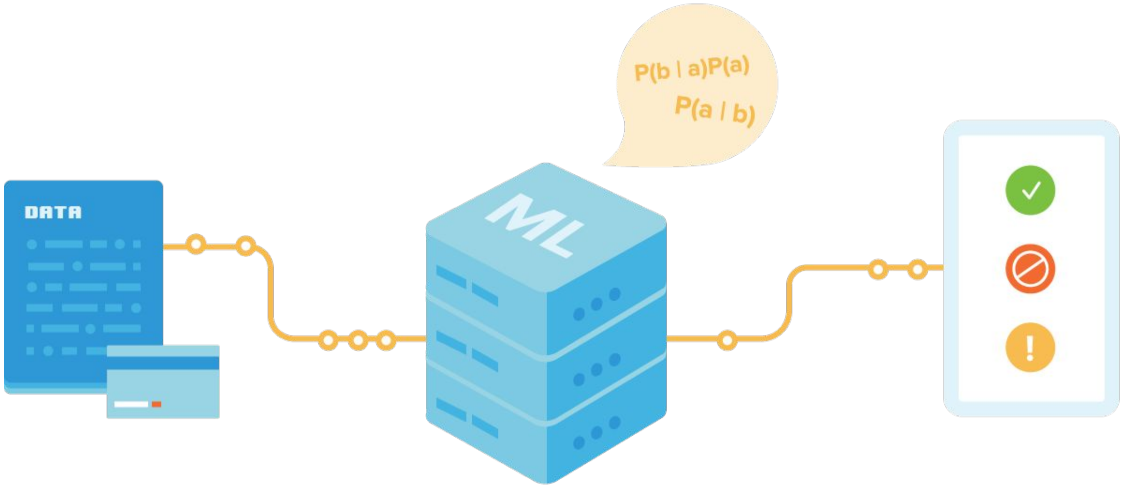
GOAL

- The huge amount of goods available makes it difficult for customers to navigate through their product of interest.
- To influence the user's buying decision and increase the company's revenue, we will create a product recommender system to provide a suggested list of items based on users history.



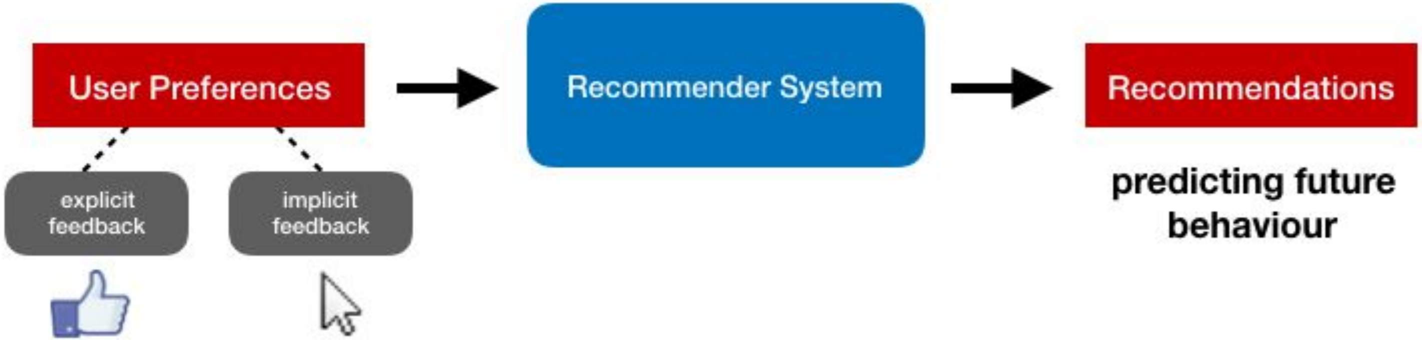
OVERVIEW

- RECOMMENDER MODELS
- DATASET
- MODELING
- DATA PREPROCESSING
- DEMO
- FUTURE WORK



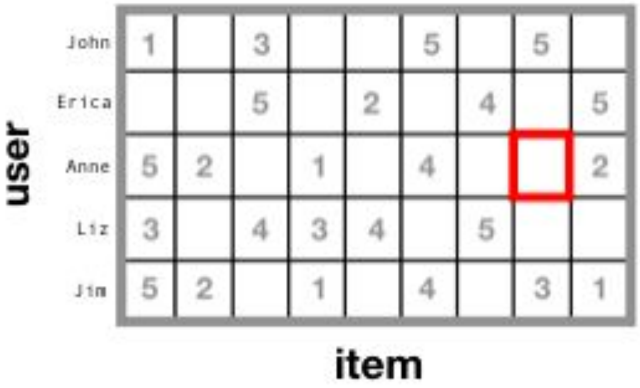
Feed data into a machine learning algorithm to help you make a decision.

RECOMMENDER SYSTEMS



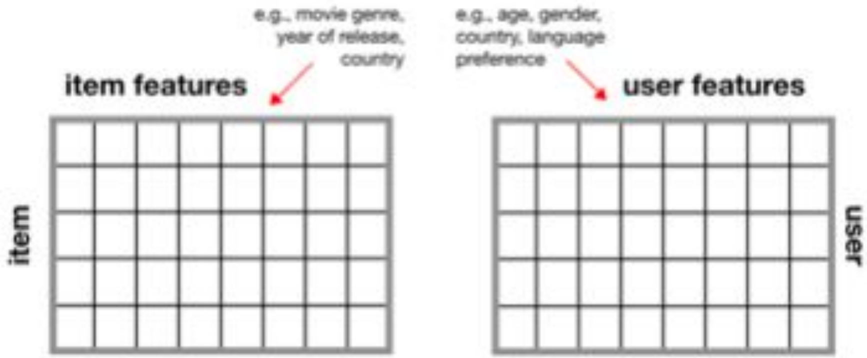
RECOMMENDER SYSTEMS

Collaborative filtering



similar users like similar things

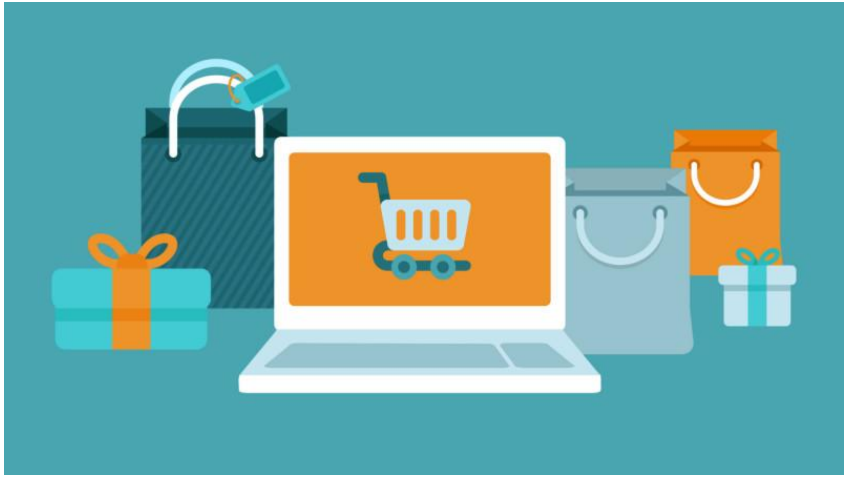
Content-based filtering



considers items/users features

DATA SET

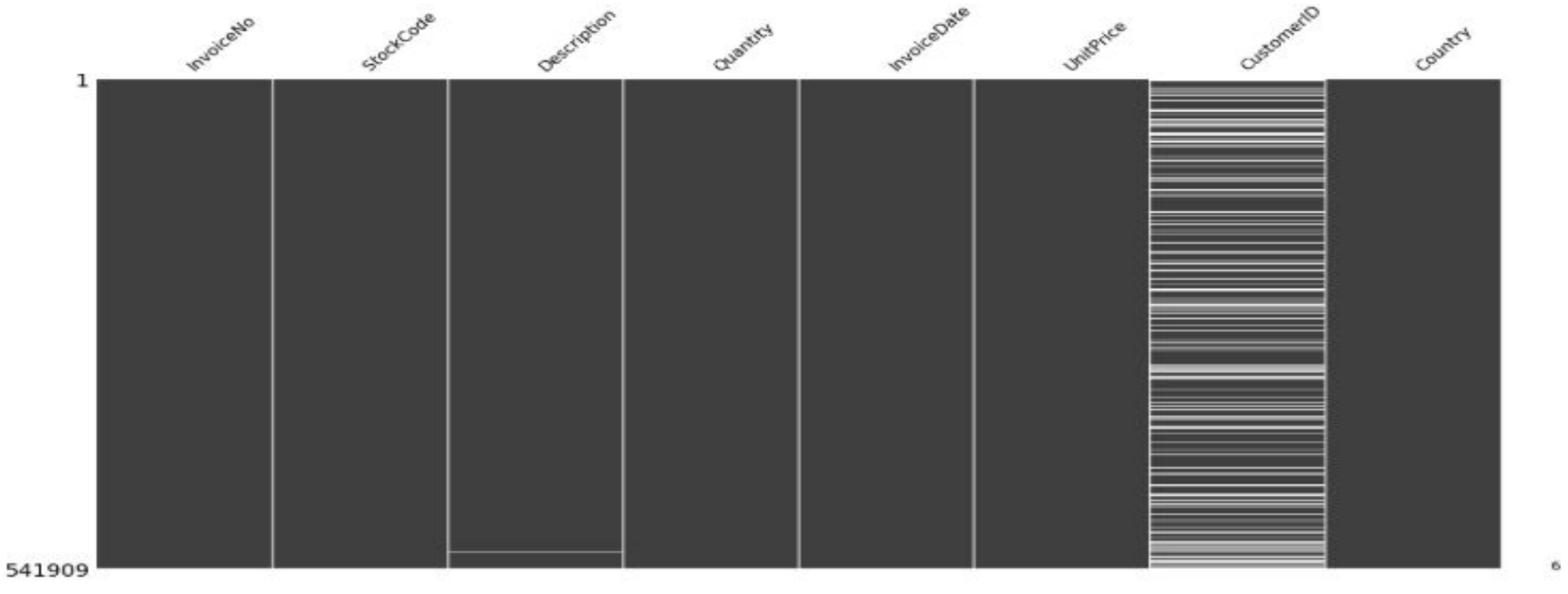
- Collected by Daqing Chen, Sai Liang Sain, and Kun Guo and available on UCI Machine Learning Repository
- Contains transactional information of an online retail company based in the UK during a year period



DATA SAMPLE

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

DATA CLEANING



25% of Customer IDs were missing and these rows were dropped

EXPLORATORY DATA ANALYSIS

CONTENT

Number of transactions: 18,535

Number of products: 3,664

Number of customers: 4,339

Number of countries: 37

	quantity	unit_price	cust_id
count	406829.000000	406829.000000	406829.000000
mean	12.061303	3.460471	15287.690570
std	248.693370	69.315162	1713.600303
min	-80995.000000	0.000000	12346.000000
25%	2.000000	1.250000	13953.000000
50%	5.000000	1.950000	15152.000000
75%	12.000000	3.750000	16791.000000
max	80995.000000	38970.000000	18287.000000

TOP CUSTOMERS

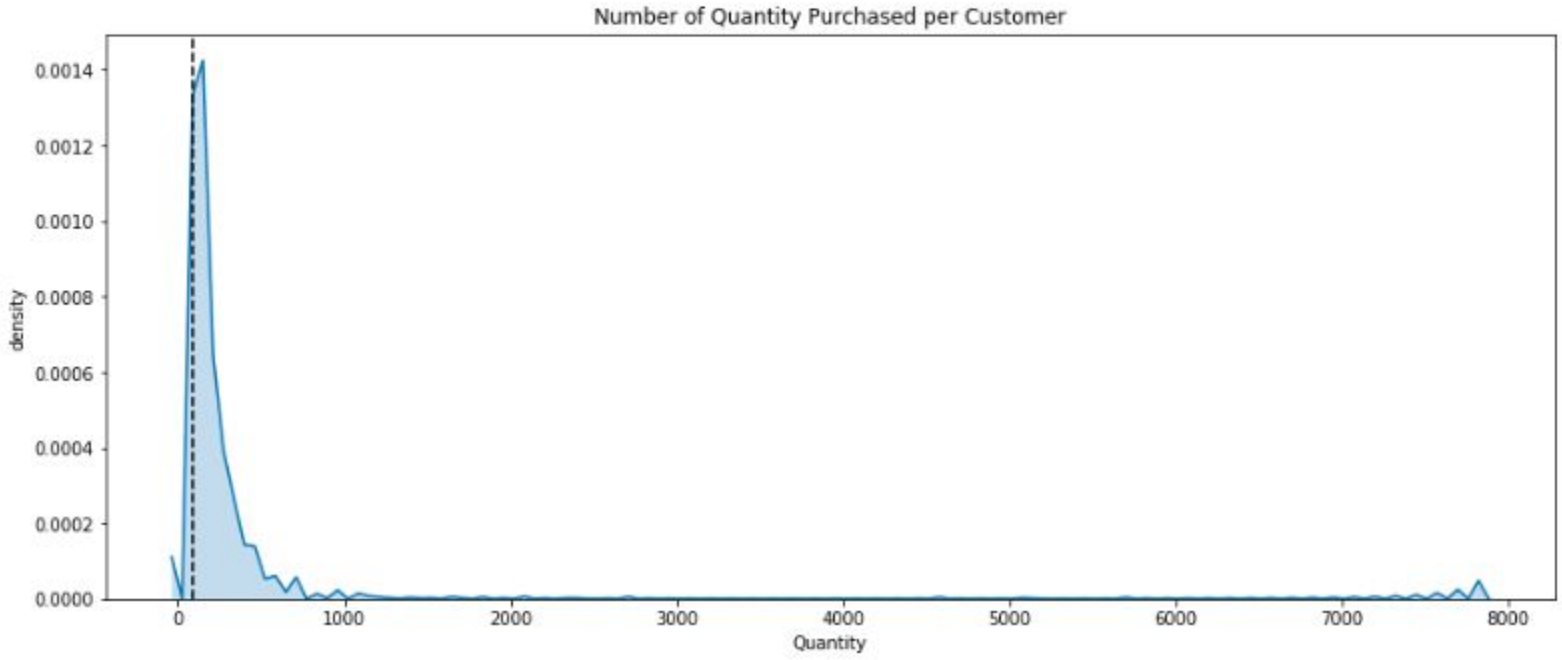
Who are the top customers with the most number of orders?

	cust_id	country	invoice_num
4019	17841	United Kingdom	7847
1888	14911	EIRE	5677
1298	14096	United Kingdom	5111
334	12748	United Kingdom	4596
1670	14606	United Kingdom	2700

Which customers spent the most money?

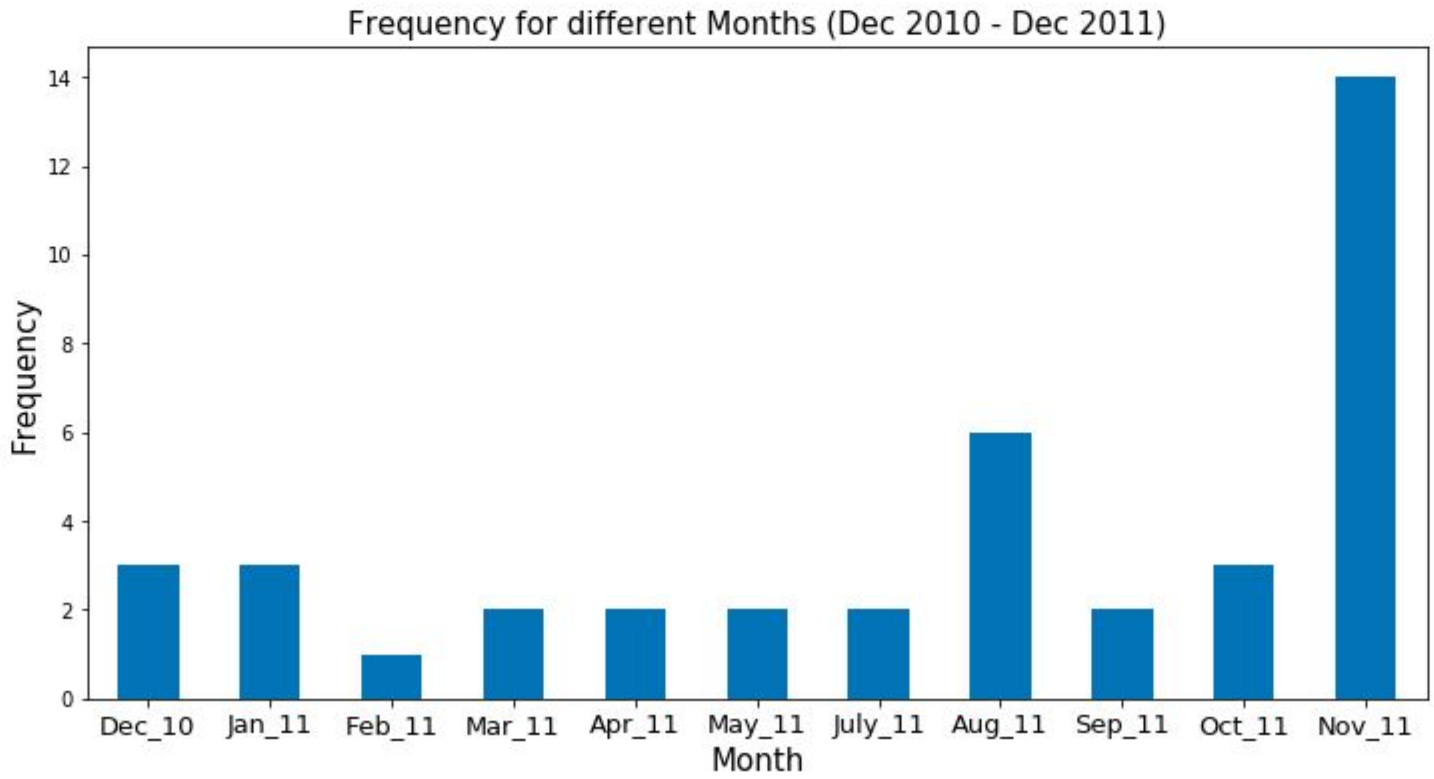
	cust_id	country	amount_spent
1698	14646	Netherlands	280206.02
4210	18102	United Kingdom	259657.30
3737	17450	United Kingdom	194550.79
1888	14911	EIRE	143825.06
57	12415	Australia	124914.53

DISTRIBUTION OF QUANTITY ORDERED

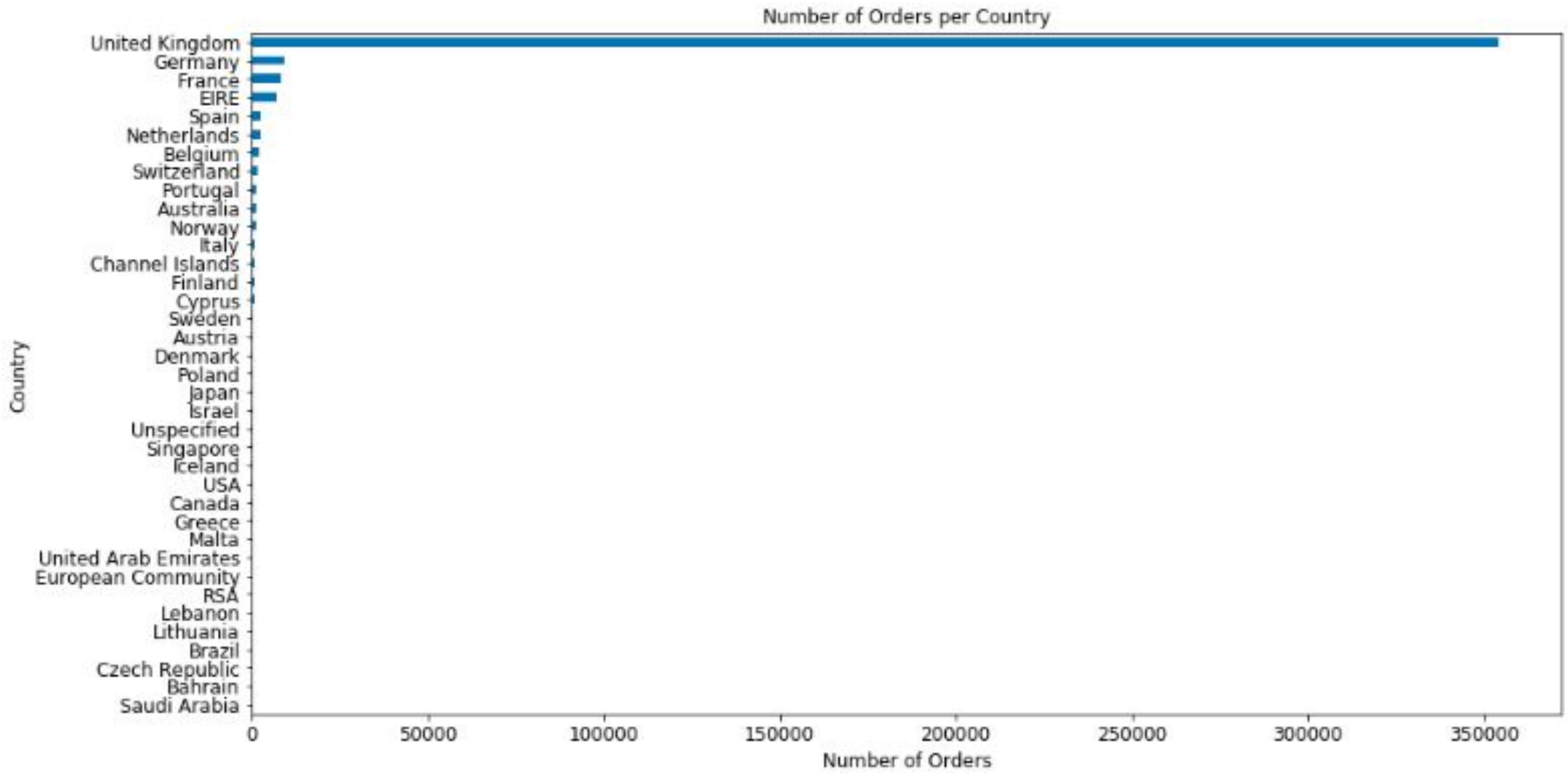


SALES REPORT

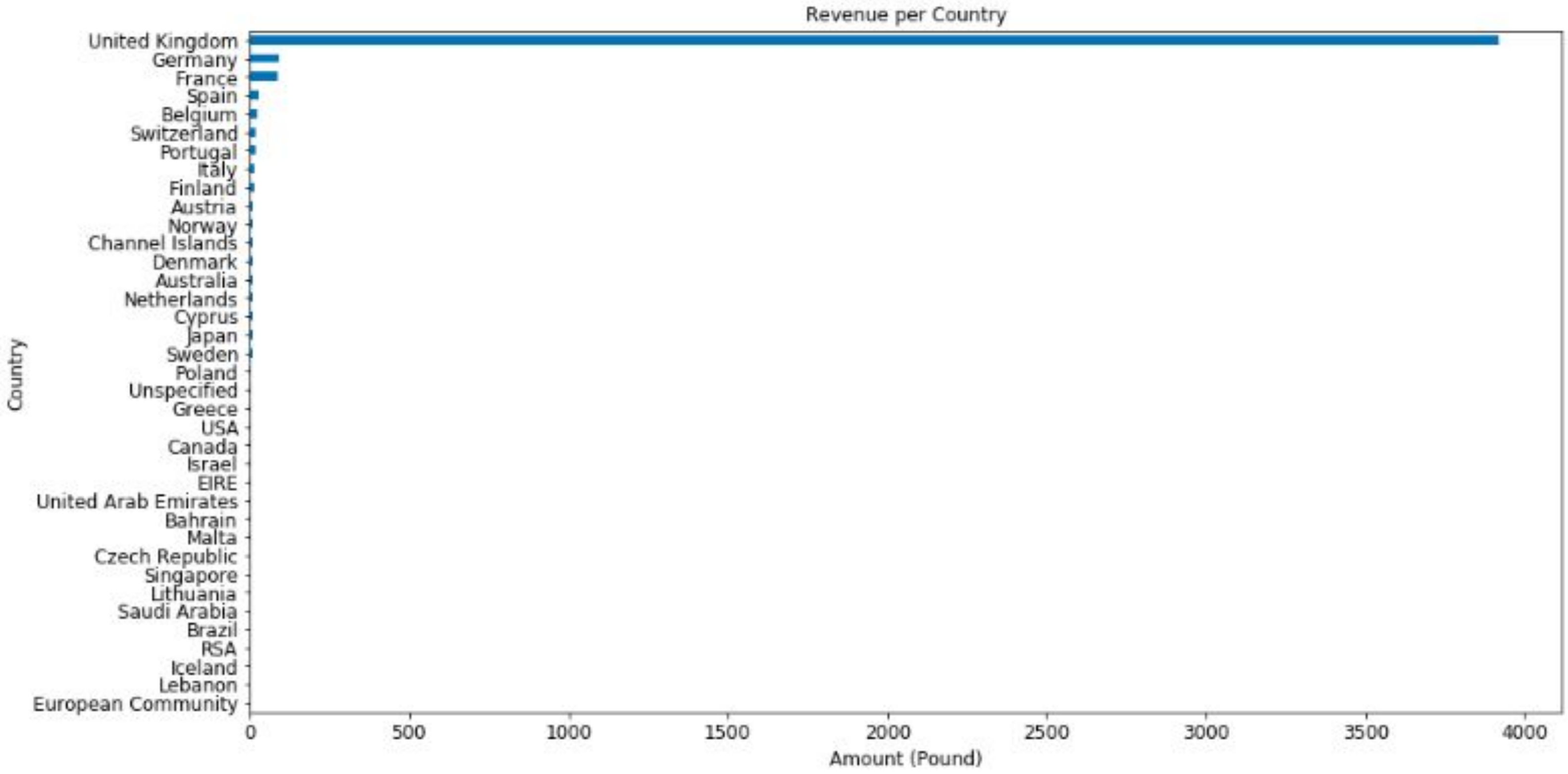
PROMOTIONAL ITEMS



COUNTRY PROFILE



COUNTRY PROFILE



PRODUCT TRENDS

BEST SELLERS

	stock_code	description	unit_price	quantity
5155	23166	medium ceramic top storage jar	1.04	76087
6699	84077	world war 2 gliders asstd designs	0.29	27528
6698	84077	world war 2 gliders asstd designs	0.21	23904
2457	22197	popcorn holder	0.72	22940
6893	84879	assorted colour bird ornament	1.69	22106
8318	85123A	white hanging heart t-light holder	2.55	19966
4930	23084	rabbit night light	1.79	19961
8291	85099B	jumbo bag red retrospot	1.79	19136
801	21212	pack of 72 retrospot cake cases	0.55	17534
3222	22492	mini paint set vintage	0.65	16888

PRODUCT TRENDS

MOST REVENUE

	stock_code	description	unit_price	amount_spent
3031	22423	regency cakestand 3 tier	10.95	93064.05
5155	23166	medium ceramic top storage jar	1.04	79130.48
8318	85123A	white hanging heart t-light holder	2.55	50913.30
3032	22423	regency cakestand 3 tier	12.75	48934.50
3245	22502	picnic basket wicker 60 pieces	649.50	39619.50
6893	84879	assorted colour bird ornament	1.69	37359.14
8894	POST	postage	18.00	36468.00
4930	23084	rabbit night light	1.79	35730.19
665	21137	black record cover frame	3.39	35161.08
8291	85099B	jumbo bag red retrospot	1.79	34253.44

SELECTING A MODELS

COLLABORATIVE FILTERING MODELS

Neighborhood Models

- Commonly used to estimate unknown ratings based on similar users historical data
- Examples: K-Nearest Neighbors, K-Clustering
- These models are not able to distinguish between user preference and the confidence for those preferences with implicit feedback

Latent Factor Models

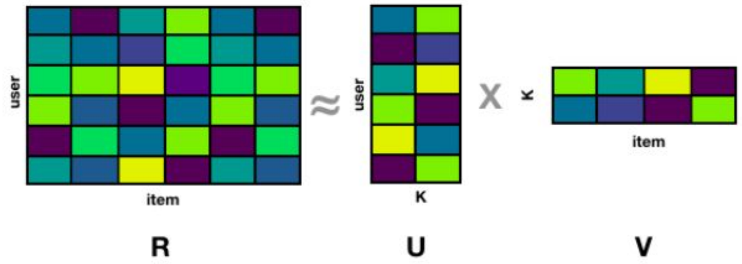
- Provides a more holistic approach to uncovering latent features that explain observed ratings
- Singular value decomposition is used to solve for user-factor vector and tem-factors vector
- Require inverting a potentially very large matrix and be computationally expensive

MODEL SELECTION: IMPLICIT FEEDBACK

Alternating Least Squares: uses a technique called “matrix factorization” to generate item recommendations for a set of users

- Based on the notion of a confidence matrix
- Involves decomposing a matrix, R, into two lower dimensional factor matrices, U and V
- Predictions are made by multiplying factor matrices U and V
- Optimization is achieved by recomputing U and then V. This process is repeated until the loss function converges

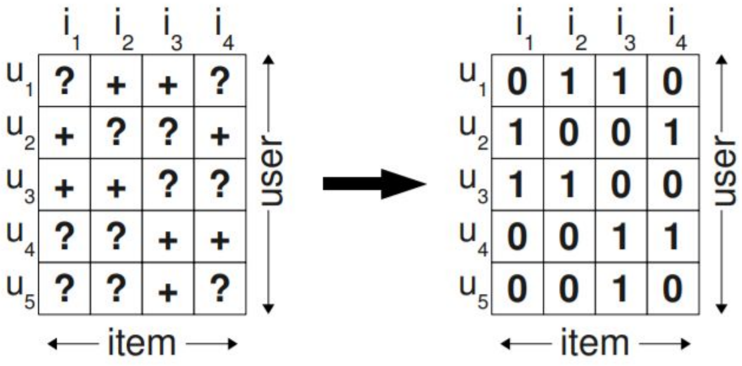
$$R_{mn} = U^T_{mk} V_{nk}$$



MODEL SELECTION: IMPLICIT FEEDBACK

Bayesian Personalized Ranking: uses a technique called “learning to rank” (LTR) to generate a personalized ranked list of items

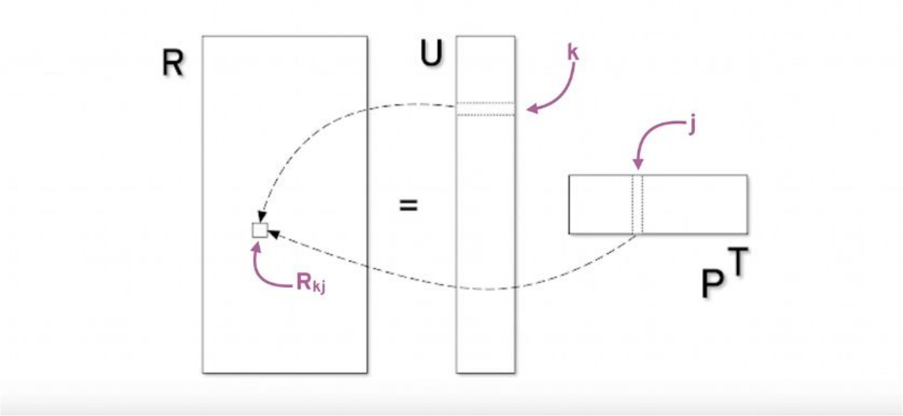
- Aims to optimize the order of a list of items
- Based on the assumption that a user prefers positive items (i.e., observed items) over non-observed items



ALTERNATING LEAST SQUARES

We ultimately chose Alternating Least Squares as our model for the recommender pipeline:

- Easier to implement
- Computationally efficient
- Generated impressive results based on a smaller proof-of-concept prototype



DATA PREPROCESSING

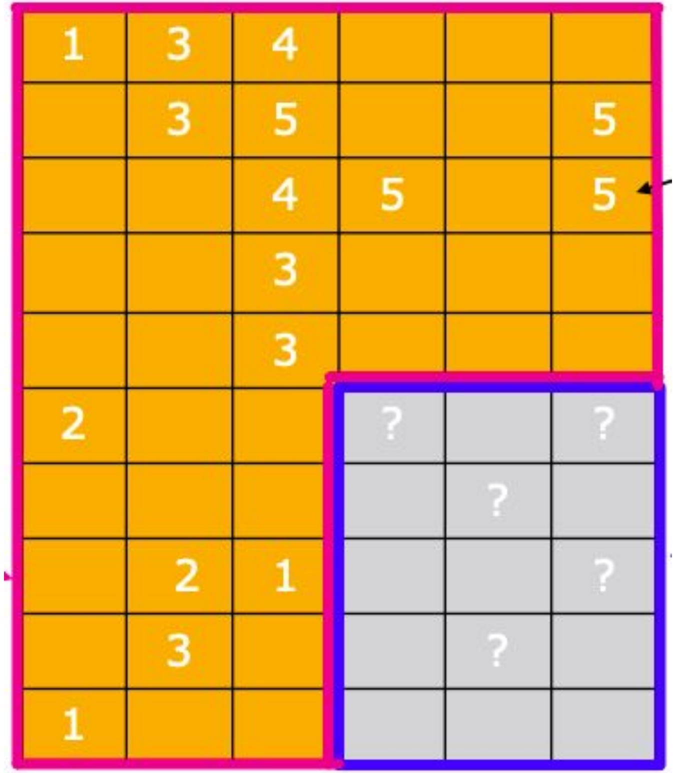
USER-ITEM MATRIX

- Filter for userID, itemID, and quantity of item
- Shape:
  - 4,339 rows (users)
  - 3,664 columns (items)
- 98.34% Sparse

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0	3	0	3	0
User 2	4	0	0	2	0
User 3	0	0	3	0	0
User 4	3	0	4	0	3
User 5	4	3	0	4	0

CREATE TRAIN AND TEST SET

- Mask certain part of training set and reset values to zero to indicate customer has not purchased the item
- Test set will contain the original values
- Return a list of altered cells



MODEL METRIC: ROC AUC

Evaluate the performance of the recommender model with ROC AUC:

- Classification task: see how many relevant items were recommended
- Relevance is defined whether the user has purchased the item or not
- Binary Variables:
    - 0: not purchased
    - 1: purchased

ALTERNATING LEAST SQUARES

- Hyperparameters:
  - Alpha = 15
  - Regularization = 0.1
  - Factors = 32
  - Iterations = 50
- Model has an auc score of 0.871
- Popularity items serve as a baseline model score of 0.813

RECOMMENDATION EXAMPLE

CUSTOMERID: 12346

	stock_code	description
31495	22258	felt farm animal rabbit

	StockCode	Description
0	22761	chest 7 drawer ma campagne
1	22264	felt farm animal white bunny
2	22247	bunny decoration magic garden
3	21654	ridged glass finger bowl
4	22694	wicker star
5	84678	classical rose small vase
6	22425	enamel colander cream
7	22393	paperweight vintage collage
8	22419	lipstick pen red
9	22782	set 3 wicker storage baskets

EXAMPLE

```
get_items_purchased(12353, product_train,
```

	stock_code	description
1087	21826	eight piece dinosaur set
3637	21158	moody girl door hanger
9102	21865	pink union jack passport cover
231161	23125	6pc wood plate set disposable

	StockCode	Description
0	84946	antique silver tea glass etched
1	85071B	red charlie+lola personal doorsign
2	22559	seaside flying disc
3	84598	boys alphabet iron on patches
4	21065	boom box speaker girls
5	22190	local cafe mug
6	22191	ivory diner wall clock
7	23186	french style storage jar cafe
8	35912B	white/pink chick decoration
9	84879	assorted colour bird ornament

CONCLUSION

SHORTCOMINGS

- Deploying an implicit model and fitting a single user-item matrix in one machine would be too memory intensive.
- A more practical method would be to implement it through Spark.
- In addition, our model is not equipped to resolve the cold start problem since we do not have previous user history. A hybrid approach would be able to incorporate feature properties and produce recommend similar items without user data.

FUTURE WORK

- Strive to improve customer experience on the website by performing customer segmentation
- Uncover insightful behavior patterns that can allow us to create a set of personalized systems for each tier of customers

REFERENCE

- https://jessesw.com/Rec-System/
- http://yifanhu.net/PUB/cf.pdf
- https://www.ethanrosenthal.com/2016/10/19/implicit-mf-part-1/
- https://implicit.readthedocs.io/en/latest/als.html