Market basket analysis

Kamel,Eliana et Ellana

Sommaire

01

Contexte

Contexte

Quantité de produit

Plusieurs Pays

Planning

To-do +

présentation

rapport EL

Do today +

Code model et requetes api EL

Streamlit

code requête bdd

voir kamel pour pas push code perso

In progress 1 / 3 +

Fast API
EL KA

Done +

Today

Wednesday, 10 May

Repo Git App KA

Veille EL EL KA

suivre tuto arturo fast api EL

Tuesday, 9 May

Repo git Api EL

choix bdd EL

mettre code bdd sur jupy EL

Connetion Mysql Workbench avec Azure

Creation BDD sur Azure EL

02
Exploration des données

Ensemble des données

Country

InvoiceNo

Quantité

StockCode

Description

CustomerID

InvoiceDate

UnitPrice

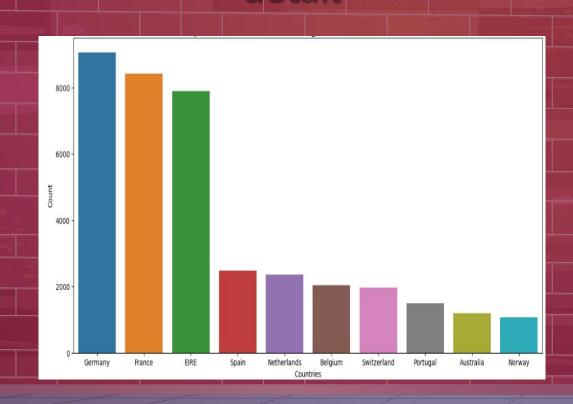# 36 pays avec 2393 transactions et 418 clients

'France', 'Australia', 'Netherlands', 'Germany', 'Norway', 'EIRE',
'Switzerland', 'Spain', 'Poland', 'Portugal', 'Italy', 'Belgium',
'Lithuania', 'Japan', 'Iceland', 'Channel Islands', 'Denmark',
'Cyprus', 'Sweden', 'Austria', 'Israel', 'Finland', 'Bahrain',
'Greece', 'Hong Kong', 'Singapore', 'Lebanon',
'United Arab Emirates', 'Saudi Arabia', 'Czech Republic',
'Canada',
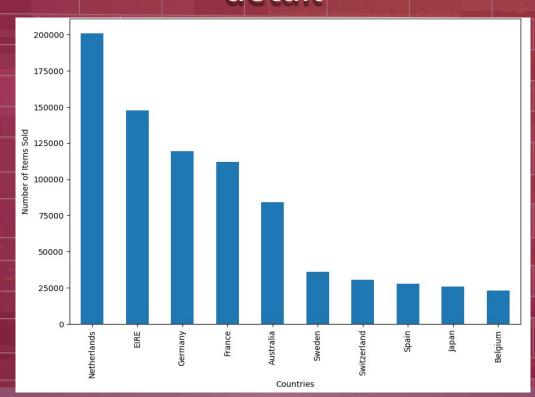'Brazil', 'USA', 'European Community', 'Malta', 'RSA'

03
Visualisation des données

Mots les plus représentés du dataset
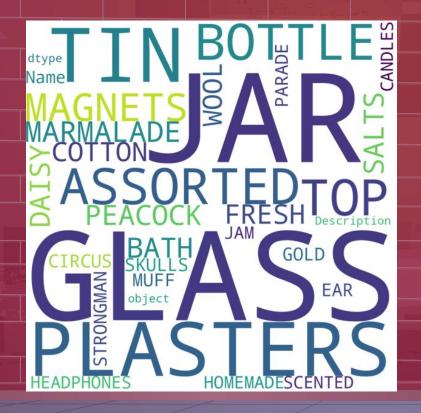
Top 10 des pays qui ont un site de vente au détail

Top 10 des pays qui ont un site de vente au détail

Mots les plus représentés pour la France
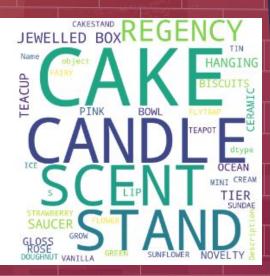
Mots les plus représentés pour la Portugal

Mots les plus représentés pour la Suède

Comparaison des mots choisi pour décrire les produits par pays (france, Portugal, Suède)

04

Modèle

Explaining Apriori metrics

- Antecedents
- Consequents
- Antecedent support
- Consequent support
- Support

- Confidence
- Lift
- Leverage
- Conviction
- Zhang's metric

Modèle choisi: Apriori (France)

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
261	(POSTAGE)	(RED TOADSTOOL LED NIGHT LIGHT, PLASTERS IN TI...	0.765306	0.051020	0.051020	0.066667	1.306667	0.011974	1.016764	1.000000
45	(POSTAGE)	(JUMBO BAG WOODLAND ANIMALS)	0.765306	0.076531	0.076531	0.100000	1.306667	0.017961	1.026077	1.000000
273	(POSTAGE)	(RED TOADSTOOL LED NIGHT LIGHT, PLASTERS IN TI...	0.765306	0.053571	0.053571	0.070000	1.306667	0.012573	1.017665	1.000000
303	(SET/6 RED SPOTTY PAPER PLATES)	(SET/6 RED SPOTTY PAPER CUPS, SET/20 RED RETRO...	0.127551	0.102041	0.099490	0.780000	7.644000	0.086474	4.081633	0.996251
305	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES, SET/20 RED RET...	0.137755	0.102041	0.099490	0.722222	7.077778	0.085433	3.232653	0.995904
345	(SET/6 RED SPOTTY PAPER PLATES)	(SET/6 RED SPOTTY PAPER CUPS, POSTAGE, SET/20 ...	0.127551	0.084184	0.081633	0.640000	7.602424	0.070895	2.543934	0.995431
347	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET/2...	0.137755	0.084184	0.081633	0.592593	7.039282	0.070036	2.247913	0.995007
24	(CHILDRENS CUTLERY DOLLY GIRL)	(CHILDRENS CUTLERY SPACEBOY)	0.071429	0.068878	0.063776	0.892857	12.962963	0.058856	8.690476	0.993846
175	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES)	0.137755	0.127551	0.122449	0.888889	6.968889	0.104878	7.852041	0.993343
87	(PACK OF 6 SKULL PAPER CUPS)	(PACK OF 6 SKULL PAPER PLATES)	0.063776	0.056122	0.051020	0.800000	14.254545	0.047441	4.719388	0.993188

Modèle choisi: Apriori (Finland)

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
198	(POSTAGE)	(SET OF 60 PANTRY DESIGN CAKE CASES, MINT KITC...	0.902439	0.073171	0.073171	0.081081	1.108108	0.007139	1.008608	1.0
66	(POSTAGE)	(PANTRY PASTRY BRUSH)	0.902439	0.073171	0.073171	0.081081	1.108108	0.007139	1.008608	1.0
106	(CHILDRENS CUTLERY POLKADOT BLUE)	(CHILDRENS CUTLERY POLKADOT PINK, CHILDRENS CU...	0.170732	0.097561	0.097561	0.571429	5.857143	0.080904	2.105691	1.0
179	(CHILDRENS CUTLERY POLKADOT BLUE, POSTAGE)	(CHILDRENS CUTLERY POLKADOT PINK, CHILDRENS CU...	0.170732	0.097561	0.097561	0.571429	5.857143	0.080904	2.105691	1.0
186	(CHILDRENS CUTLERY POLKADOT BLUE)	(CHILDRENS CUTLERY POLKADOT PINK, POSTAGE, CHI...	0.170732	0.097561	0.097561	0.571429	5.857143	0.080904	2.105691	1.0
117	(POSTAGE)	(CHILDRENS CUTLERY POLKADOT BLUE, CHILDRENS CU...	0.902439	0.146341	0.146341	0.162162	1.108108	0.014277	1.018883	1.0
0	(POSTAGE)	(60 CAKE CASES DOLLY GIRL DESIGN)	0.902439	0.073171	0.073171	0.081081	1.108108	0.007139	1.008608	1.0
4	(POSTAGE)	(60 TEATIME FAIRY CAKE CASES)	0.902439	0.073171	0.073171	0.081081	1.108108	0.007139	1.008608	1.0
6	(POSTAGE)	(BAKING SET 9 PIECE RETROSPOT)	0.902439	0.073171	0.073171	0.081081	1.108108	0.007139	1.008608	1.0
10	(POSTAGE)	(BLUE POLKADOT BEAKER)	0.902439	0.073171	0.073171	0.081081	1.108108	0.007139	1.008608	1.0

05

Conclusion / Perspectives

Conclusion / Perspectives

Comme constaté précédemment, le modèle Apriori a posé problème, il serait judicieux de tenter d'utiliser le modèle FP-Growth, qui prend moins de données que Apriori, mais est plus rapide, et est plus facilement interprétable.