

Sales data analytics_retail

Data analyst Xuefei ZHANG
06-2022

Objectif

Analyse descriptive et exploratoire de data commerciale d'une entreprise française E-commerce dans le but d'optimiser la gestion de stockage

Questions à répondre

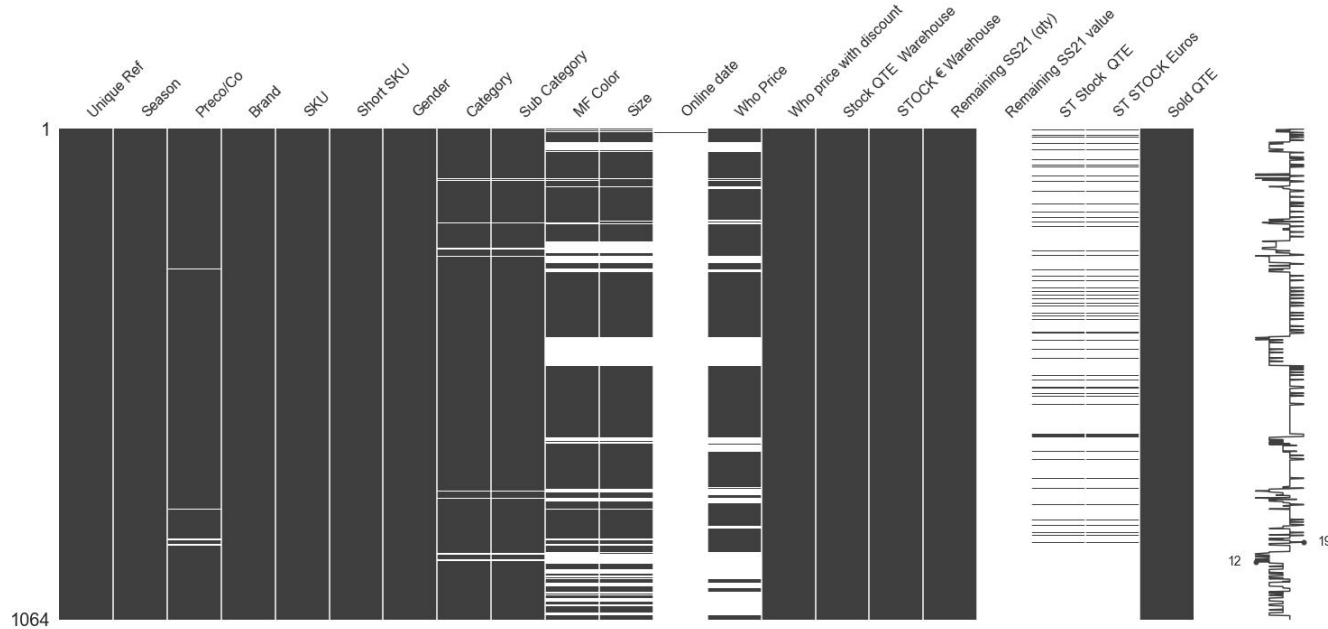
1. **Sell-Through Rate (STR)** par marque et par genre et par catégories
2. Quelle est la **marque** la plus performante ? Pourquoi ?
3. Les **catégories** plus vendues en quantité
4. Y'a-t-il des **SKU référence produit à risque** ? Critères de choix et listez-les
5. **Recommandation d'ajouts de KPIs** utiles pour l'entreprise

Aperçu de données - sales recap

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Unique Ref	Season	Preco/Co	Brand	SKU	Short SKU	Gender	Category	Sub Category	MF Color	Size	Online date	Who Price	Who price with discount	Stock QTE Warehouse	STOCK € Warehouse	Remaining SS21 (qty)
1	PNR	CO	X	UZI015000	UZI015000	UNISEX	BAGS	BACKPACKS	BLACK	FR41	#ERROR!	185 €	185 €	0	0 €	0
1	PNR	CO	X	UZI015001	UZI015001	UNISEX	BAGS	BELT BAGS & FANN'	BLACK	0	#ERROR!	150 €	150 €	24	3,600 €	1
1	OUTLET	CO	X	UZI015003	UZI015003	WOMEN	JEWELRY	NECKLACES	N/A	N/A	#ERROR!	205 €	205 €	70	14,350 €	2
1	OUTLET	CO	X	UZI015006	UZI015006	WOMEN	ACCESSORIES	TECH	N/A	N/A	#ERROR!	75 €	75 €	0	0 €	0
1	PNR	CO	X	UZI015008	UZI015008	UNISEX	BAGS	BACKPACKS	BLACK	0	#ERROR!	220 €	220 €	0	0 €	0
1	OUTLET	CO	X	UZI015303	UZI015303	WOMEN	ACCESSORIES	TECH	N/A	N/A	#ERROR!	50 €	50 €	12	600 €	6
1	OUTLET	CO	X	UZI015304	UZI015304	WOMEN	ACCESSORIES	TECH	N/A	N/A	#ERROR!	50 €	50 €	0	0 €	0
1	OUTLET	CO	X	UZI015306	UZI015306	WOMEN	BAGS	CLUTCHES	N/A	N/A	#ERROR!	285 €	285 €	0	0 €	0
1	FW21	PRECO	X	UZI016000	UZI016000	WOMEN	ACCESSORIES	BELTS	PINK	0	9/5/1900	100 €	130 €	0	0 €	0
0	FW21	PRECO	X	UZI016001000	UZI016001	UNISEX	ACCESSORIES	BELTS	BLACK	0	#ERROR!	100 €	100 €	30	3,000 €	24
1	FW21	PRECO	X	UZI016001001	UZI016001	UNISEX	ACCESSORIES	BELTS	BLACK	0	#ERROR!	100 €	100 €	0	0 €	0
0	FW21	PRECO	X	UZI016001002	UZI016001	UNISEX	ACCESSORIES	BELTS	BLACK	0	#ERROR!	100 €	100 €	0	0 €	0
0	FW21	PRECO	X	UZI016002000	UZI016002	WOMEN	ACCESSORIES	BELTS	PINK	0	#ERROR!	100 €	100 €	0	0 €	0
1	FW21	PRECO	X	UZI016002001	UZI016002	WOMEN	ACCESSORIES	BELTS	PINK	0	#ERROR!	100 €	100 €	0	0 €	0
0	FW21	PRECO	X	UZI016002002	UZI016002	WOMEN	ACCESSORIES	BELTS	PINK	0	#ERROR!	100 €	100 €	70	7,000 €	0
1	FW21	PRECO	X	UZI016003	UZI016003	UNISEX	ACCESSORIES	HATS	BLACK	0	#ERROR!	115 €	115 €	0	0 €	0
1	FW21	PRECO	X	UZI016004	UZI016004	UNISEX	ACCESSORIES	HATS	PINK	0	#ERROR!	80 €	80 €	0	0 €	0

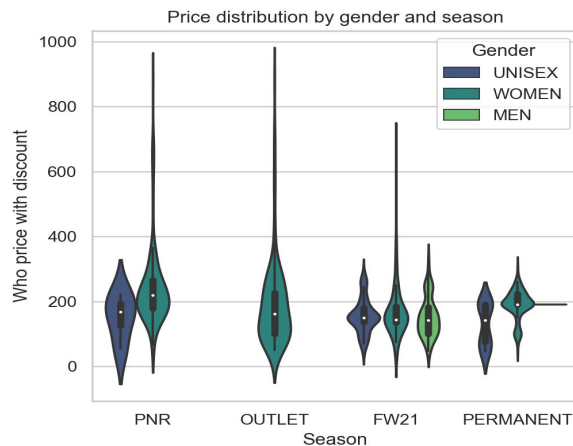
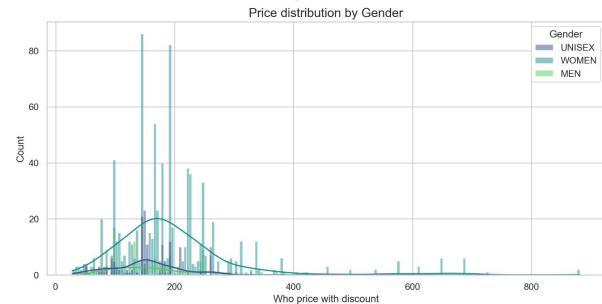
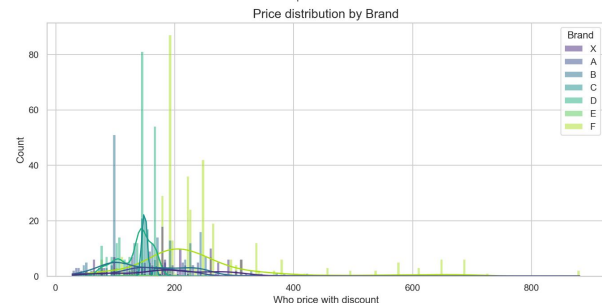
Aperçu de l'intégrité du jeu de données

*valeurs manquantes (NaN) sont représentées en blanc



Manque de valeurs dans certaines colonnes, mais cela n'impacte pas notre objectif

Comment se répartissent les prix (par catégorie/marque/sexe/saison)



- Les prix se varient en fonction de catégorie/marque/genre
- En générale, il n'y a pas grande différence en terme de prix de produits promus dans ces 4 saisons de promo.
- Alors que le prix de différents sexes se varient parmi les saisons.

Comment prix et stockage se varient (en fonction de catégorie)

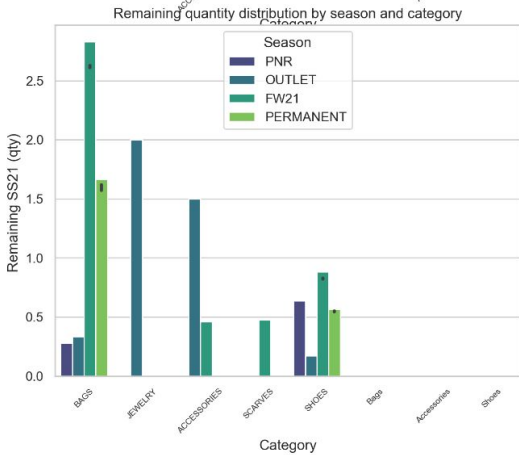
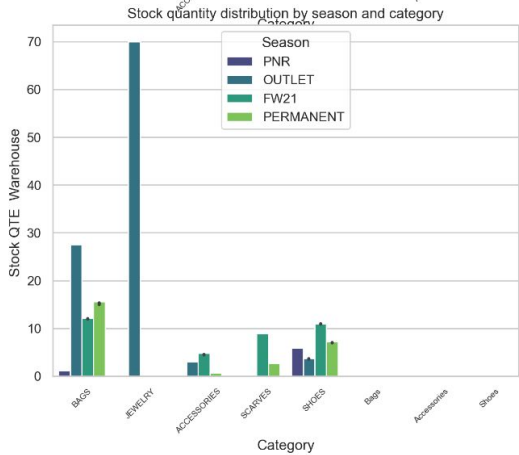
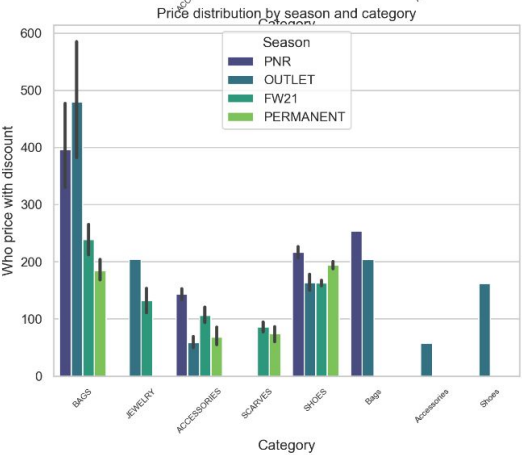
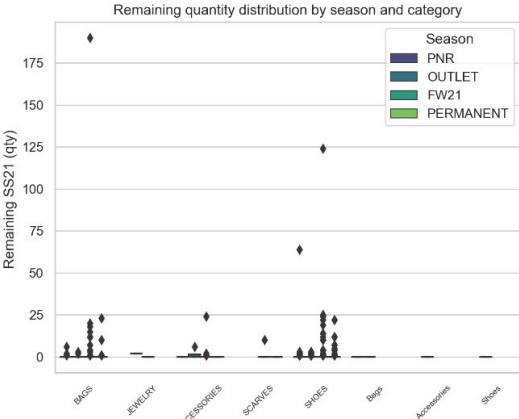
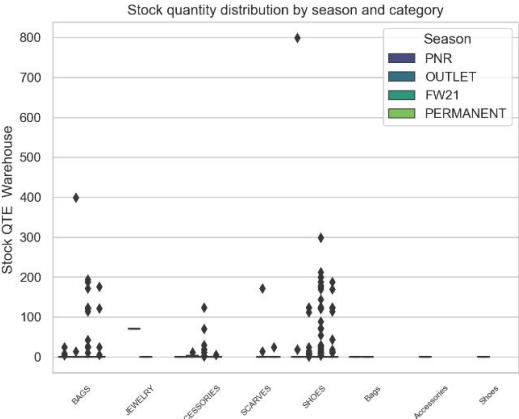
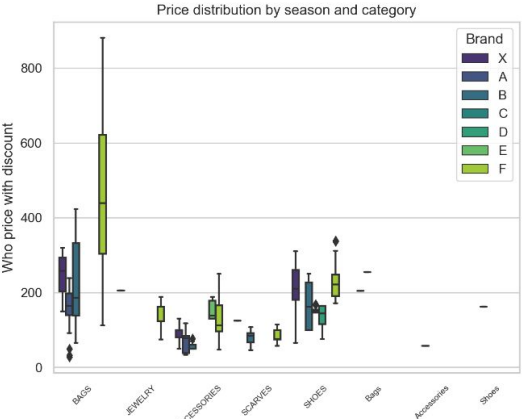


Tableau croisé dynamique - stockage par catégorie

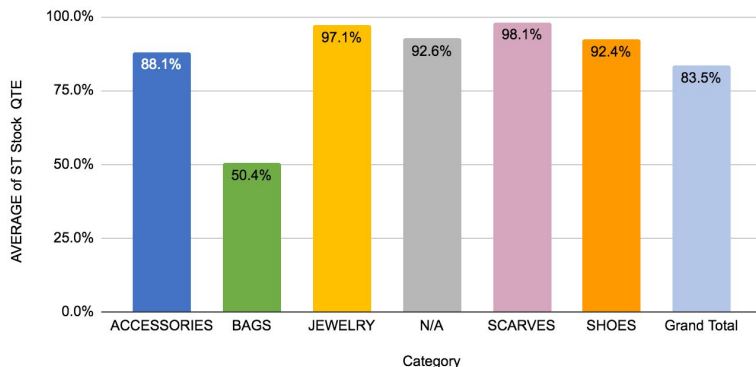
Category	Sub Category	AVERAGE of ST Stock QTE	MIN of ST Stock	MEDIAN of ST St	MAX of ST Stock
[-] ACCESSORIES	BELTS	60.0%	20.0%	60.0%	100.0%
	EYEWEAR	94.0%	66.7%	100.0%	100.0%
	HATS	97.2%	94.4%	97.2%	100.0%
	TECH	87.1%	50.0%	99.2%	100.0%
ACCESSORIES Total		88.1%	20.0%	100.0%	100.0%
[-] BAGS	BELT BAGS & FAN	95.8%	95.8%	95.8%	95.8%
	CLUTCHES	50.0%	50.0%	50.0%	50.0%
	CROSSBODIES	89.7%	53.5%	94.4%	100.0%
	HANDBAGS & PU	78.6%	78.6%	78.6%	78.6%
	SHOULDER & HC	1.4%	-691.7%	93.9%	100.0%
	TOTES & SHOPPI	25.0%	25.0%	25.0%	25.0%
BAGS Total		50.4%	-691.7%	93.0%	100.0%
[-] JEWELRY	NECKLACES	97.1%	97.1%	97.1%	97.1%
JEWELRY Total		97.1%	97.1%	97.1%	97.1%
[-] N/A	N/A	92.6%	85.7%	92.6%	99.4%
N/A Total		92.6%	85.7%	92.6%	99.4%
[-] SCARVES	SCARVES	98.1%	94.2%	100.0%	100.0%
SCARVES Total		98.1%	94.2%	100.0%	100.0%
[-] SHOES	ANKLE BOOTS	95.6%	82.0%	98.5%	100.0%
	BOOTS	80.7%	53.7%	87.6%	94.0%
	FLAT SHOES	93.0%	58.5%	99.2%	100.0%
	SANDALS	100.0%	100.0%	100.0%	100.0%
	SLIDES	86.4%	64.3%	91.0%	99.2%
	SNEAKERS	90.5%	48.1%	96.4%	100.0%
SHOES Total		92.4%	48.1%	97.8%	100.0%
Grand Total		83.5%	-691.7%	96.9%	100.0%

Category	SUM of Sold QTE	SUM of Remainii	SUM of Stock QT	AVERAGE of ST Stock QTE
ACCESSORIES	281	35	316	88.1%
BAGS	1,657	329	1986	50.4%
JEWELRY	68	2	70	97.1%
N/A	188	3	191	92.6%
SCARVES	200	10	210	98.1%
SHOES	5,601	454	6055	92.4%
Grand Total	7,995	833	8828	83.5%

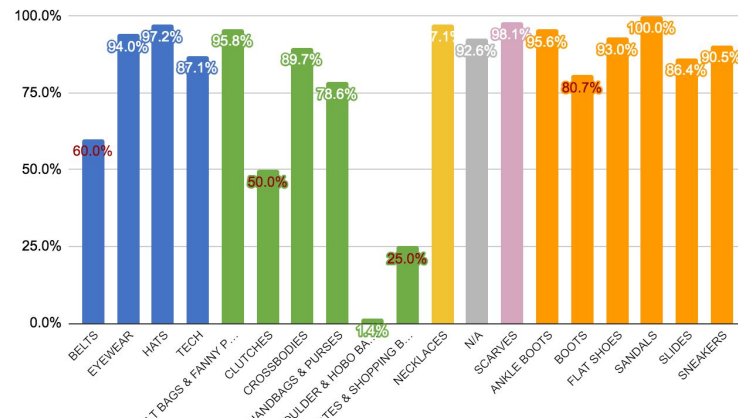
*non-exhaustif. Réf excel pour plus d'info.

STR par catégorie

1. STR per Category (average)



2. STR per sub-category (average)

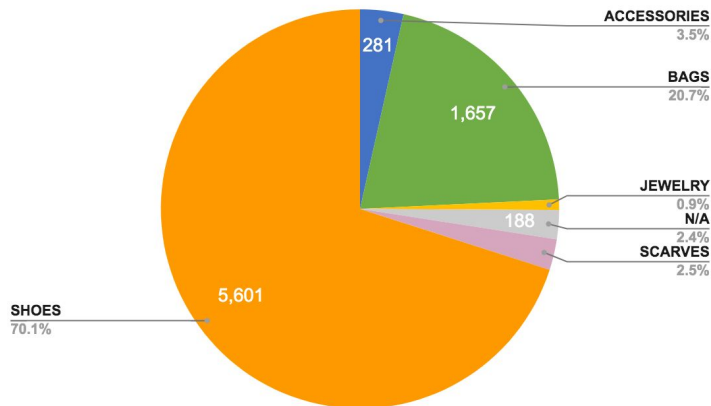


Graphes 1 et 2 Observations:

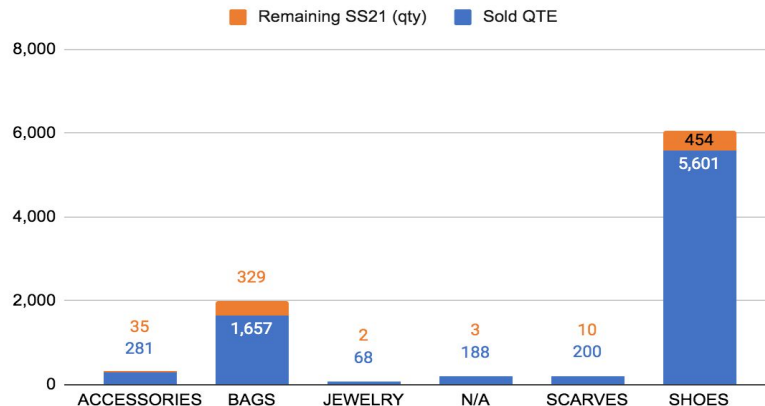
1. les ST Stock QTE varient beaucoup en fonction de Category
2. Dans la catégorie BAGS, la performance des ses Sub Category se varient beaucoup.
3. les produits de catégorie BAGS a sa performance au dessous de la moyenne de l'ensemble des produits.
4. Si l'on creuse dans les Sub Category, on constate que:
 - A. pour les BAGS, les sous-catégorie SHOULDER & HOBO BAGS (1.4%), TOTES & SHOPPING BAGS (25%), CLUTCHES (50%) et HANDBAGS & PURSES (78.7%) proportionnellement moins vendus par rapport aux stockage.
 - B. pour les ACCESSORIES, BELTS (60%) est moins performante.
 - C. pour les SHOES, sous-catégorie BOOTS (80.7%) est moins performante.

Performance (quantité vendu et restante) par catégorie

3. Sold quantity per Category



4. Sold and Remaining quantity per Category



Graphes 1 + 3 + 4 Observations:

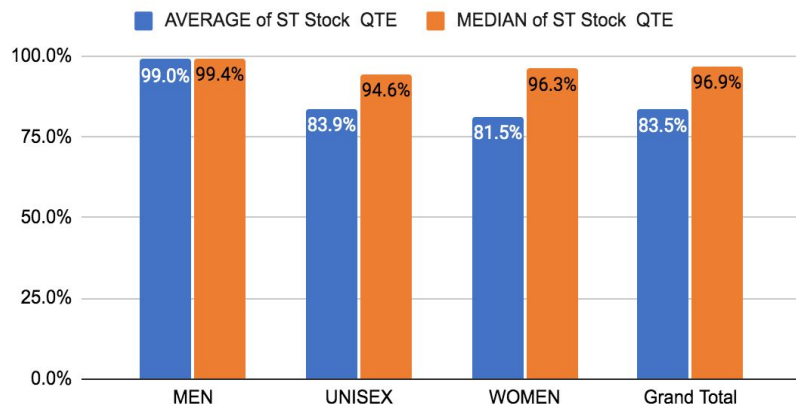
1. **SHOES est la catégorie plus vendue : 5562 durant la période qui représente 70% de vendus en terme de quantité**
2. suivi par BAGS qui a 1657 produits vendus, mais BAGS a le plus petit ST ratio - catégorie la moins performante
3. JEWELRY et SCARVES ont les ST plus élevés , mais ces deux catégories n'ont pas beaucoup de produits.

C'est ainsi qu'elles n'occupent pas bcp de proportion en terme de Sold quantity.

=> Prochainement on peut ajouter une métrique "Sold value", comme ça on peut avoir un aperçu sur la répartition de CA apportées par chaque catégorie.

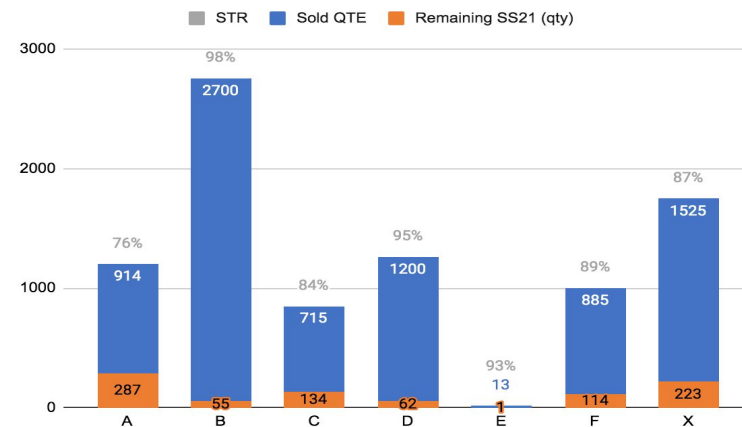
STR par sexe et marque

STR (ST Stock QTE) by Gender



MEN est plus performant,
WOMEN est le moins performant au courant de la période passée.

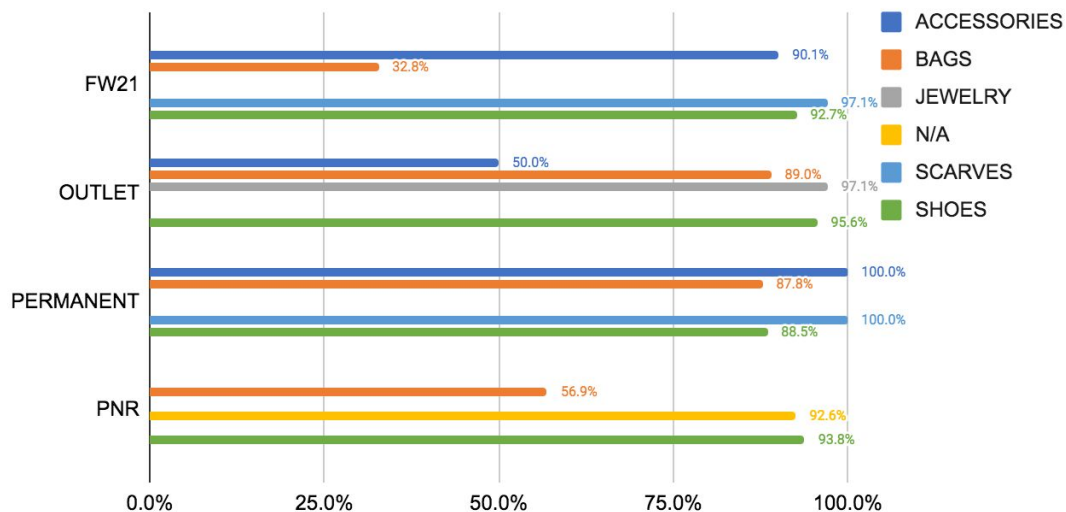
Sold and Remaining (quantity), Sell-Through Rate per Brand



1. **Brand B** a la meilleure **STR** performance - **98%**,
B est aussi la plus vendue en terme de **quantité** - 2700 sur 2755
2. le stockage de Brand A ne s'est pas bien vidé vu son ST ratio relativement bas : 76% , ctd. 23% de ses produits équivalant 287 articles sont pas vendus à la fin de saison.
3. Brand A, C, F et X ne sont pas assez satisfaisants en terme de STR

STR par catégorie et saison

1. STR of each Category per Season



1. JEWELRY ne sont que vendu durant la saison OUTLET, et elle connaît un RT élevé
2. En générale, **SHOES** connaît ses performances (STR) similaires pour tous ces saisons. Mais se varient un peu: le meilleur résultat STR durant OUTLET - 95.6%, par contre en saison PERMANENT relativement bas - 88.5%
3. ACCESSORIES et SCARVES connaissent RT élevés en saison PERMANENT et FW21
4. OUTLET ne fonctionne pas bien pour ACCESSORIES - 50%
5. BAGS a sa performance relativement bien en saison OUTLET et PERMANENT (> 87%) que en FW21(32%) et PNR (56%).

la marque la plus performante: **SHOES**

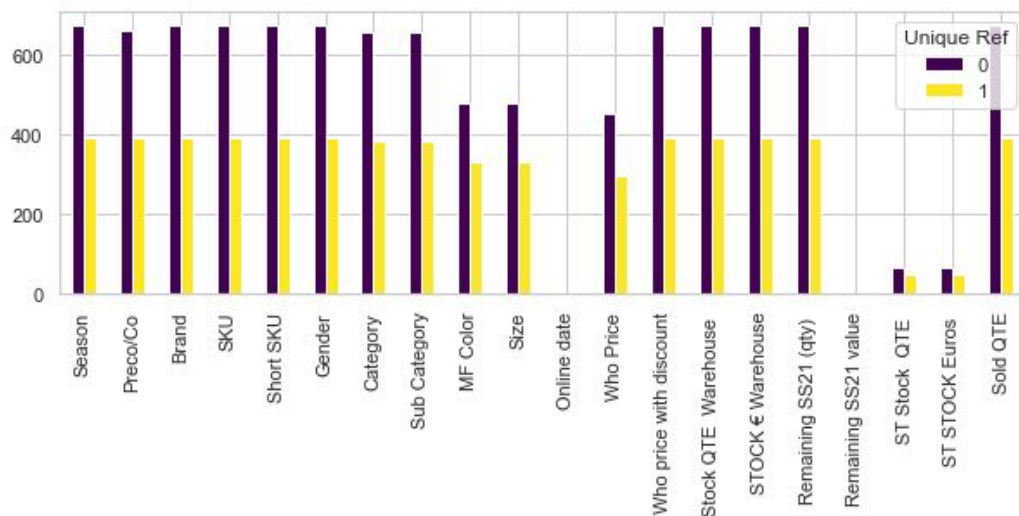
SKU référence produit à risque

0 sont des doublons, ctd. les produits qui ont seulement leur **SKU unique** mais **répétitif** quant à **Short SKU**
1 sont des produits qui ont leur **SKU et Short SKU unique**

Critère pour doublons:

Unique Ref = 0 and SKU != Short SKU

J'ai identifié **672** doublons.



Pour Category et Sub Category, le nombre de valeurs non-null sont inférieur (654,384) que ce qu'il fallait (672, 392) dans SKU, vu que SKU est la référence unique pour un produit.

Autrement dit, il y a des produits n'ayant pas de Category ni Sub Category - à risque.

Recommandation d'ajouts de KPIs et les raisons

Gross product margin = $(\text{Sold price} - \text{Purchase price}) / \text{Purchase price}$	permet de savoir la marge (%) par produit
Refund and Return Rate = $\text{Returned product quantity} / \text{Sold quantity}$	permet de mesurer le niveau de satisfaction de ce produit chez les clients.
Taille du panier moyen (Average Basket Size) = $\text{Nombre total d'articles vendus durant une période} / \text{Nombre de commandes}$	à croiser avec tableau sur transaction (commandes). clé primaire : SKU
Panier moyen (Average Basket Value) = $\text{CA} / \text{Nombre de commandes}$	valeur totale par action d'achat
Customer Lifelong Value	à utiliser croisé avec tableau sur les clients . éventuelle Clé primaire pour croisement: SKU. variables pour formule de calcul: Panier moyen, Retention rate , Discount rate, nombre de périodes à prédire
Fréquence d'achat par Gender sur une période	
Indice de Gini	mesure de concentration. ie. un Gini de 0.6 pour CA : 60% de CA est apportée par 40% de clients. A croiser avec tableau sur clients (client ID, genre, âge, pays...)



Merci