

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Partie 0 : Mise en place de l'environnement		
Importation des librairies	Import panda, plotly, seaborn, numpy, matplotlib	1
Afficher toutes les colonnes du Dataframe	pd.set_option('display.max_columns', None)	2
Charger les fichiers	Chargement des fichiers web, erp, liaison	3
Partie 1 : Analyse exploratoire des données		
Analyse du fichier erp		
Dimensions	6 colonnes, 825 observations	4
Nature et valeur des colonnes	(voir cellule)	5
Vérification doublons product_id	Pas de doublons	7
Valeurs de stock_status	instock, outofstock	8
Lien stock_status	Varie avec stock_quantity	8
Créer colonne stock_status_2	df_erp['stock_status_2'] = np.where(condition_outofstock, 'outofstock', 'instock')	9
Comparaison stock_status vs stock_status_2	2 incohérences détectées	11
Identification incohérence	ID : 4039 stock 3 signalé outofstock ID : 4885 stock 0 marqué instock	12

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Correction incohérence	Alignement stock_status sur stock_quantity (anomalies)	13
Variable Prix non renseignés, négatifs ou nuls ?	Nan : 0 Prix <= 0 : 3 (4233 à -20, 5017 à -8, 6594 à -9,1) Suppression des prix négatifs (anomalies)	14
Prix min et maximum	Min : -20 Max : 225	14
Quantité stock	Min : -10 Max : 145	15
Stock négatif	2 anomalies trouvées : - 4973 : -10 - 5700 : -1 Suppression des stocks négatifs (anomalies)	15
Valeurs uniques de onsale_web	0, 1 (1 dispo en vente sur le site, 0 pas dispo)	16
Répartitions valeurs onsale_web	- 1 : 716 - 0 : 109	16
Afficher les colonnes de erp	print(df_erp.columns)	17
Choix des colonnes à conserver	(voir cellule)	17
Suppression colonne stock_status_2	(voir cellule)	18
Prix d'achat non renseignés	0	20
Prix d'achat min et max	Min : 2,74 Max : 137,81 (Pas d'anomalies ici)	20
Analyse du fichier web		

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Dimensions	- Colonnes : 29 - Observations : 1513	21
Nature et valeurs des colonnes	(voir cellule)	22
Colonnes à conserver	(voir cellule)	23
Suppression des colonnes	(voir cellule)	24
Règles de codification du SKU	Format attendu : chiffres entre 1 et 20 Règle : regex_sku_valide = r"^\d{1,20}\$"	25
Anomalies du SKU	- 85 valeurs Nan - 2 anomalies de format : 13127-1, bon-cadeau-25-euros	25
Analyse et corrections des SKU anormaux	- Demande clarification métier pour SKU anormaux - Suppression possible après vérif des SKU Nan	26
SKU unique ?	- Non, 1513 ont des SKU dupliqués - SKU distincts qui apparaissent plus d'une fois : 714 (tous 2 fois)	27
Vérification contenu SKU non renseignés	- 83 observations entièrement vide (anomalies à supprimer) - 2 avec données partielles (besoin clarification métier)	28
Analyse du fichier liaison		
Dimensions	- Colonnes : 2 - Observations : 825	29
Nature et valeurs des colonnes	- id_web : 734 non nulle (object) - product_id : 825 non nulle (int64) Id_web format object : anomalie type de données	30
Unicité de product_id	Les valeurs de product_id sont toutes uniques	31

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Unicité de web_id	- Les valeurs non Nan de id_web sont toutes uniques - 91 valeurs manquantes	32
Partie 2 : Fusion des fichiers		
Jonction du fichier df_erp et df_liaison		
Fusion left join fichier df_erp et df_liaison : df_merge	(voir cellule)	33
Jonction du fichier df_merge et df_web		
Nettoyage des lignes vides dans df_web	Suppression des 85 lignes avec SKU = Nan	34
Dédoublonnage de df_web	Utilisation de post_type = product pour dédoublonner	34
Filtrage de df_merge	Exclusion des liens sans id_web valide	34
Jonction inner join du fichier df_merge et df_web : df_final	(voir cellule)	34
Suppression des colonnes inutiles	- Suppression colonne id_web (redondance avec SKU) - Suppression colonne temporaire merge	34
Caractéristiques df_web post traitement	- Colonnes : 7 - Observations : 715	34
Caractéristiques df_merge post filtrage	- Colonnes : 8 - Observations : 734	34
Caractéristiques df_final	- Colonnes : 15 - Observations : 714	34

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Partie 3 : Analyse de la cohérence des données		
Cohérence entre erp et liaison		
Les 91 web_id manquants correspondent ils aux produits marqués offline dans erp ?	Seulement une partie des 91 web_id manquants correspondent à offline (rappel : 109 produits marqués offline dans erp)	35
Nombre de produits supposés offline mais avec id_web ?	21 (voir cellule pour liste)	36
Produits marqués online mais sans id_web	3 anomalies : - 4594, 5070, 5075 - Explication probable : produits sans stock donc pas mis	37
Cohérence entre erp et final		
Les produits marqués offline sont ils présents dans le fichier web et final ?	Oui, 1 anomalie trouvée : - Product_id : 4200, sku : 16295	38
CA du produit offline	81,2 (à enlever de nos calculs)	39
Comparaison nombre items dans erp vs final	- Erp : 716 - Final : 714	41
Différence produits entre les deux tables	La table final a 2 anomalies : - 3 produits marqués online sont absents : 4594, 5070, 5075 - 1 produit marqué offline est présent : 4200	41
Cohérence entre erp et web		
Existence de produits web only ?	Il n'y a pas de produits web only	42
Partie 4 : Data visualisation		

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Analyse des prix		
Prix : boxplot en utilisant pandas	(voir cellule)	43
Prix : boxplot en utilisant plotly express	(voir cellule)	44
Moyenne prix	32,33€	45
Écart type prix	27,60€	45
Z-score prix	Z-score 3 = 115,12 €	46
Interval interquartile prix	(voir cellule)	47
Calcul des outliers	(voir cellule)	48
Nombre d'outliers	31 articles soit 4,34% du catalogue	49
Analyse des outliers	Pour prix élevés, produits de luxe donc prix cohérents. Pour prix négatif, correction des anomalies.	50
Analyse du chiffre d'affaires		
Chiffre d'affaire total du site web	143,680.10 €	51
Visualisation : top 20 produits par chiffre d'affaire	(voir cellule)	52
Pareto du CA	Nombre d'articles représentant environ 80% du CA : 435 Ces 435 articles représentent 60.92 % du catalogue total.	53
Analyse de la quantité de produits vendus		

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Visualisation : top 20 produits par quantité vendue	(voir cellule)	54
Pareto de la quantité de produits vendus	Nombre d'articles représentant environ 80% des ventes en quantité : 434. Ces 434 articles représentent 60.78 % du catalogue total.	56
Analyse des stocks		
Visualisation : flop 20 des produits par temps de rotation > 0	(voir cellule) 3 produits dormants : 4337, 4355, 5932 (2 champagnes un cognac)	57
Temps de rotation global des stock	3,44 mois	58
Quantité total de stock disponible en ligne	16740	59
Valorisation totale des stocks	277,328.07 €	60
Analyse des marges		
Création de la colonne taux_marge (tva 20%)	(voir cellule)	61
Taux de marge min et max	- Min : -634.99% (anomalie) - Max : 47.76%	61
Identification des produits avec taux de marge négatifs	ID : 4355, prix achat : 77,48, prix vente (ttc) : 12.65	62
Nombre de produits avec une marge positive	713	63
Marge min et max	- Min : 22.78% - Max : 47.76%	63
Visualisation : Taux de marge moyen par type de produits	(voir cellule) Nous vendons aussi de l'huile d'olive mais pas très rentable	64

Rapport d'analyse business		
Instructions [1]	Résultats	Cellule
Analyse des corrélations		
Visualisation : Heatmap corrélation stock, prix et ventes	(voir cellule)	65
Analyse des corrélations	(voir cellule)	66

[1] Autant que possible, formulez vos actions sous forme de compétences.

Comment bien formuler une compétence :

[verbe d'action] + [object de l'action] + [contexte de l'action]