

Presentation : Analyse des ventes de votre entreprise

Bilel Hemmara

Plan de la presentation :

- 1 - Détail du nettoyage des données. (10 minutes)**
- 2 - Présentation de l'analyse demandée. (10 minutes)**
- 3 - Présentez votre analyse des corrélations. (10 minutes)**
- 4 - Séance de questions-réponses éventuelle.**

Support de presentation : Streamlit

Mission 1 : Nettoyage des données

quelles valeurs aberrantes et manquantes avez-vous trouvées, comment les avez-vous traitées, avez-vous effectué d'autres nettoyages ? (10 minutes)

I - Dataset Products

Presentation du dataset

	id_prod	prix_prod	categorie_prod
0	0_1421	19.9900	0
1	0_1368	5.1300	0
2	0_731	17.9900	0
3	1_587	4.9900	1
4	0_1507	3.9900	0
5	0_1163	9.9900	0
6	1_463	36.9900	1
7	0_2157	34.9900	0
8	0_1915	16.9900	0
9	0_389	18.9900	0
10	2_228	225.1700	2

Etude du dataset (optionnel)

☐ Dataset

Vérification du dataset

```
id_prod      T_0
prix_prod    -1
categorie_prod  0
Name: 731, dtype: object
```

Valeur aberrante : Prix < 0

II - Dataset customers

Presentation du dataset

	client_id	sex_client	naissance_client	age_client
0	c_4410	Femme	1967	54
1	c_7839	Femme	1975	46
2	c_1699	Femme	1984	37
3	c_5961	Femme	1962	59
4	c_5320	Homme	1943	78
5	c_415	Homme	1993	28
6	c_285	Femme	1967	54
7	c_160	Femme	1978	43
8	c_6446	Homme	1971	50
9	c_6866	Homme	1993	28
10	c_1013	Homme	1982	39

Etude du dataset (optionnel)

☐ Dataset

Vérification du dataset

Verification OK

III - Dataset transactions

Presentation du dataset

id_prod	date	session_id	client_id
---------	------	------------	-----------

0	0_1483	2021-04-10 18:37:28.723910	s_18746	c_4450
1	2_226	2022-02-03 01:55:53.276402	s_159142	c_277
2	1_374	2021-09-23 15:13:46.938559	s_94290	c_4270
3	0_2186	2021-10-17 03:27:18.783634	s_105936	c_4597
4	0_1351	2021-07-17 20:34:25.800563	s_63642	c_1242
5	0_1085	2021-09-15 05:47:48.215162	s_90139	c_2526
6	0_1508	2021-08-29 05:39:01.055455	s_82100	c_5799
7	0_1627	2021-09-19 14:54:52.410362	s_92294	c_1422
8	0_1469	2022-02-03 07:13:22.559995	s_159252	c_2207
9	0_1453	2022-02-26 09:03:10.571665	s_171098	c_5433
10	0_423	2021-10-09 23:50:33.907269	s_102501	c_2225

Etude du dataset (optionnel)

☐ Dataset

Vérification du dataset

Nombre de doublons important dans le dataset (126)

Doublons

	id_prod	date	session_id	client_id
34387	T_0	test_2021-03-01 02:30:02.237443	s_0	ct_0
54813	T_0	test_2021-03-01 02:30:02.237412	s_0	ct_1
57261	T_0	test_2021-03-01 02:30:02.237439	s_0	ct_1
58802	T_0	test_2021-03-01 02:30:02.237429	s_0	ct_0
60170	T_0	test_2021-03-01 02:30:02.237446	s_0	ct_0
62365	T 0	test 2021-03-01 02:30:02.237412	s 0	ct 1

	-	-	-	-
71097	T_0	test_2021-03-01 02:30:02.237446	s_0	ct_0
72405	T_0	test_2021-03-01 02:30:02.237424	s_0	ct_1
75176	T_0	test_2021-03-01 02:30:02.237443	s_0	ct_1
76735	T_0	test_2021-03-01 02:30:02.237442	s_0	ct_0
80189	T_0	test_2021-03-01 02:30:02.237414	s_0	ct_1

T_0

Toutes les transactions du produit T_0

	id_prod	date	session_id	client_id
1431	T_0	test_2021-03-01 02:30:02.237420	s_0	ct_1
2365	T_0	test_2021-03-01 02:30:02.237446	s_0	ct_1
2895	T_0	test_2021-03-01 02:30:02.237414	s_0	ct_1
5955	T_0	test_2021-03-01 02:30:02.237441	s_0	ct_0
7283	T_0	test_2021-03-01 02:30:02.237434	s_0	ct_1
13745	T_0	test_2021-03-01 02:30:02.237443	s_0	ct_0
20470	T_0	test_2021-03-01 02:30:02.237442	s_0	ct_0
22347	T_0	test_2021-03-01 02:30:02.237412	s_0	ct_1
26359	T_0	test_2021-03-01 02:30:02.237439	s_0	ct_1
26407	T_0	test_2021-03-01 02:30:02.237426	s_0	ct_0
26814	T_0	test_2021-03-01 02:30:02.237447	s_0	ct_1

(200, 4)

Hypothese :

Les transactions liées au produit T_0 sont telles des transactions blanches ?

Suppression du Produit T_0 dans les tables

```
transactions = transactions[transactions ['id_prod'] != 'T_0']
```

```
products = products[products ['id_prod'] != 'T_0']
```

	id_prod	prix_prod	categorie_prod
730	0_219	13.9900	0
731	T_0	-1	0
732	0_407	15.9900	0
733	0_727	12.9900	0

	id_prod	prix_prod	categorie_prod
730	0_219	13.9900	0
732	0_407	15.9900	0
733	0_727	12.9900	0

IV - Coherence entre les datasets

Coherence entre les tables transactions/customers :

Tout les clients qui ont réalisé une transaction sont dans la table customer

Coherence entre les tables transactions/products :

Nombre de transaction : 336816

103 transactions avec des produits non repertoriés dans la table des produits :

Le produits non repertoriés dans la table des produits : 0_2245

Il y a 103 transactions avec le produit 0_2245 non present dans la table products

Ajout du Produit 0_2245

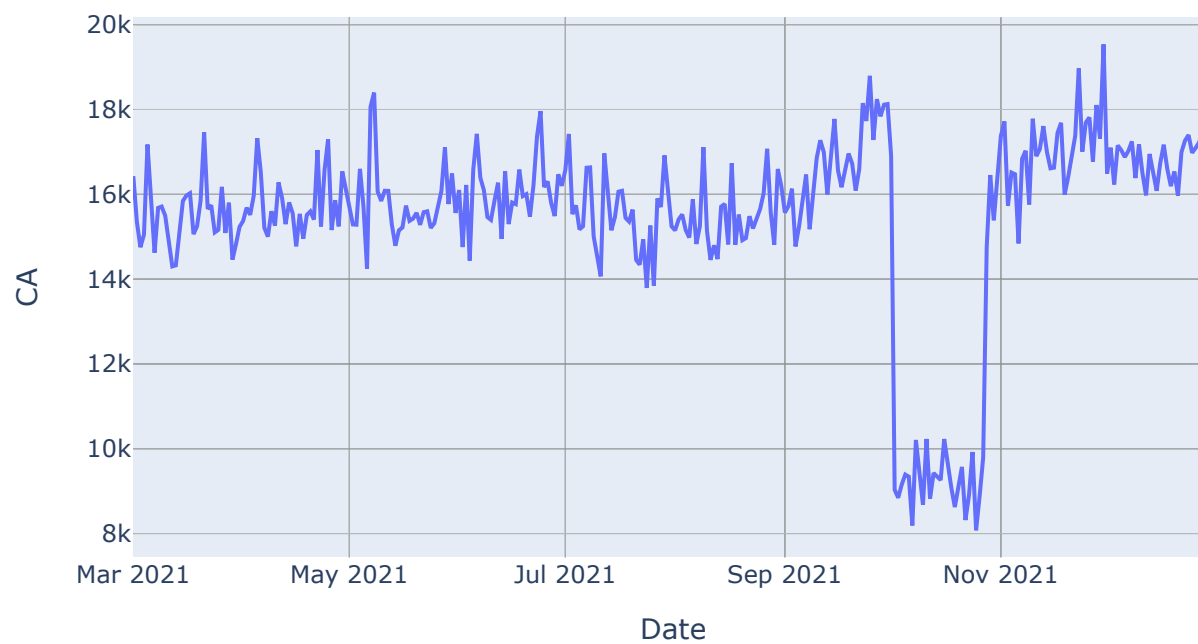
	id_prod	prix_prod	categorie_prod
3282	2_23	115.9900	2
3283	0_146	17.1400	0
3284	0_802	11.2200	0
3285	1_140	38.5600	1
3286	0_1920	25.1600	0

	id_prod	prix_prod	categorie_prod
3283	0_146	17.1400	0
3284	0_802	11.2200	0
3285	1_140	38.5600	1
3286	0_1920	25.1600	0
3287	0_2245	11.7328	0

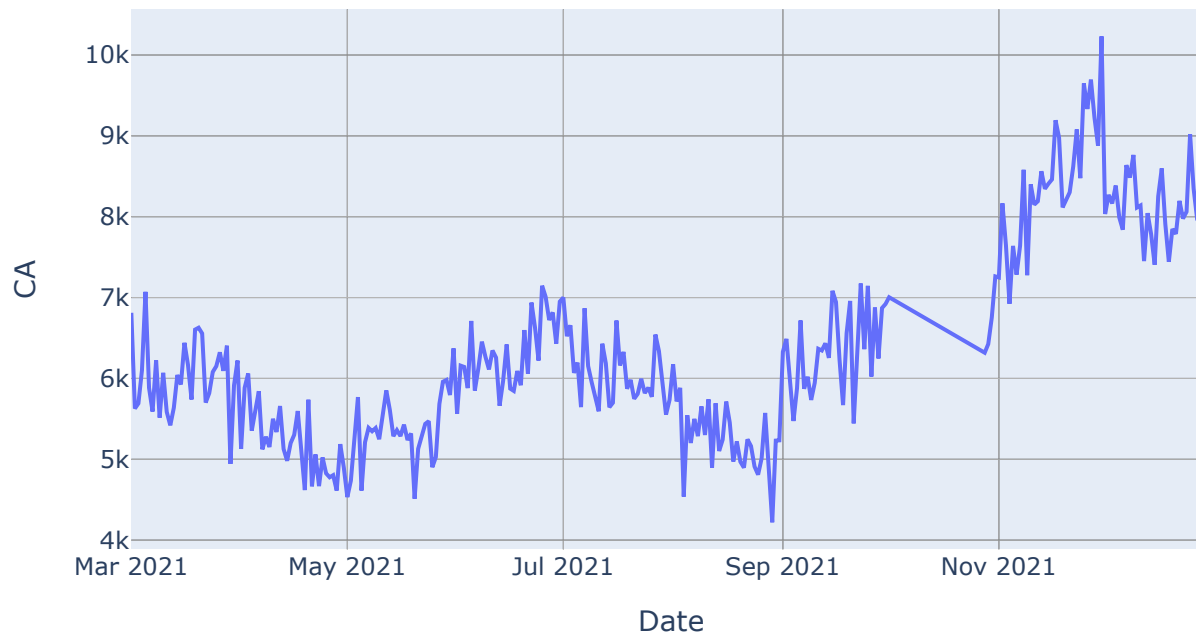
V - Creation d'un dataframe unique

	client_id	sex_client	naissance_client	age_client	session_id	id
2021-03-01	c_6584	Homme	1986	35	s_388	
2021-03-01	c_184	Homme	2004	17	s_39	
2021-03-01	c_4015	Homme	1986	35	s_218	
2021-03-01	c_7469	Homme	1956	65	s_16	
2021-03-01	c_2721	Homme	1977	44	s_322	
2021-03-01	c_5062	Homme	1988	33	s_144	
2021-03-01	c_2682	Homme	1976	45	s_132	
2021-03-01	c_6403	Homme	1980	41	s_299	
2021-03-01	c_330	Homme	1953	68	s_279	
2021-03-01	c_1769	Homme	1996	25	s_125	
2021-03-01	c_5558	Femme	1937	84	s_107	

Evolution du CA sur l'année 2021



Vente des produits catégorie 1



Il manque des informations

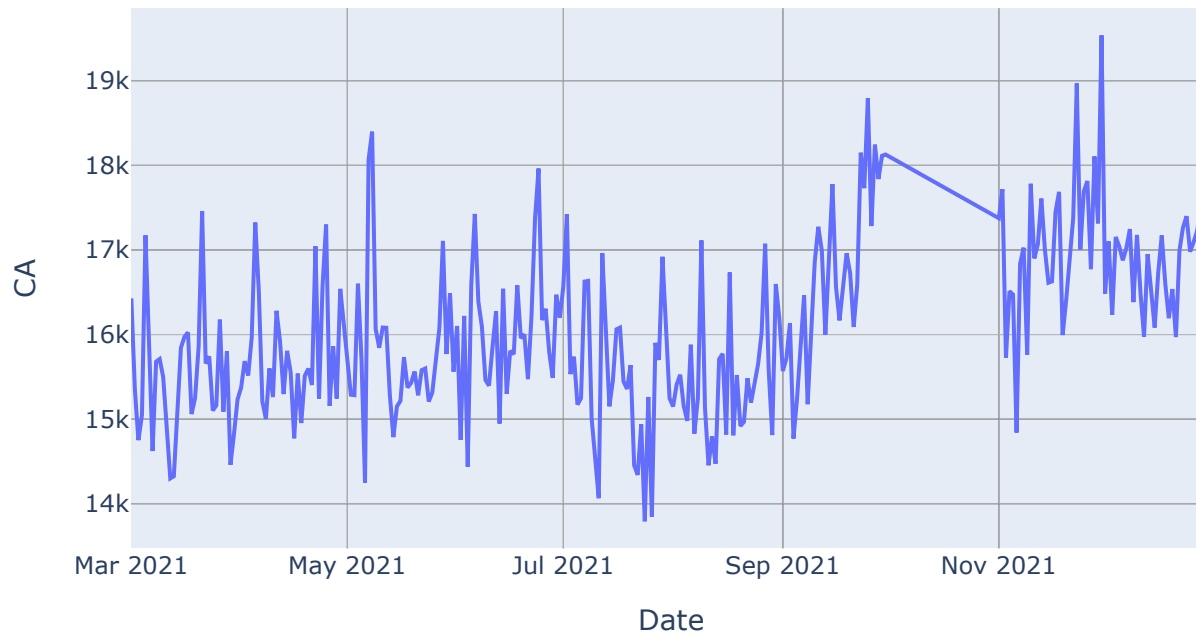
Les valeurs manquantes sont du 2 au 28 Octobre. Le mois d Octobre est supprimé.

	prix_prod
2021-10-01	7,003.7900
2021-10-28	6,317.9900
2021-10-29	6,425.1800
2021-10-30	6,753.6900
2021-10-31	7,261.6700

2021-11-01

7,248.7500

Evolution des ventes avec le mois d Octobre supprimé



Mission 2 : Analyse des données

Des indicateurs de tendance centrale et de dispersion ;

Pour la categorie 0 :

Il y a 190771.00 transactions répertoriées pour cette catégorie.

Les indicateurs de tendance centrale :

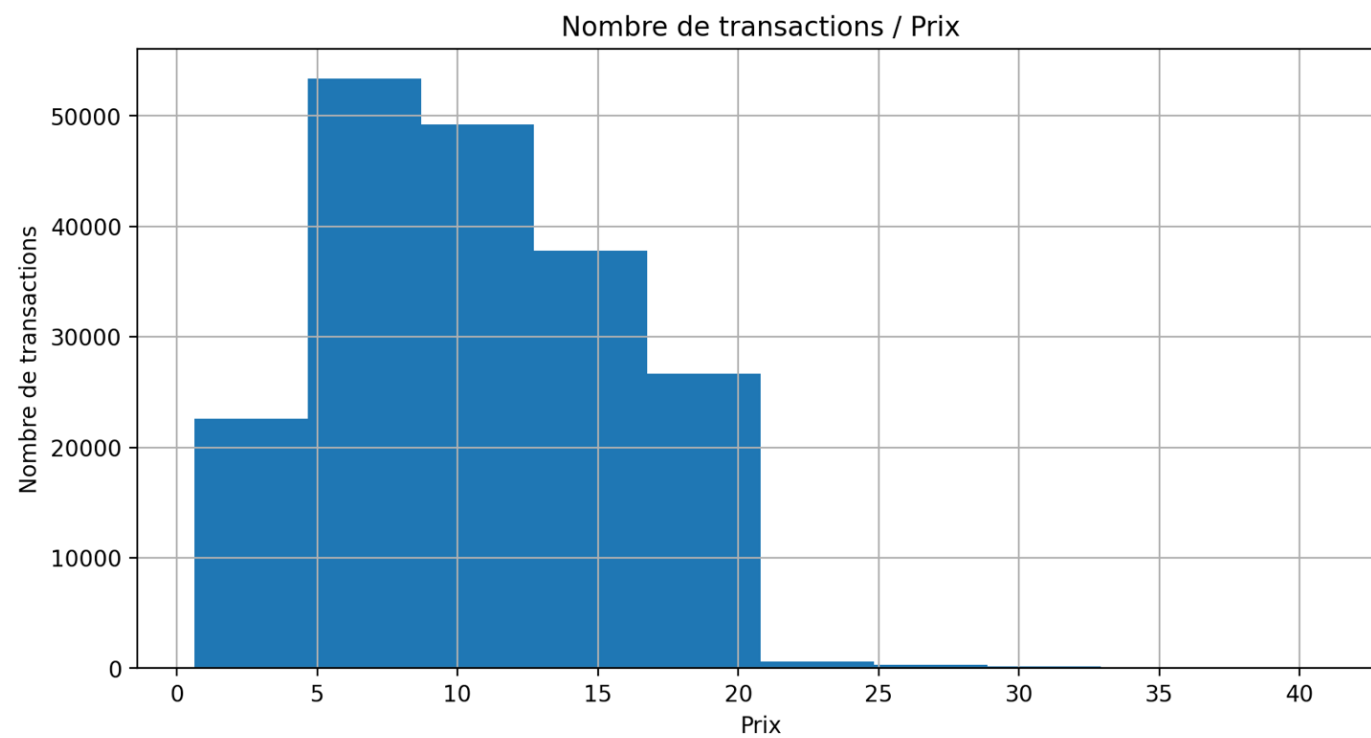
La moyenne 10.65

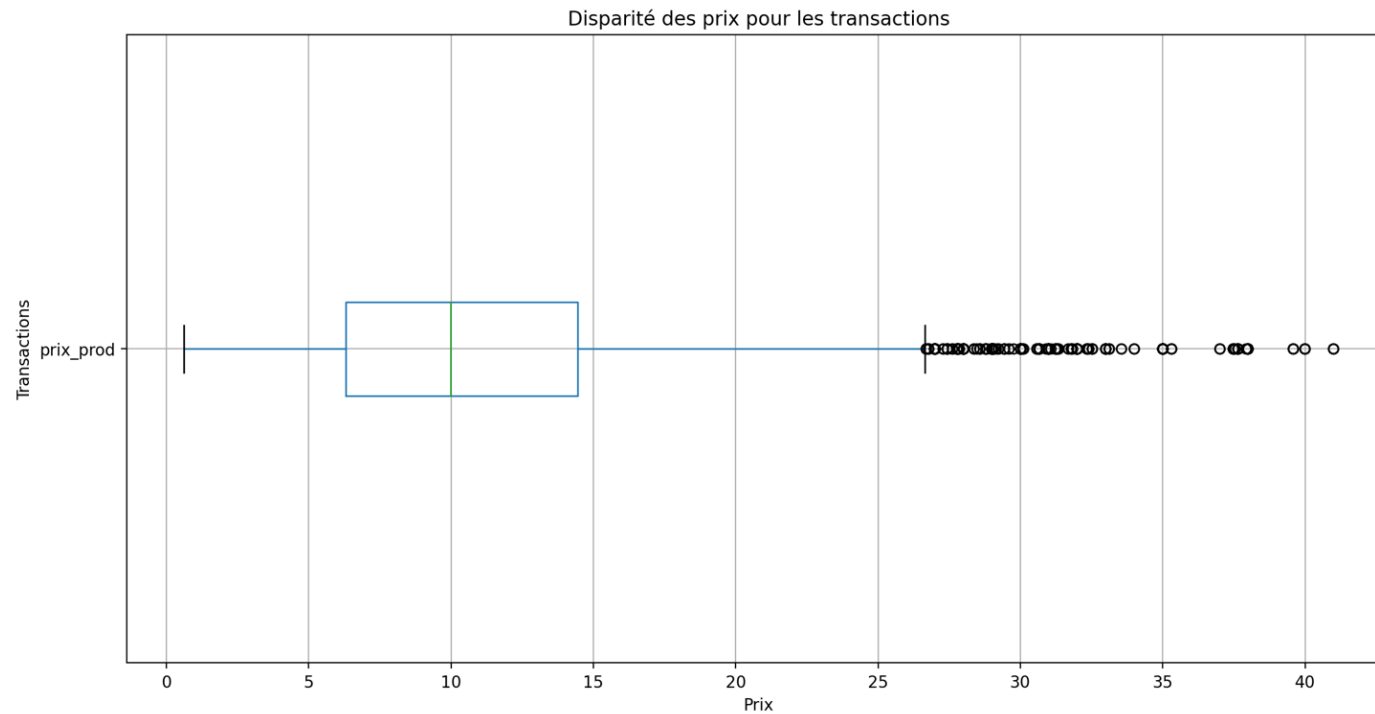
La mediane 9.99

La modalité 4.99

Les indicateurs de dispersions :

L ecart type 4.93





Pour la categorie 1 :

Il y a 108069.00 transactions répertoriées pour cette catégorie.

Les indicateurs de tendance centrale :

La moyenne 20.48

La mediane 19.08

La modalité 15.99

Les indicateurs de dispersions :

L ecart type 7.58

Made with [Streamlit](#)