



| Projet ida |

Chez nous, le frais gagne

Equipe DATA

Aline Amat, Hadel Benna,
Clara Gonçalves, Mohammed Ouahabi



L'avenir de la grande distribution se joue sur le frais...

Ida est un outil basé sur l'intelligence artificielle. La mission de ses créateurs est de lutter contre le gaspillage alimentaire de la distribution.

Problématique DATA

Optimiser les produits disponibles en rayon des magasins (minimiser la démarque).



Préparation des données

I) Nettoyage des deux datasets

1) `Ida_1` : 72 049 enregistrements et 10 colonnes concernant les ventes des 4 magasins parisiens en 2022. Absence de NaN.

- Nettoyage de la table
- Uniformisation de la colonne 'colisage' par nombre de vente
- Analyse et correction des ventes en doublons (prix de vente et quantité vendue différentes)
- Analyse et correction de la colonne 'promo'
- Compléments et réorganisation des colonnes

Au final : 69 052 enregistrements et 9 colonnes.



Préparation des données

Ida_1

AVANT :

	store	libelle	CA	date	weekday	week	PV	PCB	PROMO	sold_qty
0	NATURALIA PARIS GAMBETTA	CELERI BRANCHE COUPE PIECE	2.99	2022-08-04	JEUDI	2022-08-01	2.99	6.0		1.000000
1	NATURALIA GRENELLE	PECHE BLANCHE CAL A/B	11.00	2022-08-04	JEUDI	2022-08-01	6.99	4.0		1.573677
2	NATURALIA PARIS GAMBETTA	PRUNE REINE CLAUDE	6.91	2022-08-05	VENDREDI	2022-08-01	8.99	6.0		0.768632
3	NATURALIA ST OUEN 2	ORANGE FILET 1KG	4.58	2022-08-06	SAMEDI	2022-08-01	2.29	8.0		2.000000
4	NATURALIA GRENELLE	GROSEILLE ROUGE BQT 125G	4.99	2022-08-07	DIMANCHE	2022-08-01	4.99	5.0		1.000000
...
72044	NATURALIA GRENELLE	POMME GRISE CANADA PLATEAU 170	4.57	2022-05-05	JEUDI	2022-05-02	3.29	4.0		1.389058
72045	NATURALIA GRENELLE	POMME GRISE CANADA PLATEAU 170	4.92	2022-05-13	VENDREDI	2022-05-09	3.29	4.0		1.495441
72046	NATURALIA GRENELLE	POMME GRISE CANADA PLATEAU 170	10.51	2022-05-16	LUNDI	2022-05-16	3.29	4.0		3.194529
72047	NATURALIA PARIS GAMBETTA	POMME GRISE CANADA PLATEAU 170	5.63	2022-06-16	JEUDI	2022-06-13	3.29	4.0		1.711246
72048	NATURALIA PARIS GAMBETTA	POMME GRISE CANADA PLATEAU 170	4.81	2022-07-10	DIMANCHE	2022-07-04	3.29	4.0		1.462006

72049 rows x 10 columns

APRÈS :

	nom_magasin	nom_produit	date_vente	semaine_iso	prix_vente	PROMO	qt_vendue	montant_vente	colisage
0	NATURALIA PARIS GAMBETTA	KIWI VERT PIECE CAL 30/33	2022-02-23	8	0.59	Promo	21.423729	12.64	10.0
1	NATURALIA MONTORGUEIL	RAISIN NOIR MUSCAT 5KG	2022-09-22	38	5.49	Promo	4.952641	27.19	5.0
2	NATURALIA ST OUEN 2	ABRICOT CAL A/AA	2022-07-04	27	5.49	Promo	6.892532	37.84	5.0
3	NATURALIA GRENELLE	TOMATE POP NOIRE DE CRIMEE	2022-08-07	31	5.49	Promo	0.883424	4.85	5.0
4	NATURALIA PARIS GAMBETTA	ABRICOT CAL A/AA	2022-07-09	27	5.49	Promo	8.036430	44.12	5.0
...
69047	NATURALIA GRENELLE	OIGNON CEBETTE BOTTE	2022-04-14	15	2.99	Hors Promo	1.000000	2.99	6.0
69048	NATURALIA ST OUEN 2	OIGNON CEBETTE BOTTE	2022-04-19	16	2.99	Hors Promo	1.000000	2.99	6.0
69049	NATURALIA GRENELLE	OIGNON CEBETTE BOTTE	2022-05-17	20	2.99	Hors Promo	2.000000	5.98	6.0
69050	NATURALIA PARIS GAMBETTA	OIGNON CEBETTE BOTTE	2022-05-24	21	2.99	Hors Promo	1.000000	2.99	6.0
69051	NATURALIA GRENELLE	ASPERGE BLANCHE VRAC 22+	2022-03-15	11	24.99	Hors Promo	0.224090	5.60	2.0

69052 rows x 9 columns



Préparation des données

Création fichier `df_produit_promo.csv`, contenant les produits portant, au moins sur une vente, la mention ‘Promo’ : utile pour les analyses sur l’impact des promotions dans les ventes (CA, quantité vendue).

	nom_magasin	nom_produit	montant_vente	date_vente	prix_vente	PROMO	qt_vendue	colisage
0	NATURALIA PARIS GAMBETTA	KIWI VERT PIECE CAL 30/33	12.64	2022-02-23	0.59	Promo	21.423729	10.0
1	NATURALIA MONTORGUEIL	RAISIN NOIR MUSCAT 5KG	27.19	2022-09-22	5.49	Promo	4.952641	5.0
2	NATURALIA ST OUEN 2	ABRICOT CAL A/AA	37.84	2022-07-04	5.49	Promo	6.892532	5.0
3	NATURALIA GRENELLE	TOMATE POP NOIRE DE CRIMEE	4.85	2022-08-07	5.49	Promo	0.883424	5.0
4	NATURALIA PARIS GAMBETTA	ABRICOT CAL A/AA	44.12	2022-07-09	5.49	Promo	8.036430	5.0
...
25046	NATURALIA GRENELLE	COURGE MIX PANACHE	4.95	2022-04-11	2.99	Hors Promo	1.655518	8.0
25047	NATURALIA ST OUEN 2	POMME VERTE GRANNY PLATEAU 170	5.32	2022-09-28	2.99	Hors Promo	1.779264	4.0
25048	NATURALIA MONTORGUEIL	POMME VERTE GRANNY PLATEAU 170	2.20	2022-09-19	2.99	Hors Promo	0.735786	4.0
25049	NATURALIA GRENELLE	POMME VERTE GRANNY PLATEAU 170	6.91	2022-01-04	2.99	Hors Promo	2.311037	4.0
25050	NATURALIA MONTORGUEIL	POMME VERTE GRANNY PLATEAU 170	4.37	2022-09-24	2.99	Hors Promo	1.461538	4.0

25051 rows x 8 columns



Préparation des données

Création de catalogues

- `df_ida1_catalogue.csv` : utile pour remplir une partie des NaN de la table `ida_2`.

```
df_ida1_catalog = df_ida1[['date_vente', 'nom_produit', 'semaine_iso', 'prix_vente', 'colisage']].drop_duplicates()
df_ida1_catalog['année'] = df_ida1_catalog['date_vente'].dt.year
df_ida1_catalog.drop(columns=['date_vente'], inplace=True)
df_ida1_catalog = df_ida1_catalog[['année', 'nom_produit', 'semaine_iso', 'prix_vente', 'colisage']].drop_duplicates()
df_ida1_catalog.sort_values(by=(['nom_produit', 'semaine_iso']), inplace=True)
df_ida1_catalog.reset_index(drop=True, inplace=True)
df_ida1_catalog
```

	année	nom_produit	semaine_iso	prix_vente	colisage
0	2022	ABRICOT CAL A/AA	22	5.99	5.0
1	2022	ABRICOT CAL A/AA	23	5.99	5.0
2	2022	ABRICOT CAL A/AA	24	5.99	5.0
3	2022	ABRICOT CAL A/AA	25	5.99	5.0
4	2022	ABRICOT CAL A/AA	26	5.49	5.0
...
5061	2022	TRESSE MIX CONDIMENTS	17	4.99	8.0
5062	2022	TRESSE MIX CONDIMENTS	19	4.99	8.0
5063	2022	TRESSE MIX CONDIMENTS	20	4.99	8.0
5064	2022	TRESSE MIX CONDIMENTS	22	4.99	8.0
5065	2022	TRESSE MIX CONDIMENTS	23	4.99	8.0

5066 rows × 5 columns



Préparation des données

1) Nettoyage des deux datasets

2) `ida_2` : 38 571 enregistrements et 15 colonnes concernant les ventes de 221 magasins en Province en 2022 et janvier-avril 2023. Présence de 8 182 NaN.

- Nettoyage initial de la table ;
- Gestion des NaN par le remplissage des valeurs nulles :
 - par les valeurs les plus proches dans le temps de la table `ida_2` (colonnes 'prix_vente', 'colisage' et 'prix_achat') ;
 - par les valeurs présentes dans la table `ida_1` (colonnes 'prix_vente' et 'colisage') ;
 - calcul de la colonne 'qt_vendue' ;
- Compléments et réorganisation des colonnes.

Au final : 38 373 enregistrements et 10 colonnes.
2 956 NaN pour 3 colonnes et 6 340 pour 1 colonne.



Préparation des données

Ida_2

AVANT :

	id	store_id	sale_id	sale_date	sale_name	sale_name_ida	company_name	file_id	weekofyear	year	sale_amount	quantity_actual	pv	colisage	pa
0	ee4bc0906774591a63640...	7708b78251ebab505f31a...	1647f504c105312f75a6d...	2022-12-08	POMME BICO R.REINETTE...	2414381	naturalia	dff1dc0aa9a196a6e2d75...	49	2022	1.80	0.602	2.99	14.0	1.60
1	a864d26ab025ec18a52f0...	1c500c2c996db6cfa3a40...	cf0baa03ca0d294d4b08a...	2022-11-17	#POMME LOCALE BICO JO...	2404275	naturalia	dff1dc0aa9a196a6e2d75...	46	2022	1.91	0.639	2.99	14.0	1.60
2	3b06d294b96b9d3261bdd...	44ec10ea0ff2720c38128...	c8d2a4754a252f4bf3abb...	2022-08-10	POMME BICO IDARED 130+	2404085	naturalia	dff1dc0aa9a196a6e2d75...	32	2022	10.42	2.749	3.79	14.0	1.85
3	8cf8b91107f51f268539c...	a2f89d2f9034212c2b825...	a093c63a8c24b7c6d897d...	2023-03-03	POIVRON ROUGE 3KG	8292625	naturalia	dff1dc0aa9a196a6e2d75...	9	2023	2.15	0.308	6.99	3.0	3.71
4	104d24f763fb139054391...	790094e2c3383d76e2188...	a149ba099030e53f16f59...	2022-07-22	MYRTILLE BQT 125G FRANCE	2453728	naturalia	dff1dc0aa9a196a6e2d75...	29	2022	3.99	1.000	3.99	8.0	2.60
...
38566	0998d4a5a7eeb2a081641...	e8a83965ab1d87fcc48f2...	93a595ec6b5d1a961cca4...	2023-03-13	SALADE BQT MIX JEUNES...	2426179	naturalia	dff1dc0aa9a196a6e2d75...	11	2023	2.99	1.000	2.99	6.0	1.26
38567	bbbd3f711b3d5448f4c65...	33d4c2963df2e358a1d24...	75cd4b66d6a1f5c3fe392...	2023-03-11	MANDARINE FILET 5/6	2427052	naturalia	dff1dc0aa9a196a6e2d75...	10	2023	5.38	2.000	2.69	8.0	1.51
38568	03f32013de31879c3f542...	81dda62af11433dd70918...	f1f9d21659e691a875482...	2022-10-20	CLEMENTINE CAL 1-4	2409209	naturalia	dff1dc0aa9a196a6e2d75...	42	2022	8.49	1.417	5.99	6.0	2.65
38569	aa25d67d2bddb07096781...	8fabba03a75acbd7c2d3...	b359db5b49b6570628544...	2022-10-05	RAISIN NOIR MUSCAT 5KG	9967415	naturalia	dff1dc0aa9a196a6e2d75...	40	2022	14.70	2.630	5.59	5.0	3.40
38570	3a59f4c9fe8ec11ba107b...	06713f8f35969f712bf47...	997ab5ab4da85fe685ef3...	2023-03-29	MIX TRIO FRUITS ROUGE...	7275696	naturalia	dff1dc0aa9a196a6e2d75...	13	2023	24.66	1.898	12.99	1.0	7.80

38571 rows x 15 columns

APRÈS :

	nom_magasin	nom_produit	date_vente	année	semaine_iso	prix_vente	qt_vendue	montant_vente	colisage	prix_achat
0	Magasin_1	POMME BICO R.REINETTE 135+	2022-12-08	2022	49	2.99	0.602007	1.80	14.0	1.60
1	Magasin_2	POMME LOCALE BICO JONAGOLD	2022-11-17	2022	46	2.99	0.638796	1.91	14.0	1.60
2	Magasin_3	POMME BICO IDARED 130+	2022-08-10	2022	32	3.79	2.749340	10.42	14.0	1.85
3	Magasin_4	POIVRON ROUGE 3KG	2023-03-03	2023	9	6.99	0.307582	2.15	3.0	3.71
4	Magasin_5	MYRTILLE BQT 125G FRANCE	2022-07-22	2022	29	3.99	1.000000	3.99	8.0	2.60
...
38368	Magasin_71	SALADE BQT MIX JEUNES POUSSES	2023-03-13	2023	11	2.99	1.000000	2.99	6.0	1.26
38369	Magasin_13	MANDARINE FILET 5/6	2023-03-11	2023	10	2.69	2.000000	5.38	8.0	1.51
38370	Magasin_126	CLEMENTINE CAL 1-4	2022-10-20	2022	42	5.99	1.417362	8.49	6.0	2.65
38371	Magasin_124	RAISIN NOIR MUSCAT 5KG	2022-10-05	2022	40	5.59	2.629696	14.70	5.0	3.40
38372	Magasin_111	MIX TRIO FRUITS ROUGES PLATEAU	2023-03-29	2023	13	12.99	1.898383	24.66	1.0	7.80

38373 rows x 10 columns



Préparation des données

Ida_2

AVANT :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38571 entries, 0 to 38570
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id               38571 non-null  object
1   store_id        38571 non-null  object
2   sale_id         38571 non-null  object
3   sale_date       38571 non-null  object
4   sale_name       38571 non-null  object
5   sale_name_ida   38571 non-null  int64
6   company_name    38571 non-null  object
7   file_id         38571 non-null  object
8   weekofyear      38571 non-null  int64
9   year            38571 non-null  int64
10  sale_amount     38571 non-null  float64
11  quantity_actual 30389 non-null  float64
12  pv              30389 non-null  float64
13  colisage        30389 non-null  float64
14  pa              30389 non-null  float64
dtypes: float64(5), int64(3), object(7)
memory usage: 4.4+ MB
None
id               0
store_id         0
sale_id          0
sale_date        0
sale_name        0
sale_name_ida    0
company_name     0
file_id          0
weekofyear       0
year             0
sale_amount      0
quantity_actual  8182
pv               8182
colisage         8182
pa               8182
dtype: int64
```

APRÈS :

```
<class 'pandas.core.frame.DataFrame'>
Index: 38373 entries, 0 to 38372
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   nom_magasin     38373 non-null  object
1   nom_produit     38373 non-null  object
2   date_vente      38373 non-null  datetime64[ns]
3   année           38373 non-null  int64
4   semaine_iso     38373 non-null  UInt32
5   prix_vente      35417 non-null  float64
6   qt_vendue       35417 non-null  float64
7   montant_vente   38373 non-null  float64
8   colisage        35417 non-null  float64
9   prix_achat      32033 non-null  float64
dtypes: UInt32(1), datetime64[ns](1), float64(5), int64(1), object(2)
memory usage: 3.1+ MB
None
nom_magasin      0
nom_produit      0
date_vente       0
année            0
semaine_iso     0
prix_vente       2956
qt_vendue        2956
montant_vente    0
colisage         2956
prix_achat       6340
dtype: int64
```




Préparation des données

Création de catalogues

- produit/catégorie/famille/sous_famille.csv : création en vue de la réalisation des analyses par catégorie et famille de produits.

	A	B	C	D
1	nom_produit_clean	catégorie	famille	sous-famille
2	ABRICOT CAL A/2A DEMETER	Fruit	Fruit à noyau	Abricot
3	ABRICOT CAL A/AA	Fruit	Fruit à noyau	Abricot
4	ABRICOT LOCAL	Fruit	Fruit à noyau	Abricot
5	ABRICOT ORANGERED	Fruit	Fruit à noyau	Abricot
6	AIL	Légume	Condiment	Ail
7	AIL 3 TETES FILET	Légume	Condiment	Ail
8	AIL BLANC VRAC 10KG	Légume	Condiment	Ail
9	AIL DES OURS BOTTE	Légume	Condiment	Ail
10	AIL FRAIS	Légume	Condiment	Ail
11	AIL FRAIS PIECE J+	Légume	Condiment	Ail
12	AIL FRIT POT 70G	Légume	Condiment	Ail
13	AIL NOIR POT 60G	Légume	Condiment	Ail
14	ALOE VERA	Fruit	Fruit exotique	Aloe Vera
15	ALOE VERA PIECE	Fruit	Fruit exotique	Aloe Vera
16	AMANDE COQUE	Fruit	Fruit sec ou à coque	Amande
17	AMANDE FRAICHE	Fruit	Fruit sec ou à coque	Amande
18	AMANDES TAMARI	Fruit	Fruit sec ou à coque	Amande
19	ANANAS B9 PIECE	Fruit	Fruit exotique	Ananas
20	ANANAS EXTRA SWEET PIECE	Fruit	Fruit exotique	Ananas
21	ANANAS PAIN DE SUCRE PIECE	Fruit	Fruit exotique	Ananas
22	ANONE	Fruit	Fruit exotique	Anone
23	AROMATE BOTTE CIBOULETTE	Aromate	Aromate	Ciboulette
24	AROMATE BOTTE CORIANDRE	Aromate	Aromate	Coriandre
25	AROMATE BOTTE PERSIL FRISE	Aromate	Aromate	Persil



Analyses des données

- Vue d'ensemble
- Performance produit
- Impact Promo
- Colisage
- Fréquence et Taux de Présence

Analyses réalisées mais non pertinentes :

- Analyse par magasin
- Analyse par Produits communs aux deux datasets



Quelques conclusions

- Performance produit :
 - Produits du Top 10
 - Produits Flop 10 ;
- Impact Promo : les promotions boostent les ventes ;
- Colisage : adapter les colisages aux ventes ;
- Fréquence et Taux de Présence :
 - Produits à forte fréquence de vente et à fort volume vendu ;
 - Porter une attention aux taux de présence hebdomadaire des produits dans chaque saison



Merci !
