

# Projet Big Data

## "La Fromagerie Fêtes du Brie"

C. Alfred, F. Berrabah, D. Cormerais, C. Mougani

Diginamic  
<https://www.diginamic.fr>

February 15, 2024



## La Fromagerie *Fêtes du Brie*



« Nos clients les plus fidèles sont comme nos fromages dans une fondue : leur présence fait fondre nos cœurs et donne à chaque instant une saveur inoubliable. »

- La fromagerie Fêtes du Brie a partagé avec nous les chiffres de son programme de fidélisation dans le cadre d'un projet nommé Big Data
- Ils nous ont demandé des statistiques et analyses pour optimiser leurs opérations de fidélisation.
- leur Entrepôt de données (Datawarehouse) existe depuis 2004.

## Problématique

### ► Fidélisation Client dans un Marché Compétitif

- **Objectif:** Ajuster le programmes de fidélisation via l'extraction de données.
- **Lot 1:** (2006-2010, Dépts 53,61 et 28)- Analyse comportements préférences.
- **Lot 2:** (2011-2016, Dépts 22, 49 et 53)- Tendances régionales, meilleures commandes.
- **Lot 3:** Base NoSQL HBase pour stockage des données.
- **Lot 4:** Power BI et visualisations interactives
- **Conclusion:** Fidélisation client: pierre angulaire de la croissance

## Données Qualifiées

- Fiabilité
- Pertinence
- Complétude
- Cohérence
- Conformité réglementaire



## Exemple

- codcli
- cpcli
- date\_str
- Nbcolis
- libobj
- Poidsobj



## Processus d'importation des données

- Lecture et traitement des lignes CSV.
- Extraction des colonnes spécifiques et conversion.
- Filtrage des données.
- Génération des paires clé-valeur.
- Tri et transmission aux reducers.

## Structuration des Données

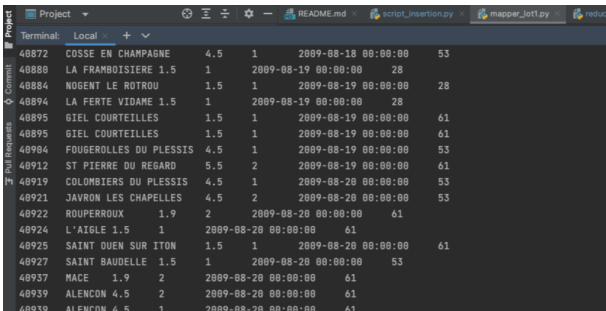
- Résumé de l'analyse des données fournies.
- Description du schéma de données conçu.
- Explication des techniques de normalisation utilisées pour réduire les redondances et les incohérences dans les données.



## Objectif

- Filtrer les données de 2006 à 2010, départements 53, 61 et 28.
- Extraire les 100 meilleures commandes avec ville, quantités d'articles, et valeur de "timbrecde".
- Exporter le résultat dans un fichier Excel.

les données entre 2006 et 2010, avec les départements 53, 61 et 28

A screenshot of a terminal window with a dark background. The window title bar shows 'Project' and several open files: 'README.md', 'script\_insertion.py', 'mapper\_lot1.py', and 'reduc...'. The terminal output is a table of data records. The first column contains IDs, the second contains location names, and the remaining columns contain numerical values and timestamps. The records are sorted by location name.

40872	COSSE EN CHAMPAGNE	4.5	1	2009-08-18 00:00:00	53
40880	LA FRAMBOISIERE 1.5	1		2009-08-19 00:00:00	28
40884	NOGENT LE ROTROU	1.5	1	2009-08-19 00:00:00	28
40894	LA FERTE VIDAME 1.5	1		2009-08-19 00:00:00	28
40895	GIEL COURTEILLES	1.5	1	2009-08-19 00:00:00	61
40895	GIEL COURTEILLES	1.5	1	2009-08-19 00:00:00	61
40904	FOUGEROLLES DU PLESSIS	4.5	1	2009-08-19 00:00:00	53
40912	ST PIERRE DU REGARD	5.5	2	2009-08-19 00:00:00	61
40919	COLOMBIERS DU PLESSIS	4.5	1	2009-08-20 00:00:00	53
40921	JAVRON LES CHAPELLES	4.5	2	2009-08-20 00:00:00	53
40922	ROUPERROUX	1.9	2	2009-08-20 00:00:00	61
40924	L'AIGLE 1.5	1		2009-08-20 00:00:00	61
40925	SAINT OVEN SUR ITON	1.5	1	2009-08-20 00:00:00	61
40927	SAINT BAUDELLÉ	1.5	1	2009-08-20 00:00:00	53
40937	MACE	1.9	2	2009-08-20 00:00:00	61
40939	ALENCON	4.5	2	2009-08-20 00:00:00	61
40939	ALENCON	4.5	1	2009-08-20 00:00:00	61

```
cat ./Input/dataw_fro04.csv |python3 ./Lot_1/mapper_lot1.py
```

Tableau des 100 meilleures commandes avec la ville, la somme des quantités des articles et la valeur de "timbrecde"

```
Project + I O + MyEnv: jupyter | myenv: jupyter | myenv: jupyter
In Project_Big_Data - Documents\Project_Big_Data
In Jupyter
> In [1]:
41 # Import de l'ensemble standard
42 # Lecture de l'ensemble standard
43 Par ligne du sys.stdin:
44     process_line(line, commandes)
45
46 # Preload le dictionnaire en fonction de la clé "cne" et "timeperiod" de maillon de maillon
47 commandes = {}
48 commandes.items()
49 commandes["cne"] = {}
50 commandes["timeperiod"] = {}
51
52 # Créez un Dataframe à partir des données
53 mydata = []
54
55 Terminal - Jupyter
~/Documents/Project_Big_Data/.jupyter/myenv: jupyter: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 2.0),
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at: https://github.com/pandas-dev/pandas/issues/54446

import pandas as pd
000000 ville ute timeperiod
0 25062 AMNOU 42 9.5
1 42997 SAINT HOMERES DES BASSILLES 22 6.4
2 25061 AMNOU 30 7.95
3 24701 TOURNAI SUR OIVE 25 5.1
4 26038 BOUCE 26 5.8
5 ... ..
95 26476 HORNBAI 9 6.4
96 26468 BILTHOVEN 9 6.4
97 35996 LE GRONIST SAINT EMIL 9 6.4
98 27433 AINIS VAL DE ROYANG 9 6.4
99 29976 ASTELLE 9 6.4

[100 rows x 4 columns]
monquicrist19@DESKTOP-Project_Big_Data %
```

```
cat ./Input/dataw_fro04.csv | python3 ./Lot_1/mapper_lot1.py | python3
./Lot_1/reducer_lot1.py
```

## Exporter le résultat dans un fichier Excel

The screenshot shows a Microsoft Excel spreadsheet with the following data:

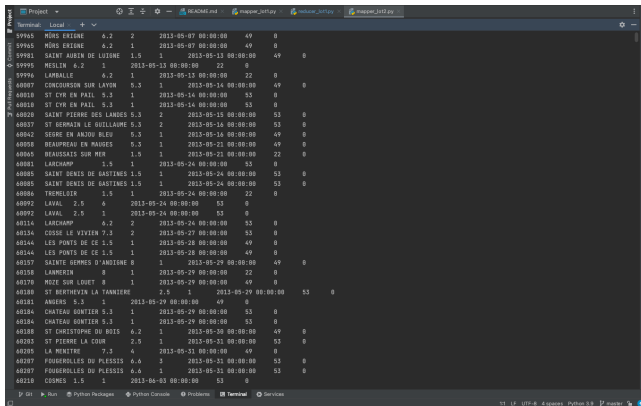
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	code	ville	qte	limbende																		
2	25882	ARRONJ	47 9.85																			
3	42997	SANT GEC	23 6.4																			
4	25861	ARRONJ	30 7.05																			
5	24701	TOURNAI	25 5.1																			
6	26620	BOUCE	20 5.8																			
7	16216	ARRONJ	18 6.4																			
8	22809	COMMUN	18 6.4																			
9	36805	LE MERLE	18 5.8																			
10	20775	SANT DEB	17 5.8																			
11	45087	DOMPRO	16 7.8																			
12	45676	LA GOMER	16 7																			
13	47308	LA GOMER	16 6.5																			
14	16234	ARRONJ	16 6.4																			
15	16236	ARRONJ	16 6																			
16	22346	VIMOUTRE	16 5.1																			
17	46341	CEROY BE	15 7.8																			
18	15346	SANT DE	14 6.4																			
19	42331	ATHYS VAL	14 7.7																			
20	41271	GOULET	14 5.8																			
21	15369	SA	13 7.15																			
22	44514	RI	13 7.1																			
23	46735	LA FERTE	13 6.5																			
24	17151	SANT MA	13 5.8																			
25	46074	CEROY BE	12 7.1																			
26	46209	LIVROIS	12 7.1																			
27	46448	LA LANDE	12 6.5																			
28	26478	MAVENNE	12 6.4																			
29	37017	SANT PRI	12 6.4																			

Figure: Légende de la première image

## Objectif

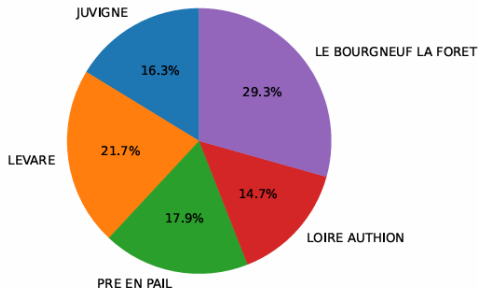
- Filtrer les données de 2011 à 2016, départements 22, 49 et 53.
- Extraire aléatoirement 5% des 100 meilleures commandes sans "timbrecli" avec la moyenne des quantités.
- Créer un PDF avec un graphe (PIE) par ville.
- Exporter le résultat dans un fichier Excel.

## Les données entre 2011 et 2016, avec les départements 22, 49 et 53



ID	Name	Age	Gender	Date	Value 1	Value 2
59905	MORS EIGONE	6.2	2	2013-05-07 00:00:00	49	0
59905	MORS EIGONE	6.2	1	2013-05-07 00:00:00	49	0
59901	SAINT AUBIN DE LUIGNE	1.5	1	2013-05-13 00:00:00	49	0
59995	MESLIN	6.2	1	2013-05-13 00:00:00	22	0
59996	LAMBALLE	6.2	1	2013-05-13 00:00:00	22	0
60007	CONCOURSUR SUR LAYON	5.3	1	2013-05-14 00:00:00	49	0
60010	ST CYR EN PAIL	5.3	1	2013-05-14 00:00:00	53	0
60010	ST CYR EN PAIL	5.3	1	2013-05-14 00:00:00	53	0
60020	SAINT PIERRE DES LANDES	5.3	2	2013-05-15 00:00:00	53	0
60037	ST GERMAIN LE GUILLAUME	5.3	2	2013-05-16 00:00:00	53	0
60042	SEGRE EN ANJOU BLEU	5.3	1	2013-05-16 00:00:00	49	0
60050	BEAUPREAU EN MAUGES	5.3	1	2013-05-21 00:00:00	49	0
60065	BEAUSAIS SUR MER	1.5	1	2013-05-21 00:00:00	22	0
60081	LANCHAMP	1.5	1	2013-05-24 00:00:00	53	0
60085	SAINT GENES DE GASTINES	1.5	1	2013-05-24 00:00:00	53	0
60085	SAINT GENES DE GASTINES	1.5	1	2013-05-24 00:00:00	53	0
60086	TRENELLOIR	1.5	1	2013-05-24 00:00:00	22	0
60092	LAVAL	2.5	6	2013-05-24 00:00:00	53	0
60092	LAVAL	2.5	1	2013-05-24 00:00:00	53	0
60114	LANCHAMP	6.2	2	2013-05-24 00:00:00	53	0
60134	COSSE LE VIVIER	7.3	2	2013-05-27 00:00:00	53	0
60144	LES PONTS DE CE	1.5	1	2013-05-29 00:00:00	49	0
60144	LES PONTS DE CE	1.5	1	2013-05-29 00:00:00	49	0
60157	SAINT GERMES D'ANGIERE	8	1	2013-05-29 00:00:00	49	0
60158	LANNERIN	8	1	2013-05-29 00:00:00	22	0
60170	MOZE SUR LOUET	8	1	2013-05-29 00:00:00	49	0
60180	ST BERTHEVIN LA TANNIERE	2.5	1	2013-05-29 00:00:00	53	0
60181	ANGERS	5.3	1	2013-05-29 00:00:00	49	0
60184	CHATEAU GONTIER	5.3	1	2013-05-29 00:00:00	53	0
60184	CHATEAU GONTIER	5.3	1	2013-05-29 00:00:00	53	0
60188	ST CHRISTOPHE DU BOIS	6.2	1	2013-05-30 00:00:00	49	0
60202	ST PIERRE LA COUR	2.5	1	2013-05-31 00:00:00	53	0
60205	LA MENETRE	7.3	4	2013-05-31 00:00:00	49	0
60207	FOUDEROLLES DU PLESSIS	6.6	3	2013-05-31 00:00:00	53	0
60207	FOUDEROLLES DU PLESSIS	6.6	1	2013-05-31 00:00:00	53	0
60216	COSNES	1.5	1	2013-06-03 00:00:00	53	0

Repartition des villes par rapport a la quantite Avec Hadoop



Extraire de façon aléatoire 5% des 100 meilleures commandes avec la ville, la somme des quantités des articles sans "timbrecli".

## Exporter le résultat dans un fichier Excel

results\_lot2\_3

Rechercher dans Excel

Fichier Accueil Insertion Partager Mise en page Formules Données Révision Affichage Aide Dessin

Commentaires Rattrapage Edition Partager

Calibri 11

Général

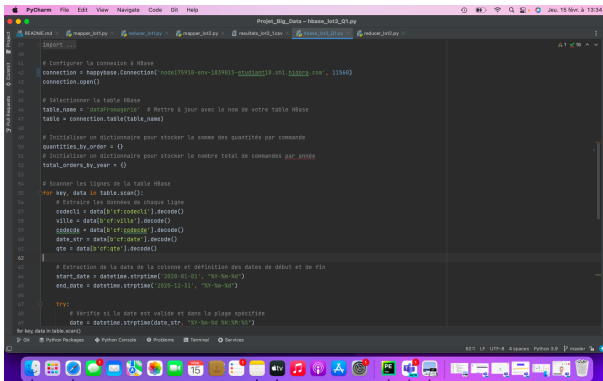
A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Code Commande	Ville	Nb. Articles sans Timbre	Moyennes Quantités de la Commande										
2	50468	PRE EN PAIL ST SAMSON	9	3										
3	51717	SAINT PIERRE DES NIDS	17	5,666666667										
4	71410	LE BURET	15	5										
5	69537	LAVAL	11	2,75										
6	71037	THUBOEUF	9	2,25										
7														
8														
9														
10														
11														
12														



## Objectif

- Mettre en place une base NoSQL HBASE.
- Réaliser des requêtes avec des scripts Python sur le contenu du fichier CSV dans HBASE.
- Création d'un programme Python (avec Pandas) pour générer des graphes en PDF ainsi que des tableaux Excel et CSV à partir de vos données importées dans HBase



```

17 import sys
18
19 # Configurer la connexion à HBase
20 connection = hbasepy.Connection(host='192.168.1.101', port=2181)
21 connection.open()
22
23 # Sélectionner la table HBase
24 table_name = 'datafromagerie' # Mettre à jour avec le nom de votre table HBase
25 table = connection.table(table_name)
26
27 # Initialiser un dictionnaire pour stocker la somme des quantités par commande
28 quantities_by_order = {}
29 # Initialiser un dictionnaire pour stocker le nombre total de commandes par année
30 total_orders_by_year = {}
31
32 # Scanner les lignes de la table HBase
33 for key, data in table.scan():
34     # Extraire les données de chaque ligne
35     codecli = data[b'cf:codecli'].decode()
36     ville = data[b'cf:ville'].decode()
37     sdate = data[b'cf:sdate'].decode()
38     date_str = data[b'cf:date'].decode()
39     qte = data[b'cf:qte'].decode()
40
41     # Extraction de la date de la colonne et définition des dates de début et de fin
42     start_date = datetime.strptime('2020-01-01', '%Y-%m-%d')
43     end_date = datetime.strptime('2020-12-31', '%Y-%m-%d')
44
45     try:
46         # Vérifier si la date est valide et dans la plage spécifiée
47         date = datetime.strptime(date_str, '%Y-%m-%d %H:%M:%S')
48     except:
49         continue
50
51     # Mise à jour des dictionnaires
52     quantities_by_order[codecli] = quantities_by_order.get(codecli, 0) + qte
53     total_orders_by_year[date.year] = total_orders_by_year.get(date.year, 0) + 1
54
55 # Affichage des résultats
56 print("Quantités par commande :")
57 for key, value in quantities_by_order.items():
58     print(key, value)
59
60 print("Nombre total de commandes par année :")
61 for year, count in total_orders_by_year.items():
62     print(year, count)
63
64 # Fermer la connexion
65 connection.close()
66
67 if __name__ == '__main__':
68     main()
69 
```

## Presentations des résultats

results\_lot3\_1

Rechercher dans Excel

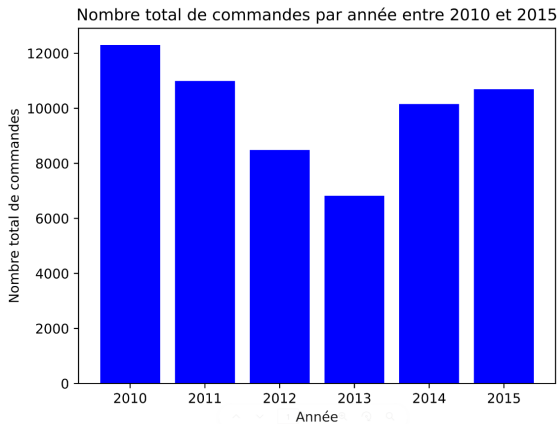
Fichier Accueil Insertion Partager Mise en page Formules Données Révision Affichage Aide Dessin

Commentaires Ratrapage Affichage Fastap

12 G

Commande, Somme des quantités

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Commande, Somme des quantités															
2	85380,6															
3	85643,1															
4	85684,1															
5	85783,4															
6	86079,1															
7	86096,1															
8	86420,1															
9	86935,1															
10	86944,3															
11	86986,1															
12	87166,2															
13	87214,5															
14	87458,5															
15	87475,2															
16	87657,1															
17																
18																



RECHERCHER UNED LATER

Fichier Accueil Insertion Partager Mise en page Formules Données Révision Affichage Aide Dessin Commentaires Rattrapage

Calibri (Corps) 11 G

11%

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	nomcli	prenomcli	nombre_commandes	somme_quantites	max_timbrecede										
2	LEMARDELE	Edith	9	9	17,9										
3															
4															
5															
6															
7															
8															



## La Fromagerie Fêtes du Brie

01/01/2004 26/07/2021

Nos clients les plus fidèles sont comme nos fromages dans une fondue : leur présence fait fondre nos cœurs et donne à chaque instant une saveur inoubliable.

ID Comma ndes	Ville	Departement	Identité	Timbre Commandes	Timbre Client	Date de Commandes	Nb Colis	Poids(g)	Nb Articles	Points
57069	BEAUFORT EN VALLEE	Maine-et-Loire	ROGER Sabrina	9,00 €	0,00 €	13 August 2012	1	16	300	0
94806	GACE	Orne	MICHEL Jocelyne	10,10 €	0,00 €	22 November 2019	1	282	58	-430
96871	ST CONTEST	Calvados	HIE Lyliane	15,40 €	0,00 €	27 July 2020	1	2316	51	-5
88599	JUVIGNY LES VALLEES	Manche	PIGEON Claude	13,20 €	0,00 €	22 January 2021	1	2191	51	440
81470	SOURDEVAL	Manche	TREOL Pascale	11,80 €	0,00 €	29 November 2018	2	355	50	-370
99819	MALICORNE SUR SARTHE	Sarthe	MAUGET Laurence	11,45 €	0,00 €	17 June 2021	1	1094	47	-230
78809	SOURDEVAL	Manche	TREOL Pascale	12,40 €	0,00 €	06 February 2018	2	1260	45	-30
87848	ECHALOU	Orne	RAGEOT Michel	17,90 €	0,00 €	26 October 2020	2	660	45	-340
88130	MALICORNE SUR SARTHE	Sarthe	MAUGET Laurence	12,35 €	0,00 €	07 December 2020	1	2642	44	270
<b>Total</b>							<b>7</b>	<b>20339976</b>	<b>199052</b>	<b>1701070</b>

Total Commandes

Max Colis

70,8...

7

Variétés Articles

Max Articles

77

300

Articles De CommandesLe Plus Demandés

Tete de me...

### Choix de Filtre:

Articles

Départements

Villes

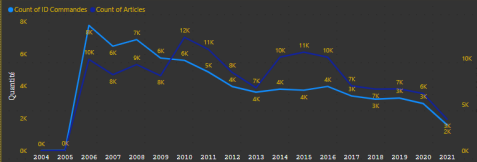
All

All

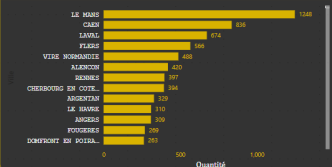
All

Effacer tous les filtres

### Evolution Des Commandes par rapport aux Articles



### Commandes Par Villes



Cette partie sera présentée sur PowerBi

## Recommandation

- Communication Proactive
- Allocation des Ressources
- Contrôle Qualité
- Respect des Délais
- Suivi Post-Implémentation



## Conclusion

- Résumer les principaux points du projet.
- Mettre en avant les réalisations et les résultats obtenus.
- Présenter les enseignements tirés du projet.

