

# Projet Big Data

## "La Fromagerie Fêtes du Brie"

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

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

February 19, 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. »

- Présentation du programme de fidélisation: 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.
- Présentation du Projet Big Data (effectuer des statistiques et analyses en lots)
- Le groupe est composé de: C. Alfred, F. Berrabah, D. Cormerais, C. Mougani

## 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

```
hadoop jar hadoop-streaming-2.7.2.jar -file mapper_lot1.py -mapper  
"python3 mapper_lot1.py" -file reducer_lot1.py -reducer "python3  
reducer_lot1.py" -input input/dataaw_fro03.csv -output  
output/output_lot1_exo1
```

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

```
hadoop jar hadoop-streaming-2.7.2.jar -file mapper_lot1.py -mapper "python3  
mapper_lot1.py" -file reducer_lot1.py -reducer "python3 reducer_lot1.py" -input  
input/dataaw_fro03.csv -output output/output_lot1_exo1
```

```
root@hadoop-master:~/data/lot1# hadoop fs -cat mapout/output_lot1_snel/part-000000
villu NEW 1089980
0 25842 APPROU 42 9.80
1 42947 SAINT GEORGES DES BRASILLERES 25 0.1
2 25841 APPROU 38 7.85
3 26781 TOURNAI SUR OYVE 25 0.1
4 26628 SOUCE 28 0.8
5 26235 APPROU 18 6.4
6 22485 COMRAE NOUVELLE D APPROU 18 6.4
7 26665 LE MEULENAULT 18 0.8
8 28075 SAINT DENMAN DU COMEIS 17 0.8
9 45887 DORPONT EN POINTE 16 7.8
10 42676 LA DONNIEURE 16 7
11 47388 LA DONNIEURE 16 6.5
12 24234 APPROU 16 6.4
13 26236 APPROU 16 6
14 22144 VIMOUTIERES 16 5.1
15 46141 CERISY BELLE ETIOLE 15 7.8
16 13548 SEINT DORVILLE 14 8.5
17 42333 ATHIS VAL DE NOUYRE 14 7.7
18 43373 SOULET 14 0.8
19 35349 SAI 13 7.15
20 44834 RE 13 7.1
21 48735 LA FERTE MACS 13 6.5
22 37351 SAINT MAURICE DU DESERT 13 5.8
23 44874 CERISY BELLE ETIOLE 12 7.1
24 49259 LESIGOTS 12 7.1
25 45648 LA LANDE PATRY 12 6.5
26 37817 SAINT PIERRE SUR OYVE 12 6.4
27 43715 ST DENMAN DU COMEIS 12 6.4
28 26478 MAYENNE 12 6.4
29 47886 REES 12 8.9
... ..
70 26724 BLOU 18 6.4
71 28873 LAVAL 18 6.4
72 23373 LA BAZOULE 18 6.4
73 43374 COMRAE NOUVELLE D APPROU 18 6.4
74 37388 LA FERTE DU OACHE 18 6.4
75 24555 ROUPPEROUX 18 6
76 26616 NORMANDEL 18 6
77 26874 NEUPRE SUR OYVE 18 6
78 47118 LA COURBE 18 8.9
79 34387 FLURE 18 6.8
80 24784 LA FERTE DU OACHE 18 6.8
81 26878 SAINT OEN LE BRIGOUT 18 5.5
82 29756 MONTCHOUVEL 18 6.5
83 25111 BALLOTS 18 5.2
84 43214 LES CHETTES 18 6.5
85 34529 VILLIERS SOUS MONTAGNE 9 8.8
86 44222 MORENT LE ROTOU 9 7
87 48719 LOUPE SUR MAINE 9 6.8
88 47142 LEBLANC LE VIGNON 9 6.5
89 47242 TORCE VIZIERS EN CHARNIE 9 6.5
90 45849 SAINT GEORGES DES BRASILLERES 9 6.5
91 47318 PRE EN DAIL 9 6.5
92 45737 TINGHOUV 9 6.5
93 49275 ALDREON 9 6.5
94 48970 RENLARS EN PERCHE 9 6.4
95 27431 ATHIS VAL DE NOUYRE 9 6.4
96 39746 SAINT PIERRE D ENTREMONT 9 6.4
97 43188 ENTREMONT 9 6.4
98 32718 LES AUPRES 9 6.4
99 24474 MORVAL 9 6.4
```

## Exporter le résultat dans un fichier Excel

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

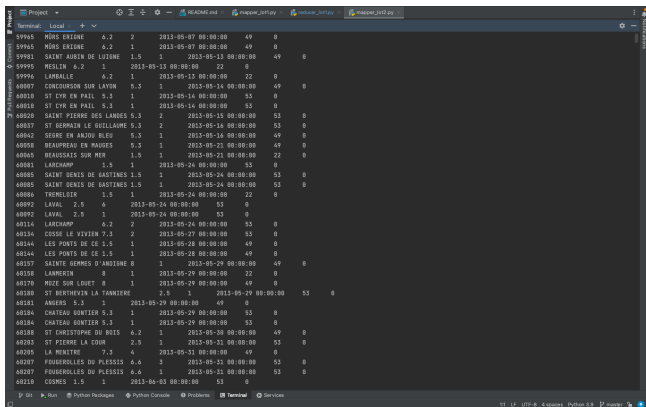
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	code	ville	qte	limbende																		
2	25882	ARRONJ	47 9.85																			
3	42997	SANT GEC	23 6.4																			
4	25861	ARRONJ	30 7.85																			
5	24701	TOURNAI	25 5.1																			
6	26620	BOUCE	20 5.8																			
7	16216	ARRONJ	18 6.4																			
8	22809	COMMUN	18 6.4																			
9	36885	LE MERLE	18 5.8																			
10	20775	SANT DEB	17 5.8																			
11	45087	DOMPRO	16 7.8																			
12	45676	LA GONNE	16 7																			
13	47308	LA GONNE	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	ATHIS 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

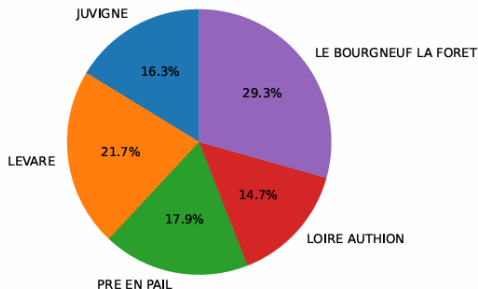
- Filtrer les données de 2011 à 2016, départements 22, 49 et 53.
- Extraire aléatoirement 5% des 100 meilleures commandes sans "timbrecli" avec la moyenne des quantités.
- Créer un PDF avec un graphe (PIE) par ville.
- Exporter le résultat dans un fichier Excel.

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



59905	MORS EAUONE	6.2	2	2013-05-07 00:00:00	49	0
59905	MORS EAUONE	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'ANGIENNE	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".



```

21 46735 LA FERTE MACÉ 13 6,5
22 17382 SAINT MAURICE DU DESERT 13 6,8
23 44874 CERISEY BELLE ÉTOILE 12 7,1
24 49584 12 7,1
25 45648 LA LANDE PATRY 12 6,5
26 27637 SAINT PIERRE SUR OISE 12 6,4
27 43715 ST GERMAIN DU CORREIS 12 6,4
28 26478 MAYENNE 12 6,4
29 47866 SEES 12 5,9
... ..
70 36724 ANNOU 18 6,4
71 38873 LAVAL 18 6,4
72 32672 LA NAYONNE 18 6,4
73 43774 COMBALE NOUVELLE D'ANNOU 18 6,4
74 37588 LA FERTE DU CACHE 18 6,4
75 24655 ROUPESROUX 18 6
76 35616 NORMANVILLE 18 6
77 36674 NEAUPHE SUR OISE 18 6
78 47118 LA COURSE 18 5,9
79 34387 FLEURE 18 5,8
80 34784 LA FERTE EN MACIE 18 5,8
81 38878 SAINT OREN LE BRISQUALT 18 5,5
82 39754 BENTHOUVEL 18 5,5
83 35111 BALLOTS 18 5,2
84 43214 LES GINETTES 18 6,5
85 34529 VILLIERS SOUS MONTAGNE 9 8,8
86 44222 MOËNT LE ROTOU 9 7
87 48719 LONGUE SUR MAISE 9 6,5
88 47242 LONLAT LE TESSON 9 6,5
89 47262 TORCE VIZIERS EN CHAMNIE 9 6,5
90 44869 SAINT GEORGES DES MOUTILLERS 9 6,5
91 47518 PRIE EN PAIL 9 6,5
92 46737 TINCHEMAY 9 6,5
93 49276 ALDONCO 9 6,5
94 48902 SCHLARD EN POUCE 9 6,4
95 27451 ATHIS VAL DE BOAISE 9 6,4
96 35286 SAINT PIERRE D'ENTREMENT 9 6,4
97 41286 ENTENNES 9 6,4
98 32758 LES AGRISES 9 6,4
99 24676 NORMAY 9 6,4

[180 rows x 4 columns]
root@hadoop-master:~/datawarehouse# hadoop fs -cat output/output_lot2_xeul/part-00000
root@hadoop-master:~/datawarehouse# hadoop fs -ls output/output_lot2_xeul/
Found 2 items
-rw-r--r-- 2 root supergroup 0 2024-02-19 14:09 output/output_lot2_xeul/_SUCCESS
-rw-r--r-- 2 root supergroup 0 2024-02-19 14:09 output/output_lot2_xeul/part-00000
root@hadoop-master:~/datawarehouse# hadoop fs -cat output/output_lot2_xeul/part-00000
Code Commande \
0 83126 LAVERGNE 18
1 17879 RESUPSE DE MACOU 18
2 43492 MONTMAY 18
3 42444 LANCHEP 12
4 83611 ST GERMAIN LE GUILLEME 18

Moyennes Quantiles de la Commande
0 2.32222
1 2.399999
2 2.488889
3 3.488887
4 2.375000
root@hadoop-master:~/datawarehouse# hadoop fs -cat output/output_lot2_xeul/part-00000@

```

## 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

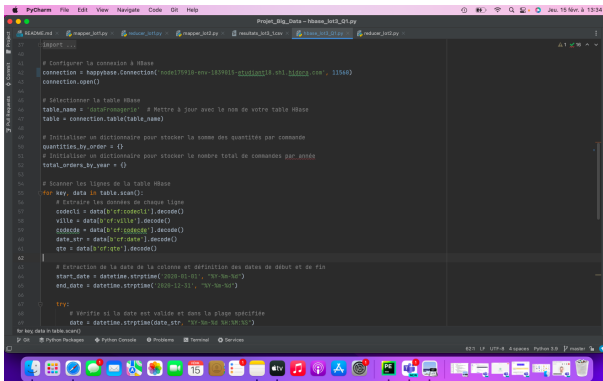
Général

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         pass
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

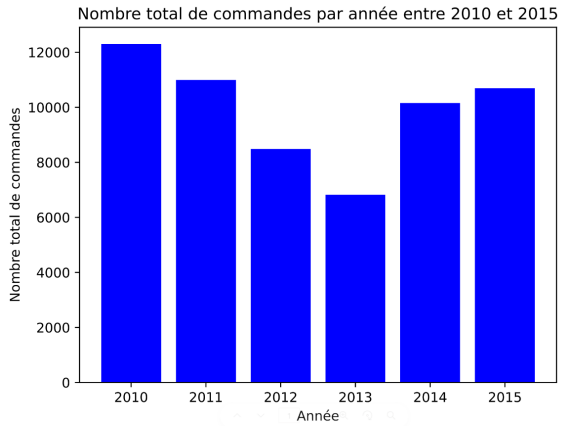
Fichier Accueil Insertion Partager Mise en page Formules Données Révision Affichage Aide Dessin

Commentaires Ratapage 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	85381,6															
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 UNID 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	11,60 €	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
87648	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
<b>70,8...</b>	<b>7</b>
Variétés Articles	Max Articles
<b>77</b>	<b>300</b>
Articles De CommandesLe Plus Demandés	
<b>Tete de me...</b>	

### Choix de Filtre:

Articles

Départements

Villes

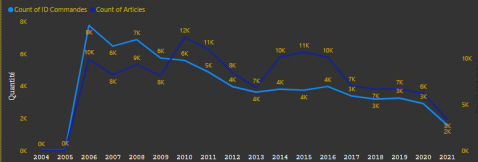
All

All

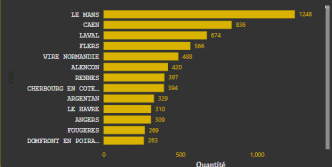
All

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.

