

SPARK

Analyse d'un jeu de données

CHABCHOUB AYMEN

EL YADINI SAID

LIN DIXUAN

YOUSFI OMAR

Introduction

2

- ▶ Nous possédons 3 jeux de données
 - ▶ Clients
 - ▶ Produits
 - ▶ Transactions
- ▶ L'objectif est de créer un reporting sur la vente des produits de l'entreprise Data's Madness, ainsi que sur leurs clients.
- ▶ Le but est de montré différente façon analysé la donnée et non de répété plusieurs fois la même requête en modifiant le mois dans la clause « where ».

Création de DataFrame

Pour créer un dataframe à partir d'un fichier de type csv, on suit les étapes suivantes :

- ▶ `spark.read.format("csv")`
 - `.option("header", "true")`
 - `.option("delimiter", ";")`
 - `.option("inferSchema", "true")`
- 1. `.csv("D:\Document D\COURSESGI\SPARK/transaction.csv")`
- 2. `.csv("D:\Document D\COURSESGI\SPARK/client.csv")`
- 3. `.csv("D:\Document D\COURSESGI\SPARK/items.csv")`

Jointure

- ▶ La création de jointure a été nécessaire afin de lier les tables des clients et des transactions ensemble ainsi que celles des clients et des transactions.
- ▶ L'ajout de colonnes représentant les jours de la semaine, les jours d'un mois, les mois et les années ont également été ajoutés afin d'avoir un rapport plus précis au niveau des chiffres.

Analyse du jeu de données


5





- Nb articles vendus :
 - ▶ 3 840 264 en 2019
 - ▶ 99 327 en 2020
 - ▶ 82 en 2021
- 10 points de ventes
- 10 000 clients réparties en 3 catégories

1 – Cas Général

Moyenne de vente € par transaction en 2019

7

Vente moyenn 
181,60 €

Shop id		Shop Name		Vente moyenne	
	1	Paris		182,17 €	
	2	Marseille		181,53 €	
	3	Lyon		181,39 €	
	4	Bordeaux		181,38 €	
	5	Nantes		181,94 €	
	6	Lille		181,57 €	
	7	Strasbourg		181,26 €	
	8	Reims		181,62 €	
	9	Toulouse		181,34 €	
	10	Grenoble		181,82 €	

Nombre moyen de transaction par jour en 2019

8

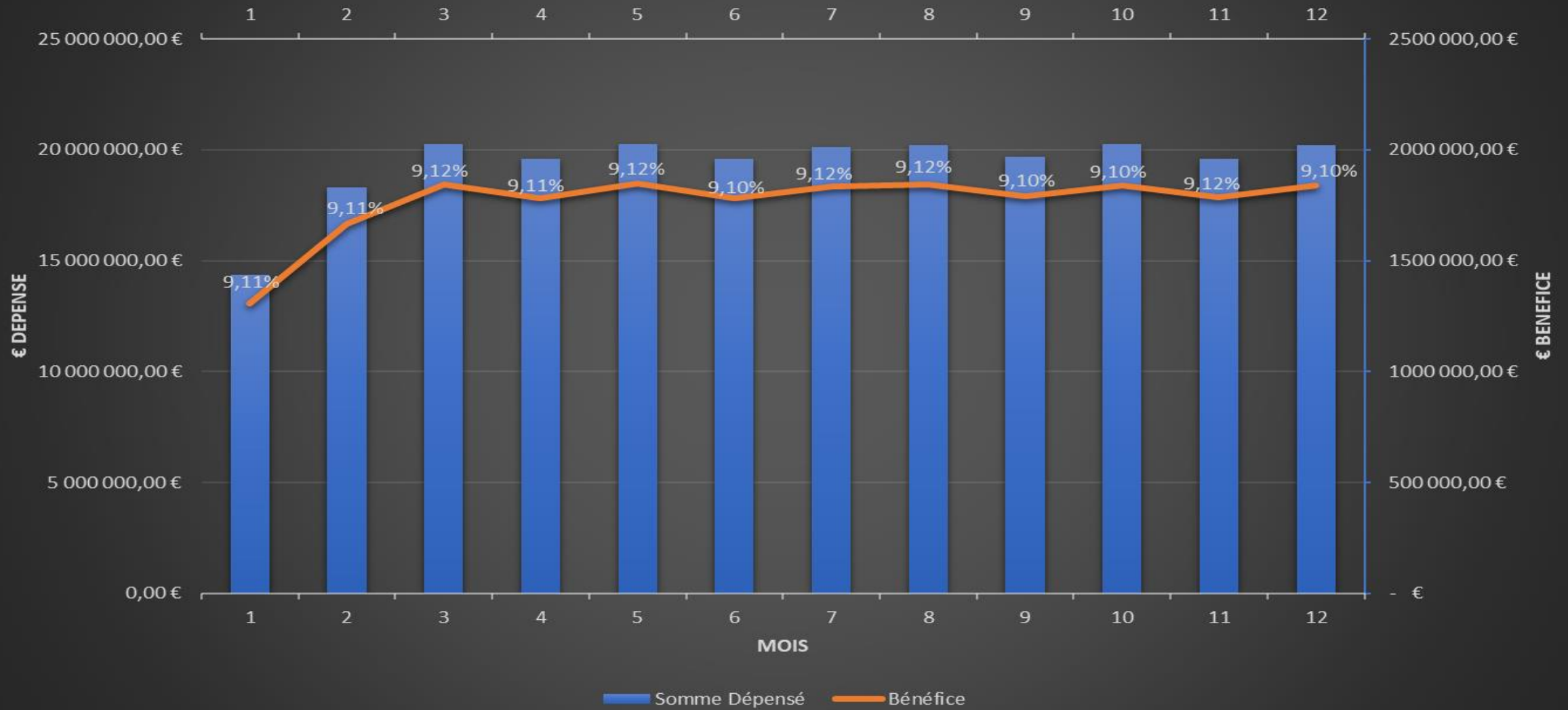
Moyenne de tra ▼
358,35

Shop Id ▼	Shop Nam ▼	Moyenne transacti ▼
1	Paris	358,02
2	Marseille	358,17
3	Lyon	358,64
4	Bordeaux	357,90
5	Nantes	357,64
6	Lille	357,63
7	Strasbourg	359,50
8	Reims	359,50
9	Toulouse	359,50
10	Grenoble	359,50

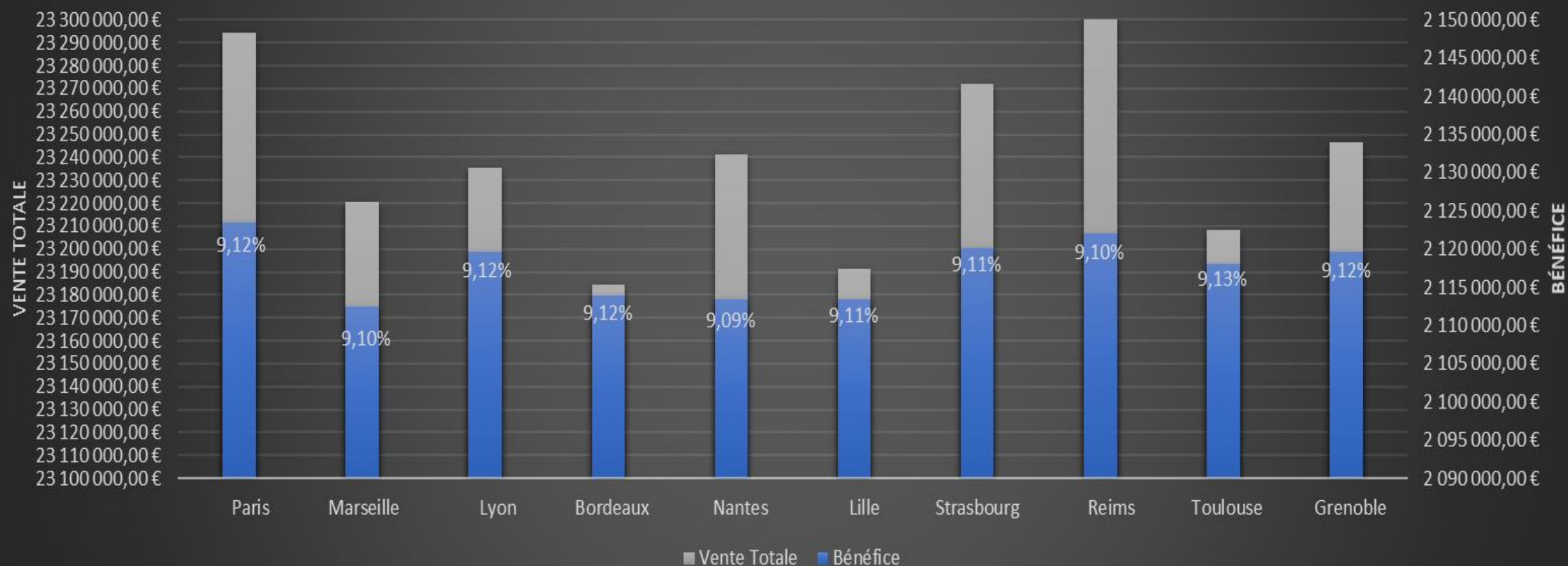
Tous les points de ventes

9

Dépense / Bénéfice 2019



Vente et Bénéfice des boutiques en 2019



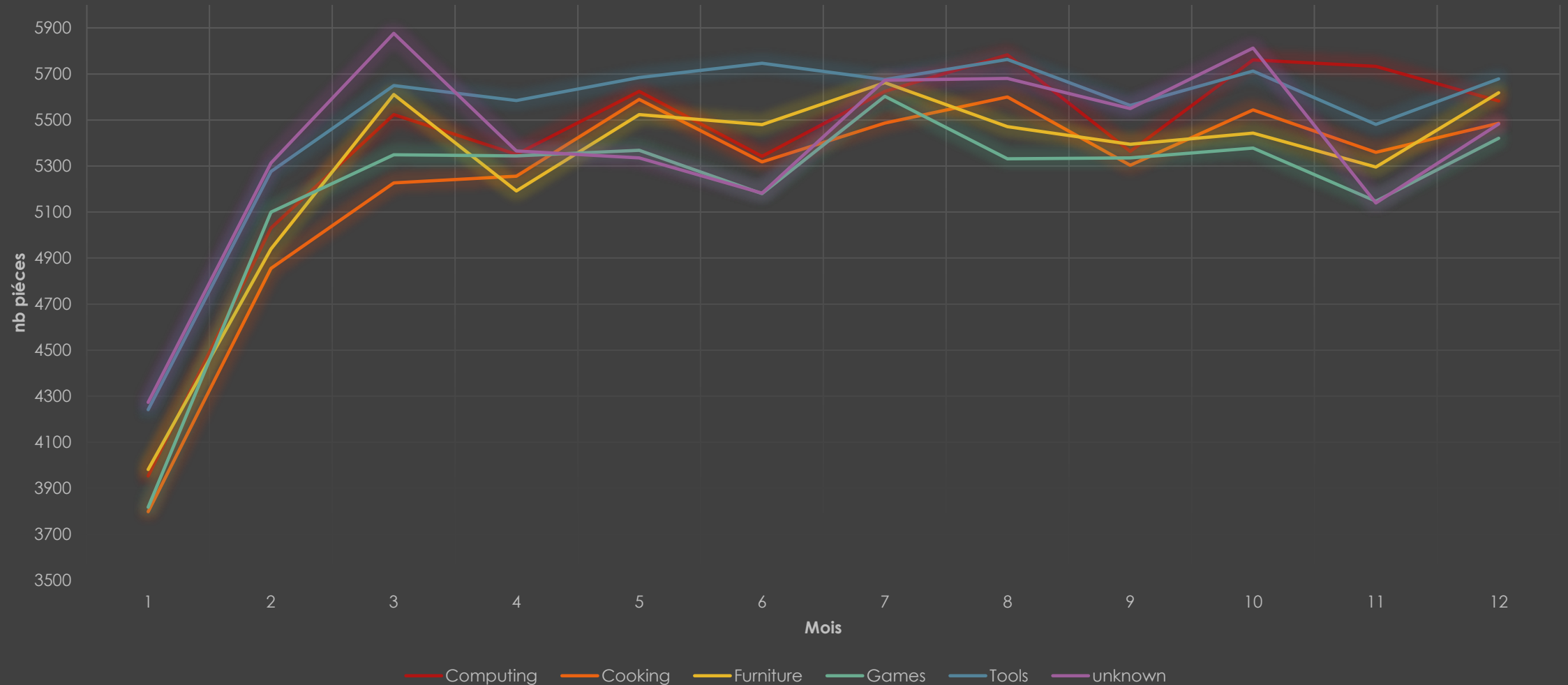
Conclusion Partie 1

11

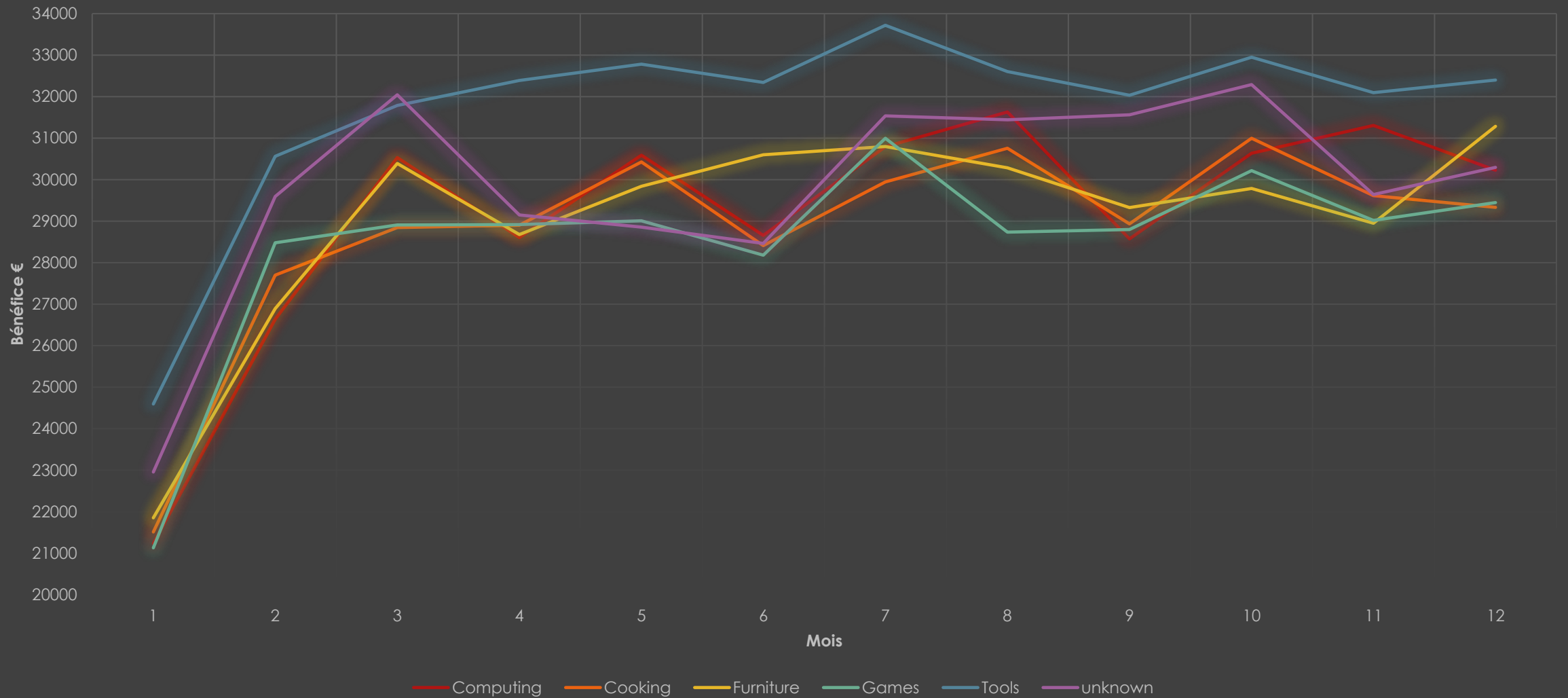
- ▶ Faible dépense des points de vente en début année (Janvier, février) puis relativement stable
- ▶ Bénéfice plus faible en début d'année
- ▶ Le CA n'est pas totalement dépendants du Capitale
- ▶ Paris ville la plus rentable

2 - Analyse du cas Paris

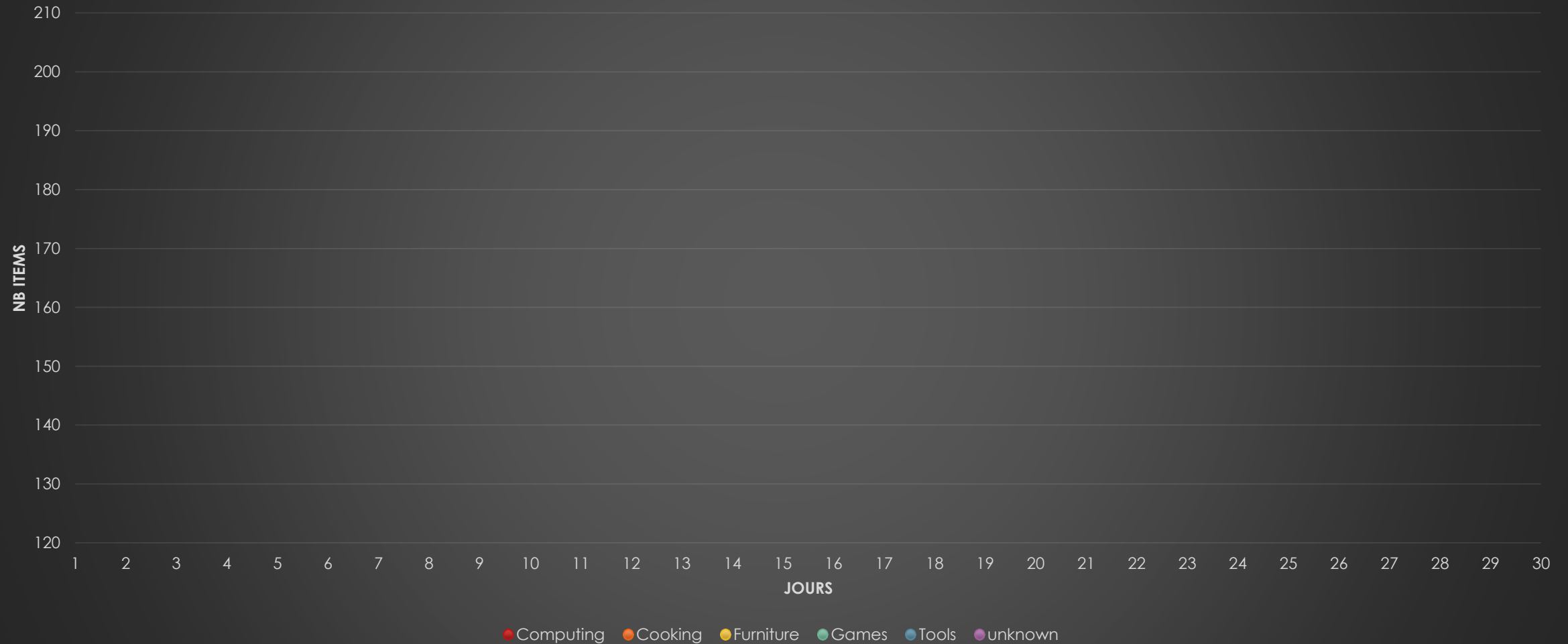
Boutique de Paris: Quantité items vendu par cartagorie par mois



Boutique de Paris: Bénéfice par catégorie d'Item vendu par mois

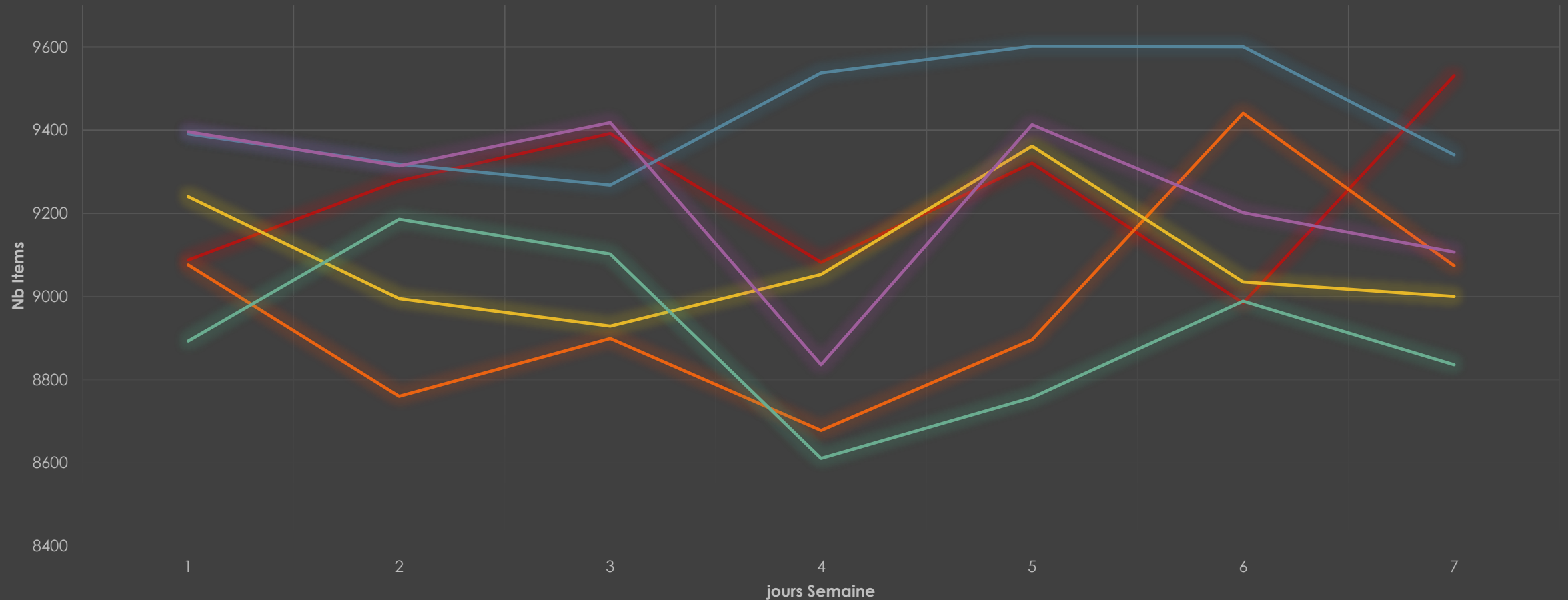


Boutique de Paris: Nombres moyens items acheté par jours en 2019



Boutique de Paris: Nombre d'Items vendu

Computing Cooking Furniture Games Tools undnown



Conclusion Partie 2

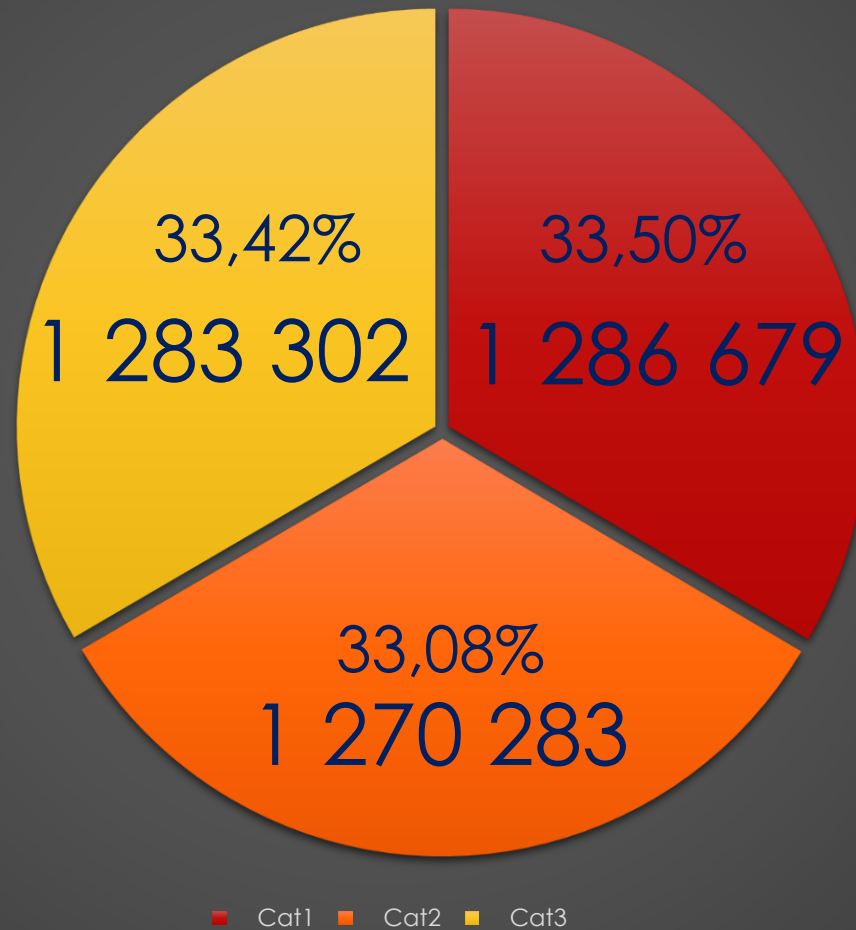
17

- ▶ L'activité augmente en générale aux alentours du 10 et diminue en fin de mois.
- ▶ La catégorie d'items et le jour de la semaine sont corrélé pour sa vente
- ▶ Tools est la catégorie d'item qui rapporte le plus argent malgré qu'elle ne soit toujours la catégorie la plus vendu

3 - Les clients

1286679

Catégorie de client en 2019 et nombres articles achetés



Nombre de transaction par catégorie item pour chaque catégorie de client en 2019

20

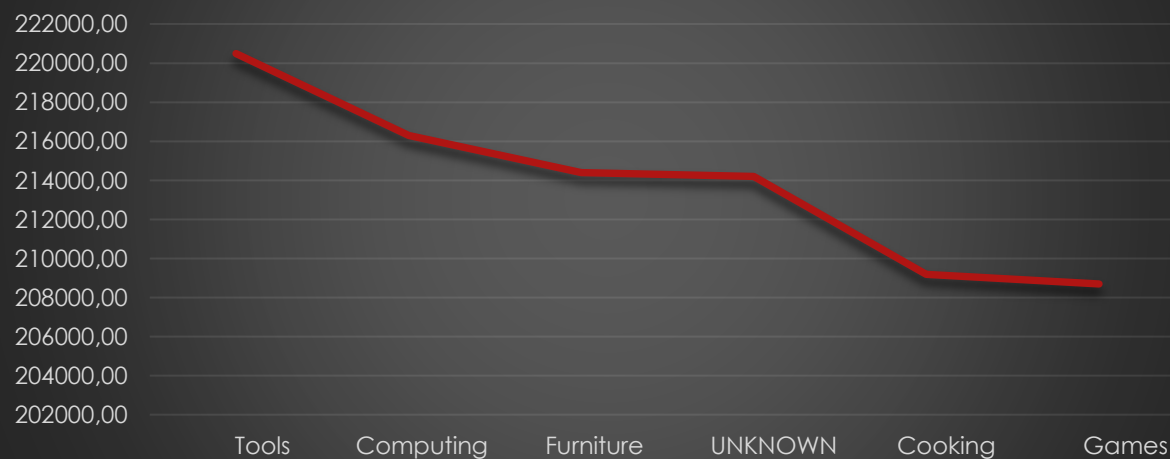
client cat 1 NB item acheter



client cat 2 NB item acheter



client cat 3 NB item acheter



La transaction les plus chères

21

id	Name	First_Name	Address	Date_Creation	Category	Item	Item_Na	Category	Item_Id	Used_Price	Selling_Price	Transaction_Id	Item_Id	Number	Year	Month	Day	Date_Transaction	Shop_Name	Shop_Id		
6684	Dos Santos	Thierry	Thierry	12 48 66 4520	09:59:46	Cat3	7204 Breton	UNKNOWN	5	99	113.85	36888	6684	7204	5	2019	6	17	22019-06-17	01:50:50	Grenoble	10
9314	Fouquet	André	André	84 43 36 8620	01:47:27	Cat3	1187 Chevallie	Furniture	5	99	113.85	37606	9314	1187	5	2019	9	30	22019-09-30	16:37:36	Marseille	2
6021	Thierry	Thibaut	Thibaut	05867343642	14:33:48	Cat2	7204 Breton	UNKNOWN	5	99	113.85	39885	6021	7204	5	2019	2	4	22019-02-04	08:02:04	Nantes	5
8245	Goncalves	Jules	Jules	84 00 91 8320	13:38:50	Cat3	6096 Germai	UNKNOWN	6	99	113.85	45447	8245	6096	5	2019	3	22	62019-03-22	03:25:41	Grenoble	10
3962	Pierre	Lucy	Lucy	17 64 27 9420	12:28:50	Cat1	7162 Loiseat	Games	1	99	113.85	50290	3962	7162	5	2019	5	22	42019-05-22	19:20:47	Reims	8
474	Gomez	Denis	Denis	34 71 07 7420	20:19:57	Cat3	6990 Rodrigue	Computing	1	99	113.85	52172	474	6990	5	2019	4	21	12019-04-21	17:50:04	Paris	1
3933	Thierry	Marie	Marie	69 03 26 1720	20:34:02	Cat2	6096 Germai	UNKNOWN	6	99	113.85	53959	3933	6096	5	2019	4	24	42019-04-24	04:55:00	Strasbourg	7
8784	Boulay	Grégoire	Grégoire	082851406320	22:19:59	Cat2	6096 Germai	UNKNOWN	6	99	113.85	58592	8784	6096	5	2019	6	8	72019-06-08	00:15:43	Bordeaux	4
44	Monnier	Lucie	Lucie	05 72 78 8320	09:51:35	Cat2	6096 Germai	UNKNOWN	6	99	113.85	59112	44	6096	5	2020	1	2	52020-01-02	19:48:12	Marseille	2
5851	Millet	Jeanne	Jeanne	05 02 66 0420	00:15:43	Cat3	7162 Loiseat	Games	1	99	113.85	63069	5851	7162	5	2019	11	26	32019-11-26	02:09:55	Lyon	3
3484	Rousseau	daisy	daisy	89 43 38 8420	05:35:03	Cat3	6990 Rodrigue	Computing	1	99	113.85	63213	3484	6990	5	2019	7	25	52019-07-25	13:00:22	Grenoble	10
1636	Etienne	Marcelle	Marc	01 70 28 3320	14:57:08	Cat3	6096 Germai	UNKNOWN	6	99	113.85	80017	1636	6096	5	2019	10	25	62019-10-25	02:16:22	Lyon	3
7178	Rolland	Hugues	Hugue	28 77 21 4620	14:29:30	Cat1	7162 Loiseat	Games	1	99	113.85	98104	7178	7162	5	2019	10	24	52019-10-24	06:32:06	Paris	1
242	Vasseur	Éric	Éric	55 75 21 5620	01:13:39	Cat3	1187 Chevallie	Furniture	5	99	113.85	120369	242	1187	5	2019	7	16	32019-07-16	01:05:57	Lille	6
9715	Blin	Bernadette	Be	00 18 83 5220	13:41:33	Cat1	7162 Loiseat	Games	1	99	113.85	129599	9715	7162	5	2019	3	21	52019-03-21	07:56:45	Lille	6
9328	Salmon	Alice	Alice	93 64 71 4720	21:07:56	Cat1	7204 Breton	UNKNOWN	5	99	113.85	132363	9328	7204	5	2019	3	9	72019-03-09	03:32:56	Paris	1
3631	Peltier	Corinne	Corin	03704486112	10:50:20	Cat3	6990 Rodrigue	Computing	1	99	113.85	143949	3631	6990	5	2019	1	23	42019-01-23	22:17:26	Lyon	3
9869	Masson	Léon	Léon	80 54 04 0120	18:49:33	Cat1	6990 Rodrigue	Computing	1	99	113.85	155383	9869	6990	5	2019	3	7	52019-03-07	05:54:47	Lyon	3
6833	Techer	Corinne	Corin	63 61 52 8720	19:01:09	Cat2	7162 Loiseat	Games	1	99	113.85	164567	6833	7162	5	2019	11	10	12019-11-10	04:11:24	Paris	1
3034	Julien	Anouk	Anouk	019333258520	22:21:55	Cat3	7204 Breton	UNKNOWN	5	99	113.85	174092	3034	7204	5	2019	8	18	12019-08-18	19:06:56	Lyon	3
1550	Buisson	Olivie	Olivie	05352198412	16:44:57	Cat1	7204 Breton	UNKNOWN	5	99	113.85	181039	1550	7204	5	2019	4	12	62019-04-12	20:02:15	Lyon	3
1303	Sauvage	Denise	Denise	79 27 55 8220	07:26:54	Cat1	6096 Germai	UNKNOWN	6	99	113.85	191726	1303	6096	5	2019	8	19	22019-08-19	07:18:39	Nantes	5
7891	Leclerc	Jeannine	Jean	22 89 28 7920	12:53:36	Cat3	7162 Loiseat	Games	1	99	113.85	192207	7891	7162	5	2019	7	6	72019-07-06	13:04:17	Reims	8

Conclusion partie 3

22

- ▶ Les 3 catégories de clients achètent approximativement la même quantité d'items.
- ▶ Les transactions les plus chères sont environs 3 fois supérieurs à la moyenne (slide 7).
- ▶ La Catégorie de client influe sur la catégorie items acheté bien que des similitudes persiste.

Conclusion

23

- ▶ Analyse sous différent angles nous permet de :
 - ▶ Prévoir les items vendu en fonction du jours de la semaine
 - ▶ Prévoir quels items proposé aux clients en face de nous
 - ▶ Prévoir les stock aux fils des jours