Data Analyst Portfolio CRM Project — SQL

Dans ce projet, nous allons explorer les principales possibilités, conseils et astuces qui permettent aux scientifiques des données de naviguer et d'analyser efficacement les grands ensembles de données à l'aide de SQL. Les données de CRM sont les données très utiles pour les dirigeants et le overview rapide pour faciliter la prise de décisions et ajuster la stratégie de l'entreprise.

L'objectif du projet

- 1. Analyser les données de ventes pour identifier les insights.
- 2. Convertir ces insights en indicateurs clés de performance (KPI).
- 3. Créer un tableau de bord intégrant les KPI définis.

La méthodologie

J'ai développé et rédigé plusieurs requêtes SQL qui répondaient à diverses demandes concernant l'ensemble de données afin de répondre aux objectifs du projet. Chaque demande était accompagnée d'un plan soigneusement réfléchi pour extraire des insights particuliers, et elle était suivie de commentaires détaillés les décrivant. J'ai utilisé les fonctions de conversion appropriées, le regroupement et le tri pour agréger et résumer les données de manière utile.

L'observation des données

Nous avons les données de ventes d'un CRM avec les tables suivantes ici:

1. accounts.csv

	T account	# revenue >	# employees V
1	Acme Corporation	1100,04	2822
2	Betasoloin	251,41	495
3	Betatech	647,18	1185
4	Bioholding	587,34	1356
5	Bioplex	326,82	1016

2. employees.csv

	# emp_id	T name	# mgr_id V	T regional_office	T status
1	1000	1 Anna Snelling	10036	Central	Current
2	1000	2 Cecily Lampkin	10036	Central	Current
3	1000	3 Versie Hillebrand	10036	Central	Current
4	1000	4 Lajuana Vencill	10036	Central	Current
5	1000	5 Moses Frase	10036	Central	Current

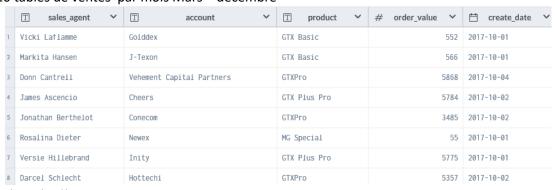
products.csv

	□ product ✓	# sales_price V
1	GTX Basic	550
2	GTX Pro	4821
3	MG Special	55
4	MG Advanced	3393
5	GTX Plus Pro	5482
6	GTX Plus Basic	1096

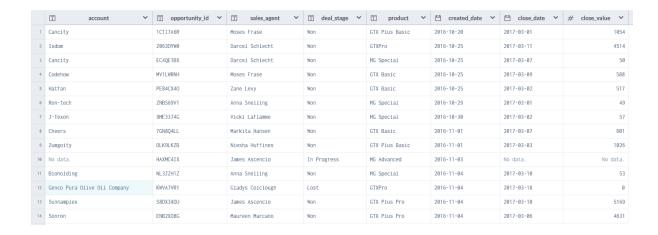
4. sales-teams.csv

			□ regional_office ✓	
1	Anna Snelling	Dustin Brinkmann	Central	Current
2	Cecily Lampkin	Dustin Brinkmann	Central	Current
3	Versie Hillebrand	Dustin Brinkmann	Central	Current
4	Lajuana Vencill	Dustin Brinkmann	Central	Current
5	Moses Frase	Dustin Brinkmann	Central	Current
6	Jonathan Berthelot	Melvin Marxen	Central	Current
7	Marty Freudenburg	Melvin Marxen	Central	Current
8	Gladys Colclough	Melvin Marxen	Central	Current
9	Niesha Huffines	Melvin Marxen	Central	Current

5. 10 tables de ventes par mois Mars – décembre



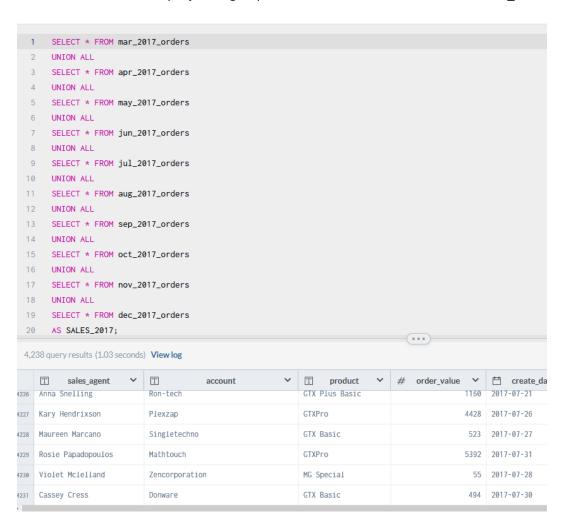
6. sales-pipeline.csv



Questions and Insight

L'analyse des ventes, par agent, accounts, mois

Afin de facilité notre analyse je vais grouper des tables des ventes à un table SALES_2017 :



1) Quelle est la performance de l'équipe commerciale ?

```
SELECT sales_agent, <u>SUM</u>(sales_2017.order_value) FROM Sales_2017

GROUP BY sales_agent

ORDER BY sales_agent DESC;
```

Faisons le groupage des CA par agent pour voir la performance de chaque agent. On peux attribuer un trophé à un meilleur agent :

```
WITH summed_sales AS (
    SELECT
        sales_agent,
        SUM(order_value) AS sum_order
   FROM
        sales_2017
    GROUP BY
        sales_agent
),
ranked_sales AS (
    SELECT
        sales_agent,
        sum_order,
        RANK() OVER (ORDER BY sum_order DESC) AS sales_rank
   FROM
        summed_sales
)
SELECT
   CASE
        WHEN sales_rank = 1 THEN '\ \ ' | sales_agent
        ELSE sales_agent
    END AS sales_agent,
    sum_order
FROM
    ranked_sales
ORDER BY
sales_rank;
```



En regardant les ventes par agent nous pouvons constater que l'agent le plus performant est Darcel Schlecht

2) Quel client est le plus rentable ?

Ici nous avons 2 valeurs order value et revenue regardons tout d'abord les ventes par accounts :

```
SELECT sales_2017.account, <u>SUM</u>(sales_2017.order_value) FROM sales_2017

GROUP BY account

ORDER BY <u>SUM</u>(sales_2017.order_value) DESC
```



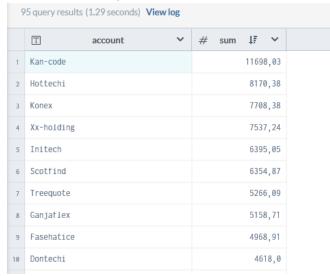
Nous avons Kan -code qui est largement devant suivie de près par le top 5.

```
SELECT account, <u>SUM</u>(accounts.revenue) FROM accounts

GROUP BY account

ORDER BY account DESC;
```

Le revenu par account :



3) Quelle est la distribution des ventes des agents par account

```
SELECT DISTINCT
sales_2017.sales_agent,
sales_2017.account,
    SUM(sales_2017.order_value) OVER(PARTITION BY sales_2017.sales_agent) AS
ag_total_sales,
    SUM(sales_2017.order_value) OVER(PARTITION BY sales_2017.account) AS
account_total_sales,
    ROUND(100.0 * SUM(sales_2017.order_value) OVER(PARTITION BY sales_2017.account)
//
```

		□ account ✓	# ag_total_sales V	# account_total_sales V	# percentage_of_total_sales V
1	Darcel Schlecht	Condax	1153214	206410	17,9
2	Darcel Schlecht	Hottechi	1153214	194957	16,91
3	Darcel Schlecht	Treequote	1153214	176751	15,33
4	Darcel Schlecht	Warephase	1153214	170046	14,75
5	Darcel Schlecht	Isdom	1153214	164683	14,28
30	Vicki Laflamme	Kan-code	478396	341455	71,37
31	Vicki Laflamme	Konex	478396	269245	56,28
32	Vicki Laflamme	Condax	478396	206410	43,15
33	Vicki Laflamme	Cheers	478396	198020	41,39
34	Vicki Laflamme	Hottechi	478396	194957	40,75
35	Vicki Laflamme	Goodsilron	478396	182522	38,15
36	Vicki Laflamme	Xx-holding	478396	169357	35,4
61	Kary Hendrixson	Kan-code	454298	341455	75,16
62	Kary Hendrixson	Konex	454298	269245	59,27
63	Kary Hendrixson	Condax	454298	206410	45,43
64	Kary Hendrixson	Cheers	454298	198020	43,59
65	Kary Hendrixson	Hottechi	454298	194957	42,91
66	Kary Hendrixson	Goodsilron	454298	182522	40,18
92	Cassey Cress	Kan-code	450489	341455	75,8
93	Cassey Cress	Konex	450489	269245	59,77
94	Cassey Cress	Condax	450489	206410	45,82
95	Cassey Cress	Mathtouch	450489	163339	36,26
16	Cassey Cress	Plussunin	450489	155195	34,45
7	Cassey Cress	Umbrella Corporation	450489	152701	33,9

Le client Kan-code représente plus de 70% de nos ventes.

4) Quel est le produit le plus vendu

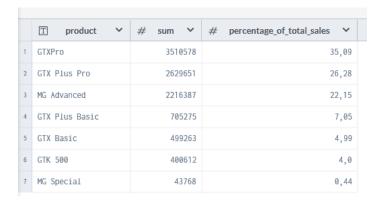
```
SELECT sales_2017.product, <u>SUM</u>(sales_2017.order_value),

<u>ROUND</u>(100.0 * <u>SUM</u>(sales_2017.order_value) / (SELECT <u>SUM</u>(sales_2017.order_value) FROM
sales_2017), 2) AS percentage_of_total_sales

FROM sales_2017

GROUP BY product

ORDER BY percentage_of_total_sales DESC
```



Analysons s'il y a la corrélation entre le prix et les ventes.

L'information de ventes par régions et le prix de ventes se trouve dans le tables sales_teams et produits.

Pour analyse plus détailler on va réunir les table sales, products et sales team :



Maintenant, nous constatons que ce ne sont pas nécessairement les produits les moins chers qui se vendent le plus. Le top 3 des ventes se situe dans une fourchette de prix moyens compris entre 1000 et 5500 euros :

```
5482
 GTX Plus Pro
                               2629651
2 MG Advanced
                               2216387
                                              3393
3 GTX Plus Basic
                                705275
                                              1096
4 GTX Basic
                                499263
                                               550
5 GTK 500
                                400612
                                              26768
6 MG Special
                                43768
                                               55
```

```
SELECT
    sales_2017_product,
    total_order_value,
    sales_price,
    RANK() OVER (ORDER BY total_order_value DESC) AS rrank
```

5) Les ventes par mois :

```
SELECT MONTH(sales_2017_union.create_date) AS MONTH,

SUM(sales_2017_union.order_value)

FROM sales_2017_union

GROUP BY MONTH(sales_2017_union.create_date)

ORDER BY MONTH(sales_2017_union.create_date);
```

	#	MONTH	~	#	sum	~	
1			3		113	4672	
2			4		72	1932	
3			5		102	5713	
4			6		133	8466	
5			7		69	6932	
6			8		105	0059	
7			9		123	5264	
8			10		73	1980	
9			11		93	8943	
0			12		113	1573	

Nous avons le pic en Juin et Septembre, on peut descendre et voir quel est produits vendu plus p.ex. en Septembre :

```
MONTH(sales_2017_union.create_date) AS MONTH,
product,
SUM(sales_2017_union.order_value) AS total_sales
FROM sales_2017_union
```

```
WHERE MONTH(sales_2017_union.create_date) = 9
GROUP BY MONTH, product
ORDER BY total_sales DESC;
```

	# M	ONTH	~	T	product	~	#	total_sales	~
1			9	GTXP	го			48	9195
2			9	GTX I	Plus Pro			32	5614
3			9	MG A	dvanced			25	0431
4			9	GTX I	Plus Basic			7	6739
5			9	GTX I	Basic			5	9155
6			9	GTK	500			2	9166
7			9	MG S	pecial				4964

La répartition des ventes en septembre semble similaire à celle observée pour les autres mois.

6) Quelle région réalise le plus de ventes ?

```
sales_2017_total.regional_office,

SUM(sales_2017_total.order_value) AS total_sales

FROM sales_2017_total

GROUP BY regional_office

ORDER BY total_sales DESC;
```

		# total_sales V
1	West	2447825
2	Central	2242184
3	East	1804947

7) Quel manager apporte le plus ?

```
SELECT

sales_2017_total.manager,

SUM(sales_2017_total.order_value) AS total_sales

FROM sales_2017_total
```

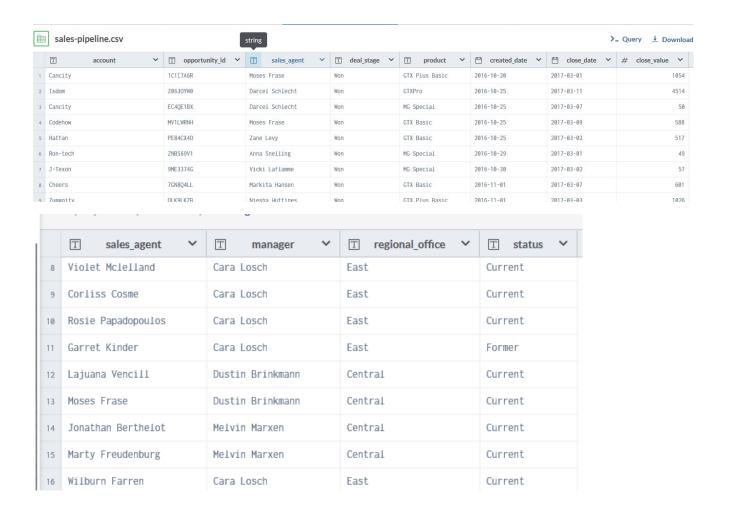
```
GROUP BY manager

ORDER BY total_sales DESC;
```

	T manager	<pre># total_sales </pre>
1	Celia Rouche	1259786
2	Summer Sewald	1188039
3	Melvin Marxen	1147821
4	Rocco Neubert	1110252
5	Dustin Brinkmann	1094363
6	Cara Losch	694695

Sales pipeline Analyse

Nous avons encore 2 tables avec l'analyse des pipelines de ventes sales-pipeline et sales team



On va les réunir pour avoir l'information total sur une seule BD :

Nous avons le pipeline avec des opportunité de ventes avec les attributs suivantes :

- Account
- Opportunity_id
- Sales agent
- Deal_stage
- Product
- Created date
- Closed date
- Close value
- Manager Regional Office
- Status

Je vais faire analyse des efficacité de ventes avec ces opportunités.

1) Quelle est la pourcentage de opportunités gagné ?

```
deal_stage,
SUM (close_value),
COUNT(*) AS num_deals,
COUNT(*) * 100.0 / (SELECT COUNT(*) FROM full_pipeline) AS percentage_deals

FROM
full_pipeline
GROUP BY
deal_stage;
```

	# sum v	# num_deals V	# percentage_deals V
In Progress	No data.	2089	23.7386
Lost	0	2473	28.1023
Won	10005534	4238	48.1591

Nous avons 48% d'opportunités gagné pour CA 10 005 534, par contre nous n'avons pas de données pour les opportunités on progress et lost.

2) Analyse des meilleurs apporteurs d'affaires

LE meilleur commerciale est Darcel Schiecnht avec 1 153 214 et 349 opportunité gagné .

Top 5 client est :

```
SELECT

account,

SUM(close_value) AS total_close_value,

COUNT (deal_stage) AS num_deal

FROM

full_pipeline

WHERE

deal_stage = 'Won'

GROUP BY

account

ORDER BY

total_close_value DESC LIMIT 5;
```

	T account V	# total_close_value V	# num_deal V
1	Kan-code	341455	115
2	Konex	269245	108
3	Condax	206410	105
4	Cheers	198020	57
5	Hottechi	194957	111

Le dispatch entre les régions est +/- égales

3 query results (0.53 seconds) View log				
	□ regional_office ✓	# total_close_value >	# num_deal V	
1	West	3568647	1438	
2	Central	3346293	1629	
3	East	3090594	1171	

Proposition d'un Dashboard CRM avec Power BI

