

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

	 account	# revenue 	# employees 
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	 name	# mgr_id 	 regional_office	 status
1	10001	Anna Snelling	10036	Central	Current
2	10002	Cecily Lampkin	10036	Central	Current
3	10003	Versie Hillebrand	10036	Central	Current
4	10004	Lajuana Vencill	10036	Central	Current
5	10005	Moses Frase	10036	Central	Current

3. products.csv

	<input type="text"/> product ▼	# sales_price ▼
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

	<input type="text"/> sales_agent ▼	<input type="text"/> manager ▼	<input type="text"/> regional_office ▼	<input type="text"/> status ▼
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

	<input type="text"/> sales_agent ▼	<input type="text"/> account ▼	<input type="text"/> product ▼	# order_value ▼	<input type="text"/> create_date ▼
1	Vicki Laflamme	Golddex	GTX Basic	552	2017-10-01
2	Markita Hansen	J-Texon	GTX Basic	566	2017-10-01
3	Donn Cantrell	Vehement Capital Partners	GTXPro	5868	2017-10-04
4	James Ascencio	Cheers	GTX Plus Pro	5784	2017-10-02
5	Jonathan Berthelot	Conecom	GTXPro	3485	2017-10-02
6	Rosalina Dieter	Newex	MG Special	55	2017-10-01
7	Versie Hillebrand	Inity	GTX Plus Pro	5775	2017-10-01
8	Darcel Schleicht	Hottechi	GTXPro	5357	2017-10-02

6. sales-pipeline.csv

	account	opportunity_id	sales_agent	deal_stage	product	created_date	close_date	# close_value
1	Cancity	1C1I7A6R	Moses Frase	Won	GTX Plus Basic	2016-10-20	2017-03-01	1054
2	Isdom	Z0630Y00	Darcel Schlecht	Won	GTXPro	2016-10-25	2017-03-11	4514
3	Cancity	EC4QE1BX	Darcel Schlecht	Won	MG Special	2016-10-25	2017-03-07	50
4	Codehow	MV1LWRNH	Moses Frase	Won	GTX Basic	2016-10-25	2017-03-09	588
5	Hatran	PE84CK40	Zane Levy	Won	GTX Basic	2016-10-25	2017-03-02	517
6	Ron-tech	ZNBS69V1	Anna Snelling	Won	MG Special	2016-10-29	2017-03-01	49
7	J-Texon	9ME3374G	Vicki Laflamme	Won	MG Special	2016-10-30	2017-03-02	57
8	Cheers	7GN0Q4LL	Markita Hansen	Won	GTX Basic	2016-11-01	2017-03-07	601
9	Zumgoity	OLK9LKZB	Niesha Huffines	Won	GTX Plus Basic	2016-11-01	2017-03-03	1026
10	No data.	HAXMC4IX	James Ascencio	In Progress	MG Advanced	2016-11-03	No data.	No data.
11	Bioholding	NL3JZH1Z	Anna Snelling	Won	MG Special	2016-11-04	2017-03-10	53
12	Genco Pura Olive Oil Company	KWVA7VR1	Gladys Colclough	Lost	GTXPro	2016-11-04	2017-03-18	0
13	Sunnamplex	S8DX3XOU	James Ascencio	Won	GTX Plus Pro	2016-11-04	2017-03-10	5169
14	Sonron	ENB2XD8G	Maureen Marcano	Won	GTX Plus Pro	2016-11-04	2017-03-06	4631

Questions and Insight

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

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

```

1  SELECT * FROM mar_2017_orders
2  UNION ALL
3  SELECT * FROM apr_2017_orders
4  UNION ALL
5  SELECT * FROM may_2017_orders
6  UNION ALL
7  SELECT * FROM jun_2017_orders
8  UNION ALL
9  SELECT * FROM jul_2017_orders
10 UNION ALL
11 SELECT * FROM aug_2017_orders
12 UNION ALL
13 SELECT * FROM sep_2017_orders
14 UNION ALL
15 SELECT * FROM oct_2017_orders
16 UNION ALL
17 SELECT * FROM nov_2017_orders
18 UNION ALL
19 SELECT * FROM dec_2017_orders
20 AS SALES_2017;

```

4,238 query results (1.03 seconds) [View log](#)

	sales_agent	account	product	# order_value	create_da
4226	Anna Snelling	Ron-tech	GTX Plus Basic	1160	2017-07-21
4227	Kary Hendrixson	Plexzap	GTXPro	4428	2017-07-26
4228	Maureen Marcano	Singletechno	GTX Basic	523	2017-07-27
4229	Rosie Papadopoulos	Mathtouch	GTXPro	5392	2017-07-31
4230	Violet Mclelland	Zencorporation	MG Special	55	2017-07-28
4231	Cassey Cress	Donware	GTX Basic	494	2017-07-30

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

```
SELECT sales_agent, SUM(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 peut attribuer un trophée 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;
```

	sales_agent	#	sum_order
1	Darcel Schlecht		1153214
2	Vicki Laflamme		478396
3	Kary Hendrixson		454298
4	Cassey Cress		450489
5	Donn Cantrell		445860
6	Reed Clapper		438336
7	Zane Levy		430068
8	Corliss Cosme		421036
9	James Ascencio		413533
10	Daniell Hammack		364229
11	Maureen Marcano		350395
12	Gladys Colclough		345674
13	Markita Hansen		328792

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, SUM(sales_2017.order_value) FROM sales_2017
GROUP BY account
ORDER BY SUM(sales_2017.order_value) DESC
```

	<input type="text"/>	account	#	sum
1		Kan-code		341455
2		Konex		269245
3		Condax		206410
4		Cheers		198020
5		Hottechi		194957
6		Goodsilron		182522
7		Treequote		176751
8		Warephase		170046
9		Xx-holding		169357
10		Isdom		164683

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

```
SELECT account, SUM(accounts.revenue) FROM accounts
GROUP BY account
ORDER BY account DESC;
```

- Le revenu par account :

95 query results (1.29 seconds) View log				
	<input type="text"/>	account	#	sum
1		Kan-code		11698,03
2		Hottechi		8170,38
3		Konex		7708,38
4		Xx-holding		7537,24
5		Initech		6395,05
6		Scottfind		6354,87
7		Treequote		5266,09
8		Ganjaflex		5158,71
9		Fasehatice		4968,91
10		Dontechi		4618,0

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)
/
```

```

SUM(sales_2017.order_value) OVER(PARTITION BY sales_2017.sales_agent), 2)
AS percentage_of_total_sales
FROM sales_2017
ORDER BY ag_total_sales DESC, account_total_sales DESC;

```

	sales_agent	account	# ag_total_sales	# account_total_sales	# percentage_of_total_sales
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
96	Cassey Cress	Plussunin	450489	155195	34,45
97	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, SUM(sales_2017.order_value),
ROUND(100.0 * SUM(sales_2017.order_value) / (SELECT SUM(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

```

	product	# sum	# percentage_of_total_sales
1	GTXPro	3510578	35,09
2	GTX Plus Pro	2629651	26,28
3	MG Advanced	2216387	22,15
4	GTX Plus Basic	705275	7,05
5	GTX Basic	499263	4,99
6	GTK 500	400612	4,0
7	MG Special	43768	0,44

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 :

```
SELECT * FROM sales_2017 AS Sales_2017_Total
JOIN
    sales_teams ON sales_2017.sales_agent = sales_teams.sales_agent
JOIN
    products ON sales_2017.product = products.product;
```

	sales_2017.sales_agent	account	sales_2017.product	# order_value	create_date	sales_teams.sales_agent	manager	regional_office	status	products.product	# sales_price
1	Boris Faz	Betasoloin	GTX Plus Basic	992	2017-04-03	Boris Faz	Rocco Neubert	East	Current	GTX Plus Basic	1096
2	Anna Snelling	dambase	GTX Plus Pro	5060	2017-04-01	Anna Snelling	Dustin Brinkmann	Central	Current	GTX Plus Pro	5482
3	Cassey Cress	Kan-code	GTX Basic	585	2017-04-01	Cassey Cress	Rocco Neubert	East	Current	GTX Basic	550
4	Versie Hillebrand	Zungoity	MG Advanced	3251	2017-04-01	Versie Hillebrand	Dustin Brinkmann	Central	Current	MG Advanced	3393
5	Violet Mclelland	Mathtouch	GTX Plus Basic	941	2017-04-01	Violet Mclelland	Cara Losch	East	Current	GTX Plus Basic	1096
6	Anna Snelling	Codehow	MG Advanced	3445	2017-04-04	Anna Snelling	Dustin Brinkmann	Central	Current	MG Advanced	3393
7	Daniell Hammack	Vehement Capital Partner	GTX Basic	540	2017-04-02	Daniell Hammack	Rocco Neubert	East	Current	GTX Basic	550
8	Versie Hillebrand	Rangreen	MG Special	52	2017-04-02	Versie Hillebrand	Dustin Brinkmann	Central	Current	MG Special	55

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 :

	sales_2017_product	# total_order_value	# sales_price	# rrank
1	GTX Plus Pro	2629651	5482	1
2	MG Advanced	2216387	3393	2
3	GTX Plus Basic	705275	1096	3
4	GTX Basic	499263	550	4
5	GTK 500	400612	26768	5
6	MG Special	43768	55	6

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



```

FROM (
    SELECT
        sales_2017_product,
        SUM(order_value) AS total_order_value,
        sales_price
    FROM
        sales_2017_total
    GROUP BY
        sales_2017_product, sales_price
) AS subquery_alias;

```

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			1134672	
2		4			721932	
3		5			1025713	
4		6			1338466	
5		7			696932	
6		8			1050059	
7		9			1235264	
8		10			731980	
9		11			938943	
0		12			1131573	

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

```

SELECT
    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;
```

	#	MONTH	product	#	total_sales
1		9	GTXPro		489195
2		9	GTX Plus Pro		325614
3		9	MG Advanced		250431
4		9	GTX Plus Basic		76739
5		9	GTX Basic		59155
6		9	GTK 500		29166
7		9	MG Special		4964

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

6) Quelle région réalise le plus de ventes ?

```
SELECT
    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;
```

	regional_office	#	total_sales
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;
```

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

```
SELECT
    sales_2017_total.regional_office,
    sales_2017_total.manager,
    SUM(sales_2017_total.order_value) OVER(PARTITION BY
sales_2017_total.regional_office) AS reg_total_sales,
    SUM(sales_2017_total.order_value) OVER(PARTITION BY sales_2017_total.manager) AS
manager_total_sales,
    ROUND(100.0 * SUM(sales_2017_total.order_value) OVER(PARTITION BY
sales_2017_total.manager) /
    SUM(sales_2017_total.order_value) OVER(PARTITION BY sales_2017_total.manager), 2)
    AS percentage_of_total_sales
FROM sales_2017_total
ORDER BY manager_total_sales DESC, reg_total_sales DESC;
```

Sales pipeline Analyse

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

sales-pipeline.csv string Query Download

	account	opportunity_id	sales_agent	deal_stage	product	created_date	close_date	#	close_value
1	Cancity	1C117A6R	Moses Frase	Won	GTX Plus Basic	2016-10-20	2017-03-01		1054
2	Isdom	Z0630YW0	Darcel Schlecht	Won	GTXPro	2016-10-25	2017-03-11		4514
3	Cancity	EC4QE1BX	Darcel Schlecht	Won	MG Special	2016-10-25	2017-03-07		50
4	Codehow	MV1LWRNH	Moses Frase	Won	GTX Basic	2016-10-25	2017-03-09		588
5	Haffan	PE84CX40	Zane Levy	Won	GTX Basic	2016-10-25	2017-03-02		517
6	Ron-tech	ZNB569V1	Anna Snelling	Won	MG Special	2016-10-29	2017-03-01		49
7	J-Texon	9ME3374G	Vicki Laflamme	Won	MG Special	2016-10-30	2017-03-02		57
8	Cheers	7GN8Q4LL	Markita Hansen	Won	GTX Basic	2016-11-01	2017-03-07		601
9	Zumosity	0IK9IK7R	Niesha Huttines	Won	GTX Plus Basic	2016-11-01	2017-03-03		1026

	sales_agent	manager	regional_office	status
8	Violet McLelland	Cara Losch	East	Current
9	Corliss Cosme	Cara Losch	East	Current
10	Rosie Papadopoulos	Cara Losch	East	Current
11	Garret Kinder	Cara Losch	East	Former
12	Lajuana Vencill	Dustin Brinkmann	Central	Current
13	Moses Frase	Dustin Brinkmann	Central	Current
14	Jonathan Berthelot	Melvin Marxen	Central	Current
15	Marty Freudenburg	Melvin Marxen	Central	Current
16	Wilburn Farren	Cara Losch	East	Current

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

```

WITH `full pipeline` AS (
    SELECT sales_teams.*,
           sales_pipeline.*
    FROM sales_teams, sales_pipeline
    WHERE (sales_pipeline.sales_agent = sales_teams.sales_agent)
)
SELECT *
FROM `full pipeline`

Or

SELECT *
FROM sales_pipeline
JOIN sales_teams ON sales_pipeline.sales_agent = sales_teams.sales_agent

```


Nous avons le pipeline avec des opportunit  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 efficacit  de ventes avec ces opportunit s.

1) Quelle est la pourcentage de opportunit s gagn  ?

```
SELECT
    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;
```

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

Nous avons 48% d'opportunit s gagn  pour CA 10 005 534 , par contre nous n'avons pas de donn es pour les opportunit s on progress et lost.

2) Analyse des meilleurs apporteurs d'affaires

LE meilleur commerciale est Darcel Schiecnht avec 1 153 214 et 349 opportunit  gagn  .

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;
```

	<input type="text"/> account ▼	# total_close_value ▼	# num_deal ▼
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				
	<input type="text"/> regional_office ▼	# total_close_value ▼	# num_deal ▼	
1	West	3568647	1438	
2	Central	3346293	1629	
3	East	3090594	1171	

Sales Analysis

Total sales
6494956

Average by month
649496

Month variation
13,14 %

BestSalesAgent
Darcel Schlecht

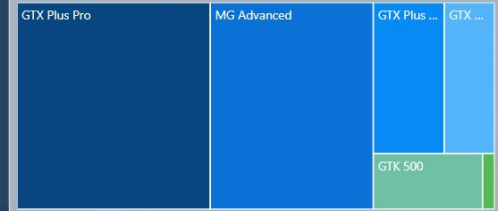
BestProduct
GTK Plus Pro

Regional office

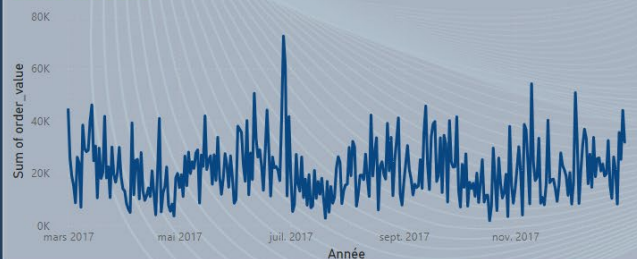
CA by Sales Agent and regional office



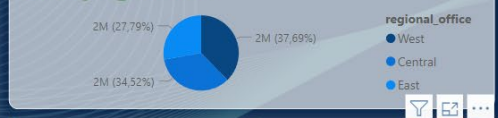
Sales by products product



Sales timeline



Sales by regiona office



SALES by manager

