



Data Analyst Specialist Graduation Project Framework

The Data Analyst Specialist Graduation Project Framework provides a comprehensive guide for students to showcase their data analysis skills acquired throughout their course. This framework outlines a structured approach to completing a successful data analysis project, from inception to final presentation.

The project encompasses various stages including problem definition, data preparation, exploratory data analysis, advanced analytical techniques, and the presentation of actionable insights. By following this framework, students can demonstrate their proficiency in using data analysis tools and techniques to solve real-world problems.

Agenda



The project begins with a clear proposal that outlines the project's title, team members, and a well-defined problem statement. Students must articulate the business question they aim to address through data analysis and propose a solution using specific analytical techniques and tools.

A comprehensive project plan is crucial, detailing key milestones, task breakdowns, and resource requirements. This plan should also include a risk management strategy to anticipate and mitigate potential challenges throughout the project lifecycle.

1 Project Proposal
Team member, Business questions

2 Project steps
Get Data , Explore Data , Prepare Data, Analyze Data.

3 Visualization
Present Finding.

Team Member

Hadeer Nabil Elsayed

Raghda Mahmoud

Aya mohamed



Made with Gamma

Business Questions



designed by  freepik.com

CRM Sales Opportunities

1- Total Sales:

- What is the total revenue generated from all sales in the current period, and how does it compare to the previous period?

2- Won Deals:

- How many deals have been successfully closed in the current period, and what percentage of total deals were won?

3- Lost Deals:

- How many deals were lost during this period, and what are the common reasons for losing them?

4- Prospecting Deals:

- How many deals are currently in the prospecting stage, and what is the conversion rate from this stage to the next?

5- Engagement Deals:

- How many deals are in the engagement phase, and what is the likelihood of these deals progressing to a won status?

6- Count of Product:

- How many unique products were involved in sales during this period, and which products contribute most to overall sales performance?

7- Count of Deal Stage by Deal Stage:

- How are deals distributed across various stages in the sales pipeline, and which stage has the highest number of deals?

8- Sales Agent Performance:

- Which sales agent has the highest total closed deal value, and what strategies are employed by top-performing agents?

9- Sum of Won Deals by Manager:

- Which manager oversees the highest number of won deals, and how do the won deal totals compare across different managers?

10- Sum of Close Value and Count of Won Deals by Quarter:

- How does the total close value change from one quarter to the next, and which quarter has the highest count of won deals?

11- Count of Account by Sector:

- Which sector has the highest number of accounts, and are there underserved sectors that could be targeted for future sales efforts?

12- Max of Revenue and Sum of Employees by Account:

- Which accounts generate the maximum revenue, and what is the average revenue generated per employee for each account?

13- Count of Account, Count of Office Location, and Sum of Revenue by Sector and Office Location:

- How is the total revenue distributed across different sectors and office locations, and are there locations that contribute significantly to specific sectors?

14- Count of Account, Count of Sector, and Count of Office Location:

- What is the total count of accounts, sectors, and office locations, and how does this reflect the company's market reach and diversity?

Data Preparation and Exploratory Data Analysis

The data preparation phase involves collecting relevant datasets from various sources, ensuring accuracy and relevance. Students must implement data cleaning techniques to handle missing data, duplicates, and inconsistencies, preparing the data for analysis.

Exploratory Data Analysis (EDA) follows, utilizing statistical tools and visualizations to explore the dataset. This process generates summary statistics and helps understand data distribution and relationships, laying the groundwork for more advanced analysis.

Business question



designed by freepik.com

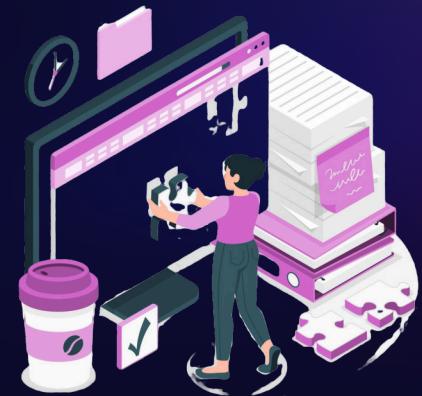
Get Data



Explore Data



Prepare Data



Detailed Data Analysis and Visualization



Insight Generation



Derive meaningful insights from the analysis and visualizations.

Apply techniques like regression, clustering, and hypothesis testing.

Accounts

A screenshot of a data visualization application. On the left, there is a table titled "accounts" with columns: account, sector, year_established, revenue, employees, and office_location. The table contains 40 rows of data. On the right, there is a sidebar titled "Data" with a search bar and a list of datasets: accounts, products, sales_pipeline, sales_teams, and Won Products. The "accounts" dataset is currently selected.

account	sector	year_established	revenue	employees	office_location
Acme Corporation	technolgy	1996	1100.04	2822	United States
Betasoloin	medical	1999	251.41	495	United States
Bioplex	medical	1991	326.82	1016	United States
Blackzim	retail	2009	497.11	1588	United States
Bluth Company	technolgy	1993	1242.32	3027	United States
Bubba Gump	software	2002	987.39	2253	United States
Cancity	retail	2001	718.62	2448	United States
Cheers	entertainment	1993	4269.9	6472	United States
Codehow	software	1998	2714.9	2641	United States
Condax	medical	2017	4.54	9	United States
Conecom	technolgy	2005	1520.66	1806	United States
Dalitechnology	software	2013	98.79	96	United States
dambase	marketing	1995	2173.98	2928	United States
Domzoom	entertainment	1998	217.87	551	United States
Doncon	technolgy	2010	587.72	1501	United States
Donquadtech	technolgy	1992	1712.68	3194	United States
Dontechi	software	1982	4618	10083	United States
Donware	marketing	1999	1197.44	2570	United States
Fasehatice	retail	1990	4968.91	7523	United States
Faxquote	telecommunications	1995	1825.82	5595	United States
Finhigh	finance	2006	1102.43	1759	United States
Finjob	employment	1988	2059.9	3644	United States
Funholding	finance	1991	2819.5	7227	United States
Gekko & Co	retail	1990	2520.83	3502	United States
Gogozoom	telecommunications	2007	86.68	187	United States

Product

A screenshot of a data visualization application. On the left, there is a table titled "products" with columns: product, series, sales_price, and won_deals. The table contains 8 rows of data. On the right, there is a sidebar titled "Data" with a search bar and a list of datasets: accounts, products, sales_pipeline, sales_teams, and Won Products. The "products" dataset is currently selected.

product	series	sales_price	won deals
GTX Basic	GTX	550	915
GTX Pro	GTX	4821	729
MG Special	MG	55	793
MG Advanced	MG	3393	654
GTX Plus Pro	GTX	5482	479
GTX Plus Basic	GTX	1096	653
GTX 500	GTX	26768	15

Sales_Pipeline

opportunity_id	sales_agent	product	account	deal_stage	engage_date	close_date	close_value	Won Deals
TCHFT25B	Darcel Schlecht	GTX Basic	Stanredtax	Lost	Monday, November 14, 2016	Wednesday, March 8, 2017	0	0
6ROE69W5	Rosalina Dieter	GTX Basic	Plexzap	Lost	Wednesday, November 16, 2016	Saturday, March 11, 2017	0	0
XK668R7C	Darcel Schlecht	GTX Basic	Dontechi	Lost	Saturday, December 3, 2016	Sunday, March 12, 2017	0	0
2LP5SUR5	Lajuana Vencill	GTX Basic	Treequote	Lost	Sunday, December 11, 2016	Thursday, March 9, 2017	0	0
ALW92XQL	Gladys Colclough	GTX Basic	Isdom	Lost	Tuesday, December 20, 2016	Saturday, March 25, 2017	0	0
TJ6ZA1M3	Kary Hendrixson	GTX Basic	Dalttechnology	Lost	Tuesday, December 20, 2016	Tuesday, March 7, 2017	0	0
2DKO9X1T	Jonathan Berthelot	GTX Basic	Rundofase	Lost	Wednesday, December 21, 2016	Friday, March 10, 2017	0	0
IBDEQVBN	Kary Hendrixson	GTX Basic	Faxquote	Lost	Wednesday, December 21, 2016	Tuesday, March 28, 2017	0	0
S91N9YBL	Vicki Laflamme	GTX Basic	Sunnamplex	Lost	Wednesday, December 21, 2016	Wednesday, March 8, 2017	0	0
LWPJEV3U	Niesha Huffines	GTX Basic	Domzoom	Lost	Saturday, December 24, 2016	Friday, March 10, 2017	0	0
GK8YD2XW	Darcel Schlecht	GTX Basic	Rangreen	Lost	Monday, December 26, 2016	Thursday, March 23, 2017	0	0
77ZECYYJ	Kary Hendrixson	GTX Basic	Finjob	Lost	Tuesday, January 3, 2017	Friday, March 17, 2017	0	0
85XNDIK1	Kary Hendrixson	GTX Basic	Blackzim	Lost	Thursday, January 5, 2017	Thursday, March 30, 2017	0	0
R0W6A9BG	Lajuana Vencill	GTX Basic	Treequote	Lost	Friday, January 6, 2017	Wednesday, May 10, 2017	0	0
P5EGQ0HI	Lajuana Vencill	GTX Basic	Zumgoity	Lost	Friday, January 6, 2017	Monday, March 20, 2017	0	0
S3XI6YB3	Versie Hillebrand	GTX Basic	Domzoom	Lost	Friday, January 6, 2017	Friday, March 24, 2017	0	0
XC4LTIEO	Lajuana Vencill	GTX Basic	Conecom	Lost	Tuesday, January 10, 2017	Tuesday, May 2, 2017	0	0
977FJ9QQ	Boris Faz	GTX Basic	The New York Inquirer	Lost	Wednesday, January 18, 2017	Wednesday, March 29, 2017	0	0
D7H1EGS9	Zane Levy	GTX Basic	Lexiqvolax	Lost	Friday, January 20, 2017	Monday, March 20, 2017	0	0
N852D2Z8	Zane Levy	GTX Basic	Scotfind	Lost	Monday, January 23, 2017	Thursday, March 30, 2017	0	0
4FAEDB16	Markita Hansen	GTX Basic	J-Texon	Lost	Wednesday, January 25, 2017	Monday, May 22, 2017	0	0
AWDOKG69	Garret Kinder	GTX Basic	Bluth Company	Lost	Friday, January 27, 2017	Wednesday, May 10, 2017	0	0
RBXWRHBJ	Jonathan Berthelot	GTX Basic	Rundofase	Lost	Sunday, January 29, 2017	Monday, March 6, 2017	0	0
W4K39GNG	Zane Levy	GTX Basic	Scotfind	Lost	Wednesday, February 1, 2017	Wednesday, March 29, 2017	0	0
ZL6TJDPG	Jonathan Berthelot	GTX Basic	Globex Corporation	Lost	Friday, February 3, 2017	Tuesday, May 16, 2017	0	0

Search

- > accounts
- > products
- > sales_pipeline
- > sales_teams
- > Won Products

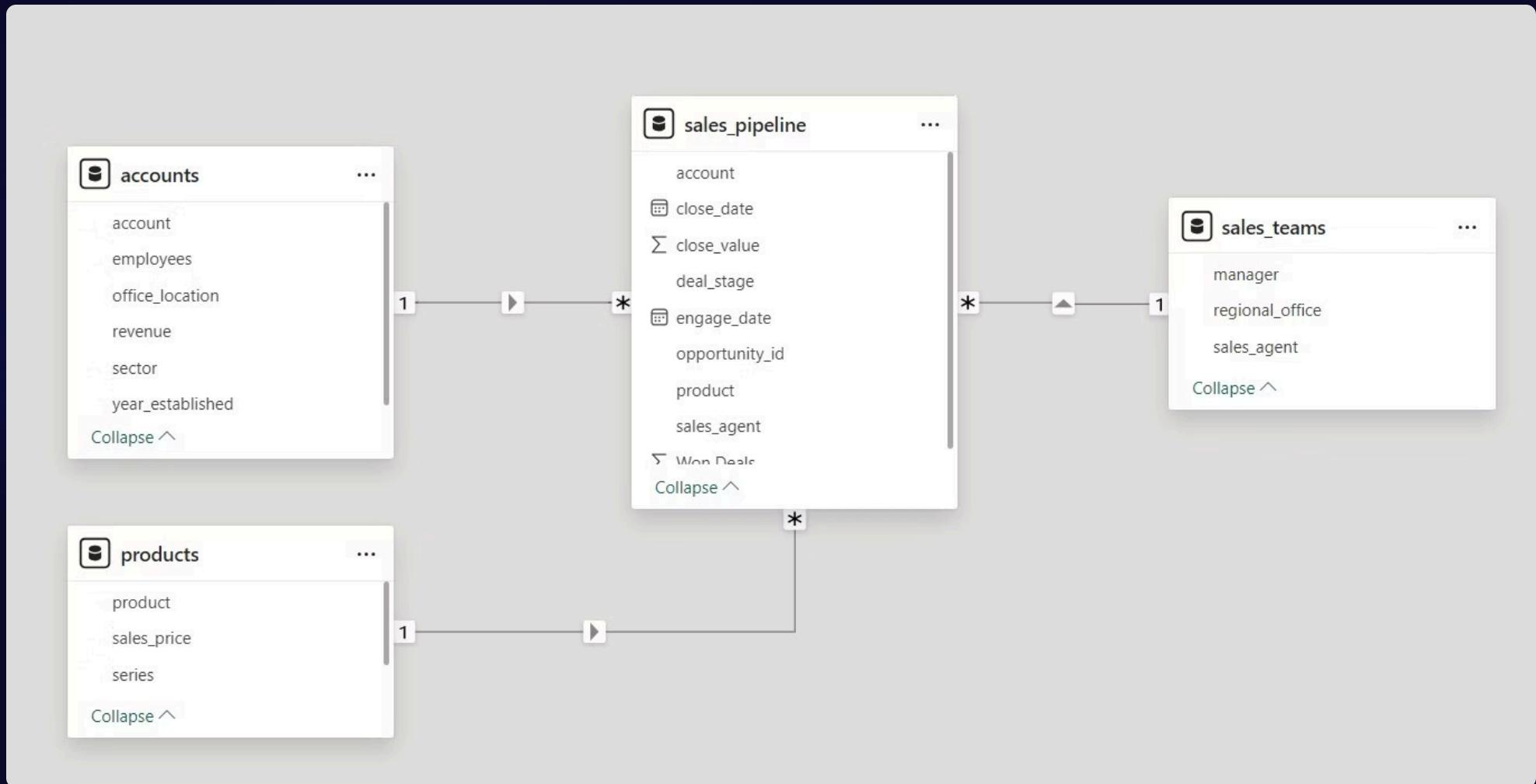
Sales_Team

sales_agent	manager	regional_office
Anna Snelling	Dustin Brinkmann	Central
Cecily Lampkin	Dustin Brinkmann	Central
Versie Hillebrand	Dustin Brinkmann	Central
Lajuana Vencill	Dustin Brinkmann	Central
Moses Frase	Dustin Brinkmann	Central
Jonathan Berthelot	Melvin Marxen	Central
Marty Freudenburg	Melvin Marxen	Central
Gladys Colclough	Melvin Marxen	Central
Niesha Huffines	Melvin Marxen	Central
Darcel Schlecht	Melvin Marxen	Central
Mei-Mei Johns	Melvin Marxen	Central
Violet Mclelland	Cara Losch	East
Corliss Cosme	Cara Losch	East
Rosie Papadopoulos	Cara Losch	East
Garret Kinder	Cara Losch	East
Wilburn Farren	Cara Losch	East
Elizabeth Anderson	Cara Losch	East
Daniell Hammack	Rocco Neubert	East
Cassey Cress	Rocco Neubert	East
Donn Cantrell	Rocco Neubert	East
Reed Clapper	Rocco Neubert	East
Boris Faz	Rocco Neubert	East
Natalya Ivanova	Rocco Neubert	East
Vicki Laflamme	Celia Rouche	West
Rosalina Dieter	Celia Rouche	West

Search

- > accounts
- > products
- > sales_pipeline
- > sales_teams
- > Won Products

Data Modeling



Using Excel TO Cleaning Data

"Using the formula =IF(E2="Won",1,0) in Excel, I was able to identify winning and losing deals."

	B	C	D	E	F	G	H	I
1	sales_agent	product	account	deal_stage	engage_date	close_date	close_value	Won Deals
2	Moses Frase	GTx Plus Basic	Cancyty	Won	10/20/2016	3/1/2017	1054	1
3	Darcel Schlecht	GTxPro	Isdom	Won	10/25/2016	3/11/2017	4514	1
4	Darcel Schlecht	MG Special	Cancyty	Won	10/25/2016	3/7/2017	50	1
5	Moses Frase	GTx Basic	Codehow	Won	10/25/2016	3/9/2017	588	1
6	Zane Levy	GTx Basic	Hatfan	Won	10/25/2016	3/2/2017	517	1
7	Anna Snelling	MG Special	Ron-tech	Won	10/29/2016	3/1/2017	49	1
8	Vicki Laflamme	MG Special	J-Texon	Won	10/30/2016	3/2/2017	57	1
9	Markita Hansen	GTx Basic	Cheers	Won	11/1/2016	3/7/2017	601	1
10	Niesha Huffines	GTx Plus Basic	Zumgoity	Won	11/1/2016	3/3/2017	1026	1
11	James Ascencio	MG Advanced	Rantouch	Engaging	11/3/2016			0
12	Anna Snelling	MG Special	Bioholding	Won	11/4/2016	3/10/2017	53	1
13	Gladys Colclough	GTxPro	Genco Pura Olive Oil Compa	Lost	11/4/2016	3/18/2017	0	0
14	James Ascencio	GTx Plus Pro	Sunnplex	Won	11/4/2016	3/10/2017	5169	1
15	Maureen Marcano	GTx Plus Pro	Sonron	Won	11/4/2016	3/6/2017	4631	1
16	Hayden Neloms	MG Advanced	Finjob	Won	11/5/2016	3/11/2017	3393	1
17	Rosalina Dieter	MG Special	Sonron	Lost	11/5/2016	3/3/2017	0	0
18	Rosalina Dieter	MG Advanced	Scotfind	Won	11/5/2016	3/6/2017	3284	1
19	Versie Hillebrand	MG Special	Treequote	Won	11/6/2016	3/5/2017	61	1
20	Daniell Hammack	GTxPro	Xx-zobam	Lost	11/7/2016	3/9/2017	0	0
21	Elease Gluck	MG Special	Rantouch	Won	11/7/2016	3/8/2017	46	1
22	James Ascencio	GTx Plus Pro	Fasehatice	Lost	11/7/2016	3/17/2017	0	0
23	Moses Frase	MG Special	Ron-tech	Won	11/7/2016	3/18/2017	50	1
24	Violet Mclelland	GTx Plus Basic	Vehement Capital Partners	Won	11/7/2016	3/11/2017	1014	1
25	Darcel Schlecht	GTx Basic	Warephase	Won	11/8/2016	3/26/2017	561	1
26	Kami Bicknell	GTx Basic	Zoomit	Won	11/10/2016	3/23/2017	590	1
27	Maureen Marcano	MG Advanced	Ganjaflex	Engaging	11/10/2016			0

Using Python

To

Explore & Clean Data



Made with Gamma

Export data

```
import pandas as pd

df = pd.read_csv('/content/sales_pipeline2.csv')
df
```

	opportunity_id	sales_agent	product	account	deal_stage	engage_date	close_date	close_value
0	1C1I7A6R	Moses Frase	GTX Plus Basic	Cancity	Won	10/20/2016	3/1/2017	1054.0
1	Z063OYW0	Darcel Schlecht	GTXPro	Isdom	Won	10/25/2016	3/11/2017	4514.0
2	EC4QE1BX	Darcel Schlecht	MG Special	Cancity	Won	10/25/2016	3/7/2017	50.0
3	MV1LWRNH	Moses Frase	GTX Basic	Codehow	Won	10/25/2016	3/9/2017	588.0
4	PE84CX4O	Zane Levy	GTX Basic	Hatfan	Won	10/25/2016	3/2/2017	517.0
...
8795	9MIWFW5J	Versie Hillebrand	MG Advanced	NaN	Prospecting	NaN	NaN	NaN
8796	6SLKZ8FI	Versie Hillebrand	MG Advanced	NaN	Prospecting	NaN	NaN	NaN
8797	LIB4KUZJ	Versie Hillebrand	MG Advanced	NaN	Prospecting	NaN	NaN	NaN
8798	18IUIUK0	Versie Hillebrand	MG Advanced	NaN	Prospecting	NaN	NaN	NaN

Cleaning Data

```
import numpy as np

unique_values = df['account'].dropna().unique()
df['account'] = df['account'].apply(lambda x: np.random.choice(unique_values) if pd.isna(x) else x)

df.tail(10)
```

	opportunity_id	sales_agent	product	account	deal_stage	engage_date	close_date	close_value
8790	FCNN6UY0	Versie Hillebrand	MG Advanced	Ganjaflex	Prospecting	NaN	NaN	NaN
8791	8M200Q8V	Versie Hillebrand	MG Advanced	Silis	Prospecting	NaN	NaN	NaN
8792	WNWI8RJU	Versie Hillebrand	MG Advanced	Streethex	Prospecting	NaN	NaN	NaN
8793	4H6DFF04	Versie Hillebrand	MG Advanced	Betasoloin	Prospecting	NaN	NaN	NaN
8794	IUM2FVA1	Versie Hillebrand	MG Advanced	Statholdings	Prospecting	NaN	NaN	NaN
8795	9MIWFW5J	Versie Hillebrand	MG Advanced	Blackzim	Prospecting	NaN	NaN	NaN
8796	6SLKZ8FI	Versie Hillebrand	MG Advanced	Bioholding	Prospecting	NaN	NaN	NaN

Using Power BI

Analyze Data

Measures We used

```
1 total sales = SUMX(products,products[sales_price]*products[won deals])
```

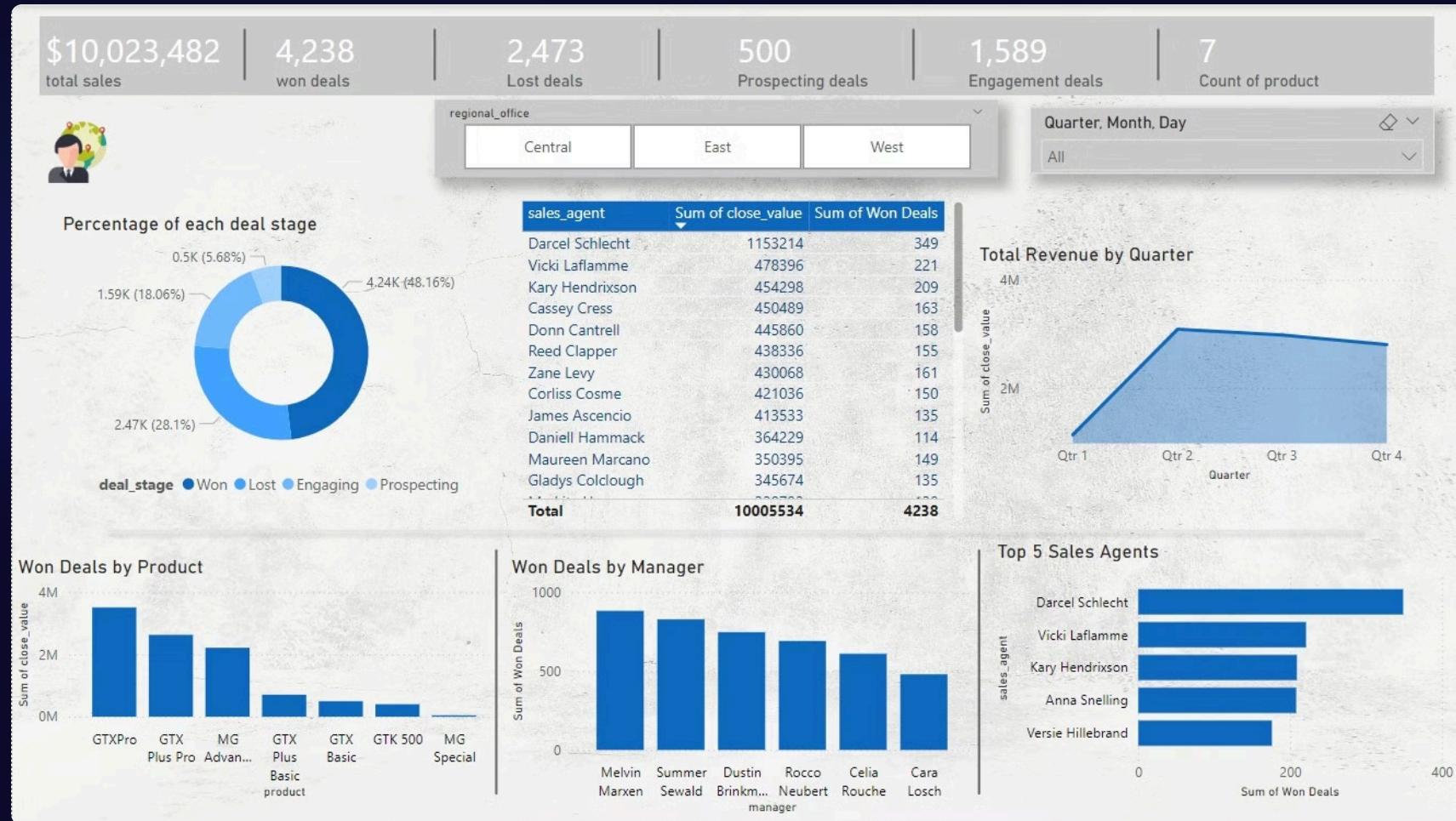
```
1 won deals = CALCULATE(COUNTROWS('sales_pipeline'), 'sales_pipeline'[deal_stage] = "won")
```

```
1 Won Rate =  
2 DIVIDE(COUNTROWS(FILTER('sales_pipeline','sales_pipeline'[deal_stage] = "won")),  
3 | COUNTROWS('sales_pipeline'), 0)
```

```
1 Lost deals = CALCULATE(COUNTROWS('sales_pipeline'), 'sales_pipeline'[deal_stage] = "LOST")
```

```
1 Prospecting deals = CALCULATE(COUNTROWS('sales_pipeline'), 'sales_pipeline'[deal_stage] = "Prospecting")
```

CRM Sales Opportunities



85

Count of account

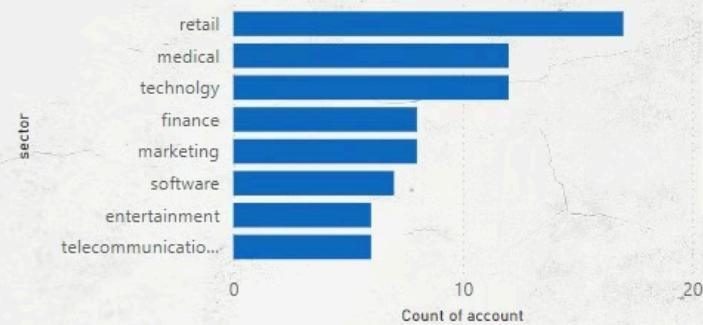
10

Count of sector

15

Count of office_location

Number of Accounts by Industry Sector



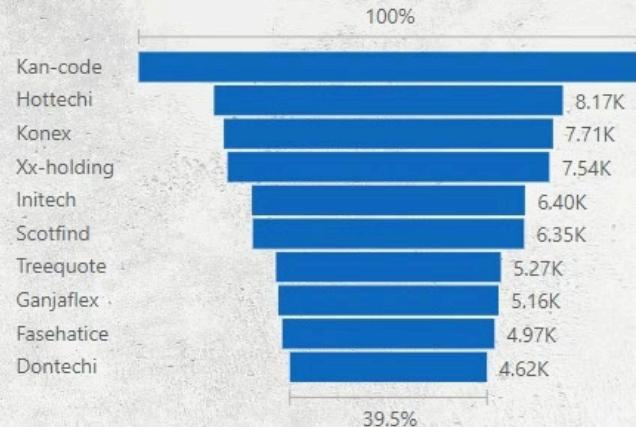
regional_office

Central

East

West

Top Ten accounts by account revenue



Accounts Sectors & Locations



Recommendations

- 1- There is a difference in the sales figures between the sales agents, so we need to provide training for the agents with low sales.
- 2- There is a product (GTK) with low sales but high revenue, so we need to focus on it more and identify the issue to achieve higher sales.
- 3- We need to focus on the retail sector and medical sector, expand into different areas, and dominate the markets with our products in the current regions.

Question time



Thank You