



Analysis of B2B Sales Opportunities Our Team





Abdelrahman Saeed

Mohamed Saied

Mohamed Ismael

Noha Mohamed





Introduction

Data Preparation

DWH Design & Data Modelling

elling

Data Analysis

Report Creation



Introduction



Introduction



This project involves the analysis of B2B sales pipeline data for a fictitious company that specializes in selling computer hardware. The objective is to gain actionable insights into the company's sales operations by examining various aspects of the data, including sales opportunities, product performance, account activities, and sales team efficiency.

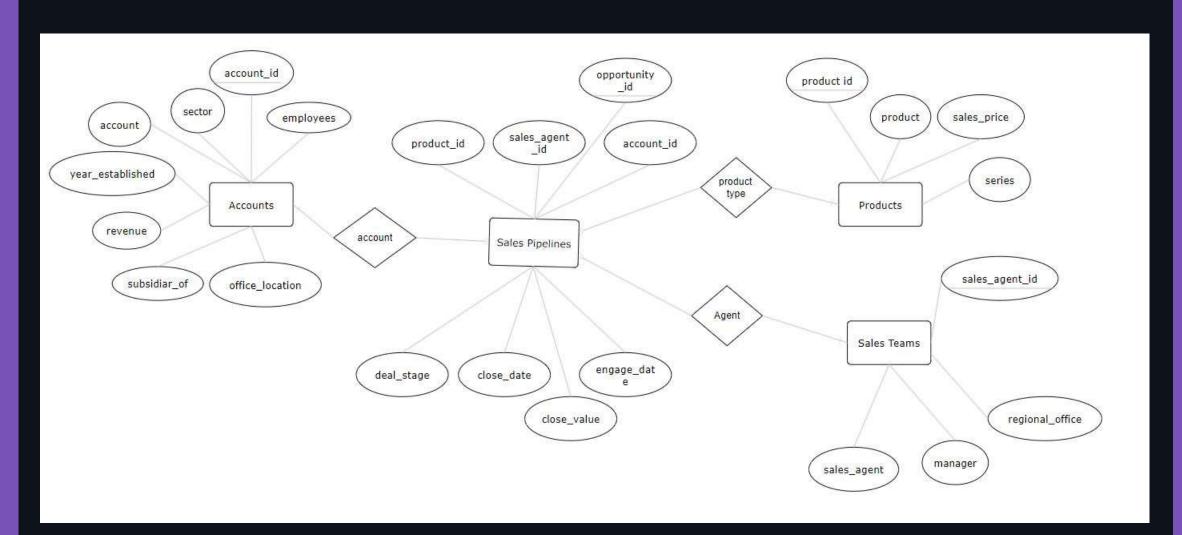
Objectives

The main goals of this project are to:

- Assess Sales Team Performance. How is each sales team performing in comparison to others? Identifying the top-performing teams and those that may need improvement.
- Evaluate Sales Agent Productivity. Are any sales agents significantly lagging behind? Determining which
 agents are underperforming and may require additional support or training.
- Identify Trends. Can any quarter-over-quarter trends be observed? Understanding patterns in sales performance.
- Analyze Product Success Rates. Do any products have better win rates than others? Investigating which
 products are more successful in closing deals.

ER Diagram







Data Preparation



Data Preparation



- 1. Add ID columns: I added unique identifier (ID) columns to the <u>products.csv</u>, <u>accounts.csv</u>, and <u>sales_teams.csv</u> files. These IDs are now referenced in the <u>sales_pipeline.csv</u> file to standardize the data and make querying more efficient.
- 2. Create mapping dictionaries: I created dictionaries to map IDs to full names for the <u>products.csv</u>, <u>accounts.csv</u>, and <u>sales_teams.csv</u> files. These dictionaries serve as lookup tables, ensuring that each ID corresponds correctly to the appropriate name in the data.
- **3. Validate data consistency**: I cross-referenced the dictionaries with the <u>sales_pipeline.csv</u> file to identify any discrepancies, such as names in <u>sales_pipeline.csv</u> that did not match those in the dictionaries. This validation step was crucial for maintaining data integrity.
- **4. Correct misspelled values**: I corrected any misspelled values in the dictionaries to ensure consistency across all files. This helped prevent potential errors during data replacement that could have led to incorrect insights in future analyses.
- 5. Replace full names with IDs: I replaced the full names in the <u>sales_pipeline.csv</u> file with the corresponding IDs for sales agents, products, and accounts, standardizing the data for optimal querying.
- **6. Save the updated data**: Finally, I saved the updated data in the <u>sales_pipeline.csv</u> file, ensuring it was properly prepared and optimized for the next phases of the project.



NOTE:

When collecting data, the data must be appropriate for the project's goal and must be from a reliable source. Therefore, data was selected from the Maven Analytics website.

You can explore it from this [link]







```
for dataframe in dataframes:
  # Construct the full path to the CSV file
  full_path = path + dataframe
  # Read the CSV file into a DataFrame
  df = pd.read_csv(full_path)
  # Find the maximum index in the DataFrame and create a new range for IDs
  max_idx = df.index.max() + 1 # Get the maximum index and add 1 for ID generation
  id_values = np.arange(1, max_idx + 1, 1) # Create an array of IDs starting from 1
  # Insert the new 'id' column at the beginning of the DataFrame
  df.insert(0, 'id', id_values)
  # Construct the save path for the modified CSV file
  save_path = modified_path + dataframe
  # Save the modified DataFrame back to a CSV file without the index
  df.to_csv(save_path, index=False)
```

至

Data before:

opportunity_id	sales_agent	product	account	deal_stage	engage_date	close_date	close_value
1C1I7A6R	Moses Frase	GTX Plus Basic	Cancity	Won	10/20/2016	3/1/2017	1054
Z063OYW0	Darcel Schlecht	GTXPro	Isdom	Won	10/25/2016	3/11/2017	4514
EC4QE1BX	Darcel Schlecht	MG Special	Cancity	Won	10/25/2016	3/7/2017	50
MV1LWRNH	Moses Frase	GTX Basic	Codehow	Won	10/25/2016	3/9/2017	588
PE84CX4O	Zane Levy	GTX Basic	Hatfan	Won	10/25/2016	3/2/2017	517
ZNBS69V1	Anna Snelling	MG Special	Ron-tech	Won	10/29/2016	3/1/2017	49
9ME3374G	Vicki Laflamme	MG Special	J-Texon	Won	10/30/2016	3/2/2017	57
7GN8Q4LL	Markita Hansen	GTX Basic	Cheers	Won	11/1/2016	3/7/2017	601
OLK9LKZB	Niesha Huffines	GTX Plus Basic	Zumgoity	Won	11/1/2016	3/3/2017	1026
HAXMC4IX	James Ascencio	MG Advanced		Engaging	11/3/2016		
NL3JZH1Z	Anna Snelling	MG Special	Bioholding	Won	11/4/2016	3/10/2017	53
KWVA7VR1	Gladys Colclough	GTXPro	Genco Pura Olive Oil Company	Lost	11/4/2016	3/18/2017	0
S8DX3XOU	James Ascencio	GTX Plus Pro	Sunnamplex	Won	11/4/2016	3/10/2017	5169
ENB2XD8G	Maureen Marcano	GTX Plus Pro	Sonron	Won	11/4/2016	3/6/2017	4631
09YE9QOV	Hayden Neloms	MG Advanced	Finjob	Won	11/5/2016	3/11/2017	3393
3F5MZNEH	Rosalina Dieter	MG Special	Sonron	Lost	11/5/2016	3/3/2017	0
M6WEJXC0	Rosalina Dieter	MG Advanced	Scotfind	Won	11/5/2016	3/6/2017	3284
6PTR7VBR	Versie Hillebrand	MG Special	Treequote	Won	11/6/2016	3/5/2017	61
902REDPA	Daniell Hammack	GTXPro	Xx-zobam	Lost	11/7/2016	3/9/2017	0
5J9CMGDV	Elease Gluck	MG Special	Rantouch	Won	11/7/2016	3/8/2017	46
JJXRR8R6	James Ascencio	GTX Plus Pro	Fasehatice	Lost	11/7/2016	3/17/2017	0
WF4HA5NW	Moses Frase	MG Special	Ron-tech	Won	11/7/2016	3/18/2017	50

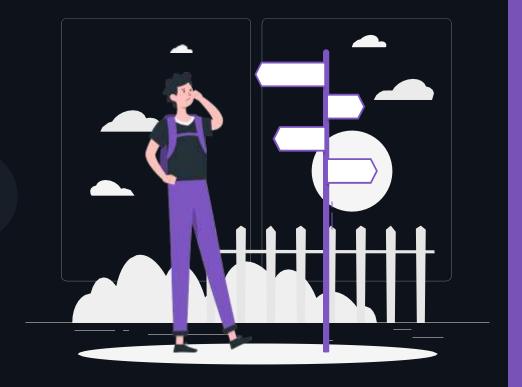


Data after:

opportunity_id	sales_agent	product	account	deal_stage	engage_date	close_date	close_value
1C1I7A6R	5	6	9	Won	10/20/2016	3/1/2017	1054
Z063OYW0	10	2	39	Won	10/25/2016	3/11/2017	4514
EC4QE1BX	10	3	9	Won	10/25/2016	3/7/2017	50
MV1LWRNH	5	1	11	Won	10/25/2016	3/9/2017	588
PE84CX4O	33	1	35	Won	10/25/2016	3/2/2017	517
ZNBS69V1	1	3	59	Won	10/29/2016	3/1/2017	49
9ME3374G	24	3	41	Won	10/30/2016	3/2/2017	57
7GN8Q4LL	27	1	10	Won	11/1/2016	3/7/2017	601
OLK9LKZB	9	6	85	Won	11/1/2016	3/3/2017	1026
HAXMC4IX	30	4		Engaging	11/3/2016		
NL3JZH1Z	1	3	4	Won	11/4/2016	3/10/2017	53
KWVA7VR1	8	2	28	Lost	11/4/2016	3/18/2017	0
S8DX3XOU	30	5	70	Won	11/4/2016	3/10/2017	5169
ENB2XD8G	34	5	65	Won	11/4/2016	3/6/2017	4631
09YE9QOV	26	4	24	Won	11/5/2016	3/11/2017	3393
3F5MZNEH	25	3	65	Lost	11/5/2016	3/3/2017	0
M6WEJXC0	25	4	61	Won	11/5/2016	3/6/2017	3284
6PTR7VBR	3	3	73	Won	11/6/2016	3/5/2017	61
902REDPA	18	2	78	Lost	11/7/2016	3/9/2017	0
5J9CMGDV	28	3	58	Won	11/7/2016	3/8/2017	46
JJXRR8R6	30	5	21	Lost	11/7/2016	3/17/2017	0



Create mapping dictionaries



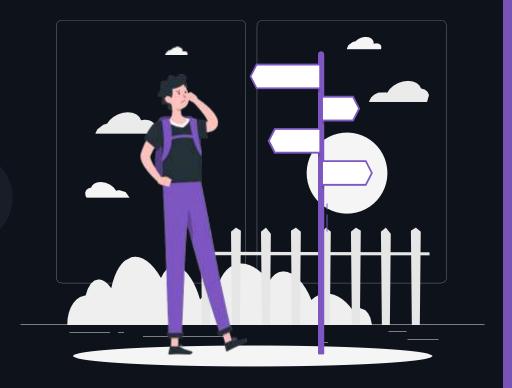
Create mapping dictionaries



```
dict_name = ['accounts_dict', 'products_dict', 'teams_dict']
for i in range(len(dict_name)):
  # Assuming modified_path and dataframe are defined elsewhere
  full_path = modified_path + dataframes[i]
  df = pd.read_csv(full_path)
  # Create the dictionary from the DataFrame
  temp_dict = df.iloc[:, 1].to_dict()
  # Increment keys by 1 and convert to string keys
  temp_dict = {key + 1: value for key, value in temp_dict.items()}
  temp_dict = {str(key): value for key, value in temp_dict.items()}
  # Swap keys and values
  temp_dict = {value: key for key, value in temp_dict.items()}
  # Store the dictionary in the global namespace using the name from dict_name
  globals()[dict_name[i]] = temp_dict
print(products_dict)
```



Validate data consistency



Validate data consistency

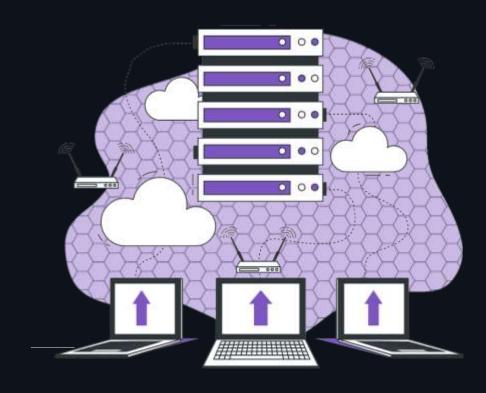
print(f'No missing values in the dictionary.')



```
sales_pipeline_path = f'{path}sales_pipeline.csv'
df_sales = pd.read_csv(sales_pipeline_path)
data = {'product': products_dict, 'account': accounts_dict, 'sales_agent': teams_dict}
for column, dict in data.items():
  print(f'Sales_pipeline column - {column}:')
  # Check for non-matching values in the sales pipeline column against the dictionary keys
  x = df_sales[~df_sales[column].isin(dict.keys())][column].unique()
  if len(x) != 0:
     # If there are non-matching values, print them
     print(f'Non matching values in sales_pipeline column: {x}.')
     # Check for any keys in the dictionary that are not present in the DataFrame's unique values
     for keys in dict.keys():
       y = df_sales[column].unique() # Get unique values from the DataFrame column
       if keys not in y:
          # If a key from the dictionary is not found in the DataFrame's unique values, print it
          print(f'Non matching values in the dictionary: {keys}.')
  else:
     # If there are no non-matching values, print a message
```



DWH Design & Data Modelling



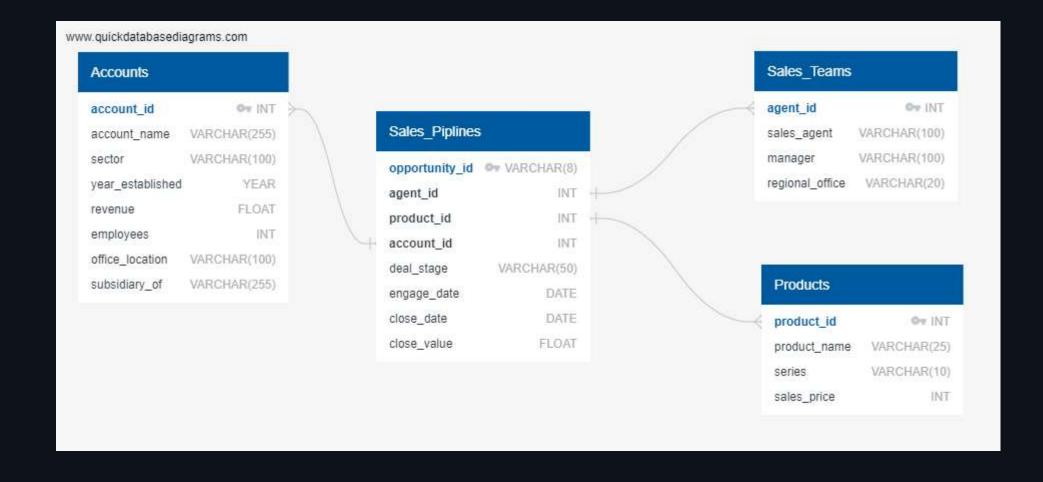


The primary goal of a data warehouse is to provide a reliable, centralized repository of integrated data that can be easily accessed and analyzed to support strategic decision-making and business intelligence activities within an organization.

By understanding the business first and then converting normalized tables in a database into denormalized tables in a data warehouse through dimension tables and a fact table. To meet the business needs of analysis for a later period



Identify dimension tables and fact table:





Example of creating tables:

```
CREATE TABLE [sales] (
    [opportunity_id] VARCHAR(8) NOT NULL,
    [agent_id] INT ,
    [product_id] INT ,
    [account_id] INT ,
    [deal_stage] VARCHAR(50) ,
    [engage_date] DATE ,
    [close_date] DATE ,
    [close value] FLOAT ,
    CONSTRAINT [PK_sales] PRIMARY KEY CLUSTERED (
        [opportunity_id] ASC
ALTER TABLE [sales] WITH CHECK ADD CONSTRAINT [FK_sales_agent_id] FOREIGN KEY([agent_id])
REFERENCES [sales_teams] ([agent_id])
ALTER TABLE [sales] CHECK CONSTRAINT [FK_sales_agent_id]
ALTER TABLE [sales] WITH CHECK ADD CONSTRAINT [FK_sales_product_id] FOREIGN KEY([product_id])
REFERENCES [products] ([product_id])
ALTER TABLE [sales] CHECK CONSTRAINT [FK_sales_product_id]
ALTER TABLE [sales] WITH CHECK ADD CONSTRAINT [FK_sales_account_id] FOREIGN KEY([account_id])
REFERENCES [accounts] ([account id])
ALTER TABLE [sales] CHECK CONSTRAINT [FK_sales_account_id]
```

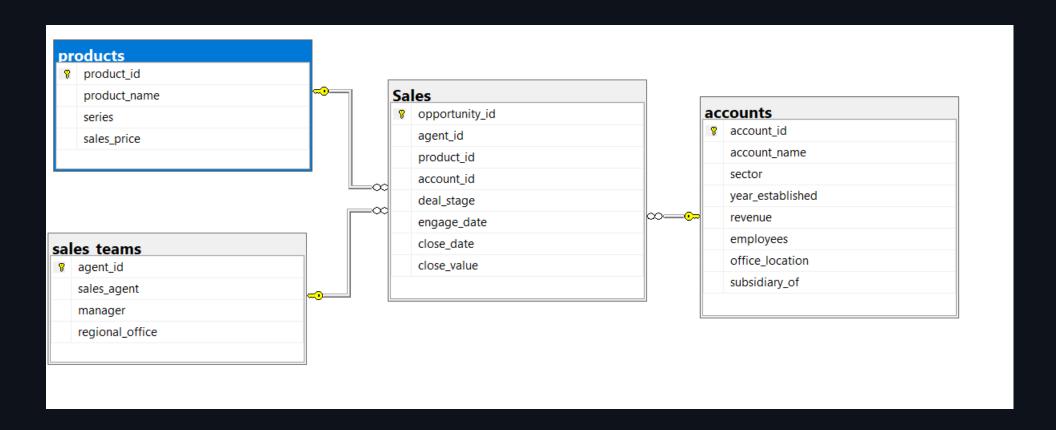


Another example of creating tables:

```
CREATE TABLE [products] (
    [product_id] INT NOT NULL ,
   [product name] VARCHAR(25) NOT NULL,
    [series] VARCHAR(10) NOT NULL,
   [sales_price] INT NOT NULL,
   CONSTRAINT [PK_products] PRIMARY KEY CLUSTERED (
       [product_id] ASC
CREATE TABLE [accounts] (
   [account_id] INT NOT NULL,
    [account_name] VARCHAR(255) NOT NULL ,
    [sector] VARCHAR(100) NOT NULL,
    [year_established] INT ,
    [revenue] FLOAT ,
    [employees] INT ,
    [office_location] VARCHAR(100) ,
   [subsidiary_of] VARCHAR(255) ,
   CONSTRAINT [PK_accounts] PRIMARY KEY CLUSTERED (
       [account_id] ASC
```



Data warehouse Diagram (Star Schema):





loading data into the tables

```
-- loading data into the sales teams table - file: sales teams.csv
BULK INSERT [dbo].[sales_teams]
FROM 'C:\Users\asmaa\Desktop\CRM Sales Opportunities project\1 - Data Preparation\modified_files\sales_teams.csv'
WITH (
    FIELDTERMINATOR = ',',
    ROWTERMINATOR = '\n',
    FIRSTROW = 2
);
SELECT TOP (5) * FROM [dbo].[sales_teams];
-- loading data into the products table - file: products.csv
BULK INSERT[dbo].[products]
FROM 'C:\Users\asmaa\Desktop\CRM Sales Opportunities project\1 - Data Preparation\modified_files\products.csv'
WITH (
    FIELDTERMINATOR = ',',
    ROWTERMINATOR = '\n',
    FIRSTROW = 2
SELECT TOP(5) * FROM [dbo].[products];
```



Analyzing data

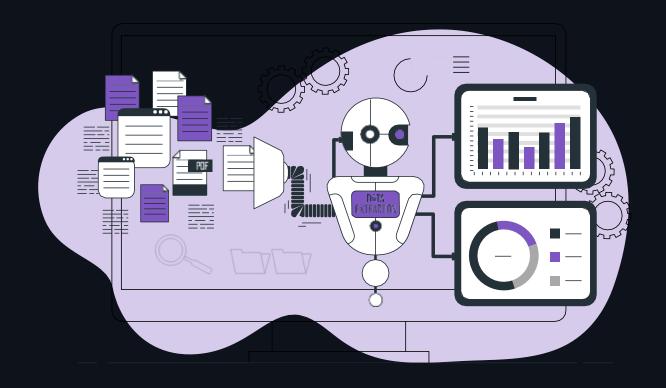


Analyzing data



We can analyze data in multiple ways, such as:

- 1. SQL server
- 2. visualization using Power BI



SQL Server



Example of product success rate:

```
-- 1) Which products have the highest success rates in closing deals?
SELECT p.product name,
        ROUND((CAST(SUM(CASE WHEN s.deal stage = 'Won' THEN 1 ELSE 0 END) AS float) /
                CAST(SUM(CASE WHEN s.deal_stage IN ('Won','Lost') THEN 1 ELSE 0 END) AS float))*100, 2) AS success_rate_pct
FROM [dbo].[sales] AS s
JOIN [dbo].[products] p
    ON s.product_id = p.product_id
GROUP BY p.product_name
ORDER BY success_rate_pct DESC;
-- 2) Which products generate the most revenue, and how do they compare to other products?
WITH products_revenue AS (
    SELECT p.product_name,
            SUM(s.close_value) AS sales_revenue
    FROM [dbo].[sales] AS s
    JOIN [dbo].[products] p
        ON s.product_id = p.product_id
    WHERE s.deal_stage = 'Won'
    GROUP BY p.product name
SELECT product_name,
        sales revenue,
        ROUND((sales revenue/SUM(sales revenue) OVER ())*100, 2) AS revenue pct
FROM products_revenue
ORDER BY sales_revenue DESC
```

SQL Server



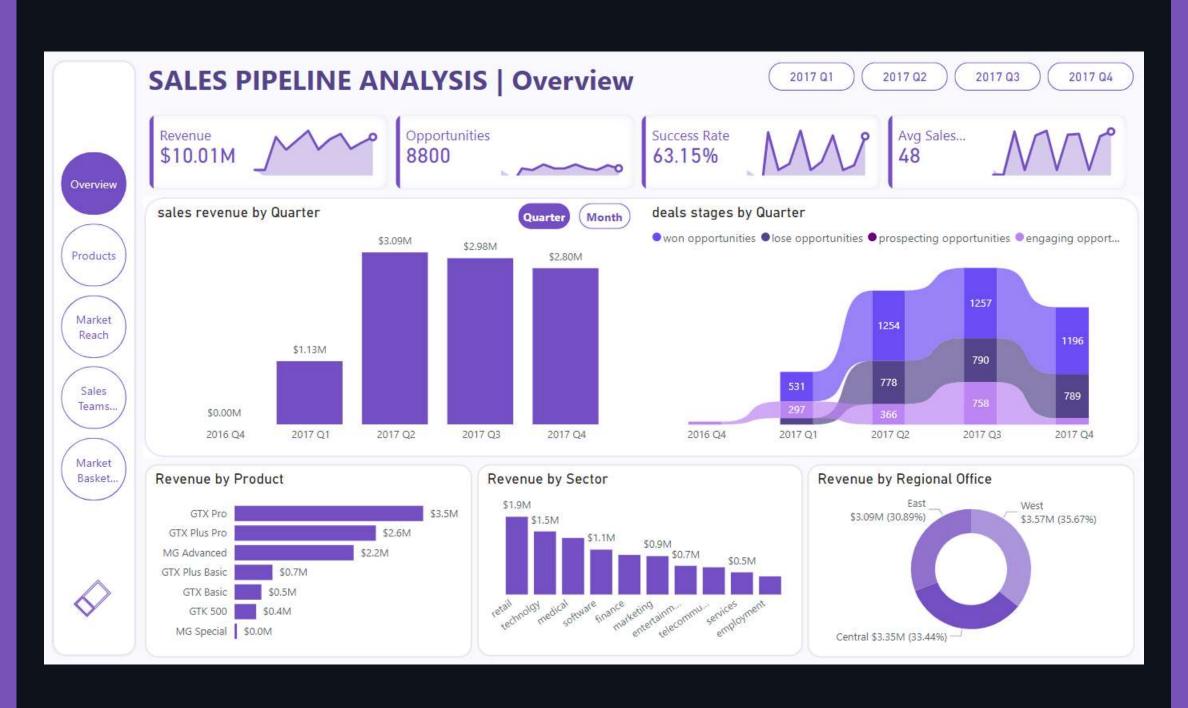
Example of sales cycle duration:

```
-- 1) What is the average sales cycle duration for won and lost opportunities?
SELECT deal_stage,
        ROUND(AVG(DATEDIFF(DAY, engage_date, close_date)), 2) AS avg_sales_cycle_days,
        MAX(DATEDIFF(DAY, engage_date, close_date)) AS max_sales_cycle_days,
        MIN(DATEDIFF(DAY, engage date, close date)) AS min sales cycle days
FROM [dbo].[sales]
WHERE deal_stage IN ('Won', 'Lost')
GROUP BY deal stage;
-- 2) How does the sales cycle duration vary by product or sector?
-- Sales cycle duration by product
SELECT p.product name,
        ROUND(AVG(CASE WHEN s.deal stage = 'Won' THEN DATEDIFF(DAY, s.engage date, s.close date) END), 2) AS avg won sales cycle days,
        ROUND(AVG(CASE WHEN s.deal stage = 'Lost' THEN DATEDIFF(DAY, s.engage date, s.close date) END), 2) AS avg lost sales cycle days
FROM [dbo].[sales] AS s
JOIN [dbo].[products] AS p
    ON s.product id = p.product id
GROUP BY p.product name
ORDER BY avg_won_sales_cycle_days ASC, avg_lost_sales_cycle_days ASC;
-- Sales cycle duration by sector
SELECT a.sector.
        ROUND(AVG(CASE WHEN s.deal_stage = 'Won' THEN DATEDIFF(DAY, s.engage_date, s.close_date) END), 2) AS avg_won_sales_cycle_days,
        ROUND(AVG(CASE WHEN s.deal_stage = 'Lost' THEN DATEDIFF(DAY, s.engage_date, s.close_date) END), 2) AS avg_lost_sales_cycle_days
FROM [dbo].[sales] AS s
JOIN [dbo].[accounts] AS a
    ON s.account_id = a.account_id
GROUP BY a.sector
ORDER BY avg_won_sales_cycle_days ASC, avg_lost_sales_cycle_days ASC;
```



Visualization using Power Bl





SALES PIPELINE ANALYSIS | Products

Top Selling Product

GTX Pro \$3,510,578.00 Top Winning Product

GTX Basic 915

Highest Success Rate

Plus

Basic

Basic

GTK

MG Special 64.84%

Fast Sales Cycle

2017 Q2

GTX Pro 46

2017 Q1

series GTK GTX

MG

2017 Q4

2017 Q3

Products

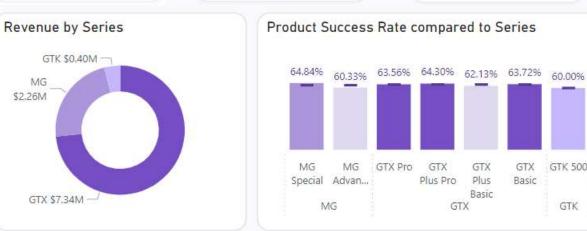
Overview

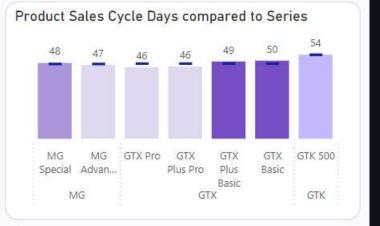
Market Reach

Sales Teams...

Market Basket...









product_name	price	avg revenue	difference po	t revenue	revenue pct
GTK 500	\$26,768	26,707.47	-0.239	\$400,612.00	4.00%
GTX Basic	\$550	545.64	-0.799	\$499,263.00	4.99%
GTX Plus Basic	\$1,096	1,080.05	-1.459	6 3705,275.00	7.05%
GTX Plus Pro	\$5,482	5,489.88	0.149	% \$2,629,651.00	26.28%
GTX Pro	\$4,821	4,815.61	-0.119	83,510,578.00	35.09%
MG Advanced	\$3,393	3,388.97	-0.129	\$2,216,387.00	22,15%
MG Special	\$55	55.19	0.359	§ \$43,768.00	0.44%

SALES PIPELINE ANALYSIS | Market Reach

2017 Q1) (2017 Q2) (2017 Q3) (2017 Q4

Overview

Products

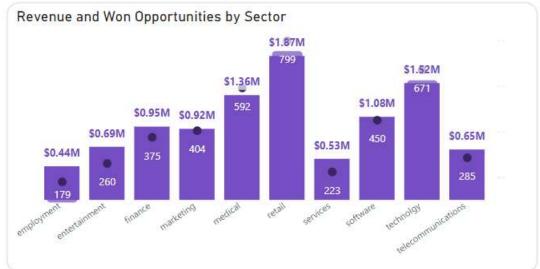
Market Reach

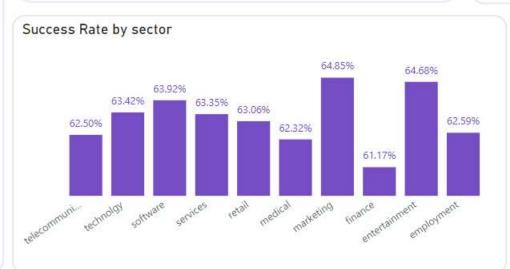
Sales Teams...

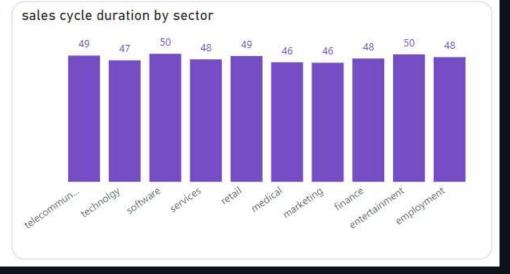
Market Basket...











SALES PIPELINE ANALYSIS | Sales Teams 2017 Q1 2017 Q2 2017 Q3 agents without deals Top 3 Managers won deals contribution revenue agent contribution revenue manager manager sales revenue Carl Lin Melvin Marxen 882 22.51% \$2,251,930,00 Carol Thompson \$2,251,930.00 □ Summer Sewald 19.64% \$1,964,750.00 Overview Summer Sewald Elizabeth Anderson Rocco Neubert 19.59% \$1,960,545.00 691 \$1,964,750.00 ⊞ Celia Rouche \$1,603,897.00 16.03% Mei-Mei Johns 610 Rocco Neubert 11.29% □ Cara Losch \$1,130,049.00 480 Natalya Ivanova Products \$1,960,545.00 Dustin Brinkmann \$1,094,363.00 747 10.94% \$10,005,534.00 Total 4238 100.00% regional office sales agent Market manager Reach Success Rate vs. Average Sales ● Won Opp. Lost Opp. Prospecting Opp. Engaging Opp. Teams... 63.02% 62.56% Central \$3,35M (33,44%) \$3.57M (35.67%) Market Basket... 0.5K East Central 1.0K Sales Cycle Days vs. Average 0.8K 0.7K 1.6K 1.4K 1.2K East \$3.09M (30.89%)

Central

East

West

Central

East

2017 Q4

100.00%

100.00%

100.00%

100.00%

100.00%

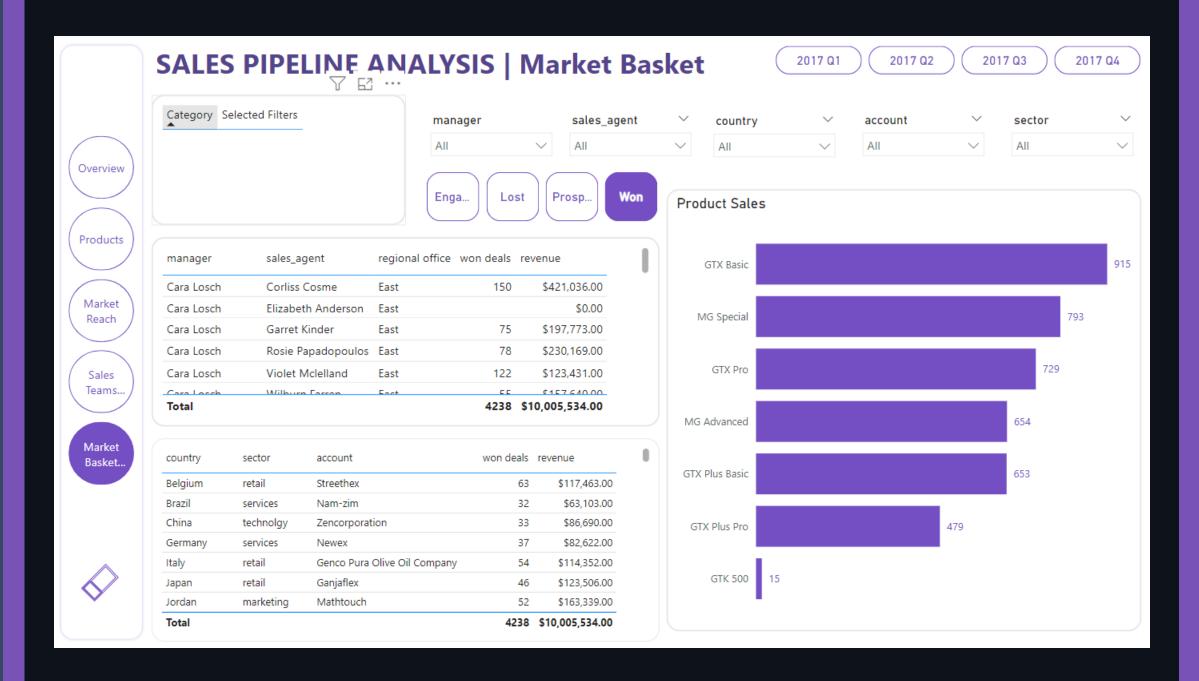
100.00%

100.00%

63.94%

West

West







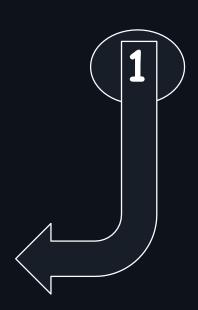




> Products:

- The top-selling product is the GTX Pro, generating \$3.51M in revenue.
- The GTX Plus Pro and MG Advanced products are also significant revenue contributors.
- The product success rates vary, with the MG Special having the highest at 64.84%.

- Analyze the sales performance of each product in more depth, looking at factors such as market share, growth rates, profit margins, and customer satisfaction.
- Identify the drivers behind the high success rate of the MG Special and explore opportunities to apply those best practices to other product lines.
- Consider conducting a product portfolio analysis to identify opportunities for product line expansion, optimization, or rationalization.

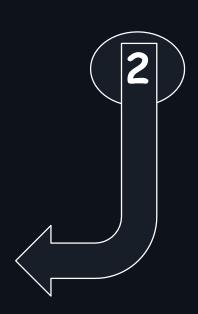




Accounts:

- The company has a presence in 15 countries, with the top 3 being the United States, Korea, and Jordan.
- The top 3 sectors are technology, retail, and medical, representing the largest customer segments.
- The top 3 accounts are Konex, Kan-code, and Condax, suggesting these are the company's largest or most strategic customers.

- Analyze the growth potential and profitability of each country and sector to identify priority markets for expansion or increased investment.
- Investigate the factors contributing to the success of the top accounts, such as customer satisfaction, loyalty, and cross-selling opportunities.
- Consider developing targeted marketing and sales strategies for the highpotential countries, sectors, and accounts.

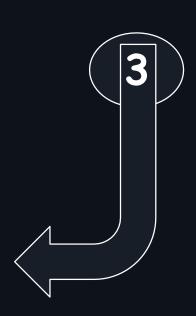




Sales Teams:

- The top-performing managers are Melvin Marxen, Summer Sewald, and Rocco Neubert, based on their sales revenue and contribution.
- The success rate and sales cycle duration vary across different regional offices, indicating potential areas for improvement or knowledge sharing.

- Conduct interviews or focus groups with the top-performing managers to understand their best practices and processes, and explore opportunities to share this knowledge with other sales teams.
- Analyze the factors contributing to the performance differences across regional offices, such as team composition, training, incentives, or customer dynamics.
- Implement a sales team performance management program to set clear goals, provide training and coaching, and foster a culture of continuous improvement.

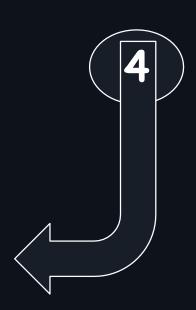




Market basket analysis for accounts:

- Analyze the most common product combinations purchased by each account.
- Determine which products are most frequently purchased together.

- Offer bundled discounts or package deals for the most commonly purchased product combinations.
- Provide targeted cross-sell and upsell recommendations to sales teams based on the account-specific product affinities.
- Incentivize sales teams to focus on promoting the high-volume product bundles through spiffs, quota adjustments, or other compensation structures.

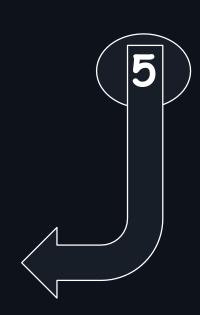




Market basket analysis for Sales Team:

- Analyze the sales performance and product knowledge of individual sales reps or teams.
- Identify the sales teams or individuals that excel at selling specific product combinations or bundles.
- Understand the factors that contribute to their success, such as training, account relationships, or product expertise.

- Assign sales teams or individuals as "product specialists" for the highvolume product combinations.
- Provide specialized training and enablement resources to help these specialists deepen their product knowledge and sales skills.
- Leverage the product specialists to create and deliver targeted sales pitches, demos, and proposals to accounts with a history of purchasing those product bundles.



Conclusion



- By integrating these market basket and sales team specialization insights, the organization can:
- Offer more compelling and personalized product packages to customers.
- Increase average order values and customer loyalty through strategic discounts and cross-selling.
- Empower the sales force to become true product experts and trusted advisors to their accounts.
- Optimize sales team performance and focus on the highest-yield product opportunities.
- ➤ Overall, the report provides a comprehensive view of the company's sales pipeline performance, highlighting the strengths and potential areas for optimization across products, sectors, regions, and sales teams. This information can be used to make informed decisions and adjust strategies to drive continued growth and success.





Thank you

