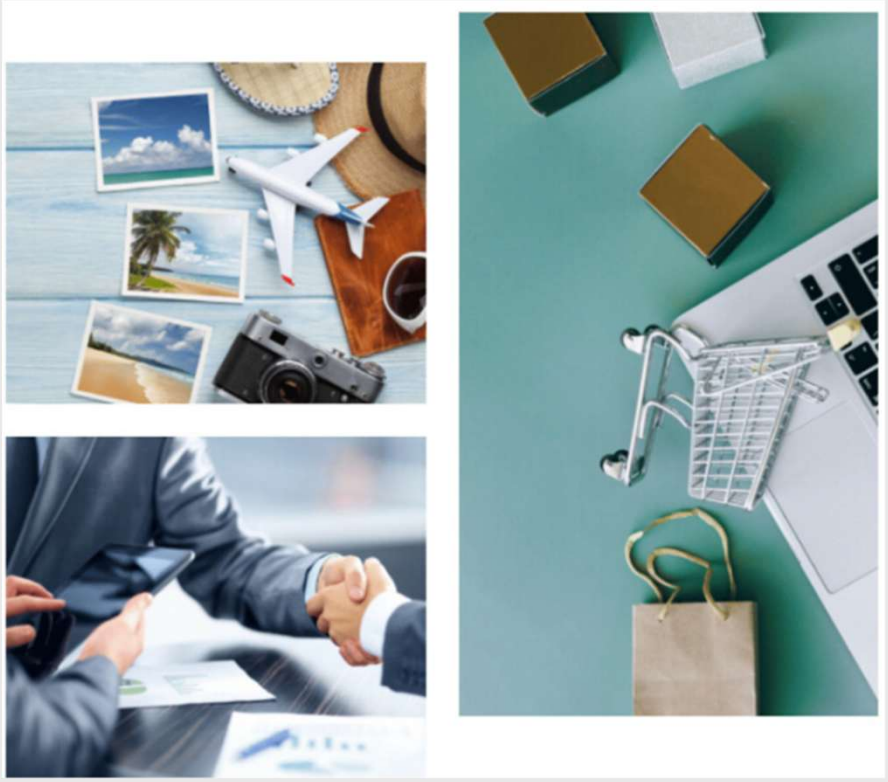


Project Title

“Proposed Datasets & Business Problem Statements for Final Data Analytics Project”

Name:
Abhishek Kumar Pandey

Date: 19/11/2025

Chandra Mouli Kotta Kota

Chief Data Scientist

Areas of Expertise:

- Big Data Analytics
- Business Analytics
- Machine Learning
- Deep Learning
- Marketing Analytics
- Risk Analytics
- Operation Analytics
- Digital Analytics
- Business Intelligence
- Artificial Intelligence

Chandra Mouli Kotta Kota is a former Business Consultant/Data Scientist and has worked with prestigious companies like McKinsey, Citigroup(E-serve), Genpact in the past 19+ years. He has worked for clients across the globe and is an expert in Business and Big Data Analytics.

Professional Experience

- Worked on Marketing Analytics(CSI, CLM & Pricing), Risk Analytics(Credit Risk), Operation Analytics, and Digital Analytics with a focus on Retail/E-Commerce, Banking, Insurance, Telecom, and Media clients in Asia, Australia, Europe, and the US

- Hands on experience in development of Marketing & CRM Models (Acquisition, LTV Models, Cross Sell & Upsell, Attrition and MROI, Price and Promotion Models), Pricing Models(price & promo) and Risk Models (PD, LGD, EAD Models for Credit Cards, Consumer Loans and Insurance Portfolios)

- Hands-on expertise in Big data and Multivariate analytical techniques, including classical & machine learning algorithms, including regression, instance-based, regularization, Decision tree, Bayesian, clustering, Association rules, ANN, Deep learning, Generative AI, and Ensemble algorithms

- Have used different statistical flat forms like SAS/SAS EM, R, Python, SPSS, SPSS Modeler, Hadoop, Spark, Tableau, Matlab, Julia, Salesforce, SQL, Excel, VBA, PowerBI, Cloud Platforms (AWS, GCP), MongoDB, Hbase.

- Trained and coached several client teams and various individuals on advanced analytics, Big data analytics tools and techniques as part of capability building programs
- Chandra Name is Featured as
  - Top 10 data scientists (2015) in India by Analytics India Magazine (https://www.analyticsindiamag.com/top-10-data-scientists-in-india-2015/)
  - Featured as expert in Data Analytics & AI by whoknowsabout. (https://whoknowsabout.com/blogpage/hidden-gems-boutiques-and-independent-consultants.html)

Academic Credentials

- Master of Science(Mathematics/Statistics): IIT-Madras, Chennai

AGENDA

Project Requirements	• Why the mentor asked for dataset selection • What must be included in the PPT • How final dataset will be selected
Overview of All Four Selected Datasets	Names, domains, and sources Summary of what each dataset represents
Dataset 1: Airlines, Airports, Cancellations & Flights	Introduction, description, business problems, KPIs Why this dataset is worthy
Dataset 2: CRM Sales Opportunities	Detailed breakdown of the CRM pipeline Sales forecasting and performance analysis
Dataset 3: Maven Fuzzy Factory — E-commerce Analytics	Digital marketing & website session analysis Conversion funnel insights
Dataset 4: Northwind Traders — Retail & Order Management	Product, order, shipment, customer insights Retail performance monitoring
Dataset Comparison	Complexity, insight potential, business relevance
Recommendation	Which dataset fits best for the final project
Next Steps	• After selection • Workflow for the final project

PROJECT REQUIREMENT

- Present **four well-structured datasets**

- For each dataset provide:
  - Introduction
  - Dataset overview
  - **Data dictionary (expanded)**
  - **Business problem (detailed)**
  - **KPIs + formulas**
  - Analytical scope

- **One dataset** will be chosen for final analysis
- Final project will include: SQL + Cleaning + Power BI dashboard + Insights

DATASET 1: AIRLINES + AIRPORTS + FLIGHTS

Dataset 1: Introduction

Airlines, Airports, Cancellation Codes & Flight Performance Dataset

This dataset is a **realistic aviation operations dataset** combining multiple tables related to:

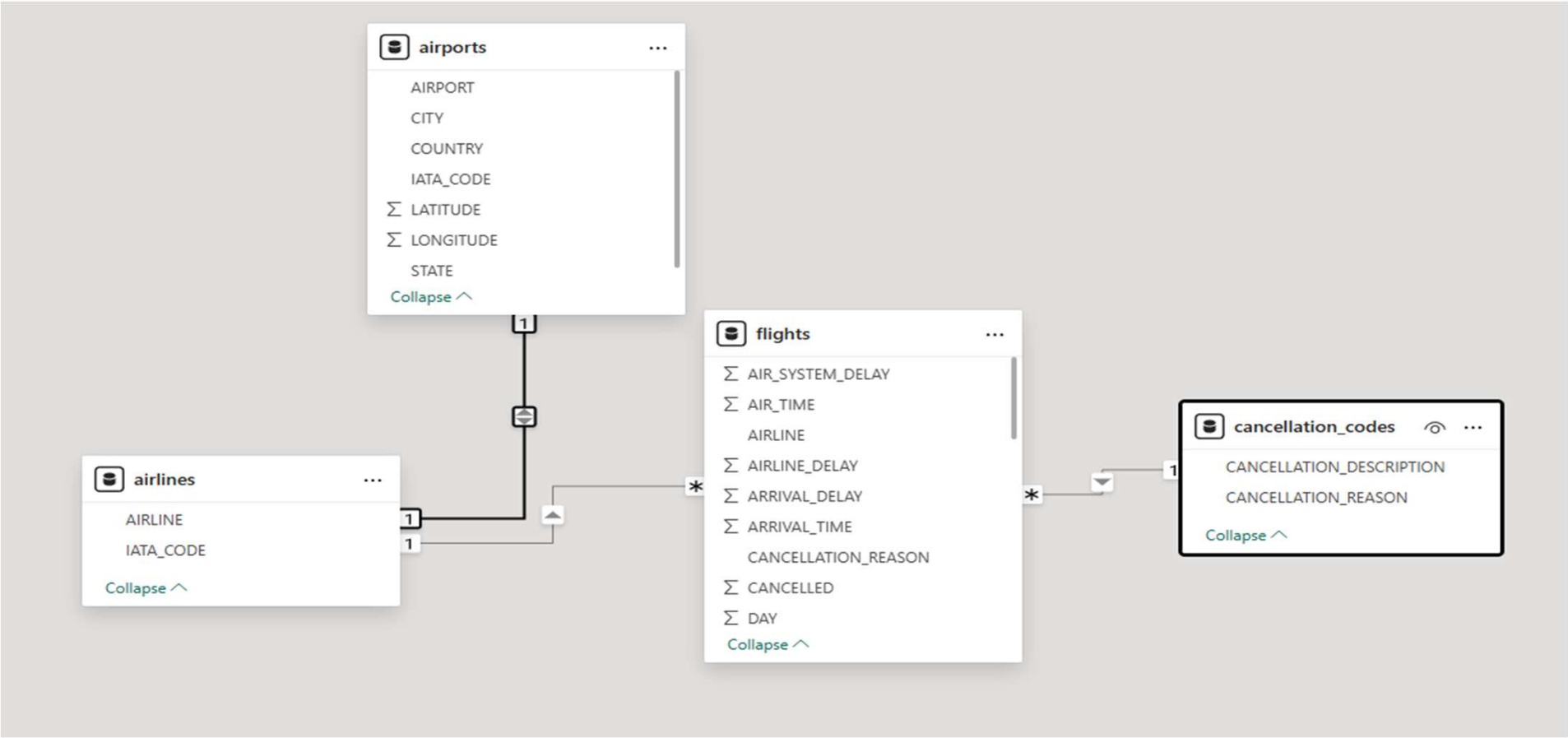
- Airline companies
- Airport infrastructure
- Daily flight schedules
- Flight delays
- Cancellation reasons
- Geographic route mapping

It provides a **complete end-to-end view** of the airline industry, allowing analysis of operational efficiency, logistics challenges, travel patterns, and performance metrics.

Why this dataset is powerful:

- High in volume, time-series rich
- Multi-dimensional (airline, route, time, distance, cancellations)
- Extremely relevant due to frequent flight delays & disruptions in India
- Ideal for operational analytics, forecasting, and reliability scoring

Dataset 1: ER- Diagram



DATASET 1: DATA DICTIONARY

	Column	Meaning
airlines.csv	Airline	Airline name (e.g. “American Airlines”, “Delta Air Lines”).
	IATA_Code	Airline code (e.g. “AA”, “DL”).

DATASET 1: DATA DICTIONARY

airports.csv	Column	Meaning
	Airport	Full airport name.
	STATE	State/region of the airport (for US, two-letter state code).
	City	City of airport
	Country	Country
	Latitude	Geo coordinate
	Longitude	Geo coordinate
	IATA	Airport code

DATASET 1: DATA DICTIONARY

	Code	Meaning
cancellation_codes.csv	A	Airline/Carrier Issue
	B	Weather-related
	C	National Airspace System (NAS)
	D	Security Issues

DATASET 1: DATA DICTIONARY

	Column	Meaning
flights.csv	YEAR, MONTH, DAY, DAY_OF_WEEK	Time period of the flight.
	AIRLINE	Airline code (links to airlines.IATA_CODE).
	FLIGHT_NUMBER	Flight number.
	TAIL_NUMBER	Aircraft tail number.
	ORIGIN_AIRPORT	Origin airport code (links to airports.IATA_CODE).
	DESTINATION_AIRPORT	Destination airport code.
	SCHEDULED_DEPARTURE	Scheduled departure time (HHMM, stored as text).
	DEPARTURE_TIME	Actual departure time (HHMM).
	Departure_Delay	Delay in minutes
	TAXI_OUT	Time when wheels left the ground (HHMM).
	WHEELS_OFF	Miles/km
	SCHEDULED_TIME	Scheduled total flight time in minutes.

DATASET 1: DATA DICTIONARY

flights.csv	ELAPSED_TIME	Meaning
	ELAPSED_TIME	Actual total flight time in minutes
	AIR_TIME	Time in the air in minutes.
	DISTANCE	Distance between origin and destination (miles).
	WHEELS_ON	Time when wheels touched down (HHMM).
	TAXI_IN	Taxi-in time in minutes.
	SCHEDULED_ARRIVAL	Scheduled arrival time (HHMM).
	ARRIVAL_TIME	Actual arrival time (HHMM).
	ARRIVAL_DELAY	Arrival delay in minutes.
	DIVERTED	1 if flight diverted, 0 otherwise.
	CANCELLED	1 if flight cancelled, 0 otherwise.
	CANCELLATION_REASON	Code for cancellation reason (joins to cancellation_codes).
	AIR_SYSTEM_DELAY	Minutes of delay caused by air system.

DATASET 1: DATA DICTIONARY

	ELAPSED_TIME	Meaning
flights.csv	SECURITY_DELAY	Minutes of delay caused by security issues.
	AIRLINE_DELAY	Minutes of delay caused by airline operations.
	LATE_AIRCRAFT_DELAY	Minutes of delay caused by late arrival of previous aircraft.
	WEATHER_DELAY	Minutes of delay caused by weather.

Dataset 1: Business Problem Statement

“How can airlines reduce delays, improve route reliability, and enhance customer experience by optimizing operations across airports, flights, and scheduling?”

Airlines today face several operational challenges:

- Increasing delays at major airports
- Weather-related disruptions
- Overcrowded routes
- Poor airline performance visibility
- High cancellation rates leading to loss of customer trust
- Inefficient scheduling across peak hours

This dataset helps answer:

- What causes the highest delays?
- Which airlines & airports are most problematic?
- What patterns occur across seasons/months?
- How cancellations affect business and customer experience?

This makes the dataset ideal for **operational analytics, forecasting, decision-making & root-cause analysis**.

Dataset 1: Key Insights & KPIs

KPIs	• On-Time Arrival Percentage • Average Delay per Airline • Total Delayed Flights per Month • Cancellation Rate by Reason • Route-Wise Delay Distribution • Distance vs Delay Correlation • Average Turn-Around Time
Insights Possible	• Identify worst-performing airlines and airports • Patterns in delays (time-of-day, season, weather) • Root cause for cancellations • Which routes need better scheduling • Whether long-distance flights face more delays • Which airports need capacity improvement

Why This Dataset Should Be Selected

- Highly realistic and applicable to real-world scenarios
- Rich in time-series and categorical data
- Ideal for forecasting, clustering, and operational KPIs
- Best dataset for building a performance dashboard
- Very strong for portfolio: Operations Analyst / Data Analyst roles

DATASET 2: CRM SALES OPPORTUNITIES

Dataset 2: Introduction

CRM Sales Pipeline / Opportunities Dataset

This dataset represents a **complete sales funnel**, tracking how leads travel across different stages of the pipeline:

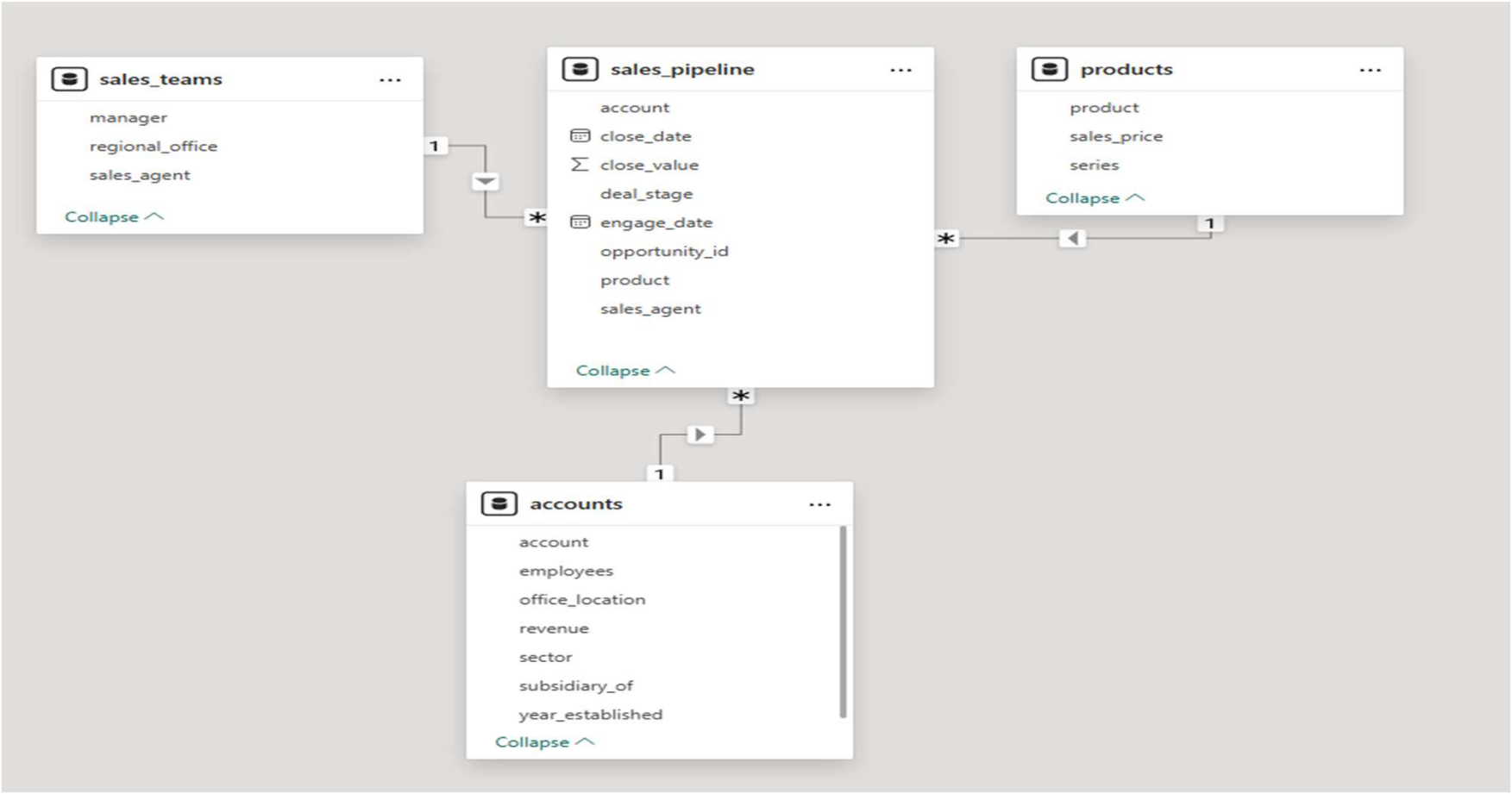
- Prospect
- Qualified Lead
- Proposal
- Negotiation
- Closed Won / Lost

It simulates the lifecycle of a lead in a B2B or B2C environment, showing:

- Revenue potential
- Stage progression
- Sales rep performance
- Deal probability

This dataset is excellent for **sales analytics, forecasting, and performance optimization**.

Dataset 2: ER- Diagram



DATASET 2: DATA DICTIONARY

	Column	Description	Example
sales_pipeline.csv This is Fact Tables **Purpose: Sales pipeline transactions**	opportunity_id	Unique deal/opportunity identifier	"1C17A6R"
	sales_agent	Name of salesperson handling the deal	"Moses Frase"
	product	Product involved in deal	"GTX Plus Basic"
	account	Customer/company name	"Cancity"
	deal_stage	Current/Final stage of deal	"Won", "Lost", "Engaging"
	engage_date	Date deal engagement started	"20-10-2016"
	close_date	Actual close date of deal	"01-03-2017"
	close_value	Revenue collected for the deal	1054

DATASET 2: DATA DICTIONARY

	Column	Description	Example
ACCOUNTS.csv	account	Name of the customer/company	"Acme Corporation"
	sector	Business sector/industry	"technology", "medical"
	year_established	Year when company was founded	1996
	revenue	Annual revenue (in millions of USD)	1100.04
	employees	Number of employees	2822
	office_location	Parent company name (nullable)	"United States"
	subsidiary_of	Sales territory	Null or "Some Group"

DATASET 2: DATA DICTIONARY

	Column	Description	Example
sales_teams.csv	sales_agent	Name of salesperson	"Anna Snelling"
	manager	Name of manager for the salesperson	"Dustin Brinkmann"
	regional_office	Region of operation	"Central"

DATASET 2: DATA DICTIONARY

products.csv	Column	Description	Example
	product	Product name	"GTX Pro"
	Series	Product category	“GTK", “GTX", “MG"
	Sales_Price	Standard price for the product	2045.00

Dataset 2: Business Problem Statement

“How can the company improve its sales conversion rates, optimize team performance, and accurately forecast revenue based on pipeline characteristics?”

This dataset helps solve key sales challenges:

- Low conversion from stages like Proposal → Closed
- High churn at negotiation stages
- Poor revenue forecasting accuracy
- Underperforming sales reps
- Unknown high-value industries or regions

Insights deeply support **strategic sales decisions**.

Dataset 2: Key Insights & KPIs

KPIs	**Stage Conversion Rate (%)**	**Definition: % of opportunities moving from one stage to the next.** **Formula:** $Conversion\ Rate = \frac{Opportunities\ in\ Next\ Stage}{Opportunities\ in\ Current\ Stage} \times 100$ **Example:** Qualification → Proposal: Qualification: 120 deals Proposal: 78 deals $= \frac{78}{120} \times 100 = 65\%$
	Win Rate / Win Ratio (%)	**Definition:** What percentage of all opportunities get closed successfully? **Formula:** $Win\ Rate = \frac{Closed\ Won}{Total\ Closed} \times 100$ **Example:** Closed Won = 40 Closed Lost = 60 $= \frac{40}{100} \times 100 = 40\%$
	Expected Revenue	**Definition:** Forecasted earnings. **Formula:** $Expected\ Revenue = Deal\ Size \times Deal\ Probability$ **Example:** Deal Size = ₹ 2,00,000 Probability = 60% (=0.6) $Expected\ Revenue = 2{,}00{,}000 \times 0.6 = ₹1{,}20{,}000$

Dataset 2: Key Insights & KPIs

KPIs	**Average Deal Size**	**Formula:** $$Average\ Deal\ Size = \frac{Sum\ of\ Deal\ Size}{Number\ of\ Deals}$$
	Pipeline Value	**Definition:** Total potential business in pipeline. **Formula:** $$Pipeline\ Value = \sum D\ eal\ Size$$
	Forecast Accuracy (%)	**Formula:** $$Forecast\ Accuracy = \frac{Actual\ Revenue}{Expected\ Revenue} \times 100$$
	Sales Cycle Length (Days)	**Formula:** $$Sales\ Cycle = Close\ Date - Created\ Date$$
	Sales Rep Performance Score	**Formula:** $$Score = (Closed\ Won \times 2) + (In\ Progress \times 1) - (Closed\ Lost \times 1)$$
	Win–Loss Ratio	Win/Loss Ratio = Closed Lost / Closed Won
	Lead Source Effectiveness	Lead Source Effectiveness=Closed Won from Source×100 / Leads from Source

Dataset 2: Key Insights & KPIs

Insights

- Which stages lose most opportunities?
- Which industries give highest revenue?
- Which regions are most profitable?
- Which sales reps need training?
- What is the ideal lead source for revenue?

Why This Dataset Should Be Selected

- Very business-centric & industry-relevant
- Great dataset for Sales Analytics roles
- Clean structure → Easy to demonstrate SQL & BI skills
- High explanatory value in presentation
- Excellent for funnel dashboards & forecasting models

DATASET 3: MAVEN FUZZY FACTORY (E-COMMERCE ANALYTICS)

Dataset 3: Introduction

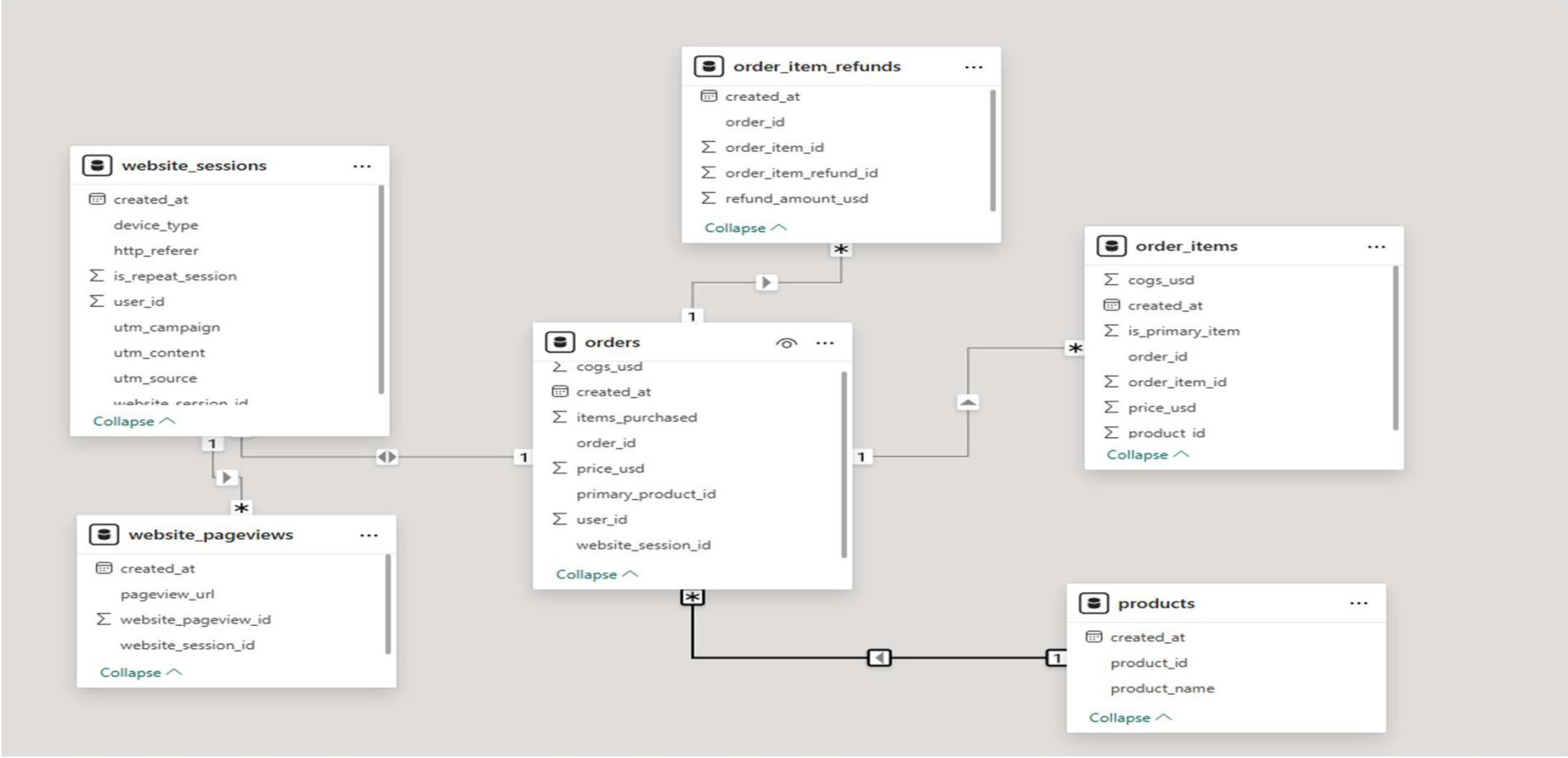
Website, Marketing, and E-commerce Transactions Dataset

This dataset simulates a real-world e-commerce brand's website activity, allowing analysis of:

- Website traffic
- User sessions
- Marketing channels
- Product performance
- Order behaviour

Excellent for corporate sales analytics.

Dataset 3: ER- Diagram



DATASET 3: DATA DICTIONARY

	Column	Description
website_sessions	website_session_id	Unique session identifier
	created_at	Timestamp of session
	user_id	Unique visitor ID
	is_repeat_session	1 = returning visitor
	utm_source	Marketing source (e.g., google, email)
	utm_campaign	Campaign name
	utm_content	Variation of ad
	device_type	Desktop, mobile, tablet
	browser	Browser used

DATASET 3: DATA DICTIONARY

	Column	Description
products	product_id	Unique product code
	product_name	Product name
	product_category	Category name
	created_at	Timestamp
	cost	Cost of product
	retail_price	Selling price
referrals	referral_id	PK
	created_at	Timestamp
	user_id	FK → users
	referrer_user_id	Who referred this user

DATASET 3: DATA DICTIONARY

	Column	Description
orders	order_id	Order ID
	created_at	Timestamp of session
	website_session_id	FK → website_sessions
	price_usd	Total revenue of order
	cogs_usd	Cost of goods sold
	website_pageview_id	FK → pageview that triggered order
website_pageviews	website_pageview_id	Primary key
	created_at	Timestamp of pageview
	website_session_id	FK → website_sessions
	pageview_url	URL visited (e.g., “/home”)

DATASET 3: DATA DICTIONARY

	Column	Description
order_items	order_item_id	PK
	order_id	FK → orders
	created_at	Timestamp
	product_id	FK → products
	is_primary_item	1 = main product
	price_usd	Price at item level
	cogs_usd	Cost at item level
inventory_items	inventory_item_id	PK
	product_id	FK → products
	created_at	Timestamp
	sold_at	Timestamp when sold
	cost	Cost of manufacturing

DATASET 3: DATA DICTIONARY

	Column	Description
users	user_id	Unique user ID
	created_at	Timestamp account created
	email	User email
	primary_product_id	First purchased product
	is_geo_valid	1 = valid IP → valid location
	state	US State of customer
	country	Country
sales	sale_id	PK
	created_at	Timestamp
	item_id	FK → inventory_items
	sale_price_usd	Price sold at

Dataset 3: Business Problem Statement

“How can the company increase its conversion rate, improve marketing ROI, and optimize user acquisition strategies?”

Problems addressed:

- High bounce rate
- Low conversion
- High customer acquisition cost
- Unknown best marketing channel
- Product performance issues

This dataset is ideal for studying user behavior and optimizing growth.

Dataset 3: KPIs

KPIs	**Conversion Rate**	$CR = \frac{Orders}{Total\ Sessions} \times 100$
	Revenue per Session	$RPS = \frac{Total\ Revenue}{Total\ Sessions}$
	Cost per Acquisition (CPA)	$CPA = \frac{Marketing\ Spend}{Orders}$
	ROI per Channel	$ROI = \frac{Revenue - Spend}{Spend} \times 100$
	Bounce Rate	$Bounce\ Rate = \frac{Single\ Page\ Sessions}{Total\ Sessions} \times 100$

Dataset 3: Insights

- Top channels for revenue generation
- Customer journey bottlenecks
- Which product categories sell best
- Campaign-level profitability
- Trends in digital traffic

Why Select This Dataset

- Perfect for e-commerce & digital marketing analytics
- Teaches modern business metrics used at Amazon, Flipkart, Meesho
- Excellent dashboard storytelling
- Highly relevant skill for interviews

DATASET 4: NORTHWIND TRADERS (RETAIL SALES & OPERATIONS)

Dataset 4: Introduction

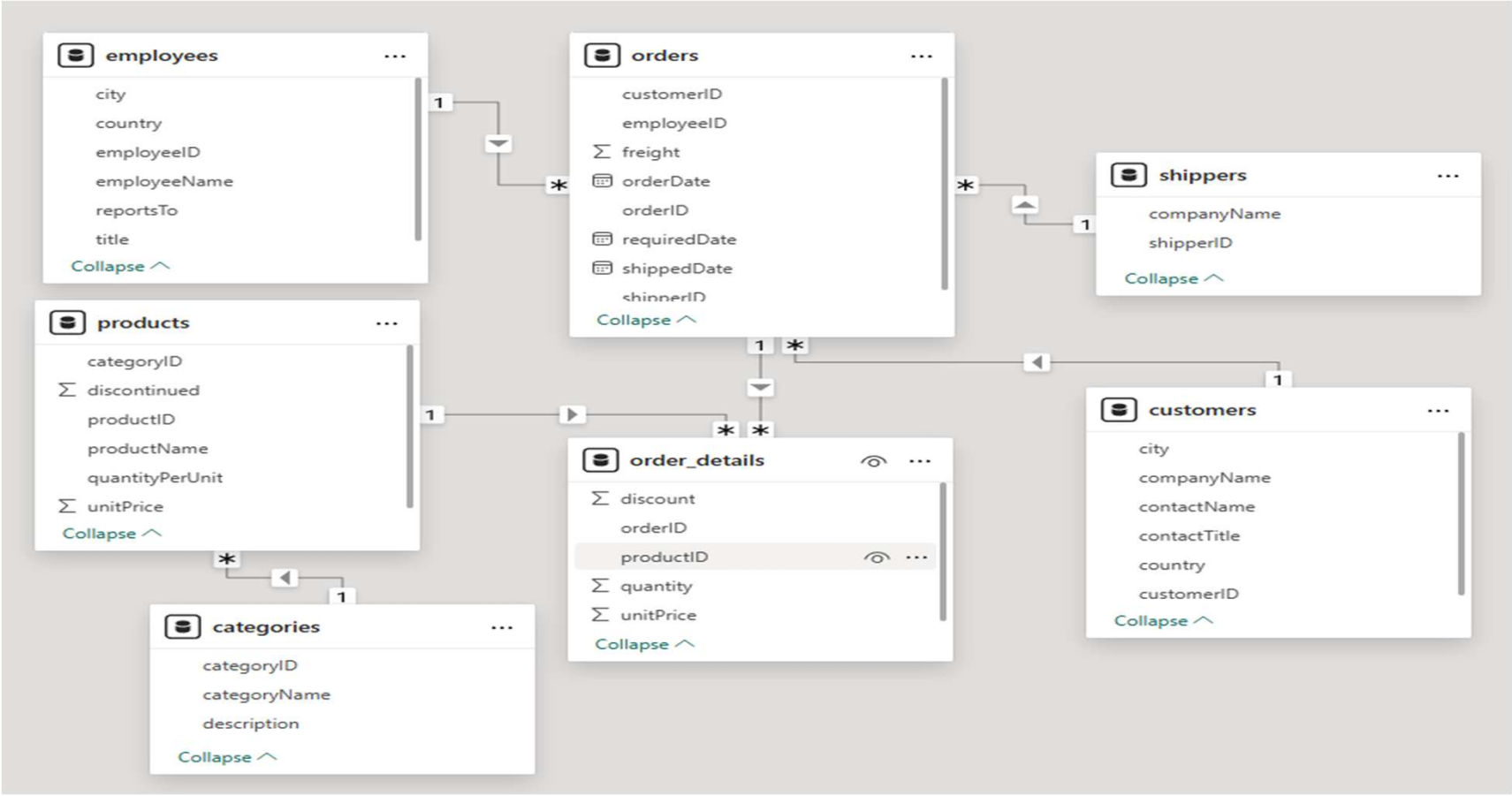
Northwind Traders — Retail Orders & Inventory Analytics

A classic retail operational dataset capturing:

- Customers
- Orders
- Order details
- Products
- Shipments
- Employees
- Inventory
- Suppliers

Widely used across the world to teach supply chain & retail analytics.

Dataset 4: ER- Diagram



DATASET 4: DATA DICTIONARY

	Column	Description
employees	EmployeeID	Unique identifier for each employee
	employeeName	Full name of the employee
	title	The employee's job title
	Title	Job title
	city	The city where the employee works
	country	The country where the employee works
	reportsTo	The ID of the employee's manager
categories	categoryID	Unique identifier for each product category
	categoryName	The name of the category
	description	A description of the category and its products

DATASET 4: DATA DICTIONARY

	Column	Description
orders	order_id	Unique identifier for each order
	customer_id	The customer who placed the order
	employee_id	Employee processing order
	order_date	Order creation date
	required_date	The date the customer requested the order to be delivered
	shipped_date	The date when the order was shipped
	shipperID	The ID of the shipping company used for the order
	freight	The shipping cost for the order (USD)
shippers	shipperID	Unique identifier for each shipper
	companyName	The name of the company that provides shipping services

DATASET 4: DATA DICTIONARY

	Column	Description
order_details	order_id	The ID of the order this detail belongs to
	product_id	The ID of the product being ordered
	unit_price	The price per unit of the product at the time of ordering (USD – discount not included)
	quantity	The number of units being ordered
	discount	The discount percentage applied to the price per unit
products	product_id	Unique ID
	product_name	Name
	quantity_per_unit	The quantity of the product per package
	unit_price	The current price per unit of the product (USD)
	discontinued	Indicates with a 1 if the product has been discontinued
	categoryID	The ID of the category the product belongs to

DATASET 4: DATA DICTIONARY

	Column	Description
customers	customerID	Unique identifier for each customer
	companyName	The name of the customer's company
	contactName	The name of the primary contact for the customer
	ContactTitle	The job title of the primary contact
	city	Customer city
	country	Country

Dataset 3: Business Problem Statement

“How can Northwind optimize inventory, improve order fulfillment, and increase revenue performance?”

Challenges solved:

- Stock shortages
- Delivery delays
- Low category profitability
- Poor customer segmentation

Dataset 4: KPIs

KPIs	Total Revenue	$Revenue = \sum (UnitPrice \times Quantity \times (1 - Discount))$
	Average Order Value	$AOV = \frac{Total\ Revenue}{Total\ Orders}$
	On-Time Delivery %	$OTD = \frac{Shipped\ On\ Time}{Total\ Orders} \times 100$
	Category Profitability	$Category\ Revenue = \sum Revenue\ (Category)$
	Stock Turnover Ratio	$Turnover = \frac{Cost\ of\ Goods\ Sold}{Average\ Inventory}$

Insights

- Seasonal demand trends
- Which suppliers deliver late?
- Which products generate maximum profit?
- Inventory levels vs demand patterns
- Region-wise and employee-wise performance

Why Select this Dataset

- Very clean and structured dataset
- Perfect for SQL, Power BI, forecasting, trends
- Similar to real retail systems used worldwide
- Strong analytics storytelling for operations & sales
- Excellent for dashboards & KPI tracking


USER ACCESS
PROFILE CUSTOM
CRM
CUSTOMER LOYALTY
COMMUNICATION
DOCUMENTATION
DATABASE & INFO
ORDER & CUSTOM
ANALYSIS & GRAPHICS
INTERACTION
PROMOTION & ADS
Customer Relationship Management

CRM SALES OPPORTUNITIES

Introduction

Business context & problem statement

Business context (from data itself)

The CRM data tracks **B2B accounts (companies)** in accounts
Products sold (products)
Sales opportunities / deals (sales_pipeline)
Sales team structure (sales_teams)
All deals flow through stages: **Prospecting** → **Engaging** → **Won / Lost** (deal_stage in sales_pipeline).
So, this is a **retail / SaaS-like CRM pipeline** dataset to analyze how products and sales teams convert opportunities into revenue.

Problem statement

Using the available CRM data, the analytics team needs to:
Understand **who we sell to** (accounts, sectors, locations).
Analyse **sales pipeline performance** (volume, win rate, revenue).
Evaluate **product performance** (which products/series drive revenue).
Evaluate **sales team performance** (agents, managers, regions).

Design **dashboards & KPIs** to be built in SQL + Power BI.

Technology stack & purpose

Excel	• Initial data inspection, manual checks, quick pivots.
SQL Server	• Import all CSVs as tables • Run data checks & cleaning • Create views / aggregated tables for dashboards.
Power BI Desktop	• Connect to SQL tables/views • Build dashboards (Sales, Pipeline, Product, Account, Sales Team).
Power BI Service	• Publish dashboards • Share with stakeholders, schedule refresh.
PowerPoint	• Present business context, data model, data issues, cleaning steps, KPIs, dashboard mockups, and learnings.

Tables overview

Table	Rows	Columns	Key fields (candidate PK)
accounts	85	7	account
products	7	3	product
sales_pipeline	8,800	8	opportunity_id
sales_teams	35	3	sales_agent

Data Dictionary

DATA DICTIONARY

	Column	Description	Distinct Rows
sales_pipeline.csv This is Fact Tables **Purpose: Sales pipeline transactions**	opportunity_id	Unique deal/opportunity identifier	8800
	sales_agent	Name of salesperson handling the deal	30
	product	Product involved in deal	7
	account	Customer/company name	86
	deal_stage	Current/Final stage of deal	Lost Engaging Prospecting Won
	engage_date	Date deal engagement started	
	close_date	Actual close date of deal	
	close_value	Revenue collected for the deal	

DATA DICTIONARY

	Column	Description	Distinct Rows
ACCOUNTS.csv	account	Name of the customer/company	85
	sector	Business sector/industry	10
	year_established	Year when company was founded	
	revenue	Annual revenue (currency unspecified)	
	employees	Number of employees	
	office_location	Parent company name (nullable)	15
	subsidiary_of	Sales territory	

DATA DICTIONARY

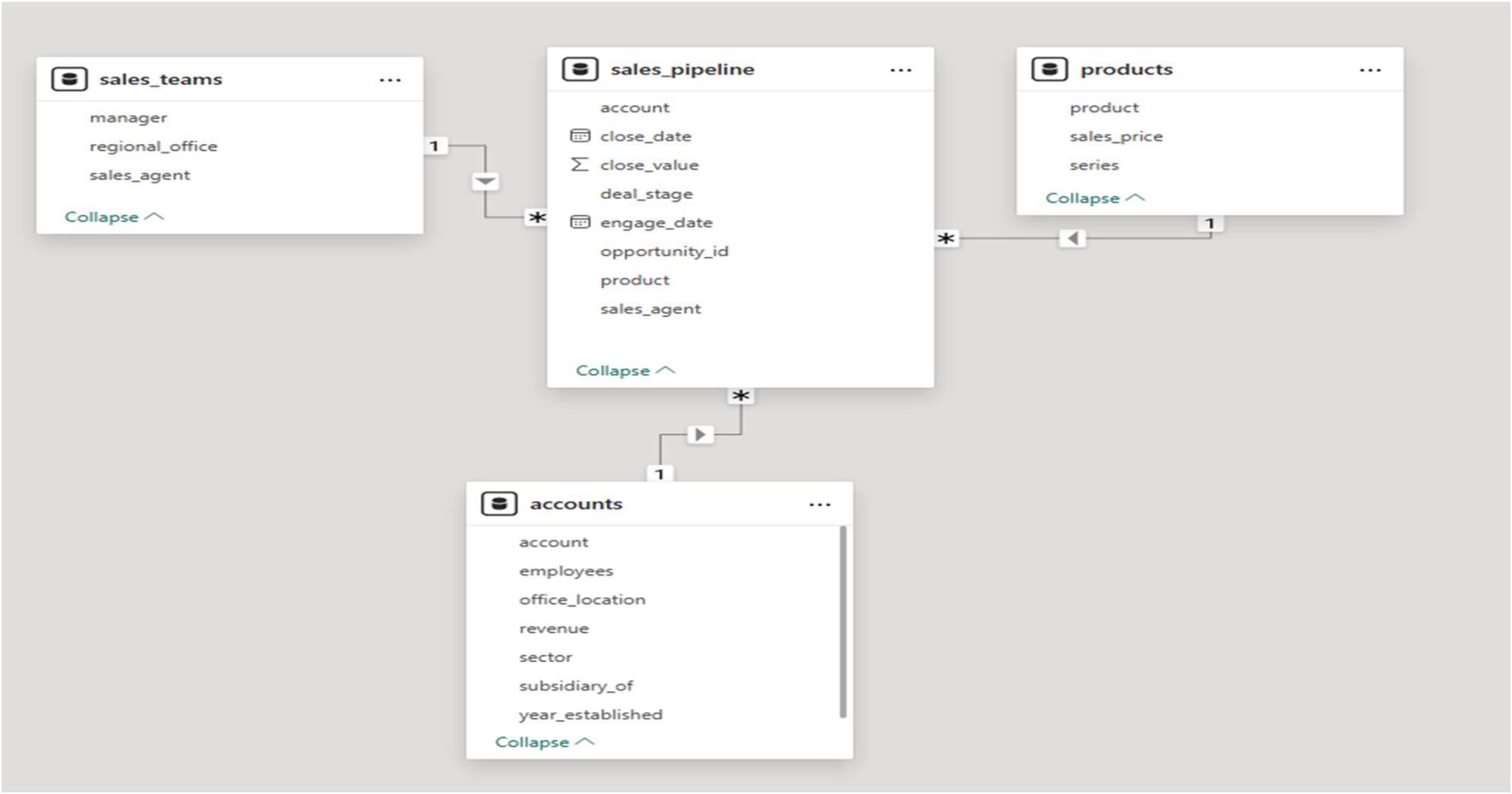
	Column	Description	Distinct rows
sales_teams.csv	sales_agent	Name of salesperson	35
	manager	Name of manager for the salesperson	6
	regional_office	Region of operation	Central East West

DATA DICTIONARY

	Column	Description	Example
products.csv	product	Product name	7
	Series	Product category	“GTK", “GTX", “MG"
	Sales_Price	Standard price for the product	

ER- Diagram

ER- Diagram


sales_teams
manager
regional_office
sales_agent
Collapse
1
*
sales_pipeline
account
close_date
close_value
deal_stage
engage_date
opportunity_id
product
sales_agent
Collapse
products
product
sales_price
series
Collapse
*
1
*
1
accounts
account
employees
office_location
revenue
sector
subsidiary_of
year_established
Collapse

ER diagram (textual)

Entities & keys:
ACCOUNTS(account PK, sector, year_established, revenue, employees, office_location, subsidiary_of)
PRODUCTS(product PK, series, sales_price)
SALES_TEAMS(sales_agent PK, manager, regional_office)
SALES_PIPELINE(opportunity_id PK, sales_agent FK, product FK, account FK, deal_stage, engage_date, close_date, close_value)
Cardinalities:
One account → many sales_pipeline rows
One product → many sales_pipeline rows
One sales_agent → many sales_pipeline rows

Table Overview

Primary key candidates

accounts.account	0 duplicate rows, 0 duplicate account.
products.product	0 duplicates of product.
sales_teams.sales_agent	35 unique agents, 0 duplicates.
sales_pipeline.opportunity_id	8,800 rows, 0 duplicate opportunity_id.

*Notes: So, we can treat these as **PKs** in SQL.*

Relationships (verified from data)

sales_pipeline.account → accounts.account	All non-null account values in sales_pipeline exist in accounts (no mismatches).
sales_pipeline.product → products.product	**One mismatching value**: GTXPro exists in sales_pipeline but in products it is GTX Pro (1480 rows).
sales_pipeline.sales_agent → sales_teams.sales_agent	All agents in pipeline appear in sales_teams.

Data Checks

Data checks / inconsistencies

Null / missing values

accounts
subsidiary_of: 70 / 85 rows are NULL (≈82%).
All other fields have 0 NULLs.
sales_pipeline
account: 1,425 NULLs
engage_date: 500 NULLs
close_date: 2,089 NULLs
close_value: 2,089 NULLs
By **deal_stage**:
Engaging: close_date & close_value are 100% NULL (expected – deal not closed).
Prospecting: also 100% NULL (no close yet).
Won and Lost: 0 NULLs in close_date and close_value.

So, missing close information is consistent with stage.

Data checks / inconsistencies

Categorical value issues

accounts.sector includes:
['technolgy', 'medical', 'retail', 'software', 'entertainment', 'marketing', 'telecommunications', 'finance', 'employment', 'services']
→ potential spelling issue: **technolgy**.
accounts.office_location includes :
Philipines → likely spelling issue.

Referential integrity

sales_pipeline.account vs accounts.account:
0 missing (good).
sales_pipeline.sales_agent vs sales_teams.sales_agent:
0 missing (good).
sales_pipeline.product vs products.product:
There is **one inconsistent product value**: GTXPro (no space) in 1,480 pipeline rows. The master products table has GTX Pro (with a space).

Data checks / inconsistencies

Date fields

engage_date, close_date are stored as text, not DATE.
Format observed: "YYYY-MM-DD" (e.g. 2016-10-25, 2017-03-11).

Numeric range checks

accounts.revenue (in millions USD):	min ≈ 4.54, max ≈ 11,698.03.
accounts.employees:	min = 9, max = 34,288.
products.sales_price:	55 to 26,768.
sales_pipeline.close_value:	• For Won deals: min = 38, max = 30,288. • For Lost deals: all **0**, which matches the dictionary meaning “revenue from the deal”.

Data Cleaning

Data Cleaning

Standardizing Parent–Subsidiary Structure	• Many accounts had **NULL values** in the *subsidiary_of* field. • NULL indicates the company is a **parent organization**. • For consistent roll-up reporting, each parent company is set to be **its own parent**. • This ensures clean hierarchy, easier grouping, and accurate parent-level aggregations.

Product Name Correction (GTXPro → GTX Pro)	• Product master lists **"GTX Pro"**, but Sales Pipeline contained **"GTXPro"** (missing space).
	• This mismatch affected:
	  • Product-level analysis
	  • Revenue grouping
	  • Product-wise KPIs
	• All occurrences of **GTXPro** were standardized to **GTX Pro** for accurate reporting.

Fixing Spelling & Standardization Issues	• Certain values contained spelling errors or inconsistent naming:
	  • technolgy → corrected to **technology**
	  • Philipines → corrected to **Philippines**
	• Standardizing categories ensures:
	  • Clean segmentation
	  • No duplicate categories
	  • Reliable visualizations & filtering in Power BI

Data Cleaning

Data Type Improvements (Dates & Numeric Fields)	• Dates in Sales Pipeline were originally stored as **text**.
	• Converted to **DATE** format for:
	  • Time-series analysis
	  • Funnel duration calculations
	  • Accurate monthly/quarterly trend charts
	• Verified that numeric fields (close_value, sales_price, revenue, employees) contain valid numeric values.

Ensuring Referential Integrity Across Tables	• Validated relationships between:
	  • Products ↔ Sales Pipeline
	  • Accounts ↔ Sales Pipeline
	  • Sales Teams ↔ Sales Pipeline
	• After cleaning:
	  • All product names align
	  • All accounts correctly map
	  • All sales agents match their team hierarchy

Handling Missing Values Intelligently	• subsidiary_of handled using hierarchy logic (as above). • Missing *account*, *close_date*, *close_value* in certain stages were validated as **legitimate**: • Prospecting & Engaging stages naturally have NULL close values/dates. • No imputation was required — values represent real pipeline state.

Introduction: 360° Analytical Tables

360° Analytical Tables

Objective:	Create consolidated 360° views that combine master data + pipeline performance for deeper insights.
Why 360 Tables?	• Provide a **single source of truth** for each entity • Simplify reporting & dashboard creation • Enable deeper analysis at **Account, Product, and Sales Agent** levels • Reduce repetitive joins in reports • Improve accuracy and performance of Power BI dashboards
360 Tables Created:	**1. Account_360** **2. Product_360** **3. Sales_Agent_360**

Why We Created 360 Tables

- Raw tables contain information scattered across multiple datasets (Accounts, Products, Sales Teams, Sales Pipeline)
- For meaningful business insights, data must be **consolidated**
- 360 tables enrich master data with aggregated KPIs such as:
  - Total deals
  - Won/Lost counts
  - Revenue generated
  - Win rates
  - Sales cycle duration
  - Unique customers / products
- These tables reduce complexity of analysis and maintain consistency

ACCOUNT 360 Overview

Overview

Purpose:	• Provide a complete, unified view of each account (customer).
Core Components:	• Company profile attributes • Pipeline performance linked to the account • Revenue generated + buying patterns • Sales cycle behavior
Key Metrics Included:	• Total opportunities • Won, Lost & Open opportunities • Revenue from Won deals • Average deal value • Win rate • Distinct products purchased • First engagement & last closure dates • Average sales cycle duration • Account size indicators (annual revenue, employees)
Business Value:	• Identify high-value customers • Segment accounts based on behavior & potential • Improve targeted selling strategies • Support Account Management dashboards

SQL Logic Used (Account 360)

Steps Executed:	• **Dropped existing table** (if present) to avoid duplicates
	• **Joined Accounts with Sales Pipeline**
	• **Performed aggregations** like sum, count, averages
	• **Calculated advanced KPIs**
	  • Win rate
	  • Avg sales cycle
	  • Distinct products purchased
	• **Saved results in a permanent table** using SELECT INTO

Outcome:	• A clean, enriched table—**account_360**—ready for dashboard usage.

PRODUCT 360 Overview

Overview

Purpose:	• Provide a complete performance view for each product across the pipeline.
Core Components:	• Product master information • Opportunity distribution • Conversion metrics • Product-specific buying behavior
Key Metrics Included:	• Total, Won, Lost, Open opportunities • Revenue generated from Won deals • Average winning deal value • Win rate per product • Number of distinct accounts buying product • Sales cycle patterns
Business Value:	• Identify top-performing and low-performing products • Understand demand patterns • Evaluate pricing vs value realization • Support product strategy & revenue forecasting

SQL Logic Used (Product_360)

Steps Executed:	• **Dropped old table** (product_360)
	• **Joined Products with Sales Pipeline**
	• **Aggregated pipeline performance metrics**
	• **Calculated product-level KPIs**
	  ◦ Win rate
	  ◦ Avg deal value
	  ◦ Distinct buyers
	• **Created permanent product_360 table** using SELECT INTO

Outcome:	• A unified dataset for all product-level analysis.

SALES AGENT 360 Overview

Overview

Purpose:	• Provide a complete view of individual sales agent performance.
Core Components:	• Agent hierarchy & region • Opportunity handling • Conversion behavior • Revenue contribution • Sales efficiency
Key Metrics Included:	• Total, Won, Lost, Open opportunities • Revenue generated • Average deal size • Win rate • Distinct accounts handled • Distinct products sold • First engagement to last closure • Avg sales cycle duration
Business Value:	• Identify top & bottom performers • Understand training needs • Optimize territory/agent assignment • Support Sales Manager & Regional dashboards

SQL Logic Used (Sales_Agent_360)

Steps Executed:	• Dropped previous sales_agent_360 table • Joined Sales Teams with Sales Pipeline • Calculated all performance metrics • Added product & account diversity metrics • Inserted results into permanent table via SELECT INTO
Outcome:	• A powerful dataset for Sales Performance dashboards.

Benefits and where to use

Benefits of 360° Consolidated Tables

- Reduced need for complex joins in SQL or Power BI
- Faster dashboard refresh times
- High data accuracy & consistency
- Unified KPIs across all reports
- Easier stakeholder communication
- Seamless segmentation & drill-down capabilities

Where These Tables Are Used (Dashboards)

Account 360 Dashboard	• Customer profiling • Revenue contribution • Win rate analysis • Product penetration • Sales cycle tracking
Product 360 Dashboard	• Product performance • Revenue by product • Deal size comparison • Demand segmentation
Sales Agent 360 Dashboard	• Agent performance • Regional comparison • Pipeline ownership • Sales effectiveness metrics

Final Outcome

We now have THREE strong, analytical tables:

- ❑ **ACCOUNT_360** – Customer Insight Layer
- ❑ **PRODUCT_360** – Product Insight Layer
- ❑ **SALES_AGENT_360** – Sales Performance Layer

These datasets are now ready to directly power:

- ✓ Power BI Dashboards
- ✓ KPI Monitoring
- ✓ Executive Reporting
- ✓ Advanced Analytics (Forecasting, Segmentation, etc.)

EDA plan – descriptive & diagnostic analysis

Descriptive analysis (what is happening?)

At account level (accounts)	• Count of accounts by:
	  • sector
	  • office_location
	  • subsidiary_of
	• Distribution of:
	  • revenue (min, max, quartiles)
	  • employees
	• Age of the company (current year − year_established).

At product level (products + sales_pipeline)	• Number of opportunities and won deals by product and series. • Total close_value (revenue) by product and series. • Average realized deal size vs sales_price for each product.
At sales pipeline level (sales_pipeline)	• Count of opportunities by deal_stage. • Win / loss counts and rates. • Distribution of close_value for Won deals. • Volume of opportunities by engage_date (month, quarter, year). • Average time from engage_date to close_date.
At sales team level (sales_teams + sales_pipeline)	• Number of opportunities & won deals by sales_agent, manager, regional_office. • Revenue (sum of close_value for Won) by agent/manager/region.

Diagnostic analysis (why is it happening?)

- Win rate by:
  - product
  - sector
  - office_location
  - sales_agent, manager, regional_office
- Average deal size by:
  - product / series
  - sector
- Conversion funnel:
  - Count of opportunities progressing Prospecting → Engaging → Won/Lost (using deal_stage snapshots).
- Deal cycle length:
  - Average days between engage_date and close_date by product, sector, and region.
- Impact of account size on performance:
  - Compare win rate & deal size across revenue bands and employee bands.

Dashboard Layout

EXECUTIVE SALES OVERVIEW DASHBOARD

Overview

Dashboard Objective	• Provide top-level visibility into overall sales performance • Monitor revenue trends, win/loss efficiency, opportunity pipeline health • Help leadership identify growth areas and performance gaps
Stakeholders	• CXO / Executive Management • VP Sales • Strategy & Planning team • Business Intelligence team
Key KPIs	• Total Revenue (Won Deals) --(Sum of close_value where deal_stage = 'Won'.) • Total Opportunities --(Count of all rows in the sales_pipeline table.) • Won Opportunities --(Count opportunities where deal_stage = 'Won'.) • Lost Opportunities --(Count opportunities where deal_stage = 'Lost'.) • Win Rate (%) --(Won / (Won + Lost).) • Open Opportunities (Prospecting + Engaging) --(Count where stage IN ('Prospecting', 'Engaging').) • Average Deal Value --(Average of close_value where deal_stage = 'Won'.) • Active Accounts --(Count distinct accounts present in sales_pipeline.)
Visuals	• Revenue Trend (Line chart) – monthly/quarterly revenue from Won deals • Revenue by Product (Bar chart) – identifies top-performing products • Opportunities by Deal Stage (Bar chart / Funnel) – pipeline health • Revenue by Sector (Bar chart / Treemap) – industry contribution • Revenue by Region (Bar chart) – regional performance • Top 10 Accounts (Horizontal bar) – key revenue drivers

Company Logo

EXECUTIVE SALES OVERVIEW

Navigating Buttons

Slicers

- Date Range
- Product
- Region

KPIs(Total Revenue , Total Opportunities, Won Opportunities, Lost Opportunities, Win Rate (%), Open Opportunities (Prospecting + Engaging), Average Deal Value, Avg Sales Cycle Duration)

Revenue Trend (Line chart)

Revenue by Product (Bar chart)

Opportunities by Deal Stage (Bar chart / Funnel)

Revenue by Sector (Bar chart / Treemap)

Revenue by Region (Bar chart)

Top 10 Accounts (Horizontal bar)

ACCOUNT 360 DASHBOARD

Overview

Dashboard Objective	• Provide a holistic, 360° view of each customer • Understand buying behavior, product mix, and revenue contribution • Support account planning & cross-sell strategy
Stakeholders	• Account Managers • Customer Success Team • Sales Directors • Retention & Loyalty Teams
Key KPIs	• Total Opportunities per Account --(Count opportunities where account = selected account.) • Won, Lost & Open Opportunities --(Count of deal_stage = 'selected' for that account.) • Revenue Won --(Sum close_value where stage = Won for that account.) • Win Rate --(Won ÷ (Won + Lost) for that account.) • Avg Deal Value --(Average close_value for Won deals of that account.) • Distinct Products Purchased --(Count distinct products in Won deals of that account.) • Avg Sales Cycle Duration --(Average of (close_date – engage_date) for Won deals.) • First & Last Engagement Dates --(MIN(engage_date) and MAX(close_date).)
Visuals	• Account Overview (Bar Chart – Sector, Location, Revenue) • Opportunity Stage Summary (Stacked Bar Chart) • Revenue by Product (Bar chart) • Engagement & Close Trend (Line chart) • Product Mix (Donut chart) • Opportunities Over Time (Line or Column Chart)

Company Logo

ACCOUNT 360 DASHBOARD

Navigating Buttons

Slicers

- Date Range
- Account
- Sector

KPIs(KPIs(Total Opportunities per Account, Won, Lost & Open Opportunities, Revenue Won, Win Rate, Avg Deal Value, Distinct Products Purchased, Active Accounts, First & Last Engagement Dates)

Sector Dominance by Country (100% Stacked Bar Chart)

Account`s Annual Revenue (Bar chart)

No of Opportunities per Account (Clustered Bar Chart)

Win Rate by Sector (Column Chart)

Avg Sales Cycle (Line Chart)

Avg deal value per sector (line chart)

PRODUCT 360 DASHBOARD

Overview

Dashboard Objective	• Provide complete insight into product performance • Understand revenue patterns, win rate, sector adoption, and pricing behavior • Help product managers and leadership plan product strategy
Stakeholders	• Product Managers • Sales Directors • Marketing Team • Revenue Strategy & Pricing Team
Key KPIs	• Total Opportunities by Product --(Count opportunities where product = selected product.) • Won & Lost Opportunities --(Count deal_stage = 'Won' for that product.) • Win Rate (Product) --(Won ÷ (Won + Lost).) • Revenue Won (Product) --(Sum close_value for Won deals of that product.) • Average Deal Value (Product) --(Avg close_value for Won deals of that product.) • Realization Gap (Sales Price vs Deal Value) --(Avg(close_value) – sales_price.) • Distinct Accounts Purchasing Product --(Count distinct accounts purchasing that product.) • Avg Sales Cycle by Product --(Avg(close_date – engage_date) for Won deals with that product.)
Visuals	• Revenue by Product (Bar chart) • Win Rate by Product (Column chart) • Avg Deal Value vs Sales Price (Scatter/Line) • Revenue Trend by Product (Line chart) • Product Adoption by Sector (Heatmap) • Detailed Product Opportunity Table

Company Logo

PRODUCT 360 DASHBOARD

Navigating Buttons

Slicers

- Product
- Sector
- Range

KPIs(Total Opportunities by Product, Won & Lost Opportunities, Win Rate (Product), Revenue Won (Product), Average Deal Value (Product), Distinct Accounts Purchasing Product, Avg Sales Cycle by Product)

Product Revenue(Column chart)

No of Opportunities per Product(Clustered Bar chart)

Avg Deal Value by Product (Bar Chart)

Win Rate By Product (Line Chart)

Product Adoption (Distinct Customers) (Donut Chart)

Avg Sales Cycle by Product (Line chart)

SALES AGENT 360 DASHBOARD

Overview

Dashboard Objective	• Provide full performance visibility of each sales agent • Highlight top performers and identify coaching opportunities • Understand agent-specific funnel health and revenue contribution
Stakeholders	• Sales Managers • Regional Heads • HR & Training Team • Director of Sales
Key KPIs	• Total Opportunities handled --(Count opportunities where sales_agent = selected agent.) • Won, Lost & Open Opportunities --(Count deal_stage = 'Won/Lost/Open' for that agent.) • Win Rate % (Agent) --(Won ÷ (Won + Lost).) • Revenue Won --(Sum close_value where deal_stage = 'Won' AND agent = selected agent.) • Avg Deal Value --(Avg close_value for Won deals of that agent.) • Distinct Accounts handled --(Count distinct accounts linked to that agent.) • Distinct Products sold --(Count distinct products in Won deals of that agent.) • Avg Sales Cycle Duration --(Avg(close_date – engage_date) for Won deals handled by that agent.)
Visuals	• Sales Agent Profile Card (Manager, Region) • Opportunities by Deal Stage (Bar) • Revenue by Sales Agent (Bar chart) • Win Rate by Agent (Column chart) • Revenue Trend by Agent (Line chart) • Product Mix Sold by Agent (Donut chart)

Company Logo

SALES AGENT 360 DASHBOARD

Navigating Buttons

Slicers

- Sales Agent
- Region
- Date Range

KPIs(Total Opportunities handled, Won- Lost & Open Opportunities, Win Rate % (Agent), Revenue Won, Avg Deal Value, Distinct Accounts handled, Avg Sales Cycle Duration)

Revenue by Agent (Bar Chart)

Opportunity Breakdown (Won, Lost, Open)(Clustered Column Chart)

Avg Deal Value by Agent (Bar chart)

Win Rate by Agent(Line chart)

Account Coverage by Sales Agent(Donut Chart)

Avg Sales Cycles (Line chart)

COHORT ANALYSIS DASHBOARD

Overview

Dashboard Objective	• Analyze customer retention, repeat engagement, and product lifecycle • Identify which customer acquisition months deliver long-term value • Support long-term forecasting and repeat revenue strategy
Stakeholders	• Strategy & Growth Team • Marketing • Customer Success • Business Analysts
Key KPIs	• Account Acquisition Cohorts --(First Won deal date (Year-Month) for each account.) • Cohort Size --(Count of accounts who's first Won date falls in that month.) • Retention % per Cohort --(Accounts returning in later months ÷ cohort size.) • Cohort Revenue --(Sum close_value of accounts belonging to that cohort.) • Average Number of Repeat Deals --(Count extra Won deals after the first for that cohort.) • Product Adoption per Cohort --(Distinct products purchased by cohort members.)
Visuals	• Cohort Heatmap (Year/Month vs Retention %) • Cohort Revenue Trend (Line chart) • Deal Count per Cohort (Bar chart) • Product Adoption by Cohort (Heatmap)

Company Logo

COHORT ANALYSIS DASHBOARD

Navigating Buttons

Slicers
Cohort Month/ Retention Month

KPIs(Total Cohort Revenue, Avg Month to Repeat, Repeat Purchase, Repeat Customers, Cohort Size)

Cohort Size (column chart)

Cohort Revenue Trend (line chart)

Retention Curve (Line chart)

Avg Month to Repeat by Cohort (Scatter Plot)

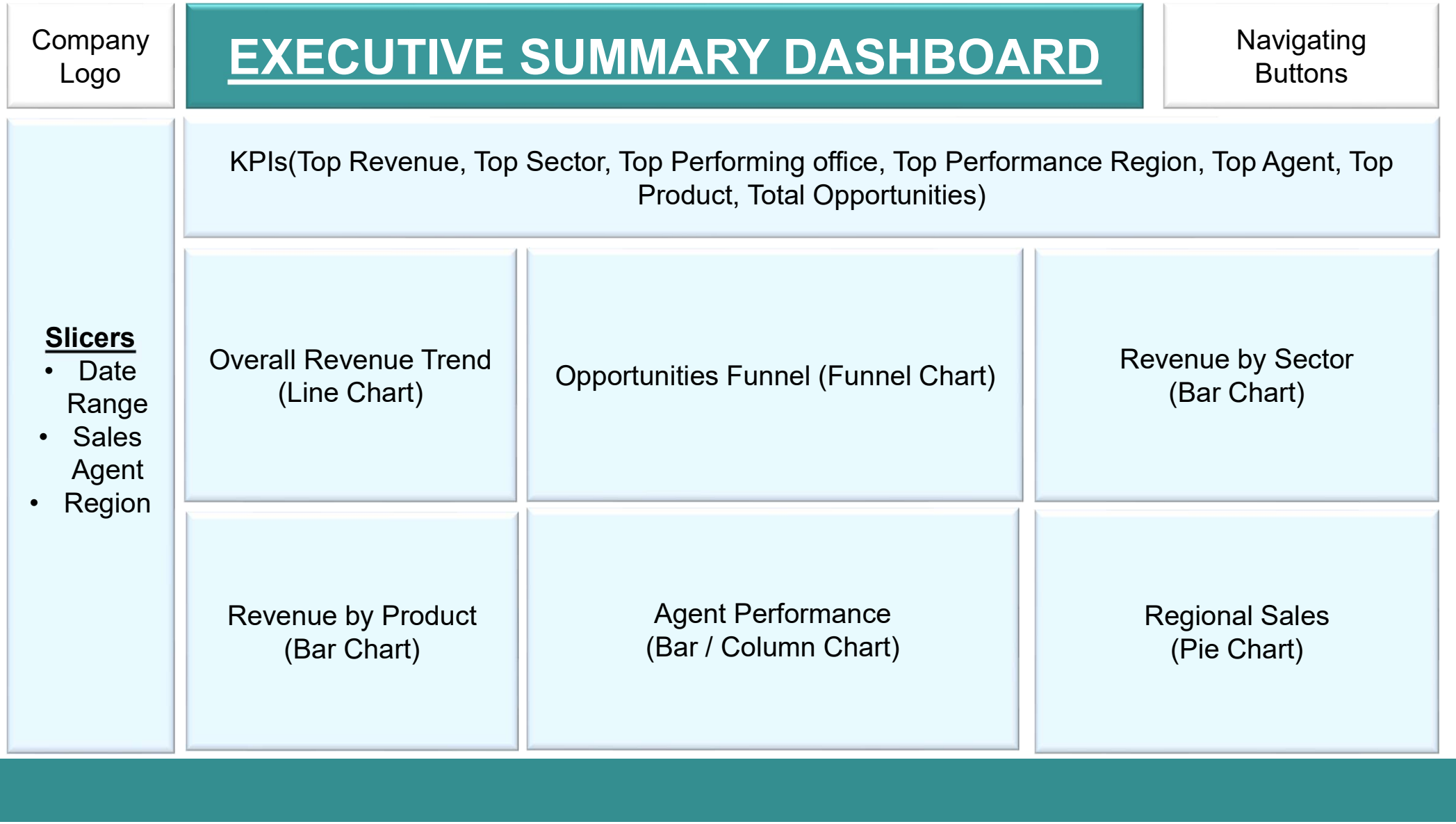
Repeat Purchase Analysis (Column Chart)

Cohort Analysis (Matrix with formatting)

EXECUTIVE SUMMARY DASHBOARD (Final Slide)

Overview

Dashboard Objective	• Combine highlights from all dashboards • Provide a concise overview for top management • Enable fast decision-making with minimal complexity
Stakeholders	• CEO • CXO • VPs • Board-Level View
Key KPIs	• Total Revenue • Win Rate • Top Product • Top Sector • Best Sales Agent • Top Account • Monthly Revenue Trend • Pipeline Snapshot
Visuals	• Overall Revenue Trend (Line Chart) • Top Products (Bar Chart) • Top Sectors (Bar / Column Chart) • Top Agents (Bar Chart) • Revenue by Region (Bar Chart) • Opportunity Stage Summary (Stacked Column Chart)


Company Logo
EXECUTIVE SUMMARY DASHBOARD
Navigating Buttons
KPIs(Top Revenue, Top Sector, Top Performing office, Top Performance Region, Top Agent, Top Product, Total Opportunities)
Slicers
• Date Range
• Sales Agent
• Region
Overall Revenue Trend (Line Chart)
Opportunities Funnel (Funnel Chart)
Revenue by Sector (Bar Chart)
Revenue by Product (Bar Chart)
Agent Performance (Bar / Column Chart)
Regional Sales (Pie Chart)

360° Business Performance Story

Executive Summary

- **Topline Performance**

- **Total Revenue:** ₹10.01M
- **Total Opportunities:** 8,800
- **Win Rate:** ~48%
- **Average Deal Value:** ₹2,360
- **Top Performing Sector:** Retail
- **Top Account:** Acme Corporation
- **Top Product:** GTX Pro
- **Top Region:** West
- **Top Sales Agent:** Darcel Schlecht

- **High-Level Insight**

Despite a strong revenue base and high win rate, the *business shows dependency on a small set of accounts, agents, and SKUs*, exposing the system to concentration risk.
Growth is being supported primarily by *acquisitions*, while *cohort retention is declining*, signaling weak repeat engagement.

Key Insight and Business Story

- **Strong Market Activity (8,800 Total Opportunities)**

Across **14 business months**, the team generated **8,800 opportunities**, indicating consistent demand and healthy engagement from the market.
This reflects:
- Strong brand presence
- Steady inflow of prospects
- Effective lead-generation channels

- **High Conversion Momentum (4,238 Won Deals)**

Winning **4,238 deals** shows a robust **48%-win rate**.
This signals:
- Strong product-market fit
- Effective sales execution
- High closing capability across the team

- **Deal Losses Highlight Improvement Areas (2,473 Lost Deals)**

The **2,473 lost deals** reveal:
- Competitive challenges
- Pricing and feature objections
- Potential gaps in nurturing or qualification

These represent **targeted opportunities** for process optimization.

Key Insight and Business Story

- **Healthy Near-Term Pipeline (2,089 Open Deals)**

With **2,089 open deals**, we have a strong forward pipeline.
This means:
- Significant future revenue potential
- Deals that need active follow-up
- Visibility into next-quarter forecasting

- **Team & Capacity Outlook (30 Sales Agents)**

Our **30 agents** handled all CRM interactions, leading to:
- ~293 opportunities per agent
- ~141 wins per agent on average

This forms the baseline for evaluating **productivity and load distribution**.

- **Portfolio Breadth Supports Cross-Sell (7 Products)**

The presence of **7 products** offers:
- Multiple entry points into accounts
- Upsell and cross-sell potential
- Opportunities to build product-level performance dashboards

Key Insight and Business Story

- **Focused but High-Value Account Base (85 Companies)**

Selling to **85 companies** shows:
- A concentrated but high-value customer base
- Scope to deepen wallet share
- Potential for strategic account-based selling

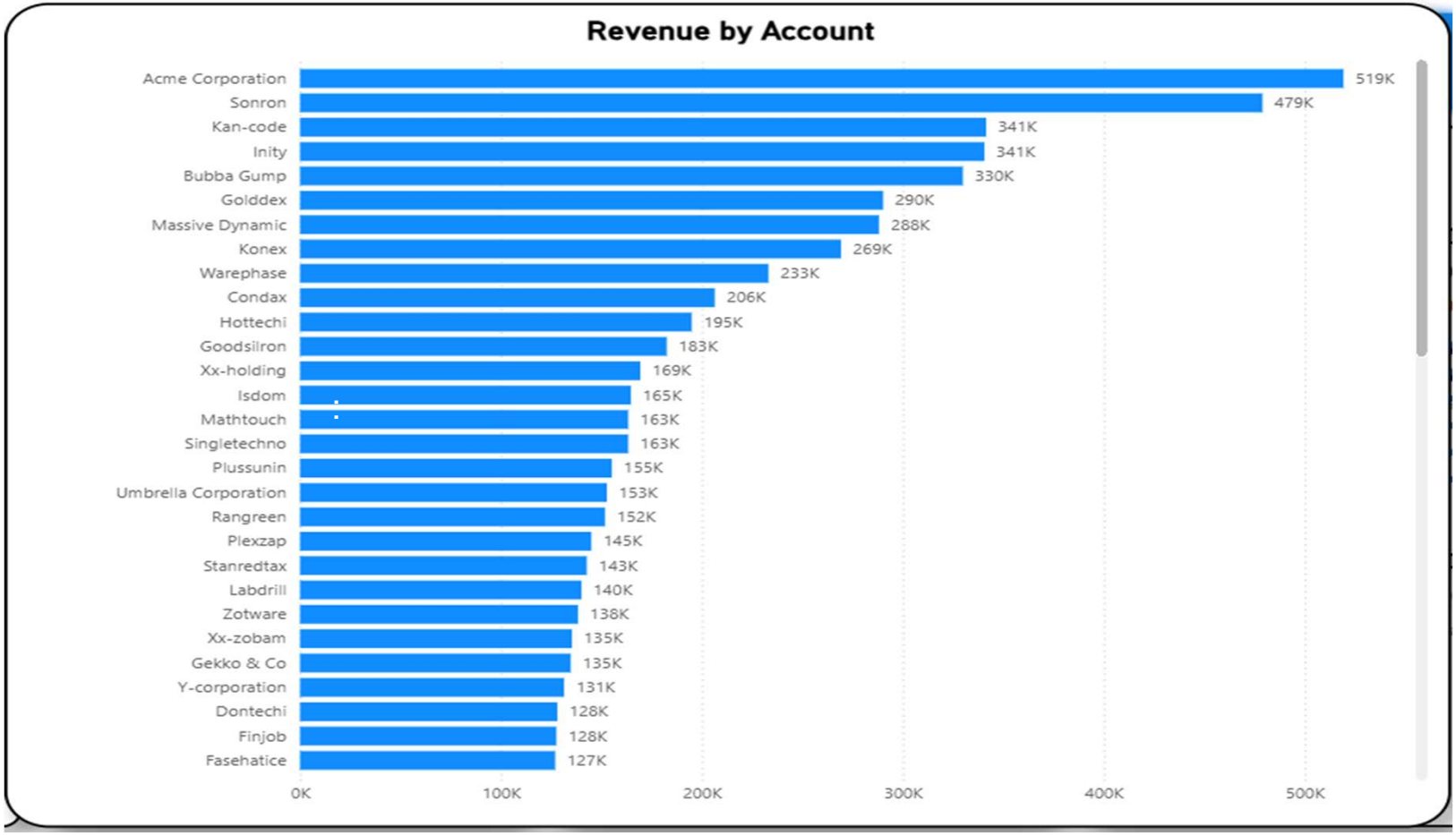
- **14 Months of Business Trends Enable Cycle Analysis**

The **14 business months** allow us to identify:
- Seasonal patterns
- Sales cycle duration
- Monthly performance variations
- Trends in wins, losses, and opportunity creation

This turns the CRM data into a **time-series story**, helping leadership understand momentum, dips, and growth phases.

"Over the last 14 business months, our CRM has recorded 8,800 opportunities, reflecting consistently strong market engagement. With 4,238 wins and 2,473 losses, we maintain a healthy win rate while identifying clear areas for process refinement. A forward pipeline of 2,089 open deals positions us well for upcoming quarters.
Our 30-agent team manages a diverse 7-product portfolio across 85 companies, highlighting both depth and breadth in customer relationships. The 14-month trend data provides a complete view of our sales cycle, performance trajectory, and strategic opportunities for growth."

Account level Insights (Revenue Distribution)



Account level Insights (Revenue Distribution)

- **Revenue is highly concentrated at the top**

- **Acme Corporation (519K)** and **Sonron (479K)** alone contribute a **major share of the total revenue**.
- These two accounts form the **strategic core** of the portfolio and should receive priority in relationship management, retention strategy, and dedicated account plans.

- **Strong mid-tier accounts support stability**

- Companies like **Kan-code, Inity, Bubba Gump, Golddex, Massive Dynamic**, and **Konex** each generate **250K–340K**.
- This mid-tier segment provides **steady recurring revenue** and offers high potential for **upsell and cross-sell**, especially given your 7-product portfolio.

- **Long-tail accounts show diversification but low penetration**

More than 20 accounts fall in the **120K–200K** revenue range.
These accounts are important for:

- Reducing over-dependence on top customers
- Identifying hidden growth opportunities
- Expanding product adoption (low product penetration likely)

Account level Insights (Revenue Distribution)

- **Opportunity for deepening wallet share**

- Many accounts appear to have **moderate spend** despite long-term engagement.
- There is immediate potential to:
  - Improve share of wallet
  - Introduce bundled offerings
  - Target renewals + expansions with account-based marketing (ABM)

- **Risk Management Insight**

- The top 2 accounts contribute disproportionately high revenue.
- Revenue concentration risk exists — losing even one top client would significantly impact performance.
- Recommended:
  - Strengthen customer success engagement
  - Create retention plans
  - Monitor account health KPIs (NPS, ticket volume, deal activity)

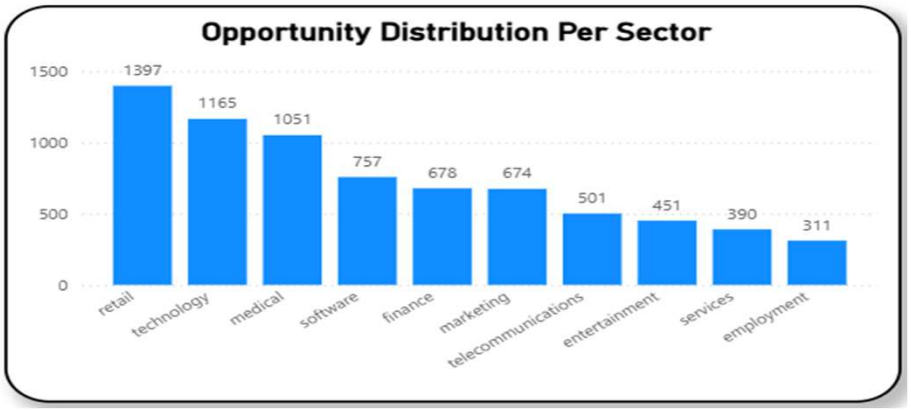
Business Interpretation

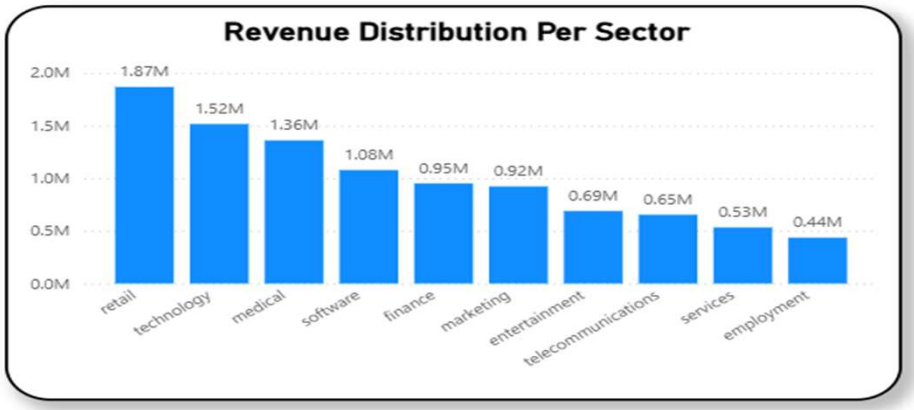
“Revenue is concentrated across a few strategic accounts, with Acme Corporation and Sonron leading by a significant margin. Mid-tier accounts contribute stable revenue and represent strong cross-sell opportunities. The long-tail accounts provide diversification but show limited penetration, indicating room for growth. Strengthening top-account relationships while scaling mid-tier value will maximize revenue expansion and reduce concentration risk.”

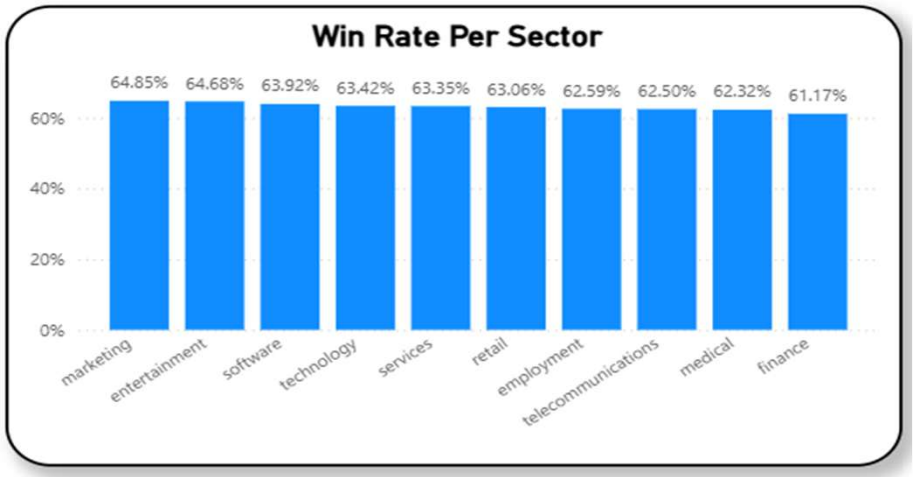
Win Rate Insights (48%)

- **High Performance:** A 48%-win rate means we convert nearly two-thirds of qualified opportunities.
- **Above Benchmark:** This is between typical industry win rates (30–50%).
- **Strong Product-Market Fit:** Indicates customers clearly see value in our offering.
- **Effective Sales Execution:** Shows strong qualification, follow-ups, and deal management by the sales team.
- **Healthy Funnel Efficiency:** We close most well-qualified deals, reducing leakage in the pipeline.
- **Strategic Focus Ahead:** Maintain this high conversion rate while increasing the volume of quality opportunities.

Sector Level Behavior







Sector Level Behavior

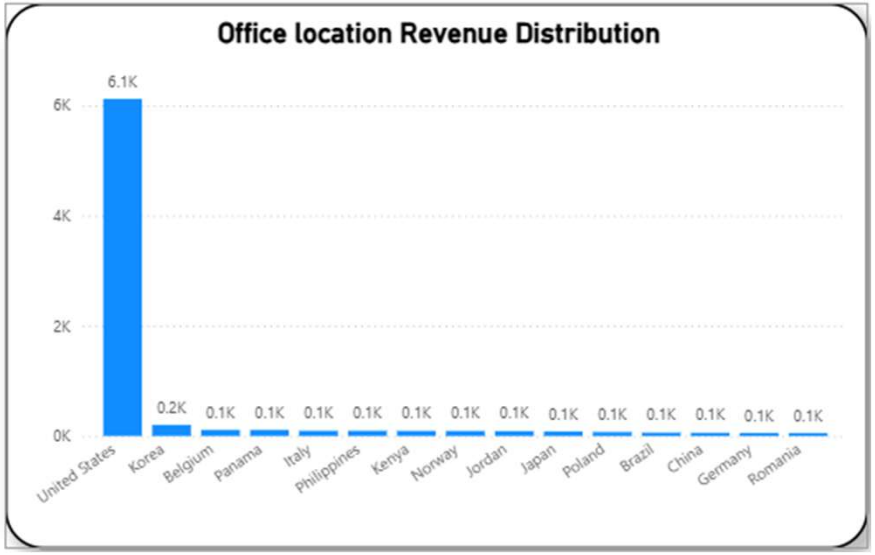
- **Holistic Sector Performance Analysis**

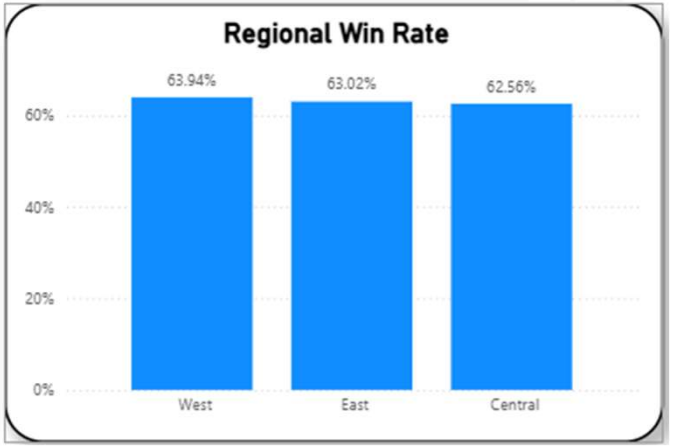
- The business is **heavily driven by Retail, Technology, and Medical sectors**, which contribute the largest share of both revenue and opportunity volume. These sectors form the **core commercial foundation** of the company.
- However, **conversion efficiency does not always match pipeline size**. Some high-volume sectors convert at only moderate rates, indicating potential improvement areas in pricing, messaging, or sales approach.
- A surprising insight is that **Marketing, Entertainment, and Software sectors show very high win rates**, despite having low opportunity volume. These sectors represent **high-return, under-penetrated markets** that could grow rapidly with increased lead generation.
- Meanwhile, sectors like **Medical and Finance** show strong demand but comparatively lower win rates, suggesting **revenue leakage** and a need for strategic refinement.

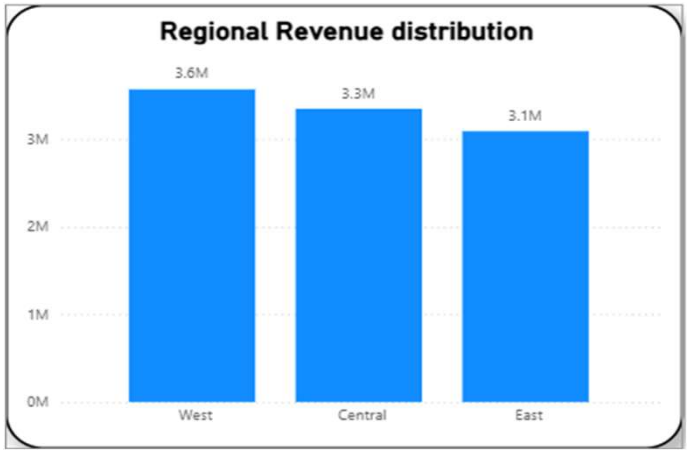
- **Business Interpretation**

The company's growth is anchored in a few large sectors, but hidden opportunities exist in smaller, high-efficiency markets. Improving conversion in Medical/Finance and increasing outreach in high-win sectors can significantly boost revenue without increasing overall acquisition cost.

Regional & Office Performance







Regional & Office Performance

- **Office Location Revenue Distribution – Highly Concentrated Revenue**

- **US alone contributes the majority of total revenue (≈8.4M)** — far ahead of all other locations.
- Remaining countries contribute **very small, fragmented values (<0.2M each)**.

Meaning:
The business is over-dependent on the US market. Revenue risk is high if US demand slows. There is also large growth potential in under-penetrated international offices.

- **Regional Win Rate – Consistent but Slightly Higher in West**

- Win rates across regions are **very close** (East ~63%, Central ~62.6%, West ~63.9%).
- West is marginally the strongest performer.

Meaning:
Sales capability is **uniform across regions**. The West has a slightly more efficient conversion engine — worth studying for best practices.

- **Regional Revenue Distribution – West Generates the Most Revenue**

Revenue is highest in the **West (~3.6M)**, followed by **Central (~3.3M)** and **East (~3.1M)**.

Meaning:
The West's higher revenue is **volume-driven**, not conversion-driven (since win rates are similar). This indicates:

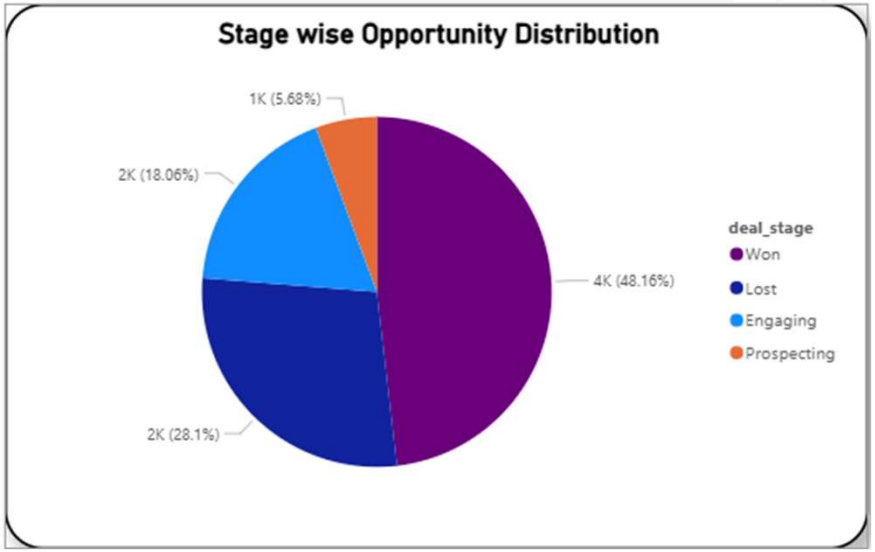
- Stronger customer base density
- More mature accounts
- Higher-value deals

Regional & Office Performance

- Business Interpretation

- **US market is the revenue backbone**, contributing the bulk of global income.
- **West Region is the commercial hub**, driving the largest revenue volume with slightly higher efficiency.
- **East and Central have strong win rates**, but **lower deal volume**, indicating untapped sales potential.
- **Opportunity:** expand pipelines in East & Central while replicating the West's sales playbook to scale revenues.

Opportunity Dynamics



Stage wise Opportunity Distribution

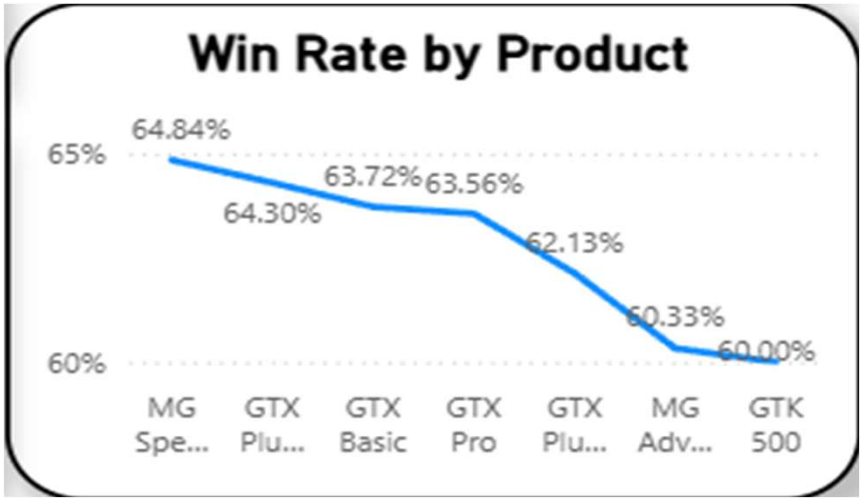
- **Insight**

- Nearly **half of all opportunities (48%) convert to Won**, reflecting strong sales effectiveness.
- However, **28% Lost deals** indicate a meaningful drop-off, suggesting issues in competitive positioning or objection handling.
- **18% are still Engaging**, representing the primary segment for near-term conversions.
- **Prospecting is only 6%**, revealing a weak top-of-funnel pipeline.

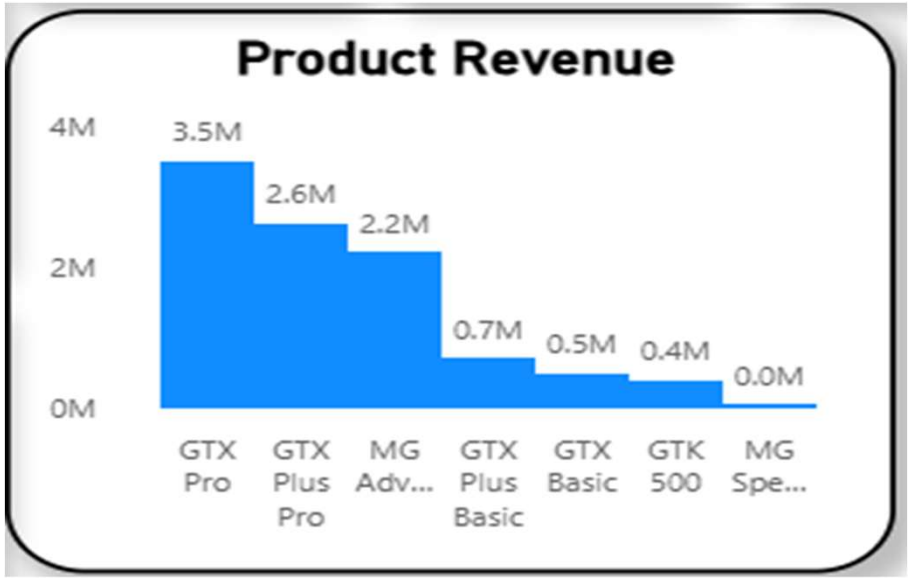
- **Business Interpretation**

- High win rate signals **strong product–market fit and sales execution**, but the high loss volume implies **inconsistent deal closure discipline** or **late-stage disqualification**.
- The low Prospecting share is a red flag: without more top-funnel opportunities, **future revenue may decline even if win rates remain strong**.
- Leadership should prioritize **lead generation**, **early qualification**, and **improving win–loss strategies** to reduce leakage and stabilize long-term pipeline health.

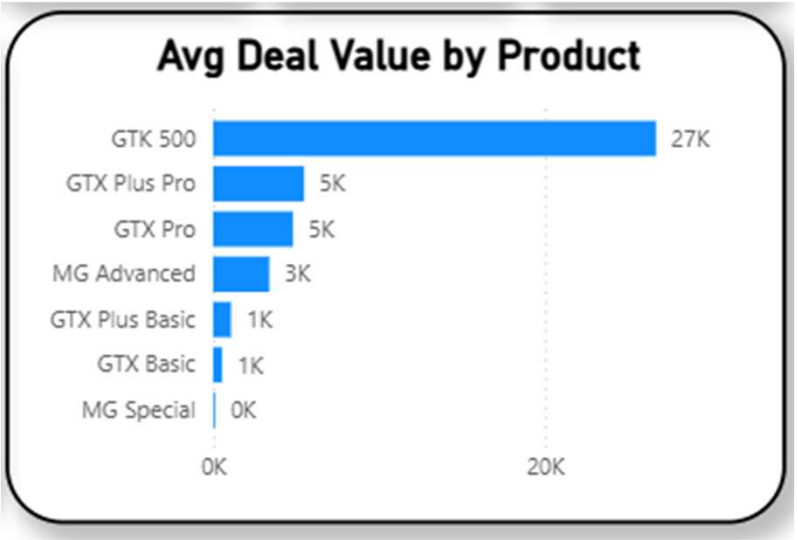
Product Win Rate



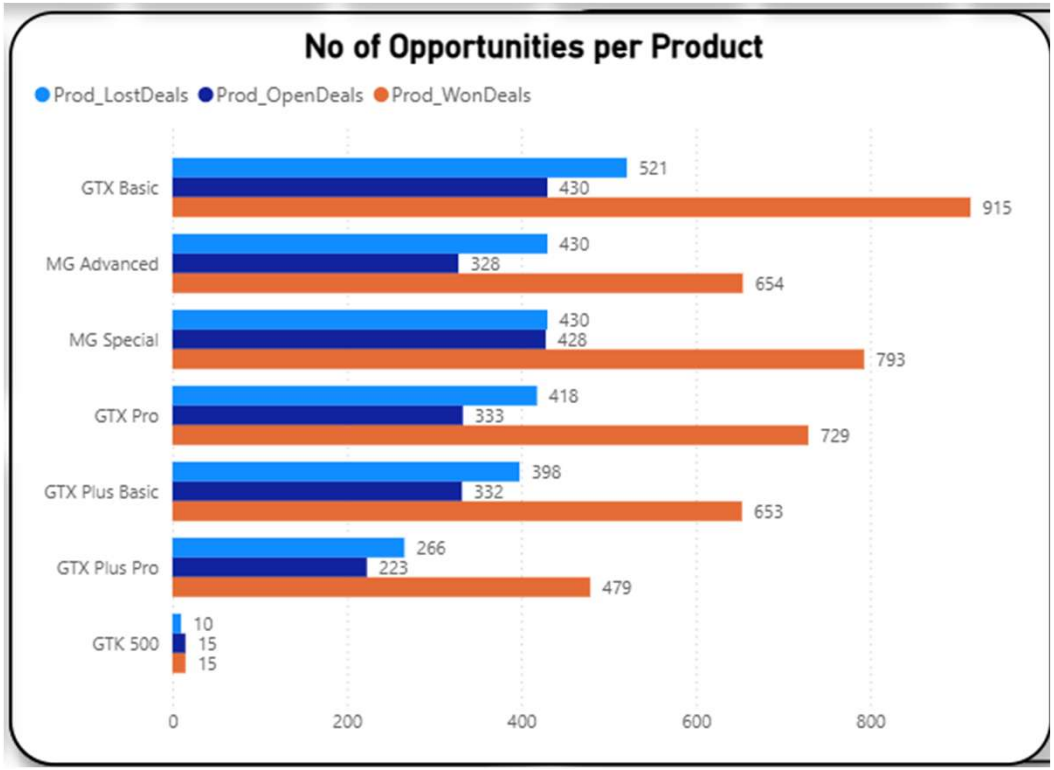
Product Performance



Product Avg Deal Value



Product Pipeline Performance



Product Performance Overall Insight

- **GTX Pro & GTX Plus Pro are the Core Revenue Engines**

- **GTX Pro generates the highest revenue (~$3.5M)** despite being *3rd* in total opportunities, indicating **high conversion efficiency**.
- **GTX Plus Pro follows with ~$2.6M**, also showing strong monetization.
- These two products **contribute more than 55–60% of total revenue**, making them the **primary growth drivers** in the portfolio.

- **Win Rates are Strong but Declining Across Higher-Volume Products**

- **MG Special leads with a 64.8%-win rate**, followed closely by GTX Plus Basic and GTX Basic.
- The win rate **declines with higher-priced or premium models** (MG Advanced, GTK 500 at ~60%).

- **GTX Basic Generates the Highest Opportunity Volume but Lower Revenue**

- GTX Basic has **915 opportunities**, the highest among all products.
- However, its revenue is **significantly lower (~$0.5M)** due to lower pricing.

- **MG Special & MG Advanced Are Underperforming Despite Strong Win Rates**

- Both MG variants show **healthy conversion rates**, but their **total revenue contribution is low**.
- This suggests:
    - Low pricing or narrow target segment
    - Weak market demand relative to the GTX series

Product Performance Overall Insight

- GTK 500 Is a Red Flag Product

- Lowest revenue ($0M–very low).
- Very few opportunities (≈15).
- Win rate is **(~48%)**.
- Highest average deal value (27K), but negligible sales volume.

- Opportunity Distribution Shows a Heavy Skew

Products with high opportunity volume:
- GTX Basic (915)
- GTX Pro (729)
- GTX Plus Basic (653)

Products with low volume:
- GTK 500 (15)
- MG Special (430)
- MG Advanced (430)

Business Interpretation

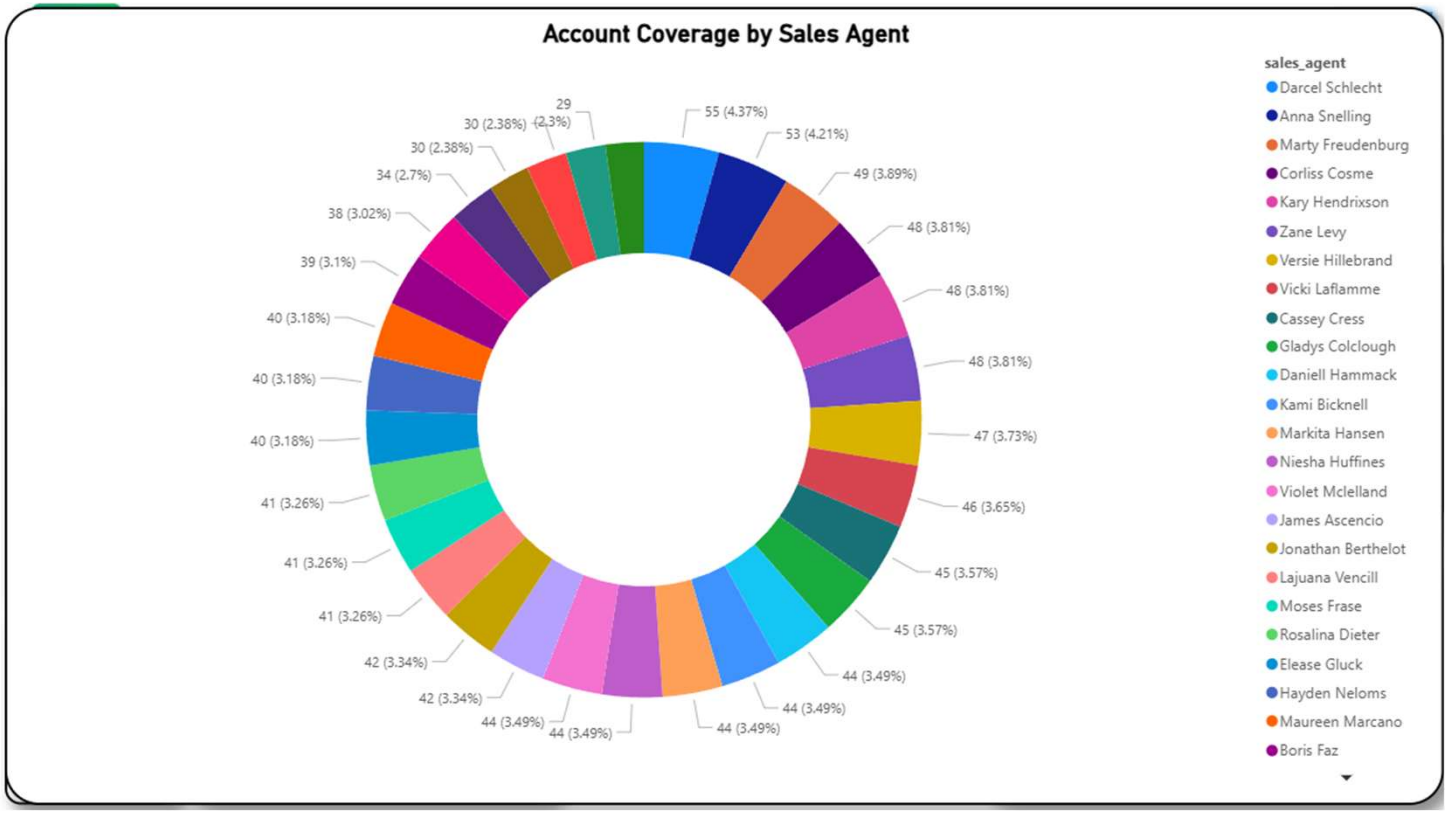
The product portfolio analysis reveals a structural imbalance in revenue contribution and market traction across the offering suite. The GTX family—particularly **GTX Pro** and **GTX Plus Pro**—continues to anchor overall business performance by delivering the highest revenue and strong win efficiency, despite not commanding the largest share of total opportunities. Conversely, mass-market products such as **GTX Basic** and **GTX Plus Basic** attract significant customer engagement but convert into relatively low-value deals, diluting overall revenue productivity.
Premium models, notably **MG Advanced** and **GTK 500**, demonstrate limited commercial momentum, reflected in both lower win rates and smaller deal pipelines. This suggests a potential misalignment between product positioning, perceived customer value, and pricing strategy.

Taken together, the findings highlight an urgent need for **portfolio optimization**:

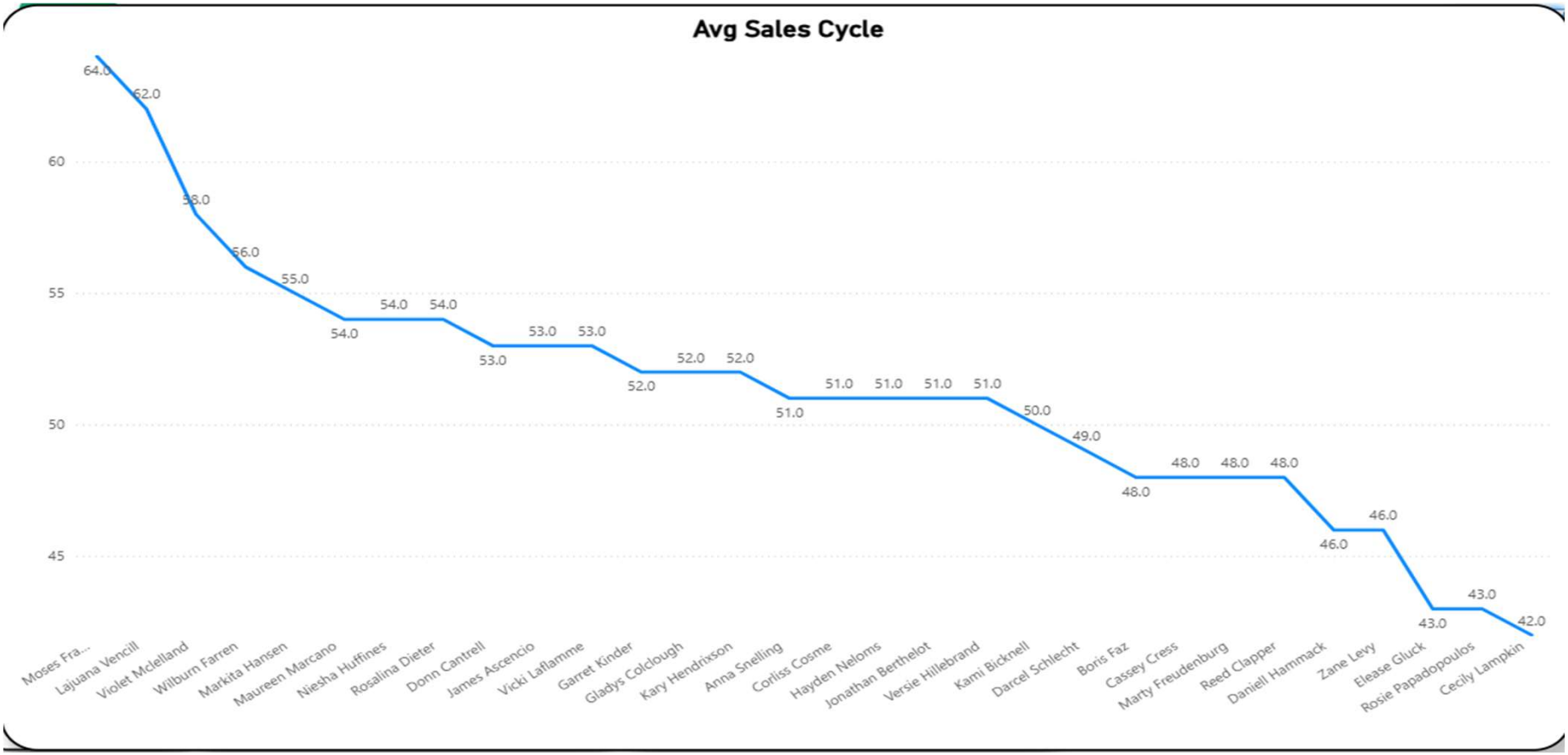
- reinforcing commercial investment behind high-performing GTX variants,
- repositioning or re-engineering underperforming premium products, and
- deploying a structured cross-sell and value-upgrade strategy to migrate high-volume customers toward higher-margin offerings.

This insight underscores a strategic opportunity to improve revenue quality, strengthen product-market fit, and increase overall sales efficiency.

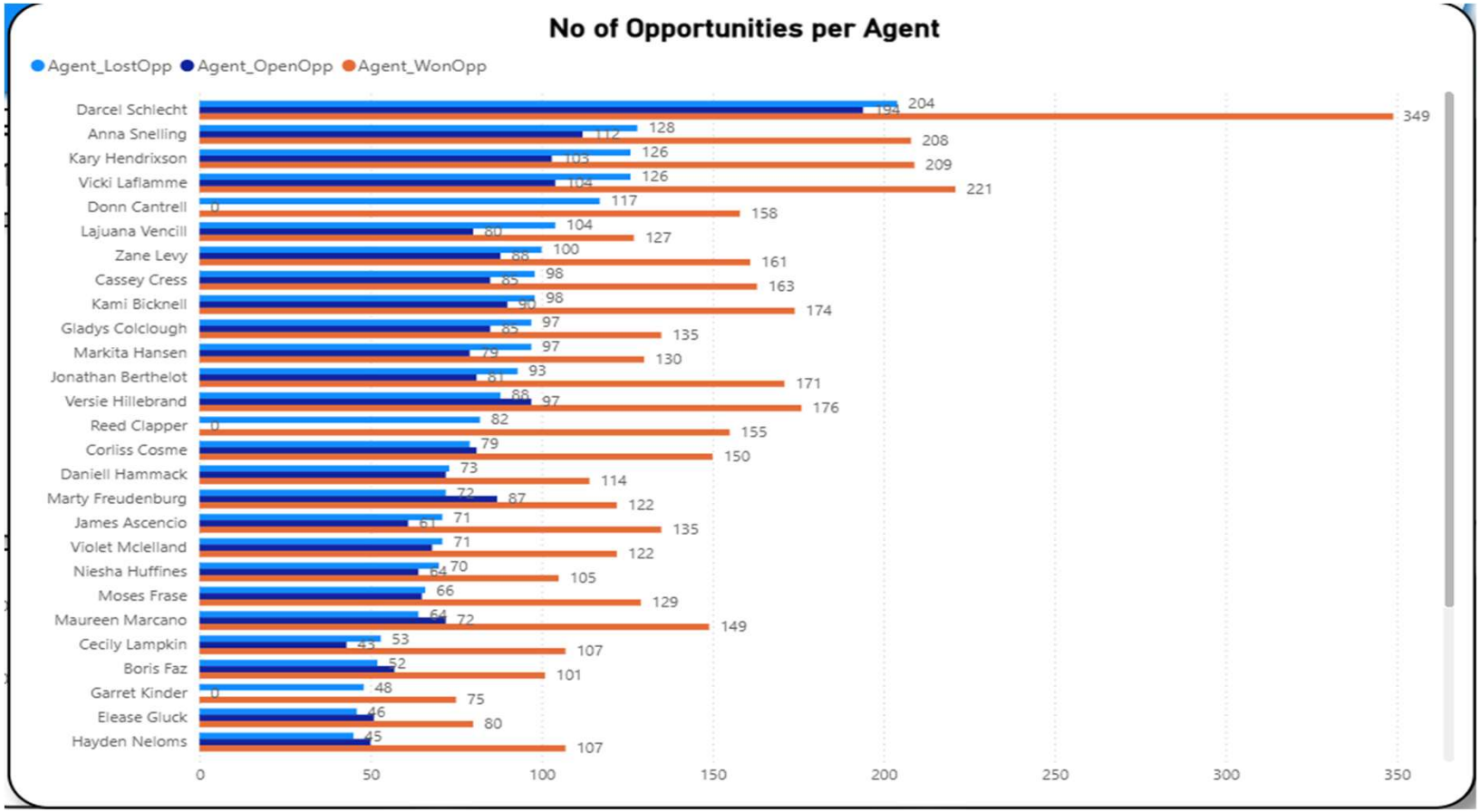
Sales Agent Performance



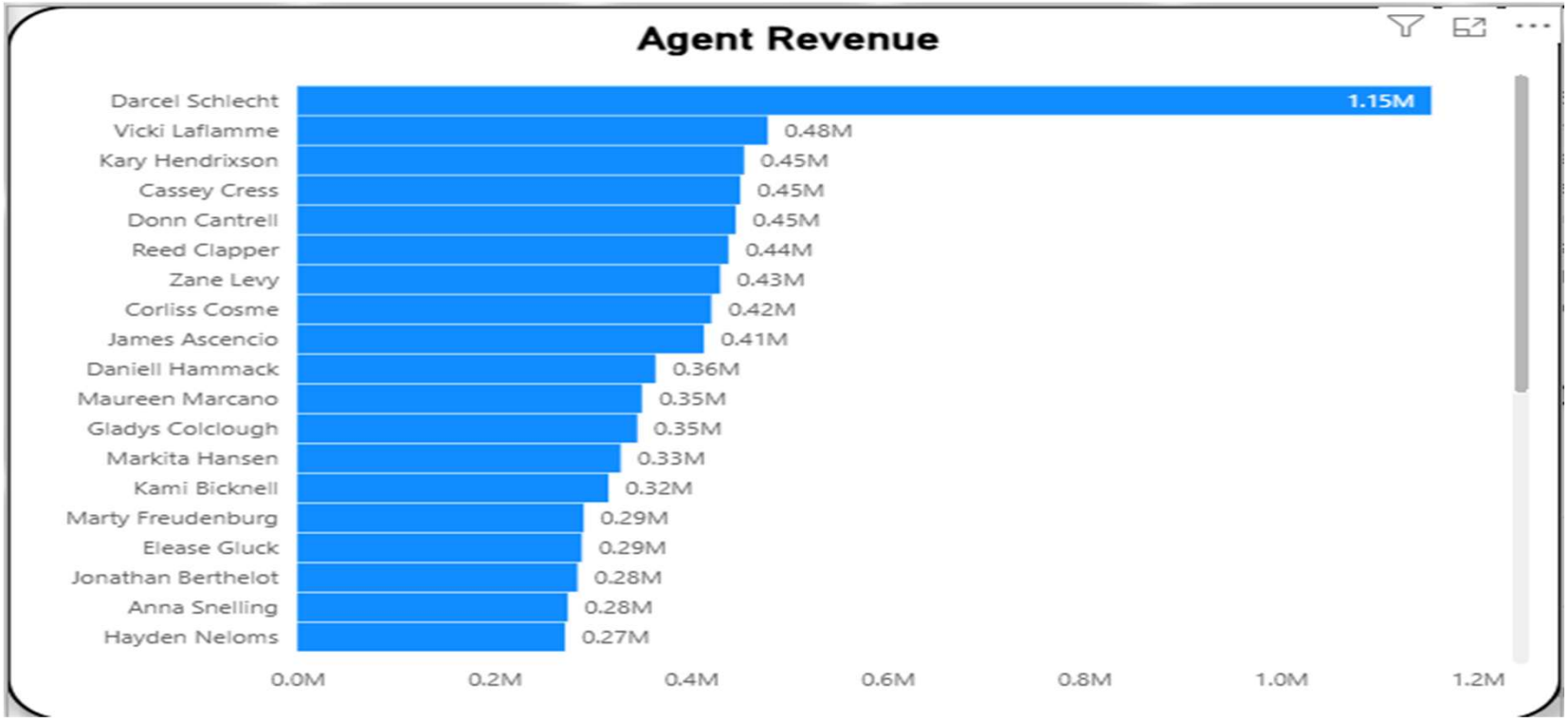
Sales Agent Performance



Sales Agent Performance



Sales Agent Performance



Executive Business Interpretation – Sales Agent Performance Insights

The sales-agent performance analysis reveals a **highly balanced distribution of account coverage**, with most agents managing **40–55 accounts each**, indicating equitable workload allocation across the team. Despite this uniform distribution, the **effectiveness of account conversion varies sharply**.

Agents such as **Darcel Schlecht, Anna Snelling, Kary Hendrixson, and Vicki Laflamme consistently outperform peers**, leading significantly in **won opportunities** while also contributing the **highest revenue volumes**, with Darcel notably exceeding **$1.15M**, far ahead of the rest of the team.

A deeper look at the **average sales cycle** shows a strong correlation between **shorter cycle times and higher revenue productivity**. Top performers exhibit cycle times in the range of **42–50 days**, whereas underperforming agents take **60+ days**, signaling opportunity for targeted coaching and process optimization. This suggests that reducing cycle friction directly enhances win outcomes and revenue acceleration.

The **opportunity mix across agents** also highlights strategic coaching opportunities. While high performers maintain a balanced ratio of open, lost, and won deals, several mid-tier agents—despite similar account loads—show disproportionately high open opportunity backlogs and comparatively lower win counts, indicating **pipeline stagnation** and the need for **better prioritization or lead qualification practices**.

Finally, the distribution of revenue confirms a **clear Pareto pattern**: a small subset of agents is responsible for a disproportionate share of organizational revenue. This concentration emphasizes the importance of **replicating top-performer behaviors**, strengthening enablement processes, and deploying targeted training for the long tail of the team.

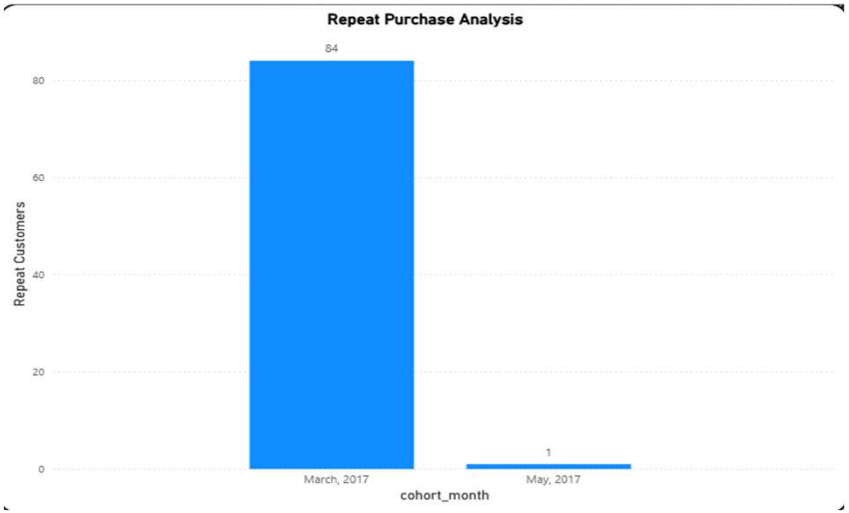
Cohort Analysis

Cohort Heatmap

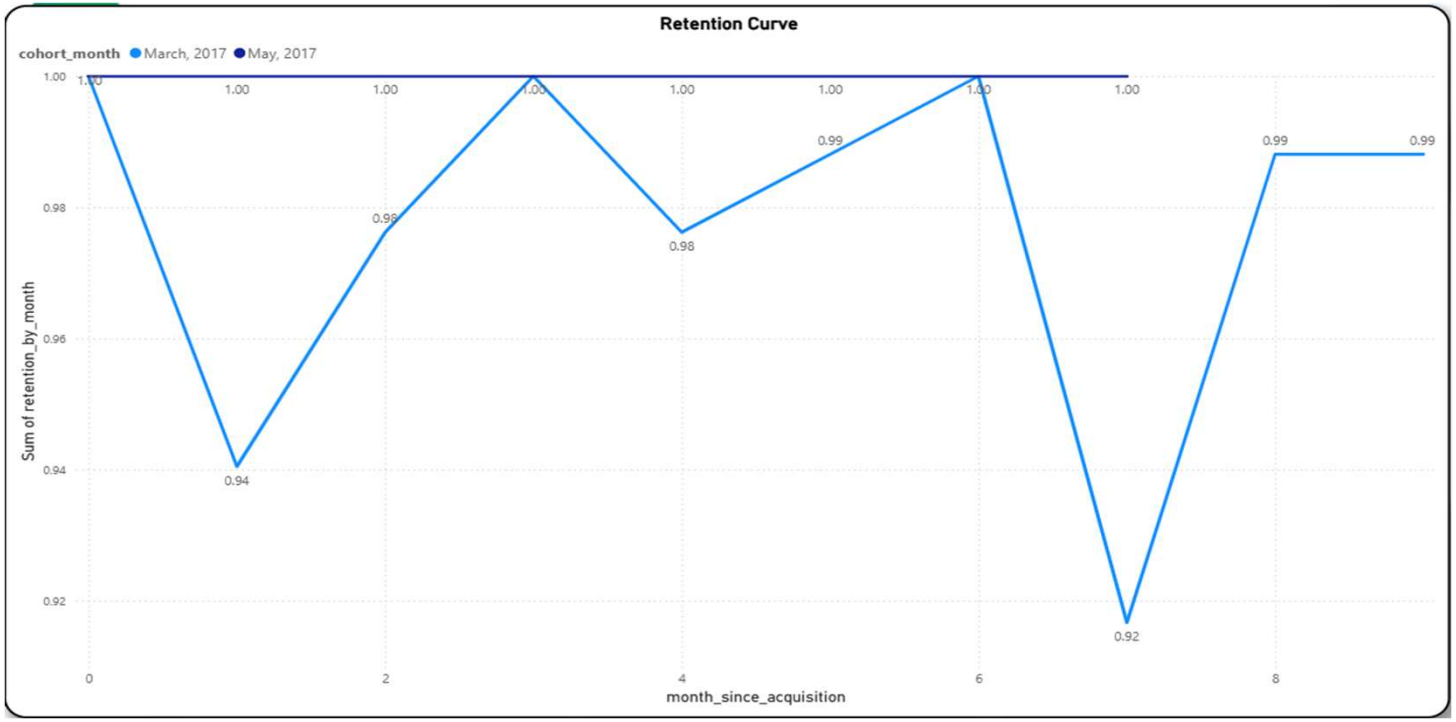
Cohort Analysis

cohort_month	0	1	2	3	4	5	6	7	8	9
May, 2017	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
March, 2017	1.00	0.94	0.98	1.00	0.98	0.99	1.00	0.92	0.99	0.99

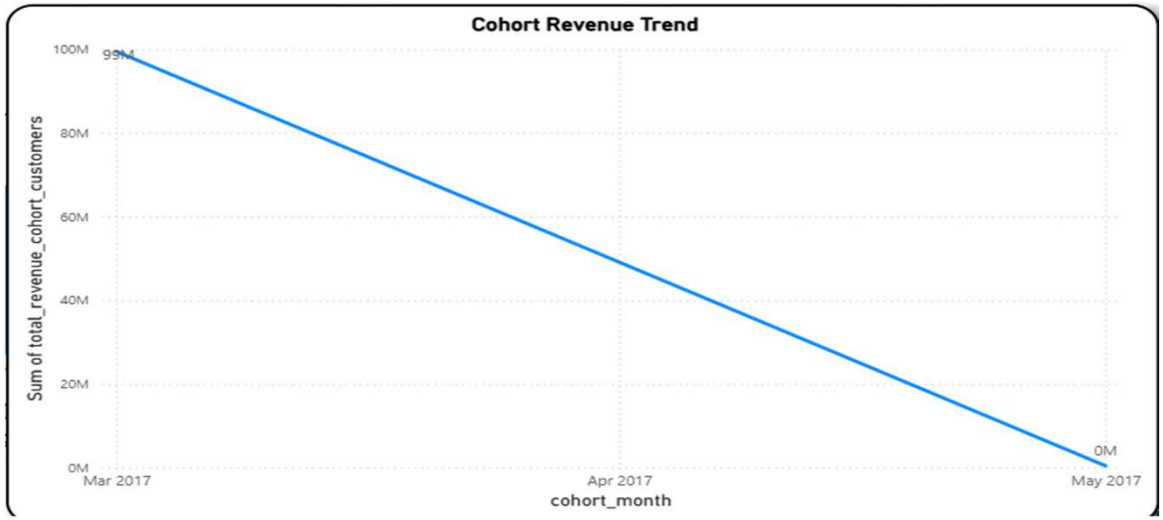
Repeat Purchase Analysis



Retention Curve



Cohort Revenue Trend



Cohort Analysis Executive Summary

• Cohort Engagement & Repeat Purchase Behavior

Repeat Purchases Are Extremely Concentrated

- March 2017 cohort generated 84 repeat customers, while May 2017 generated only 1 → a 99% drop.
- Indicates that customers who joined later are not progressing to repeat buying cycles.

Executive Interpretation:

- The business is highly dependent on a single large cohort (March 2017) for recurring revenue.
- Newer customers are not nurtured enough to move beyond the first purchase.

• Revenue Sustainability & Cohort Contribution

Cohort Revenue Trend Shows Sharp Decline

- March 2017 cohort contributed **~99M in cumulative revenue**.
- May 2017 contributed **near-zero** (almost no repeat activity).

Executive Interpretation:

- Revenue is disproportionately skewed toward older cohorts.
- **Recent cohorts are not monetizing**, creating a future revenue cliff unless retention improves.
- This threatens long-term revenue stability.

Cohort Analysis Executive Summary

• Retention Curve Insights

Retention curve shows:

- **Month 0 retention = 100%** for both cohorts—indicating solid acquisition.
- **Month 1 dip to 94%** for March 2017 → early disengagement.
- **Retention stabilizes between 98–100%** in mid-cycle for March 2017.
- **Severe drop at Month 7 (92%)** → churn accelerates late in the cycle.
- May 2017 stays at **100% across most months** simply because *the cohort is extremely small*—not because retention is strong.

Executive Interpretation:

- Early engagement (Month 1–2) is the **critical drop-off point**.
- Later churn (Month 7) suggests customers see **limited long-term value**, or product/service lifecycle is short.
- May 2017 looks stable numerically, but **cohort size is too small to draw meaningful conclusions**.

• Cohort Heatmap Interpretation

Insights:

- March 2017 shows **clear lifecycle behaviour**: strong acquisition → mild mid-cycle stability → late-cycle churn.
- May 2017's perfect scores are misleading—due to **small sample size**, not true retention.

Executive Interpretation:

- Retention problems are masked in smaller cohorts.
- Larger cohorts exhibit **real behavioural decline** indicating structural retention weakness.

Strategic Diagnosis

- **Acquisition is strong, but retention is weak**

- You acquire customers effectively (100% Month 0 retention),
- But **fail to convert them into repeat buyers** (only 1 repeat in May cohort).

- **The business is over-reliant on legacy customers**

- March 2017 drives almost all recurring revenue.
- New cohorts contribute almost nothing beyond initial purchase.

- **Customer lifecycle management is missing**

Patterns suggest:

- No structured onboarding
- No follow-up touchpoints
- No engagement campaigns between Month 1–3
- No renewal / upsell strategy

- **Revenue risk is growing**

If new cohorts do not develop into revenue-generating customers:

- Revenue will decline
- CAC will rise
- Sales ROI will drop

Executive Recommendations

- **Fix Early-Cycle Retention (Month 1–2)**

- Deploy onboarding workflows
- Automated reminders
- Product usage or value communication within first 30 days

- **Launch a Repeat Purchase Program**

Examples:
- 10–15% discount on second purchase
- Upgrade bundles
- Loyalty credits for long-term users

- **Build a Month 6–8 Re-Activation Campaign**

To address the Month 7 drop:
- Renewal email campaigns
- Service review calls
- Exclusive upgrade offers

- **Cohort-Specific Playbooks**

- For large cohorts → dedicated account management
- For small cohorts → automated CRM engagement

Conclusion and Recommendation

Conclusion

The CRM analytics project reveals a business with clear strengths — strong acquisition, high-performing products, and a capable core sales team. However, performance is heavily concentrated among a few cohorts, products, accounts, and agents, creating long-term risk.

The most significant challenge is the **extremely low repeat purchase rate**, indicating a gap in customer lifecycle management rather than product acceptance. Similarly, inconsistent win rates, unbalanced sales performance, and underperforming sectors reveal operational inefficiencies rather than market limitations.

By implementing focused retention initiatives, optimizing sales execution, prioritizing high-potential markets, and refining product strategy, the organization can shift from an acquisition-driven model to a **sustainable, retention-led growth engine**. This transition will improve revenue stability, reduce risk, and unlock scalable long-term performance.

Final Recommendation

- **Strengthen Customer Retention (Highest Priority)**
  - Introduce structured onboarding and post-purchase engagement workflows.
  - Launch renewal and upgrade programs to convert one-time buyers into repeat buyers.
  - Implement automated follow-ups between Month 0–2, where most customer drop-off occurs.
- **Optimize Sales Performance**
  - Provide focused coaching for agents with low win rates and long sales cycles.
  - Reassign accounts to balance workload across agents.
  - Introduce incentive programs tied to conversion rates, not just activity volume.
- **Improve Pipeline Quality & Conversion**
  - Implement opportunity scoring to prioritize high-probability deals.
  - Audit lost-deal reasons to identify gaps in pricing, pitch, or product fit.
  - Strengthen follow-up cadence in mid-to-late pipeline stages.
- **Product Portfolio Strategy**
  - Double down on the GTX Pro and GTX Plus lines — your strongest revenue and win-rate drivers.
  - Reposition lower-performing SKUs (MG & Basic series) with revised pricing or bundled offers.
  - Evaluate whether premium products like GTK 500 need targeted marketing or a revised value proposition.
- **Account Segmentation & Market Focus**
  - Prioritize high-performing sectors such as Retail, Technology, and Software.
  - Develop targeted strategies for underperforming sectors (Finance, Entertainment).
  - Expand in regions with strong adoption while developing localized strategies for weak markets.
- **Executive Monitoring & Analytics**
  - Track key health indicators: retention curve, repeat transactions, cohort revenue, agent win rate, and product profitability.
  - Use dashboards to identify early warning signals in opportunity leakage and agent performance.

THANK YOU!