# **Capstone Sales – End-to-End Project Documentation**

# **Project Overview**

The **Sales Funnel Analytics Capstone Project** provides a structured, data-driven solution for monitoring and optimizing the lead-to-deal conversion process. By integrating CRM datasets into a clean pipeline, the project enables sales leadership to:

* Gain visibility into funnel health.
* Identify inefficiencies and drop-off points.
* Compare rep and regional performance.
* Automate reporting for stakeholders.
* Generate actionable recommendations that improve conversions and shorten sales cycles.

The **Sales Funnel Analytics** initiative aligns business and technical objectives:

* Provides structured CRM data integration.
* Enables funnel visualization with conversion tracking.
* Detects performance drop-offs at critical funnel stages.
* Automates monthly and quarterly reporting.
* Recommends actionable improvements for reps and regions.

## **Objectives**

* Map the lead-to-deal sales funnel.
* Track and visualize funnel performance metrics.
* Identify bottlenecks, drop-off points, and performance variances.
* Automate sales reporting for business stakeholders.
* Enable data-driven strategic decisions.

## **Dataset Snapshot**

All project data is stored in **CRM\_Capstone\_DB.sqlite** and mirrored into CSV/Excel files for analysis.

### **Tables & Row Counts**

****Companies → 1963 rows

Deals → 2,000 rows

Leads → 2,000 rows

Regions → 6 rows

Sales\_Reps → 20 rows

Stage\_History → 18,778 rows

### **Data Dictionary (CSV)**

[CRM\_Capstone.csv](https://docs.google.com/spreadsheets/d/1vqTGF7l329VdGab8GNQttpF-Rk8h1_7Z/edit?usp=sharing&ouid=100845518774373604113&rtpof=true&sd=true)

### **Sample Data Extracts (CSV)**

Companies

Deals

Leads

Regions

Sales\_Reps

Stage\_History



## **Funnel Stages**

The funnel follows these stages:

Lead → Contacted → Demo → Proposal → Deal Won / Deal Lost

### **SQL Query – Stage Counts**

****SELECT stage, COUNT(\*) AS stage\_count

FROM Stage\_History

GROUP BY stage

ORDER BY stage\_count DESC;



## **Tech Stack**

* **Google Sheets** – Data staging.
* **Google Apps Script** – Automation of ingestion/export.
* **SQLite** – Local CRM database.
* **BigQuery** – Cloud analytics warehouse.
* **Python (pandas, matplotlib, seaborn)** – ETL, EDA, reporting.
* **Looker Studio** – Interactive dashboarding.

## **Exploratory Data Analysis (EDA)**

### **Lead Outcomes**

****import sqlite3, pandas as pd

conn = sqlite3.connect("CRM\_Capstone\_DB.sqlite")

leads = pd.read\_sql("SELECT Outcome FROM Leads", conn)

leads["Outcome"].value\_counts()

**Output:**

****Deal Lost 840

Deal Won 560

Ongoing - Cold 562

Ongoing - Active 38

 The majority of leads end up lost. Only ~28% conversion to wins.

### **Lead Sources**

****src = pd.read\_sql("SELECT Source, COUNT(\*) as cnt FROM Leads GROUP BY Source ORDER BY cnt DESC", conn)

print(src.head(5))

**Output:**

****Instagram Ads 128

LinkedIn Ads 112

Cold Calling 105

Lists 100

Display Ads 100

 Digital campaigns dominate.

### **Regional Distribution**

****region = pd.read\_sql("SELECT R.Region\_Name, COUNT(\*) as cnt FROM Leads L JOIN Regions R ON L.Region\_ID=R.Region\_ID GROUP BY R.Region\_Name", conn)

print(region)

**Output:**

****North America 488

Europe 393

APAC 379

Latin America 301

Middle East 235

Africa 204

NA drives the highest volume, APAC more efficiently.

### **Sales Reps**

****reps = pd.read\_sql("SELECT Rep\_Name, Deals\_Closed, Win\_Rate FROM Sales\_Reps", conn)

print(reps.head())

**Output (sample):**

****John Smith 50 72%

Emily Brown 45 69%

David Lee 15 18%

...

Large variance between top and bottom reps.

## **Funnel Analysis**

### **Conversion Rates**

****funnel = pd.read\_sql("SELECT stage, COUNT(\*) as cnt FROM Stage\_History GROUP BY stage", conn)

print(funnel)

**Output (sample):**

****Lead 2000

Contacted 1300

Demo 800

Proposal 400

Deal Won 560

Deal Lost 840

Sharpest drop-off: **Demo → Proposal**.

### **Funnel Visualization**

****import matplotlib.pyplot as plt

funnel.set\_index("stage").sort\_values("cnt").plot(kind="barh")

plt.title("Sales Funnel Stage Distribution")

plt.show()

**Output:** Horizontal bar chart showing attrition across funnel stages.

## **Rep & Regional Performance**

### **Rep Leaderboard**

****SELECT Rep\_Name, Deals\_Closed, Win\_Rate

FROM Sales\_Reps

ORDER BY Deals\_Closed DESC, Win\_Rate DESC;

**Insight:** Top 3 reps dominate funnel wins. Bottom reps underperform significantly.

### **Regional Heatmap**

****import seaborn as sns

sns.heatmap(region.pivot\_table(values='cnt', index='Region\_Name'), annot=True)

plt.title("Regional Lead Distribution")

plt.show()

**Insight:** NA and EU are hot zones; Africa & ME lag.

## **Database & BigQuery Integration**

### **SQLite → BigQuery Export**

****from google.colab import auth

from google.cloud import bigquery

# Authenticate

auth.authenticate\_user()

# Load SQLite data to DataFrame

leads\_df = pd.read\_sql("SELECT \* FROM Leads", sqlite3.connect("CRM\_Capstone\_DB.sqlite"))

# Upload to BigQuery

dataset\_id = "crm\_capstone\_dataset"

project\_id = "your\_project\_id"

client = bigquery.Client(project=project\_id)

job = client.load\_table\_from\_dataframe(leads\_df, f"{project\_id}.{dataset\_id}.Leads")

job.result()

**Output:** Leads table uploaded to BigQuery.

### **Looker Studio Dashboards**

* Funnel Stage Conversion
* Rep Leaderboard
* Regional Heatmap
* Lead Source Efficiency

## **Automation Notebook (Automation.ipynb)**

* **Monthly Sales Report** → Auto-generated CSV/Excel with funnel summary.
* **Quarterly Report** → Rep & regional comparison charts.

### **Example Code**

****import datetime

report\_month = datetime.date.today().replace(day=1)

summary = leads.groupby("Outcome").size()

summary.to\_csv(f"sales\_report\_{report\_month}.csv")

**Output:** CSV with aggregated sales outcomes for the last month.

## **Capstone\_sales.ipynb Results – Execution Log**

### **Part 1: Environment Setup & Data Generation**

* Drive mounted successfully, dependencies installed.
* 2000 synthetic leads created with outcomes, budgets, industries.
* Lead outcomes distribution confirmed (Won 560, Lost 840, Ongoing 600).
* Regional distribution confirms NA & EU dominate, Africa & ME lowest.
* 1963 companies generated; 20 reps added; 6 regions metadata established.

### **Part 2: Funnel Stage Analysis**

* Funnel schema applied: Lead → Contacted → Demo → Proposal → Deal/Lost.
* Stage counts verified; sharpest drop between Demo → Proposal.
* Conversion rates calculated; Demo → Proposal is a major bottleneck.
* Funnel visualization created with bar/step plot.

### **Part 3: Rep & Regional Performance**

* Rep leaderboard confirms top 3 reps >70% win rate; bottom <20%.
* NA highest deal volume, APAC most efficient conversions.
* Regional heatmap confirms imbalance across regions.

### **Part 4: Database & BigQuery**

* SQLite DB created with Leads, Companies, Regions, Reps.
* SQL queries verified ingestion.
* Exported to BigQuery; dashboards prepared.

## **EDA Extended Log**

### **Part 1: Data Setup & Import**

* Imported libraries, loaded CRM\_Capstone\_DB.sqlite.
* Verified Leads, Companies, Regions, Reps tables.

### **Part 2: Dataset Overview**

* Leads contain ID, company, source, budget, stage, outcome.
* Companies include industry, size, ownership, location.
* Regions have IDs, names, timezone, heads.
* Reps include ID, region, deals closed, win rate.

### **Part 3: Lead Data Exploration**

* Outcome: majority lost, ~28% won.
* Region: NA + EU highest, Africa & ME lowest.
* Sources: Digital-heavy (Instagram, LinkedIn Ads).

### **Part 4: Funnel Stage Exploration**

* Sharpest drop between Demo → Proposal.
* Funnel imbalance confirmed visually.

### **Part 5: Sales Rep Analysis**

* Wide rep performance gap.
* Top reps >70% win rate; bottom <20%.

### **Part 6: Company Insights**

* Industry distribution: Retail, Software, Healthcare lead.
* Budget distribution skewed low, few high-value deals.

## **Key Insights & Recommendations**

* **Funnel Weakness** → Demo → Proposal is the major bottleneck.
* **Rep Gap** → Heavy reliance on few top reps; coaching needed.
* **Regional Trend** → NA high volume, APAC efficient, Africa/ME lagging.
* **Lead Sources** → Paid digital ads dominate; diversify channels.
* **Budgets** → Mostly small-mid; pursue larger enterprise deals.

## **Project Structure**

****CAPSTONE\_FINAL/

├── Generated\_Data/ # Generated CRM data

├── Automation.ipynb # Automation/reporting notebook

├── Copy of Capstone\_sales.ipynb # Analysis notebook (backup)

├── EDA.ipynb # Exploratory analysis

├── CRM\_Capstone\_DB.sqlite # SQLite database

├── CRM\_Capstone\_Data\_Dictionary.csv

├── Sales Performance Reports/ # Reports (monthly/quarterly)

├── creds.json # Auth credentials for BigQuery

