Project 1: Retail - Customer Behavior & RFM Analysis

➢ Project Title

Consumer360 – Retail Analytics & Customer Churn Intelligence System

1. Project Overview

- Consumer360 is an end-to-end retail analytics project developed during my internship at **Infotact Solutions**. The project converts raw retail transaction data (100,000+ rows) into actionable business insights using SQL, Python, and Power BI.
- The system performs customer behavior analysis using **RFM segmentation**, identifies
- **churn-risk customers**, and provides interactive dashboards for sales and customer
- intelligence. The pipeline is automated using Python and Task Scheduler to simulate a real-world analytics workflow.

2. Objectives

- Analyze retail sales and customer data:
  - The first objective of the project is to systematically analyze raw retail transaction data to understand customer purchasing behavior and business performance. This involves examining sales trends, order frequency, revenue contribution, and product demand patterns.
  - Through structured querying and aggregation, the organization can identify:
    - Who buys most frequently
    - Which regions generate higher revenue
    - What products drive profitability
    - Seasonal or monthly sales variations

- Segment customers using RFM model:
  - A core goal of the solution is to divide customers into meaningful groups using the Recency, Frequency, Monetary (RFM) framework.
    - Recency measures how recently a customer made a purchase.
    - Frequency measures how often they purchase.
    - Monetary measures how much money they spend.
  - Each customer is assigned a score (commonly 1–5) for R, F, and M.
    These scores are combined to create segments such as:
    - Champions
    - Potential Loyalists
    - At Risk / Hibernating

- Identify churn-risk customers:
  - Using RFM outputs, customers with:
    - Low recency (haven't purchased recently)
    - Low frequency
    - Decreasing spending
    - are flagged as churn risks.
  - The system helps the business:
    - Detect early warning signs

- Launch retention offers
- Trigger re-engagement campaigns
- Prioritize intervention for high-value customers

➢ Build interactive BI dashboards:

♦ The project aims to transform analytical outputs into **easy-to-understand visual insights** for business users.
♦ Power BI dashboards will include:

- Sales performance trends
- Revenue by geography
- Top products
- Distribution of RFM segments
- Cohort retention analysis

➢ Implement Row Level Security (RLS):

♦ Since the organization operates across multiple regions, it is essential to restrict data visibility.
♦ **Row Level Security (RLS)** ensures that:

- Regional managers only see customers from their assigned territory
- Sensitive financial or customer data is protected
- Reports remain secure while still being accessible

➢ Automate RFM pipeline using Python:

♦ To make the solution production-ready, the analytics workflow must run automatically without manual intervention.
♦ Automation includes:

- Scheduled data extraction from SQL
- Automatic RFM calculation
- Segment refresh
- Output file generation
- Dashboard update

❖ Tools & Technologies

- Python (Pandas)
- SQL (PostgreSQL)
- Power BI

- DAX
- Star Schema Modeling
- RFM Analysis
- Task Scheduler Automation
- Row Level Security

---

3. Dataset Description

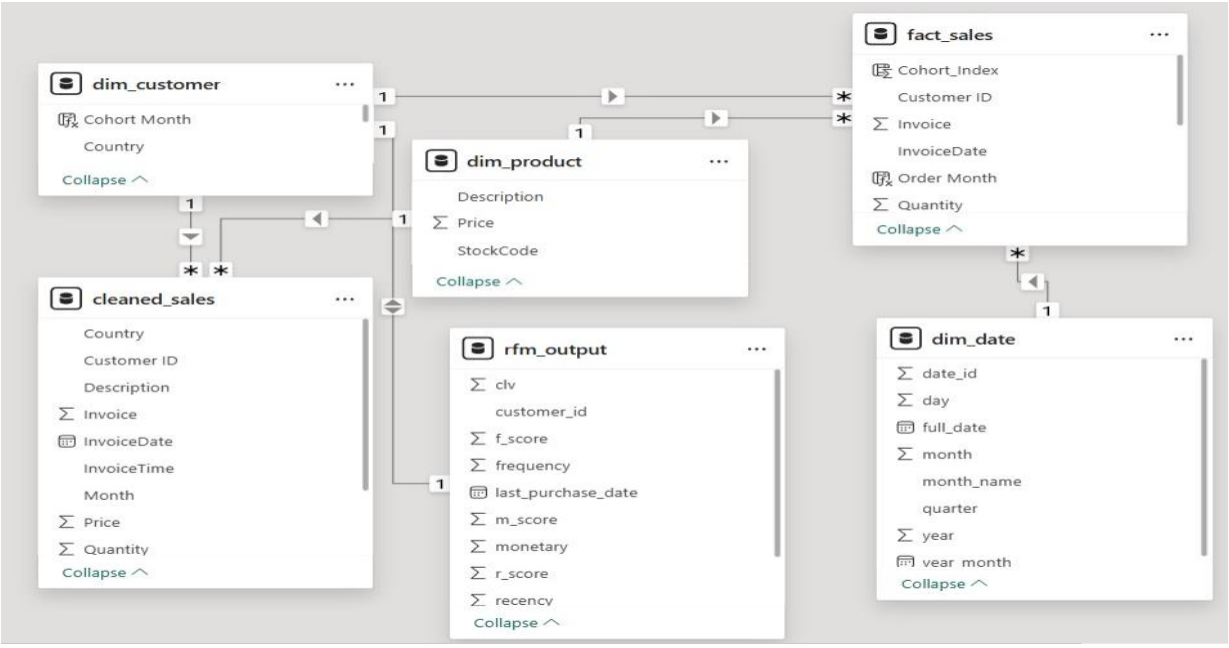
- Source: Retail transaction dataset
- Size: ~100,000 rows
- Fields: Orders, Customers, Products, Amount, Region, Dates
- Data Type: CSV → PostgreSQL → Power BI

---

4. Data Modeling (SQL Layer)

- A **Star Schema** model was created in PostgreSQL

- Fact Table: Sales
    - Dimension Tables: Customer, Product, Region
    - Relationships created for analytics optimization
- This structure improves query performance and dashboard clarity.



5. Data Processing & RFM Analysis (Python)

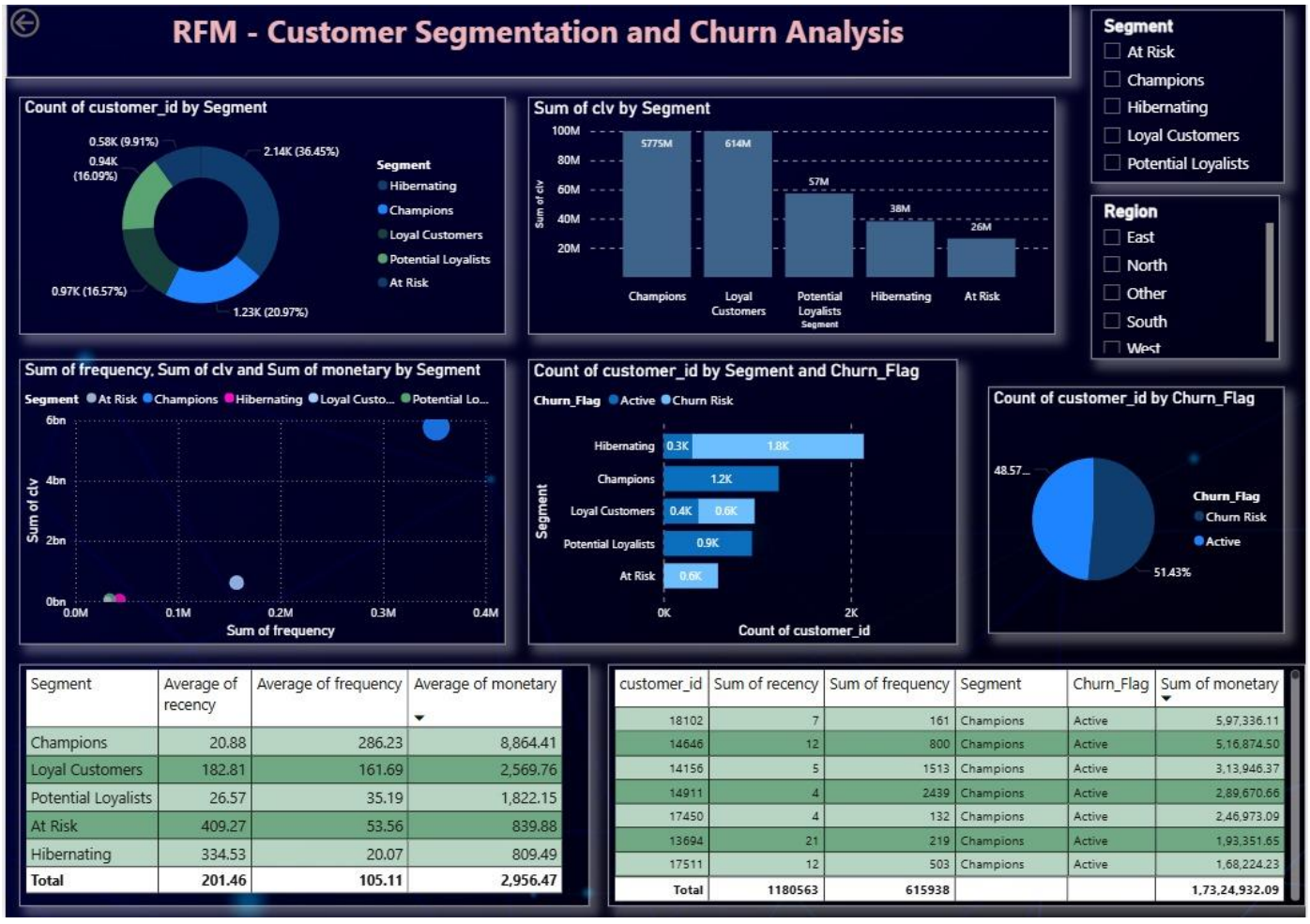
- Python was used to:
    - Clean and standardize columns
    - Convert date fields
    - Calculate Recency, Frequency, Monetary
    - Generate RFM scores
    - Create customer segments:
        - Champions
        - Loyal
        - At Risk
        - Regular
- Identify churn-risk customers
- Export processed file for Power BI

```
rfm.head()
rfm.describe()
```

	customer_id	last_purchase_date	frequency	monetary	recency
count	5860.000000	5860	5860.000000	5860.000000	5860.000000
mean	15316.832423	2011-05-17 12:55:46.894198016	105.108874	2956.473053	201.461263
min	12346.000000	2009-12-01 00:00:00	0.000000	2.900000	1.000000
25%	13836.750000	2010-11-23 00:00:00	18.000000	344.492500	26.000000
50%	15315.500000	2011-08-25 00:00:00	48.000000	883.595000	102.000000
75%	16799.250000	2011-11-09 00:00:00	125.000000	2284.647500	377.000000
max	18287.000000	2011-12-04 00:00:00	2982.000000	597336.110000	734.000000
std	1714.531194	NaN	165.150724	14352.382490	207.496838

6. Power BI Dashboard

- Interactive dashboards were created with:
    - KPI cards (Orders, Customers, Sales)
    - Sales by Category C Region
    - Customer Segmentation charts
    - RFM analysis visuals
    - Churn risk distribution
    - Drill-down (Country → Region)
    - Slicers C filters
    - DAX measures
    - Row Level Security (RLS)

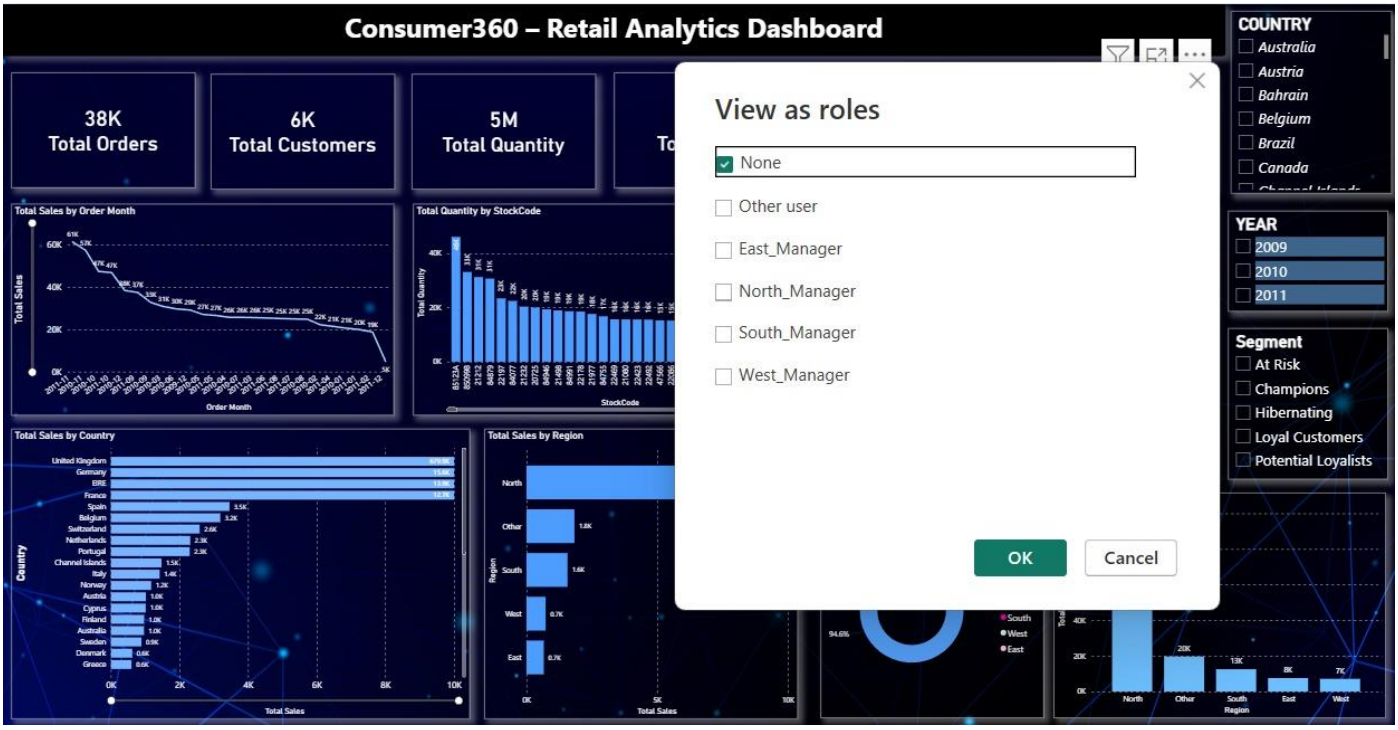


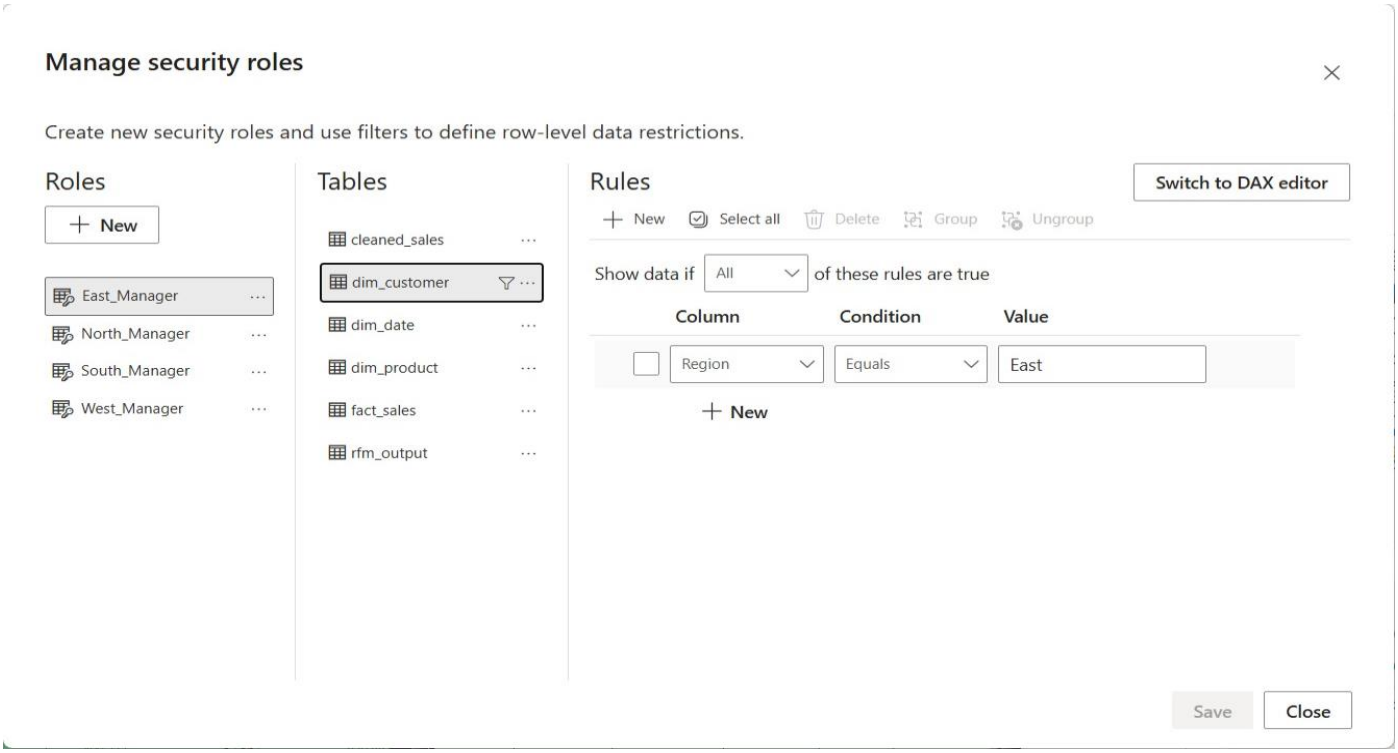
Segment	Average of recency	Average of frequency	Average of monetary
Champions	20.88	286.23	8,864.41
Loyal Customers	182.81	161.69	2,569.76
Potential Loyalists	26.57	35.19	1,822.15
At Risk	409.27	53.56	839.88
Hibernating	334.53	20.07	809.49
Total	**201.46**	**105.11**	**2,956.47**

customer_id	Sum of recency	Sum of frequency	Segment	Churn_Flag	Sum of monetary
18102	7	161	Champions	Active	5,97,336.11
14646	12	800	Champions	Active	5,16,874.50
14156	5	1513	Champions	Active	3,13,946.37
14911	4	2439	Champions	Active	2,89,670.66
17450	4	132	Champions	Active	2,46,973.09
13694	21	219	Champions	Active	1,93,351.65
17511	12	503	Champions	Active	1,68,224.23
Total	**1180563**	**615938**			**1,73,24,932.09**

7. Row Level Security (RLS)

- RLS was implemented to restrict dashboard access based on region.
  Different users can view only their assigned country data.





8. Automation Pipeline

- Automation was implemented using:
    - Python ETL script (rfm_pipeline.py)
    - Automated RFM recalculation
    - Output file refresh
    - Windows Task Scheduler job
    - Log file generation for execution tracking
- This ensures updated analytics without manual effort.

---

9. Key Insights

- High revenue comes from limited customer segments (Champions C Loyal)
- At-Risk customers identified using RFM recency patterns
- Category sales concentration visible
- Region-wise performance differences
- Customer retention opportunities identified

---

10. Business Impact

- Helps target high-value customers
- Supports churn prevention strategies
- Enables data-driven marketing
- Reduces manual reporting via automation
- Improves decision-making visibility

---

11. Business Impact

- Helps target high-value customers
- Supports churn prevention strategies
- Enables data-driven marketing
- Reduces manual reporting via automation
- Improves decision-making visibility
