Capstone Retail Business Analysis

-

Capstone Project

Objective :

To analyze sales, customer behavior, and profitability to generate actionable insights that improve decision-making.

6 Layers Workflow

Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6

Capstone Project

Click Here :

GitHub

Tools Used :

Data Structuring : MySQL, CSV, MySQL Workbench

SQL Layer : Joins, Aggregations, RFM

Python Layer : Pandas, Matplotlib, Seaborn

Excel Layer : Cleaning, formatting

Visualization : Power BI

Insight Report : AI,/ PPT

Capstone Project

Layer 1 : Data Collection & Storage

- Setting Jupiter Notebook
- establish three main tables Customers, Products, Sales
- creating Unique columns and setting tables
- making connection with mysql using alchemy engine
- testing connection using engine.connect
- using superset dataset from kaggle
- Import .csv or .xlsx files into SQL database

Tools used : Jupiter (pandas, mysqlconnector, alchemy)

Layer 2 : SQL Layer

- Setting up database and tables
- Changing and creating new organised date column
- Join data from multiple tables
- Aggregate sales monthly/yearly
- Identify top/bottom performing products
- Segment customers by RFM (Recency, Frequency, Monetary)

Tools used : mysql, joins)

Layer 3 : Python

- ETL: Clean data, fill NAs, normalize values
- Perform EDA:
- Sales over time
- Profit by product/region
- Correlation heatmaps
- Outlier detection

Tools used : Jupiter (pandas, mysqlconnector, alchemy)

Layer 4 : Excel

- Data standardization
- Import summary tables into Excel
- Pivot tables
- Conditional formatting
- Financial metrics
- Setting columns with correct formats

Tools used : MsExcel, pivot tables, formatting

Layer 5 : Visualization Layer

- KPI cards (Revenue, Profit, Customer Retention)
- Filters by region, month, category
- Customer segment visuals
- Drill-down charts (Region → Store → SKU)

Tools used : Power Bi, Dax, power query

Layer 5 : Insights Generation

- Which regions are underperforming?
- Are discounts reducing profit?
- Who are the most valuable customers?
- What inventory should be optimized?

Tools used : SQL, Excel, Power Bi, Dax

Capstone Project

Underperforming Region Analysis

SQL Output with Code

```sql
select region, sum(sales) as total_sales, sum(profit) as total_profit,
sum(profit)/nullif(sum(sales),0) as profit_margin
from sales s
join customers c on s.customer_id=c.customer_id
group by region
order by total_profit desc;
```

region	total_sales	total_profit	profit_margin
West	5306522.057299955	682912.6673999871	0.12869307995441825
East	4522990.0624999255	552491.1379999905	0.12215174704465749
Central	3853691.953099923	510508.7983999942	0.13247265339652775
South	2706437.5374999535	321804.96159999905	0.11890352433452551

Although the West region leads in total profit and sales, South region is underperforming with the lowest profit and profit margin (0.1189), despite generating over 2.7M in sales. Strategic intervention (cost reduction, pricing optimization) is needed in the South.

Capstone Project

Most Valuable customers

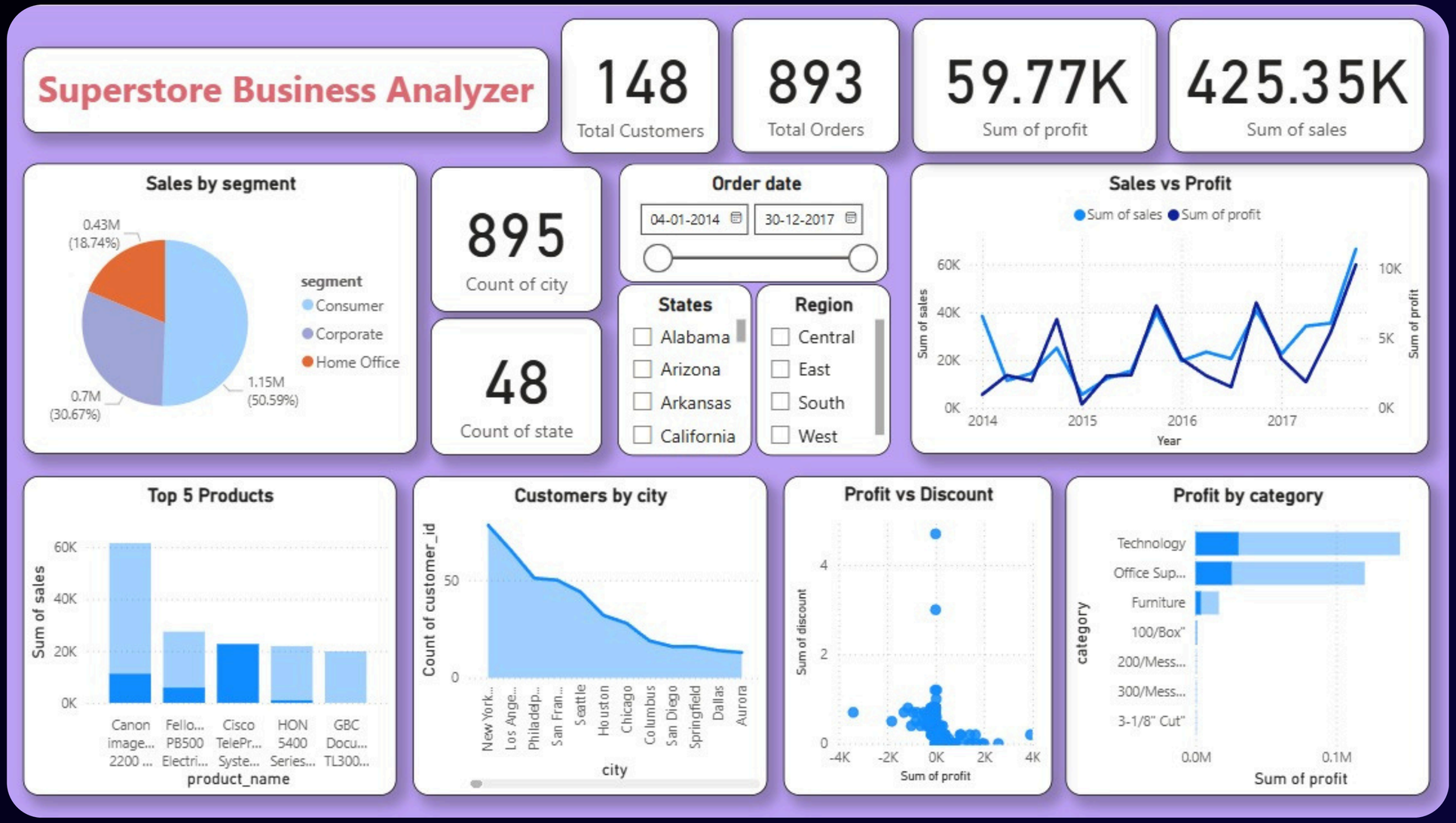
```sql
select customer_id, max(order_date) as last_order,
count(order_id) as frequency,
sum(sales) as monetary
from sales
group by customer_id
order by monetary desc;
```

customer_id	last_order	frequency	monetary
SM-20320	8/21/2015	15	25043.05
TC-20980	9/20/2015	12	19052.21799999999
RB-19360	9/25/2017	18	15117.339
TA-21385	10/22/2017	10	14595.62
AB-10105	9/25/2016	20	14473.57099999998
KL-16645	7/31/2016	29	14175.229
SC-20095	9/23/2014	22	14142.33399999999
HL-15040	6/24/2016	11	12873.29799999999
SE-20110	9/22/2014	19	12209.43800000000
CC-12370	8/15/2016	11	12129.072

Customers like SM-20320, TC-20980, and RB-19360 are high-value customers based on their total purchase value, even if their order frequency is lower than others. These should be prioritized for loyalty or upselling campaigns.

Capstone Project

BI Dashboard



Thank You