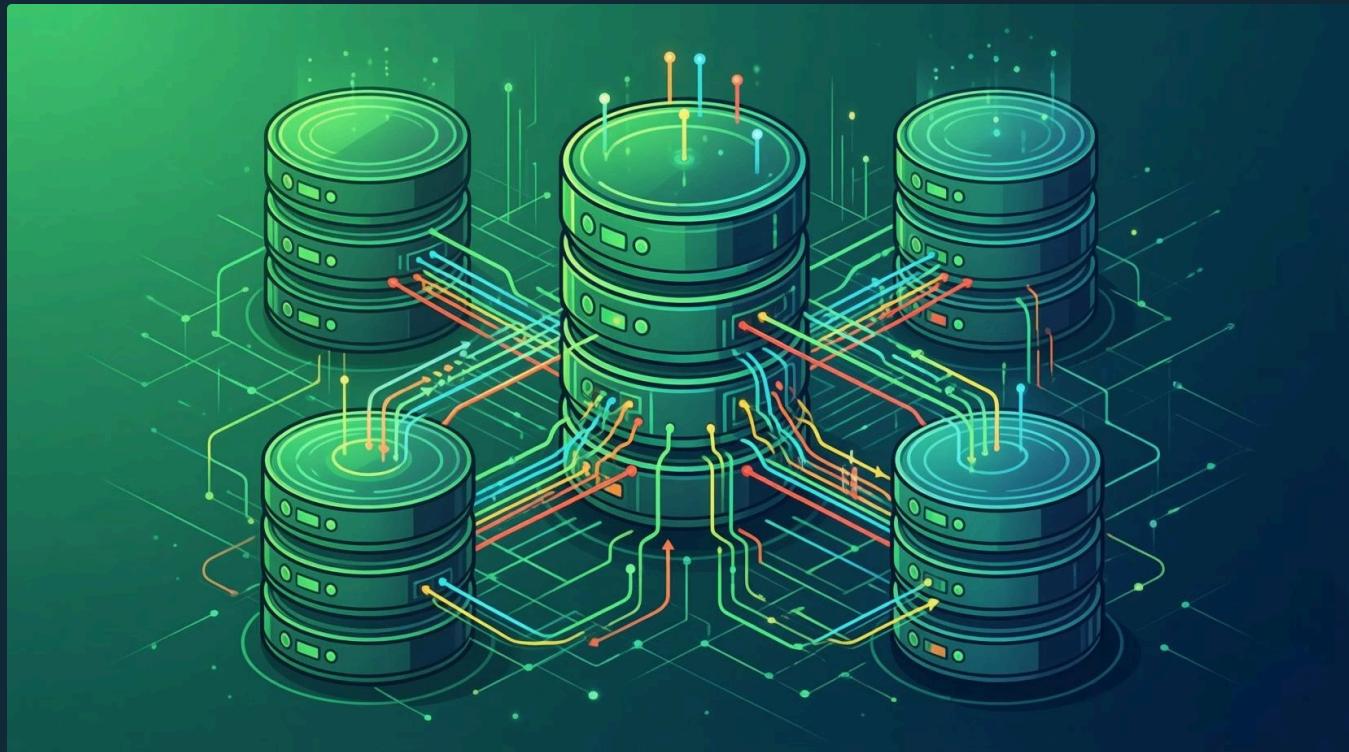




Target Dataset Exploration: An SQL and Python Approach

Mastering e-commerce data analysis through practical SQL queries and Python visualization techniques

Navigating the E-Commerce Data Landscape



Our comprehensive e-commerce dataset consists of seven interconnected files that capture the complete customer journey from browsing to purchase completion.

O1

Customer Demographics

Location and profile data

O2

Orders & Items

Purchase history and details

O3

Products & Sellers

Inventory and vendor information

O4

Payments & Location

Transaction and geographic data

Dataset Files Overview



customer s.csv

Customer demographic details including location, city, and state information for segmentation analysis



orders.cs v

Complete order history with timestamps, status tracking, and customer relationships



products. csv

Product catalog with categories, dimensions, and pricing information



payments .csv

Payment details including installments, methods, and transaction values



sellers.csv

Seller information with location and performance metrics



geolocati on.csv

Geographic coordinates for precise location-based analysis and mapping

Basic SQL Queries: Foundation Skills

1

Unique Customer Cities

Extract distinct cities from customer data

```
SELECT DISTINCT customer_city  
FROM customers;
```

2

Orders Count by Year

Filter orders from 2017 using date functions

```
SELECT COUNT(*) FROM orders  
WHERE YEAR(order_date)=2017;
```

3

Sales by Category

Join tables and aggregate product sales

```
SELECT p.category, SUM(oi.price)  
FROM order_items oi  
JOIN products p  
ON oi.product_id=p.product_id  
GROUP BY p.category;
```

Building Your SQL Foundation

These fundamental queries teach essential SQL operations: filtering with WHERE clauses, extracting unique values with DISTINCT, and performing basic aggregations with COUNT and SUM.

Master these patterns to understand data distribution, count records efficiently, and calculate payment installment percentages. Each query builds toward more complex analysis techniques.



↗ INTERMEDIATE LEVEL

Intermediate Queries: Advanced Aggregations

Monthly Order Trends

Track order patterns across 2018 by extracting month from order dates and grouping results

```
SELECT  
MONTH(order_date),  
COUNT(*)  
FROM orders  
WHERE  
YEAR(order_date)=2018  
GROUP BY  
MONTH(order_date);
```

Revenue Distribution

Calculate each category's percentage of total revenue using subqueries and mathematical operations

```
SELECT  
p.category,  
SUM(oi.price)  
)*100/  
(SELECT  
SUM(price)  
FROM  
order_items)  
FROM  
order_items  
oi  
JOIN  
products p  
GROUP BY  
p.category;
```

Seller Performance Ranking

Identify top-performing sellers with revenue totals and competitive rankings using window functions

```
SELECT  
seller_id,  
SUM(price),  
RANK()  
OVER(ORDER  
BY  
SUM(price)  
DESC)  
FROM  
order_items  
GROUP BY  
seller_id;
```



ADVANCED LEVEL

Advanced SQL: Window Functions & Analytics



Mastering Window Functions

Advanced queries leverage powerful SQL window functions for sophisticated time-series analysis. Calculate moving averages to smooth customer order trends, track cumulative monthly sales growth, and measure year-over-year performance changes.

These techniques enable retention analysis by identifying repeat customers within 6-month windows and ranking top spenders by year using `PARTITION BY` and `RANK` functions.

Python Analysis Setup



Getting Started with Python

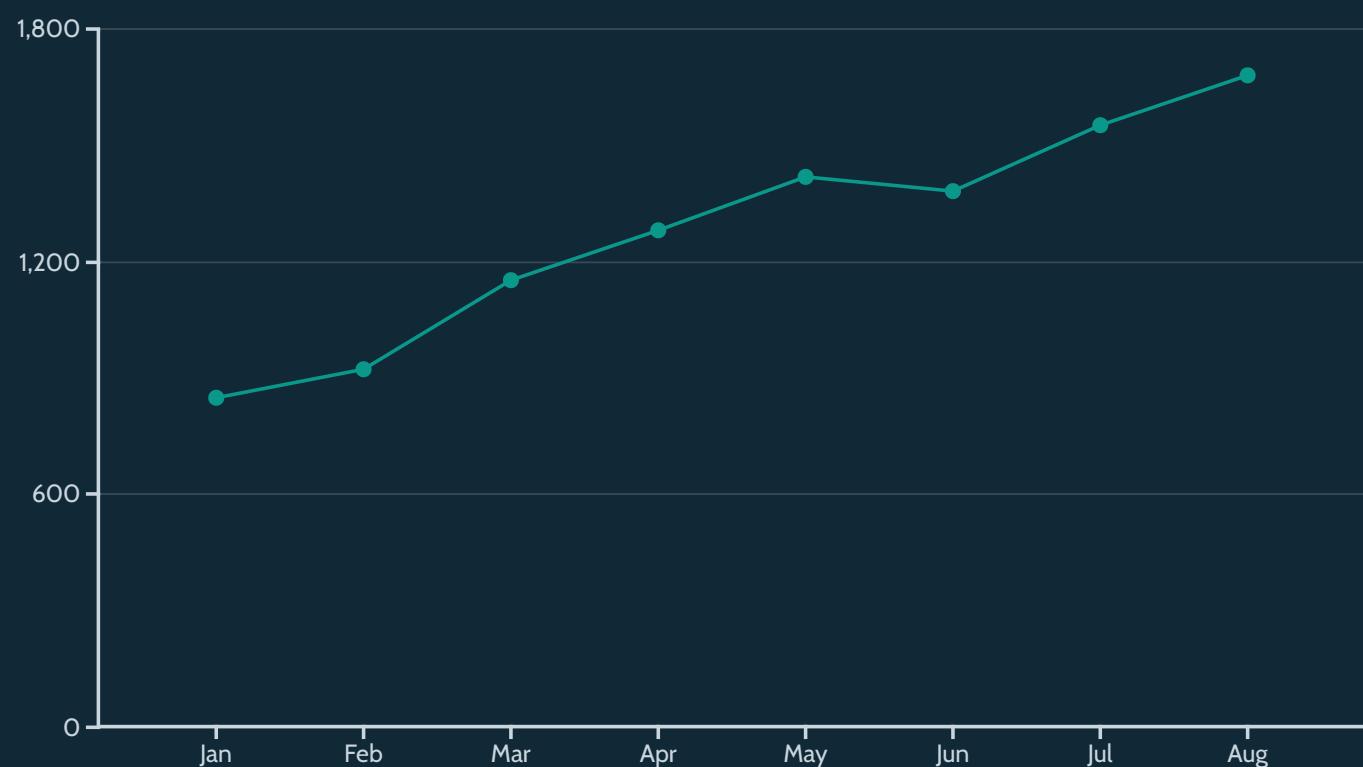
Load all dataset files efficiently using pandas and prepare your environment for comprehensive data analysis and visualization

```
import pandas as pd
import matplotlib.pyplot as plt

# Load all dataset files
customers =
pd.read_excel("customers.xlsx")
orders = pd.read_excel("orders.xlsx")
order_items =
pd.read_excel("order_items.xlsx")
products =
pd.read_excel("products.xlsx")
payments =
pd.read_excel("payments.xlsx")
sellers = pd.read_excel("sellers.xlsx")
geolocation =
pd.read_csv("geolocation.csv")
```

Pro tip: Convert order timestamps to datetime format immediately after loading for seamless date-based filtering and analysis throughout your workflow.

Python Data Analysis: Key Insights



Visualizing Monthly Trends

Python's matplotlib library transforms raw order data into clear visual insights. This line chart reveals steady growth in monthly orders throughout 2018, with a notable acceleration during summer months.

```
orders_2018['month'] =  
  
orders_2018['order_purchase_timestamp'].dt.month  
monthly_orders = orders_2018.groupby('month').size()  
plt.plot(monthly_orders.index, monthly_orders.values)
```

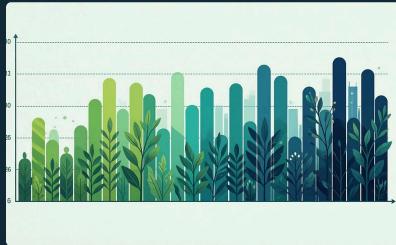
Advanced Python Analytics



Price-Purchase Correlation

Analyze the relationship between product pricing and purchase frequency using pandas aggregation and correlation methods to optimize pricing strategies

```
price_analysis = order_items  
    .groupby('product_id')  
    .agg(purchase_count=  
        ('product_id','count'),  
        avg_price=('price', 'mean'))
```



Top Seller Performance

Identify and visualize the top 10 revenue-generating sellers to understand marketplace dynamics and partner performance patterns

```
seller_revenue = (order_items  
    .groupby('seller_id')['price']  
    .sum().sort_values(ascending=False)  
    .head(10))
```



Category Sales Analysis

Merge order items with product data to calculate total sales by category, revealing which product types drive the most revenue

```
merged =  
    order_items.merge(products,  
        on='product_id', how='left')  
category_sales = merged  
    .groupby('product_category')  
    ['price'].sum()
```



Your E-Commerce Analysis Toolkit



SQL Mastery

15 progressive queries from basic SELECT statements to advanced window functions for comprehensive data extraction



Python Visualization

Pandas and matplotlib techniques for transforming raw data into actionable business insights through charts and analysis



Business Intelligence

Practical analytics for revenue tracking, customer retention, seller performance, and trend identification in e-commerce

This comprehensive guide equips you with the essential skills to explore, analyze, and visualize e-commerce datasets effectively. Start with basic queries, progress through intermediate techniques, and master advanced analytics to drive data-informed business decisions.