بازار

Launchpad 2024 - Analytics Case Study

Case Study Brief:

Welcome to the Analytics case study! In this exercise, you will be working with a sample dataset representing sales transactions from a retail company. Your task is to demonstrate your SQL querying, analytical thinking, and data visualization skills to derive insights and recommendations for improving business performance.

Dataset Description:
The **attached dataset** contains information about sales transactions including order date, order number, warehouse ID, store ID, item ID, amount per unit, ordered quantity, item discount, and expected delivery date.

Tasks:

1. Data Exploration 30 minutes):

    - Write SQL queries for the following tasks and provide a brief summary of key findings including:
    - Total number of orders ● Total sales revenue
    - Average order quantity
    - Distribution of orders by warehouse and store
    - Top selling items

Queries:

```
-- Total number of orders

SELECT COUNT(DISTINCT order_number) AS total_orders

FROM sales_transactions;
```

بازار

```
-- Total sales revenue

SELECT SUM(amount_per_unit * ordered_quantity * (1 - item_discount)) AS total_revenue

FROM sales_transactions;


-- Average order quantity

SELECT AVG(ordered_quantity) AS avg_order_quantity

FROM sales_transactions;


-- Distribution of orders by warehouse and store

SELECT warehouse_id, store_id, COUNT(DISTINCT order_number) AS order_count

FROM sales_transactions

GROUP BY warehouse_id, store_id;


-- Top selling items

SELECT item_id, SUM(ordered_quantity) AS total_quantity_sold

FROM sales_transactions

GROUP BY item_id

ORDER BY total_quantity_sold DESC

LIMIT 5;
```

بازار

2. Analytical Questions 60 minutes):

- Answer the following analytical questions using SQL queries:
    1. What is the overall discount rate (average discount per item sold)?
    2. Which warehouse has the highest average order value (total sales revenue divided by number of orders)?
    3. What is the total revenue generated by each store?
    4. Identify the top 5 customers (based on total amount spent).
    5. Calculate the month-over-month growth rate of sales revenue.
    6. Determine the percentage of orders that were canceled.

Queries:

```
-- 1. Overall discount rate (average discount per item sold)

SELECT AVG(item_discount) AS avg_discount_rate

FROM sales_transactions;



-- 2. Warehouse with the highest average order value

SELECT warehouse_id, SUM(amount_per_unit * ordered_quantity * (1 - item_discount)) AS
total_revenue, COUNT(DISTINCT order_number) AS order_count, (SUM(amount_per_unit *
ordered_quantity * (1 - item_discount)) / COUNT(DISTINCT order_number)) AS
avg_order_value

FROM sales_transactions

GROUP BY warehouse_id

ORDER BY avg_order_value DESC

LIMIT 1;
```

بازار

```
-- 3. Total revenue generated by each store

SELECT store_id, SUM(amount_per_unit * ordered_quantity * (1 - item_discount)) AS
total_revenue

FROM sales_transactions

GROUP BY store_id;



-- 4. Top 5 customers based on total amount spent

SELECT customer_id, SUM(amount_per_unit * ordered_quantity * (1 - item_discount)) AS
total_amount_spent

FROM sales_transactions

GROUP BY customer_id

ORDER BY total_amount_spent DESC

LIMIT 5;



-- 5. Month-over-month growth rate of sales revenue

WITH monthly_revenue AS (

  SELECT DATE_TRUNC('month', order_date) AS month, SUM(amount_per_unit *
ordered_quantity * (1 - item_discount)) AS revenue

  FROM sales_transactions

  GROUP BY DATE_TRUNC('month', order_date)
```

بازار

ORDER BY month

)

SELECT month, revenue,

(revenue - LAG(revenue, 1) OVER (ORDER BY month)) / LAG(revenue, 1) OVER (ORDER BY month) AS mom_growth_rate

FROM monthly_revenue;

-- 6. Percentage of orders that were canceled

SELECT ROUND(100.0 * SUM(CASE WHEN status = 'Canceled' THEN 1 ELSE 0 END) / COUNT(*), 2) AS percent_canceled

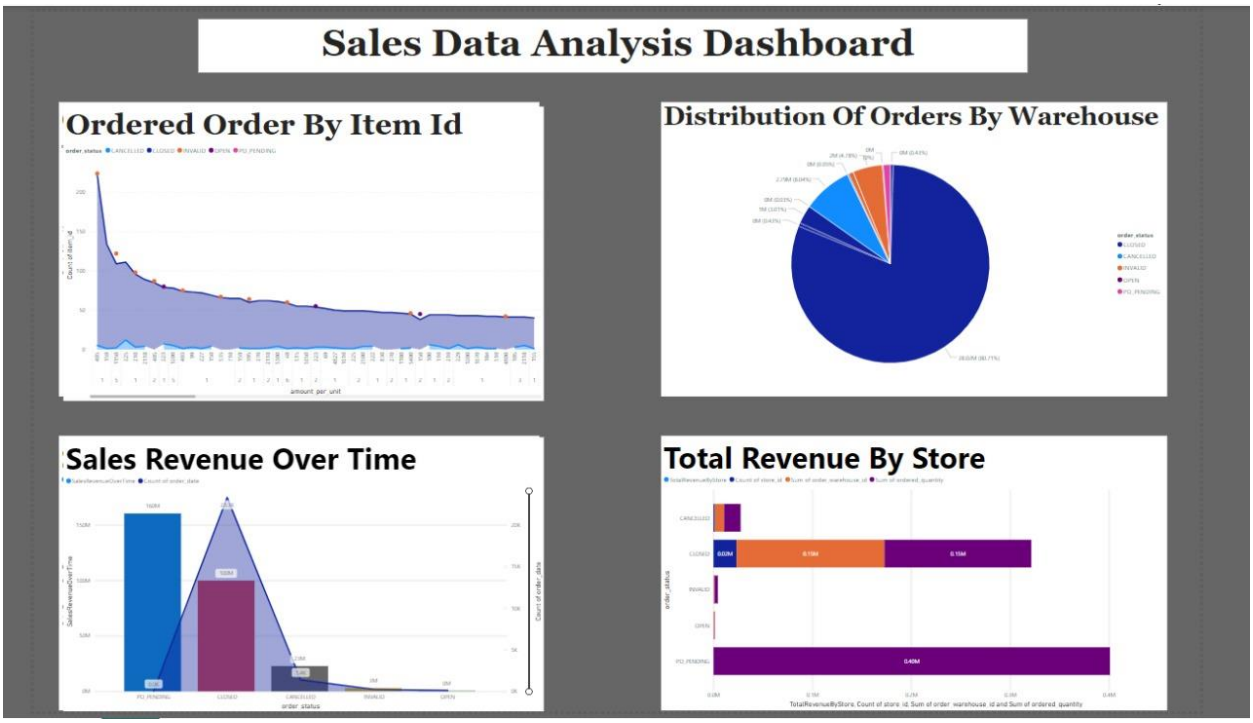
FROM sales_transactions;

3. Data Visualization 90 minutes):

- Choose a data visualization tool (e.g., Tableau, Power BI, or Python libraries like Matplotlib or Seaborn).
- Create visualizations to present insights from the data exploration and analytical questions.
- Visualizations should include:
- Time series plot of sales revenue over time.
- Bar chart showing total revenue by store.
- Pie chart illustrating the distribution of orders by warehouse.
- Any additional visualizations you deem relevant to showcase insights.

بازار

Sales Data Analysis Dashboard By Using Power BI:



Sales Data Analysis By Using Python (MatplotLib & Seaborn):

[9]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

df = pd.read_csv('/kaggle/input/new-sales-data/Sales_Data_New.csv')
df.head()
```

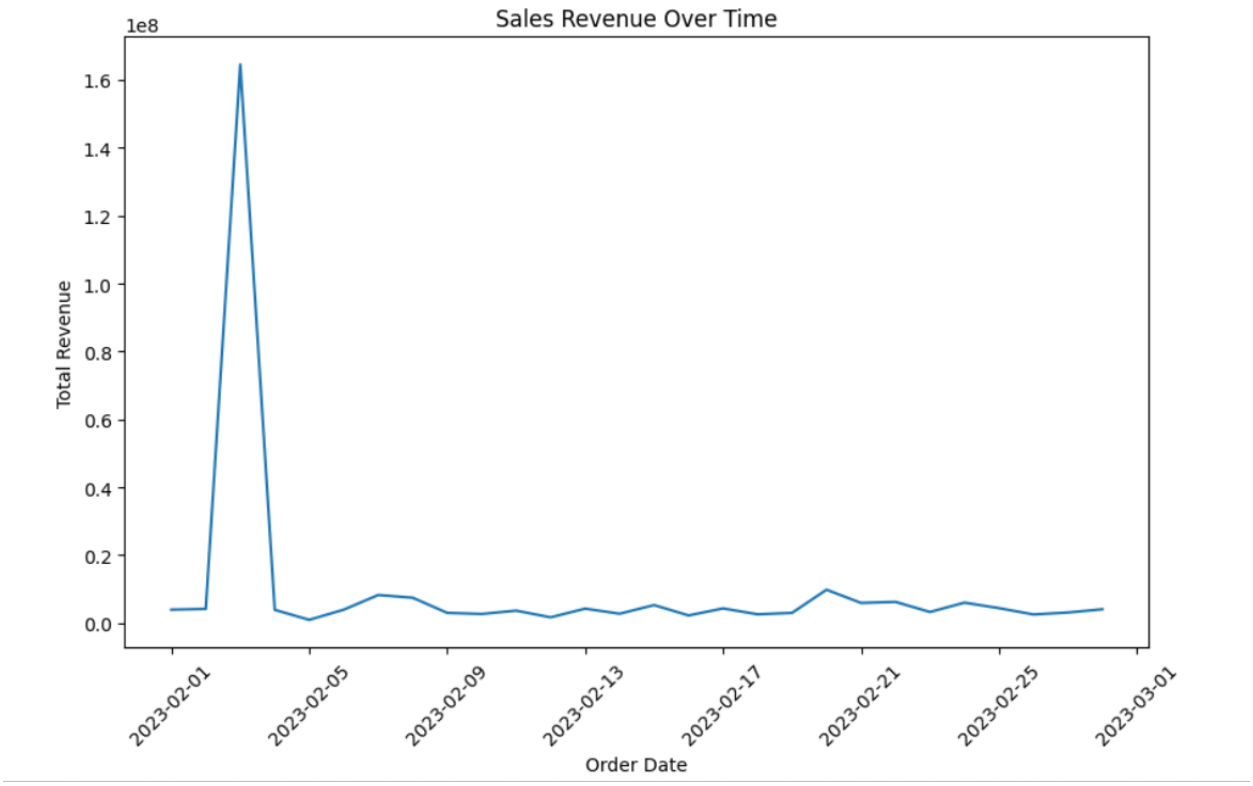
[9]:

	order_date	order_number	order_warehouse_id	store_id	order_status	item_id	amount_per_unit	ordere
0	14/02/2023	23045429724	9.0	pY85faVkR6OCPZQ0CefuH	CLOSED	P437048134760065132	281.0	
1	19/02/2023	23050463841	21.0	6FTIDJ65jt1arkhJm2bbmL	CLOSED	P5801744203932	75.0	
2	20/02/2023	23051024482	4.0	7JQvKYYMMCSzfni42DKw5a	CLOSED	P810622842427393343	5420.0	
3	06/02/2023	23037635371	4.0	6tjBYLZoWt70PUXokHPEit	CANCELLED	P957287950450245066	2600.0	
4	15/02/2023	23046508076	15.0	1w0FMk0J5K43sIqSZv9b5t	CLOSED	P862686710671848219	85.0	

بازار

Sales Revenue Over Time

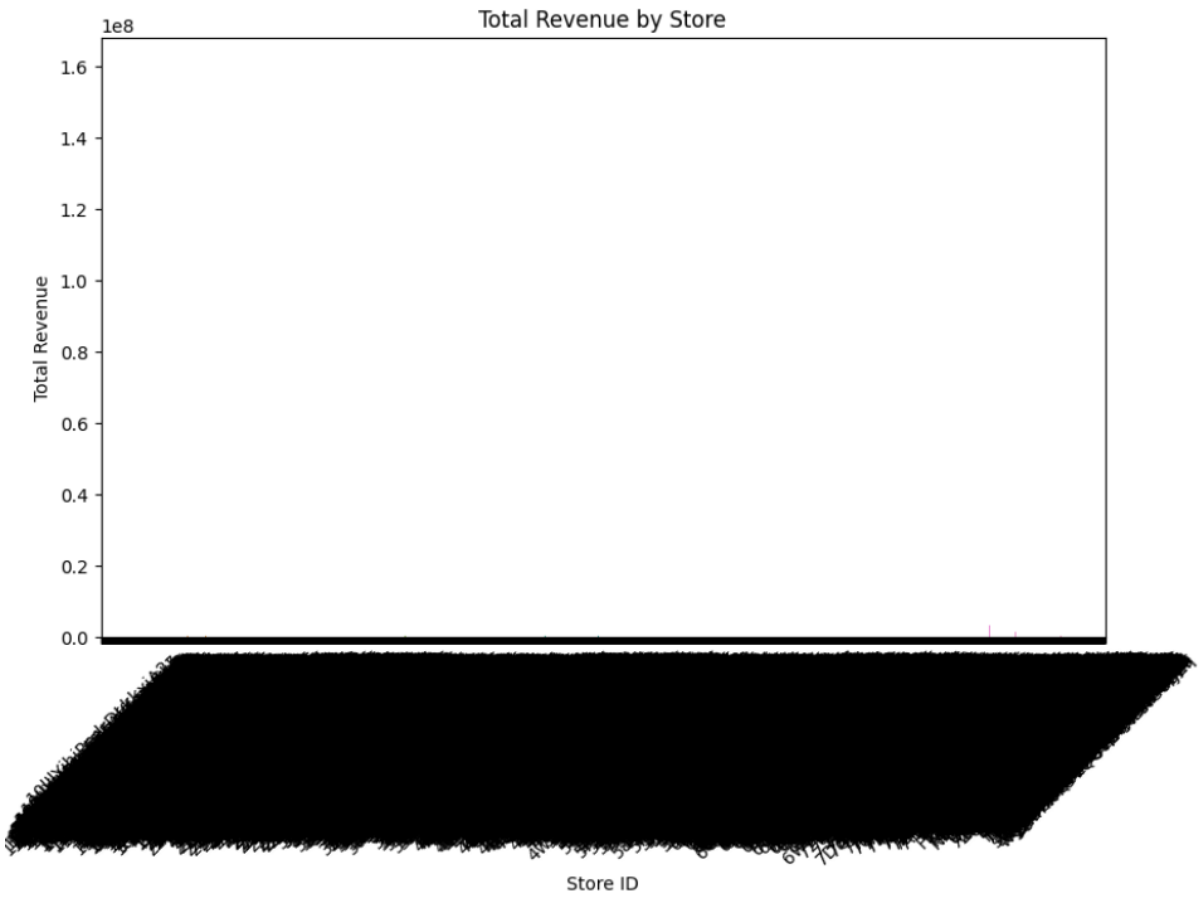
```
# Assuming your data is stored in a DataFrame called 'df'

# Sales Revenue Over Time
df['order_date'] = pd.to_datetime(df['order_date'])  # Convert order_date to datetime
df['total_revenue'] = df['amount_per_unit'] * df['ordered_quantity']  # Calculate total revenue f
revenue_over_time = df.groupby('order_date')['total_revenue'].sum().reset_index()  # Group by ord
plt.figure(figsize=(10, 6))
sns.lineplot(data=revenue_over_time, x='order_date', y='total_revenue')
plt.title('Sales Revenue Over Time')
plt.xlabel('Order Date')
plt.ylabel('Total Revenue')
plt.xticks(rotation=45)
plt.show()
```



بازار

Total Revenue by Store

```
# Total Revenue by Store
total_revenue_by_store = df.groupby('store_id')['total_revenue'].sum().reset_index()  # Group by store_id an
plt.figure(figsize=(10, 6))
sns.barplot(data=total_revenue_by_store, x='store_id', y='total_revenue')
plt.title('Total Revenue by Store')
plt.xlabel('Store ID')
plt.ylabel('Total Revenue')
plt.xticks(rotation=45)
plt.show()
```

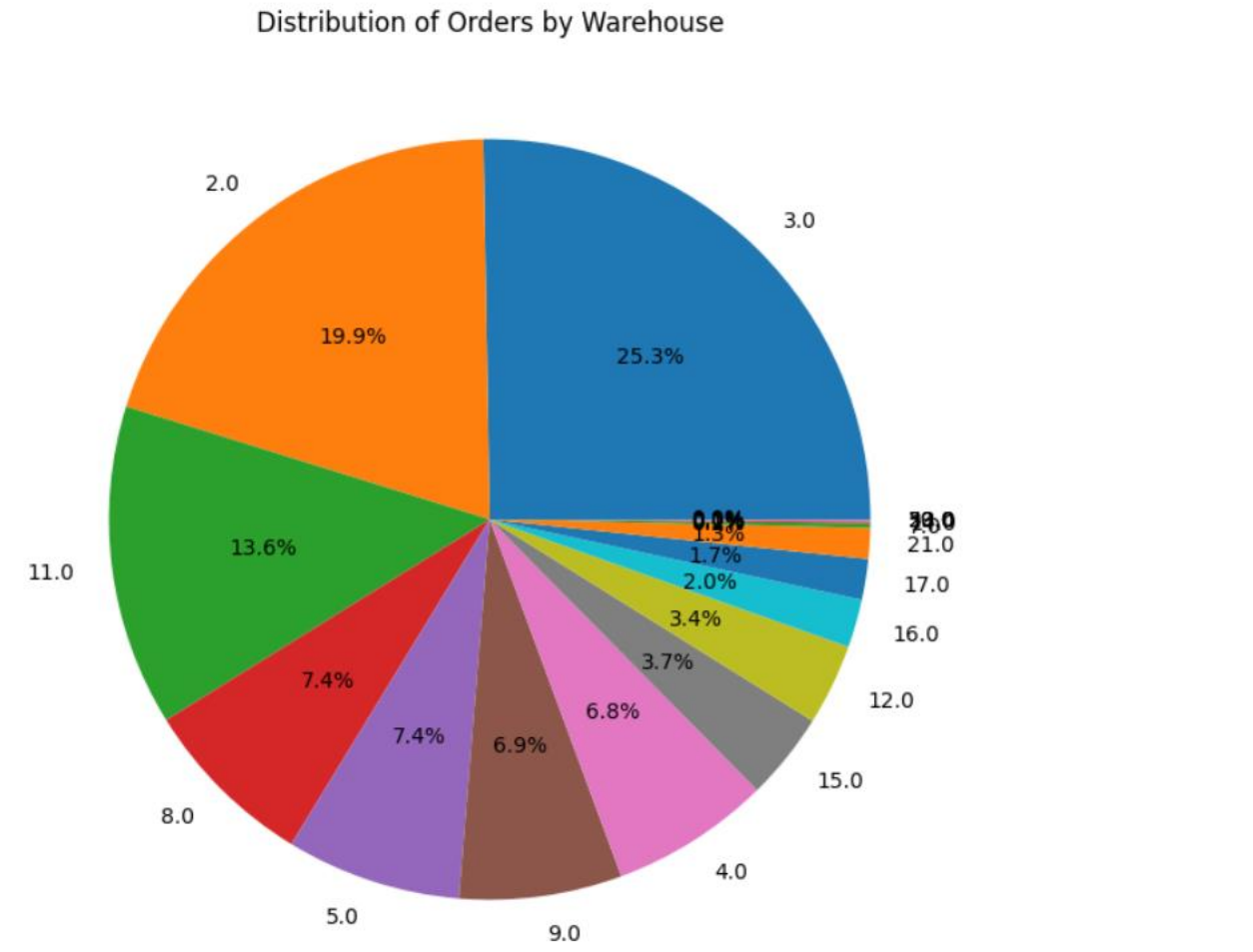


بازار

Distribution of Orders by Warehouse

```
# Distribution of Orders by Warehouse
orders_by_warehouse = df['order_warehouse_id'].value_counts().reset_index()  # Count the number of orders for each warehouse
orders_by_warehouse.columns = ['Warehouse ID', 'Order Count']
plt.figure(figsize=(8, 8))
plt.pie(orders_by_warehouse['Order Count'], labels=orders_by_warehouse['Warehouse ID'], autopct='%1.1f%%')
plt.title('Distribution of Orders by Warehouse')
plt.show()
```

Distribution of Orders by Warehouse



بازار

4. Insights and Recommendations 30 minutes):

- Based on your findings from the data exploration, analytical questions, and visualizations, provide insights into the business performance.
- Suggest actionable recommendations for the company to optimize sales and improve overall performance.
- Consider factors such as warehouse/store performance, customer behavior, and sales trends.

Insights:

The total number of orders, sales revenue, and average order quantity provide an overview of the business performance.

The distribution of orders by warehouse and store helps identify high-performing and low-performing locations.

The top-selling items can guide inventory management and promotional strategies.

The overall discount rate and the warehouse with the highest average order value highlight pricing and operational efficiency.

The total revenue generated by each store identifies top-performing and underperforming stores.

The top customers based on total amount spent represent valuable segments for targeted marketing and loyalty programs.

The month-over-month growth rate of sales revenue reveals sales trends and potential seasonality.

The percentage of canceled orders indicates potential issues with order fulfillment or customer satisfaction.

بازار

Recommendations:

Optimize inventory levels and replenishment strategies for top-selling items to ensure availability and maximize sales.

Analyze the factors contributing to the success of high-performing warehouses and stores, and implement best practices across all locations.

Review pricing strategies and consider adjusting discounts or promotions based on the overall discount rate and customer behavior.

Develop targeted marketing campaigns and loyalty programs for the top customers to retain and further engage these valuable segments.

Investigate the reasons behind canceled orders and implement measures to improve order fulfillment and customer satisfaction.

Monitor sales trends and seasonality patterns to plan for inventory, staffing, and promotional activities accordingly.

Continuously analyze data and adjust strategies based on evolving customer preferences and market conditions.

بازار

Conclusion:

Congratulations on completing the case study! We hope you found it challenging and insightful. This exercise tested your SQL querying, analytical thinking, and data visualization skills, which are crucial for success in the field of analytics. We look forward to reviewing your solutions and discussing them further during the interview process.

Submission Instructions:

Please submit your SQL queries, analysis, visualizations, and insights/recommendations in a presentation format. Ensure your submission is clear, well-organized, and includes all relevant details.

If there are any questions and concerns, please feel free to email us at **shafaq.masood@bazaartech.com or duaa.kashif@bazaartech.com**. Good luck with the process!