OVERVIEW OF THE PROJECT

Here's a quick overview of the datasets:

1. **List of Orders**:
- Columns: Order ID, Order Date, CustomerName, State, City
- Contains customer and order-level details.
2. **Order Details**:
- Columns: Order ID, Amount, Profit, Quantity, Category, Sub-Category
- Provides detailed information about the items in each order, including financial metrics like Amount and Profit.
3. **Order Details**:
- Columns: Month of Order Date, Category, Target
- Contains sales targets for various categories over time.

OBJECTIVES

Customer segmentation using SQL

Perform customer segmentation by querying and analyzing data to:

- Identify customer purchase patterns.
- Group customers into segments
- Extract actionable customer groups for targeted marketing strategies.

Using Power Bi

Visualize customer segments to:

- Display key metrics.
- Create interactive dashboards showing geographical trends (State/City) of customer segments.
- Highlight top customers, products, and categories contributing to each segment.

Using Python for Sales Forecasting

- Aggregate sales data over time to identify trends and seasonality.
- Train and evaluate forecasting models (e.g., ARIMA, Prophet Model).
- Provide actionable insights for category-specific sales planning.

REPORT SUMARRY

SQL

Data Cleaning, Normalization and Transformation

Data Column (Orderdate) was cleaned and reformatted into sql date format

Count of rows in each table:

Sales Target Table: Has 3 rows, 3 columns.

Order Details: Has 1500 rows, 6 columns.

List of Orders: Has 500 rows and 6 columns.

Checking for Null Values:

In the three tables none has empty cell.

Merging related table together

Merging list of orders and order details together. Creating connections in the two tables for easy access by adding primary and foreign keys.

Customer segmentation using SQL

Question:

1. **Identify customer purchase patterns**. (analyzing customer name, total quantity ordered, sum of total purchase, customer's purchase start day and the last purchase day).

Insights:

1. Customers who recently purchased, buy frequently, and spend significantly. Prioritize them with loyalty programs.

2. Customers who are frequent but slightly less recent buyers. Offer targeted promotions to re-engage.

Output: from the analysis result, Abhishek has the highest quantity ordered of (89) and has order id placed in 5 times. Abhishek first ordered date started on '2018-03-12' and his last ordered date's '2018-09-24' which gives insights that he has ordered items from the company in three months 12 days using his first ordered and last ordered date.

From the result 'Yaanvi' is the customer with the highest Total purchase amount in the company, his first ordered date started on '2018-01-05' and his last purchased date is '2019-10-03'. I total he has 71 Total quantity ordered. Total profit made on his purchase is 488.

'Ashwin' has the lowest purchase amount from the result, he has 2 total quantity ordered, his first ordered date is '2018-03-08' and last ordered date is '2018-03-08'.

2. Group customers into segments

Customer is grouped by Totalamount and given a threshold of <5000 ‘low sales’, >5000 ‘high sales’.

Insights: Marketing Analysis strategies.

Output: From the result we have

- High Sales Customers: These are valuable customers who contribute significantly to revenue. They are prime targets for loyalty programs and premium offers.
- Low Sales Customers: These customers spend less and may require engagement strategies like discounts or personalized recommendations to increase their spending.

3. Extract actionable customer groups for targeted marketing strategies.

Segment Customers by Total Spending (e.g., Low, Medium, High Sales)

Segment Customers by Frequency of Purchases

Segment Customers by Product Categories Purchased

Insights: Gain insights into customer behavior, helping in strategic planning and inventory management.

Output: From the result we have three segment for the customers purchase which are:

- **Low Spenders**: These customers contribute minimally to revenue. They might be new customers or occasional buyers. To increase their spending, you could provide them with introductory offers, discounts, or bundled deals.
- **Medium Spenders**: This group represents an opportunity for upselling or cross-selling. They might respond well to personalized recommendations or loyalty programs.
- **High Spenders**: These are your most valuable customers. Retaining them is critical. Offering them VIP rewards, early access to sales, or premium services can ensure their loyalty.

USING PYTHON FOR SALES FORECASTING

Question:

1. Aggregate sales data over time to identify trends and seasonality.

Insights: Aggregating sales data over time provides actionable insights into sales trends and seasonality, enabling businesses to optimize operations, forecast demand, and align strategies with customer behavior. By aggregating sales data (e.g., monthly or yearly), you can identify whether sales are increasing, decreasing, or remaining constant over time.

Output: From the result, 2018 vs. 2019:

- In 2018, total monthly sales consistently exceeded 5,000 for most months, with peaks in January, August, and October.
- In 2019, total monthly sales dropped significantly, especially in the later months (e.g., December had only 259 in sales).
  Insight: There's a significant decline in sales performance from 2018 to 2019, which suggests either:
- A reduction in customer demand.
- External factors such as competition or economic downturn.

PEAKS AND TROUGHS
2018 Peaks:

- January (18,035) and August (20,331) are the strongest months.
- Suggests demand surges early in the year and during the late summer.

2018 Troughs:

- February (6,566) and September (4,861) reflect relatively weaker sales periods.

2019 Peaks:

- April (11,079) and October (14,147) are the highest-performing months.
- Indicates some recovery in the second quarter and fall of 2019.

2019 Troughs:

December (259) and January (1389) reflect relatively weaker sales periods in 2019.

A line chart is created to display the output.

2. Train and evaluate forecasting models (using Prophet Model).

Output:

Chart Interpretation

Deep Blue Line (Predicted Sales):

This line represents the forecasted values for both historical and future data. The model uses past sales data and identifies trends and seasonality to predict sales for the next 6 months.

Sky Blue Shaded Area (Uncertainty Intervals):
This represents the confidence intervals around the forecast. A wider area indicates that the model has lower confidence in the future predictions for those periods. The narrower the area, the higher the certainty about the forecast.

The model considers factors like volatility, seasonality, and historical patterns, and the uncertainty reflects how much the model expects the future to vary from its forecast.

Key Insights:

- If the deep blue line is trending upwards or downwards, that gives an idea of the expected direction of sales (growth or decline).
- Wider uncertainty intervals suggest that the model is less certain about the future sales in those periods, so there might be fluctuations in sales due to unpredictable factors or volatile trends.
- If the sky blue area is narrow and close to the deep blue line, the model is confident that the predicted sales will fall within a small range.

A new file is created to showcase the previous and forecasted sales.

Forecasted Values for 6 Interpretation

- **ds**: This column represents the date for each prediction.
- **yhat**: This is the predicted value of sales for that month.
- **yhat_lower**: The lower bound of the confidence interval, which represents the minimum expected sales value.
- **yhat_upper**: The upper bound of the confidence interval, representing the maximum expected sales value.

Insights from the Predicted Values

The forecast shows negative sales values for the first four months (January to April), indicating a decline in sales. The large negative values in January, March, and February may suggest specific issues such as low demand**,** inventory shortages, or external economic factors (e.g., seasonal fluctuations). However, there is a turnaround in May, with a positive sales forecast, followed by moderate growth in June. This indicates that the market could recover, potentially due to factors like marketing efforts**,** new product launches, or seasonal demand.

Actionable Insights:

- For the months of January to April, you might want to prepare for a decline in sales. This could involve cost-cutting measures, revising marketing strategies, or adjusting production plans.
- For May and June**,** the predicted recovery suggests you might want to plan for a rebound, adjusting your strategy to leverage any improving market conditions.
- Due to the high uncertainty in the forecast (wide confidence intervals), it is important to stay flexible with plans and continuously monitor actual sales to make adjustments as needed.

3. Provide actionable insights for category-specific sales planning.

Actionable Insights:

- **High-Performing Categories**: Categories with the highest total sales are clear opportunities for focus, marketing, and resource allocation.
- **Seasonal Trends**: The line chart helps to identify categories with seasonal peaks or troughs, enabling us to plan inventory and promotions.

- **Predicted Sales**: The forecast model (Prophet) provides an outlook for the next 6 months.
- **Underperforming Categories**: Categories with low sales or declining trends may need attention, such as new marketing campaigns or adjustments in pricing.

POWER BI INTERPRETATION

Visualizing Top 10 customer with the highest sales:

Chart type: Bar chart

Interpretation: From the result, top 10 customers was filtered and Yaanvi happens to be the customer with the highest sales while from the top 10 Sarita is the customer with the lowest sales.

Visualizing Total sales by State:

Chart type: Map chart

Interpretation: Map chart is used to visualize total sales by state to generate state with the highest sales.

Visualizing Total sales by Category

Chart type: Bar chart

Interpretation: From the result Electronic has highest sales followed by clothing while Furniture has lowest sales.

Visualizing Category by profit by Quantity

Chart type: Ribbon chart

Interpretation: From the result, Clothing has highest profit and quantity sold while Furniture has lowest profit and quantity sold.

Slicers were added to the dashboard for filtering records and the data added are: category, state, city for easy filtering.

Card chart was also added to summarize total amount, total profit, total quantity sold, total category and total sub categories.

From the result, the total sales made was 432,000, 24,000 profit, 5615 quantity sold, 3 categories, 17 sub categories were observed.

Summary

The analysis was performed on datasets containing order-level details, product information, and sales targets to achieve three objectives: customer segmentation, sales forecasting, and visualizing key metrics. Using SQL, customer purchase patterns were identified, and customers were segmented into groups based on spending and frequency. Python was utilized to forecast sales trends, uncovering significant seasonality and declining trends. Power BI visualizations highlighted top customers, states, and product categories, providing insights into performance across regions and segments.

Key Insights

1. **Customer Segmentation (SQL)**:

- Customers like Abhishek and Yaanvi are high-value contributors due to frequent and significant purchases.
- Low spenders, such as Ashwin, contribute minimally and require engagement strategies to improve their contribution.
- Segmentation identified three groups: Low Spenders, Medium Spenders, and High Spenders, each requiring tailored marketing strategies.

2. **Sales Forecasting (Python)**:

- Sales peaked in early 2018 but showed a consistent decline by 2019, with significant drops in December.
- Forecasting (using the Prophet model) indicated negative sales for the initial months of the next period, followed by recovery in May and June.
- Seasonal peaks were observed in January and August of 2018, guiding promotional planning.

3. **Visualization of Key Metrics (Power BI)**:

- Yaanvi emerged as the top customer by sales, while Sarita had the lowest among the top 10.
- Electronics was the top-performing category by sales, while Furniture lagged in both profit and quantity sold.
- Geographical analysis identified top-performing states, aiding targeted regional strategies.

Conclusion

The analysis highlights the variability in customer spending behavior, category performance, and geographical contributions. Sales trends showed significant seasonality, while forecasting suggests potential challenges in the near term. High spenders drive a majority of revenue, emphasizing the need to nurture this segment. The visualizations provide a comprehensive understanding of business performance, highlighting key opportunities and challenges.

Recommendations

1. **Customer Retention and Engagement**:
- Reward loyal customers with personalized loyalty programs and exclusive offers.
- Re-engage low spenders with targeted promotions, discounts, and introductory offers.
2. **Sales Growth Strategies**:
- Focus marketing efforts on Electronics, leveraging its strong performance.
- Address underperforming categories like Furniture through pricing adjustments or better promotion.
3. **Seasonal Planning**:
- Align marketing campaigns and inventory with seasonal peaks (e.g., early-year and late-summer surges).
- Prepare for the forecasted decline in early months by optimizing resources and exploring cost-cutting measures.
4. **Regional Optimization**:
- Prioritize high-performing states for regional campaigns.
- Explore opportunities in underperforming states through localized promotions.
5. **Forecast Monitoring**:
- Continuously monitor actual sales against forecasted values to refine strategies.
- Use insights from predicted recovery to plan for mid-year growth opportunities.

These insights and recommendations can help drive data-driven decision-making, improving customer satisfaction and overall business performance.