REPORT Final Project

Carerha – Data Analysis Track

Maria Achraf Sabry Attallah

1-SQL (Data exploration and manipulation)

- **Total overview:**
  Total Number of Invoices, Unique Customers, and Product Categories:
  - Total Invoices: calculates the total number of distinct orders.
  - Unique Customers: It determines the number of distinct customers.
  - Product Categories: It counts the distinct product lines/categories.

- **Customer analysis:**
  - Unique Customers by Product Line and Product Code:
  - The number of unique customers for each combination of product line and product code is identified and ordered by the number of unique customers in descending order.

- **Sales analysis:**
  Total Sales Amount:
  - The total sales amount is calculated for all orders excluding those with statuses 'Cancelled', 'Disputed', or 'On Hold'.
    Average Order Value:
  - The average order value is determined by dividing the total sales by the number of distinct orders, excluding 'Cancelled', 'Disputed', or 'On Hold' statuses.

- **Product analysis:**
  Most Frequently Purchased Items:
  - The frequency of purchases for each product line is counted and ordered in descending order of frequency.

  Purchase Patterns by Country:
  - For each country and product line, the project calculates:
  - Total number of distinct orders.
  - Total sales amount.

- Average quantity ordered.
- The results are ordered by total sales in descending order.

Number of Sub-Categories for Each Category:

- The number of distinct product codes for each product line is counted.

Most Sold Sub-Category:

- For each product line and product code, the project calculates:
- Total number of distinct orders.
- Total sales amount.
- The results are ordered by total sales in descending order.

- **Location analysis:**

Top 20 Cities by Total Sales for Each Product Line:

- For each city and product line, the project calculates:
- Total number of distinct orders.
- Total sales amount.
- Average quantity ordered.
- The results are ordered by total sales in descending order, limited to the top 20 cities.

2-Python (data cleaning, analysis, and visualization)

Data Cleaning

- Libraries: Pandas.
- Cleaned File:Loaded `sales_data_sample (1).csv`.
- Changing data type :Converted `**ORDERDATE**` to datetime format for consistency.
- Column Operations:
  - Combined `**CONTACTFIRSTNAME**` and `**CONTACTLASTNAME**` into `**CONTACTNAME**`.
  - Removed unnecessary columns (`**ADDRESSLINE2**`, `**POSTALCODE**`, `**PHONE**`, `**CONTACTFIRSTNAME**`, `**CONTACTLASTNAME**`).
  - Filled missing values in `**TERRITORY**` with 'Na' and in `**STATE**` with `**CITY**`.

- Merged `**Vintage Cars**` and `**Classic Cars**` into a single category called `**Cars**`.

Analysis and visualization

Libraries: **Pandas**, **Matplotlib** and **seaborn**

Customer Segmentation

1. **Metrics Calculation**: Aggregated data to determine purchase frequency and total spend per customer.
2. **Segmentation**: Classified customers into four segments based on median purchase frequency and total spend:
   - **High Value**: High purchase frequency and high total spend.
   - **High Frequency**: High purchase frequency and low total spend.
   - **High Spend**: Low purchase frequency and high total spend.
   - **Low Value**: Low purchase frequency and low total spend.
3. **Visualization**: Created bar plots to illustrate customer segments based on purchase frequency and total spend.

Customer Lifetime Value (CLTV)

1. **Metrics Calculation:** Computed total revenue, order count, average order value, and customer lifespan.
2. **CLTV Formula:**
   CLTV = Average Order value X Purchase Frequency X Customer Lifespan
3. **Visualization:** Created bar plots for the top 20 and bottom 20 customers by CLTV.
4. **Average CLTV:** Calculated and displayed the average CLTV for all customers.

Sales Trends Over Time

1. **Monthly and Quarterly Trends:** Aggregated sales data by month and quarter to identify trends.
2. **Visualization:** Created line plots to show sales trends over months and quarters, highlighting peak periods and seasonal variations**.**

Product Line Sales Distribution

1. **Sales Calculation:** Summarized total sales by product line.

2. **Visualization:** Created bar plots to show sales distribution across different product lines, highlighting the dominance of Cars.

Deal Size Analysis

1. **Sales by Deal Size:** Analyzed sales distribution based on deal sizes.
2. **Visualization:** Created bar plots to show how sales vary with deal sizes, emphasizing the significance of large deals.

Customer Distribution by State

1. **Customer Count:** Counted unique customers per state.
2. **Visualization:** Created bar plots to illustrate the number of customers in each state, identifying regions with high and low customer concentrations.

3-Power BI (data visualization and storytelling)

Overview Page

- **Total Sales Card**: Shows the total sales amount.
- **Number of Orders Card**: Shows the total number of orders.
- **Number of Customers Card**: Shows the total number of customers.
- **Number of Products Card**: Shows the total number of different products.
- **Growth Rate Card**: Shows the sales growth from 2004 to 2005.

DAX

```
Growth Rate =
VAR Sales2005 = CALCULATE(SUM('data'[SALES]), 'data'[YEAR_ID] = 2005)
VAR Sales2004 = CALCULATE(SUM('data'[SALES]), 'data'[YEAR_ID] = 2004)
RETURN DIVIDE(Sales2005 - Sales2004, Sales2004)
```

- **Average Order Value Card**: Shows the average sales amount per order.

DAX

```
Average Order Value = DIVIDE(SUM(data[SALES]),
COUNT(data[ORDERNUMBER]))
```

- **Bar Chart**: Shows total sales for each year.

- **Line Chart**: Shows the average order value for each year.
- **Combo Chart**: Shows the number of orders and the number of customers for each year.

Date Page

- **Line Chart**: Shows total sales by month and year.
- **Bar Chart**: Shows total sales by product line for each year.
- **Waterfall Chart**: Shows the quantity of orders for each year.
- **MTD Sales**: Shows month-to-date sales.

DAX
```
MTD Sales = TOTALMTD(SUM('data'[SALES]), 'data'[ORDERDATE])
```

- **YTD Sales**: Shows year-to-date sales.

DAX
```
YTD Sales = TOTALYTD(SUM(data[SALES]), 'data'[ORDERDATE])
```

Customer Page

- **Pie Chart**: Shows total sales by customer segments.

DAX
```
PurchaseSegment =
SWITCH( TRUE(), 'data'[PurchaseCount] > 40 , "High-frequency",
    AND('data'[PurchaseCount] > 13 , 'data'[PurchaseCount] <= 40) , "Medium-
   frequency",
    'data'[PurchaseCount] < 40 , "Low-frequency",
    "Non-purchaser"
)
```

- **Line Chart**: Shows total purchases by each customer.

DAX
```
   PurchaseCount = CALCULATE( COUNT('data'[ORDERLINENUMBER]),
      ALLEXCEPT('data', 'data'[CUSTOMERNAME])
   )
```

- **Bar Chart**: Shows average order value by customer segments.
- **Bar Chart**: Shows total sales for each customer.
- **Map**: Shows where customers are located.

Product Page

- **Doughnut Chart**: Shows average order value by product line.
- **Line Chart**: Shows total sales by product line and deal size.
- **Bar Chart**: Shows total sales by product line.
- **Annual and Quarterly Sales**: Shows total sales by product line for each year and quarter.
- **Table**: Shows the count of product codes by product line.
- **Area Chart**: Shows the quantity of orders by product line.

Location Page

- **Map**: Shows total quantities sold by country.
- **Map**: Shows total sales by country.

4-Business questions, insights and recommendations

Business Questions

1. What strategies can we implement to reverse the decline in total sales and achieve positive growth rate?
2. How can we increase the average order value and leverage the current trend?
3. What actions should we take to attract new customers while retaining existing customers?
4. Which product line we should focus on improving?
5. How can we optimize our sales strategy to improve performance in underperforming geographic regions?

Insights

1. Total sales stand at $10.03M, with a growth rate of -0.62%, indicating a slight decline.
2. The Average Order Value (AOV) is $3.55K and shows an increasing trend.
3. Sales peaked in 2004 at $4M but dropped to $2M in 2005.
4. The number of customers is 92, with a decreasing trend, suggesting retention or acquisition issues.

5. Low-frequency customers account for 62% of sales, while medium-frequency customers have the highest AOV at $4M.
6. Product performance varies, with motorcycles and cars leading in sales percentages.
7. Overall sales by product line are decreasing.
8. The number of orders is 307 and increasing, indicating growing transaction volumes.
9. Monthly and yearly sales are $457.86K MTD and $2M YTD, respectively.
10. Despite increasing order quantities, the total purchase by customers is decreasing.
11. Recommendations include focusing on customer retention, improving engagement with medium-frequency customers, optimizing the product portfolio, and addressing the declining growth rate.

Recommendations:

1.Reverse Decline in Sales:
- Diversify product offerings and expand marketing efforts.
- Improve customer experience and optimize pricing strategies.
- Strengthen sales channels.

2. Increase Average Order Value:
- Implement upselling and cross-selling techniques.
- Create product bundles and introduce a loyalty program.
- Provide personalized recommendations.

3.Attract and Retain Customers:
- Use targeted advertising and referral programs.
- Collect and act on customer feedback.
- Engage customers through email marketing and social media.

4. Focus on Product Line:
- Improve motorcycles and cars, which have the highest sales percentages.
- Invest in R&D and conduct market analysis.
- Gather and address customer feedback.

5. Optimize Sales Strategy in Underperforming Regions:
- Tailor localized marketing campaigns and form regional partnerships.
- Analyze regional data and adapt product offerings.
  Increase local presence with dedicated sales teams and distribution channels.