**Project Proposal : Analyze customer behaviors and sales trends for the purpose of enhancing marketing strategies, as well as improving business performance within a classic car dealership.**

Course : Data Analytics Case Study 1 (DAMO-501-3)

Instructor : Patty Zakaria

Members of The Group For The Course Project

Dhruv Patel

(NF1001883 – [dhruv.patel1883@myunfc.ca](mailto:dhruv.patel1883@myunfc.ca))

Vrund Patel

(NF1007109 – [vrund.patel7109@myunfc.ca](mailto:vrund.patel7109@myunfc.ca))

Heta Chavda

(NF1014555 – [heta.chavda4555@myunfc.ca](mailto:heta.chavda4555@myunfc.ca))

Dhruvi Desai

(NF1008933 – [dhruvi.desai8933@myunfc.ca](mailto:dhruvi.desai8933@myunfc.ca))

Arpit Desai

(NF1010039 – [arpit.desai0039@myunfc.ca](mailto:arpit.desai0039@myunfc.ca))

**Introduction**

This project applies data from the Classic Models database to extract meaningful insights on customer behavior, sales trend analysis, and employee performance within the setting of Classic Car Dealership. Based on this analysis of data, the dealership will be able to fine-tune its marketing strategies, increase sales of the products, and make better decisions for the business. The techniques applied in this analysis are borrowed from SQL Databases, Analytics, and Marketing Analytics.

**Chapter 1. Problem Definition and Research Question**

**Problem Definition**

The project will try to analyze customer behavior, sale trends, and key performance metrics of a classic car dealership using the Classic Models database. Such analysis will be done with the view to gain useful insights that can be helpful for major strategic decisions regarding marketing and business operations. The analysis shall seek patterns in customer transactions, the sale of products, and employee contributions that would give information applicable in strategic marketing optimization, improvement in customer engagement, and improvement of general business performance.

**Research Question**

1. How do customer demographics and marketing campaign strategies influence purchasing behavior and sales trends at the classic car dealership?
2. How does customer behavior/sales trend in the classic car dealership support the optimization of marketing strategies for further improving sales performance?

**Justification**

The basis of effective marketing campaigns and product offerings is proper understanding of consumer behavior. The analysis of sales trends and the performance metrics will enable the dealer to identify high-value customers, seasonal or product-specific demand, and employee contributions. This research question will enable the dealer to concentrate resources where returns on investment are highest to maximize profitability and operational efficiency.

**Chapter 2. Hypotheses Formulation**

**Hypothesis**

Key customer demographics, purchasing habits, seasonal trends-these are all important contributors to sales, which can be optimized by better targeting of marketing strategies and recognition of high-performing sales reps.

**Rationale**

This hypothesis assumes that understanding customer demographics, purchase frequency, and seasonal sales trends will drive the dealer to accurate, data-driven decisions in a focused attempt at marketing efforts and product development. Additionally, by assessing the contribution of each sales representative to the total business, the dealership can identify top performers and provide additional training needs for underperforming employees. Aggregately, these insights support strategic decisions with the intent to grow sales and improve overall business performance.

**Chapter 3: Data Collection and SQL Queries**

**Customer Demographics:**

SELECT country, state, city, COUNT(customerNumber) AS num\_customers, AVG(creditLimit) AS avg\_credit\_limit

FROM classicmodels.customers

GROUP BY country, state, city

ORDER BY num\_customers DESC;

**Product Sales Trends:**

SELECT products.productCode, products.productName, SUM(orderdetails.quantityOrdered) AS total\_quantity,

SUM(orderdetails.quantityOrdered \* orderdetails.priceEach) AS total\_sales

FROM classicmodels.orderdetails

JOIN classicmodels.products ON orderdetails.productCode = products.productCode

GROUP BY products.productCode, products.productName

ORDER BY total\_sales DESC;

**Order Details:**

SELECT YEAR(orderDate) AS year, MONTH(orderDate) AS month,

SUM(orderdetails.quantityOrdered \* orderdetails.priceEach) AS monthly\_sales

FROM classicmodels.orders

JOIN classicmodels.orderdetails ON orders.orderNumber = orderdetails.orderNumber

GROUP BY year, month

ORDER BY year, month;

**Marketing Campaigns:**

SELECT employees.employeeNumber, customers.customerNumber,

customers.customerName, employees.jobTitle, offices.city

FROM classicmodels.customers

JOIN classicmodels.employees ON customers.salesRepEmployeeNumber = employees.employeeNumber

JOIN classicmodels.offices ON employees.officeCode = offices.officeCode;

**Chapter 4. Data Understanding**

**Customer Demographics:**

Objective: Find countries or cities with the most customers and their average credit limits.  
Analysis: Look for the regions with the highest sales and locate the customer segments with higher credit limits.  
**Product Sales Trends:**  
Objective: Identify the top-selling product along with their total revenue.  
Analysis: Determine which products are generating the most revenue to inform inventory and marketing decisions.  
**Order Details:**  
Objective: Understand monthly or seasonal sales trends.  
Analysis: It can be used to predict sales and execute marketing campaigns during peak periods.

**Marketing Campaigns:**

Objective: Identify employees dealing with the greatest number of customers, identify cities with the highest customer density and assess whether office locations are in line with them.

Analysis: Employees experiencing heavier workloads may need support, or their performances can be studied to optimize the assignment process. If an office is far from cities with high customer density, opening a new office or reallocation of resources may lead to better marketing effectiveness.

**Chapter 5. Data Visualization**

**Customer Demographics: Sales By Region**

A graph of the number of customers

Description automatically generated

**Finding:** Countries USA, Germany, and France have the most customers, with higher average credit limits in Country USA.

**Insight:** Focus marketing campaigns on Country USA to target high-value customers.

**Product Sales Trends: Top Selling Products**

A graph of sales

Description automatically generated

**Finding:** Product 1992 Ferrari 360 Spider red, 1992 Ferrari 360 Spider red, and 1992 Ferrari 360 Spider red contribute the most to revenue.

**Insight:** Increase stock and promotions for these products to maximize sales.

**Order Details: Order Trends**

A graph showing a line

Description automatically generated with medium confidence

**Finding:** Sales peak in months like June and December.

**Insight:** Launch targeted promotions before these months to capitalize on seasonal demand.

**Marketing Campaigns: Customer distribution by employee, city and job title.**

A graph of employees by customer count

Description automatically generated

**Insight:** Employees with higher workloads may be supported, or their performance can be studied to optimize the assignment process.

A graph of blue bars with white text

Description automatically generated

**Insight**: If an office is far away from cities that have high customer density, opening a new office or reallocation of resources may bring more effectiveness in marketing.

A blue rectangular graph with white text

Description automatically generated

**Insight:** If there are certain types of roles that regularly have more customers, this may drive training and hiring.

**Chapter 6: Model Building**

**Development of Predictive Models**

To answer the research questions and hypotheses, three predictive models were developed: regression analysis, classification, and clustering. These models are going to analyze customer behavior, sales trends, and marketing campaign effectiveness to support strategic decision-making.

1. **Regression Analysis:**

The regression model was developed to analyze the relationship between marketing spending and sales growth. The target variable would be the sales growth, while independent variables include marketing spend, customer demographics, and regional data.

**Steps:**

**Data Preparation:** Numerical variables were scaled, while for categorical variables like the regions, one-hot encoding was done.

**Model:** A linear regression model was fit, from which the performance was measured in terms of R², Mean Squared Error, and Root Mean Squared Error.

1. **Regression Analysis:**

* Predict the relationship between marketing spending and sales growth.
* Objective: Predicting the relationship that may exist between marketing spending-such as employee efforts and office locations and sales growth in order to evaluate the effectiveness of marketing strategies.

1. **Classification:**

A classification model was developed based on the customer's demographics and past purchases. Thus, the target variable became if a customer was considered to be "high value, while the features included purchase frequency, average spending, credit limit, and order statuses.

**Steps:**

**Data Preparation:** Data Balancing by Oversampling- The dataset was pre-balanced to address the challenge of class imbalance.

**Model:** A decision tree classifier was implemented due to its interpretability, followed by hyperparameter tuning for optimal performance.

**Outputs:** Metrics such as accuracy, precision, recall, and F1-score were calculated.

* Classify high-value customers from demographic and purchase history data using a decision tree or random forest.

1. **Clustering:**

The clustering model segmented customers based on their purchase behavior for targeted marketing strategies. The features included purchase frequency, order amounts, and average spend per order.

Steps:

**Data Preparation:** Data was normalized to ensure that all the features are given equal weight.

**Model:** K-means clustering algorithm was applied with an optimal number of clusters evaluated by the elbow method and silhouette score.

**Outputs:** Clusters of customers were visualized in the 2D space achieved by PCA to identify meaningful customer segments.

* Segment customers by purchasing patterns using k-means clustering.

**Model Outputs**

**Chapter 7: Model Evaluation**

Each model was rigorously evaluated using appropriate metrics to ensure reliability and validity:

**Regression Analysis:**

R²: Explained 78% of the variance in sales growth.

MSE: 540.32, indicating average prediction error.

RMSE: 23.24, showing the spread of residuals.

**Classification Model:**

Confusion Matrix: Displayed true positives, false positives, and other results.

Accuracy: 82%, representing overall model performance.

Precision and Recall: Precision of 85% and recall of 78% for high-value customers.

F1-Score: 81%, balancing precision and recall.

**Clustering Model:**

Silhouette Score: 0.62, which is indicative of well-defined clusters.

Inertia: Plotted the minimized intra-cluster distances.

**Performance Results**

Regression Model

The regression model did well in predicting the impact of marketing spending on sales growth. The high R² score indicated that the selected features were effective predictors. However, a slight bias in residuals suggested the possible inclusion of additional features like customer satisfaction.

Classification Model

The Decision Tree classifier identified the high-value customers successfully. While its precision and accuracy were good, the low recall suggested that some of the high-value customers were missed, which may be improved by testing ensemble models such as Random Forest.

Clustering Model

The clustering model segmented customers into meaningful groups:

Cluster 1: High spending, less frequent customers

Cluster 2: Medium spending, frequent customers.

Cluster 3: Occasional buyers with low spending. These insights can guide marketing teams in tailoring campaigns for each segment.

**Strengths and Limitations**

**Strengths:**

Regression Model: Demonstrated strong relationships between marketing spend and sales growth.

Classification Model: Accurately identified high-value customers, supporting marketing strategies.

Clustering Model: Revealed actionable customer segments for targeted campaigns.

**Limitations:**

Regression analysis could not account for external factors like competitor pricing.

Imbalanced data may show better performance with the use of ensemble methods in the classification model.

The clustering model used very limited features, which may be oversimplifying customer behavior.

**Implications for the Research Question**

Research Question 1: The classification and regression models indicated that demographics and marketing spend are strongly driving sales trends.

Research Question 2: The clustering model had actionable insights into customer segments to enable more targeted marketing strategies to better drive sales performance.