Assignment Submission

You can check the code and detailed analysis on my GitHub: https://github.com/s7raaj/customer-purchasing-behaviour/tree/main

Datixity is working on analyzing customer purchasing behaviors. Using the provided dataset, perform the following tasks:

1. Build a machine learning model to predict whether a customer will spend above or below the average transaction value (Total column).

2. Use appropriate evaluation metrics (accuracy, precision, recall, and F1-score) to assess the model's performance.

3. Provide a brief report on the most important features (e.g., Product line, Branch, Customer type, etc.) that influence customer spending behaviors.

4. Suggest strategies for increasing customer transaction amounts based on your findings.

Introduction:

1. This analysis aims to predict whether a customer's transaction will be above or below the average transaction value.
2. We applied Logistic Regression and Decision Tree models to classify transactions.
3. The analysis evaluated the performance of these models using metrics such as accuracy, precision, recall, and F1 score.
4. Additionally, feature importance analysis helped identify the most relevant factors influencing customer spending.

Dataset:

5. The dataset includes key details such as Invoice ID, Branch, City, Customer Type, Gender, Product Line, Unit Price, Quantity.
6. Tax, Total, Date, Time, Payment Method, COGS, Gross Margin Percentage, Gross Income, and Rating. These features were used to explore customer purchasing behavior and determine if their spending was above or below the average transaction value.

Average Transaction Value:

7. The average transaction value calculated from the dataset was approximately **322.97**. Based on this value, customers were classified into two categories: those who spent above the average and those who spent below it.

Logistic Regression Results:

8. Logistic Regression was applied to the dataset and produced strong results. It achieved a high accuracy of **97.5%.**

9. The model's precision and recall for above-average transactions were **1.0** and **0.94**, respectively.
10. The F1 score, which balances precision and recall, was **0.97** for these transactions.

Decision Tree Results:

11. The Decision Tree model achieved perfect accuracy (100%). However, the feature importance analysis showed that 'Tax 5%' was the only feature with any significance.
12. Other features like Unit Price, Quantity, and Gross Income had little to no effect on the model's decisions.

Cross-Validation Results:

13. Cross-validation was used to further test the stability of both models.
14. Logistic Regression achieved a mean accuracy of **98.1%**, while the Decision Tree model had an impressive **99.8%** mean accuracy, showing both models perform reliably across multiple data splits.

Conclusion and Recommendations:

15. The analysis shows that the Decision Tree model is highly effective at predicting whether customers will spend above or below the average.
16. Interestingly, 'Tax 5%' is the most significant feature driving this prediction. Based on these insights, personalized marketing strategies can be developed to target customers more effectively.
17. For example, businesses can promote additional services or products to customers who tend to purchase lower-value items. Additionally, using loyalty programs or offering discounts in underperforming branches could encourage larger transactions.