Intro to Data Science Project Fall 2024

Muhammad Azmat | 23i-2651

Abdullah Ilyas | 23i-2603

DS-C

REPORT:

Firstly, in **Module 1**, we cleaned the data by performing various data cleaning methods, followed by the removal and filling of missing values. After ensuring the cleanliness of our dataset, we proceeded to perform Exploratory Data Analysis (EDA) i.e. **Module 2**, which helped us understand the relationships between different features and uncover correlations among the variables, and thus understand what we're working with. This step was crucial in identifying key patterns and trends that would inform our modeling process.

In **Module 3**, the major focus was on building a **Linear Regression model**. The target variable, **average spending per month**, was dependent on various independent features (X factors). We carefully cleaned the data, removed outliers, and scaled the features to ensure that the model could make accurate predictions. Thanks to these efforts and handy python libraries, we were able to achieve an exceptionally accurate model that predicted average spending with great precision.

Moving on to **Decision Tree Analysis** in the next phase, we leveraged a decision tree model, which is essentially a flowchart-like structure that makes decisions based on multiple conditions (or "if" statements). This model once again explored the relationship between the dependent and independent variables, but this time through a non-linear approach. We achieved an accuracy of **73%**, which provided valuable insights into how different features influence customer behavior and purchasing patterns.

In **Module 4**, we shifted gears to **K-Means Clustering**, a technique used for segmenting customers into distinct groups based on their purchase behavior and preferences. By choosing the optimal number of clusters, we were able to

categorize customers and identify distinct groups with similar characteristics. The clustering provided a deeper understanding of customer segmentation, which can be instrumental in targeting specific customer groups with tailored marketing strategies.

Finally, in **Module 5**, we compared the performance of the three models we had built: the **Linear Regression model**, **Decision Tree**, and **K-Means Clustering**. Each model provided unique insights, and the comparison allowed us to understand the strengths and weaknesses of each approach in the context of customer behavior analysis. Based on our findings, we concluded that while the Linear Regression model was ideal for predicting numerical values like average spending, the Decision Tree model excelled at making categorical decisions (e.g., predicting purchase behavior). K-Means Clustering, on the other hand, provided powerful segmentation capabilities, allowing for targeted marketing strategies.

Actionable Recommendations for the Electronics Section:

Based on the results from all three models, we can make several actionable recommendations for the electronics section:

1. **Targeted Marketing for High Spend Customers**: By leveraging the Linear Regression model, we can identify customers with high average spending and target them with personalized offers on premium products.
2. **Behavioral Segmentation**: The insights from the Decision Tree and K-Means Clustering models can help segment customers based on their purchasing behavior. These segments can be targeted with specialized promotions to maximize engagement and sales.
3. **Optimizing Product Inventory**: Based on the clusters identified through K-Means, the electronics section can better align its product offerings with

the preferences of each customer segment, optimizing inventory management and sales strategies.

4. **Customer Retention**: For customers identified as high-value, loyalty programs or personalized discounts could be introduced to retain them and increase repeat purchases.

In conclusion, this analysis has provided us with a deeper understanding of customer behavior, and with these insights, the electronics section can adopt more data-driven strategies that will likely result in increased customer satisfaction, retention, and overall sales growth.

The End!