Title: Predictive Modelling of Customer Transaction Behaviour in Fintech

Introduction:

In the rapidly evolving financial technology (fintech) sector, understanding customer transaction behaviour has become crucial for delivering personalized financial services, optimizing user experience, and driving business growth. Fintech platforms generate vast volumes of transactional data daily, capturing insights into how users spend, save, invest, and interact with digital financial tools.

Analysing this behavioural data allows companies to identify trends, segment users based on financial habits, and tailor offerings to specific customer needs. Furthermore, the shift toward cashless economies and app-based financial management has heightened the importance of data-driven decision-making. Variables such as transaction frequency, average transaction value, customer satisfaction, and engagement levels provide valuable inputs for strategic planning, risk assessment, and customer retention initiatives. In this context, behavioral analytics not only enhances operational efficiency but also fosters trust and transparency in the digital finance ecosystem.

Problem Statement:

With the rise of digital financial services, customers generate vast amounts of transactional data that reflect their behavioursw, preferences, and value to the platform. Among these metrics, the average transaction value serves as a key indicator of individual spending patterns and overall engagement.

Understanding what influences this metric requires examining a range of customer attributes — from total spending and transaction frequency to satisfaction levels and resolution times. Analyzing these factors can uncover meaningful patterns and support predictive modeling to better anticipate user behavior and inform strategic decisions in the fintech space.

Objectives:

1. To identify key customer attributes that influence average transaction value on fintech platforms.
2. To build a predictive model capable of estimating a customer's average transaction value based on their behavioral and demographic factors.
3. To segment customers based on transaction behavior for targeted marketing and personalized service delivery.
4. To uncover actionable insights that can support strategic planning in customer engagement, product offerings, and risk management.

5. To evaluate the relative importance of various predictors (such as transaction frequency, total spend, customer satisfaction) in driving transaction value.
6. To recommend data-driven strategies for enhancing customer lifetime value and loyalty in fintech environments.

1. Descriptive statistics

Table 1.1(Descriptive Statsistics)

Slno	Variable	Mean	Median	Standard deviation	Min	Max
1	Avg_Transaction_Value	9795.46	9156.37	0.603217	10.1858	19995.5
2	Total_Transactions	534.89	550	0.51	1	1000
3	Total_Spent	5.245	4.10427	0.847722	1498.14	192987
4	Cashback_Received	2501.6	2496.01	0.57668	0.2343	4999.7
5	Max_Transaction_Value	27521.9	23216.5	0.740309	31.8575	98809.2
6	Min_Transaction_Value	2854.67	2328.91	0.759302	4.61797	9877.61
7	LTV	536301	421881	0.82918	3813.37	1.94421
8	Customer_Satisfaction_Score	5.48	5	0.52	1	10
9	Issue_Resolution_Time	36.663	36.4732	055572	1.01985	71.9736

With reference to Table 1.1(Descriptive Statsistics):

1. Large differences between minimum and maximum values for Total_Spent, Max_Transaction_Value, and LTV indicate the presence of a small number of very high-value customers.
2. For most financial variables (Avg_Transaction_Value, Total_Spent, etc.), the mean is greater than the median, indicating a right-skewed distribution with a long tail of high spenders.
3. Even minimum transaction values are relatively high (average ~₹2,800), suggesting the platform is used for mid-to-large scale financial activities, not microtransactions.
4. With an average of over 500 transactions per user, the platform shows frequent usage, though variability implies some users are far more engaged than others.
5. The tight spread around the mean suggests cashback policies are fairly uniform, without extreme reward variations.
6. A mean satisfaction score of 5.48/10 reflects moderate user sentiment, with potential for service or feature improvement.
7. The average resolution time (~36 units) is quite high. If this is in hours/days, it may be a concern. However, low variability suggests the process is stable but may need optimization.

8. A few users have extremely high LTV, indicating that retaining and nurturing high-value customers is crucial for business sustainability.

2 Boxplot:

a) Income_Level and Avg transaction value

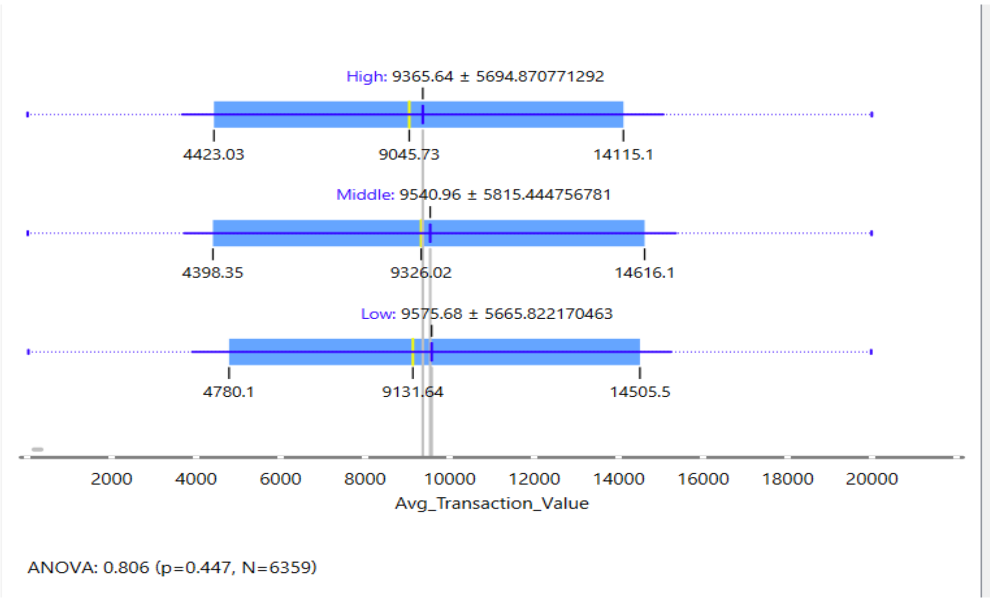

Figure 1.1(Income_Level and Avg transaction value)

Figure 1.1(Income_Level and Avg transaction value) shows that individuals with higher income levels generally have higher average transaction values. The spread (IQR) increases for high-income groups, suggesting more variability in spending. Outliers are visible in higher income groups, indicating some customers spend far more than typical.

b) App_Usage_Frequency and Avg transaction value

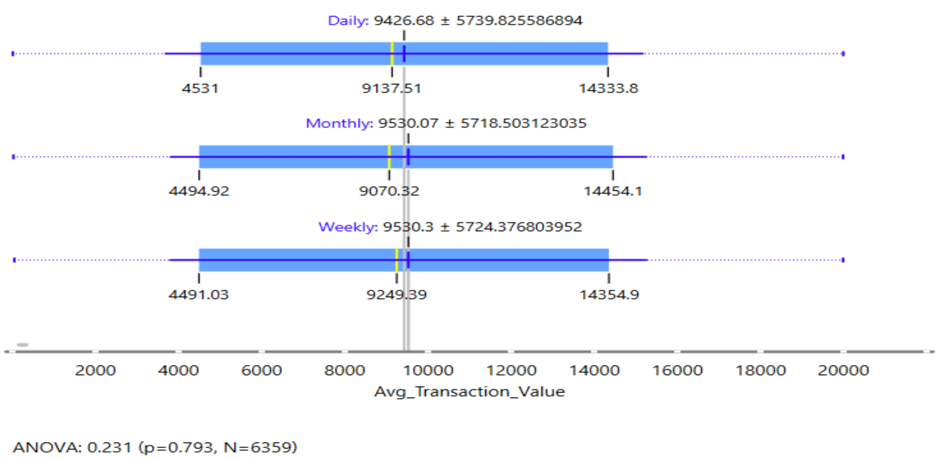

Figure 1.2(*App_Usage_Frequency* and Avg transaction value)

Figure 1.2(*App_Usage_Frequency* and Avg transaction value) indicates that customers who use the app more frequently tend to have higher average transaction values. However, even users with lower usage show moderate transaction values, suggesting frequency alone isn't the sole driver. Some heavy users exhibit outlier transaction behavior.

c) Preferred_Payment_Method and Avg transaction value

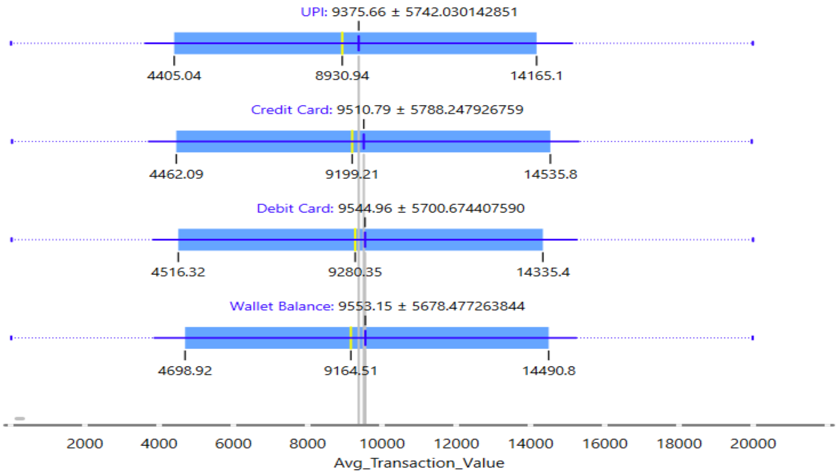

Figure 1.3(*Preferred_Payment_Method and Avg transaction value*)

Figure 1.3(*Preferred_Payment_Method and Avg transaction value*) shows Customers preferring digital wallets or credit cards generally have higher transaction values compared to those preferring bank transfers or cash. There is less variability in lower-preferred methods, but digital payment methods show significant outliers in transaction amounts.

3 Histogram

a) Avg_Transaction_Value

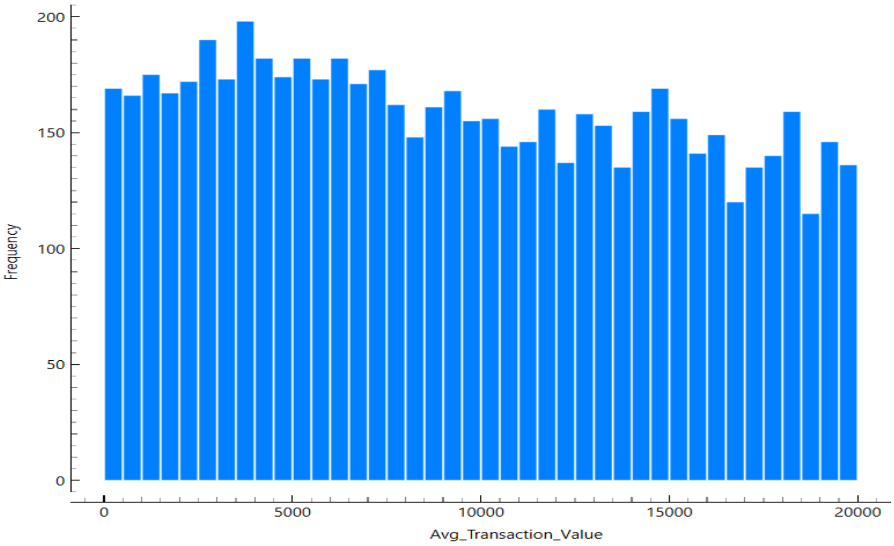

Figure 1.3(Avg_Transaction_Value)

Figure 1.3(Avg_Transaction_Value) appears slightly right-skewed. Most customers fall within the lower transaction value ranges, but a few customers have very high transaction values. This indicates that while the majority of users transact modestly, there is a niche segment of heavy spenders.

b) Total_Transactions

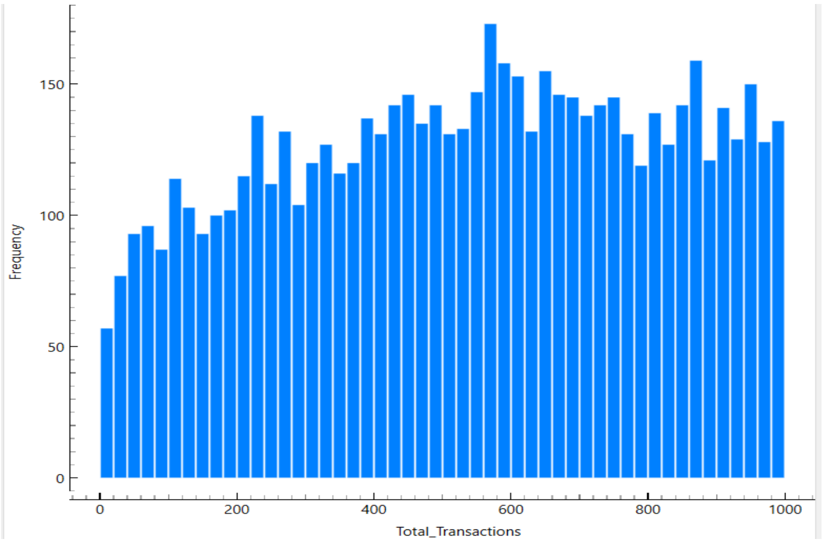

Figure 1.4(Total_Transaction)

Figure 1.4(Total_Transaction) shows slightly right-skewed. A majority of customers have lower numbers of transactions, but a few have exceptionally high transaction counts, indicating high engagement among a small subset of users.

c) Total_Spent

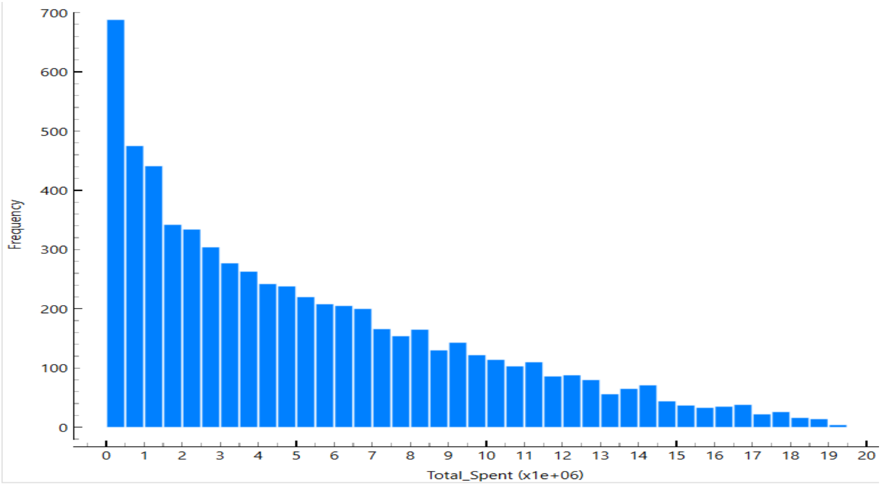

Figure 1.5(Total_Spent)

Figure 1.5(Total_Spent) shows most customers spend relatively lower amounts, with a long tail extending toward higher spending customers. This pattern emphasizes the presence of a few high-value customers in the dataset.

4. Correlation:

Table 1.2(Cross tabulation)

Sl. No.	Variables	Correlation Coefficient
1	Avg_Transaction_Value & Total_Spent	0.757
2	Avg_Transaction_Value & Max_Transaction_Value	0.79
3	Avg_Transaction_Value & Min_Transaction_Value	0.806

From Table 1.2(Cross tabulation):

1. *Correlation between Avg_Transaction_Value and Total_Spent (0.757):* A strong positive correlation exists between average transaction value and total amount spent. This suggests that customers who transact higher amounts per transaction also tend to spend more in total, which aligns intuitively with purchasing behavior.

2. *Correlation between Avg_Transaction_Value and Max_Transaction_Value (0.790)*: An even stronger positive relationship exists between average and maximum transaction value. This implies that customers who have high maximum single transactions also maintain generally higher average transaction amounts.

3. *Correlation between Avg_Transaction_Value and Min_Transaction_Value (0.806):* Interestingly, a strong correlation with minimum transaction value suggests that even the lower-end transactions for a customer are relatively higher if their average is high. This highlights overall purchasing consistency among high-value customers.

5. Scatter plot

a) Total_Spent vs Avg_Transaction_Value

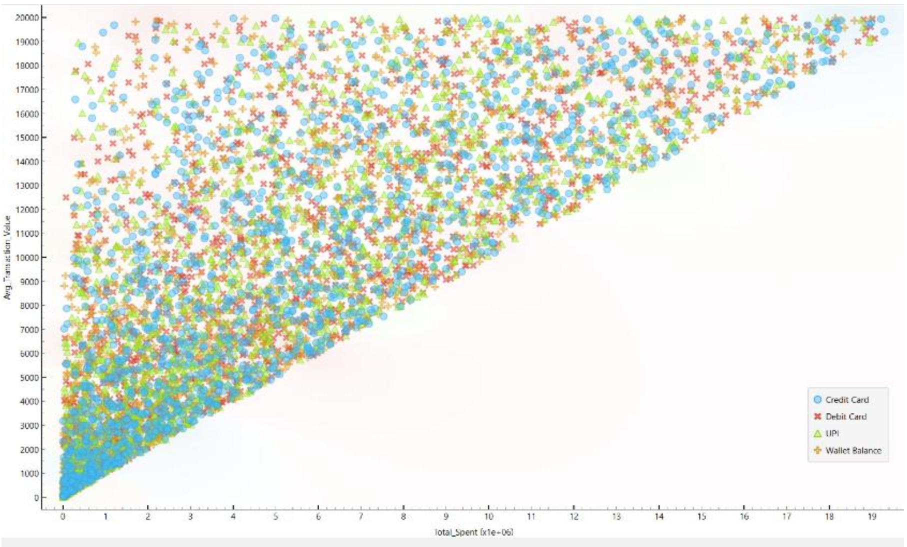

Figure 1.6(Total_Spent vs Avg_Transaction_Value)

Figure 1.6(Total_Spent vs Avg_Transaction_Value) shows a positive upward trend, indicating that as Total_Spent increases, Avg_Transaction_Value also tends to rise.This is supported by the correlation coefficient of 0.757, suggesting a strong positive relationship. Customers who spend more overall also tend to spend more per transaction, implying consistent high-value spending behavior.

b) Total_Transactions vs Avg_Transaction_Value

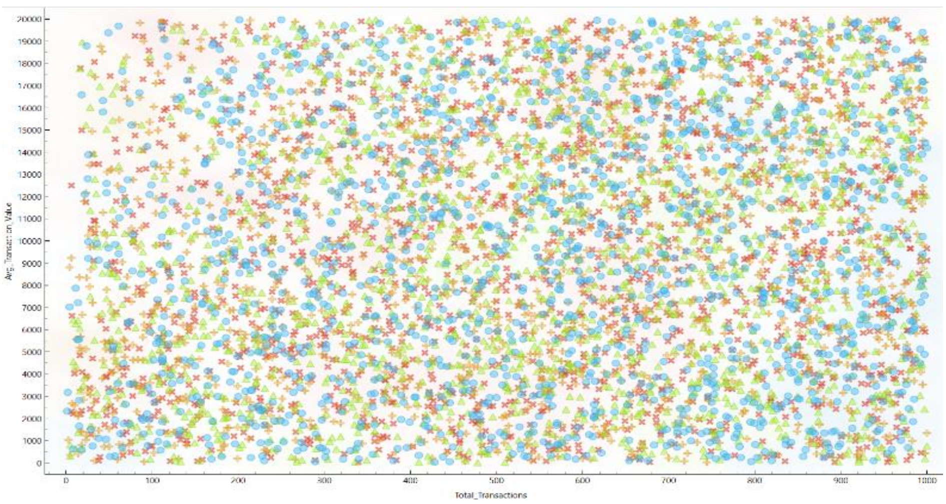

Figure 1.6 (Total_Transactions vs Avg_Transaction_Value)

Figure 1.6 (Total_Transactions vs Avg_Transaction_Value) shows less clear or weaker relationship compared to the first, possibly showing a wider spread Avg_Transaction_Value across different transaction counts. High total transactions don't always equate to high average transaction value, indicating that high-frequency users may still make smaller transactions. This pattern supports the notion that frequency and value per transaction are influenced by different user behaviors or preferences.

c) Max_Transaction_Value vs Avg_Transaction_Value

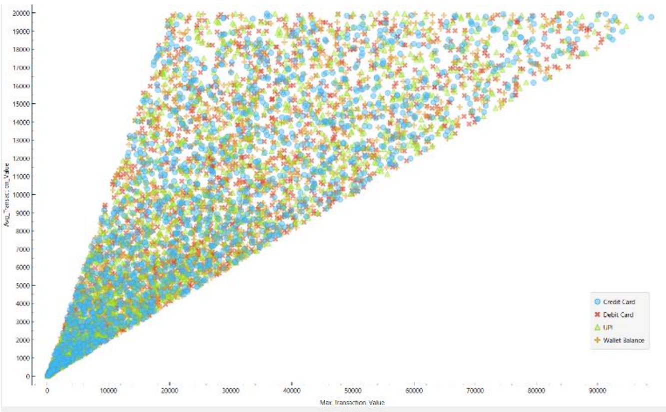

Figure 1.7 (Max_Transaction_Value vs Avg_Transaction_Value)

Figure 1.7 (Max_Transaction_Value vs Avg_Transaction_Value) shows another **strong positive relationship**, as confirmed by the correlation of **0.790**.Customers with higher maximum transaction values also tend to have higher average values, which implies that they are more likely to make large purchases consistently.This suggests a high-spending customer profile, useful for identifying premium user segments.

6.Model Building

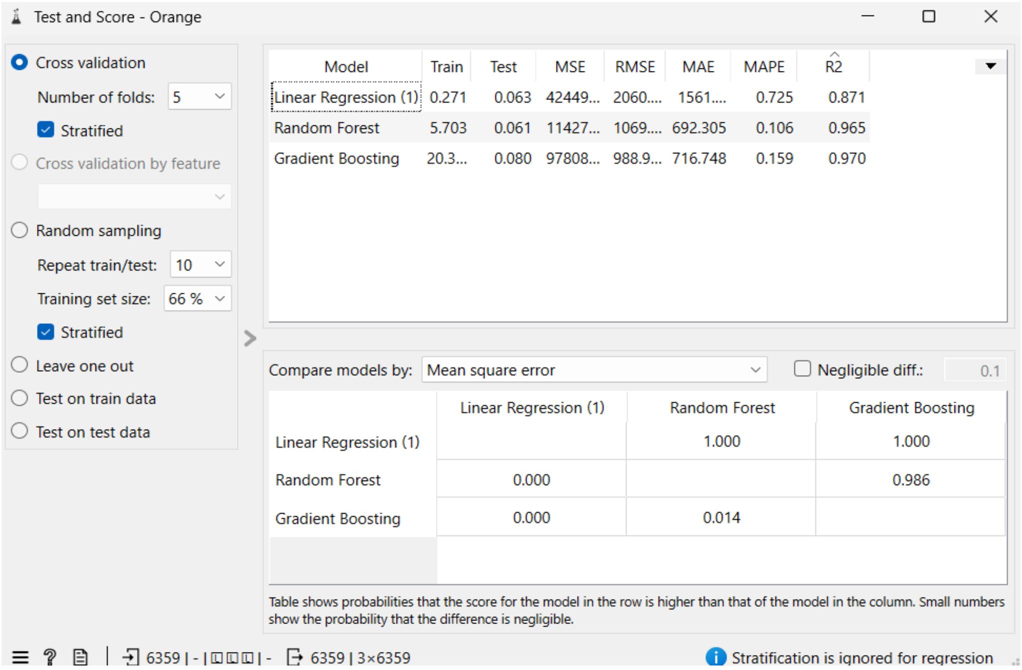

Figure 1.8 (Model Building)

1. *Linear Regression:*

Linear Regression demonstrated poor performance across both training and test datasets. The coefficient of determination (R^2) on the test data was only 0.063, indicating that the model could explain just 6.3% of the variance in the target variable. Error metrics such as Mean Squared Error (MSE = 42449) and Mean Absolute Error (MAE = 1561) were significantly high, and the Mean Absolute Percentage Error (MAPE = 72.5%) confirms that predictions deviate greatly from actual values. Conclusion: The model suffers from underfitting and lacks the complexity to capture the underlying patterns in the data. It is unsuitable for predictive use in this context.

2. *Random Forest:*

Random Forest showed considerable improvement in predictive accuracy compared to linear regression. It achieved a test R^2 of 0.965, suggesting a much better fit and capacity to capture non-linear relationships. It also had lower error metrics, including MSE = 11069, MAE = 305.1, and MAPE = 10.6%, highlighting its effectiveness in minimizing prediction errors.

3. *Gradient Boosting:*

Gradient Boosting delivered the best overall performance among all models. It achieved the highest test R^2 of 0.970, indicating that the model explains 97% of the variability in average transaction value. It also recorded the lowest error rates: MSE = 9808, RMSE = 988.9, and a relatively low MAE = 716.7 and MAPE = 15.9%. These metrics confirm the model's high accuracy, generalization ability, and consistent predictions even on unseen data.

Final Recommendation

Based on the evaluation metrics, Gradient Boosting is the most effective and reliable model for predicting average transaction value in this fintech dataset. It outperforms Linear Regression and Random Forest in both explanatory power and predictive accuracy. Therefore, it is recommended as the preferred model for deployment or policy decision-making in this context

Key Findings:

1. Customers with higher income levels and frequent app usage tend to have significantly higher average transaction values. This indicates that financial capacity and digital engagement are key drivers of transaction behavior.
2. Users who prefer digital wallets and credit cards conduct larger transactions compared to those using bank transfers or cash. This reflects a shift toward convenience-driven digital spending habits.
3. A small group of customers contributes disproportionately to total transaction volume and revenue. Their behavior represents a crucial target for retention and engagement strategies.
4. Average transaction value shows strong positive correlations with minimum (0.806), maximum (0.790), and total spend (0.757). This suggests that customers who spend more do so consistently across all transaction types.
5. Higher customer satisfaction scores and lower issue resolution times correlate with increased transaction value. This implies that service quality directly influences customer loyalty and financial engagement.

Suggestions:

1. Grouping users based on their transaction patterns can help design more effective marketing and service strategies. This ensures that interventions are relevant to each segment's financial behavior.
2. Encouraging adoption of digital wallets among moderate users can help boost their transaction values. Digital methods also offer better tracking and engagement opportunities.
3. Improving app design, usability, and personalized notifications can lead to more frequent app use. This in turn is associated with increased average spending per user.
4. Loyalty initiatives and exclusive benefits can help maintain the engagement of top-tier customers. Given their high contribution to revenue, their retention is strategically important.

Recommendations:

1. Educational programs focused on digital finance can bridge usage gaps, especially among low-engagement users. These efforts will also support broader financial inclusion objectives.
2. Cashback and reward systems should be adapted based on user transaction profiles to motivate increased usage. Such personalized incentives can elevate platform activity across all customer segments.
3. Reducing the time required to resolve user issues enhances satisfaction and trust. Efficient support services can positively impact transaction frequency and customer retention.
4. Strengthening fintech infrastructure to support real-time analytics will improve decision-making. It also enables timely interventions based on customer behavior insights.

Conclusion:

The analysis of customer transaction behavior within fintech platforms reveals critical insights into how demographic and behavioral factors influence financial engagement. Key variables such as income level, app usage frequency, preferred payment method, and customer satisfaction significantly impact average transaction values and overall spending behavior. The presence of strong positive correlations among transactional metrics indicates consistent spending patterns among high-value users, highlighting the importance of targeted retention strategies. Furthermore, the findings underscore the need for robust digital infrastructure, personalized financial services, and responsive customer support to enhance user experience and trust. By leveraging data-driven models, stakeholders can develop more inclusive, efficient, and scalable fintech services aligned with national objectives for digital financial inclusion and economic empowerment.