Financial Technology and Innovation

♦ ISSN: 2836-5372 Vol: 05, No. 01, 2025 ♦ PP. 24–30
Article DOI: https://doi.org/10.54216/FinTech-I.050103

FULL LENGTH ARTICLE

Credit-Card Engagement Segmentation for Embedded FinTech Product Strategy: Evidence from Expenditure Microdata

Samandarboy Sulaymanov[1,*] **Olimjonov Abbosjon Olimjonovich**[1]

[1] Tashkent State University of Economics, Tashkent, Uzbekistan

Emails: sulaymanovsamandarboy@gmail.com · abbos.olimjonovv@gmail.com

Received: January 03, 2025 Revised: March 01, 2025 Accepted: June 04, 2025 ⋆ Corresponding author

ABSTRACT

Digital card products create business value only when issued accounts translate into sustained, responsible use. This paper develops an expenditure-based segmentation model for embedded FinTech card strategy using a real credit-card micro-dataset. Average monthly card expenditure is treated as an observable engagement outcome and is examined alongside income, age, and home-ownership status. The empirical design combines descriptive portfolio profiling, robust regression, cross-validated prediction, and product-action mapping. The results show that income is the strongest observed driver of monthly spend, but the relationship is nonlinear and does not fully explain customer heterogeneity. A quartile-based segmentation separates low-use, developing-use, active-use, and premium-use customers, with mean monthly expenditure increasing from 39.92 to 666.35 across the four operating segments. The analysis argues that card engagement should be managed as a portfolio state rather than as a simple activation metric. The study contributes a transparent business-analytics framework that links observed card expenditure to embedded-finance decisions, including activation support, limit calibration, reward design, and repayment-aware engagement monitoring.

Keywords: Financial technology ▪ Credit-card analytics ▪ Embedded finance ▪ Customer engagement ▪ Behavioral segmentation

1. INTRODUCTION

Card-based financial technology is no longer defined only by the issuance of payment instruments. Digital card providers, marketplace lenders, banking-as-a-service platforms, and embedded-finance vendors increasingly compete by translating transaction and profile data into product decisions. A card platform must decide which customers require activation support, which customers can be offered additional services, and which customers may be approaching a level of engagement that should be monitored carefully. These decisions are commercial, but they are also risk-sensitive because card engagement combines payment convenience, credit exposure, and household liquidity behaviour.

This paper studies card engagement as a measurable business phenomenon. Instead of modelling a binary approval decision or a default event, it focuses on the level of monthly credit-card expenditure. This perspective is useful for FinTech product strategy because digital card value is created after issuance. Issuing a card with no use has little strategic value; encouraging excessive use may create repayment pressure; and treating all active customers alike misses important differences in income capacity and usage intensity.

The analysis uses Bill Greene's credit-scoring data as distributed in the statsmodels datasets module [1]. The dataset is modest in size but valuable for method development because it contains observed monthly card expenditure together with income, age, and home ownership. These variables allow a transparent analysis of spend intensity without relying on proprietary customer identifiers or opaque digital traces. The study therefore positions the dataset as a laboratory for business analytics rather than as a full-scale production credit model.

The paper makes three contributions. First, it defines card engagement intelligence as the translation of observed credit-card expenditure into operational segments. Second, it evaluates both explanatory and predictive models so that the findings can support business interpretation as well as forecasting. Third, it presents portfolio actions tied to the empirical segments, linking quantitative evidence to embedded-finance strategy.

2. RELATED WORK

Research on consumer credit has long connected statistical modelling with business decisions. Greene's credit-card data and econometric examples provide a useful foundation because they connect customer attributes, card expenditure, and credit decisioning in a compact applied setting [1,2]. This tradition treats customer characteristics not only as predictors, but also as signals that help interpret product use, affordability, and credit exposure.

Machine-learning research extended this foundation by showing how richer nonlinear methods can improve consumer credit-risk assessment. Khandani, Kim, and Lo [3] demonstrated that machine-learning models applied to consumer credit data can improve forecasts of delinquency and default. FinTech credit studies broadened the agenda further by emphasizing digital distribution, alternative signals, and platform-based lending. Jagtiani and Lemieux [4] found that fintech lenders can serve areas that are relatively underserved by traditional banking channels, while Berg et al. [5] showed that digital footprints can complement conventional credit information in consumer credit scoring.

A second stream focuses on market structure and digital financial inclusion. Cornelli et al. [6] documented the emergence of fintech and big-tech credit as distinct forms of platform-based lending, while the Global Findex evidence highlights the role of digital payments and financial-service use in inclusion and resilience [7]. These studies are relevant because card engagement is part of a wider digital-finance portfolio: adoption, usage, risk monitoring, and customer value must be interpreted together. Fraud and conduct-risk research also supports the need for responsible analytics, as digital-payment systems can amplify risk when product expansion is not matched by monitoring capacity [8].

Recent credit-market evidence confirms the strategic importance of machine learning while also highlighting distributional and fairness considerations in credit settings [9]. However, much of the literature remains concentrated on approval, fraud, and default prediction. The present paper examines a different but complementary business problem: how an issuer or embedded-finance provider should interpret observed card expenditure after the customer is already active. This shifts the emphasis from acceptance decisions to engagement, portfolio development, and responsible product management.

Table 1 clarifies the contribution of the paper. Existing work provides strong evidence on credit approval, credit scoring, market expansion, and payment risk. The present analysis uses that foundation to address a narrower product-management question: how can observed card use be converted into actionable engagement segments without losing sight of responsible credit behaviour?

3. RESEARCH DESIGN AND DATA

The empirical dataset contains 72 observations and five variables: average monthly credit-card expenditure, age, income, squared income, and home ownership. Income is measured in units of USD 10,000. Average monthly expenditure is used as the primary engagement outcome. The analysis also computes an engagement-rate proxy by dividing monthly expenditure by annual income.

Let y_i denote average monthly card expenditure for customer i, x_i annual income in USD 10,000 units, a_i age, and o_i home ownership. A transparent baseline model is written as:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 a_i + \beta_3 o_i + \varepsilon_i. \tag{1}$$

Because expenditure may rise with income but flatten at higher levels, a nonlinear income model is also estimated:

$$y_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \gamma_3 a_i + \gamma_4 o_i + u_i. \tag{2}$$

The business segmentation layer uses expenditure quartiles to define four operating segments: low use, developing use, active use, and premium use. These labels are not normative judgments; they are portfolio-management categories intended to guide product actions.

Table 2 separates the original variables from the derived business measures. The key managerial distinction is between expenditure level and engagement rate. A high absolute spend level may represent valuable card activity, but a high expenditure-to-income rate may also indicate a need for repayment monitoring or controlled product design.

4. EMPIRICAL PROFILE OF CARD ENGAGEMENT

Table 3 reports the descriptive statistics. Average monthly card expenditure is 262.53, while the median is 158.32, indicating a right-skewed engagement distribution. Income averages 3.44 in USD 10,000 units, and 37.5% of the sample are home owners. The engagement-rate proxy averages 0.72%, with a maximum of 3.95%.

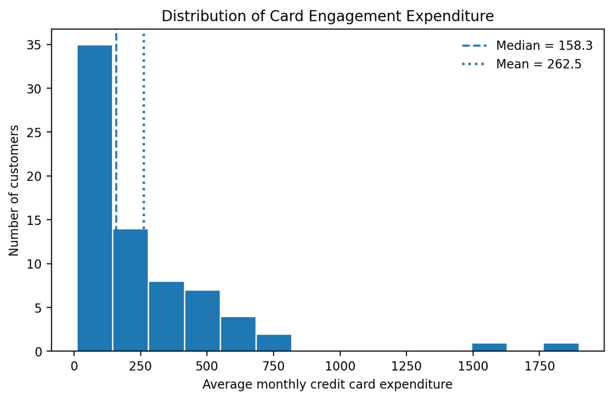

Figure 1. Distribution of average monthly card expenditure.

Table 1. Validated studies and data sources informing card engagement segmentation.

Source	Focus	Relevant contribution	Use in this paper
Statsmodels / Greene [1,2]	Credit-card expenditure and credit scoring	Provides a real econometric credit-card dataset with expenditure, income, age, and home ownership.	Supplies the empirical basis for expenditure segmentation and robust modelling.
Khandani et al. [3]	Consumer credit-risk analytics	Shows that machine-learning methods can improve consumer credit-risk forecasting.	Motivates comparison between interpretable statistical models and predictive models.
Jagtiani and Lemieux [4]	FinTech lending and market reach	Examines whether fintech lenders reach borrowers in underserved areas.	Supports the business relevance of digitally mediated credit access and product scaling.
Berg et al. [5]	Digital footprints and credit scoring	Shows that digital traces can complement conventional credit-bureau information.	Frames usage behaviour as a valuable product signal rather than a simple accounting measure.
Cornelli et al. [6]	FinTech and big-tech credit markets	Documents the expansion of platform-based credit and the need to distinguish business models.	Places card engagement analytics within a broader platform-credit context.
Demirgüç-Kunt et al. [7]	Digital financial inclusion	Documents access, use, digital payments, and resilience across economies.	Connects card engagement with the wider use dimension of financial inclusion.
Chang et al. [8]	Digital-payment fraud analytics	Evaluates fraud-detection methods for digital payment environments.	Supports responsible monitoring when engagement growth increases transaction exposure.
Fuster et al. [9]	Machine learning in credit markets	Examines how machine-learning methods affect credit allocation and pricing outcomes.	Supports combining prediction with transparent, responsible managerial interpretation.

Table 2. Dataset and derived measurement variables.

Variable	Meaning	Use in the analysis
Average expenditure	Monthly card expenditure	Primary engagement outcome
Age	Customer age	Lifecycle profile variable
Income	Annual income in USD 10,000 units	Capacity and spend driver
Income squared	Nonlinear income term	Tests saturation or curvature
Home ownership	Owner/renter indicator	Financial stability proxy
Engagement rate	Expenditure-to-income percentage	Intensity signal for business interpretation
Spend band	Quartile-based expenditure segment	Product-strategy grouping

Figure 1 confirms the skewness visible in Table 3. Most customers are clustered in lower expenditure ranges, while a smaller number of customers generate much higher activity. From a FinTech product perspective, this implies that engagement strategy should not be designed around the mean customer. The median customer requires activation and habit-building, while high-spend customers require stronger monitoring of utilization pressure and value-added services.

5. SEGMENT-LEVEL RESULTS

The expenditure quartile segmentation produces four equally sized operating groups. Table 4 shows that the average monthly spend rises from 39.92 in the low-use group to 666.35 in the premium-use group. The same progression is visible in income and home ownership: the premium-use group has average income of 4.94 and an owner share of 66.7%.

Table 4 suggests that card engagement is not purely a product of age or housing status. Income rises clearly across the segments, but the developing-use group is younger and has the lowest owner share. This pattern is important for embedded-finance design: younger or renter-heavy customers may not need aggressive credit expansion; they may need product education, low-friction payment use cases, and transparent repayment cues.

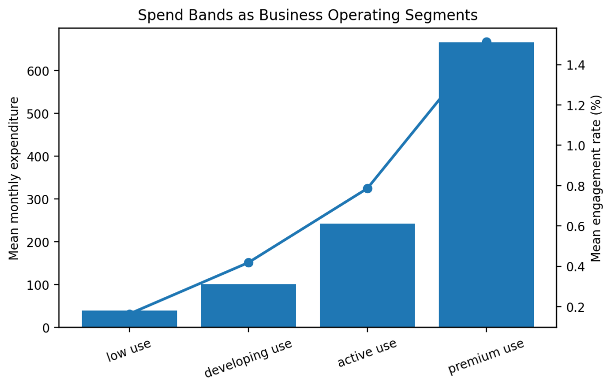

Figure 2. Spend-band operating profile: mean expenditure and engagement rate.

Figure 2 highlights the managerial difference between absolute expenditure and engagement rate. The premium-use segment has much higher monthly expenditure, but it also has the highest expenditure-to-income percentage. That combination is valuable for revenue but requires responsible monitoring. A fintech card manager should therefore treat the premium group as both a growth segment and a conduct-risk segment.

Table 5 shows that ownership does not have a uniform effect across income bands. Owners in lower-middle and upper-middle income groups spend more than renters, while high-income renters show the highest average spend. This should be interpreted carefully because the segment sizes are small, but it illustrates a practical point: portfolio segmentation should combine demographic and behaviour variables rather than rely on one profile indicator.

Figure 3 visualizes the non-uniform ownership pattern. In the middle-income bands, owner status is associated with higher mean expenditure. In the high-income band, renters in this sample show higher activity, which may reflect lifestyle or liquidity differences. The figure reinforces the need for behaviour-based segmentation: a fintech cannot infer engagement capacity from housing status alone.

Table 3. Descriptive statistics of the card engagement dataset.

Variable	Mean	SD	Min	Q1	Median	Q3	Max	Missing
Average monthly expenditure	262.53	318.05	9.58	67.60	158.32	323.48	1898.03	0
Age	31.28	7.15	20.00	26.00	30.00	36.00	55.00	0
Income	3.44	1.70	1.50	2.40	3.00	3.97	10.00	0
Annual income (USD)	34370.83	16994.52	15000.00	24000.00	30000.00	39700.00	100000.00	0
Engagement rate (%)	0.72	0.70	0.03	0.27	0.53	0.91	3.95	0
Home ownership	0.38	0.49	0.00	0.00	0.00	1.00	1.00	0

Table 4. Card engagement segments based on expenditure quartiles.

Spend band	N	Mean spend	Median spend	Income	Age	Owner share	Engagement rate
Low use	18	39.92	41.91	2.75	31.33	0.28	0.16
Developing use	18	101.03	97.13	2.55	28.72	0.17	0.42
Active use	18	242.84	242.15	3.51	32.61	0.39	0.79
Premium use	18	666.35	547.27	4.94	32.44	0.67	1.51

Table 5. Income and home-ownership segments.

Income band	Status	N	Mean spend	Engagement rate	Age
Low income	Owner	4	113.82	0.53	33.25
Low income	Renter	16	97.14	0.50	26.94
Lower-middle	Owner	5	263.17	1.02	32.40
Lower-middle	Renter	14	103.53	0.37	28.36
Upper-middle	Owner	4	477.80	1.32	38.00
Upper-middle	Renter	11	295.87	0.90	30.00
High income	Owner	14	434.64	0.73	36.14
High income	Renter	4	719.23	1.50	35.25

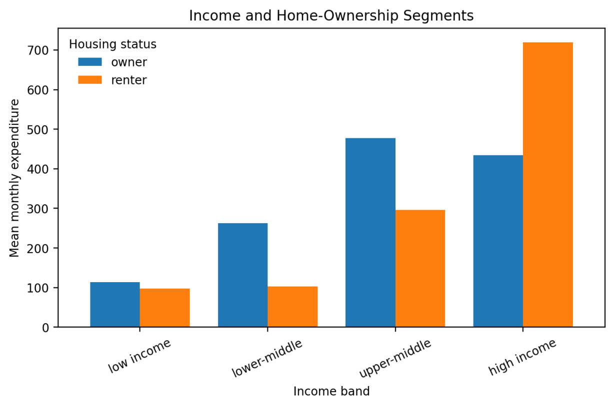

Figure 3. Mean card expenditure by income band and home-ownership status.

6. MODEL-BASED EVIDENCE

Table 6 reports the robust regression results. In the linear engagement model, income is the only clearly significant predictor, with a coefficient of 79.84. The nonlinear income model indicates a positive income term and a negative squared-income term, suggesting that monthly expenditure rises with income but with diminishing marginal increase. Age and home ownership are not statistically strong in these models.

The regression results support a conservative conclusion. Income matters, but it does not fully explain card engagement. The nonlinear model improves adjusted fit from 0.163 to 0.198, which is a useful but modest improvement. This is exactly why fintech card strategy should combine model-based indicators with segmentation rules and ongoing behavioural observation.

Table 7 confirms that the nonlinear income specification has stronger explanatory value. However, the adjusted R^2 remains below 0.20. From a business analytics viewpoint, this is informative: profile variables can guide product strategy, but they cannot replace transaction history, repayment behaviour, and real-time account signals.

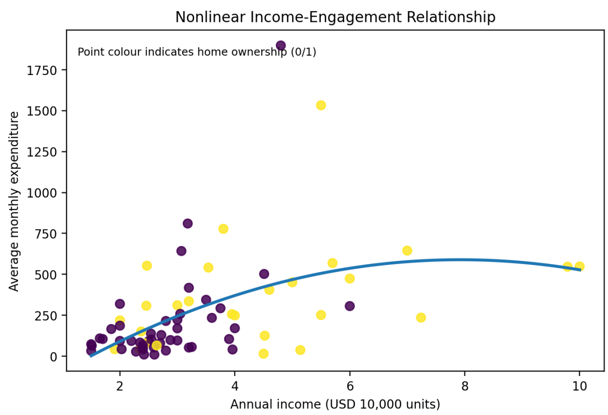

Figure 4. Nonlinear relationship between income and card expenditure.

Figure 4 shows the same pattern visually. Expenditure generally increases with income, but dispersion remains high and the fitted curve flattens. A fintech product team should therefore avoid deterministic rules such as raising limits solely by

Table 6. Robust regression results for monthly card expenditure.

Model	Term	Coef.	Robust SE	t	p-value
Linear engagement model	Constant	0.40	140.09	0.00	0.998
Linear engagement model	Income	79.84	27.47	2.91	0.004
Linear engagement model	Age	-0.78	4.27	-0.18	0.856
Linear engagement model	Home owner	32.09	99.23	0.32	0.746
Nonlinear income model	Constant	-237.15	229.57	-1.03	0.302
Nonlinear income model	Income	234.35	95.48	2.45	0.014
Nonlinear income model	Income squared	-15.00	7.48	-2.01	0.045
Nonlinear income model	Age	-3.08	3.60	-0.85	0.393
Nonlinear income model	Home owner	27.94	99.31	0.28	0.778

Table 7. Regression fit summary.

Model	R^2	Adj. R^2	AIC	BIC
Linear engagement model	0.198	0.163	1025.18	1034.29
Nonlinear income model	0.244	0.198	1022.98	1034.36

income tier. Income should be treated as a capacity signal, not a complete representation of the customer's engagement or repayment behaviour.

7. PREDICTIVE ASSESSMENT

Table 8 evaluates four predictive models using grouped cross-validation. Ridge regression provides the best overall performance in this small dataset, with MAE of 185.12 and RMSE of 304.46. More complex models do not improve performance; gradient boosting performs worst in this setting. This is a useful finding for fintech product analytics because complexity does not automatically create operational value when the feature set is compact.

Table 8 should not be read as a failure of predictive analytics. Rather, it demonstrates the boundary between profile-level forecasting and operational card intelligence. Prediction is feasible, but the small set of profile variables leaves considerable unexplained variation. A production fintech environment would enrich this model with transaction frequency, repayment punctuality, merchant diversity, channel use, and digital wallet behaviour.

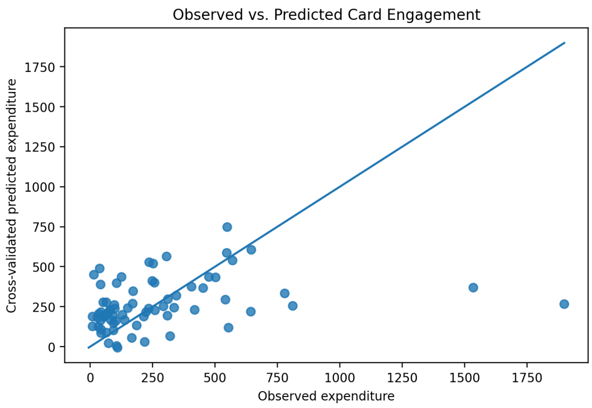

Figure 5. Cross-validated observed and predicted card expenditure.

Figure 5 shows that the model captures the central tendency but struggles with high-expenditure outliers. The diagonal line represents perfect prediction. The wider dispersion at higher expenditure values indicates that high-value engagement requires richer behavioural features. For business practice, this supports a two-stage approach: use simple models for early segmentation, then use transaction-level analytics once usage data accumulate.

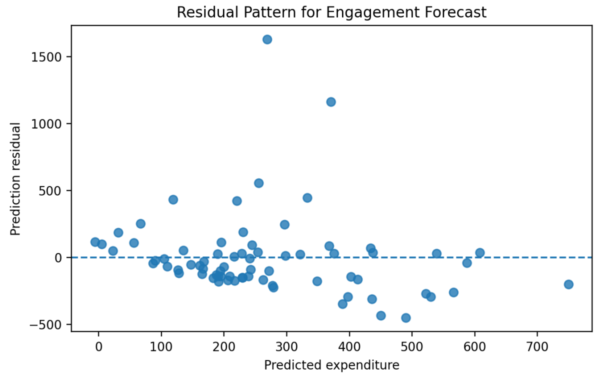

Figure 6. Residual pattern for the card-expenditure prediction model.

Figure 6 further shows that prediction errors are larger among higher-expenditure customers. This residual pattern reinforces the need to treat premium card engagement as a distinct operating state rather than as a simple extension of average usage.

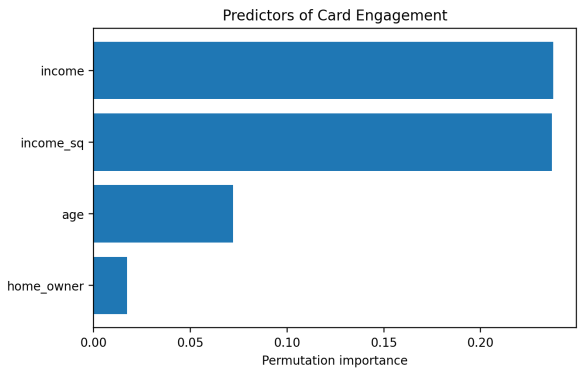

Figure 7. Permutation importance of engagement predictors.

Figure 7 confirms income as the dominant predictor, with income and squared income far above age and home ownership. This aligns with the regression results and reinforces the interpretation that card engagement is primarily income-sensitive in the available data. The practical implication is not that fintechs should use income alone, but that income

Table 8. Cross-validated prediction of monthly card expenditure.

Model	MAE	RMSE	R^2	Observed-predicted correlation
Ridge regression	185.12	304.46	0.071	0.316
Polynomial ridge	188.57	306.39	0.059	0.257
Random forest	188.29	318.06	-0.014	0.302
Gradient boosting	225.85	376.68	-0.422	0.186

should anchor the first-generation engagement model before deeper behavioural signals are added.

8. BUSINESS INTERPRETATION

Table 9 translates the spend bands into operating rules. The low-use group is best addressed through activation nudges and low-risk merchant offers. The developing-use group can be monitored for habit formation and controlled limit calibration. Active-use customers are candidates for rewards and repayment reminders. Premium-use customers may justify premium services, but the higher engagement rate means that responsible monitoring remains essential.

Table 9. Business segmentation rules derived from expenditure bands.

Segment	Rule	N	Mean spend	Suggested product action
Low use	Expenditure $\leq$ Q1	18	39.92	Activate through low-risk digital nudges and merchant offers.
Developing use	Q1 $<$ expenditure $\leq$ median	18	101.03	Increase controlled limit and monitor engagement rate.
Active use	Median $<$ expenditure $\leq$ Q3	18	242.84	Bundle rewards and add repayment reminders.
Premium use	Expenditure $>$ Q3	18	666.35	Offer premium services while controlling utilization strain.

Table 9 is deliberately written as a product decision table. It helps convert a statistical segmentation into practical management actions. The main point is that customer engagement should be encouraged differently across segments. A low-use cardholder does not require the same intervention as a premium-use customer, and the same expenditure level may imply different risk depending on income and repayment behaviour.

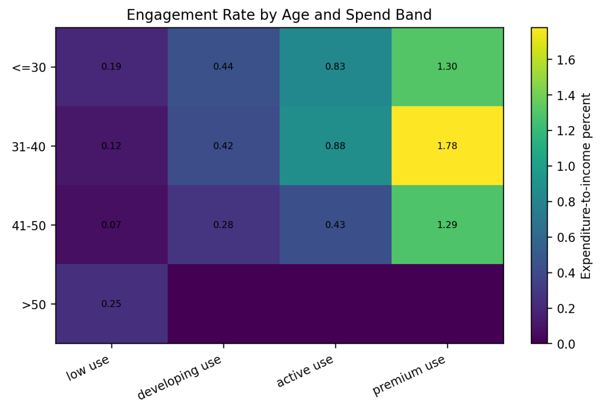

Figure 8. Engagement rate by age group and expenditure band.

Figure 8 adds lifecycle interpretation. The engagement rate is concentrated in the higher spend bands, especially among customers in the early and middle age ranges. For embedded finance, this suggests that high engagement is not simply a sign of loyalty. It can also signal higher wallet share or payment dependence. Product teams should combine engagement growth with repayment prompts and transparent spending controls.

9. MANAGERIAL IMPLICATIONS

The findings lead to three managerial implications. First, embedded-card providers should measure engagement as a portfolio state rather than as a simple activation metric. The customer journey does not end with card issuance; it begins there. Expenditure bands reveal whether a customer is dormant, developing, active, or premium in business terms.

Second, income-based decisioning should be nonlinear. The regression and visualization results show that expenditure rises with income but does not do so in a perfectly linear way. This means that automated limit and offer engines should not rely on a single income multiplier. A more responsible design uses income as an anchor and then updates the engagement profile as real usage and repayment data arrive.

Third, the commercial and conduct-risk sides of engagement must be linked. Premium usage generates revenue opportunities, but it also raises the need for repayment visibility. Fintech innovation is therefore not only about increasing transaction volume; it is also about designing digital card experiences that help customers manage the consequences of that volume.

10. LIMITATIONS AND FUTURE RESEARCH

The dataset is small and contains only a compact set of customer variables. This limits predictive performance and prevents direct modelling of repayment outcomes, merchant categories, mobile-wallet use, or digital-channel behaviour. The value of the dataset is not scale but transparency: it allows the mechanics of card engagement segmentation to be shown without proprietary data restrictions.

Future research should connect expenditure segmentation to repayment quality, delinquency risk, card-fee economics, and customer-level digital interaction histories. Larger datasets would also allow separate modelling of revolvers and transactors, which is important for card profitability and conduct risk. Another useful extension would be fairness analysis, because engagement-based personalization can unintentionally produce unequal product treatment if it is not audited.

11. CONCLUSION

This paper developed a behavioural segmentation model for card engagement intelligence in embedded fintech product strategy. Using real credit-card expenditure data, it showed that income is the strongest observable driver of monthly spend, but that expenditure remains heterogeneous within income and ownership groups. The analysis supports a practical shift from one-size-fits-all card management toward segment-specific engagement strategies. Low-use customers need activation, developing-use customers need cautious growth, active-use customers can be served with value-added bundles, and premium-use customers require both premium treatment and responsible monitoring. The broader conclusion is that fintech card innovation should not optimize for spend alone; it should optimize for sustainable engagement that combines business value, customer utility, and repayment-aware design.

REFERENCES

[1] Statsmodels Developers, "Bill Greene's credit scoring data," *statsmodels Datasets Documentation*, version 0.14.4, 2024.

[2] W. H. Greene, *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall, 2003.

[3] A. E. Khandani, A. J. Kim, and A. W. Lo, "Consumer credit-risk models via machine-learning algorithms," *Journal of Banking & Finance*, vol. 34, no. 11, pp. 2767–2787, 2010.

[4] J. Jagtiani and C. Lemieux, "Do fintech lenders penetrate areas that are underserved by traditional banks?" *Journal of Economics and Business*, vol. 100, pp. 43–54, 2018.

[5] T. Berg, V. Burg, A. Gombović, and M. Puri, "On the rise of FinTechs: Credit scoring using digital footprints," *The Review of Financial Studies*, vol. 33, no. 7, pp. 2845–2897, 2020.

[6] G. Cornelli, J. Frost, L. Gambacorta, R. Rau, R. Wardrop, and T. Ziegler, "Fintech and big tech credit: A new database," BIS Working Papers, no. 887, 2020.

[7] A. Demirgüç-Kunt, L. Klapper, D. Singer, and S. Ansar, *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. Washington, DC: World Bank, 2022.

[8] V. Chang, P. Doan, A. Di Stefano, Z. Sun, and G. Fortino, "Digital payment fraud detection methods in digital ages and Industry 4.0," *Computers and Electrical Engineering*, vol. 100, article 107734, 2022.

[9] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, "Predictably unequal? The effects of machine learning on credit markets," *The Journal of Finance*, vol. 77, no. 1, pp. 5–47, 2022.