Marketing Analytics Problem Assignment: Part 2

Parminder Kaur (NFI030012)

Data Analytics Program, University of Niagara Falls, Canada

Marketing Analytics (DAMO-520-21)

Mohamad Makki

November 22, 2025

Executive Summary

In this report, the critical competencies of diversity in payments by consumers have been addressed. This discussion examines the aspect of the most powerful sources that affect the number of various payment methods that consumers employ, aiming to inform more effective marketing and segmentation efforts of financial institutions. Based on a comprehensive survey that began with 4,362 participants and was refined to 2,094 quality observations through rigorous data cleaning, this researcher utilizes a linear regression to establish that variables of payment behavior are stronger predictors of payment method diversity compared to demographic characteristics.

Key Findings:

Digital payment adoption has raised the number of payment methods by 2.35. The use of traditional payment alone adds +2.35 to payment method diversity. One example is that, Demographic factors (income and age) showed limited statistical significance compared to the strong behavioral predictors when controlling for payment behavior. In this study, simple and multiple linear regressions were evaluated, results were validated by square-root transformation, and robust standard errors were used in dealing with the heteroscedasticity to ensure that the results are reliable in statistical inference.

Recommendations:

Recommend shifting from demographic to behavior-based customer segmentation. The optimal solution would be taking the digital-first approach to acquisition and maintaining the traditional option. Target Comprehensive Users- cross-selling- both digital and traditional. Invest in internet infrastructure and payment behaviour analytics. Final model accounts for 42.3 percent of payment diversity variation; it is statistically significant, hence a valid basis of undertaking data-driven decision making in marketing of financial services.

Introduction

Background Context

The economic reality is that it is a rapid digital revolution where customers are rapidly executing their transactions using mobile wallets, contactless payment, and online banking. The pressure on a transition into a specific mode of payment is of great worth and will become essential to the financial institutions in their struggles to optimise the customer experience and resource allocation.

Problem Statement

This analysis is aimed at comparing the best behavioral and demographic variables that predict the payment method diversity using regression modeling. This is aimed at identifying which model, between linear and non-linear models, is more suitable and which variable is a stronger predictor of the other.

Analytical Approach

Linear regression was chosen because the preliminary correlation analysis revealed mostly linear values, and the residual diagnostics revealed heteroscedasticity instead of non-linearity as the most important issue.

This has been done by a step-by-step approach by modeling development until reaching the diagnostic testing stage, and then to business interpretation through the linear regression modeling on survey data (James et al., 2013).

It is necessary to comprehend the variables that impact the number of payment methods that customers employ since payment diversity has a powerful impact on the rate of conversion, customer satisfaction at the checkout stage, and profitability of the channel. With the digital and traditional payment systems developing, companies have to find which behavioral and demographic factors influence customer tastes significantly. Predictive modelling aids marketers to not just describe these relationships but also predict which client sets are likely to utilize various payment options so that a more specific strategy can be formulated.

Methodology

Data Cleaning, Preparation & Variables

- Applied targeted filtering to retain observations with complete Income and AgeGroup data.
- Eliminated untidy demographic documents (age or income not provided).
- Categorical variables (AgeGroup, Income) were converted to factors.
- Guaranteed stability in the 15 indicators of payment methods.

Complete-case analysis reduced the sample size of 4362 to 2094 good observations, which contained only full demographic information (Income and AgeGroup). This methodology was selected so as to have consistency in the models and not be distorted by non-random missingness patterns in demographic variables. Although this led to the loss of 52 percent of the original sample, the final data of 2,094 observations gives it a strong statistical power with 174 observations per predictor variable, which is far more than the recommended limit. Categorical balance was checked to make sure that there was sufficient representation of all levels of factors and that there were no inflated standard errors. The variables of payment behavior showed a high completion rate of 95% and +, and they maintain the integrity of the key analytical constructs. All these steps will ultimately guarantee a representative, reliable, and dataset to be used in the linear modeling.

Data quality assessment: Over 95% completion rate across payment variables indicates high reliability.

Variable Specification:

- Dependent Variable: TotalPayments (amount of 15 payment methods used)
- Independent Variables:
    - UsesDigital: Digital payment adoption binary indicator.
    - UsesTraditional: Traditional payment usage binary indicator
    - Income: Nominal income variable with 5 categories.
    - AgeGroup: 7-categorical variable.

Selection of variables was done based on data analysis, where the primary predictors (payment behaviors) were selected based on correlation analysis, which indicated a significantly stronger correlation with TotalPayments (r=0.53 digital, r=0.33 traditional) than demographic variables (r=0.12 income, r=-0.20 age).

Exploratory Analysis for Variable Selection

The evaluation was done on both linear and non-linear specifications. Patterns of correlation indicated linear relationships, whereas exploratory residual plots indicated that heteroscedasticity is the main problem and not non-linearity. Hence, the final model was a linear regression model that was chosen due to robust standard errors, and a square-root transformation was used as a second validation.

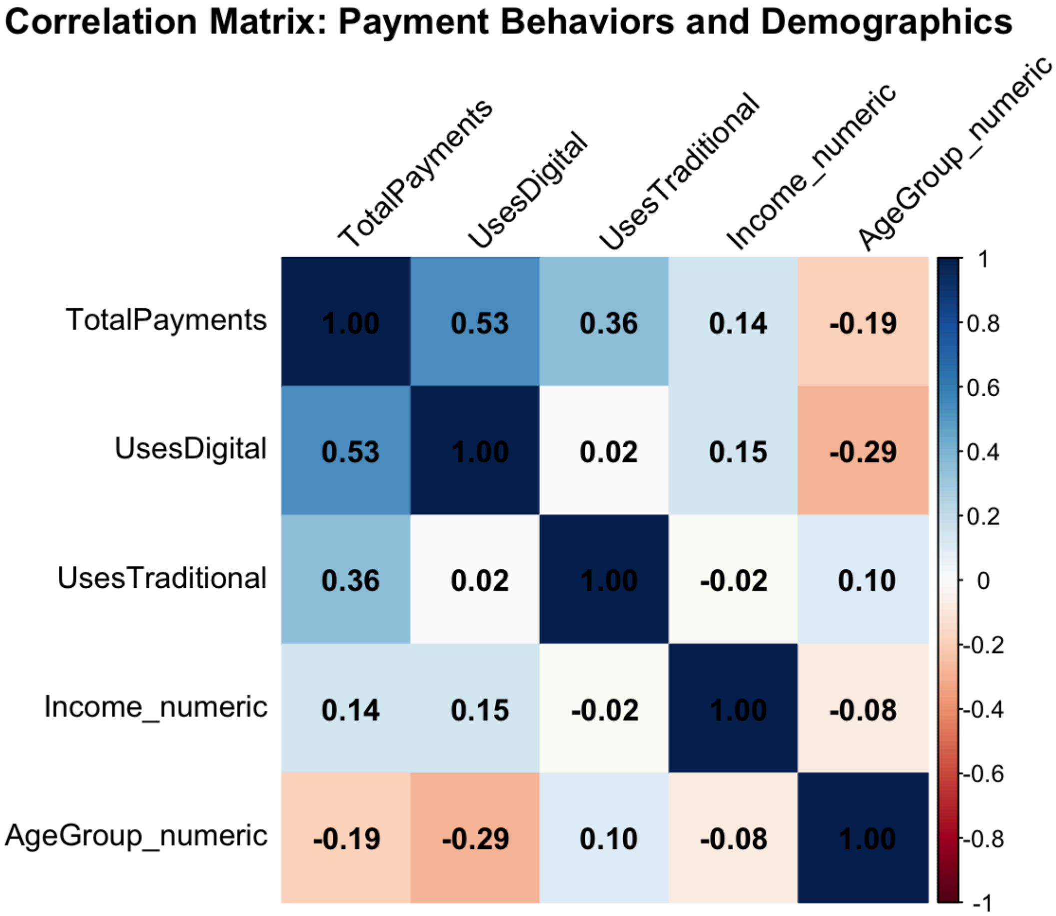

Figure 1: Color-coded correlation table indicating correlation among major variables.

Correlation Matrix Insights & Justification:

- The UsesDigital and TotalPayments (r=0.53) have a high level of correlation.
- UsesTraditional has a moderately correlated (r = 0.36) relationship.
- Demographic variables are weakly related (Income: r=0.14, Age: r= -0.19).
- No dependency exists between the digital and the traditional payment (r= -0.01).

Although correlation gives a preliminary picture of the direction and the strength, it does not imply causality, nor is it able to capture the effects of interaction between variables.

Nevertheless, the patterns of correlation assisted in the selection of the model that indicated the predictors that were significant and linear to the dependent variable. These understandings aided in the justification of the Digital and Traditional payment behavior as important predictors in the following regression models.

Regression Analysis

Model 1: Simple Regression

Purpose: To evaluate whether digital payment adoption predicts payment method diversity.

Formula Used: TotalPayments ~ UsesDigital

Coefficients:

Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.64130	0.07106	51.24 <2e-16 ***
UsesDigital	2.55457	0.08824	28.95 <2e-16 ***
R-squared: 0.286 | F-statistic: 838.1 (p < 0.001)			

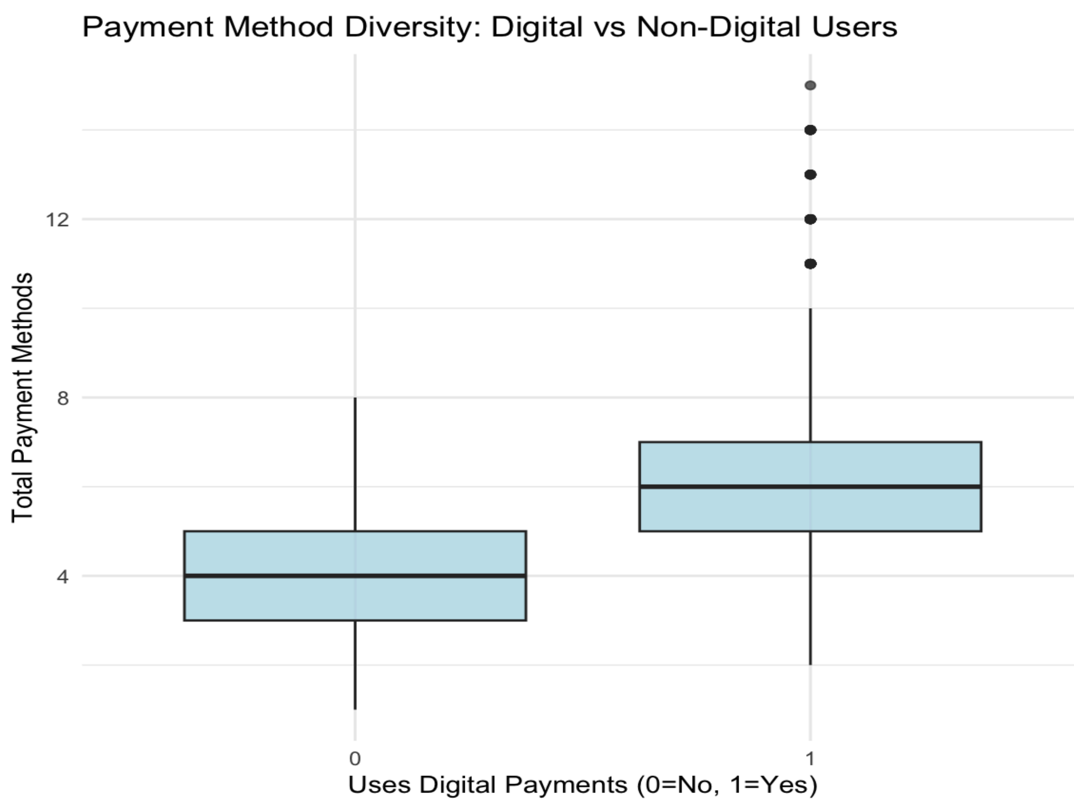

Figure 2: Box plot Total Payments of digital users and non-users.

Interpretation: The average number of payment methods used by digital users is 2.55 higher than that of non-users. Model 1 describes 28.6 percent of the difference in the diversity of payment methods.

Outlier Analysis:

The distribution insight based on visual inspection had anticipated outliers that were depicting high-engagement customers. The analysis of formal influence based on Cook distance revealed 57 influential observations (2.7% of the sample) with an average of 10.6 payment methods, who are comprehensive users with 98 percent digital adoption. These premium customers are the real market segments and not the data quality problems. Since these are business-relevant as they are the target customer profile for financial institutions, they were kept in all analyses to ensure the analytical integrity of identifying valuable customer segments.

Model 2: Multiple Regression

Purpose: To assess how behavioral and demographic factors jointly explain payment diversity.

The formula used: TotalPayments ~ UsesDigital + UsesTraditional + Income + AgeGroup

Key Coefficients (with Robust Standard Errors):

Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.37431	0.21811	6.301 3.60e-10 ***
UsesDigital	2.35427	0.06941	33.916 < 2e-16 ***
UsesTraditional	2.34956	0.09017	26.056 < 2e-16 ***

R-squared: 0.4232 | Adjusted R-squared: 0.4199

F-statistic: 127.2 ($p < 0.001$)

Note: Robust standard errors address heteroscedasticity

Key Insights:

- Digital and traditional payments both have strong and independent impacts.
- The statistics of the demographic variables are not significant.
- The model is an excellent predictor of behavioral data 42.3% variance explained.
- Explanatory power is dominated by behavioral variables.
- Comparison metrics of the models ensure better achievement through the 9 percent lower prediction error and a 424.79-point improvement in AIC.

Note: The demographic variables were included to establish that the effects of behaviors were still strong even after the background characteristics were controlled.

Diagnostic Testing

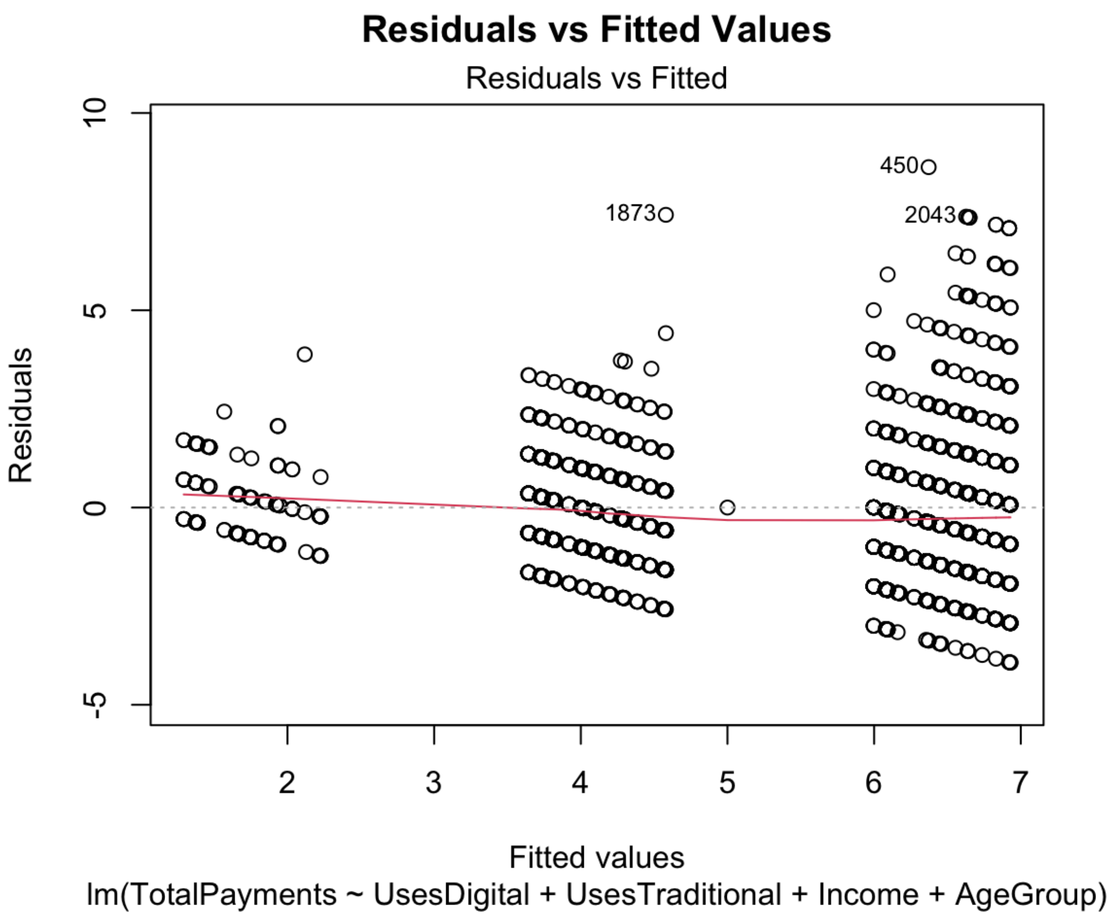

Figure 3:Scatterplot of the residuals with a funnel pattern that reflects heteroscedasticity.

Interpretation: This funnel-shaped heteroscedasticity, where the residual variance increases with the fitted values, is typical of behavioral data and suggests that more variability occurs among the high-engaged customers.

Statistical Tests:

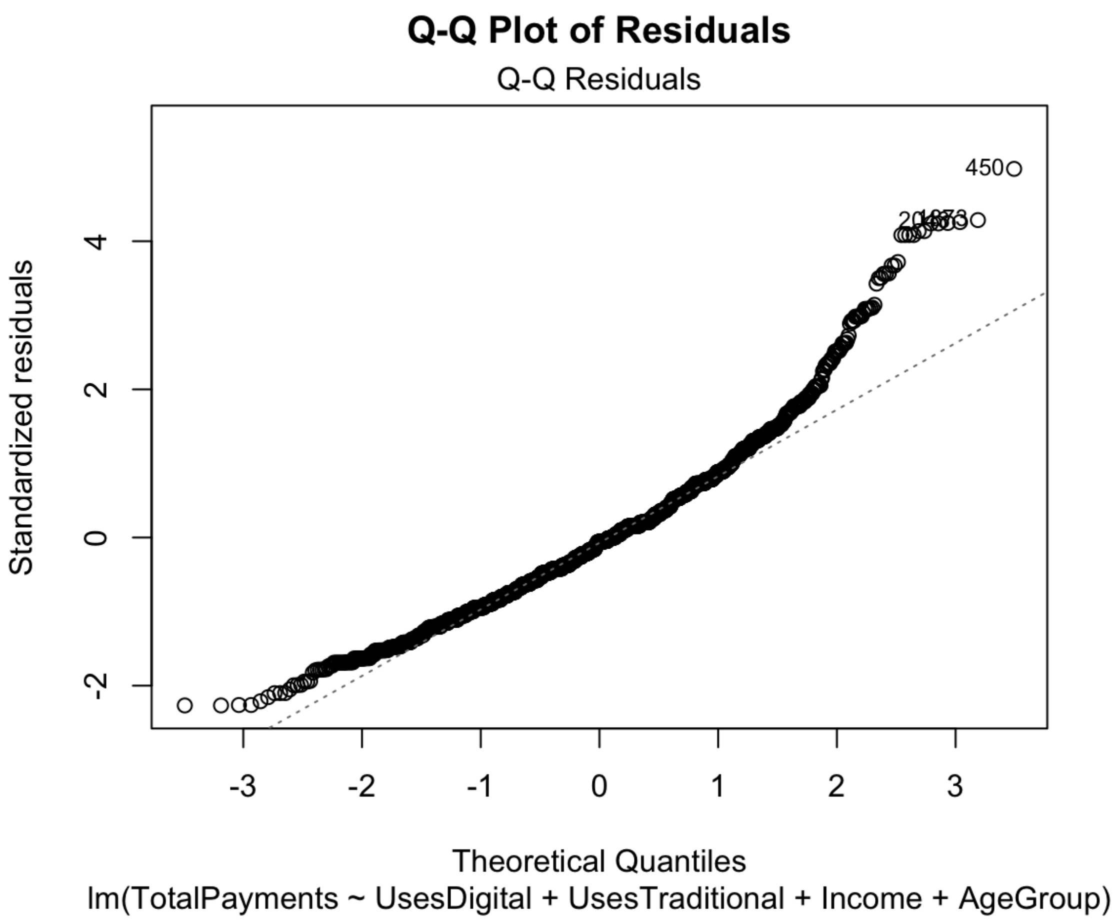

Figure 4: Q-Q plot of residuals. where points generally follow a diagonal line, confirming normality.

Homoscedasticity: The test of heteroscedasticity (BP = 113.23, $p < 0.001$) BP test shows that there is a high level of heteroscedasticity ($p < 0.001$).

Multicollinearity: VIF for all values below 1.23 (way lower than the standard of 5).

Independence: Durbin-Watson = 1.98 ($p = 0.36$) = no autocorrelation.

The influential point analysis revealed 57 observations of highly engaged Comprehensive Users (on average 10.6 payment methods, out of which 98% digital usage). These high-valued customers automatically are moderate in influence without showing signs of data quality problems.

Diagnostic Conclusion:

Robust standard errors have been used to address the heteroscedasticity to make reasonable statistical inferences. The models possess strong statistical characteristics that can be applied in business decision-making.

Transformation Justification:

To fix the heteroscedasticity, a log transformation is usually employed, although it did not suit the case since the dependent variable (TotalPayments) contains zeros, and log(0) is undefined. Instead, use the square root, as it is appropriate for small-range count variables and is effective in stabilizing the variance. Once √TotalPayments had been applied, the model exhibited a little better residual pattern, and the degree of heteroscedasticity was reduced, which is why the square-root transformation is the most appropriate to use in this case.

Square-Root Model Results

In order to assess the performance of the square-root transformation, a second regression was performed, and the dependent variable was √TotalPayments. The results of the transformed model were the following key ones:

R-squared: 0.4831

Adjusted R-squared: 0.4802

AIC: 1650.57 (vs. 8270.77 in the original model; lower is better)

Breusch–Pagan Test: BP = 81.59, $p < 0.001$ (heteroscedasticity reduced but not fully eliminated)

Interpretation:

There is a significant improvement in the AIC value and slightly better residual behavior that is observed in the square-root model, which proves that the transformation improved the stability of variance. Nevertheless, since heteroscedasticity was still there, strong standard errors were the most suitable way of correction. Thus, business interpretations remain based on the initial model estimated with strong standard errors.

Note: It is not possible to directly compare the AIC values of transformed and non-transformed models because of their differences in scale. Thus, the reformed model is not tested based on the overall model superiority, but based on diagnostic improvement.

Assumption Summary

Linearity: Satisfied

Independence: Satisfied (DW = 1.96)

Normality: Appears to be satisfied.

Homoscedasticity: BP test violates — robust SE correction.

Multicollinearity: Not preserved (VIF < 1.23)

Coefficient estimates might be affected by any apparent violation of the assumption. Heteroscedasticity, as an example, can lead to inflated standard errors, which may cause some predictors of variable values to be less important than they should be. These factors were put into consideration in the interpretation of the model results and assisted in determining that there was a need to transform or to make changes in the specification.

Model performance comparison

Error Metrics Analysis:

MODEL COMPARISON:

Metric	Model1	Model2
1 R-squared	0.286	0.423
2 RSE	1.928	1.737
3 AIC	8695.60	8270.80

R-squared: Model 1: 0.286 vs Model 2: 0.4232 (48.0% improvement in explanatory power)

Residual Standard Error: Model 1: 1.928 vs Model 2: 1.737 (9.9% improvement)

AIC: Model 1: 8,695.60 vs Model 2: 8,270.80, significant improvement of 424.79 points

Interpretation:

- Model 2 shows better performance on all the metrics with:
- Minimization of prediction error (residual standard error decreased by 0.1905 of methods of payment)
- Better fit to the model: AIC declined by 424.79 (significantly better fit)
- Explanatory power is also significantly enhanced, and the level of R-squared grows to 42.3% from 28.6%.

Hence, significantly higher AIC and the reduced residual error prove that the multiple regression model is more appropriate and provides more accurate forecasts in comparison to the simple regression model.

Business Implications

Customer Segmentation

On the basis of this analysis, four key segments can be pointed out:

Segment 1: Digital Adopters (UsesDigital = 1)

- 6.65 payment methods on average
- Financially active, tech-savvy.
- Perfect with new products and high-end markets.

Segment 2. Traditional Users (UsesTraditional = 1)

- 3.64 average payment methods
- Traditional values familiarity.
- Demand education, slow introduction to digital.

Segment 3. Comprehensive Users (UsesDigital = 1 & UsesTraditional = 1)

- 7+ payment methods
- Most valuable segment
- Best prospects for cross-selling and loyalty programs.

4. Minimal Users (UsesDigital = 0 & UsesTraditional = 0)

- <3 payment methods
- Monetarily uninvolved or hampered.
- Basic financial education and simple solutions are needed.

Strategic Recommendations

Acquisition of Digital Strategy.

Priority: The digital payment infrastructure and mobile platform must be given top priority.

Target Segments: Technologically minded individuals, urban dwellers, and young adults.

Key Tactics:

- Easy boarding procedures on the Internet.
- Mobile-first payment solutions.
- An electronic loyalty program and reward program.
- Influencer marketing and social evidence.
- Traditional Preservation Planning.

Traditional Preservation Strategy

Priority: Maintaining the conventional options, improving digital transformation.

Target Segments: age-conscious consumers, rural consumers, and cash-preferred consumers.

Key Tactics:

- Gradual programmes of digital introduction.
- Instructional content: advantages of online payment.
- Hybrid solutions: electronic cash receipts.
- Enhance ATM and branch access.

Resource Allocation

- High Impact Initiatives (Recommendations on Investment):
- Educating customers on the advantages of payment diversity.

- Digital adopter cross-selling programs.
- Segmentation data analytics of payment behavior.
- Creation and enhancement of online payment systems.

Efficiency Measures (Optimize Existing):

- Traditional and smooth payment processing.
- Legacy systems: phased maintenance of legacy systems.
- Campaign optimization in demographics.
- Conversion of branches to advisory.

Limitations & Future Research

Methodological Limitations

- **Binary simplification:** payment behaviour is simply reduced to use/non-use without lacking intensities and frequencies.
- **Cross-sectional data:** Incapable of causation, nor of behavioral change tracking.
- **Self-Reported Data:** This has the potential to be biased by social desirability in digital payment information.

Non-linear transformations were also taken into account, but they did not have any substantive value over a linear specification that has robust standard errors. Moreover, the data is cross-sectional and this limits the possibility of making inferences in the long run regarding how behavioral changes can be made over time. Because it is self-report data, there is the possibility of recall and social desirability bias. Future model accuracy would be enhanced by including longitudinal data or records of behavior at the transactional level.

Data Limitations

- Lacking psychographic data: attitudes, preferences, comfort with technology.
- Limited contextual factors-geographic, cultural influences
- No transaction frequency or volume metrics

Future Research Directions

- **Advanced Modeling:** Poisson Regression for Count Data, Machine Learning Segmentation

- **Longitudinal Tracking:** Development of payor behavior over time
- **Improved Variables:** scores of technology adoption, real transaction data
- **Experimental Designs:** A/B testing of payment adoption interventions

Conclusion

This analysis therefore proves that payment patterns go beyond the classic demographic groups in terms of the prediction of the diversity of the payment methods. The independent high dependency of both digital and traditional payment adoption is a strong challenge to traditional segmentation strategies and provides a clear strategic path.

Key Contributions

- **Empirical Justification:** There is very strong statistical evidence that behavior surpasses demographics with respect to payment diversity.
- **Strategic Framework:** Well-defined segments and suggestions as to how resources should be allocated.
- **Methodological Rigor:** Long Diagnostic testing to back the validity of results.

Strategic Imperatives

Immediate Actions (0-6 months):

- Target the marketing campaigns based on payment behavior(Segmentation).
- Create internet payment training and orientation.
- Develop cross-sale initiatives among digital adopters.

Medium-term Activities: (6-18 months):

- Establish extensive payment diversity figures.
- Definitely apply advanced analytics to foresee payment behavior.
- Establish payment behavior analytics organizational capacity.

Long-term Vision (18+ months):

- Introduce the payment diversity as a key performance indicator.
- Develop the industry leadership in payment behaviour knowledge.
- Create customer-driven innovations based on deep insights.

Final Recommendation

The financial institutions ought to be investing more effort and money in learning and using payment patterns to take advantage of them instead of the conventional segmentation methods of demographics. Adoption of document payment, with a high degree of connection with financial involvement, presents gigantic customer acquisition, retention, and revenue enhancement opportunities with plans that are based on facts and not hypothetical considerations.

References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jonker, N. (2017). What drives the use of cash in the Netherlands? *International Journal of Central Banking, 13*(4), 225-264.

Kamakura, W. A., & Wedel, M. (1995). Life-style segmentation with tailored interviewing. *Journal of Marketing Research, 32*(3), 308-317.

Liébana-Cabanillas, F., Sánchez-Fernández, J., & Muñoz-Leiva, F. (2018). The role of age in the adoption of mobile payment systems. *Journal of Behavioral and Experimental Finance, 19*, 1-11.