ROBERT H. SMITH
SCHOOL OF BUSINESS

BUDT 704: Data Processing and Analysis

Exploring Patterns and Predictors of Transaction Fraud Across Customer and Order Behavior

Team Project Report By

Isaac Ajanaku
Kyoungmin Lee
Yilin Qian
Nayab Safdar
Junyuan Xu
Esteban Zhu

EXECUTIVE SUMMARY

This project investigates some of the drivers of transaction fraud to help highlight which factors involved contribute to the risk and which don't add much value once key behavioral controls are applied. We focused on producing results that are both interpretable and operationally relevant to the project by using regression analysis and other interaction tests.

Across all analyses conducted, transaction amount proved to be the strongest predictor of fraud. Fraud risk tends to remain low for the majority of transactions but faces a sharp increase for high-value purchases, particularly in the upper tail of the distribution. This pattern is consistent across product categories, devices, payment methods, and address configurations. Additionally, address mismatch increases the overall risk, but its effect is still modest in comparison to the transaction amount.

In contrast, customer age and product category provide little and insignificant predictive value. We found no meaningful nonlinear effects or interaction patterns involving age. While the device type and payment method do not change the central relationship between transaction value and fraud, the time of day does show a moderate effect, with early morning transactions presenting a higher risk.

From an operational perspective, these findings suggest that fraud detection should prioritize clear behavioral factors like transaction amount and address consistency rather than focusing on demographic characteristics. This strategy would support accurate risk identification while remaining both scalable and aligned with fairness considerations.

INTRODUCTION

Transaction fraud can be a major operational challenge for many e-commerce platforms since fraudulent orders can lead to direct financial losses and can hurt customer trust when legitimate transactions are falsely flagged. As fraud mostly tends to happen at a large scale, even small mistakes can quickly snowball, requiring good understanding of not only which factors contribute to the risk, but also how they combine and interact with each other.

Our project goal is to provide a clear, practical view of which signals matter most when detecting fraud and how they can help create better models.

Our work investigates eight research questions:

1. **Independent effects:** Do age, amount, address match, and product category each show meaningful standalone effects on fraud?

2. **Nonlinear effects:** Does fraud risk rise sharply at very high amounts, or change at specific age levels?

3. **Amount × Category:** Do some product categories become riskier as the transaction amount grows?

4. **Amount × Address mismatch:** When billing and shipping addresses differ, does high transaction amount amplify fraud risk?

5. **Age interactions:** Are younger or older customers more exposed in certain categories or under address mismatch?

6. **Device / payment / time moderation:** Do device type, payment method, or time-of-day meaningfully shift the amount–fraud relationship?

7. **Error concentration:** Where do false positives and false negatives cluster, and what simple rules could reduce these blind spots?

8. **Fairness:** Does the model perform similarly across age bands, or would age-specific adjustments improve fairness?

By conducting targeted research, our project sets up a clear understanding of how fraud risk behaves and which factors provide the strongest signals.

DATA AND PREPROCESSING

The analysis uses an e-commerce transactional dataset including factors like customer attributes, transaction details, product categories, payment methods, and timing information.

1. Data Cleaning and Feature Engineering

Preprocessing the data ensured that we work with quality data and produce comparable results. The target variable "Fraudulent Transaction (Yes/No)" was converted into a binary integer format (1 for fraud, 0 for legitimate).

Categorical variables, including product category, payment method, and device used were all transformed using one-hot encoding. Baseline categories were established as clothing (product), PayPal (payment), and desktop (device). Continuous variables such as "Transaction Amount ($)" and "Customer Age (Years)" were normalized to ensure the model maintained stable estimation.

We then created specific features to test for threshold effects:

- **Address Consistency:** A binary variable "is_address_mismatch" (1 if addresses differ).
- **Amount Bins:** Quantile-based bins (e.g., Very Low to Very High) to visualize risk.

- **Time Segments:** Early Morning (0–5), Business Hours (6–17), and Late Night (18–23).

2. Data Splitting Strategy

The dataset was further split for testing and training based on transaction day: the first 60% for training, the next 20% for validation, and the final 20% for testing. This ensures that our model adequately learns from past data to predict future transactions, mirroring real-world workflows.

METHODOLOGY

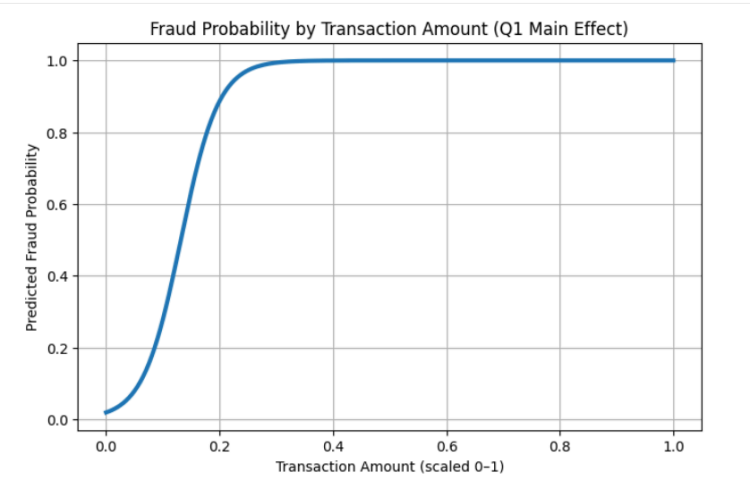
We utilized a multi-stage analytical approach:

1. **Logistic Regression:** To quantify the independent effect of each variable while controlling for confounders.
2. **Interaction Modeling:** To test if risk factors behave differently across different subgroups (e.g., Age multiplied by Product Category).
3. **Non-Linearity Testing:** To detect if risk accelerates disproportionately at higher values using quadratic terms.
4. **Error Analysis:** Using a Decision Tree to identify sections clustering False Positives and False Negatives.
5. **Fairness Assessment:** Evaluating performance across customer age bands using calibration curves and True Positive Rates.

RESULTS

1. Independent Effects on Fraud Probability

We first examined whether core variables can predict fraud independently after controlling for context. A baseline logistic regression model showed that **Transaction Amount** is the strongest predictor ($\beta = 30.12, p < 0.001$). As the transaction value increases, fraud also probability rises sharply.



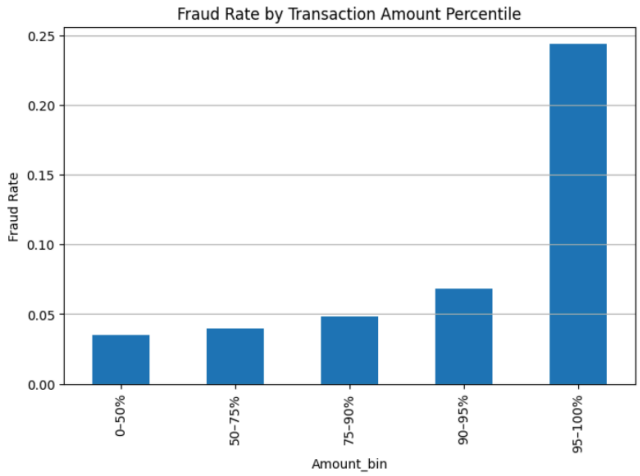
In contrast, other variables did not show significant independent effects:

- **Customer Age:** p-value of 0.305, suggesting no meaningful difference across age groups.
- **Address Consistency:** p-value of 0.718, indicating address matching alone is not a strong independent driver once amount is controlled.
- **Product Categories:** None were statistically significant ($p > 0.05$).

This suggests that demographic and product related factors lose their predictive power when transaction amount is controlled.

2. Nonlinearity and Threshold Effects in Transaction Amount

We also tested if fraud risk tends to accelerate at higher amounts. The regression model confirmed a nonlinear relationship, with a highly significant squared term for Transaction Amount ($\beta = 200.7, p < 0.001$), hinting there is an accelerating risk curve.



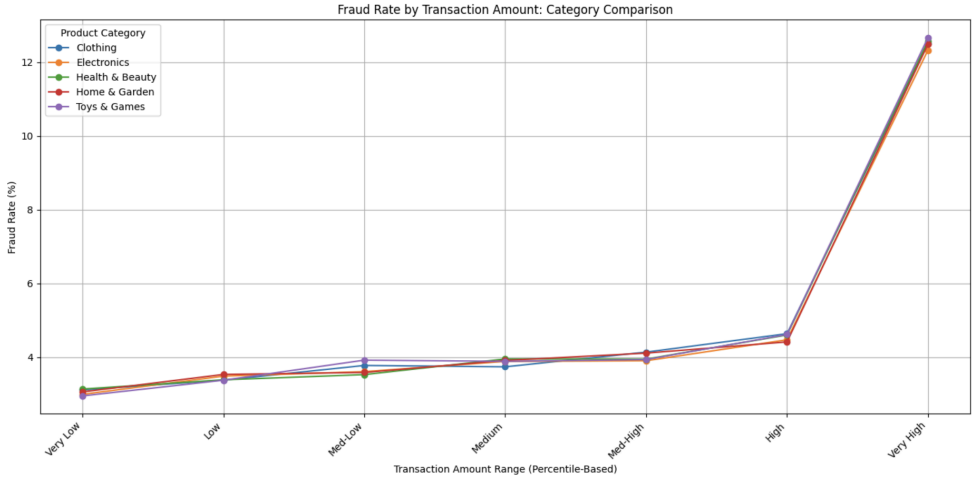
Further percentile analysis revealed a distinct threshold effect. Fraud rates in the 0–90th percentile range remain stable (3–4%), but jump to over 12% when approaching the top 5% range. Meanwhile, customer age did not exhibit any nonlinear behavior ($p > 0.8$).

3. Transaction Amount by Product Category

We evaluated if certain product categories can become risky faster as the price increases. The analysis revealed striking consistency.

- **Visual Trends:** Risk curves for all five categories are nearly parallel, showing a gradual increase followed by a simultaneous spike at the "Very High" bin.
- **Slope Estimation:** Risk slopes ranged narrowly from 18.53 to 19.19 (only 3.5% variation).

- **Interaction Testing:** Interaction terms between Amount and Category were not statistically significant.

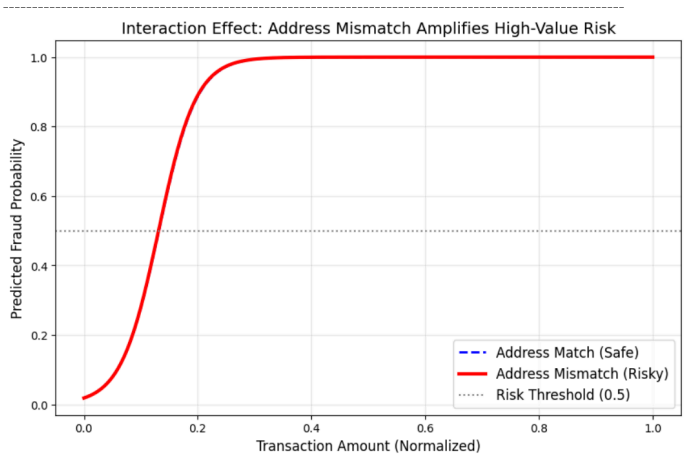


High-value transactions are risky regardless of the product purchased.

4. Address Mismatch and High-Amount Transactions

We examined if mismatched addresses amplify risk at high amounts. The regression highlighted that the interaction term between transaction amount and address mismatch was not statistically significant ($p = 0.625$).

- **Main Driver:** Transaction amount remains the dominant predictor.
- **Additive Effect:** Mismatch adds a small, consistent risk increase but does not alter the curve's slope.

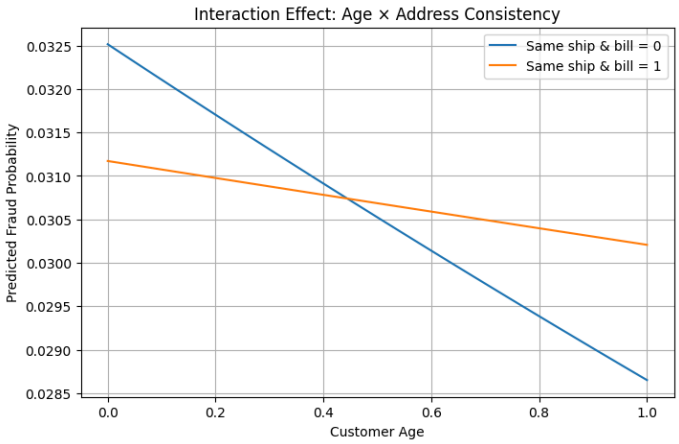


Visually the risk curves for matched and mismatched addresses are nearly identical. A high-value transaction is risky even if addresses match.

5. Customer Age Interactions

We tested if customer age interacts with product category or address consistency.

- **Age x Product Category:** No interaction terms were significant ($p > 0.05$).
- **Age x Address Consistency:** The interaction was insignificant ($p = 0.476$).



These null results confirm that fraud risk is not effected by demographic details, indicating that strategies based on age segmentation may only add complexity without actually improving accuracy.

6. Moderators: Device, Payment, and Time

We checked if context changed the Transaction Amount risk curve.

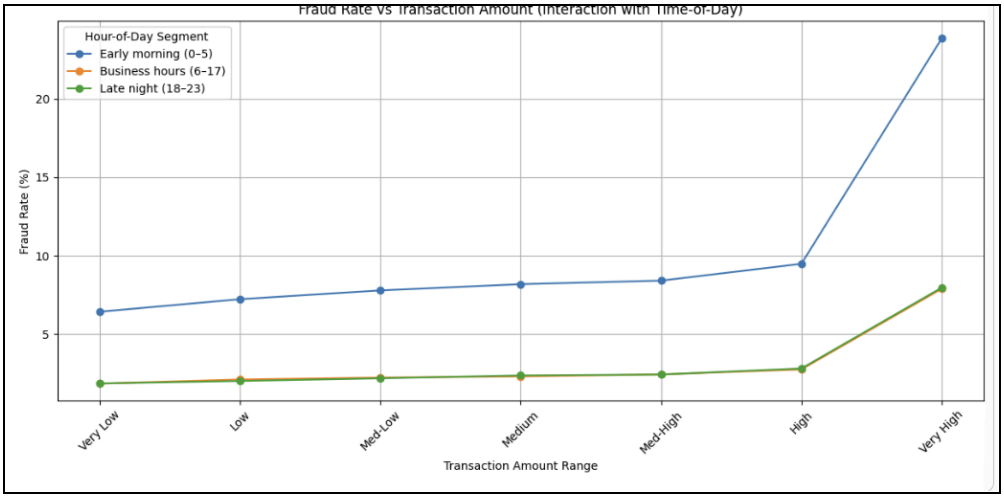
- **Device & Payment:** Risk profiles are almost identical across all devices and payment methods (PayPal, Cards, etc.).
- **Time of Day:** This moderator showed a meaningful effect. Early Morning (0–5 AM) transactions had a significantly steeper risk curve compared to Business Hours (6–17).

"When" a customer pays is a more critical context than "how" they pay when evaluating high-value orders.

ERROR ANALYSIS: MODEL BLIND SPOTS

Our team analyzed the model's mistakes using a decision tree to categorize the predictions into three categories: correct, false positive (FP), and false negative (FN).

Concentration by Transaction Amount



The model is more "anxious" at high values.

- **False Negatives:** Evenly distributed across amount bins.
- **False Positives:** Increase drastically with amount. The model aggressively flags expensive orders, producing false alarms.

Concentration by Address Consistency

The model revealed:

- **Mismatched Addresses:** ~6,500 False Positives.
- **Matched Addresses:** ~59,400 False Positives.
- **Concentration by product category:** Error rates were consistent across all categories, reinforcing that category is a weak predictor.

Same ship and bill	error_type	
0	Correct	22595
	FN	378
	FP	6511
1	Correct	206583
	FN	3863
	FP	59388

dtype: int64

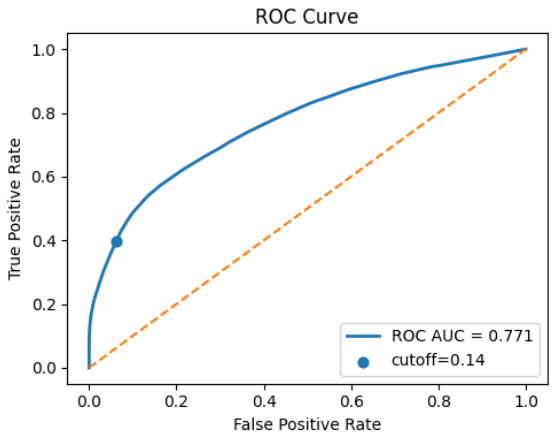
While mismatched addresses are riskier per transaction, the volume of legitimate matched transactions causes the bulk of customer friction to occur in the "safe" matched-address section.

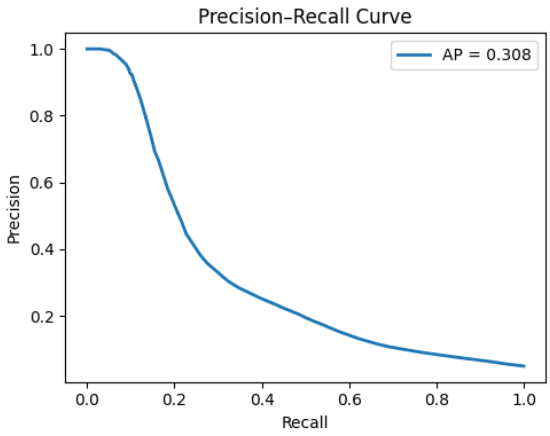
FAIRNESS AND CALIBRATION

We evaluated the model's fairness using a global decision threshold of 0.14 on the test set. We divided the customers into five age bands and calculated the True Positive Rate (TPR), False Positive Rate (FPR), and Expected Calibration Error (ECE).

Results:

- **Overall Performance:** ROC-AUC of 0.771 and precision of 0.285.
- **Stability:** Metrics were stable across all age bands.
    - **TPR Gap:** 2.3 percentage points.
    - **FPR Gap:** 0.6 percentage points.
    - **ECE Gap:** 1.5 percentage points.





These gaps are not significant, indicating we can safely use one global threshold (0.14), because the model already behaves the same across age groups..

OPERATIONAL IMPLICATIONS

Based on our team's analysis, we recommend three adjustments:

1. Implement a "High-Value" Velocity Circuit Breaker

Since the risk rises rapidly in the top 5% of amounts, relying on a linear score is insufficient. We recommend adding an automatic extra verification step (like 2FA) for any transaction above the 90th percentile in amount.

2. Refine "Matched Address" Review Logic

Almost 60,000 false positives occurred with matched addresses. We recommend not simply assuming matching addresses mean a transaction is safe. If an order is high-value and happens in

the early-morning hours, it should still be reviewed, even when the shipping and billing addresses match.

3. Simplify Data Collection

Customer age and product category added almost no predictive value. Removing these features makes the system easier to manage and avoids unnecessary privacy concerns, and it doesn't hurt performance. Resources are better spent on clear behavioral signals like how fast someone conducts a transaction, how much they spend, and when they buy.

LIMITATIONS

- **Proxy Limitations:** Customer age may not perfectly correlate with Account tenure.
- **Data Uniformity:** The consistency of risk slopes across product categories suggests that the dataset may be highly normalized, potentially lacking the specific fraud patterns seen in real-world data.
- **Geospatial Data:** Our data lacked geospatial features like IP distance, which are often potent predictors.

CONCLUSION

This project demonstrates that fraud in this context is driven almost exclusively by transaction magnitude and timing, rather than demographics details or product types.

The defining characteristic is the "hockey stick" relationship between the amount and risk. Fraud tends to accelerate rapidly in the top decile of amounts. Address mismatch does add some risk, but it doesn't change how the overall risk patterns work. A high-value transaction is inherently risky regardless of other factors or influences.

Conclusively, the predictive model is fair across age groups. Having a single global threshold would allow for consistency and accuracy for all customers. The best way to improve fraud detection is to keep things simple; focus on checking high-value orders, especially those made at unusual hours, and avoid using complicated demographic details and rules.

REFERENCES

- **Analysis Codebase:** Google Colab (Isaac Ajanaku, Kyoungmin Lee, Yilin Qian, Nayab Safdar, Junyuan Xu, Esteban Zhu).
- **Dataset:** "Processed_data_FD.csv".
- **Methods:** Logistic Regression, Decision Tree Classifier, Calibration Curves.