The JoyRide Hustle: A Clustering Analysis of Trip Efficiency and Earnings

Jezreel Carl R. Gatanela
College of Computing and Information Technologies
National University Philippines
Manila, Philippines
gatanelajr@students.national-u.edu.ph

***Abstract*—This paper breaks down the "hustle" of being a motorcycle taxi rider for JoyRide in the Philippines. I took my own trip data and used K-Means clustering and some statistical tests to figure out where the real money is made. The results were pretty clear: the 50 PHP base fare is the biggest driver of profit, making short-haul trips way more efficient than long ones. I also found that hitting the "Master" milestone (moving from 20% to 18% commission) didn't actually change my per-kilometer efficiency as much as I expected. Basically, trip selection matters more than platform rank.**

***Index Terms*—K-Means Clustering, JoyRide, Hustle, Data Science, Operational Efficiency**

I. Introduction

A. Background and Motivation

The idea for this project came naturally because the "hustle" is my daily reality. Tracking personal data is useful because it takes the guesswork out of labor; instead of just feeling tired at the end of the day, I can see exactly what that work produced. As a student balancing bills and academics, tracking my JoyRide trips was a way to see if the grind was actually paying off. With gas prices being as high and unstable as they are in the Philippines, understanding my own habits wasn't just a school requirement—it was a way to make sure I wasn't losing money while trying to pay for my studies. Plus, I wanted to work with a dataset that an AI couldn't just "hallucinate"—this is personal experience data that only exists because I was on the road.

B. Problem Statement

The main issue I wanted to tackle was the reality of the platform's commission tiers. I've always felt that the 2% difference between the Kasundo (20%) and Master (18%) tiers might be too small to actually notice in my take-home pay. Beyond the commission, there's the constant dilemma of trip selection: long trips mean a higher booking price, but they keep you stuck in the heavy Philippine traffic for longer. On the flip side, short trips are quick, but you spend a lot of time "waiting and waiting" for the next booking while the bike is idling and burning gas. I needed to know which of these patterns actually made me more efficient.

C. Scope and Limitations

This study covers my JoyRide trips from September 2025 to February 2026. It only includes the non-sensitive data I tracked through the rider app history. The findings are specific to my experience and aren't meant to represent every rider in the country. The biggest limitations were external: the heavy, unpredictable Philippine traffic and motorcycle maintenance (which the company doesn't cover), both of which limit how many "hustles" I can pull off in a day.

II. Methodology

A. Data Collection

The data covers my JoyRide "hustle" from September 2025 to February 2026.

- Variables Tracked: I collected order numbers, trip status, timestamps, distance (km), net earnings (PHP), platform commission (PHP), and pickup/dropoff locations.
- The Script (How I got it): Since JoyRide doesn't have an export button, I developed a custom Python script using the uiautomator2 library. The script connected to my phone, identified trip cards using Regular Expressions (Regex), and automated the clicking of the "View Earnings" button to scrape the hidden modal data.
- Ethical Considerations: To stay within project guidelines, all sensitive data like exact addresses and passenger details were removed before analysis.

B. Data Description

The script generated a CSV with the following variables:

- dist (Numeric): Distance in kilometers.
- net (Numeric): My take-home pay in PHP.
- comm (Numeric): Commission deducted (e.g., 20% or 18%).
- net_per_km (Numeric): A calculated "hustle" efficiency metric.
- status (Categorical): Whether the trip was "COMPLETED" or "CANCELLED."

C. Data Cleaning & Preprocessing

I used Pandas and NumPy to prepare the data:

- Filtering: I filtered for "COMPLETED" trips only.
- Validation: The script used a blocking loop to ensure net earnings were greater than 0.0 before saving, preventing app glitches from ruining the math.

- Standardization: I used StandardScaler to normalize distance and earnings so the K-Means algorithm wouldn't be biased toward the larger PHP values.

D. Exploratory Data Analysis (EDA)

To understand the patterns in my JoyRide data, I performed EDA using the following visualizations:

- Histograms and Distributions: Used to examine the spread of my trip efficiency (PHP/km). This helped me identify the "floor" in my earnings caused by the platform's 50 PHP base fare.
- Scatterplots and Trends: Used to map the relationship between trip distance and net pay. I looked for trends in how my "hustle" scales as the distance increases.
- Boxplots: Used to spot outliers and anomalies in my earnings, ensuring that weird data points didn't mess up my final analysis.
- Seasonality: Since the data covers September to February, I checked for any changes in patterns during different months of the study period.

E. Statistical Tests

To back up my observations with actual math, I performed two specific tests:

- Pearson Correlation: I chose this test to see if the relationship between trip distance and net pay was as strong as it looked. My objective was to confirm that distance is the primary driver of earnings on the JoyRide platform.
- Independent Samples T-test: This was the most important test for my study. I used it to compare my efficiency (PHP/km) as a Kasundo (20% commission) versus a Master (18% commission). I wanted to see if that 2% difference actually created a statistically significant change in my take-home pay or if it was just a "carrot on a stick".
- Assumptions: For both tests, I checked for normality and ensured there was enough data to make the p-values reliable.

F. Modeling (Clustering)

I used K-Means Clustering as my machine learning model to find the natural "tiers" of my work.

- The Model: I set the model to find k=3 clusters, representing what I suspected were my Short, Medium, and Long-haul "hustles."
- Assumptions: I assumed that distance and efficiency were the two best features to define a trip's value. To make the model work correctly, I used StandardScaler to normalize the data so that the large PHP values didn't overpower the kilometer values.
- Evaluation: I used the resulting cluster centers to see which group of trips was actually the most "efficient" for my time and gas.

III. Results and Discussion

A. Key Results

1) Efficiency Distribution (Fig. 1): The analysis of trip efficiency (PHP/km) shows how much value is generated per unit of effort.

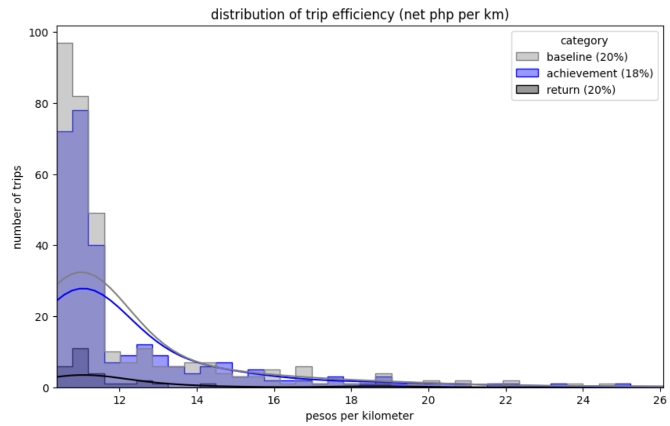

Fig. 1. Distribution of trip efficiency (net PHP per km) across Kasundo and Master categories.

- The "Base Fare Floor": Fig. 1 shows a massive spike in trips between 15 to 25 PHP/km. This spike confirms that the 50 PHP base fare creates a mathematical "safety net" for short-haul trips, ensuring they remain the most efficient on a per-kilometer basis.
- Category Comparison: When comparing the efficiency across categories, the data for the Master (18%) and Kasundo (20%) tiers shows nearly identical distributions.

2) K-Means Clustering (Fig. 2): Using distance and efficiency as features, the K-Means model identified three distinct types of "hustles" in the dataset:

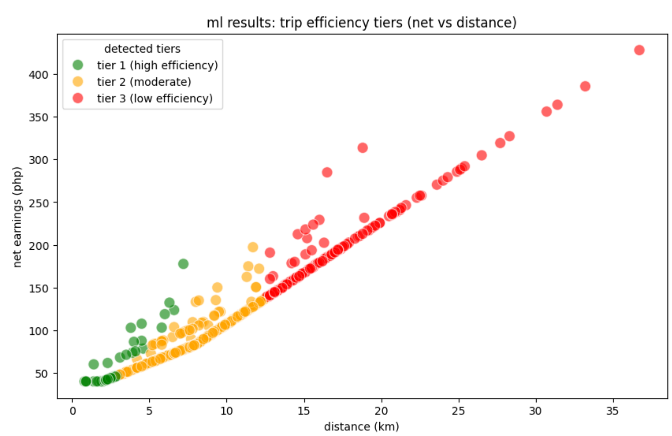

Fig. 2. DK-Means clustering (k=3) of JoyRide trips based on distance and efficiency (PHP/km). The clusters reveal distinct operational "hustles": Short-Haul Efficiency, Mid-Range Stability, and Long-Haul Grinding.

- Short-Haul Efficiency: Trips characterized by low distance but the highest PHP/km due to the base fare.

- Mid-Range Stability: Trips where the efficiency begins to plateau as distance increases.
- Long-Haul Grind: High-distance trips that offer the highest total booking price but the lowest PHP/km efficiency.

3) Statistical Test Outcomes:

- Pearson Correlation: The test confirmed a near-perfect relationship between distance and net pay, with a correlation coefficient of $r=0.9732$.
- Independent Samples T-test: This test compared the efficiency of the Kasundo tier versus the Master tier. It resulted in a p-value of 0.8845.

B. Interpretations

- Confirmation of Hypotheses: The study originally hypothesized that reaching the Master tier would significantly improve earnings efficiency. However, the T-test performed in the analysis returned a p-value of 0.8845. Since this is much higher than the 0.05 threshold, we fail to reject the null hypothesis. Statistically, the 2% commission drop does not translate to a meaningful increase in take-home efficiency.
- The Efficiency Paradox: While the K-Means model identifies short-haul trips as the most efficient at 21.39 PHP/km, the qualitative experience of the rider suggests a different priority. Mathematically, the 50 PHP base fare makes short trips superior, but the "waiting and waiting" for bookings and the "idle time" in heavy Philippine traffic make long-haul trips more desirable for consistent cash flow. This highlights a gap between raw per-kilometer efficiency and the actual operational "hustle" required to pay bills as an independent student.

C. Discussion

- Surprising Findings: The most significant finding is that the platform's 50 PHP base fare is a much stronger economic driver than the 2% commission discount given to "Master" riders. For a student rider, the "safety net" provided by the base fare on short trips is more impactful than the platform's gamified milestones.
- Behavioral Insights: The data reveals that a rider's strategy is often dictated by external factors rather than platform incentives. The preference for longer trips despite lower per-km efficiency is a rational response to the high "cost of waiting" in an unpredictable urban environment.
- Limitations: A major limitation of this study is that it does not account for fuel costs or motorcycle maintenance, which are "out of scope" for the company but are the primary expenses for a rider. Additionally, the unstable and high gas prices in the Philippines mean that "true" efficiency is likely lower than what the app-scraped data suggests.

IV. Conclusion and Future Work

A. Conclusion

This study transformed a personal "hustle" into a data-driven evaluation of platform economics and rider strategy. By analyzing 641 completed trips through a custom automation script, the research identified that the 50 PHP base fare acts as a powerful economic "floor" that dictates earnings efficiency.

While the initial objective was to determine if reaching the Master rank (18% commission) would significantly increase profitability, the statistical results—specifically an Independent Samples T-test p-value of 0.8845—proved that the commission drop did not create a statistically significant difference in efficiency.

Furthermore, the study uncovered an Efficiency Paradox: while short-haul trips are mathematically superior in terms of pesos-per-kilometer, the qualitative experience of the rider favors long-haul trips. This is because long-haul trips provide more consistent cash flow and bypass the "idle time" and psychological cost of "waiting and waiting" for new bookings in heavy traffic. Ultimately, the data confirms that for an independent student rider, the "hustle" is more about managing time and mental fatigue than chasing platform milestones.

B. Future Work

To build on this research, I propose the following improvements:

- Tracking Fuel and Maintenance: Future studies should include "out-of-scope" variables like unstable gas prices and motorcycle maintenance to calculate true net profit.
- Longer Study Period: Extending the study beyond six months would allow for better analysis of seasonal trends, such as increased demand during holidays or rainy seasons in the Philippines.
- Logging Idle Time: While the scraper script successfully automated data extraction, it does not capture the "wait time" between bookings. Integrating a timer to track idle hours would provide a more accurate picture of hourly efficiency.
- Advanced Predictive Modeling: Instead of just clustering past data, future work could use Time-Series Forecasting or Neural Networks to predict the most profitable hours to be on the road based on historical traffic and demand patterns.