

REPORT

Data Analysis for Improving Upfront Pricing Precision

Prepared for:
Blinkit Data Analysis Team

Prepared by:
Kritagya Kashyap
Data Analyst

OVERVIEW

01.

Summary and Introduction

An overview of the report's purpose and a summary of the key findings and recommendations.

Univariate Analysis

02.

Examination of critical individual variables such as distance, duration, metered price and upfront price to identify patterns and trends that impact pricing precision.

03.

Bivariate Analysis

Analysis of relationships between two variables, such as distance vs. upfront price and duration vs. metered price, to uncover insights on price discrepancies.

Discrepancy Analysis & Correlation Insights

04.

A detailed investigation into the deviations between upfront and metered prices, with correlation findings that highlight key factors contributing to pricing inaccuracies.

05.

Recommendations and Conclusions

Data-driven recommendations to improve upfront pricing accuracy, followed by concluding remarks summarizing the key opportunities for enhancement.

Summary

This report provides an in-depth analysis of pricing precision for ride-hailing services, focusing on discrepancies between upfront and metered prices. The analysis utilizes univariate and bivariate methods, discrepancy assessments, and correlation studies to identify key areas for improvement.

Key Findings:

Accuracy of Upfront Pricing: Generally accurate for short rides but less precise for longer distances and durations.

Price Variability: Significant variability in metered prices underscores the need for better real-time adjustments and model calibration.

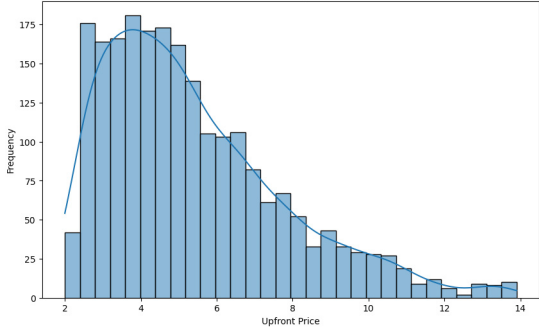
Opportunities: Focus on improving duration and distance predictions, and enhancing real-time data integration to reduce pricing deviations.

Introduction

Upfront pricing in ride-hailing services offers customers an estimated cost before the ride begins. The actual metered price, which reflects the true cost based on distance and duration, can differ from the upfront estimate, potentially leading to customer dissatisfaction. This report analyzes the factors affecting pricing precision and proposes improvements to enhance accuracy

Univariate Analysis

Distribution of Upfront Price



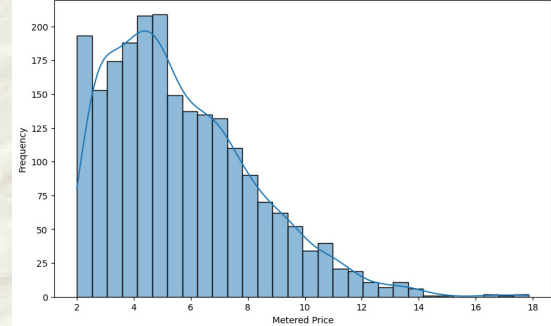
Upfront Price

The upfront price distribution is right-skewed with peaks between \$3 and \$6, indicating most rides are priced lower. There's a sharp decline in frequency beyond \$6, suggesting the model underestimates higher prices.

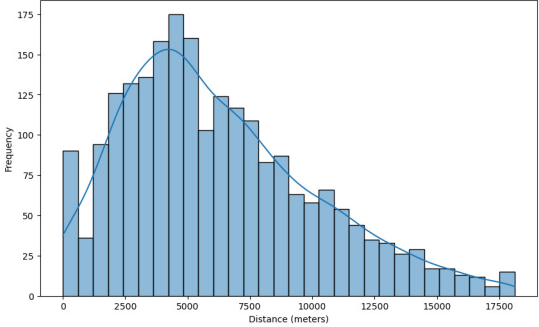
Metered Price

The metered price distribution is also right-skewed but shows higher peak values, especially between \$4 and \$6, compared to upfront prices. The broader tail suggests greater variability due to factors like delays and traffic.

Distribution of Metered Price



Distribution of Distance



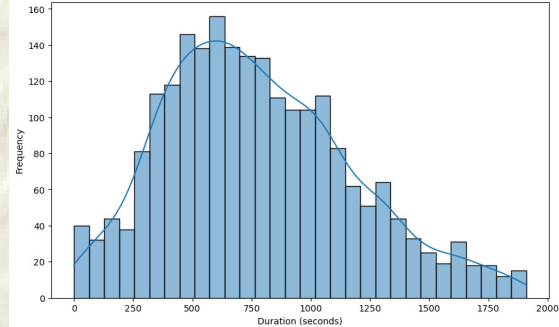
Distance

The distance distribution is right-skewed, with most rides between 2,500 and 7,500 meters, indicating a predominance of shorter rides. This suggests that the upfront pricing model, calibrated for shorter distances, may perform less accurately for longer rides.

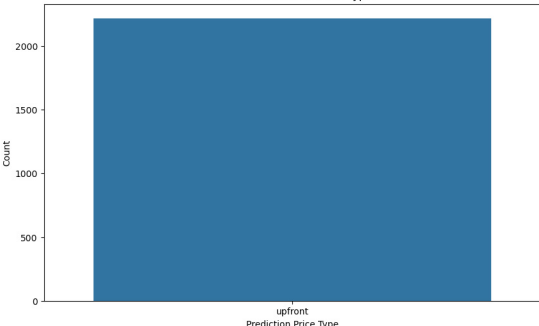
Duration

The duration distribution is right-skewed, with most rides occurring between 300 and 1,100 seconds. This indicates a higher frequency of shorter to moderate duration rides. The skew suggests that the pricing model may need adjustment for longer-duration rides.

Distribution of Duration



Count of Prediction Price Type

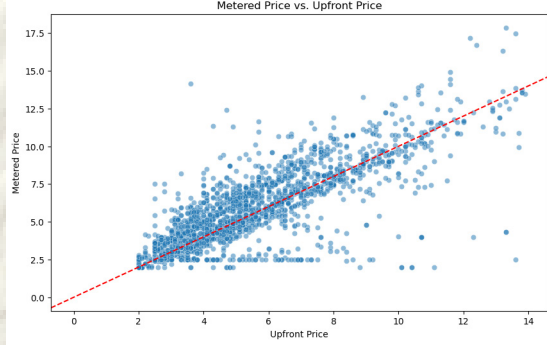


Prediction Price Type

All predictions are upfront, showing a gap in capturing real-time changes affecting ride costs.

Bivariate Analysis

Metered Price vs. Upfront Price

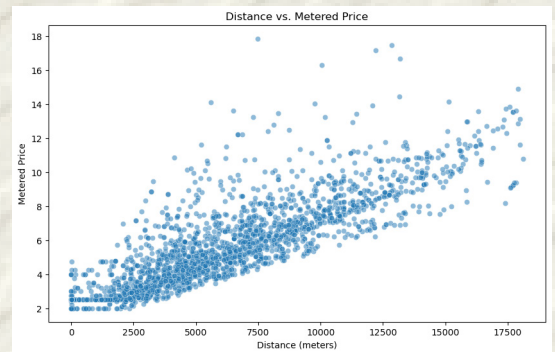


Metered vs Upfront Price

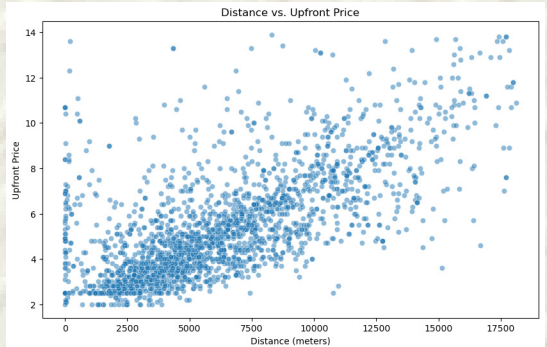
positive correlation between duration and upfront price, with prices generally increasing as duration lengthens. Outliers indicate occasional discrepancies in pricing, suggesting that duration-based pricing adjustments could improve accuracy.

Distance vs Metered Price

Strong linear correlation between distance and metered price, similar to upfront pricing. Despite this, there is significant variability in metered prices for similar distances, likely due to factors like traffic or GPS issues. This variability indicates potential inefficiencies in the pricing model's accuracy.



Distance vs. Upfront Price

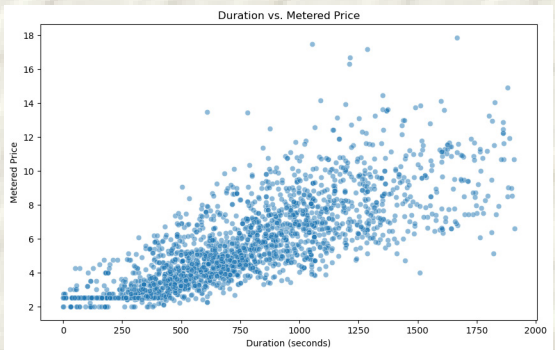


Distance vs Upfront Price

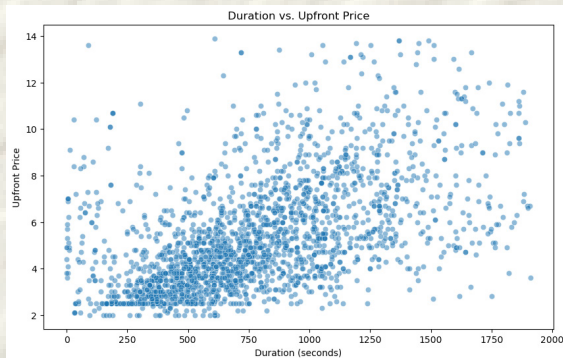
The scatter plot shows a clear linear relationship between distance and upfront price, with longer distances leading to higher prices. However, there's a moderate spread in pricing for similar distances, suggesting that real-time factors like traffic or road conditions could improve pricing precision.

Duration vs Metered Price

positive correlation between duration and metered price, with longer durations generally leading to higher prices. Despite this, similar durations can have significantly different prices, indicating factors like traffic or route efficiency influence price variations.



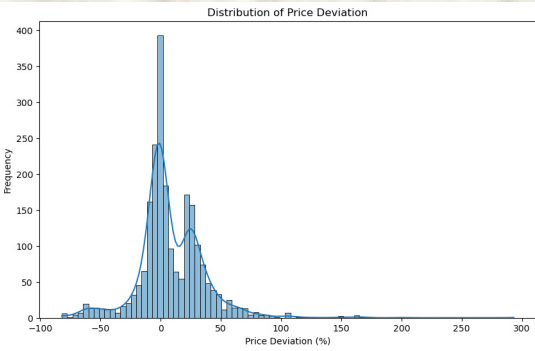
Duration vs. Upfront Price



Duration vs Upfront Price

positive correlation between duration and upfront price, with higher durations typically leading to higher prices. Data points are denser at shorter durations and lower prices, while variability increases with longer durations. Outliers show significant deviations, indicating diverse pricing for extended durations.

Discrepancy Analysis & Correlation Insights

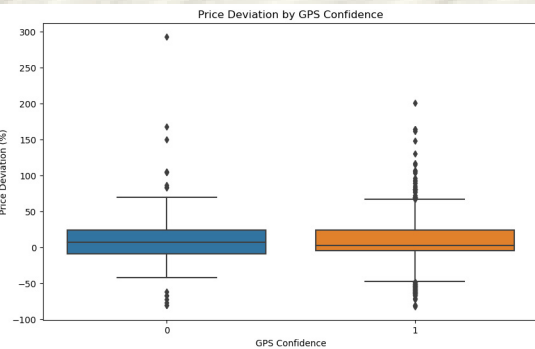
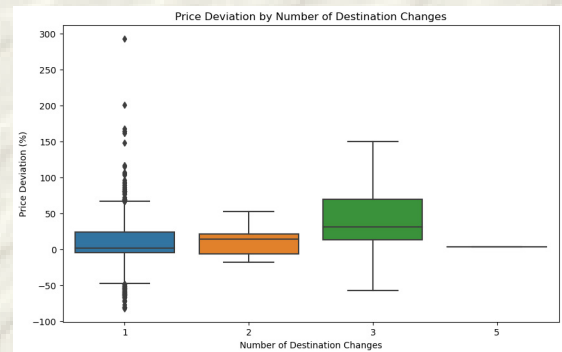


Distribution of Price Deviation

The price deviation distribution is centered around 0%, with a right-skewed tail indicating occasional large price increases. There are frequent minimal deviations and smaller peaks for both negative and positive deviations, with outliers mostly reflecting higher positive deviations.

Price Deviation by no. of Destination change

Price deviation is highly variable with one destination change and decreases with more changes. The median deviation varies, increasing with more changes. Limited data for five changes suggests a need for more data to better understand pricing impacts.



Price Deviation by GPS COnfidence

The median price deviation is similar across GPS confidence levels, around 0%. Lower GPS confidence levels show greater variability and more extreme outliers, indicating pricing inconsistencies. Improving GPS accuracy could help reduce these discrepancies.

Metered Price vs. Distance/Duration: Strong positive correlation, indicating proportional increases with ride length.

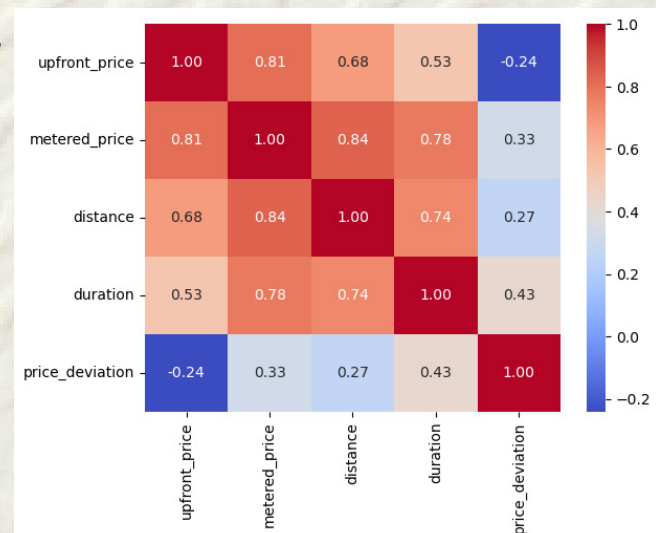
Upfront Price vs. Distance/Duration: Moderate positive correlation, showing some accuracy, but weaker than metered pricing.

Price Deviation vs. Duration: Moderate positive correlation, indicating longer rides have more significant price deviations.

Upfront Price vs. Price Deviation: Weak negative correlation, suggesting smaller deviations at higher upfront prices.

Metered Price vs. Price Deviation: Moderate positive correlation, with higher metered prices leading to larger deviations from upfront prices.

Correlation Matrix



Recommendations

Top 2 Opportunities to Improve Upfront Pricing Precision:

Enhance Duration Estimation:

Since ride duration has a moderate correlation with price deviation, refining duration predictions can help reduce pricing discrepancies. Incorporating real-time traffic data, peak-time adjustments, and machine learning models will improve accuracy, particularly for longer rides. This will lead to more consistent upfront pricing and fewer deviations from metered prices.

Refine Upfront Pricing for High-Cost Rides:

High metered price scenarios often show greater discrepancies, affecting customer satisfaction. A more dynamic upfront pricing model that adapts to real-time traffic, route changes, and rider behavior can better reflect actual ride costs. This will improve pricing precision for higher-cost rides, ensuring more accurate upfront estimates.

Conclusions

- To improve upfront pricing precision, the most impactful steps include enhancing duration predictions and refining the model for high metered price rides.
- By integrating more real-time data and adjusting for dynamic ride conditions, Blinkit can reduce discrepancies between upfront and metered prices, leading to increased pricing accuracy and better customer satisfaction.
- Additionally, improving the reliability of GPS data will further reduce pricing variability.