Optimizing Upfront Pricing Accuracy: Insights and Opportunities from Ride-Hailing Data Analysis

DIVYANSH JAIN

➢ Introduction:

In the highly competitive ride-hailing industry, accurate **upfront pricing** is crucial for providing a seamless and transparent customer experience. Upfront pricing allows customers to see the estimated cost of their ride before booking, based on predicted distance and time. However, when the actual ride cost significantly deviates from the predicted price—typically by more than 20%—customers may be charged the metered price instead, leading to potential dissatisfaction and complaints.

This report analyzes key factors that contribute to discrepancies between the predicted upfront prices and the actual metered prices. By examining a dataset of ride-hailing transactions, we identify the main causes of these discrepancies and highlight actionable opportunities to enhance pricing precision. The **goal** is to **improve prediction accuracy**, **reduce customer complaints**, and ensure more consistent pricing for both riders and drivers.

The findings presented are intended for both technical teams focused on **refining pricing algorithms** and business stakeholders interested in **improving customer satisfaction** and **operational efficiency**.

• Data Analysis:

The dataset used in this analysis consists of ride-hailing transaction records, including details of both upfront price predictions and the actual metered prices. Each record represents a unique ride and contains variables relevant to the pricing model, ride details, and customer interactions.

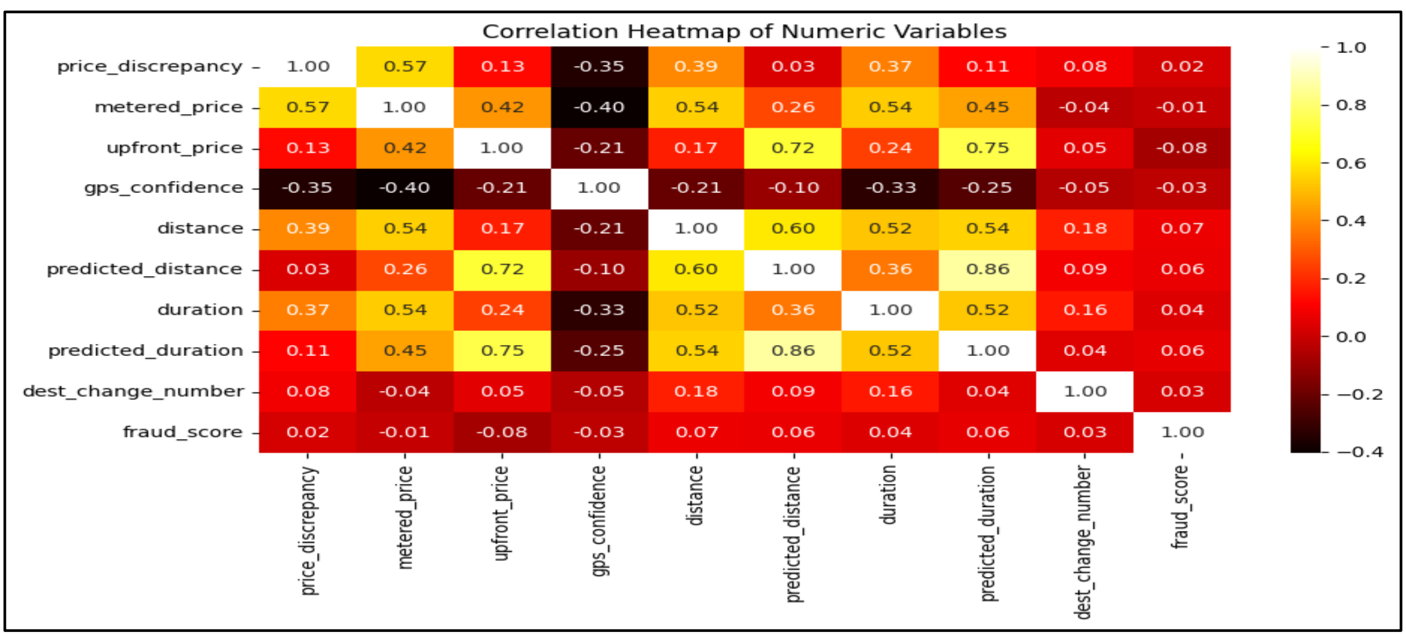
➢ Data Statistics:

Variable	Mean	Median	Mode	Standard Deviation
Price Disrepancy (%)	22%	5%	0%	54%
Upfront Price	4160	7	3	17016
Metered Price	7998	13	6000	15816
Distance	9769	7140	0	10912
Predicted Distance	8823	6918	0	10549
Duration	1566	1054	0	1650
Predicted Duration	1107	939	764	806
GPS Confidence (0 or 1)	0.8	1	1	0.4
Destination Changes	1.12	1	1	0.5

The **mean metered price** is much **higher** than the **mean upfront price**, indicating that, on average, actual ride costs tend to exceed the predicted prices, thus resulting in **>20% mean price discrepancy**. GPS signal was reliable for 80% of rides, but **poor GPS quality** may have contributed to pricing inaccuracies in **20% of cases**. The **mean of destination changes** is **1.12**, with most rides having minimal destination changes.

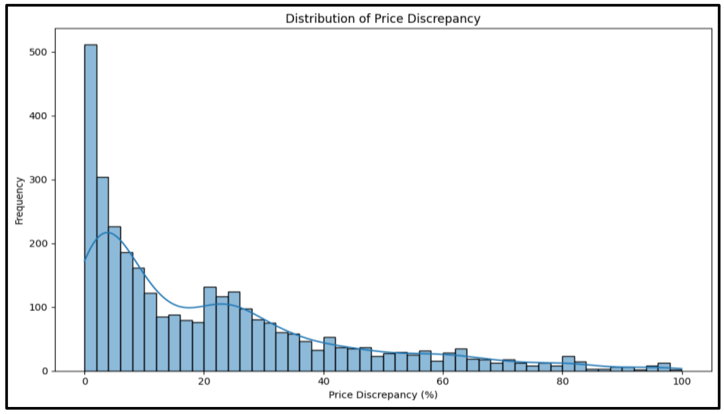
➢ Analysis of Discrepancies:

This **Correlation Heatmap** visualizes the relationship between various numerical variables in the dataset. The color intensity and the values in each cell represent the strength (0 to1) and direction (positive or negative) of the correlation between two variables.

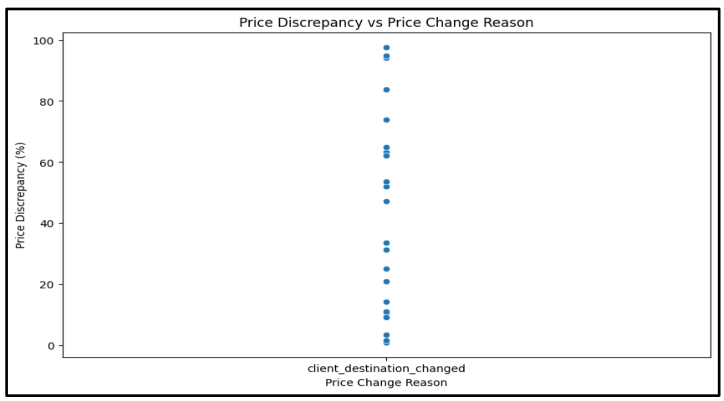


There is a **negative correlation** between **price discrepancy** and **GPS confidence**, indicating that **poor GPS signals** are linked to **higher pricing discrepancies**, likely due to route inefficiencies. Trips with **lower GPS confidence** tend to have **longer Durations**. **Fraud risk** shows **no significant impact** on pricing errors, while **Destination** changes do **slightly increase price discrepancies**, the effect is weak and not a major contributor.

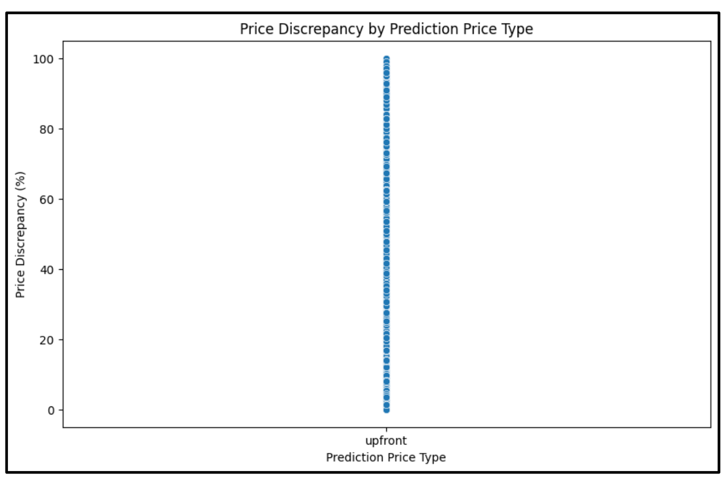
A **Histogram of price discrepancy** (limited to discrepancies less than 100%, with other values treated as outliers) shows that the distribution is **skewed to the right**, indicating that in many cases, the metered price tends to be in range of 20% of the upfront price, leading to the 22% mean price discrepancy.



The **Scatter plot** analyzing price discrepancies (limited to discrepancies less than 100%, with other values treated as outliers) reveals a clear trend: price changes are predominantly associated **with alterations in the destination made by the client.**



The Scatter plot examining price discrepancies (restricted to those under 100%, with outliers excluded) shows that they are primarily linked to the upfront prediction price type. This finding suggests that the algorithm used to determine upfront prices may need to be reviewed and adjusted to improve accuracy.



- ## Opportunities for Improvement:

  Based on the analysis of price discrepancies and factors influencing them, key opportunities for improving upfront pricing precision have been identified:

➢ ### Improving Real-Time GPS Data and Refining Traffic Predictions for Long-Distance and Peak-Hour Rides:

Issue: Low GPS confidence affects 20% of rides and is linked to higher price discrepancies and route inefficiencies. Longer rides and peak-hour trips show greater price discrepancies, with a correlation of 0.39 between distance and discrepancy. This suggests that the current model, which heavily relies on predicted distance and duration, underperforms in high-traffic conditions or when actual trip duration exceeds predictions.

Opportunity: Incorporating real-time GPS and traffic data, along with historical congestion patterns and advanced machine learning algorithms, can significantly improve the accuracy of predicted durations and distances. These enhancements would reduce price discrepancies, especially during peak hours and for long-distance trips, leading to more reliable upfront pricing.

➢ ### Improving Price Adjustments for Destination Changes and Enhancing Customer Communication:

Issue: Although the correlation between destination changes and price discrepancies is weak, rides with multiple destination changes show a noticeable increase in discrepancies, indicating that mid-ride changes pose challenges to the upfront pricing model.

Opportunity: Introducing a flexible, transparent pricing adjustment mechanism for destination changes and providing real-time fare updates can enhance customer satisfaction. Clear communication about potential price variations, especially for longer rides or during poor GPS conditions, can manage expectations and reduce overpayment complaints.

- ## Conclusions and Recommendations:

  Pricing discrepancies are mainly influenced by GPS confidence, ride length, and peak traffic while destination changes and fraud have minimal impact. To improve accuracy, integrate real-time GPS and traffic data, adjust for signal issues, refine predictive models, and implement dynamic pricing for destination changes.

  Enhance communication about fare variations and regularly update algorithms to ensure precision.