# Objective

The main objective of this research project is to understand the causes of overpayment incidents in the ride-hailing services. It is with this understanding that correlations between ride attributes are to be analyzed in their relation to overpaid ride tickets in order to pinpoint areas which need to be addressed in an effort to minimize overcharging occurrences.

# Features Selected for Analysis

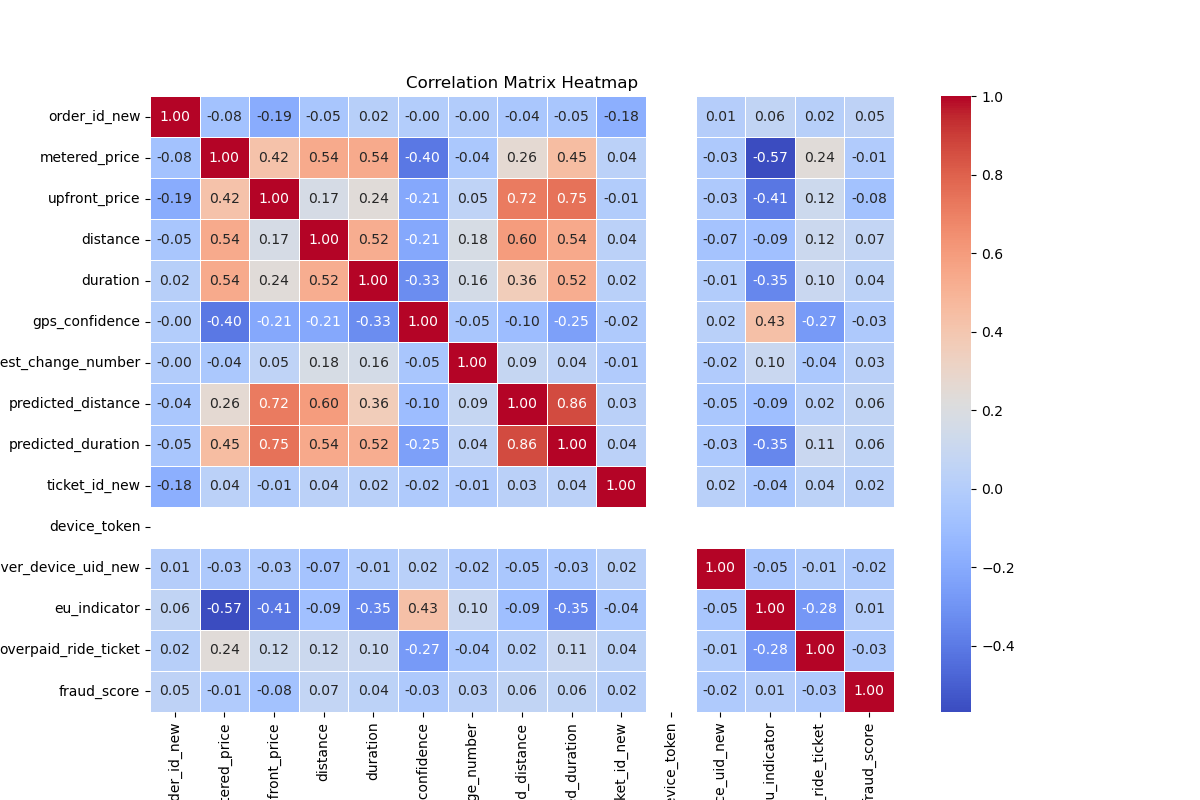
This report focuses on analyzing the following features to understand and address overpayment incidents in ride-hailing services:

* **Metered Price:** The final charge for a ride.
* **Upfront Price:** The initial price charged upon selecting the car.
* **Distance:** The total distance of the ride.
* **Duration:** The total duration of the ride.
* **Predicted Distance:** The estimated distance of the ride.
* **Predicted Duration:** The estimated duration of the ride.
* **GPS Confidence:** The confidence level in the GPS data accuracy.
* **Destination Change Number:** The number of destination changes during the ride.
* **Fraud Score:** A score indicating the likelihood of fraudulent activity.

These features were selected based on their relevance to ride pricing and the potential impact on overpayment incidents.

# Analysis using Correlation Matrix

The correlation matrix makes it easier to understand the quantitative levels of pairwise linear connection between different variables in a given data set. These correlations can vary between -1 and 1 depending on what some of the perfect correlations mean; a perfect positive correlation being 1, a perfect negative correlation is -1 while an absence of correlation is 0.



The matrix signifies mere and complex indefinite relations between metered price scheme, distance, time taken, GPS precision, and the probability of overpaid ride tickets. Recognition of these relationships is important in finding out the source of overpayment in ride-hailing services.

# Key Findings

## Metered Price

The metered price, which is the final charge for a ride, shows significant positive correlations with several key variables:

### Distance (0. 539)

As expected, distance and metered price are positively related suggesting that the longer the distance of a ride, the higher the charges. This is as it should be with most distance based pricing paradigms.

### Duration (0. 536)

In the same way, it shows that the time spent on a ride is positively related to a metered price, this means that longer times contributing to the fare price.

### Upfront Price (0. 421)

There is still a moderate positive association between the initial price that one is charged upon selecting the car and the metered price meaning that there exists some degree of accuracy with a possibility of divergence.

In fact, metered price has a very strong negative correlation with GPS confidence (- 0.404). This implies that where confidence in GPS is low, the metered prices could be higher this could be as a result of wrong estimates of the distance and time of the ride offered. This could also lead to wrong fare determination which may lead to over-charging of the passengers.

## Overpaid Ride Tickets

The overpaid ride ticket variable, which indicates whether a rider has complained about overpayment, reveals several important correlations:The overpaid ride ticket variable, which indicates whether a rider has complained about overpayment, reveals several important correlations:

### Metered Price (0. 236)

It can be observed that there is direct proportion between the metered price and overpaid ride tickets. This implies that rates which are higher than the metered prices are likely to attract complains of over-payment. As you can see, the correlation is not very high, but it proves that with increasing fares customers show discontent and can turn into disputes.

### Distance (0. 117) and Duration (0. 100)

Distance and duration of a single journey show only a weak positive relationship with overpaid ride tickets. This essentially means that chances of overpayment complaint are slightly higher with longer distance or time travel. These correlations though weak suggest that the complexity or number of events involved in a ride maybe a factor in pricing mistakes or dissatisfaction that riders experience.

### GPS Confidence (-0. 269)

It is also evident that GPS confidence is strongly negatively related with overpaid ride tickets. This is an important discovery because it indicates that rides for which GPS has low confidence level will likely lead to overpayment disagreement. Travel through GPS may sometimes have an inaccurate measurement in terms of distance and therefore alter effective fares for longer distances as well as may result in overcharging.

## GPS Confidence

GPS confidence is a key factor that influences both metered price and the likelihood of overpayment:

### Negative Correlation with Metered Price (-0. 404)

As described before, metered prices increase as the GPS confidence decreases. This means that even when GPS tracking is used, it is not accurate and therefore over-charging may result due to wrong calculations.

### Negative Correlation with Overpaid Ride Tickets (-0. 269)

The negative correlation between GPS confidence and overpaid ride tickets is showing that the accurate GPS data is very essential. Old GPS tracking may also contribute to the incidents of overpayment as it tends to produce mistakes in distance and time estimates of a ride.

## Destination Changes

The number of destination changes during a ride is also loosely connected to metered price (-0. 036) and overpaid ride tickets (-0. 037). This implies that changes in destination do not affect price or effect overpayment complaints in any way. Although, the data reveals that destination changes affect overpayment insignificantly while complicating the pricing formula.

# Results and discussions

The correlation matrix highlights several key insights that are crucial for understanding and addressing overpayment issues in ride-hailing services.

## Impact of GPS Confidence on Overpayment

As a result, GPS confidence is found to be significant in influencing the accuracy of the ride pricing. The negative correlations with GPS confidence of metered price and overpaid ride ticket suggest that the smartphone GPS inaccuracy is the main factor behind the overpayment cases. In such cases the GPS data can get skewed and the system can get confused as to exactly how long and how far a ride is, and therefore offer a wrong estimate of the fare. This may lead to higher metered prices and also increases the chances that overpayment complains from the riders may emerge.

## Correlation Between Price and Overpayment

The evidence that the relation between metered price and overpaid ride tickets is positive but weak which means that high fares do not solely lead to overpayment claims. Nevertheless, there could be influences that stem from other factors for example GPS precision and the nature of the trip in question. The relatively moderate positive correlation between the anteriority price and the metered price also suggests that while the prediction model is highly reliable most of the time it can be off sometimes especially when GPS data is sketchy.

## Role of Ride Complexity

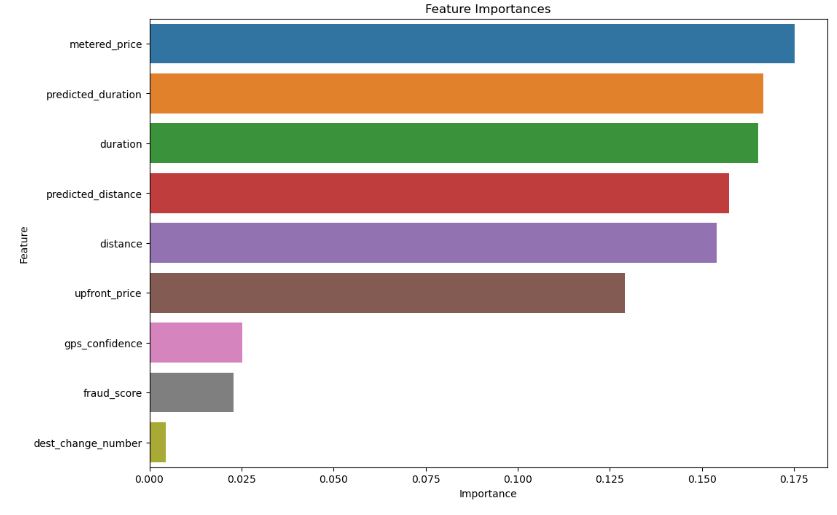
Weak correlation coefficients of distance, duration, and overpaid ride tickets indicate that longer rides may result in overpayment, but the extent is negligible. These observations also suggest that potentially, pricing errors are even higher when simpler rides are not involved – changes of destinations are multiple or the routes detected with the GPS have low signal quality. Solving these problems may entail better solutions to GPS and new and enhanced algorithms for handling peculiar ride conditions.

# Analysis using Random Forest Model

Random Forest is an algorithm that is developed from the concept of taking several decision trees to perform better as well as to be more accurate in classification. During the training process, it creates many decision trees and combines the results of these trees for better performance. This technique proves useful especially in projects such as ours that require determination of feature importance. Specifically in the context of our ride-hailing overpayment analysis Random Forests can easily work with numerous features and their interdependencies and a total measure of ‘feature importance’ is given which enlightens us about features on which aspects overpayment cases heavily rely.

## Reults

|  |  |
| --- | --- |
| **Feature** | **Importance** |
| Metered\_price | 0.175 |
| Duration | 0.165 |
| Distance | 0.154 |
| Predicted\_duration | 0.167 |
| Predicted\_distance | 0.157 |
| Upfront\_price | 0.129 |
| Gps\_confidence | 0.025 |
| Destination\_change\_number | 0.004 |
| Fraud\_score | 0.023 |



## Discussion

This bar plot represents the feature importance of the Random Forest Classifier which gives an idea about the importance of various features which might be useful for ticket categorisation related to overpayment. Based on the results, metered price, predicted duration and duration are quite give the higher importance score out of one, which means metered price has higher impact on click through rate. This suggests that fluctuations in metered price of a ride are the sole determinant of overpayment more than any other factor. There is an almost identical pattern between Predicted\_duration and duration suggesting that deviation in expected and actual durations also strongly affect overpayment cases. On the other hand, we can come across low-impact features as dest\_change\_number and fraud\_score for approximately possessing importance scores of 0. 004 and 0. 023, respectively. The main opportunities of the graph are to decide which of the features should be utilized when addressing overpayment potential issues in the ride-hailing context, thus, orientating interventions to the maximal overcharge eliminations.

# Measures that Must be Taken in Order to Minimize Overpayment According to the Conclusion

Based on the comprehensive analysis of correlations and feature importance from the Random Forest model, several targeted recommendations can be made to reduce the incidence of overpayment in ride-hailing services:

## Enhance GPS Accuracy

### Recommendation

Spend more in developing GPS and enhancing the accuracy of the system.

### Rationale

The study showed that GPS confidence is highly negatively related with metered prices and overpaid ride tickets. The odds of having overpayment and imprecisions in the fare estimations are highly related to having low GPS confidence. You may increase GPS accuracy as the best way of increasing distance and duration measurement certainty to prevent overpayments.

## Improve Fare Prediction Models

### Recommendation

Implement smarter algorithms that will be used to predict the duration of a ride and the number of miles to be traveled.

### Rationale

Further, the results showed that predicted duration and predicted distance should be included into overpayment influential factors. Solving for the parameters that enhance these predictions will narrow down their gap with actual values to minimise deviations in fare estimations. This form of improvement could include using more complex algorithms in the predictive models, or including more data to the model for calculations.

## The pricing mechanisms of an organisation should be reviewed and adjusted for a number of reasons.

### Recommendation

It is necessary to conduct constant appraisals and modifications of the price determination procedures in order to reflect the real attributes of rides.

### Rationale

Regarding overpayment, the metered price was identified to be a key factor. Eliminating inconsistencies that can be caused by setting the metered price is possible if the price is set to match the real cost of the ride, which is accurately estimated by the system. Thus, the problem of firms setting ‘unreasonable’ fares can be mitigated by integrating automated checks and balances into the pricing system to make changes based on real-time data.

## Enhance Transparency and Communication

### Recommendation

There should be clear information regarding the fares which are charged and clear communication of the likelihood of a fare being hiked.

### Rationale

A moderate correlation between the initial cost and the metered cost means that disparities between these values could cause customers’ dissatisfaction and further, they may cause disputes about overpayment. Giving clear information of how fares are computed and informing the passengers in the right manner in the event of rate changes will also prove beneficial.

## Monitor and Address Outliers

### Recommendation

Use overview procedures to check for irregularities in the journey and place corresponding measures to correct them, when needed.

### Rationale

Some of the factors that seem to have a negligible effect on over payment which we may not necessarily monitor are factors such as Destination changes When creating the system it is however useful to monitor for such anomalies in ride data. It may be beneficial to approach these particularities as threats in advance, because their effect, even if it is negligible at the moment, may bring about factors that can lead to overpayment.

By considering such areas of GPS accuracy, the fare prediction models, pricing policies, transparency and monitoring in particular, the ride hailing services can significantly cut the chances of overpayment and increase the rate of consumer satisfaction.