**Ride-hailing Analysis report**

**Step 1: Understanding the Problem:**

**The main objective is to improve the upfront pricing precision for ride-hailing services to minimize the difference between the upfront price (the predicted price) and the metered price (the actual price). Key factors contributing to the price discrepancy will be identified to propose actionable insights**

**Step 2: Exploring the Data:**

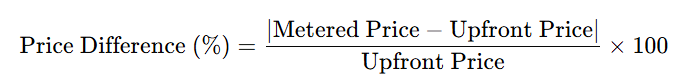
**Variables of interest include:**

* **order\_id\_new and order\_try\_id\_new: Unique identifiers for rides.**
* **calc\_created: Time of order creation.**
* **metered\_price: Actual ride price.**
* **upfront\_price: Promised price.**
* **distance and duration: Actual ride details.**
* **predicted\_distance and predicted\_duration: Predicted ride details.**
* **gps\_confidence: Indicator of GPS accuracy (which could affect price prediction accuracy).**
* **dest\_change\_number: Number of destination changes during the ride (could contribute to price variance).**
* **prediction\_price\_type: Indicates when the prediction happened (before or after destination change).**
* **fraud\_score: Indicator of rider's fraud risk (could impact overpayment or price manipulation).**

**Step 3: Analysing Key Metrics**

**3.1 Calculate Price Discrepancy**

**Calculate the percentage difference between the upfront and metered prices:**

****

**Filter out cases where the price difference exceeds 20% (the defined threshold)**

**3.2 Segment Analysis:**

* **Segment by GPS Confidence:** Determine whether rides with poor GPS confidence lead to higher price discrepancies. Poor GPS signals could affect distance and duration prediction accuracy.
* **Analysis**: Compare average price difference for rides with good vs. bad GPS confidence.
* **Destination Change Analysis:** **I**nvestigate the impact of destination changes on price discrepancies**.**
* **Compare rides where the destination was changed (i.e., dest\_change\_number > 1) vs. those that weren't changed.**
* **Look at how upfront vs. after-destination-change predictions differ in accuracy.**

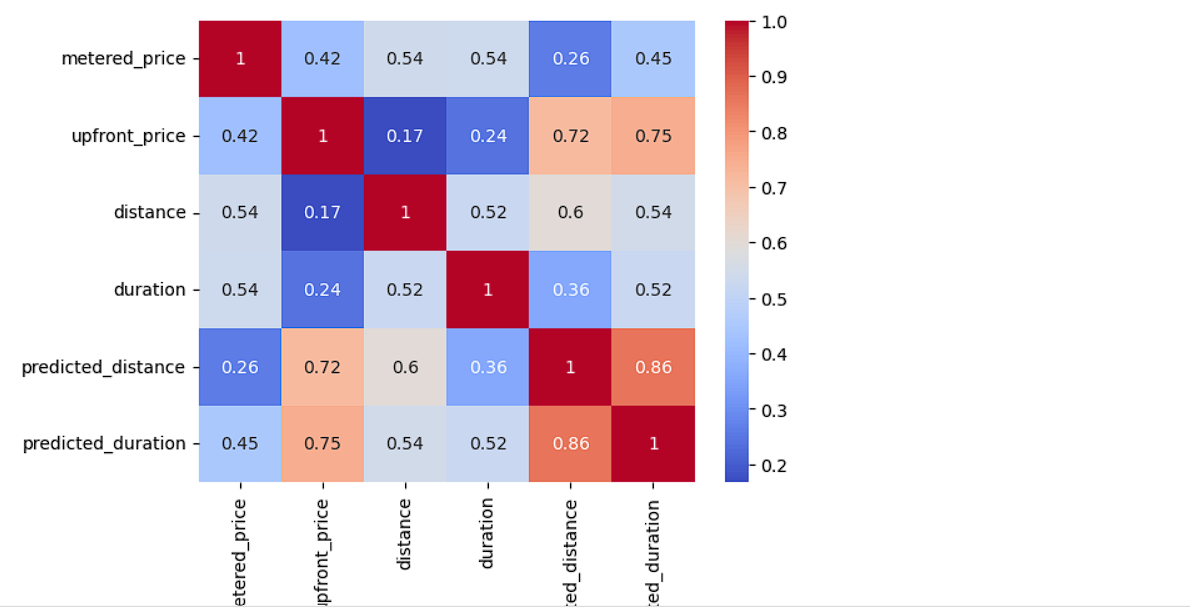
**3.3 Fraud Score Analysis**

* **Overpayment and Fraud**: Investigate if a higher fraud score correlates with overpayment complaints and higher price discrepancies.

**Analysis**:

* + Correlate fraud score with price discrepancies and overpaid ride tickets to check for patterns

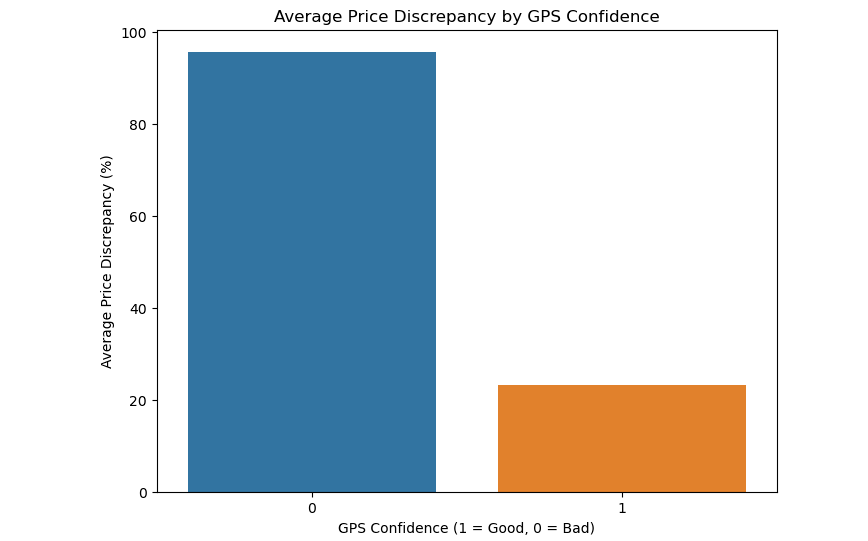
**Steps:4 Visualising the datasets.  
 Finding the Correlation between the data variable:**

****

From this correlation diagram we can conclude that Actual price and predicted price is positively correlated distance and duration is highly positively corelated .

We will conclude that actual pricing and predicated pricing is corelated **40%**

**GPS Confidence Analysis:**

****

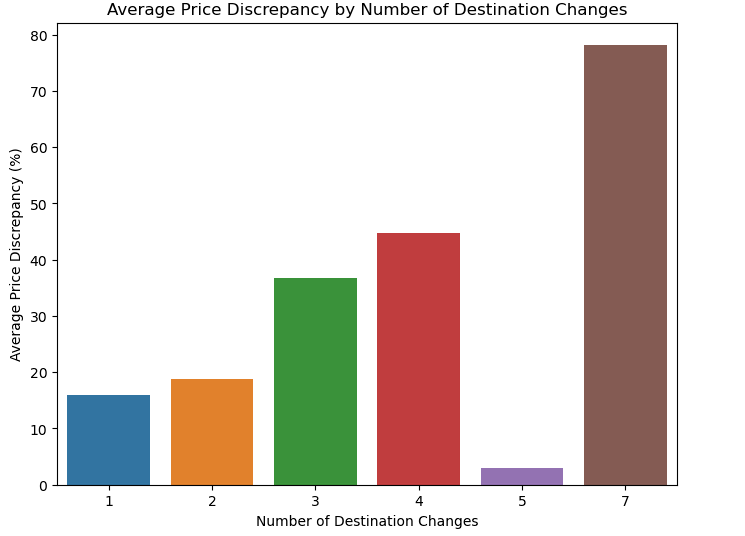
**We will conclude that Bad GPS Confidence can create 95% Price Variance between actual price and preditcted price.**

**gps\_confidence Avg\_price\_diff\_pct**

**0 Bad 95%**

**1 Good 23.5%**

**Destination Change Impact:**

****

We will conclude that by looking the average price varicenace is highly affected due to destination changes as the user changes **7 times** due to which it’s price varies **80% .**

**dest\_change\_number price\_diff\_pct**

**0 1 16.024845**

**1 2 18.873239**

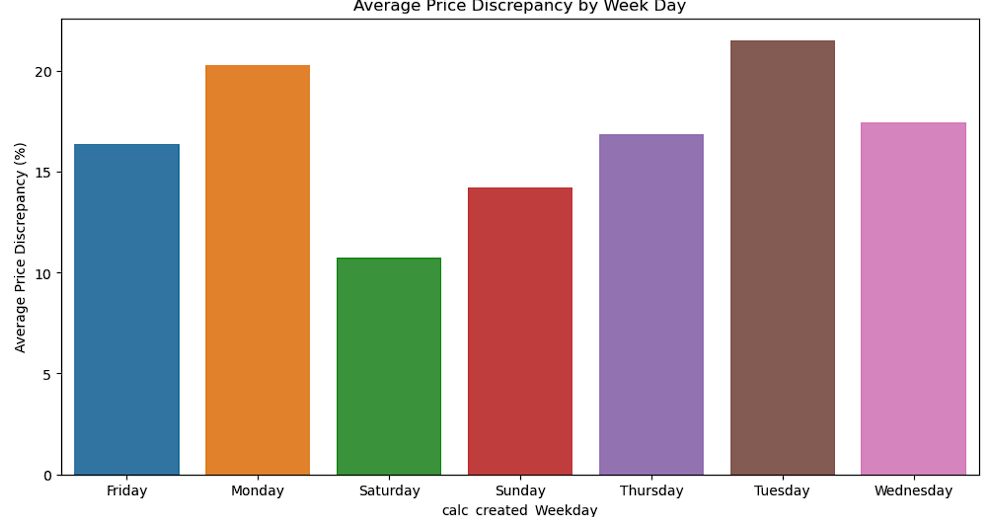
**2 3 36.763728**

**3 4 44.790761**

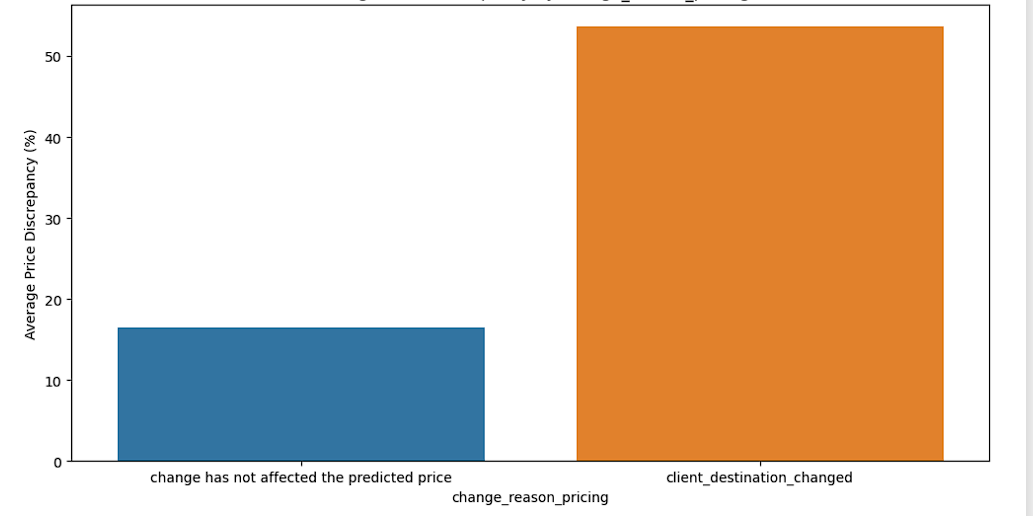
**4 5 2.893103**

**5 7 78.221525**

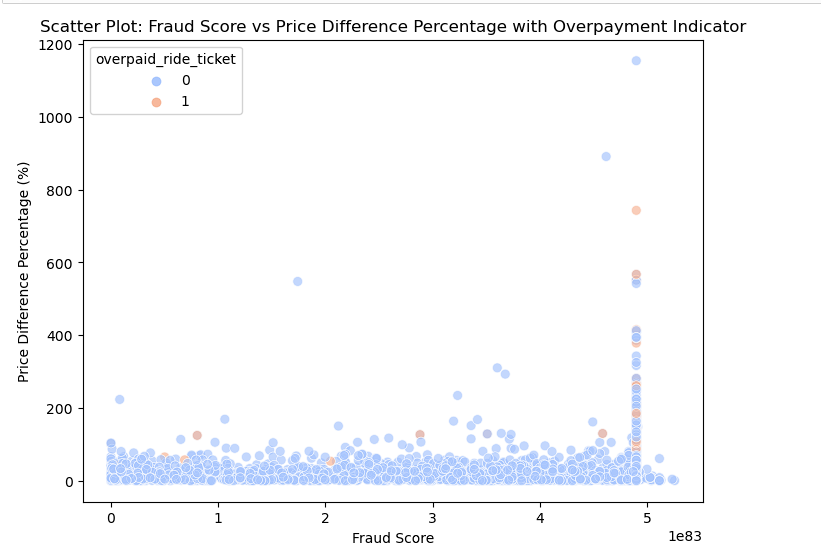
**Week Day Impact:**

**  
As we can see week day does not create any significant impact on price variance between actual and predicted price**

**Change Reason Impact:**

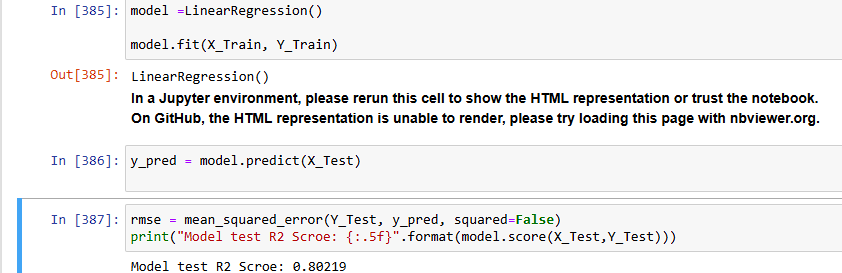
****

As we have check that the **change destination** is occur by the client and which cause the  **price variance upto 55%**

**Fraud Score Impact:  
**

**The graph shows that most transactions are likely non-fraudulent, the outliers with high price differences and fraud scores highlight potential areas for further investigation to prevent and detect fraud effectively. The data suggests that both metrics should be considered together to better identify and understand potentially fraudulent transactions.**

**The analysis of the graph indicates that while most transactions show low fraud scores and minimal price differences, outliers with high price differences and fraud scores suggest potentially fraudulent activity, warranting further investigation.**

**Data points with a higher fraud score (closer to 5 or 5000 on the scale) tend to have a higher Price Difference Percentage. This trend suggests a correlation where higher fraud scores might be associated with greater price variances  
  
Model Score  
  
  
  
Improvements  
  
A. Identifying Opportunities for Pricing Precision**

* **Analyze Discrepancy Patterns: Identify patterns in large price discrepancies (e.g., destination changes or poor GPS signal) and propose model improvements.**
* **Customer Impact: Highlight cases where customers are paying more or less than the upfront price, impacting customer satisfaction.**

**B. Recommendations**

* **Destination Changes: Encourage customers to finalize destinations early or provide upfront warnings about the price impact of changes.**
* **GPS Improvement: Invest in improving GPS confidence, especially in areas where poor signals lead to inaccurate predictions.**

**Github code\_link:**