**Data Enrichment Recommendations**

As part of our data strategy for the Used Vehicles platform, it's important to consider the potential areas of additional data that can significantly enhance the value of our analyses. By incorporating these datasets, we can gain deeper insights into the temporal patterns of vehicle listings and provide more informed recommendations for pricing, demand-supply dynamics, and market trends. Below are a few key areas of additional data that we recommend integrating into our analyses:

1. **Vehicle Features and Specifications Data:**

**Why it's essential:** Having access to detailed information about vehicle features and specifications, such as engine type, horsepower, interior/exterior features, and safety options, would empower us to conduct more granular analyses and understand how specific features influence pricing and demand.

**Analyses enabled:**

*- Feature Correlation Analysis*: Explore correlations between individual vehicle features and their impact on listing prices.

*- Segmentation by Features*: Segment vehicle listings based on specific features to identify market preferences and demand trends.

*- Price Prediction Modelling*: Develop predictive models that estimate vehicle prices based on comprehensive specifications.

1. **Seller Ratings and Reviews:**

**Why it's essential:** Integrating seller ratings and reviews from the platform can offer insights into the credibility of sellers and the satisfaction of users, contributing to informed decision-making.

**Analyses enabled:**

*- Seller Impact on Sales:* Examine how seller ratings and reviews influence sales and pricing dynamics.

*- User Sentiment Analysis:* Perform sentiment analysis on seller reviews to understand user experiences and identify potential issues affecting sales.

**Issue: Unbounded "Description" Column Length**

One of the issues I identified was the presence of an unbounded character length in the "description" column. This issue led to challenges where the content within this column extended beyond its designated area, causing interference with adjacent columns. Consequently, I had to exclude a portion of the data, resulting in a loss of approximately 20% of valuable information.

**Suggested Rectification**

To mitigate this issue and optimize data quality, I recommend the following rectification:

- *Column Character Limit:* Implement a character limit for the "description" column to prevent content from exceeding its designated space. This limit should be based on an analysis of the average and maximum description lengths to ensure adequate coverage while avoiding interference with neighbouring columns.

*- Data Validation:* Apply data validation mechanisms during data ingestion to ensure that new entries adhere to the defined character limit for the "description" column.