BLACK PEAK TECHNOLOGIES®

Data Scientist Technical Assessment

Dataset: Please use the dataset publicly available here.
Note: *Unless otherwise stated, please use 'price' as the target variable.*

Problem statement:

- Perform exploratory data analysis and select three key plots, explain your takeaways. Please note that there is no need to explain how you have generated your plots, only the inference and findings need to be explained.
- Perform multiple correspondence analysis (MCA) on the dataset and briefly describe your findings.
- Build a model for predicting the price of the automobile. Explain your choices on encoding, model selection etc.
- Explain the high-level plan of action if the above model is to be used as the backend for a web application for B2B use at an automobile manufacturing firm. Please write in bullet points the steps needed and explain your technological choices.
    - *Bonus points if you can make an actual web application using MLFlow, Streamlit (or any framework of your choice) which takes key data points as inputs and provides the predicted price as the output to the user.*
- For a simple LLM (large language model) use case, create an LLM application where you train the given data and ask it to create new rows.

Please submit your code using a Git repository with an appropriate Readme.