**Introduction**

Thank you for your interest in our data science position supporting BCA Engineering Analytics. In order to help us get to know your technical capabilities and style better, we invite you to submit a response to the problem posed in these materials. Because much of Boeing’s proprietary data requires domain specific knowledge to understand fully, the problem proposed here uses third-party data and is an application from the used car industry. Despite this, the underlying skills involved in this application are highly relevant to our position and we invite you to use this opportunity to demonstrate to us what you can do!

**Problem Statement**

Included in these materials please find two comma separated values (.csv) files: Training\_dataset.csv and Test\_dataset.csv

Included in the training dataset is information on used cars previously sold. Each row corresponds to one used car listing. The first column of the data contains a unique identifier for the listing. The next twenty-six columns contain information on parameters relevant to the transaction, with those parameters described in more detail in the appendix attached. Finally, the last two columns of the “Training\_dataset.csv” contain information on “Vehicle Trim” and “Dealer Listing Price”, which describe the trim of the vehicle involved in the sale, and the price at which the vehicle was listed by the dealer.

Your challenge is to build one or more models, through whatever means you find most appropriate, capable of **predicting vehicle trim and dealer listing price** given the other twenty-six variables provided.

**Instructions**

1. Model the problem using whatever means you consider best. **Your work is expected to be entirely your own**. You may consult any resource or reference of your choosing to aide in solving the problem, but the work must be entirely yours. Please reference any resources you use in the write-up covered in step 5.
2. If you use a software package to assist you, please include **ALL** of your original source code in its entirety, and submit it to us **EXACTLY** following the instructions in Steps 5 and 6. Please also include information about which package you used and why in your brief problem write-up.
3. Do not use or add data from any third-party sources, such as internet car estimating tools, to the data provided. At your discretion, some or all of the provided data in “Training\_dataset.csv” may be used, omitted or manipulated in any way during modeling, but no additional data may be added from outside sources.
4. Once your model is built, use it to make predictions on **EACH** of the 1,000 vehicle listings included in the “Test\_dataset.csv” file. Your output should be a comma separated values (.csv) file with one-thousand rows and three columns. The first column should be the unique identifier for the listing. The second column should be your predicted value for vehicle trim. The third column should be your predicted value for dealer listing price.
5. Please submit a brief write-up of no more than 500 words describing the approach you selected and why. Please save your response as a PDF if possible. Please copy any source code from Step 2 and paste it as text into an appendix at the end of your write-up.
6. Return your submission to us by replying back to the original email before the date and time specified in that email. Please attach:
   1. The CSV containing your predictions from Step 4
   2. The PDF containing your write-up
   3. Source code from step 5 to the email.

**Evaluation**

Your models will be evaluated on several objective criteria such as overall R-squared and Area Under the Curve (as appropriate). We will also be looking at some qualitative aspects of the code considering the overall approach and style. The submitted approach and outcomes will be discussed during the technical interview with the panel.

Good luck and thank you for your interest in our position!

**Appendix Data: Parameter descriptions and background data on the car models in question.**

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Type** | **Description** |
| ListingID | int64 | Unique key that identifies each listing |
| SellerCity | object | Seller city |
| SellerIsPriv | bool | Boolean that indicates if the listing if from a private seller |
| SellerListSrc | object | Seller listing source identifier |
| SellerName | object | Seller name |
| SellerRating | float64 | Seller rating (continuous over [0,5] with 5 being a favorable rating) |
| SellerRevCnt | int64 | Seller review count |
| SellerState | object | Seller state |
| SellerZip | float64 | Seller zip code |
| VehCertified | bool | Boolean that indicates if the listing has a manufacturer certification (generally indicates extended warranty) |
| VehColorExt | object | Vehicle exterior color |
| VehColorInt | object | Vehicle interior color |
| VehDriveTrain | object | Vehicle drivetrain (rear/front/all wheel drive) |
| VehEngine | object | Vehicle engine (generally includes displacement size, whether it is turbocharged, sometimes includes fuel type) |
| VehFeats | object | Vehicle features as listed by the seller in a semi-structured list format |
| VehFuel | object | Vehicle fuel type |
| VehHistory | object | Vehicle ownership history in a semi-structured format that may also indicate if there is buy-back protection, previous commercial use, accidents, or potential title problems |
| VehListdays | float64 | Duration (in days) the vehicle listing has been active |
| VehMake | object | Vehicle make (manufacturer) |
| VehMileage | float64 | Vehicle mileage |
| VehModel | object | Vehicle model |
| VehPriceLabel | object | A classification label applied by the listing site that indicates if the listing price is a good deal or not |
| VehSellerNotes | object | Unstructured text the seller has entered that provides additional details on the vehicle |
| VehSellerStockNum | object | Vehicle seller stock number |
| VehTransmission | object | Vehicle transmission type |
| VehYear | int64 | Vehicle model year (not necessarily the year it was manufactured) |
| Vehicle\_Trim | object | Vehicle trim |
| Dealer\_Listing\_Price | float64 | Vehicle listing price, dependent variable to be predicted. |

Model Information

<https://en.wikipedia.org/wiki/Cadillac_XT5>

<https://en.wikipedia.org/wiki/Jeep_Grand_Cherokee_%28WK2%29>