Suppliers Onboarding process

Igor Levecharov

Agenda

Introduction

Key Facts of Input / Output Data

Changes made to the input data

Potential changes to the input data

Conclusion

Introduction

An e-commerce shop would like to onboard new suppliers efficiently. Integrate product data from suppliers in various formats and styles into the pre-defined data structure of their e-commerce shop application.

Primary goals

Transform the supplier data so that it could be directly loaded into the target dataset without any other changes.
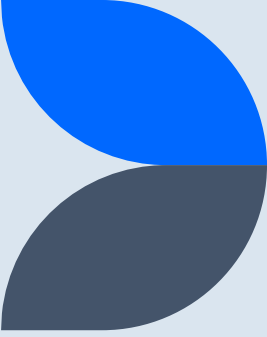
Key Facts of Input / Output Data

Input data

➢ JSON file with 21906 rows and 9 columns

➢ One ID column

➢ Five columns describing Make, Model and Type of the vehicle

➢ Two column for Attributes. Key => Values pairs.

➢ One unique identity column

Output data

✓ Excel file with 18 columns

✓ One group (3 columns): Make, Model, Model variant

✓ Attributes group (15 columns)

Data tranformation pipeline

1

EDA

Exploratory Data Analysis

2

Pre-processing

Split, Pivot and merge data

3

Normalization

Normalizing column values

4

Integration

Transform to a specific data schema

5

Final Output

Save the final output into excel file

Changes made to the input data

Pre-processing

1. Split to two dataset:

✓ One with columns make, model and type
(1153 rows x 5 columns)

✓ Second one: using pivot to reshape the given data based on attributes
(1153 rows x 19 column)

2. Merge two data sets into one:
(**1153 rows x 24 columns)**

Normalization

✓ Map the pre-processed data with the target data **(11 out of 18 columns)**

Normalize the attributes values for color, car_type, condition (using translation) and fuel_consumption_unit.

✓ For translation we use Deep Translator tool

Integration

Transform the Supplier Data into a new dataset that have the same data schema as Target Data **18 columns**

Potential changes to the input data

Auto populate values from GEO API's, city as input, output values:

➢ Currency

➢ Country

➢ Zip

Example:

https://github.com/DenisCarriere/geocoder

Not mapped values, need more info:

▪ drive

▪ mileage_unit

▪ price_on_request

▪ type

Example:
A car in Switzerland can have right hand drive and miles as mileage unit.

Summary

Valid JSON needs to be provided. Validation test before delivery.

Check for missing attributes. Vehicle with ID = 824 have 18 attributes instead of 19.

More info needed for fields like:

- drive

- mileage_unit

- price_on_request

- type

Thank you

Igor Levecharov

igor.levecarov@gmail.com

GitHub code:

https://github.com/igorlevecarov/Onedot