

Praxisprojekt: Programmierung in Python

Einleitung

Die Projektarbeit "Programmierung in Python" soll in einem Team von 3 - 4 Studierenden bearbeitet werden. Für ein erfolgreiches Bestehen des Kurses sind folgende Voraussetzungen zu erfüllen:

- Aufgabe 1 sowie Aufgabe 2 sind vollständig zu bearbeiten.
- Gemeinsame Vorstellung der Projektergebnisse im letzten Praxistermin (ca. 10 - 15 Minuten pro Team)
- Beantwortung der Fragen zum Projekt und zum Vorlesungsinhalt (ca. 5 Minuten pro Person)
- Abgabe des dokumentierten Projektergebnisses in Form einer (PowerPoint)-Präsentation oder (Latex-)Skriptes* pro Team

Gemeinsame Vorstellung der Projektergebnisse im letzten Praxistermin

Beschreibung

Die Daten, welche in dieser Challenge verwendet werden sollen, finden Sie in der Datei `vehicle_data.xlsx`. Bitte beachten Sie, dass die Daten zunächst aufbereitet werden müssen, also noch nicht in einer Qualität vorliegen, die eine finale Insight-Generierung ermöglicht.

Die Quelldatei `vehicle_data.xlsx` beinhaltet Daten zu 500 Fahrzeugen:

Im Reiter `sales code` finden Sie 500 Trucks deren eindeutige Fahrzeugidentifikationsnummer (FIN) als hash in der Spalte `h_vehicle_hash` dargestellt ist. Jedes Fahrzeug ist aufgeschlüsselt nach Produktionsdatum, Land in welches das Fahrzeug verkauft wurde und `sales_code_array`. Letztere Spalte beinhaltet Codes, welche die genaue Zusammensetzung des Fahrzeuges[†] beschreiben. Im Reiter `engines` finden Sie 9 sales Codes, die Aufschluss darüber geben, welcher Motor[‡] im entsprechenden Fahrzeug verbaut ist.

*Weitere Anforderungen an die schriftliche Abgabe sind der Orga.pdf Datei zu entnehmen

[†]verbauter Motor, Leistungsklasse, Kabinentyp und weitere Merkmale

[‡]OM924, OM926, etc ...

Mit der Tabelle `vehicle_hash` können Sie die gehashte FIN zurück in die originale Darstellung[§] mappen.

Bevor Sie starten, bietet es sich an sich zunächst mit den Datenätzen vertraut zu machen. Verschaffen Sie sich einen Überblick und erkunden Sie die Zusammenhänge der Datensätze.

[§]17 stellige Buchstaben-Zahlenkombination

Aufgabe 1: Data Engineering

Schreiben Sie eine ETL Pipeline zur Datenaufbereitung. Gehen Sie dabei wie folgt vor.

- Daten laden
- Daten bereinigen und aufbereiten
- Daten zusammenführen
- Gesamttabelle bestehend aus folgenden Spalten abspeichern:
 - `fin`
 - `production_date`
 - `country`
 - `sales_code_array`

Aufgabe 2: Data Science

Analysieren Sie die Daten, indem Sie folgende Fragestellungen auswerten. Visualisieren Sie ihre Ergebnisse.

- Welches sind die top drei Länder, in die wir zwischen 01.01.2014 und 31.12.2020 am meisten Fahrzeuge verkauft haben.
- In welchem dieser Jahre haben wir insgesamt am meisten Fahrzeuge verkauft?

Aufgabe 2: Data Science (optional)

- Welche FIN hat das zeitlich erste verkaufte Fahrzeug.
- Wie viele Fahrzeuge wurden zwischen 01.01.2017 und 01.01.2021 mit OM934, OM936, OM470 und OM471 Motoren verkauft.
- Welche Fahrzeuge (FIN) wurden zwischen 01.01.2017 und 01.01.2021 und mit OM936 Motor nach Neuseeland verkauft.