

Praxisprojekt – Python für Datenanalysten

Für dieses Praxisprojekt stehen fünf Tage zur Bearbeitung und ein Präsentation zur Verfügung. Wieder gibt es mehrere Projektvorschläge, von denen nur einer bearbeitet werden muss (aber weitere bearbeitet werden dürfen, falls die Zeit reicht).

Wählt als Team eines der Projekte aus und sprecht Euch ab, wer sich um welchen Teil des Projekts kümmert. Dokumentiert dabei, wer was zum Projekt beigetragen hat. Denkt an eure Softskills und setzt diese geschickt ein, um glücklich und effizient im Team zu arbeiten!

Übersicht über die Projekte:

Aufgabe 1 – Autodaten analysieren

Schwerpunkte: Datenanalyse | Aggregation | Visualisierung

Aufgabe 2 – Titanic Analyse

Schwerpunkte: Datenanalyse | Aggregation | Visualisierung

Aufgabe 3 - Restaurant Bestell-Tool

Schwerpunkte: Funktionen | DataFrames | Modularisierung

Aufgabe 1 – Explorative Analyse von Autodaten

Euch liegt [ein Datensatz](#) mit Daten eines großen Autohändlers vor. Führt am Datensatz mit Euren neuen Pandas-Kräften eine explorative Datenanalyse durch. Folgende Fragen können Euch interessieren:

- Wie viele Fahrzeuge umfasst der Datensatz? Was ist das teuerste, was ist das billigste Fahrzeug?
- Welche Marke macht den größten Anteil am Umsatz aus? Macht sie auch den größten Anteil an verkauften Stückzahlen aus? Was sind die zehn wichtigsten Marken?
- Was sind die top Fünf beliebtesten Modelle?
- Welchen Einfluss haben Angebotstyp, Art des Antriebs, Pferdestärke, Kilometerzahl und Baujahr auf den Preis von Fahrzeugen?
- Welche deutschen Marken gibt es im Datensatz? Welchen Anteil machen diese aus? Gibt es hier Besonderheiten?
- Selektiere einen Ausschnitt mit deutschen Fahrzeugen und untersuche diesen tiefer.
- Welche Marken aus Asien gibt es im Datensatz? Welchen Anteil machen diese aus?
- Es gibt Luxus-Fahrzeuge von Ferrari & Co., die ganz pro Stück ganz schön viel einbringen. Welchen Anteil macht das Luxus-Segment aus?
- Was sind die billigsten Fahrzeuge? Welchen Anteil machen diese aus?
- Gibt es Dinge im Datensatz, die keine Autos sind?
- ...

Nutzt bei Eurer Arbeit Selektionen, Aggregationen und Visualisierungen, um interessante Einsichten zu gewinnen. Entwickelt eigene Fragestellungen. Überlegt Euch, wem Ihr Eure Ergebnisse präsentiert. Soll es der Autohändler selbst sein, der sich einen Bericht wünscht? Ist es das Konkurrenz-Unternehmen, das sich etwas anschauen will oder Lücken des Gegners für die eigene Strategie nutzen will? Geht es um Empfehlungen für Preise künftiger Fahrzeuge, die verkauft werden sollen? Oder was ganz anderes? In allen Fällen solltet Ihr mit schicken Tabellen und Grafiken in Eurer Präsentation überzeugen. Powerpoint ist Pflicht, was nicht heißen soll, dass Ihr nicht besondere Stellen aus Eurem Code zeigen soll oder – falls Ihr etwas programmiert – Eurer Empfehlungstool präsentiert.

Aufgabe 2 – Explorative Analyse des Titanic-Datensatzes

Euch liegen die Daten zu den Passagieren der Titanic vor. Die wichtigste Spalte ist natürlich die mit dem Überlebensstatus. Aber gleichzeitig spielen alle eine Rolle, denn Ihr habt die große Aufgabe, ein für alle Mal verbindlich in einer explorativen Datenanalyse zu klären, wer welche Überlebenswahrscheinlichkeiten auf diesem Schiff hatte. Von Interesse können Fragen sein wie:

- Welche Rolle hat die Klasse gespielt?
- Welche das Geschlecht?
- Welche das Alter? (Was war die Altersverteilung überhaupt?)
- Wie steht es um verheiratete und unverheiratete Frauen?
- Welche Rollen hatten Geschwister, Ehepartner, Eltern, Kinder dabei?
- Wie war die Sterblichkeit von Kindern? Gibt es da Unterschiede je nach Alter?
- Lässt sich etwas aus den Kabinennummern ableiten?
- ...

Erstellt zur Untermauerung aller Eurer Einsichten klare Tabellen und eindrucksvolle Grafiken. Malt mit diesen Mitteln ein lebhaftes Bild von der Titanic und beantwortet am Ende die Frage: Vertreter welcher Gruppen hatten die geringsten Aussichten zu überleben und bei wem war eine Rettung am wahrscheinlichsten? Überlegt Euch, welches Publikum Ihr adressiert.

Aufgabe 3 - Restaurant Bestelltool

Das Restaurant "golden seagull" benötigt ein Programm zur Abwicklung von Bestellungen im Restaurant und beauftragt Euch als selbstständiges, kleines Developer-Team mit der Entwicklung der Software. Zusammen habt Ihr folgende Anforderungen ausformuliert:

a.

Eine Speisekarte sollte im `.csv` Format abgelegt werden und als DataFrame eingelesen werden. Die Speisen-Nummer soll der Zeilen-Index sein. Mit einer Funktion soll die Speisekarte geladen werden. Mit einer anderen Funktion soll man auch die Speisekarte anzeigen können.

b.

Die Liste der Bestellungen soll ebenfalls in einem DataFrame festgehalten werden. Jede Bestellung soll eine ID, ein Datum, eine Tischnummer, SpeiseID, Menge und Status enthalten. Man soll mit einer Funktion in der Lage sein, neue Bestellungen zu erzeugen, und dabei die Tischnummer sowie mehrere SpeiseIDs mit zugehörigen Mengen eingeben können. Das Datum, die ID, und der Status "offen" sollen beim Erstellen automatisch vergeben werden. Es soll durch die Erstell-Funktion ein DataFrame mit Zeilen für jede Speise der Bestellung erzeugt werden. Dieses DataFrame soll an die Gesamtliste der Bestellungen angefügt werden.

c.

Datenvalidierung ist hier sehr wichtig! Eine Funktion sollte Bestellungen prüfen, und True zurückgeben, wenn alle Gerichte in der Speisekarte sind, sonst False. Jede neue Bestellung sollte überprüft werden und Fehler entsprechend behandelt werden, bevor die neue Bestellung zur Gesamtliste hinzugefügt wird.

d.

Man soll per Funktion in der Lage sein, Bestellungen zu stornieren. Dabei soll der Bestellstatus in der Liste auf "storno" gesetzt werden.

e.

Mit einer Bezahlungsfunktion soll der Bestellstatus der Bestellung auf "Beahlt" gesetzt werden und eine Rechnung auf dem Bildschirm ausgegeben werden. Die Rechnung umfasst die Gerichte, Nettopreis, Menge, MwSt., Bruttopreis, sowie evtl. Trinkgeld. Die Preise der Gerichte sollten aus der Speisekarten-Datei erhältlich sein.

Das bildet das Grundgerüst der Bestell-App. Wie immer können beliebige Zusatzfunktionen und Details eingebaut werden (z.B. logging, Bestellhistorie, Bestellstatistiken, Umsatzzahlen...). Auch das Interface ist wieder frei wählbar – Ob Kommandozeile, GUI, Notebook oder Widgets, oder als eigenständiges Modul. Als Abschluss-Präsentation eignet sich eine Live-Demonstrationen zu eurer PowerPoint. Viel Erfolg!