

Praxisprojekt - Modul 2.2 - Python für Datenanalysten

04.08.2023

Für dieses Praxisprojekt stehen vier Tage zur Bearbeitung (04.-09.08.) und ein ganzer Präsentationstag (10.08.) zur Verfügung. Wieder gibt es mehrere Projektvorschläge, von denen nur einer bearbeitet werden muss.

Wieder sind die Gestaltung des Projektes und auch die Präsentation frei wählbar. Ihr könnt also eine oder mehrere .py Dateien erstellen oder mit .ipynb Notebooks arbeiten. Ihr könnt euer Programm oder eure Analyse oder euren Code live und interaktiv vorführen, und/oder auch auf PowerPoint zurückgreifen.

Denkt wieder an eure Soft Skills zurück, und setzt diese geschickt ein um glücklich und effizient im Team zu arbeiten!

Wieder wird das Praxisprojekt bewertet nach den gewohnten Kriterien Datenauswertung/Programmierung – Projekt-design/architektur – Visualisierung – Präsentation.

Zu den einzelnen Aufgabenstellungen kommst du hier:

Aufgabe 1 - Shopauswertung

Schwerpunkte: Datenanalyse | Input/Output | Aggregation | Visualisierung

Aufgabe 2 - Sternzeichen Profilbild Generator

Schwerpunkte: Bilddbearbeitung | Arrays | Dateien schreiben/lesen

Aufgabe 3 - Online-Teaching Klassenstruktur

Schwerpunkte: OOP | Vererbung | Modularisierung

Aufgabe 4 - Restaurant Bestell-Tool

Schwerpunkte: Funktionen | DataFrames | Modularisierung

Aufgabe 5 - IMDB Daten

Schwerpunkte: Datenanalyse | Aggregation | Visualisierung | Vorhersage

Aufgabe 1 - Shopdaten Analyse

Bei deinem neuen 1-Monat-Projekt kommst du bei “El Puente” unter – einem Fair-Trade Dienstleister mit Sitz im wunderschönen Nordstemmen. El Puente betreibt neben dem deutschen Standort noch einen weiteren Standort, aus welchem die Verkäufe ins restliche Europa abgewickelt werden. Von diesen beiden Standpunkten liegen dir die Daten (DE in Excel-Datei, restl. Europa in CSV-Datei) vor.

Fertige eine Datenanalyse zu diesem Datensatz an. Dazu gehören Import, möglicherweise Aufbereitung / Cleaning und explorative Datenanalyse. Mögliche Fragen die dabei beantwortet werden könnten sind:

- Wie sind die Daten abgelegt und welche Besonderheiten gibt es eventuell bei der Betrachtung?
- Welche Produkte verkaufen sich in den Shops wie gut oder schlecht? Hinweis: Du kannst davon ausgehen, dass "A" = Produkt 1 bedeutet (und so weiter).
- Gibt es Unterschiede im zeitlichen Verlauf (wöchentlich, monatlich, jährlich) oder zwischen den einzelnen Shops?
- Ist ein Produkt besonders wichtig oder unwichtig?
- Welche Empfehlungen würdest du El Puente geben, um seinen Umsatz in der Zukunft noch weiter zu steigern?

Das Ergebnis soll eine gut dokumentierte Analyse sein. In der Präsentation müssen Außenstehende in der Lage sein die Daten und die benötigten Analyseschritte nachzuvollziehen. Die Präsentation kann mit PowerPoint erfolgen, oder mithilfe eines gut formatierten, eventuell interaktives Notebooks, welches die wichtigsten Ergebnisse zusammenfasst. Viel Erfolg!

Aufgabe 2 - Sternzeichen Profilbild Generator

Nachts um 03:28 Uhr spricht dir ein Freund eine geniale Geschäftsidee auf die Mailbox: Er sagte irgendwas darüber, dass er eine App anbieten will, mit welcher man einem Profilbild das persönliche Sternzeichen hinzufügen können. Du sollst dich dringend jetzt am nächsten Morgen ans Werk machen, um den Code zu schreiben. Du findest auf deinem Handy aber nur noch folgende Nachricht:

„Heeeeeeeeeeeeeeeeeey Superstar! 1. Nutzer soll den Funktionsnamen „Sternenbildner“ eingeben und dabei als Argumente nur sein Profilbild (mit mindestens 800x800 Abmessung), seinen Geburtstag und seinen Geburtsmonat angeben. 2. Nutzer bekommt danach sein Profilbild wieder ausgegeben, in welchem in der oberen rechten Ecke mit 100x100 das Bild des Sternzeichens halbtransparent in schwarz eingeblendet wird 3. Und jetzt feier weiter!“

Anforderung an die Lösung:

- Du willst deinen Code deinem Freund aus der Kneipe erklären können, kommentiere diesen also in seinen Bestandteilen.
- Mach den Code idiotensicher; die gängigsten Fehler bei der Eingabe der Funktionsparameter durch den Nutzer soll abgefangen und wenn möglich direkt innerhalb des Codes verbessert werden (Beispiele: Wenn keine Dateiendung angegeben wird, sollen die Top 5 häufigsten Dateiendungen von Bildern probiert werden; Wenn Monat als Text angegeben wird, soll er in eine Zahl umgewandelt werden). Fehler, die du nicht behandeln kannst (z.B.: fehlende Datei), sollen dem Nutzer möglichst klar und deutlich kommuniziert werden.
- Es ist egal, ob das Sternzeichen irgendwas in dem Originalbild überdeckt, es soll immer in der Abmessung 100x100 in der oberen rechten Ecke erscheinen! Die Bilder für die Sternbilder sind alle in der Word Datei vorgegeben und müssen nicht durch dich gesucht werden.

Auch für dieses Projekt kann der Code wieder live vorgeführt werden, zum Beispiel mit ein bis zwei Beispielbildern. Eine PowerPoint Präsentation geht natürlich auch, ist aber kein Muss. Viel Erfolg!

Aufgabe 3 - Online-Teaching Klassenstruktur

Du gründest deine eigene Coding-Schule namens PyWizards of the Coast. Die Schule soll online stattfinden und braucht dazu eine eigene Software. Um die Daten aller beteiligten Personen zu managen machst du dich daran eine Klassenstruktur dafür zu entwickeln, die alle nötigen Funktionen bereitstellt. Du hast schon einen groben Plan dafür ausgearbeitet:

Teil 1: Die Personen: Person, Student und Dozent

Jeder Kontakt der Online-Schule, egal ob Student oder Dozent, ist eine Person: - Jede Person hat einen Namen - Jeder Person sollte es möglich sein, sich vorstellen (zB: „Hallo, mein Name ist Eric“)

Jede Person ist entweder ein Student oder ein Dozent: - Jeder Student hat einen Grund zur Teilnahme (zB: „Ich will coden lernen“) - Jeder Dozent hat eine Biografie (zB: „Ich codiere mit R und Python und liebe die Lehre“) - Jeder

Dozent hat Fertigkeiten (zB: „Python“, „R“, „SAS“ ...) - Jeder Dozent kann neue Fertigkeiten mit `add_skill` erlernen (`add_skill(„C++“)` fügt z.B. den Fertigkeiten des Dozenten diese neue Fertigkeit hinzu).

Teil 2: Workshop

PyWizards of the Coast bietet natürlich Workshops an. Jeder Workshop hat: - Ein Start- und Enddatum - Ein Thema - Einen oder mehrere Dozenten - Mehrere Studenten - Eine Funktion `add_participants`, die nur Personen als Argument erlaubt. Ist die Person Dozent, wird sie der Liste der Dozenten hinzugefügt. Ist die Person Student, wird sie der Liste der Studenten hinzugefügt.

Teil 3: Äußere Funktionalitäten

Diese Bestandteile könnten über Methoden, Funktionen, oder andere Interfaces implementiert werden:

- Über `init` Methoden soll man neue Personen und Workshops erstellen können.
- Ermögliche es dem Nutzer die Informationen aller Studenten in einem DataFrame und aller Dozenten in einem zweiten DataFrame darzustellen.
- Ermögliche es dem Nutzer alle Informationen eines Workshops darzustellen.
- Ermögliche es dem Nutzer beide obigen Funktionen zusammen durchführt und alle Informationen darstellt.

Diese Bestandteile sollen von dir implementiert (programmiert) werden und bilden das Grundgerüst für das Personendaten-Management von PyWizards of the Coast. Natürlich sind kleine Abweichungen vom Plan erlaubt. Das Konzept ist beliebig erweiterbar, solltet ihr vor Ende der Projektzeit damit fertig sein. Toll wäre z.B. in einer oder mehreren Dateien geeigneten Formats die Workshop und Personendatenbank zu speichern und zu laden. Support für andere Personengruppen, eine Funktion zum anzeigen aller Workshops an einem bestimmten Termin oder auch ein Interface zur Ein- und Ausgabe der Informationen wären nützliche Zusatzfunktionen.

Als Präsentation kann wie immer eine live Demonstration des Codes erfolgen, oder auch eine PowerPoint als Unterstützung genutzt werden. Viel Erfolg!

Aufgabe 4 - 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 Speisenummer 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 eignen sich sowohl live Demonstrationen als auch PowerPoints. Viel Erfolg!

Aufgabe 5 - IMDB Top 250 Daten Analyse

Diese Aufgabe kennt ihr schon aus dem Excel Projekt. Nun habt ihr mit euren erweiterten Datenanalyse Fähigkeiten und Python Programmierskills eine weitere Chance diesen Datensatz noch tiefer und effektiver zu analysieren.

In der bereitgestellten Excel Datei liegen Daten über die 250 am besten bewerteten Filme auf imdb.com vorliegen (Erscheinungsjahr, Laufzeit, Genre, Director/Regisseur, und die vier Haupt-Schauspieler). Zusätzlich habt ihr die Information, ob Nils den jeweiligen Film gesehen hat (Spalte „Gesehen“ 1 = gesehen, 0 = nicht gesehen) und ob er ihn richtig gut fand (Spalte „Nils' Top Filme“: 1 = richtig gut, 0 = nicht richtig gut).

1. Erstellt als Gruppe aus diesen Daten eine komplette Fallanalyse zu a) den Top 250 Filmen, b) den Filmen die Nils gesehen hat und c) den Filmen, die Nils gut fand. Unter anderem folgende Fragen könnten hierbei beantwortet werden:
 - Welches Genre ist wie häufig vertreten?
 - Gibt es ein Lieblingsgenre?
 - Gibt es Besonderheiten bei den vertretenen Jahresangaben?
 - Gibt es Unterschiede zwischen den Top 250 selbst und den Nils-Spalten?
 - Sind kürzere Filme oder längere Filme beliebt?
 - Sind bestimmte Schauspieler besonders häufig vertreten?
2. Nach Durchsicht und unter Einbezug ALLER Informationen über Nils' Top Filme: Welchen Film von der Top 250 Liste, den Nils noch nicht gesehen hat, sollte er sich unbedingt anschauen? Mit dieser Folie / Information sollte eure Präsentation abschließen und diese Frage **muss** beantwortet werden.

Eurer Kreativität sind keine Grenzen gesetzt. Erarbeitet gemeinsam, welche Fragen ihr beantworten wollt und durch welchen Lösungsweg ihr dorthin gelangt. Wichtig ist: IHR müsst erklären können, was ihr in eurer Datenanalyse wie gemacht habt, wieso ihr etwas eventuell nicht gemacht habt, und welche Informationen und Schlussfolgerungen ihr aus eurer Analyse ziehen könnt. Viel Erfolg!
