

Hausarbeit „Datenanalyse mit Python“

Thema: Mobile Device Data

Formalien:

Nutzen Sie für die untenstehenden Aufgabenstellungen die Ihnen zur Verfügung gestellte **Datendatei mobile_device_data.csv**. Entwerfen Sie ein **Python-Script**, welches über die geforderten Funktionalitäten verfügt. Für die Abgabe ist eine **Dokumentation als .pdf** zu erstellen und die **.py Datei** mit abzugeben. In der Dokumentation sind **alle Teile** Ihres Python-Scripts **begründet zu erläutern** und die Funktionalität mit **Screenshots von Ausgaben** zu belegen. Das Python-Script ist mit **Kommentaren** zu den einzelnen funktionalen Blöcken zu versehen.

Laden Sie die Dokumentation und die .py Datei in Ilias als **Übung**  in der **Gruppe**  „Hausarbeitabgabe“ bis zum Prüfungstermin hoch. Es genügt, wenn eine/einer in der Gruppe einen Upload vornimmt.

Bewertungskriterien:

- Vollständigkeit (der Funktionalität, Aufgabenstellung, Dokumentation)
- Effizienz (bei der Umsetzung)
- Lauffähigkeit des Skripts und Umsetzung der Funktionalitäten
- Begründung und Kommentare

Aufgabenstellung:

- a) Lesen Sie die Ihnen zur Verfügung gestellte .csv-Datei ohne die Spalte „id“ ein und lassen Sie sich als Nachweis dieser Funktionalität die ersten 12 Zeilen ausgeben.
Bemerkung: drei Spalten beinhalten nur 1 und 0, wobei 1 = yes und 0= no gilt.
- b) Bereinigen Sie den Datensatz, indem Sie auf Fehleinträge (NaN und Leerzellen) prüfen. Zeilen mit diesen Fehleinträgen werden für die Auswertung gelöscht. Der Datensatz sollte noch aus 200 Zeilen bestehen, prüfen Sie diese Bedingung!
- c) Erzeugen Sie eine Tabelle, die folgende Kennzahlen einer univariate Analyse der metrischen Merkmale vornimmt und speichern Sie die Tabelle in einer .csv Datei: Maximum, 80% Perzentil, Median, arithmetisches Mittel, Standardabweichung, 20%-Perzentil und Minimum.
- d) Erzeugen Sie eine Abbildung, in der gruppierte Balkendiagramme zu den Spalten „bluetooth“, „dual_sim“ und „4G“ angezeigt werden.
- e) Nehmen Sie eine Korrelationsanalyse nach Pearson vor und führen Sie eine lineare Regression der beiden höchst korrelierten Werte vor. Präsentieren Sie die Ergebnisse in einem Scatter-Plot einschließlich Regressionsgerade.
- f) Skalieren Sie die Werte der drei metrischen Merkmale in das Intervall [0, 1].
- g) Geben Sie die Streuung der drei metrischen Merkmale anhand von Boxplots mit Ausreißern aus.

Viel Erfolg !!!