

Exploratory Data Analysis (EDA) – Multivariate Abhängigkeit

Daten	Kreditkartenabrechnung - credit.csv
Inhalte	In diesem Datensatz sind Informationen über 400 Kreditkarteninhaber und ihre demografischen Merkmale wie Alter, Geschlecht, Familienstand, ethnische Herkunft, etc. enthalten.
Informationen	Der Datensatz dient der Kundenstammanalyse.
Fragestellung	Mit welcher Genauigkeit lassen sich aus den persönlichen Kundendaten deren Kreditkarten-Abrechnungen vorhersagen und was sind die stärksten Einflussfaktoren?
Vokabular	<ul style="list-style-type: none"> • Income: Einkommen des Kunden (in USD) • Limit: Verfügbares Kreditkarten-Limit des Kunden • Rating: Bewertung Kreditwürdigkeit • Cards: Anzahl der Kreditkarten des Kunden • Age: Alter (Jahre) • Education: Bildungslevel des Kunden (in Anzahl der Ausbildungsjahre) • Gender: Geschlecht des Kunden (male / female) • Student: Student (True / False) • Married: Verheiratet (True / False) • Ethnicity: Ethnische Herkunft des Kunden • Balance: Kreditkarten-Rechnung des Kunden (monatlicher Mittelwert)
Quelle	<ul style="list-style-type: none"> • https://www.statlearning.com/

- 1) Erstellen Sie ein Python-Programm, mit dem Sie den Datensatz für ihre weitere Analyse mit Python einlesen und analysieren können
 - 2) Verschaffen Sie sich einen Überblick über den Datensatz. Wie viele Datensamples sind enthalten, wie viele Attribute gibt es? Gibt es fehlende Daten? Gibt es redundante Daten? Zeigen Sie die ersten Zeilen des Datensatzes an!
 - 3) Bereinigen Sie den Datensatz wenn notwendig!
 - 4) Berechnen Sie statistische Kennzahlen für die numerischen Variablen – Mittelwert, Standardabweichung und Median für den Gesamt-Datensatz. Erstellen Sie ein Histogramm für die Verteilung der Kreditkarten-Balance. Was fällt auf? Werden weitere Bereinigungen notwendig?
 - 5) Untersuchen Sie Abhängigkeiten zwischen den Merkmalen:
Bivariat: z.B. balance vs. Limit; balance in Abhängigkeit vom gender, balance in Abhängigkeit vom Attribut „Student“
Multivariat: z.B. Korrelationskoeffizienten aller Variablen
 - 6) Welche Merkmale haben anhand dieser Untersuchungen wahrscheinlich signifikanten Einfluss auf die Kreditkartenabrechnung? Fassen Sie ihre Ergebnisse kurz zusammen!
- **Abgabe bis Donnerstag 30.01. 16:00 Uhr zwei Folien pro Gruppe (pdf!!), im Namen des Files bitte Gruppennummer + Nummer der Übung verwenden (z.B. GruppeXX_uebung-a2.pdf)**
 - **in Verzeichnis Upload_DASC hochladen: <https://nc.ufz.de/s/39zCZKFAYZKDQ2z>**
 (Passwort: ?solution_DASC25)

- Folie 1: Grafik zu Aufgaben 4
Folie 2: Grafik auswählen aus Aufgabe 5 + kurze Zusammenfassung der Ergebnisse aus der Analyse
- **5-Minuten Pitch: Jede Gruppe präsentiert kurz ihre Ergebnisse (5 min maximal) zu Beginn der nächsten Veranstaltung am 31.01.25!**