

Testat Aufgabe 1

Sie sollen für die Personalabteilung der NANUAG die Altersverteilung der Beschäftigten analysieren. In der Datei **sr_aufg_1_xy.txt** im Unterverzeichnis **Endziffer_xy** („xy“ steht für die beiden Endziffern Ihrer Matrikelnummer) finden Sie die dafür erforderlichen Daten. Diese Datei enthält folgende fünf Felder, die durch Semikolons getrennt sind: Personalnummer, Anrede („Herr“ oder „Frau“), Vorname, Nachname, Geburtsdatum.

1. Lesen Sie Daten in **Python** bzw. ein **Jupyter Notebook** ein. Speichern Sie sie in einem **Pandas DataFrame** ab.
2. Führen Sie eine Datenbereinigung durch. Listen Sie in einer Tabelle auf, welche Sätze Sie gelöscht und welche Sätze Sie wie korrigiert haben. Sie können hierzu die Python Funktionen **drop()** und **at()** verwenden.
3. Berechnen Sie das Alter der Beschäftigten zum Stichtag 31.12.2005.
Mit „Alter“ ist hier die Anzahl der vollendeten Lebensjahre gemeint. Beispiel: Wer am 30.06.1975 geborgen ist, ist zum Stichtag 30 Jahre alt. (Niemand würde von sich sagen, er wäre „30,5 Jahre alt“.) Folglich wird diese Person in der Alterklasse 21 – 30 Jahre mitgezählt.
In Python ist hierzu die Bibliothek **datetime** und ihre Funktion **to_datetime()**, sowie **dt.year** nützlich.
4. Stellen Sie die Altersverteilung der Beschäftigten in einem gruppierten Säulendiagramm dar, in dem Männer und Frauen getrennt ausgewiesen werden. Es sollen die Altersklassen bis (einschließlich) 20 Jahre, 21 – 30 Jahre, 31 – 40 Jahre, 41 – 50 Jahre und über 50 Jahre gewählt werden.
Auf der nächsten Seite ist ein verkleinertes Muster angegeben. Sie benötigen hierfür die Python Bibliothek **Matplotlib** und **NumPy**. sowie aus **Pandas** die Funktion **value_counts()**
5. Stellen Sie die Altersverteilung der männlichen Beschäftigten und der weiblichen Beschäftigten in je einem Kreisdiagramm („Tortendiagramm“) dar. Die Kreissektoren sind mit den zugehörigen Prozentzahlen zu beschriften. Auf der nächsten Seite ist ein verkleinertes Muster (mit anderen Zahlenwerten) angegeben.
6. Berechnen Sie Mittelwert, empirische Varianz, empirische Standardabweichung, Median und Spannweite der Alterszahlen
->für alle Beschäftigten
->getrennt für Männer und Frauen

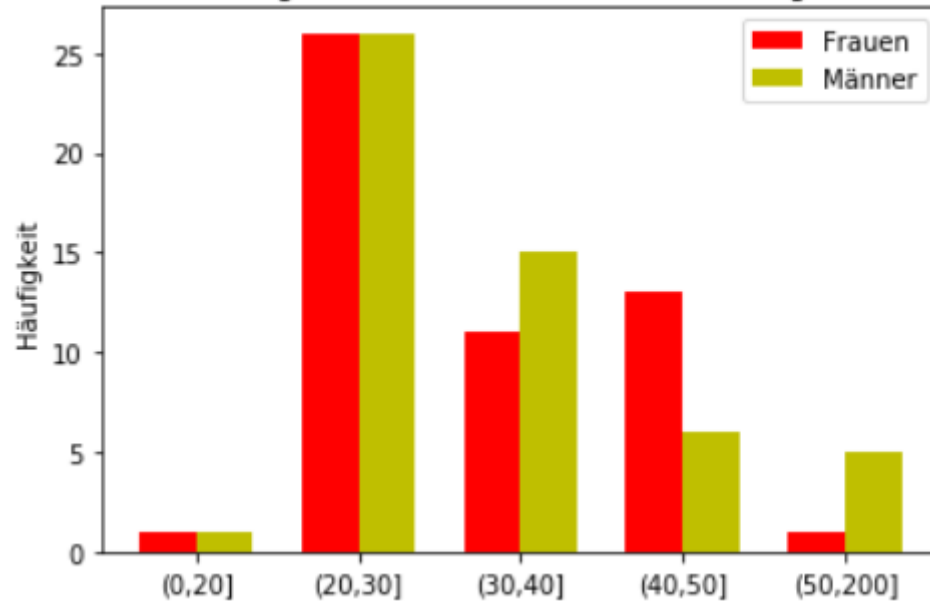
Beispielhaft einige Teile, wie Ihre Abgabe aussehen könnte:

Personal-nr.	Anrede	Vorname	Nachname	Geburts-Datum	Daten-satz?	Typ des Fehlers
025	Herr	Klaus	Schneider	im Januar 1941	Korrektur	Monat reicht für Alter
033	Frau	Kathrin	Kleber	UVGoJxpkci	Löschen	Unsinn

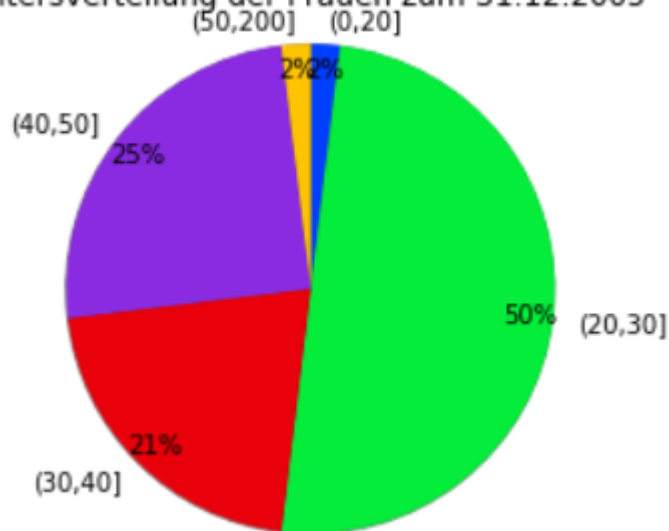
Kennzahlen

	gesamt	Frauen	Männer
Mittelwert	43,5196	43,6863	43,3529
Varianz	146,5887	153,2596	142,7929
Std.-Abw.	12,1074	12,379806	11,9496
Median	44,5	46	43
Spannweite	46	46	46

Altersverteilung Männer und Frauen zum Stichtag 31.12.2005



Altersverteilung der Frauen zum 31.12.2005



Achten Sie bei den Diagrammen auch auf die richtige Reihenfolge der einzelnen Klassen.

Bitte überprüfen Sie, ob Sie alle geforderten Teile liefern.

Bitte überprüfen Sie Ihre Resultate vor der Abgabe Ihre Ergebnisse mit den Teilergebnissen aus der Datei ***Teilergebnisse_xy.txt***. In der Datei bekommen Sie jeweils die ersten beiden Nachkommastellen des Durchschnittsalters.

Bitte drucken Sie auf keinen Fall die Liste mit den 100 Datensätzen aus!

Die Unterlagen sind als **pdf-Format** in Moodle hochgeladen abzugeben. Richtige Lösungen werden nicht mehr an Sie zurückgegeben. Eine Abgabe per Email ist nicht möglich.

Bitte fügen Sie die folgende Seite als Deckblatt vor Ihre Lösung an und geben es mit Ihrer Lösung zusammen ab.

Statistik Rechnerübung SWB 2 Sommersemester 2022, Aufgabe 1

Nachname:

Vorname:

Matrikelnummer:

Semester:

Email-Adresse:

Abgabe-Schlusstermin: Donnerstag, 21.04.2022