

Name: _____

Aufwand (h): _____

Punkte: _____

Die Aufgaben sollen vom Code optimiert einfach und schnell, bevorzugt mit den Elementen der 5. Lehreinheit, gelöst werden.

Aufgabe 1 (3 Pkt): Forbes 2000

Forbes ist eine englischsprachige Zeitschrift und eines der erfolgreichsten Wirtschaftsmagazine weltweit. Jährlich werden Listen mit den reichsten Personen und Unternehmen veröffentlicht. Öffnen Sie die Liste mit den 2000 teuersten Unternehmen „*Forbes2000.csv*“ und speichern Sie die Einträge als **data.table**. Untersuchen Sie den Inhalt der Datei nach deskriptiven Werten:

- Überblick über Anzahl der Einträge
- Min, Max, Mittelwerte, Standardabweichung, etc.
- Die ersten 5 und die letzten 5 Einträge

Setzen Sie geeignete „keys“ um die folgenden Aufgaben optimiert (einfach und schnell mit Hilfe der data.table) beantworten zu können:

- alle österreichischen Finanzinstitute (insurance und banking)
- Die Mittelwerte des Marktwertes je Land
- Die Mittelwerte des Marktwertes je Land und Kategorie
- Die Mittelwerte des Marktwerts und des Verkaufserlöses
- Die Anzahl der US-amerikanischen Unternehmen, deren mittlerer Marktwert sowie Erlös – jeweils pro Kategorie.

Es sollen folgende Dateien abgegeben werden:

- Markdown-Datei: Dateiname_NameAutor.Rmd
- R-Skript mit extrahierten Befehlen: Dateiname_NameAutor.R
- Wenn nötig weitere Skripte / Dateien

Aufgabe 2 (3 Pkt): Tips

Laden Sie die Datei „*tips.csv*“, in der die Trinkgelder angegeben sind, welche von Gästen eines Restaurants mit begleitenden Parametern (wann, Anteil am Essen, Geschlecht der Gäste, etc.) angegeben sind. Ermitteln Sie mit `plyr`-Funktionen das mittlere Trinkgeld, die Standardabweichung und den Prozentsatz an Trinkgeld, jeweils für die Kombination aus Geschlecht, Tag und Zeit. Stellen Sie die Ergebnisse in einer Tabelle (exakt, also auch z.B. die Spaltenausrichtung und Nachkommastellen, die (deutschsprachige) Beschriftung) derart dar. Sortieren Sie absteigend nach dem Anteil.

	Geschlecht	Tag	Zeit	n	Endbetrag	Tip MW(SD)	Anteil
2	Weiblich	Fr	Mittag	4	13.94	2.75 (0.64)	20 %
1	Weiblich	Fr	Abend	5	14.31	2.81 (1.21)	19.9 %
4	Weiblich	So	Abend	18	19.87	3.37 (1.14)	18.2 %
8	Männlich	Fr	Mittag	3	11.39	1.9 (0.31)	17.4 %
11	Männlich	Do	Mittag	30	18.71	2.98 (1.35)	16.5 %
10	Männlich	So	Abend	58	21.89	3.22 (1.27)	16.2 %
5	Weiblich	Do	Abend	1	18.78	3 (NaN)	16 %
6	Weiblich	Do	Mittag	31	16.65	2.56 (1.12)	15.7 %
3	Weiblich	Sa	Abend	28	19.68	2.8 (1.23)	15.6 %
9	Männlich	Sa	Abend	59	20.80	3.08 (1.79)	15.2 %
7	Männlich	Fr	Abend	7	23.49	3.03 (1.21)	13 %

Es sollen folgende Dateien abgegeben werden:

- Markdown-Datei: `Dateiname_NameAutor.Rmd`
- R-Skript mit extrahierten Befehlen: `Dateiname_NameAutor.R`
- Wenn nötig weitere Skripte / Dateien

Aufgabe 3 (8 Pkt): Datensatz „Marketing“

Öffnen Sie den Datensatz „*Marketing.csv*“. Für Einkaufsmärkte (LocationID) wurden verschiedene Promotion-Strategien (Promotion 1-3) zur Einführung eines neuen Produktes angewandt. Über vier Wochen (Week 1-4) wurde der Verkauf des Produktes erhoben.

Programmieren Sie mit den Funktionen des dplyr- und tidyr-Packages so effizient wie möglich die folgenden Aufgaben. Fassen Sie dazu beispielsweise die Einträge für jeden Markt in einer Zeile zusammen.

Erstellen Sie eine passende Auswertung für einen potentiellen Auftraggeber, in dem Sie die Auswertungen ansprechend darstellen und die Werte erklären.

- a. Marktgröße: Erstellen Sie im Datensatz eine neue (kategoriale) Variable „Size“. Codieren Sie dafür die MarketSize um
 - >500: Supermarkt
 - >300-500: Markt
 - <=300: Geschaeft
- b. Geben Sie tabellarisch und graphisch die oben ermittelte Verteilung sowie die Verteilung nach den Promotion-Strategien aus.
- c. Verkaufserfolge: Berechnen Sie die Änderung der Verkaufszahlen innerhalb der ersten Woche (also Woche 2-Woche1) und im gesamten Zeitraum (bis Woche 4) und speichern Sie die Werte in geeigneter Weise im Datensatz ab.
- d. Verkäufe: Ermitteln Sie mit Funktionen der dplyr-Familie deskriptive statistische Werte (in der Form wie angegeben, also z.B. mit „±“).
 - i. Alle Verkäufe zu Beginn der Studie (also in Woche 1): „Mittelwert“
 - ii. Alle Verkäufe zu Beginn der Studie: „Mittelwert ± SD“. Runden Sie die Werte auf 0 Nachkommastellen.
 - iii. Alle Verkäufe am Ende der Studie „Mittelwert ± SD“. Runden Sie die Werte auf 0 Nachkommastellen.
 - iv. Alle Verkäufe zu Beginn der Studie: „Median, Iqr, Spannweite“
 - v. Mittelwert aller numerischen Werte.
 - vi. Mittelwert aller Verkäufe zu Beginn der Studie nach Geschäftsgröße (a)
- e. Analyse: Erstellen Sie eine Tabelle mit den Mittelwerten für folgende Spalten/Variablen:
 - i. MarketSize, Verkäufe zu Beginn (Sales), Erfolg nach 1 Woche (also in Woche 2) und Erfolg nach 3 Wochen (also in Woche 4) in Abhängigkeit von der kategorialen Groesse (Variable \$Size, siehe a)
 - ii. MarketSize, Verkäufe zu Beginn (Sales), Erfolg nach 1 Woche (also in Woche 2) und Erfolg nach 3 Wochen (also in Woche 4) in Abhängigkeit von der Werbung (Variable \$Promotion)
 - iii. Der Erfolg der Promotion in Abhängigkeit von den kategorialen Daten Size und Promotion.
- f. Zeigen Sie ob die Unterschiede in 3ei und 3eii statistisch signifikant sind.