

1. Praktikumsaufgabe IVDA

a) Lesen Sie den Datensatz „baeckerei.csv“ in R ein und untersuchen Sie den Zweck der Daten sowie die Datenqualität. Identifizieren Sie mögliche Probleme in den Daten und beheben Sie die Fehler, falls möglich. Dokumentieren Sie Ihr Vorgehen.

b) Visualisieren Sie die Daten aus a) mittels einer Shiny Applikation in R, indem Sie die Verteilungen darstellen:

- Verteilung der Items
- Transaktionen aggregiert über den Tagesverlauf
- Transaktionen über Zeit (Gesamt, Täglich, Wöchentlich)
- Anzahl Items pro Transaktion

c) Bauen Sie bei allen Darstellungen Interaktionsmöglichkeiten zum Filtern der Daten ein. Sollten die Daten Ausreißer haben, passen Sie Ihre Visualisierung dahingehend an.

d) Benutzen Sie die Library arules (<http://r-statistics.co/Association-Mining-With-R.html>), um mögliche Assoziationsregeln aus den Daten zu extrahieren. Stellen Sie den itemFrequencyPlot dar und erstellen Sie mit der apriori Funktion (support=0.0045, confidence=.5) Regeln. Nutzen Sie die inspect Funktion, um die Regeln zu untersuchen. Stellen die Verteilung nach support, confidence und lift dar.

Die Abgabe der Praktikumsaufgabe ist bis zum 19. November 2019. Vereinbaren Sie hierfür rechtzeitig einen Termin mit Ihrem Betreuer.

Die Gruppen A, B, C, D und E werden von Yves Annanias (annanias@informatik.uni-leipzig.de) betreut.

Die Gruppen F, G, H, I und K werden von Dr. Daniel Wiegrefe (daniel@informatik.uni-leipzig.de) betreut.

Die Gruppen L, M, O, und P werden von Dr. Dirk Zeckzer (zeckzer@informatik.uni-leipzig.de) betreut.