**Data Mining Exercise 2**

Magdalena König

01455794

1. Product Orders

* Ein Bild, das Text, Diagramm, parallel, Reihe enthält.

  Automatisch generierte BeschreibungTask 1: Data cleaning & exploration

Ein Bild, das Text, Screenshot, Diagramm, Schrift enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Screenshot, Schrift, Diagramm enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

* Task 2 – Association Rules
  + Min\_support and min\_confidence were generally chosen so that there are 5-10 rules as a result (in a common range of support and confidence 🡪 not too many outliers)
  + Frequent purchase of products in a subcategorie per order (min\_support=0.023, min\_threshold=0.15)

Ein Bild, das Text, Screenshot, Schrift, Design enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Karte Menü, Schrift enthält.

Automatisch generierte Beschreibung

* + - Lifts are gernally all under 1 – so purchasing a certain subcategory doesn’t increase to purchase the other (except storage and art). The confidence and support is rather small. So is the Kulszynski coefficent – so there is only a very low connection appliceable. The imbalance ratio shows a quite balanced occurance of the subcategories.

* + Ein Bild, das Text, Screenshot, Schrift, Karte Menü enthält.

    Automatisch generierte BeschreibungFrequent purchase of products in a subcategorie per customer (0.7;0.9)

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

* + - Lifts are gernally all over 1 – so purchasing a certain subcategory from one customer does increase the purchase the other. The confidence is high. So is the Kulszynski coefficent – so there is a high connection appliceable. The imbalance ratio shows a very balanced occurance of the subcategories.
  + Products bought together on the same day (0.74;0.9)

Ein Bild, das Text, Screenshot, Schrift, Design enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift, Karte Menü enthält.

Automatisch generierte Beschreibung

* + - -- Lifts are gernally all over 1 – so purchasing a certain subcategory on one day does increase the purchase the other. The confidence is high. So is the Kulszynski coefficent – so there is a high connection appliceable. The imbalance ratio shows a very balanced occurance of the subcategories.
* Task 3: Multiassociation-Rules
  + Ein Bild, das Text, Screenshot, Karte Menü, Schrift enthält.

    Automatisch generierte BeschreibungAssociation Rules from Subcategories to Categories (0.06, 0.5)

Ein Bild, das Text, Schrift, Screenshot, Typografie enthält.

Automatisch generierte Beschreibung

* + - The lift varies greatly, sometimes a subcategory increases the likehood of purchasing a product in a specific category, sometimes it decreases. The confidence is medium.
    - Discovered rules are interesting to see the purchase of which subcategories lead to the purchase of which categegories and how theres two purchases are connected.
  + Association Rules from Products to Subcategories (0.0015, 0.1)

Ein Bild, das Text, Screenshot, Schrift, Informationen enthält.

Automatisch generierte Beschreibung

* + - Positive Lift (except the last one) – so the product Staples increases the likehood of a purchase in the subcategory art, binders,…

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Schrift, Screenshot enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Schrift, Screenshot, Typografie enthält.

Automatisch generierte Beschreibung

* Discovered rules are interesting to see the purchase of which products lead to the purchase of which subcategories and how theres two purchases are connected. However as seen here, neither the support not the confidence are high. And after the redandency check only one rule remains (all the other consequents were already covered by the SubCategory:Paper and therefore redandent)
* Exercise 2 – Health data
  + Task 1: Data cleaning and exploration

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Diagramm, Schrift enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

* Task 2 - Association Rules
  + 🡪 Min\_support and confidence was generally choose to get between 3 and 10 rules in a common range of support and confidence (not too many outliers)
  + Symptoms that occur together (0.1; 0.5)

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Schrift, Screenshot, Design enthält.

Automatisch generierte Beschreibung

* + Ein Bild, das Text, Schrift, Screenshot, Design enthält.

    Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift enthält.

    Automatisch generierte BeschreibungCondition that occur together (0.05; 0.4)
  + Ein Bild, das Text, Screenshot, Schrift enthält.

    Automatisch generierte BeschreibungTreatments that occur together (0.03; 0.05)

Ein Bild, das Text, Schrift, Screenshot, Design enthält.

Automatisch generierte Beschreibung

* + - Positive lift – so the likehood of one condition rises when the other is here. However Kulzynski signalises a low – mid connection, so ist he imbalance ratio (number of overall occurance)
* Multi-Association Rules
  + Ein Bild, das Text, Screenshot, Schrift enthält.

    Automatisch generierte BeschreibungCondition 🡪 Treatment (0.014; 0.5)

🡪 There are high lifts and mid - high confidence! So the likehood to receive a certain treatment rises when a patient has a certain condition

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

* + Ein Bild, das Text, Screenshot, Schrift enthält.

    Automatisch generierte BeschreibungTreatment 🡪 Condition (0.014; 0.9)

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Screenshot, Schrift, Typografie enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Schrift, Typografie enthält.

Automatisch generierte Beschreibung

🡪 There are high lifts and high confidence! So the likehood of a certain condition rises when a patient receives a certain treatment.

* + Ein Bild, das Text, Screenshot, Schrift enthält.

    Automatisch generierte BeschreibungWeather 🡪 Condition (0.06; 0.2)
    - Ein Bild, das Text, Schrift, Screenshot enthält.

      Automatisch generierte BeschreibungEvery rule (63) has theh same support, confidence and lift. Positive Lift 🡪 so it seems like many weather condition increse the likehood of fibromyalgia
  + Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

    Automatisch generierte BeschreibungFood 🡪 Condition (0.005; 0.2)
    - The likehoof of the appearance of fibromyalgia decreases, than when a patient drinks water.
  + Condition 🡪 Food (-)

Couldn’t be computed!