

# Predictive Analytics:

## Ablauf der Vorlesung



# Vorlesungstermine

## 28.10.2019

- Besprechung der Übungsaufgabe
- Types of data
- Data preprocessing
  - Missing data
  - Outliers
  - Handling categorical variables
  - Splitting data
  - Scaling

## 04.11.2019

- Regression
  - Simple linear regression
  - Multiple linear regression
  - Polynomial regression

## 25.11.2019

- Classification
  - Logistic regression
  - Decision trees
- Clustering
  - K-means clustering

## 02.12.2019

- Vorstellung der Gruppenarbeiten

# Predictive Analytics:

## Übungsaufgabe



# Gruppenaufteilung für die Übungsaufgabe

## **Gruppe WZ1 - Gruppe WZ9 (Wetzlar)**

- 5 Studenten/innen

## **Gruppe WZ10 - Gruppe WZ12 (Wetzlar)**

- 4 Studenten/innen

## **Gruppe VIL (Bad Vilbel)**

- 4 Studenten/innen

## **Gruppe WIL (Bad Wildungen)**

- 4 Studenten/innen

## **Gruppe FKB (Frankenberg)**

- 5 Studenten/innen

## **Gruppe HEF (Bad Hersfeld)**

- 4 Studenten/innen

# Gruppenaufteilung für die Übungsaufgabe

## **Gruppe WZ1 - Gruppe WZ9 (Wetzlar)**

- 5 Studenten/innen

## **Gruppe WZ10 - Gruppe WZ12 (Wetzlar)**

- 4 Studenten/innen

## **Gruppe VIL (Bad Vilbel)**

- 4 Studenten/innen

## **Gruppe WIL (Bad Wildungen)**

- 4 Studenten/innen

## **Gruppe FKB (Frankenberg)**

- 5 Studenten/innen

## **Gruppe HEF (Bad Hersfeld)**

- 4 Studenten/innen



**Moodle**

# Übungsaufgabe

Jede Gruppe erhält einen Datensatz, der in Gruppenarbeit zu bearbeiten ist. Beispielhafte Fragen die Sie sich bei der Bearbeitung der Übung stellen sollten sind:

1. Um was für Daten handelt es sich?
2. Wie sehen meine Daten überhaupt aus?
3. Muss ich meine Daten vor der Analyse vorverarbeiten?
4. Welche Analysen sind für meinen Datensatz geeignet und was will ich damit herausfinden?
5. Wie stelle ich die Ergebnisse meiner Analysen da, damit auch meine Kommilitonen etwas daraus lernen können? (in **max. 15 min**)

## Idee:

Bearbeitung eines realen Datensatzes (keine konstruierten Daten, die immer ein schönes Ergebnis erzeugen).

# Übungsaufgabe

Jede Gruppe erhält einen Datensatz, der in Gruppenarbeit zu bearbeiten ist. Beispielhafte Fragen die Sie sich bei der Bearbeitung der Übung stellen sollten sind:

1. Um was für Daten handelt es sich?
2. Wie sehen meine Daten überhaupt aus?
3. Muss ich meine Daten vor der Analyse vorverarbeiten?
4. Welche Analysen sind für meinen Datensatz geeignet und was will ich damit herausfinden?
5. Wie stelle ich die Ergebnisse meiner Analysen da, damit auch meine Kommilitonen etwas daraus lernen können? (in **max. 15 min**)

## Idee:

Bearbeitung eines realen Datensatzes (keine konstruierten Daten, die immer ein schönes Ergebnis erzeugen).

Es gibt im Prinzip kein richtig oder falsch. Die Ergebnisse werden nicht benotet, sind aber **klausurrelevant**.

# Daten für die Übungsaufgabe



## Frankfurt Car Parking Data

The Frankfurt Open Data provides dynamic data on over 20 car parking facilities in Frankfurt am Main. This data was collected on a minute - by minute basis from 17.02.2018 until 09.03.2018.

**Data:** ffm\_parkdaten.csv

**More Information:**

<https://www.kaggle.com/orgesleka/frankfurt-car-parking-data>

## Hessen House Prices Dataset

Over 22000 Houses with Pricing Information from Hessen Germany

**Data:** houses.json

**More Information:**

<https://www.kaggle.com/orgesleka/hessen-house-prices-dataset>



## DBahn Travels Captures

Data captured from trains and travels in different stations of Germany

**Data:** 19.06.20\_travels\_Frankfurt.csv, 19.06.20\_travels\_Mainz.csv

**More Information:**

<https://www.kaggle.com/chemamengibar/dbahn-travels-captures>



# Daten für die Übungsaufgabe



## Vehicle Insurance Customer Data

This is customer data with their vehicle insurance policies. Details about customers and the insurance taken for their vehicles are provided.

**Data:** AutoInsurance.csv

**More Information:**

<https://www.kaggle.com/ranja7/vehicle-insurance-customer-data>

## Bakery Sales for Food & Drinks

**Data:** BakerySales.csv

**More Information:**

<https://www.kaggle.com/alexanderjohn/bakery-sales-for-fooddrinks>



## Forbes Top 2000 Companies of 2017

The World's Largest Public Companies of 2017

**Data:** ForbesTop2000\_2017.csv

**More Information:**

<https://www.kaggle.com/sudhanshuaware/forbes-top2000>

# Übungsaufgabe

Gruppe	Daten	Gruppe	Daten
Gruppe WZ1	Vehicle Insurance Customer Data	Gruppe WZ9	Hessen House Prices Dataset
Gruppe WZ2	DBahn Travels Captures Frankfurt	Gruppe WZ10	Bakery Sales for Food & Drinks
Gruppe WZ3	Hessen House Prices Dataset	Gruppe WZ11	Forbes Top 2000 Companies of 2017
Gruppe WZ4	Forbes Top 2000 Companies of 2017	Gruppe WZ12	DBahn Travels Captures Mainz
Gruppe WZ5	Frankfurt Car Parking Data	Gruppe VIL	DBahn Travels Captures Frankfurt
Gruppe WZ6	DBahn Travels Captures Mainz	Gruppe WIL	Frankfurt Car Parking Data
Gruppe WZ7	Vehicle Insurance Customer Data	Gruppe FKB	Vehicle Insurance Customer Data
Gruppe WZ8	Frankfurt Car Parking Data	Gruppe HEF	Bakery Sales for Food & Drinks

## 02.12.2019 – Vorstellung der Gruppenarbeiten

Jede Gruppe zwischen 10 - 15 min (**max. 15 min**)