



# PROJECT INFORMATION

Statistics I

winter term 2021/22



# Why doing a project?

to be able to **prepare data**, **analyze data** and provide **meaningful answers** to given questions based on the statistical analysis results

# Parts of the Project

- data collection
- data understanding
- data preparation
- data description
- data analysis (hypothesis testing)

# 1. Data Collection

- kaggle

- <https://www.kaggle.com/>

- own research

- UCI Machine Learning Repository

- <https://archive.ics.uci.edu/ml/index.php>

## 2. Data Understanding

- one of the most critical parts
- necessary for further data analysis and hypothesis testing
- both qualitative and quantitative variables need to be included

# 3. Data Preparation

- variables coding
- value labels
- data transformation/recoding
- missing values analysis
- extremes and outliers

# 4. Data Description

- descriptive statistics
- graphs

# 5. Data Analysis (hypothesis testing)

- comparing an average value between male/female, age categories, regions
  - test scores between male/female
  - expenditures between families from different regions
  - time spent on social networks between age categories
- min of 2 hypotheses (1 member in the team)
- min of 3 hypotheses (2 and 3 members in the team)



# General Information

- individually or in groups (max three students)
- will be presented and defended as a part of the final exam
- to be uploaded to Moodle before the exam
  
- extra points for presentation during the semester