

Business Analytics & Machine Learning

Tutorial sheet 0: Data IO

Prof. Dr. Martin Bichler, Prof. Dr. Jalal Etesami
Julius Durmann, Markus Ewert, Johannes Knörr, Yutong Chao
October 25, 2023

Exercise T0.1 *Oktoberfest Beer Data*

In this exercise we will use pandas to analyze a dataset of Oktoberfest beer consumption and prices. To get a quick overview of pandas basics, have a look at: https://pandas.pydata.org/docs/user_guide/10min.html

- a) The provided data is contained in the file "Oktoberfest.csv". Read the data into a pandas dataframe.
- b) Take a look at the dataframe's column names.
- c) What was the price of a beer in 1995?
- d) Which year is the earliest recorded entry of the dataset?
- e) Determine the value range of "Visitors_Total".
- f) Plot the beer consumption over the years.
- g) Create a scatterplot that shows the number of visitors per year. Calculate the correlation between "Visitors_Total" and "Beer_Consumption".
- h) Compute the mean and the variance of beer price during the period from 2000 to 2007.
- i) Add a new column to the dataframe, describing the difference of beer prices between two consecutive years. Plot this difference against "Year".