

Business Analytics & Machine Learning

Tutorial sheet 0: Data IO

Prof. Dr. Martin Bichler

Julius Durmann, Markus Ewert, Yutong Chao, Dr. Mete Ahunbay

Exercise T0.1 *Oktoberfest Beer Data*

In this exercise we will use pandas to analyze a dataset of Oktoberfest beer consumption and prices. The data for this exercise is contained in *visitors.csv* and *beer_price.csv*.¹ To get a quick overview of pandas basics, have a look at the "*10 minutes to pandas*" guide².

- a) Read the data from *visitors.csv* into a pandas dataframe.
- b) Take a look at the dataframe's column names. Display the ten top columns of the dataframe.
- c) Select and display ...
 - i) ... only the column of visitors.
 - ii) ... only the fifth row of your dataframe.
- d) How many visitors were there in 1995?
- e) What is the value range of the attribute "Beer consumption (million liters)"?
- f) What was the year with the greatest beer consumption? How much beer was consumed?
- g) Plot the beer consumption over the years. Turn on the grid of your plot.
- h) Compute the correlation matrix of the dataframe. Which conclusions can you draw?
- i) Load the data from *beer_price.csv* into a second dataframe. Then merge the two dataframes based on the year with an outer join. Observe which values you get for years that are not listed in some of the data sets.
- j) Until 2001, the beer price is reported in DEM ("Deutsche Mark"), the former German currency. Adjust the beer prices (min and max) to give values in EUR. You may assume that 1 EUR = 1.05583 DEM.
- k) Compute the mean and variance of "Min price" during the period from 2000 to 2007.
- l) Add a new column to the merged dataframe, describing the difference in minimum beer prices between two consecutive years in percent. Plot this difference against "Year".
- m) Create a bar plot displaying the minimum and maximum prices over the years.
- n) Compute estimates (lower and upper bound) of the beer revenue and visualize them by a bar plot.

¹Sources: Statista.com - Visitors | Statista.com - Beer Consumption | Statista.com - Beer Price

²https://pandas.pydata.org/docs/user_guide/10min.html