Prof. Dr. Jörg Müller
Arthur Fleig

AI VIII – Serious Games

UNIVERSITÄT BAYREUTH

Exercises for **Programming, Data Analysis, and Deep Learning in Python** (SoSe 2021)

Exercise Sheet no. 9, *Deadline*: Monday, June 21, 10:15

Notes

- Pay attention to the notes on the previous sheet.

Exercise 26 Star Wars: Part 1 (programming exercise) (8 points)

This is the first part of a trilogy of exercises. Answer the following questions using Python commands. Manual work, e.g., counting by hand, is allowed only in the header. Note that the first two rows belong to the header. Do not change anything in the csv file.

a) Import the csv file[1] with only the first line as header and with the first column being the index. (Treat the second header line as data for now.) Make sure the index column ("RespondentID") is of type `'Int64'`.[2]

b) The column names are too long. Change them to
`['seen_any_film', 'fan_sw', 'seen_ep1',...,'seen_ep6', 'rank_ep1',...,`
`'rank_ep6', 'like_char1',...,'like_char14', 'first_shot',`
`'know_exp_univ', 'exp_univ_fan', 'fan_st', 'gender', 'age', 'h_income',`
`'education', 'location']`.
Use the first row (the second header line) to create a dictionary of film names with `'seen_ep1'`,...,`'seen_ep6'` as keys and to create a dictionary of character names with `'like_char1'`,...,`'like_char14'` as keys. Then drop the first row (the second header line) from the data frame.

c) How many respondents are in the data set? How many are male, how many are female, and how many did not answer this question?

[1] Source: "star-wars-survey" at `https://data.fivethirtyeight.com`.
[2] Check the "dtype" parameter in the `read_csv` method: `https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html`

Exercise 27 Star Wars: Part 2 (programming exercise) (8 points)

This is the second part of a trilogy of exercises. Import the associated csv file[3] with the first column as index of type `'Int64'` and answer the following questions using Python commands. Do not change anything in the csv file.

a) In the columns `'seen_ep1'`,...,`'seen_ep6'`, the data entries contain the respective title in case the respondent has seen this film and are empty otherwise. Replace these values with "Yes" or "No", depending on whether the respondent has seen the respective film.

b) In the columns `'like_char1'`,...,`'like_char14'`, replace the entries "Very unfavorably" by `-2`, "Somewhat unfavorably" by `-1`, "Neither favorably nor unfavorably (neutral)" by `0`, "Somewhat favorably" by `1`, and "Very favorably" by `2`. (All numbers are integer values, not strings.) Leave the other entries in these columns as they are.

c) How many respondents are neither a fan of the Star Wars franchise (column `'fan_sw'`) nor of the Star Trek franchise (column `'fan_st'`)? What percentage of respondents who answered both questions is neither a fan of the Star Wars franchise nor of the Star Trek franchise?

d) Calculate the percentage of Star Wars fans who have seen exactly one out of the six films in total.

Exercise 28 Star Wars: Part 3 (programming exercise) (8 points)

This is the third part of a trilogy of exercises. Import the associated csv file[4] with the first column as index of type `'Int64'` and answer the following questions using Python commands. Do not change anything in the csv file.

a) Plot a histogram of the age of all respondents who have seen at least one Star Wars film according to the `seen_any_film` column. Plot a histogram of the age of all respondents who have seen all of the latter episodes (IV to VI) but none of the first three (I to III). What observation do you make?

b) How many of those who have seen at least one of the six films according to the `seen_any_film` column are unfamiliar ("Unfamiliar (N/A)") with at least one character? Give the absolute number and the percentage.

c) How many respondents have seen all six films? Determine which film is the least favorite according to the respondents who watched all six films. To this end, calculate the column-wise mean and the median for the required columns (and rows). The respondents ranked the films from "1" (best) to "6" (worst).

d) Which character is liked least by men (average score)? Which character is liked most by women (average score)? (Instead of the character name you may provide the column name, e.g., `like_charN`.)
Hint: You may need to convert some columns to numeric values. To this end, applying `to_numeric` to these columns might be helpful.

[3]Modified from source: "star-wars-survey" at `https://data.fivethirtyeight.com`.
[4]Modified from source: "star-wars-survey" at `https://data.fivethirtyeight.com`.