Work on project. Stage 5/5: Visualize it!

Project: Data Analysis for Hospitals

Visualize it!

Hard 1 hour **202** users solved this stage. Latest completion was **3 minutes ago**.

Description

Are you ready to catch sight of your data?

Graphics are arguably the most accessible way to represent the data and its structure. Sometimes, it can help to find the main data patterns and deviations. We will use data visualization methods to conclude our dataset.

In the last stage, you need to create data visualization to answer the following questions:

1. What is the most common age of a patient among all hospitals? Plot a histogram and choose one of the following age ranges: 0-15, 15-35, 35-55, 55-70, or 70-80
2. What is the most common diagnosis among patients in all hospitals? Create a pie chart
3. Build a violin plot of height distribution by hospitals. Try to answer the questions. What is the main reason for the gap in values? Why there are two peaks, which correspond to the relatively small and big values? No special form is required to answer this question
   There is a comprehensive explanation of violin plots by Eryk Lewinson.

19 / 19 Prerequisites

- ✓ Basics of sets 2
- ✓ Basics of functions 2
- ✓ Properties of functions 1
- ✓ Properties of numerical functions 1
- ✓ Cartesian coordinates 1

Show all

Hint

Please note that the answers are independent of each other.

At this stage, we recommend using `pandas` visualization tools. However, feel free to use `seaborn`, `matplotlib` or any other library.

The tests to check graph content are very limited and we are sure that you can easily answer the questions without plotting any charts. Despite this, please be curious to answer them using graphs. It is a very valuable skill for a data scientist to plot and interpret the data.

Objectives

Steps 1-8 are the same as in the previous stage. The fifth stage requires completing the following steps:

1. Read the CSV files with datasets
2. Change the column names. The column names of the sports and prenatal tables must match the column names of the general table
3. Merge the dataframes into one. Use the `ignore_index=True` parameter and the following order: `general`, `prenatal`, `sports`.
4. Delete the `Unnamed: 0` column
5. Delete all the empty rows
6. Correct all the gender column values to `f` and `m` respectively
7. Replace the `NaN` values in the gender column of the prenatal hospital with `f`
8. Replace the `NaN` values in the `bmi`, `diagnosis`, `blood_test`, `ecg`, `ultrasound`, `mri`, `xray`, `children`, `months` columns with zeros
9. Answer questions 1-3. Output the answers in the specified format. The answers to the first two questions should be formatted as in the examples. No special form is required to answer the third question

If you have corrupted CSV files, please download them and unzip in your working directory.

Example

The input is 3 CSV files, `test/general.csv`, `test/prenatal.csv`, and `test/sports.csv`.