DA/AB

Fake Piscine

Week 3

"Panda, Panda,
Panda, Panda, Panda, Panda, Panda"

INSTRUCTIONS

This week we will dig into Pandas, a software library written for the Python programming language for data manipulation and analysis.

In #subject you will find the dataset required in order to make the following exercises. You will need to check and, eventually, fix any problem in how the dataset is initially presented to you.

Again, create a new block notes in Google Colab and write your code in there. Once you are done, share it with gscala(slack or discord).

week3.ipynb

Import "Salaries.csv" in your Block Notes. Then:

1. Check the head of the dataset
2. Use the .info() method to find out how many entries there are
3. Check the highest amount of OvertimePay in the dataset ?
4. What is the job title of JOSEPH DRISCOLL ?
5. How much does JOSEPH DRISCOLL make (including benefits)?
6. What is the name of highest paid person (including benefits)?
7. What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid?
8. What was the average (mean) BasePay of all employees per year? (2011-2014) ?
9. How many unique job titles are there?

10.What are the top 5 most common jobs?
11. How many Job Titles were represented by only one person in 2013? (e.g. Job Titles with only one occurence in 2013?)
12. How many people have the word Chief in their job title?

BONUS

netflix&chill

Using the dataset "titles.csv", answer these questions graphically - that is - by creating visualizations through matplotlib.

1)Which are the years with the highest number of movies released on Netflix?
2)Is there a majority of tv shows or movies on Netflix?
3)What are the most popular genres on Netflix and the ones with the highest imbd rating?
4)Which is the year with most popular US movies?

HYPERBONUS

scrapit

You will need to extract, clean and organize a dataset ready to be manipulated with Pandas (you can use Salaries.csv or titles.csv as examples). In order to do so, you will need to pick a place on the internet which sparks your interest and then build a simple web scraper which will extract useful informations from it, all well organized in a .csv file. You could start by looking at Beautiful Soup. Be sure to detail with text every passage of your code.

DO NOT DO ANYTHING ILLEGAL, BE SURE THAT WHATEVER YOU ARE DOING IS PERFECTLY ALLOWED. IF YOU ARE UNSURE, DO NOT DO IT.