Exploring Chocolate Consumption and Happiness Scores

Introduction

In this exercise, you will analyze data on chocolate consumption and happiness scores across different countries and try to identify patterns and connections between the two. This exercise will guide you through using some of the popular python packages, such as pandas (to manage data) and matplotlib (to create visualizations). By the end, you will be able to share your findings with our students of the school.

Setup

Before beginning the exercise, please make sure that you:

1. Access the CERNBox service if you do not have a CERNBox account yet (https://cernbox.cern.ch). This is needed to access SWAN;
2. Start a session in the SWAN service at https://swan.cern.ch. The session should be started with the default parameters and use the "LCG 105a" software stack;
3. Create a new SWAN project by cloning the git repository with the contents and notebooks that will be used during this exercise:

   `https://github.com/swan-cern/it-csc.git`

   Hint: use the button to "Download project from git" in the Projects view of the classic UI.

Exercise

Project

Once the content of the repository is cloned and the project is created for you, analyze its structure and its files (including the ones inside the "data"

directory), both from the file explorer (under the "Projects" tab) and from the session terminal (**Hint:** SWAN projects are located in `$HOME/SWAN_projects`).

Part 1: Chocolate Consumption by Country

First, we will analyze chocolate consumption around the world to identify which countries have the highest chocolate intake per capita.

Instructions

1. Open the jupyter notebook present in the project;
2. Import the dataset with pandas to the first cell of the notebook. (**Hint:** you can use the "read_csv" method from the pandas library (Pandas Documentation));
3. Run the first cell of the notebook (either by pressing the "start" button in the menu or by pressing SHIFT and ENTER) and observe the result. Which is the country that consumes the highest amount of chocolate per capita?

Part 2: Happiness Scores by Country

Now, we will explore the happiness scores by country to understand which countries report the highest levels of happiness.

Instructions

1. Analyze the second and third code cells of the jupyter notebook present in the project;
2. Import the dataset with pandas, as you did for the previous part;
3. Sort the data to find the top 10 happiest countries (**Hint:** You can use the "sort_values" method from the pandas library (Pandas Documentation));
4. Run the third code cell of the notebook and observe the result. Which are the top 10 happiest countries? Write your answer in a markdown cell below the output.

Part 3: Exploring the relationship between Chocolate and Happiness

For this part, we will examine whether there is any visible connection between the chocolate consumption and happiness scores. We will combine both datasets to check for any patterns.

Instructions

1. Merge the two datasets on `Country` to analyze the relationship, by using the "merge" method on the datasets (Pandas Documentation);

2. Create a **scatter plot** (Matplotlib Documentation) where chocolate consumption is on the x-axis, happiness score is on the y-axis, and each country is represented as a point. For this, complete the last cell of the jupyter notebook with the missing functions (**Hint:** Take a look at how the plots of the previous two parts are created and use the functions needed to create the plot. The scatter plot can be created by resorting to the "scatter" method of matplotlib);

3. Analyzed the scatter plot that was displayed. Is there a connection between Chocolate and Happiness? Write the answer to this question in a new markdown cell at the end of the notebook.

Part 4: Share your results

To complete the exercise, share your notebook (which contains the resulting plots) with the students near you, by asking for their usernames. You can do this by going to the project folder and press the "Share" button at the top right menu, just above the files list. Moreover, you can also check the projects that have been shared with you under the "Share" tab of the classic UI.