Mini Project 1 – Where do I fly next?

Note:
The mini projects are individual assignments. You could and it is encouraged to brainstorm the assignments with your classmates, but you need to have individual codes and reports. All written reports and codes are going to be checked for plagiarism.

Task description

Step1 - Webscraping

- Choose a city in the world to fly to from Helsinki, for a trip, on the condition that there are at least 50 flights available to choose from;
- Choose the date for the trip;
- Scrape the data about the flights from Helsinki to that city from three booking websites of your choosing, e.g., Skyscanner, Kayak, Momondo, Expedia, Booking.com etc.
- What are the challenges in scraping the data? How do you store the scraped data for a further easy use?

Step 2 – _Data processing

- What are the data processing steps for having a proper dataset to be used in the next step?

Step 3 – Exploratory data analysis (EDA)

Use the appropriate visualisation method to provide a comprehensive EDA, e.g., price ranges, number of stops, airlines, departure and arrival times, duration of the flight, layover time (if the flight is not direct), etc.

Step 4 – Interaction with the user

Your code should be able to perform the following task to give the user the list of "ideal" flight to the chosen city:

- The user inputs certain preferences for the flight, e.g., price ranges, trip duration ranges (how long from departure to destination), whether the flight is direct or has one or more stops (if it has stops then layover duration), flight company from available options etc.
- The user can choose the criteria based on which the flights should be sorted.
- The code returns the sorted list of all flights based on the above criteria.
- The user should be able to input the earliest convenient time of departure on the departure day from your list, maximum travel time (trip duration from departure to destination) and choose a preferred airline company from available options and get the following information:

- Cheapest flight for the preselected data with the following details:
  - Total duration of the flight?
  - Whether the cheapest flight is direct or has one or more stops?
  - If it is not direct how long is the layover and in which city? If it has multiple stops, print all layovers.
  - Type of the aircraft where available, etc.

- Fastest flight for the preselected data with the same details as above.

Reporting

Write a scientific report which includes:

- Introduction (What is the problem you are solving?)
- Data collection (What is the data you are collecting and what is your method of choice?)
- Data analysis (How did you visualise the data and how do you interpret the data visualisation? What are the interesting observations you could report?)
- Conclusion (What were the "scientific" bottlenecks? How did you overcome them?)

You will also be graded on the overall quality of your report. To give you an idea of a proper project report, a sample is uploaded in Moodle.

You need to use Google Colab for your code. Remember to include the installation steps in the code if you use any specific package. Upload the code and the written report into Moodle!