Travel recommendation project: Analyzing the thought process and reviewing the code.

Made by Abdelrahman Hesham (320210232)

Explanation

The code starts with parsing the data from the three csv files and converting them into three pandas data frames called flights, hotels, and users respectively.

The first two function are merely helper functions. The first one: showData is to show the first five rows of each data frame just so I could look on the data types in each row and the exact names of columns. The second one is to compare between two lists, it returns true if the length and content of the two lists are equal and false otherwise.

Then we initialize three empty lists which are going to hold the unvisited cities, visited cities, and unique cities.

The next two function are meant two fill the previously mentioned three lists. The first one: getAllCities(flights) it takes the flights data frame and loops over all the values and only appends the unique cities into the unique cities list. The second function: getUnvisitedCities(userCode, flights) use loops over all the data in flights and checks the cities that a user with the given user code has visited and appends them to the visited cities list. The final list which is unvisited cities is computed by getting the difference between the visited and unique cites lists.

The final function called recommendation is where the magic happens. It accepts the parameters: user code, number of cities and budget. It initializes the previous two functions using the input parameters, creates an empty recommendation list, then checks if the two lists: visited cities, and unique cities are equal using the previously defined compare function and if they are, it just tells the user that he has seen all the cities. If not, it goes on to check if the number of cities he wants to see is more than the cities left to be visited which if it is true it tells the user that he exceeded the number of cities he can visit. If not, it initializes a min variable with an

arbitrary really high number then it loops over the whole hotels data frame and checks three things.

1. if the place of the hotel is in an unvisited city
2. The price per night in that hotel does not exceed the budget
3. That that place has not been already added to the recommendation array.

If they are all true, it checks if the price is less than the max which we set to 99999 for example. If so, it sets that price as the new minimum and saves the row index in a variable called index. After the whole loop is over it subtracts the price from the budget and adds the place to the recommendation array.

The complexity here is O(n^2) since the recommendation function has a nested for loop.