

Exam scientific programming (DS) 2022-23

PART A

11 November 2022

Please read the following instructions carefully:

You should name your exam files as follows:

name_surname_id.py

For instance (Antonio_Longa_123456.py)

SUBMISSION INSTRUCTIONS:

Load to EXAMINA your solution.

DOCS:

Here you can find the list of accessible websites:

- numpy.org, pandas.pydata.org, matplotlib.org, docs.python.org, antoniolonga.github.io

Myntra is an Indian fashion e-commerce. You are given a dataset (**Clothings.csv**) of 10000 sold items. For each item you have the following information:

- **Product_id**: it is an integer that univocally represent a product, (each product has an unique product id).
- **BrandName**: it reports the brand name of the item
- **cod_Category**: it is a code representing the macro category (cod_0, cod_1,...)
- **category_by_Gender**: it can be an item for males or for females (men, women)
- **Price**: the price of the product (in €)
- **Ratings**: the rate of the product given by customers.

Here you can find an example of the dataset (**Clothings.csv**):

	Product_id	BrandName	cod_Category	category_by_Gender	Price	Ratings
0	2296012	Roadster	Cod_0	Men	17.988	3.9
1	13780156	LOCOMOTIVE	Cod_0	Men	13.788	4.0
2	11895958	Roadster	Cod_1	Men	16.788	4.3
3	4335679	Zivame	Cod_2	Women	15.540	4.2
4	11690882	Roadster	Cod_3	Women	7.188	4.2

You are also given a second dataset (**Category_conversion.csv**) which contains the conversion from cod_category to the actual category. In particular, it contains two columns:

- **cod_Category**: it is a cod representing the macro category (cod_0, cod_1,...)
- **Category**: the actual category (Bottom Wear, Topwear, Lingerie ...)

Here you can find an example of the dataset (**Category_conversion.csv**):

```
cod_Category,Category
Cod_0,Bottom Wear
Cod_1,Topwear
Cod_2,Lingerie & Sleep Wear
Cod_3,Western
Cod_4,Sports Wear
```

QUESTIONS:

1. Build a function to print the Product_id and the price of the most expensive item
2. Build a function to extract the rating of the following items, given the following product_id [12335810,4335679,12134590]
3. Build a function to print the maximum rating of the item with cod_category = Cod0 and category_by_Gender = Men
4. Print the category of the product with id = 11535928
5. Build a function that takes in input the dataset and a minimum price, then compute the average rank of those items with a price greater than the minimum price! the default value for the minimum price must be 30. Then use the function with a minimum price equal to 30 and then equal to 50.
example:

```
>>> compute_mean(dataset,minimum_price=30)  
>>> 4.130418250950572
```
6. Make a unique plot with distributions **plt.hist(list)** of Category_by_Gender, Price and Ratings. Then in the last plot, you have to show the scatter plot between Rate and price, using **plt.scatter(price,ratings)**. You should also limit the y axes from 0 to 5 (**plt.ylim(0,5)**).
NOTE: remember to put titles, and for the scatter plot even y and x labels.
You should obtain a plot like the one below:

