Scientific Programming with Python - Exercise 2

Attached to this exercise is a CSV dataset, "mobile_prices_1.csv" pertaining to mobile phone prices. The dataset contains specifications of mobile phones, each with an id, a set of features, and a price. You will be tasked with some data manipulations as well as some questions regarding the behavior of the data. You must show your work (code used) for any question answered.

Your submission must be a Google Colab Notebook (.ipynb file). Make sure your answers are ordered correctly (as the questions) and there is no residual test code.

Following are the feature descriptions for this data set:

battery_power: mAh capacity of the battery (numerical)
m_dep: the thickness of the phone (numerical)
mobile_wt: the weight of the phone in grams (numerical)
px_height: screen height in pixels (numerical)
px_width: screen width in pixels (numerical)
ram: amount of available RAM in MB (numerical)
sc_h: screen height in centimeters (numerical)
sc_w: screen width in centimeters (numerical)
talk_time: length of maximum call time in hours for 100% charged battery (numerical)
bluetooth: does the device have Bluetooth (categorical – Yes/No)
gen: what maximum generation network is supported (categorical – 2/3/4)
cores: core architecture (categorical – single/dual/triple/quad/penta/hexa/hepta/octa)
speed: what is the processor speed level (categorical - low/medium/high)
sim: is the sim support single or dual (categorical - Single/Dual)
f_camera: megapixel quality of the front camera (numerical, empty if no camera)
camera: megapixel quality of the camera (numerical, empty if no camera)
memory: the internal memory in MB (numerical)
screen: the screen type (categorical – Touch/LCD)
wifi: the top wifi standard supported (categorical – none/b/a/g/n)

Task 1

1. Load the data into a Pandas Dataframe.
2. Which of the categorical features are nominal and which are ordinal?
3. Add a column that holds the total screen resolution for each device. Name it resolution.
4. Add a column that holds the DPI (dots per inch) of the screen width and name it DPI_w. **DPI of a digital image is calculated by dividing the total number of dots wide by the total number of inches wide. Do not leave NaN/Infinite values.**

5. Add a column that holds the ratio battery_power/talk_time and name it call_ratio. 6. Do not leave NaN/Infinite values.
Change the memory column to hold the memory in GB instead of MB.
7. Include the output of the `describe()` function of the dataframe.
8. Convert the following features into categorical series in the Dataframe: speed,screen,cores

Task 2

1. How many phones do not have a camera at all (front or back)?
2. What is the average battery power for single-sim phones that have a camera or front camera with a higher resolution than 12 megapixels?

3. What is the ID and price of the most expensive phone that has no wifi, a touch screen and weighs more than 145 grams?
4. Create a pivot table that shows the percentage of phones with Bluetooth per generation, pivoted around the phone generation and split by "ram" quartiles. (i.e. the rows are the generation number and the columns are 4 quartiles of ram size).
5. Create a new Dataframe based on the original that has the following features: [id, battery_power, ram, talk_time, Bluetooth, cores, sim, memory, price], and contains a random sampling of half of the medium speed phones.
6. Using this new dataset, what is the maximum total talk time you can achieve if you use 3 phones, and which 3 phones will you use?

Do not forget to write clean, readable code, and document your functions!

Do not forget to title your graphs and add legends and color bars!

Grading:

60 points for correctness.

20 points for efficiency (correctly using Pandas features).

20 points for code cleanliness and readability.