A3: Laptop Data Cleaning and Preparation

The dataset used in this exercise was downloaded from Kaggle at the following URL: https://www.kaggle.com/datasets/jacksondivakarr/laptop-price-prediction-dataset?resource=download. A brief description of each included column is provided below. Perform the tasks listed below and answer the associated questions. Deliver the results as an HTML webpage generated from an R Markdown or Quarto file. Use headers or text to differentiate each component of the assignment. Make sure to include both the code and the results in your submission. Hint: the forcats, stringr, and dplyr packages will be needed. When reading in the data, make sure all character columns are treated as factors.

brand: laptop brand name
name: name of laptop
price: price in US Dollars×100 (divide by 100 to get price)
spec_rating: specification score (0 to 100)
processor: processor name
CPU: central processing unit (CPU) specs
Ram: amount of installed RAM
Ram_type: type of RAM
ROM: Size of hard disk
ROM_type: type of hard disk (SSD or Hard-Disk)
GPU: installed graphics processing unit (GPU)
display_size: size of display in inches
resolution_width: resolution in width dimension in pixels
resolution_height: resolution in height dimension in pixels
OS: operating system
warranty: length of warranty in years

Q1: How many factor levels of RAM are differentiated? (2 Points)
T1: Write code to recode the Ram factor levels as follows and convert to a numeric type (Original → New): "12GB" = "12", "16GB" = "16", "2GB" = "2", "32GB" = "32", "4GB" = "4", "64GB" = "64", "8GB" = "8". (3 Points)
Q2: What percentage of the computers have at least 16GB of RAM? (2 Points)
T2: Write code to recode the ROM factor levels as follows and convert to a numeric type (Original → New): "128GB" = "128", "1TB" = "1000", "256GB" = "256", "2TB" = "2000", "32GB" = "32", "512GB" = "512", "64GB" = "64". (3 Points)
Q3: What percentage of the computers have at least 512 GB of ROM? (2 Points)
T3: Write code to create a field that indicates whether the machine has an Intel processor. (3 Points)
Q4: What percentage of the computers have an Intel processor? (2 Points)
T4: Write code to create a field that indicates whether the machine has an AMD processor. (3 Points)
Q5: What percentage of the computers have an AMD processor? (2 Points)
T5: Write code to create a field that indicates whether the machine has an NVIDIA GPU. (3 Points)
Q6: What percentage of the computers have an NVIDIA GPU? (2 Points)
T6: Write code to create a single column that differentiates between Intel and AMD processors. Any other type should be coded as "Other". (3 Points)
T7: Write code to create a single column that will differentiate between i3, i5, i7, and i9 Intel processors. All other processors should be coded to "Other". (4 Points)
Q7: What percentage of the computers has an i7 or i9 processor? (2 Points)
Q8: What percentage of computers have an AMD processor, at least 16GB of RAM, and at least 1TB of ROM? (2 Points)
Q9: What percentage of the computers have an Intel i7 processor, at least 1TB of ROM, at least 32GB of RAM, and a solid-state hard drive (SSD). (2 Points)