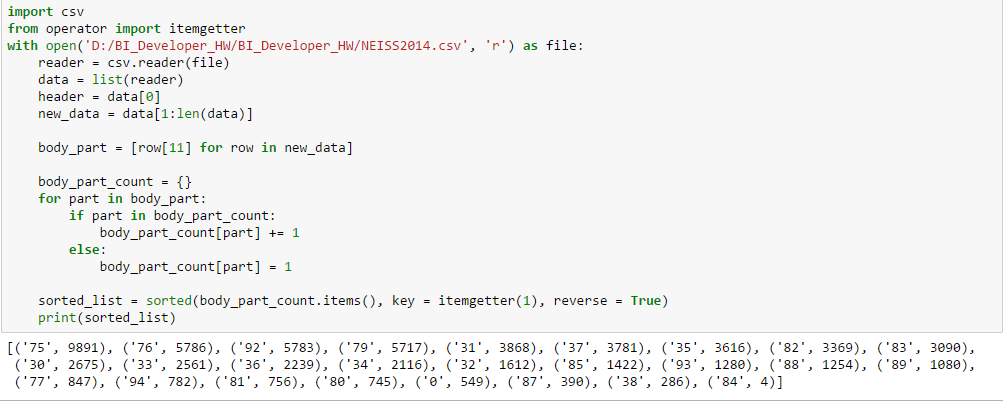
**ANSWERS**

**QUESTION 1**

1. ***What are the top 3 body parts most frequently represented in the dataset?***
2. ***What are the top 3 body parts that are least frequently represented?***



* As you can see from the findings above,

The top 3 body parts represented are: Head (75), Face (76) and Finger (92)

The top 3 body parts least frequently represented are: Not recorded (87), Pubic region (38) and 25 – 50% of body (84)

* To answer this question, I read the csv file and converted that into a list. After removing the header, I created a list comprehension to iterate the body\_part column and created a dictionary to count each body part that occurred in the dataset.
* After looping through the entire column, the challenge I faced was to get the highest and lowest body parts. In python, we cannot sort a dictionary, so I converted the dictionary into a list and sorted that based on its values rather than its keys.

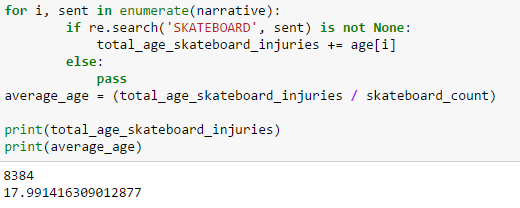
**QUESTION 2**

1. ***How many injuries in the dataset involve a skateboard?***
2. ***Of those injuries what percentage were male and what percentage were female?***



* As you see above, the number of injuries involving a skateboard are 466 and 384 out of those are male whereas 82 are female. 82.40 is the percentage of male injuries involving a skateboard and for females it is 17.59%.
* To get the result, my biggest ordeal was to get ‘skateboard’ from narrative column. I used the regular expression module to extract it. Now, to get the number of male and female injuries involved in a skateboard I used the enumerate function and performed an iteration over the narrative column.
* In the same loop, I extracted the string ‘SKATEBOARD’ alongside determining whether the sex is ‘male’ or ‘female’ and created a counter for the number of skateboard, male and female.
* After finding the total count, I found the percentage of male and female injuries involving a skateboard

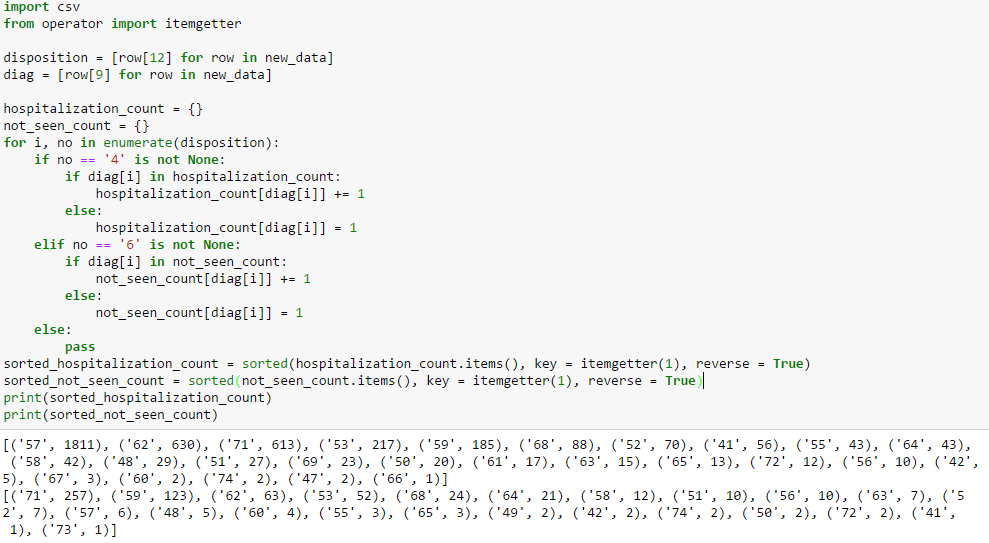
1. ***What was the average age of someone injured in an accident involving a skateboard?***



* The average age of someone injured in an accident involving a skateboard is 17.99
* The total age of all the patients is 8384.
* Here, I extracted the string ‘SKATEBOARD’ and found the corresponding age of the patients. Summed up all ages in the variable total\_age\_skateboard\_injuries. Divided that by the skateboard\_count to get the average age of a patient.

**QUESTION 3**

1. ***What diagnosis had the highest hospitalization rate?***
2. ***What diagnosis most often concluded with the individual leaving without being seen?***



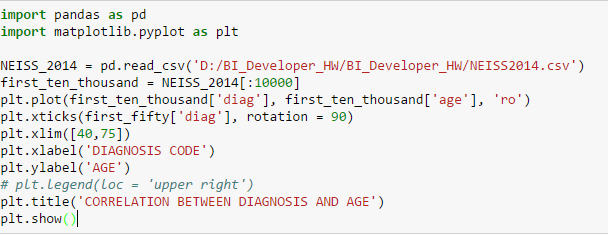
* The diagnosis with the highest hospitalization rate is ‘Fracture (57)’ and the diagnosis most often concluded with the individual leaving without being seen is ‘Other/Not Stated (71)’.
* Here, I used the enumerate function to iterate over the disposition column and created 2 conditions where the disposition code was ‘4’ and ‘6’, under those conditions I found the diagnosis that occurred and stored their values in a dictionary.
* Again, since dictionaries cannot be sorted I converted them into a list and sorted their values from high to low.

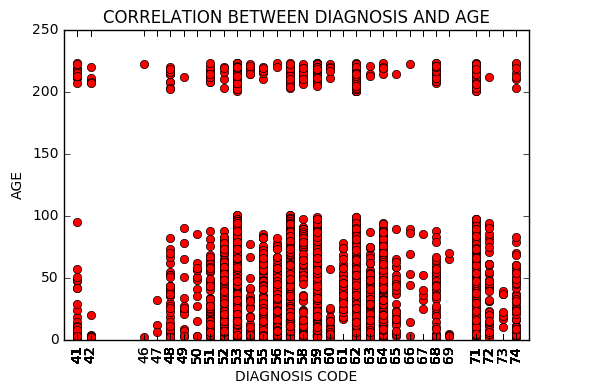
1. ***Briefly discuss your findings and any caveats you’d mention when discussing this data***

* The highest reported injuries are on Head (9891), next highest are on Face (5786); A difference of ~4100 between the first two. The lowest reported injuries are 4 and those are on (25 – 50% of body).
* Only 82 out of 466 skateboard injuries are female, which is not a high number, this shows not a lot of female skateboard or not many females get injured while skateboarding.
* Average age of someone involved in a skateboard injury is ~18, which means majority of skateboarders are young and teenagers.
* Only one patient diagnosed with Hemorrhage was treated and hospitalized and only one patient that had burns or affected by radiation left without being seen / left against medical advice.

**QUESTION 4**

1. ***Visualize any existing relationship between age and reported injuries.***

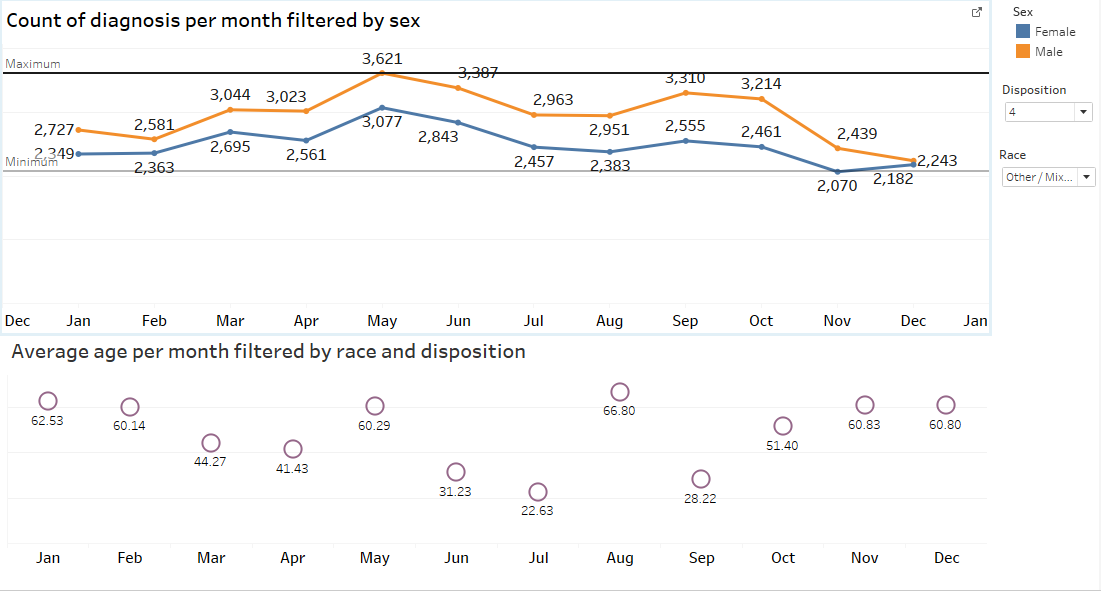




* After visualizing the first 10,000 rows of the dataset I noticed that the highest diagnosis is for
  + Contusions, Abrasions (53)
  + Fracture (57)
  + Laceration (59)
  + Internal organ injury (62)
  + Strain or Sprain (64)
  + Other / Not Stated (71)
* Diagnosis is sporadic from the age 0 - 100 and there has been absolutely no diagnosis from the age 100 – 200 then some from age 200 – 223.

**QUESTION 5**

1. ***Investigate the data however you like and discuss any interesting insights you can find in the data. (Don’t feel pressured to spend hours on this)***



* I’ve created a tableau dashboard where I count the number of diagnosis per month based on sex, as you can see the count of diagnosis for males have been consistently higher than that of females and the highest diagnosis was recorded at 3621 number of diagnosis in the month of May whereas the lowest was that for females at 2070 number of diagnosis in November.
* I also calculated the average age of patients per month filtered by the race and disposition. I noticed that the average age for white people is stable and between 40 – 48 if the disposition filter is set to all. Also, there are absolutely no American Indian / Alaska Native when the disposition is Fatality, including DOA, died in the ED (8).