**Peleg’s Assignment**

In this following document I’ll present the Workflow of my assignment, the project was done in Python – Jupyter Notebook.

1. ***Identify different device models from the data, how many different models can you identify?***

The file that was given to me contains 3 columns: mac, timestamp, user\_agent.  
  
In order to identify the devices I have researched online and found the **user\_agents** library. I also import pandas library in order to analyze the data.



My next step was to import the file from the assignment – data.



I defined the column user\_agent as var



The following line is parsed the column user\_agent and put the outcome into a new column Device Mode.



I’ve extracted the device model using the function device.model from user\_agents library.



The final step was to drop duplicates based on Device Mode column, because we want to know how many different models I have extracted from the data. Then I used the function len in order to know the exact number.



I’ve identified **469** different devices out of the file – data.

***2. What is the most popular model you can find?***

I have imported 3 libraries - pandas, user\_agents (same reason as (1)) and matplotlib to create a bar chart.



My next step was to import the file from the assignment – data.



I’ve drop duplicates from the column mac. Since the data is raw, there can be duplicates mac addresses.



The following line is parsed the column user\_agent and put the outcome into new column Device Mode.

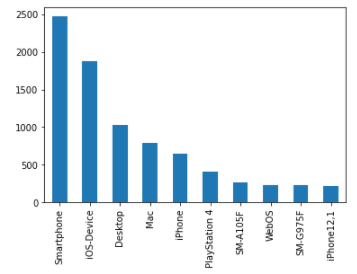


I’ve extracted the device model using the function device.model from user\_agents library.



My final step was to present the data in a bar chart, I used value\_counts function in order to count the devices per model, head(10) to present top 10 models and plot.bar in order to present the outcome in a Bar Chart.

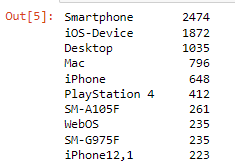




As we can see at the chart, **PlayStation 4** was the most popular device that I can find. If we are looking for the most popular device model it is **SM-A105F** (Samsung A10).

***3. Insight about Apple device***

According to the data it seems like it is more difficult to identify an Apple product.



As we can see here (top 10 most popular models), I couldn’t identify 2474 Smartphones, but if we investigate Apple products, we can see that 2 out of the top 4 are apple products. If we sum up the apple devices that weren’t identify by the code, we will get 2520.

***4. How can we estimate, from the data alone. When was it collected?***

I’ve researched online how can I extract the timestamp based on user agent, but I couldn’t find anything.

I can assume that since we’ve exported the data by using a protocol, we have a documentation when we did it. In that way we can know the time stamp because we know when we sent the protocol.