**Mobile App Statistics (Apple iOS app store)**

URL: <https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>

The ever-changing mobile landscape is a challenging space to navigate. The percentage of mobile over desktop is only increasing. Android holds about 53.2% of the smartphone market, while iOS is 43%. To get more people to download your app, you need to make sure they can easily find your app. Mobile app analytics is a great way to understand the existing strategy to drive growth and retention of future user.

With million of apps around nowadays, the following data set has become very key to getting top trending apps in iOS app store. This data set contains more than 7000 Apple iOS mobile application details. The data was extracted from the iTunes Search API at the Apple Inc website. R and linux web scraping tools were used for this study.

Data collection date (from API): July 2017.

Dimension of the data set: 7197 rows and 16 columns.

Variables:

1. "id" : App ID
2. "track\_name": App Name
3. "size\_bytes": Size (in Bytes)
4. "currency": Currency Type
5. "price": Price amount
6. "ratingcounttot": User Rating counts (for all version)
7. "ratingcountver": User Rating counts (for current version)
8. "user\_rating" : Average User Rating value (for all version)
9. "userratingver": Average User Rating value (for current version)
10. "ver" : Latest version code
11. "cont\_rating": Content Rating
12. "prime\_genre": Primary Genre
13. "sup\_devices.num": Number of supporting devices
14. "ipadSc\_urls.num": Number of screenshots showed for display
15. "lang.num": Number of supported languages
16. "vpp\_lic": Vpp Device Based Licensing Enabled

**Please consider the following tasks**

1. Describe the dataset (number of observations and variables, missing values, data types).

2. Conduct a descriptive data analysis (frequency data analysis, central tendency and dispersion measures, variables’ distribution characteristics, graphical data analysis, pivot tables). Interpret the results. Note that data analysis involves the following steps:

- research task setting and selection of variables,

- selection of the approproate analysis method,

- calculations,

- interpretation of results.

Please note that if you calculate something in pandas, but don’t explain why you do the calculations and don’t interpret the results of the calculations, the task is estimated at 0 points. It will be considered that the **data** **analysis** was not done.

3. Proposed questions to consider (you can formulate your own questions):

- Which primary genres are more and less popular?

- How do you visualize price distribution of paid apps, please, descripe the shape of it?  
- How does the price distribution get affected by category?

- As the size of the app increases do they get pricier?  
- How are the apps distributed category wise?

- Find the most popular genres in each content rating group.

- Describe the main characteristics of the apps with the biggest size.

- Compare the characteristics of free and non-free apps.

Please send the Jupyter Notebook with comments and answers to amelikyan@hse.ru from your **personal** e-mail. The task can be done individually or in a group of 2-3 students. If the task is done in a group, then one member of the group can send the task indicating the names of the group members in the subject of the email.