Report on learning practice # 1

Analysis of univariate random variables

Performed by:

Grandilevskii Aleksei

J4133c

Dubinin Ivan

Sorokin Mikhail

J4132c

Saint-Petersburg

2020

**Table of contents:**

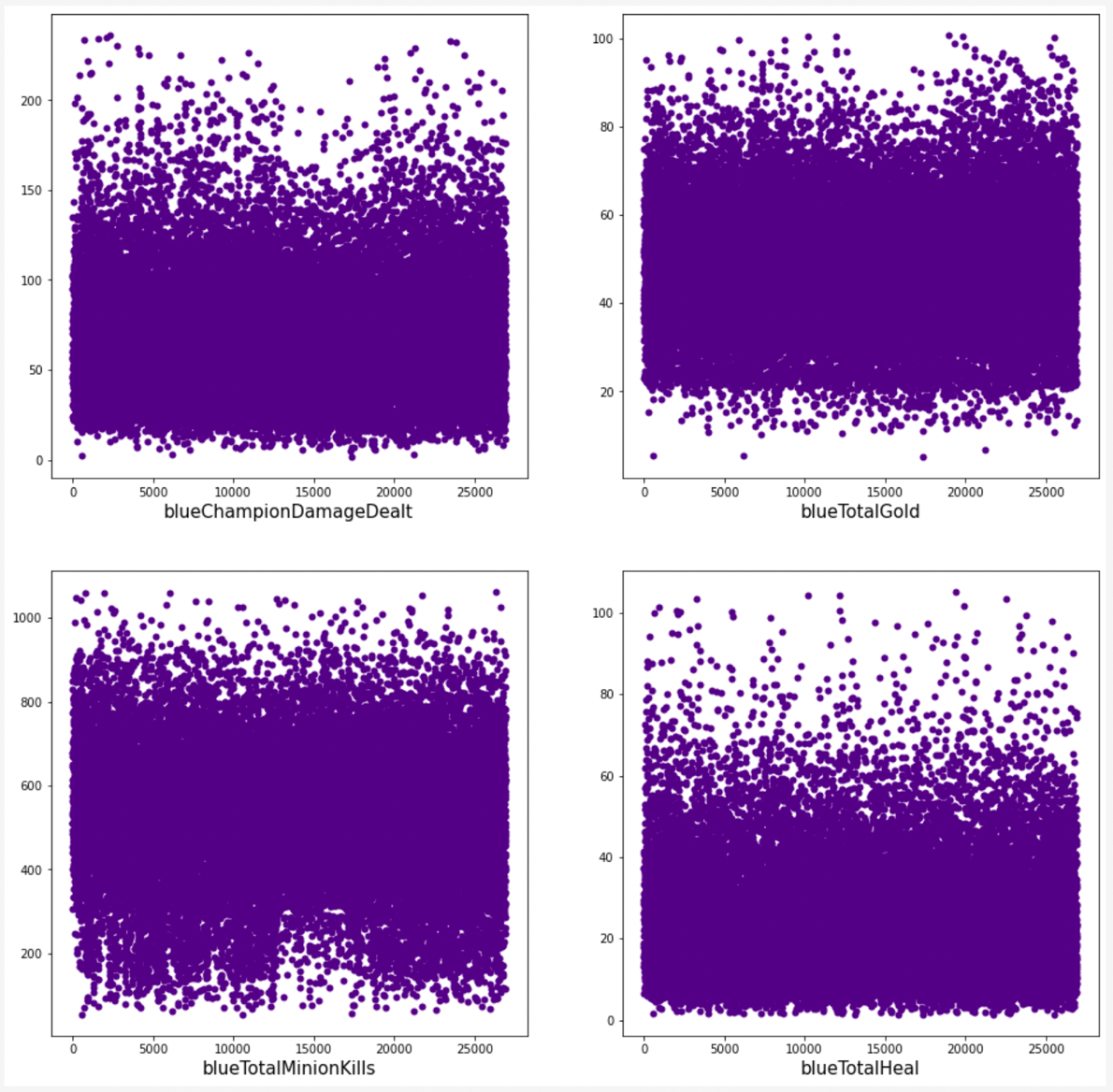
**1. Substantiation of chosen subsample:**

Изображение выглядит как стол

Автоматически созданное описание

*Pic.1. Short DS pandas presentation.*

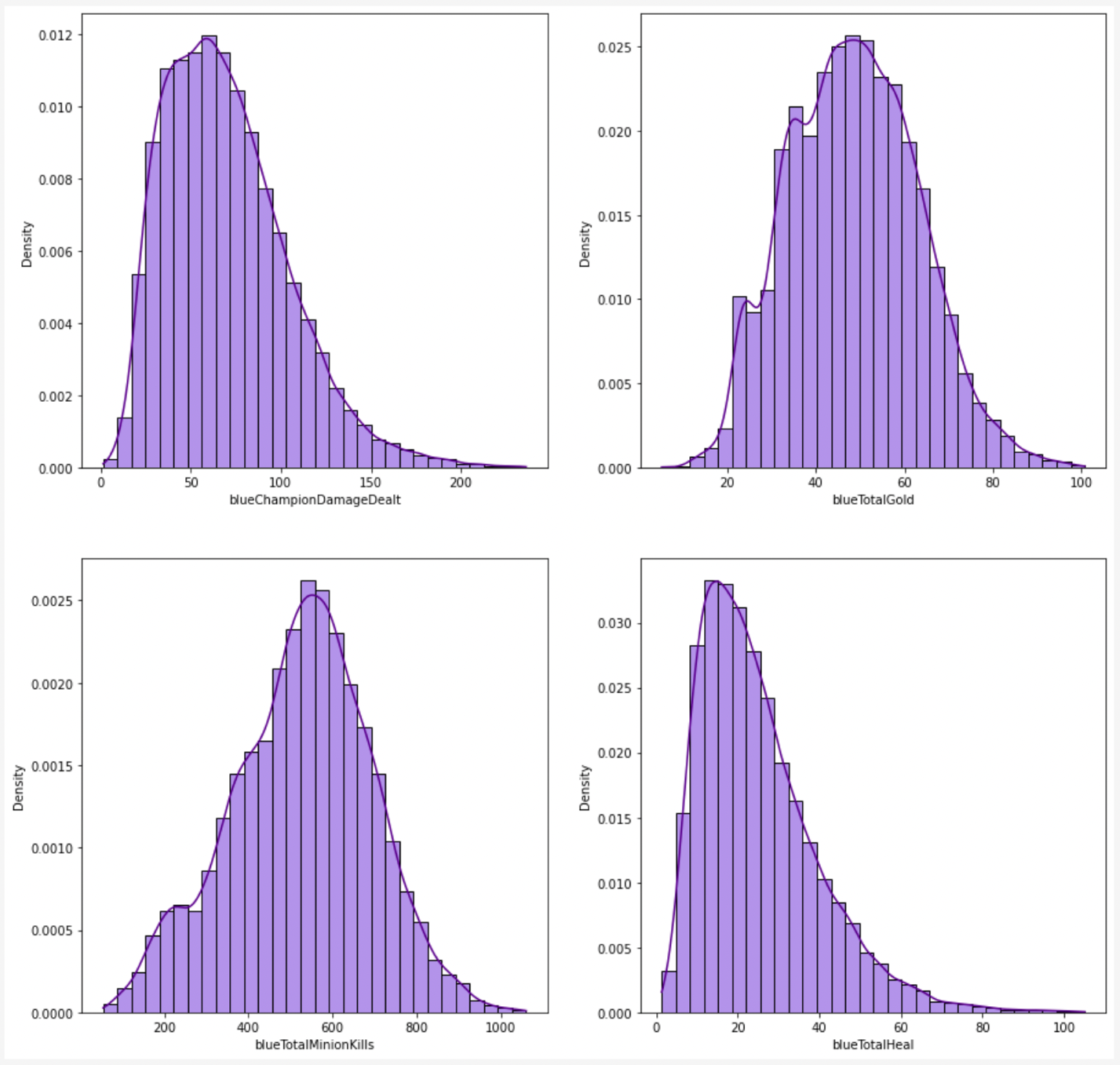
Our dataset is game statistics data from the League of Legends game for 2020 from rated games in the "challenger" rank. The dataset is built using Riot.API (open public API for various in-game parameters from online games from Riot Games). The dataset contains many statistical parameters of past matches, including damage done, in-game currency earned, data on victories and defeats, etc. More details can be found in the README.MD file in the datasets folder.



*Pic.2. Display Lab 1 used data.*

This Lab four of variables have been used. In our laboratory work, we analyze only the statistics of the "blue team".

**2. Plotting a non-parametric estimation of PDF in form of a histogram and using kernel density function (all chosen variables are continuous in our case)**



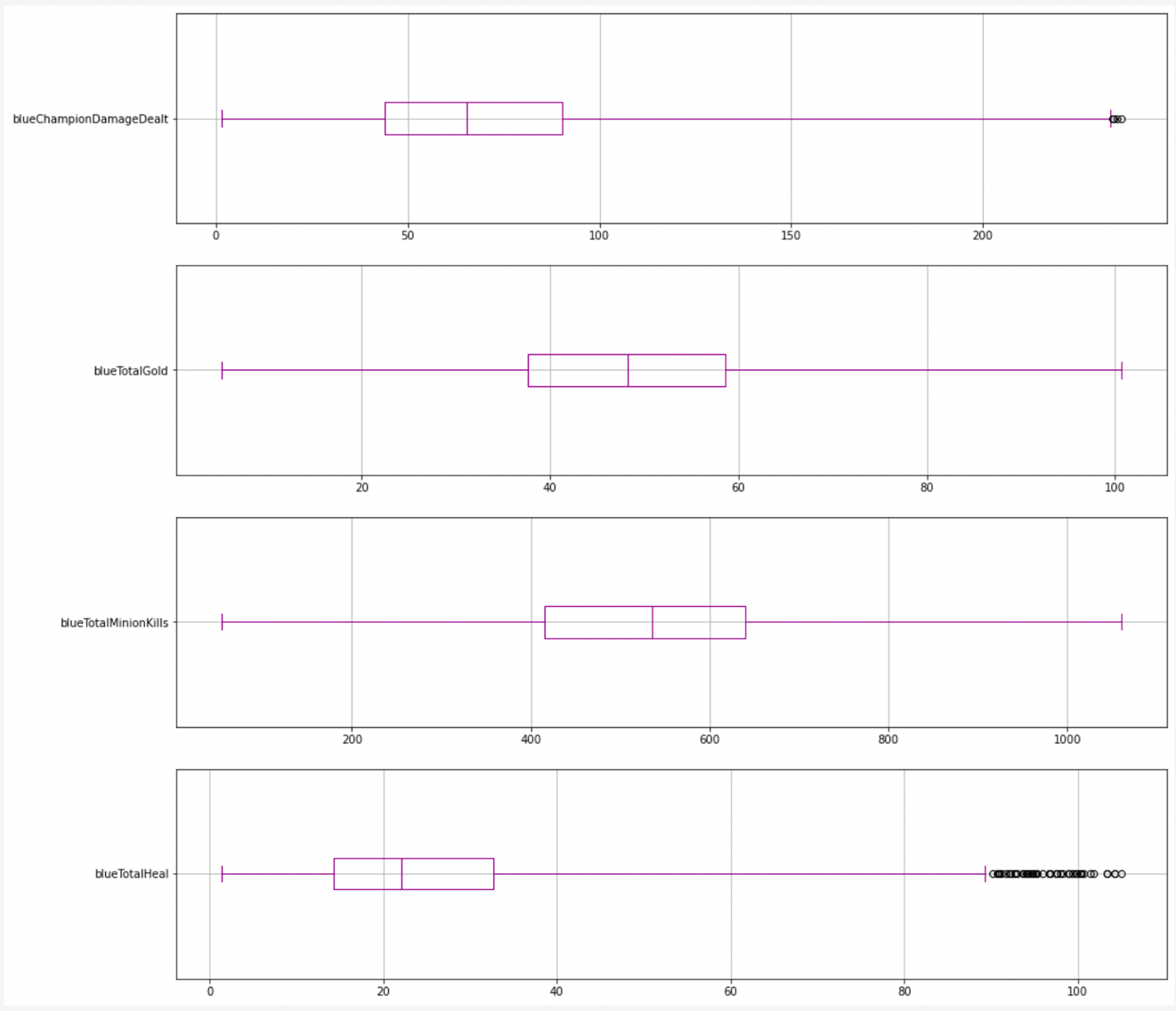
*Pic.3. Histogram and KDE.*

**3. Order statistics estimation and its representation as “box with whiskers” plot.**

Изображение выглядит как стол

Автоматически созданное описание

*Pic.4. Statistics estimation.*



*Pic.5. Variables’ boxplot visualization.*

*As one can see boxplot for blueTotalHeal shows noticablenumber of outliers. Nethertheless, this number is actually about 1/1000 of all values and does not have bad influence on distribution parameters estimation.*

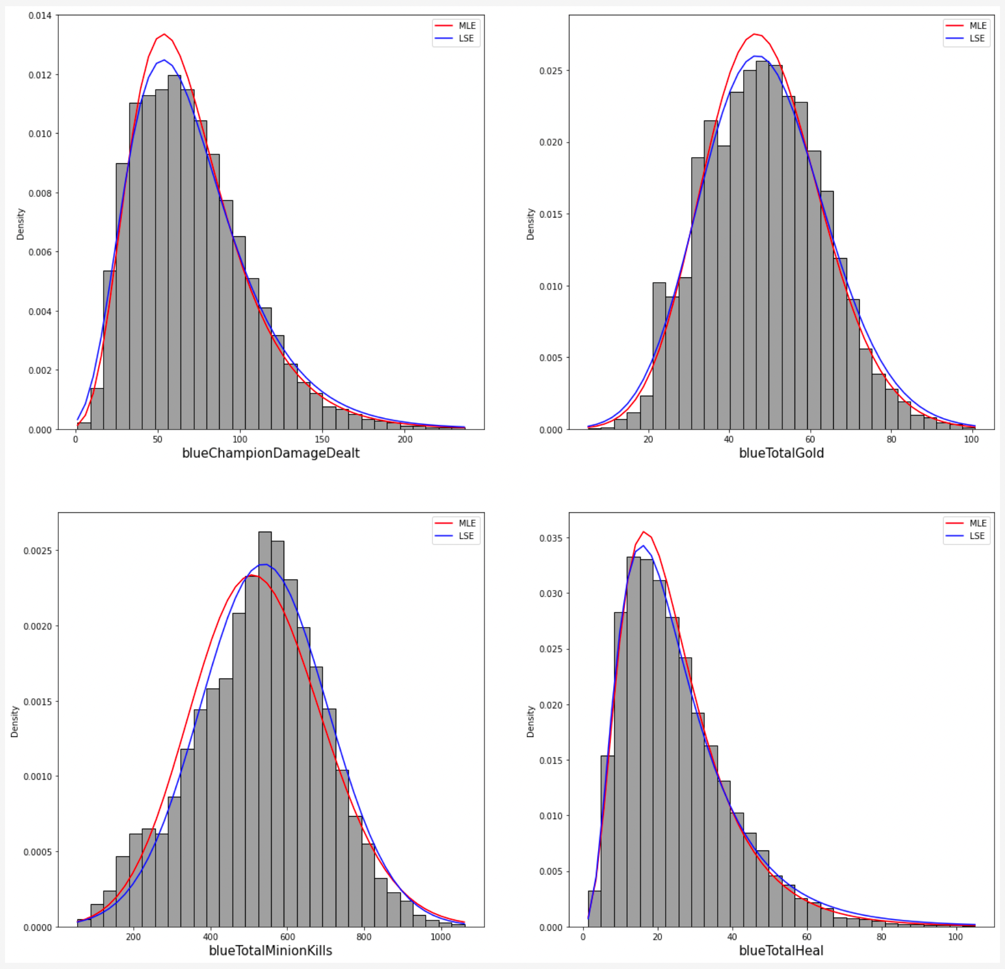
**4. Selection of theoretical distributions that best reflect empirical data and 5. Estimation of random variable distribution parameters using maximum likelihood technique and least squares methods:**

Изображение выглядит как стол

Автоматически созданное описание

*Pic.6. Parametric representation of MLE and LSE.*

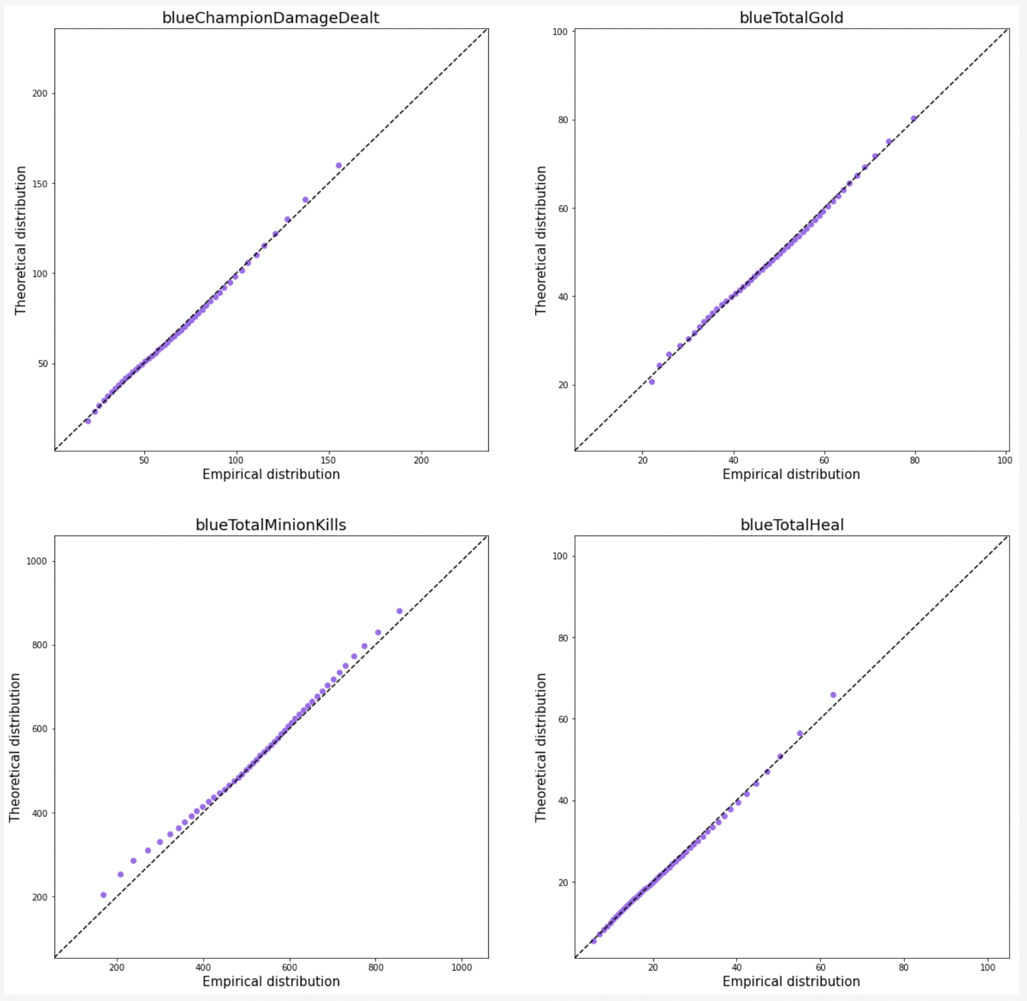
This table shows the distribution parameters calculated using the MLE and LSE algorithms. For MLE of parameters scipy.stats distribution fitting functions were used. For least squares method minimize function from scipy.optimize was used. X and Y data for optimization problem were got from histograms. As one can see on the picture below for *blueChampionDamageDealt* and *blueTotalHeal* variables **lognormal** distributionwas chosen. For *blueTotalGold* and *blueTotalMinionKills* **chi-squared** one showed the best results*.*

On the picture blue line represents results obtained by Least Squares method. PDF with parameters, estimated by MLE is represented by red line.

*Pic.7. Hist MLE and LSE visualization.*

**6. Validation of empirical and theoretical distributions using quantile biplots.**

On this plot one can see, that tails of all QQ plots are quite good. At the same time *for blueTotalMinionKills* noticeable deviations for the lower percentiles can be found. This is due to rather sharp form of the appropriate histogram.



*Pic.8. QQ biplots.*

**7. Statistical tests (2 at least).**

For statistical test we used Kstest and CramerVonMises functions. For these both MLE and LSE estimated parameters are used. The gained results are shown in the table below.

Actually the obtained results do not favor the acceptance of the hypothesis that theoretical distribution correctly fit our data. Only for *blueTotalHeal the hypothesis* can be accepted with good significance level.

Изображение выглядит как стол

Автоматически созданное описание

*Pic.9. Statistical tests.*

**Source code:**

* The full repository with all the labs: <https://github.com/vandosik/M-M-MSA>
* The repo with Datasets and additional used Data info: <https://github.com/vandosik/M-M-MSA/tree/master/Datasets>
* The Lab 1 ipynb file: <https://github.com/vandosik/M-M-MSA/blob/master/Lab_1/lab_1.ipynb>

We recommend to use the first link because our GitHub project has README file with similar links and instructions which is really easy to use.

Изображение выглядит как текст

Автоматически созданное описание