Report on learning practice # 2

Analysis of multivariate random variables

Performed by:

Grandilevskii Aleksei

J4133c

Dubinin Ivan

Sorokin Mikhail

J4132c

Saint-Petersburg

2021

**Table of contents:**

Dataset preparation:

Изображение выглядит как стол

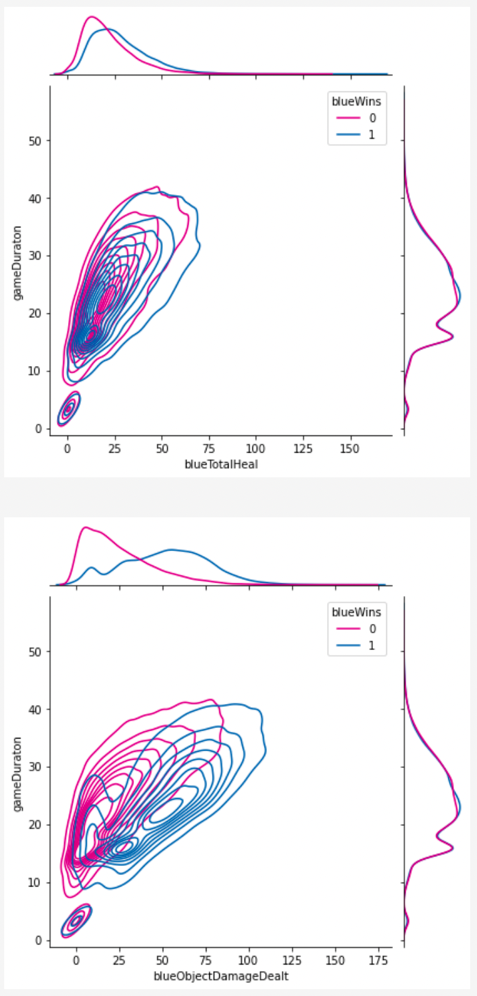
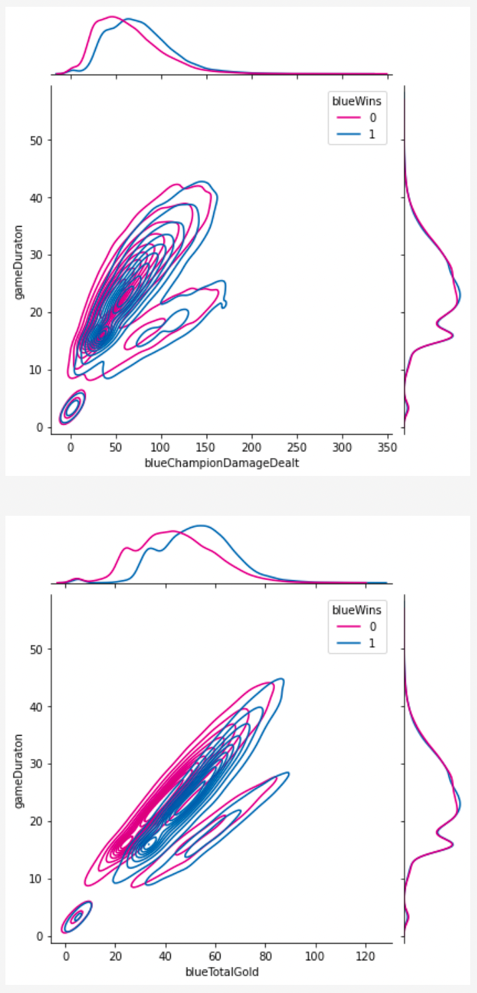
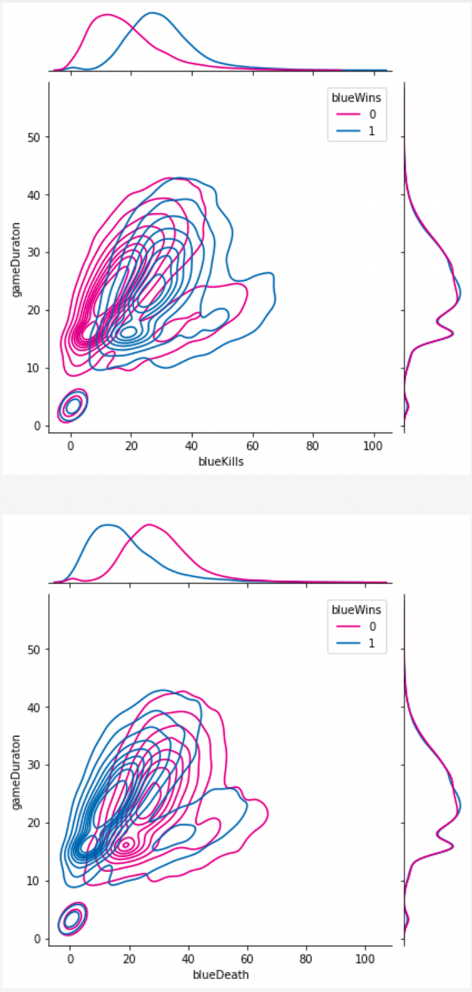
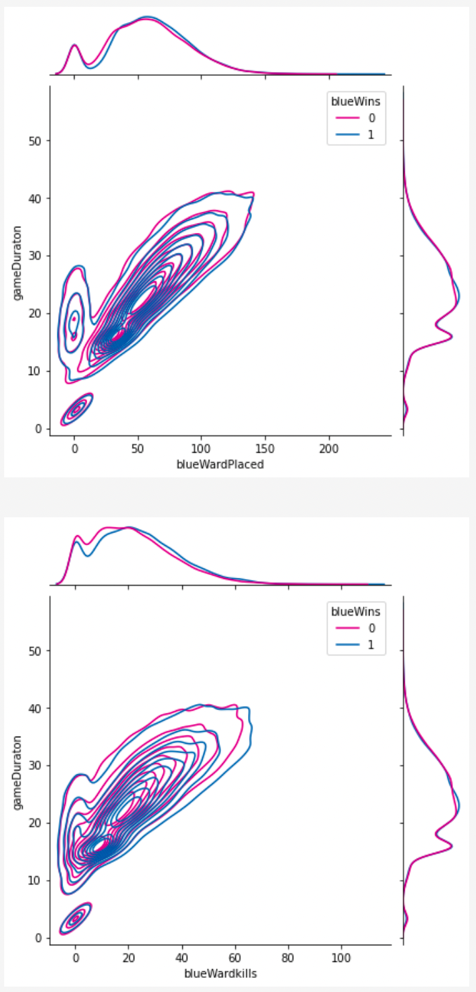
Автоматически созданное описание

*Pic.1. Dataset preparation.*

In this lab, 12 dimensions will be considered, 1 target for predictive analysis, one dimension for categorization, and 10 predictor values.

**From Lab 1:** *“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.”*

1. Plotting a non-parametric estimation of PDF in form of a histogram and kernel density function for MRV (or probability law in case of discrete MRV).



*Pic.2. KDF plots.*

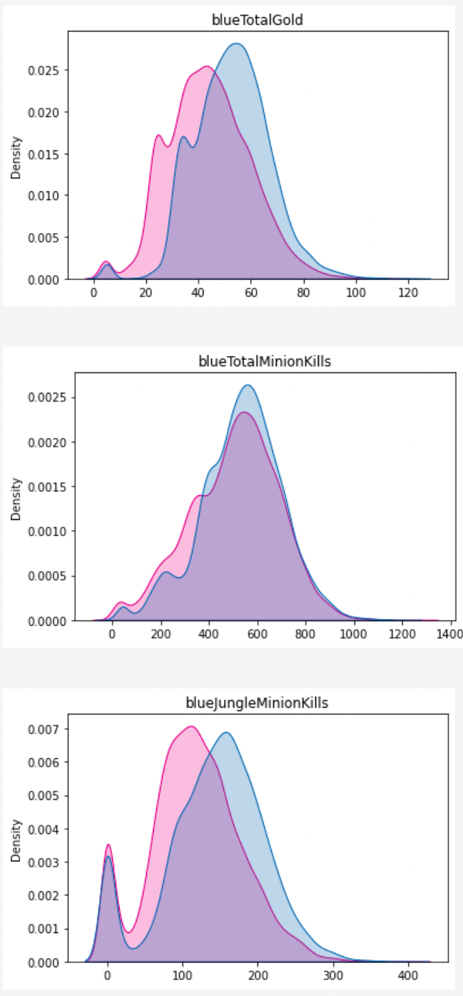
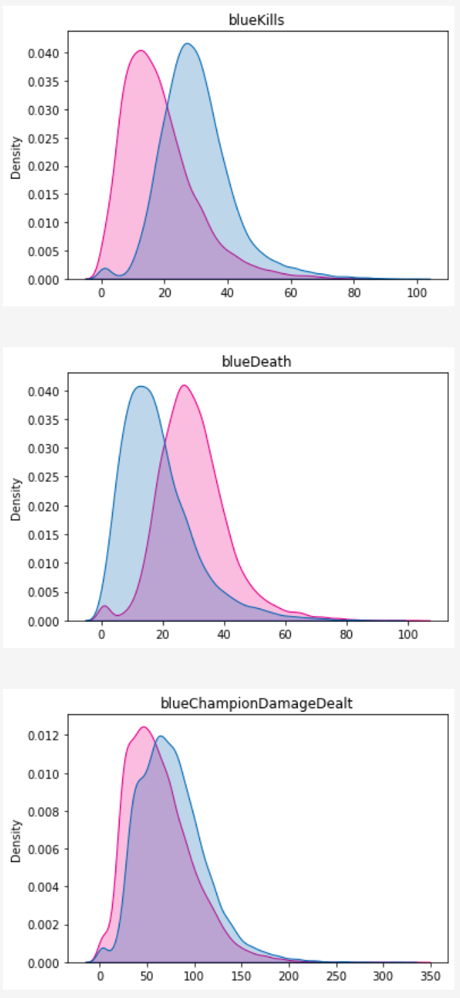
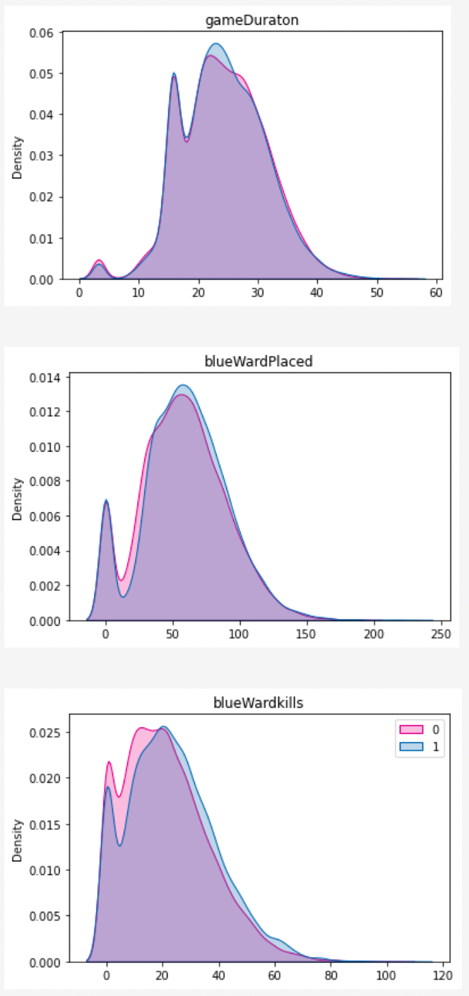
2. Estimation of multivariate mathematical expectation and variance.

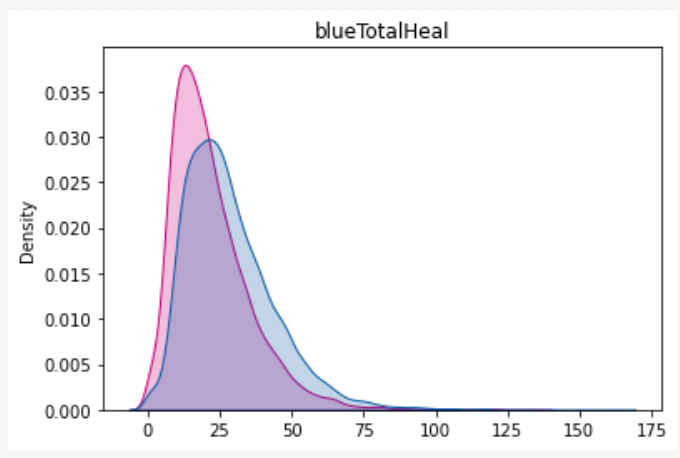
Изображение выглядит как стол

Автоматически созданное описание

*Pic.3. Multivariate m.e. and var*

3. Non-parametric estimation of conditional distributions, mathematical expectations and variances.





*Pic.4. NPE visualization.*

*Изображение выглядит как стол

Автоматически созданное описание*

*Pic.5. NPE results.*

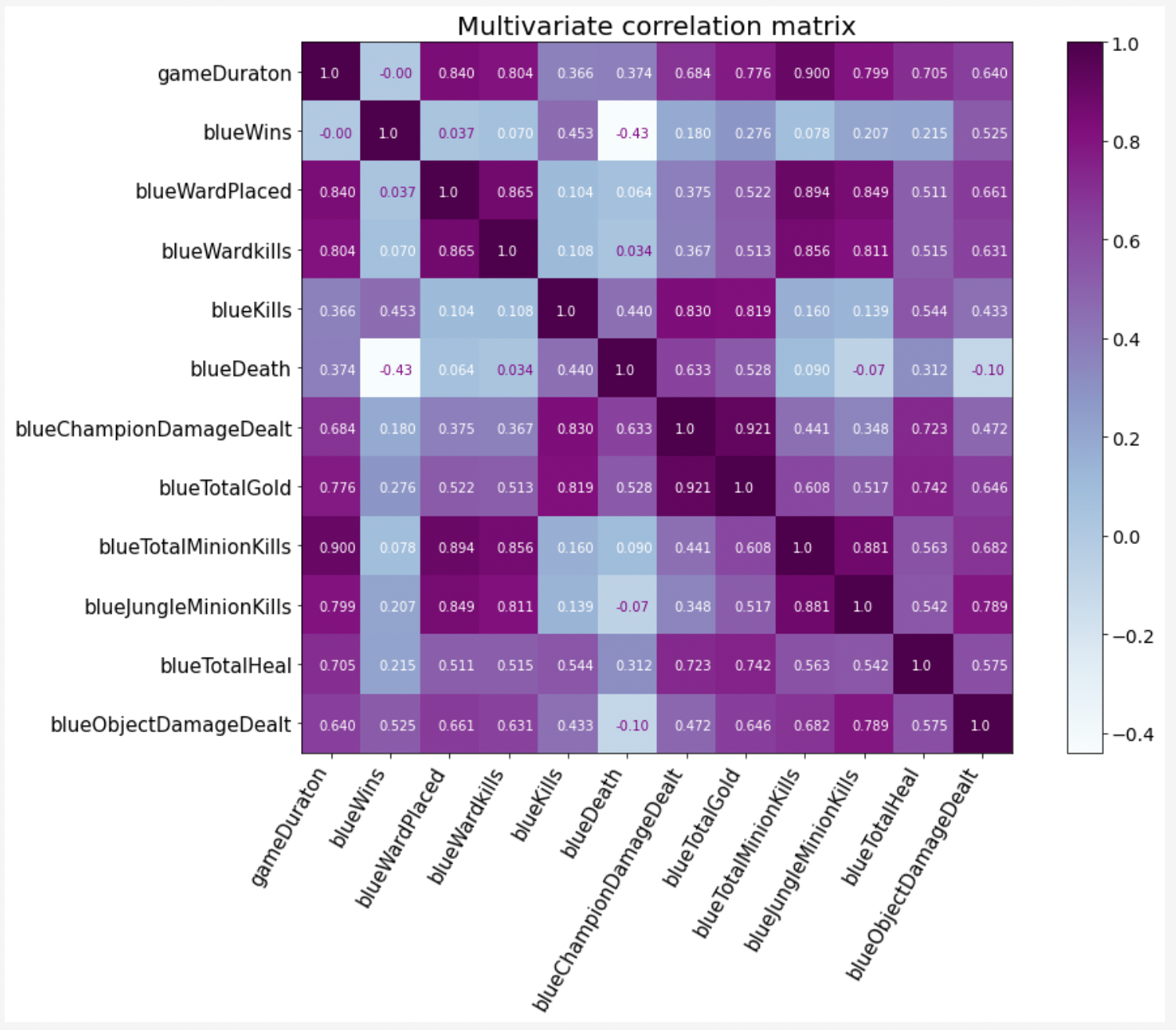
4. Estimation of pair correlation coefficients, confidence intervals for them and significance levels.

Изображение выглядит как стол

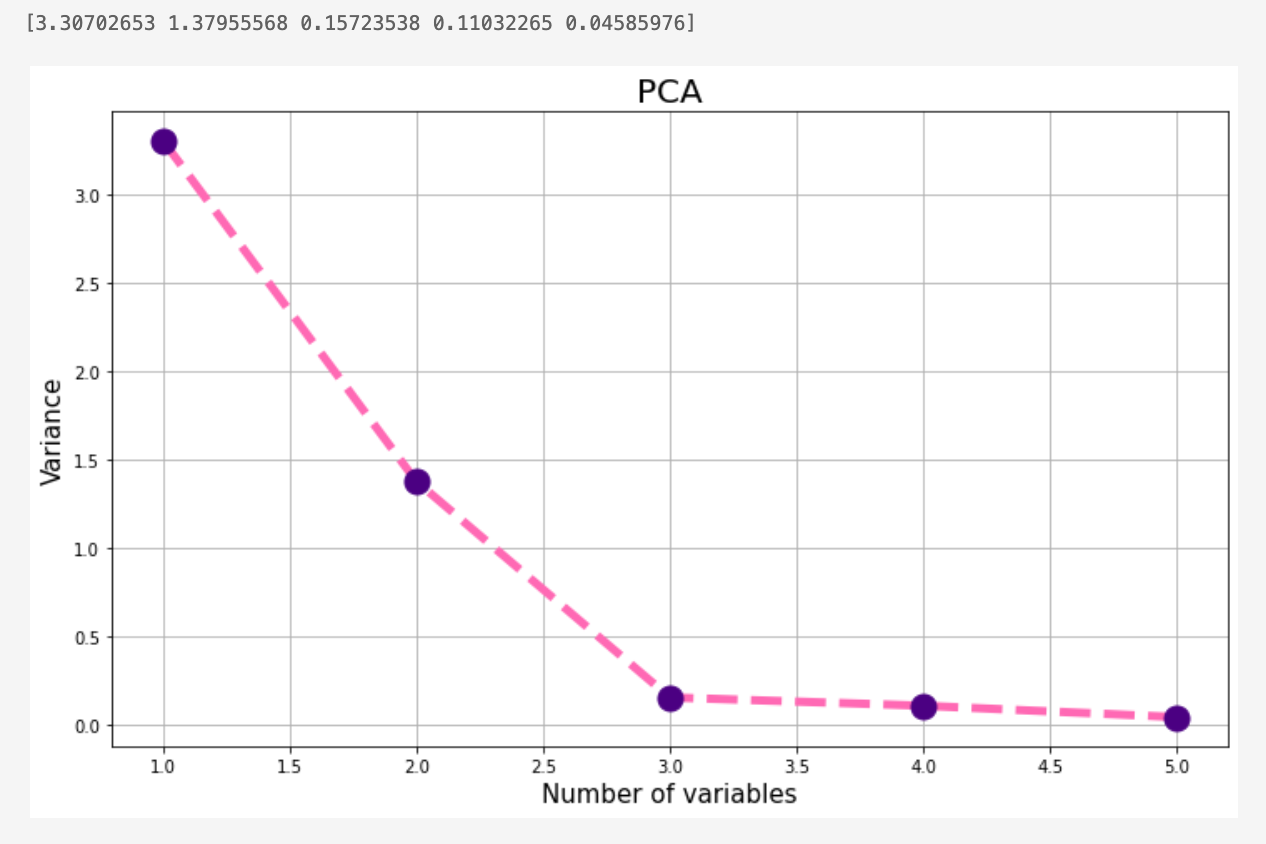
Автоматически созданное описание

*Pic.6. Pair coefficients results.*

5. Task formulation for regression, multivariate correlation.



*Pic.7. MVCM.*

**

*Pic.8. PCA analysis.*

PCA algorithm was used in order to reduce feature dimensionality. When the number of components goes from 1 to 3, the decrease in the variance is significant and more variables are not descriptive.

So, the number of chosen variables for the regression problem should be 3. (Further analysis corrected in accordance with the corrected PCA analysis)

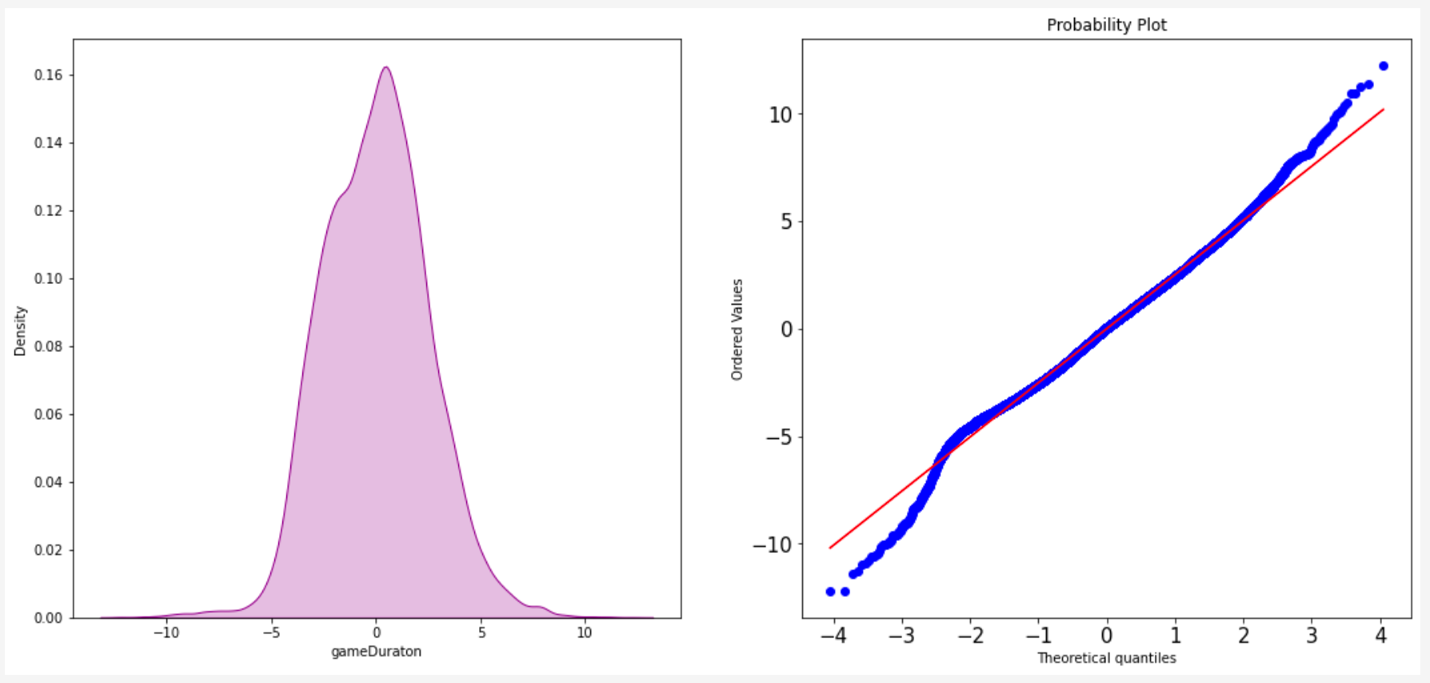
6. Regression model, multicollinearity and regularization (if needed).

Изображение выглядит как стол

Автоматически созданное описание

*Pic.9. LSM, Lasso and Ridge models.*

7. Quality analysis.



*Pic.10. Results visual analysis.*

*Изображение выглядит как стол

Автоматически созданное описание*

*Pic.11. Mathematical results.*

**Sourcecode:**

* 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_2/lab_2.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