Report on learning practice # 4

Stationarity of the processes

Performed by:

Performed by:

Grandilevskii Aleksei

J4133c

Dubinin Ivan

Sorokin Mikhail

J4132c

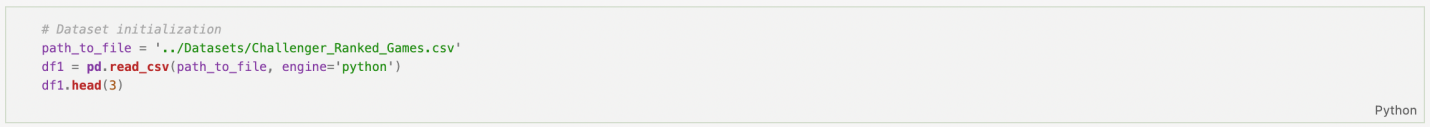
Saint-Petersburg

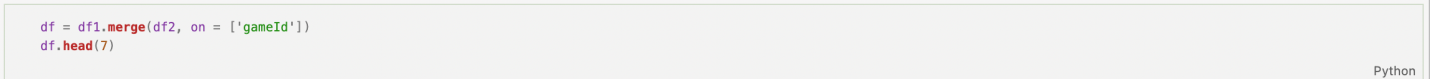
2021

**Table of contents:**

1. Substantiation of chosen sampling.

This Lab we used to have timestamps in our dataset. Dataset we used in the previous works hasn’t them. But the specialized dataset from labs 1-3 is a processed squeeze from the raw data obtained using the Riot.API. (<https://www.kaggle.com/gyejr95/league-of-legendslol-ranked-games-2020-ver1#challenger_match_V2.csv>). The original dataset contains 7Gb of practically raw data from the API. We wrote our own script for parsing timestamps from raw data and already processed dataset.

Изображение выглядит как текст

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание

*Pic.1. Timestamps parsing.*

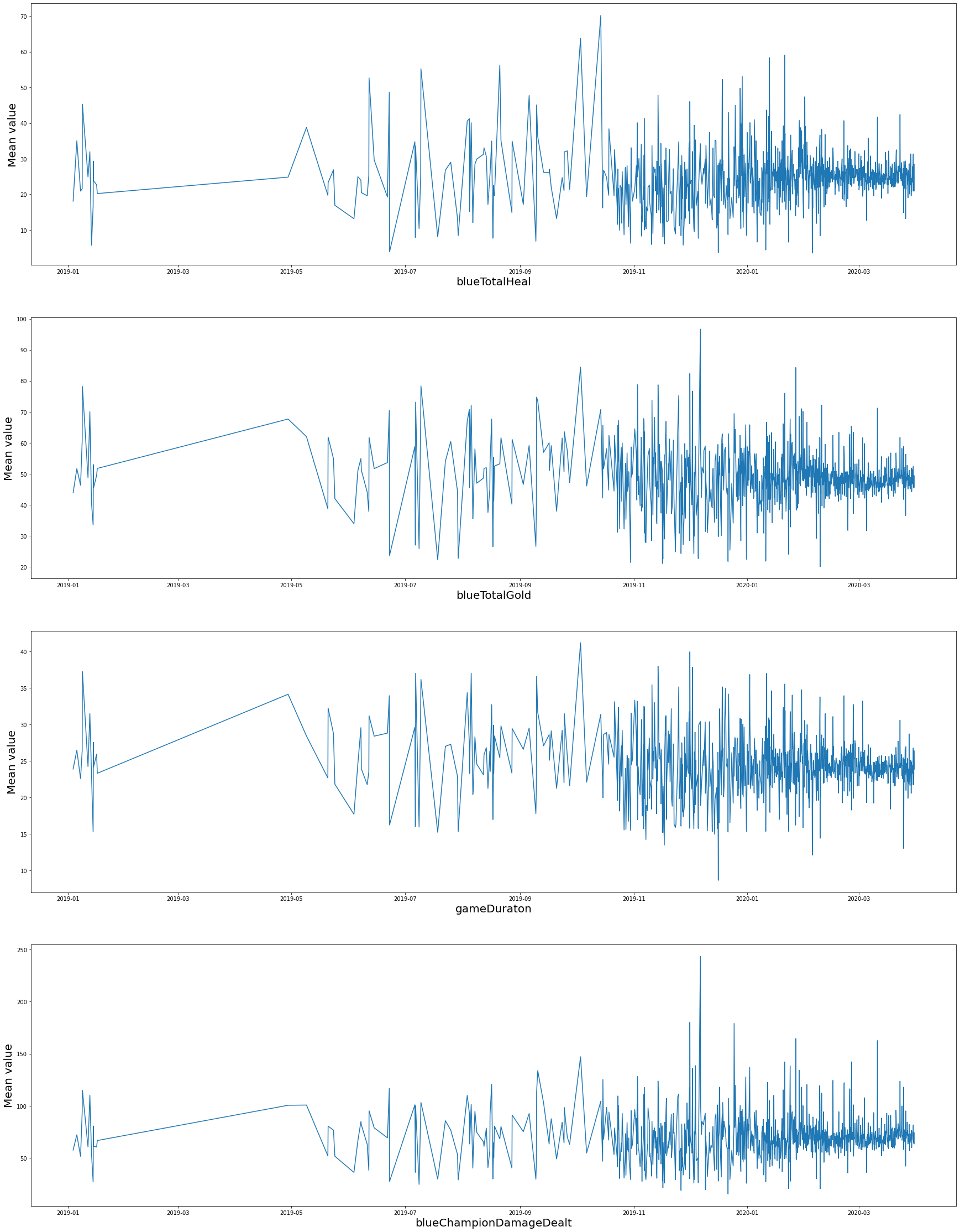
As You can see in the code timestamps are timestamps are represented as Unix-ts, so our script also converts them into default Python 3 timestamps.

Изображение выглядит как стол

Автоматически созданное описание

*Pic.2. Working dataset for lab 4 visualization.*

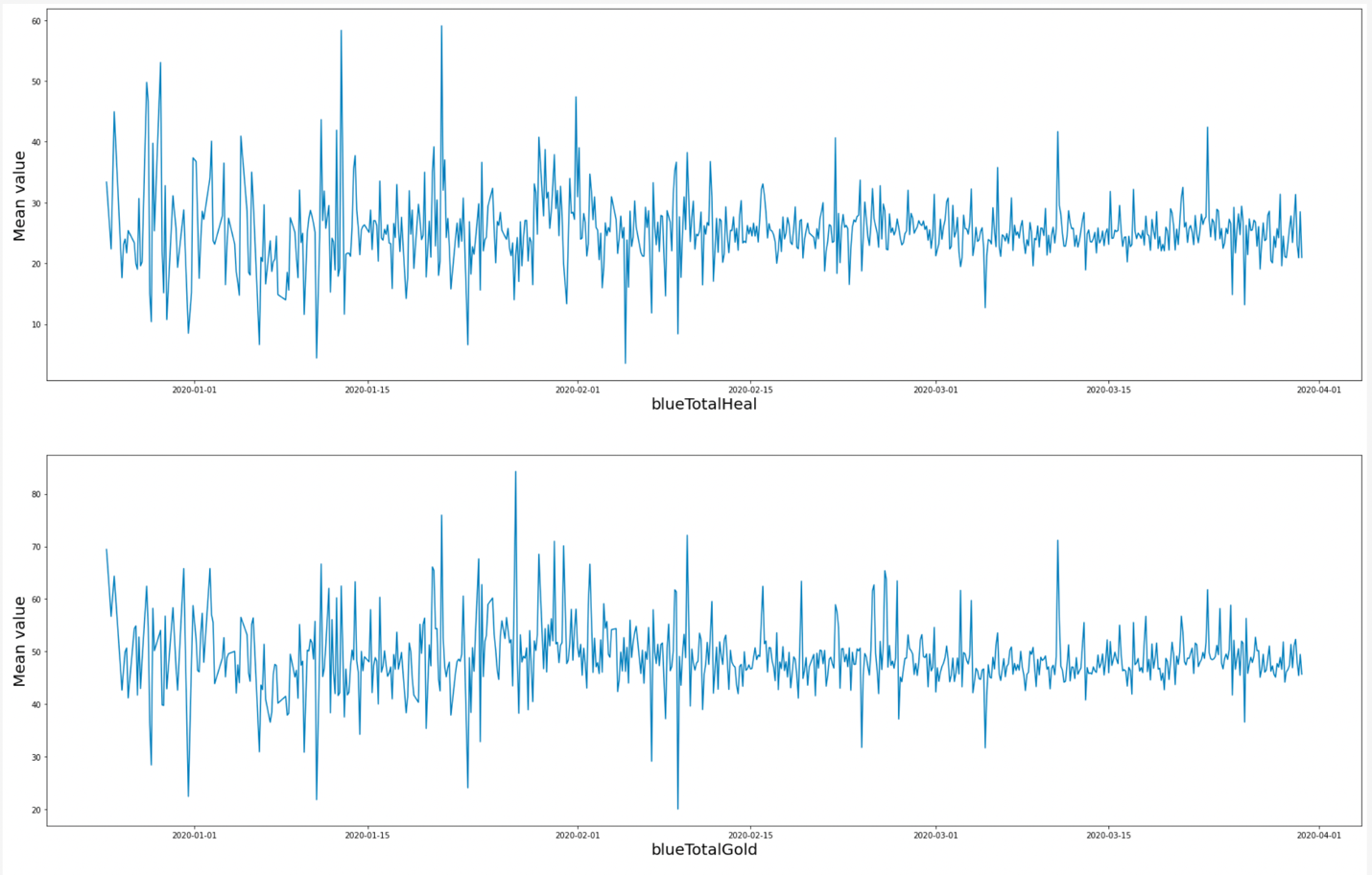
For this task were chosen these target variables: *blueTotalHeal, blueTotalGold* and following predictors: *gameDuraton, blueChampionDamageDealt. timeStamp* variable contains timestamps for our data. Time series for all chosen variables aggregated by 3 hours with mean value are shown on the picture below.



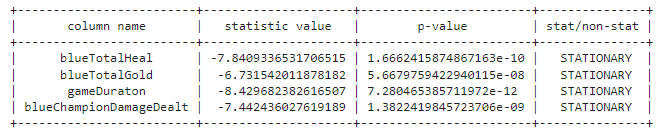
Pic.2 – Initial time series for all variables

2. Stationary analysis.

As one can easily see gained time series represent non-stationary processes. To make them more stationary space part till 2020-01-01 was removed. The gained time series are presented on picture 3. All this time series passed Augmented Dickey-Fuller Test and can be considered stationary. The Results of Dickey-Fuller Test are presented for each variable on the table below.



*Pic.3. Gained stationary process.*

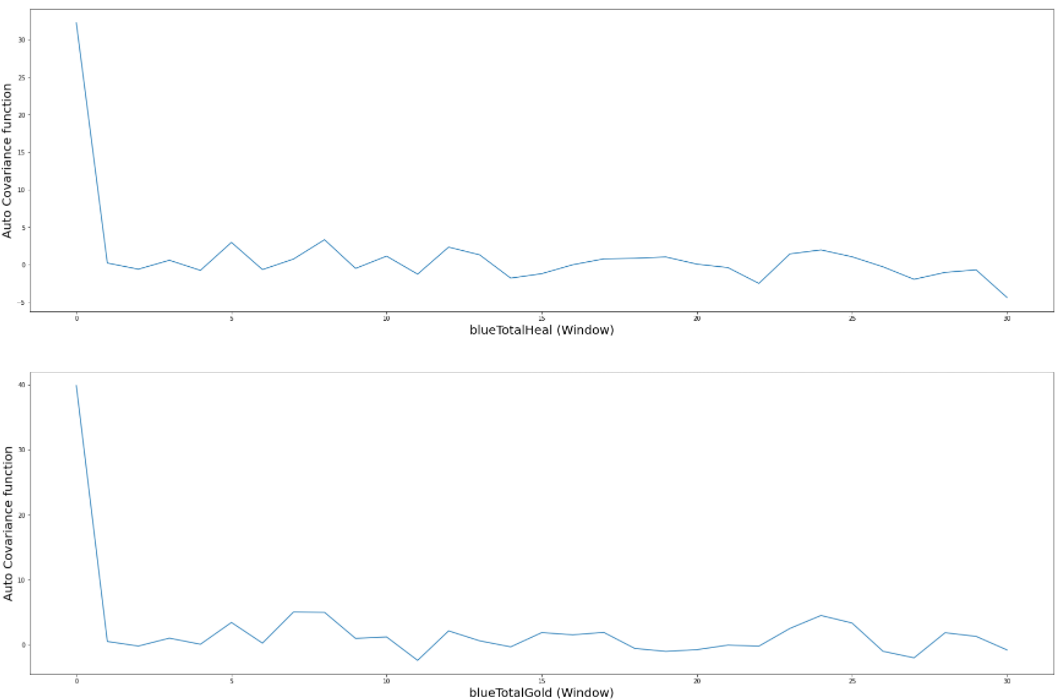


*Pic.4. ACF*

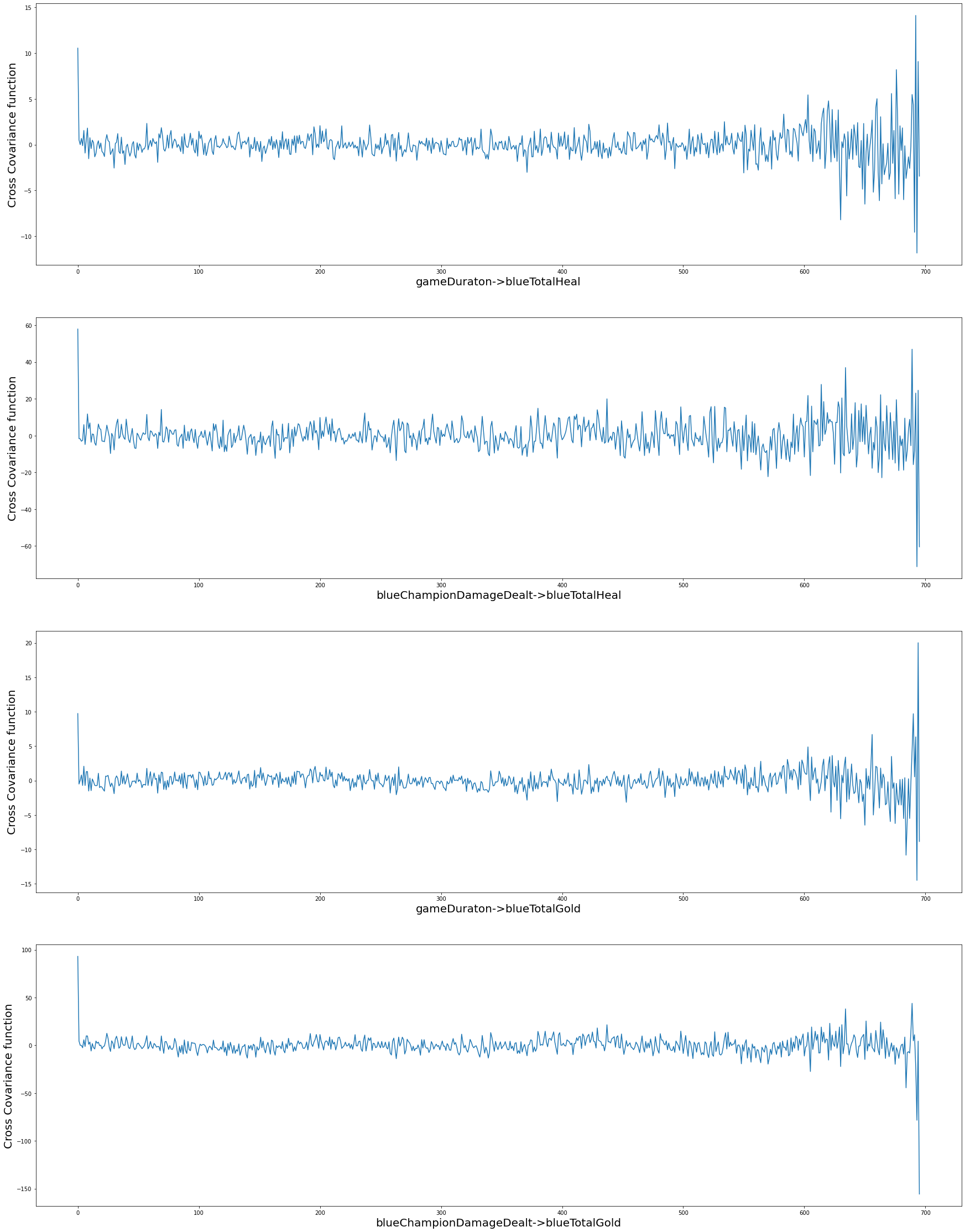
3. Covariance or correlation function analysis.

Picture 4 represents Auto-covariance function for target variables. On picture 5 Auto-covariance function gained using window of 30 items is shown. These functions look like their processes are stationary.

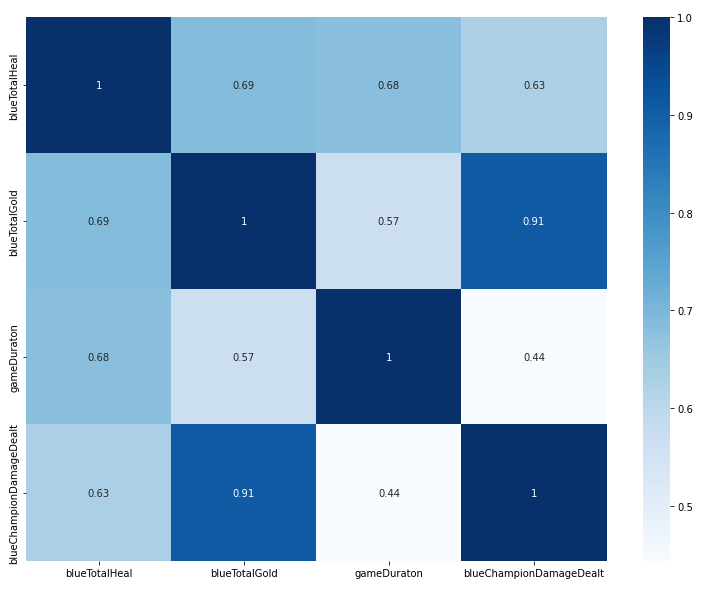
Изображение выглядит как текст

Автоматически созданное описание

Picture 6 represents cross-covariance (mutual correlation) function between each target and predictor variable

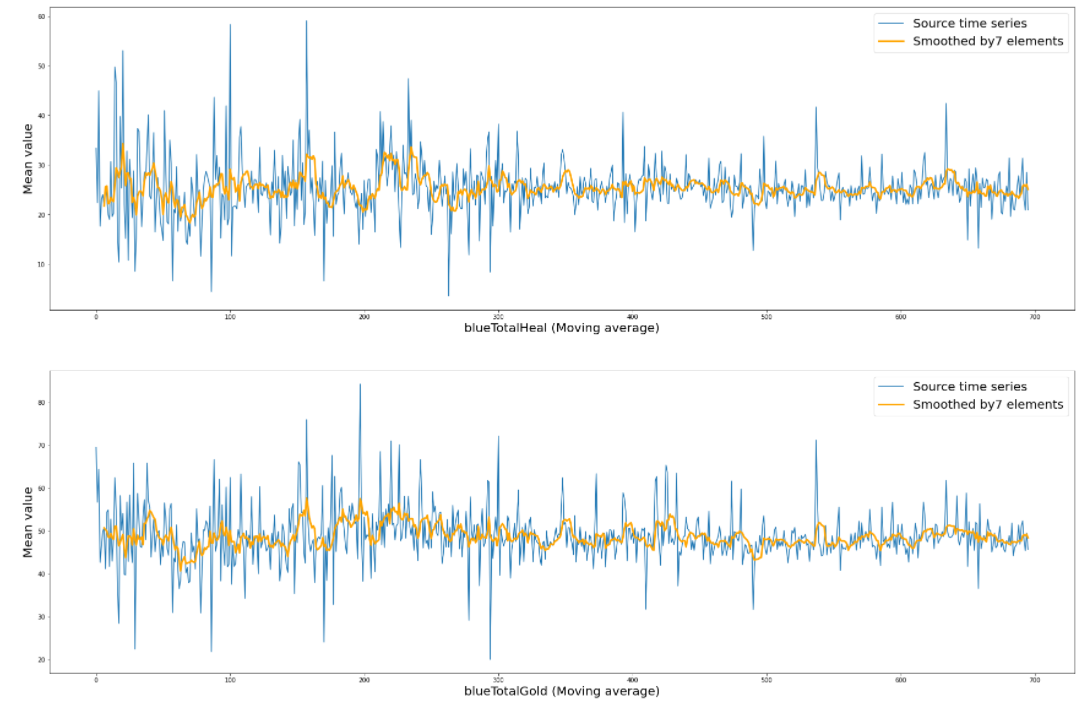


*Pic.6. Cross-covariance functions.*

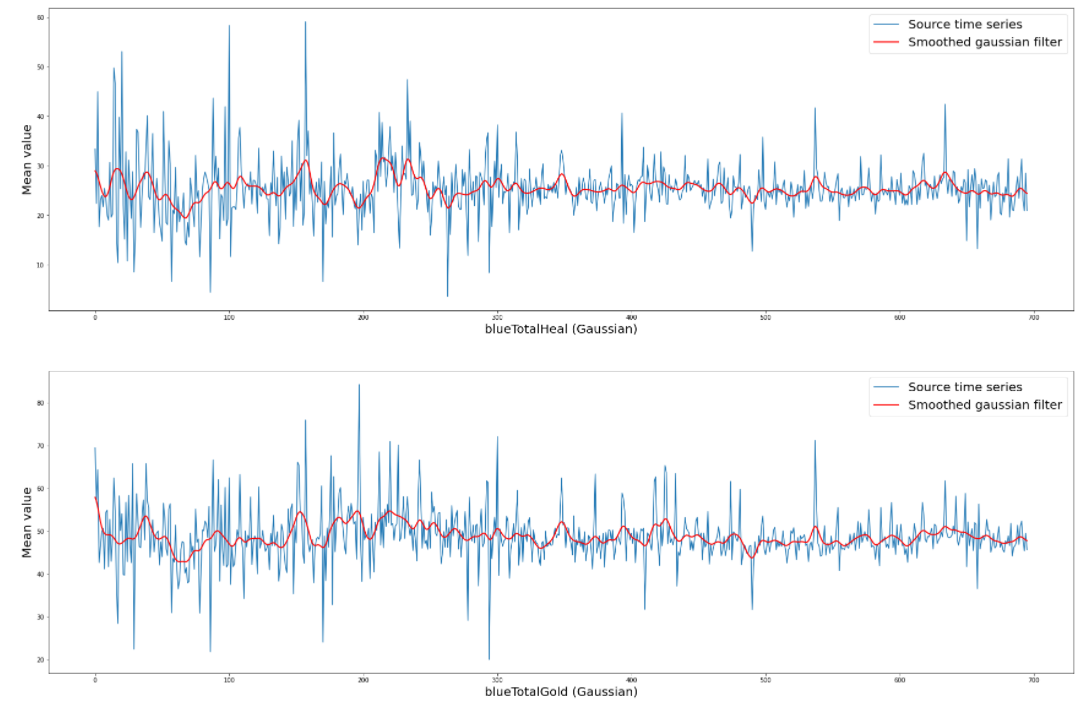
Cross correlation matrix for variables is presented on picture 7. As one can see, target and predictor variables are highly correlated, especially BlueTotalGold is correlated with blueChampionDamageDealt.

*Pic.7. Cross correlation matrix.*

4. Noise filtration.

For filtering high frequencies Moving average filter (Pic.8) and Gaussian filter (Pic.9) from FEDOT framework were used.

*Pic.8. Using Moving average filter.*

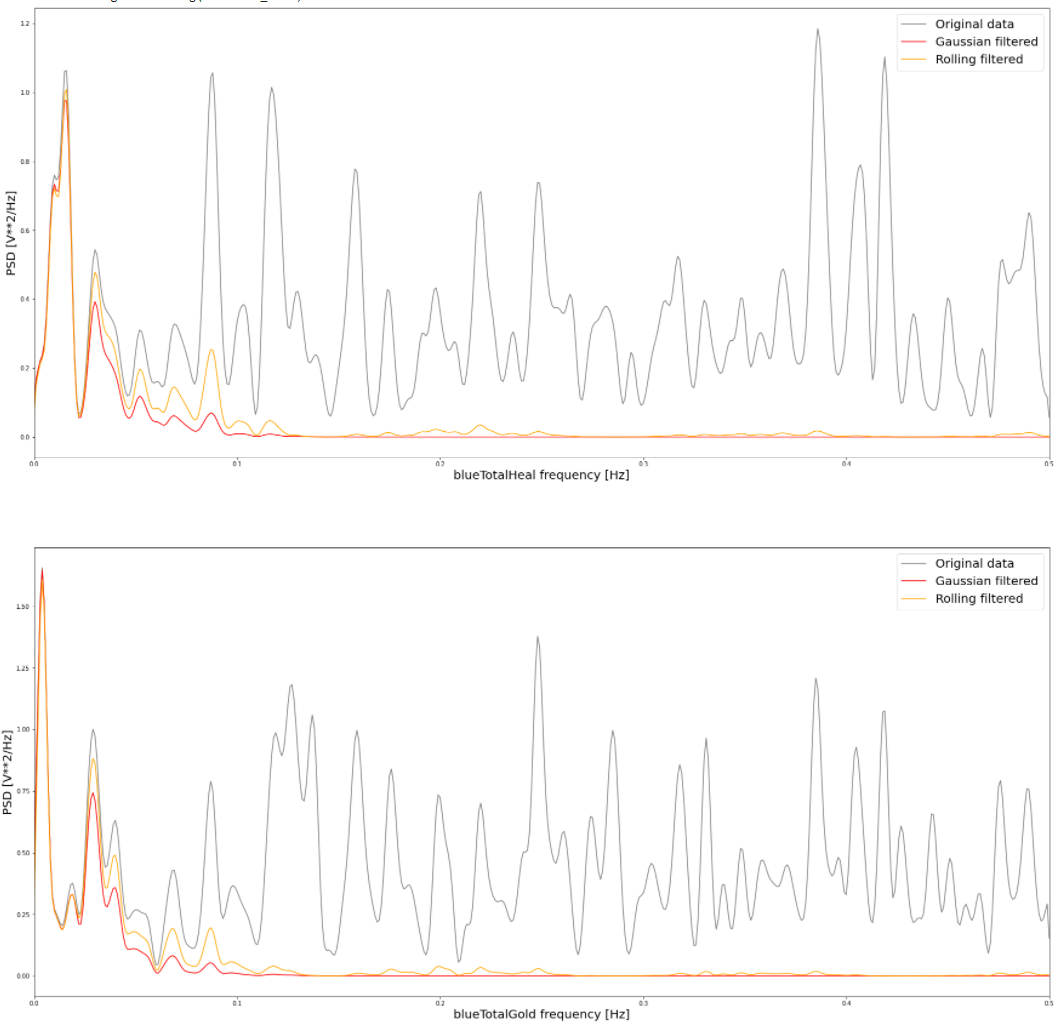
*Pic.9. Using Gaussian filter.*

5. Estimation of spectral density function.

Spectral density using was estimated Welch’s method from scipy.signal module. The results for both non-filtered and filtered time series are presented on picture 10. There are two plots for each target variable. Two different filters are shown using different colors: orange for Moving average and red for Gaussian filter.

As one can see both filters removed high frequencies from spectral density function.

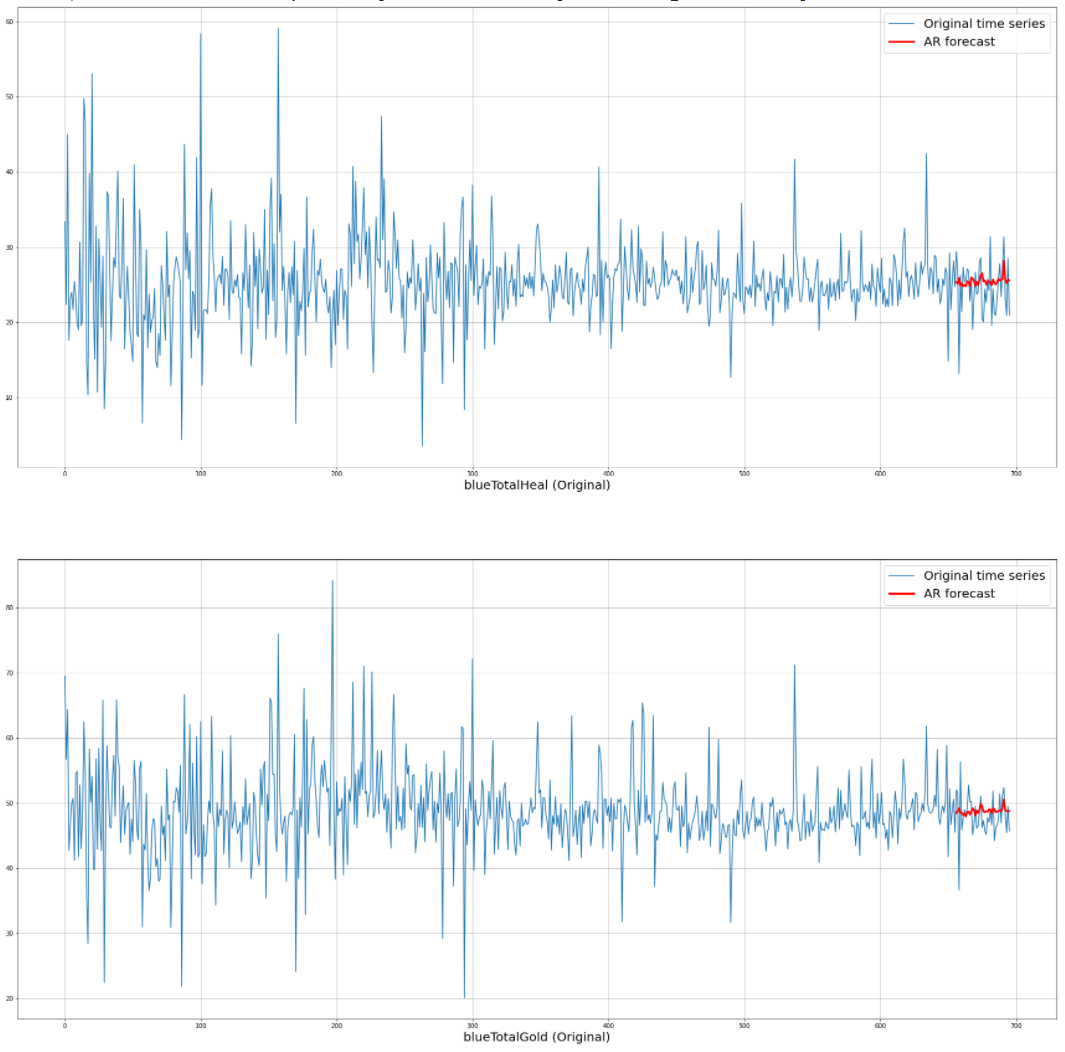
*Pic.10. Spectral density functions.*



6. Auto-regression model.

For building Auto-regression model AR Pipeline from FEDOT framework was used. The generated auto-regression model was used to forecast the future values of time series. All in all there were built four AR models – for both filtered and non-filtered time series of both target variables. The result of forecasting for non-filtered time series are shown on picture 11, for filtered data – on picture 12.

As one can see the generated forecasts are not accurate.



*Pic.11. AR model forecasting non-filtered time series.*

Изображение выглядит как текст, доска

Автоматически созданное описание

*Pic.12. AR model forecasting filtered time series.*

**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 4 ipynb file: <https://github.com/vandosik/M-M-MSA/blob/master/Lab_4/lab_4.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