

Рубежный контроль №1

Писарчук Надежда ИУ5-22М

Вариант - 7

```
In [ ]: # This Python 3 environment comes with many helpful analytics libraries insta
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will li

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that g
# You can also write temporary files to /kaggle/temp/, but they won't be save

pd.set_option('max_colwidth', 800)
pd.set_option('display.max_columns', None)
```

```
In [ ]: data = pd.read_csv(
        '/kaggle/input/kepler-exoplanet-search-results/cumulative.csv',
        sep=",")
```

```
In [ ]: data.head()
```

Задача №7.

Для набора данных проведите устранение пропусков для одного (произвольного) числового признака с использованием метода заполнения медианой.

Я буду заполнять признак `koi_score`, принимающий значение от 0 до 1, имеющий 16% пропусков

```
In [ ]: data.isna().sum()['koi_score']
```

```
In [ ]: data.koi_score.median()
```

```
In [ ]: data['koi_score_nonnan'] = data['koi_score'].fillna(
        data['koi_score'].median())
data.koi_score_nonnan.isna().sum()
```

Доп задание

Для произвольной колонки данных построить гистограмму

Для визуализации заполнения пропусков построю гистограмму для koi_score и koi_score_nonnan

```
In [ ]: data[['koi_score_nonnan', 'koi_score']].hist()  
data[['koi_score_nonnan', 'koi_score']].plot.hist()
```

Задача №27.

Для набора данных для одного (произвольного) числового признака проведите обнаружение и замену (найденными верхними и нижними границами) выбросов на основе 5% и 95% квантилей. Обнаружение и замена выбросов будет производиться для признака koi_steff

```
In [ ]: import matplotlib.pyplot as plt  
import seaborn as sns  
  
sns.displot(data, x="koi_steff")
```

```
In [ ]: q5 = data.koi_steff.quantile(0.05)  
q95 = data.koi_steff.quantile(0.95)  
print(q5, q95)  
  
data['koi_steff_comp'] = np.where(data.koi_steff < q5, q5,  
                                np.where(data.koi_steff > q95, q95, data.koi_steff))
```

```
In [ ]: sns.displot(data, x="koi_steff_comp")
```