Test Task for Tutors in Data Science

Hi,

we are very happy to have you among the candidates for the position of a tutor!

In order to be objective and to make sure that the scope of your knowledge and skills corresponds to the requirements of this job, we invite you to take a test.

This is a test task for data science specialists with experience working with Python. Some tasks are easy, while others are a bit more difficult. If they seem a little too simple for an IT company like Yandex, don't go looking for the catch. The main objective of our course is to prepare beginners for work; keep this in mind. After you have completed the test task, we'll get in touch with you to discuss the results of this stage of the selection process.

Task 1. Working with data

movie_metadata.csv 1566.1KB

To complete this task, use the data set in the attached file. Indicate the answer to each of the following steps and time to complete the entire task.

1.1. Download the data set `movie_metadata.csv`, which contains data about films from IMDb (Internet Movie Database).

1.2. The `duration` column contains data on the film length. How many missing values are there in this column?

1.3. Replace the missing values in the `duration` column with the median value for this column.

1.4. What is the average film length? Give the answer as a floating-point figure rounded to two decimal places.

1.5. How many films between 90 minutes and two hours long were released in 2008?

1.6. The `budget` column contains the film's budget. What is the median budget for all the films listed? Give the answer as an integer.

Task 2. Answering student questions

How would you answer the student's question below? Your task is to get your message across in such a way that a beginner can understand your explanation. You can do this any way you want (pictures, GIFs, metaphors, anything) so long as it makes your explanation clear. Indicate how much time you spent completing this task.

> **What is the difference between DataFrame and Series?**

Task 3. Task on probability theory

You are given two random variables X and Y.

E(X) = 0.5, Var(X) = 2

E(Y) = 7, Var(Y) = 3.5

cov (X, Y) = -0.8

Find the variance of the random variable Z = 2X - 3Y

Task 4. Task for Data Science course

Omer trained a linear regression model and tested its performance on a test sample of 500 objects. On 400 of those, the model returned a prediction higher than expected by 0.5, and on the remaining 100, the model returned a prediction lower than expected by 0.7.

What is the MSE for his model?

Limor claims that the linear regression model wasn't trained correctly, and we can do improve it by changing all the answers by a constant value. What will be her MSE?

You can assume that Limor found the smallest error under her constraints.

Return two values - Omer's and Limor's MSE.

Task 5. Please make a short video (under 5 minutes) and tell us more about yourself, your experience in DS/DA and the reasons you're interested in becoming a tutor with Practicum. If you're allowed to, tell us about the most interesting projects you have worked on.

To complete this task please follow the link

https://praktikumyandex.vcv.ru/r/tutor_data_science

We wish you good luck and looking forward to hearing from you soon 😃