

(https://profile.intra.42.fr)

SCALE FOR PROJECT PISCINE PYTHON DATA SCIENCE (/PROJECTS/PISCINE-PYTHON-DATA-SCIENCE) / DAY 06 (/PROJECTS/PISCINE-PYTHON-DATA-SCIENCE-DAY-06)

You should evaluate 1 student in this team



Git repository

`git@vogsphere.kzn.21-school.ru:vogsphere/intra-uuid-f824e`



Introduction

The methodology of School 21 makes sense only if peer-to-peer assessments are done seriously. This document will help you to do it properly.

- Please, stay courteous, polite, respectful and constructive in all communications during this assessment. The bond of trust between community 21 and you depends on it.
- Highlight possible malfunctions of the work done by the person and take the time to discuss and debate it.
- Keep in mind that sometimes there can be differences in interpretation of the tasks and the scope of features. Please, stay open-minded to the vision of the other.

Guidelines

- Evaluate only the files that are on the GIT repository of the student or group.
- Doublecheck that the GIT repository is the one corresponding to the student or the group as long as to the project.
- Meticulously check that nothing malicious has been used to mislead you and have you assess something except the content of the official repository.

- If you have not finished the project yet, it is compulsory to read the entire instruction before starting the review.

- Use the special flags in the scale to report an empty or non-functional solution as long as a case of cheating. In these cases, the assessment is completed and the final grade is 0 (or in a case of cheating is -42). However, except for a case of cheating, you are encouraged to continue reviewing the project to identify the problems that caused the situation in order to avoid them for the next assessment.

- You must stop giving points from the first wrong exercise even if the following exercises are correct.

Attachments

 subject.pdf (<https://cdn.intra.42.fr/pdf/pdf/49870/en.subject.pdf>)

 checking-logs.sqlite.sqlite copy (/uploads/document/document/8789/checking-logs.sqlite.sqlite_copy)

Preliminaries

Respect the rules

- The repository contains the work of the student (or group).
- The student is able to explain their work at any time during the assessment.
- The general rules and any rules specific to the day are respected throughout the assessment.

 Yes

 No

Piscine Python | Data Science D06

Any hardcoded result is worth zero for the exercise.

Exercise 00 – Select

- Run all the cells in the notebook, they should work without errors
 - Run `pageviews.info()`, it should include:
DatetimeIndex: 987 entries, 2020-04-26 21:53:59.624136 to 2020-05-21 16:36:40.915488
Data columns (total 1 columns):
Column Non-Null Count Dtype

0 uid 987 non-null object
dtypes: object(1)
- The connection is closed?
- In all other cases, the test is failed.

✓ Yes

✗ No

Exercise 01 – Subquery

- Run all the cells in the notebook, they should work without errors
- Run checkers, it should include:

cnt

0 985

- The connection is closed?

In all other cases, the test is failed.

✓ Yes

✗ No

Exercise 02 – Join

- Run all the cells in the notebook, they should work without errors
- Run datamart.info(), it should include:

RangeIndex: 146 entries, 0 to 145

Data columns (total 4 columns):

Column Non-Null Count Dtype

0 uid 140 non-null object

1 labname 140 non-null object

2 first_commit_ts 140 non-null datetime64[ns]

3 first_view_ts 59 non-null datetime64[ns]

- Run test.info(), it should include:

Int64Index: 59 entries, 3 to 145

Data columns (total 4 columns):

Column Non-Null Count Dtype

0 uid 59 non-null object

1 labname 59 non-null object

2 first_commit_ts 59 non-null datetime64[ns]

3 first_view_ts 59 non-null datetime64[ns]

- Run control.info(), it should include:

Int64Index: 87 entries, 0 to 143

Data columns (total 4 columns):

Column Non-Null Count Dtype

0 uid 81 non-null object

1 labname 81 non-null object

2 first_commit_ts 81 non-null datetime64[ns]

3 first_view_ts 81 non-null datetime64[ns]

- The connection is closed?

- Open the connection and run test = pd.io.sql.read_sql('SELECT * FROM test', conn),
it should run without errors

- Open the connection and run control = pd.io.sql.read_sql('SELECT * FROM control', conn),
it should run without errors

In all other cases, the test is failed.

 Yes No

Exercise 03 – Aggregations

- Run df_min, it should be:
uid MIN(diff)
0 user_30 -202

- Run df_max, it should be:
uid MAX(diff)
0 user_25 -2

- Run df_avg, it should be:
AVG(diff)
0 -89.125

- Run views_diff.corr(), it should be:
avg_diff pageviews
avg_diff 1.000000 -0.062967
pageviews -0.062967 1.000000

- The connection is closed?

In all other cases, the test is failed.

 Yes No

Exercise 04 – A/B-testing

- Run all the cells in the notebook, they should work without errors

- Run test_results, it should be:
time AVG(diff)
0 after -99.523810
1 before -66.047619

- Run control_results, it should be:
time AVG(diff)
0 after -99.322222
1 before -98.033333

- The connection is closed?


- Ask if the hypothesis might be true, the answer should be yes, because
in the test group we see the difference and in the control group we do not see

In all other cases, the test is failed.

 Yes No

Ratings

Don't forget to check the flag corresponding to the defense

 Ok Outstanding project Empty work No author file Invalid compilation Norme Cheat Crash Leaks Forbidden function

Conclusion

Leave a comment on this evaluation

Отличный код, все
работает, пир разбирается

Finish evaluation

- Terms of use for video surveillance (<https://profile.intra.42.fr/legal/terms/1>)
- Rules of procedure (<https://profile.intra.42.fr/legal/terms/4>)
- Declaration on the use of cookies (<https://profile.intra.42.fr/legal/terms/2>)
- General term of use of the site (<https://profile.intra.42.fr/legal/terms/6>)
- Legal notices (<https://profile.intra.42.fr/legal/terms/3>)
- Privacy policy (<https://profile.intra.42.fr/legal/terms/5>)