UDACITY

< Return to Classroom

Predict Customer Churn with Clean Code

REVIEW

CODE REVIEW

HISTORY

Requires Changes

12 specifications require changes

Hello,

You are off to a good start to your program. Your work demonstrates your understanding of the concepts introduced in this module. Well done! 👏👏 👍 These are concepts you will find useful in later sections of this program as well as in your career.

Here's a tip that has helped me write better tests for my own code: read the tests of open source packages. It helps you see how the community is thinking about tests and how they are using/extending tools like Pytest.

You might enjoy Gleb Tocarenco's Clean Code with Python – Stop Feeding the Spaghetti Monster. Apart from its awesome title, it includes code snippets that help elucidate his points. Give a read :)

You have some changes to make and I have included notes in the relevant sections. Be sure to raise a Knowledge Question if you need support or clarification.

Looking forward to your submission!

Code Quality

All the code written for this project should follow the PEP 8 guidelines. Objects have meaningful names and syntax. Code is properly commented and organized. Imports are correctly ordered.

Running the below can assist with formatting.

```
autopep8 --in-place --aggressive --aggressive script.py
```

Then students should aim for a score exceeding 7 when using `pylint`

```
pylint script.py
```

pylint scored your scripts below 7, you need to make sure to follow PEP 8 Guidelines;

```
churn_script_logging_and_tests.py:127:2: C0103: Variable name "df" doesn't conform to snake_case naming styl
e (invalid-name)
churn_script_logging_and_tests.py:130:2: C0103: Variable name "X_train" doesn't conform to snake_case naming
 style (invalid-name)
churn_script_logging_and_tests.py:130:11: C0103: Variable name "X_test" doesn't conform to snake_case naming
 style (invalid-name)
churn_script_logging_and_tests.py:133:10: R1729: Use a generator instead 'all(col in X_train.columns for col
 in keep_cols)' (use-a-generator)
churn_script_logging_and_tests.py:134:10: R1729: Use a generator instead 'all(col in X_test.columns for col
in keep_cols)' (use-a-generator)
churn_script_logging_and_tests.py:130:19: W0612: Unused variable 'y_train' (unused-variable)
churn_script_logging_and_tests.py:130:28: W0612: Unused variable 'y_test' (unused-variable)
churn_script_logging_and_tests.py:151:1: C0103: Variable name "df" doesn't conform to snake_case naming styl
e (invalid-name)
churn_script_logging_and_tests.py:153:1: C0103: Variable name "X_train" doesn't conform to snake_case naming
 style (invalid-name)
churn_script_logging_and_tests.py:153:10: C0103: Variable name "X_test" doesn't conform to snake_case naming
 style (invalid-name)
churn_script_logging_and_tests.py:5:0: C0411: third party import "import joblib" should be placed before "im
port churn_library as cls" (wrong-import-order)
churn_script_logging_and_tests.py:1:0: R0801: Similar lines in 2 files
==churn_library:[134:142]
==churn_script_logging_and_tests:[117:124]
    keep_cols = ['Customer_Age', 'Dependent_count', 'Months_on_book',
                 'Total_Relationship_Count', 'Months_Inactive_12_mon',
                 'Contacts_Count_12_mon', 'Credit_Limit', 'Total_Revolving_Bal',
                 'Avg_Open_To_Buy', 'Total_Amt_Chng_Q4_Q1', 'Total_Trans_Amt',
                 'Total_Trans_Ct', 'Total_Ct_Chng_Q4_Q1', 'Avg_Utilization_Ratio',
                 'Gender_Churn', 'Education_Level_Churn', 'Marital_Status_Churn',
                 'Income_Category_Churn', 'Card_Category_Churn']
 (duplicate-code)

------------------------------------------------------------------
Your code has been rated at 0.71/10 (previous run: 8.21/10, -7.50)
```

The file contains a summary of the purpose and description of the project. Someone should be able to run the code by reading the README.

[Changes]

- You need to write a README that contains a summary of the purpose and description of the project. Someone should be able to run the code by reading the README.

All functions have a document string that correctly identifies the inputs, outputs, and purpose of the function. All files have a document string that identifies the purpose of the file, the author, and the date the file was created.

[Changes]

- You need to add a document string that identifies the purpose of the file, the author, and the date the file was created.

```
"""
This is the Python Test for the churn_library.py module.

This module will be used to test
    1. import_data
    2. peform_eda
    3. encode_data
    4. perform_feature_engineering
    5. train_test_model

Author: Your name
Date: date file was created
"""
```

Testing & Logging

Each function in `churn_script_logging_and_tests.py` is complete with tests for the input function.

There is an error in your test script that you need to address;

```
/workspace/home$ python *s.py
Traceback (most recent call last):
  File "/mnt/c/Users/USER/Downloads/submit-17bdaba5-89c0-41d5-9cd0-e5fd5a46f84b/workspace/home/churn_script_
logging_and_tests.py", line 176, in <module>
    test_train_models(cls.import_data, cls.perform_feature_engineering, cls.train_models)
  File "/mnt/c/Users/USER/Downloads/submit-17bdaba5-89c0-41d5-9cd0-e5fd5a46f84b/workspace/home/churn_script_
logging_and_tests.py", line 153, in test_train_models
    X_train, X_test, y_train, y_test = perform_feature_engineering(df)
  File "/mnt/c/Users/USER/Downloads/submit-17bdaba5-89c0-41d5-9cd0-e5fd5a46f84b/workspace/home/churn_library
.py", line 149, in perform_feature_engineering
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=random_state
)
NameError: name 'y' is not defined
```

Each function in `churn_script_logging_and_tests.py` is complete with logging for if the function successfully passes the tests or errors.

This would be reviewed after you have addressed the error in your test script.

All log information should be stored in a `.log` file, so it can be viewed post the run of the script.

This would be reviewed after you have addressed the error in your test script.

The log messages should easily be understood and traceable that appear in the `.log` file.

This would be reviewed after you have addressed the error in your test script.

The README should inform a user how they would test and log the result of each function.

Something similar to the below should produce the `.log` file with the result from running all tests.

```
ipython churn_script_logging_and_tests_solution.py
```

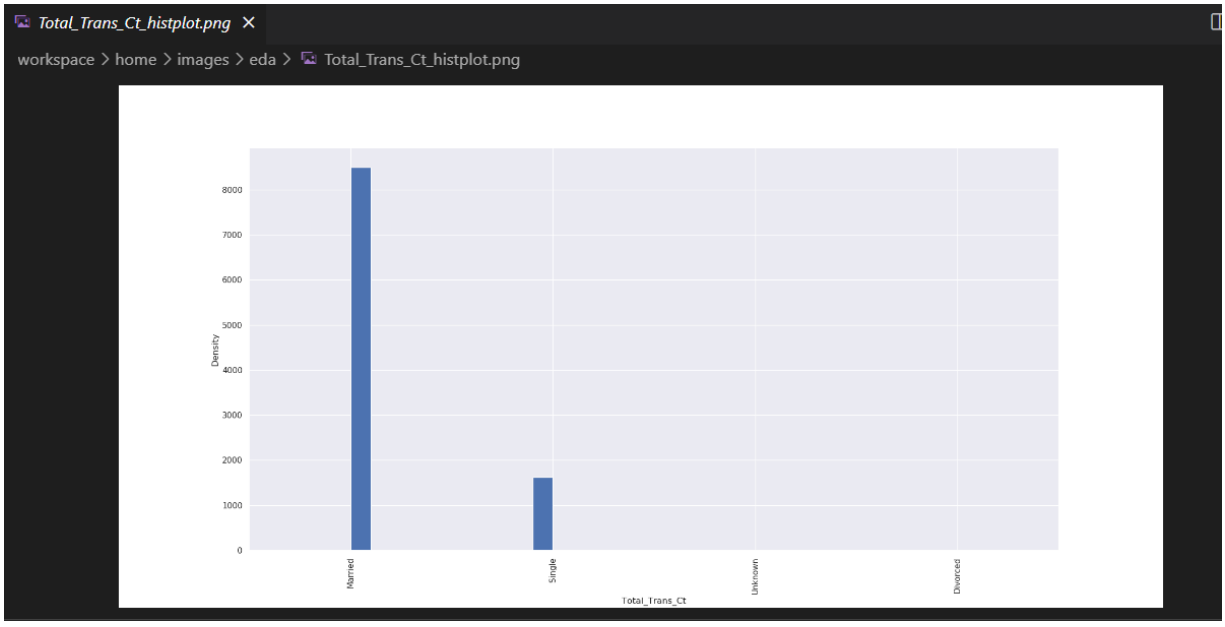
Your README should inform a user how they would test and log the result of each function.

Save Images & Models

Store result plots including at least one:

1. Univariate, quantitative plot
2. Univariate, categorical plot
3. Bivariate plot

Your `eda` plots are not correctly implemented, here is one of the incorrect plots;



Store result plots including:

1. ROC curves
2. Feature Importances

You also need to plot the ROC curve for your Logistic Regression model

Store at least two models. Recommended using `joblib` and storing models with `.pkl` extension.

✓✓ You persisted both models using `joblib` and stored them with a `pkl` extension in the model folder.

Problem Solving

Code in `churn_library.py` completes the process for solving the data science process including:

1. EDA
2. Feature Engineering (including encoding of categorical variables)
3. Model Training
4. Prediction
5. Model Evaluation

This would be reviewed after you have addressed the issues stated above.

Use one-hot encoding or mean of the response to fill in categorical columns. Currently, the notebook does this in an inefficient way that can be refactored by looping. Make this code more efficient using the same method as in the notebook or using one-hot encoding. Tip: Creating a list of categorical column names can help with looping through these items and create an easier way to extend this logic.

This would be reviewed after you have addressed the issues stated above.

RESUBMIT

DOWNLOAD PROJECT

Learn the best practices for revising and resubmitting your project.

RETURN TO PATH

Rate this review

START