










<a href="#">Previous</a>	 	 			<a href="#">Next</a> 
--------------------------	---	---	---	---	--

## Case Study 7, Part 1 Homework: Exercises 5-7

 Bookmark this page

## Exercise 5

1/1 point (graded)

Some variables in the dataset are already numeric and perhaps useful for regression and classification. In Exercise 5, we will store the names of these variables for future use. We will also take a look at some of the continuous variables and outcomes by plotting each pair in a scatter plot. Finally, we will evaluate the skew of each variable.

### Instructions

- Call `plt.show()` to observe the plot generated by the code given below. Which of the covariates and/or outcomes are correlated with each other?
- Call `skew()` on the columns `outcomes_and_continuous_covariates` in `df`. Is the skew above 1 for any of these variables?

Here is the code to get you started:

```
continuous_covariates = ['budget', 'popularity', 'runtime', 'vote_count', 'vote_average']
outcomes_and_continuous_covariates = continuous_covariates + [regression_target, classification_target]
plotting_variables = ['budget', 'popularity', regression_target]

axes = pd.plotting.scatter_matrix(df[plotting_variables], alpha=0.15, \
    color=(0,0,0), hist_kwds={"color":(0,0,0)}, facecolor=(1,0,0))
# show the plot.

# determine the skew.
```

Which continuous covariate appears to be the most skewed?

☐ budget

☒ popularity

☐ runtime

☐ vote\_count

☐ vote\_average

☐ revenue

☐ profitable



Submit

You have used 1 of 2 attempts

✓ Correct (1/1 point)

## Exercise 6

1/1 point (graded)

It appears that the variables `budget`, `popularity`, `runtime`, `vote_count`, and `revenue` are all right-skewed. In Exercise 6, we will transform these variables to eliminate this skewness. Specifically, we will use the `np.log10()` method. Because some of these variable values are exactly 0, we will add a small positive value to each to ensure it is defined; this is necessary because  $\log(0)$  is negative infinity.

### Instructions

For each above-mentioned variable in `df`, transform value `x` into `np.log10(1+x)`.

What is the new value of `skew()` for the covariate `runtime`? Please provide the answer to 3 decimal points.



0.530

Submit

You have used 1 of 10 attempts

✓ Correct (1/1 point)

## Exercise 7

1/1 point (graded)

Now we're going to save our dataset to use in Part 2 of this case study.

### Instructions

Use `to_csv()` to save the `df` object as `movies_clean.csv`.

What is the correct way to save the `df` object?



`pd.to_csv(df)`



`df.to_csv("movies_clean.csv")`



`pd.to_csv("movies_clean.csv")`



`np.full((3,3), dtype=int)`



Submit

You have used 1 of 2 attempts

✓ Correct (1/1 point)

< Previous

Next >



# edX

- [About](#)
- [Affiliates](#)
- [edX for Business](#)
- [Open edX](#)
- [Careers](#)
- [News](#)

## Legal

- [Terms of Service & Honor Code](#)
- [Privacy Policy](#)
- [Accessibility Policy](#)
- [Trademark Policy](#)
- [Sitemap](#)

## Connect

- [Blog](#)
- [Contact Us](#)
- [Help Center](#)
- [Media Kit](#)
- [Donate](#)



© 2021 edX Inc. All rights reserved.  
深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)