



<a href="#">&lt; Previous</a>	<div><div>📄 ✓</div></div>	<div><div>📝</div></div>	<div><div>📝</div></div>	<div><div>📄</div></div>	<a href="#">Next &gt;</a>
-------------------------------	---------------------------	-------------------------	-------------------------	-------------------------	---------------------------

Case Study 7, Part 1 Homework: Exercises 1-4

🔖 Bookmark this page

Homework due Jul 14, 2021 05:59 +06

## Exercise 1

1/1 point (graded)

First, we will import several libraries. **scikit-learn** ( `sklearn` ) contains helpful statistical models, and we'll use the `matplotlib.pyplot` library for visualizations. Of course, we will use `numpy` and `pandas` for data manipulation throughout.

### Instructions

Read and execute the given code, then call `df.head()` to take a look at the data.

Here's the import code:

```
import pandas as pd
import numpy as np

from sklearn.model_selection import cross_val_predict
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import r2_score

import matplotlib.pyplot as plt

df = pd.read_csv("https://courses.edx.org/asset-v1:HarvardX+PH526x+2T2019+type@asset+block@movie_data.csv")

# Enter code here.
```

What is the title of the first movie in this dataset?

Enter the title exactly as it appears.



Submit

You have used 1 of 10 attempts

✓ Correct (1/1 point)

## Exercise 2

1/1 point (graded)

In Exercise 2, we will define the regression and classification outcomes. Specifically, we will use the `revenue` column as the target for regression. For classification, we will construct an indicator of profitability for each movie.

### Instructions

- Create a new column in `df` called `profitable`, defined as 1 if the movie revenue ( `revenue` ) is greater than the movie budget ( `budget` ), and 0 otherwise.
- Next, define and store the outcomes we will use for regression and classification. Define `regression_target` as the string `'revenue'`. Define `classification_target` as the string `'profitable'`.

Hide Notes

How many movies in this dataset are defined as profitable (value 1)?



2585

Submit

You have used 1 of 10 attempts

✓ Correct (1/1 point)

## Exercise 3

1/1 point (graded)

For simplicity, we will proceed by analyzing only the rows without any missing data. In Exercise 3, we will remove rows with any infinite or missing values.

### Instructions

- Use `df.replace()` to replace any cells with type `np.inf` or `-np.inf` with `np.nan`.
- Drop all rows with any `np.nan` values in that row using `df.dropna()`. Do any further arguments need to be specified in this function to remove rows with any such values?

How many movies are left in the dataset after dropping any rows with infinite or missing values?



1406

Submit

You have used 1 of 10 attempts

✓ Correct (1/1 point)

## Exercise 4

1/1 point (graded)

Many of the variables in our dataframe contain the names of genre, actors/actresses, and keywords. Let's add indicator columns for each genre.

### Instructions

- Determine all the genres in the `genre` column. Make sure to use the `strip()` function on each genre to remove trailing characters.
- Next, include each listed genre as a new column in the dataframe. Each element of these genre columns should be 1 if the movie belongs to that particular genre, and 0 otherwise. Keep in mind that a movie may belong to several genres at once.
- Call `df[genres].head()` to view your results.

How many genres of movies are in this dataset?



20

Submit

You have used 1 of 10 attempts

Hide Notes

✓ Correct (1/1 point)

◀ Previous

Next ▶

© All Rights Reserved



## 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](#)

Hide Notes