datacamp   Guides ● ▭ ⊘

[Instructions | Code]

▭ Code-Along Video   ▭ ⌄
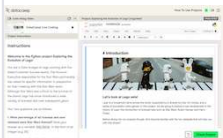
DataCamp Live Coding   ▶

Project Instructions

Instructions

Data visualization is often a great way to start exploring your data and uncovering insights. In this notebook, you will initiate this process by creating an informative plot of the episode data provided to you. In doing so, you're going to work on several different variables, including the episode number, the viewership, the fan rating, and guest appearances. Here are the requirements needed to pass this project:

1. Create a `matplotlib` **scatter plot** of the data that contains the following attributes:

   - Each episode's **episode number plotted along the x-axis**
   - Each episode's **viewership (in millions) plotted along the y-axis**
   - A **color scheme** reflecting the **scaled ratings** (not the regular ratings) of each episode, such that:
     - Ratings < 0.25 are colored `"red"`
     - Ratings >= 0.25 and < 0.50 are colored `"orange"`
     - Ratings >= 0.50 and < 0.75 are colored `"lightgreen"`
     - Ratings >= 0.75 are colored `"darkgreen"`
   - A **sizing system,** such that episodes with guest appearances have a marker size of `250` and episodes without are sized `25`
   - A **title,** reading `"Popularity, Quality, and Guest Appearances on the Office"`
   - An **x-axis label** reading `"Episode Number"`
   - A **y-axis label** reading `"Viewership (Millions)"`

2. Provide the name of one of the guest stars (hint, there were multiple!) who was in the most watched Office episode. Save it as a string in the variable `top_star` (e.g. `top_star = "Will Ferrell"`).

Important!

To test your `matplotlib` plot, you will need to initalize a `matplotlib.pyplot` `fig` object, which you can do using the code `fig = plt.figure()` (provided you have imported `matplotlib.pyplot as plt`). In addition, in order to test it correctly, **please make sure to specify your plot (including the type, data, labels, etc) in the same cell as the one you initialize your figure (`fig`)!** *You are still free to use other cells to load data, experiment, and answer Question 2.*

In addition, if you want to be able to see a larger version of your plot, you can set the figure size parameters using this code (provided again you have imported `matplotlib.pyplot as plt`):

```
plt.rcParams['figure.figsize'] = [11, 7]
```

Bonus Step!

Although it was not taught in Intermediate Python, a useful skill for visualizing different data points is to use a different marker. You can learn more about them via the **Matplotlib documentation** or via our course **Introduction to Data Visualization with Matplotlib**. Thus, as a bonus step, try to differentiate guest appearances not just with size, but also with a star!

All other attributes still apply (data on the axes, color scheme, sizes for guest appearances, title, and axis labels).

[⏻ Take Hint]