**Investigating Netflix Movies and Guest Stars in The Office**

**Python projects on** [**DataCamp**](https://app.datacamp.com/learn/projects)

*Apply the foundational skills in Introduction to Python and Intermediate Python courses to manipulate and visualize movie and TV data.*

**🛠️ Description**

In these projects, I’ll apply the skills I learned in Introduction to Python and Intermediate Python to solve a real-world data science problem. I’ll press “watch next episode” to discover if Netflix’s movies are getting shorter over time and which guest stars appear in the most popular episode of "The Office", using everything from lists and loops to pandas and matplotlib.

This project helps me gain experience in an essential data science skill — exploratory data analysis, which allows me to perform critical tasks such as manipulating raw data and drawing conclusions from plots I create of the data.

**📺 Projects**

**Guided project**

*Dig into a real-world Netflix movie dataset using everything from lists and loops to pandas and matplotlib.*

**Project tasks**

1. Loading a friend's data into a dictionary
2. Creating a DataFrame from a dictionary
3. A visual inspection of our data
4. Loading the rest of the data from a CSV
5. Filtering for movies!
6. Creating a scatter plot
7. Digging deeper
8. Marking non-feature films
9. Plotting with color!
10. What next?

**Unguided project**

*Bring together a wide variety of skills from Intermediate Python to prepare and plot data on the history of the sitcom The Office.*

**Project task**

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)"

1. Provide the name of one of the guest stars (hint, there were multiple!) who was in the most watched Office episode.