**Plotting Linear Regressions with Seaborn**

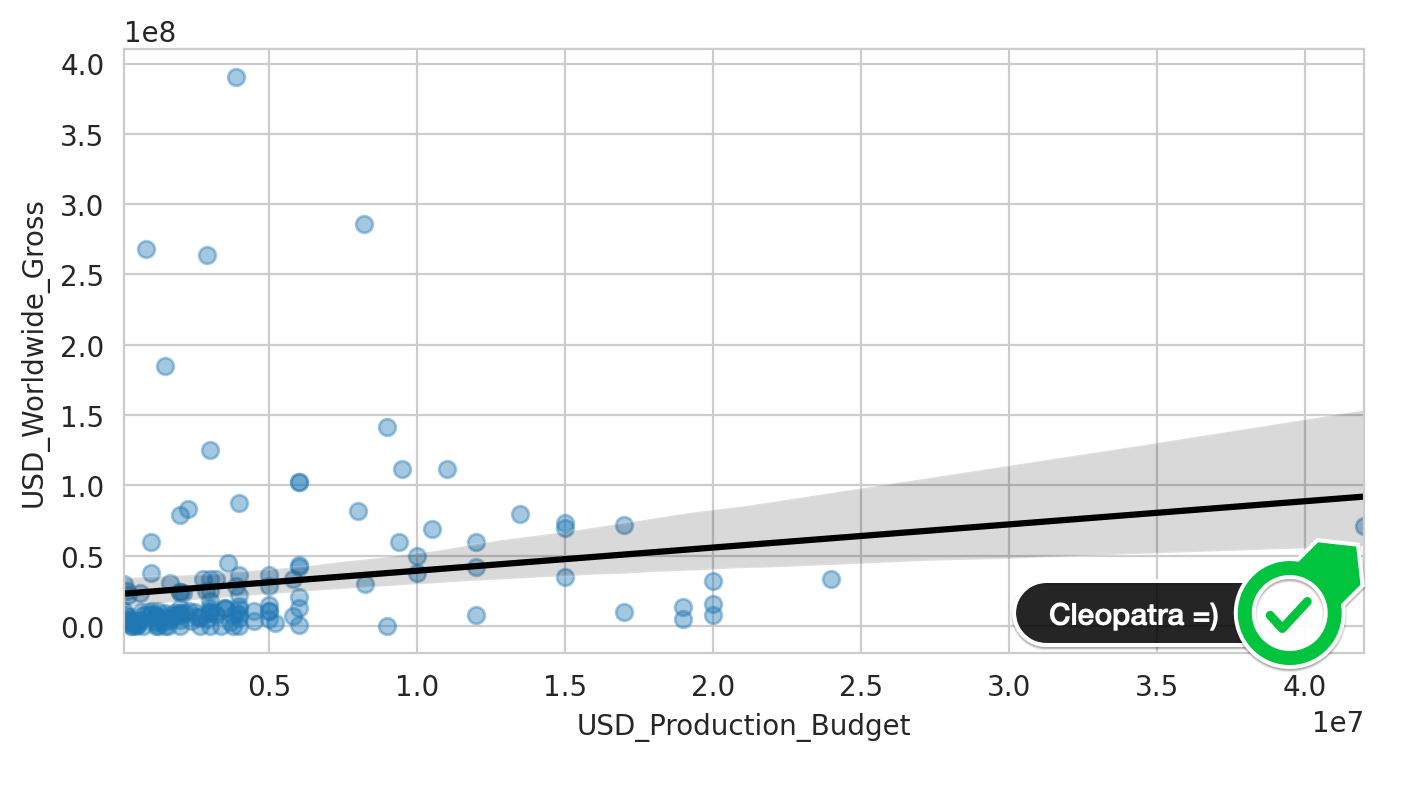
Let's visualise the relationship between the movie budget and the worldwide revenue using linear regression. Seaborn makes this incredibly easy with the [.regplot()](https://seaborn.pydata.org/generated/seaborn.regplot.html#seaborn.regplot) function.

1. sns.regplot(data=old\_films,
2. x='USD\_Production\_Budget',
3. y='USD\_Worldwide\_Gross')

This creates a scatter plot and draws a linear regression line together with the confidence interval at the same time.

To style the chart further, we can once again, drop into the Matplotlib layer and supply keyword arguments as dictionaries. We can customise the scatter plot (e.g., by changing the transparency of the dots) and the regression line itself (e.g., by changing the colour).

1. plt.figure(figsize=(8,4), dpi=200)
2. with sns.axes\_style("whitegrid"):
3. sns.regplot(data=old\_films,
4. x='USD\_Production\_Budget',
5. y='USD\_Worldwide\_Gross',
6. scatter\_kws = {'alpha': 0.4},
7. line\_kws = {'color': 'black'})



What do we see here? Well, first off we can spot Cleopatra on the far right. But also, we see that many lower budget films made much more money! The relationship between the production budget and movie revenue is not very strong. Many points on the left are very far away for the line, so the line appears not to capture the relationship between budget and revenue very well at all!

But does the same hold true for the newer films?

Challenge

Use Seaborn's .regplot() to show the scatter plot and linear regression line against the new\_films.  
  
Style the chart

* Put the chart on a 'darkgrid'.
* Set limits on the axes so that they don't show negative values.
* Label the axes on the plot "Revenue in $ billions" and "Budget in $ millions".
* Provide HEX colour codes for the plot and the regression line. Make the dots dark blue (#2f4b7c) and the line orange (#ff7c43).

Interpret the chart

* Do our data points for the new films align better or worse with the linear regression than for our older films?
* Roughly how much would a film with a budget of $150 million make according to the regression line?

.

.

..

...

..

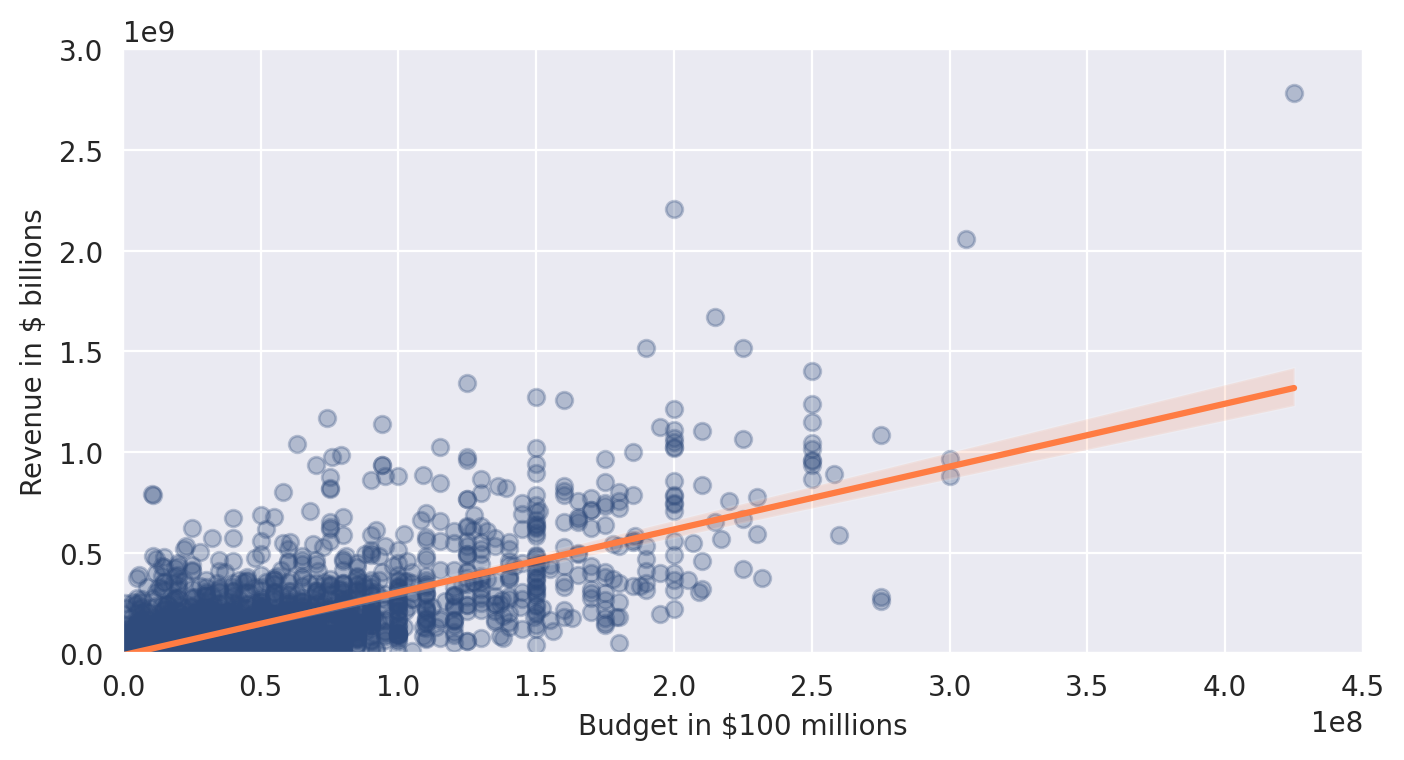
.

.

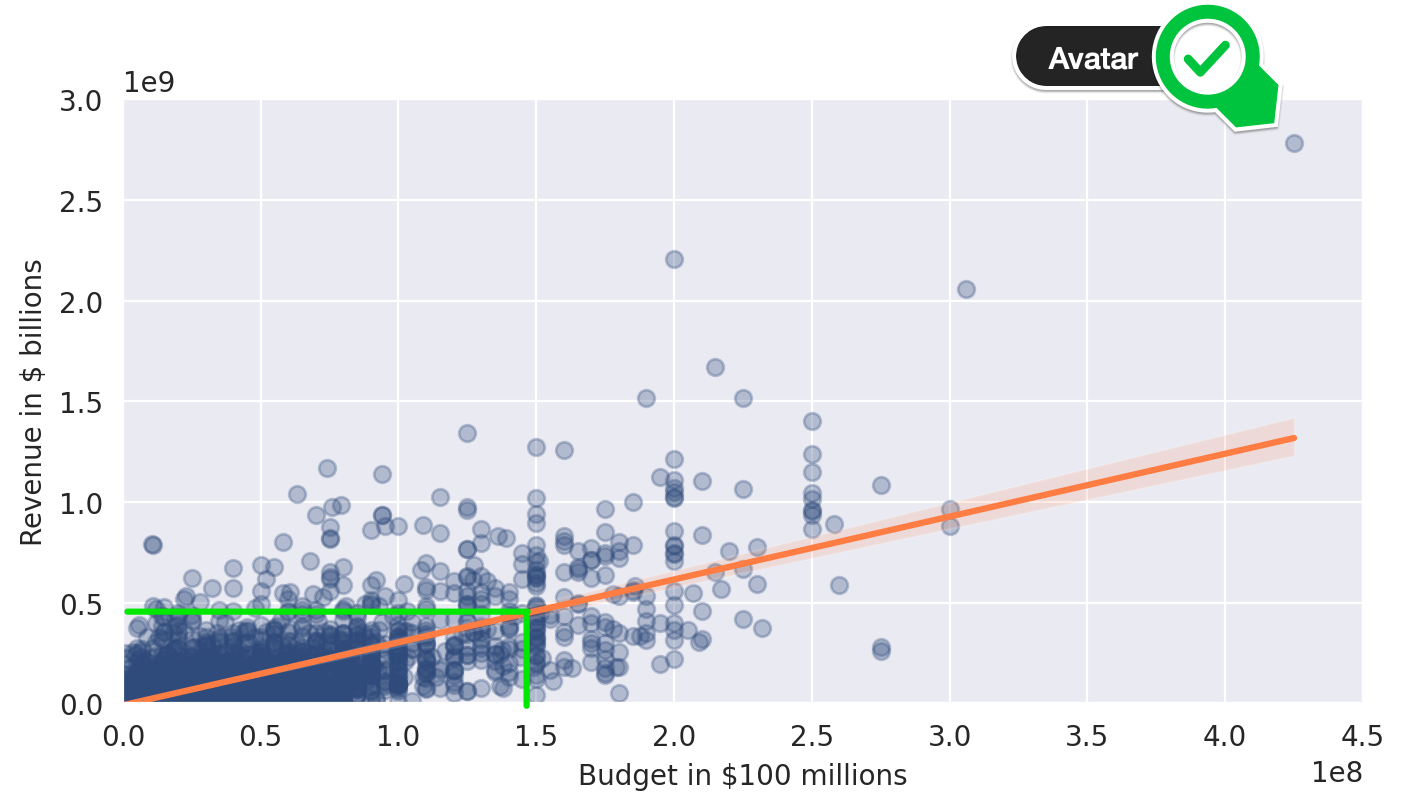
**Solution: Plotting a regression against the newer films**

To style the chart we can use the same techniques as before: providing values for the .regplot() function, as well as making use of the Matplotlib Axes object to fine-tune the limits, labels, and general style.

1. plt.figure(figsize=(8,4), dpi=200)
2. with sns.axes\_style('darkgrid'):
3. ax = sns.regplot(data=new\_films,
4. x='USD\_Production\_Budget',
5. y='USD\_Worldwide\_Gross',
6. color='#2f4b7c',
7. scatter\_kws = {'alpha': 0.3},
8. line\_kws = {'color': '#ff7c43'})
10. ax.set(ylim=(0, 3000000000),
11. xlim=(0, 450000000),
12. ylabel='Revenue in $ billions',
13. xlabel='Budget in $100 millions')



How do we interpret our chart? This time we are getting a much better fit, compared to the old films. We can see this visually from the fact that our data points line up much better with our regression line (pun intended). Also, the confidence interval is much narrower. We also see that a film with a $150 million budget is predicted to make slightly under $500 million by our regression line.



All in all, we can be pretty confident that there does indeed seem to be a relationship between a film's budget and that film's worldwide revenue.

But how much of the variation in revenue does the budget actually explain? And how much extra revenue can we expect for an additional $1 increase in the budget? To find out, we need to dive into the numbers underlying our regression model.