**SI649 Fall 2016 – Lab 1-answer-sheet**

**Sep 12/13, 2016  -- DUE SUNDAY, SEPTEMBER 18th at MIDNIGHT**

Name: tzuilee

Unique Name:

Lab Partner’s Name:

* Please answering the following questions with screenshots and descriptions.
* When taking screenshots, please include the ENTIRE tableau window.
* Bonus tasks are optional and give you extra credit.
* When answering question 10 and 11, please document steps that you used to generate the visualization. Also describe your solutions in words.

You will upload the following items to the canvas:

⬜ the pdf version of the answer sheet

⬜ the tableau book of the group project (even if you complete these individually)

⬜ the tableau book of the individual project

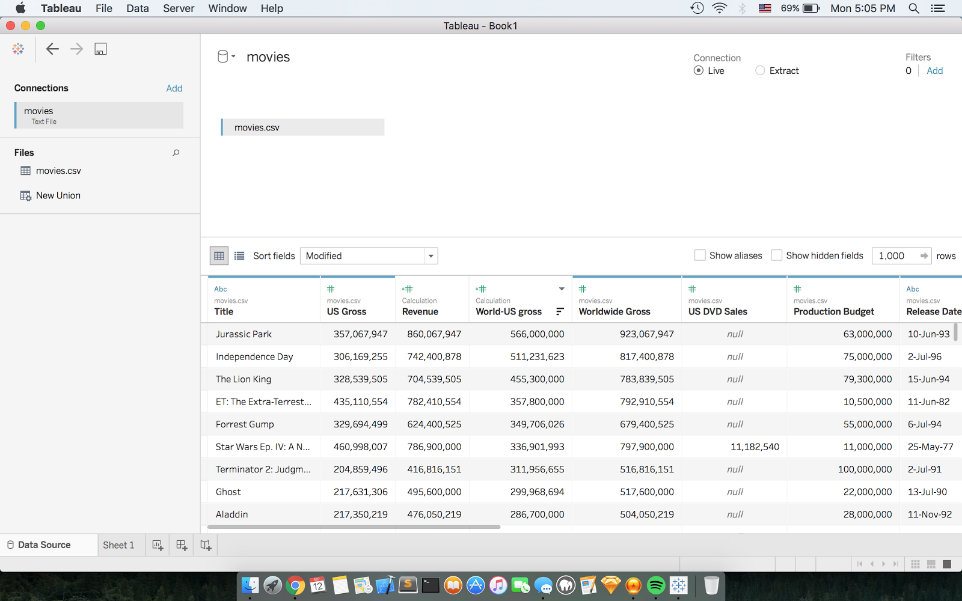
⬜ data used for the individual project (if not embedded in the Tableau workbook)

----------------------------------------------------Group Task------------------------------------------------------

***Question 3:*Which is the most profitable movie?**

⬜    3.3 Which has the largest difference between worldwide gross and U.S. gross?

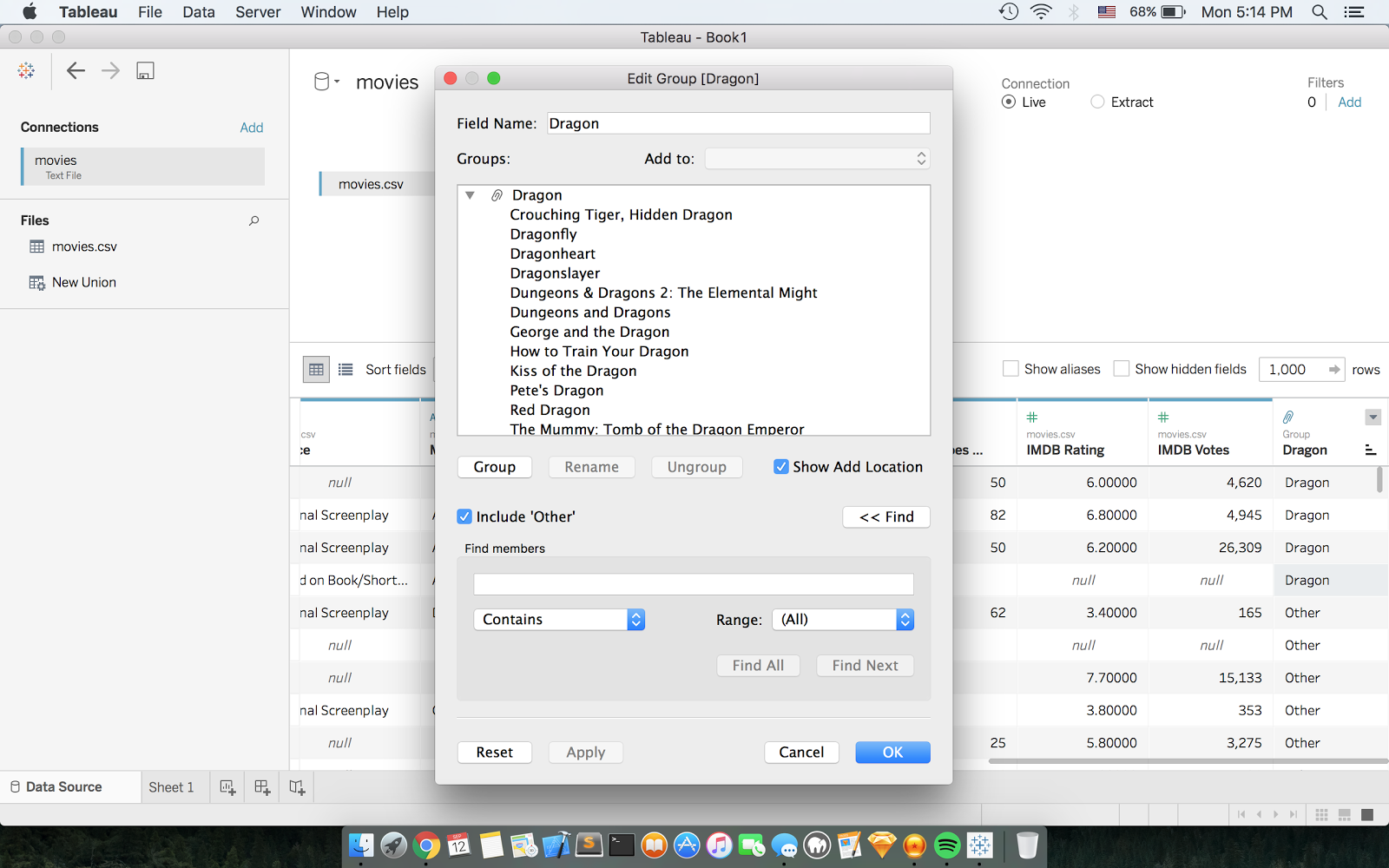
Movie name: Jurassic Park



***Question 4:* How do I find all movies that have contain specific string in their names?**

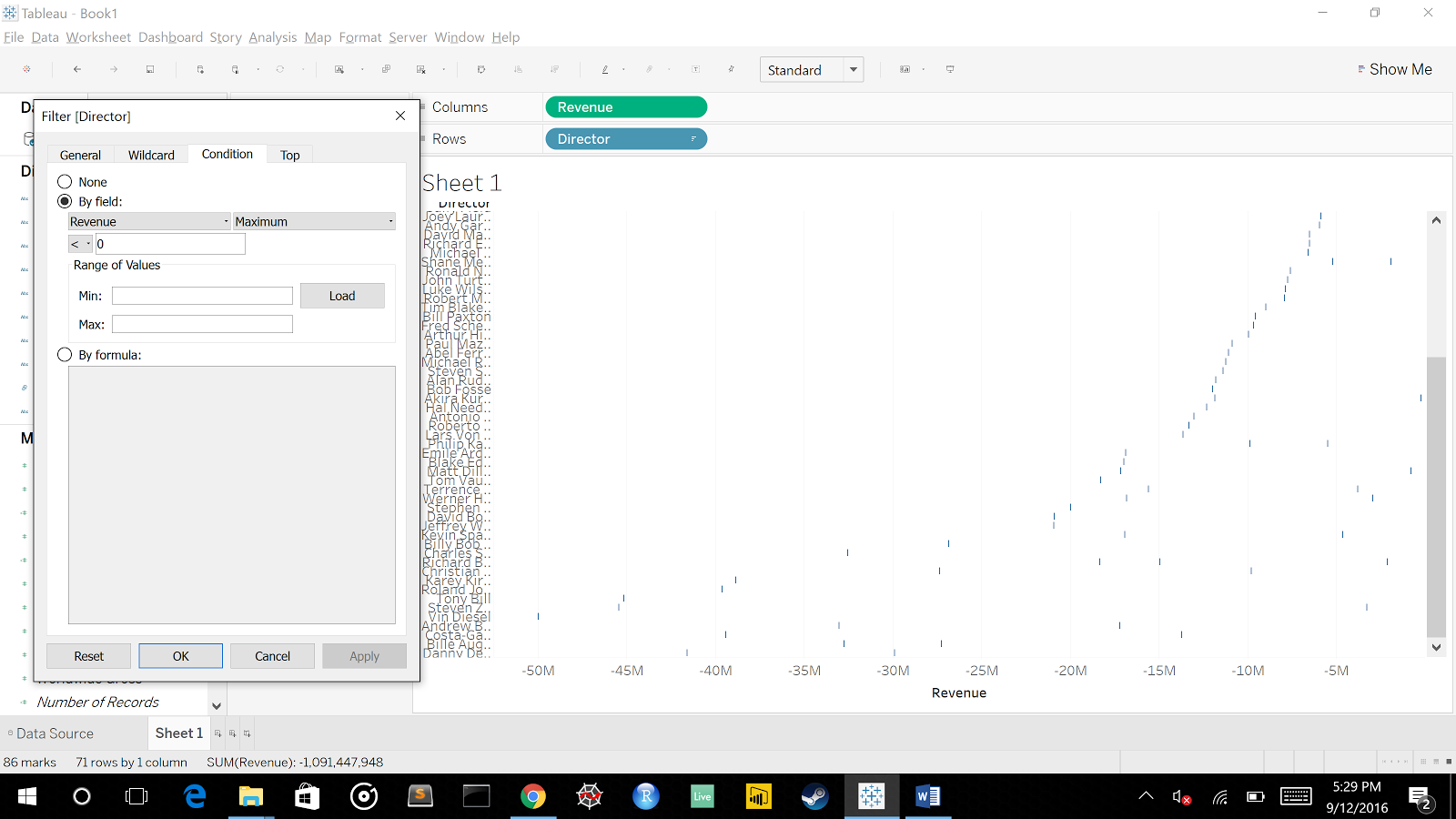
⬜    4.2 Create a group with all titles that contains the string “dragon”

Paste screenshot here of you creating the group (we don’t need the full list):



***Question 7:* Is there any director who has never directed any profitable movie?**

⬜  7.6 Are there directors who have never directed profitable movies? Who are they? (Hint, besides using revenue as a dimension, you can also use filters to complete this task) (You don’t need to write down the whole list, just show us how to find them.)

Screenshot:

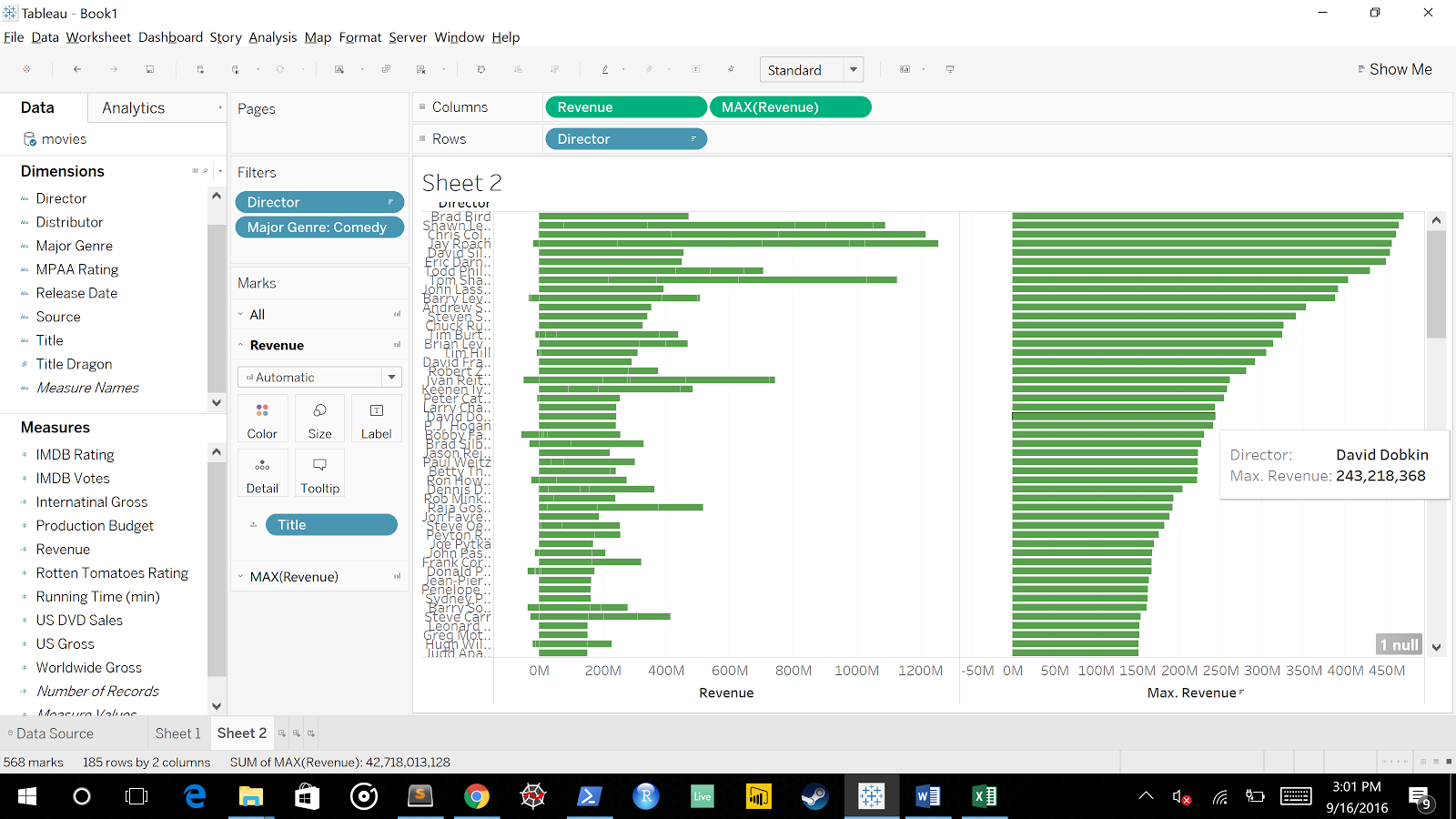
Answer: Yes

**Question 8: Can I get more information about a specific genre?**

⬜     8.3 If you want to find out which director has the highest average revenue for action movies, do you want to use pages shelf or filters shelf? Why?

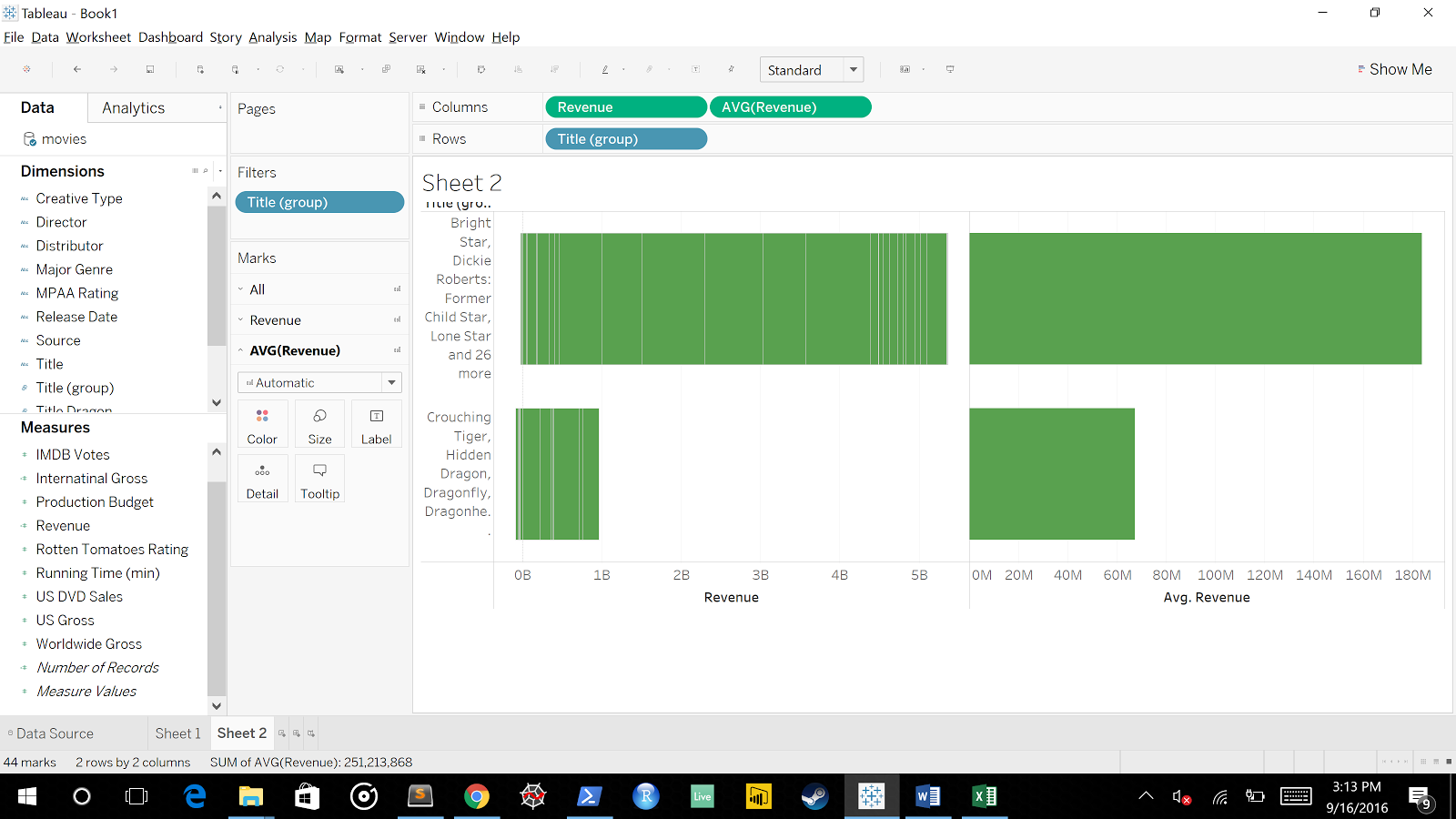
You want to use the filter shelf because if you use pages it sorts based on the total of all genres. If you want to find the Director with the greatest average for a specific genre you have to filter out all the other genres.

⬜   8.4 Which director has directed the most profitable Comedy movie? (Hint: use https://lh3.googleusercontent.com/uxn3qhjjR8uUP2E71kY_SpAWjEdU76a6HXOpq0UHf8-WaQ10_la5HEEIZOq2TQ7G423SsMJs_nouKnd-z7qZAo3AN7r-d_DTDaM3hCvnD61IO4XMvjrqdWlwyr99t1vf6x-5bONland  filters/pages shelf)

Paste your screenshot:

Director name: Brad Bird for “Ratatouille”

⬜   8.5  Bonus Question: do movies that contain the name “star” have higher average revenue than movies contain the name “dragon”?

Screenshot of visualization: 

Answer: Movie’s with the string ‘star’ have higher average rating than movies with ‘dragon’.

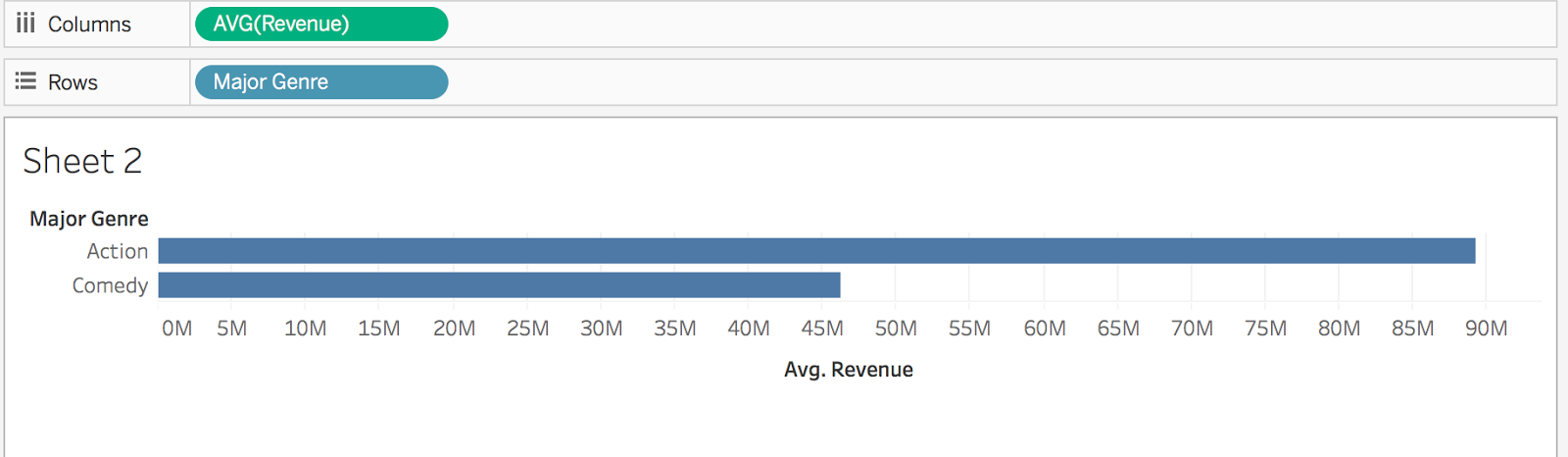
***Question 11*: Is it better to invest in action movies or in comedy movies? Which director would you invest in?**

Our group want to invest in action movies because the average revenue of action movie is higher than comedy. Our group want to invest in James Cameron because he is the director whose average movie revenue is the highest.

I obtained this solution by completing the following steps:

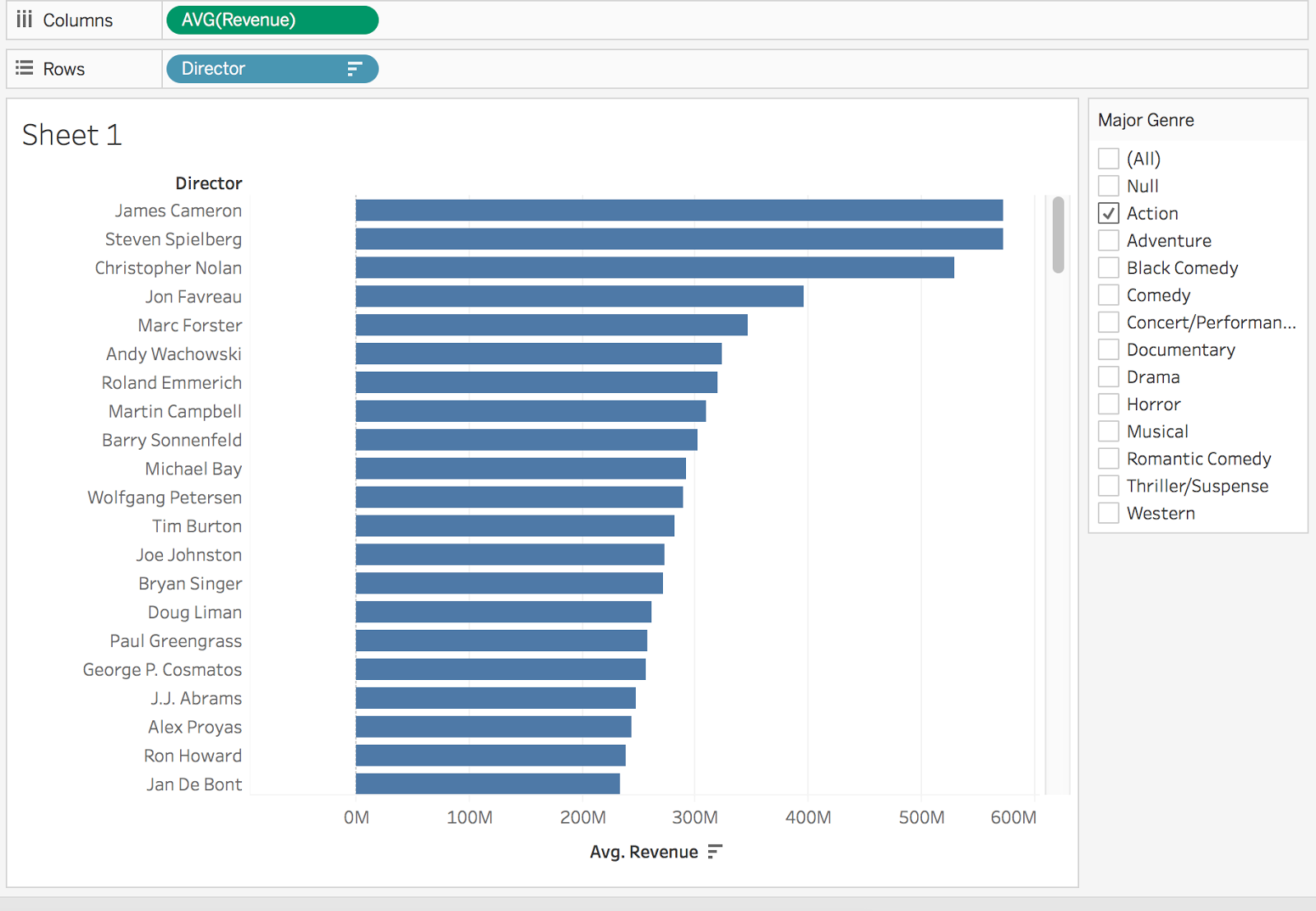
Step1: Put AVG(revenue) at columns and major genre as rows.

Step 2: Filter out genre to that it make it only show action and comedy movie.



Step3: Create another figure to show average revenue of action movie per director.

Step4: Sort avg(revenue) to display highest to lowest.



***Question 12*: What’s the relationship between IMDB ratings and the revenue?Are highly-rated movies more profitable?  ( Please write down the steps that you used to create the visualizations, and describe how these visualizations help you to make your decision)**

Ratings and the revenue is statistically correlated based on the regression model. So, yes, we could say that highly-rated movies are more profitable.

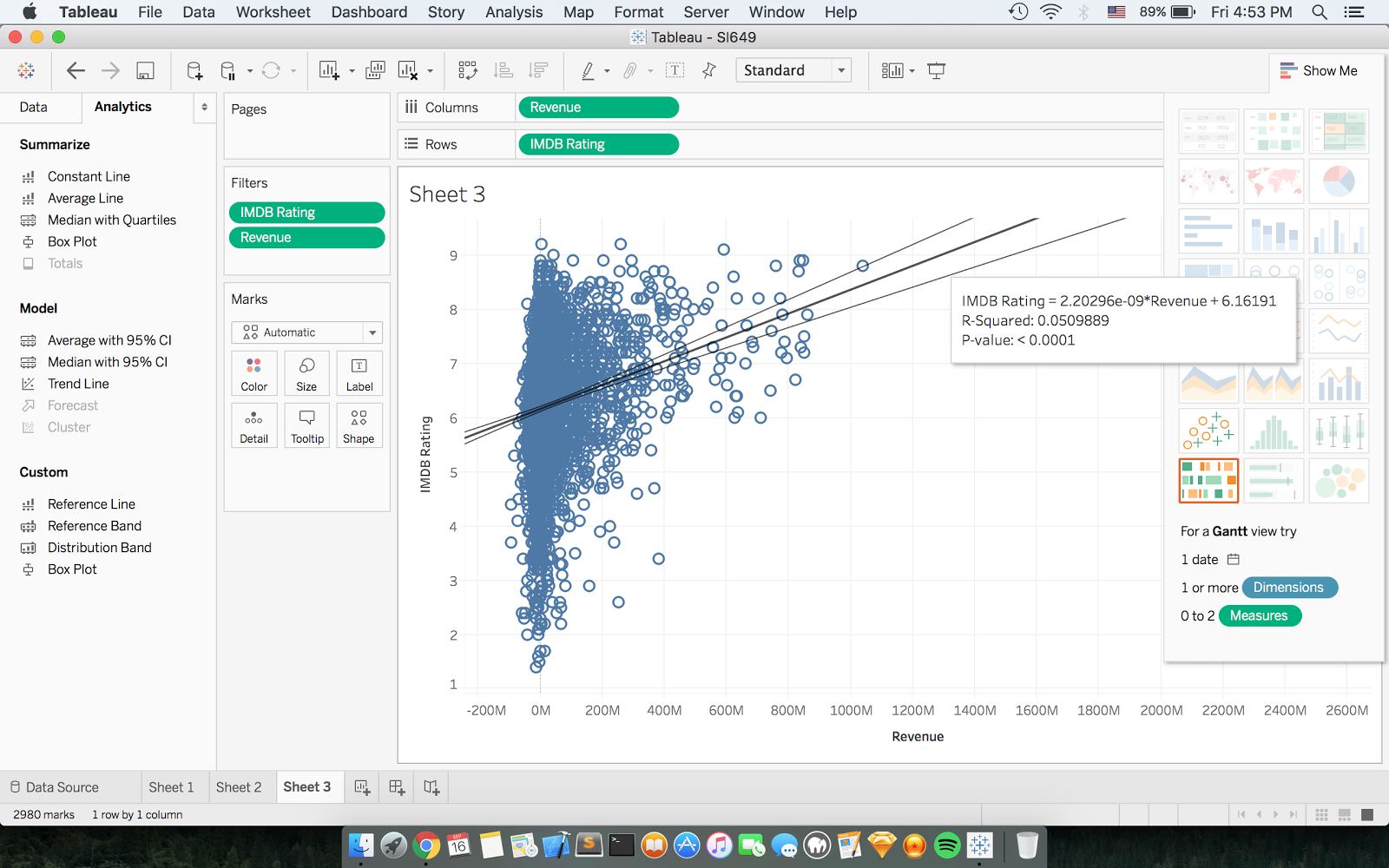
I obtained this solution by completing the following steps:

Step1: Create scatter plot, row: revenue, rating rows.

Step2: Exclude null

Step3: Add trend line to the plot.

Step4: the p-value < 0.00001 which means they are positively correlated.

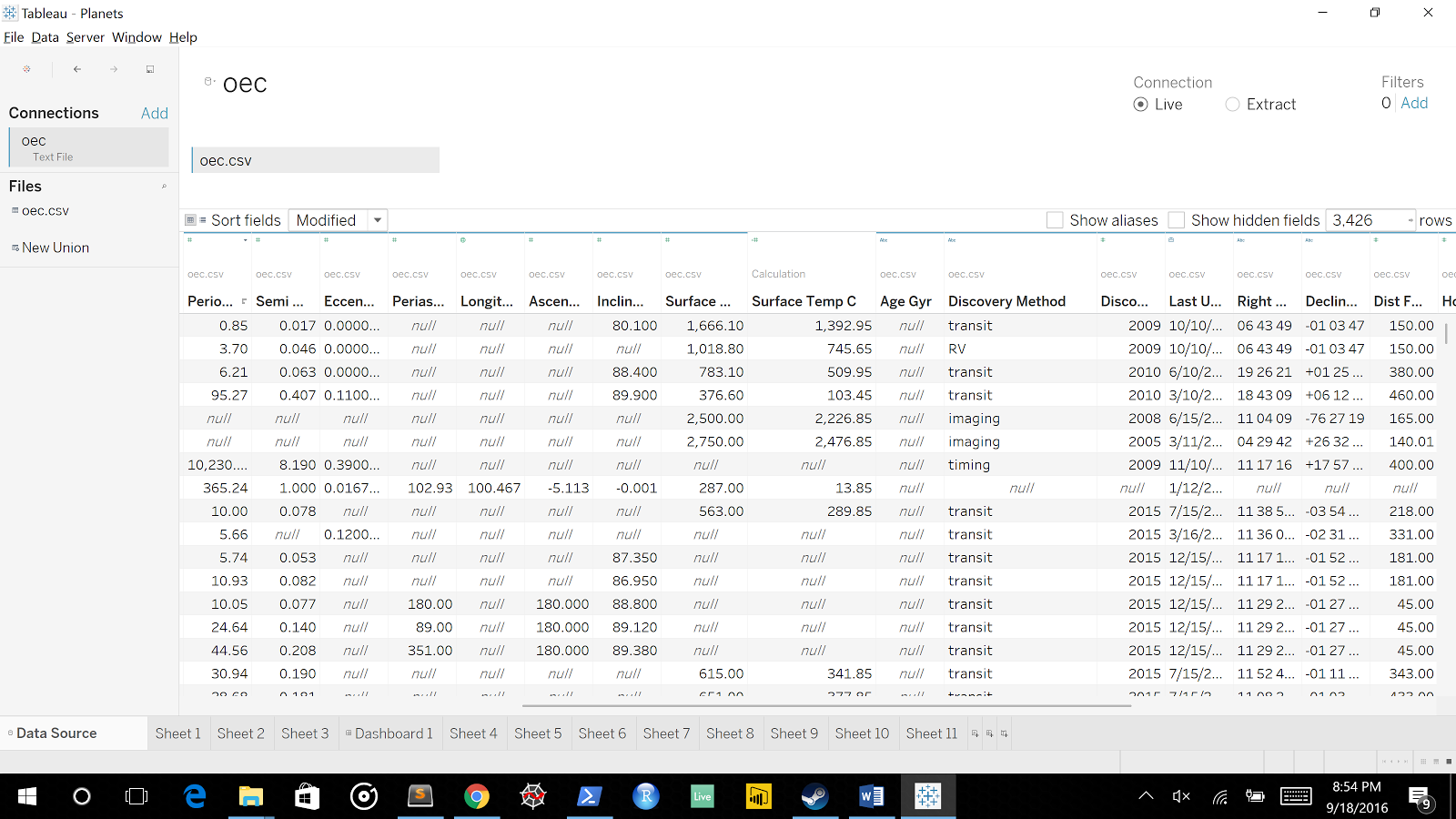


------------------------------------------------------------Individual Task------------------------------------------------------

⬜ Step 1. The domain that I am interested in is:

Cosmology and Astrophysics

⬜ Step 1.1. The dataset that I selected is: The open-exoplanet-catalogue

⬜ a screenshot of the sample data-table

⬜ The link to the dataset

https://www.kaggle.com/mrisdal/open-exoplanet-catalogue

⬜ include the dataset that we need to open your tableau book in the canvas submission

⬜ Step 2. The  initial question(s) that I would like to answer is (are):

Have we found any planets that could be settled?

⬜ Step 3. Documentwhat you had to do to construct the visualizations and how the questions evolved. Please include screenshot of the visualizations that you have created. **Each insight visualization should be well constructed and labeled (put annotations on the images if you want to call attention to specific findings).  You should have at least 2-3 insight visualizations and 9-10 experiments**. **Aesthetics count here, please make sure you make good choices in colors, fonts, layouts, etc.**Here are some aspects that you might want to include in your documentation:

⬜ **1.** Describe any transformations or rearrangements of the dataset that you needed to perform;

Fortunately all the transformations could be done in Tableau using calculated fields. I had to calculate Gravity and converted Kelvin into Celsius

⬜ **2.** Describe how you got the data into the format needed by the visualization system.

I just opened it using their tool.

⬜ **3**. how it answers the question you posed

I decided to look at the gravity and the temperature of the planet's surface. It turns out only a couple planets are within the freezing and boiling points of water. Of the two planets one has Several Times the gravity that earth has, so that is out. The other has the lower gravity than earth but it has a larger radius suggesting that it doesn’t have a rocky core like earth. While this could mean that it’s a big ball of water it has the added problem of having an average temperature of 66C or 150F!