Exercise 02

This should be completed individually.

Section 1: Querying a database in python

1. Connect to your local postgres instance from python (the pyscopg2 library is relatively straightforward).

2. Write a query to return a dataset with film, rental, and payment information. Your dataset should have multiple rows per film, one for each time the film was rented and the amount spent on each rental. Create a dataframe with this information

3. Create a dataframe from the customer table

Section 2: Manipulating dataframes

1. Create a column for customer_name that has the first name and last name in the same column.
2. Remove any inactive customers from the dataframe
3. Change the email addresses to be 'joe.person@wustl.edu', but only when their store_id is an even number

Section 3: Visualizations

1. How much do customers tend to spend on rentals in aggregate?
   Clarification: You want to first create total spend by customer, then you want to visualize that distribution, each customer being an observation. A box and whisker plot would be a good visualization
2. What does the distribution of film revenue look like?
   Clarification: You want to first sum the revenue by film (multiply the rental rate by how many times the film was rented), then use each film as an observation. A histogram would be a good visualization

Section 4: Analysis

1. On average, do some movie ratings generate more revenue than others?
2. Do distributions of counts of rented films by category_id look the same for the 2 stores?
3. Do films that have a 'husband' (check the 'fulltext' field) generate the same amount of revenue as films that feature a 'dinosaur'?