Worksheet-5a in R

R Analytics

Worksheet for R Programming

Instructions:

- This is a group activity

- Use RStudio or the RStudio Cloud accomplish this worksheet.

- Create a new repo and name it as **R Analytics** with a folder *Worksheet#5* in the repo you created. Only one member will create this and other members need to fork the said repo.

- Save the R script as Rmd as *RWorksheet#5_group(lastnames of members).Rmd*, together with the knitted pdf in the *Worksheet#5*.

- Push the folder *Worksheet#5* to the new repo you have created.

Note: + Team members need to fork the repo of the member who created and pushed the repo. + Do not forget to give the repo to the class mayor together with the names of members + Accomplish this worksheet by answering the questions being asked and writing the code manually.

- In scraping a website, make sure to be polite all the times. Read the notes prior to scraping.

Extracting TV Shows Reviews

1. Each group needs to extract the top 50 tv shows in Imdb.com. It will include the rank, the title of the tv show, tv rating, the number of people who voted, the number of episodes, the year it was released.

It will also include the number of user reviews and the number of critic reviews, as well as the popularity rating for each tv shows.

2. From the 50 tv shows, select at least 5 tv shows to scrape 20 user reviews that will include the reviewer's name, date of reviewed, user rating, title of the review, the numbers for "is helpful" and "is not helpful", and text reviews.

3. Create a time series graph for the tv shows released by year. Which year has the most number of tv shows released?

Extracting Amazon Product Reviews

4. Select 5 categories from Amazon and select 30 products from each category.

5. Extract the price, description, ratings and reviews of each product.

6. Describe the data you have extracted.

7. What will be your use case for the data you have extracted?

8. Create graphs regarding the use case. And briefly explain it.

9. Graph the price and the ratings for each category. Use basic plotting functions and `ggplot2` package.

10. Rank the products of each category by price and ratings. Explain briefly.