

r/jobs data science exploration project overview and management

Monday, November 2, 2020 4:26 PM

NOTES:

1. <https://towardsdatascience.com/scraping-reddit-data-1c0af3040768>
 - a. How to use PRAW to scrape data from subreddit
 - b. There I can just export all these notes on project management
 - c. Grading skills - you are graded on the quality of your data story,
 - i. Not on fancy code
 - ii. Can you go through the steps of data analysis
 - iii. Setting out the research question
 - iv. Finding the data relevant to that
 - v. And exploring it correctly
 - vi. Shape of data, what rows mean, understanding the data, distributions, etc.
 - vii. How do I start to answer some of my questions using this data?
 - viii. Asking the right questions - can I sequence the questions correctly?
 - 1) Ex: at which periods of time has there been peaks in cases ? From an overall covid data set
 - 2) Can you systematically explore these in a notebook using markdown
 - 3) Break down your big question into sub

GUIDES:

1. Typology of Data Stories
 - a. <https://datajournalism.com/read/handbook/one/understanding-data/data-stories>
 - b. Always link to where the data can be found - for yourself and your readers, link to any codebooks, etc.
2. Refer to `use_of_data_in_stories` notebook for general outline
 - a. Need to subset data frame using row filtering
 - b. Then answer questions regarding those rows

TO-DO	Sub-Tasks	Due Date	Notes
<input checked="" type="checkbox"/> Set up PRAW using the data science article		11-2-2020	
<input checked="" type="checkbox"/> Figure out git pull/push to and from comm 318 folder and laptop		11-2-2020	
<input checked="" type="checkbox"/> Get all previous work down from comm 318 folder		11-3-2020	
<input type="checkbox"/> Create CSV with top 1000 posts of all time		11-11-2020	- This part was

from r/jobs <input type="checkbox"/> Then we can investigate each of the top 100 threads? <input checked="" type="checkbox"/> How can I hide the client secret key? Figure it out - hide it in separate module <input type="checkbox"/> Making a csv from observations			interesting
<input type="checkbox"/> Re-do data exploration in Jupyter for appropriate notebook		11-11-2020	
	<input type="checkbox"/> Write down any interesting observations		
	<input type="checkbox"/> Write down new questions to take the investigation in		
	<input type="checkbox"/> Compare trends across years		
	<input type="checkbox"/> Compare trends from march 2020 onwards, and most recently with biden elect		
<input type="checkbox"/> Think of appropriate visualizations			