

EPPS 7v81 Advanced Data Programming

The Project Proposal is responsible for 30% of the final grade. This proposal should demonstrate:

1. How the research question can be formulated and investigated in data science research context
2. How you use different programming techniques covered in class to solve a research problem
3. Applications of data science skills in four aspects:
 - a. Data collection
 - b. Data visualization
 - c. Data management
 - d. Data modeling

You will use this proposal as a framework you use to develop into the final report covering the product of data programming in terms of applications, actionable suggestions and/or other deliverables.

It should include but not be limited to the following elements:

Area/Topic:

An area or a topic you intend to develop expertise on after this course. Adopt a data collection method to get relevant and original data for this area or topic. Data sources could be social media, internet websites, API's or document databases.

Problem Statement

Outline what the problem you intend to solve. You may first start by asking an interesting question.

Data Programming:

Design a method or a mixed method approach in solving the problem. You may use multiple programming platforms/techniques but a majority of the programs or tools must be covered in class. If you want to use other methods, you must consult with instructor first.

Reference

Cite all resources and websites you are to use or adapt from.

Due date: 6/24/2020

Appendix: Data API and links (*)

Description	url
Open data network	http://www.opendatanetwork.com/
Awesome public datasets by disciplines and governments	https://github.com/caesar0301/awesome-public-datasets
FiveThirtyEight	https://github.com/fivethirtyeight/data
U.S. Government open data	http://www.data.gov/
Texas Government open data	https://data.texas.gov/
Seattle Government open data	https://data.seattle.gov/
Indian Government Open Data	https://data.gov.in/
Canadian Government Open Data	https://open.canada.ca/en/open-data
Worldbank open data	http://data.worldbank.org/
Kaggle data	https://www.kaggle.com/datasets
Data on Reddit	https://www.reddit.com/r/datasets
Links to datasets on Quora	https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public
Global health data	https://data.medicare.gov/ http://ghdx.healthdata.org/ihme_data
Instagram API	https://www.instagram.com/developer/
Reddit Data Extractor	https://www.reddit.com/r/reddit_data_extractor/
Twitter API	https://developer.twitter.com/en
Investment data service Quandl	https://www.quandl.com/
Zillow	https://www.zillow.com/research/data/
COVID19	https://covid19.healthdata.org/united-states-of-america

(*) Credits: Built upon Michael Freeman collections