

Practicum project topics can be defined using industry sponsored projects, faculty research projects, or projects that you have an interest in.

If you have employment with a company who has a project for you or you have secured a faculty project, be sure to discuss the project with your sponsor and determine if it will provide you the opportunity to demonstrate a range of the skills you have learned in the program. Remember, that as a data scientist, you have the knowledge and skills to enhance/extend a basic project such that it is a rich and deep data science project. You can meet the needs of your sponsor while also meeting the academic needs of the practicum.

Defining your own project can be equally rewarding. Is there a topic outside of your normal responsibilities that you are interested in? This is a great opportunity to merge your academic knowledge and skills with a project that sparks your interest. Be sure to think through your project (from start to finish) before you jump in. How can you scope the project such that it satisfies the academic requirements of the practicum? To help you get started with defining your practicum project, we are providing some simple guidance.

1. Identify an area of interest. Maybe you are interested in the spread of disease, the impact of the pandemic on oil and gas supply and pricing, or the geographic patterns among social media narrative about the pandemic. Formulate some very general questions.
2. Identify source(s) of data that are available that can help you address your question. Does that data have the depth and breadth to be able to address your question? Examine the potential datasets; are they large enough? Are they reliable and valid in their collection and construction? Are the fields sufficiently defined? Will the data satisfy your requirements?/
3. Refine your question based on the data you have available.

Here are various datasets that are available online. You may find others that suit your question(s) as well.

General Data Sets

- Google Public Data Explorer - <https://www.google.com/publicdata/directory>
- Registry of Open Data on AWS - <https://registry.opendata.aws/>
- Kaggle - <https://www.kaggle.com/datasets>
- UCI Machine Learning Repository - <https://archive.ics.uci.edu/ml/index.php>
- Data.gov - <https://www.data.gov/>
- GitHub - <https://github.com/awesomedata/awesome-public-datasets>
- Natural Language Processing Datasets - <https://github.com/niderhoff/nlp-datasets>
- freeCodeCamp Open Data - <https://github.com/freeCodeCamp/open-data>
- DBpedia - <https://wiki.dbpedia.org/>
- DataScience Dojo - <https://code.datasciencedojo.com/datasciencedojo/datasets>

Health Data Sets

- World Health Organization – <https://www.who.int/gho/database/en/>
- Unicef - <https://data.unicef.org/>

- HealthData.gov - <https://healthdata.gov/>
- Centers for Medicare and Medicaid Services - <https://www.cms.gov/>
- Big Cities Health Coalition - <https://www.bigcitieshealth.org/city-data/>
- DrugBank <https://www.drugbank.ca/>

Energy Data Sets

- OpenEI (open energy data initiative) - <https://openei.org/datasets/dataset>
- Oil and Gas Authority Open Data - <https://data-ogauthority.opendata.arcgis.com/>
- Data World - <https://data.world/datasets/oil>
- FracTracker - <https://www.fractracker.org/data/>

Other Topical Data Sets

- Global Development Data - World Bank – <https://data.worldbank.org/>
- Demographic and Economic Indicators - US Census Bureau - <https://www.census.gov/data.html>
- Journalism Data - FiveThirtyEight - <https://data.fivethirtyeight.com/>
- Business, review, and users - Yelp - <https://www.yelp.com/dataset>
- Financial and Economic – Quandl - <https://www.quandl.com/>