

DS 3000 – Topic Proposals

The first deliverable of the project is a topic proposal for your DS project. You should start by thinking about a real-world problem that interests you and also has clear connections to an area to which data science can contribute. Each group will submit two or three topic proposals (depending on whether your group consists of two or three members), and your professor and TAs will help evaluate which topic seems most promising for this project.

Here are some ideas for the project. You are by no means limited to these ideas and may propose anything else that would be similar to these examples in scope (requiring approximately 30-60 hours of work).

- ✓ Collect visual data from medical diagnostic tools (MRI, CT Scans, ultrasound, etc.) store image files in a file format, and retrieve the information in a manner useful for physician/patient. Retrieve data for descriptive and predictive analysis.
- ✓ Retrieve data from social media web APIs, e.g. Twitter, Facebook, Instagram, etc., clean the data, and create an alternative data schema for a particular file format and retrieve data for descriptive and predictive analysis.
- ✓ Collect publicly available economic data via web APIs offered by Yahoo Finance, US Bureau of Economic Analysis, and Quandl Resource Hub, etc., filter it for a particular industry/task, store in a file, and retrieve data for descriptive and predictive analysis.
- ✓ Collect publicly available data produced by manufacturing and production systems with the aim of increasing efficiency for a certain manufacturing process. Try Amazon Public Datasets, DBpedia, or World Bank for dataset sources.

A well-designed project is one that takes a relatively complex dataset from some source(s), cleans it, stores it in a file, and retrieves the data from the file. Your project should provide a description of the dataset both statistically and visually. All visualizations must be interpreted. **At least one modeling technique needs to be applied to the data for a classification or a prediction.** A collection of models need to be created using the technique as well as an evaluation of the models.

Please refer to the sample project papers posted on Blackboard (Final Project > Sample Projects).

What to submit

- ✓ Complete the following sections for each topic idea.
- ✓ Before submitting your proposals, delete this first page containing the instructions.
- ✓ Submit your completed document on Blackboard.

Topic Idea 1: <Insert title here>

1. Problem Statement

- ✓ Describe the problem you would like to tackle.
- ✓ What is the topic of your project?
- ✓ What do you want to learn about it?
- ✓ What are some keywords for your topic to use in dataset search?

2. Significance of the Problem

- ✓ Why is it important to tackle this problem in your project?
- ✓ In what ways could the insights from this project be useful?

3. Potential Datasets

- ✓ Look into some potential datasets using the online resources provided on Blackboard or other data sources you find online.
- ✓ Look through the datasets to identify any potential matches.
- ✓ The dataset should include at least 5-10 columns with at least 100-200 rows.
- ✓ Provide links to potential datasets here.
- ✓ If you are planning to scrape your own data from online resources, describe your data and online resources here. Provide links to the online pages/documents.

Topic Idea 2: <Insert title here>

1. Problem Statement

- ✓ Describe the problem you would like to tackle.
- ✓ What is the topic of your project?
- ✓ What do you want to learn about it?
- ✓ What are some keywords for your topic to use in dataset search?

2. Significance of the Problem

- ✓ Why is it important to tackle this problem in your project?
- ✓ In what ways could the insights from this project be useful?

3. Potential Datasets

- ✓ Look into some potential datasets using the online resources provided on Blackboard or other data sources you find online.
- ✓ Look through the datasets to identify any potential matches.
- ✓ The dataset should include at least 5-10 columns with at least 100-200 rows.
- ✓ Provide links to potential datasets here.
- ✓ If you are planning to scrape your own data from online resources, describe your data and online resources here. Provide links to the online pages/documents.

Topic Idea 3: <Insert title here>

1. Problem Statement

- ✓ Describe the problem you would like to tackle.
- ✓ What is the topic of your project?
- ✓ What do you want to learn about it?
- ✓ What are some keywords for your topic to use in dataset search?

2. Significance of the Problem

- ✓ Why is it important to tackle this problem in your project?
- ✓ In what ways could the insights from this project be useful?

3. Potential Datasets

- ✓ Look into some potential datasets using the online resources provided on Blackboard or other data sources you find online.
- ✓ Look through the datasets to identify any potential matches.
- ✓ The dataset should include at least 5-10 columns with at least 100-200 rows.
- ✓ Provide links to potential datasets here.
- ✓ If you are planning to scrape your own data from online resources, describe your data and online resources here. Provide links to the online pages/documents.