

Final Project



Presentation Due: Sunday, August 6th, 11:59 pm Pacific Time on Canvas.

Introduction: The Final project is a creative exploration of Data Science and its connection to art, medicine, finance, geography, biology, or any topic in the natural world. In this project, you will put together the techniques, skills, and insights you have learned in this course to answer questions that are important to you. It has two components:

1. **Final Presentation:** Submitting 20-minute video presentation.
2. **Final Report:** Submitting your project files, including codes and PowerPoint slides (in case you use slides as well)

Data Science Projects:

After a data scientist selects a dataset for analysis, they typically follow a structured workflow to extract insights and derive meaningful conclusions. Here are the general steps involved in the data analysis process:

1. Data Exploration: The data scientist starts by exploring the dataset to better understand its structure, contents, and quality. They examine the dimensions, variables, data types, and missing values present in the dataset.
2. Data Cleaning and Preprocessing: In this step, the data scientist handles missing data, removes duplicates, and resolves inconsistencies in the dataset. They may also perform data transformations, such as scaling, normalization, or encoding categorical variables, to prepare the data for further analysis. This step ensures that the dataset is in a suitable format for modeling and analysis.
3. Feature Engineering: Feature engineering involves creating new features or transforming existing ones to enhance the predictive power of the dataset. Data scientists may extract meaningful information from text, derive new variables from existing ones, or perform dimensionality reduction techniques to reduce the number of features while preserving relevant information.
4. Data Modeling and Analysis: This step involves applying appropriate machine learning or statistical techniques to analyze the dataset and extract insights you will learn later in machine learning courses.
5. Communication and Visualization: Data scientists communicate their findings and insights to stakeholders once the analysis is complete. They prepare visualizations, reports, or presentations to convey the results effectively. Clear and concise explanations and visual representations help stakeholders understand the implications and make informed decisions based on the analysis.

Final Project:

Codes: You will apply all the above processes in your final project through these specific steps:

1. You will work on a large data set (with more than 10k records) that you choose from UC Irvine Repository, Not Kaggle datasets, and Not Seaborn datasets.
2. Design 25 exciting questions based on your dataset, adhering to the following format:
 - (a) At least five questions should be solved using functions you define (such as lambda-map functions)
 - (b) For at least ten questions, you should utilize Pandas and NumPy.
 - (c) For at least five questions, you should create visualizations using Matplotlib and Seaborn.
 - (d) Design at least five questions that reveal valuable and exciting insights from your dataset using any techniques you like (Scipy, Pandas, Data Vis, etc..)

Video Presentation:

It is a 20-minute video presentation that could be PowerPoint slides along with your Python scripts, Videos, etc. Here are the main requirements:

1. The video should effectively communicate the key aspects of the project to the audience clearly and concisely.
2. Begin with a brief introduction that sets the project's context and objective, highlighting its significance and potential impact.
3. Break the project into smaller components or stages, highlighting the critical tasks involved.
4. Describe the data science techniques, algorithms, or libraries used to analyze and visualize the main questions (or maybe all the questions).
5. Ensure that you analyze the answer to each of your questions and effectively communicate them through visual aids, such as slides, charts, or diagrams, to enhance the presentation and make it more interesting.
6. For the conclusion, summarize the key points discussed in the presentation, emphasizing the importance of the project and its outcomes.
7. Remember to keep the video presentation focused, well-structured, and visually appealing.

Please note: In the end, you are free to choose the format of your presentation, anything you like; just make it informative, clear, and very interesting. You can be as creative as you want or simply join a Zoom or Teams meeting, share your screen, and record your presentation. Practice your delivery to ensure a confident and exciting presentation and ensure that your voice's quality is good.

Submission:

You will submit all your files: codes, video presentation, and slides into
Canvas-->Assignment-->Final Project.

Grading

The **Video Presentation** is 40 points and the **Codes** is 60 points.

Please feel free to message me or come to my office hour if you have any questions regarding the project.