**Capstone Project**

The capstone project involves an analysis of the students choosing relating to their field of interest or current industry, creating a visualization (static or interactive) using relevant data and presentation to their classmates. You may use any dataset you can find, if the data is publicly available, or you have permission from the data’s owner to work with it and share it. Datasets can also be selected from the sample project list. A short 2-3 paragraph proposal on why this dataset is relevant to your industry is required.

You may work with a team of up to 3-4 people, or individually. Each team member is responsible for understanding and being able to explain all portions of the project.

You should expect to share your work with classmates, your instructor, as well as your employer.

Your project should include the following guidelines:

* Describe your motivation for performing this analysis
* Describe your data sources
* Project includes multiple data transformation (example: converting columns to different data type)
* Project includes at least one statistical analysis and one graphic that supports your data.
* One graph that supports your conclusions.
* Each section of your project should be accompanied by an explanation. For example, a 1-2 paragraph explaining what is happening in that section or graph – so the reader understandings the story being told.
* Presentation 10-15 minutes.
* Presentation: did the audience come away with a clear understanding of why you undertook this project?
* Presentation: did the audience come away with a clear understanding of one insight or conclusion you’ve found?
* Delivered code and data should run without errors.
* Code should be hosted on github.

Sample Project datasets

* How does diabetes and other medical indicators impact patient length of stay within hospitals?
  + <https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008>
* Has NYC Vision zero made an impact on the safety of NYC? What kind of recommendations can be presented to NYC to improve the program. What seems to be working or not working?
  + <https://data.cityofnewyork.us/Public-Safety/Vision-Zero-View-Data/v7f4-yzyg>
* Citi Bike within NYC is one of the largest bike sharing systems in the US. Perform exploratory data analysis on the data set. Suppose you’d like to create a marketing campaign – who would you focus on and why?
  + <https://s3.amazonaws.com/tripdata/index.html>
* How has the number of marriages and divorces changed over time? Are people likely to get a divorce in 2019 vs. another year? Are there any variables that people who get a divorce or stay married have in common?
  + <https://ftp.cdc.gov/pub/health_Statistics/nchs/nvss/marriage-divorce/>
* The NYC school system is continually changing how teachers instruct students. Are there any patterns or insight that can be drawn from looking at students standardized test scores?
  + <https://data.cityofnewyork.us/Education/2014-15-To-2016-17-School-Level-NYC-Regents-Report/csps-2ne9>
* Analysis of college teacher salaries. How does your title, state, or college effect your pay? How does it vary on the state and college level?
  + <https://github.com/trendct/datasets/blob/master/0415-ct-professors/ct-professor-salaries.csv>
* What medical indicator shows a positive association with cardiovascular or other diseases?
  + <https://data.world/cdc/us-chronic-disease-indicators>

Students can also select their own passion projects. If there is a project you feel passionate about, and it directly relates to your current career path. Please discuss with the instructor for approval. Projects will be approved on a case-by-case basis; however, we welcome students to take ownership of the learning and explore and find value from meaningful datasets.

Sites such as:

* <https://opendata.cityofnewyork.us/>
* <https://data.world/>
* <https://archive.ics.uci.edu/ml/index.php>
* <https://www.kaggle.com/datasets>
* <https://www.data.gov/>

Offer almost endless amounts of meaningful datasets waiting to be explored. These websites act as essential tools in helping you find projects to explore during class. All projects should relate to your current line of work.

**Your final product should include:**

* Python Jupyter notebook
* Tableau dashboard
* 1-2-page paper

**Python:**

* Your Python analysis should be in a Jupyter notebook.
  + This analysis should be like the ones done previously
  + Include a research question – what is the goal of your analysis/project? What are you trying to understand?
  + Descriptive and inferential statistics regarding the data in your dataset
  + Data cleaning/transformations
  + Data visualizations to support your findings
  + A conclusion on your analysis

**Tableau**:

* Your project should also include a tableau dashboard
  + Your dashboard should show some visualizations from your data
    - It makes sense to show data that an administrator, office manager, etc. would think is important.

**Paper**

* You should write a 1-2-page paper, describing your analysis and findings. Think of this paper as: your company has asked you to investigate this dataset for information. This paper is a summary of what you did and what you found.
* Sections that should be included:
  + Introduction – what is the goal of your project? What are you investigating and why is it important?
  + Data sources – where did you find your data, how many rows/columns are in your data?
  + Methodologies – what tools/techniques did you use to analyze this data? Python libraries, types of visualizations, data cleaning, transformations, etc.
  + Findings and conclusions – what did you learn from the data? Was it what you expected?
  + References: list the resources you used to aid in your project.

**PowerPoint Presentation**

* A presentation should be developed in PowerPoint.
* The presentation should last about 10-15 mins.

**Capstone guidelines**

* Your project should be relevant to your current industry! This is a project you can share with your employer – make it count!
* Ask your self – are there any issues or ways I can help my organization grow or improve now? This should be the type of project you want to do!
* The choices are endless – it’s up to you to find a meaningful project and create an analysis.
* The best analytics professionals are the ones that can come up with questions that need to be solved – and \***try** to solve them!
  + \*not everything is solvable! Or at least – in a short timeframe!
* The instructor is at your disposal to assist you!

**Capstone Agenda**

* Selection and methodologies (4hrs) –
  + Overview of capstone
  + Data selection
  + Proposal
* Project Development (4hrs)
  + Work on portions of your project
* Project Development (4hrs)
  + Work on your project
  + A sample draft of your project should be uploaded to your github account by the end of this lesson
* Project Development (4hrs)
* Presentations of final projects (2-4hrs)