NYC DATA SCIENCE ACADEMY

NYC Data Science Bootcamp

Capstone Project

Introduction

You've drudged through proving you can analyze data numerically and visually. You've slaved through displaying your data collection and scraping abilities. You've labored through demonstrating your knowledge of both supervised and unsupervised machine learning techniques. You've conquered exhibiting both your descriptive and predictive analytical skills. Soon you'll face all these challenges in the realm of big data.

What We're Looking For

It's time to unleash all that you've learned during the bootcamp and start answering questions and constructing solutions as though you are a true data scientist. Data are your oysters, and R, Python, and machine learning are your shucks. So, one last time, what story will you tell?

For this project, we open the floor to you. You are now prepared to take the stage as a blossoming data scientist and flourish in an undertaking of your own aspiration. The primary task is to create a well-rounded research project of which you are proud. How you go about this is up to you. While this project is more free-form, we only ordain that you must still include both visual and numerical analyses, alongside some unsupervised and/or supervised machine learning techniques. The lens through which you present is up to you.

Successful projects will encompass, but are not limited to, the following:

- Submission in respect of the deadline.
- Background knowledge of dataset(s).
- Communication of motivation: why do we care?
- Research questions of interest: what do you want to find out?

- Answers to research questions: what have you uncovered?
- Presentation skills.
- Time management (not going over the allotted time).
- Ability to answer audience questions effectively and efficiently.
- Balance of complexity and simplicity.
- Explanation of future work: what would you do if given more time, data, etc.?
- Demonstration of EDA skills:
    - Numeric methodology.
    - Graphic methodology.
- Demonstration of machine learning skills:
    - Supervised methodology.
    - Unsupervised methodology.
- Ability to assess model weaknesses and identify improvements.
- Ability to manage a team workflow (if applicable).

The Details

Your project proposal declaration is due by the beginning of the Pulse Check on **Friday, March 10**. **No exceptions.** You must declare your project on the project proposal document and give a short background on some initial research questions in a sentence or two. These may change completely as you proceed with your analysis – this is ok.

This is either **an individual or team project** in respect to the final deliverable; you may decide to work alone or as a part of a group. We encourage collaboration and knowledge generation, so it is possible that individuals and/or groups may decide to merge and/or specialize after the proposal deadline.

All code, data, etc. used to generate your graphics and/or Shiny app and any slides, markdown files, etc. intended for your presentation are due to the project GitHub repository uniformly by **Tuesday, March 28 at 11:59pm**. **No exceptions.** Should you be working with partners, make sure that all partners contribute their code and slides to the repository.

You will be required to deliver a **10 - 20 minute presentation** dependent upon group size (up to 10 minutes for groups of size ≤ 2, up to 15 minutes for groups of size > 2 and size ≤ 4, up to 20 minutes for groups of size ≥ 5) and respond to any audience questions. Every team member must speak for a portion of the presentation. Time slots will be randomly assigned on this calendar, so all projects must be submitted on time.

An associated blog post will be due by **Friday, April 7 at 11:59pm**. **No exceptions.** Remember, this is a living and breathing document. You may continue to develop and edit your project far beyond the deadline, as no project will ever truly be complete. If working in a team, you may choose to co-author a single blog post describing your whole project, or submit individual blog posts highlighting your own personal project workflow (e.g., if your team specifically delineated responsibilities, etc.).

For inspiration, take a look at our previous students' blog posts (here's a link specifically to the Capstone category).

For any lingering questions, please do not hesitate to reach out; we are always here to help!

Good luck!