AI Boot Camp

Project 1, Week 1

Module 9 Day 1

Class Objectives

By the end of class, you will be able to:

1. List the requirements for Project 1.
2. Describe the purpose of your project and the problems you aim to solve.

Time to divide into **teams**!

Introduction to **Project 1**

Choosing a Project Track

This project gives you the ability to focus your efforts on a specific dataset, and therefore in a specific industry. Below are a few examples, but feel free to find your own!

Finance

Healthcare

Technology

Exploratory Data Analysis (EDA)

What is EDA?	The use of analytical tools and visualizations to explore a dataset, usually for the benefit of the analyst. EDA can start with strictly defined goals or be used to broadly explore a dataset for the first time. EDA is blanket term for the process of discovery using data analytics.
Who else utilizes EDA?	Everybody! Any industry that uses data benefits from EDA.

Project Goal

Use EDA to explore and summarize a dataset, detailing any interesting findings and next steps.

Private Investigator

EDA example.

01

Use aggregate crime data from different police precincts in a city to uncover patterns in criminal activity.

02

Most crime in New York City takes place in the summer.

Can you uncover similar patterns in your city?

03

What do your results suggest about how police should plan their patrols?

What do your results suggest about how best to distribute law enforcement resources over the calendar year?

Uber Rides and Weather

EDA example.

01

No one likes to walk in sub-zero temperatures or scorching heat. Do people use Uber more when the weather is uncomfortable?

02

Using **Uber ride data from Kaggle** and **climate data from the National Centers for Environmental Information**, find out if people take Ubers more during summer and winter, and if there are relationships between daily temperature and ride frequency.

03

What do the results tell you about surge-pricing strategies and commuter habits?

Requirements

1. Effectively use GitHub for version control. (10 points)
2. Create documentation for your project, including a well formatted README. (10 points)
3. Deliver strong analysis and conclusions. (30 points)
4. Build six to eight visualizations. (20 points)
5. Present findings in a slideshow presentation. (30 points)

Detailed requirements are available in Canvas

Project Tips

01

Start by finding a dataset. Take care to select one that you can import quickly! Using too much project time just to import your data could be a recipe for a rushed analysis.

02

Examine your dataset for basic information as a team. What columns are there? What unique values do they contain? Are any columns especially suitable for grouping? Are there any interesting questions you have based on the columns available?

03

Focusing on a few questions based on your initial exploration, divide the remaining EDA among team members. Everyone should create a few visualizations; remember, you don't need to present every visual the team creates!

Presentation Tips

01

Consider your classmates your audience, and avoid slides full of code! Your presentation should be on your dataset, not on your code.

02

Tell a story! Start with a contextual introduction to the dataset, briefly state the interesting questions you initially thought the data might answer, then use visualizations to detail what you found. Make sure to include any conclusions or follow up research you think is relevant!

03

Visualizations should be easy for the audience to understand on their own. If you feel the need to over explain a visualization in your presentation, consider editing it to be more intuitive.

Time to work with your teams!

Break

15 mins

Time to work with your teams!

Questions?

The End