**GADSI CAPSTONE**

[**Capstone, Part 2: Problem Statement + EDA**](https://git.generalassemb.ly/DSI-EAST-1/capstone/blob/master/part_02)

For Part 2, provide an overview of your approach to solving the problem you've chosen. Summarize your objectives, goals & success metrics, and any risks & assumptions. Outline your proposed methods and models, perform your initial EDA, and summarize the process. **Your data should be in hand by this point in the process!**

**Goal**: Describe your proposed approach and summarize your initial EDA in a code submission to your local instructor (submit a link to your repository [here](https://goo.gl/forms/W4cU06SjDmX1MMnd2)) **Due**: November 15, 2017 (Weds.)

[**Capstone, Part 3: Progress Report + Preliminary Findings**](https://git.generalassemb.ly/DSI-EAST-1/capstone/blob/master/part_03)

In Part 3, you'll create a progress report of your work in order to get feedback along the way. Describe your approach, initial results, and any setbacks or lessons learned so far. Your report should include updated visual and statistical analysis of your data. You’ll also meet with your local instructional team to get feedback on your results so far!

**Goal**: Discuss progress and setbacks, include visual and statistical analysis, review with instructor. Submit your progress report [here](https://goo.gl/forms/52JPjyuazVumztZM2) **Due**: November 29, 2017 (Wednesday)

[**Capstone, Part 4: Report Writeup + Technical Analysis**](https://git.generalassemb.ly/DSI-EAST-1/capstone/blob/master/part_04)

By now, you're ready to apply your modeling skills to make machine learning predictions. Your goal for Part 4 is to develop a technical document that can be shared among your peers.

Document your research with a summary, explaining your modeling approach as well as the strengths and weaknesses of any variables in the process. You should provide insight into your analysis, using best practices like cross validation or applicable prediction metrics.

**Goal**: Detailed report and code with a summary of your statistical analysis, model, and evaluation metrics. **Due**: December 18, 2017 (Monday)

[**Capstone, Part 5: Presentation + Recommendations**](https://git.generalassemb.ly/DSI-EAST-1/capstone/blob/master/part_05)

Whether during an interview or as part of a job, you will frequently have to present your findings to business partners and other interested parties - many of whom won't know anything about data science! That's why for Part 5, you'll create a presentation of your previous findings with a non-technical audience in mind. You should already have the analytical work complete, so now it's time to clean up and clarify your findings. Come up with a detailed slide deck or interactive demo that explains your data, visualizes your model, describes your approach, articulates strengths and weaknesses, and presents specific recommendations. Be prepared to explain and defend your model to an inquisitive audience!

**Goal**: Detailed presentation deck that relates your data, model, and findings to a non-technical audience. **Due**: December 20, 2017 (Wednesday)

APD INCIDENT REPORT DATABASE

[**Capstone, Part 2: Problem Statement + EDA**](https://git.generalassemb.ly/DSI-EAST-1/capstone/blob/master/part_02)

For Part 2, provide an overview of your approach to solving the problem you've chosen. Summarize your objectives, goals & success metrics, and any risks & assumptions. Outline your proposed methods and models, perform your initial EDA, and summarize the process. **Your data should be in hand by this point in the process!**

**Goal**: Describe your proposed approach and summarize your initial EDA in a code submission to your local instructor (submit a link to your repository [here](https://goo.gl/forms/W4cU06SjDmX1MMnd2)) **Due**: November 15, 2017 (Weds.)

Dataset:

<https://data.austintexas.gov/Public-Safety/2016-Annual-Crime-Data/8iue-zpf6>

Questions:

1) 2016 (12 month-36k rows way more complete)

2) tie in weather data?

3) predict crime type and location?

| **Column Name** | **Description** | **Type** |
| --- | --- | --- |
| GO Primary Key |  | Number |  |
| Council District |  | Number |  |
| GO Highest Offense Desc |  | Plain Text |  |
| Highest NIBRS/UCR Offense Description |  | Plain Text |  |
| GO Report Date |  | Plain Text |  |
| GO Location |  | Plain Text |  |
| Clearance Status |  | Plain Text |  |
| Clearance Date |  | Plain Text |  |
| GO District |  | Plain Text |  |
| GO Location Zip |  | Plain Text |  |
| GO Census Tract |  | Number |  |
| GO X Coordinate |  | Number |  |
| GO Y Coordinate |  | Number |  |

Goal : Illustrate map-based crime stats

Target

Next steps:

Dummify months column

Start model creation

Refine model

Add in other sources(inspections/etc?)

profit