

**Milestone 2: Exploratory Data Analysis**

**Deadline: September 28th 11:59 pm EST**

**Word from Us Content Creators:**

Hello Everyone! We hope you are doing well. Congratulations on completing your proposal for your team project. We really enjoyed reading about the awesome ideas the teams came up with. This milestone will act as the first step to diving into data science.

**What it Takes:**

When it comes to learning, persistence is key. In a real world project, change is inevitable. Strong teams adapt to change. As long as you keep trying and are willing to ask for help, you will be in good shape :)

**Objective:**

During Milestone 1, your team researched and identified viable datasets for the project. Exploring the dataset is an important step in the data science project lifecycle as this will help your team make informed decisions and help get acclimated to the dataset. Your team will be responsible for using techniques/concepts covered from Workshop 2 of bootcamp and beyond to explore your chosen datasets (eg: distribution of certain features, interesting statistics, basic data visualizations, etc). Relevant visualizations are **highly recommended** to be incorporated in your team’s analysis.

**Deliverables:**

The two major deliverables are the code and a writeup.

1. **Code:** Each team should make a **Github Repository titled with your team name** (eg: If your team name is “Team Buzz”, your Github Repository should be named “Team Buzz”). Your repository should contain the code you used to explore your dataset (preferably an interactive Python notebook such as Jupyter Notebooks or Google Colaboratory since output would be displayed in the same environment as your code). If you would like to put a Google Colaboratory notebook on Github, you can download the .ipynb file from Google Colaboratory (File → download .ipynb) and upload it to your team’s Github repository.
2. **Writeup:** The writeup should consist of an explanation on how your team’s dataset was explored and why it’s useful to your team project (1-2 pages). Some questions to consider in your writeup include (but not limited to):
   1. What was the most interesting insight you uncovered from exploring your dataset? (ie: most prominent aspect you noticed from your dataset)
   2. How did the exploration of the dataset influence your direction with your team project?

**Tutorial on setting up a Github Repository:** [**https://drive.google.com/file/d/1GrAFpkprT-6ZftIneSXBHaULSpRHT\_Lv/view?usp=sharing**](https://drive.google.com/file/d/1GrAFpkprT-6ZftIneSXBHaULSpRHT_Lv/view?usp=sharing)

**Link to Install Git:** <https://git-scm.com/downloads>

**Basic Tutorial on Exploratory Data Analysis with the Kaggle Housing dataset (inspiration for some ideas):** <https://www.kaggle.com/spscientist/a-simple-tutorial-on-exploratory-data-analysis>