

# PROFESSIONAL CERTIFICATE IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

## Module 11

### Practical Applications II

Office Hours with Viviana Márquez  
November 16, 2023


## AGENDA

- Review required activities Mod 11
- How to have a successful capstone meeting?
- Practical Applications II Project
- Questions


## AGENDA

- Review required activities Mod 11
- How to have a successful capstone meeting?
- Practical Applications II Project
- Questions

## Required Activities for Module 11

- Practical Application Assignment 11.1 
- Capstone Project 11.1: Initial Question and Data

## AGENDA

-  Review required activities Mod 11
- How to have a successful capstone meeting?
- Practical Applications II Project
- Questions

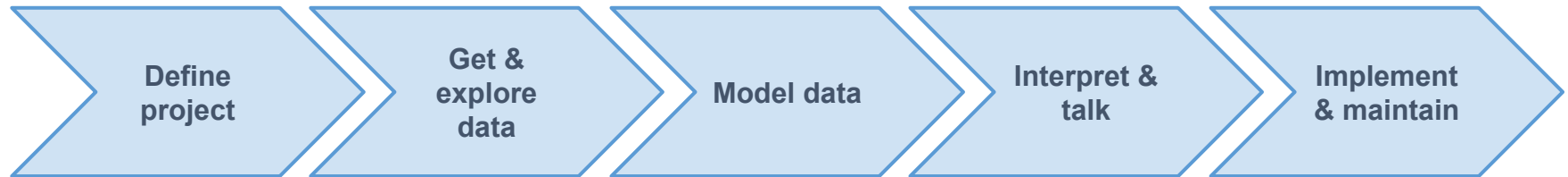
## CAPSTONE PROJECT

- **Deliverables**
  - A predictive model using supervised or unsupervised learning techniques on a dataset of your choosing
  - A technical write-up in Jupyter Notebook posted on your GitHub repository
  - An non-technical README describing your findings posted on your GitHub repository

## CAPSTONE PROJECT

- More info and examples:
  - [https://classroom.emeritus.org/courses/3424/pages/capstone-project-overview?module\\_item\\_id=1170722](https://classroom.emeritus.org/courses/3424/pages/capstone-project-overview?module_item_id=1170722)

## The Machine Learning pipeline



### Define project

- Specify business problem
- Acquire domain knowledge

### Get and explore data

- Find appropriate data
- Exploratory Data Analysis
- Clean and pre-process data
- Feature engineering

### Model data

- Determine ML task
- Build candidate models
- Select model based on performance metrics

### Interpret & talk

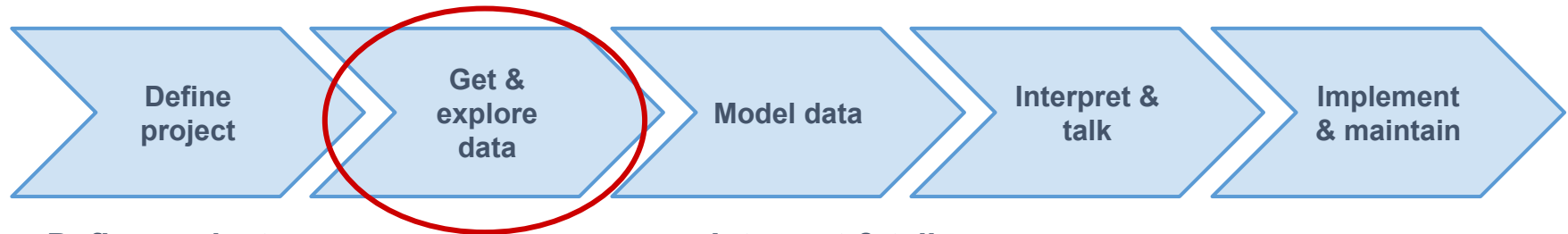
- Interpret model
- Communicate model insights

### Implement & maintain

- Set up function to predict on new data
- Document process
- Monitor and maintain model



## The Machine Learning pipeline



### Define project

- Specify business problem
- Acquire domain knowledge

### Get and explore data

- Find appropriate data
- Exploratory Data Analysis
- Clean and pre-process data
- Feature engineering

### Model data

- Determine ML task
- Build candidate models
- Select model based on performance metrics

### Interpret & talk

- Interpret model
- Communicate model insights

### Implement & maintain

- Set up function to predict on new data
- Document process
- Monitor and maintain model

## CAPSTONE PROJECT

- Where to get data?

### Open data repositories:

[OpenML.org](https://openml.org) (<https://openml.org>)  
[Kaggle.com](https://kaggle.com/datasets) (<https://kaggle.com/datasets>)  
[PapersWithCode.com](https://paperswithcode.com/datasets) (<https://paperswithcode.com/datasets>)  
UC Irvine Machine Learning Repository (<https://archive.ics.uci.edu/ml>)  
Amazon's AWS datasets (<https://registry.opendata.aws>)  
TensorFlow datasets (<https://tensorflow.org/datasets>)  
Google's data search engine: (<https://datasetsearch.research.google.com/>)

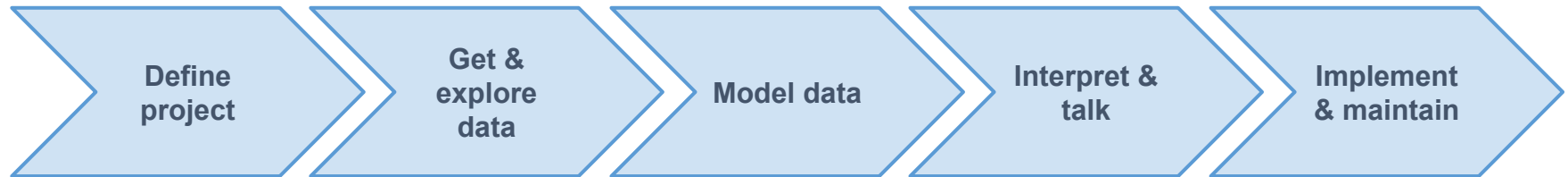
### Meta portals and other pages listing datasets:

[DataPortals.org](https://dataportals.org/) (<https://dataportals.org/>)  
[OpenDataMonitor.eu](https://opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex) (<https://opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex>)  
Wikipedia's list of machine learning datasets  
([https://en.wikipedia.org/wiki/List\\_of\\_datasets\\_for\\_machine-learning\\_research](https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research))  
Quora's list (<https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>)  
Reddit's dataset (<https://www.reddit.com/r/datasets>)  
GitHub \* (<https://github.com/>)

### Location-specific:

San Francisco Open Data (<https://datasf.org/opendata/>)  
NYC Open Data (<https://opendata.cityofnewyork.us/>)  
You can also google "open data + location" to get data about your desired location

## The Machine Learning pipeline



### Define project

- Specify business problem
- Acquire domain knowledge

### Get and explore data

- Find appropriate data
- Exploratory Data Analysis
- Clean and pre-process data
- Feature engineering

### Model data

- Determine ML task
- Build candidate models
- Select model based on performance metrics

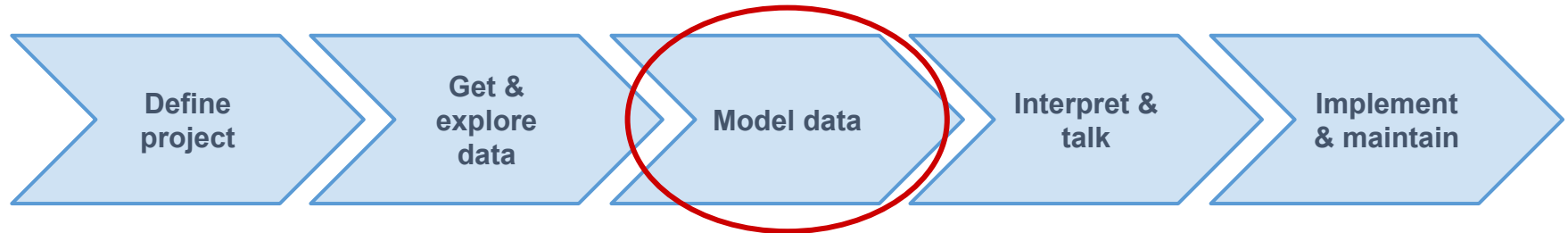
### Interpret & talk

- Interpret model
- Communicate model insights

### Implement & maintain

- Set up function to predict on new data
- Document process
- Monitor and maintain model

## The Machine Learning pipeline



### Define project

- Specify business problem
- Acquire domain knowledge

### Get and explore data

- Find appropriate data
- Exploratory Data Analysis
- Clean and pre-process data
- Feature engineering

### Model data

- Determine ML task
- Build candidate models
- Select model based on performance metrics

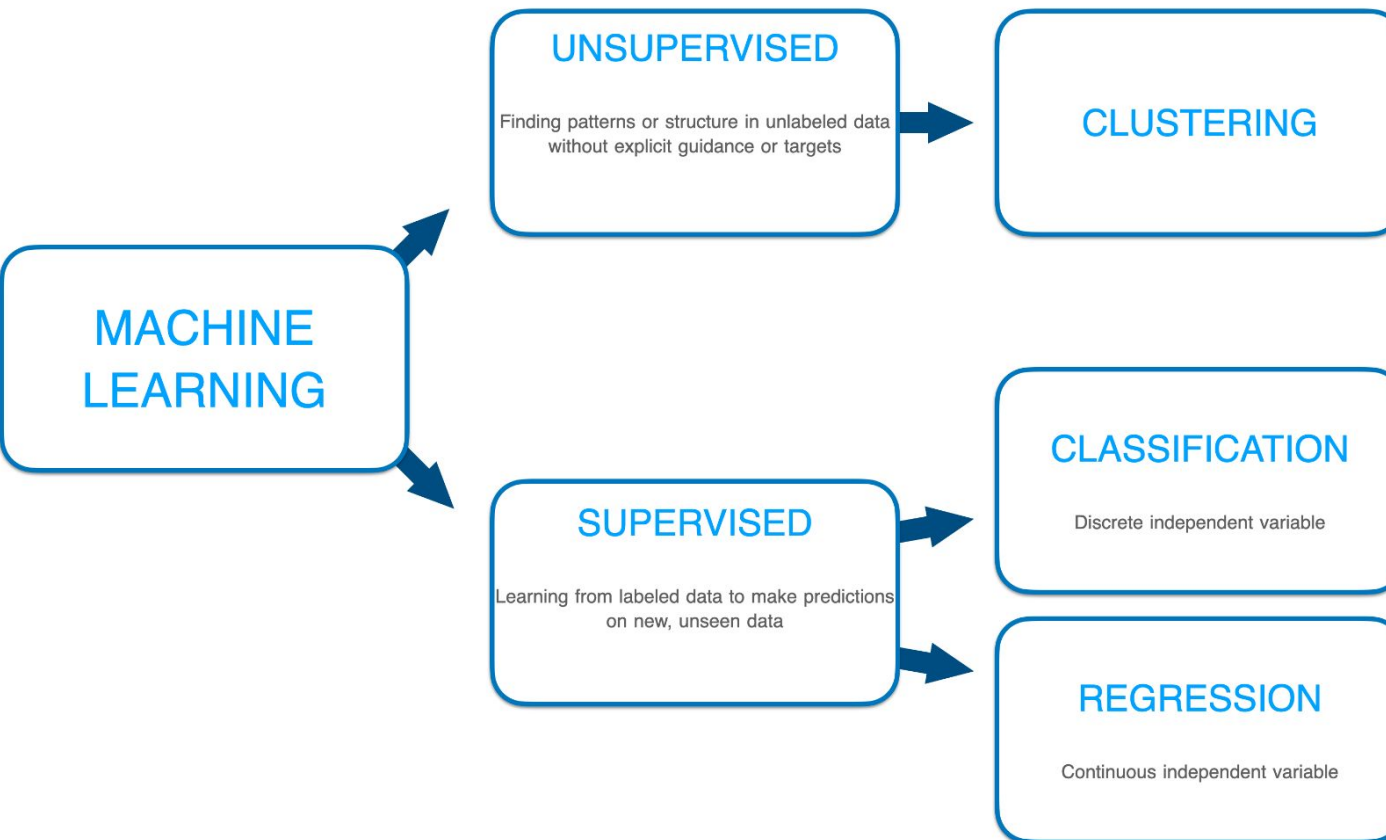
### Interpret & talk

- Interpret model
- Communicate model insights

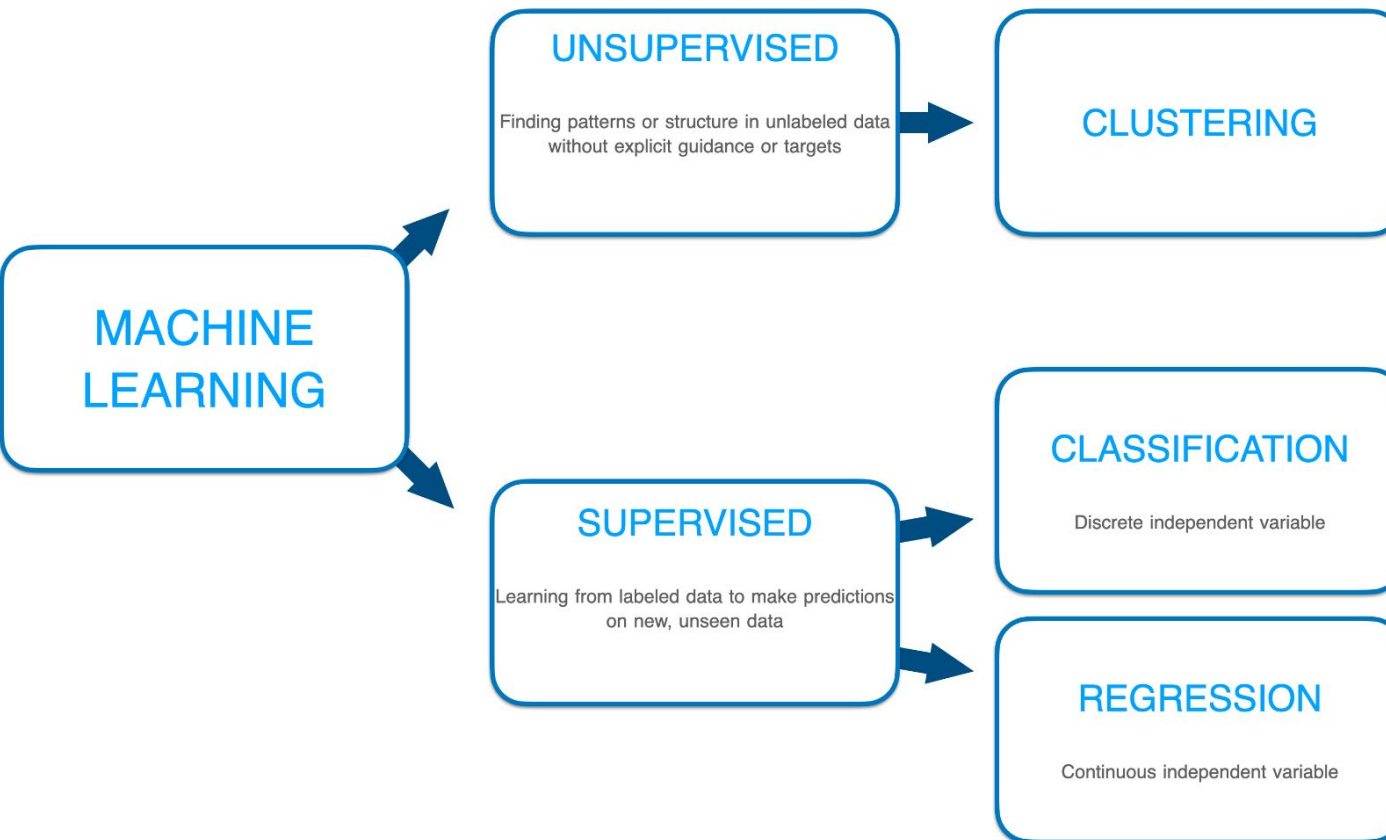
### Implement & maintain

- Set up function to predict on new data
- Document process
- Monitor and maintain model

## Do we have labels? Is my target variable discrete?



## Do we have labels? Is my target variable discrete?



### Clustering:

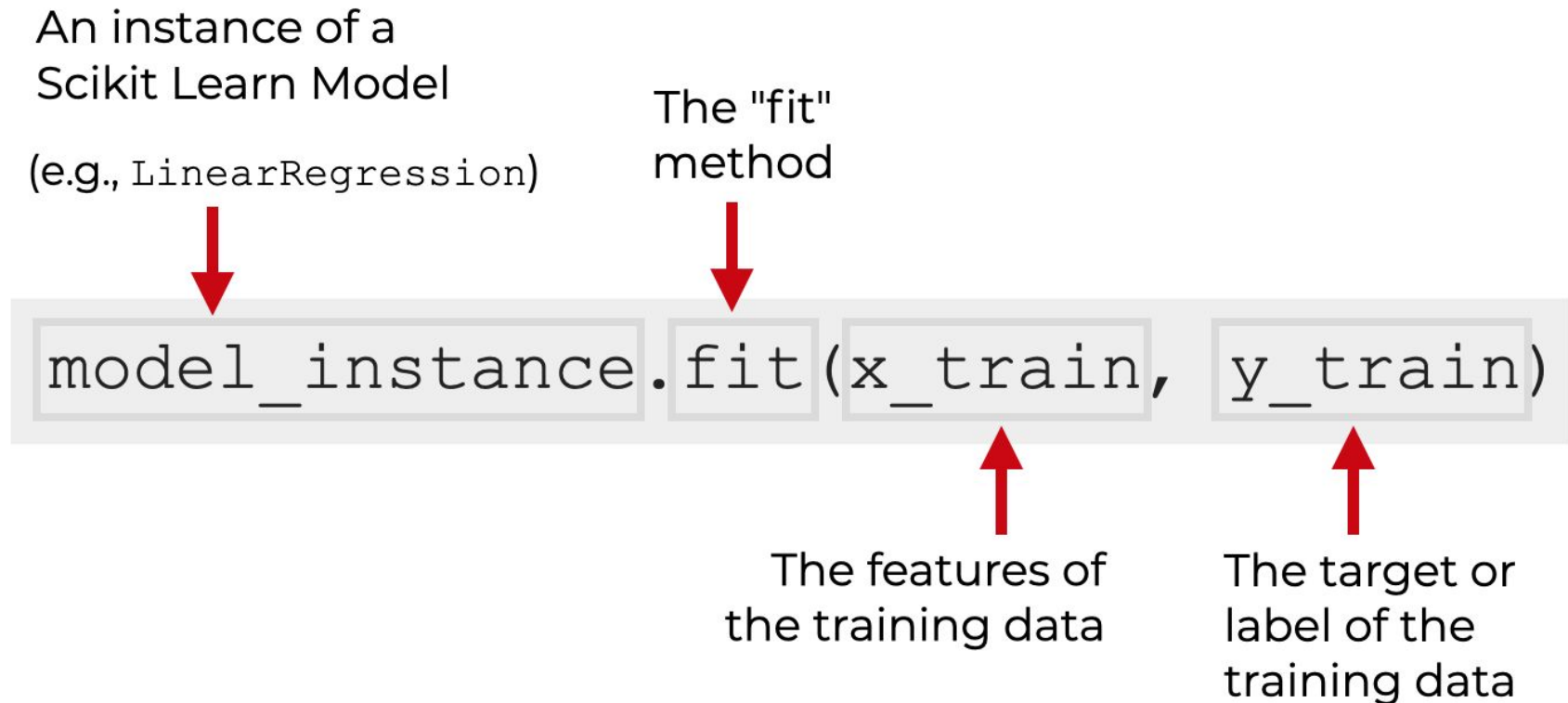
- K-Means Clustering
- Hierarchical Clustering
- DBSCAN
- Gaussian Mixture Models (GMM)
- Spectral Clustering

### Classification:

- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machines (SVM)
- Naive Bayes
- K-Nearest Neighbors (KNN)
- Neural Networks
- Gradient Boosting Machines (GBM)
- AdaBoost (Adaptive Boosting)

### Regression:

- Linear Regression
- Polynomial Regression
- Ridge Regression
- Lasso Regression
- Elastic Net Regression
- Support Vector Regression (SVR)
- Decision Tree Regression
- Random Forest Regression
- Gradient Boosting Regression
- Neural Networks



## Scikit Learn Unsupervised Learning

- [https://scikit-learn.org/stable/unsupervised\\_learning.html](https://scikit-learn.org/stable/unsupervised_learning.html)

## Scikit Learn Supervised Learning

- [https://scikit-learn.org/stable/supervised\\_learning.html](https://scikit-learn.org/stable/supervised_learning.html)

## CAPSTONE PROJECT

- **Roadmap**
  - **Module 11**  
Define your problem statement and develop a prospectus of the project
  - **Modules 12 to 15**  
First 1:1 With Your Learning Facilitator
  - **Module 17**  
Problem Statement
  - **Module 20**  
Initial Report and EDA
  - **Modules 21 to 23**  
Second 1:1 With Your Learning Facilitator
  - **Module 24**  
Final Analysis and Report



## CAPSTONE PROJECT

- **Schedule a 1:1 with your Learning Facilitator**


- **Viviana Marquez (Section A):**  
<https://calendly.com/vivianamarquez/bh-pcmlai-23-08-capstone-consultation-1?month=2023-11>
- **Matilde D'Amelio (Section B):**  
<https://calendly.com/matilde-damelio/bh-pcmlai-23-08-capstone-consultation-session-1?month=2023-11>
- **Muhammad Chaudhry (Section C):**  
<https://calendly.com/ali8/bh-pcmlai-23-08-capstone-consultation-1?month=2023-11>
- **Jessica Cervi (Section D):**  
<https://calendly.com/jessicacervi/bh-pcmlai-23-08-capstone-consultation-1>


- **Link:**


[https://classroom.emeritus.org/courses/3424/discussion\\_topics/304577](https://classroom.emeritus.org/courses/3424/discussion_topics/304577)


## CANVAS


Berkeley  
Haas


Account

Dashboard

Courses

Calendar

Inbox

Help

≡ Professional Certificate in Machine Learning and Artificial I... > Modules

63 Student View

Home

Modules

Live Sessions

Announcements

People

Grades

Assignment Extension

Q&A


Support

Collapse All


View Progress

+ Module


▸ Program Orientation

Complete All Items  + ⋮

▸ Pre-program Learner Expectations

Complete All Items  + ⋮

▸ Python Refresher

Complete All Items  + ⋮

Know your section!

Ask for help!

## CAPSTONE PROJECT

- **How to have a great 1:1?**
    - Book as soon as possible!
    - Come prepared to the office hour
      - Have datasets, examples, ideas, questions
    - If possible, share what you have so far with your learning facilitator beforehand
    - Only one 30 min consultation per learner in Mod12-Mod15
- 29 Nov 2023 – 9 Jan 2024 (Excluding break weeks– 12/20/2023- 1/2/2024)



## CAPSTONE PROJECT

- If you want to go the extra mile...



Dash  
byplotly



Linked 



 Medium



github:pages



## AGENDA

-  Review required activities Mod 11
-  How to have a successful capstone meeting?
- Practical Applications II Project
- Questions

## PRACTICAL APPLICATIONS II PROJECT

- Assignment:  
[https://classroom.emeritus.org/courses/3424/assignments/171432?module\\_item\\_id=1170720](https://classroom.emeritus.org/courses/3424/assignments/171432?module_item_id=1170720)

### Overview:

In this application, you will explore a dataset from kaggle. The original dataset contained information on 3 million used cars. The provided dataset contains information on 426K cars to ensure speed of processing. Your goal is to understand what factors make a car more or less expensive. As a result of your analysis, you should provide clear recommendations to your client -- a used car dealership -- as to what consumers value in a used car.

## PRACTICAL APPLICATIONS II PROJECT

- If you want to go the extra mile...



Dash  
byplotly



Linked 

 Medium



github:pages



## PRACTICAL APPLICATIONS II PROJECT

- If you want to go the extra mile...








## PRACTICAL APPLICATIONS II PROJECT

- **If you want to go the extra mile...**
- GitHub Pages is a static site hosting service offered by GitHub. It allows users to transform their GitHub repositories into websites
- By default, the URL of your GitHub Pages site will be in the format `username.github.io/repository-name`. However, you can also set up a custom domain if you want a more professional or personalized URL
- GitHub Pages is free for public repositories, but there are some limitations
- Documentation: <https://docs.github.com/en/pages/getting-started-with-github-pages/creating-a-github-pages-site>
- Examples: <https://github.com/collections/github-pages-examples>







## AGENDA

-  Review required activities Mod 11
-  How to have a successful capstone meeting?
-  Practical Applications II Project
- Questions

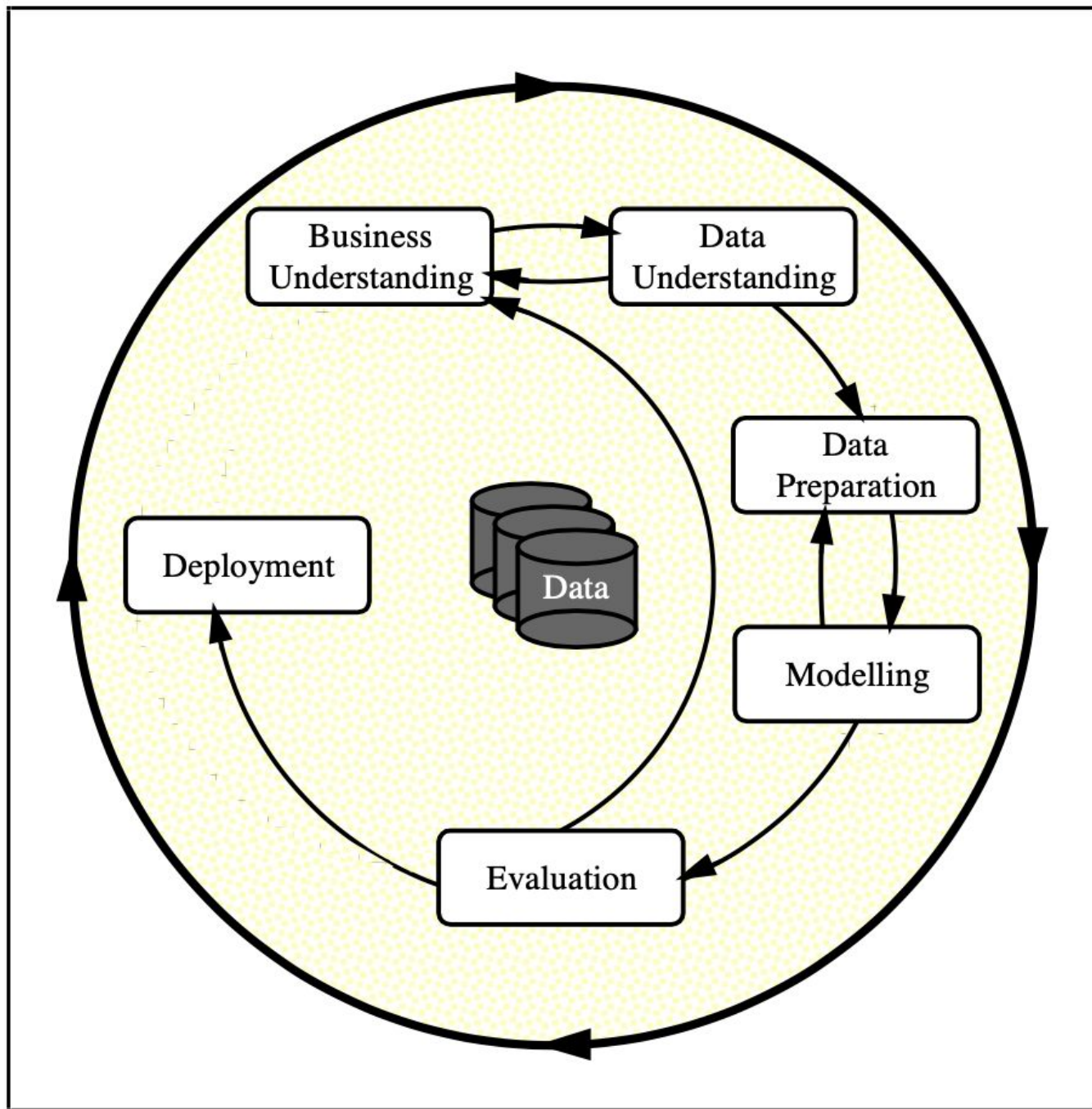
## QUESTIONS?



## AGENDA

-  Review required activities Mod 11
-  How to have a successful capstone meeting?
-  Practical Applications II Project
-  Questions

# APPENDIX



## Content review Module 5: Practical Applications I

When to use `.py` and `.ipynb`?

- **`.py` Files**

- Production Code (they can be integrated into larger applications or workflows)
- Large-Scale Projects (better modularity and code reusability)
- Automation
- Software development

- **`.ipynb` Files**

- Exploratory Data Analysis (EDA)
- Documentation and Training
- Presentations and Reporting
- Collaborative Work

## Content review Module 5: Practical Applications I

- GitHub authentication errors:  
<https://ginnyfahs.medium.com/github-error-authentication-failed-from-command-line-3a545bfd0ca8>
- GitHub green squares:  
<https://vivianamarquez.medium.com/you-probably-dont-know-you-have-this-problem-but-you-ll-want-to-fix-it-github-contributions-bf54c1878f9f>