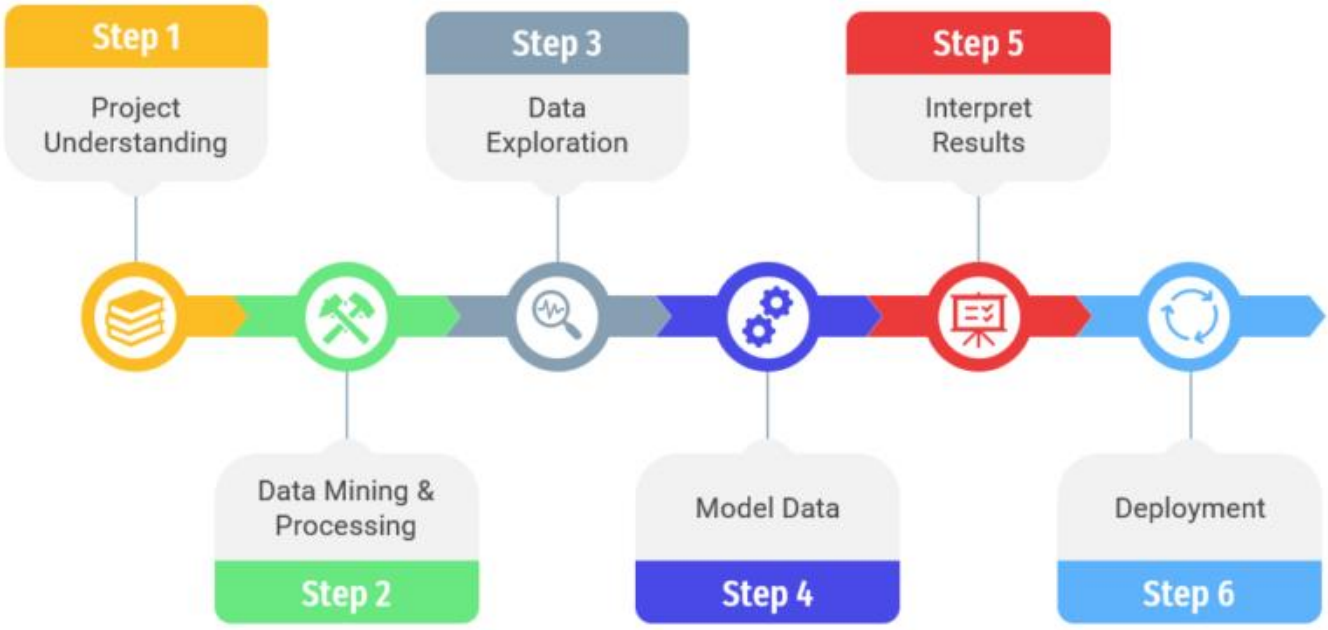
FINAL PROJECT

And here we are at the end of the road, let's make it the best end 😀

Project Overview:

- **Data is meaningless until its conversion into valuable information. And that's the power of Data Science, it helps businesses monitor, manage, and collect performance measures to improve decision-making across the organization.**
- **In this project you will solve a real-life problem using all what you have learned.**

Project Steps



Dataset (optional):

- **you can search for a dataset in any domain you interested in and prepare a proposal to be reviewed and find if it's suitable for this project (you can search for your dataset here)**

  Note: The data you chose must be the same as the recommended data in terms of
  1. **Size (Number of columns and rows).**
  2. **Quality (Don't choose a cleaned data because that's your mission).**
  3. **Simulates a real-time problem.**
  4. **Your proposal must show the description of your dataset and prove that it's not a clean one .**

- **Also , There are more than 10 datasets that we recommend for you, choose any of them and start your project!**
- **You will loss 5 grades by default if you didn't search for a dataset for your project and just pickup one from our dataset.**

Project criteria (The standard by which we judge):

1- Code Functionality

Does the code work?	**The whole code must be functional and produces no errors when run.** **-The given code is sufficient to reproduce the results described.**
Does the project use good coding practices?	-The code makes use of functions to avoid repetitive code. -The code contains descriptive comments and variable names, making it easy to read.

2- Quality of Analysis

Is the question clearly posed?	**-The project clearly states one or more questions, then addresses those questions in the rest of the analysis.**

3- Data Cleaning Phase

Is the data cleaning well documented?	**-The project documents any changes that were made to clean the data, such as merging multiple files, handling missing values, etc.**

4- Exploration Phase

Is the data explored in many ways?	**-The project investigates the stated question(s) from multiple angles.** **-At least six variables are investigated using both univariate and bivariate analysis**
Does a variety of relevant visualizations and statistical summaries exist?	-The project's visualizations are varied and show multiple comparisons and trends. Relevant statistics are computed throughout the analysis when an inference is made about the data. -At least five kinds of plots should be included as part of the explorations.

5- Visualization

Does the data visualize use appropriate plots and parameter choices?	**Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted.**

6- Feature Engineering and Selection

Create new features	**At least one new feature is implemented.**
Features selection	Using one of the following techniques: - Filter Methods - Wrapper Methods - Embedded Methods

7- Pick and Tune an Algorithm

Pick an algorithm	**At least three different algorithms are attempted, and their performance is compared, with the best performing one used in the final analysis.**
Discuss parameter tuning and its importance.	Response addresses what it means to perform parameter tuning and why it is important.
Tune the algorithm	At least one important parameter tuned with any of technique for example: (GridSearchCV)

8- Validate and Evaluate

Usage of Evaluation Metrics	**At least two appropriate metrics are used to evaluate algorithm performance (e.g., precision and recall)**
Discuss validation and its importance.	Response addresses what validation is and why it is important.
Algorithm Performance	When tester.py is used to evaluate performance, precision and recall are both at least 0.3.

9- Project deployment

Deploy model on web app	**You must provide a link with your web application**

Important Notes for each student for submission:

1- Must submit a notebook including all his work.
2- Must submit the app.py file that contains the deployment of the project.
3- Must submit the dataset, the saved transformers, and the saved model.
4- Must submit a video that explain everything done in the project.
5- Must create a repo for his project and mention the main Epsilon AI repo.

Important Notes for each instructor for submission:

1- Must follow the grading template.
2- Must hold a discussion with students.

"Keep it Simple, BUT significant 🧐"