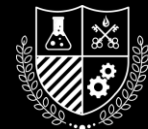


Creating your own project



Mass Street
University
PER EDUCATIONEM PROGRESSUS

Capstone project

- Combine what you have learned into one project to show your data science skills and problem-solving skills.
- Must involve machine learning and some form of data.
- Must create visualizations and tell a data story with your project.



What to do?

- Solve a problem that actually exists in society or business.
- Data at:
 - <https://github.com/awesomedata/awesome-public-datasets>
 - <https://github.com/datasets/awesome-data>
 - <https://github.com/academic/awesome-datascience#data-sets>
- You don't have to use dirty data.



Examples

- Modeling flight paths and figuring out when and where cancelations will happen, Trading stocks, traffic planning, police strategy, better gaming AI, etc.



Advice

- Good data is hard to find and usually costs a lot of money.
- GPUs are a must for deep learning. Training time is several times faster on GPU vs CPU.
- You can use cloud computing methods such as AWS or Google Colab instead of buying your own GPU. Colab also offers TPU support.

Visualizations and Testing

- Make sure visualizations are clear. Show it to a lay person that you know and explain your project. If they don't understand your visualizations, then they are bad.
- Make your code reproducible and testable for other people. If other people can't follow what you did and get the same results people won't trust or listen to you.

