Final Project Guidelines

Building End-to-End Data science Project

Identify and solve problems with **data**, to deliver value.

DO & DON’T

- * ✕ DON’T add a project that uses a generic dataset and models to do a generic task (ex. sentiment classification on news reviews or predicting mortality on the Titanic). These are great to learn from initially but not great to showcase for a job. ✓ DO pick a unique dataset (or go web-scrape your own!) and apply simple → complex models unique to your task
- * ✕ DON’T merely dump a notebook full of code. ✓ Provide necessary details when posting your project (objectives, highlight and takeaways). This is also the place to showcase your product/business sense which is key for a data scientist.
- * ✕ DON’T just apply a set of complex models on random datasets ✓ DO work on projects in industries you’re applying to with a class of models that make sense. The familiarity in the space will come across as major advantage. (Uses salesforce app exchange, shopify app store, zendesk app marketplace ,stripe partners, slack etc to get B2B ideas) *
- ✕ DON’T just focus on modeling in your projects. ✓ DO incorporate full-stack components (streamlit , databases, apis, devops, etc.) to show you can create a product and not just run code on a jupyter notebook.

TIPS

- Use Real DataTry to do something with real data rather than Kaggle or other pre-cleaned data. Data cleaning, prep and transformation is a real part of any data job.
- Scrape Your Own DataBeautifulSoup or Scrapy in Python are absurdly easy to use. If you can see it on the web, you can get it!
- Ask for Data from your oragnization
- Use Publicly Accessible APIs
- Pick Interesting Data Best portfolio projects are less about doing fancy modeling and more about working with interesting data.
- Pick Something You're Curious About, Not Something You Hope Will Be ImpressivePick an Analysis That is Interesting Regardless of What You Find
- Perfect the Visuals
- Make Your Data Interactiveuse Plotly Maxbox and Dash https://app.flourish.studio/templates
- Put the Code on Github Comment it and organize it well. Try to make the whole exercise, from downloading data to the visualizations
- Productionize your Analysis or APP You get a lot of bonus credit for productionizing any model or data product. People thought our baseball analysis was cool, but they were really impressed by our Twitterbot that made predictions in real time. Similarly, Slackbots, Facebook chat bots, and Reddit bots use flask ,Dash , Streamlit

Practical Guide

We will Build End to End Machine learning Project with PyCaret, Deploy with FastAPI , Streamlit & Heroku