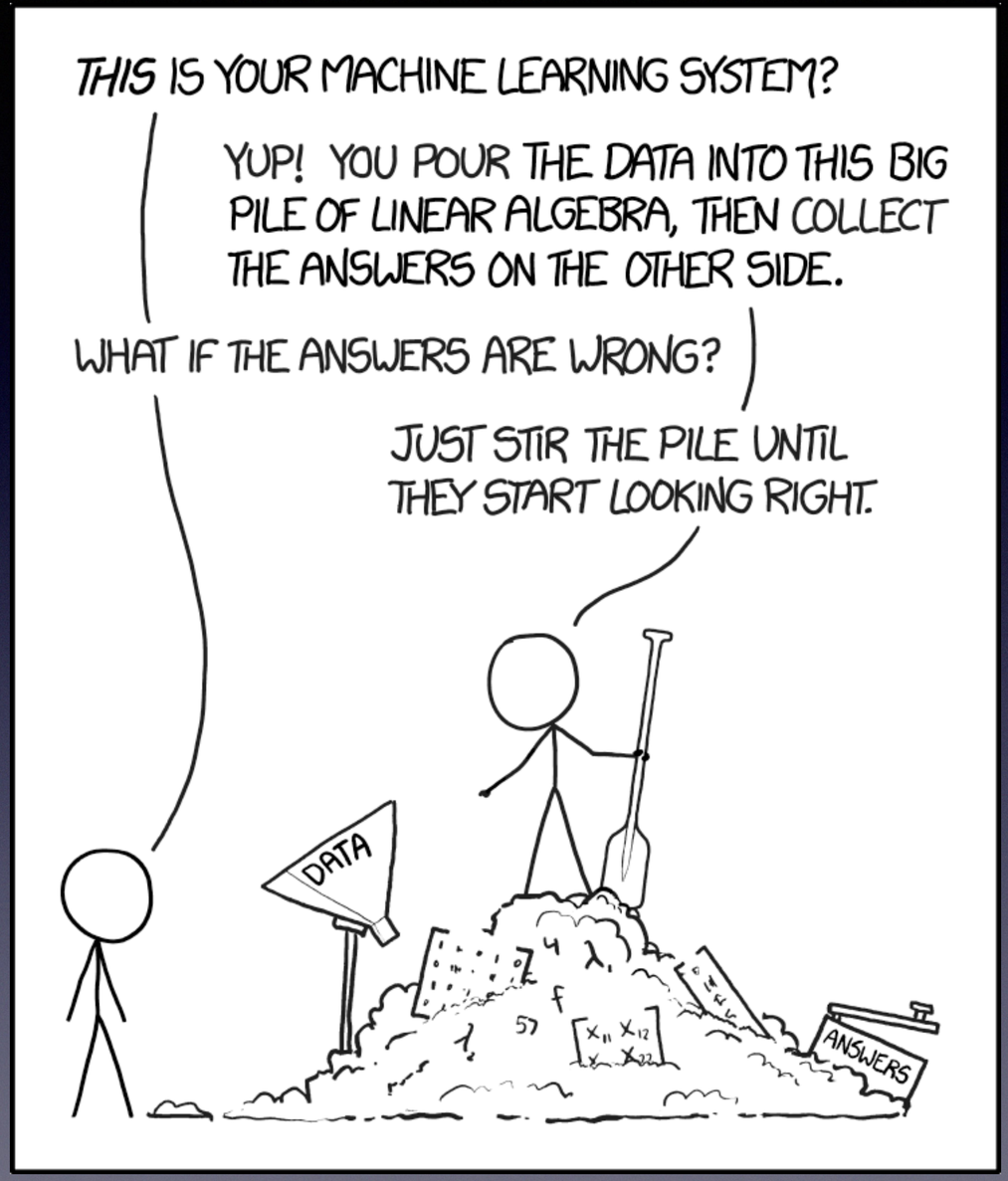
Machine Learning Workflows

Brendan Herger, https://www.hergertarian.com/

Slides: TODO



https://xkcd.com/1838/

Intro
Workflow
Demo & pro-tips

Intro

Why?

Why?

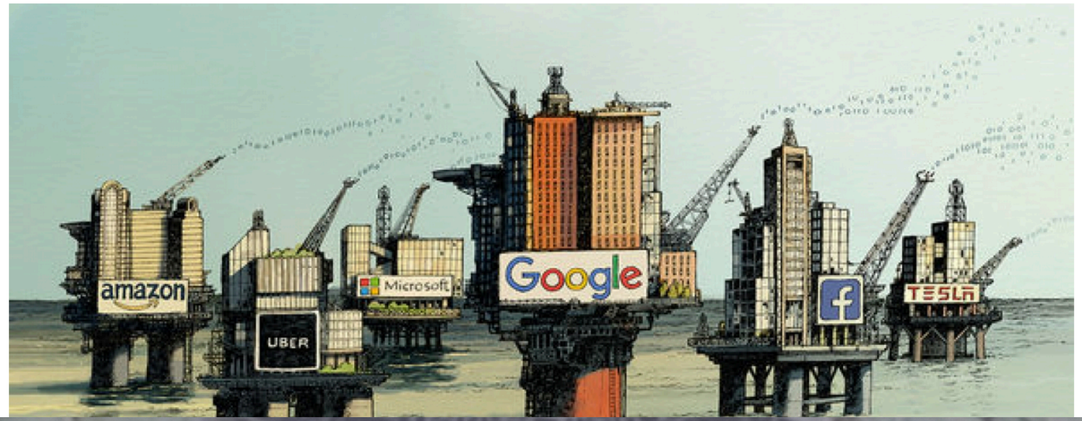
The Economist

Topics Current edition More Subscribe Log i Mana

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules



Tools

- **Pandas:** Tabular data store
- **Scikit-Learn:** Data science & machine learning library

Terms

- **DataFrame:** Tabular data store (i.e. like a SQL or excel table)
- **Transformer:** Accepts data, transforms it (e.g. fills null values)
- **Machine learning model:** Accepts data, makes a numerical prediction (regression) or categorical prediction (classification)
- **Pipeline:** A list of transformers / ML models

Workflow

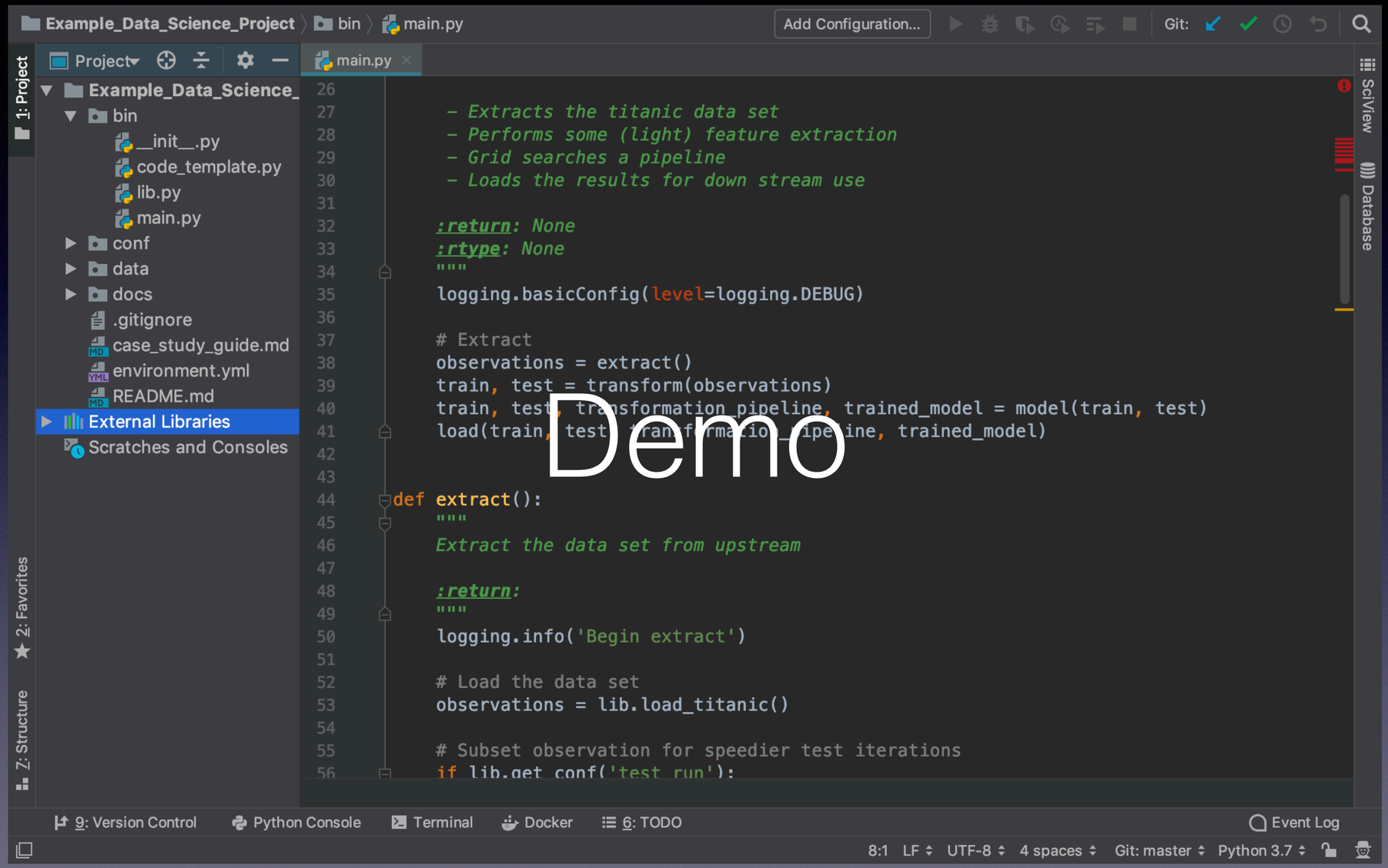
ET(M)L

- **Extract**
- **Transform**
- **(Model)**
- **Load**

ET(M)L

- **Extract:** Pull anything you need from the world
- **Transform:** Feature engineering
- **(Model):** Model the data
- **Load:** Store anything the world needs from you

Demo & pro-tips



https://github.com/bjherger/Example_Data_Science_Project

Pro tips

- **Consistency:** If your process is an assembly line, you can write robust code quickly. Create a repo with your project ‘template’
- **Iteration:** Set time aside to iterating on your workflow, not just the task at hand
- **Accept inconsistency:** Know when and why to break the rules

Intro
Workflow
Demo & pro-tips

Thanks!

Brendan Herger, https://www.hergertarian.com/
Slides: TODO