

# Machine Learning Workflows

Brendan Herger, <https://www.hergertarian.com/>  
Slides: <https://bit.ly/2WWPs79>

**PLEASE STAND BY**



Convo Careers | Convoy

https://convoy.com/convoy-careers/

Carriers Shippers About Convoy Careers Blog Contact Us Sign Up Log In

JOIN A TEAM THAT'S MOVING THE WORLD FORWARD

View Featured Job Openings

## WELCOME TO CONVOY

Our passionate team is committed to our mission of transporting the world with endless capacity and zero waste. Founded in 2015, Convoy has quickly grown to over

C Convoy Careers | Convoy

C Convoy Careers | Convoy

Convoy, Flexport automating th

https://www.cnbc.com/2019/05/15/convoy-flexport-automating-the-shipping-industry.html

SIGN IN PRO WATCHLIST MAKE IT ↗  
MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV USA • INTL SEARCH QUOTES

Emerging tech for emerging markets. Nasdaq REWRITE TOMORROW → MARKET TECHNOLOGIST

CNBC DISRUPTOR 50 Nasdaq

# The next Uber will come from the \$1.2 trillion global truck and shipping industry

PUBLISHED WED, MAY 15 2019 • 8:32 AM EDT | UPDATED WED, MAY 15 2019 • 4:11 PM EDT

Joel Dreyfuss

SHARE f t in e ...

---

KEY POINTS • The U.S. trucking market generates \$260 billion in revenue, 20% of a \$1.2 trillion global market

Nasdaq

WSJ Digital Freight Startup Convoy Convoy, Flexport automating th

<https://www.wsj.com/articles/digital-freight-startup-convoy-raises-185-million-surpasses-1-billion-in-value-1537524002>

DOW JONES, A NEWS CORP COMPANY

DJIA 25851.19 0.67% ▲ Nasdaq 7789.55 1.13% ▲ U.S. 10 Yr -1/32 Yield 2.433% ▼ Crude Oil 62.97 -0.21% ▼ Euro 1.1161 -0.05% ▼

# THE WALL STREET JOURNAL.

U.S. Edition | May 21, 2019 | Print Edition | Video

Home World U.S. Politics Economy **Business** Tech Markets Opinion Life & Arts Real Estate WSJ. Magazine 

## LOGISTICS REPORT

 SHARE  TEXT  2

BUSINESS | LOGISTICS REPORT | WSJ LOGISTICS REPORT

### Digital Freight Startup Convoy Raises \$185 Million, Surpasses \$1 Billion in Value

CapitalG, investment arm of Google parent Alphabet Inc., led a funding round aimed at backing the business in the growing field of digital freight booking



Waiting for tr.snapchat.com...

recruiting@convoy.com  
b@convoy.com

## WELCOME TO CONVOY

Our passionate team is committed to our mission of transporting the world with endless capacity and zero waste. Founded in 2015, Convoy has quickly grown to over

**PLEASE STAND BY**



THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG  
PILE OF LINEAR ALGEBRA, THEN COLLECT  
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL  
THEY START LOOKING RIGHT.



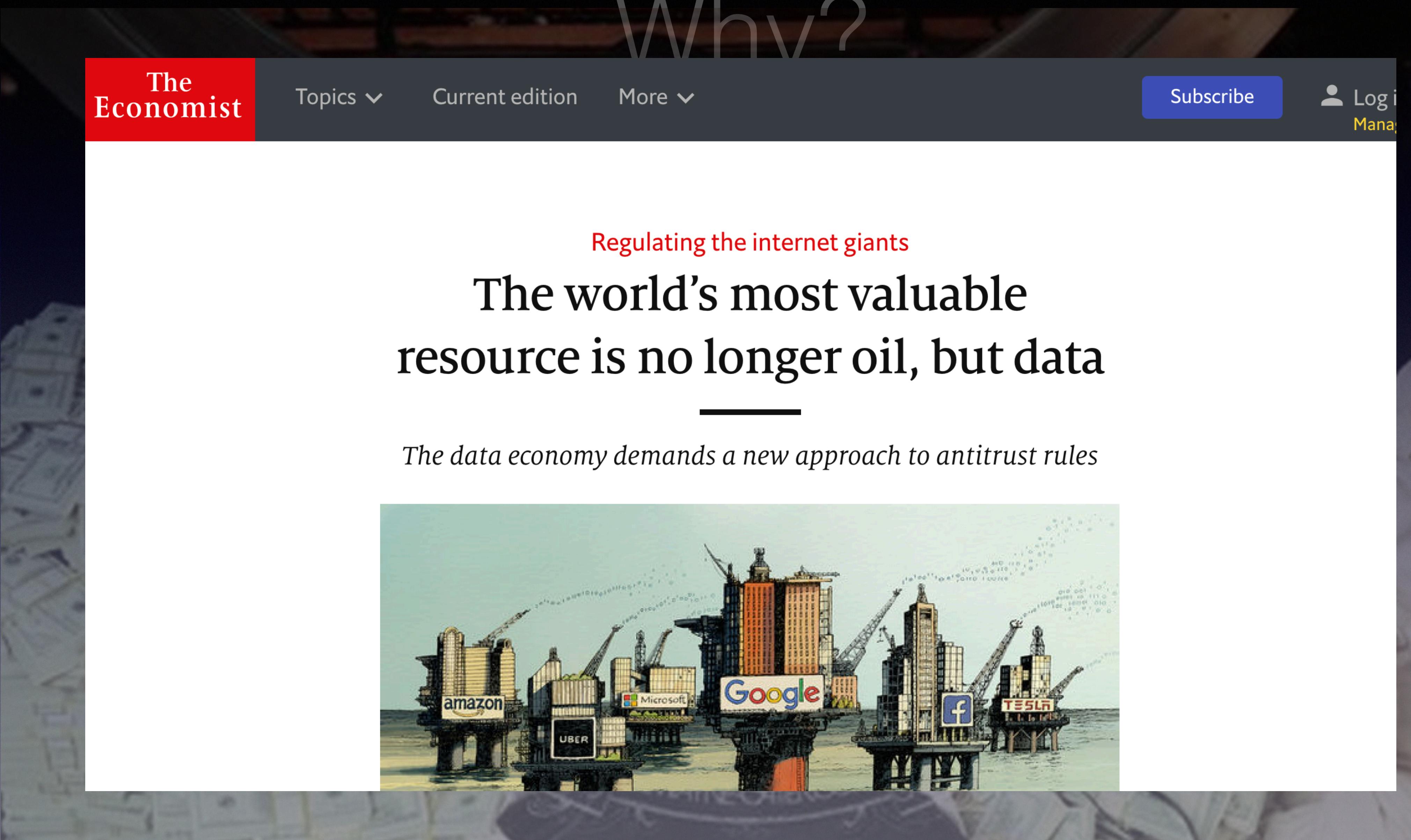


Intro  
Workflow  
Demo & pro-tips

# Intro

# Why?





# 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

Example\_Data\_Science\_Project > bin > main.py

Add Configuration... ▶ ⚙️ ⚡ ⚡ ⚡ ⚡ ⚡ Git: ✘ ✓ ⏴ ↻ 🔍

Project 1: Project Example\_Data\_Science\_ Example\_Data\_Science\_ bin \_\_init\_\_.py code\_template.py lib.py main.py conf data docs .gitignore case\_study\_guide.md environment.yml README.md External Libraries Scratches and Consoles

SciView Database

26 - Extracts the titanic data set  
27 - Performs some (light) feature extraction  
28 - Grid searches a pipeline  
29 - Loads the results for down stream use  
30  
31  
32 :return: None  
33 :rtype: None  
34 """  
35 logging.basicConfig(level=logging.DEBUG)  
36  
37 # Extract  
38 observations = extract()  
39 train, test = transform(observations)  
40 train, test, transformation\_pipeline, trained\_model = model(train, test)  
41 load(train, test, transformation\_pipeline, trained\_model)  
42  
43  
44 def extract():  
45 """  
46 Extract the data set from upstream  
47  
48 :return:  
49 """  
50 logging.info('Begin extract')  
51  
52 # Load the data set  
53 observations = lib.load\_titanic()  
54  
55 # Subset observation for speedier test iterations  
56 if lib.get\_conf('test run'):

# Demo

9: Version Control Python Console Terminal Docker 6: TODO Event Log

8:1 LF UTF-8 4 spaces Git: master Python 3.7

[https://github.com/bjherger/Example\\_Data\\_Science\\_Project](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: <https://bit.ly/2WWPs79>