How To Choose A Project Topic

Agenda

- **Where Do I Start?**

- **Criteria**

- **Avoid**

- **Scoping & Planning**

- **Where Do I Find My Dataset**

Where Do I Start...?

Professional Experience
- Finance
- Retail/Commerce
- Real Estate
- Transportation

Personal Passion
- Sports
- Travel
- Weather
- Food

Criteria

- 50K rows min for Mod 5

- Must have at least 8 continuous features prior to feature engineering.

- Dataset cannot be more than 50% categorical features

- For Mod 5 project avoid DIY datasets.
  - Just because you can doesn't mean you'll have time!

- What format? CSV, Sqlite, JSON or API?

WHATEVER

I DO WHAT I WANT

ICANHASCHEEZBURGER.COM

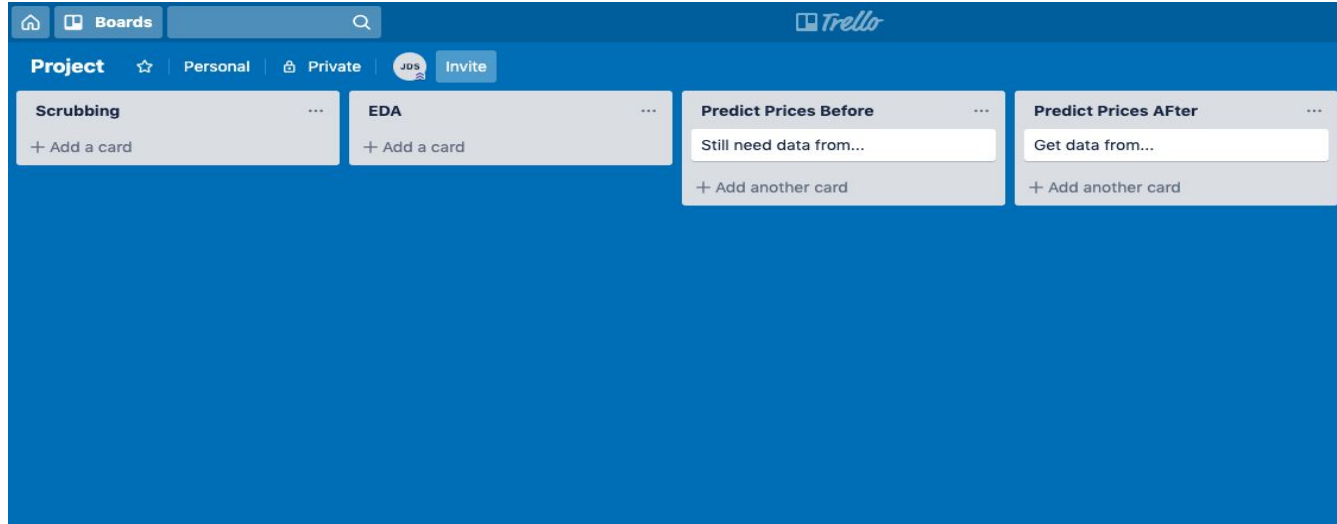# Avoid These Data types & Sectors...or else!

Source: Reddit

Data-types:

- NLP
- Computer Vision
- Review data
- DIY datasets (combining existing is ok)
- Anything that requires human input

Sectors:
- Education
- Government Programs
- Science (unless you have a background in it)

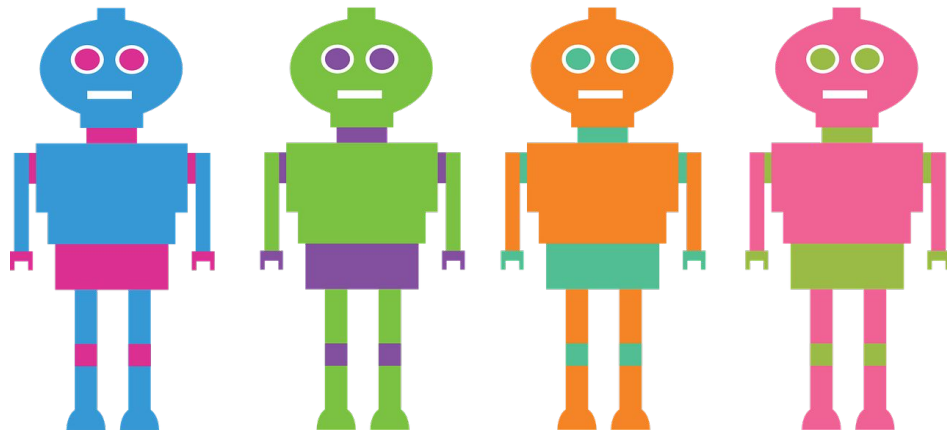# *Project Scoping & Planning*

- Again..for Mod 5 avoid DIY datasets.
  - Just because you can doesn't mean you'll have time!

- You will always have something more you could analyze

- Save project clean-up and additional info during break or at the end of the course

Where Do I Get My Data?

Stick to Machine Learning Data Aggregation Sites

- Most are good

- Someone has organized the data already

Source: Pixabay

- UCI Machine Learning Datasets Repository
- Kaggle Datasets
- Awesome Datasets Repo on Github
- New York City Open Data Portal
- Inside AirBNB
- data.world
- Google Dataset Search

Questions?