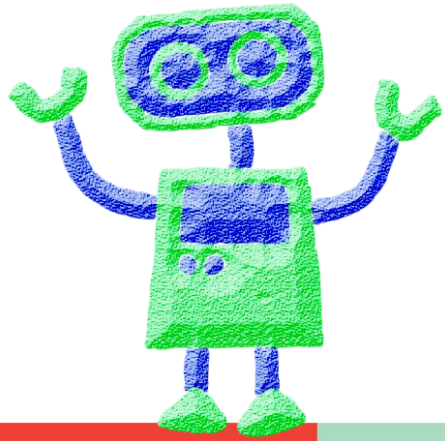
Friendly
Data Science

Data Science made fun, easy, and worthwhile



About me

Where you can find me:

- **Email:** ellfae@gmail.com
- **Website:** elvissaravia.com
- **Github:** @omarsar
- **Twitter:** @omarsar0
- **Communities:** Medium / CS-X (slack group -- invite only)

PhD scholar with focus on Data Mining, NLP, and Machine Learning

Motivations

Best Practices

A Humanistic Approach to Data Science

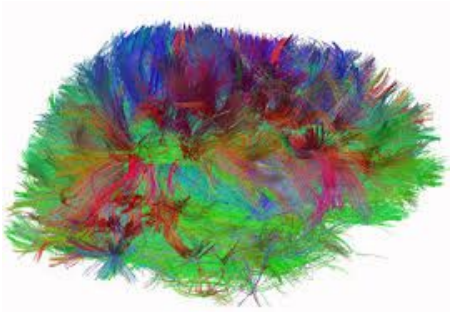
Data Science is about fun; it is about exploration and creativity

Learning Approaches:

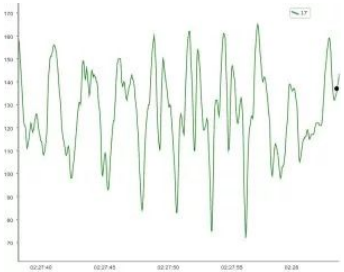
- Let’s start from scratch
- Theoretical and Practical approach

Best Practices (1)

- **Real-world Data Science**
  - MRI brain scans
  - Reddit conversations
  - Kaggle datasets
  - Sensor data
  - Photos
  - Social media data
  - ...and much more



kaggle



Best Practices (2)

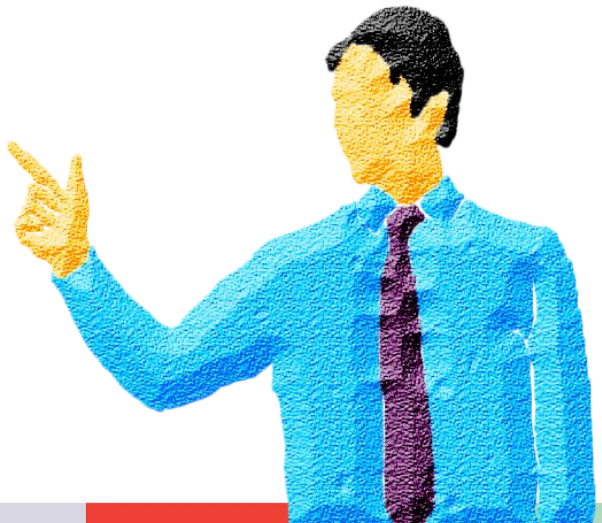
- **Testing your code for bugs**
  - To avoid misleading interpretations
  - To avoid biased results
  - To handle missing values and duplicate records
  - To handle other discrepancies in the dataset
  - ...and many others

```
In [39]: dummy_duplicate_dict = [{
                                    'text': 'dummy record',
                                    'category': 1,
                                    'category_name': "dummy category"
                                },
                                {
                                    'text': 'dummy record',
                                    'category': 1,
                                    'category_name': "dummy category"
                                }]
```

Injecting dummy duplicated records

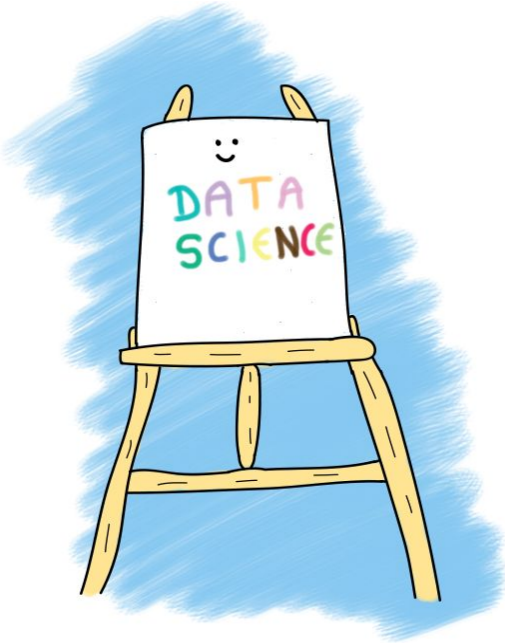
Best Practices (3)

- **How to store and retrieve data efficiently**
  - Database (SQL / Document-based)
  - Raw files (.txt, .csv)
  - Pandas dataframes
  - Search engine (e.g., Elasticsearch)
  - ...and much more



Humanistic approach to Data Science

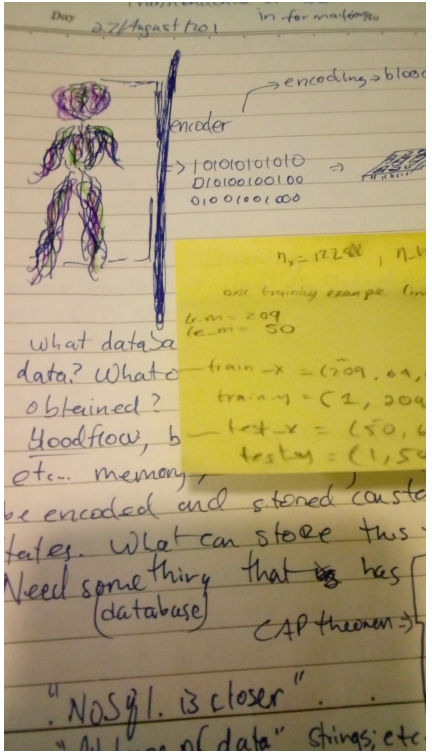
- **How to have fun doing Data Science**
  - Explore data responsibly (morals, ethics, etc.)
  - Maintaining your health
- **How to keep Motivated**
  - Dealing with bottlenecks and challenges
  - Community involvement
  - Open Source?
- **How to become and stay innovative**
  - GitHub, Opensource, forums, MOOCs, writing, etc.



Let’s have fun while learning

Fun techniques for applying Data Science:

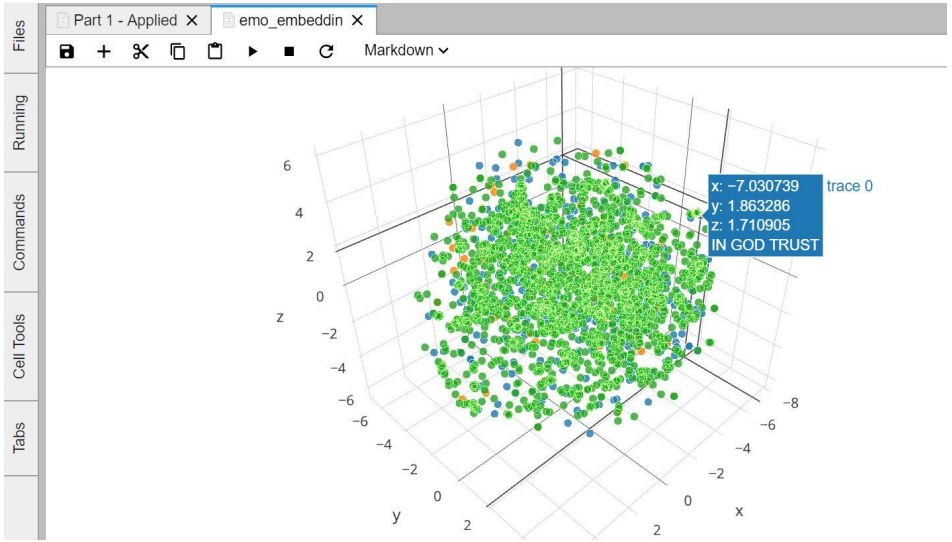
- Doodling
- Daydreaming
- Researching
- Reading
- Chatting (video/text-based)
- etc.



Doodle of a dream I had

Theoretical and Practical Approach

- **Learn the intuition and mathematical proof of concepts**
  - Principal Component Analysis (PCA)
  - Singular-value decomposition (SVD)
  - Clustering
  - Deep Neural Networks
  - ...and much more
- **Implementing from Scratch**
  - Numpy
  - Python
  - Scikit Learn
  - ...and much more



See you soon!