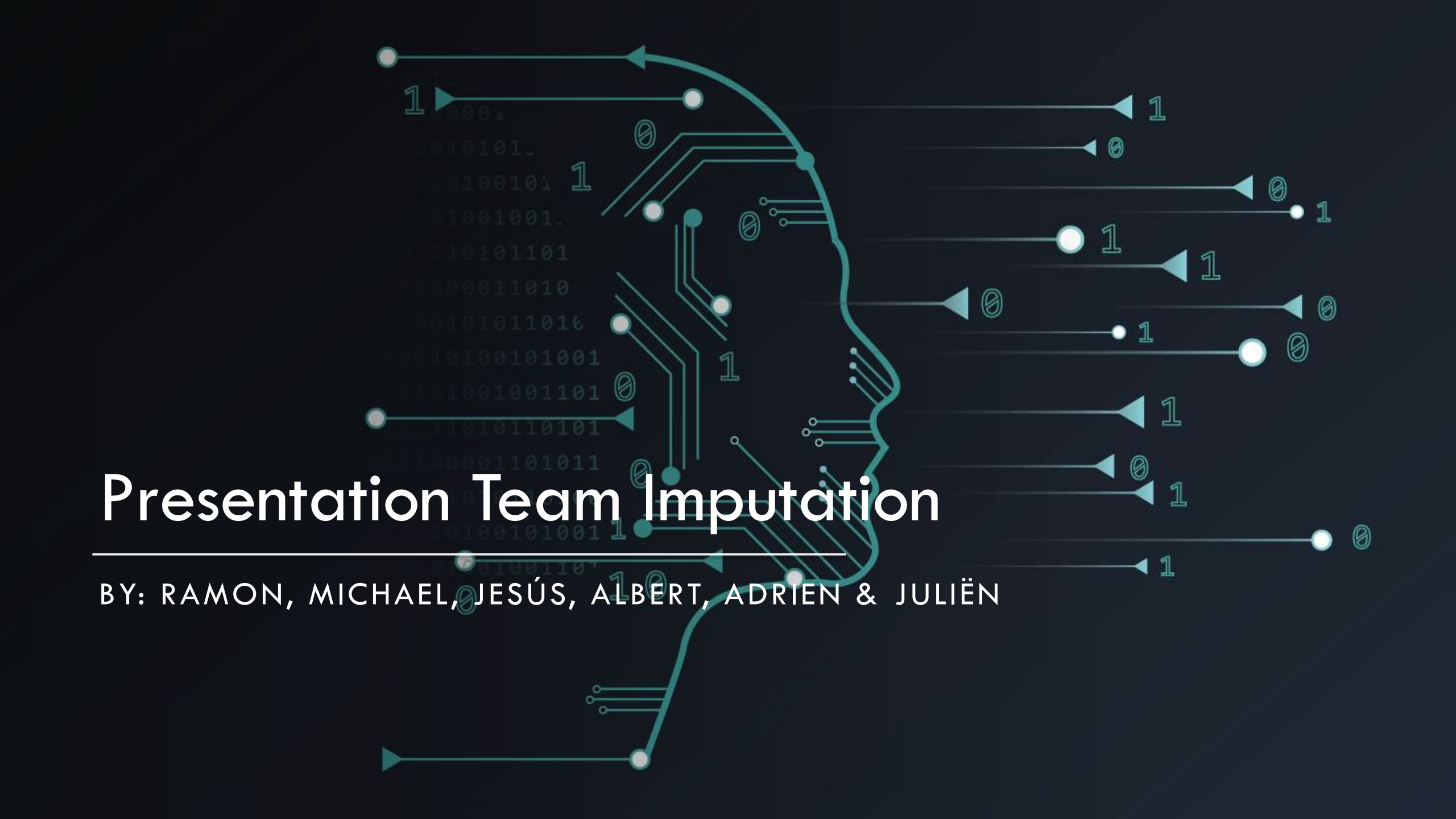


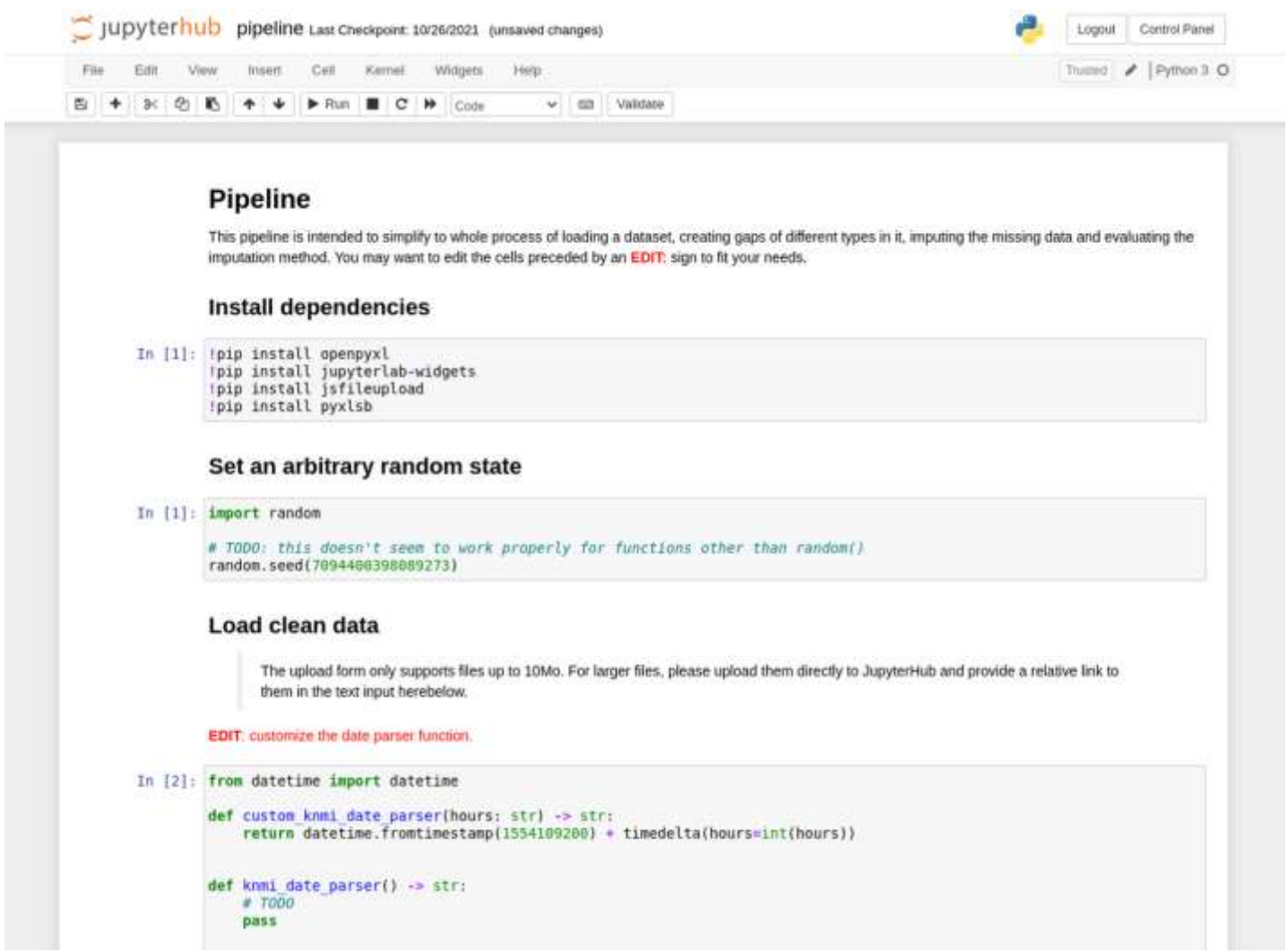
Presentation Team Imputation

BY: RAMON, MICHAEL, JESÚS, ALBERT, ADRIEN & JULIËN

Table of contents

Pipeline

- Uniform workflow
- File statistics
- Gap creation
- Multiple imputation methods
- Evaluation



Evaluation methods

- Bias
- Sum of error
- Distribution of error
- Maximum error

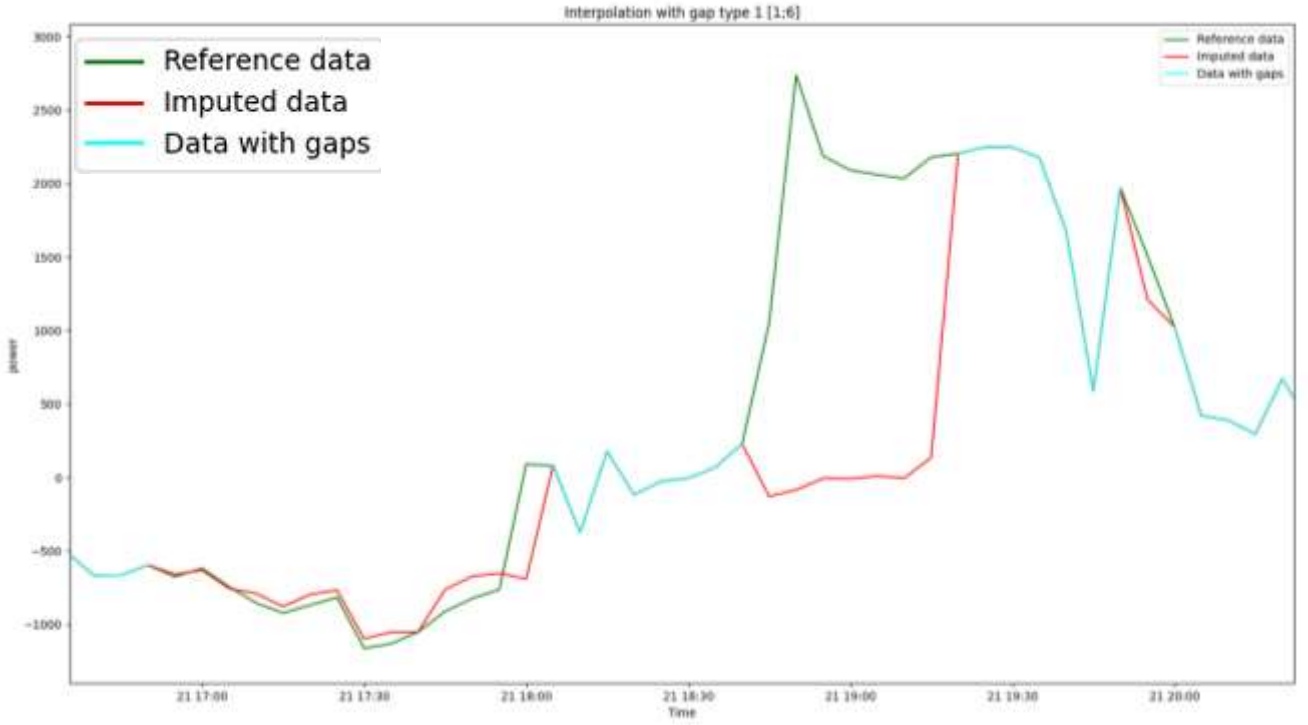
- Kurtosis : measures "tailedness"
- Skewness: measures asymmetry in probability distribution

```
Error distribution with gap type 1 [1;6]
        Mean Squared Error (lower is better)
                357117.4727332921
        Raw Bias
                -108.59066446031686
        Absolute Raw Bias (lower is better)
                628.6086817371577
        Percent Bias (bellow 5% is acceptable)
                15.819181324715185
        Sum (lower is better)
                12225181.642424243
        Maximum (lower is better)
                6379.900000000001
        Variance (lower is better)
                954858.1722140332
```

Pipeline evaluation metrics for Hot Deck with random donor pool.

Imputation methods

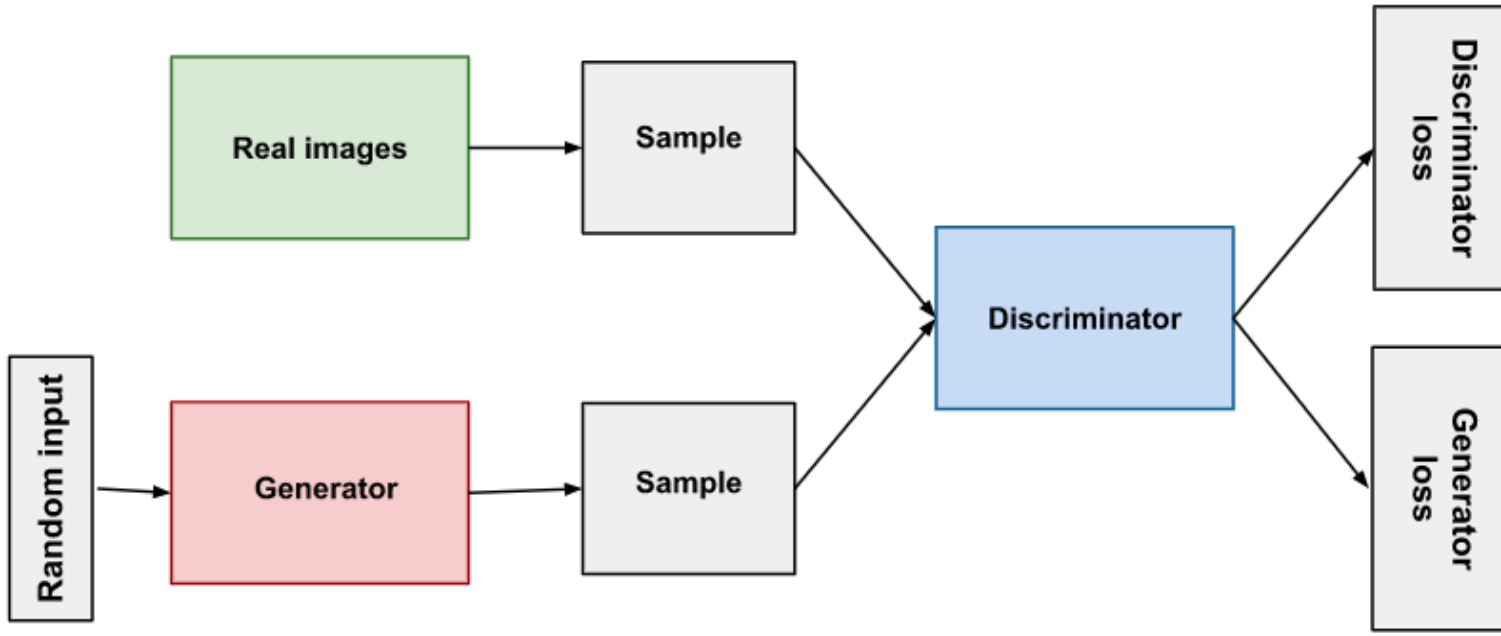
- Mode , Median & Mean
- KNN Nearest Neighbour
- Hot deck
  - Donor pool selection



Hot deck imputation result with a random donor

Neural Networks

- Recurrent Neural Networks
- GAN
- Convolutional Neural Networks



Research (paper)

- Worked on introduction for official paper
- CLIMA conference

What we will be working on

- Evaluation metrics in pipeline
- Other imputation methods
- Neural networks
- Research paper

- Focus right now: data collection

Questions & Feedback

- Questions we have for you: