

MD3 KAGGLE MATCAD

Data Analysis Project Suicide Rates Overview 1985 to 2016

**An in-depth study about how different socio-economic factors
might influence suicide rates**

Pau Blasco Roca - 1600959
<https://github.com/Nerocraft4>

Project structure



Step 1

Prepare the workspace

Set up the GitHub and the python environment.

Step 2

Data exploration and cleaning

General outlook of our dataset, spot interesting relations.
Clean the data.

Step 3

Feature selection

Find the significant variables in our model.

Step 4

Model learning and assessing

Try out several models and approaches to analyze data.

Step 5

Final analysis and conclusion

Note the main points of our results and compare it to other sources.

Setting up the repo + tools used



Data Exploring and Cleaning

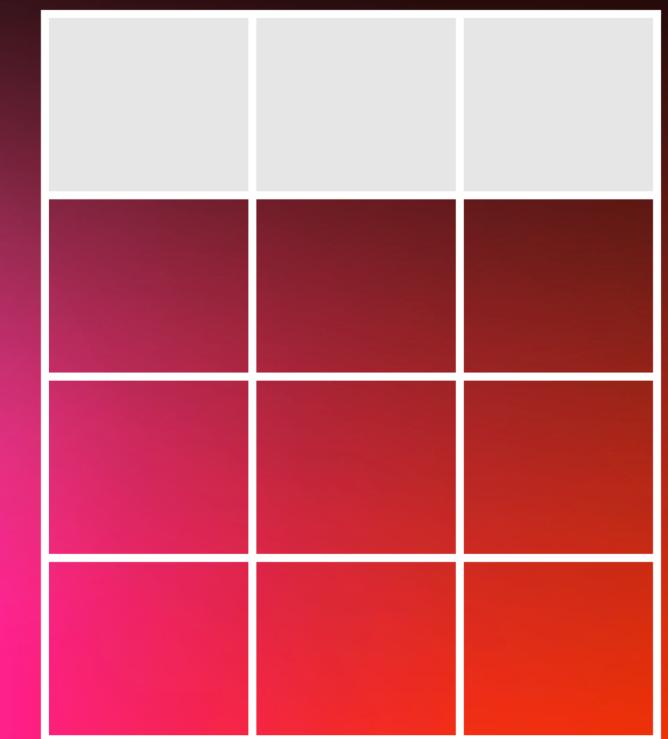
Previous study &
documentation



Data exploration,
early trends



Data scaling and
normalization



Feature selection

Pearson
correlation

Low variance
detection

Lasso feature
selection

RFE feature
selection

Model learning & evaluation

Ridge-Tikhonov
Regularization

0.54

Multi-linear
Regression

0.54

Lasso
Regression

0.50

Country-isolated
study

0.76

Final Analysis and Conclusions

**Consistently
correlated variables**

**Inconsistently
correlated variables**