DataSets

<https://www.kaggle.com/kemical/kickstarter-projects/version/7#ks-projects-201801.csv>

<https://webrobots.io/kickstarter-datasets/>

Orange:

[https://scikit-learn.org/stable/tutorial/machine\_learning\_map/index.html](https://meet.google.com/linkredirect?authuser=0&dest=https%3A%2F%2Fscikit-learn.org%2Fstable%2Ftutorial%2Fmachine_learning_map%2Findex.html)

[https://orange.biolab.si](https://orange.biolab.si/)

Fluxograma:

Análise do problema;

Pesquisa de dados;

Análise exploratória dos dados;

Pré-processamento:

* Eliminar dados indesejados;
  + Estados que são pertinentes para o projeto como:
    - Indefinido;
    - Cancelado;
    - Suspenso;
* Transformação dos dados;
  + Criação de novas features;
    - Percentagem de arrecadação;
    - Percentagem de dias faltantes;
    - Vetorização das palavras presentes nos nomes dos projetos;
      * Eliminar StopWords;
  + Categorizar campos
    - category
    - main\_category;
    - currency
    - country
    - state
  + Eliminar dados indesejados;
    - Estados que são pertinentes para o projeto como:
      * Indefinido;
      * Cancelado;
      * Suspenso;
  + Transformação dos dados;
    - Criação de novas features;
      * Percentagem de arrecadação;
      * Percentagem de dias faltantes;
      * Vetorização das palavras presentes nos nomes dos projetos;
        + Eliminar StopWords;
    - Categorizar campos
      * category
      * main\_category;
      * currency
      * country
      * state

Anomalias:

0 backers e ter algum valor investido sucesso/falha

191 anomalias

Outliers:

goal com valores muito pequenos ex.: 0,01;1

failed 3D Printing 326

Academic 589

Accessories 1667

Action 514

Animals 166

Animation 1531

Anthologies 216

Apparel 4751

Apps 4909

Architecture 463

Art 4101

Art Books 1113

Audio 252

Bacon 143

Blues 117

Calendars 162

Camera Equipment 158

Candles 323

Ceramics 143

Children's Books 3850

Childrenswear 323

Chiptune 6

Civic Design 162

Classical Music 793

Comedy 1156

Comic Books 879

Comics 2075

Community Gardens 205

Conceptual Art 564

Cookbooks 325

...

successful Shorts 6673

Small Batch 557

Software 371

Sound 274

Space Exploration 120

Spaces 295

Stationery 77

Tabletop Games 7866

Taxidermy 5

Technology 1317

Television 156

Textiles 77

Theater 4405

Thrillers 200

Translations 35

Typography 63

Vegan 185

Video 51

Video Art 51

Video Games 2386

Wearables 388

Weaving 27

Web 443

Webcomics 358

Webseries 1698

Woodworking 292

Workshops 50

World Music 922

Young Adult 172

Zines 179