











Artigo

[Yuri Marx](#) · Jan. 13 5min de leitura

[Open Exchange](#)

Dez conjuntos de dados abertos em saúde em apenas uma aplicação do IRIS

10 REAL HEALTH DATASETS

 Diabetes Predict Diabetes 768 rows and 9 columns (features)	 Breast Cancer Predict Breast Cancer 569 rows and 31 columns (features)	 Heart Disease Predict Heart Disease 270 rows and 14 columns (features)	 Dementia Predict Dementias 373 rows and 13 columns (features)	 Hepatitis Predict Hepatitis Death Risk 155 rows and 20 columns (features)
 Kidney Disease Predict Kidney Disease Risk 388 rows and 25 columns (features)	 Hospital Mortality Predict Hospital Mortality Risk 1177 rows and 50 columns (features)	 Maternal Risk Predict Maternal Risk 1014 rows and 8 columns (features)	 Life Expectancy Predict Life Expectancy by social and health indicators 373 rows and 13 columns (features)	 Pollution Deaths Deaths by fossil fuel pollution 190 rows and 5 columns (features)

Oi comunidade,

Em um trabalho intenso de curadoria e qualidade de dados, a aplicação "health dataset" entrega os conjuntos de dados acima.

Esses conjuntos de dados podem ser utilizados no seu modelo ou aplicação de Machine Learning, AutoML e de aplicações analíticas. Veja mais detalhes aqui:

Instalação

1. Clone/git pull no repositório em qualquer diretório local

```
$ git clone https://github.com/yurimarx/automl-heart.git
```

2. Abra o terminal no diretório da aplicação e execute:

```
$ docker-compose build
```

3. Execute o IRIS container:

```
$ docker-compose up -d
```

4. Faça um select no HeartDisease dataset:

```
SELECT  
age, bp, chestPainType, cholesterol, ekgResults, exerciseAngina, fbsOver120, heartDis  
ease, maxHr, numberOfVesselsFluro, sex, slopeOfSt, stDepression, thallium  
FROM dc_data_health.HeartDisease
```

5. Faça um select no Kidney Disease dataset:

```
SELECT  
age, al, ane, appet, ba, bgr, bp, bu, cad, classification, dm, hemo, htn, pc, pcc, pc  
v, pe, pot, rbc, rc, sc, sg, sod, su, wc  
FROM dc_data_health.KidneyDisease
```

6. Faça um select no Diabetes dataset:

```
SELECT  
Outcome, age, bloodpressure, bmi, diabetespedigree, glucose, insulin, pregnancies, sk  
inthickness  
FROM dc_data_health.Diabetes
```

7. Faça um select no Breast Cancer dataset:

```
SELECT  
areamean, arease, areaworst, compactnessmean, compactnesse, compactnessworst, concav  
epointsmean, concavepointsse, concavepointsworst, concavitymean, concavityse, concavi  
tyworst, diagnosis, fractaldimensionmean, fractaldimensionse, fractaldimensionworst,  
perimetermean, perimeterse, perimeterworst, radiusmean, radiusse, radiusworst, smooth  
nessmean, smoothnesse, smoothnessworst, symmetrymean, symmetryse, symmetryworst, tex  
turemean, texturese, textureworst  
FROM dc_data_health.BreastCancer
```

8. Faça um select no Maternal Health Risk dataset:

```
SELECT  
BS, BodyTemp, DiastolicBP, HeartRate, RiskLevel, SystolicBP, age  
FROM dc_data_health.MaternalHealthRisk
```

9. Faça um select no Hospital Mortality dataset:

```
SELECT  
age, aniongap, atrialfibrillation, basophils, bicarbote, bloodcalcium, bloodpotassium  
, bloodsodium, bmi, chdwithnomi, chloride, copd, creatinekise, creatinine, deficiency  
anemias, depression, diabetes, diastolicbloodpressure, ef, gendra, glucose, "group",  
heartrate, hematocrit, hyperlipemia, hypertensive, inr, lacticaacid, leucocyte, lymph  
ocyte, magnesiumion, mch, mchc, mcv, neutrophils, ntprobnp, outcome, pco2, ph, plate  
lets, pt, rbc, rdw, relfailure, respiratoryrate, spo2, systolicbloodpressure, tempera  
ture, ureanitrogen, urineoutput
```

```
FROM dc_data_health.HospitalMortality
```

10. Faça um select no Life Expectancy dataset:

```
SELECT
AdultMortality, Alcohol, BMI, Country, Diphtheria, GDP, HIVAIDS, HepatitisB, IncomeCo
mpositionOfResources, InfantDeaths, LifeExpectancy, Measles, PercentageExpenditure, P
olio, Population, Schooling, Status, Thinness1To19Years, Thinness5To9Years, TotalExpe
nditure, UnderFiveDeaths, Year
FROM dc_data_health.LifeExpectancy
```

11. Faça um select no Pollution Deaths dataset:

```
SELECT
Country, CountryCode, DeathYear, ExcessMortality
FROM dc_data_health.PollutionDeaths
```

12. Faça um select no Dementia dataset:

```
SELECT
ASF, Age, CDR, EDUC, Genre, Hand, MMSE, MRDelay, Outcome, SES, Visit, eTIV, nWBV
FROM dc_data_health.Dementia
```

13. Faça um select no Hepatitis Death risk dataset:

```
SELECT
age, albumin, alkphosphate, anorexia, antivirals, ascites, bilirubin, fatigue, histol
ogy, liverbig, liverfirm, malaise, outcome, protime, sex, sgot, spiders, spleenpalpab
le, steroid, varices
FROM dc_data_health.Hepatitis
```

Para instalar no ZPM

A aplicação também pode ser instalada com o ZPM:

```
zpm "install dataset-health"
```

Licenças/Créditos dos Datasets

- MIT License para esta aplicação
- CC BY-NC-SA 4.0 License para o o Conjunto de Dados de Câncer do Coração
 - Fonte de dados: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
 - Arquivo dentro do app: /opt/irisapp/data/breast-cancer.csv
 - Classe Persistente: dc.data.health.BreastCancer
- CC0: Public Domain para o o Conjunto de Dados de Diabetes
 - Fonte de dados: <https://www.kaggle.com/mathchi/diabetes-data-set>
 - Arquivo dentro do app: /opt/irisapp/data/diabetes.csv
 - Classe Persistente: dc.data.health.Diabetes
- CC0: Public Domain para o o Conjunto de Dados de Doença do Coração
 - Fonte de dados: <https://data.world/informatics-edu/heart-disease-prediction>
 - Arquivo dentro do app: /opt/irisapp/data/heart-disease.csv
 - Classe Persistente: dc.data.health.HeartDisease

- CC0: Public Domain para o Conjunto de Dados de Risco Maternal
 - Fonte de dados: <https://www.kaggle.com/yasserhessein/classification-maternal-health-5-al...>
 - Arquivo dentro do app: /opt/irisapp/data/maternal_health_risk.csv
 - Classe Persistente: dc.data.health.MaternalHealthRisk
- CC0: Public Domain para o Conjunto de Dados de Expectativa de Vida
 - Fonte de dados: <https://www.kaggle.com/kumarajarshi/life-expectancy-who> - The data was collected from WHO and United Nations website with the help of Deeksha Russell and Duan Wang.
 - Arquivo dentro do app: /opt/irisapp/data/life_expectancy.csv
 - Classe Persistente: dc.data.health.LifeExpectancy
- CC0 1.0 Universal (CC0 1.0) Public Domain Dedication para o Conjunto de Mortalidade em Hospital
 - Fonte de dados: <https://www.kaggle.com/saurabhshahane/in-hospital-mortality-prediction> (Zhou, Jingmin et al. (2021), Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database, Dryad, Dataset, <https://doi.org/10.5061/dryad.0p2ngf1zd>)
 - Arquivo dentro do app: /opt/irisapp/data/hospital_mortality.csv
 - Classe Persistente: dc.data.health.HospitalMortality
- CC0 1.0 Universal (CC0 1.0) Public Domain para o Conjunto de Dados de Mortes por Poluição
 - Fonte de dados: <https://www.kaggle.com/mathurinache/pollution-deaths>
 - Arquivo dentro do app: /opt/irisapp/data/pollution-deaths-from-fossil-fuels.csv
 - Classe Persistente: dc.data.health.PollutionDeaths
- Attribution-NonCommercial-ShareAlike 3.0 IGO (CC BY-NC-SA 3.0 IGO) para o Conjunto de Dados de Demência
 - Fonte de dados: <https://www.kaggle.com/shashwatwork/dementia-prediction-dataset>
 - Arquivo dentro do app: /opt/irisapp/data/dementia.csv
 - Classe Persistente: dc.data.health.Dementia
- CC0 1.0 Universal (CC0 1.0) Public Domain para o Conjunto de Dados de Hepatite
 - Fonte de dados: <https://www.kaggle.com/codebreaker619/hepatitis-data>
 - Arquivo dentro do app: /opt/irisapp/data/hepatitis.csv
 - Classe Persistente: dc.data.health.Hepatitis
- CC0: Public Domain para o Conjunto de Dados de Rim
 - Fonte de dados:
 - @misc{Dua:2019 , author = "Dua, Dheeru and Graff, Casey", year = "2017", title = "{UCI} Machine Learning Repository", url = "<http://archive.ics.uci.edu/ml>", institution = "University of California, Irvine, School of Information and Computer Sciences" }
 - Arquivo dentro do app: /opt/irisapp/data/kidney_disease.csv
 - Classe Persistente: dc.data.health.KidneyDisease

[#Analytics](#) [#Importação e Exportação de Dados](#) [#Machine Learning](#) [#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)
[Confira o aplicativo relacionado no InterSystems Open Exchange](#)

URL de origem: <https://pt.community.intersystems.com/post/dez-conjuntos-de-dados-abertos-em-sa%C3%BAde-em-apanas-uma-aplica%C3%A7%C3%A3o-do-iris>