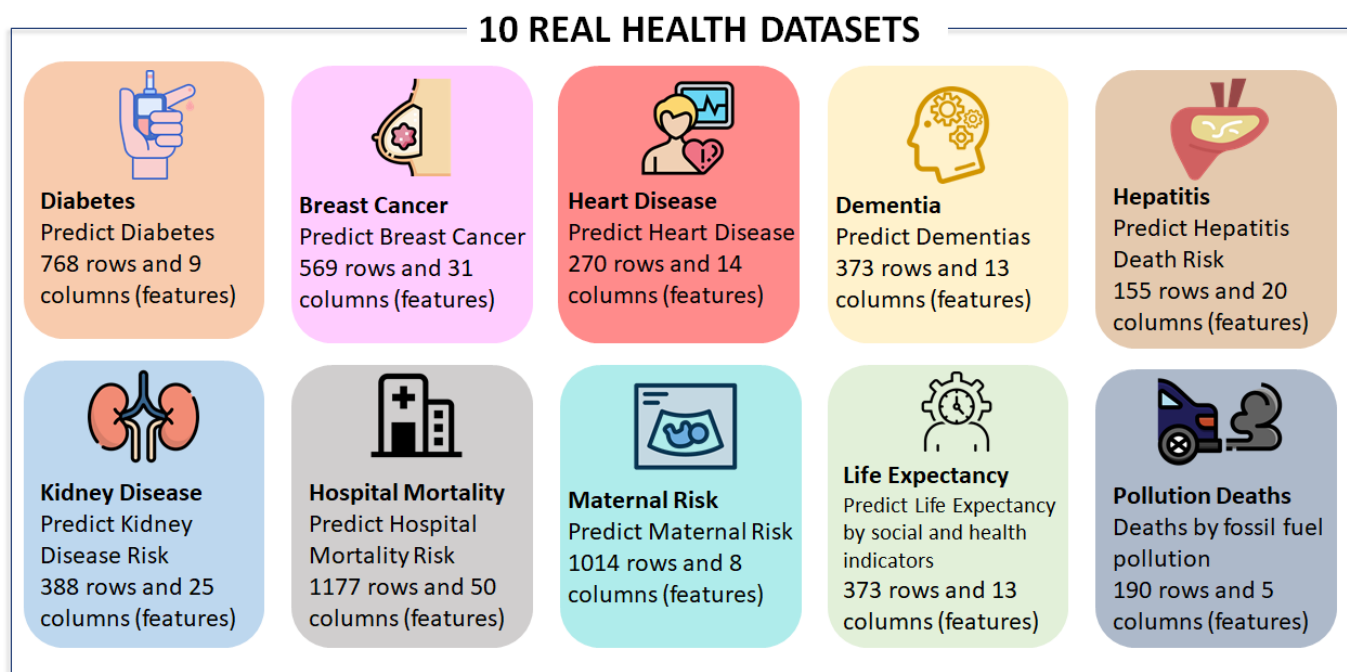


Article

[Yuri Marx](#) · Jan 9, 2022 5m read

[Open Exchange](#)

Ten real health datasets in a just one OEX application



Hi Community,

In a hard work of data curation and data quality, the Health Dataset application deploy to you the above datasets.

These datasets can be used in your ML applications/models, AutoML and analytics projects. See more details [here](#):

Installation

1. Clone/git pull the repo into any local directory

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

2. Open a Docker terminal in this directory and run:

```
$ docker-compose build
```

3. Run the IRIS container:

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

4. Do a Select to the HeartDisease dataset:

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

5. Do a Select to the 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. Do a Select to the Diabetes dataset:

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

7. Do a Select to the Breast Cancer dataset:

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

8. Do a Select to the Maternal Health Risk dataset:

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

9. Do a Select to the Hospital Mortality dataset:

```
SELECT
age, aniongap, atrialfibrillation, basophils, bicarbote, bloodcalcium, bloodpotassium
, bloodsodium, bmi, chdwithnomi, chloride, copd, creatinekise, creatinine, deficiency
anemias, depression, diabetes, diastolicbloodpressure, ef, gendera, glucose, "group",
heartrate, hematocrit, hyperlipemia, hypertensive, inr, lacticaacid, leucocyte, lymp
hocyte, 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. Do a Select to the 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. Do a Select to the Pollution Deaths dataset:

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

12. Do a Select to the Dementia dataset:

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

13. Do a Select to the 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
```

To install with ZPM

It's packaged with ZPM so it could be installed as:

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

Dataset Licenses and sources/credits

1. MIT License for this Application
2. CC BY-NC-SA 4.0 License for the Breast Cancer Dataset
 - Original Source: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
 - File into the app: /opt/irisapp/data/breast-cancer.csv
 - Persistent Class: dc.data.health.BreastCancer
3. CC0: Public Domain for Diabetes Dataset
 - Original Source: <https://www.kaggle.com/mathchi/diabetes-data-set>
 - File into the app: /opt/irisapp/data/diabetes.csv
 - Persistent Class: dc.data.health.Diabetes
4. CC0: Public Domain for Heart Disease
 - Original Source: <https://data.world/informatics-edu/heart-disease-prediction>
 - File into the app: /opt/irisapp/data/heart-disease.csv
 - Persistent Class: dc.data.health.HeartDisease
5. CC0: Public Domain for Maternal Health Risk

- Original Source:
<https://www.kaggle.com/yasserhessein/classification-maternal-health-5-algorithms-ml/data>
 - File into the app: /opt/irisapp/data/maternalhealthrisk.csv
 - Persistent Class: dc.data.health.MaternalHealthRisk
6. CC0: Public Domain for World Life Expectancy
- Original Source: <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.
 - File into the app: /opt/irisapp/data/lifeexpectancy.csv
 - Persistent Class: dc.data.health.LifeExpectancy
7. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication for Hospital Mortality
- Original Source: <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>)
 - File into the app: /opt/irisapp/data/hospitalmortality.csv
 - Persistent Class: dc.data.health.HospitalMortality
8. CC0 1.0 Universal (CC0 1.0) Public Domain for Pollution Deaths dataset
- Original Source: <https://www.kaggle.com/mathurinache/pollution-deaths>
 - File into the app: /opt/irisapp/data/pollution-deaths-from-fossil-fuels.csv
 - Persistent Class: dc.data.health.PollutionDeaths
9. Attribution-NonCommercial-ShareAlike 3.0 IGO (CC BY-NC-SA 3.0 IGO) for Dementia dataset
- Original Source: <https://www.kaggle.com/shashwatwork/dementia-prediction-dataset>
 - File into the app: /opt/irisapp/data/dementia.csv
 - Persistent Class: dc.data.health.Dementia
10. CC0 1.0 Universal (CC0 1.0) Public Domain for Hepatitis Death Risk dataset
- Original Source: <https://www.kaggle.com/codebreaker619/hepatitis-data>
 - File into the app: /opt/irisapp/data/hepatitis.csv
 - Persistent Class: dc.data.health.Hepatitis
11. CC0: Public Domain for Kidney Disease
- Original Source:
 - @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" }
 - File into the app: /opt/irisapp/data/kidneydisease.csv
 - Persistent Class: dc.data.health.KidneyDisease

[#Analytics](#) [#Data Import and Export](#) [#Machine Learning \(ML\)](#) [#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)
[Check the related application on InterSystems Open Exchange](#)

Source URL: <https://community.intersystems.com/post/ten-real-health-datasets-just-one-oex-application>