Possible datasets are: 1

Possible datasets are:

1. IMDB dataset

IMDB 5000 Movie dataset. It contains information on various movies & television shows. The attributes include: Director's name, Number of Critic Reviews, Duration, Secondary Actor's Name, Gross, Genres, Primary Actor's Name, Movie Title, Number of Votes, Number or User Reviews, Country, Content Rating, Budget, Title Year and IMDB score. The dataset is in CSV format and can be found at: https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset

2. Game of Thrones dataset

There are a few datasets related to the Game of Thrones books available at
https://www.kaggle.com/mylesoneill/game-of-thrones

The "character-deaths.csv" details information about characters who have appeared in the "A Song of Ice and Fire" book series, the books that Game of Thrones is based off of. There are 13 different attributes and 918 records.
The attributes are as follows:
Name- The characters name
Allegiances- Which noble house the character has sworn allegiances too
Death year- The year a character died in (If they have died)
Book of Death- What book they died in (If they have died)
Death Chapter- What chapter in the respective book they died in (If they have died)
Book Intro- What book they were introduced in

Gender- What the characters gender is (1 for male, 0 for female)
Nobility- If the particular character is from some form of noble house
The books the character has appeared in (1 for appears, 0 for doesn't appear):
GoT = A Game of Thrones
CoK = A Clash of Kings
SoS = A Storm of Swords
FfC = A Feast for Crows
DwD = A Dance with Dragons

3. Pokemon stats dataset

URL: https://www.kaggle.com/abcsds/pokemon
This Pokemon stats data set contains 721 rows in total, with 7 attributes.

ID (int)
Name (Varchar)
Type1 (Varchar)
Type2 (Varchar)
Total (Int)
HP (Int)
Attack (Int)
Defense (Int)
Sp. Att (Int)
Sp, Def (Int)
Speed (Int)
Generation (Int)
Legendary (Boolean / Int)

Airplane crashes

https://www.kaggle.com/saurograndi/airplane-crashes-since-1908

The attributes are as follows:
Date: The date that the crash occurred on.
Time: The time that the crash occurred at if the time was recieved. Some of the values are null.
Location: The location at which the crash occurred.
Operator: The company/airline/military that the crashed plane belonged to.
Flight #: The flight number of the flight if it was available. Some values are null.
Route: The destination for the flight if available. Some values are null.
Type: The make and possibly model of the crashed aircraft.
Registration: The registration number for the crashed aircraft.
cn/In: We are not sure what this value represents, but it might be the construction number.
Aboard: The number of people on the aircraft at takeoff.
Fatalities: The number of people aboard the aircraft who were killed
Ground: Number of fatalities on the ground where the aircraft crashed.
The dataset is a .csv file

4. UFO sightings

The dataset is a set of UFO reports from the 'national UFO reporting center' for the state of Colorado.
Link : http://www.nuforc.org/webreports/ndxlCO.html
The column headings are:
Date/time(of event), city, state, shape, duration, summary(text description of event), Posted(date of report)

5. Titanic dataset

Dataset picked is passenger data from the Titanic.

Source: https://www.kaggle.com/c/titanic/data
It contains 12 attributes :PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
"PassengerId" is the passenger ID number. "Survived" shows weather they survived (0 =No, 1 - Yes). "Pclass" shows is what class the passenger was in 1st class, 2nd class etc..., "Name" is the passenger name. "Sex" is the gender. "Age. "SibSp" shows the number of siblings or spouse aboard. "Parch" shows the number of parents/children aboard. "Ticket" is the ticket number. "Fare" is the fare price. "Cabin" is the cabin number they were in. "Embarked" shows what locations they got on board the ship from.

6. Datascience courses

https://www.kaggle.com/sriharirao/datascience-universities-across-us

(Some fields will be trimmed - location, latitude and longitude), CSV file

Datascience Universities across US.

This dataset contains datascience programs offered and location data by universities.

It includes the University name, State, City, How it's delivered(online or on campus), what year it was available, world ranking, Department, and duration of program (in credits or months).

This dataset has a mix of integers, strings, hyperlinks. Their identifier will be the University name, name of program, and/or department. (will be decided later)

7. Volcanos

https://public.tableau.com/s/sites/default/files/media/Resources/significantvolcanoeruptions.xlsx

Attributes are:

Year
Month
Day
Associated Tsunami?
Associated Earthquake?
Name
Location
Country
Latitude
Longitude
Elevation
Type
Status
Time
Volcano Explosivity Index (VEI)
Agent
DEATHS
DEATHS_DESCRIPTION

MISSING
MISSING_DESCRIPTION
INJURIES
INJURIES_DESCRIPTION
DAMAGE_MILLIONS_DOLLARS
DAMAGE_DESCRIPTION
HOUSES_DESTROYED
HOUSES_DESTROYED_DESCRIPTION
TOTAL_DEATHS
TOTAL_DEATHS_DESCRIPTION
TOTAL_MISSING
TOTAL_MISSING_DESCRIPTION
TOTAL_INJURIES
TOTAL_INJURIES_DESCRIPTION
TOTAL_DAMAGE_MILLIONS_DOLLARS
TOTAL_DAMAGE_DESCRIPTION
TOTAL_HOUSES_DESTROYED
TOTAL_HOUSES_DESTROYED_DESCRIPTION

8. Games over the years

https://www.kaggle.com/egrinstein/20-years-of-games
It contains 18625 lines with the fields like the release date, it's platform and IGN's score. All the lines are fully filled.
In 20 years, the gaming industry has grown and sophisticated. By exploring this dataset, one is able to find trends about the industry, compare consoles against eachother, search through the most popular genres and more.
with the following headings :
Number ,score_phrase, title , url , platform ,score ,genre ,editors_choice ,release_year ,release_month ,release_day
Allowing for us to search through platforms and genres or release years etc, also allowing for searching this users score of the games
File type: .csv
File size 2mb

Database Description:

A complete dataset of 20 years' worth of video game reviews from the website IGN.com. This .csv file composes attributes of the videogame, its review score (both numeric and a phrase score i.e Amazing, good, bad, awful etc…), the url location of it on the IGN website, the release date, each gaming platform said game was released on (i.e. PC, PlayStation, Xbox etc..) and lastly whether or the videogame received an editor's choice recommendation. This dataset holds 11 attributes and while it exceeds the 3000 rows by having 18625 rows it was discussed in a lab not downsize in case of major data lose and that it's small enough to run on our own servers. This was downloaded from
Kaggle: https://www.kaggle.com/egrinstein/20-years-of-games
Attributes are:
ID: Originally the attribute was not named on download, by just giving it an 'id' name for the attribute one can use it as an ID identification for each video game. This is a numeric data value.
Score_Phrase: A varchar data value that is simply the score phrase given to the videogame based on the numeric score i.e. if review is between 8-8.9 it gets a 'great' or above 9 it's an 'amazing' etc.…
Title: A varchar datatype for simply giving the title of the videogame.
url: Another varchar that holds the url address of the videogame review on the IGN.com website.
Platform: A varchar that tells what platform the game was released on.
Score: A numeric data type that gives the games review score.
Editor Choice: A char data type that tells either yes or no, the videogame got an editor's recommendation.
Release month, year & date: All numeric data types for the release date of the video game.

9. Consumer Complaints

URL:
https://catalog.data.gov/dataset/consumer-complaint-database
Description:
Complaints about financial products and services
Format:
CSV
Attributes:
Date, Product, Sub Product, Issue, Sub Issue, Consumer, Company Public Response, Company, State, ZIP Code, Tags, Consumer, Submitted, Date Sent, Company Response, Timely Response, Consumer Disputed?, Compaint ID

Number of Rows: 636815
Number of Columns: 8

10. Olymic Games dataset

This dataset contains information on medal winners over a period of time. The attributes are:

The city the game was in
Edition (year of games)
Sport (specific set of events)
Discipline (more generic sport area)
Athlete
NOC (home country TLA)
Gender
Event
Event_gender
Medal

https://public.tableau.com/s/sites/default/files/media/Resources/Summer_Olympic_medallists_1896-2008.xlsx

11. Business in the US dataset

The Inc. 5000 is Inc. Magazine's annual list of the 5000 fastest growing private companies in the United States. The list is compiled by measuring each company's percentage revenue growth over a four-year period.
https://public.tableau.com/s/sites/default/files/media/Data%20Set-%20Inc5000%20Company%20List_2014.csv

The attributes are:

    _num
    _widgetName
    _source
    _resultNumber
    _pageUrl
    id

rank
workers
company
url
state_l
state_s
city
metro
growth
revenue
industry
yrs_on_list

12. Eurovision songs, 1998 to 2012

https://public.tableau.com/s/sites/default/files/media/Resources/eurovision_1998%20to%202012.xlsx

	Song.Quality
Year	Normalized.Points
Country	energy
Region	duration
Artist	acousticness
Song	danceability
Artist.gender	tempo
Group.Solo	speechiness
Place	key
Points	liveness
Home.Away.Country	time_signature
Home.Away.Region	mode
Is.Final	loudness
Semi.Final.Number	valence
Song.In.English	Happiness