

Data Science Hackathon

Pick one of the following datasets and using your knowledge of python (or other tools) prepare a presentation demonstrating some of the insights you have found.

(Hint: a good project has a solid framework around answering a question)

1. Olympics

This dataset provides an opportunity to ask questions about how the Olympics have evolved over time, including questions about the participation and performance of athletes, different nations, and different sports and events.

2. Netflix Movies and TV Shows

Some of the interesting questions which can be performed on this dataset

- Understanding what content is available in different countries
- Identifying similar content by matching text-based features
- Network analysis of Actors / Directors and find interesting insights
- Is Netflix increasingly focusing on TV rather than movies in recent years?

3. COVID-19 World Vaccination Progress

Some potential questions to be asked:

- Which country is using what vaccine?
- In which country the vaccination programme is more advanced?
- Where are vaccinated more people per day? But in terms of per cent of the entire population?

4. The Museum of ModernArt (MoMA) Collection

Artworks Dataset

The Artworks dataset contains 138,151 records, representing all of the works that have been accessioned into MoMA's collection and catalogued in our database. It includes basic metadata for each work, including title, artist, date made, medium, dimensions, and date acquired by the Museum. Some of these records have incomplete information and are noted as "not Curator Approved."

Artists Dataset

The Artists dataset contains 15,222 records, representing all the artists who have work in MoMA's collection and have been catalogued in our database. It includes basic metadata for each artist, including name, nationality, gender, birth year, death year, Wiki QID, and Getty ULAN ID.

5. Chinook Database - SQL Database.

The Chinook data model represents a digital media store, including tables for artists, albums, media tracks, invoices and customers. Media related data was created using real data from an iTunes Library. Customer and employee information was manually created using fictitious names, addresses that can be located on Google maps, and other well-formatted data (phone, fax, email, etc.). Sales information is auto-generated using random data for a four year period.

You can access this database using pgadmin. Please follow the instructions below:

URL: <http://delivery-pgadmin.multiverse.io/>

Username: pgadmin4@docker

Password: pgadmin4

