

DATA PROJECT

#8

NETFLIX



NETFLIX DATA
MOVIE & TV-SHOWS



FABIAN WERKMEISTER

DATA

WHAT data

&

WHERE to get it



DATA SOURCE

**Dataset of different movies
& TV shows on the Netflix
streaming platform**

- 
- 
- Director & cast
 - Country of production
 - Release year
 - Genre
 - Short description



DATA SOURCE

You can find & download
the dataset on Kaggle

A title	A director	A cast	A country
Title of the Movie / Tv Show	Director of the Movie	Actors involved in the movie / show	Country where the / show was produced
8807 unique values	[null] 30% Rajiv Chilaka 0% Other (6154) 70%	[null] 9% David Attenborough 0% Other (7963) 90%	United States India Other (5017)
Dick Johnson Is Dead	Kirsten Johnson		United States
Blood & Water		Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dillon Windvogel, Natasha Thahane, Arno Gree...	South Africa
Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Noureddin...	
Jailbirds New			

Summary

- 1 file
- 12 columns



LINK IN NOTION



FABIAN WERKMEISTER

DATA TRACKS

SELECT ONE or COMBINE



I - Data Engineering

II - Data Analytics

III - Data Science



TRACK I

DATA ENGINEERING

Feature Engineering

- Categorize release years into decade of release
- Derive versatility of a movie/TV show by counting no. of genres
- Derive continent from country of production



TRACK I DATA ENGINEERING



Some shows only have 1
season - why?!

Did they got cancelled or
were they intentionally
designed as a mini-show?

Find this information!

TRACK II

DATA ANALYTICS

Exploratory Data Analysis

- Distribution of movies and TV shows across country, release year & genre
- Calculate the average no. of seasons of TV shows across genres
- Are certain genres more prevalent during different years of production?



TRACK II

DATA ANALYTICS



Find the top directors and
actors based on the
number of movies

Which are the most
frequent pairs of director
and actor?



FABIAN WERKMEISTER

TRACK III

DATA SCIENCE

Predict the genre of a movie

- Which features could be relevant for this?
- Does the same model also work for TV shows?
- How can you make the duration comparable between movies & shows?



TRACK III

DATA SCIENCE



◀ Predict the genre solely
based on the plot summary
and the title ▶

After that, use only the cast
and the director(s)



TUTORIAL IN NOTION



FABIAN WERKMEISTER

LEARNINGS



why YOU

should do

this project



FABIAN WERKMEISTER

LEARNINGS

Track I - Data Engineering

Feature engineering, Applying domain knowledge on the data



Track II - Data Analytics

Exploratory data analysis, Finding hidden relationships in text data

Track III - Data Science

Multi-label classification with numerical and text features



**WHAT WOULD
YOU DO WITH
THIS DATA?**

**SEE YOU IN
THE COMMENTS**



FABIAN WERKMEISTER

HI, I'M FABIAN



**I HELP YOU
TO BECOME
FLUENT IN DATA**



FOR PROJECT IDEAS EACH FRIDAY

+ Follow

& hit the 



FABIAN WERKMEISTER

 **BONUS** 



FREE Data



Notion Space



FABIAN WERKMEISTER

DATA NOTION SPACE

Fabian Werkmeister

I help you to become fluent in Data and enable you to translate between business & tech ★ Data Scientist @FUNKE ★  Python
 Data Science  Data Engineering

Talks about #data, #python, #mindset, #datascience, and #productivity

Hamburg, Hamburg, Germany · [Contact info](#)

[FREE Data Notion Space](#) 



✓ All Project Resources

✓ Project Management Tools

✓ Tutorials & Learning Roadmap



FABIAN WERKMEISTER

DATA NOTION SPACE

Simply click "Duplicate" and get
your very own Notion Data space

The screenshot shows a Notion workspace interface. At the top, there is a search bar with the placeholder 'Search' and a 'Duplicate' button, both of which are circled with a hand-drawn style orange marker. To the right of the search bar is the Notion logo. A large orange arrow points downwards from the search bar area towards a grid of nine cards. On the far left, there is a circular icon with a left-pointing arrow. The grid of cards includes:

- 1 Project Structure**: An image of a large building under construction with cranes.
- 2 Data Collection**: An image of hands holding chestnuts next to a bucket.
- 3 Data Cleaning**: An image of a hand holding a yellow spray bottle.
- 4 Data Exploration & Analysis**: An image of people looking at a map outdoors.
- 5 Preprocessing & Feature Engineering**: An image of various tools like pliers and a hammer on a table.
- 6 Modeling & Evaluation**: An image of several colorful toy robots.
- 7 Productionize Model**: An image of a large cargo ship.
- 8 Showcase Results**: An image of a clapperboard being held up.
- 9 Summary & Retrospective**: An image of a road sign against a sunset sky.

