

The background features a complex network diagram on the left side, consisting of numerous black dots (nodes) connected by thin grey lines. Some nodes are larger than others. Scattered across the right side of the slide are several thin, light-grey triangles of various sizes and orientations. The overall background is a light grey color.

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Movie Search using IMDb Dataset

Carlos Veríssimo 201907716
Chloe Vidal 202202860
Joaquim Monteiro 201905257

MEIC - PRI 2022/20023
Group 36



Theme **01**

Data Sources and
Datasets **02**

Pipelines **03**



TABLE OF CONTENTS

04 Data
Characterization

05 Conceptual Model

06 Prospective Search
Tasks

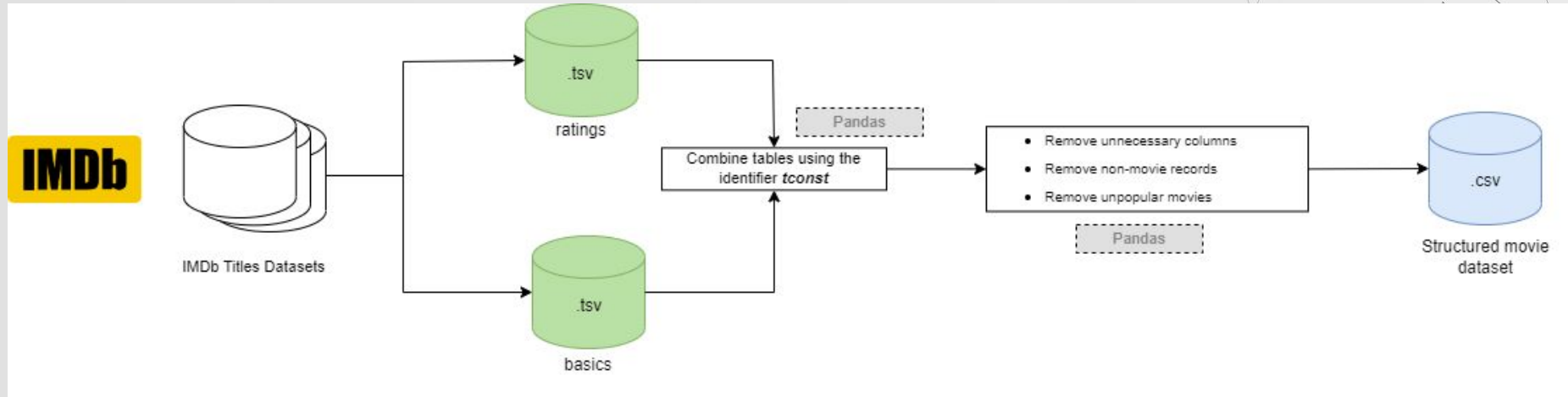
Data Sources / Datasets

The IMDb logo consists of the letters "IMDb" in a bold, black, sans-serif font. The letters are centered within a bright yellow rounded rectangle. In the top-left corner of the overall image, there is a faint, light gray network diagram with several nodes and connecting lines.

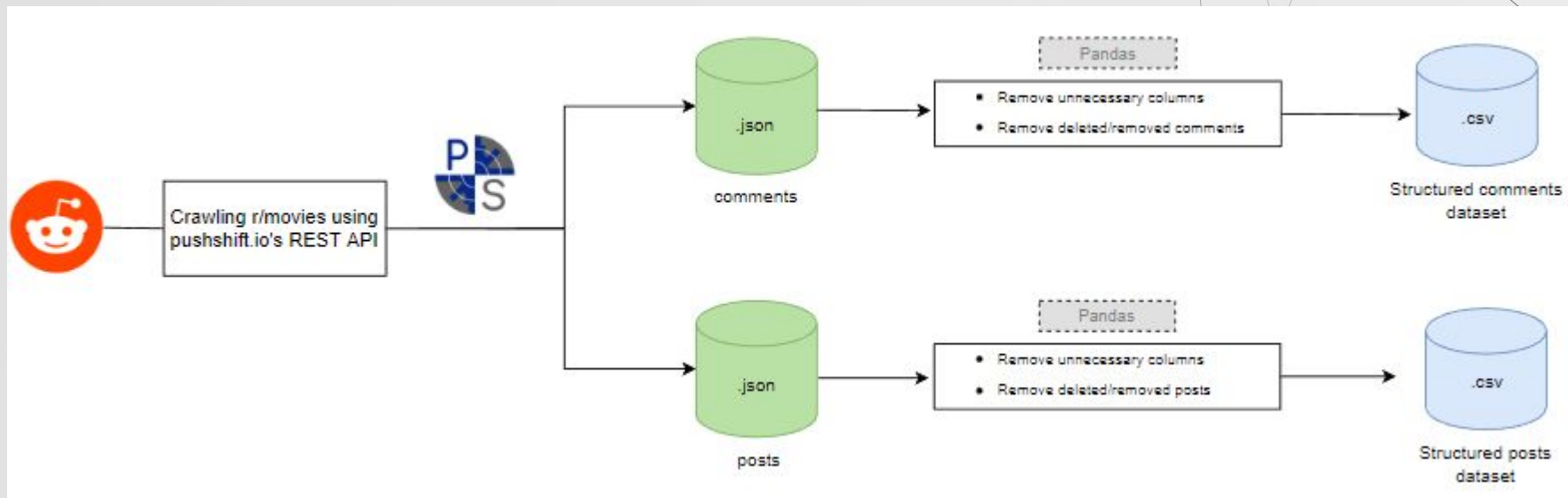
IMDb



Data Processing Pipelines 1/2

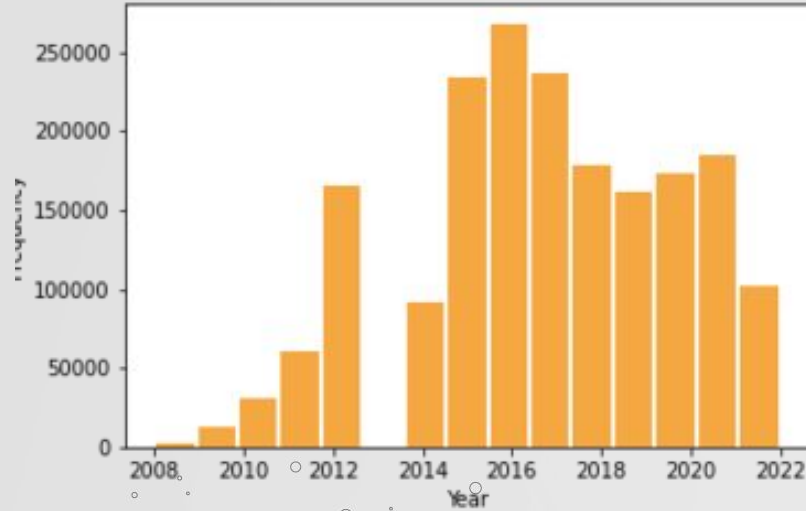


Data Processing Pipelines 2/2

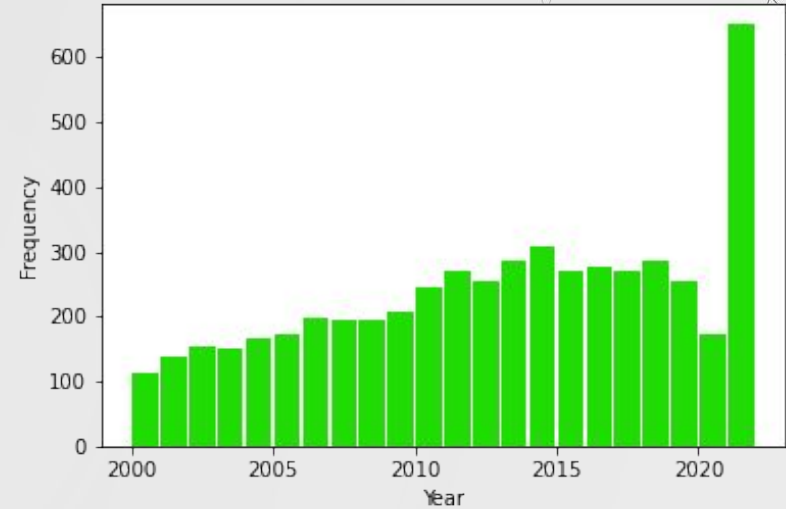


Data Characterization 1/2

Histogram of Posts By Year



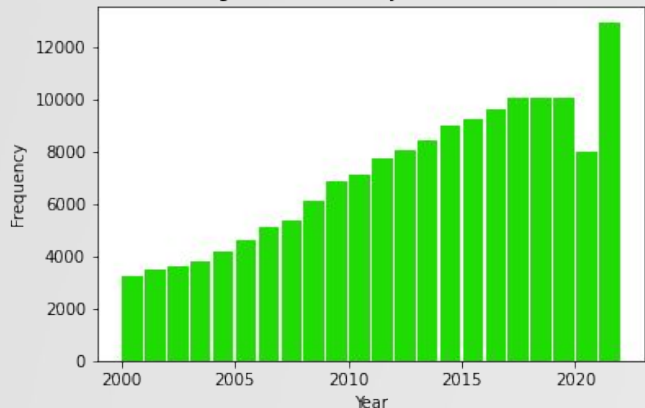
Histogram of Popular Movies by Year (after 1999)



[illegible]

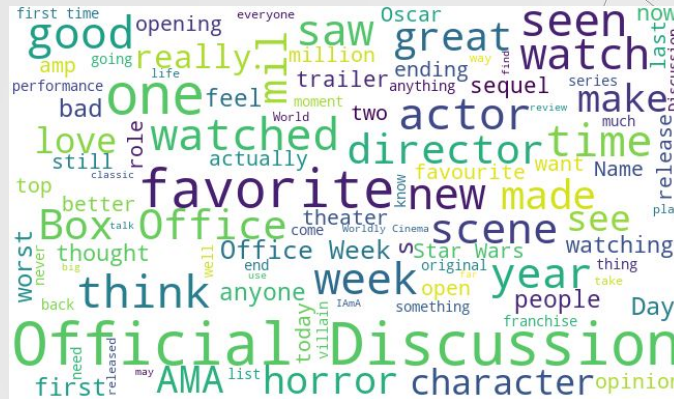
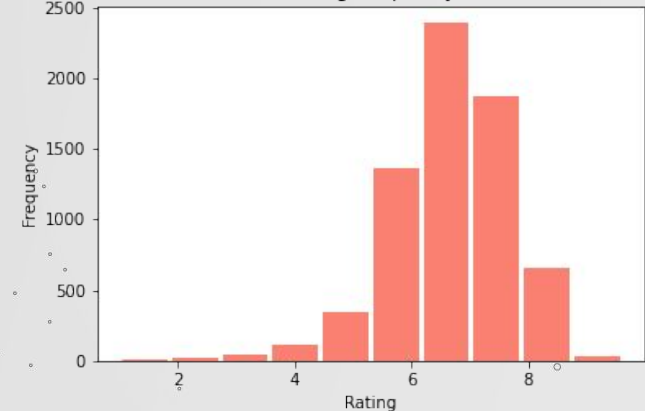
A bar chart illustrating the frequency of tweets per year from 2000 to 2022. The x-axis represents the year, and the y-axis represents the frequency, ranging from 0 to 12,000. The frequency shows a steady increase over the years, with a notable dip in 2021 followed by a sharp rise in 2022.

Year	Frequency
2000	3200
2001	3500
2002	3600
2003	3800
2004	4000
2005	4500
2006	5100
2007	5300
2008	6100
2009	6800
2010	7000
2011	7700
2012	8100
2013	8400
2014	9100
2015	9300
2016	9600
2017	10000
2018	10000
2019	10000
2020	8000
2021	12500
2022	12800

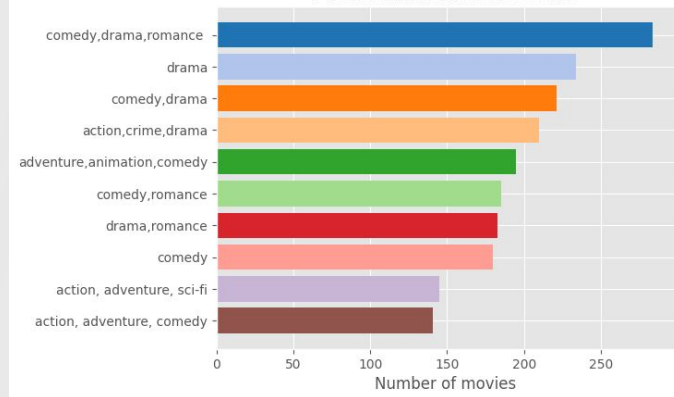


A histogram showing the frequency distribution of ratings. The x-axis is labeled 'Rating' and ranges from 1 to 10. The y-axis is labeled 'Frequency' and ranges from 0 to 2500. The distribution is unimodal and slightly right-skewed, with the highest frequency occurring at a rating of 7 (approximately 2400).

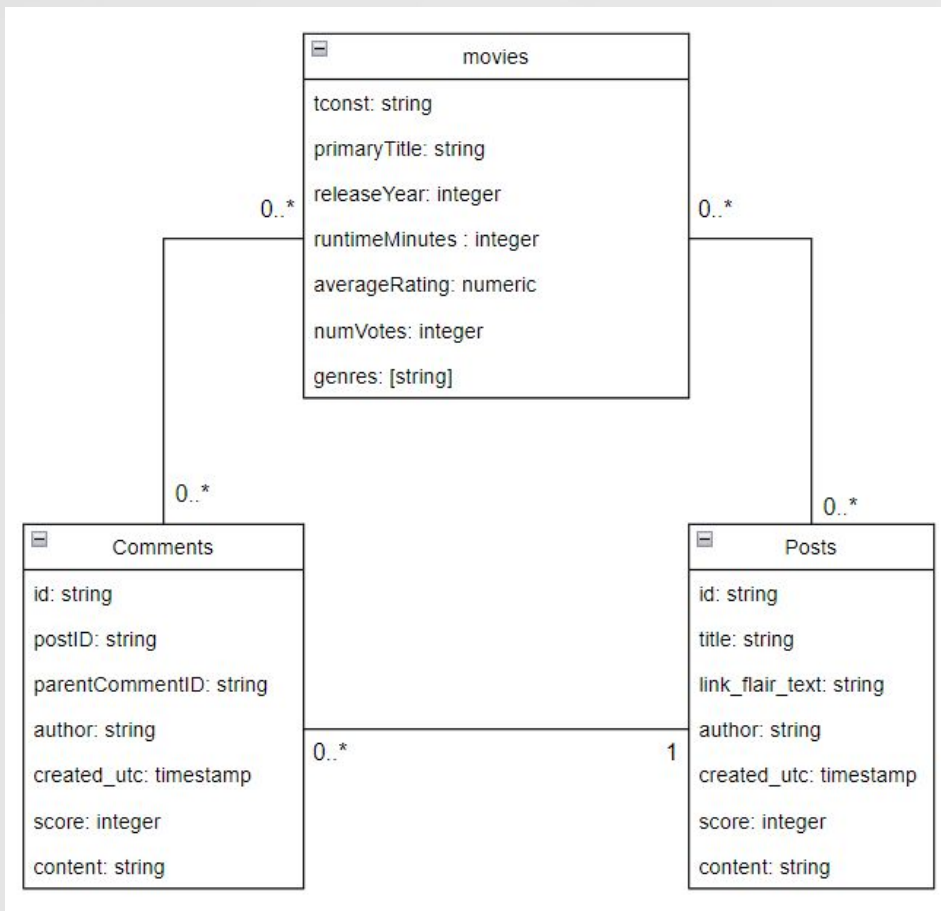
Rating	Frequency
1	10
2	20
3	50
4	100
5	350
6	1350
7	2400
8	1850
9	650
10	50



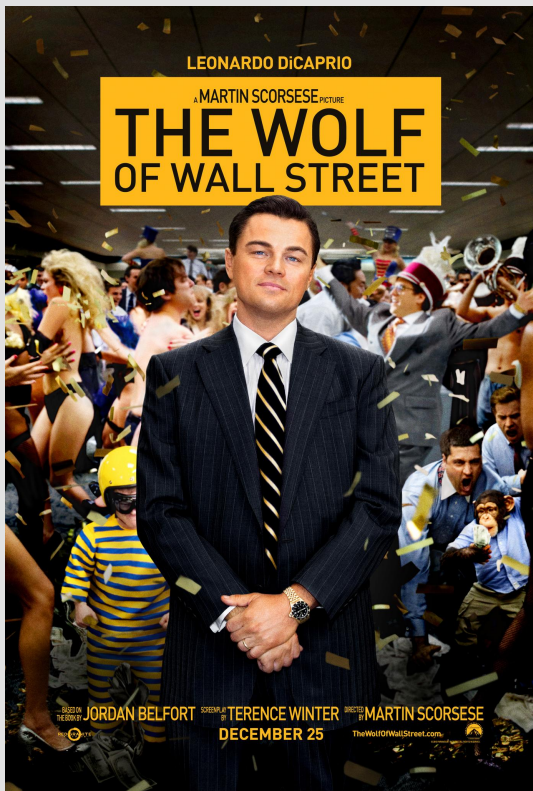
Genre Combination	Number of movies
comedy,drama,romance	280
drama	235
comedy,drama	220
action,crime,drama	210
adventure,animation,comedy	195
comedy,romance	185
drama,romance	185
comedy	180
action,adventure,sci-fi	145
action,adventure,comedy	140



Conceptual Model



Prospective Search Tasks



The Wolf of Wall Street



Quaalude crawl. Enough said.

I loved it, it really didn't feel like three hours.

"Those are Rookie numbers."

I enjoyed it. Saw the double cross at the end coming a mile away though.

I'm going to have to go see that again, without my mom.

Load More