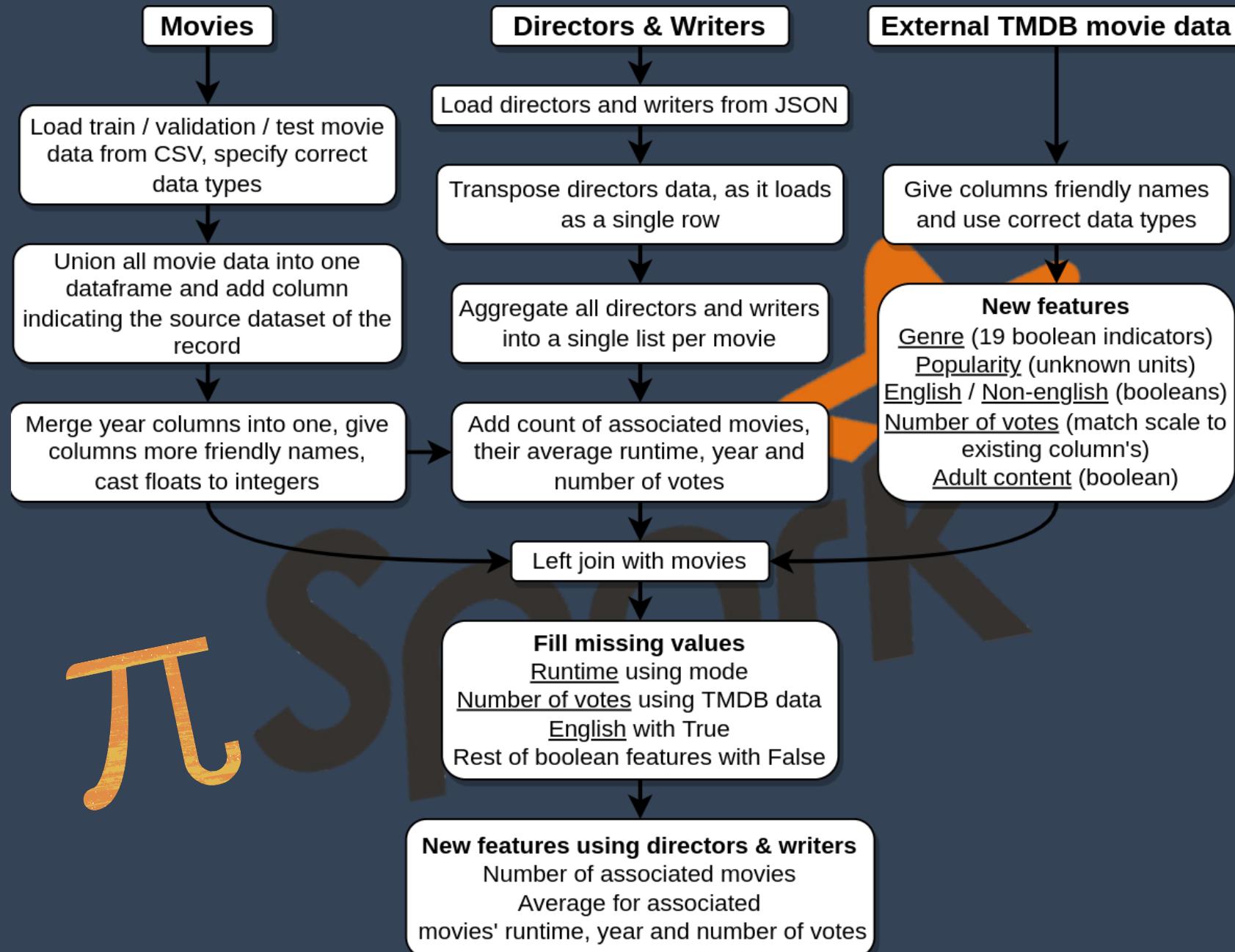




CHAPTER 1 PIPELINE DESIGN



NOW
SHOWING

CHAPTER 2 DATA CLEANING

- STARRING MOVIES DATA
 - MISSING YEARS, RUNTIME MINUTES AND NUMBER OF VOTES
 - SPECIAL CHARACTERS IN MOVIE TITLES
- DIRECTORS & WRITERS
 - NESTED FORMAT OF FILES
- EXTERNAL TMDB DATA
 - FETCHED EXTERNAL DATA USING TMDB API FOR EACH MOVIE ID
 - DID NOT ALLOW TO LEAK DEPENDENT VARIABLE

CHAPTER 3 FEATURE ENGINEERING

- FEATURING ONE-HOT ENCODED MOVIE GENRES
- TOTAL NUMBER OF MOVIES ASSOCIATED WITH DIRECTORS & WRITERS
- AVERAGE RUNTIME, YEAR & NUMBER OF VOTES OF MOVIES ASSOCIATED WITH DIRECTORS & WRITERS
- SENTIMENT POLARITY OF MOVIE OVERVIEWS

CHAPTER 4 MODEL VALIDATION SCORE (SPOILER ALERT)

CASTING

WITHOUT EXTERNAL DATA 79%

WITH TMDB EXTERNAL DATA 81%

WITH DIRECTORS/WRITERS 82%

SENTIMENT POLARITY 84%

HYPERPARAMETER TUNING 85%

