

Data-Analytics-Mid-Bootcamp-Project

Bollywood data set



By Akshaykumar Halkude

Insights and Information from Data

► Bollywood Actor Ranking

	actorId	actorName	movieCount	ratingSum	normalizedMovieRank	googleHits	normalizedGoogleRank	normalizedRating
0	373	Aamir Khan	11	1170	9.44862	2460000	7.34283	10.00000

► Bollywood Director Ranking

	directorId	directorName	movieCount	ratingSum	normalizedMovieRank	googleHits	normalizedGoogleRank	normalizedRating
0	1	Rajkumar Hirani	3	440	10.00000	146000	2.07796	10.00000

► Bollywood Movie Detail

	imdbId	title	releaseYear	releaseDate	genre	writers	actors	directors	sequel	hitFlop
0	tt0118578	Albela	2001	20 Apr 2001	Romance	Honey Irani (screenplay) Honey Irani (story)...	Govinda Aishwarya Rai Bachchan Jackie Shro...	Deepak Sareen	0.0	2

Cleaning and standardizing data

► Problem:

- data set contain missing values with different data types
- headers are not in standard format
- data has na values
- meaningless information

► Solution:

- Standardized header names by making all header in lower case.
- Handling missing values by taking median for integer data type and rest data dropped
- drop na, writers and release date also sequel

actor.isna().sum()

actorid	0
actorname	0
moviecount	0
ratingsum	0
normalizedmovierank	2
googlehits	0
normalizedgooglerank	112
normalizedrating	0
dtype: int64	

director.isna().sum()

directorid	0
directorname	1
moviecount	0
ratingsum	0
normalizedmovierank	0
googlehits	0
normalizedgooglerank	0
normalizedrating	0
dtype: int64	

movie.isna().sum()

imdbid	0
title	0
releaseyear	0
releasedate	53
genre	2
writers	119
actors	3
directors	4
sequel	3
hitflop	0
dtype: int64	

sequel

0.0

0.0

0.0

0.0



Asked interesting questions

- ▶ looking which movie title has hit and flop
- ▶ actor name and their movie count
- ▶ Top actor who has highest movie count
- ▶ Top 5 actor who has highest movie count
- ▶ Top 5 director who has highest movie count
- ▶ Top director who has highest movie count
- ▶ List of movies released in 2014
- ▶ Top 5 actor who has highest google hits
- ▶ Top actor who has highest google hits
- ▶ Top 5 director who has highest google hits
- ▶ Top director who has highest google hits



Dashboards

► Actor:

<https://public.tableau.com/app/profile/akshaykumar.halkude/viz/ActorDashboard/ActorDashboard?publish=yes>

► Director:

<https://public.tableau.com/app/profile/akshaykumar.halkude/viz/Directordashboard/Directorsdashboard?publish=yes>

► Movie:

https://public.tableau.com/app/profile/akshaykumar.halkude/viz/Moviedashboard_16541802722080/Moviedashboard?publish=yes



Summary:

- Akshay kumar is top actor who has highest movie count
 - Priyadarshan is top director who has highest movie count
 - Salman Khan is top actor who has highest google hits
 - Raj Kumar Gupta is top director who has highest google hits
- 

