SQPSK mini project

Objective: The purpose of this project is to analyze box office collections from **INDIAN BOX OFFICE** using SQL and PySpark. You are provided with 4 datasets as outlined below:

- Boxoffice_Fact
- Director_dim
- Genere_dim
- Language_dim

A detailed data dictionary provided, please refer that for better understanding. Your task is to write SQL and PySpark queries to extract meaningful insights, perform data aggregation, and generate reports based on the datasets.

Note: *There we lot of special characters, duplicates and trailing spaces in few columns. You have to clean them then you start answering the following questions.*

Questions to be answered:

SL#	Question	SQL	Python
1	Write a query to get Total films released ?	Yes	Yes
2	Write a query to get Total budget ?	Yes	Yes
3	Write a query to get Total worldwide collection ?	Yes	Yes
4	Write a query to get Total First day collection worldwide ?	Yes	Yes
5	Write a query to get Total Overseas collection ?	Yes	Yes
6	Write a query to get Total India Gross collection ?	Yes	Yes
7	Top 10 filmsbased on world wide collections. Display films,collections	Yes	Yes
8	Total Number of films released by year	Yes	Yes
9	Top 10 filmsbased on india collections.Display films,collections	Yes	Yes
10	Top 10 filmsbased on overses collections.Display films,collections	Yes	Yes
11	Top 10 filmsbased on firstday collections.Display films,collections	Yes	Yes
12	Weekday wise films released,Display week name and no of films released	Yes	Yes
13	Write a query to get OTT platofrm wise movies count	Yes	Yes
14	Top 10 Directors by films released	Yes	Yes
15	Top 10 directors by world wide collection	Yes	Yes
16	Top10 lead actors by world wide collection	Yes	Yes
17	Top 10 movies by IMDb rating	Yes	Yes
18	Bottom 10 movies by IMDb rating	Yes	Yes
19	Write a query to get 5 longest run time movies	Yes	Yes
20	Write a query to get 5 shortest run time movies	Yes	Yes
21	Top7 movies by world wide collection in Bollywood	Yes	Yes
22	Top7 movies by world wide collection in Tollywood	Yes	Yes

SL#	Question	SQL	Python
23	Top7 movies by world wide collection in Kollywood	Yes	Yes
24	Top7 movies by world wide collection in Sandalwood	Yes	Yes
25	Top7 movies by world wide collection in Mollywood	Yes	Yes
26	Write query to display industry and verdict wise films count	Yes	Yes
27	Write query to get films based on budget in Bollowood	Yes	Yes
28	Write query to get films based on budget in Tollywood	Yes	Yes
29	Write query to get films based on budget in Kollywood	Yes	Yes
30	Write query to get films based on budget in Sandalwood	Yes	Yes
31	Write query to get films based on budget in Mollywood	Yes	Yes
32	Top 5 movies by IMDb rating from Bollowood	Yes	Yes
33	Top 5 movies by IMDb rating from Kollywood	Yes	Yes
34	Top 5 movies by IMDb rating from Tollywood	Yes	Yes
35	Top 5 movies by IMDb rating from Sandalwood	Yes	Yes
36	Top 5 movies by IMDb rating from Mollywood	Yes	Yes
37	Write a query to get language wise budget?	Yes	Yes
38	Write a query to get language wise how many directors are there?	Yes	Yes
39	Write a query to get language wise worldwide total collection ?	Yes	Yes
40	Write a query to get language, lead actor/actress wise films they acted?	Yes	Yes
41	Write a query to get language, year wise films released?	Yes	Yes
42	Write a query to get films which was not released on overseas?	Yes	Yes
43	Write a query to get language wise top 3 longest runtime moves?	Yes	Yes
44	Write a query to get language wise bottom 5 shortest runtime moves?	Yes	Yes
45	Write a query to get language wise top 5 films based first day collections?	Yes	Yes
46	Write a query to get language wise top 5 films based India gross collections?	Yes	Yes
47	Write a query to get language, Director wise films count?	Yes	Yes
48	Write a query to get language wise OTT platofrm wise fims count?	Yes	Yes
49	What are top 10 fimns based language and first day collection?	Yes	Yes
50	Write a query to get director wise number of fims released in from year 2017 to 2019	Yes	Yes
51	Write a query to get director wise world wide collections?	Yes	Yes
52	Write a query to get director wise first day world wide collections?	Yes	Yes
53	Write a query to get director wise India gross collections?	Yes	Yes
54	Write a query to get director wise overseas collections?	Yes	Yes
55	Write a query to get director, lead actor/actress and number of films?	Yes	Yes
56	Write a query to get films which is having budget on between 150 crores and 277 crores?	Yes	Yes
57	Write a query to get director, week name wise films released?	Yes	Yes
58	Write a query to get OTT platofrm and director wise films released?	Yes	Yes
59	Write a query to get director wise films released on Friday only?	Yes	Yes
60	Write a query to get films based on IMDb reating between 6.5 and 7.7?	Yes	Yes
61	Write a query to get director,films and IMDb ratings?	Yes	Yes
62	Write a query to get films with highest budget based flop verdict?	Yes	Yes

SL#	Question	SQL	Python
63	Write a query to get total number of directors?	Yes	Yes
64	Write a query to get vedridct wise total films released?	Yes	Yes
65	Write a query to get top 10 directors based number of films?	Yes	Yes
66	Write a query to get top 5 directors based on world wide collections and also industry wise?	Yes	Yes
67	Write a query to get top 3 directors based on India gross collections in Bollowood industry?	Yes	Yes
68	Write a query to get top 3 directors based on India gross collections in Tollowood industry?	Yes	Yes
69	Write a query to get top 3 directors based on India gross collections in Kollowood industry?	Yes	Yes
70	Write a query to get top 3 directors based on India gross collections in Mollowood industry?	Yes	Yes
71	Write a query to get top 3 directors based on India gross collections in Sandalwood industry?	Yes	Yes
72	Write to get total number of generes?	Yes	Yes
73	Write query to get director,language,genere films count?	Yes	Yes
74	Write a query to genere wise budget?	Yes	Yes
75	Write a query to get genere wise first day worldwide collections?	Yes	Yes
76	Write a query to get genere wise overseas collections?	Yes	Yes
77	Write a query to get genere wise India gross collections?	Yes	Yes
78	Write a query to get genere wise top 2 longest run time movies?	Yes	Yes
79	Write a query to get genere wise bottom shortest runtime movies?	Yes	Yes
80	Write a query to get verdict, genere wise films released?	Yes	Yes
81	Write a query to get genere, OTT platform wise films count?	Yes	Yes
82	Write a query to get genere wise films count?	Yes	Yes
83	Write a query to get genere wise films count in Tollowood Industry?	Yes	Yes
84	Write a query to get genere wise films count in Kollowood Industry?	Yes	Yes
85	Write a query to get genere wise films count in Mollowood Industry?	Yes	Yes
86	Write a query to get genere wise films count in Bollowood Industry?	Yes	Yes
87	Write a query to get genere wise films count in Sandalwood Industry?	Yes	Yes
88	Write a query to get lead actors/actress wise,genere and films count?	Yes	Yes
89	Write a query to get 5th rank movie based on Worldwide total collections?	Yes	NA
90	Write a query to get 5th rank movie by industry wise based on First day worldwide collections?	Yes	NA
91	Write a query to get 3rd rank movie by industry wise based on IMDb Ratings?	Yes	NA
92	Write a stored procedure to get top 5 films based on India gross colelctions?	Yes	NA
93	Write a stored procdure to get industry Budget and Worldwide total collections? If we call any industry it will display industry,budget and wordwide total collections?	Yes	NA
94	Calculate YoY% Budget growth?	Yes	Yes
95	Calculate YoY% Worldwide total collelctions growth?	Yes	Yes
96	Calculate YoY% Indian Gross colelctions growth?	Yes	Yes
97	Create a stored procedure to get Total films count, world wide total colelctions?	Yes	NA

All the best!