

⚡ Powerplay Assessment ⚡

Done by : **KASI**

```
In [1]: import pandas as pd
```

◆ Set 1:-

Use data set to answer the following questions using SQL :-

https://raw.githubusercontent.com/rashida048/Datasets/master/movie_dataset.csv

(Read Json data using Python / R Script and create Movie table in database)

```
In [2]: df = pd.read_csv('movie_dataset.csv')

df['release_date'] = pd.to_datetime(df['release_date'])

ndf = df[['budget', 'title', 'production_companies', 'production_countries', 'release_date', 'runtime', 'vote_average', 'popularity', 'revenue']]

ndf['runtime'] = ndf['runtime'].fillna(ndf['runtime'].mode()[0])
ndf['release_date'] = ndf['release_date'].fillna(ndf['release_date'].mode()[0])

display(ndf.isna().sum())
```

C:\Users\KASI Profession\AppData\Local\Temp\ipykernel_22140\959624700.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
ndf['runtime'] = ndf['runtime'].fillna(ndf['runtime'].mode()[0])
C:\Users\KASI Profession\AppData\Local\Temp\ipykernel_22140\959624700.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
ndf['release_date'] = ndf['release_date'].fillna(ndf['release_date'].mode()[0])
```

```
budget          0
title           0
production_companies  0
production_countries  0
release_date     0
runtime         0
vote_average    0
popularity      0
revenue         0
dtype: int64
```

```
In [3]: movie_data = ndf.copy()

# Converting the 'production_companies' and 'production_countries' columns from JSON string to python List
# json.loads didnt work
movie_data['production_companies'] = movie_data['production_companies'].apply(eval)
movie_data['production_countries'] = movie_data['production_countries'].apply(eval)

movie_data
```

Out[3]:

	budget	title	production_companies	production_countries	release_date	runtime	vote_average	popularity	revenue
0	237000000	Avatar	[{'name': 'Ingenious Film Partners', 'id': 289...	[{'iso_3166_1': 'US', 'name': 'United States o...	2009-12-10	162.0	7.2	150.437577	2787965087
1	300000000	Pirates of the Caribbean: At World's End	[{'name': 'Walt Disney Pictures', 'id': 2}, {'...	[{'iso_3166_1': 'US', 'name': 'United States o...	2007-05-19	169.0	6.9	139.082615	961000000
2	245000000	Spectre	[{'name': 'Columbia Pictures', 'id': 5}, {'nam...	[{'iso_3166_1': 'GB', 'name': 'United Kingdom'...	2015-10-26	148.0	6.3	107.376788	880674609
3	250000000	The Dark Knight Rises	[{'name': 'Legendary Pictures', 'id': 923}, {'...	[{'iso_3166_1': 'US', 'name': 'United States o...	2012-07-16	165.0	7.6	112.312950	1084939099
4	260000000	John Carter	[{'name': 'Walt Disney Pictures', 'id': 2}]	[{'iso_3166_1': 'US', 'name': 'United States o...	2012-03-07	132.0	6.1	43.926995	284139100
...
4798	220000	El Mariachi	[{'name': 'Columbia Pictures', 'id': 5}]	[{'iso_3166_1': 'MX', 'name': 'Mexico'}, {'iso...	1992-09-04	81.0	6.6	14.269792	2040920
4799	9000	Newlyweds	[]	[]	2011-12-26	85.0	5.9	0.642552	0
4800	0	Signed, Sealed, Delivered	[{'name': 'Front Street Pictures', 'id': 3958}...	[{'iso_3166_1': 'US', 'name': 'United States o...	2013-10-13	120.0	7.0	1.444476	0
4801	0	Shanghai Calling	[]	[{'iso_3166_1': 'US', 'name': 'United States o...	2012-05-03	98.0	5.7	0.857008	0
4802	0	My Date with Drew	[{'name': 'rusty bear entertainment', 'id': 87...	[{'iso_3166_1': 'US', 'name': 'United States o...	2005-08-05	90.0	6.3	1.929883	0

4803 rows × 9 columns

In [4]:

```
movie_data['production_companies'] = movie_data['production_companies'].apply(lambda x: [company['name'] for company in x])
movie_data['production_countries'] = movie_data['production_countries'].apply(lambda x: [country['name'] for country in x])

movie_data.head()
```

Out[4]:

	budget	title	production_companies	production_countries	release_date	runtime	vote_average	popularity	revenue
0	237000000	Avatar	[Ingenious Film Partners, Twentieth Century Fo...	[United States of America, United Kingdom]	2009-12-10	162.0	7.2	150.437577	2787965087
1	300000000	Pirates of the Caribbean: At World's End	[Walt Disney Pictures, Jerry Bruckheimer Films...	[United States of America]	2007-05-19	169.0	6.9	139.082615	961000000
2	245000000	Spectre	[Columbia Pictures, Danjaq, B24]	[United Kingdom, United States of America]	2015-10-26	148.0	6.3	107.376788	880674609
3	250000000	The Dark Knight Rises	[Legendary Pictures, Warner Bros., DC Entertai...	[United States of America]	2012-07-16	165.0	7.6	112.312950	1084939099
4	260000000	John Carter	[Walt Disney Pictures]	[United States of America]	2012-03-07	132.0	6.1	43.926995	284139100

In [5]:

```
df = movie_data.copy()
```

In [6]:

```
unnesting = ['production_companies', 'production_countries']
for column in unnesting:
    df = df.explode(column)
```

In [7]:

```
df
```

Out[7]:

	budget	title	production_companies	production_countries	release_date	runtime	vote_average	popularity	revenue
0	237000000	Avatar	Ingenious Film Partners	United States of America	2009-12-10	162.0	7.2	150.437577	2787965087
0	237000000	Avatar	Ingenious Film Partners	United Kingdom	2009-12-10	162.0	7.2	150.437577	2787965087
0	237000000	Avatar	Twentieth Century Fox Film Corporation	United States of America	2009-12-10	162.0	7.2	150.437577	2787965087
0	237000000	Avatar	Twentieth Century Fox Film Corporation	United Kingdom	2009-12-10	162.0	7.2	150.437577	2787965087
0	237000000	Avatar	Dune Entertainment	United States of America	2009-12-10	162.0	7.2	150.437577	2787965087
...
4800	0	Signed, Sealed, Delivered	Muse Entertainment Enterprises	United States of America	2013-10-13	120.0	7.0	1.444476	0
4801	0	Shanghai Calling	NaN	United States of America	2012-05-03	98.0	5.7	0.857008	0
4801	0	Shanghai Calling	NaN	China	2012-05-03	98.0	5.7	0.857008	0
4802	0	My Date with Drew	rusty bear entertainment	United States of America	2005-08-05	90.0	6.3	1.929883	0
4802	0	My Date with Drew	lucky crow films	United States of America	2005-08-05	90.0	6.3	1.929883	0

22780 rows × 9 columns

In [8]: df.isna().sum()

Out[8]: budget 0
title 0
production_companies 394
production_countries 184
release_date 0
runtime 0
vote_average 0
popularity 0
revenue 0
dtype: int64

In [9]: df['production_companies'] = df['production_companies'].fillna('unknown')
df['production_countries'] = df['production_countries'].fillna('unknown')

In [10]: df.isna().sum()

Out[10]: budget 0
title 0
production_companies 0
production_countries 0
release_date 0
runtime 0
vote_average 0
popularity 0
revenue 0
dtype: int64

In [11]: df.to_csv('powerplay_movie_cleaned.csv',index=False)

In [12]: df.shape

Out[12]: (22780, 9)

📁 Loaded to Bigquery

⚡ 1. Which is the movie(s) with the smallest runtime?

```
SELECT DISTINCT
  title AS smallest_movie,
  runtime AS duration_in_mins
FROM
  `powerplay.powerplay_movie_cleaned`
WHERE
  runtime = (
    SELECT MIN(runtime)
    FROM `powerplay.powerplay_movie_cleaned`
  )
GROUP BY
  1, 2;
```

12 SELECT DISTINCT

13 title AS smallest_movie,

14 runtime AS duration_in_mins

15 FROM

16 `powerplay.powerplay_movie_cleaned`

17 WHERE

18 runtime = (

19 SELECT MIN(runtime)

20 FROM `powerplay.powerplay_movie_cleaned`

21)

22 GROUP BY

23 1, 2;

Press Alt

Query results

SAVE RESULTS EX

JOB INFORMATIONRESULTSCHARTJSONEXECUTION DETAILSEXECUTION GRAPH

Row	smallest_movie	duration_in_mins
1	Should've Been Romeo	0.0
2	Sharkskin	0.0
3	Hum To Mohabbat Karega	0.0
4	Dil Jo Bhi Kahey...	0.0
5	Harrison Montgomery	0.0
6	Light from the Darkroom	0.0
7	America Is Still the Place	0.0

Results per page: 501 – 35 of 35

🔥 Which is the movie(s) with the highest runtime?

```
SELECT DISTINCT
  title AS Lengthiest_movie,
  runtime AS duration_in_mins
FROM
  `powerplay.powerplay_movie_cleaned`
WHERE
  runtime = (
    SELECT MAX(runtime)
    FROM `powerplay.powerplay_movie_cleaned`
  )
GROUP BY
  1, 2;
```

powerplay-set1

RUN

SAVE QUERY

SHARE

SCHEDULE

MORE

29

30 SELECT DISTINCT

31 title AS Lengthiest_movie,

32 runtime AS duration_in_mins

33 FROM

34 `powerplay.powerplay_movie_cleaned`

35 WHERE

36 runtime = (

37 SELECT MAX(runtime)

38 FROM `powerplay.powerplay_movie_cleaned`

39)

40 GROUP BY

41 1, 2;

Query results

SAVE RESULTS

JOB INFORMATIONRESULTSCHARTJSONEXECUTION DETAILSEXECUTION GRAPH

Row	Lengthiest_movie	duration_in_mins
1	Carlos	338.0

- ◆ Insight:
- There are 35 movies with duration as 0.0 mins of runtime.
 - Carlos movie is lengthiest movie with duration of 338 minutes.

🔥 2.Take the top 5 production houses (by budget) and list their top 5 most popular movies, their revenue and vote_average.

```
WITH top_production_houses AS (  
  SELECT DISTINCT  
    production_companies,  
    DENSE_RANK() OVER (ORDER BY MAX(budget) DESC) AS rnk  
  FROM  
    `powerplay.powerplay_movie_cleaned`  
  GROUP BY 1  
  ORDER BY 2  
  LIMIT 5  
)  
SELECT production_house, movie_name, revenue, vote_average  
FROM (  
  SELECT  
    p.production_companies as production_house,  
    p.title AS movie_name,  
    p.revenue,  
    p.vote_average,  
    ROW_NUMBER() OVER (PARTITION BY p.production_companies ORDER BY p.popularity DESC) AS popularity_rnk  
  FROM  
    `powerplay.powerplay_movie_cleaned` p  
  JOIN  
    top_production_houses tp  
  USING (production_companies) ) as c  
WHERE  
  popularity_rnk <= 5  
ORDER BY  
  production_house;
```

powerplay-set1

RUNSAVE QUERYSHARESHARESCHEDULEMORE

```
43 -- 2. Take the top 5 production houses (by budget) and list their top 5 most popular movies, their revenue and vote_average.  
44  
45 WITH top_production_houses AS (  
46   SELECT DISTINCT  
47     production_companies,  
48     DENSE_RANK() OVER (ORDER BY MAX(budget) DESC) AS rnk  
49   FROM  
50     `powerplay.powerplay_movie_cleaned`  
51   GROUP BY 1  
52   ORDER BY 2  
53   LIMIT 5  
54 )  
55 SELECT production_house, movie_name, revenue, vote_average  
56 FROM (  
57   SELECT  
58     p.production_companies as production_house,  
59     p.title AS movie_name,  
60     p.revenue,  
61     p.vote_average,  
62     ROW_NUMBER() OVER (PARTITION BY p.production_companies ORDER BY p.popularity DESC) AS popularity_rnk  
63   FROM  
64     `powerplay.powerplay_movie_cleaned` p  
65   JOIN  
66     top_production_houses tp  
67   USING (production_companies) ) as c  
68 WHERE  
69   popularity_rnk <= 5  
70 ORDER BY  
71   production_house;
```

Output:

Query results

SAVE RESULTS

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	production_house	movie_name	revenue	vote_average		
1	Jerry Bruckheimer Films	Pirates of the Caribbean: The C...	655011224	7.5		
2	Jerry Bruckheimer Films	Pirates of the Caribbean: Dead ...	1065659812	7.0		
3	Jerry Bruckheimer Films	Pirates of the Caribbean: Dead ...	1065659812	7.0		
4	Jerry Bruckheimer Films	Pirates of the Caribbean: Dead ...	1065659812	7.0		
5	Jerry Bruckheimer Films	Pirates of the Caribbean: Dead ...	1065659812	7.0		
6	Marvel Studios	Guardians of the Galaxy	773328629	7.9		
7	Marvel Studios	Guardians of the Galaxy	773328629	7.9		
8	Marvel Studios	Captain America: Civil War	1153304495	7.1		
9	Marvel Studios	The Avengers	1519557910	7.4		

Results per page: 501 – 25 of 25

◆ Insights:

Top Production houses by budget:

- Walt Disney Pictures

- Moving Picture Company (MPC)
- Jerry Bruckheimer Films
- Second Mate Productions
- Marvel Studios

and each production houses top 5 productions are listed.

⚡ 3. List the production house for every year from 2000-2016 which has released the most number of movies in that year.

```
WITH yearlymoviecounts AS (
  SELECT
    EXTRACT(YEAR FROM release_date) AS release_year,
    production_companies,
    COUNT(DISTINCT title) AS num_of_movies
  FROM
    `powerplay.powerplay_movie_cleaned`
  WHERE
    EXTRACT(YEAR FROM release_date) BETWEEN 2000 AND 2016
    AND production_companies != 'unknown'
  GROUP BY
    release_year, production_companies
  ORDER BY
    release_year, num_of_movies DESC
),topproductioncompanies AS (
  SELECT
    *,
    ROW_NUMBER() OVER (PARTITION BY release_year ORDER BY num_of_movies DESC) AS rn
  FROM
    yearlymoviecounts
)
SELECT
  release_year,production_companies,num_of_movies
FROM
  topproductioncompanies
WHERE
  rn = 1
ORDER BY
  release_year;
```

```
77 WITH yearlymoviecounts AS (
78   SELECT
79     EXTRACT(YEAR FROM release_date) AS release_year,
80     production_companies,
81     COUNT(DISTINCT title) AS num_of_movies
82   FROM
83     `powerplay.powerplay_movie_cleaned`
84   WHERE
85     EXTRACT(YEAR FROM release_date) BETWEEN 2000 AND 2016
86     AND production_companies != 'unknown'
87   GROUP BY
88     release_year, production_companies
89   ORDER BY
90     release_year, num_of_movies DESC
91 ),topproductioncompanies AS (
92   SELECT
93     *,
94     ROW_NUMBER() OVER (PARTITION BY release_year ORDER BY num_of_movies DESC) AS rn
95   FROM
96     yearlymoviecounts
97 )
98 SELECT
99   release_year,production_companies,num_of_movies
100 FROM
101   topproductioncompanies
102 WHERE
103   rn = 1
104 ORDER BY
105   release_year;
```

Output:

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	release_year	production_companies	num_of_movies			
1	2000	Warner Bros.	14			
2	2001	Warner Bros.	15			
3	2002	Paramount Pictures	15			
4	2003	Universal Pictures	14			
5	2004	Warner Bros.	14			
6	2005	Universal Pictures	15			
7	2006	Warner Bros.	17			
8	2007	Warner Bros.	14			
9	2008	Relativity Media	20			
10	2009	Relativity Media	19			
11	2010	Relativity Media	18			
12	2011	Columbia Pictures	14			

Results per page: 501 – 17 of 17

◆ Insights:

- For all the 17 years from 2000 to 2016 , every years Highest no.of movies producing Productions are listed with no.of productions.
- Most of years Warner Bros has made many movies followed by Relative Media production House.

⚡ 4. You are going to invest all your life's savings in a production company. You have two choices: "Marvel Studios" and "DC Comics". Which company would you bet on? This is an open ended question. Define your own metrics to measure which one is a better investment opportunity and defend your analysis.

- If I had to pick between *Marvel Studios* and *DC Comics* for my investments, I'd go with **Marvel** .

Here's why:

I'm a fan of Comics and I love both Marvel and DC Theoretically but Cinematically I would prefer **Marve1** because of the cinematic experience, VFX, post credit scenes and writer of Marvel comics **STAN LEE** 's cameo presence makes it enjoyable than DC.

Marvel movies are fun and family-friendly, reaching a big audience, even families. DC tends to be darker, so it might not appeal to as many people. Marvel keeps trying new flavors and changes based on what people like. DC can be slower to do that.

Marvel is also good at exploring new things. They introduce new characters and listen to what fans like. DC, while sticking with its famous heroes, doesn't change as quickly.

Marvel makes a superhero world where all the characters are connected. This makes fans stick around and keeps them excited for what's next. DC, on the other hand, usually keeps its superheroes separate, which means a bad movie doesn't affect the others, but a good one doesn't boost them much either.

Marvel operates like a seamless universe, not just a collection of franchises. This interconnected approach cultivates fan loyalty and sustains repeat viewership. In contrast, DC tends to treat its superheroes as standalone franchises, minimizing the impact of a failure but also diluting the momentum of a success.

Think of superhero movies like a candy vending machine. Marvel is like a big, well-run machine giving you a variety of flavors that all connect. If one flavor isn't great, you're still excited about the next one. DC is like several machines with just one flavor each. If one flavor doesn't work, that machine stops, and even if one is great, it doesn't really help the others.

In short, Marvel knows how to keep things exciting and adapt to what people want & thereby yielding a profit and based on the BoxOffice collections for sure we can say Marvel studios Kevin Feigi & his team is Pro in it.

conclusion:

- So, if I want my life savings to be safe and successful, I'd go with **Marvel** and invest in **Marvel Studios** to keep my principle safe and high yield of shares and profit.***

◆ ◆ ◆ ◆

◆ ◆ ◆ ◆

◆ ◆ ◆ ◆