Data Exploration of Netflix movies and Tv shows

```
create table netflix
 (show_id varchar(50),
 type varchar(50),
 title varchar(100),
 director varchar(100),
 cast varchar(100),
 country varchar(100),
 date_added varchar(50),
 release_year int,
 rating varchar(50),
 duration varchar(50),
 listed_in varchar(200),
 Description varchar(200));


 load data local infile 'C:/Users/FUTECH COMPUTER/Desktop/dataset/netflix/Netflix.CSV'
into table netflix
fields terminated by ','
enclosed by '"'
lines terminated by '\r\n'
ignore 1 rows;

 ##First thing i usually would do is to have a look at dataset
```

This phase is when i clean the dataset, check for duplicates and null values

```
select * from netflix;
```

```
describe netflix;
```

```
select count(*) from netflix
;
```

having done that lets begin exploration

```
show tables
```

```
SELECT avg(release_year) from netflix;
```

we see here that the average release year is 2012

```
select release_year, count(*) as stats From netflix
group by release_year
order by stats desc
limit 10;
```

more movies and tv shows were released in 2017 than any other year

```
SELECT show_id, title, release_year, date_added

FROM netflix

ORDER BY release_year DESC

LIMIT 10;


SELECT CONCAT('There was a total of ', COUNT(*), ' ', type, ' shows released in ', release_year) AS 'Total Movies and TV Shows per Year'

FROM netflix

GROUP BY type, release_year

ORDER BY COUNT(*) DESC

limit 10;



SELECT CONCAT('There was a total of ', COUNT(*), ' ', type, ' shows released in ', release_year) AS 'Total Movies and TV Shows per Year'

FROM netflix

WHERE type = 'TV Show'

GROUP BY type, release_year

ORDER BY COUNT(*) DESC;


SELECT CONCAT('There was a total of ', COUNT(*), ' ', type, ' shows released in ', country, ' in ', release_year) AS 'Total Movies and TV Shows in different countries per Year'

FROM netflix

WHERE type IN ('Movie', 'TV Show')

GROUP BY type, country, release_year

ORDER BY COUNT(*) DESC;
```

```
SELECT CONCAT('There was a total of ', COUNT(*), ' ', type, ' shows released in ',country,' in ', release_year) AS 'Total Movies and TV Shows per Year'

FROM netflix

GROUP BY type, release_year

ORDER BY COUNT(*) DESC;


## china canada and chicago had the highest number of released movies and tv shows in 2017


SELECT listed_in, COUNT(*) AS `top 10 genre`

FROM netflix

GROUP BY listed_in

ORDER BY `top 10 genre` DESC

;


## here we see the top 10 genre


SELECT

  type,

  COUNT(*) AS count,

  CONCAT(ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM netflix), 2), '%') AS percentage

FROM netflix

GROUP BY type;
```

here we see the distribution of movies and tv shows by percentage

```
SELECT rating, COUNT(*) AS `movies and TV show ratings`
FROM netflix
GROUP BY rating
ORDER BY `movies and TV show ratings` DESC;
```

here we see the ratings of movies and tv shows from the most rated to the least rated