
EXPLORATORY DATA ANALYSIS OF NETFLIX ORIGINALS

The main purpose of this project is to analyze the given dataset of the Netflix Original Movies from 2014 to 2021, with the help of several Queries, and to summarize the main characteristics of the dataset in the form of visual Bar Graphs, pie charts etc.

The Data Set contains a total of 584 records and 7 columns like:

serial_number	Title	Genre	Premiere	Runtime	IMDB	Language
1	Enter the Anime	Documentary	August 5, 2019	58	2.5	English/Japanese
2	Dark Forces	Thriller	August 21, 2020	81	2.6	Spanish
3	The App	Science fiction/Drama	December 26, 2019	79	2.6	Italian
4	The Open House	Horror thriller	January 19, 2018	94	3.2	English
5	Kaali Khuhi	Mystery	October 30, 2020	90	3.4	Hindi
6	Drive	Action	November 1, 2019	147	3.5	Hindi
7	Leyla Everlasting	Comedy	December 4, 2020	112	3.7	Turkish
8	The Last Days of American Crime	Heist film/Thriller	June 5, 2020	149	3.7	English
9	Paradox	Musical/Western/Fantasy	March 23, 2018	73	3.9	English
10	Sardar Ka Grandson	Comedy	May 18, 2021	139	4.1	Hindi
11	Searching for Sheela	Documentary	April 22, 2021	58	4.1	English
12	The Call	Drama	November 27, 2020	112	4.1	Korean
13	Whipped	Romantic comedy	September 18, 2020	97	4.1	Indonesian
14	All Because of You	Action comedy	October 1, 2020	101	4.2	Malay
15	Mercy	Thriller	November 22, 2016	90	4.2	English
16	After the Raid	Documentary	December 19, 2019	25	4.3	Spanish
17	Ghost Stories	Horror anthology	January 1, 2020	144	4.3	Hindi
18	The Last Thing He Wanted	Political thriller	February 21, 2020	115	4.3	English
19	What Happened to Mr. Cha?	Comedy	January 1, 2021	102	4.3	Korean
20	Death Note	Horror thriller	August 25, 2017	100	4.4	English

Following Queries were Analyzed for this Data Set:

- 1) Average IMDB ratings of Genres with Occurance>10
- 2) Top 10 most Occuring Languages in the Netflix Originals
- 3) Top 10 movies with Highest IMDB Ratings
- 4) 5 Languages with lowest IMDB Ratings
- 5) Total Runtime of Each Year
- 6) Average Runtime of Top 3 Languages
- 7) Number of Movies Premiered Each Year
- 8) IMDB Ratings of English Documentaries Premiered Between June 2019-June 2020

1) Average IMDB ratings of Genres with Occurance>10

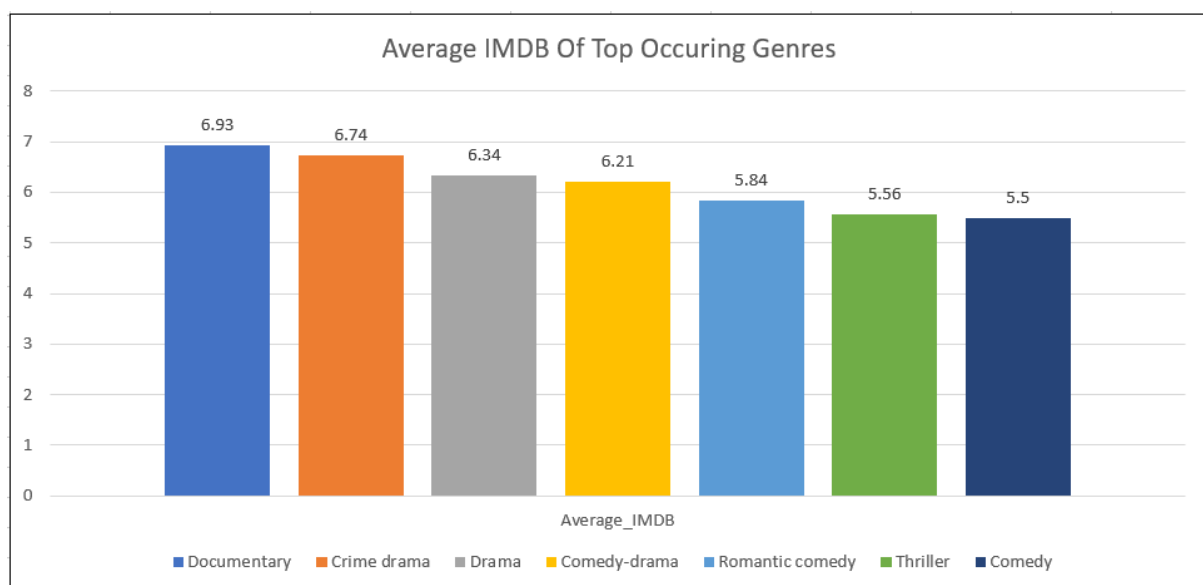
Code:

```
1      /* Average IMDB ratings of Genres with Occurance>10 */
2
3  •    SELECT genre, COUNT(genre) AS Occurance, AVG(imdb) AS Average_IMDB
4        FROM netflix.netflixoriginals_dataset
5        GROUP BY genre HAVING occurrence >10
6        ORDER BY Average_IMDB DESC
```

Top Occurring Genres with there respective occurrence were as follows:

- 1) Documentary (159)
- 2) Drama (77)
- 3) Comedy (49)
- 4) Romantic Comedy (39)
- 5) Thriller (33)
- 6) Comedy-Drama (14)
- 7) Crime Drama (11)

Result:

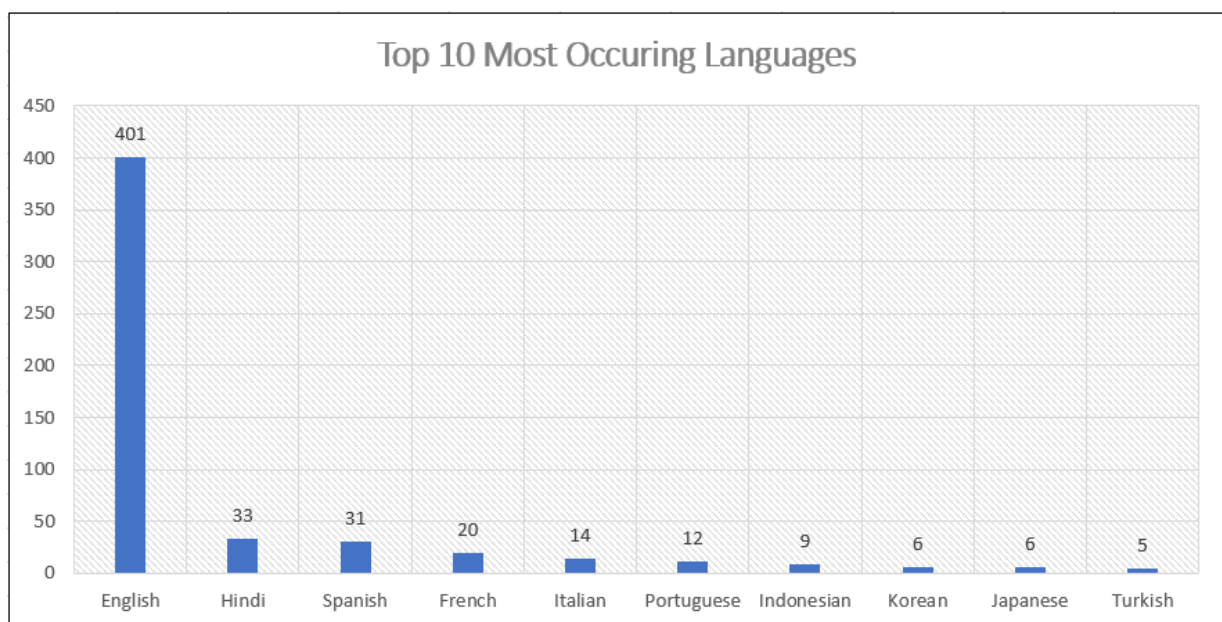


2) Top 10 most Occuring Languages in the Netflix Originals:

Code:

```
1  /* Top 10 most Occuring Languages in the Netflix Originals */
2
3  • SELECT language, COUNT(language) as OccuranceValue
4     FROM netflix.netflixoriginals_dataset
5     GROUP BY language
6     ORDER BY OccuranceValue DESC LIMIT 10
```

Result:



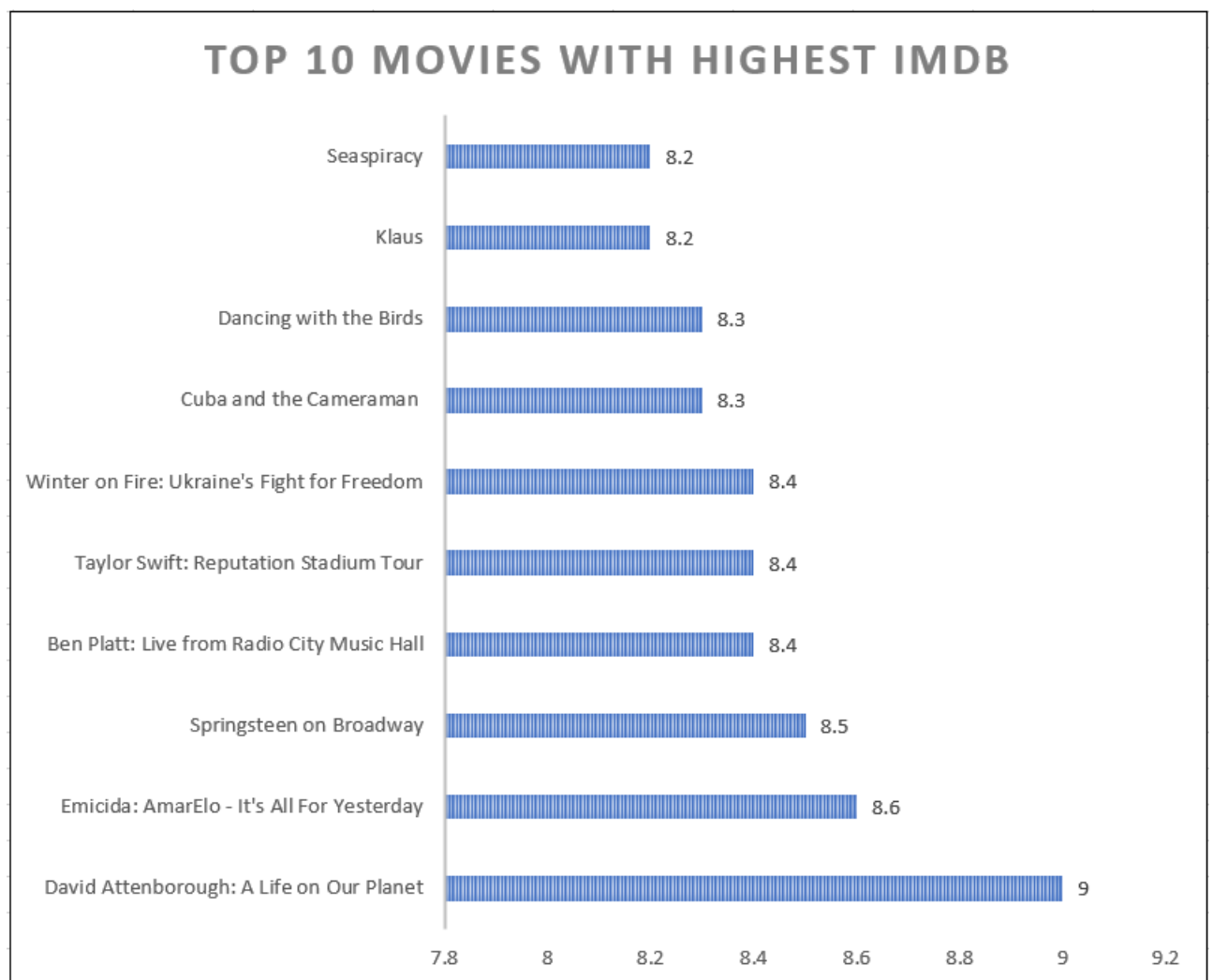
Therefore, English was the most Occuring Language followed by Hindi and then Spanish.

3) Top 10 movies with Highest IMDB Ratings:

Code:

```
1  /* Top 10 movies with Highest IMDB */
2
3  • SELECT title, IMDB
4     FROM netflix.netflixoriginals_dataset
5     ORDER BY IMDB DESC LIMIT 10
```

Result:

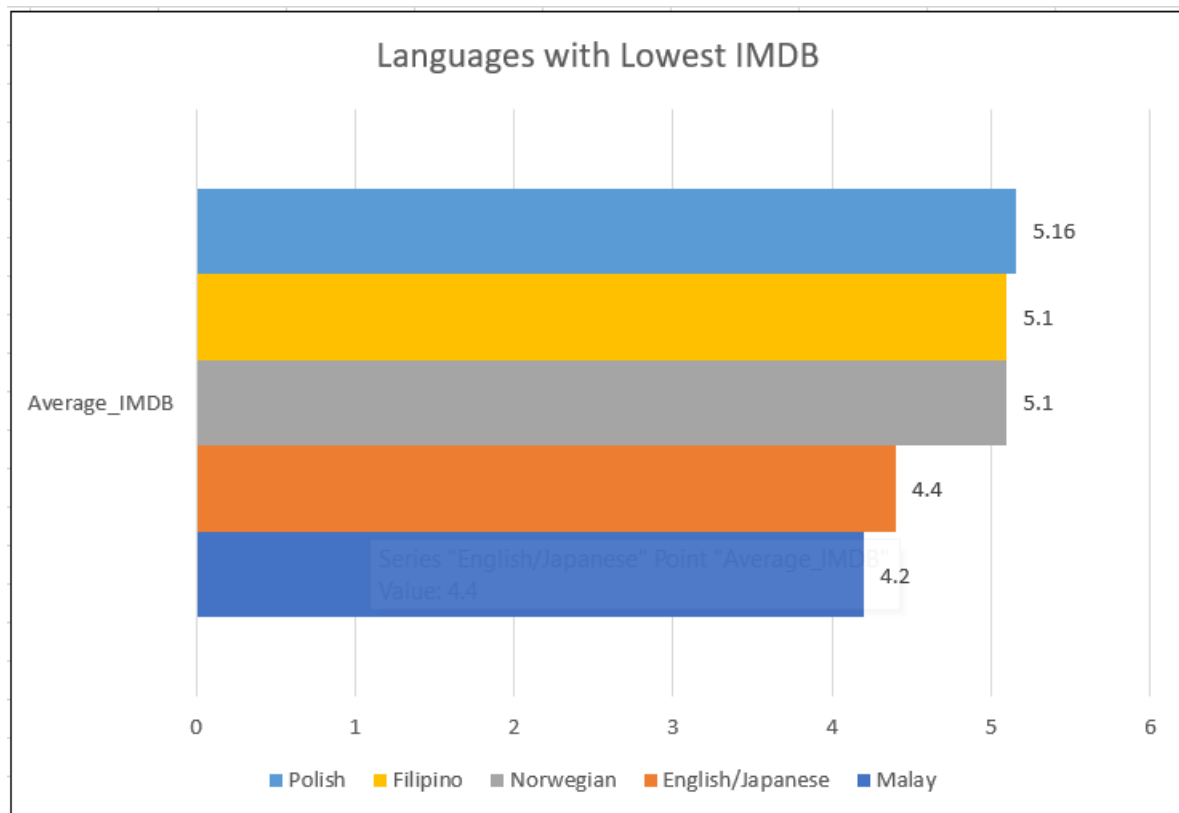


4) 5 Languages with lowest IMDB Ratings:

Code:

```
1  /* 5 Languages with lowest IMDB Rating */
2
3  •  SELECT language, AVG (IMDB) AS Average_IMDB
4     FROM netflix.netflixoriginals_dataset
5     GROUP BY language
6     ORDER BY Average_IMDB LIMIT 5
```

RESULT:



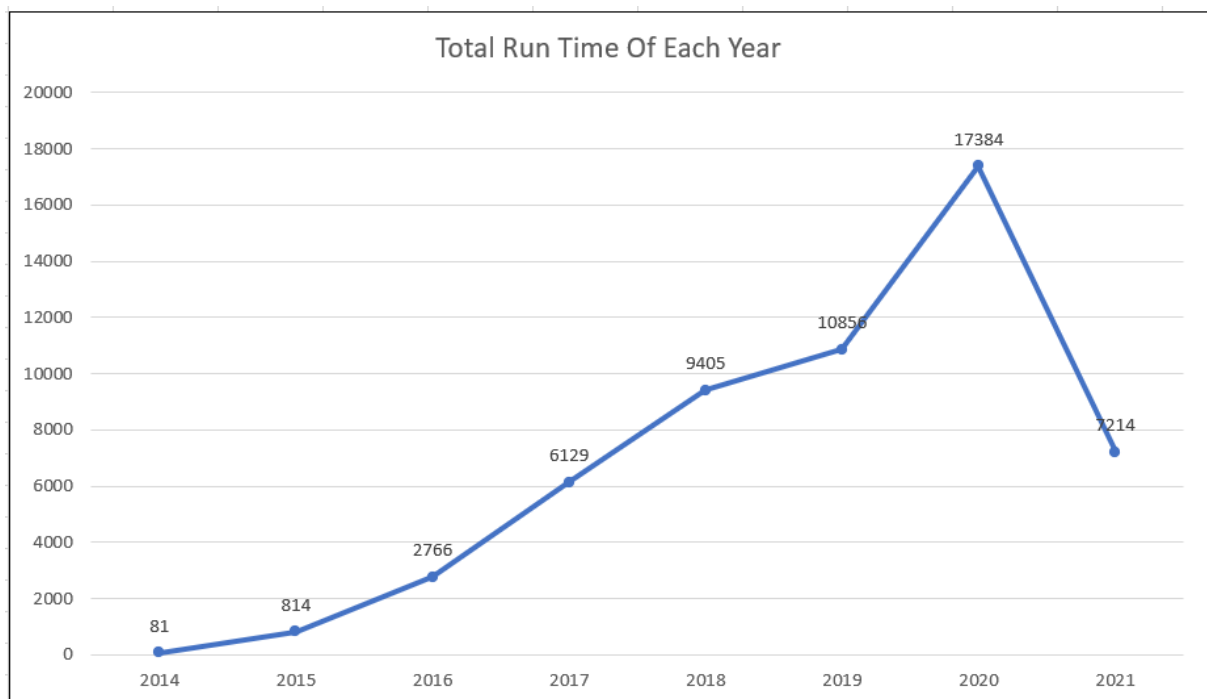
Malay has Lowest IMDB of all the languages followed by English/Japanese and Norwegian

5) Total Run-Time of Each Year:

Code:

```
1  /* Total Run-Time of Each Year */
2
3  •  SELECT RIGHT(premiere, 4) AS Year,
4     SUM(Runtime) AS Total_Runtime
5     FROM netflix.netflixoriginals_dataset
6     GROUP BY year ORDER BY year
```

Result:



Year 2020 has the Highest Total Run time while 2014 has the lowest Total Run Time.

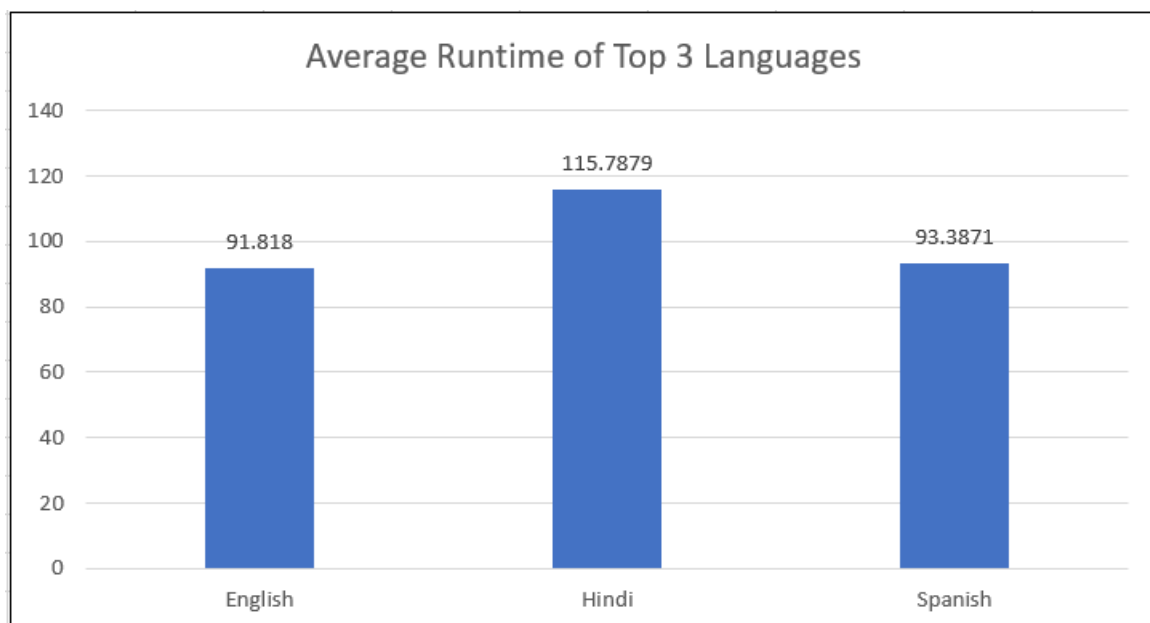
6) Average Runtime of Top 3 Languages:

Code:

Top 3 Languages were = English, Hindi and Spanish, hence

```
1  /* Average Runtime of Top 3 Languages */
2
3  •  SELECT language, AVG(Runtime) AS Average_Runtime
4     FROM netflix.netflixoriginals_dataset
5     WHERE language = 'English'
6     UNION
7     SELECT language, AVG(Runtime) AS Average_Runtime
8     FROM netflix.netflixoriginals_dataset
9     WHERE language = 'hindi'
10    UNION
11    SELECT language, AVG(Runtime) AS Average_Runtime
12    FROM netflix.netflixoriginals_dataset
13    WHERE language = 'Spanish'
```

Result:

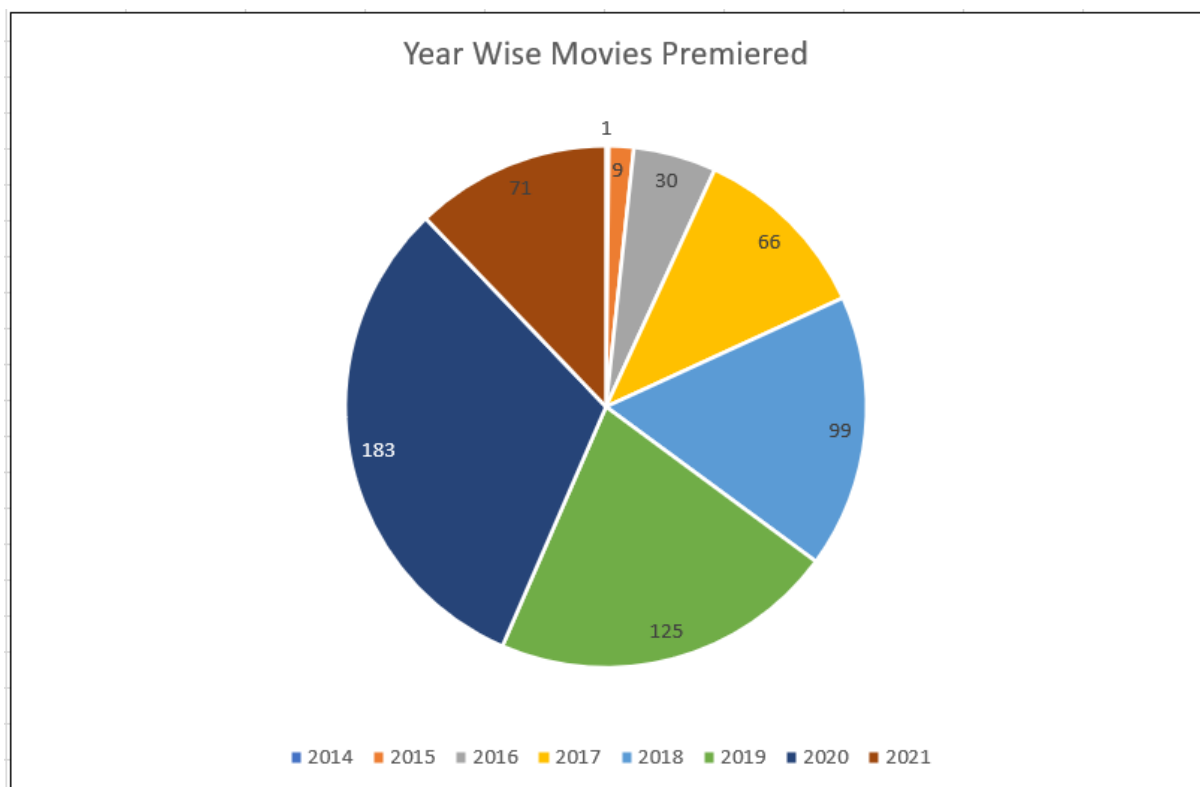


7) Number of Movies Premiered Each Year:

Code:

```
1  /* Movies Premiered Each Year */
2
3  • SELECT RIGHT (premiere, 4) AS year,
4     COUNT(Premiere) AS Movies_Premiered
5     FROM netflix.netflixoriginals_dataset
6     GROUP BY year ORDER BY year
```

Result:



Most Movies were Premiered in 2020 followed by 2019 and then 2018, whereas least movie was premiered in year 2014.

8) IMDB Ratings of English Documentaries Premiered Between June 2019-June 2020:

Code:

```
1  /* IMDB Ratings of English Documentaries Premiered Between June 2019 and June 2020 */
2
3  • SELECT title, IMDB, premiere, genre, language
4  FROM netflix.netflixoriginals_dataset
5  WHERE genre = 'Documentary' AND language = 'english' AND
6  (premiere LIKE '%2019%' OR premiere LIKE '%2020%') AND
7  premiere NOT LIKE '%january%2019%' AND premiere NOT LIKE '%february%2019%' AND premiere NOT LIKE '%march%2019%' AND premiere NOT LIKE '%april%2019%' AND
8  premiere NOT LIKE '%may%2019%' AND premiere NOT LIKE '%july%2020%' AND premiere NOT LIKE '%august%2020%' AND premiere NOT LIKE '%september%2020%' AND
9  premiere NOT LIKE '%october%2020%' AND premiere NOT LIKE '%november%2020%' AND premiere NOT LIKE '%december%2020%'
```

Result:

