PROJECT ON APPLE APP STORE

AGENDA

- INTRODUCTION
- SYSTEM DESIGN
- IMPLEMENTATION
- CONCLUSION

INTRODUCTION

- Apple. is an American multinational technology company headquartered in Cupertino, California.
- electronics company famous for creating the iPhone, iPad and Macintosh computers.
- The App Store is an app marketplace developed and maintained by Apple Inc., for mobile apps on its iOS and iPadOS operating systems.

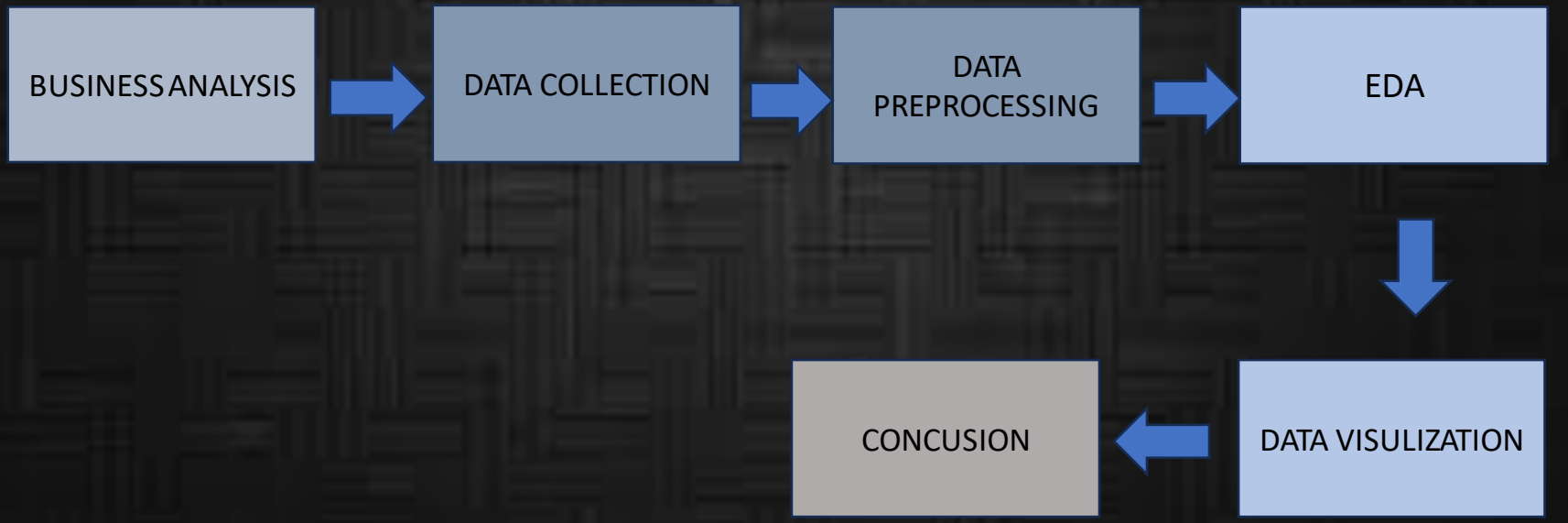
DATA SET:

The data is categorized into 21 columns and 19lakh rows:

- Apple_Id:Each application listed have unique ID
- App_Name: name of the app.
- Primary_genre: category.
- Content_Rating: age limit for each app.
- Size_Bytes: storage of the app.
- Required_IOS_version: version for the app
- Released: when was app released.
- Update: when was app updated.
- Version: version of app
- Price: price for app

- ❑Currency: which type of currency(USD)
- ❑Free: the app is free or not.
- ❑Developer_ID: app developer ID
- ❑Developer: name of the developer
- ❑Review: number of review on app

SYSTEM DESIGN



IMPLEMENTATION

IMPORT THE REQUIRED LIBRARY:

In [217]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
```

Export the data from external source :

In [267]:
```
df=pd.read_csv("apple.csv")
df
```

Out[267]:

	App_Id	App_Name	AppStore_Url	Primary_Genre	Content_Rating	Size_Bytes	Required
0	com.hkbu.arc.apaper	A+ Paper Guide	https://apps.apple.com/us/app/a-paper-guide/id...	Education	4+	21993472.0	
1	com.dmitriev.abooks	A-Books	https://apps.apple.com/us/app/a-books/id103157...	Book	4+	13135872.0	
2	no.terp.abooks	A-books	https://apps.apple.com/us/app/a-books/id145702...	Book	4+	21943296.0	
3	fr.antoinettefleur.Book1	A-F Book #1	https://apps.apple.com/us/app/a-f-book-1/id500...	Book	4+	81851392.0	
4	com.imonstersoft.azdictionaryios	A-Z Synonyms Dictionary	https://apps.apple.com/us/app/a-z-synonyms-dic...	Reference	4+	64692224.0	

DATA PREPROCESSING

Head() function is used to fetch the first 5 rows of data set:

```
1 df.head()
```

	App_Id	App_Name	AppStore_Url	Primary_Genre	Content_Rating	Size_Bytes	Required_IOS_Version	Released	
0	com.hkbu.arc.apaper	A+ Paper Guide	https://apps.apple.com/us/app/a-paper-guide/id...	Education	4+	21993472.0	8.0	2017-09-28T03:02:41Z	21
1	com.dmitriev.abooks	A-Books	https://apps.apple.com/us/app/a-books/id103157...	Book	4+	13135872.0	10.0	2015-08-31T19:31:32Z	23
2	no.terp.abooks	A-books	https://apps.apple.com/us/app/a-books/id145702...	Book	4+	21943296.0	9.0	2021-04-14T07:00:00Z	30
3	fr.antoinettefleur.Book1	A-F Book #1	https://apps.apple.com/us/app/a-f-book-1/id500...	Book	4+	81851392.0	8.0	2012-02-10T03:40:07Z	29
4	com.imonstersoft.azdictionaryios	A-Z Synonyms Dictionary	https://apps.apple.com/us/app/a-z-synonyms-dic...	Reference	4+	64692224.0	9.0	2020-12-16T08:00:00Z	18

5 rows × 21 columns

Info() function to get the information about data set:

```
In [220]: 1 df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1230376 entries, 0 to 1230375
Data columns (total 21 columns):
 #   Column                  Non-Null Count    Dtype
---  ------                  --------------    -----
 0   App_Id                  1230376 non-null  object
 1   App_Name                1230375 non-null  object
 2   AppStore_Url            1230376 non-null  object
 3   Primary_Genre           1230376 non-null  object
 4   Content_Rating          1230376 non-null  object
 5   Size_Bytes              1230152 non-null  float64
 6   Required_IOS_Version    1230376 non-null  object
 7   Released                1230373 non-null  object
 8   Updated                 1230376 non-null  object
 9   Version                 1230376 non-null  object
 10  Price                   1229886 non-null  float64
 11  Currency                1230376 non-null  object
 12  Free                    1230376 non-null  bool
 13  DeveloperId             1230376 non-null  int64
 14  Developer               1230376 non-null  object
 15  Developer_Url           1229267 non-null  object
 16  Developer_Website       586388 non-null   object
 17  Average_User_Rating     1230376 non-null  float64
 18  Reviews                 1230376 non-null  int64
 19  Current_Version_Score   1230376 non-null  float64
 20  Current_Version_Reviews 1230376 non-null  int64
dtypes: bool(1), float64(4), int64(3), object(13)
memory usage: 188.9+ MB
```

Isnull()function is used to check null values:

```
In [221]: 1 df.isnull().sum()

Out[221]: App_Id                    0
          App_Name                  1
          AppStore_Url              0
          Primary_Genre             0
          Content_Rating            0
          Size_Bytes              224
          Required_IOS_Version      0
          Released                  3
          Updated                   0
          Version                   0
          Price                   490
          Currency                  0
          Free                      0
          DeveloperId               0
          Developer                 0
          Developer_Url          1109
          Developer_Website    643988
          Average_User_Rating       0
          Reviews                   0
          Current_Version_Score     0
          Current_Version_Reviews   0
          dtype: int64
```

DATA CLEANING

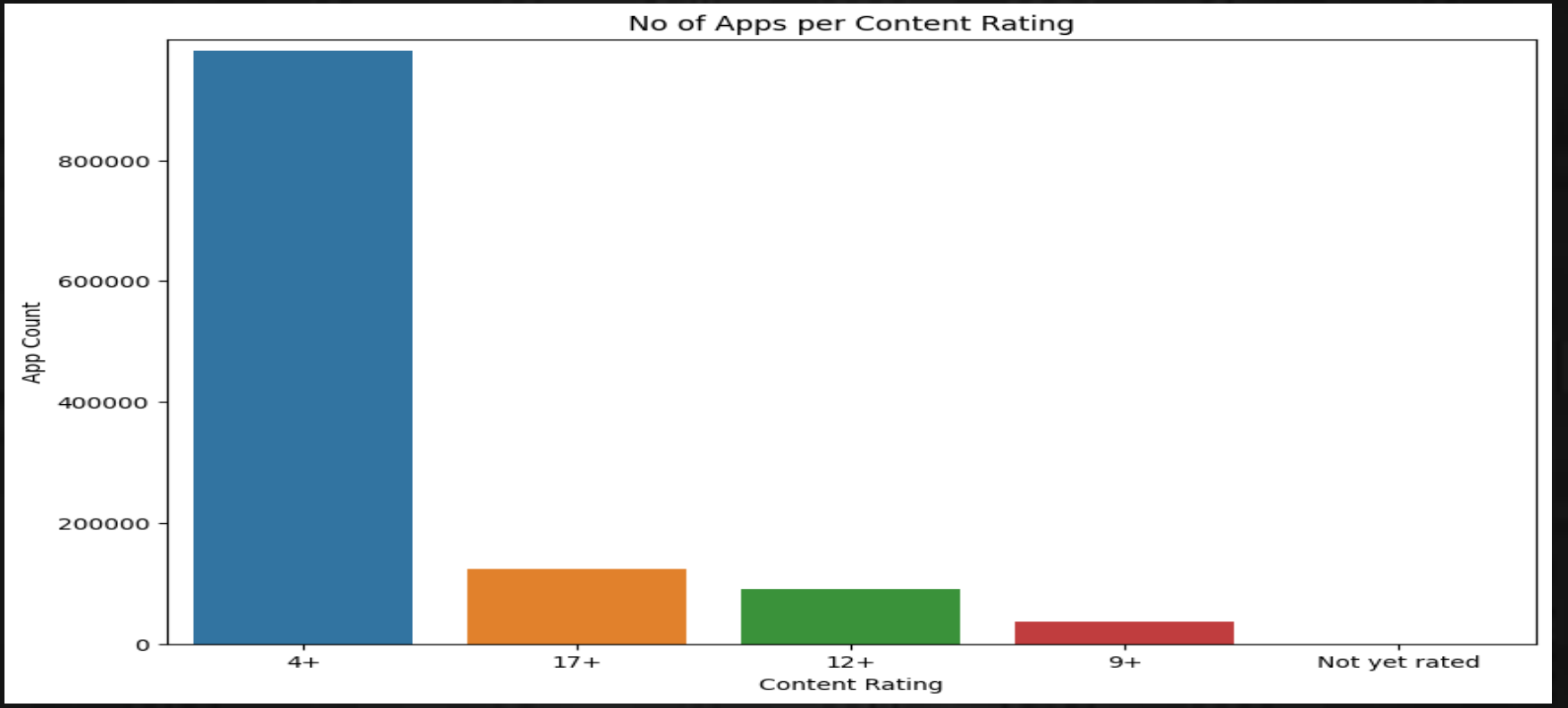
- In my data set I have observed several null values are present ,hence data cleaning are done:
- In data set App_name consist of null values ,hence it is replaced by unknown values.

```
In [223]: 1 df['App_Name']=df['App_Name'].fillna("unknown")
```

- In data set price consist of null values ,hence it is replaced by mean values.

```
]: 1 df['Price']=df['Price'].fillna("mean")
```

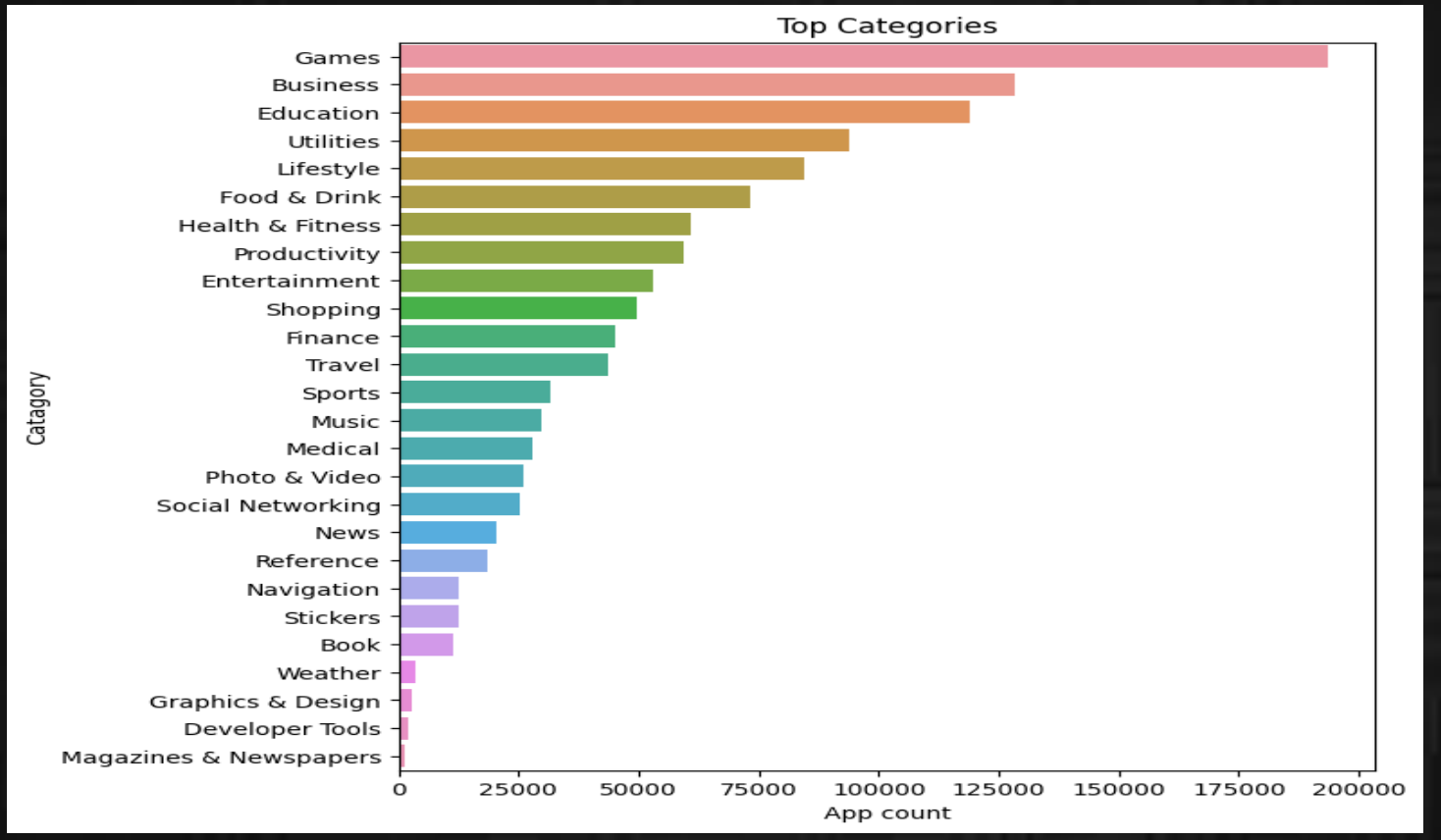
1. Content Rating



- Apps rated 4+ contain no objectionable material.
- Apps rated 9+ may contain instances of content that may not be suitable for children under the age of 9.
- Apps rated 12+ may contain instances of content that may not be suitable for children under the age of 12.
- Apps rated 17+ may contain instances of content that may not be suitable for minors under the age of 17.

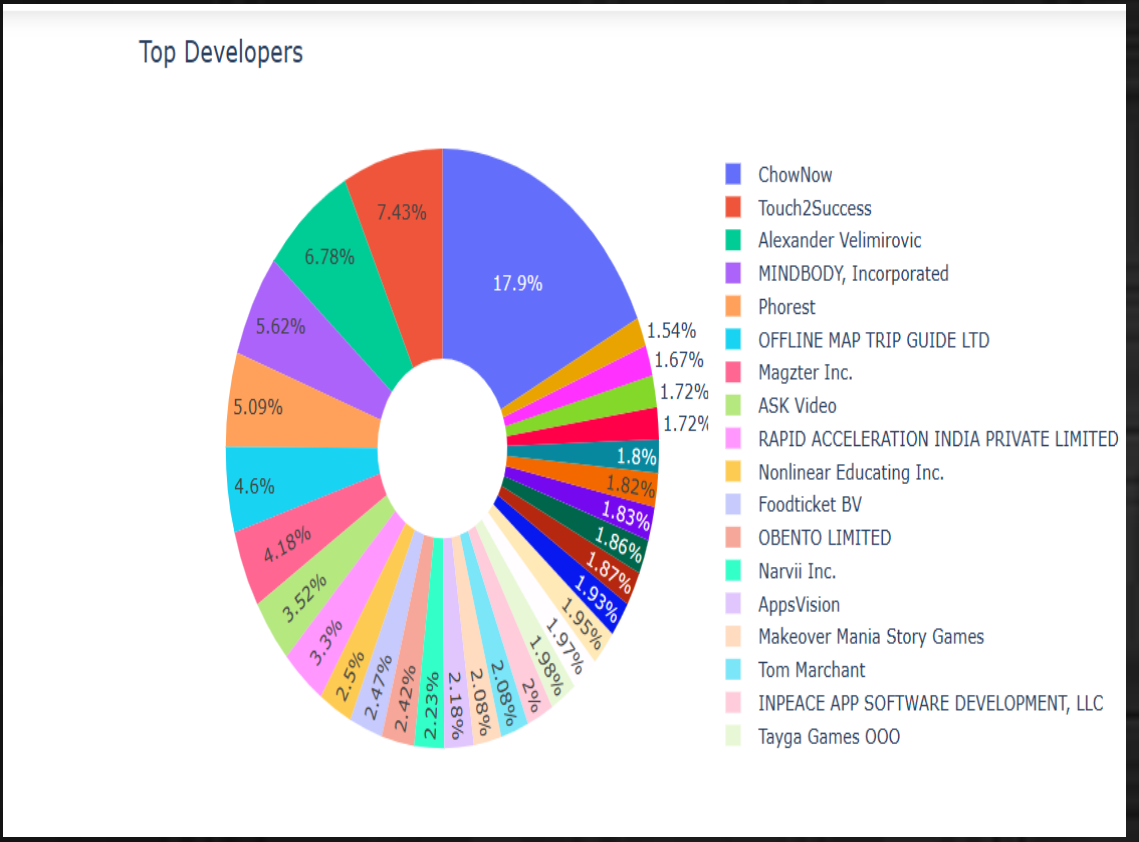
- Large amount of the apps are present in 4+ content rating

2.Top category(genre)



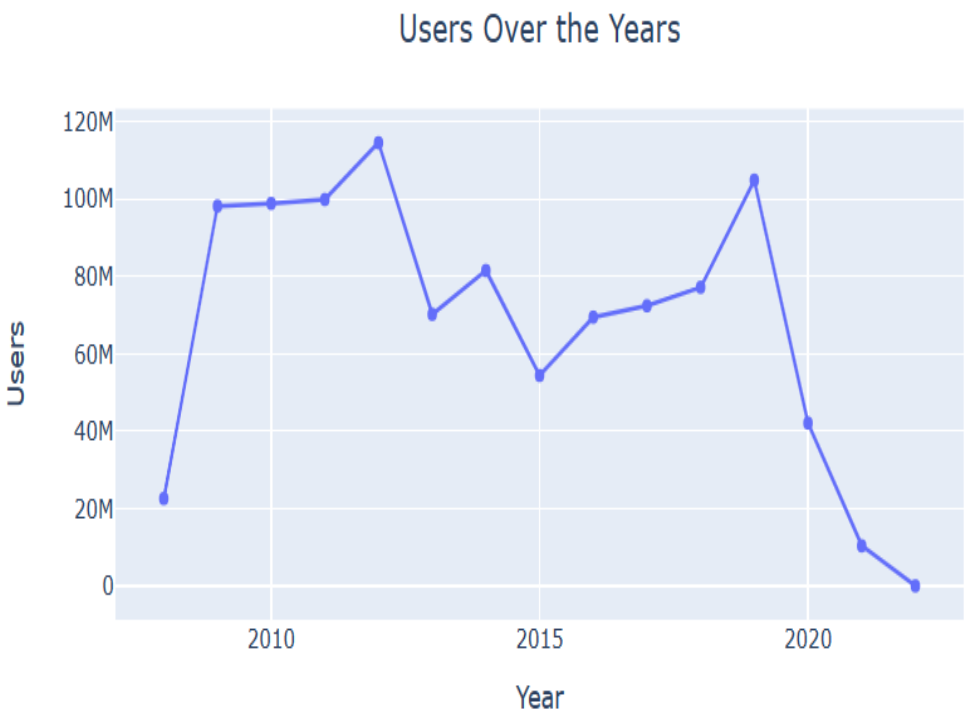
Games topped from list of categories.
Business & Education application follows and least is Magazines and Newspaper.

3.Top developers



ChowNow has published almost 4 thousand apps in appstore. Touch2Success, Alexander Velimirovic, Mindbody, Phorest, Offline map trip guide ltd these developers have published more than a thousand applications.

4.What is the number of users over the years?

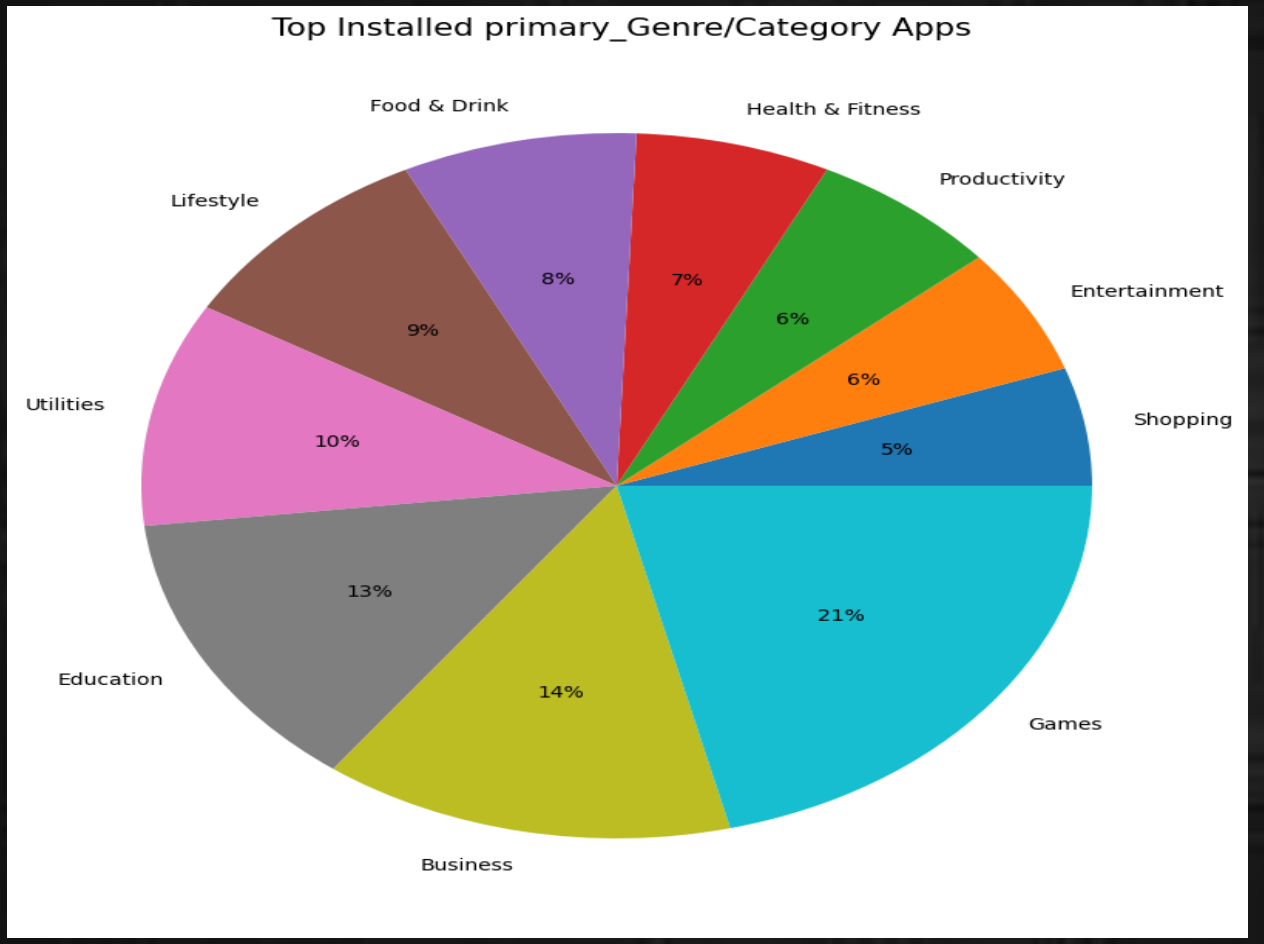


2009 to 2012 there is steady growth in users and dropped heavily in 2013,and in year 2020 their is huge drop.

5.What is percent of Free and Paid Games

Seems like there are only 8.2% of paid apps present in 19 lakh applications.

6.top 10 Primary_Genre/Categories that are installed from the Apple Store?



Here we can see that top 10 installed apps in apple store,here we can say that Games category are installed more and followed by business and education and least is shopping and entertainment

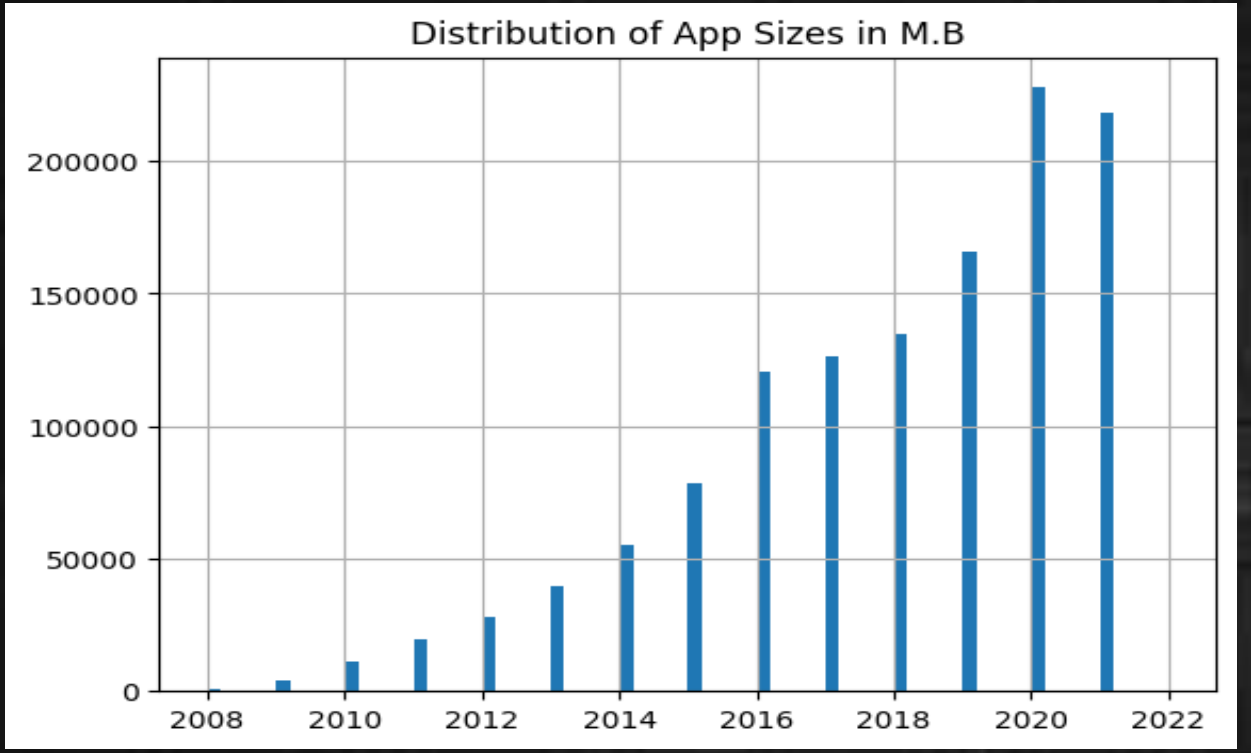
7.What is the Distribution of App Size in M.B?

Here to get the distribution of app size over year we required year column ,so we have take year separetly from the data set and replaced in data set

```
In [250]: 1 df['Released year']=df['Released'].dt.year
          2 inplace=True
```

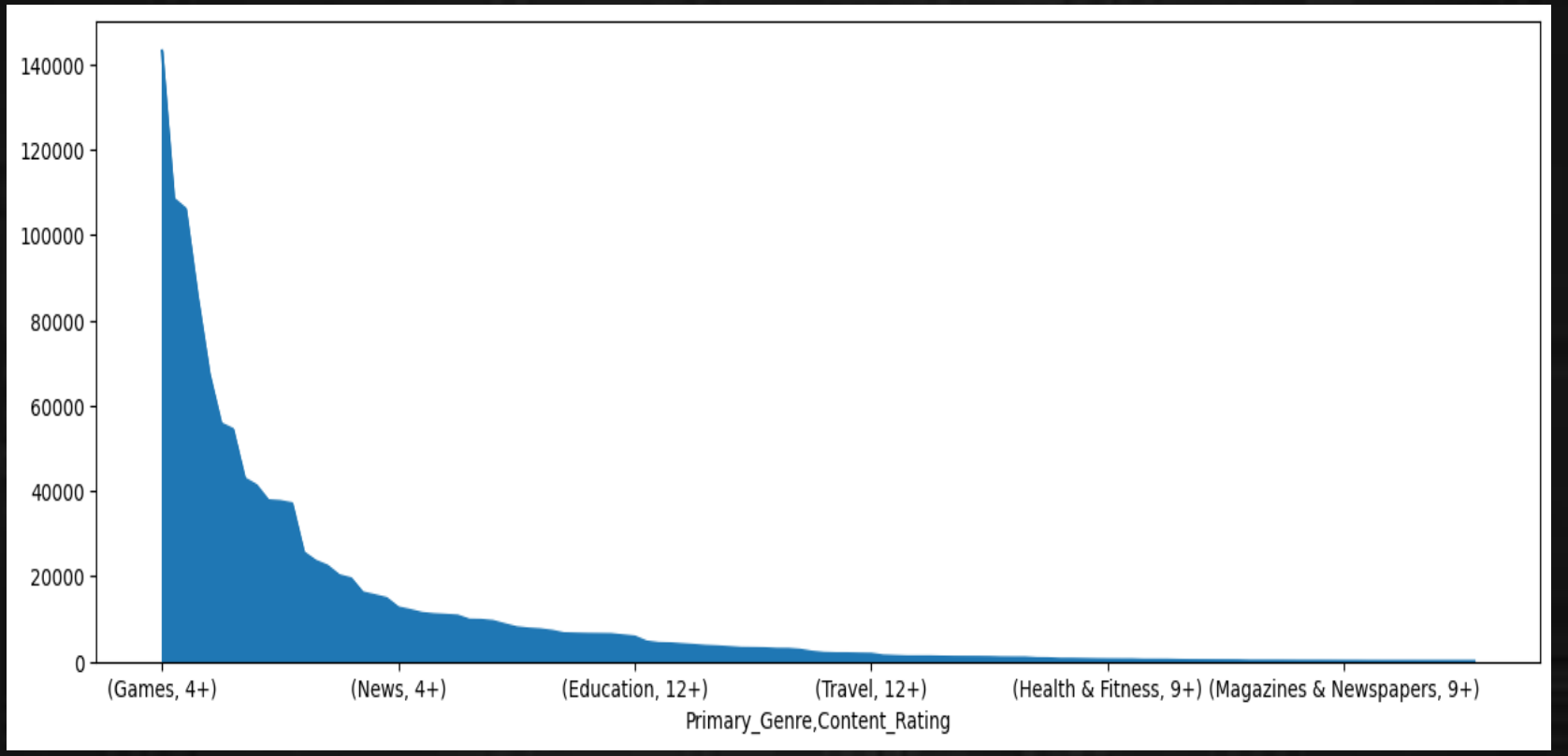
```
In [251]: 1 df['updated year']=df['Updated'].dt.year
          2 inplace=True
```

eviews	Size_MB	Revenue	Released year	updated year
0	20.974609	0.0	2017.0	2018
1	12.527344	0.0	2015.0	2019
0	20.926758	0.0	2021.0	2021
0	78.05957	0.0	2012.0	2019
0	61.695312	0.0	2020.0	2020
...	...	...	...	...
143	15.894531	0.0	2020.0	2020
1500	37.208984	0.0	2019.0	2020
1	16.425781	0.0	2018.0	2018
0	54.088867	0.0	2021.0	2021
0	81.191406	0.0	2018.0	2019

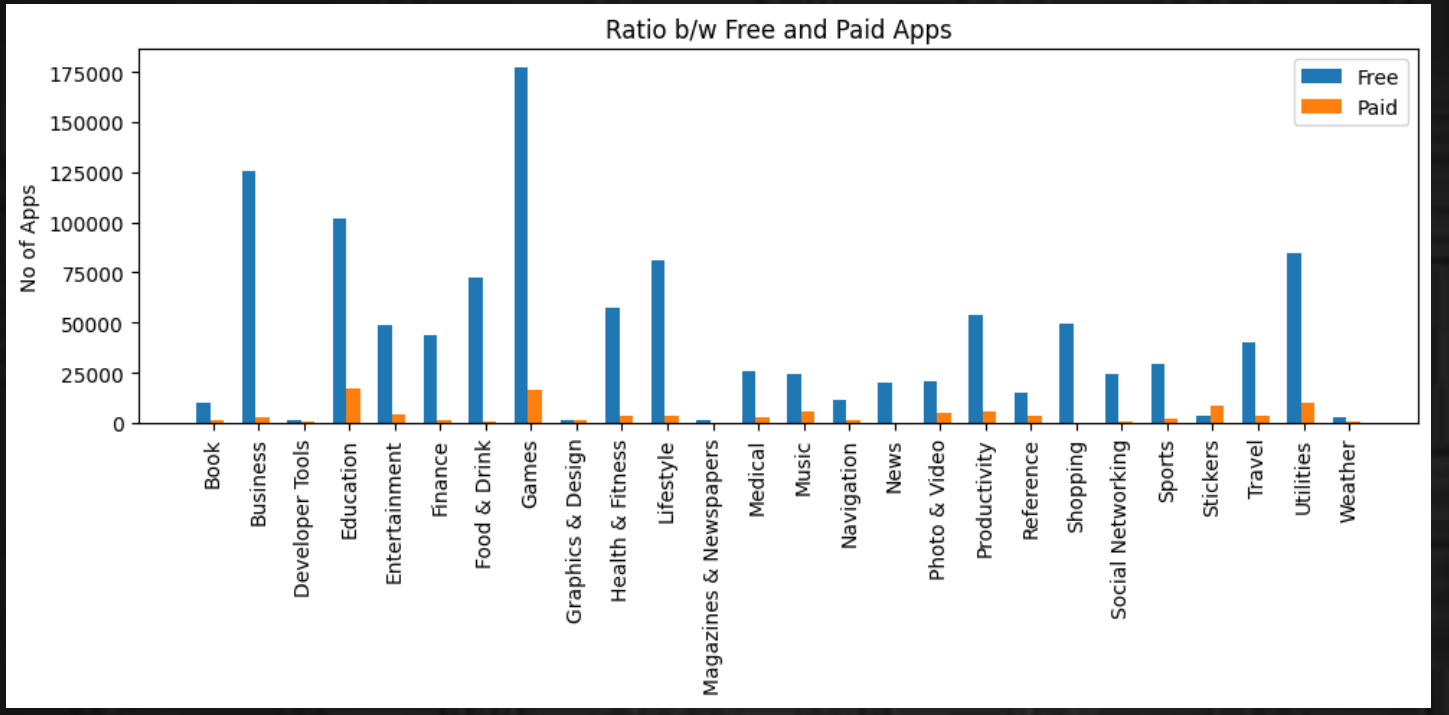


Here we can tell that in the year 2020 the size of app has increased rapidy compared to other year

8.what is the age of people and what they are watching?

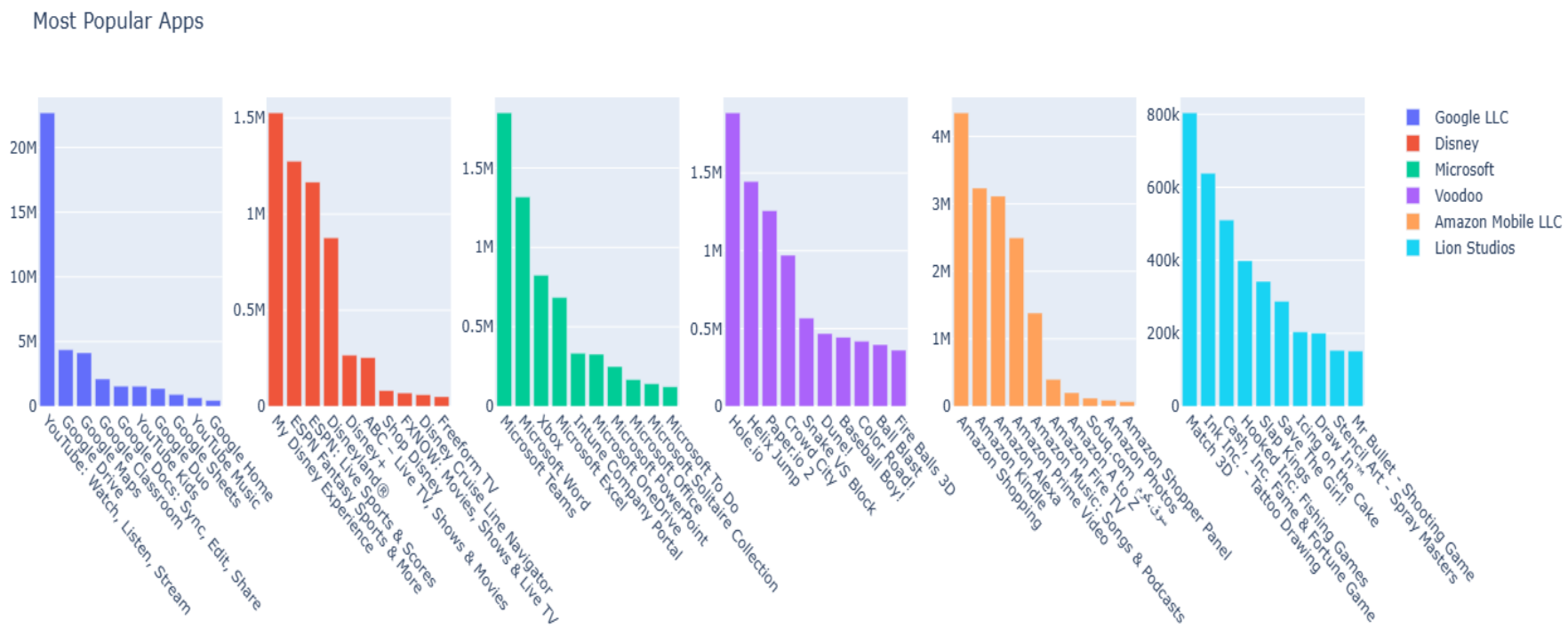


9.Ratio of the paid apps and free apps



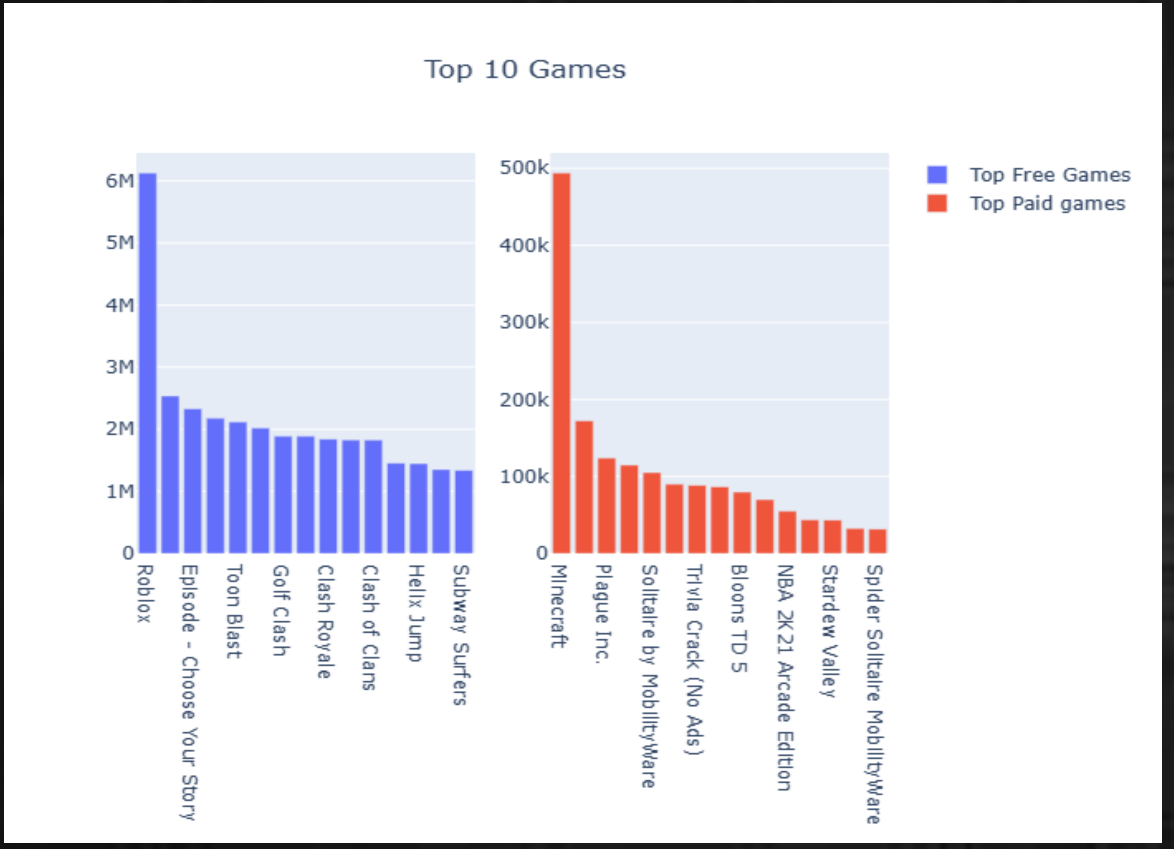
Seems like there way more less paid apps in all categories. Education and Games have almost same amount of paid apps.

10.Which app has popularity dominance from any Top 20?



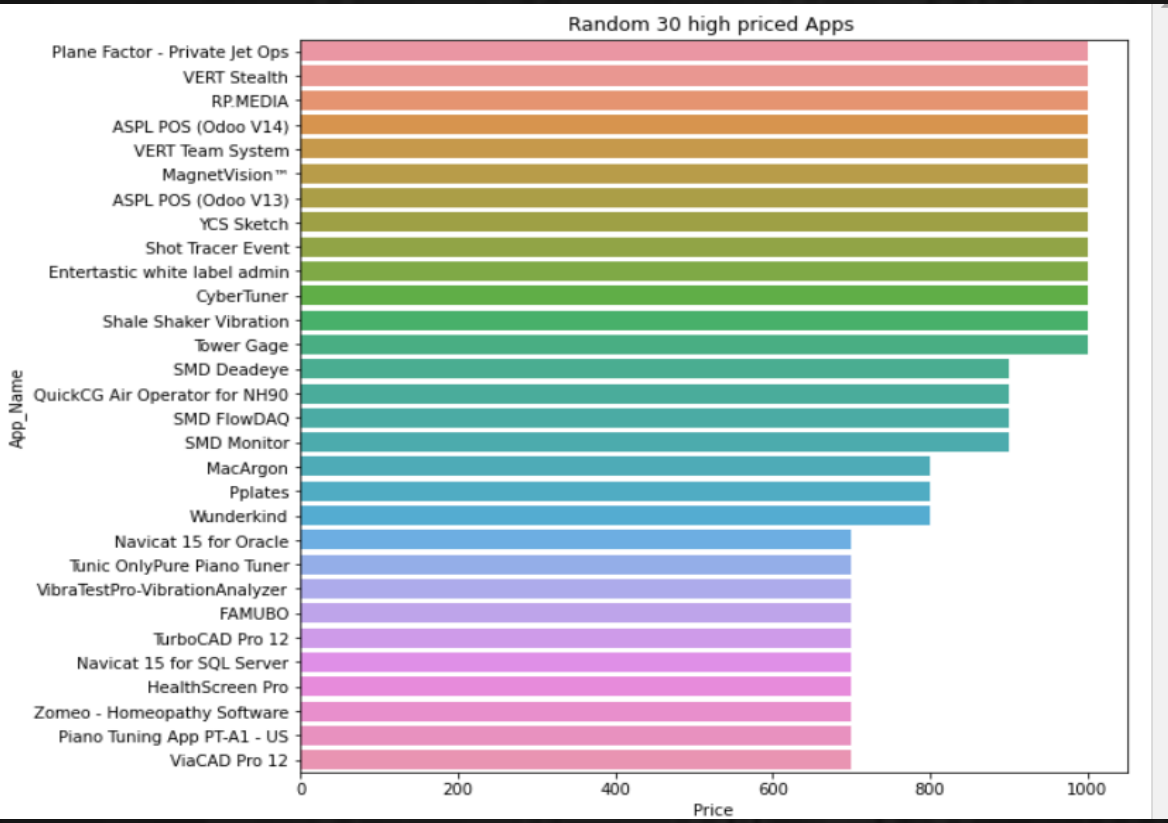
- Youtube is the dominant app from Google LLC with 20+ million users, that's the total users of below top 9 apps.
- My Disney Experience and ESPN have a large amount of users.
- Microsoft Teams is the dominant app from Microsoft.
- From Voodoo a games developer Hele.io has more popularity(users).
- Amazon has AMZN shopping app which is the dominant with almost 4.5 million users
- Lion Studio has Match 3D app with 800+ thousand users.

11.What is percent of Free and Paid Games and What are the top 10 games that most users use?



Roblox has 6 million users in free apps Minecraft is the most selling app and have around 500 thousand users.

12.High Priced Apps



CONCLUSION

- Google LLC, Instagram, Inc., Spotify Ltd., Voodoo, AMZN Mobile LLC, etc. are the top 5 developers with most popularity.
- 2009 to 2012 have is steady growth with 100M users yearly and dropped heavily in 2013 by 30M users. 2013-2020 had ups and downs(mostly downs). In the year 2020 and 2021 there has been a massive drop in users and app releases across all the categories
- Seems like there are only 8.2% of paid apps present in 19.3 lakh applications. Roblox has 6 million users in free apps. Minecraft is the most selling app and have around 500 thousand users.
- Chownow has published 3699 apps in total.
- Interesting analysis can be done by comparing Google playstore and Apple appstore.


THANK YOU!