

Top Google Play Store Statistics for 2022

05.12.2022

SHREYA GARLAPATI

G01379895

PROJECT ASSIGNMENT-4

INFS 580 DL2



TABLE OF CONTENT

Abstract	3
Introduction	4
Problem identification	4
Research Questions	4
Literature Review	4
DATASET	5
DATA ANALYSIS	6
DATA IN SQL	6
Distribution of Rating	7
Normal	7
Log-transformed	8
Distribution of Ratings	9
Distribution of type of application	9
Distribution of the number of installs	10
Distribution of reviews as a function of the number of installs	10
Distribution of types of applications	11
TOP TYPE	11
BOTTOM TYPE	12
Distribution of applications last update	13
By Month	13
By Year	14
Review Distribution	14
RESULTS	15
TOP APPLICATION REVIEWS	15
MOST INSTALLS	15
TOP APPS BY PRICE DISTRIBUTION	16
CONCLUSION	16
	2

Abstract

This research paper focuses on examining the Top Google Play Store Statistics for 2022 and how they affect users and businesses throughout the world. In this world which has a major drastic change in technology everything we use is a few taps away. Many engineers have developed the new apps and have changed the world in different presentable ways with easier technologies and one such company named *google* has become the top of the tech sector in developing the apps (Sharma, 2022).

Introduction

Google Play Store is called a digital distribution service which allows the android users to download the apps of many different types. And also it offers the new app developers to develop and showcase their talent and attract a larger audience. This research paper shows the statistics of total number of apps , total number of apps released, apps downloaded, Rating of the apps (Sharma, 2022). In 2022, more than 6 billion individuals will own mobile devices worldwide, predicts Statista. This indicates that around 70% of people on the planet own mobile devices. In the month of february it is known that the Instagram app has the most downloads in the playstore. And snapchat was the second app which had the most downloads. This research provides a play store dataset analysis. This research contributes to the analysis on the Google Play Store Statistics which is important in today's world, whose dependency is the smart phones applications(Olmstead & Atkinson, 2022).

Problem identification

The data set selected for this report is taken from the website “ <https://www.kaggle.com/> ” which is a platform for accurate datasets to be analyzed, so there will not be any incorrect data found. The data which has been obtained is used to understand the pros and cons of the apps. From the data, the number of apps which are for free , number of installations of the app, developer id , developer website can also be identified.

Research Questions

1. Which App has the best reviews overall?
2. Which App category has the most number of installs?
3. Which among the Play Store's top 10 most costly apps?

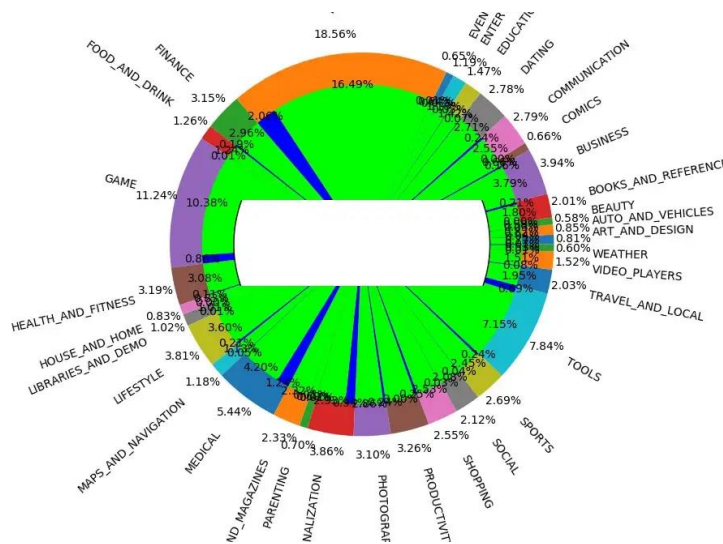
The benefits of researching these questions are; describe your ideas to an audience who is unfamiliar with the google play store statistics. Clarification of the precise goal of the research and the issue tackled by working on the research.

Literature Review

In this research they have gathered detailed information about the google play store apps and its features. In the beginning most of the apps were free to purchase and this research shows the pictorial presentation of free apps vs the paid apps. They also gathered the data about the apps which contain ads or not (Olmstead & Atkinson, 2022).



Here in this research the data is explored using data science and machine learning techniques. In this the researchers have extracted the data from Kaggle and explored it. The main observations were how the app is performing and how well it can be improved using the reviews. First they imported the libraries and then started cleaning the data. They have given the pictorial representation of the code including the distribution of apps in categories (Thakre, 2019). The three research topics which I have chosen relate to my research questions in many ways. Every research topic has the complete details regarding the number of apps installed and the ratings given to it. All the researchers have cited the clear percentage and statistics of the apps. Some questions are like what the cost of the apps and how many apps have been installed in a year were different from mine (*Analysis of apps in the google play store*, 2022).



Source: [Data Science Tutorial: Analysis Of The Google Play Store Dataset | by Abhimanyu Thakre | The Research Nest | Medium](#)

DATASET

The data set which I have chosen is "google play store apps" which is taken from "Kaggle.com". This dataset contains application data of 600k apps and also some following attributes like app name, category, rating etc. in today's world people make use of google app store more as it has all the apps related to our daily lives. So the data which has been collected doesn't have any ethical or legal issues. Even though the information collected has the ratings and reviews of the apps, it is a platform where people can share their ideas or pros and cons about the app so the app developers can use it for further development and there is no harm for the society or the people. So this data set is downloaded in CSV format which consists of data of all the apps in 24 columns and it has all the NOIR analytical data types. This data set contains all the information regarding the apps and their review, rating, number of installs, the category, maximum and minimum installs etc.

DATA ANALYSIS

DATA IN SQL

Results Messages											
	App(nominal)	Category	Rating(ordinal)	Reviews(ratio)	Size	Installs	Type	Price(interval)	Content Rating	Genres	Last Updated
1	Gyft - Mobile Gift Card Wallet	SHOPPING	4.1	9701	14M	*500	000+*	Free	0	Everyone	Shopping
2	*Twice: Buy	Sell Clothing*	SHOPPING	4.2	1558	21M	*100	000+*	Free	0	Everyone
3	*Shopkick: Free Gift Cards	Shop Rewards & Deals*	SHOPPING	4.3	213735	43M	*10	000	000+*	Free	0
4	Slice: Package Tracker	SHOPPING	4.2	16966	34M	*1	000	000+*	Free	0	Everyone
5	The Coupons App	SHOPPING	4.5	181990	Varies with device	*10	000	000+*	Free	0	Everyone
6	*Shopular: Coupons	Weekly Ads & Shopping Deals*	SHOPPING	4.7	57920	Varies with device	*1	000	000+*	Free	0
7	Wish - Shopping Made Fun	SHOPPING	4.5	6211039	15M	*100	000	000+*	Free	0	Everyone
8	*Carousel: Snap-Sell	Chat-Buy*	SHOPPING	4.3	125783	Varies with device	*10	000	000+*	Free	0
9	LightInTheBox Online Shopping	SHOPPING	4	41986	26M	*5	000	000+*	Free	0	Teen
10	Walmart	SHOPPING	4.4	441473	Varies with device	*10	000	000+*	Free	0	Everyone
11	Best Buy	SHOPPING	4.4	186116	34M	*5	000	000+*	Free	0	Everyone
12	JackThreads: Men's Shopping	SHOPPING	4	13085	2.7M	*1	000	000+*	Free	0	Teen
13	*Ibotta: Cash Back Savings	Rewards & Coupons App*	SHOPPING	4.5	315908	52M	*10	000	000+*	Free	0
14	*AliExpress - Smarter Shopping	Better Living*	SHOPPING	4.6	5916569	Varies with device	*100	000	000+*	Free	0
15	LivingSocial - Local Deals	SHOPPING	4.1	28523	29M	*5	000	000+*	Free	0	Everyone
16	zulily - Shop Daily Deals in Fashion and Home	SHOPPING	4.5	28560	Varies with device	*1	000	000+*	Free	0	Everyone
17	MinilnTheBox Online Shopping	SHOPPING	3.9	34171	Varies with device	*1	000	000+*	Free	0	Everyone
18	Zappos &C Shoe shopping made simple	SHOPPING	4.5	44588	20M	*5	000	000+*	Free	0	Everyone

There are 10841 rows

SELECT count(*)
FROM [master].[dbo].[gps]

100 %

Results Messages

(No column name)

1 10841

Most frequent categories

<pre> Select [Category], count(*) from master.dbo.gps group by [Category] order by count(*) desc </pre>		
100 %		
Results Messages		
	Category	(No column name)
1	FAMILY	1935
2	GAME	1138
3	TOOLS	805
4	MEDICAL	452
5	BUSINESS	449
6	PRODUCTIVITY	391
7	PERSONALIZATION	376
8	LIFESTYLE	368
9	COMMUNICATION	368
10	SPORTS	348
11	FINANCE	344
12	HEALTH_AND_FITNESS	316
13	PHOTOGRAPHY	300
14	SOCIAL	260
15	NEWS_AND_MAGAZINES	253
16	TRAVEL_AND_LOCAL	238
17	SHOPPING	225
18	BOOKS_AND_REFERENCE	224
19	DATING	192
20	VIDEO_PLAYERS	163
21	EDUCATION	142
22	ENTERTAINMENT	139
23	MAPS_AND_NAVIGATION	125

Distribution of Rating

Normal

```
... Index(['App(nominal)', 'Category', 'Rating(ordinal)', 'Reviews(ratio)', 'Size',
         'Installs', 'Type', 'Price(interval)', 'Content Rating', 'Genres',
         'Last Updated', 'Current Ver', 'Android Ver'],
        dtype='object')
```

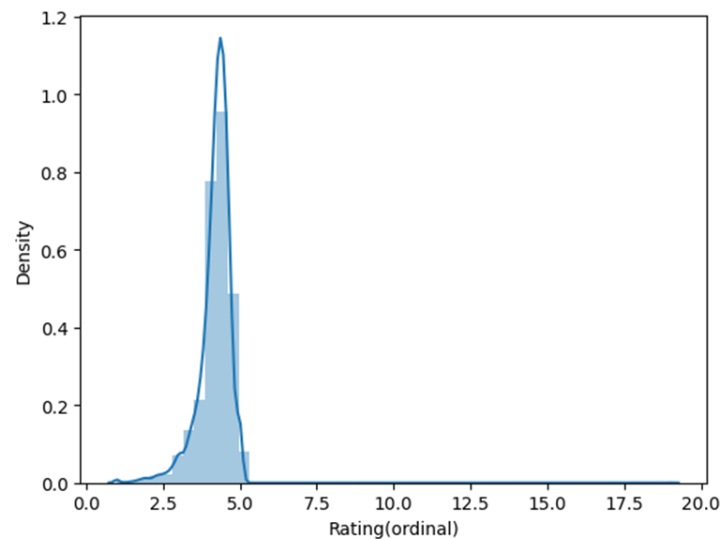
10841

```
sns.distplot(Data['Rating(ordinal)'])
```

[4]

Python

```
... <AxesSubplot: xlabel='Rating(ordinal)', ylabel='Density'>
```



Log-transformed

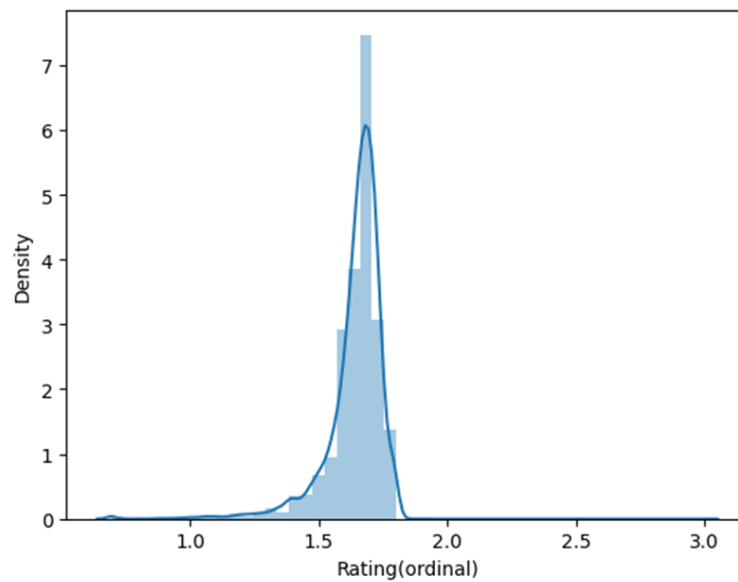
▷ ✓

```
sns.distplot(np.log(Data['Rating(ordinal)'] + 1))
# A log transformation makes the distribution normal.
```

[6]

Python

```
... <AxesSubplot: xlabel='Rating(ordinal)', ylabel='Density'>
```

Distribution of Ratings

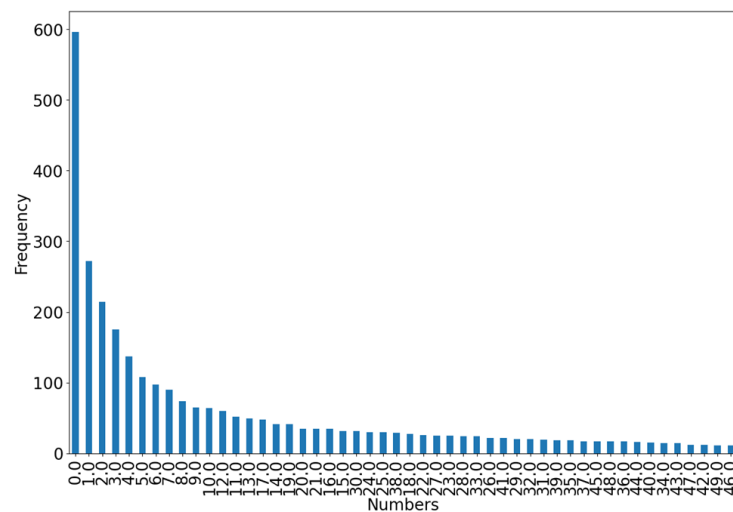
```

Data['Reviews(ratio)2'] = Data['Reviews(ratio)'].str.replace(r'\D', '')
Data['Reviews(ratio)2'] = Data['Reviews(ratio)2'].astype(float)
fig, ax = plt.subplots(figsize=(15, 10))
Data['Reviews(ratio)2'][Data['Reviews(ratio)2'] < 50].value_counts().plot(ax=ax, kind='bar', fontsize=20)
plt.xlabel('Numbers', fontsize=20)
plt.ylabel('Frequency', fontsize=20)
plt.show()

```

[54]

Python



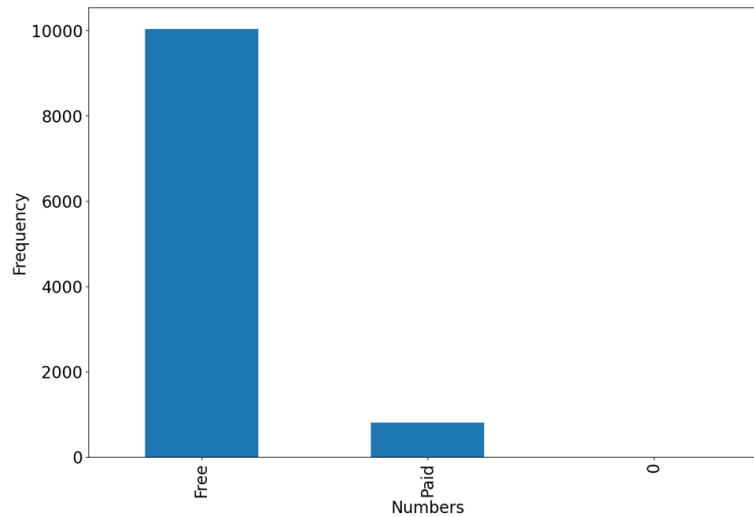
Distribution of type of application

```

fig, ax = plt.subplots(figsize=(15, 10))
Data['Type'].value_counts().plot(ax=ax, kind='bar', fontsize= 20)
plt.xlabel('Numbers', fontsize=20)
plt.ylabel('Frequency', fontsize=20)
plt.show()

```

[55] Python



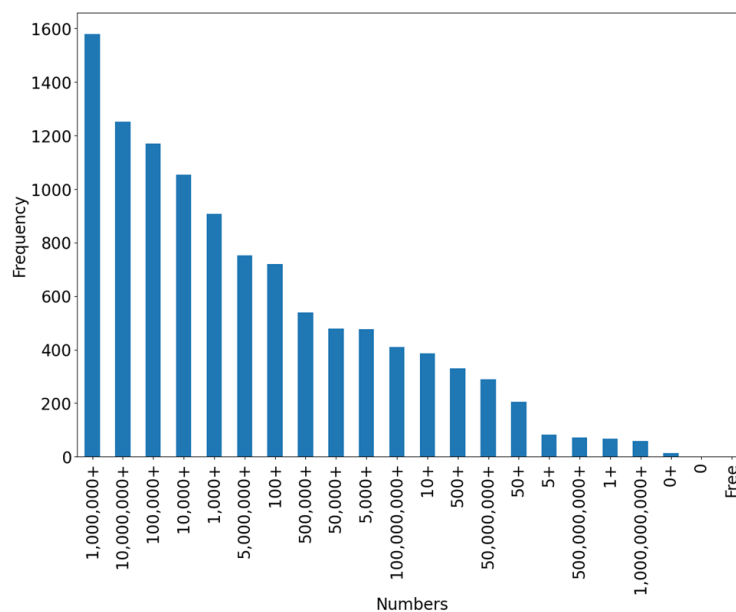
Distribution of the number of installs

```

fig, ax = plt.subplots(figsize=(15, 10))
Data['Installs'].value_counts().plot(ax=ax, kind='bar', fontsize= 20)
plt.xlabel('Numbers', fontsize=20)
plt.ylabel('Frequency', fontsize=20)
plt.show()

```

[56] Python

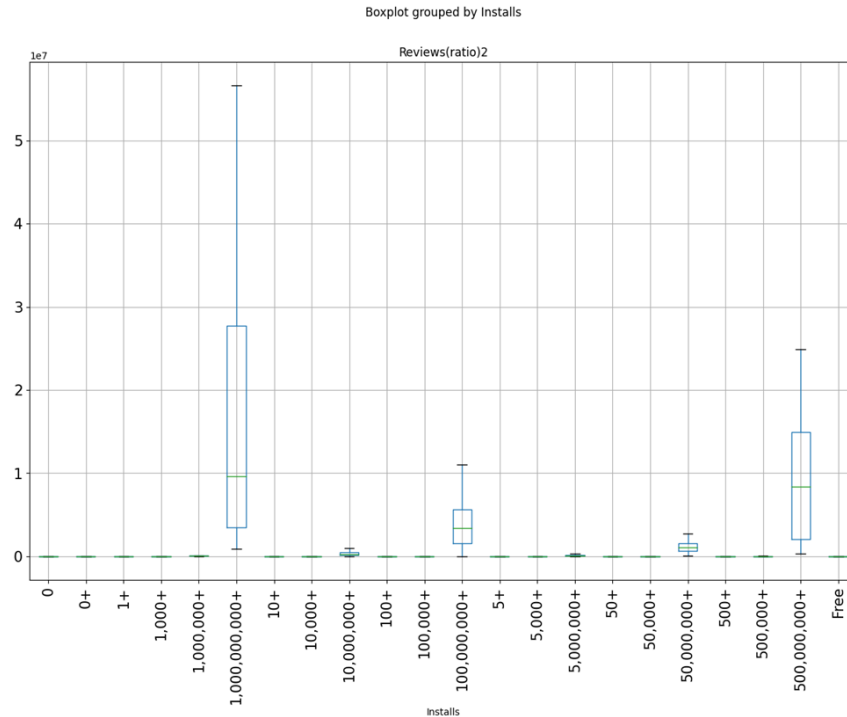


Distribution of reviews as a function of the number of installs

```

boxplot = Data.boxplot(column = 'Reviews(ratio)2', by = 'Installs', rot=90, fontsize=15, showfliers=False, figsize=
[59]
Python
...
Boxplot grouped by Installs

```



Distribution of types of applications

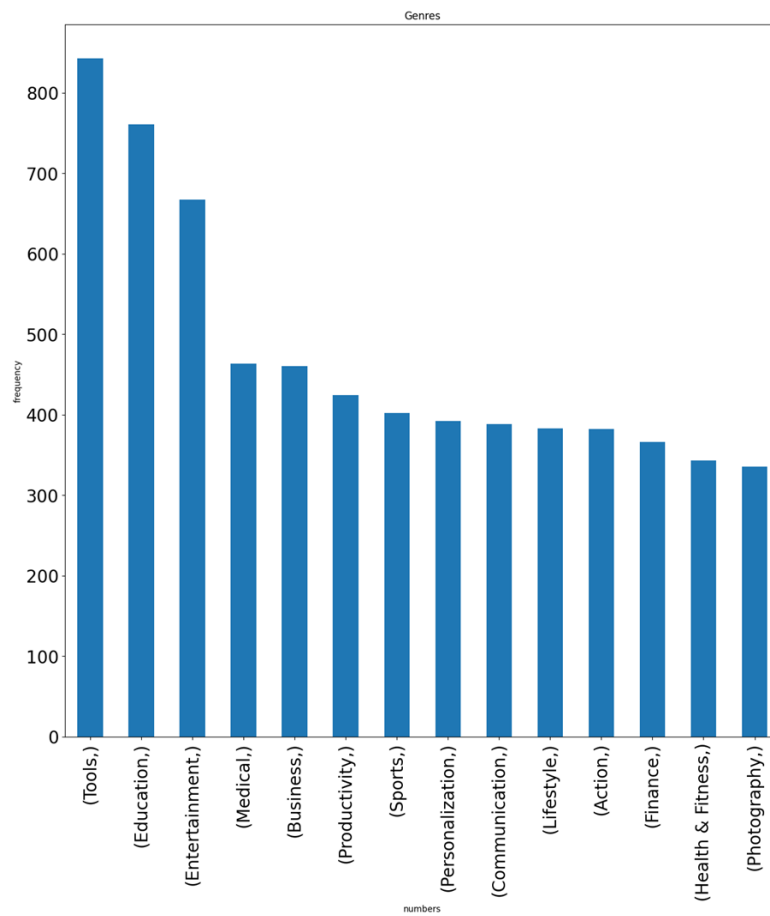
TOP TYPE

```

list_genre = []
for ii in Data['Genres']:
    list_genre.extend(ii.split(";"))

fig, ax = plt.subplots(figsize=(15, 15))
list_genre = pd.DataFrame(list_genre).value_counts()
list_genre[list_genre>300].plot(ax=ax, kind='bar', xlabel='numbers', ylabel='frequency', title="Genres", fontsize=
plt.show()
[110]
Python

```



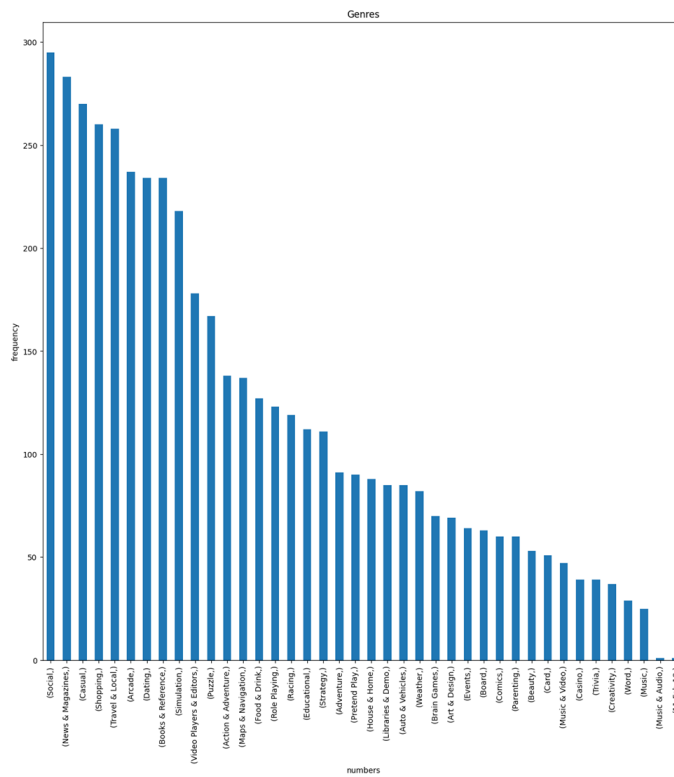
BOTTOM TYPE

```

Data['Price(interval)2'] = Data['Price(interval)'].str.replace(r'\D', '')
Data['Price(interval)2'] = pd.to_numeric(Data['Price(interval)2'], errors='coerce')
#Data['Price(interval)2'] = Data['Price(interval)2'].astype(float, errors='ignore')
fig, ax = plt.subplots(figsize=(15, 10))
Data['Price(interval)2'][Data['Price(interval)2']>0].value_counts().plot(ax=ax, kind='bar', fontsize= 10)
plt.xlabel('Numbers', fontsize=20)
plt.ylabel('Frequency', fontsize=20)
plt.show()
```

[78]

Python



Distribution of applications last update

By Month

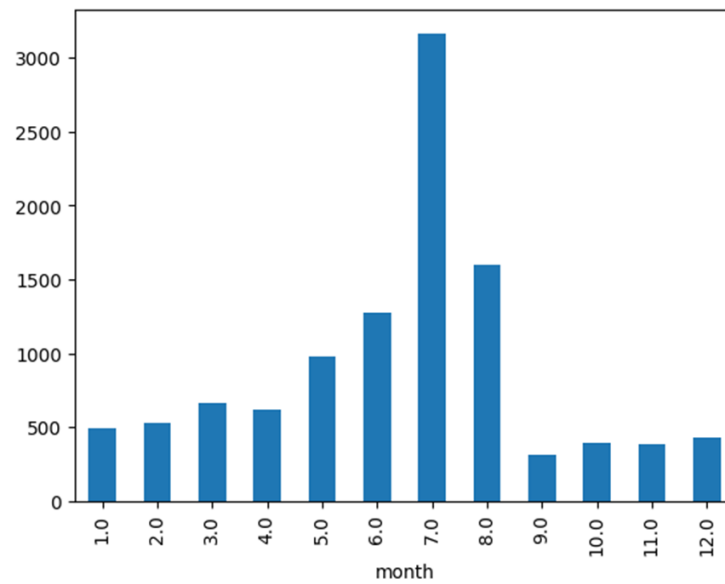
▷ ∨

```
Data['Last Updated2'] = pd.to_datetime(Data['Last Updated'], errors='coerce')
Data['year'] = Data["Last Updated2"].dt.year
Data['month'] = Data["Last Updated2"].dt.month
# Last Updated
Data['month'].groupby(Data['month']).count().plot(kind="bar")
```

[106]

Python

... <AxesSubplot: xlabel='month'>



By Year

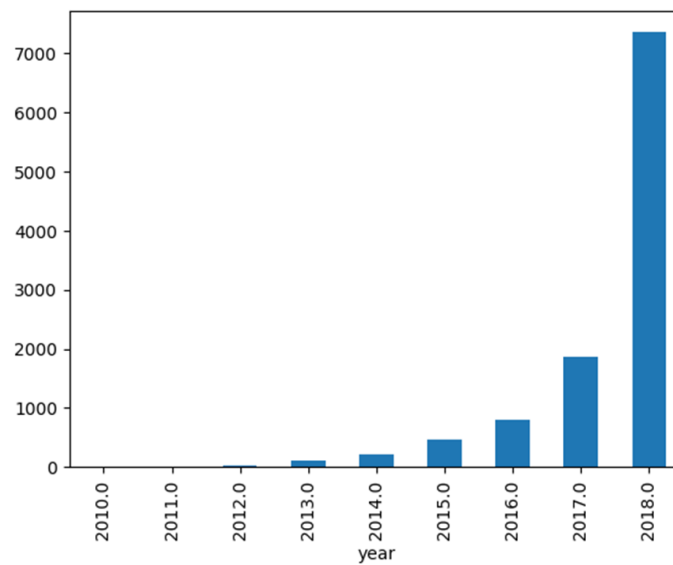
▷ ▾

```
Data['year'].groupby(Data['year']).count().plot(kind="bar")
```

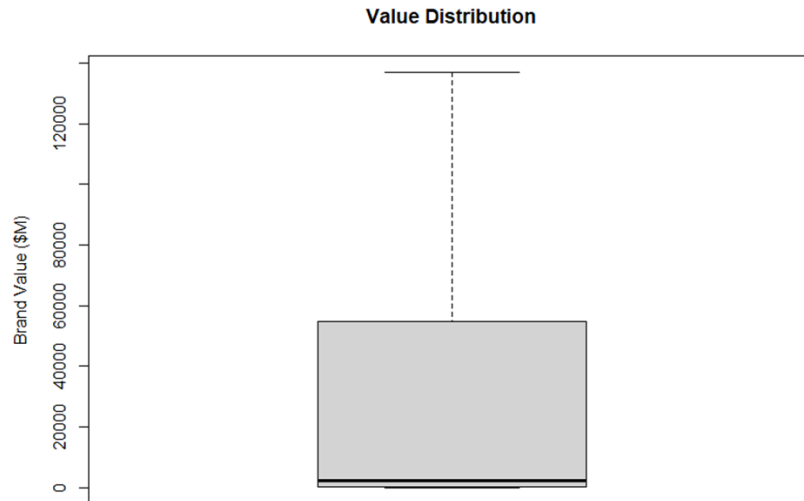
[107]

Python

... <AxesSubplot: xlabel='year'>



Review Distribution



RESULTS

Our research study contributes to the resulting finding which shows that facebook is the top most application having the most reviews on google play store as seen in the figure 09. The most installed applications are shown in Figure 10. Our third resulting answer shows that Star Wars is the most paid app on the Google Play Store following Pocket AC till the ten as shown in the figure 11.

TOP APPLICATION REVIEWS

	App(nominal)	Category	Rating(ordinal)	Reviews(ratio)	Size	Installs	Type	Price(interval)	Conte Ratin
1	Facebook	SOCIAL	4.1	76158306	Varies with device	1,000,000,000+	Free	0	Teen
2	Facebook	SOCIAL	4.1	76126208	Varies with device	1,000,000,000+	Free	0	Teen
3	WhatsApp Messenger	COMMUNICATION	4.4	69119316	Varies with device	1,000,000,000+	Free	0	Every
4	WhatsApp Messenger	COMMUNICATION	4.4	69119316	Varies with device	1,000,000,000+	Free	0	Every
5	WhatsApp Messenger	COMMUNICATION	4.4	69109672	Varies with device	1,000,000,000+	Free	0	Every

Figure 09: Top Application Reviews

MOST INSTALLS

```
# Q2 Most installs
x = c('1.2<', '>4.5', '3+.2', '-1d0', '2aaddab2', '1.3h')
dataset$Installs = gsub("+", "", dataset$Installs, fixed=TRUE)
dataset$Installs = gsub(",", "", dataset$Installs, fixed=TRUE)
dataset$Installs = strtoi(dataset$Installs, base=0L)
ggplot(dataset, aes(Category, Installs)) + geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust = 1))
```

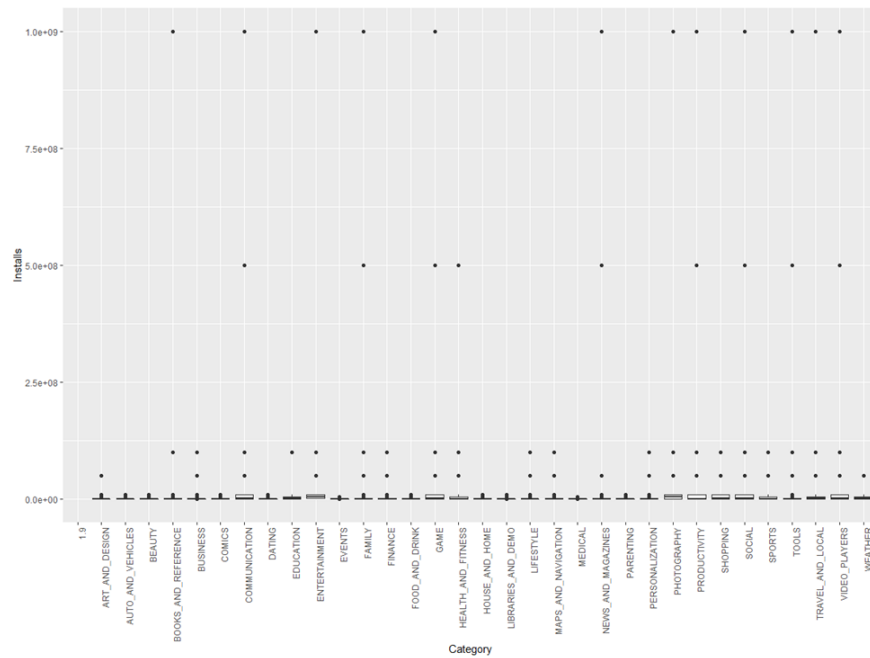


Figure 10: The most installed apps from Google Play Store

TOP APPS BY PRICE DISTRIBUTION

```

25 # Q3
26 dataset$`Price(interval)` = gsub("$", "", dataset$`Price(interval)` , fixed=TRUE)
27 dataset$`Price(interval)` = gsub("Everyone", "0", dataset$`Price(interval)` , fixed=TRUE)
28
29 dataset$`Price(interval)` = strtoi(dataset$`Price(interval)` , base=0L)
30

```

	App(nominal)	Category	Rating(ordinal)	Reviews(ratio)	Size	Installs	Type	Price(interval)
1	Star Wars™: DIRTY	GAME	4.5	38207	15M	100,000+	Paid	9.99
2	ASCCP Mobile	MEDICAL	4.5	63	25M	10,000+	Paid	9.99
3	Diabetes & Diet Tracker	MEDICAL	4.6	395	19M	1,000+	Paid	9.99
4	Critical Care Paramedic Review	MEDICAL	4.4	17	1.8M	1,000+	Paid	9.99
5	InfantRisk Center HCP	MEDICAL	2.6	41	14M	1,000+	Paid	9.99
6	ASCCP Mobile	MEDICAL	4.5	63	25M	10,000+	Paid	9.99
7	Diabetes & Diet Tracker	MEDICAL	4.6	395	19M	1,000+	Paid	9.99
8	BIG Launcher	PERSONALIZATION	4.3	681	5.9M	10,000+	Paid	9.99
9	FORD V SERIES CALC - NO LIMIT	AUTO_AND_VEHICLES	NaN	2	17M	50+	Paid	9.99
10	Pocket AC	PHOTOGRAPHY	4.8	130	4.4M	1,000+	Paid	9.99

Figure 11: Top apps by price distribution

CONCLUSION

The research paper contributes that Google Play Store has an impact on people and companies all over the world. In our era of rapid technological advancement, everything we need is just a few clicks away. Many engineers have created new applications that have transformed the world in various presentable ways with easier technology and contributing to billions of dollars on a

daily basis as shown in the figure 11 , and one such firm, Google, has risen to the top of the tech industry in terms of app development (Thakre, 2019).

REFERENCES

1. Sharma, A. (2022, July 26). *Key google play store statistics in 2022 you must know*. Appinventiv. Retrieved December 5, 2022, from <https://appinventiv.com/blog/google-play-store-statistics/>
2. Thakre, A. (2019, March 24). *Data science tutorial: Analysis of the google play store dataset*. Medium. Retrieved December 5, 2022, from <https://medium.com/the-research-nest/data-science-tutorial-analysis-of-the-google-play-store-dataset-c720330d4903>
3. *Students projects: NYC data science academy*. Data Science Blog. (n.d.). Retrieved December 5, 2022, from <https://nycdatascience.com/blog/>
4. Olmstead, K., & Atkinson, M. (2022, March 8). *Chapter 2: An analysis of apps in the google play store*. Pew Research Center: Internet, Science & Tech. Retrieved December 5, 2022, from <https://www.pewresearch.org/internet/2015/11/10/an-analysis-of-apps-in-the-google-play-store/>
5. *Analysis of apps in the google play store*. Data Science Blog. (n.d.). Retrieved December 5, 2022, from <https://nycdatascience.com/blog/student-works/web-scraping/analysis-of-apps-in-the-google-play-store/>

