



Data Engineering Programming Assignment

Solutions

Description of Project

1. Name of Data - Google Playstore Data
2. Format - xls or csv file format
3. Data Source - provided by company

Software used



From ~

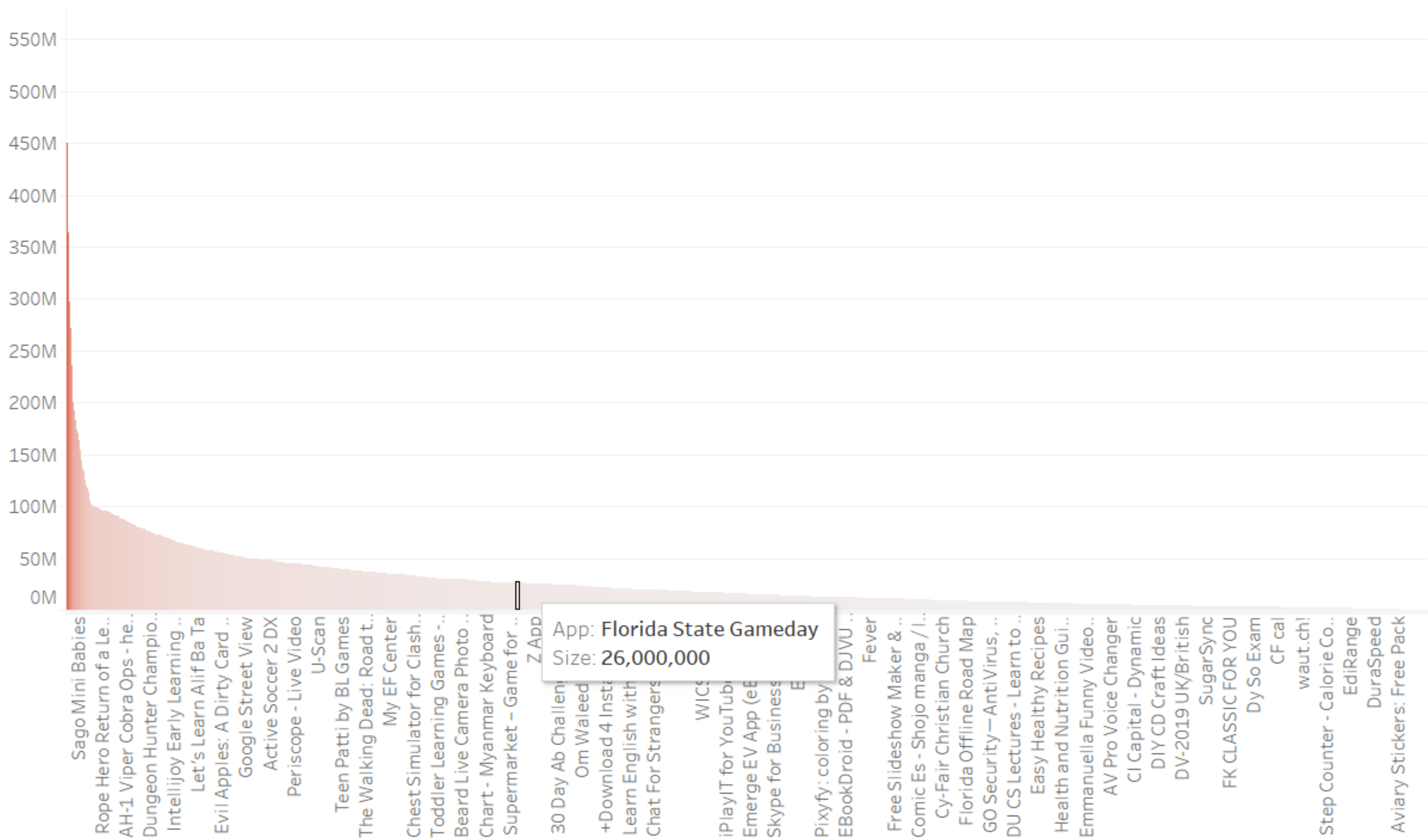
Anurag Singh

Data Scientist

B-Tech Information Technology

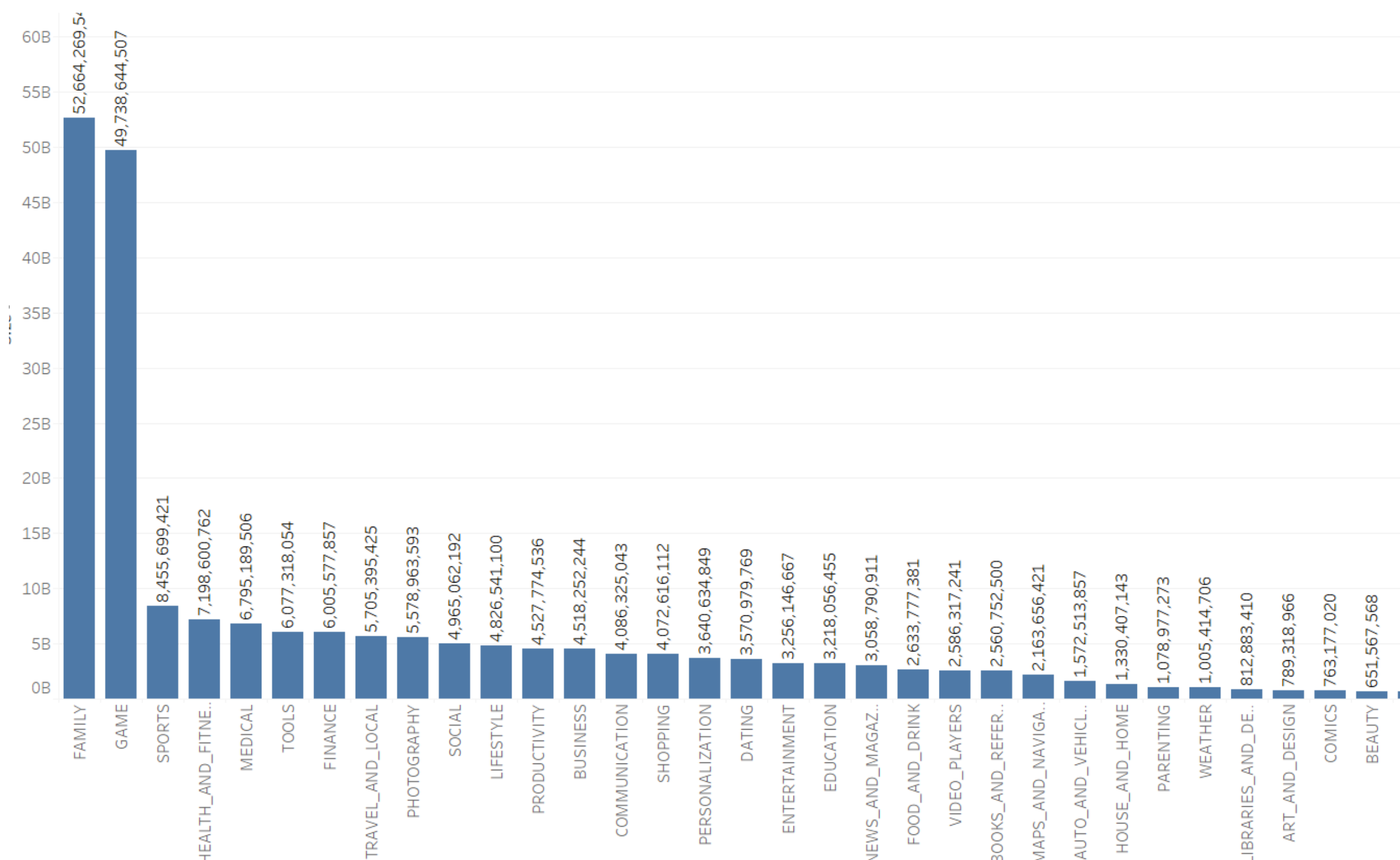
1) What is the range of the size of most of the Apps ? (2)

Below graph shows the relations between size of the Apps and App names



From above distribution we observe that -

- 1. We conclude that % of long size app is not so dominant as 450M to 100M there is long steep , which shows its percentage is less.
- 2. Above distribution prove that size below 100M to 10M occupy more area , representing maximum percentage of cover in the market.



- 3. Figure 2 shows category wise distribution of of Apps in Size
- 4. From Figure 2 we conclude that Family Apps and Gaming apps have maximum size
- 5. More than 80% of Apps lies in between the 100Mb - 10Mb size

2. What approach you will implement to fill the missing values of the Size column & why ? (2)

We are going to fill missing or NaN data with **mean** as Size is numerical variable and mean is best representator of any missing or datapoint. in this case.

```
#filling all NaN value with mean
data['Size'].fillna(data.groupby('Category')['Size'].transform('mean'),inplace = True)
```

D:\anaconda1\lib\site-packages\pandas\core\series.py:4463: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

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

```
return super().fillna(
```

Rating	
count	9367.000000
mean	4.193338
std	0.537431
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	19.000000

Having Standard deviation(std) equal to 0.537431 also tell us that that data is with in around mean and replacing it with mean is good decision

Having 25 %=4 , 50% =4.3 and 75% = 4.5 shows our distribution around size is normal hence mean is good decition

3. What is the app name having max price? (1)

```
#App Having Maximum Price
data[data.Price==400]
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Installforplot
4367	I'm Rich - Trump Edition	LIFESTYLE	3.6	275	7300000.0	10000	Paid	400.0	Everyone	Lifestyle	May 3, 2018	1.0.1	4.1 and up	8

App Having Maximum Price is **I'm Rich Trump Edition** whose price is 400 Doller.

Category - Lifestyle

Rating- 3.6

Installs - 10000M

all details are in above screen shot of this app

4) List top 20 apps in Game Category (Based on Number of Installation) (1)

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Google Play Books	BOOKS_AND_REFERENCE	3.9	1433233	1.438625e+07	1000000000	Free	0.0	Teen	Books & Reference	2018-08-03	Varies with device	Varies with device
Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56642847	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-01	Varies with device	Varies with device
WhatsApp Messenger	COMMUNICATION	4.4	69119316	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-03	Varies with device	Varies with device
Google Chrome: Fast & Secure	COMMUNICATION	4.3	9642995	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-01	Varies with device	Varies with device
Gmail	COMMUNICATION	4.3	4604324	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-02	Varies with device	Varies with device
Hangouts	COMMUNICATION	4.0	3419249	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-07-21	Varies with device	Varies with device
WhatsApp Messenger	COMMUNICATION	4.4	69119316	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-03	Varies with device	Varies with device
Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56646578	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-01	Varies with device	Varies with device

Hangouts	COMMUNICATION	4.0	3419433	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-07-21	Varies with device	Varies with device
Skype - free IM & video calls	COMMUNICATION	4.1	10484169	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-03	Varies with device	Varies with device
Google Chrome: Fast & Secure	COMMUNICATION	4.3	9643041	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-01	Varies with device	Varies with device
Gmail	COMMUNICATION	4.3	4604483	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-02	Varies with device	Varies with device
Hangouts	COMMUNICATION	4.0	3419513	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-07-21	Varies with device	Varies with device
Skype - free IM & video calls	COMMUNICATION	4.1	10484169	1.245831e+07	1000000000	Free	0.0	Everyone	Communication	2018-08-03	Varies with device	Varies with device
Google Play Games	ENTERTAINMENT	4.3	7165362	2.185333e+07	1000000000	Free	0.0	Teen	Entertainment	2018-07-16	Varies with device	Varies with device
Subway Surfers	GAME	4.5	27722264	7.600000e+07	1000000000	Free	0.0	Everyone 10+	Arcade	2018-07-12	1.90.0	4.1 and up
Subway Surfers	GAME	4.5	27723193	7.600000e+07	1000000000	Free	0.0	Everyone 10+	Arcade	2018-07-12	1.90.0	4.1 and up
Subway Surfers	GAME	4.5	27724094	7.600000e+07	1000000000	Free	0.0	Everyone 10+	Arcade	2018-07-12	1.90.0	4.1 and up
Subway Surfers	GAME	4.5	27725352	7.600000e+07	1000000000	Free	0.0	Everyone 10+	Arcade	2018-07-12	1.90.0	4.1 and up
Subway Surfers	GAME	4.5	27725352	7.600000e+07	1000000000	Free	0.0	Everyone 10+	Arcade	2018-07-12	1.90.0	4.1 and up

- Google Play Books
- Messenger – Text and Video Chat for Free
- WhatsApp Messenger
- Google Chrome
- Fast and Secure
- Gmail
- Hangouts
- Skype - free IM & video calls
- Google Play Games
- Subway Surfers
- Facebook
- Instagram
- Google+
- Google Photos
- Maps - Navigate & Explore

- Google
- Google Drive
- YouTube
- Google Play Movies & TV
- .Google News

5) Is it better that developers use a concise word or words to name their app? (Justify your answer with Data Analysis) (4)

concise - means giving lot of information in few words

1. From below data we confirm that writing whole sentence as a name of the App is not good decision of Developer

2. As when we compare App name with the installation in decreasing order we found most of the App having less Installation have name as full sentence of meaningless sort form words

App	Installs
Mu.F.O.	1
Ra Ga Ba	1
KBA-EZ Health Guide	1
CB Heroes	5
EB Cash Collections	5
Endurance Lifestyle	500
Differential Dx	500
G Theme 2 for LG V30, LG G6	500
Ek Vote	500
Navi Radiography Pro	500
El国际	1000
Simple Blood Pressure log	1000
17th Edition Cable Sizer	1000
Walk Freely (Ec Shlire)	1000
DT Driving Test Theory	10000
HTC Sense Input - CZ	10000
Lite Messenger for Facebook Lite	10000
Name Art DP - Focus n Filter Text 2018	10000
DRAGON QUEST III	10000

App	Reviews
Food-Aw - Order Food Online in Aruba	1
LC-DB	1
DN Employee	1
ES Billing System (Offline App)	1
Ruler	126
Manga AZ - Manga Comic Reader	126
I am Millionaire - Richest guy in the town	126
Plug-in app (System AC)	126
MomentCam Cartoons & Stickers	1260143
DQ Texas	336
AC Remote for Haier - NOW FREE	336
Safe Notes - Secure Ad-free notepad	3362
Resume Builder Free, CV Maker & Resume Templates	20921
FutureNet your social app	2093
Droid Zap by Motorola	2093
Teladoc Member	2094

2. Not only this when we compare App name with Review of people ,result amazed as Review also have downfall with no meaning words

Yes I confirm that it better that developers use a concise word or words to name their app-

Below are the list of Top Apps on Playstore along with their Installation and Reviews

App	Reviews	Installs
Gmail	4604483	1000000000
Gmail	4604324	1000000000
WhatsApp Messenger	69119316	1000000000
Google Chrome: Fast & Secure	9642995	1000000000
Gmail	4604324	1000000000
Microsoft Word	2084126	500000000
Facebook Lite	8595964	500000000
My Talking Tom	14891223	500000000
Temple Run 2	8118937	500000000
LINE: Free Calls & Messages	10790092	500000000
Tapatalk - 100,000+ Forums	285816	10000000
LOVOO	852455	10000000
Jaumo Dating, Flirt & Live Video	900064	10000000
textPlus: Free Text & Calls	382121	10000000
Gmail	4604483	1000000000
Google News	877635	1000000000
Google Photos	10858538	1000000000

4. Above data clear our doubt that Apps which had captured the market are the leader and leaders apps have concise name .
5. We have seen Review are also at its peak which guarantee the preference of app among the people

Conclusion!

1. Apps having concise and meaning full name increases its **Search Engine** due to which these apps are easily searchable in sea of content on internet
2. Easy , Concise name as a name of App increases its human to human **referrals**. as it is on mouth of people.
3. In our Analysis we have seen that **top brands** like google ,Facebook and so many others uses concise words to name their apps. (Even Facebook declared their new name as meta)
4. When we make list of app which are as part of our life have very concise and **meaningful name** - like PubG (popular among Indian youth) , WhatsApp , Facebook , Canva , Chrome , Telegram youtubeso on