

Capstone Project

Play Store App Review Analysis

by

Roshan Jamthe

Points for Discussion

- Data Summary
- Rating
- User Reviews
- App Size
- Minimum Android Version support
- Category
- Type
- Conclusion

Data Summary

Following datasets are the raw datasets I got for the Play Store App Review analysis EDA project,

- `play_store_df` : It contains all the information regarding app, category, rating, reviews, size, installs, type, price, android version and other information of Play store apps.
- `user_reviews_df` : It contains all the information regarding app, review comment, sentiment, polarity, subjectivity of User reviews on Play store apps.

Data Summary (contd.)

Following datasets are the few important derived datasets used at various sub-levels in the project,

- rating_with_Sentiment_Score_df: Information regarding rating and sentiment score
- rating_wise: Information regarding app, reviews, installs and size in accordance to rating
- size_wise: Information regarding app, reviews and installs in accordance to size
- Android_Ver_wise: Information regarding app, reviews, installs and size in accordance to min android version
- category_wise: Information regarding app, reviews and more in accordance to category
- category_earnings_through_paid_app : Information in accordance to paid type app

Rating

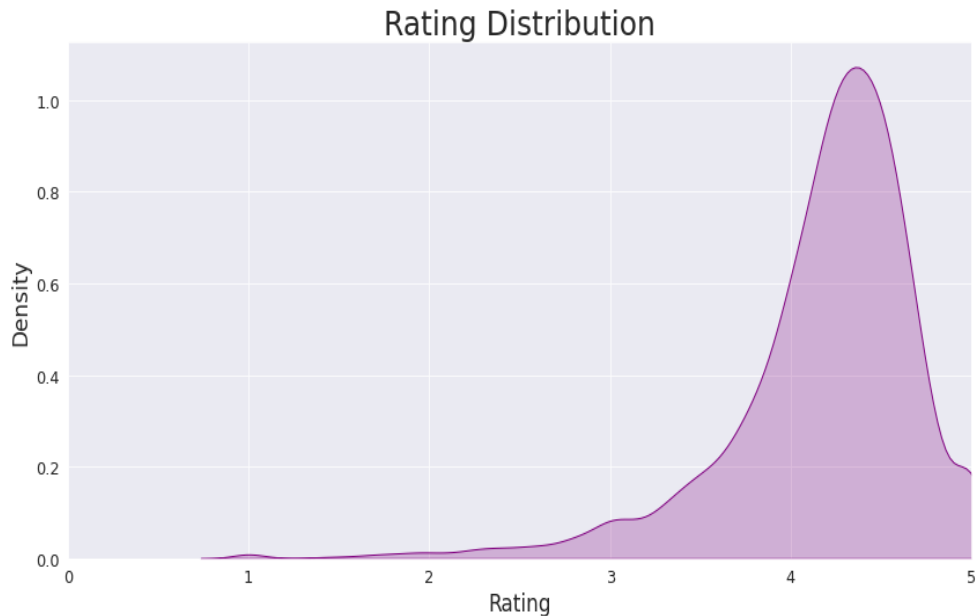
I analyzed features in accordance to Rating, to answer the following questions,

1. How is Rating feature distributed?
2. Relation between rating and sentiment score.
3. Relation between rating and number of installs.
4. Top 5 highest rated apps having 10 million+ installs.

Rating Distribution

The observations and inferences are,

1. Rating distribution is negatively skewed having a peak at 4.2.
2. The mean rating value is 4.17, which is pretty good.
3. Inference is clear that apps are rated pretty good on average.



Rating vs Sentiment Score, Number of Installs

Used 5 pointer classification of Rating to obtain clear conclusions,

[0.0 - 1.0] terming it as 'Very Bad' Rating

[1.0 - 2.0] terming it as 'Bad' Rating

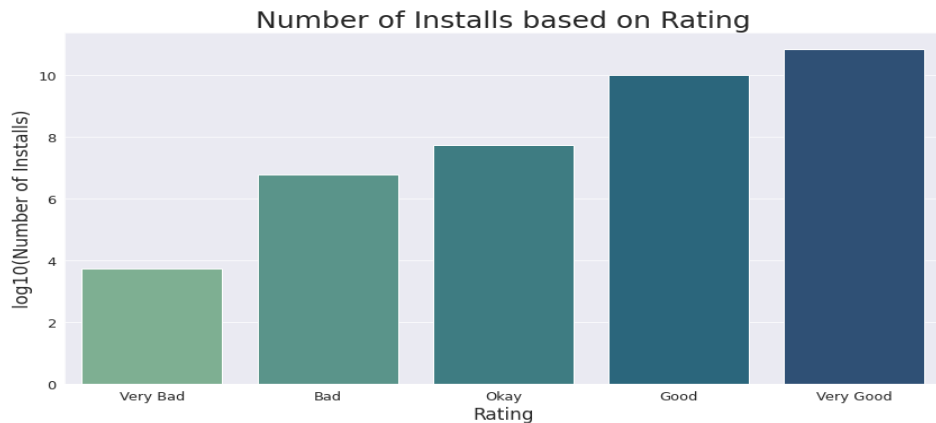
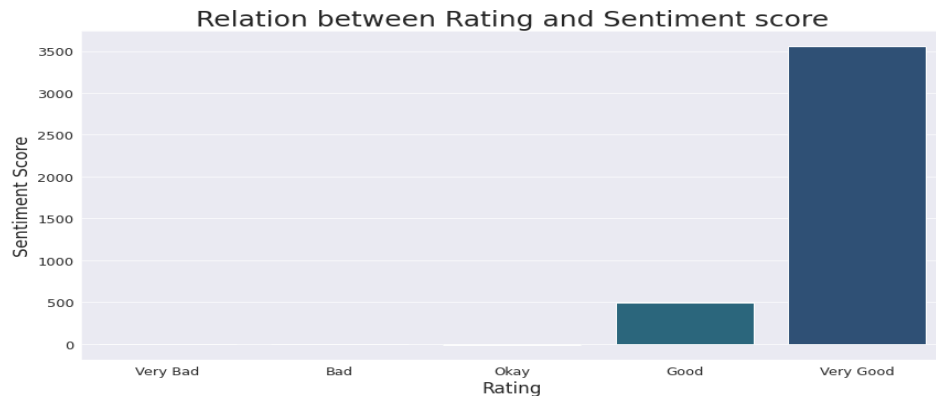
[2.0 - 3.0] terming it as 'Okay' Rating

[3.0 - 4.0] terming it as 'Good' Rating

[4.0 - 5.0] terming it as 'Very Good' Rating

The observations and inferences are,

1. Sentiment score is higher for higher rating category.
2. Number of Installs dramatically increases as rating improves.



Top 5 highest rated apps having 10 million+ installs

App	Category	Rating	Reviews	Type	Installs
JW Library	BOOKS_AND_REFERENCE	4.9	922752	Free	10000000
Six Pack in 30 Days - Abs Workout	HEALTH_AND_FITNESS	4.9	272337	Free	10000000
Period Tracker - Period Calendar Ovulation Tra...	HEALTH_AND_FITNESS	4.8	4559407	Free	100000000
DU Recorder – Screen Recorder, Video Editor, Live	VIDEO_PLAYERS	4.8	2588730	Free	50000000
Eternium	FAMILY	4.8	1506783	Free	10000000

- These are top 5 highest rated apps which are successful and engaging.
- But rating alone is a poor indicator to decide whether the app is successful.
- It has to be supported by other features as well. Let's find out which other features...

App	Rating	Installs
CL REPL	5.0	1000
Cb browser	5.0	50
CJ'S TIRE AND AUTO INC.	5.0	100
30WPM Amateur ham radio Koch CW Morse code tra...	5.0	10
CI 174 Gray Icon Pack	5.0	10

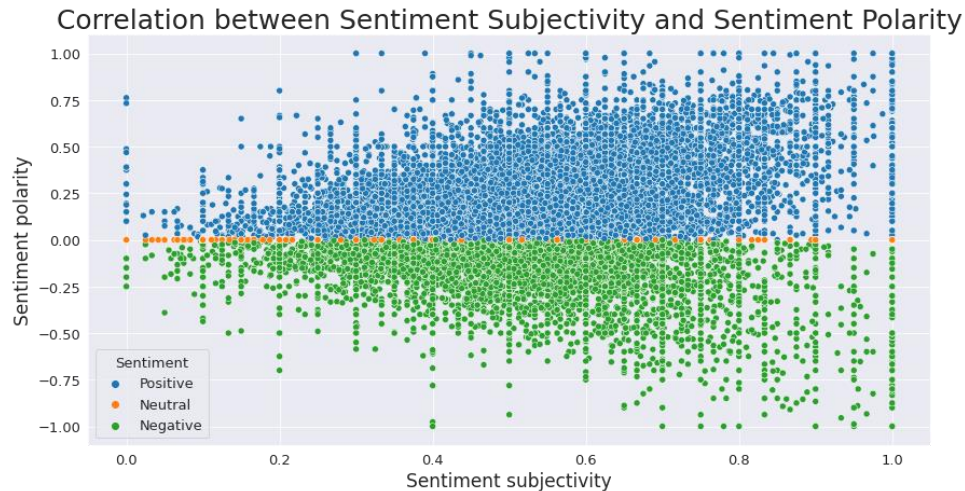
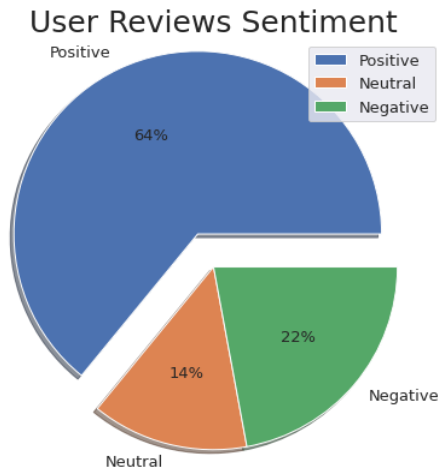
User Reviews

I analyzed `user_reviews_df`, to get answers for the following questions,

1. How is Sentiment Distributed?
2. How is Subjectivity Distributed?
3. Does there exist a correlation between Sentiment Subjectivity and Sentiment Polarity?

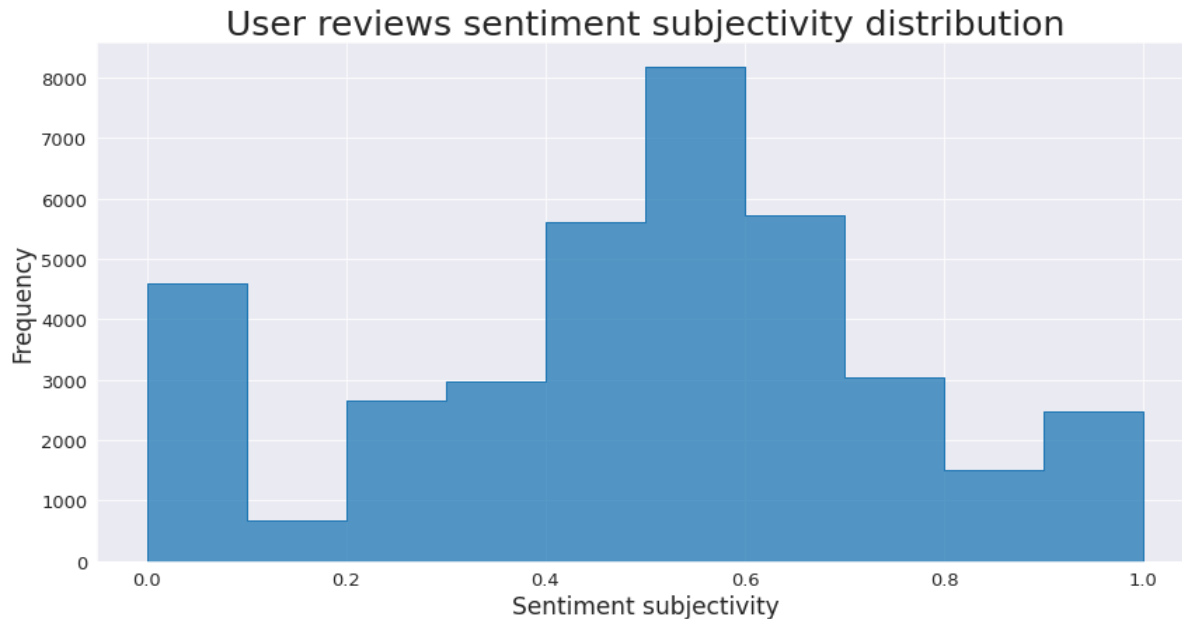
Sentiment Distribution

- Observing the pie chart, it can be inferred that Positive sentiment fairly dominates.
- Also, healthy 20% negative sentiment shows that there are areas to improve according to users.
- There appears to be no correlation between subjectivity and polarity (figure on right).



Subjectivity Distribution

- Sentiment Subjectivity lies mostly in range (0.4 - 0.7), i.e. most number of users gives reviews based on their experience. So, reviews can be held as a key factor.



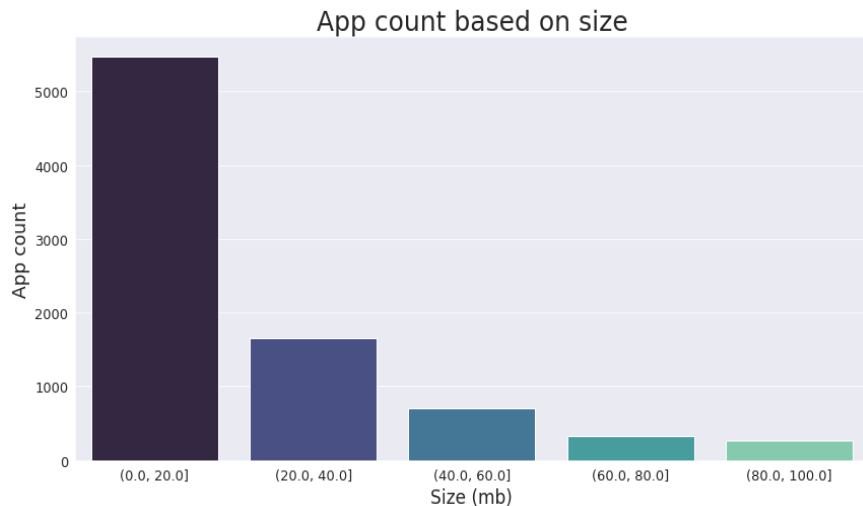
App Size

I analyzed features in accordance to App Size, to get answers for the following questions,

1. App count based on App Size
2. Does small app size engages more users?

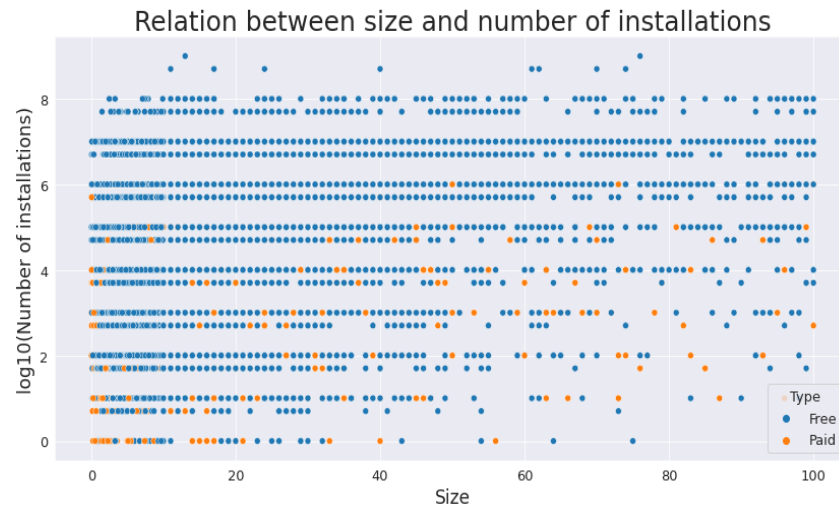
App Size vs App Count

- Most apps are below 20mb app size.
- Also, there are lesser and lesser apps with higher app sizes.



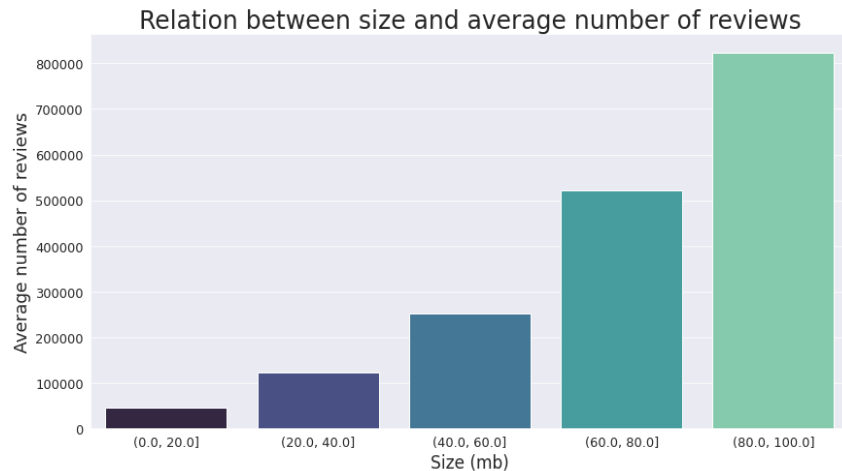
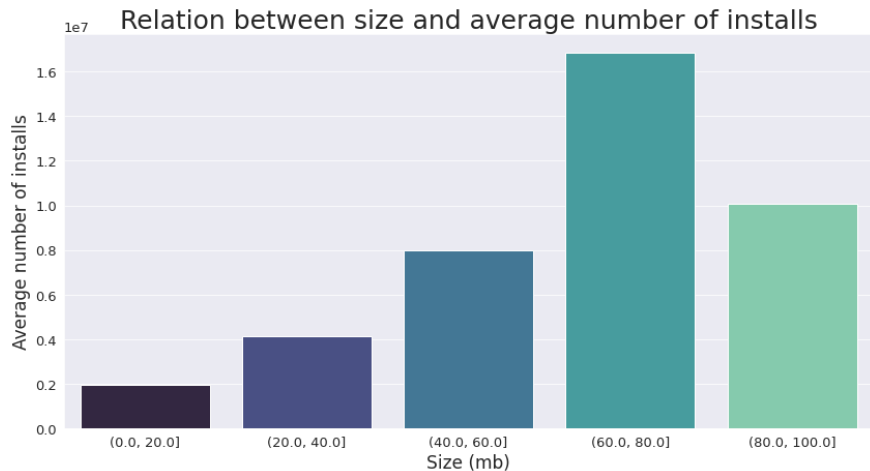
App Size vs Total Installs

- It appears that, Number of Installs decreases as app size increases.
- Free apps are far more than paid apps.



Does small app size engages more users?

- More features make an app bulky.
- Although there are high number of small size apps in the market, its evident from plots below that app with more features attracts more users.



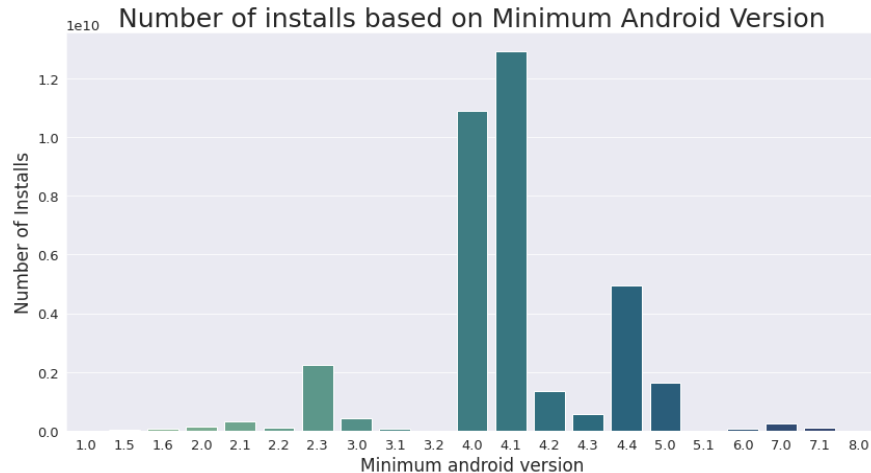
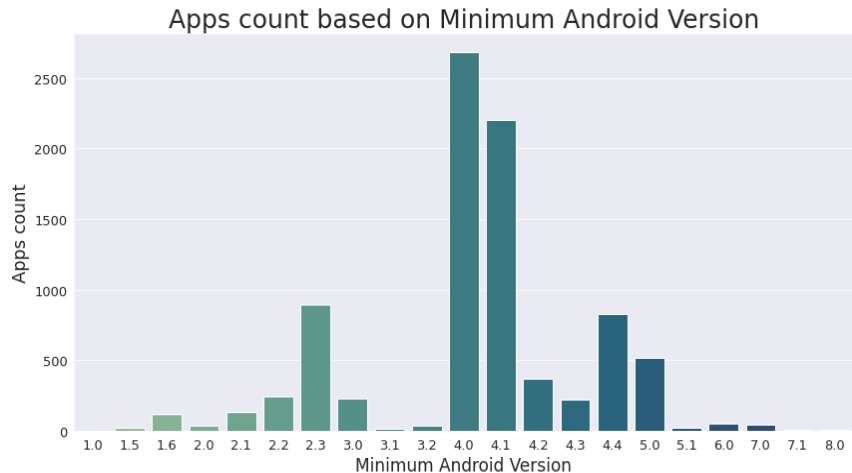
Minimum Android Version support

I analyzed features in accordance to Android Version, to get answers for the following questions,

1. What is that Minimum Android Version support the app makers should target?
2. Does the App Size increases with decreased Minimum Android Version support?

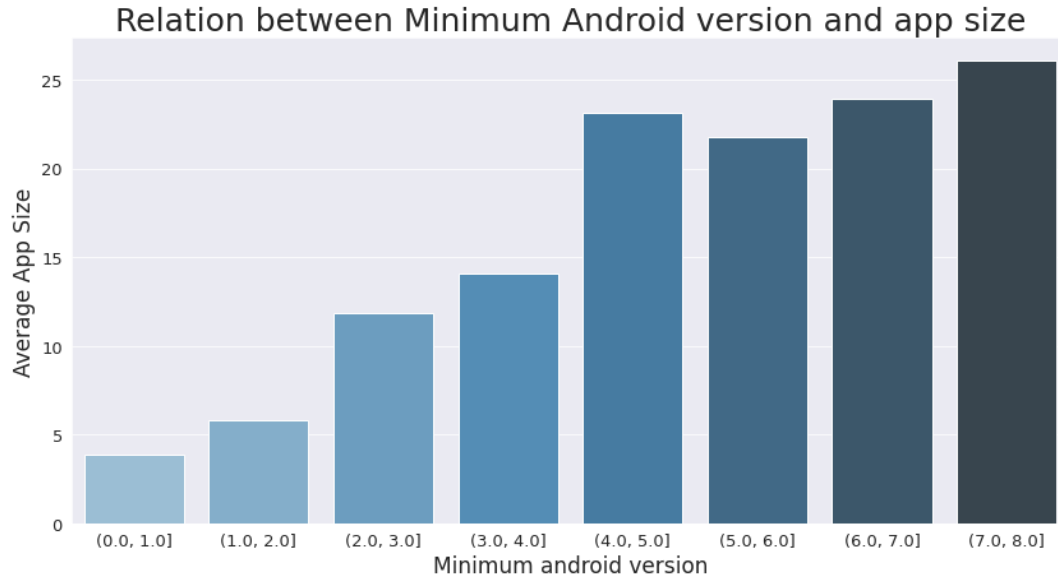
What is that Minimum Android Version support the app makers should target?

- It's important to have that optimum android version till which you want to extend your app support, since it directly affects the cost to make the app, maintenance cost and time etc.
- I conclude that, the support should be extended till android version 4.0.



Does the Average App Size increases with decreased Minimum Android Version support?

- Yes, the average app size increases with decreased minimum android version support.



Category

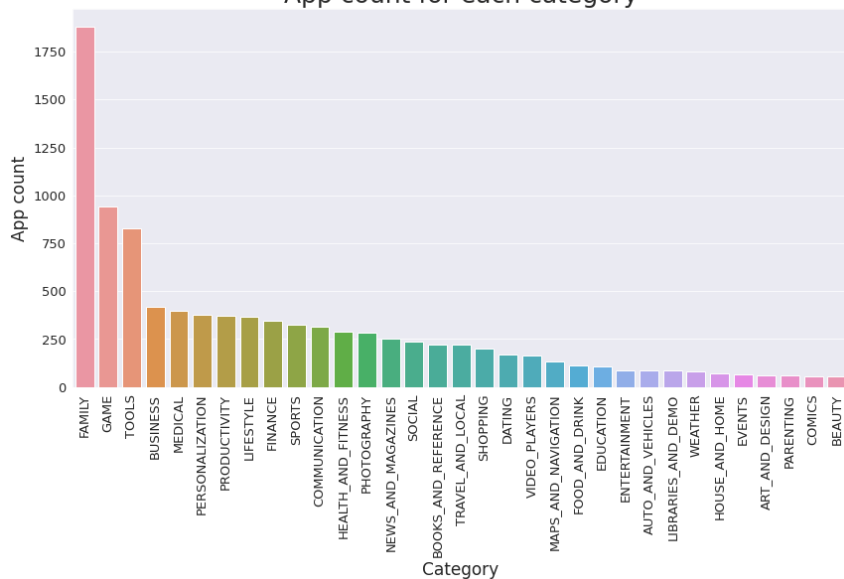
I analyzed features in accordance to Category, to get answers for the following questions,

1. App count for each category
2. Number of Installs for each category
3. Number of reviews for each category
4. Average App Size for each category
5. Most Positively Sentimental Category
6. What are the possible 'Next 5 Popular Category'?

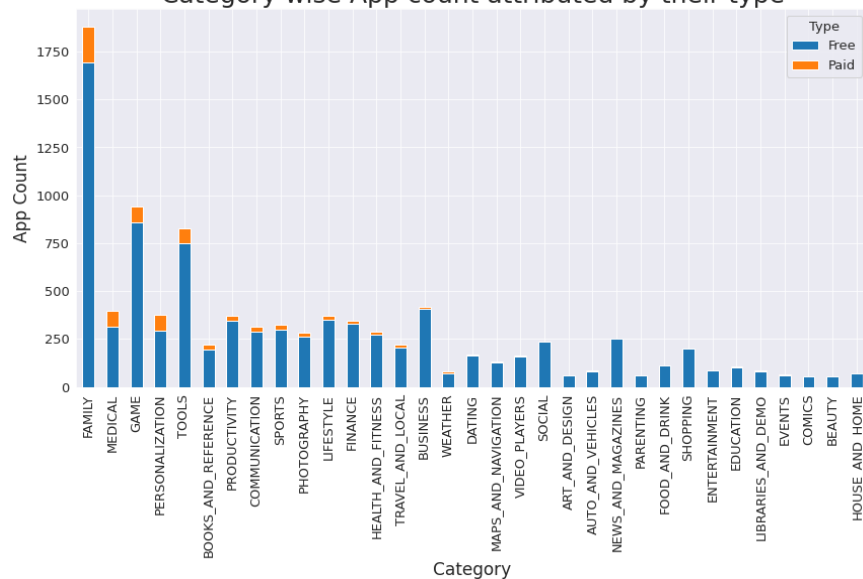
App count for each category

- Top 7 categories namely Family, Game, Tools, Business, Medical, Personalization and Productivity contributes to 54% of total apps on the play store.
- Family category has the most number of apps compared to all the other categories.

App count for each category

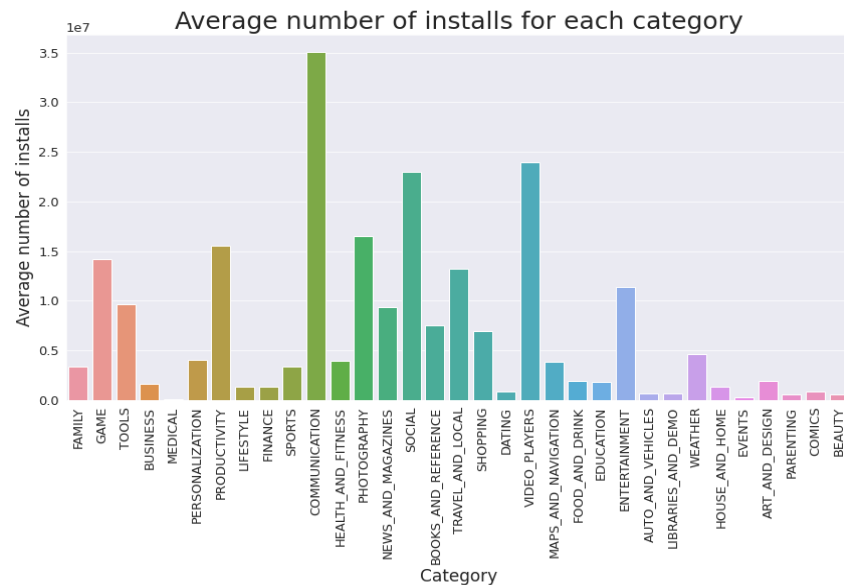
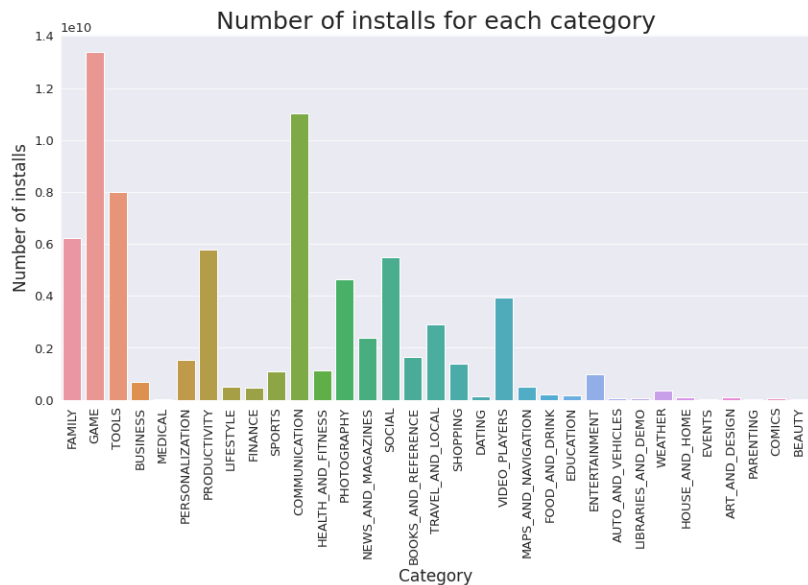


Category wise App count attributed by their type



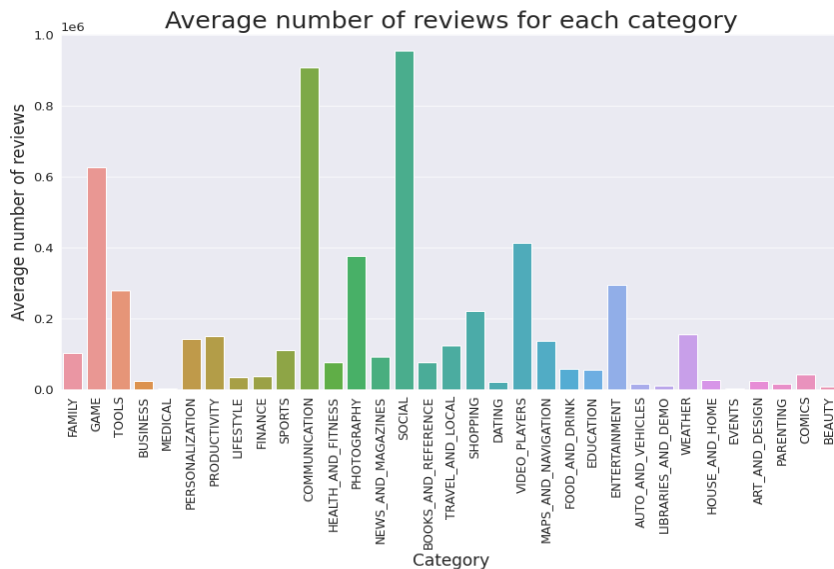
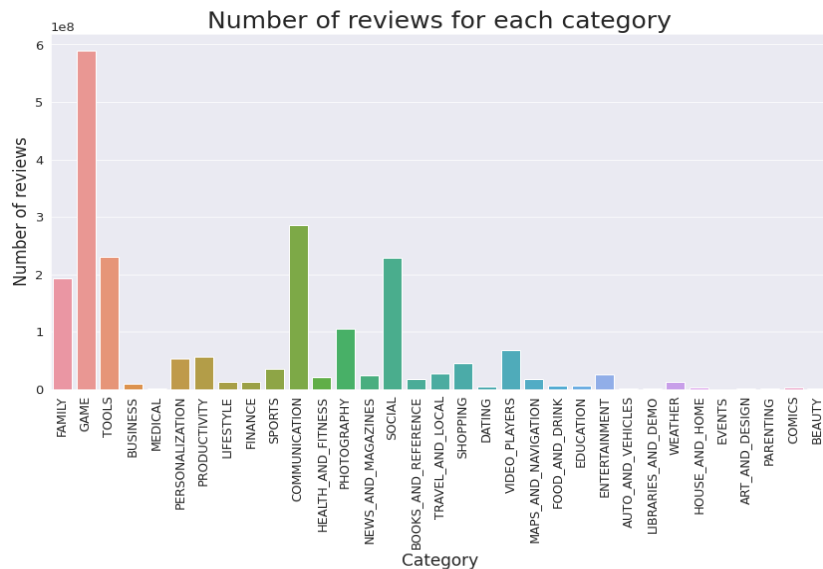
Total and avg. no. of installs for each category

- 7 categories namely Game, Communication, Tools, Family, Productivity, Social, Video players and Photography contributes to 77% of total number of installations.
- Game has the highest total installs while communication has the highest average installs.



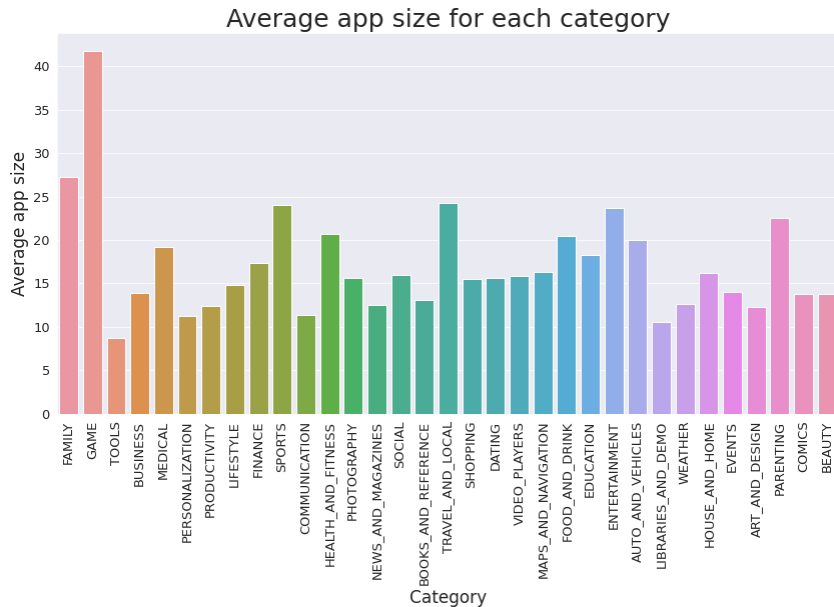
Total and avg. no. of reviews for each category

- Apps from 5 categories Game, Communication, Photography, Social and Video players were able to engage users the most compared to other categories.



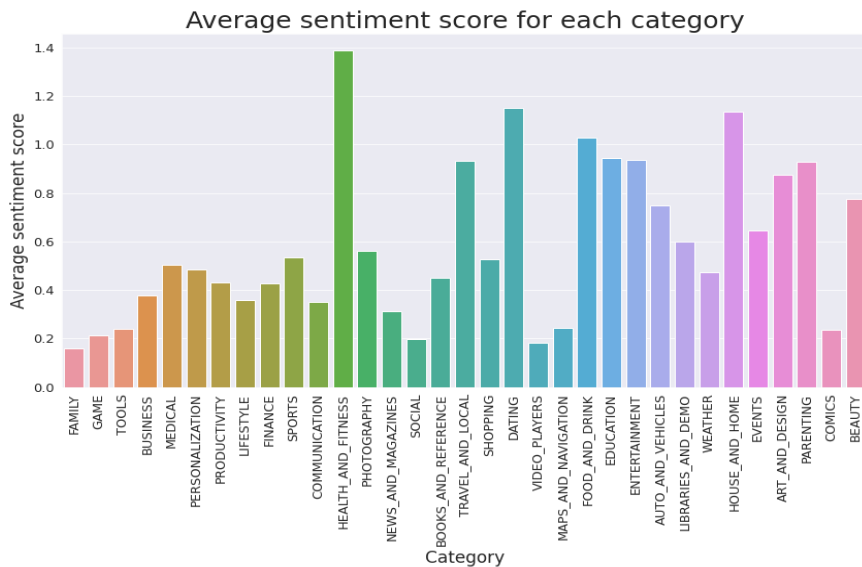
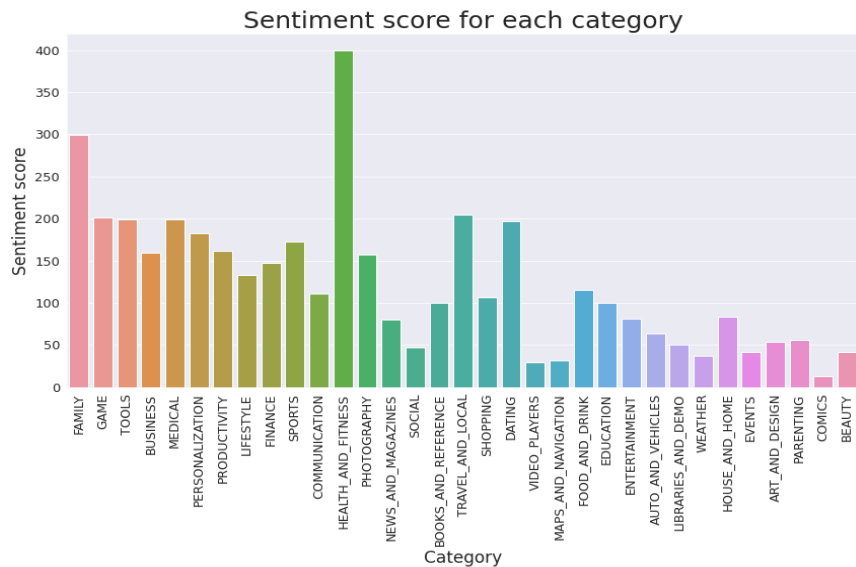
Average app size for each category

- The average app size of game category is significantly higher than the average app size of other categories. It is already established that game is one of the most engaging category.



What is the most positively sentimental category?

- All the categories have positive sentiment score. But there appears to be a lot of scope for improvement in Social, Video Players, Maps and Navigation and comics categories according to users.
- Health and Fitness category has the highest avg. as well as total sentiment score.



What are the possible 'Next 5 Popular Category'?

Category	Sentiment Score	Rating	App	Reviews	Installs	Size	Avg. Installs	Avg. Reviews	Avg. Sentiment Score
HEALTH_AND_FITNESS	399.2	4.2	288	21365720	1144022512	20.67	3.972300e+06	74186	1.386172
DATING	196.4	3.9	171	3623899	140926107	15.66	8.241293e+05	21192	1.148725
FOOD_AND_DRINK	115.3	4.1	112	6326064	211798751	20.49	1.891060e+06	56482	1.029364
EDUCATION	100.1	4.3	106	5788464	189451000	18.33	1.787274e+06	54608	0.944390
TRAVEL_AND_LOCAL	204.1	4.0	219	26821349	2894887147	24.20	1.321866e+07	122471	0.931908

- Avg. sentiment score represents what people are reacting towards the most and it favors the developing and undeveloped.
- Travel and local, Health and fitness, Food and drink, Education and finally Dating may be the part of next popular categories.

Type

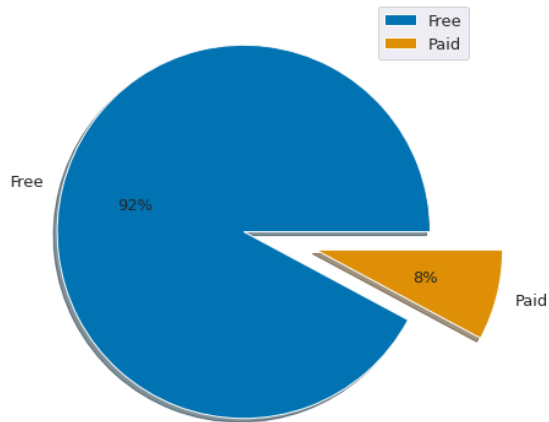
I analyzed features in accordance to Type, to get answers for the following questions,

1. Percentage of Paid type vs Free type apps
2. Average Number of Installs attributed by App type
3. Earnings from installs for each category
4. Top 10 Paid Apps (based on earnings from installs)

% of Paid type apps

- Free type app dominates the play store.
- Only 8% of total apps on the play store are paid apps and the rest are free apps.

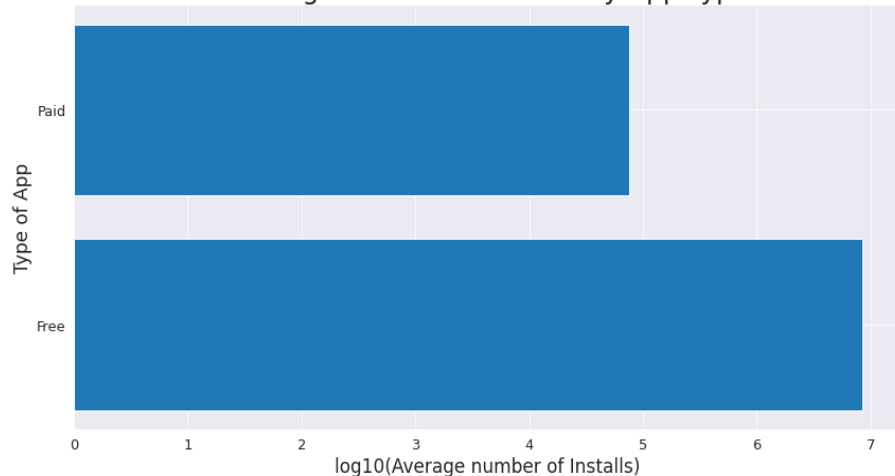
Percentage of Free vs Paid apps



Avg. Installs based on Type

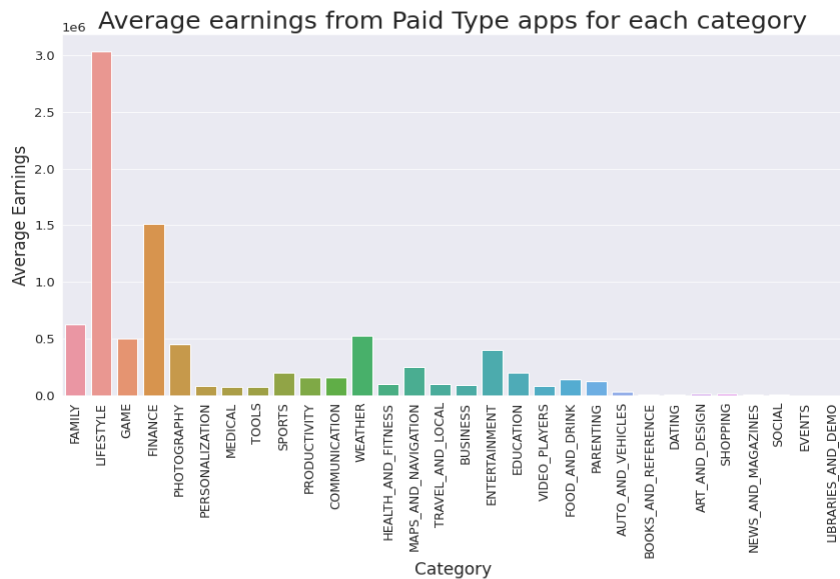
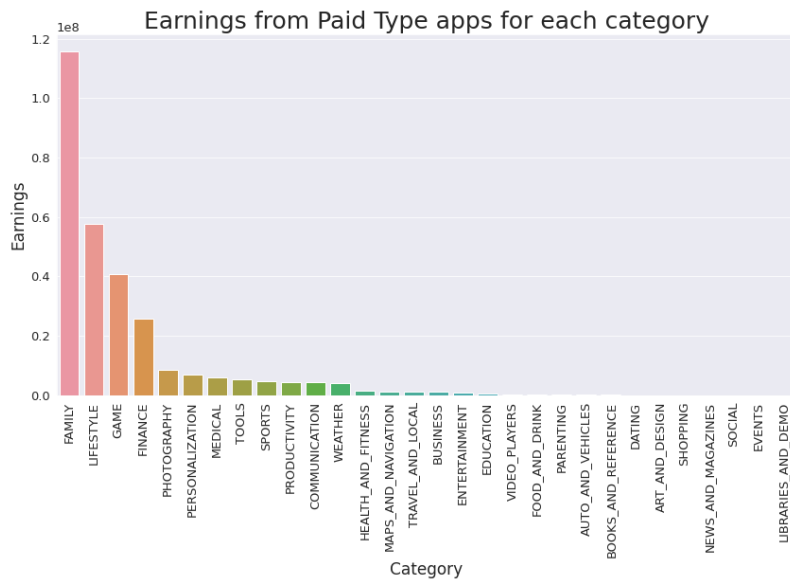
- Avg. number of installs is lower by factor of 100 for paid apps when compared to free apps.

Average Installs attributed by App Type



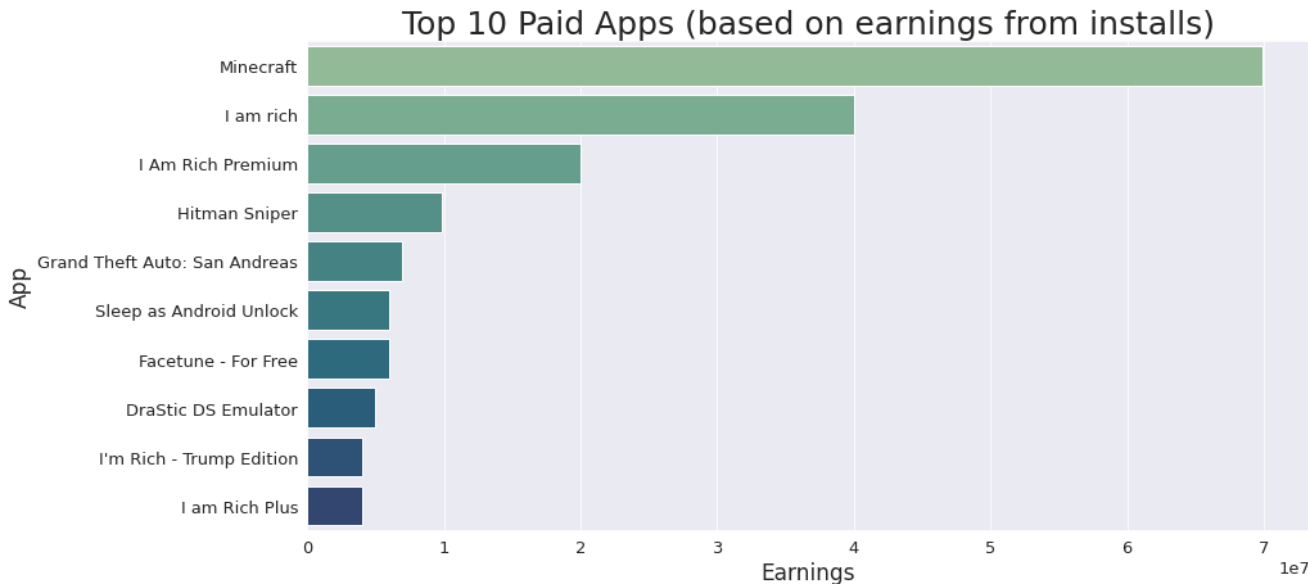
Category wise earnings from installs

- Family category leads in the total earnings, but it also has the most paid type apps.
- Based on avg. earnings, lifestyle category earns the most on average through installs.



Top 10 Paid Apps (based on earnings from installs)

- These apps earn 1 million+ dollars just through installs, that is pretty handsome earnings.
- The list is topped by Minecraft app from family category. Minecraft costs 6.99\$ and it has 10 million+ installs.



Conclusions

After analyzing the dataset I have made several conclusions which are presented below,

- Most apps are rated good on average.
- Higher the rating category higher the sentiment score and higher rated apps attracts more users.
- Using rating alone, it can't be confirmed whether the app is successful. Hence, rating is one of the supporting feature.
- Positive sentiment dominates, also most users review the app based on their experience. Hence, reviews is one of the key factor.
- Games category has average app size significantly higher than other categories.
- Although most app makers prefers to make small size apps, the bulky apps engages more users.

Conclusions (contd.)

- The avg. app size increases with decreased minimum android version support.
- To have max profitability, app makers should extend the apps support till android version 4.0.
- 7 categories namely Family, Game, Tools, Business, Medical, Personalization and Productivity contributes to 54% of total apps on the play store.
- Family is the app maker's preferred category with the most number of free and paid type apps.
- 7 categories namely Game, Communication, Tools, Family, Productivity, Social, Video players and Photography contributes to 77% of total number of installations.
- Game has the highest total installs while communication has the highest average installs.
- Apps from 7 categories namely Game, Communication, Photography, Social and Video players were able to engage users the most compared to other categories.

Conclusions (contd.)

- Although, all categories have positive sentiment score, there appears to be a lot of scope for improvement in categories like Social, Video Players, Maps and Navigation and comics according to users.
- Health and Fitness category has the highest average as well as total sentiment score.
- Travel and local, Health and fitness, Food and drink, Education and finally Dating may be the part of next popular categories.
- Free type apps dominates the play store and avg. number of installs is lower by factor of 100 for paid apps when compared to free apps (1 lakh for paid while 1 crore for free apps).
- Minecraft app from family category is the top paid app with earnings of 70 million+ through installs alone.

Thank you!