**Understanding the similarities and differences in the key characteristics of popular apps between the US and Pakistani Google Playstore**

Hamza H. Mooraj

Love For Data

Abstract

This analysis delves into a comprehensive dataset of mobile applications from the Google Play Store, focusing on apps available in both the United States and Pakistan. The study aims to understand and compare key characteristics that contribute to app popularity within these regions. Factors such as app ratings, installations, user engagement metrics are used to gauge the popularity of an app; which would be influenced by in-app purchases, content ratings, genres and many other characteristics and features of the app. Through analysis, the study sheds light on the nuanced differences and similarities between the two markets, offering insights into user preferences, trends, and potential strategies for app developers seeking to optimize their offerings. The findings underscore the dynamic interplay between app attributes and user adoption.

Keywords: *United States, Pakistan, Characteristics, App Popularity, Trends, Compare*

Contents

[Abstract 2](#_Toc144379981)

[App popularity in the Google Playstore 5](#_Toc144379982)

[Popularity 5](#_Toc144379983)

[The Formula 6](#_Toc144379984)

[Understanding the characteristics of popular in the US Google Playstore 8](#_Toc144379985)

[Feature Importance 8](#_Toc144379986)

[Genre 10](#_Toc144379987)

[Content Rating 11](#_Toc144379988)

[Understanding the characteristics of popular in the Pakistani Google Playstore 13](#_Toc144379989)

[Feature Importance 13](#_Toc144379990)

[Genre 14](#_Toc144379991)

[Offers In-App Purchases 15](#_Toc144379992)

[Content Rating 16](#_Toc144379993)

[Similarities and Differences between popular apps in Pakistan and US 18](#_Toc144379994)

[Genre 18](#_Toc144379995)

[Offers In-App Purchases 19](#_Toc144379996)

[Footnotes 21](#_Toc144379997)

[Tables 22](#_Toc144379998)

[Figures title: 23](#_Toc144379999)

# App popularity in the Google Playstore

Before I dive into the features and characteristics of an app that make it popular, I first need to formulate the popularity of an app. How can we quantify popularity? What must we consider in order to do so?

## Popularity

With popularity being such a subjective concept, calculating the popularity of an app, and thus differentiating between popular and not-so popular apps was a task that required logic as well as intuition. First I will provide the formula used, and then I will break it down and explain it. The formula is as below:

Where,

## The Formula

The Formula given above is broken into three parts. , , and finally, .

#### Function f

Function takes an argument , which is the number of installations. Since the value of can range from 0 to hundreds of millions of installations, I applied a logarithmic function (with a base of 10) to so that I can have a more standardised value for the formula.

##### **Function g**

Function takes the arguments and . The trend noticed between these two values is that although an app may have hundreds of millions of installations, only 1% to 3% of those users who installed the app actually reviewed and scored it. This is the user engagement factor, that would help determine the popularity of an app. Once again, since can be a very large number, I standardise it using a logarithmic function of base 10. As calculates the ratio of , I also had to apply the logarithmic equation to .

#### Function h

Function is used to adjust the score given to an app. Firstly, I have to adjust the range of the score from a range of 1 to 5 into a range of -2 to 2 by doing . I do this so the score has a negative or positive impact on the product. I then adjust that score by the number of reviews. can be a very large value, and so I apply a base 10 logarithmic function. The Gini coefficient allows us to identify apps where the star ratings are skewed, thus deeming the app less reliable than an app with a more even distribution of star ratings.

#### Weightages

Each function has a weightage assigned to it to 1. Keep the resulting product within a range, and 2. To determine which function of the formula takes priority in influencing the popularity\_index. After tweaking the weightages through a trial and error process, the combination that made the most intuitive sense was the following split:

This implies that the number of installations considered to be most influential in deciding the popularity of an app, followed by the user engagement, and finally the adjusted score rating.

# Understanding the characteristics of popular in the US Google Playstore

To fully analyse and understand the similarities and differences between popular apps in the US and Pakistani Playstore, I need to identify the characteristics of popular apps in each Playstore first; starting with the US Playstore.

## Feature Importance

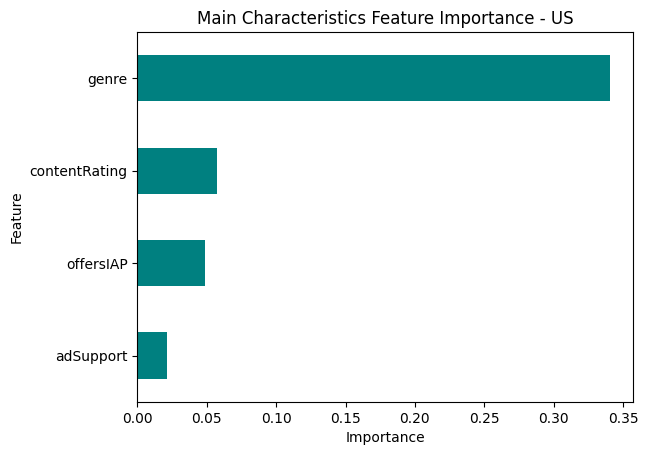
To determine key characteristics that influence the popularity of an app, it would be best to measure the importance of each feature when it comes to influencing the popularity. I did this by using a built in feature importance function in Python.

#### Random Forest Regression

Firstly I had to decide on my independent and dependant variables. I took all the features of the app (genre, content rating, offers IAP, supports advertisements, and permissions) as my independent variables, and the calculated popularity\_index as my dependant variable. Next I had to clean up the data. Using the pandas and sci-kit learn libraries in python, I encoded the categorical data ‘Genre’ and ‘Content Rating’ using One-Hot Encoding (with the OneHotEncoder class in the sci-kit learn library). I then made sure that the rest of the data was in 1s and 0s since they were all binary data. After doing the necessary encoding, I split the dataset into a training and test set, to train the model on.

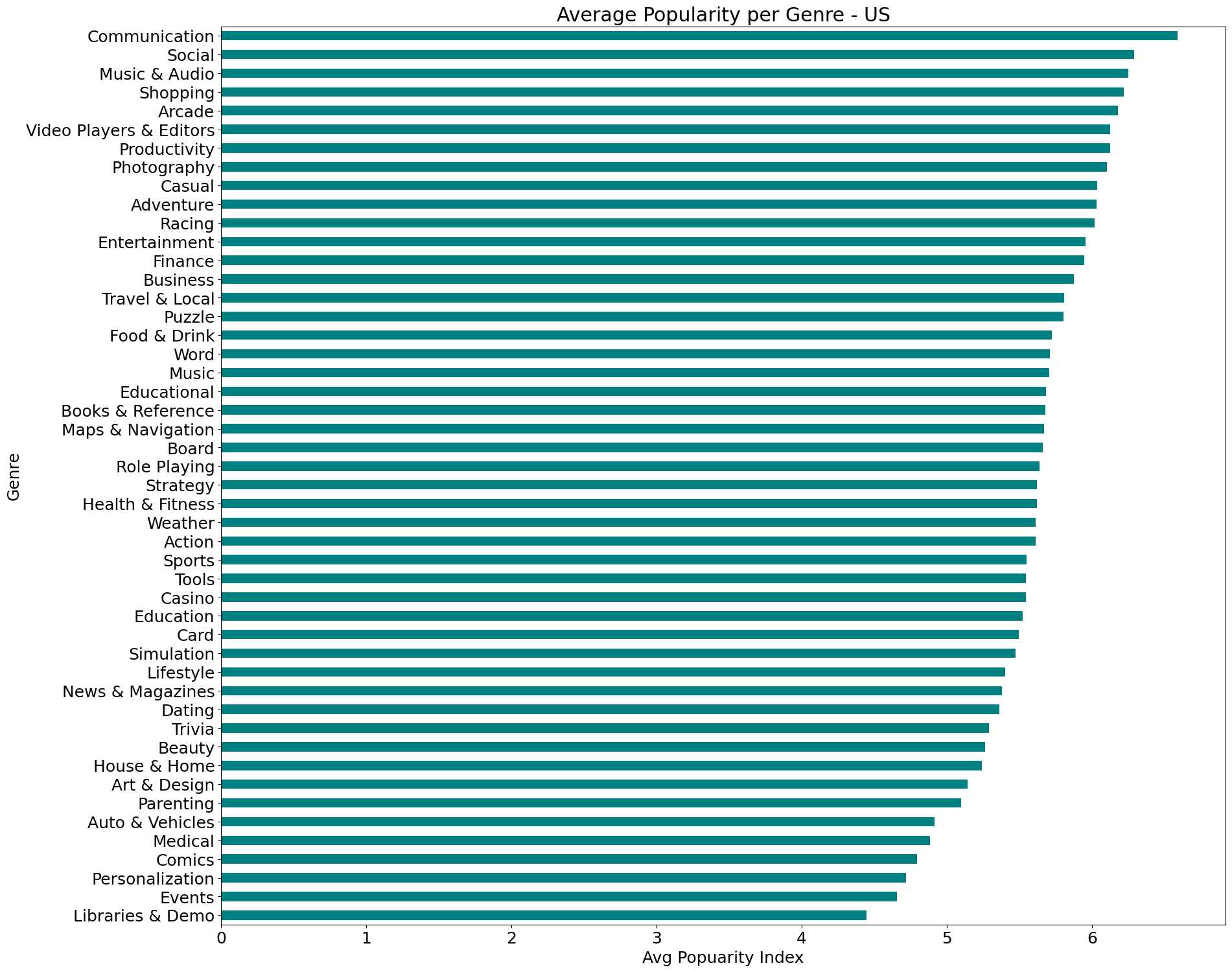
I built a Random Forst Regression model using the RandomForestRegressor class from the sci-kit learn’s ensemble library. After training the model, we can implement the built-in function, ‘feature\_importances\_’ to extract the values of importance each feature has.

#### Result

Considering all of the features of an app is not as feasible as it seems, since there are 67 features in total. I did, however, intuitively determine which features would be more influential to the popularity as well as thought of by the user. For example, the user would think of the genre of the app when installing rather than the permissions it requires (arguably included in characteristics of an app). I decided to focus on the genre, the content rating (age demographic), whether or not the app offers in app purchases, and whether or not the app supports advertisements. The following graphs out my findings:

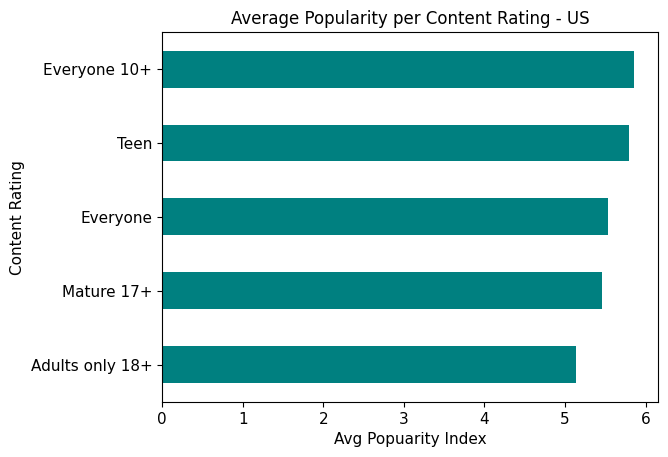
As displayed in the above diagram, the genre of the app does seem to be one of the more influential features of an apps popularity, followed by the content rating (age demographic).

## Genre

Considering the above findings (that genre is the most influential with regards to popularity), I wanted to explore the genre aspect of an app’s popularity further. This exploration is to find which genres are considered popular as well as unpopular. I grouped the dataset by the genre column and calculated the mean of the popularity\_index for each genre. The below graph plots out my findings:

The above graph shows that apps of the ‘Communication’ genre tend to be more popular than others by a significant amount. With apps like WhatsApp Messenger, Facebook Messenger, Snapchat etcetera falling under the genre of Communications, the above graph does intuitively make sense. Another observation from the above graph is that in the US, social media and utility apps have higher popularity index averages than other genre groups such as games and other miscellaneous groups.

## Content Rating

The second most influential feature from the feature importance calculations is content rating, or age demographic. While the influence of content rating on popularity may not be as great as genre, it is still worth exploring a bit further to see which age demographics provide more influence to the popularity of an app. My findings are as below:

The above graph shows that the demographic of ‘Everyone 10+’ holds the highest mean popularity. However, the difference in mean popularity between all demographics, except for ‘Adults only 18+’, is very small. The content rating ‘Adults only 18+’ has the lowest mean popularity; this could be that the higher the content rating, the smaller the audience. In spite of the previous statement, the demographic ‘Everyone’ has the largest audience, but only the third largest popularity mean; this could be because a lot of apps that may be in this rating would be directed to children specifically (educational games and apps) and would thus be less popular.

# Understanding the characteristics of popular in the Pakistani Google Playstore

Next I will analyse and try to understand the characteristics of popular apps in the Google Playstore in Pakistan.

## Feature Importance

As done with the apps in the US Playstore, it is important to determine which features are most considered when installing an app.

#### Random Forest Regression

I implemented the same algorithm to the Pakistani Playstore dataset, that I implemented on the US Playstore data to find out which features of the app most influence an app’s popularity.

#### Result

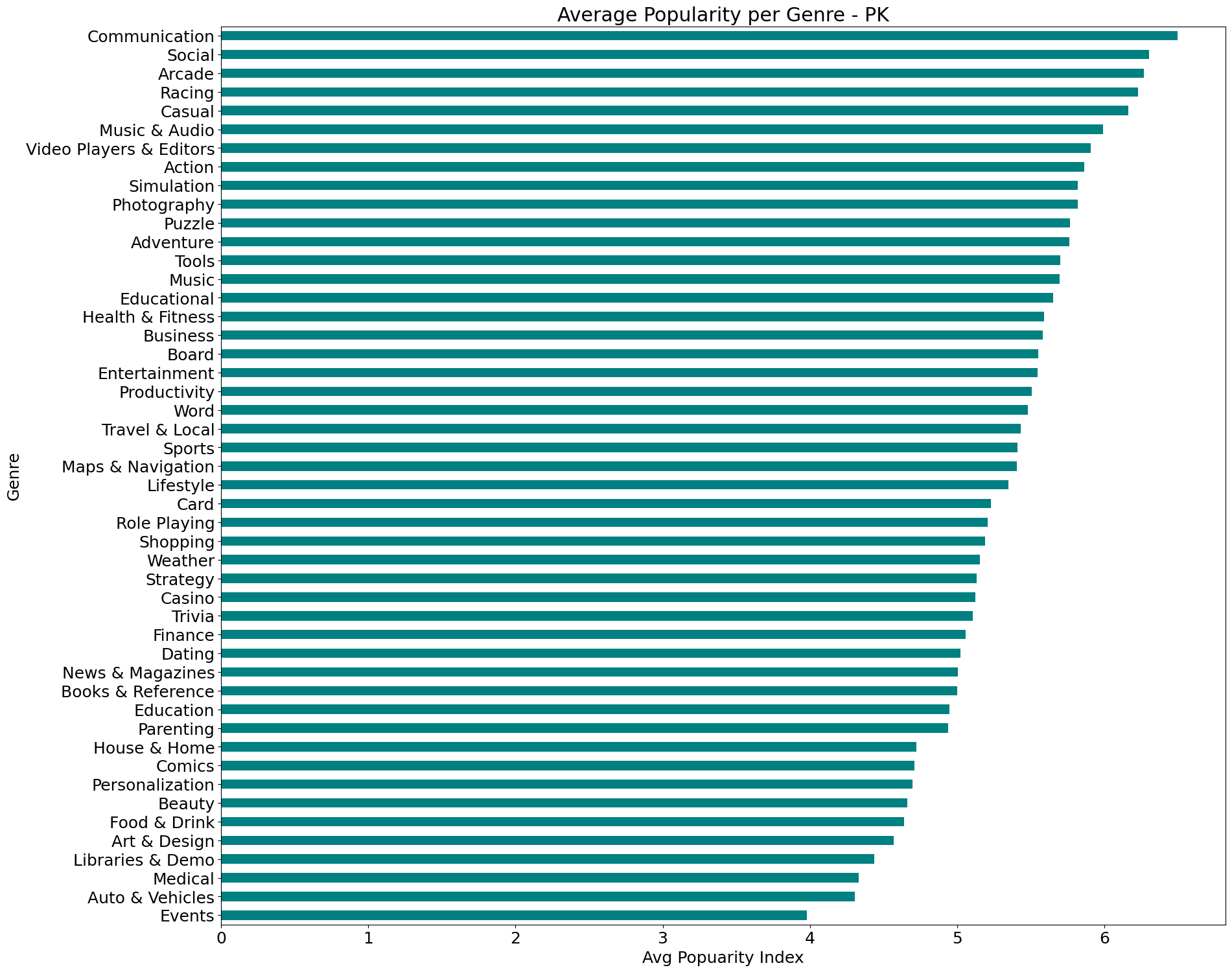
Similarly, as with the US data, it is not ideal to look at every feature since most of them would be permissions and thus not be a obvious factor in an app’s popularity for most users. The following graphs out my findings:

A bar graph with text and numbers

Description automatically generated

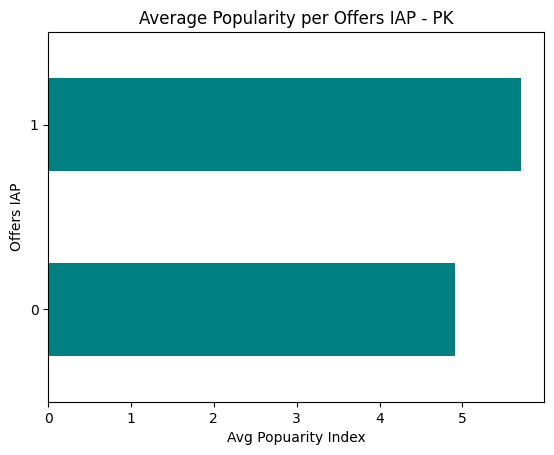
As displayed in the above diagram, the genre of the app influences the popularity the most. Genre is followed by whether or not the app offers in-app purchases and then content rating. I will explore all three features since content rating is a more obvious feature of an app than whether or not an app offers in-app purchases.

## Genre

While it is understood that many users would not decide on whether or not to install an app based on a genre, it is still important to see what it is about apps that make some users install it. I got the below graph by doing the same thing as with the US data; I grouped the dataset by the genre column and found the mean of the popularity index of all apps in one genre:

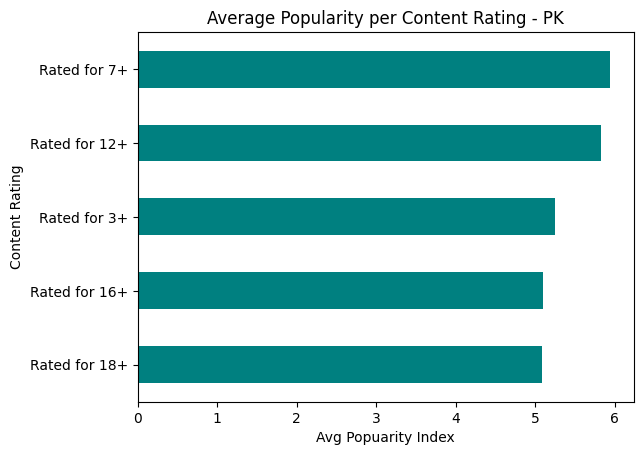
The above graph shows that social media applications are more popular than others. This can be seen with the genres ‘Communication’ and ‘Social’ having the highest mean popularity. In addition to this, gaming applications have higher popularities than utility applications (‘Medical’, ‘Shopping’, ‘Food & Drink’, etc.)

## Offers In-App Purchases

The second most influential feature from the feature importance calculations is if the app offers in-app purchases. This feature is not as obvious to users as content rating, but from analysing the genres and having gaming applications have a higher mean popularity would mean that in-app purchases would directly influence the popularity of an app (since most gaming applications do offer in-app purchases. My findings are as below:

The above graph shows that apps that do offer in-app purchases have a higher popularity mean than apps that do not. This again could be indirectly affected by gaming applications taking preference (in terms of popularity) in the Pakistani Playstore.

## Content Rating

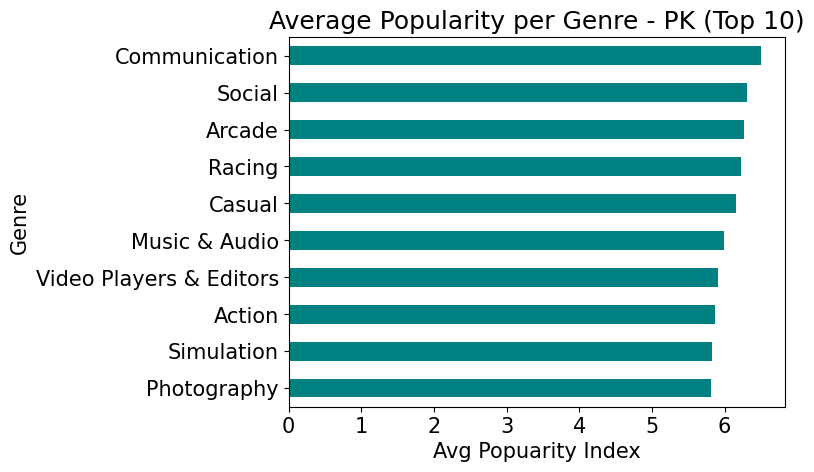
Even though content rating is not as important as whether or not an app offers in-app purchases (as per the feature importance calculations), it is still worth exploring as it is an obvious feature to the users when installing an app. My findings are as below:

The above graph shows that the demographic of ‘Rated for 7+’ holds the highest mean popularity. The difference in mean popularity between ‘Rated for 7+’ and ‘Rated for 12+’ is very small. This could be because there is not much of a difference in the age demographic that would be interested in apps with these content ratings. The higher content ratings, such as ‘Rated for 16+’ and ‘Rated for 18+’ has the lowest mean popularity of the group; this could be that the audience is much smaller than other content rating’s audience. The content rating of ‘Rated for 3+’ is the lowest of the group, and therefore would have the largest audience. However, it does not hold the highest popularity, as most apps in this content rating would be directed to young children, most of which would probably not have access to their own personal device (as Pakistani culture would stereotype), and therefore would not be as popular as other apps in higher content ratings.

# Similarities and Differences between popular apps in Pakistan and US

After analysing the popularities of apps in both Pakistan and US Google Playstores, it is now possible to compare both data’s analytics and understand what is the difference between the audiences in the US and Pakistan for mobile applications. This analysis will be split up into app features and then will be summarised at the end.

## Genre

To understand the similarities and differences between genre popularity in Pakistan and the US, I will look at the top 10 genres (based on popularity) from both regions:

A bar chart with text

Description automatically generated

From the above graphs we can see that in both regions, the genres ‘Communication’ and ‘Social’ hold the highest mean popularity. However, down the list of the top ten genres, the genres differ between the two genres. In Pakistan the top genres (after ‘Communication’ and ‘Social’) are ‘Arcade’, ‘Racing’ and ‘Casual’; all three are gaming applications. On the other hand, in the US, ‘Music & Audio’, ‘Shopping’ and ‘Arcade have higher mean popularities. The trend noticed is that in the US, utility applications are more popular than gaming applications, whereas the inverse is true in Pakistan. This trend is consistent with the lower places in the above graphs.

## Offers In-App Purchases

The trend discovered in the above exploration of the Genre category between the two regions also explains (as stated above) the importance of in-app purchases in app popularity in Pakistan vs the US. Since gaming applications are more popular in Pakistan, in-app purchases are affecting the popularity of an app, whereas in the US, the offer of in-app purchases does not have such an affect on app popularity in the US

## Content Rating

A bar chart with numbers and a number of people

Description automatically generated with medium confidenceA bar chart with text and numbers

Description automatically generatedLooking at the popularity of content ratings in both regions:

The content ratings of both regions can be considered the same. The way content is rated in both regions is different (thus the different titles), but when compared with each other, ‘Everyone’ is in line with ‘Rated for 3+’, ‘Everyone 10+’ would come under the same demographic as ‘Rated for 7+’, ‘Teen’ is the same as ‘Rated for 12+’, despite the one year difference, ‘Mature 17+’ is equal to ‘Rated for 16+’, and finally the highest content ratings for both regions both consist of the demographic of people 18+. With the above comparison, the difference in popularity of apps in the stated content ratings are the same. The reasons for this may be similar as well (as discussed above).

## Summary of Comparison

When we look at the top apps in both regions, the top three apps in Pakistan are ‘Google Maps’, ‘Google Photos’, and ‘YouTube’ (in that order). The top apps in the US are ‘YouTube’, ‘Google Photos’, ‘WhatsApp Messenger’ (in that order). Two out of the three apps are the same. This, however, does not tell the whole story as it is noticed above that gaming applications are much more popular in Pakistan than they are in the US. It can also be seen that utility apps in the US are more popular than in Pakistan. In fact, utility applications have some of the lowest mean popularities in Pakistan, with ‘Shopping’ and ‘Food & Drink’ hold 28th and 43rd positions in the genre popularity graph (respectively). Whereas in the US, ‘Shopping’ is in the 4th position, and ‘Food & Drink’ is in the 17th position. The feature of an app offering an in-app purchase is not as relevant in the US as it is in Pakistan due to gaming applications taking lower positions in mean popularity in the US. Finally, the popularity of content ratings in both regions are in line with each other and there are no differences.

# Conclusion

The purpose of this report is to discuss and analyse the differences and similarities between the audience in Pakistan and the audience in the US for mobile applications. After analysing and discussing the data for the US and data for Pakistan, there are some clear differences and similarities between the audience’s preferences in both regions. A formula for an application’s success could not be derived from this report, but what can be derived from this report is what audiences in each region looks for in an application. Gaming applications tend to do better than utility applications in Pakistan whereas the inverse is true for the US. Apps that have a content rating of at least 10+ have a larger audience, and can contain more genres than other content ratings. The top apps in each region are similar, but more differences can be noticed going down the list of apps (based on the popularity index).