

Analytics for mobile app store

Using Power BI



November 11, 2018

Neha Shashidhara Guli

California State University Los Angeles

**Table of Contents**

**Dataset URL** ------------------------------------------------------------------------------------------- **3**

**Web URL ----------------------------------------------------------------------------------------------- 3**

**Data Description -------------------------------------------------------------------------------------- 3**

**Physical properties of the data** **--------------------------------------------------------------------- 4**

**Data Cleaning ------------------------------------------------------------------------------------------ 5**

**Relationship between the tables--------------------------------------------------------------------- 9**

**Measure / Formula ------------------------------------------------------------------------------------ 10**

**Visuals---------------------------------------------------------------------------------------------------- 11**

**Dashboard---------------------------------------------------------------------------------------------- 17**

**Storytelling--------------------------------------------------------------------------------------------- 17**

**Purpose Map------------------------------------------------------------------------------------------- 20**

**References---------------------------------------------------------------------------------------------- 22**

# **Dataset URL**

<https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/home>

# **Web Link**

<https://en.wikipedia.org/wiki/App_Store_(iOS)#Number_of_iOS_applications>

# **Data Description**

Development of mobile apps in the recent has been very dynamic. We see that every other day new apps are being released to cater to our day to day requirements. The percentage of mobile over desktop is only increasing. With these millions of apps nowadays, the following dataset has become very key to getting top trending apps in the IOS app store.

This dataset contains 7200 apple IOS mobile application details. The data was extracted from the iTunes Search API at the Apple Inc website.

Mobile app analytics is a great way to understand the existing strategy to drive growth and retention of future user.

The App store continues its rapid growth, with lots of apps being added each year. To get more people to download your app, you need to make sure they can easily find your app. Also, mobile app analytics help the application developers understand the market needs based on user ratings and reviews. “For any app developer the ratings their users provide via the store are of critical importance. They are the primary mechanism for receiving user feedback, and an app's rating, whether good or bad, will play a critical part in determining whether a user will download an app when it appears in search listings”. (developer)

App analytics is a major driving force in the mobile industry. App analytics can help monitor crashes, improve usability, increase user retention and will optimize mobile marketing and advertising. “Other than provide visual representation of data, analytics could do little to help with the difficult, overwhelming, and guesswork-laden decision-making process”. (analytics)

“Mobile strategists need to ensure that mobile analytics relevant to their business goals are incorporated and used with each mobile app in order to maximize the value and return of each app project. Mobile analytics are an essential part of the mobile software development life cycle (SDLC) and continue to provide benefits throughout the life of a live app in production by providing data that can be used to improve user engagement, increase user retention and identify specific functions to enhance the overall user experience (UX)”. (strategists)

# **Physical properties of Data**

The Dataset consists of two excel files. The following are the description of the columns in the file

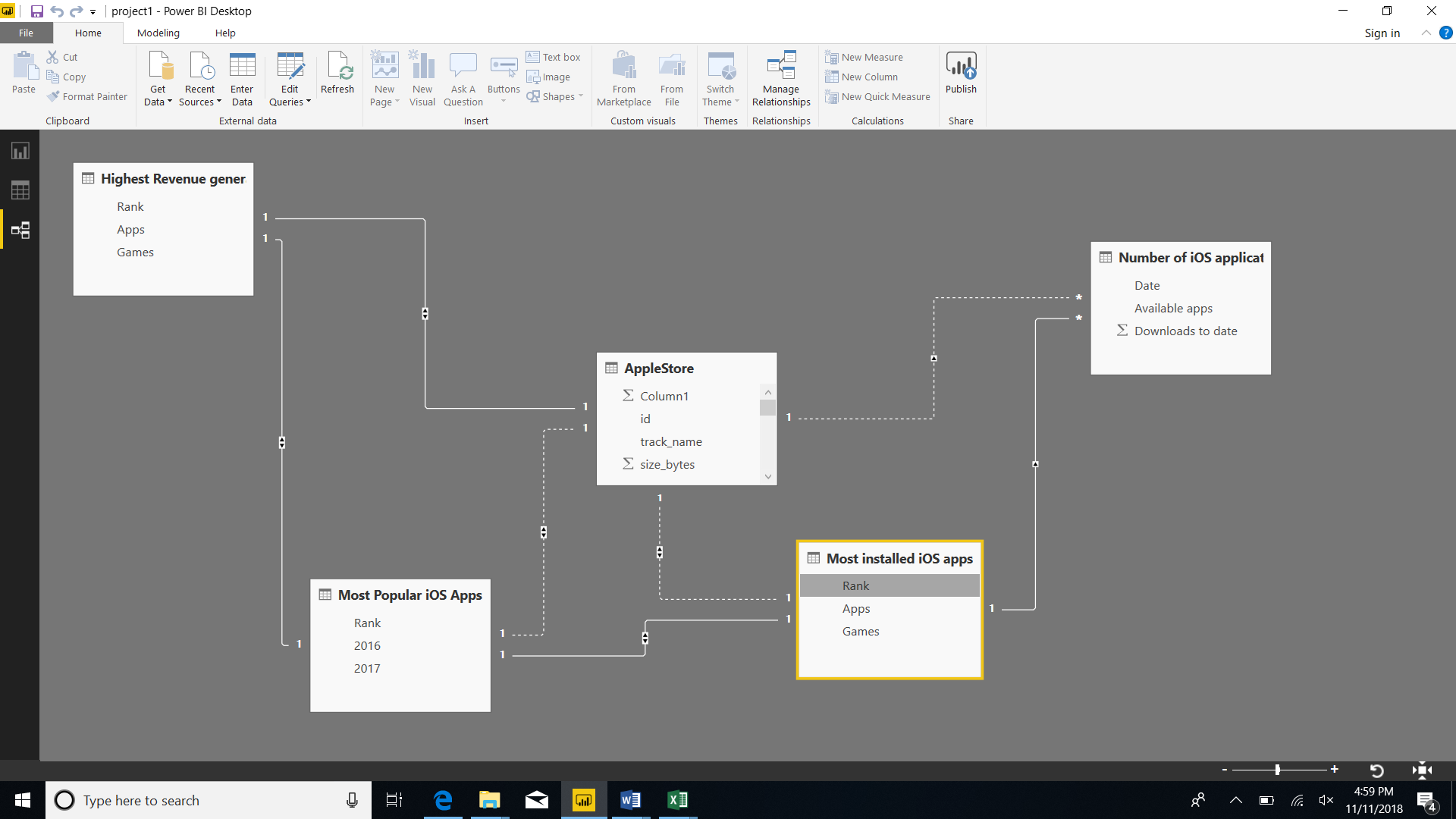
appleStore.csv

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| id | track\_name | size\_bytes | currency | price | rating\_count\_tot | rating\_count\_ver | user\_rating | user\_rating\_ver | ver | cont\_rating | Prime\_genre | sup\_devices.num | ipadSc\_urls.num | lang.num | vpp\_lic |
| App ID | App Name | Size in bytes | Currency type | amount | User rating counts (for all version) | User rating counts (for current version) | User rating value (for all version) | User rating value (for current version) | Latest version code | Content rating | Primary genre | Number of supporting devices | Number of screenshots showed for display | Number of supported languages | Vpp Device Based Licensing Enabled |

# **Data Cleaning**

|  |  |
| --- | --- |
| **Problem** | **Un-cleaned and Cleaned Data** |
| Missing values | This includes empty cells assumed as being of no value (zero/nothing) or no measurement (n/a, null). In my data set there was no missing data |
| Leading and trailing | There was no leading and trailing data |
| Erroneous values | Removing the first column as it’s of no value to the data set |
| Uncommon system characters or line breaks | All the names that contains special characters were removed by filtering and using remove rows function |
| Inconsistencies | The available apps and number of downloads was in decimal data type which was changed to whole numbers |
| Duplicate records | No repeated rows in any of the tables |
| Out of date | They are the values that might have expired in accuracy, like someone’s age or any stastic that would be reasonably expected to have subsequently changed  No out of date data |
| Date Issues around format | No date format issues |

# **Relationship between the tables**

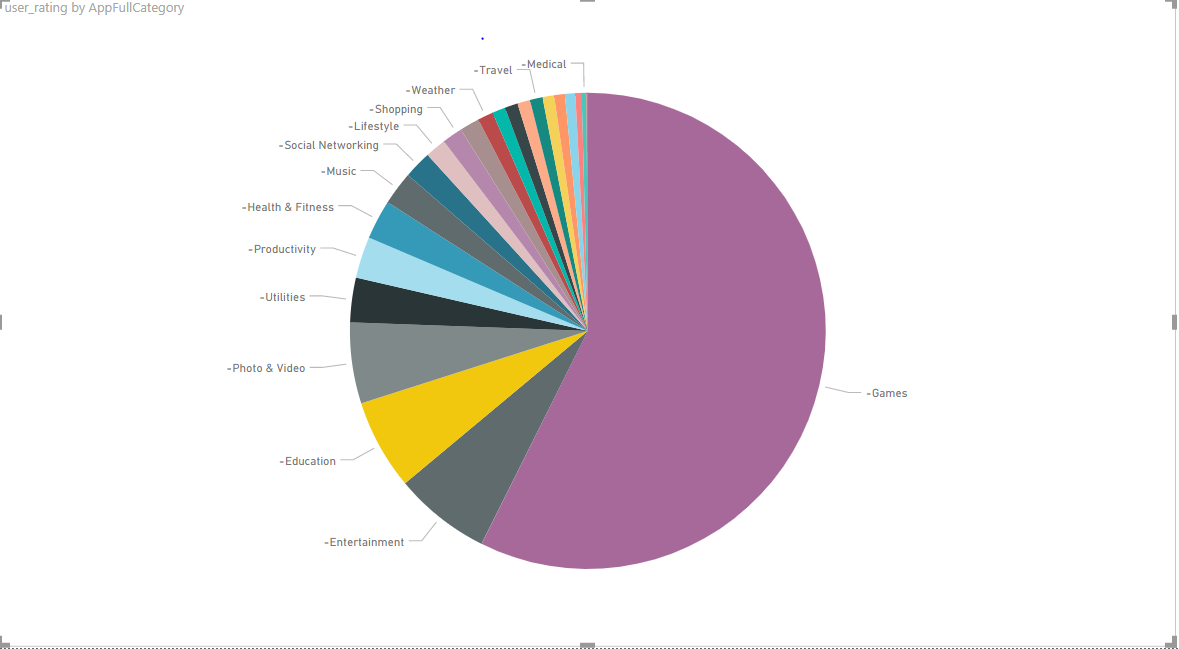


Relationships between the tables are necessary in-order to accurately calculate results and display the correct information in the reports. Power BI Desktop makes creating those relationships easy. I have combined 4 web datasets and 1 csv dataset for my visualization. Id is the primary key for the App store dataset set and Rank is the foreign key.

# **New Column / Formula**

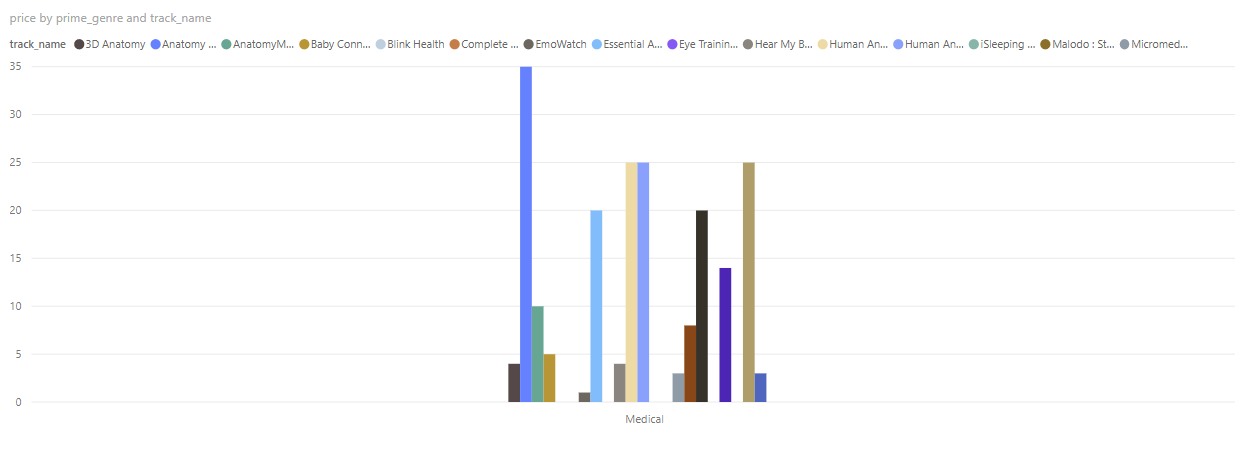
Formula used: AppFullCategory = RELATED('Most Popular iOS Apps'[Rank])&"-"&[prime\_genre]

For this visualization, I have used Column formula to create a new column which has the rank of all the popular apps merged with the primary genre of the Apps. Average user rating value is noted for all the categories in the AppFullCategory column. Pie chart can be used to show the different Genre and the corresponding user rating value for that particular genre. A pie chart therefore helps to determine the relationship between two columns of the dataset, with one data set being dependent on the other set. As per the chart, we induce that games category apps have the highest user rating with 57.38% compared to other categories and catalogs category has the lowest with 0.09% of the total.



# **Visuals**

1. Price distribution for all the apps that comes under the primary Genre – Medical

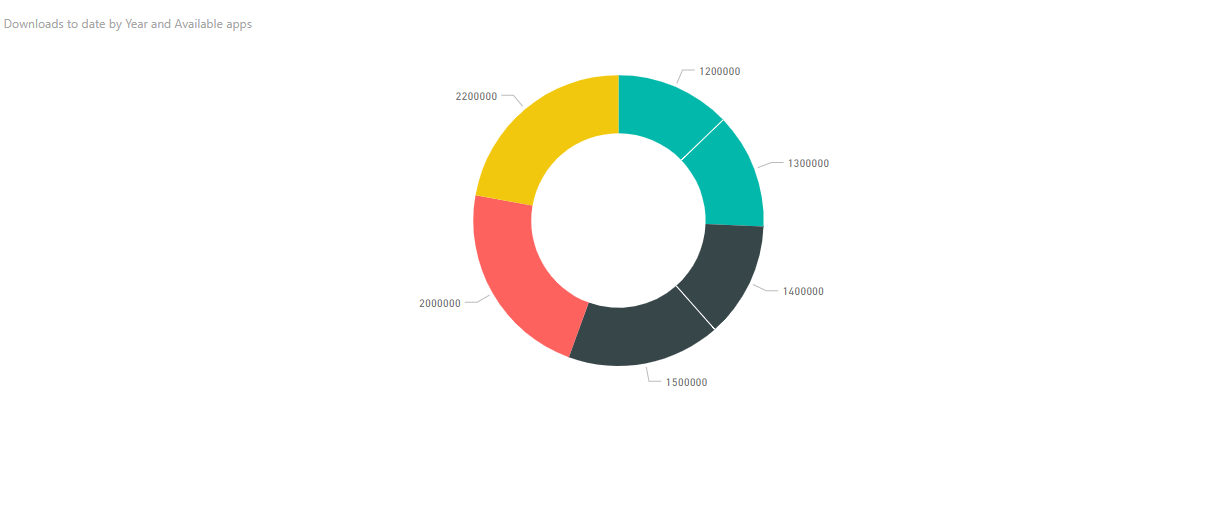


The Line and clustered chart here is used to know the price distribution of different apps that comes under medical category

From this visualization we can infer that the App named Anatomy & Physiology : Body structures and function is the most expensive App with the price of $34.99.

EmoWatch App is the cheapest App with the price of $0.99.

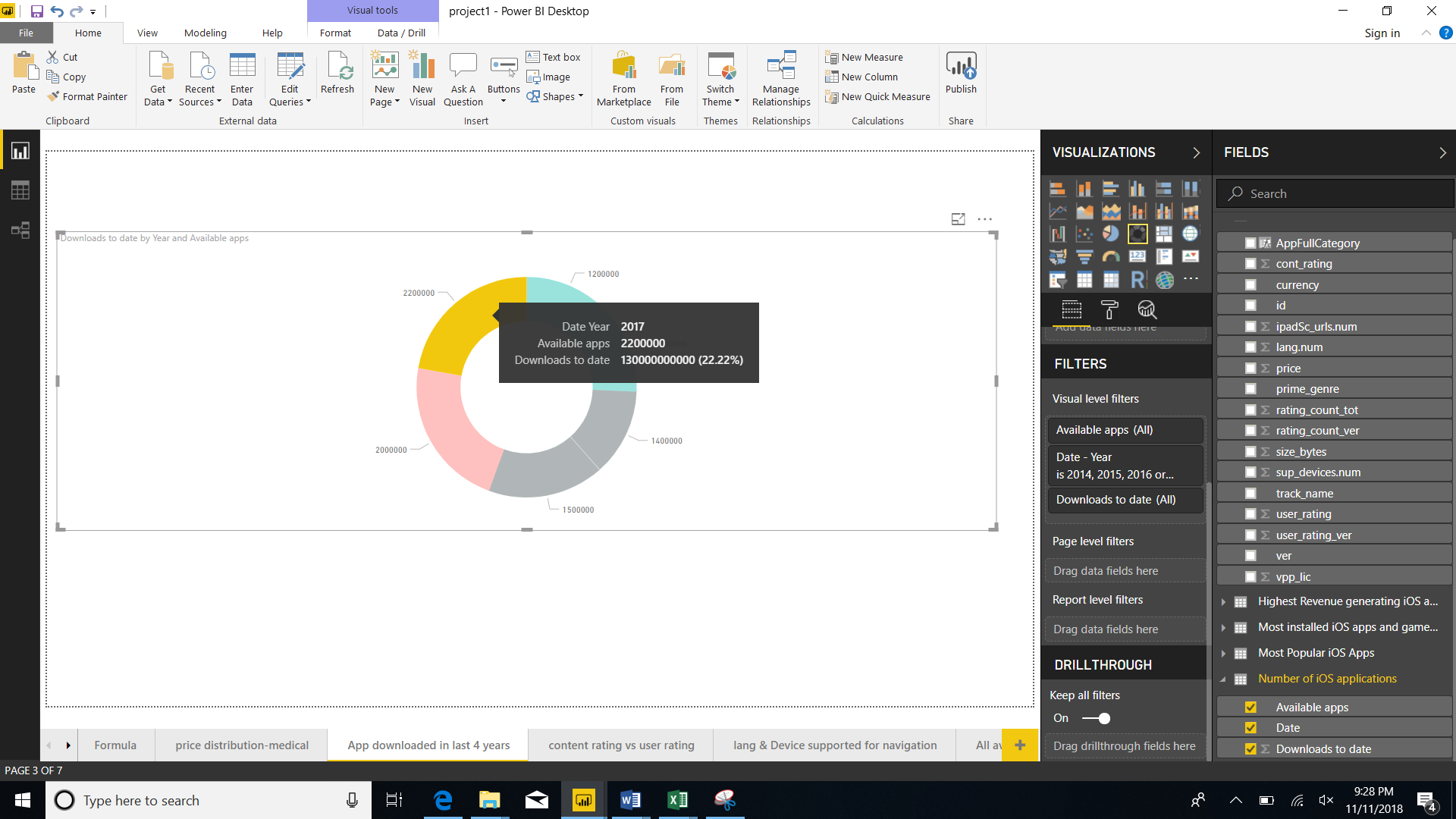
2. Total Available Apps in the past 4 years and how many downloads have happened for those available apps (2014, 2015, 2016, 2017)



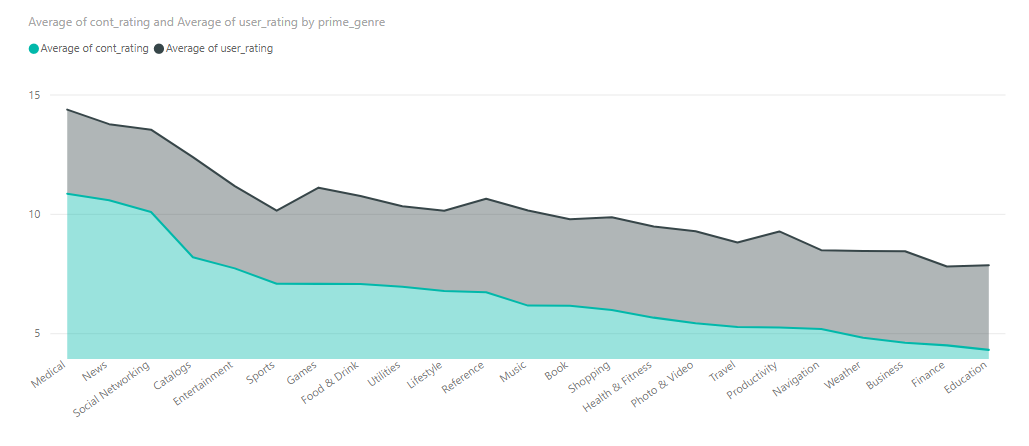
Donut Chart here is used to see the total number of available apps and the total number of apps that have been downloaded in the last 4 years that is 2014,2015,2016 & 2017.

The different colors display four different years.

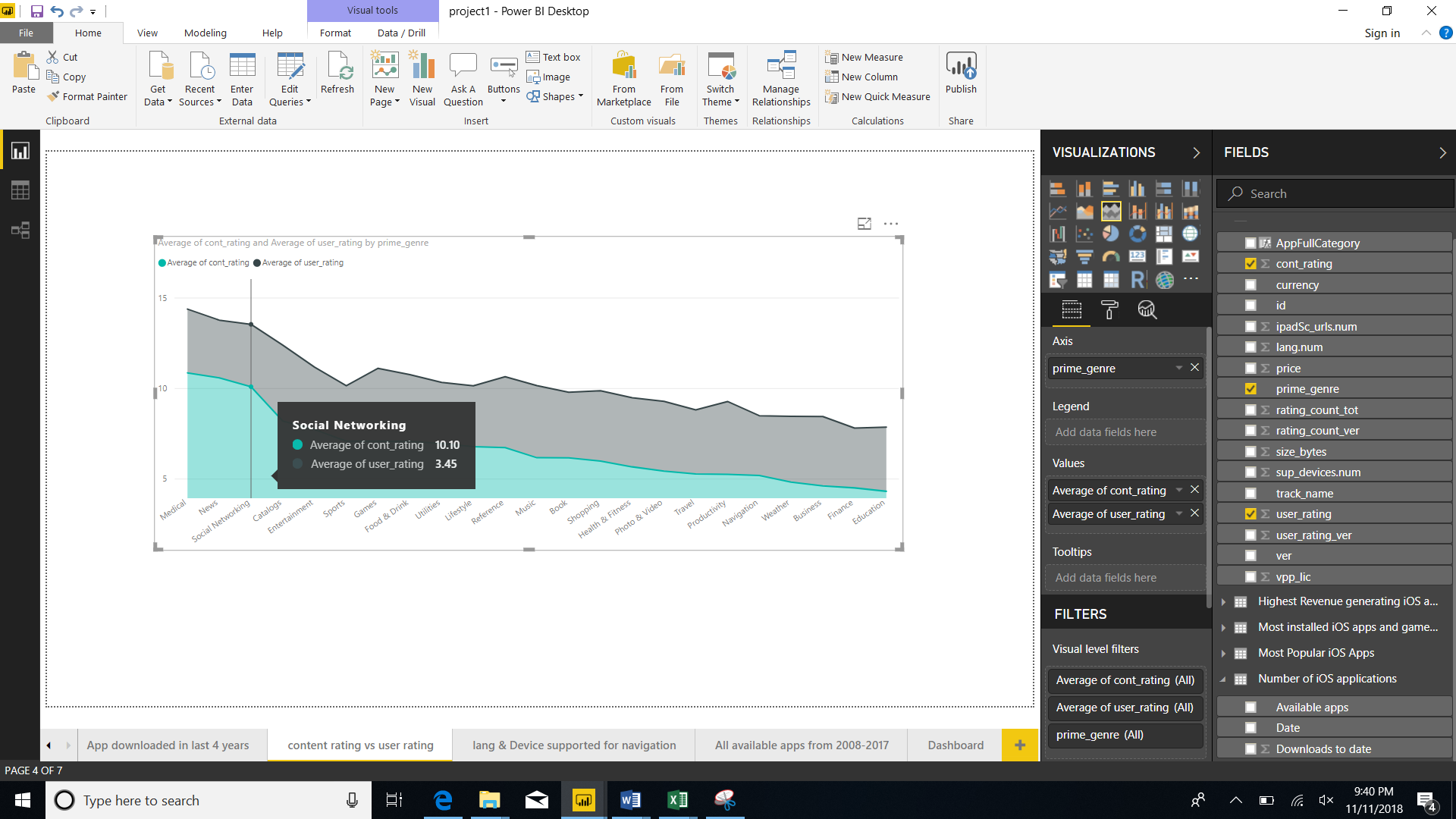
The following figure shows the total number of apps and total downloads in the year 2017



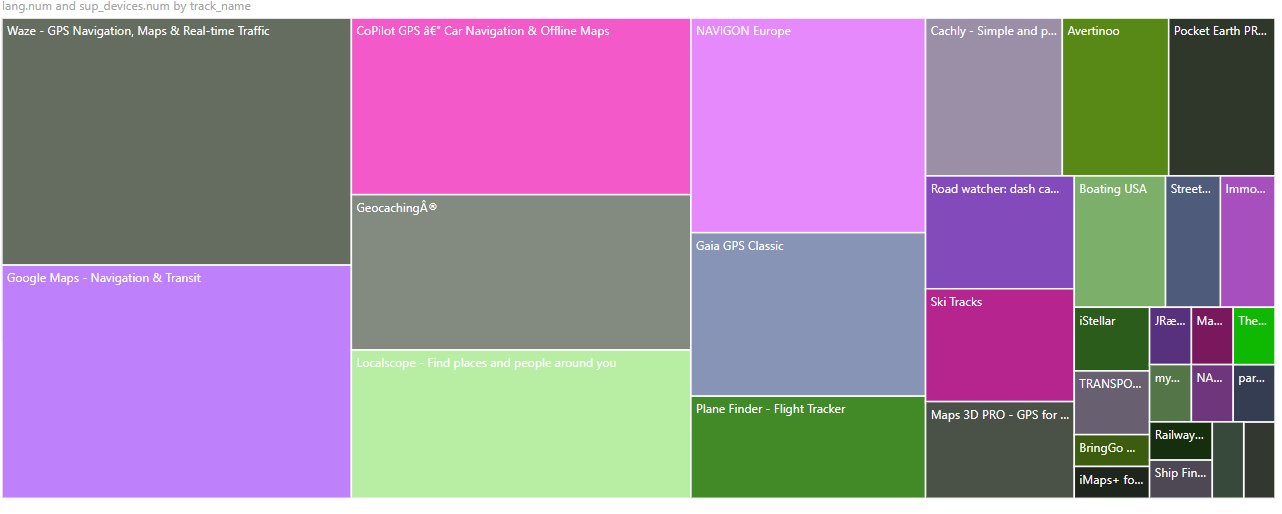
3. Content Rating Vs User Rating for different App categories



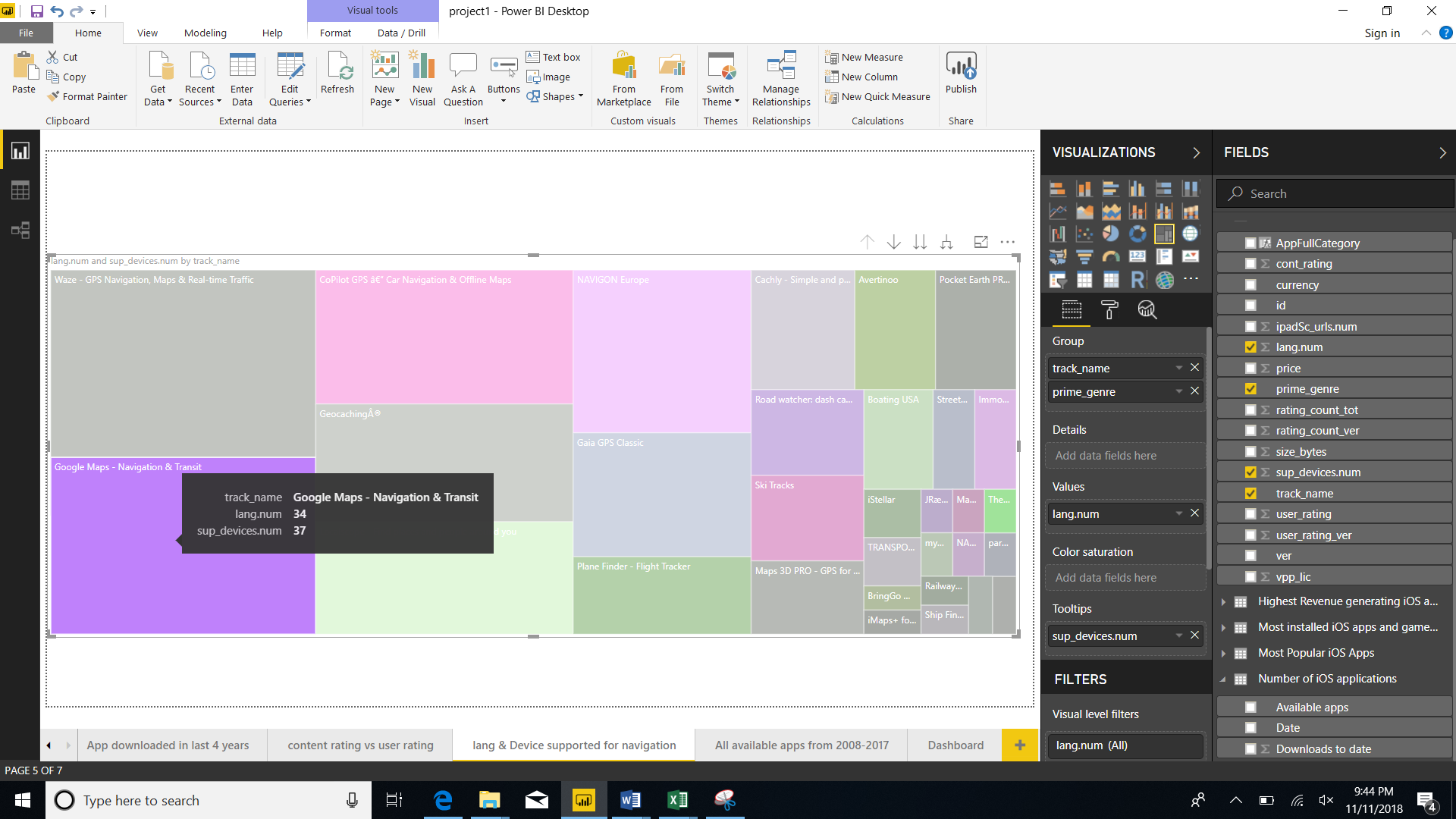
The Stacked Area Chart here displays the content rating and user rating for all the primary Genre of the application. The Average rating of each of the category is considered to display the rating for all the Genre. For example, the following figure shows the average for user and content rating that belongs to the social networking category:



4. Number of languages and devices supported for the apps that belongs to the ‘navigation’ category

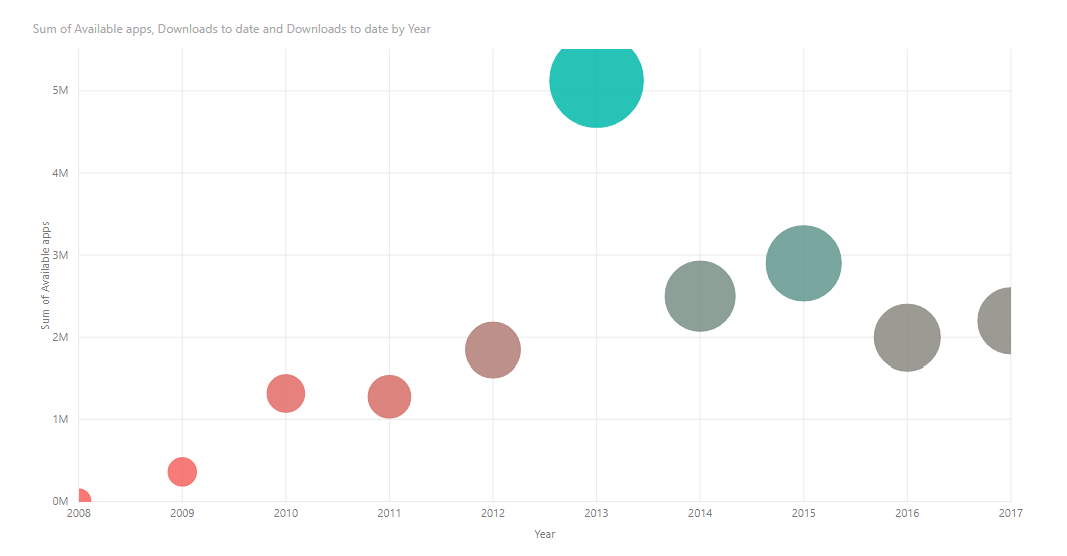


The Tree Map here shows the number of languages and devices that are supported by all the Apps that belong to the category ‘navigation’



For example, the google maps app supports 34 languages and 37 devices

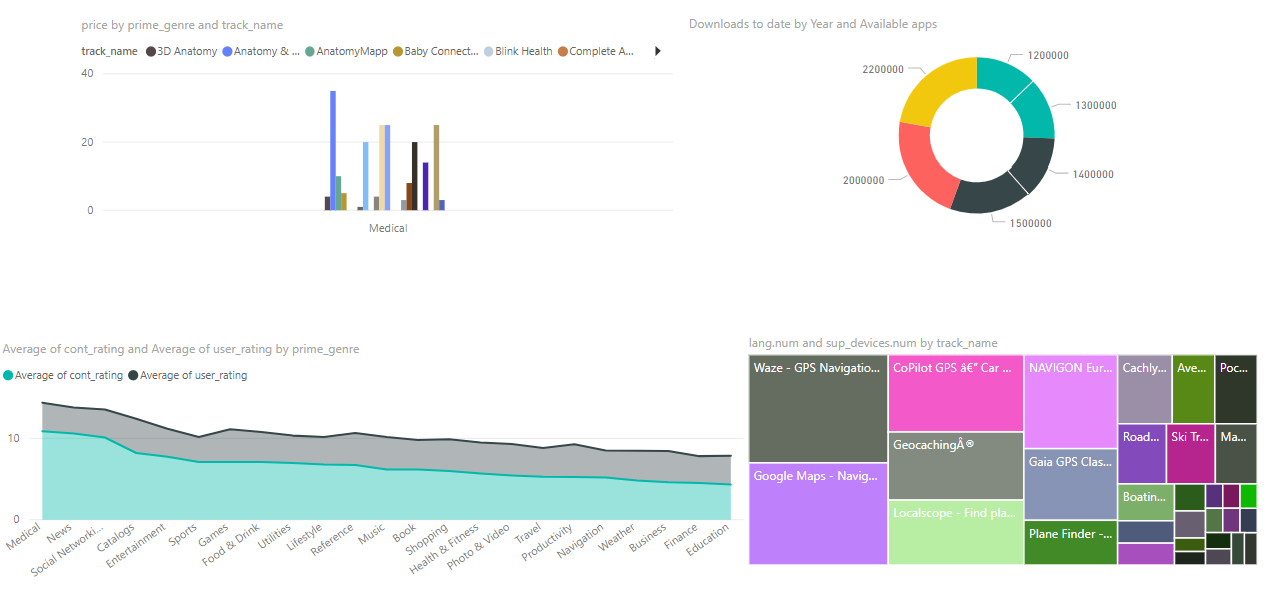
5. Summary of all the available apps from 2008-2017 that were downloaded



The scatter plot here shows all the apps that were available and all the apps that were downloaded from the year 2008 until 2017.

For example in the year 2008, total number of available apps were 4300 and total number of downloads were 110000000.

# **Dashboard**



# **StoryTelling**

Creating more engaging and interactive app is always a challenge for the app developers and often improving an app on such crucial aspects the developer team seems to be ready to tread heaven and earth. In consideration to the huge competitive challenge every app faces in app stores, it is quite natural to take such concerns seriously.

Mobile app analytics is a great way to understand the existing strategy to drive growth and retention of future user.

The App store continues its rapid growth, with lots of apps being added each year. To get more people to download your app, you need to make sure they can easily find your app. Also, mobile app analytics help the application developers understand the market needs based on user ratings and reviews. “For any app developer the ratings their users provide via the store are of critical importance. They are the primary mechanism for receiving user feedback, and an app's rating, whether good or bad, will play a critical part in determining whether a user will download an app when it appears in search listings”. (developer)

App analytics is a major driving force in the mobile industry. App analytics can help monitor crashes, improve usability, increase user retention and will optimize mobile marketing and advertising. “Other than provide visual representation of data, analytics could do little to help with the difficult, overwhelming, and guesswork-laden decision-making process”. (analytics)

“Mobile strategists need to ensure that mobile analytics relevant to their business goals are incorporated and used with each mobile app in order to maximize the value and return of each app project. Mobile analytics are an essential part of the mobile software development life cycle (SDLC) and continue to provide benefits throughout the life of a live app in production by providing data that can be used to improve user engagement, increase user retention and identify specific functions to enhance the overall user experience (UX)”. (strategists)

Here the analysis done in this project gives the insight on four main research questions.

1. Most popular Genre

Genre is simply categorized due to the same or similar conventions being used to portray a particular genre over and over again. However, this hasn’t meant that these conventions have stayed the same since apps have been developed. There are different genre that all the apps have been classified into. In my dataset there are 23 different categories in which many apps are being placed. For example, a primary genre as Food & drink has many apps such as The perfect Egg timer, DoorDash - Food Delivery etc.. based on this learning of classifying apps into different genre, a pie chart has been made to understand the user rating for different app categories.

2. Price distribution by Primary genre

The price distribution for different apps that belong to a particular genre helps in effectively analyzing which app is priced highly among 1000’s of other apps in a similar field. From this visualization we can infer that the App named Anatomy & Physiology: Body structures and function is the most expensive App with the price of $34.99. EmoWatch App is the cheapest App with the price of $0.99.

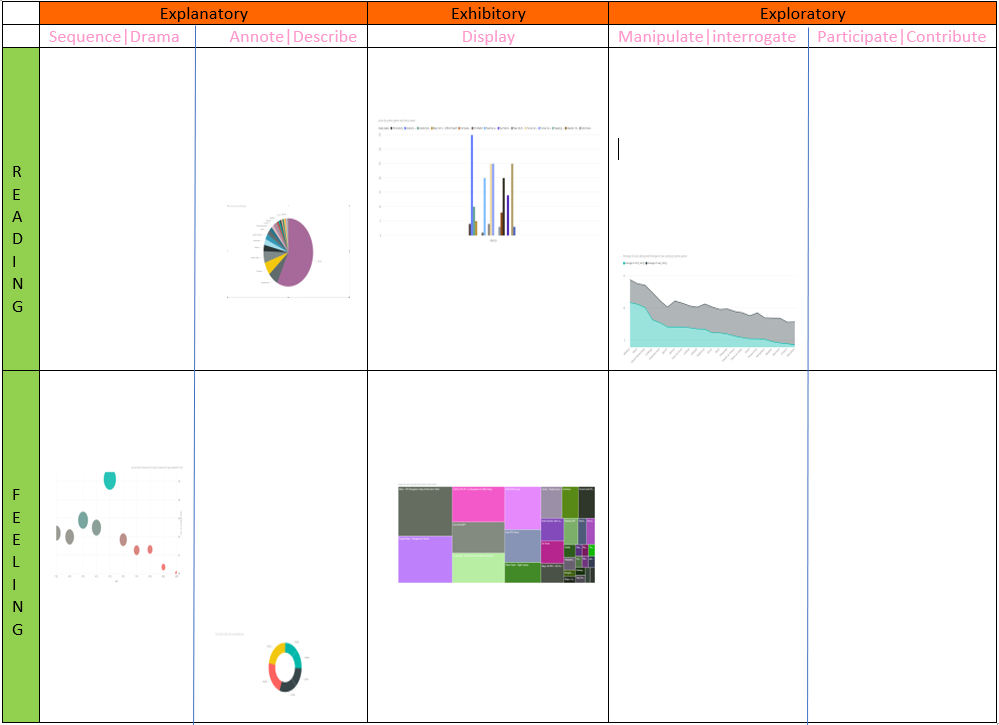
3. Content Rating and user rating based on different App categories

The average content rating and user rating helps in determining which category has been used more by the population. The average content rating and user rating helps in determining which category has been used more by the population. The Average rating of each of the category is considered to display the rating for all the Genre. For example, in the social networking genre the average content rating is 10.10 and the average user rating is 3.4.

4.Languages and devices supported by different apps for a particular category

In this visualization different apps that belongs to a particular category that is ‘navigation’ is considered. The total number of languages that are supported by a particular app say, google maps can be inferred from this kind of visualization. For example, the google maps app supports 34 languages and 37 devices

# **Purpose Map**



1. EXPLANATORY

Explanatory as visualizers provides the viewer with a visual portrayal of the subject’s data and will also take some responsibility to bring key insights to the surface, rather than leave the prospect of interpreting the meaning of the information entirely to the viewer. Pie Chart comes under explanatory because it is easily understandable by the user. From the visualization, viewers can make out which App genre is highly rated by users. Overall, it can be concluded that apps that belong to Games category has the highest user rating when compared to apps that belong to catalog category. The visualizer here is attempting to assist the viewers process of understanding as much as possible, in particular with the interpretation, drawing out the meaning of the data.

2. EXPLORATORY

Exploratory visualizations differ from explanatory visualizations in that they are focused more on helping the viewer or more specifically in this case the user find their own insights. Stacked area chart comes under exploratory because the user first needs to understand what the content rating means and what user rating mean, what is the relation between the two terms. Basically, the user needs to explore by himself about the related terms and then try to understand what the visualization is. This project is simply a visual window into the analysis of this data that lets users perceive the data values and interact with the different dimensions offered.

3. EXHIBITORY

Exhibitory are neither explanatory nor exploratory. With exhibitory visualizations the viewers have to do the work to interpret meaning, relying on their own capacity to make sense of the display of the data and the context of the subject matter. Line and clustered chart is exhibitory because the viewer must be able to understand the content of a display as well as the context of the display. Exhibitory rely entirely on and make assumptions about the capacity of and interest among the target audience. From the visualization, firstly user should know on what basis the apps are categorized and what each type mean. Once the user is able to understand all the concepts about the app category and how they are classified then it would be very easy for the viewers to interpret the visualization. This helps the user to draw their own conclusion after understanding the visual.

4. READING

At the top of a purpose map the tone of the visualization design choices is geared towards optimizing the ease with which viewers can accurately estimate the magnitude of and the relationship between the values. The reading tone is the best fit approach when the purpose of work requires to facilitate understanding with a high degree of precision and detail. In this scenario there is no need to seduce an audience through aesthetic treatment. Pie chart comes under reading because the pie chart is easily readable by the audience and also, they are familiar with the terms like rating and genre. Hence the visualization is very easy for the viewers to interpret the meaning.

5. FEELING

Feeling tone gives more emphasis on determining the gist of the big, medium and small values and a general sense of the relationships that exist. Sometimes an ‘at-a-glance’, high level view is the most suitable way to portray a subject’s value. Tree map comes under feeling tone because it cannot remotely estimate the relative proportions but can estimate the gist of the scales involved.

# **References**

1. Kumar, Rupesh. “How Visual Analytics Help Mobile Apps to Be More Engaging?” *Free Web Resources*, 5 May 2015, [www.bestfreewebresources.com/visual-analytics-mobile-apps](http://www.bestfreewebresources.com/visual-analytics-mobile-apps).

2. Ramanathan. “Mobile App Store ( 7200 Apps).” *RSNA Pneumonia Detection Challenge | Kaggle*, 10 June 2018, [www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/home](http://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/home).

3.Schonfeld, Erick (March 17, 2009). ["Apple Opens Up More Ways To Get Paid On The iPhone, Adds Key New Features. Apps Hit 800 Million Downloads"](https://techcrunch.com/2009/03/17/phone-apps-hit-800-million-downloads/). [*TechCrunch*](https://en.wikipedia.org/wiki/TechCrunch). [AOL](https://en.wikipedia.org/wiki/AOL). Retrieved March 29, 2017.

4.["Apple's App Store Downloads Top Two Billion"](https://www.apple.com/pr/library/2009/09/28Apples-App-Store-Downloads-Top-Two-Billion.html). Apple Press Info. [Apple Inc.](https://en.wikipedia.org/wiki/Apple_Inc.) September 28, 2009. Retrieved March 29, 2017.