**DSCI 5360 Final Project Proposal**

G. Brint Ryan College of Business, University of North Texas

DSCI 5360.002 - Data Visualization

Dr. Vess Johnson

**Project Group 14**:

Emeka Ugwu

Sai Akshith Garikipati

Joel Madaram

Ram Chand Perala

**Dataset**

For this final Data Visualization project, we have assembled a group consisting of Emeka Ugwu, Sai Akshith Garikipati, Joel Madaram, and Ram Chand Perala under project group 14. Our chosen dataset was derived from Kaggle.com and is titled “Google Play Store Apps” (<https://www.kaggle.com/datasets/lava18/google-play-store-apps>).

**Description**

The“Google Play Store Apps” dataset sets out to address the lack of Google Play data. While many public datasets provide Apple App Store data, there’s not much for Android due to the complexity of its web scraping. The iTunes App Store’s structure allows for easy web scraping, while its counterpart has to do a bit more to achieve that. Google Play Store employs sophisticated modern-day techniques (such as dynamic page load) with jQuery, making scraping more difficult. This data is obtained from the Google Play Store. Without it, this app information would not be available. The dataset describes the apps that are available in it in detail.

**Content**

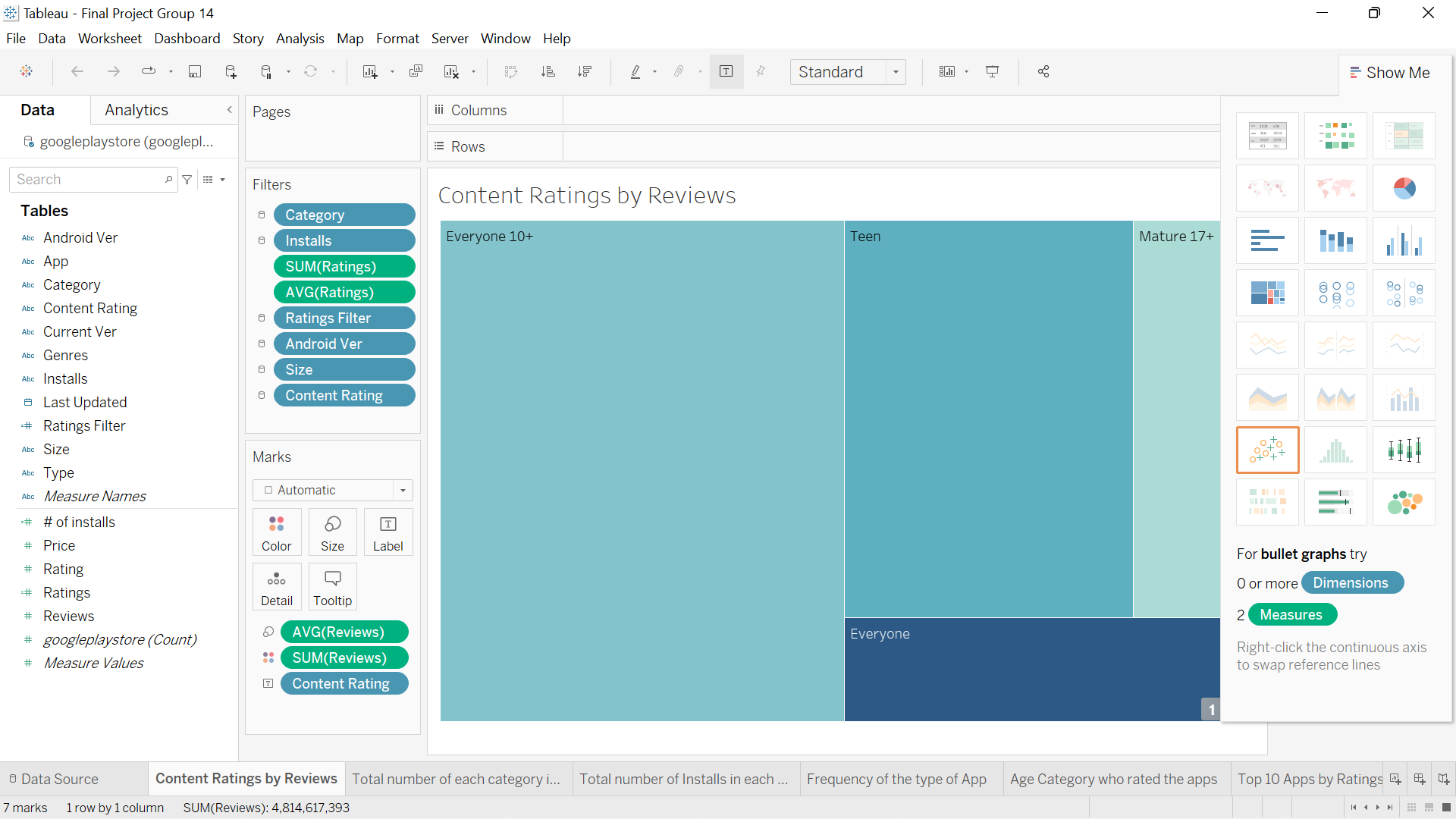
* Each app (row) has values for the following attributes: category, rating, reviews, size, installs, type, price, content, genres, last updated, current version, and Android version.
  + **Category:** Here it displays the category to which the mentioned app belongs to. There are different categories like beauty, family, business, etc.
  + **Rating:** It presents the star ratings for a total of 5.
  + **Reviews:** It enhances the reviews given by the users and gives us the total number of reviews.
  + **Size:** The app’s built size is specified in this category.
  + **Installs:** It depicts the total number of users that installed it on their devices.
  + **Type:** Here it displays whether the app is available for free or for purchase.
  + **Price:** If it is for purchase then here it displays the amount or else if it is free then it is displayed as ‘0’.
  + **Content:** The application's content is enhanced here whether it is teen based, for everyone etc.
  + **Genres:** It unveils which type of genre the app belongs to.
  + **Last updated:** It shows the app’s recent updated date.
  + **Current Version:** The current version of the application is depicted here.
  + **Android Version:** It has the device's current android version.
* We are exploring the following in the dataset:
  + Content ratings
  + Top 10 apps by ratings
  + Top 10 apps by price
  + Top 10 apps free by install

**Problem**

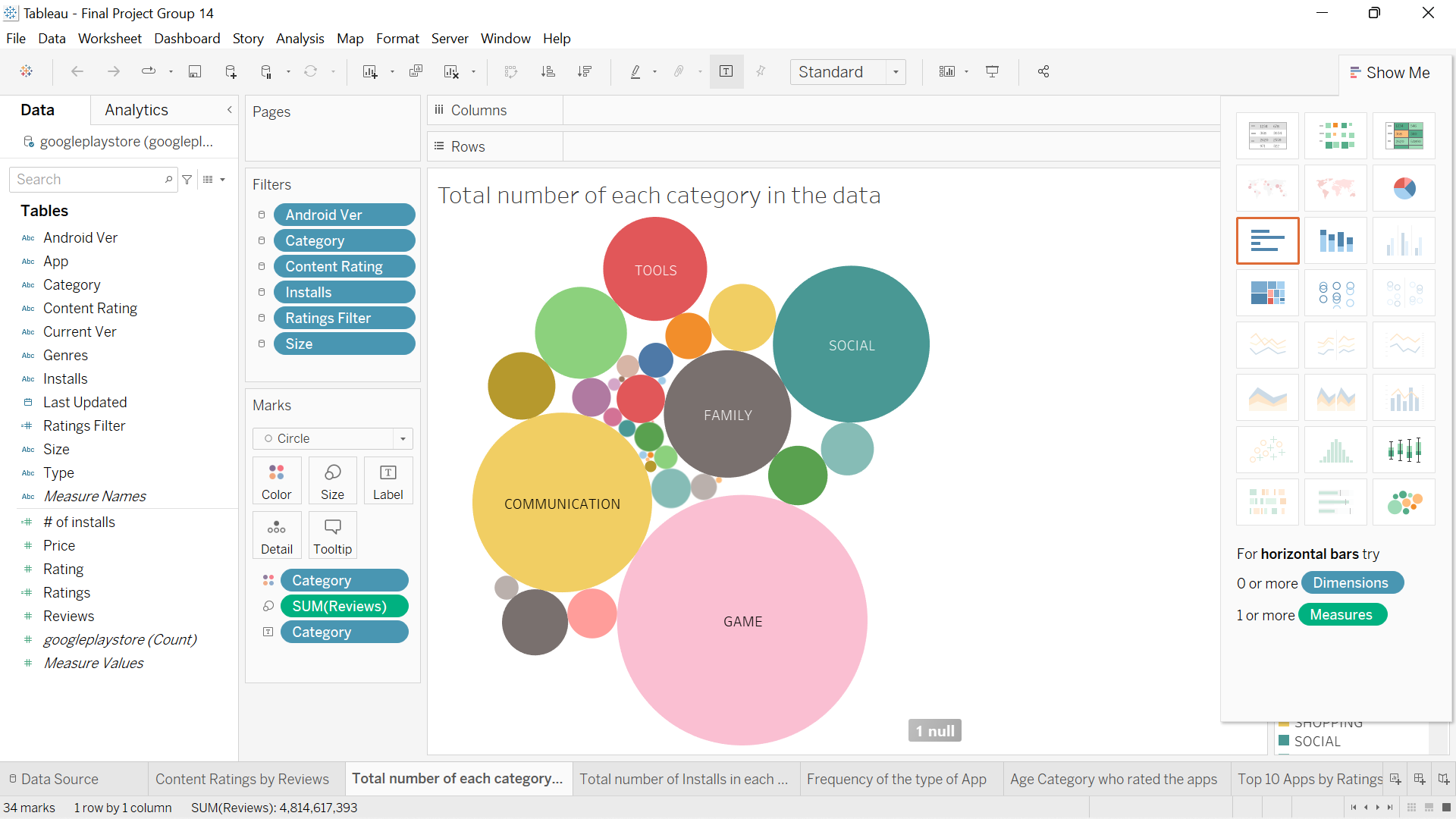
The investigation is carried out on the basis of content ratings by reviews, top apps by ratings, top apps by price, and top apps free by install. So, by comparing all of these, we can gather data for the top applications in the Google Play store. Using that data, we can determine the genres and categories that are missing in all of the apps, as well as the factors that make the apps stand out. With these, we can identify the most successful applications and develop them to achieve even greater success.

**Results**

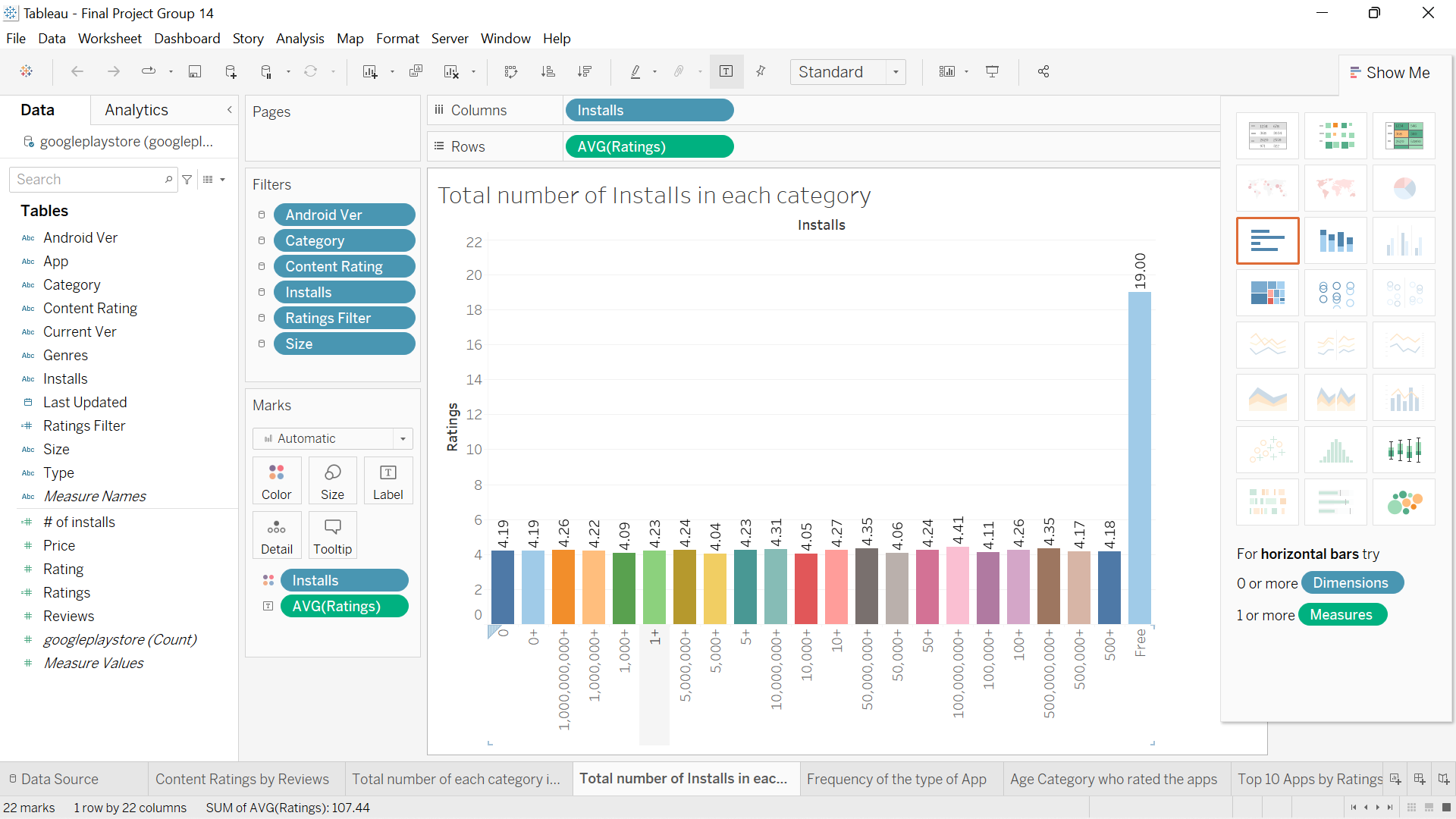
**Content Ratings by Review**

****

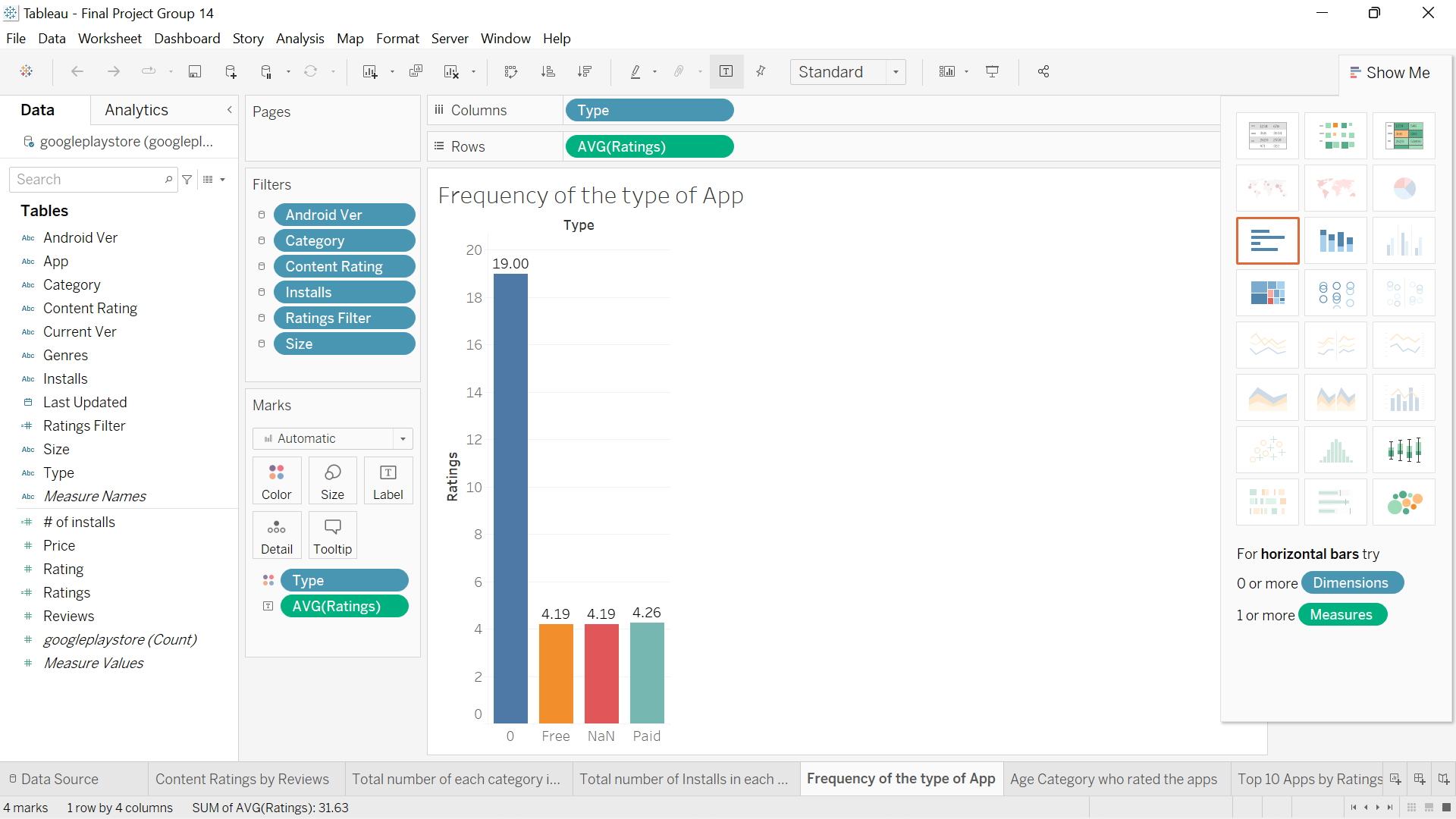
**Total number of each category in the date**

****

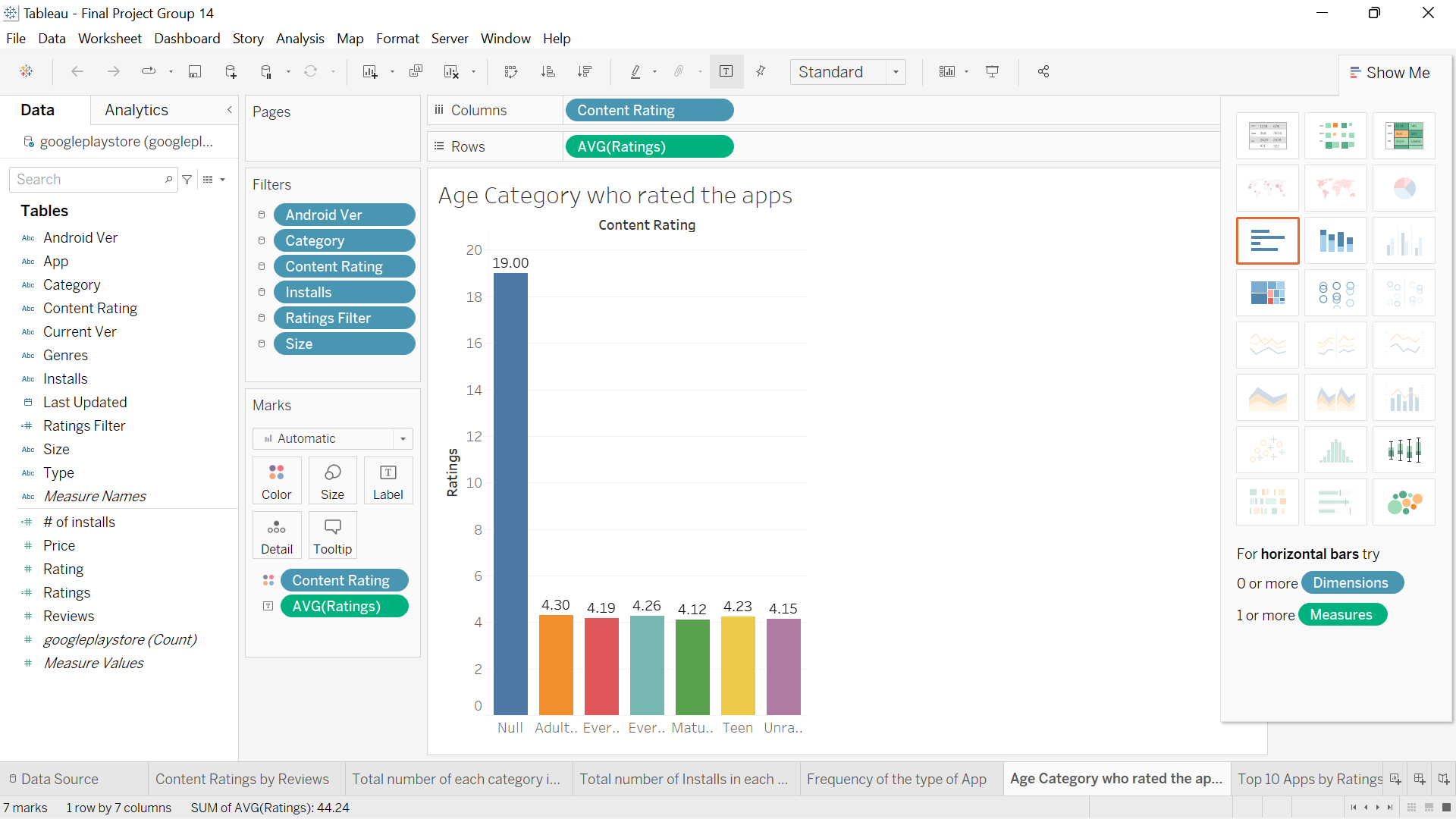
**Total number of installs in each category**

****

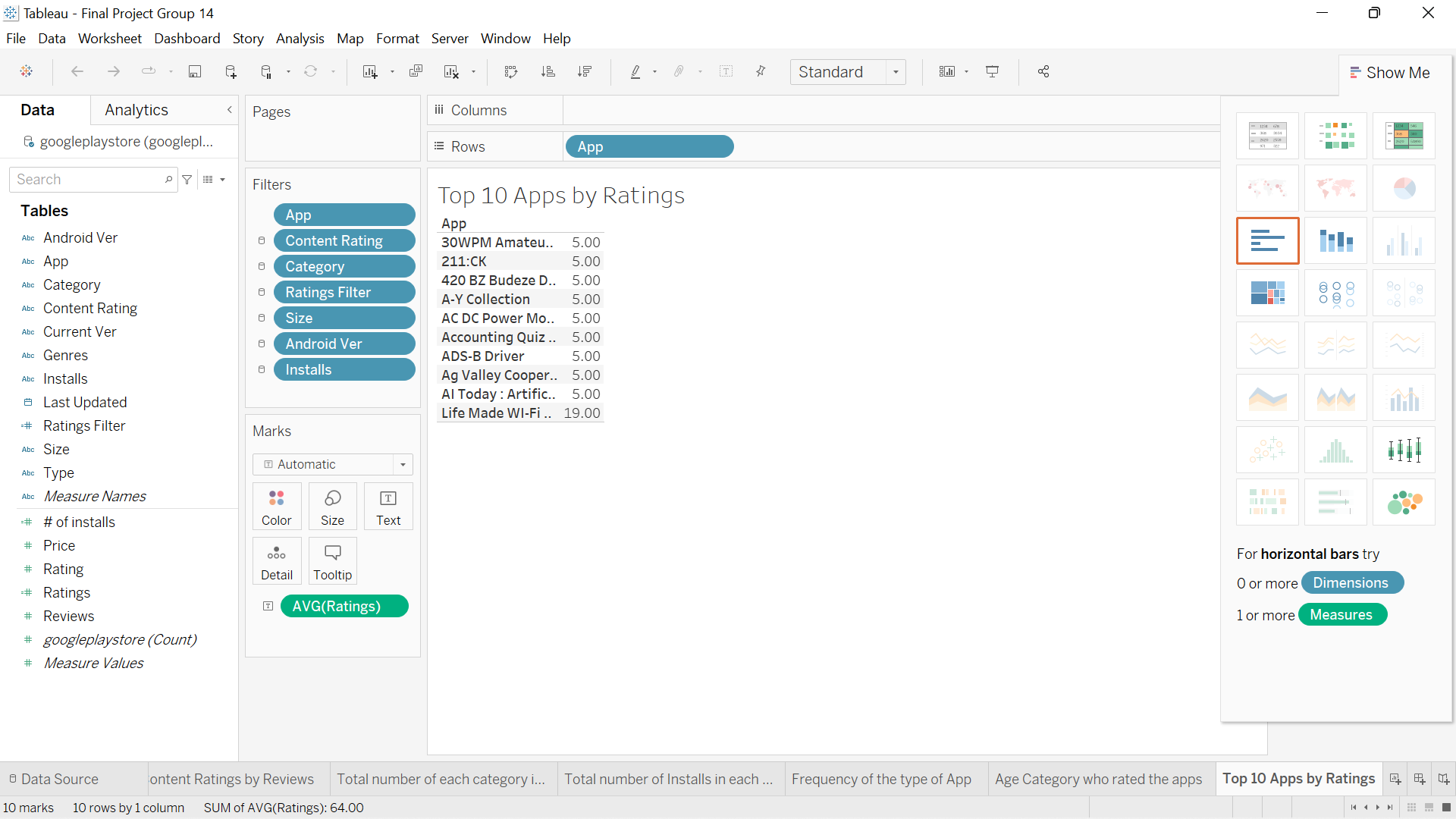
**Frequency of the type of App**

****

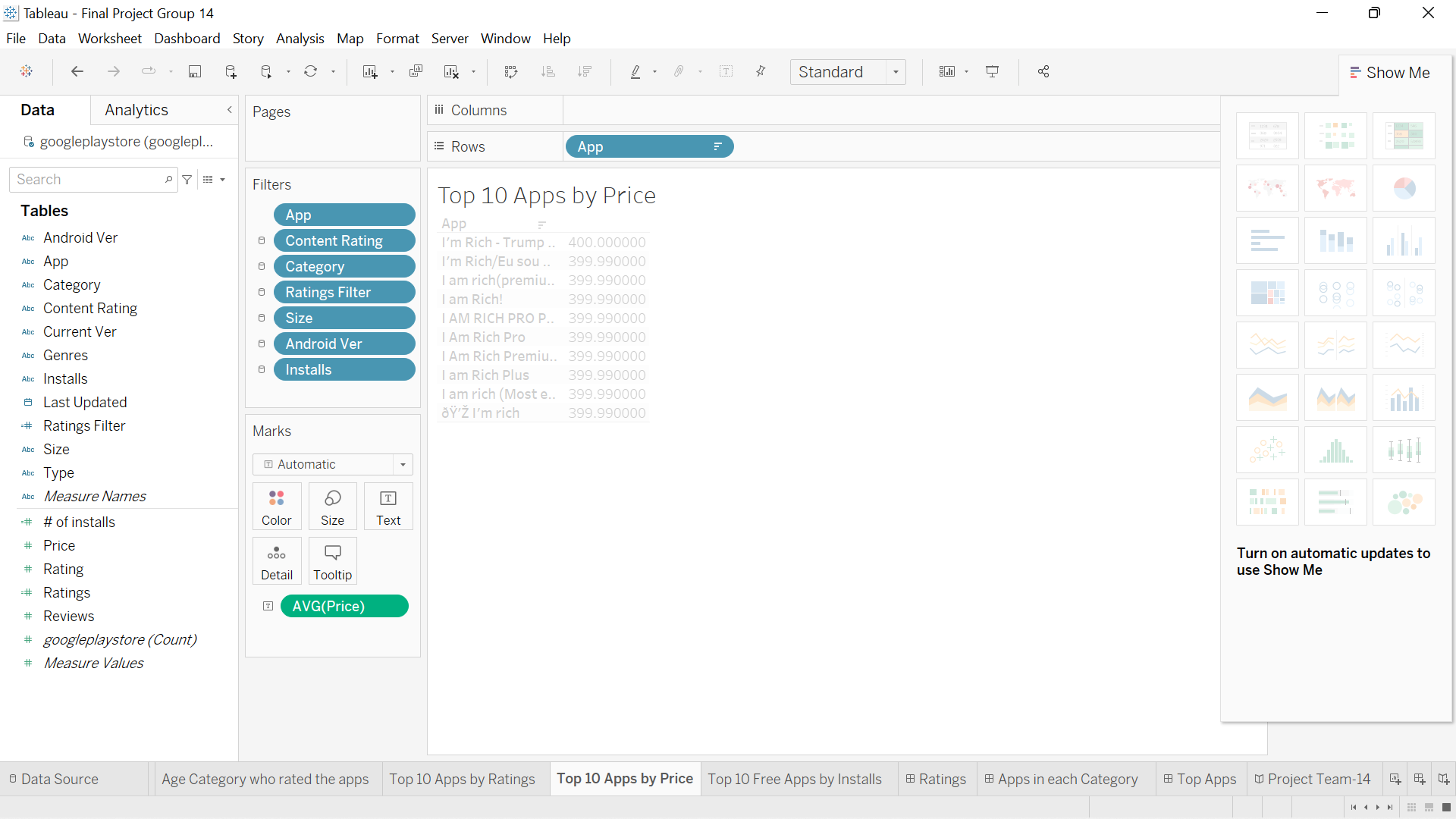
**Age category who rated the apps**

****

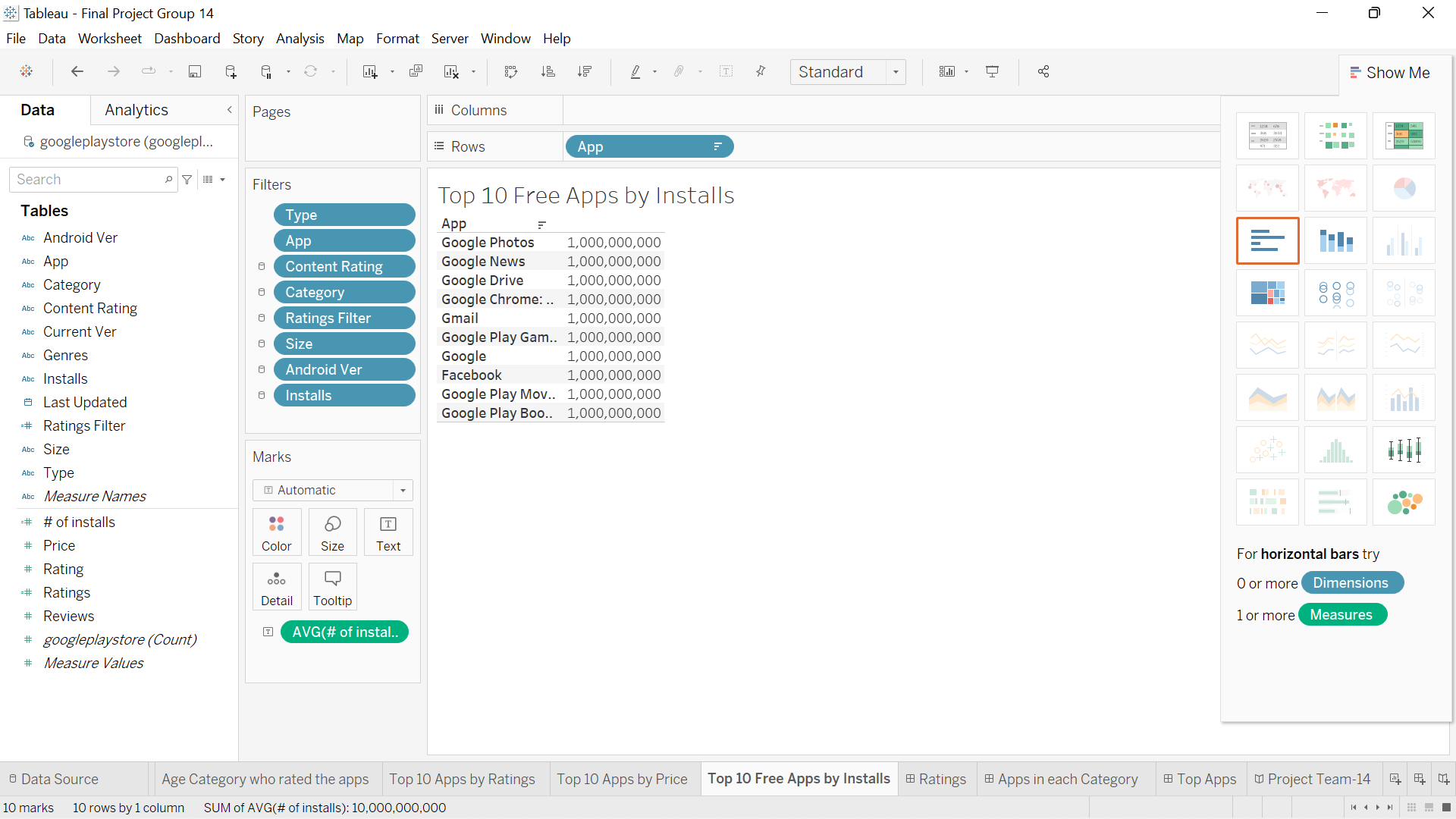
**Top 10 Apps by rating**

****

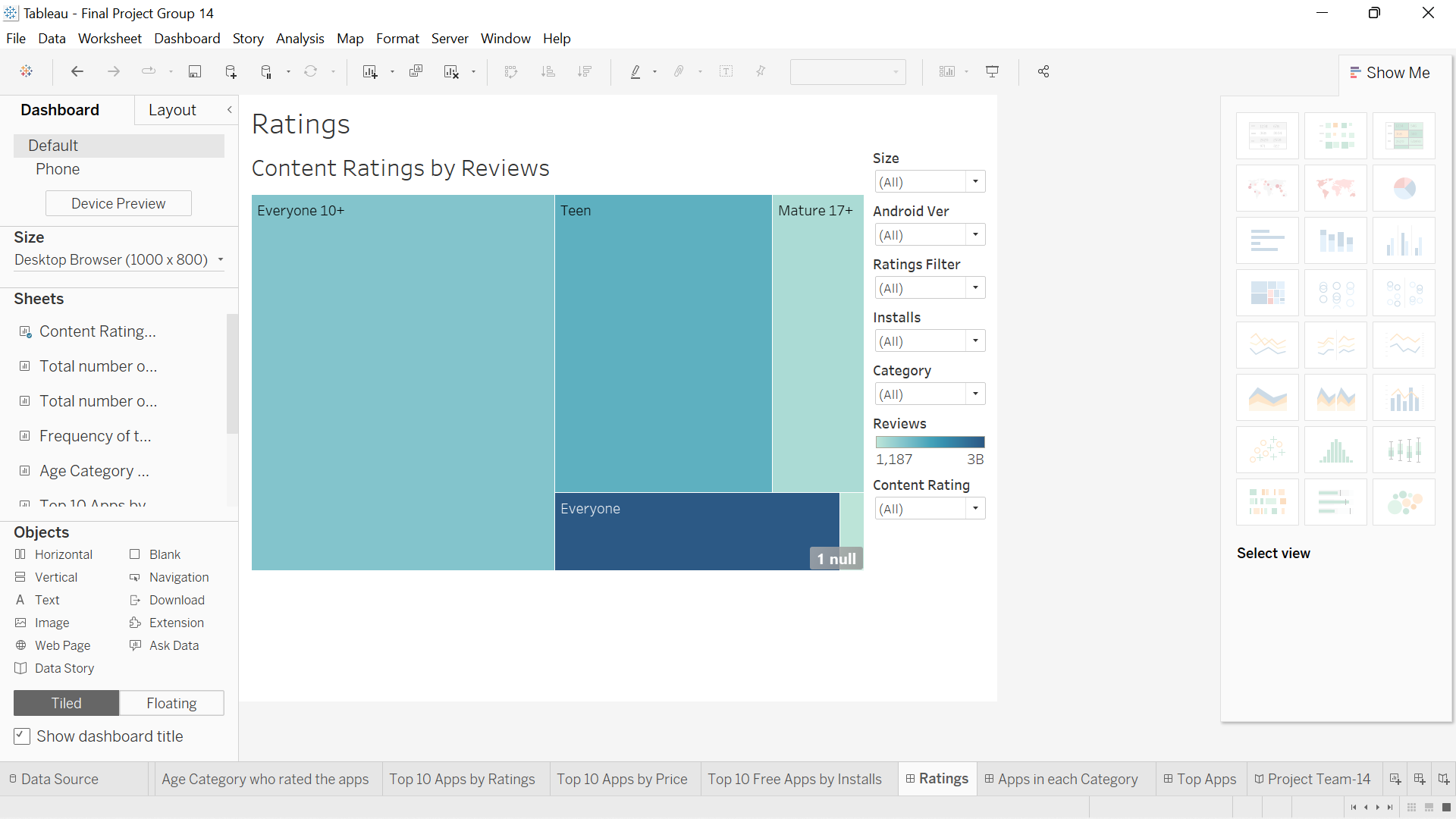
**Top 10 Apps by Price**

****

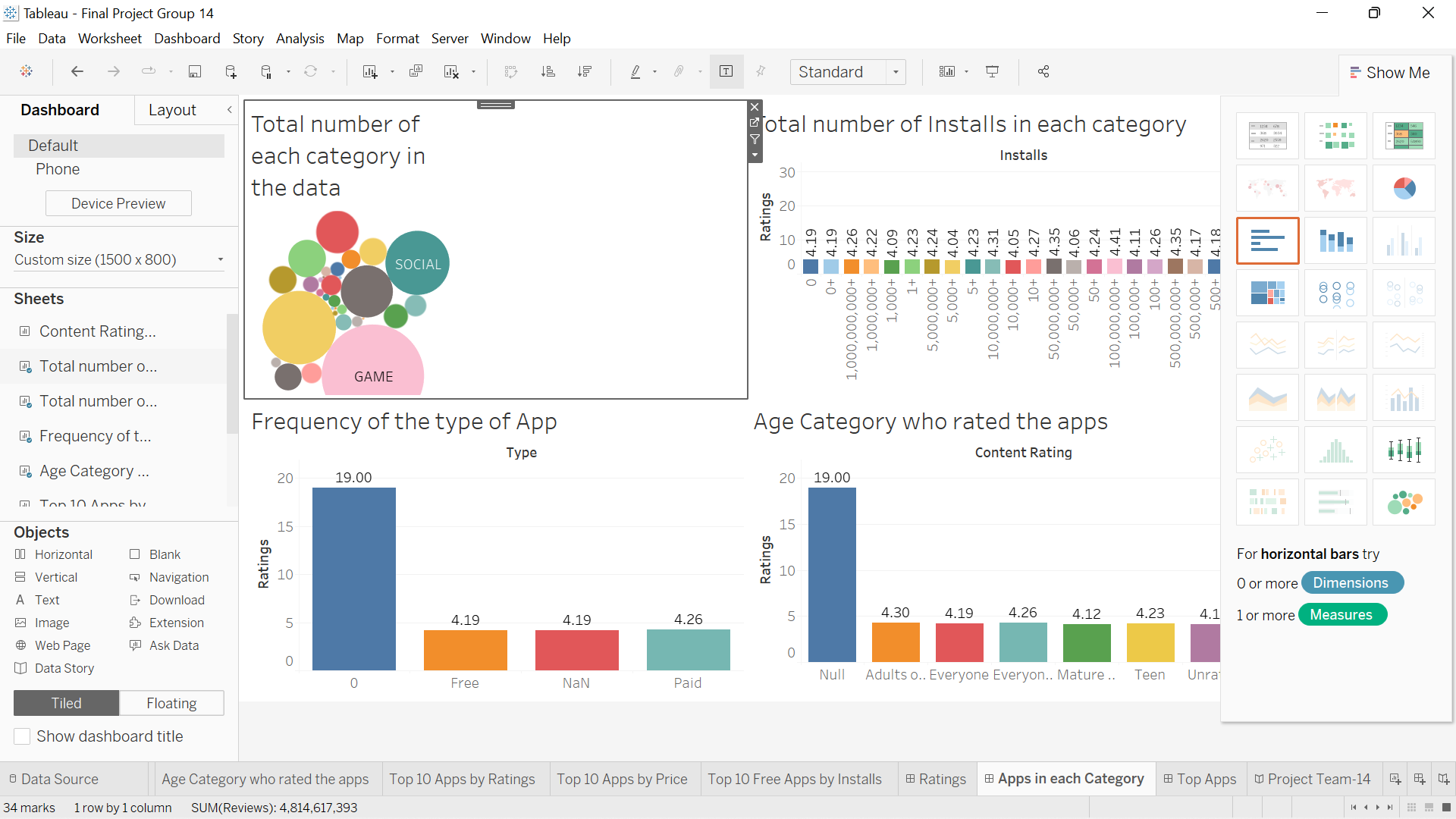
**Top 10 Free Apps by Installs**

****

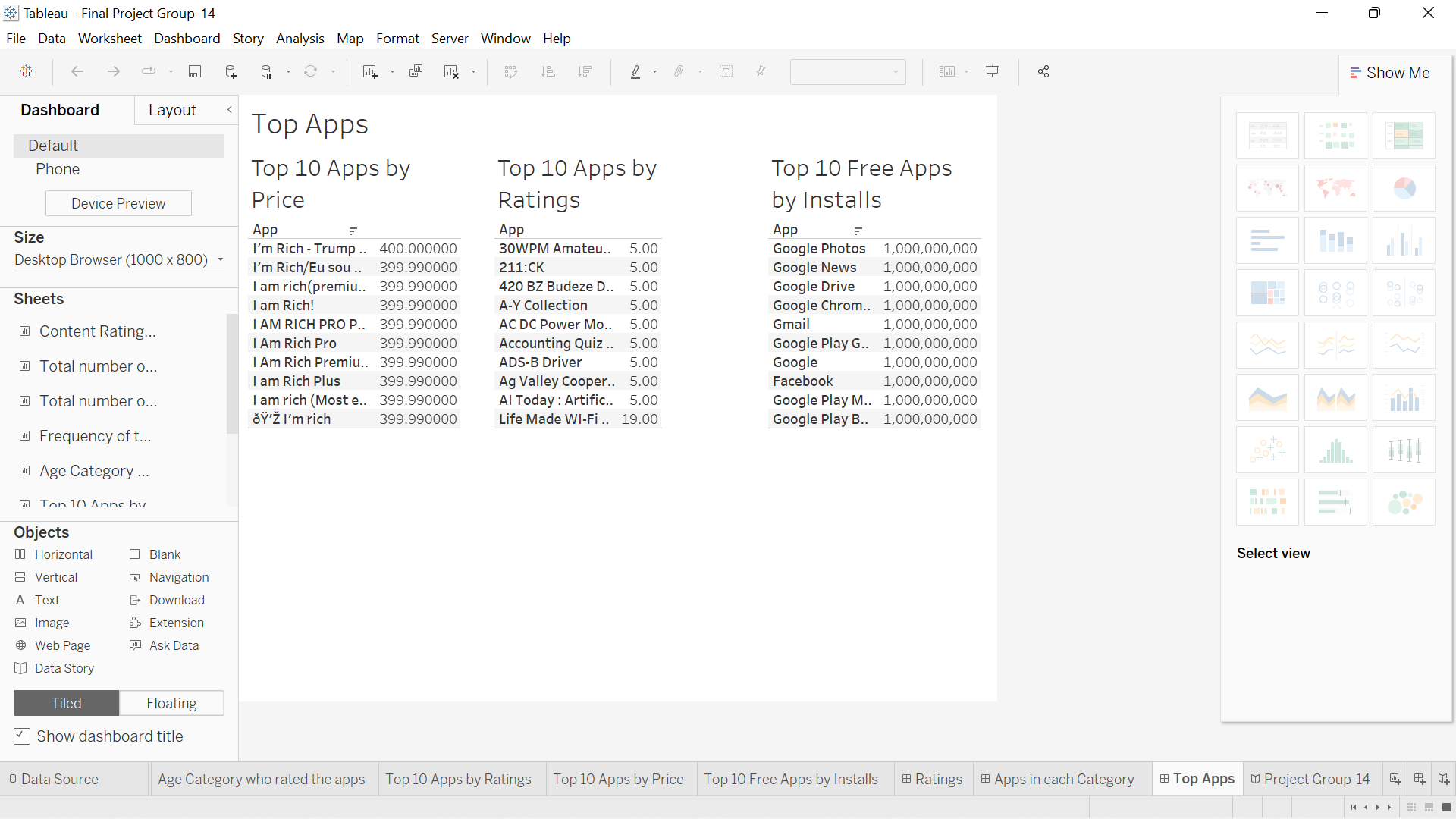
**Ratings Dashboard**

****

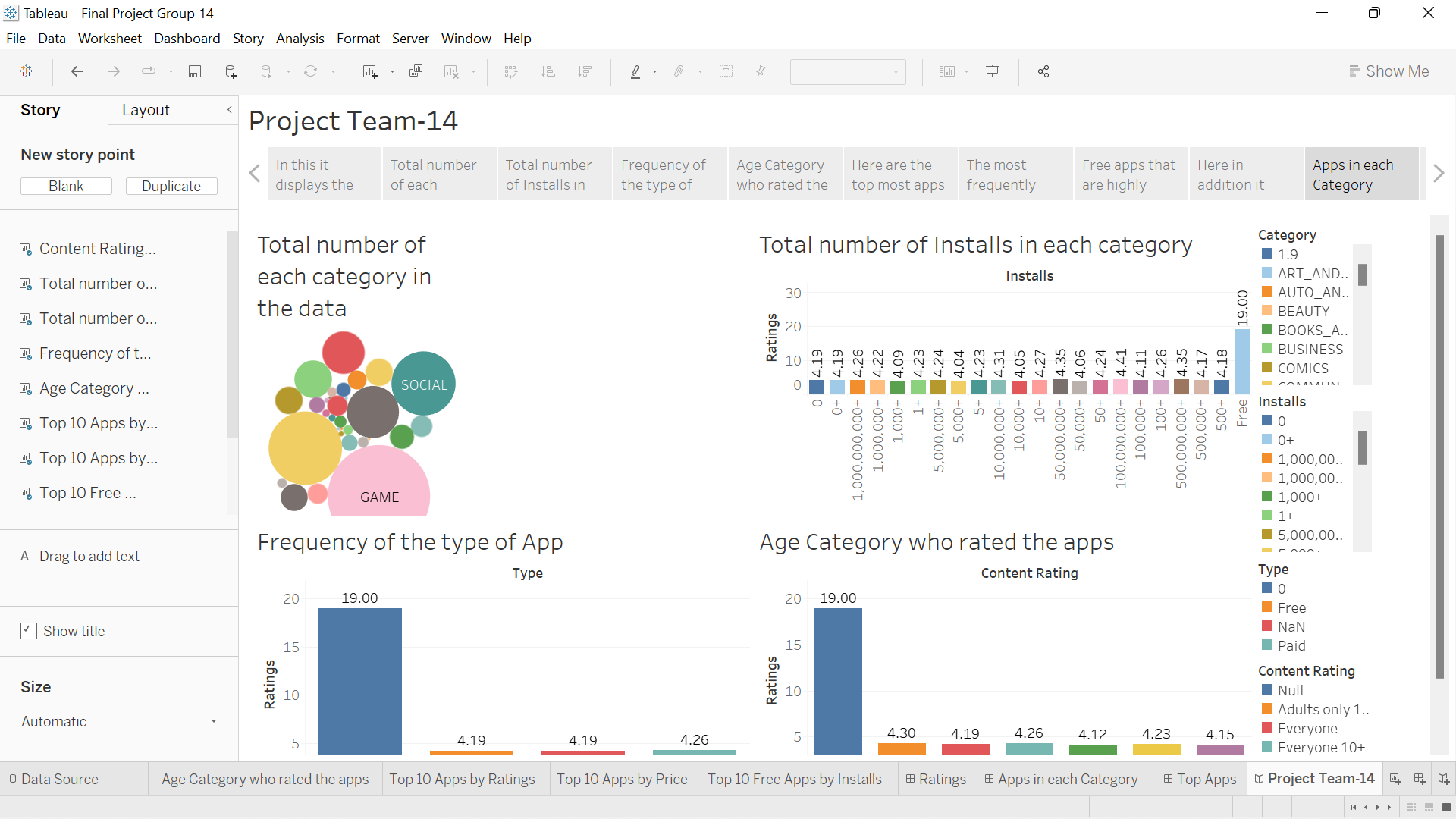
**Apps in each Category Dashboard**

****

**Top Apps**

****

**Overview**

****