

Mod#1 Case Study

--EDA of Google Play Store Data

Communicator: Tia Plagata

Team member: Wolfgang Wallach, Jiarong Chen

Question#1

—

Best Rating App Category

What is 'best'?

From dictionary, BEST is the most excellent, effective, or desirable type or quality.

In our App Store data, we can use 'rating' to define our 'best'.



Data cleaning

```
: #check the info of those erroneous
google_play_df.loc[google_play_df['Rating'] == 19]
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	
10472	Life Made WI-Fi Touchscreen Photo Frame		1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN	February 11, 2018	1.0.19	4.0 and up	NaN

```
In [12]: #check that if every value unique comparing the total entries
len(google_rating_df['App'].unique()), google_rating_df.shape
```

```
Out[12]: (8196, (9366, 13))
```

```
In [13]: #drop dulpicates and double check
google_rating_df.drop_duplicates(subset=['App'], inplace=True)
google_rating_df.shape
```

```
Out[13]: (8196, 13)
```

```
: #check the na percentile
google_rating_df['Rating'].isna().sum() / google_rating_df.shape[0]
```

```
: 0.13597785977859778
```

EDA

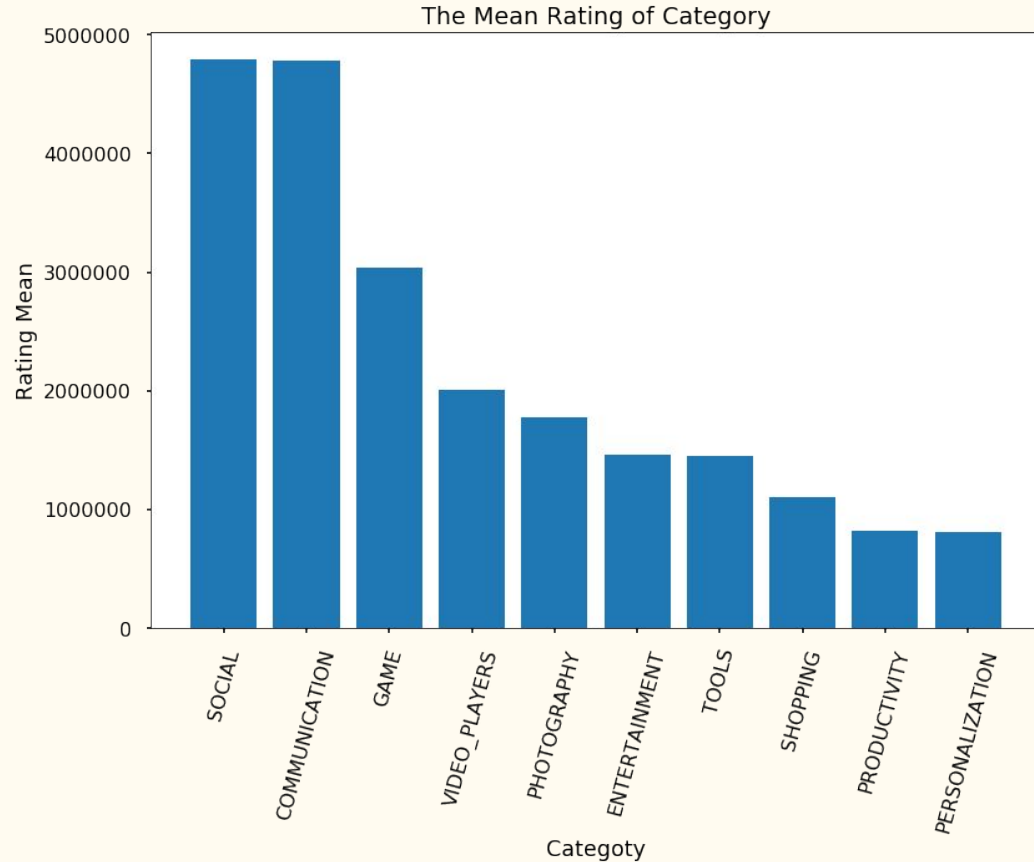
$$\text{Total_Rating} = \text{Rating} * \text{Reviews}$$

Out[20]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Total_Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up	651.9
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up	3771.3
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up	411297.0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up	970398.0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up	4158.1

Visualization

Category	Total_Rating_Mean
SOCIAL	4.789059e+06
COMMUNICATION	4.782785e+06
GAME	3.039761e+06
VIDEO_PLAYERS	2.008252e+06
PHOTOGRAPHY	1.772957e+06



Findings & Recommendations

The best category is:

Social : 4789059

*For app managers:

Import more social apps into app market

*For consumers:

There are lots of best social apps in the store, try to use different

Question#2

—

Relationship between 'Size' and 'Rating'

Data cleaning

```
array(['19M', '14M', '8.7M', '25M', '2.8M', '5.6M', '29M', '33M', '3.1M',  
      '28M', '12M', '20M', '21M', '37M', '2.7M', '5.5M', '17M', '39M',  
      '31M', '4.2M', '23M', '6.0M', '6.1M', '4.6M', '9.2M', '5.2M',  
      '11M', '24M', 'Varies with device', '9.4M', '15M', '10M', '1.2M',  
      '26M', '8.0M', '7.9M', '56M', '57M', '35M', '54M', '201k', '3.6M',  
      '5.7M', '8.6M', '2.4M', '27M', '2.5M', '7.0M', '16M', '3.4M',  
      '8.9M', '3.9M', '2.9M', '38M', '32M', '5.4M', '18M', '1.1M',  
      '2.2M', '4.5M', '9.8M', '52M', '9.0M', '6.7M', '30M', '2.6M',  
      '7.1M', '22M', '6.4M', '3.2M', '8.2M', '4.9M', '9.5M', '5.0M',  
      '5.9M', '13M', '73M', '6.8M', '3.5M', '4.0M', '2.3M', '2.1M',  
      '42M', '9.1M', '55M', '23k', '7.3M', '6.5M', '1.5M', '7.5M', '51M',  
      '41M', '48M', '8.5M', '46M', '8.3M', '4.3M', '4.7M', '3.3M', '40M',  
      '7.8M', '8.8M', '6.6M', '5.1M', '61M', '66M', '79k', '8.4M',  
      '3.7M', '118k', '44M', '695k', '1.6M', '6.2M', '53M', '1.4M',
```

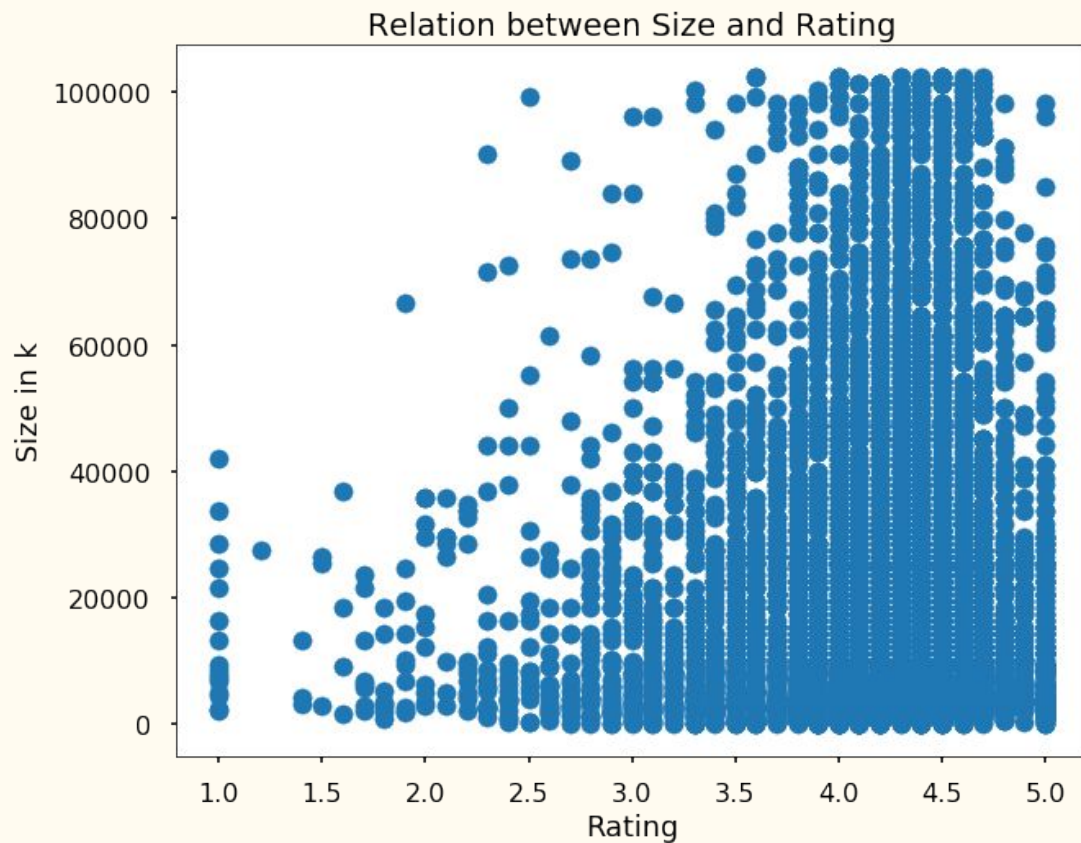
EDA

Measure it by using size in k

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Total_Rating	Size_in_k
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up	651.9	19456.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up	3771.3	14336.0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up	411297.0	8908.8
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up	970398.0	25600.0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up	4158.1	2867.2

Visualization

	Rating	Reviews	Total_Rating	Size_in_k
Rating	1.000000	0.067589	0.068383	0.063070
Reviews	0.067589	1.000000	0.999691	0.180521
Total_Rating	0.068383	0.999691	1.000000	0.178531
Size_in_k	0.063070	0.180521	0.178531	1.000000



Findings & Recommendations

The relationship between size and rating is

0.063070

The relationship between size and rating is weak.

*For app managers:

Do need to care about the size of app.

*For consumers:

Can try more different size of app. You will find surprise.

Question 3

Genres in a chosen Category

We chose: 'GAME'

```
In [26]: # All Categories
print(google_play_df['Category'].unique())

# New DF with just 'Game' Category
game_category = google_play_df.loc[google_play_df['Category'] == 'GAME']

['ART_AND_DESIGN' 'AUTO_AND_VEHICLES' 'BEAUTY' 'BOOKS_AND_REFERENCE'
'BUSINESS' 'COMICS' 'COMMUNICATION' 'DATING' 'EDUCATION' 'ENTERTAINMENT'
'EVENTS' 'FINANCE' 'FOOD_AND_DRINK' 'HEALTH_AND_FITNESS' 'HOUSE_AND_HOME'
'LIBRARIES_AND_DEMO' 'LIFESTYLE' 'GAME' 'FAMILY' 'MEDICAL' 'SOCIAL'
'SHOPPING' 'PHOTOGRAPHY' 'SPORTS' 'TRAVEL_AND_LOCAL' 'TOOLS'
'PERSONALIZATION' 'PRODUCTIVITY' 'PARENTING' 'WEATHER' 'VIDEO_PLAYERS'
'NEWS_AND_MAGAZINES' 'MAPS_AND_NAVIGATION']
```



Download Meme Center App to start the game of JUMANJI also available for Android now :
<https://play.google.com/store/apps/details?id=com.memecenter&hl=en>

How many 'Genres' in 'GAME'?

```
In [27]: # Unique Genres in Game Category  
game_category['Genres'].unique()
```

```
Out[27]: array(['Adventure;Action & Adventure', 'Arcade', 'Casual', 'Card',  
                'Casual;Pretend Play', 'Action', 'Strategy', 'Puzzle', 'Sports',  
                'Music', 'Word', 'Racing', 'Casual;Creativity',  
                'Casual;Action & Adventure', 'Simulation', 'Adventure', 'Board',  
                'Trivia', 'Role Playing', 'Simulation;Education',  
                'Action;Action & Adventure', 'Casual;Brain Games',  
                'Simulation;Action & Adventure', 'Casino'], dtype=object)
```

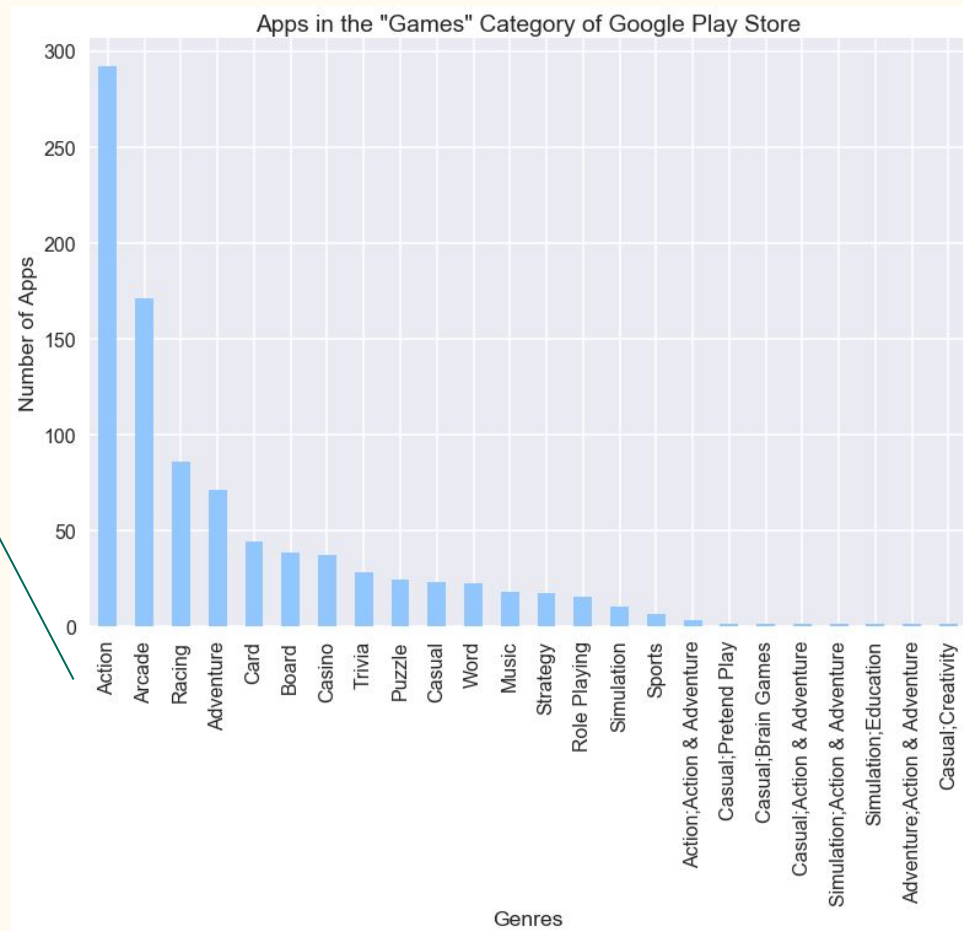
```
In [28]: # Number of Unique Genres in Game Category  
len(game_category['Genres'].unique())
```

```
Out[28]: 24
```

24!

Let's break it down:

Genres	
Action	292
Arcade	171
Racing	86
Adventure	71
Card	44
Board	38
Casino	37
Trivia	28
Puzzle	24
Casual	23
Word	22
Music	18
Strategy	17
Role Playing	15
Simulation	10
Sports	6
Action;Action & Adventure	3
Casual;Pretend Play	1
Casual;Brain Games	1
Casual;Action & Adventure	1
Simulation;Action & Adventure	1
Simulation;Education	1
Adventure;Action & Adventure	1
Casual;Creativity	1



Findings & Recommendations

The 3 genres with the most apps

1. 'Action' with 292 apps
2. 'Arcade' with 171 apps
3. 'Racing' with 86 apps.

Mean: 38.0

Median: 17.5

St. Dev: 64.7

*For app managers:

-How do you want to list your app?

-Go with trend or be unique

*For consumers:

-Most games are listed in these genres



Question 4

—

Relationship between Price and Installs

Can We Find a Relationship Between App Price
and it's Total Downloads?



Making the Data Useable

```
1 # Check types of the values
2 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9658 entries, 0 to 9658
Data columns (total 8 columns):
App                9658 non-null object
Category          9658 non-null object
Reviews           9658 non-null object
Size              9658 non-null object
Installs           9658 non-null object
Price             9658 non-null object
Genres            9658 non-null object
Last Updated      9658 non-null object
dtypes: object(8)
memory usage: 679.1+ KB
```

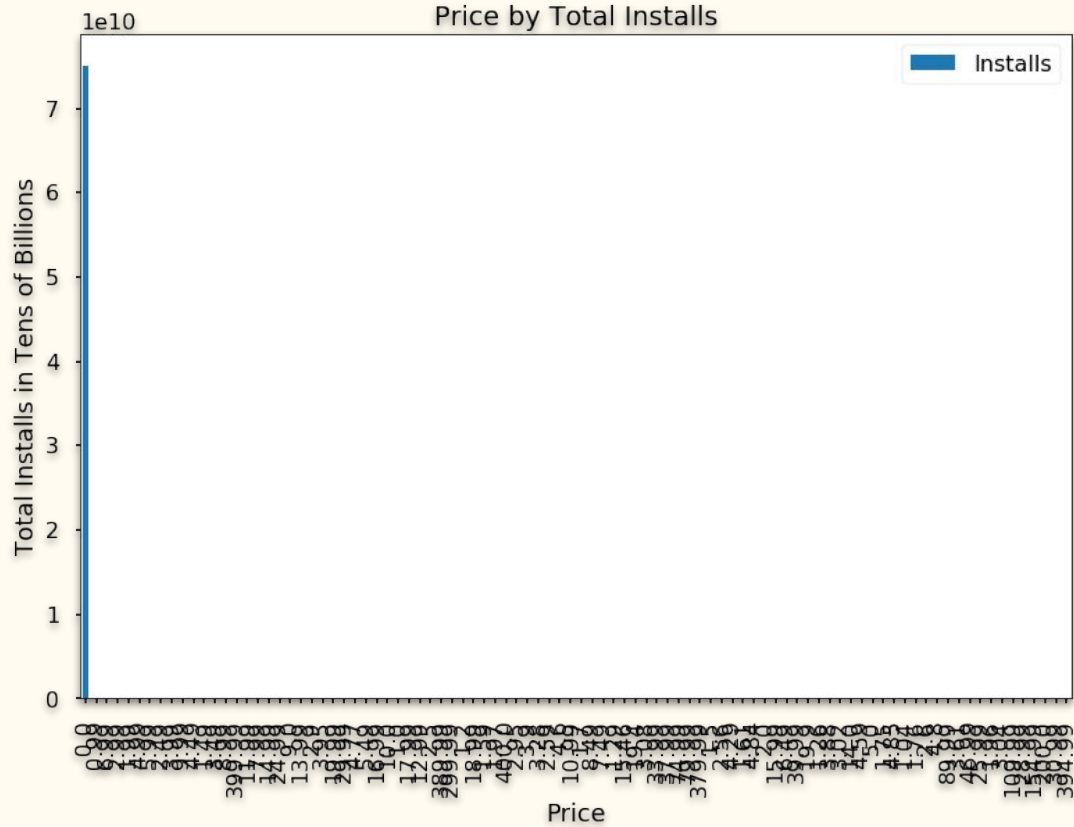
```
1 # Convert Installs to Integers
2 df.Installs = df.Installs.map(lambda x: x.replace(",", ""))
3 df.Installs = df.Installs.map(lambda x: x.replace("+", ""))
4 df.Installs = df.Installs.map(lambda x: x.strip())
5 df['Installs'] = df['Installs'].astype(int)
6 df['Installs'].value_counts()
```

```
1000000    1417
100000     1112
10000      1031
10000000    937
1000       888
100        710
5000000     607
500000     505
50000      469
5000       468
10         385
500        328
50         204
50000000    202
100000000   188
5          82
1          67
500000000   24
1000000000  20
0          14
Name: Installs, dtype: int64
```

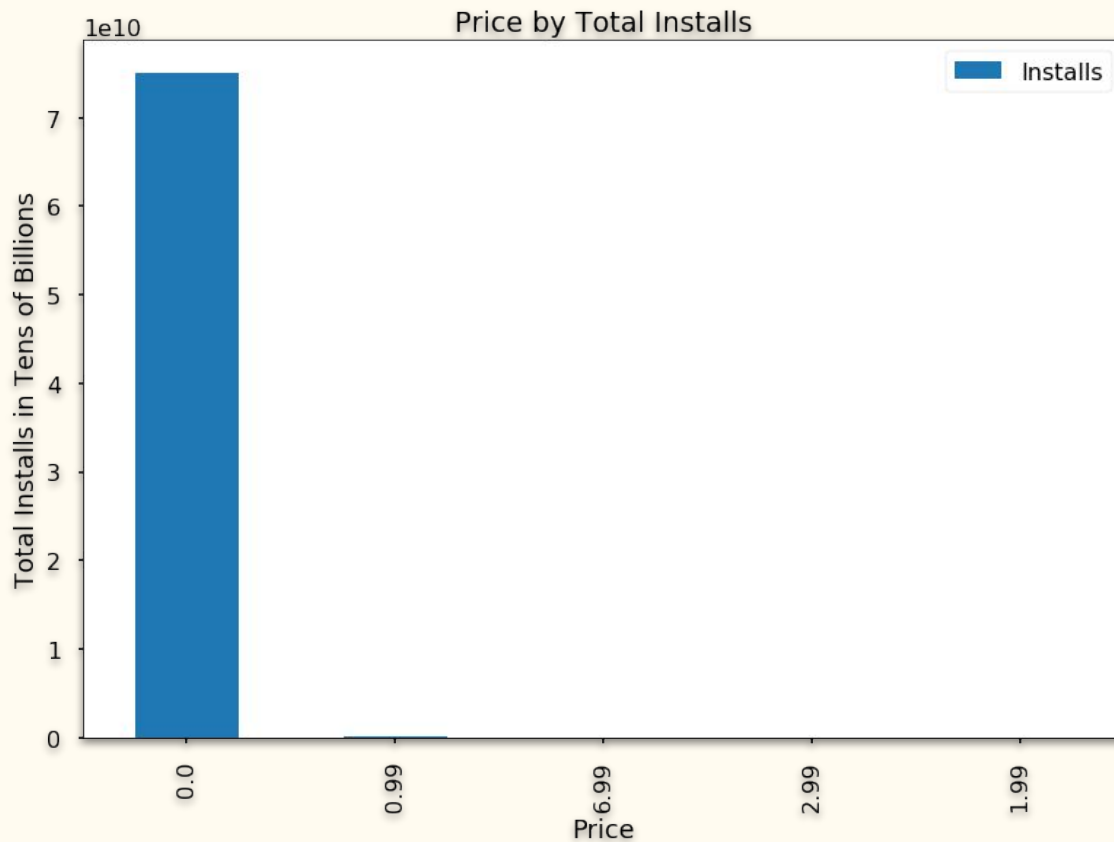
```
1 # Convert Prices to Floats
2 df.Price = df.Price.map(lambda x: x.replace("$", ""))
3 df.Price = df.Price.map(lambda x: x.strip())
4 df['Price'] = df['Price'].astype(float)
5 df['Price'].value_counts()
```

```
0.00    8902
0.99     145
2.99     124
1.99      73
4.99      70
...
1.59         1
46.99        1
19.90        1
28.99        1
1.04         1
Name: Price, Length: 92, dtype: float64
```

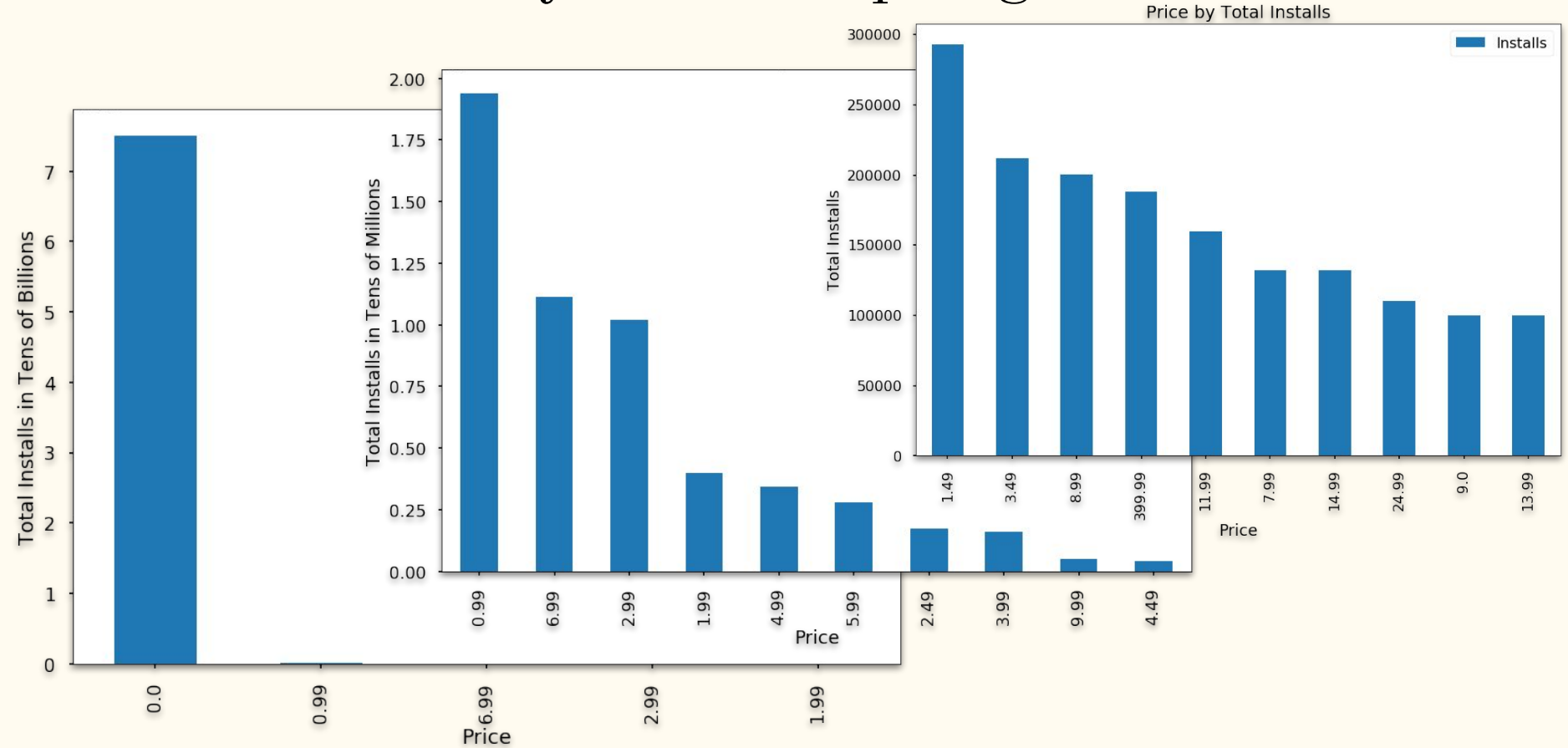
Graphing it Out!



Still Graphing it Out!



Finally Done Graphing it Out!



Findings and Recommendations

- If you plan to release an app on the Android store, your best bet for reaching a wide audience is to make the app free!
 - With in app purchasing data, this analysis could go further
- If you must put a price on your app, you'll be best served by keeping the price under \$10

