

Apple store Data Sets

```
In [2]: 1
        2 filepath='DataFiles/AppleStore.csv'
        3 import pandas as pd
        4 def readCSVdata(filepath):
        5     return pd.read_csv(filepath)
        6 readCSVdata(filepath)
```

Out[2]:

	Unnamed: 0	id	track_name	size_bytes	currency	price	rating_count_tot	rating_c
0	1	281656475	PAC-MAN Premium	100788224	USD	3.99	21292	
1	2	281796108	Evernote - stay organized	158578688	USD	0.00	161065	
2	3	281940292	WeatherBug - Local Weather, Radar, Maps, Alerts	100524032	USD	0.00	188583	
3	4	282614216	eBay: Best App to Buy, Sell, Save! Online Shop...	128512000	USD	0.00	262241	

```
In [6]: 1 Appstore=readCSVdata(filepath)
2         filepath='DataFiles/AppleStore.csv'
3         def printDataFrameColumns(df):
4             columns=df.columns
5             for column in columns:
6                 print(column)
7             return
8         printDataFrameColumns(Appstore)
```

```
Unnamed: 0
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
```

```
In [4]: 1 Appstore=readCSVdata(filepath)
2         # filepath='DataFiles/AppleStore.csv'
3         li=[]
4         def Prime_Genres(df):
5             for i in range(len(df.values)):
6                 for j in range(12,len(df.columns)-4):
7                     d=df.values[i][j]
8                     if d not in li:
9                         li.append(d)
10             print(li,end=" ")
11         Prime_Genres(Appstore)
```

```
['Games', 'Productivity', 'Weather', 'Shopping', 'Reference', 'Finance', 'Music', 'Utilities', 'Travel', 'Social Networking', 'Sports', 'Business', 'Health & Fitness', 'Entertainment', 'Photo & Video', 'Navigation', 'Education', 'Lifestyle', 'Food & Drink', 'News', 'Book', 'Medical', 'Catalogs']
```

```

In [7]: 1 Appstore=readCSVdata(filepath)
        2 li={}
        3 def highestnumberapps(df):
        4     for i in range(len(df.values)):
        5         for j in range(12,len(df.columns)-4):
        6             d=df.values[i][j]
        7             if d not in li.keys():
        8                 li[d]=1
        9             else:
        10                 li[d]+=1
        11     print(li)
        12     u=sorted(li.values(),reverse=True)
        13     print(u)
        14     print('\n')
        15     max1=max(u)
        16     for item in li.items():
        17         if item[1]==max1:
        18             print(item[0],':',max1)
        19 highestnumberapps(Appstore)
        20
        21
        22

```

```

{'Games': 3862, 'Productivity': 178, 'Weather': 72, 'Shopping': 122, 'Reference': 64, 'Finance': 104, 'Music': 138, 'Utilities': 248, 'Travel': 81, 'Social Networking': 167, 'Sports': 114, 'Business': 57, 'Health & Fitness': 180, 'Entertainment': 535, 'Photo & Video': 349, 'Navigation': 46, 'Education': 453, 'Lifestyle': 144, 'Food & Drink': 63, 'News': 75, 'Book': 112, 'Medical': 23, 'Catalogs': 10}
[3862, 535, 453, 349, 248, 180, 178, 167, 144, 138, 122, 114, 112, 104, 81, 75, 72, 64, 63, 57, 46, 23, 10]

```

Games : 3862

```

In [8]: 1 Appstore=readCSVdata(filepath)
        2 li={}
        3 def lowestnumberapps(df):
        4     for i in range(len(df.values)):
        5         for j in range(12,len(df.columns)-4):
        6             d=df.values[i][j]
        7             if d not in li.keys():
        8                 li[d]=1
        9             else:
        10                 li[d]+=1
        11     print(li)
        12     u=sorted(li.values(),reverse=True)
        13     print(u)
        14     print('\n')
        15     min1=min(u)
        16     for item in li.items():
        17         if item[1]==min1:
        18             print(item[0],':',min1)
        19 lowestnumberapps(Appstore)
        20

```

```

{'Games': 3862, 'Productivity': 178, 'Weather': 72, 'Shopping': 122, 'Reference': 64, 'Finance': 104, 'Music': 138, 'Utilities': 248, 'Travel': 81, 'Social Networking': 167, 'Sports': 114, 'Business': 57, 'Health & Fitness': 180, 'Entertainment': 535, 'Photo & Video': 349, 'Navigation': 46, 'Education': 453, 'Lifestyle': 144, 'Food & Drink': 63, 'News': 75, 'Book': 112, 'Medical': 23, 'Catalogs': 10}
[3862, 535, 453, 349, 248, 180, 178, 167, 144, 138, 122, 114, 112, 104, 81, 75, 72, 64, 63, 57, 46, 23, 10]

```

Catalogs : 10

```

In [10]: 1 Appstore = readCSVdata(filepath)
          2
          3 def ci(appdf,key):
          4     for i in range(len(appdf.values)):
          5         if appdf.columns[i]==key:
          6             return i
          7
          8 def ci1(appdf,key1):
          9     for i in range(len(appdf.values)):
         10         if appdf.columns[i]==key1:
         11             return i
         12
         13 def Highest_User_Rating(appdf,key,key1):
         14     li=[]
         15     CI=ci(appdf,key)
         16     CI1=ci1(appdf,key1)
         17     for i in appdf.values:
         18         li.append(i[CI])
         19     m=max(li)
         20     print(m)
         21     s=[]
         22     for i in appdf.values:
         23         if m==i[CI]:
         24             s.append(i[CI1])
         25     print(set(s))
         26     print(len(set(s)))
         27
         28 Highest_User_Rating(Appstore,'user_rating','prime_genre')

```

5.0

```

{'Utilities', 'Navigation', 'Education', 'News', 'Shopping', 'Music', 'Health &
Fitness', 'Food & Drink', 'Productivity', 'Book', 'Weather', 'Lifestyle', 'Cata
logs', 'Medical', 'Entertainment', 'Reference', 'Social Networking', 'Finance',
'Sports', 'Games', 'Travel', 'Photo & Video', 'Business'}

```

23

In [11]:

```
1  ### App with highest downloads
2
3  Appstore = readCSVdata(filepath)
4  def ci(appdf, key):
5      for i in range(len(appdf.values)):
6          if appdf.columns[i]==key:
7              return i
8
9
10 def ci1(appdf, key1):
11     for i in range(len(appdf.values)):
12         if appdf.columns[i]==key1:
13             return i
14
15 def Highest_Downloads(appdf, key, key1):
16     li=[]
17     CI=ci(appdf, key)
18     CI1=ci1(appdf, key1)
19     for i in appdf.values:
20         li.append(i[CI1])
21     m=max(li)
22     print(m)
23     s=[]
24     for i in appdf.values:
25         if m==i[CI1]:
26             s.append(i[CI])
27     print(set(s))
28     print(len(set(s)))
29
30 Highest_Downloads(Appstore, 'track_name', 'rating_count_tot')
```

2974676

{'Facebook'}

1

```
In [12]: 1  ### Category with highest average rating count
2
3  Appstore = readCSVdata(filepath)
4  def ci(appdf, key):
5      for i in range(len(appdf.values)):
6          if appdf.columns[i]==key:
7              return i
8
9  def ci1(appdf, key1):
10     for i in range(len(appdf.values)):
11         if appdf.columns[i]==key1:
12             return i
13
14  def Highest_Avg_Ratingcount(appdf, key, key1):
15     li=[]
16     CI=ci(appdf, key)
17     CI1=ci1(appdf, key1)
18     for i in appdf.values:
19         li.append(i[CI])
20     m=max(li)
21     print(m)
22     s=[]
23     for i in appdf.values:
24         if m==i[CI]:
25             s.append(i[CI1])
26     print(set(s))
27     print(len(set(s)))
28
29  Highest_Avg_Ratingcount(Appstore, 'rating_count_tot', 'prime_genre')
```

2974676

{'Social Networking'}

1

In [13]:

```
1 ##### Average user rating for free apps
2
3 Appstore = readCSVdata(filepath)
4
5 def ci(appdf,key):
6     for i in range(len(appdf.values)):
7         if appdf.columns[i]==key:
8             return i
9
10 def ci1(appdf,key1):
11     for i in range(len(appdf.values)):
12         if appdf.columns[i]==key1:
13             return i
14 def Average_User_Rating_forFreeApps(appdf,key,key1):
15     li=[]
16     CI=ci(appdf,key)
17     CI1=ci1(appdf,key1)
18     for i in appdf.values:
19         li.append(i[CI])
20     m=min(li)
21     print(m)
22     s=[]
23     for i in appdf.values:
24         if m==i[CI]:
25             s.append(i[CI1])
26     a=sum(s)/len(s)
27     print(a)
28 Average_User_Rating_forFreeApps(Appstore,'price','user_rating')
```

0.0

3.3767258382642997


```
In [14]: 1  ### Average user rating for paid apps
2
3  Appstore = readCSVdata(filepath)
4
5  def ci(appdf, key):
6      for i in range(len(appdf.values)):
7          if appdf.columns[i] == key:
8              return i
9
10 def ci1(appdf, key1):
11     for i in range(len(appdf.values)):
12         if appdf.columns[i] == key1:
13             return i
14 def Average_User_Rating_forPaidApps(appdf, key, key1):
15     li = []
16     CI = ci(appdf, key)
17     CI1 = ci1(appdf, key1)
18     for i in appdf.values:
19         if i[CI] > 0:
20             li.append(i[CI1])
21     print(sum(li)/len(li))
22
23 Average_User_Rating_forPaidApps(Appstore, 'price', 'user_rating')
```

3.720948742438714

```

In [15]: 1  ### Category with highest average user rating for paid apps
2  Appstore = readCSVdata(filepath)
3
4  def ci(appdf, key):
5      for i in range(len(appdf.values)):
6          if appdf.columns[i] == key:
7              return i
8
9  def ci1(appdf, key1):
10     for i in range(len(appdf.values)):
11         if appdf.columns[i] == key1:
12             return i
13
14  def ci2(appdf, key2):
15     for i in range(len(appdf.values)):
16         if appdf.columns[i] == key2:
17             return i
18  def Average_User_Rating_forPaidApps(appdf, key, key1, key2):
19     li = []
20     CI = ci(appdf, key)
21     CI1 = ci1(appdf, key1)
22     CI2 = ci2(appdf, key2)
23     for i in appdf.values:
24         if i[CI] > 0:
25             li.append(i[CI1])
26     m = max(li)
27     print(m)
28     s = []
29     for i in appdf.values:
30         if m == i[CI1]:
31             s.append(i[CI2])
32     print(set(s))
33
34  Average_User_Rating_forPaidApps(Appstore, 'price', 'user_rating', 'prime_genre'
35
36
37

```

5.0

```

{'Utilities', 'Navigation', 'Education', 'News', 'Shopping', 'Music', 'Health &
Fitness', 'Food & Drink', 'Productivity', 'Book', 'Weather', 'Lifestyle', 'Cata
logs', 'Medical', 'Entertainment', 'Reference', 'Social Networking', 'Finance',
'Sports', 'Games', 'Travel', 'Photo & Video', 'Business'}

```

```

In [16]: 1  ### Most frequent Price point >0
          2
          3
          4 Appstore = readCSVdata(filepath)
          5
          6 def ci(appdf,key):
          7     for i in range(len(appdf.values)):
          8         if appdf.columns[i]==key:
          9             return i
         10
         11 def Frequent_Price(appdf,key):
         12     li=[]
         13     d={}
         14     C=ci(appdf,key)
         15     for i in appdf.values:
         16         if i[C]>0:
         17             li.append(i[C])
         18     for i in li:
         19         if i not in d:
         20             d[i]=1
         21         else:
         22             d[i]+=1
         23     print(d)
         24     print('\n')
         25     v=max(d.values())
         26     for item in d.items():
         27         if item[1]==v:
         28             print("most frequent price point is",item[0],':',v)
         29
         30
         31 Frequent_Price(Appstore,'price')
         32

```

```

{3.99: 277, 0.99: 728, 9.99: 81, 4.99: 394, 7.99: 33, 2.99: 683, 1.99: 621, 5.99: 52, 12.99: 5, 21.99: 1, 249.99: 1, 6.99: 166, 74.99: 1, 19.99: 13, 8.99: 9, 24.99: 8, 13.99: 6, 14.99: 21, 16.99: 2, 47.99: 1, 11.99: 6, 59.99: 3, 15.99: 4, 27.99: 2, 17.99: 3, 299.99: 1, 49.99: 2, 23.99: 2, 20.99: 2, 39.99: 2, 99.99: 1, 29.99: 6, 34.99: 1, 18.99: 1, 22.99: 2}

```

most frequent price point is 0.99 : 728

```

In [3]: 1 ##### Compare average user rating for paid vs free gaming apps
2
3
4 Appstore = readCSVdata(filepath)
5
6 def ci(appdf,key):
7     for i in range(len(appdf.values)):
8         if appdf.columns[i]==key:
9             return i
10
11 def ci1(appdf,key1):
12     for i in range(len(appdf.values)):
13         if appdf.columns[i]==key1:
14             return i
15
16 def ci2(appdf,key2):
17     for i in range(len(appdf.values)):
18         if appdf.columns[i]==key2:
19             return i
20 def Compare_freeGamingapps(appdf,key,key1,key2):
21     CI=ci(appdf,key)
22     CI1=ci1(appdf,key1)
23     CI2=ci2(appdf,key2)
24     sum1=0
25     count=0
26     s=0
27     c=0
28     for i in range(len(appdf.values)):
29         if appdf.values[i][CI]=='Games':
30             if appdf.values[i][CI1]>0:
31                 sum1=sum1+appdf.values[i][CI2]
32                 count=count+1
33             else:
34                 s=s+appdf.values[i][CI2]
35                 c=c+1
36     avg=sum1/count
37     avg1=s/c
38     print(avg,avg1)
39     print('\n')
40     if avg>avg1:
41         print("Paid apps are having the highest than free apps",':',avg)
42     else:
43         print("Free are greater than paid apps",':',avg1)
44
45
46 Compare_freeGamingapps(Appstore,'prime_genre','price','user_rating',)
47
48
49
50
51
52

```

3.9049844236760123 3.5285777580859548

Paid apps are having the highest than free apps : 3.9049844236760123

In []:	1	
---------	---	--