google-play-store-app-analysis

November 11, 2024

---

1. IMPORTING LIBRARIES

```
[162]: import pandas as pd
```

---

2. IMPORTING DATASET

```
[163]: data = pd.read_csv(r"H:\DA. Python\13. Google Play Store App␣
        ↪Analysis\googleplaystore.csv")
```

---

3. DISPLAY THE TOP 5 ROWS OF DATASET

```
[164]: data.head(5)
```

```
[164]:                                                  App        Category  Rating  \
       0      Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
       1                                 Coloring book moana  ART_AND_DESIGN     3.9
       2  U Launcher Lite – FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
       3                               Sketch - Draw & Paint  ART_AND_DESIGN     4.5
       4               Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3

         Reviews  Size     Installs  Type Price Content Rating  \
       0     159   19M      10,000+  Free     0       Everyone
       1     967   14M     500,000+  Free     0       Everyone
       2   87510  8.7M   5,000,000+  Free     0       Everyone
       3  215644   25M  50,000,000+  Free     0           Teen
       4     967  2.8M     100,000+  Free     0       Everyone

                             Genres      Last Updated         Current Ver  \
       0                  Art & Design   January 7, 2018               1.0.0
       1    Art & Design;Pretend Play  January 15, 2018               2.0.0
       2                  Art & Design    August 1, 2018               1.2.4
       3                  Art & Design      June 8, 2018  Varies with device
       4      Art & Design;Creativity     June 20, 2018                 1.1

           Android Ver
```

```
0  4.0.3 and up
1  4.0.3 and up
2  4.0.3 and up
3    4.2 and up
4    4.4 and up
```

---

4. DISPLAY THE TOP 5 ROWS OF DATASET

[165]:
```
data.tail(5)
```

[165]:
```
                                                 App             Category  \
10836                               Sya9a Maroc - FR               FAMILY
10837               Fr. Mike Schmitz Audio Teachings               FAMILY
10838                         Parkinson Exercices FR              MEDICAL
10839                  The SCP Foundation DB fr nn5n  BOOKS_AND_REFERENCE
10840  iHoroscope - 2018 Daily Horoscope & Astrology            LIFESTYLE

      Rating Reviews                Size     Installs  Type Price  \
10836    4.5      38                 53M       5,000+  Free     0
10837    5.0       4                3.6M         100+  Free     0
10838    NaN       3                9.5M       1,000+  Free     0
10839    4.5     114  Varies with device       1,000+  Free     0
10840    4.5  398307                 19M  10,000,000+  Free     0

     Content Rating             Genres      Last Updated         Current Ver  \
10836       Everyone          Education     July 25, 2017                1.48
10837       Everyone          Education      July 6, 2018                 1.0
10838       Everyone            Medical  January 20, 2017                 1.0
10839     Mature 17+  Books & Reference  January 19, 2015  Varies with device
10840       Everyone          Lifestyle     July 25, 2018  Varies with device

              Android Ver
10836          4.1 and up
10837          4.1 and up
10838          2.2 and up
10839  Varies with device
10840  Varies with device
```

5.. GET INFORMATION ABOUT THE DATASET LIKE TOTAL NUMBER OF ROWS AND COLUMNS, DATATYPES OF EACH COLUMN AND MEMORY REQUIREMENT

[166]:
```
data.shape
```

[166]:
```
(10841, 13)
```

[167]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   App             10841 non-null  object
 1   Category        10841 non-null  object
 2   Rating          9367 non-null   float64
 3   Reviews         10841 non-null  object
 4   Size            10841 non-null  object
 5   Installs        10841 non-null  object
 6   Type            10840 non-null  object
 7   Price           10841 non-null  object
 8   Content Rating  10840 non-null  object
 9   Genres          10841 non-null  object
 10  Last Updated    10841 non-null  object
 11  Current Ver     10833 non-null  object
 12  Android Ver     10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```

---

6. GET OVERALL STATISTICS OF THE DATAFRAME

[168]:

```
data.describe(include = 'all')
```

[168]:

```
          App Category       Rating Reviews                Size     Installs  \
count   10841    10841  9367.000000   10841               10841        10841
unique   9660       34          NaN    6002                 462           22
top    ROBLOX   FAMILY          NaN       0  Varies with device   1,000,000+
freq        9     1972          NaN     596                1695         1579
mean      NaN      NaN     4.193338     NaN                 NaN          NaN
std       NaN      NaN     0.537431     NaN                 NaN          NaN
min       NaN      NaN     1.000000     NaN                 NaN          NaN
25%       NaN      NaN     4.000000     NaN                 NaN          NaN
50%       NaN      NaN     4.300000     NaN                 NaN          NaN
75%       NaN      NaN     4.500000     NaN                 NaN          NaN
max       NaN      NaN    19.000000     NaN                 NaN          NaN

         Type  Price Content Rating Genres    Last Updated  \
count   10840  10841          10840  10841           10841
unique      3     93              6    120            1378
top      Free      0       Everyone  Tools  August 3, 2018
freq    10039  10040           8714    842             326
mean      NaN    NaN            NaN    NaN             NaN
```

```
std       NaN    NaN            NaN    NaN             NaN
min       NaN    NaN            NaN    NaN             NaN
25%       NaN    NaN            NaN    NaN             NaN
50%       NaN    NaN            NaN    NaN             NaN
75%       NaN    NaN            NaN    NaN             NaN
max       NaN    NaN            NaN    NaN             NaN

               Current Ver Android Ver
count                10833       10838
unique                2832          33
top     Varies with device  4.1 and up
freq                  1459        2451
mean                   NaN         NaN
std                    NaN         NaN
min                    NaN         NaN
25%                    NaN         NaN
50%                    NaN         NaN
75%                    NaN         NaN
max                    NaN         NaN
```

---

7. TOTAL NUMBER OF APP TITLES CONTAINS ASTROLOGY.

[169]:
```
data.columns
```

[169]:
```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
```

[170]:
```
len(data[data['App'].str.contains('Astrology', case = False)])
```

[170]:
```
3
```

TOTAL NUMBER OF APP TITLES CONTAINS ASTROLOGY - 3

---

8. FIND AVERAGE APP RATING

[171]:
```
data.columns
```

[171]:
```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
```

[172]:
```
data['Rating'].mean()
```

```
[172]: np.float64(4.193338315362443)
```

AVERAGE APP RATING IS 4.1933

---

9. FIND TOTAL NUMBER OF UNIQUE CATEGORY.

```
[173]: data.columns
```

```
[173]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
              'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
              'Android Ver'],
             dtype='object')
```

```
[174]: data['Category'].nunique() #nunique counts the unique values
```

```
[174]: 34
```

TOTAL NUMBER OF UNIQUE CATEGORIES IS 34

---

10. WHICH CATEGORY GETTING THE HIGHEST AVERAGE RATINGS.

```
[175]: data.columns
```

```
[175]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
              'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
              'Android Ver'],
             dtype='object')
```

```
[176]: data.groupby('Category')['Rating'].mean().sort_values(ascending = False)
```

```
[176]: Category
       1.9                    19.000000
       EVENTS                  4.435556
       EDUCATION               4.389032
       ART_AND_DESIGN          4.358065
       BOOKS_AND_REFERENCE     4.346067
       PERSONALIZATION         4.335987
       PARENTING               4.300000
       GAME                    4.286326
       BEAUTY                  4.278571
       HEALTH_AND_FITNESS      4.277104
       SHOPPING                4.259664
       SOCIAL                  4.255598
       WEATHER                 4.244000
       SPORTS                  4.223511
       PRODUCTIVITY            4.211396
```

```
HOUSE_AND_HOME         4.197368
FAMILY                 4.192272
PHOTOGRAPHY            4.192114
AUTO_AND_VEHICLES      4.190411
MEDICAL                4.189143
LIBRARIES_AND_DEMO     4.178462
FOOD_AND_DRINK         4.166972
COMMUNICATION          4.158537
COMICS                 4.155172
NEWS_AND_MAGAZINES     4.132189
FINANCE                4.131889
ENTERTAINMENT          4.126174
BUSINESS               4.121452
TRAVEL_AND_LOCAL       4.109292
LIFESTYLE              4.094904
VIDEO_PLAYERS          4.063750
MAPS_AND_NAVIGATION    4.051613
TOOLS                  4.047411
DATING                 3.970769
Name: Rating, dtype: float64
```

---

11. FIND TOTAL NUMBER OF APPS HAVING 5 STAR RATING

```
[177]: data.columns
```

```
[177]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
              'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
              'Android Ver'],
             dtype='object')
```

```
[178]: len(data[data['Rating'] == 5.0])
```

```
[178]: 274
```

TOTAL NUMBER OF 5 STAR RATING APPS ARE 274

---

12. FIND AVERAGE VALUE OF REVIEWS.

```
[179]: data.columns
```

```
[179]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
              'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
              'Android Ver'],
             dtype='object')
```

[180]:
```
data['Reviews'].dtypes
```

[180]:
```
dtype('O')
```

[181]:
```
#converting data type from INT or FLOAT

data['Reviews'].astype(float)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[181], line 3
      1 #converting data type from INT or FLOAT
----> 3 data['Reviews'].astype(float)

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\generic.
 ↪py:6643, in NDFrame.astype(self, dtype, copy, errors)
   6637     results = [
   6638         ser.astype(dtype, copy=copy, errors=errors) for _, ser in self.
 ↪items()
   6639     ]
   6641 else:
   6642     # else, only a single dtype is given
-> 6643     new_data = self._mgr.astype(dtype=dtype, copy=copy, errors=errors)
   6644     res = self._constructor_from_mgr(new_data, axes=new_data.axes)
   6645     return res.__finalize__(self, method="astype")

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\internals
 ↪py:430, in BaseBlockManager.astype(self, dtype, copy, errors)
    427 elif using_copy_on_write():
    428     copy = False
--> 430 return self.apply(
    431     "astype",
    432     dtype=dtype,
    433     copy=copy,
    434     errors=errors,
    435     using_cow=using_copy_on_write(),
    436 )

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\internals
 ↪py:363, in BaseBlockManager.apply(self, f, align_keys, **kwargs)
    361         applied = b.apply(f, **kwargs)
    362     else:
--> 363         applied = getattr(b, f)(**kwargs)
    364     result_blocks = extend_blocks(applied, result_blocks)
    366 out = type(self).from_blocks(result_blocks, self.axes)
```

```
File c:
  ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\internals
  ↪py:758, in Block.astype(self, dtype, copy, errors, using_cow, squeeze)
    755         raise ValueError("Can not squeeze with more than one column.")
    756     values = values[0, :]  # type: ignore[call-overload]
--> 758 new_values = astype_array_safe(values, dtype, copy=copy, errors=errors)
    760 new_values = maybe_coerce_values(new_values)
    762 refs = None

File c:
  ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\dtypes\as
  ↪py:237, in astype_array_safe(values, dtype, copy, errors)
    234     dtype = dtype.numpy_dtype
    236 try:
--> 237     new_values = astype_array(values, dtype, copy=copy)
    238 except (ValueError, TypeError):
    239     # e.g. _astype_nansafe can fail on object-dtype of strings
    240     #  trying to convert to float
    241     if errors == "ignore":

File c:
  ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\dtypes\as
  ↪py:182, in astype_array(values, dtype, copy)
    179     values = values.astype(dtype, copy=copy)
    181 else:
--> 182     values = _astype_nansafe(values, dtype, copy=copy)
    184 # in pandas we don't store numpy str dtypes, so convert to object
    185 if isinstance(dtype, np.dtype) and issubclass(values.dtype.type, str):

File c:
  ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\dtypes\as
  ↪py:133, in _astype_nansafe(arr, dtype, copy, skipna)
    129     raise ValueError(msg)
    131 if copy or arr.dtype == object or dtype == object:
    132     # Explicit copy, or required since NumPy can't view from / to object.
--> 133     return arr.astype(dtype, copy=True)
    135 return arr.astype(dtype, copy=copy)

ValueError: could not convert string to float: '3.0M'
```

[ ]:

```
#Serching for 3.0M

data[data['Reviews'] == "3.0M"]
```

[ ]:

```
                                          App Category  Rating Reviews  \
10472  Life Made WI-Fi Touchscreen Photo Frame      1.9    19.0    3.0M
```

```
         Size Installs Type     Price Content Rating             Genres  \
10472  1,000+     Free    0  Everyone            NaN  February 11, 2018

      Last Updated Current Ver Android Ver
10472       1.0.19  4.0 and up         NaN
```

```
[ ]: data['Reviews'] = data['Reviews'].replace('3.0M',3.0) #replace 3.0M to 3.0
```

```
[ ]: data['Reviews'] = data['Reviews'].astype(float) #converting the data type to float
```

```
[ ]: data['Reviews'].dtypes #checking the data type
```

```
[ ]: dtype('float64')
```

```
[ ]: data['Reviews'].mean()
```

```
[ ]: np.float64(444111.9265750392)
```

AVERAGE VALUE OF REVIEWS IS 444111.93

---

13. FIND THE TOTAL NUMBER OF FREE AND PAID APPS

```
[ ]: data.head(1)
```

```
[ ]:                                               App        Category  Rating  \
     0  Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1

        Reviews Size Installs  Type Price Content Rating       Genres  \
     0    159.0  19M  10,000+  Free     0       Everyone  Art & Design

           Last Updated Current Ver   Android Ver
     0  January 7, 2018       1.0.0  4.0.3 and up
```

```
[ ]: data['Type'].value_counts()
```

```
[ ]: Type
     Free     10039
     Paid       800
     0            1
     Name: count, dtype: int64
```

TOTAL NUMBER OF FREE APPS ARE 10039 AND PAID APPS ARE 800

---

14. WHICH APP HAS THE MAXIMUM REVIEWS

```
[ ]: data.columns
```

```
[ ]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
            'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
            'Android Ver'],
           dtype='object')
```

```
[ ]: data[data['Reviews'].max()==data['Reviews']]['App']

     # ==data['Reviews'] - this filters .max() version from original data frame
```

```
[ ]: 2544    Facebook
     Name: App, dtype: object
```

FACEBOOK HAS THE HIGHEST REVIEWS OF 2544

---

15. DISPLAY TOP 5 APPS HAVING HIGHEST REVIEWS

```
[ ]: data.columns
```

```
[ ]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
            'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
            'Android Ver'],
           dtype='object')
```

```
[ ]: index = data['Reviews'].sort_values(ascending=False).head(5).index

     # need indexes to retrieve all information about this maximum reviews
     # lets assign to one variable "index"
```

```
[ ]: data.iloc[index]['App']

     # data.iloc[index] - we have to pass this index to iloc method to retrieve rows␣
      ↪having maximum reviews
     # ['App'] - to filter apps
```

```
[ ]: 2544              Facebook
     3943              Facebook
     381     WhatsApp Messenger
     336     WhatsApp Messenger
     3904    WhatsApp Messenger
     Name: App, dtype: object
```

5 APPS HAVING THE HIGHEST REVIEWS ARE FACEBOOK, AND WHATSAPP MESSENGER

---

16. FIND AVERAGE RATING OF FREE AND PAID APPS

```
[ ]: data.head(2)
```

```
[ ]:
                                                App        Category  Rating  \
0  Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
1                             Coloring book moana  ART_AND_DESIGN     3.9

   Reviews Size  Installs  Type Price Content Rating  \
0    159.0  19M   10,000+  Free     0       Everyone
1    967.0  14M  500,000+  Free     0       Everyone

                     Genres      Last Updated Current Ver   Android Ver
0              Art & Design   January 7, 2018       1.0.0  4.0.3 and up
1  Art & Design;Pretend Play  January 15, 2018       2.0.0  4.0.3 and up
```

```
[ ]: data.groupby('Type')['Rating'].mean()
```

```
[ ]: Type
     0       19.000000
     Free     4.186203
     Paid     4.266615
     Name: Rating, dtype: float64
```

AVERAGE RATING OF FREE APPS ARE 4.186 AND PAID APPS IS 4.266

---

17. DISPLAY TOP 5 APPS HAVING MAXIMUM INSTALLS

```
[ ]: #fisrt we have to preprocess the data and later we have to apply operation
```

```
[ ]: data.head(1)
```

```
[ ]:
                                              App        Category  Rating  \
0  Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1

   Reviews Size Installs  Type Price Content Rating        Genres  \
0    159.0  19M  10,000+  Free     0       Everyone  Art & Design

      Last Updated Current Ver   Android Ver
0  January 7, 2018       1.0.0  4.0.3 and up
```

```
[ ]: data['Installs'].dtypes
```

```
[ ]: dtype('O')
```

```
[187]: # Removing both commas and plus signs, and updating 'Installs_1'
```

```
data['Installs_1'] = data['Installs'].replace({',': '', '\+': ''}, regex=True)
```

```
<>:3: SyntaxWarning: invalid escape sequence '\+'
<>:3: SyntaxWarning: invalid escape sequence '\+'
C:\Users\Ebad\AppData\Local\Temp\ipykernel_5312\2214709168.py:3: SyntaxWarning:
invalid escape sequence '\+'
  data['Installs_1'] = data['Installs'].replace({',': '', '\+': ''}, regex=True)
```

[188]:

```
data.tail(1)
```

[188]:

```
                                                App   Category  Rating  \
10840  iHoroscope - 2018 Daily Horoscope & Astrology  LIFESTYLE     4.5

      Reviews Size     Installs  Type Price Content Rating     Genres  \
10840  398307  19M  10,000,000+  Free     0       Everyone  Lifestyle

        Last Updated         Current Ver         Android Ver Installs_1
10840  July 25, 2018  Varies with device  Varies with device   10000000
```

[189]:

```
data['Installs_1'] = data['Installs_1'].astype('int')
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[189], line 1
----> 1 data['Installs_1'] = data['Installs_1'].astype('int')

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\generic.
 ↪py:6643, in NDFrame.astype(self, dtype, copy, errors)
   6637     results = [
   6638         ser.astype(dtype, copy=copy, errors=errors) for _, ser in self.
 ↪items()
   6639     ]
   6641 else:
   6642     # else, only a single dtype is given
-> 6643     new_data = self._mgr.astype(dtype=dtype, copy=copy, errors=errors)
   6644     res = self._constructor_from_mgr(new_data, axes=new_data.axes)
   6645     return res.__finalize__(self, method="astype")

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\internals
 ↪py:430, in BaseBlockManager.astype(self, dtype, copy, errors)
    427 elif using_copy_on_write():
    428     copy = False
--> 430 return self.apply(
    431     "astype",
    432     dtype=dtype,
    433     copy=copy,
    434     errors=errors,
    435     using_cow=using_copy_on_write(),
    436 )

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\internals
 ↪py:363, in BaseBlockManager.apply(self, f, align_keys, **kwargs)
    361         applied = b.apply(f, **kwargs)
    362     else:
--> 363         applied = getattr(b, f)(**kwargs)
    364     result_blocks = extend_blocks(applied, result_blocks)
    366 out = type(self).from_blocks(result_blocks, self.axes)

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\internals
 ↪py:758, in Block.astype(self, dtype, copy, errors, using_cow, squeeze)
    755         raise ValueError("Can not squeeze with more than one column.")
    756     values = values[0, :]  # type: ignore[call-overload]
--> 758 new_values = astype_array_safe(values, dtype, copy=copy, errors=errors)
    760 new_values = maybe_coerce_values(new_values)
    762 refs = None

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\dtypes\as
 ↪py:237, in astype_array_safe(values, dtype, copy, errors)
    234     dtype = dtype.numpy_dtype
    236 try:
--> 237     new_values = astype_array(values, dtype, copy=copy)
    238 except (ValueError, TypeError):
    239     # e.g. _astype_nansafe can fail on object-dtype of strings
    240     #  trying to convert to float
    241     if errors == "ignore":

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\dtypes\as
 ↪py:182, in astype_array(values, dtype, copy)
    179     values = values.astype(dtype, copy=copy)
    181 else:
--> 182     values = _astype_nansafe(values, dtype, copy=copy)
    184 # in pandas we don't store numpy str dtypes, so convert to object
    185 if isinstance(dtype, np.dtype) and issubclass(values.dtype.type, str):

File c:
 ↪\Users\Ebad\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\dtypes\as
 ↪py:133, in _astype_nansafe(arr, dtype, copy, skipna)
    129     raise ValueError(msg)
    131 if copy or arr.dtype == object or dtype == object:
    132     # Explicit copy, or required since NumPy can't view from / to object.
--> 133     return arr.astype(dtype, copy=True)
```

```
    135 return arr.astype(dtype, copy=copy)

ValueError: invalid literal for int() with base 10: 'Free'
```

[192]:
```
data[data['Installs_1'] == 'Free'] # we have to convert this "Free" inside␣
 ↪installs to 0.
```

[192]:
```
                                       App Category  Rating Reviews  \
10472  Life Made WI-Fi Touchscreen Photo Frame      1.9    19.0    3.0M

         Size Installs Type     Price Content Rating             Genres  \
10472  1,000+     Free    0  Everyone            NaN  February 11, 2018

      Last Updated Current Ver Android Ver Installs_1
10472       1.0.19  4.0 and up         NaN       Free
```

[205]:
```
data.head(2)
```

[205]:
```
                                             App        Category  Rating  \
0  Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
1                            Coloring book moana  ART_AND_DESIGN     3.9

  Reviews Size  Installs  Type Price Content Rating  \
0     159  19M   10,000+  Free     0       Everyone
1     967  14M  500,000+  Free     0       Everyone

                     Genres      Last Updated Current Ver   Android Ver  \
0              Art & Design   January 7, 2018       1.0.0  4.0.3 and up
1  Art & Design;Pretend Play  January 15, 2018       2.0.0  4.0.3 and up

   Installs_1
0       10000
1      500000
```

[193]:
```
data['Installs_1'] = data['Installs_1'].str.replace('Free','0')
```

[195]:
```
data['Installs_1'] = data['Installs_1'].astype('int')
```

[196]:
```
data['Installs_1'].dtypes
```

[196]:
```
dtype('int64')
```

[203]:
```
index = data['Installs_1'].sort_values(ascending=False).head(5).index
```

[204]:
```
data.iloc[index]['App']
```

```
[204]: 5856    Google Play Games
       5395        Google Photos
       2853        Google Photos
       2884        Google Photos
       4170         Google Drive
       Name: App, dtype: object
```

TOP 5 APPS HAVING MAXIMUM INSTALLS ARE GOOGLE PLAY GAMES, GOOGLE PHOTOS, AND GOOGLE DRIVE