In [187]:
```
import pandas as pd
```

In [188]:
```
import matplotlib.pyplot as plt
import numpy as np
```

In [189]:
```
android_df=pd.read_csv("googleplaystore.csv")
```

In [190]:
```
android_df.head()
```

Out[190]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4	159	19M	10,000+	Free	0	Everyone	
1	Coloring book moana	ART_AND_DESIGN	4	967	14M	500,000+	Free	0	Everyone	De
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	5	87510	8.7M	5,000,000+	Free	0	Everyone	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4	215644	25M	50,000,000+	Free	0	Teen	
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4	967	2.8M	100,000+	Free	0	Everyone	Des

In [191]:
```
android_df["Category"].value_counts()
```

Out[191]:
```
FAMILY                 1972
GAME                   1144
TOOLS                   843
MEDICAL                 463
BUSINESS                460
PRODUCTIVITY            424
PERSONALIZATION         392
COMMUNICATION           387
SPORTS                  384
LIFESTYLE               382
FINANCE                 366
HEALTH_AND_FITNESS      341
PHOTOGRAPHY             335
SOCIAL                  295
NEWS_AND_MAGAZINES      283
SHOPPING                260
TRAVEL_AND_LOCAL        258
DATING                  234
BOOKS_AND_REFERENCE     231
VIDEO_PLAYERS           175
EDUCATION               156
ENTERTAINMENT           149
MAPS_AND_NAVIGATION     137
FOOD_AND_DRINK          127
HOUSE_AND_HOME           88
LIBRARIES_AND_DEMO       85
AUTO_AND_VEHICLES        85
WEATHER                  82
ART_AND_DESIGN           65
EVENTS                   64
PARENTING                60
COMICS                   60
BEAUTY                   53
1.9                       1
Name: Category, dtype: int64
```

In [192]:
```
android_df[android_df["Category"] =="1.9"]
```

Out[192]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genre
10472	Life Made WI-Fi Touchscreen Photo Frame	1.9	19	3.0M	1,000+	Free	0	Everyone	NaN	Februa 11, 201

In [193]:
```
android_df[android_df["Category"] =="1.9"
          ].values
```

Out[193]:
```
array([['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]], dtype=object)
```

In [194]:
```
clean_lst=['Life Made WI-Fi Touchscreen Photo Frame','LIFESTYLE                '
        '1,000+', 'Free', '0', 'Everyone', 'LIFESTYLE', 'February 11, 2018',
        '1.0.19', '4.0 and up']
clean_lst
```

Out[194]:
```
['Life Made WI-Fi Touchscreen Photo Frame',
 'LIFESTYLE                ',
 'LIFESTYLE',
 19.0,
 '3.0M',
 '1,000+',
 'Free',
 '0',
 'Everyone',
 'LIFESTYLE',
 'February 11, 2018',
 '1.0.19',
 '4.0 and up']
```

In [195]:
```
android_df[android_df["Category"]=="1.9"] = clean_lst
```

In [196]:
```
android_df["Category"].value_counts()
```

Out[196]:
```
FAMILY                 1972
GAME                   1144
TOOLS                   843
MEDICAL                 463
BUSINESS                460
PRODUCTIVITY            424
PERSONALIZATION         392
COMMUNICATION           387
SPORTS                  384
LIFESTYLE               382
FINANCE                 366
HEALTH_AND_FITNESS      341
PHOTOGRAPHY             335
SOCIAL                  295
NEWS_AND_MAGAZINES      283
SHOPPING                260
TRAVEL_AND_LOCAL        258
DATING                  234
BOOKS_AND_REFERENCE     231
VIDEO_PLAYERS           175
EDUCATION               156
ENTERTAINMENT           149
MAPS_AND_NAVIGATION     137
FOOD_AND_DRINK          127
HOUSE_AND_HOME           88
LIBRARIES_AND_DEMO       85
AUTO_AND_VEHICLES        85
WEATHER                  82
ART_AND_DESIGN           65
EVENTS                   64
PARENTING                60
COMICS                   60
BEAUTY                   53
LIFESTYLE                 1
Name: Category, dtype: int64
```

In [197]:
```
app_count=android_df["App"].value_counts()
```

In [198]:
```
app_count
```

Out[198]:
```
ROBLOX                                               9
CBS Sports App - Scores, News, Stats & Watch Live    8
ESPN                                                 7
Duolingo: Learn Languages Free                       7
Candy Crush Saga                                     7
                                                    ..
Meet U - Get Friends for Snapchat, Kik & Instagram   1
U-Report                                             1
U of I Community Credit Union                        1
Waiting For U Launcher Theme                         1
iHoroscope - 2018 Daily Horoscope & Astrology        1
Name: App, Length: 9660, dtype: int64
```

In [199]:
```
app_count[app_count>1]
```

Out[199]:
```
ROBLOX                                                9
CBS Sports App - Scores, News, Stats & Watch Live     8
ESPN                                                  7
Duolingo: Learn Languages Free                        7
Candy Crush Saga                                      7
                                                     ..
Transenger – Ts Dating and Chat for Free              2
Random Video Chat                                     2
Clover Dating App                                     2
Docs To Go™ Free Office Suite                         2
English Dictionary - Offline                          2
Name: App, Length: 798, dtype: int64
```

In [200]:
```
"Instagram" in app_count[app_count > 1].index
```

Out[200]:
```
True
```

In [201]:
```
"ROBLOX " in app_count[app_count > 1].index
```

Out[201]:
```
False
```

In [202]:
```
android_df[android_df["App"]=="Instagram"]
```

Out[202]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
2545	Instagram	SOCIAL	4	66577313	Varies with device	1,000,000,000+	Free	0	Teen	Socia
2604	Instagram	SOCIAL	4	66577446	Varies with device	1,000,000,000+	Free	0	Teen	Socia
2611	Instagram	SOCIAL	4	66577313	Varies with device	1,000,000,000+	Free	0	Teen	Socia
3909	Instagram	SOCIAL	4	66509917	Varies with device	1,000,000,000+	Free	0	Teen	Socia

In [203]:
```
duplicate_app_df=android_df[android_df.duplicated(subset=["App"],keep=False)]
duplicate_app_df[duplicate_app_df["App"]=="Instagram"]
```

Out[203]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
2545	Instagram	SOCIAL	4	66577313	Varies with device	1,000,000,000+	Free	0	Teen	Socia
2604	Instagram	SOCIAL	4	66577446	Varies with device	1,000,000,000+	Free	0	Teen	Socia
2611	Instagram	SOCIAL	4	66577313	Varies with device	1,000,000,000+	Free	0	Teen	Socia
3909	Instagram	SOCIAL	4	66509917	Varies with device	1,000,000,000+	Free	0	Teen	Socia

In [204]:
```
num_duplicated_app=duplicate_app_df['App'].nunique()
```

In [205]:
```
num_duplicated_app
```

Out[205]: 798

In [206]:
```
duplicate_app_df=android_df[android_df.duplicated(subset=["App"],keep=False)]
num_duplicated_app=duplicate_app_df['App'].nunique()
num_duplicated_app
```

Out[206]: 798

In [207]:
```
duplicate_app_df.shape
```

Out[207]: (1979, 13)

In [208]:
```
android_df.shape
```

Out[208]: (10841, 13)

In [209]:
```
10841-1979
```

Out[209]: 8862

In [210]:
```
review_max=android_df.groupby('App')['Reviews'].max()
```

In [211]:
```
review_max
```

Out[211]:
```
App
"i DT" Fútbol. Todos Somos Técnicos.                    27
+Download 4 Instagram Twitter                         40467
- Free Comics - Comic Apps                              115
.R                                                      259
/u/app                                                  573
                                                       ...
뽕티비 - 개인방송, 인터넷방송, BJ방송                                  414
💎 I'm rich                                              718
💔 WhatsLov: Smileys of love, stickers and GIF         22098
📏 Smart Ruler ⇔ cm/inch measuring for homework!          19
🔥 Football Wallpapers 4K | Full HD Backgrounds 😍      11661
Name: Reviews, Length: 9660, dtype: object
```

In [212]:
```
int(review_max["Instagram"])
```

Out[212]:
```
66577446
```

In [213]:
```
android_clean=[]
already_added=[]
for index,row in android_df.iterrows():
    name=row["App"]
    n_reviews=row["Reviews"]

    if(review_max[name]==n_reviews)and (name not in already_added):
        android_clean.append(row)
        already_added.append(name)
```

In [214]:
```
android_clean=pd.DataFrame(android_clean)
android_clean[android_clean["App"]== "Ebook Reader"]
```

Out[214]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
164	Ebook Reader	BOOKS_AND_REFERENCE	4	85842	37M	5,000,000+	Free	0	Everyone

In [215]:
```
duplicate_app_df= android_clean[android_clean.duplicated(subset=["App"],keep=Fa
                                ]
duplicate_app_df
```

Out[215]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Curren Ve

In [216]:
```
android_clean.shape
```

Out[216]:
```
(9660, 13)
```

In [217]:
```
already_added
```

Out[217]:
```
['Photo Editor & Candy Camera & Grid & ScrapBook',
 'U Launcher Lite – FREE Live Cool Themes, Hide Apps',
 'Sketch - Draw & Paint',
 'Pixel Draw - Number Art Coloring Book',
 'Paper flowers instructions',
 'Smoke Effect Photo Maker - Smoke Editor',
 'Infinite Painter',
 'Garden Coloring Book',
 'Kids Paint Free - Drawing Fun',
 'Text on Photo - Fonteee',
 'Name Art Photo Editor - Focus n Filters',
 'Tattoo Name On My Photo Editor',
 'Mandala Coloring Book',
 '3D Color Pixel by Number - Sandbox Art Coloring',
 'Learn To Draw Kawaii Characters',
 'Photo Designer - Write your name with shapes',
 '350 Diy Room Decor Ideas',
 'FlipaClip - Cartoon animation',
 'ibis Paint X',
```

In [218]:
```
def is_english1(app_name):
    lst=[]
    for i in app_name:
        if ord(i)>127:
            lst.append(False)
        else:
            lst.append(True)
            check=set(lst)
            if False in check:
                return False
            else:
                return True
```

In [219]:
```
is_english1("Instagram")
```

Out[219]:
```
True
```

In [220]:
```
def is_english1(app_name):
    lst = []
    for i in app_name:
        if ord(i) > 127:
            lst.append(False)
        else:
            lst.append(True)
            non_ascii = 0
            for j in lst:
                non_ascii += 1
                if non_ascii > 3:
                    return False
                else:
                    return True
```

In [221]:
```
is_english1('INSTAGRAM 😊😊😊😊')
```

Out[221]:
```
True
```

In [222]:
```
android_english1 = android_clean[android_clean['App'].apply(is_english1)]
```

In [223]:
```
android_english1
```

Out[223]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4	159	19M	10,000+	Free	0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	5	87510	8.7M	5,000,000+	Free	0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4	215644	25M	50,000,000+	Free	0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4	967	2.8M	100,000+	Free	0
5	Paper flowers instructions	ART_AND_DESIGN	4	167	5.6M	50,000+	Free	0
...	...	...	...	...	...	...	...	...
10836	Sya9a Maroc - FR	FAMILY	4	38	53M	5,000+	Free	0
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5	4	3.6M	100+	Free	0
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4	114	Varies with device	1,000+	Free	0
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4	398307	19M	10,000,000+	Free	0

9660 rows × 13 columns

In [224]:
```
android_english1.head()
```

Out[224]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4	159	19M	10,000+	Free	0	Everyone	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	5	87510	8.7M	5,000,000+	Free	0	Everyone	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4	215644	25M	50,000,000+	Free	0	Teen	
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4	967	2.8M	100,000+	Free	0	Everyone	Des
5	Paper flowers instructions	ART_AND_DESIGN	4	167	5.6M	50,000+	Free	0	Everyone	

In [225]:
```
android_english1['Price'].unique()
```

Out[225]:
```
array(['0', '$4.99', '$3.99', '$6.99', '$1.49', '$2.99', '$7.99', '$5.99',
       '$3.49', '$1.99', '$9.99', '$7.49', '$0.99', '$9.00', '$5.49',
       '$10.00', '$11.99', '$79.99', '$16.99', '$14.99', '$1.00',
       '$29.99', '$12.99', '$2.49', '$24.99', '$10.99', '$1.50', '$19.99',
       '$15.99', '$33.99', '$74.99', '$39.99', '$3.95', '$4.49', '$1.70',
       '$8.99', '$2.00', '$3.88', '$25.99', '$399.99', '$17.99',
       '$400.00', '$3.02', '$1.76', '$4.84', '$4.77', '$1.61', '$2.50',
       '$1.59', '$6.49', '$1.29', '$5.00', '$13.99', '$299.99', '$379.99',
       '$37.99', '$18.99', '$389.99', '$19.90', '$8.49', '$1.75',
       '$14.00', '$4.85', '$46.99', '$109.99', '$154.99', '$3.08',
       '$2.59', '$4.80', '$1.96', '$19.40', '$3.90', '$4.59', '$15.46',
       '$3.04', '$4.29', '$2.60', '$3.28', '$4.60', '$28.99', '$2.95',
       '$2.90', '$1.97', '$200.00', '$89.99', '$2.56', '$30.99', '$3.61',
       '$394.99', '$1.26', '$1.20', '$1.04'], dtype=object)
```

In [226]:
```
android_final=android_english1[android_english1['Price']=="0"]
```

In [227]:
```
android_final.shape
```

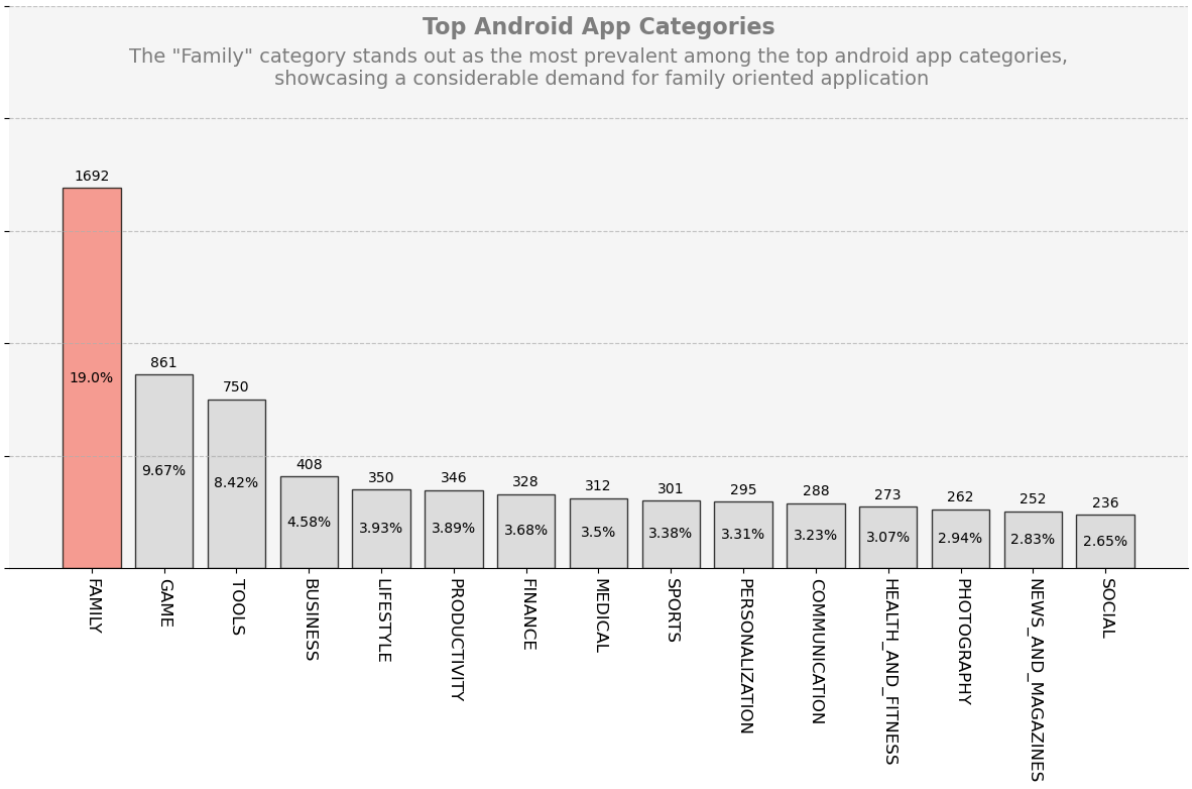
Out[227]:
```
(8904, 13)
```

In [228]:
```
android_final["Category"].value_counts(normalize= True)*100
```

Out[228]:
```
FAMILY                 19
GAME                   10
TOOLS                   8
BUSINESS                5
LIFESTYLE               4
PRODUCTIVITY            4
FINANCE                 4
MEDICAL                 4
SPORTS                  3
PERSONALIZATION         3
COMMUNICATION           3
HEALTH_AND_FITNESS      3
PHOTOGRAPHY             3
NEWS_AND_MAGAZINES      3
SOCIAL                  3
TRAVEL_AND_LOCAL        2
SHOPPING                2
BOOKS_AND_REFERENCE     2
DATING                  2
VIDEO_PLAYERS           2
MAPS_AND_NAVIGATION     1
FOOD_AND_DRINK          1
EDUCATION               1
ENTERTAINMENT           1
LIBRARIES_AND_DEMO      1
AUTO_AND_VEHICLES       1
HOUSE_AND_HOME          1
WEATHER                 1
EVENTS                  1
ART_AND_DESIGN          1
PARENTING               1
COMICS                  1
BEAUTY                  1
LIFESTYLE               0
Name: Category, dtype: float64
```

In [258]:

```
import matplotlib.pyplot as plt

categories = android_final["Category"].value_counts().index[:15]
counts = android_final["Category"].value_counts().values[:15]
percentage = round(android_final["Category"].value_counts(normalize=True) * 100

# Create a stylish bar chart
plt.figure(figsize=(12, 8))
bars = plt.bar(categories, counts, color='lightgray', alpha=0.75, edgecolor='bl
plt.xticks(rotation=-90, fontsize=12)
plt.yticks(fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.grid(axis='x', linestyle="")
plt.xticks(fontsize=12) # Customized tick labels
plt.yticks(range(0, 3000, 500), [], fontsize=12) # Customized tick labels

# Find the category with the highest count
max_count_category = categories[counts.argmax()]

# Highlight the bar for the category with the highest count
max_count_index = list(categories).index(max_count_category)
bars[max_count_index].set_color("salmon")
bars[max_count_index].set_edgecolor('black')

# Adding data labels and percentages inside each bar
for bar, perc in zip(bars, percentage):

    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height + 20, "%d" % int(height)
    plt.text(bar.get_x() + bar.get_width() / 2, height / 2, f'{perc}%', ha="cer

# Adding a background color
ax = plt.gca()
ax.set_facecolor("#f7f7f7")

# Adding CHART TITLE inside the chart
plt.text(0.5 , 0.95, 'Top Android App Categories', horizontalalignment='center'

# Adding a text annotation
plt.text(0.5, 0.86, 'The "Family" category stands out as the most prevalent amo
         horizontalalignment="center", fontsize=14, transform=plt.gca().transAx

# Removing spines
for i in ["top", "right", "left"]:
    plt.gca().spines[i].set_visible(False)

plt.tight_layout()
plt.savefig('bar.png')
plt.show()
```



In [230]:
```
android_final["Installs"].value_counts(normalize=True)*100
```

Out[230]:
```
1,000,000+       16
100,000+         12
10,000,000+      10
10,000+          10
1,000+            8
100+              7
5,000,000+        7
500,000+          6
50,000+           5
5,000+            5
10+               4
500+              3
50,000,000+       2
100,000,000+      2
50+               2
5+                1
1+                1
500,000,000+      0
1,000,000,000+    0
0+                0
0                 0
Name: Installs, dtype: float64
```

In [231]:
```
android_final["Installs_int"]= android_final["Installs"].str.replace(",","").st
```

```
C:\Users\Powad\AppData\Local\Temp\ipykernel_7712\2898442273.py:1: FutureWarni
ng: The default value of regex will change from True to False in a future ver
sion. In addition, single character regular expressions will *not* be treated
as literal strings when regex=True.
  android_final["Installs_int"]= android_final["Installs"].str.replace
(",","").str.replace("+","").astype(int)
C:\Users\Powad\AppData\Local\Temp\ipykernel_7712\2898442273.py:1: SettingWith
CopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  android_final["Installs_int"]= android_final["Installs"].str.replace
(",","").str.replace("+","").astype(int)
```

In [232]:
```
android_final["Installs_int"]
```

Out[232]:
```
0           10000
2         5000000
3        50000000
4          100000
5           50000
           ...
10836        5000
10837         100
10838        1000
10839        1000
10840    10000000
Name: Installs_int, Length: 8904, dtype: int32
```

In [233]:
```
install_frq=android_final["Installs_int"].value_counts().sort_index()
install_frq=install_frq[install_frq.index >500]
install_frq
```

Out[233]:
```
1000           751
5000           403
10000          913
50000          429
100000        1031
500000         494
1000000       1398
5000000        608
10000000       932
50000000       203
100000000      188
500000000       24
1000000000      20
Name: Installs_int, dtype: int64
```

In [234]:
```
install_frq=round(android_final["Installs_int"].value_counts(normalize=True)*10
install_frq=install_frq[install_frq.index >500]
install_frq
```

Out[234]:
```
1000          8
5000          5
10000        10
50000         5
100000       12
500000        6
1000000      16
5000000       7
10000000     10
50000000      2
100000000     2
500000000     0
1000000000    0
Name: Installs_int, dtype: float64
```

In [235]:
```
def alph_unit(value):
    if value >= 1e9:
        return f'{value / 1e9:.0f}B'
    elif value >= 1e6:
        return f'{value / 1e6:.0f}M'
    elif value >= 1e3:
        return f'{value / 1e3:.0f}K'
    else:
        return f'{value:.0f}'
```

In [236]:
```
len(install_frq.index)
```

Out[236]:
```
13
```

In [237]:
```
install_frq.index=install_frq.index.map(alph_unit)
install_frq
```

Out[237]:
```
1K      8
5K      5
10K    10
50K     5
100K   12
500K    6
1M     16
5M      7
10M    10
50M     2
100M    2
500M    0
1B      0
Name: Installs_int, dtype: float64
```

In [ ]:

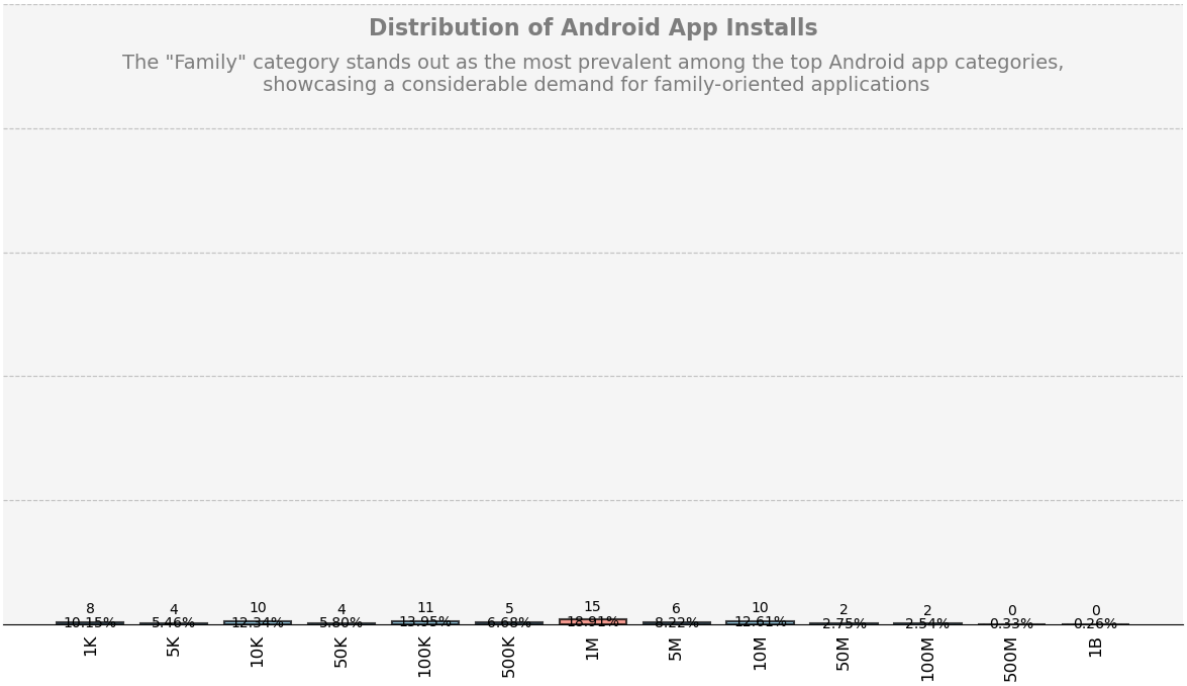
In [ ]:

In [ ]:

In [ ]:

In [238]:

```
categories = install_frq.index
counts = install_frq.values
percentage1 = counts / counts.sum() * 100  # Calculating percentages

# Define a factor to increase bar height
bar_height_factor = 1.2

# Create stylish bar chart
plt.figure(figsize=(12, 7))
bars = plt.bar(categories, counts * bar_height_factor, color='skyblue', alpha=0
plt.xticks(rotation=90, fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.grid(axis='x', linestyle='')
plt.xticks(fontsize=12)
plt.yticks(range(0, 3000, 500), [], fontsize=12)
plt.tick_params(bottom=0, left=0)

# Find the category with the highest count
max_count_category = categories[counts.argmax()]

# Highlight the bar for the category with the highest count
max_count_index = list(categories).index(max_count_category)
bars[max_count_index].set_color("salmon")
bars[max_count_index].set_edgecolor('black')

# Adding data labels and percentages inside each bar
for bar, perc in zip(bars, percentage1):
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height + 20 * bar_height_factor
    plt.text(bar.get_x() + bar.get_width() / 2, height / 2, f'{perc:.2f}%', ha=

# Adding a background color
ax = plt.gca()
ax.set_facecolor("#f7f7f7")

# Adding CHART TITLE inside the chart
plt.text(0.5, 0.95, 'Distribution of Android App Installs', horizontalalignment

# Adding a text annotation
plt.text(0.5, 0.86, 'The "Family" category stands out as the most prevalent amo
         horizontalalignment="center", fontsize=14, transform=plt.gca().transAx

# Removing spines
for i in ["top", "right", "left"]:
    plt.gca().spines[i].set_visible(False)

plt.tight_layout()
plt.show()
```



In [239]:
```
categories_android= android_final["Category"].unique()
categories_android
```

Out[239]:
```
array(['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',
       'LIFESTYLE                ''], dtype=object)
```

In [240]:
```
android_final
```

Out[240]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4	159	19M	10,000+	Free	0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	5	87510	8.7M	5,000,000+	Free	0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4	215644	25M	50,000,000+	Free	0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4	967	2.8M	100,000+	Free	0
5	Paper flowers instructions	ART_AND_DESIGN	4	167	5.6M	50,000+	Free	0
...	...	...	...	...	...	...	...	...
10836	Sya9a Maroc - FR	FAMILY	4	38	53M	5,000+	Free	0
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5	4	3.6M	100+	Free	0
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4	114	Varies with device	1,000+	Free	0
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4	398307	19M	10,000,000+	Free	0

8904 rows × 14 columns

In [241]:
```
pd.pivot_table(android_final,values = "Installs_int",index = "Category",aggfunc
```

Out[241]:

Category	Installs_int
ART_AND_DESIGN	1952105
AUTO_AND_VEHICLES	647318
BEAUTY	513152
BOOKS_AND_REFERENCE	8587352
BUSINESS	1708216
COMICS	803235
COMMUNICATION	38322626
DATING	854029
EDUCATION	1812857
ENTERTAINMENT	11640706
EVENTS	253542
FAMILY	3667548
FINANCE	1387692
FOOD_AND_DRINK	1924898
GAME	15524883
HEALTH_AND_FITNESS	4188822
HOUSE_AND_HOME	1331541
LIBRARIES_AND_DEMO	638504
LIFESTYLE	1436127
LIFESTYLE	1000
MAPS_AND_NAVIGATION	3993340
MEDICAL	120616
NEWS_AND_MAGAZINES	9401636
PARENTING	542604
PERSONALIZATION	5183851
PHOTOGRAPHY	17737668
PRODUCTIVITY	16738958
SHOPPING	7001693
SOCIAL	23253652
SPORTS	3638640
TOOLS	10668059
TRAVEL_AND_LOCAL	13984078
VIDEO_PLAYERS	24573948
WEATHER	5074486

In [242]:
```
pd.options.display.float_format = '{:.0f}'.format
```

In [243]:
```
categories_installs = pd.pivot_table(android_final, values="Installs_int", inde
categories_installs = categories_installs.sort_values(by="Installs_int", ascend
```

In [244]:
```
categories_installs
```

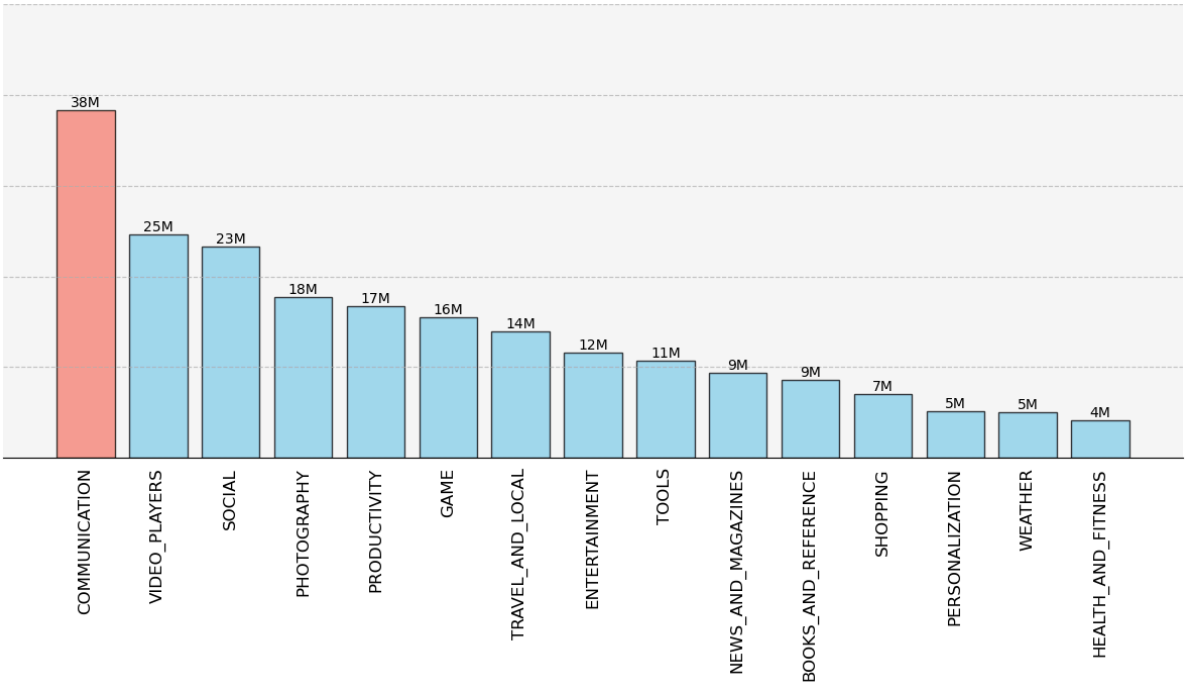
Out[244]:
```
Category
COMMUNICATION               38322626
VIDEO_PLAYERS               24573948
SOCIAL                      23253652
PHOTOGRAPHY                 17737668
PRODUCTIVITY                16738958
GAME                        15524883
TRAVEL_AND_LOCAL            13984078
ENTERTAINMENT               11640706
TOOLS                       10668059
NEWS_AND_MAGAZINES           9401636
BOOKS_AND_REFERENCE          8587352
SHOPPING                     7001693
PERSONALIZATION              5183851
WEATHER                      5074486
HEALTH_AND_FITNESS           4188822
MAPS_AND_NAVIGATION          3993340
FAMILY                       3667548
SPORTS                       3638640
ART_AND_DESIGN               1952105
FOOD_AND_DRINK               1924898
EDUCATION                    1812857
BUSINESS                     1708216
LIFESTYLE                    1436127
FINANCE                      1387692
HOUSE_AND_HOME               1331541
DATING                        854029
COMICS                        803235
AUTO_AND_VEHICLES             647318
LIBRARIES_AND_DEMO            638504
PARENTING                     542604
BEAUTY                        513152
EVENTS                        253542
MEDICAL                       120616
LIFESTYLE                       1000
Name: Installs_int, dtype: float64
```

In [245]:
```
categories_installs_unit= categories_installs.map(alph_unit)
categories_installs_unit
```

Out[245]:
```
Category
COMMUNICATION                38M
VIDEO_PLAYERS                25M
SOCIAL                       23M
PHOTOGRAPHY                  18M
PRODUCTIVITY                 17M
GAME                         16M
TRAVEL_AND_LOCAL             14M
ENTERTAINMENT                12M
TOOLS                        11M
NEWS_AND_MAGAZINES            9M
BOOKS_AND_REFERENCE           9M
SHOPPING                      7M
PERSONALIZATION               5M
WEATHER                       5M
HEALTH_AND_FITNESS            4M
MAPS_AND_NAVIGATION           4M
FAMILY                        4M
SPORTS                        4M
ART_AND_DESIGN                2M
FOOD_AND_DRINK                2M
EDUCATION                     2M
BUSINESS                      2M
LIFESTYLE                     1M
FINANCE                       1M
HOUSE_AND_HOME                1M
DATING                      854K
COMICS                      803K
AUTO_AND_VEHICLES           647K
LIBRARIES_AND_DEMO          639K
PARENTING                   543K
BEAUTY                      513K
EVENTS                      254K
MEDICAL                     121K
LIFESTYLE                     1K
Name: Installs_int, dtype: object
```

In [259]:

```
categoreis = categories_installs.index[:15]
counts = categories_installs.values[:15]

# Create stylish bar chart
plt.figure(figsize=(12, 7))
barss = plt.bar(categoreis, counts, color='skyblue', alpha=0.75, edgecolor='bla

plt.xticks(rotation=90, fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.grid(axis='x', linestyle='')
plt.xticks(fontsize=12)
plt.yticks(range(0, 60000000, 10000000), [], fontsize=12)
plt.tick_params(bottom=0, left=0)

# Find the category with the highest count
max_count_category = categoreis[counts.argmax()]

# Highlight the bar for the category with the highest count
max_count_index = list(categoreis).index(max_count_category)
barss[max_count_index].set_color("salmon")
barss[max_count_index].set_edgecolor('black')

# Adding data labels and percentages inside each bar
for bar, unit in zip(barss, categories_installs_unit.values):
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height + 25, unit, ha="center",


# Adding a background color
ax = plt.gca()
ax.set_facecolor("#f7f7f7")

# Adding CHART TITLE inside the chart
##plt.text(0.5, 0.95, 'Average Distribution of Android App Installs by Category

# Adding a text annotation
##plt.text(0.5, 0.8, '"COMMUNICATION" apps lead the pack with a staggering 38.5

# Removing spines
for i in ["top", "right", "left"]:
    plt.gca().spines[i].set_visible(False)

plt.tight_layout()
plt.savefig('bar2.png')
plt.show()
```



In [247]:
```
category_group = android_final.groupby("Category")
communication = category_group.get_group("COMMUNICATION").sort_values(by="Insta
communication.head()
```

Out[247]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Conter Ratin
336	WhatsApp Messenger	COMMUNICATION	4	69119316	Varies with device	1,000,000,000+	Free	0	Everyon
391	Skype - free IM & video calls	COMMUNICATION	4	10484169	Varies with device	1,000,000,000+	Free	0	Everyon
411	Google Chrome: Fast & Secure	COMMUNICATION	4	9643041	Varies with device	1,000,000,000+	Free	0	Everyon
382	Messenger – Text and Video Chat for Free	COMMUNICATION	4	56646578	Varies with device	1,000,000,000+	Free	0	Everyon
451	Gmail	COMMUNICATION	4	4604483	Varies with device	1,000,000,000+	Free	0	Everyon

In [248]:
```
def alph_unit(value):
    if value >= 1e9:
        return f'{value / 1e9:.0f}B'
    elif value >= 1e6:
        return f'{value / 1e6:.0f}M'
    elif value >= 1e3:
        return f'{value / 1e3:.0f}K'
    else:
        return f'{value:.0f}'
```

In [249]:
```
categories_installs.index[:15]
```

Out[249]:
```
Index(['COMMUNICATION', 'VIDEO_PLAYERS', 'SOCIAL', 'PHOTOGRAPHY',
       'PRODUCTIVITY', 'GAME', 'TRAVEL_AND_LOCAL', 'ENTERTAINMENT', 'TOOLS',
       'NEWS_AND_MAGAZINES', 'BOOKS_AND_REFERENCE', 'SHOPPING',
       'PERSONALIZATION', 'WEATHER', 'HEALTH_AND_FITNESS'],
      dtype='object', name='Category')
```

In [250]:
```
communication = category_group.get_group("COMMUNICATION").sort_values(by="Insta
df = communication[["App", "Installs_int"]].head(15)
df["Installs_int_unit"] = df["Installs_int"].map(alph_unit)
df
```

Out[250]:

	App	Installs_int	Installs_int_unit
336	WhatsApp Messenger	1000000000	1B
391	Skype - free IM & video calls	1000000000	1B
411	Google Chrome: Fast & Secure	1000000000	1B
382	Messenger – Text and Video Chat for Free	1000000000	1B
451	Gmail	1000000000	1B
464	Hangouts	1000000000	1B
403	LINE: Free Calls & Messages	500000000	500M
383	imo free video calls and chat	500000000	500M
4676	Viber Messenger	500000000	500M
420	UC Browser - Fast Download Private & Secure	500000000	500M
371	Google Duo - High Quality Video Calls	500000000	500M
395	GO SMS Pro - Messenger, Free Themes, Emoji	100000000	100M
412	Firefox Browser fast & private	100000000	100M
369	Android Messages	100000000	100M
4106	Messenger Lite: Free Calls & Messages	100000000	100M

In [251]:
```
communication = category_group.get_group("VIDEO_PLAYERS").sort_values(by="Insta
df = communication[["App", "Installs_int"]].head(15)
df["Installs_int_unit"] = df["Installs_int"].map(alph_unit)
df
```

Out[251]:

	App	Installs_int	Installs_int_unit
3665	YouTube	1000000000	1B
3687	Google Play Movies & TV	1000000000	1B
3711	MX Player	500000000	500M
4696	VideoShow-Video Editor, Video Maker, Beauty Ca...	100000000	100M
4688	VivaVideo - Video Editor & Photo Movie	100000000	100M
3672	Motorola Gallery	100000000	100M
3675	VLC for Android	100000000	100M
4032	Dubsmash	100000000	100M
10647	Motorola FM Radio	100000000	100M
5612	Ringdroid	50000000	50M
4685	HD Video Downloader : 2018 Best video mate	50000000	50M
4049	KineMaster – Pro Video Editor	50000000	50M
3686	Vigo Video	50000000	50M
3693	LIKE – Magic Video Maker & Community	50000000	50M
3691	Samsung Video Library	50000000	50M

In [252]:
```
communication = category_group.get_group("SOCIAL").sort_values(by="Installs_int
df = communication[["App", "Installs_int"]].head(15)
df["Installs_int_unit"] = df["Installs_int"].map(alph_unit)
df
```

Out[252]:

	App	Installs_int	Installs_int_unit
2544	Facebook	1000000000	1B
2554	Google+	1000000000	1B
2604	Instagram	1000000000	1B
2610	Snapchat	500000000	500M
2546	Facebook Lite	500000000	500M
3945	Tik Tok - including musical.ly	100000000	100M
2592	Tango - Live Video Broadcast	100000000	100M
6373	VK	100000000	100M
2552	Pinterest	100000000	100M
3951	BIGO LIVE - Live Stream	100000000	100M
2621	LinkedIn	100000000	100M
2548	Tumblr	100000000	100M
2588	Badoo - Free Chat & Dating App	100000000	100M
2636	Zello PTT Walkie Talkie	50000000	50M
2595	ooVoo Video Calls, Messaging & Stories	50000000	50M

In [253]:
```
communication = category_group.get_group("PHOTOGRAPHY").sort_values(by="Install
df = communication[["App", "Installs_int"]].head(15)
df["Installs_int_unit"] = df["Installs_int"].map(alph_unit)
df
```

Out[253]:

	App	Installs_int	Installs_int_unit
2884	Google Photos	1000000000	1B
2908	Retrica	100000000	100M
2921	Photo Editor Pro	100000000	100M
2945	YouCam Perfect - Selfie Photo Editor	100000000	100M
2942	Z Camera - Photo Editor, Beauty Selfie, Collage	100000000	100M
2949	Camera360: Selfie Photo Editor with Funny Sticker	100000000	100M
2847	Sweet Selfie - selfie camera, beauty cam, phot...	100000000	100M
2939	Photo Collage Editor	100000000	100M
2938	PicsArt Photo Studio: Collage Maker & Pic Editor	100000000	100M
2937	BeautyPlus - Easy Photo Editor & Selfie Camera	100000000	100M
4574	S Photo Editor - Collage Maker , Photo Collage	100000000	100M
5057	AR effect	100000000	100M
2833	YouCam Makeup - Magic Selfie Makeovers	100000000	100M
2943	PhotoGrid: Video & Pic Collage Maker, Photo Ed...	100000000	100M
8082	Cymera Camera- Photo Editor, Filter,Collage,La...	100000000	100M

In [ ]:

In [ ]:

In [ ]: