```
#level 1 task3
#segment analysis
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.graph_objs as go
import plotly.express as px
import seaborn as sns

# Load the CSV file into a pandas DataFrame
data_duplicates = pd.read_csv('C:/Users/rahul/OneDrive/Desktop/python/task-3/apps.csv')

# Display the first few rows of the DataFrame
print("first few rows")
print(data_duplicates.head())
apps = data_duplicates.drop_duplicates()
print("Total numbers of Apps : ", apps[ 'App' ].count())
apps.head()
output –
nalysis.py
  Unnamed: 0  ...   Android Ver
0          0  ...  4.0.3 and up
1          1  ...  4.0.3 and up
2          2  ...  4.0.3 and up
3          3  ...    4.2 and up
4          4  ...    4.4 and up

[5 rows x 14 columns
# Display the first few rows of the DataFrame
print("first few rows")
print(data_duplicates.head())
```

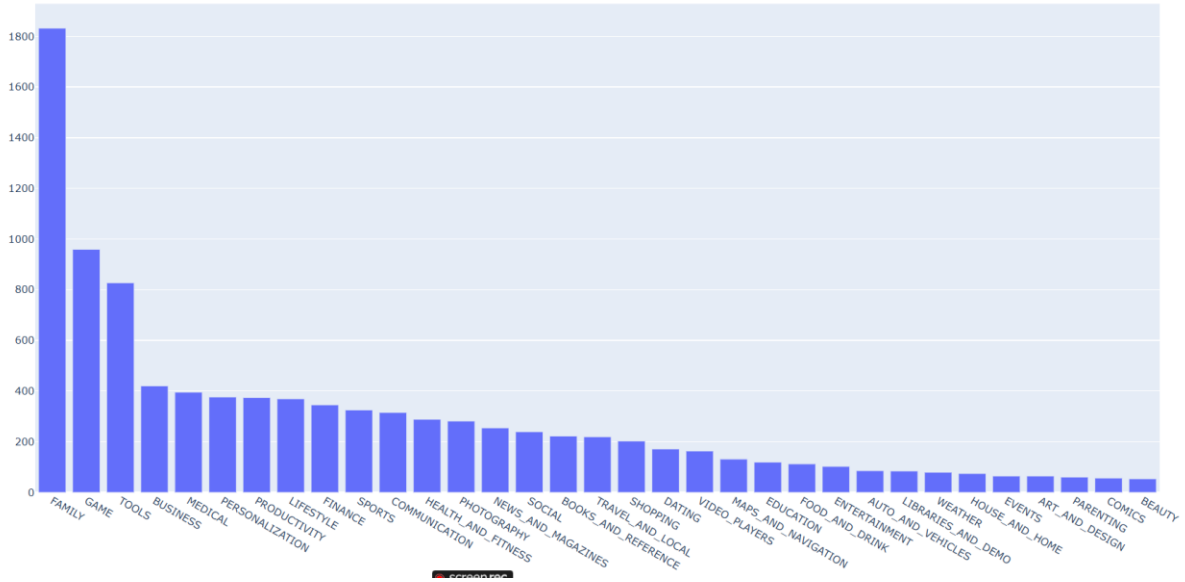
```
apps = data_duplicates.drop_duplicates()
print("Total numbers of Apps : ", apps[ 'App' ].count())
apps.head()]
```

output –

```
Total numbers of Apps :  9659
```

```
ch_remove = ('+' , '$' , 'M' , ',' )
col_clean = ('Size','Installs','Price')
for col in col_clean:
    for char in ch_remove:
        apps[col] = apps[col].str.replace(char, '')
    apps[col] = pd.to_numeric(apps[col])
n_categories = len(set(apps['Category']))
print('Number of categories = ', n_categories)
n_apps_in_category = apps['Category'].value_counts().sort_values(ascending=False)
data1 = go.Bar(
    x = n_apps_in_category.index,
    y = n_apps_in_category.values
)
fig = go.Figure(data1)
fig.show()
```

output—


1800
1600
1400
1200
1000
800
600
400
200
0
FAMILY
GAME
TOOLS
BUSINESS
MEDICAL
PERSONALIZATION
PRODUCTIVITY
LIFESTYLE
FINANCE
SPORTS
COMMUNICATION
HEALTH_AND_FITNESS
PHOTOGRAPHY
NEWS_AND_MAGAZINES
SOCIAL
BOOKS_AND_REFERENCE
TRAVEL_AND_LOCAL
SHOPPING
DATING
VIDEO_PLAYERS
MAPS_AND_NAVIGATION
EDUCATION
FOOD_AND_DRINK
ENTERTAINMENT
AUTO_AND_VEHICLES
LIBRARIES_AND_DEMO
WEATHER
HOUSE_AND_HOME
EVENTS
ART_AND_DESIGN
PARENTING
COMICS
BEAUTY