Dataset: https://drive.google.com/file/d/1FuQr0wM8tczVQvtO8_Nn82C3E4gQTgYF/view

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings

warnings.filterwarnings("ignore")

%matplotlib inline


df=pd.read_csv("googleplaystore.csv")

df[~df['Reviews'].str.isnumeric()]

df_copy=df.copy()

df_copy=df_copy.drop(df_copy.index[10472])

df_copy["Reviews"]=df_copy["Reviews"].astype('int')

df_copy["Size"]=df_copy["Size"].str.replace('M','000')

df_copy["Size"]=df_copy["Size"].str.replace("Varies with device", str(np.nan))

for i in df_copy['Size']:
    if i < 10:
        df_copy['Size']=df_copy['Size'].replace(i,i*1000)


chars_to_remove=['+',',','$']
cols_to_clean=['Installs',"Price"]
for item in chars_to_remove:
    for col in cols_to_clean:
        df_copy[col]=df_copy[col].str.replace(item,'')

df_copy["Installs"]=df_copy["Installs"].astype('int')

df_copy["Price"]=df_copy["Price"].astype('float')

df_copy["Last Updated"]=pd.to_datetime(df_copy["Last Updated"])

df_copy["day"]=df_copy["Last Updated"].dt.day
df_copy["month"]=df_copy["Last Updated"].dt.month
df_copy["year"]=df_copy["Last Updated"].dt.year

df_copy.to_csv("google_cleaned.csv",index=False)
```

pd.read_csv("google_cleaned.csv").info()

https://drive.google.com/file/d/1RqYkSLY1aw0hc2Jmvdmiju-0hYBFhmbW/view?usp=sharing