

#Importing Libraries

```
import pandas as pd
import numpy as np
import datetime
```

Loading CSV File

```
data = pd.read_csv('/content/googleplaystore(impure).csv')
```

Decoding Data Information

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 14 columns):
#   Column              Non-Null Count  Dtype
---  -
0    Unnamed: 0           10841 non-null  int64
1    App                  10841 non-null  object
2    Category             10841 non-null  object
3    Rating               9367 non-null   float64
4    Reviews              10841 non-null  object
5    Size                 10841 non-null  object
6    Installs              10841 non-null  object
7    Type                 10840 non-null  object
8    Price                10841 non-null  object
9    Content Rating       10840 non-null  object
10   Genres                10841 non-null  object
11   Last Updated         10841 non-null  object
12   Current Ver          10833 non-null  object
13   Android Ver          10838 non-null  object
dtypes: float64(1), int64(1), object(12)
memory usage: 1.2+ MB
```

Checking For Null Values

```
data.isnull().sum()
```

	0
Unnamed: 0	0
App	0
Category	0
Rating	1474
Reviews	0
Size	0
Installs	0
Type	1
Price	0
Content Rating	1
Genres	0
Last Updated	0
Current Ver	8
Android Ver	3

dtype: int64

Removing Null Values

```
data.dropna(inplace=True)
```

data

	Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
0	0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone		Art & Design	7-Jan-18	1.0.0	4.0.3 and up
1	1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play		15-Jan-18	2.0.0	4.0.3 and up
2	2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design		1-Aug-18	1.2.4	4.0.3 and up
3	3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design		8-Jun-18	Varies with device	4.2 and up
4	4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity		20-Jun-18	1.1	4.4 and up
...
10834	10834	FR Calculator	FAMILY	4.0	7	2.6M	500+	Free	0	Everyone	Education		18-Jun-17	1.0.0	4.1 and up
10836	10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education		25-Jul-17	1.48	4.1 and up
10837	10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education		6-Jul-18	1	4.1 and up
10839	10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference		19-Jan-15	Varies with device	Varies with device
10840	10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle		25-Jul-18	Varies with device	Varies with device

9360 rows x 14 columns

Dropping Unnecessary Columns

```
data.drop(columns='Unnamed: 0', inplace=True)
```

```
# Changing Data Types
```

```
data['Last Updated'] = pd.to_datetime(data['Last Updated'], errors='coerce')
```

```
/tmp/ipython-input-3556908263.py:3: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.  
data['Last Updated'] = pd.to_datetim(data['Last Updated'], errors='coerce')
```

```
data['Last Updated'] = data['Last Updated'].dt.strftime('%Y-%m-%d')
```

```
data['Last Updated'] = pd.to_datetime(data['Last Updated'], errors='coerce')
```

```
# Cleaning Dirty Columns
```

```
data['Installs'] = data['Installs'].str.replace('+', '')  
data['Installs'] = data['Installs'].str.replace(',', '')  
data['Installs'] = data['Installs'].astype(int)
```

```
data.drop(data[data['Reviews']=='3.0M'].index,inplace=True)
```

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

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Index: 9360 entries, 0 to 10840  
Data columns (total 13 columns):  
#   Column          Non-Null Count  Dtype  
---  -  
0   App              9360 non-null   object  
1   Category         9360 non-null   object  
2   Rating           9360 non-null   float64  
3   Reviews          9360 non-null   int64  
4   Size             9360 non-null   object  
5   Installs         9360 non-null   int64  
6   Type             9360 non-null   object  
7   Price            9360 non-null   object  
8   Content Rating   9360 non-null   object  
9   Genres           9360 non-null   object  
10  Last Updated     9360 non-null   datetime64[ns]  
11  Current Ver      9360 non-null   object  
12  Android Ver      9360 non-null   object  
dtypes: datetime64[ns](1), float64(1), int64(2), object(9)  
memory usage: 1023.8+ KB
```

```
# Cleaning Price Column
```

```
data['Price'] = data['Price'].str.replace('$','')  
data['Price'] = data['Price'].astype(float)
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Index: 9360 entries, 0 to 10840  
Data columns (total 13 columns):  
#   Column          Non-Null Count  Dtype  
---  -  
0   App              9360 non-null   object  
1   Category         9360 non-null   object  
2   Rating           9360 non-null   float64  
3   Reviews          9360 non-null   int64  
4   Size             9360 non-null   object  
5   Installs         9360 non-null   int64  
6   Type             9360 non-null   object  
7   Price            9360 non-null   float64  
8   Content Rating   9360 non-null   object  
9   Genres           9360 non-null   object  
10  Last Updated     9360 non-null   datetime64[ns]  
11  Current Ver      9360 non-null   object  
12  Android Ver      9360 non-null   object  
dtypes: datetime64[ns](1), float64(2), int64(2), object(8)  
memory usage: 1023.8+ KB
```

```
data.isnull().sum()
```

	0
App	0
Category	0
Rating	0
Reviews	0
Size	0
Installs	0
Type	0
Price	0
Content Rating	0
Genres	0
Last Updated	0
Current Ver	0
Android Ver	0

```
dtype: int64
```

```
# Coverting Clean Data to CSV Format
```

```
data.to_csv('appstore.csv', index=False)
```

```
cleaned_data = pd.read_csv('/content/appstore.csv')
```

```
cleaned_data
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	0.0	Everyone		Art & Design	2018-01-07	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0.0	Everyone	Art & Design;Pretend Play		2018-01-15	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	0.0	Everyone		Art & Design	2018-08-01	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0.0	Teen		Art & Design	2018-06-08	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0.0	Everyone	Art & Design;Creativity		2018-06-20	1.1	4.4 and up
...
9355	FR Calculator	FAMILY	4.0	7	2.6M	500	Free	0.0	Everyone		Education	2017-06-18	1.0.0	4.1 and up
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5000	Free	0.0	Everyone		Education	2017-07-25	1.48	4.1 and up
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100	Free	0.0	Everyone		Education	2018-07-06	1	4.1 and up
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1000	Free	0.0	Mature 17+		Books & Reference	2015-01-19	Varies with device	Varies with device
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10000000	Free	0.0	Everyone		Lifestyle	2018-07-25	Varies with device	Varies with device

9360 rows × 13 columns