

# EDA Implementation with the help of Automated EDA Tools:-

```
In [1]: # importing the data sets
import pandas as pd
df=pd.read_csv(r"C:\Users\Mukul\Downloads\googleplaystore.csv")
```

```
In [2]: df.head() # top 5 datasets
```

Out[2]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	/
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Des
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	/
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	/
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Desig

```
In [3]: df.tail() #Last 5 datasets
```

Out[3]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	E
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	E
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	E
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	E



# Performing the EDA with help of Pandas-

1 Profiling AutomatedEDA Tools

```
In [4]: import pandas as pd
from pandas_profiling import ProfileReport
profile = ProfileReport(df,title='Report')
profile
```

Summarize dataset:	29/29 [00:07<00:00, 2.63it/s,
100%	Completed]
Generate report structure:	1/1 [00:04<00:00,
100%	4.39s/it]
Render HTML: 100%	1/1 [00:01<00:00, 1.50s/it]

Current ver has a high cardinality: 2832 distinct values	High cardinality
Installs is highly correlated with Category and 3 other fields (Category, Rating, Type, Price)	High correlation
Type is highly correlated with Category and 3 other fields (Category, Rating, Installs, Price)	High correlation
Category is highly correlated with Rating and 4 other fields (Rating, Installs, Type, Price, Content Rating)	High correlation
Price is highly correlated with Category and 3 other fields (Category, Rating, Installs, Type)	High correlation
Rating is highly correlated with Category and 3 other fields (Category, Installs, Type, Price)	High correlation
Content Rating is highly correlated with Category	High correlation
Rating has 1474 (13.6%) missing values	Missing
App is uniformly distributed	Uniform

Reproduction

Analysis started	2022-10-20 11:10:12.991605
Analysis finished	2022-10-20 11:10:20.573403
Duration	7.58 seconds
Software version	pandas-profiling v3.3.0 ( <a href="https://github.com/pandas-profiling/pandas-profiling">https://github.com/pandas-profiling/pandas-p</a>
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Report%22%2C%20%22dataset%22%3A%20%22Google%20Play%20Store%22%2C%20%22fields%22%3A%5B%22Category%22%2C%22Rating%22%2C%22Type%22%2C%22Price%22%2C%22Content%20Rating%22%5D%2C%22missing%22%3A%7B%22Rating%22%3A%7B%22count%22%3A%201474%2C%22percentage%22%3A%2013.6%7D%7D)

Variables

Out[4]:

2. Performance EDA with the helpof Sweetviz Automated Tools

```
In [5]: import sweetviz as sv
analyze_report = sv.analyze(df)
analyze_report.show_notebook()
```

Done! Use 'show' commands to display/save.

[100%] 00:00 -> (00:00 left)



### 3. Performance EDA with the help of Autoviz



In [23]: !pip install dtale

Requirement already satisfied: dtale in c:\users\mukul\anaconda3\lib\site-packages (2.8.1)  
Requirement already satisfied: cycler in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.11.0)  
Requirement already satisfied: statsmodels in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.13.2)  
Requirement already satisfied: openpyxl in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (3.0.9)  
Requirement already satisfied: matplotlib in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (3.5.1)  
Requirement already satisfied: networkx in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2.7.1)  
Requirement already satisfied: xlrd in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2.0.1)  
Requirement already satisfied: strsimpy in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.2.1)  
Requirement already satisfied: plotly>=5.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (5.6.0)  
Requirement already satisfied: missingno<=0.4.2 in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.4.2)  
Requirement already satisfied: future>=0.14.0 in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.18.2)  
Requirement already satisfied: squarify in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.4.3)  
Requirement already satisfied: Flask in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2.2.2)  
Requirement already satisfied: dash-bootstrap-components in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.2.1)  
Requirement already satisfied: seaborn in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.11.2)  
Requirement already satisfied: et-xmlfile in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.1.0)  
Requirement already satisfied: kaleido in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.2.1)  
Requirement already satisfied: itsdangerous in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2.0.1)  
Requirement already satisfied: Flask-Compress in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.13)  
Requirement already satisfied: dash>=2.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2.6.2)  
Requirement already satisfied: numpy in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.21.5)  
Requirement already satisfied: xarray in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.20.1)  
Requirement already satisfied: lz4 in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (4.0.2)  
Requirement already satisfied: dash-daq in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.5.0)  
Requirement already satisfied: requests in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2.27.1)  
Requirement already satisfied: scipy in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.9.3)  
Requirement already satisfied: pandas in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.4.2)

Requirement already satisfied: scikit-learn in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.0.2)

Requirement already satisfied: six in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.16.0)

Requirement already satisfied: dash-colorscales in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.0.4)

Requirement already satisfied: flask-ngrok in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (0.0.25)

Requirement already satisfied: certifi in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (2021.10.8)

Requirement already satisfied: dash-core-components==2.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from dash>=2.0.0->dtale) (2.0.0)

Requirement already satisfied: dash-html-components==2.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from dash>=2.0.0->dtale) (2.0.0)

Requirement already satisfied: dash-table==5.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from dash>=2.0.0->dtale) (5.0.0)

Requirement already satisfied: Werkzeug>=2.2.2 in c:\users\mukul\anaconda3\lib\site-packages (from Flask->dtale) (2.2.2)

Requirement already satisfied: Jinja2>=3.0 in c:\users\mukul\anaconda3\lib\site-packages (from Flask->dtale) (3.0.3)

Requirement already satisfied: importlib-metadata>=3.6.0 in c:\users\mukul\anaconda3\lib\site-packages (from Flask->dtale) (4.11.3)

Requirement already satisfied: click>=8.0 in c:\users\mukul\anaconda3\lib\site-packages (from Flask->dtale) (8.0.4)

Requirement already satisfied: colorama in c:\users\mukul\anaconda3\lib\site-packages (from click>=8.0->Flask->dtale) (0.4.4)

Requirement already satisfied: zipp>=0.5 in c:\users\mukul\anaconda3\lib\site-packages (from importlib-metadata>=3.6.0->Flask->dtale) (3.7.0)

Requirement already satisfied: MarkupSafe>=2.0 in c:\users\mukul\anaconda3\lib\site-packages (from Jinja2>=3.0->Flask->dtale) (2.1.1)

Requirement already satisfied: tenacity>=6.2.0 in c:\users\mukul\anaconda3\lib\site-packages (from plotly>=5.0.0->dtale) (8.0.1)

Requirement already satisfied: brotli in c:\users\mukul\anaconda3\lib\site-packages (from Flask-Compress->dtale) (1.0.9)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\mukul\anaconda3\lib\site-packages (from matplotlib->dtale) (1.3.2)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\mukul\anaconda3\lib\site-packages (from matplotlib->dtale) (4.25.0)

Requirement already satisfied: pyparsing>=2.2.1 in c:\users\mukul\anaconda3\lib\site-packages (from matplotlib->dtale) (3.0.4)

Requirement already satisfied: pillow>=6.2.0 in c:\users\mukul\anaconda3\lib\site-packages (from matplotlib->dtale) (9.0.1)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\mukul\anaconda3\lib\site-packages (from matplotlib->dtale) (2.8.2)

Requirement already satisfied: packaging>=20.0 in c:\users\mukul\anaconda3\lib\site-packages (from matplotlib->dtale) (21.3)

Requirement already satisfied: pytz>=2020.1 in c:\users\mukul\anaconda3\lib\site-packages (from pandas->dtale) (2021.3)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\mukul\anaconda3\lib\site-packages (from requests->dtale) (1.26.9)

Requirement already satisfied: charset-normalizer~2.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from requests->dtale) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in c:\users\mukul\anaconda3\lib\site-packages (from requests->dtale) (3.3)

Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\mukul\anaconda3\lib\site-packages (from scikit-learn->dtale) (2.2.0)

Requirement already satisfied: joblib>=0.11 in c:\users\mukul\anaconda3\lib\site-packages (from dtale) (1.0.1)





```
import dtale
dtale.show(df)
```

13		App		Category	Rating
10841					