

Week 1 Deliverables — Data Acquisition (ScreenSense: Kids' ScreenTime Visualization)

1. Project Overview

This project analyses kids' screen time patterns in India to uncover trends by age, gender, device, location (urban/rural), and activity type using data visualization. The goal is to generate insights for parents, educators, and policymakers.

2. Week 1 Objectives

- Load and explore the dataset.
- Understand schema, datatypes, and structure.
- Identify missing values and anomalies.
- Note assumptions and initial observations.

3. Dataset Description

- **File name:** Indian_Kids_Screen_Time.csv
- **Source:** Kaggle — Indian Kids Screentime 2025
- **Size:** 9712(rows) x 8(columns)
- **Main columns:** Age, Gender, Location Type, Device Used, ScreenTime Hours, Activity Category, Day of Week.

Name: Harshit Pandey
Email: pandeyharshit7277@gmail.com

4. Data Exploration Code

```
[19]: import pandas as pd

[2]: #Load dataset
df = pd.read_csv("Indian_Kids_Screen_Time.csv")

[6]: #display first few rows
df.head()
```

	Age	Gender	Avg_Daily_Screen_Time_hr	Primary_Device	Exceeded_Recommended_Limit	Educational_to_Recreational_Ratio	Health_Impacts	Urban_or_Rural
0	14	Male	3.99	Smartphone	True	0.42	Poor Sleep, Eye Strain	Urban
1	11	Female	4.61	Laptop	True	0.30	Poor Sleep	Urban
2	18	Female	3.73	TV	True	0.32	Poor Sleep	Urban
3	15	Female	1.21	Laptop	False	0.39	NaN	Urban
4	12	Female	5.89	Smartphone	True	0.49	Poor Sleep, Anxiety	Urban

```
•[7]: #checking for null values in each column
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9712 entries, 0 to 9711
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                    9712 non-null   int64
1   Gender                                9712 non-null   object
2   Avg_Daily_Screen_Time_hr              9712 non-null   float64
3   Primary_Device                        9712 non-null   object
4   Exceeded_Recommended_Limit            9712 non-null   bool
5   Educational_to_Recreational_Ratio     9712 non-null   float64
6   Health_Impacts                        6494 non-null   object
7   Urban_or_Rural                        9712 non-null   object
dtypes: bool(1), float64(2), int64(1), object(4)
memory usage: 540.7+ KB

[9]: #checking dataset shape(no. of rows and columns)
df.shape

[9]: (9712, 8)
```

```
[10]: #verifying column names
print(list(df.columns))

['Age', 'Gender', 'Avg_Daily_Screen_Time_hr', 'Primary_Device', 'Exceeded_Recommended_Limit', 'Educational_to_Recreational_Ratio', 'Health_Impacts', 'Urban_or_Rural']

[12]: #Summarise the dataset stats
df.describe(include='all')
```

	Age	Gender	Avg_Daily_Screen_Time_hr	Primary_Device	Exceeded_Recommended_Limit	Educational_to_Recreational_Ratio	Health_Impacts
count	9712.000000	9712	9712.000000	9712	9712	9712.000000	6494
unique	NaN	2	NaN	4	2	NaN	15
top	NaN	Male	NaN	Smartphone	True	NaN	Poor Sleep
freq	NaN	4942	NaN	4568	8301	NaN	2268
mean	12.979201	NaN	4.352837	NaN	NaN	0.427226	NaN
std	3.162437	NaN	1.718232	NaN	NaN	0.073221	NaN
min	8.000000	NaN	0.000000	NaN	NaN	0.300000	NaN
25%	10.000000	NaN	3.410000	NaN	NaN	0.370000	NaN
50%	13.000000	NaN	4.440000	NaN	NaN	0.430000	NaN

```
[14]: #Checking for missing values in each column
print(df.isnull().sum())

Age          0
Gender       0
Avg_Daily_Screen_Time_hr  0
Primary_Device  0
Exceeded_Recommended_Limit  0
Educational_to_Recreational_Ratio  0
Health_Impacts 3218
Urban_or_Rural  0
dtype: int64

[15]: #verifying datatypes
print(df.dtypes)

Age          int64
Gender       object
Avg_Daily_Screen_Time_hr  float64
Primary_Device  object
Exceeded_Recommended_Limit  bool
Educational_to_Recreational_Ratio  float64
Health_Impacts  object
Urban_or_Rural  object
dtype: object

[16]: #Unique Value per categorical(where datatype is object) column
for col in df.select_dtypes(include='object').columns:
    print(f'{col}:{df[col].nunique()} Unique Values')

Gender:2 Unique Values
Primary_Device:4 Unique Values
Health_Impacts:15 Unique Values
Urban_or_Rural:2 Unique Values
```

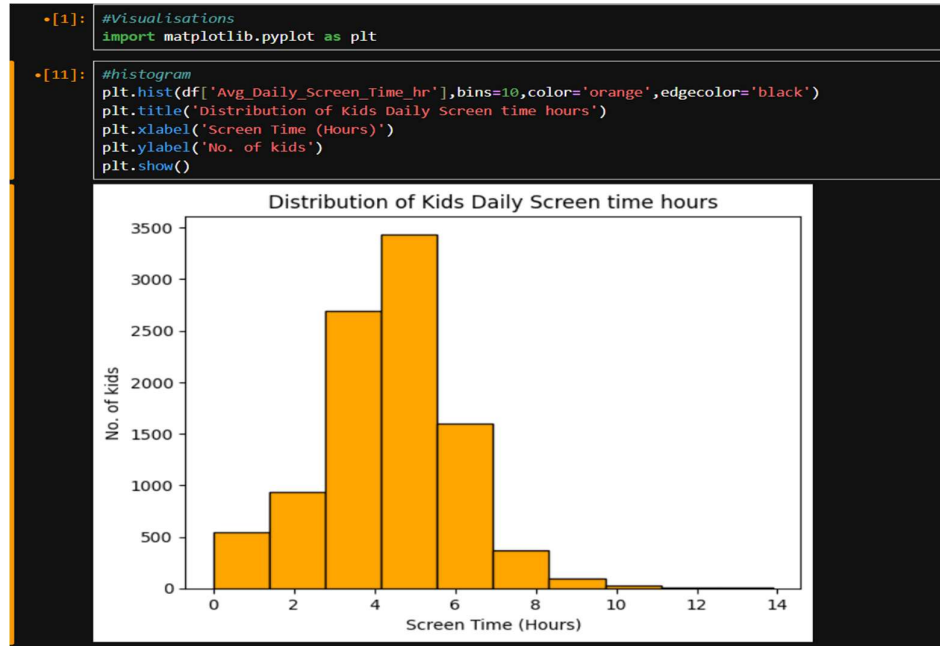
5. Observations

After doing data exploration I found:

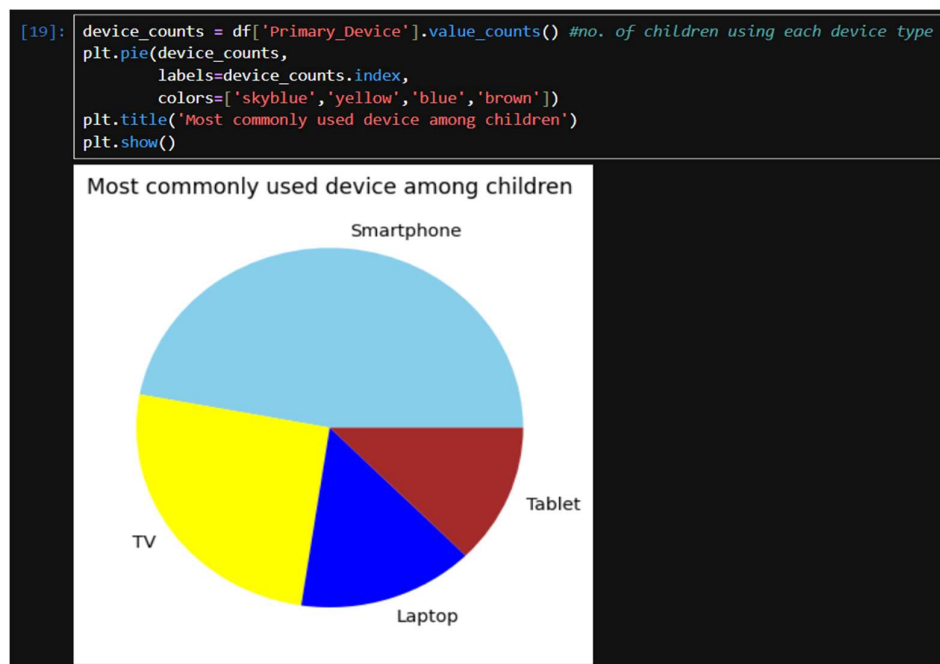
- Dataset has **9712 rows and 8 columns**.
- Some columns like “Health_Impacts” or “Urban_or_rural” contain categorical data.
- Missing values were found in Health_Impacts columns.
- Age distribution and avg daily screen time seems numeric.
- No date columns were present (so time-based trends may be based on “Day of Week”).
- There are 4,15,2 unique values in Primary_device,Health_Impacts and Urban_or_Rural respectively.
- Avg age is approximately 13 and minimum age is 8.

6.My Ideas for visualizations

- Histogram can be used to show how kids' daily screen time hours are distributed.

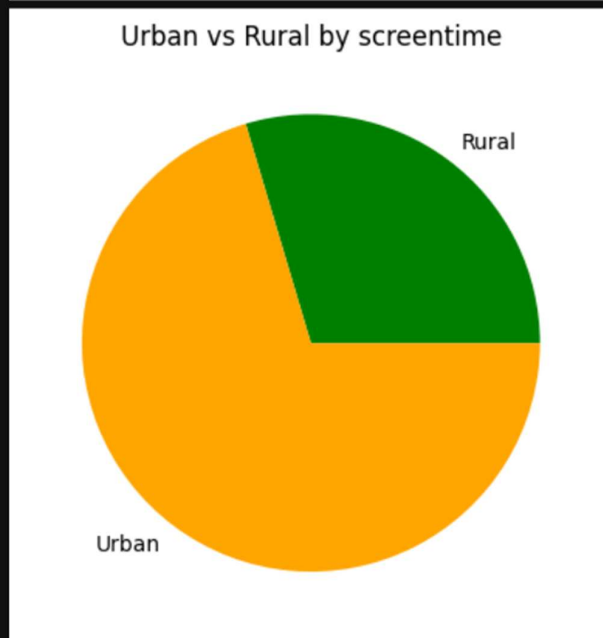


- Pie Chart can be used to display which device type is most commonly used by kids.



- Pie Chart can again be used to Compare urban area kids with rural area kids based on screentime.

```
[16]: #pie chart
screen_time_by_location = df.groupby('Urban_or_Rural')['Avg_Daily_Screen_Time_hr'].sum()
plt.pie(screen_time_by_location,
        labels=screen_time_by_location.index,
        colors=['green','orange'])
plt.title('Urban vs Rural by screentime')
plt.show()
```



7. Conclusion

In Week 1, the dataset was successfully loaded and explored in Jupyter Notebook.

Basic checks were done to understand its structure, data types, and missing values.

From this exploration, we got a clear idea about the columns, such as age, gender, device type, and screen time hours.

Key visualizations like the histogram, pie chart, and bar chart were created to get an initial view of kids' screen time patterns.