

# Screen Sense - Visualizing Kid's Screen Time

## Project Overview: Screen Sense –

The **Screen Sense** project aims to analyse and visualize children's screen time behavior to uncover meaningful trends by **age, gender, location type** (urban/rural), **device type**, **day of the week**, and **activity category** (educational or recreational). The primary goal is to convert raw screen usage data into actionable insights that can help **parents, educators, and policymakers** promote balanced and healthy digital habits among children.

## Project Statement

The project focuses on discovering patterns in children's daily screen time and understanding how these vary across demographics and contexts. Using **data visualization and exploratory data analysis (EDA)**, the aim is to identify:

- How screen time differs by **age and gender**.
- Whether **urban or rural** environments affect device usage.
- How **device type** (mobile, tablet, TV, Laptop etc.) influences time spent.
- Differences in screen activity across **weekdays vs weekends**.
- The proportion of time dedicated to **educational vs recreational** use.

These insights are communicated through visual dashboards and summary reports designed for **non-technical stakeholders**, ensuring the results are interpretable and actionable.

---

## Expected Outcomes

By the end of this project, the following deliverables are expected:

1. **Data Understanding & Preprocessing** – Clean and preprocess the dataset by handling missing values, duplicates, and inconsistencies.
  2. **Exploratory Analysis** – Explore data patterns related, Health Impacts, device usage, and other categories.
  3. **Data Visualization** – Create visual representations such as **bar charts**, **heatmaps**, and **distribution plots** to reveal key insights.
  4. **Stakeholder Insights** – Summarize findings in an easy-to-understand **visual report/dashboard** suitable for presentation.
  5. **Final Presentation** – Deliver key findings supported by data visuals and interpretation.
-

# Milestone 1: Data Foundation and Cleaning

## Week 1: Project Initialization and Dataset Setup :

- Define goals and workflow
- Load the dataset
- Explore schema, data types, size, and nulls
- Capture initial notes on quality and assumptions

## Week 2: Preprocessing and Feature Engineering :

- Handle missing values and inconsistent categories
  - Create derived fields: age bands, weekday/weekend flags, device/activity shares
  - Format any date/time fields
  - Save preprocessed data for reuse; document logic Deliverables: Cleaned dataset, preprocessing summary, feature dictionary.
- 

## Week (1) / Week (2) :

### 1. Week 1 – Project Initialization and Dataset Setup

- Defined goals & workflow.
- Structured: Data loading → Exploration → Cleaning → Visualization.

### 2. Loading the Dataset

- Used Python (`pandas`, `matplotlib`, `seaborn`).
- Checked structure & size with `.info()` and `.describe()`.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style("whitegrid")

file_path = "Indian_Kids_Screen_Time.csv"
df = pd.read_csv(file_path)
```

```
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
```

```
⌚ df.describe()
```

	Age	Avg_Daily_Screen_Time_hr	Educational_to_Recreational_Ratio
count	9712.000000	9712.000000	9712.000000
mean	12.979201	4.352837	0.427226
std	3.162437	1.718232	0.073221
min	8.000000	0.000000	0.300000
25%	10.000000	3.410000	0.370000
50%	13.000000	4.440000	0.430000
75%	16.000000	5.380000	0.480000
max	18.000000	13.890000	0.600000

### 3. Handling Duplicates & Null Values

- Formula to identify nulls: `df.isnull().sum()`
- Remove duplicates: `df.drop_duplicates()`

```
df.isnull().sum()

Σ      0
      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
```

```
duplicate_count = df.duplicated().sum()
print(f"Total duplicate rows found: {duplicate_count}")

if duplicate_count > 0:
    df_cleaned = df.drop_duplicates(keep='first').copy()
    print(f"Successfully removed {duplicate_count} duplicate rows.")
else:
    df_cleaned = df.copy()
    print("No duplicates found. Using the original DataFrame.")

print(f"\nCleaned DataFrame 'df_cleaned' size: {df_cleaned.shape[0]} rows.")

Σ Total duplicate rows found: 44
Successfully removed 44 duplicate rows.

Cleaned DataFrame 'df_cleaned' size: 9668 rows.
```

## 4. Core Insights Found

- Basic descriptive statistics and outliers identified.

```
print("\n-----  Core Descriptive Insights  -----\\n")  
  
overall_avg_screen_time = df_cleaned['Avg_Daily_Screen_Time_hr'].mean()  
print(f"1. Overall Average Daily Screen Time: {overall_avg_screen_time:.2f} hours")  
  
exceeded_limit_count = df_cleaned['Exceeded_Recommended_Limit'].sum()  
total_kids = df_cleaned.shape[0]  
percentage_exceeded = (exceeded_limit_count / total_kids) * 100  
  
print(f"2. Percentage of kids exceeding recommended limit: {percentage_exceeded:.2f}%")
```

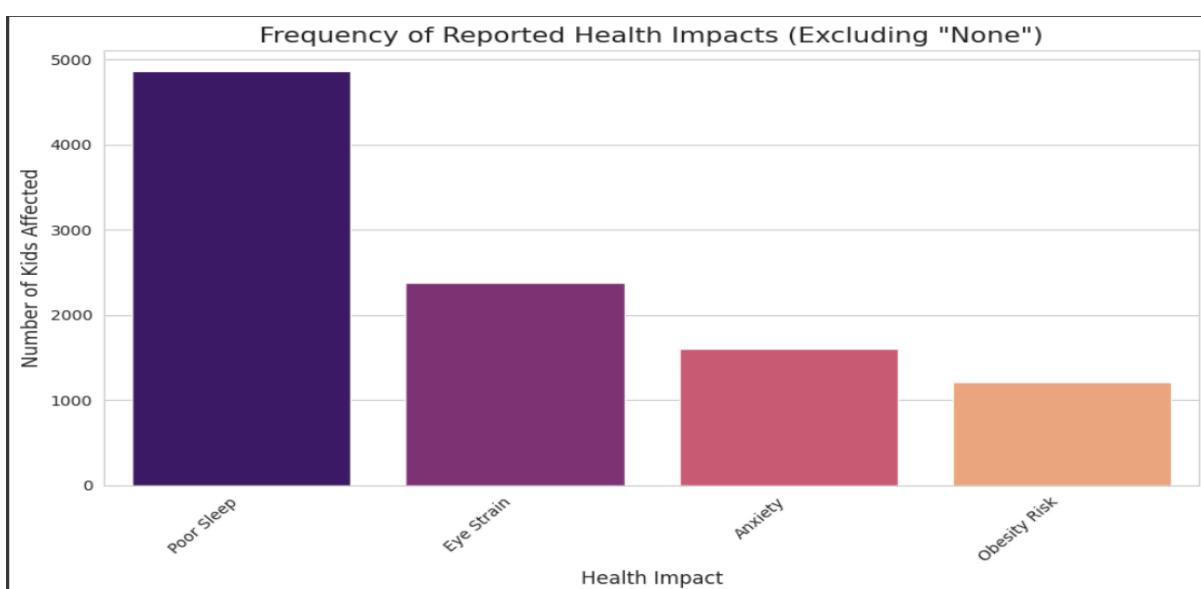


----- Core Descriptive Insights -----

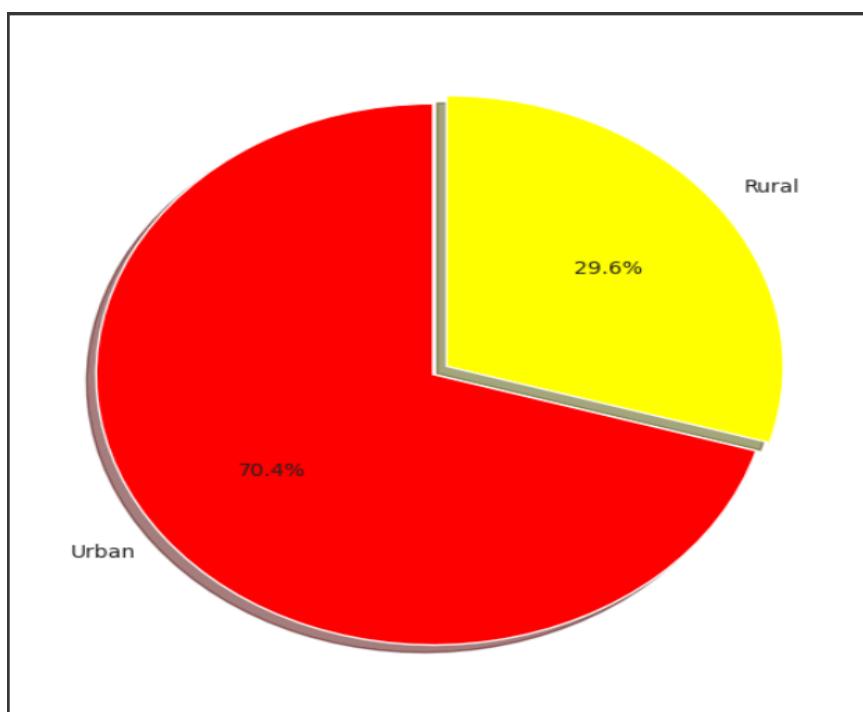
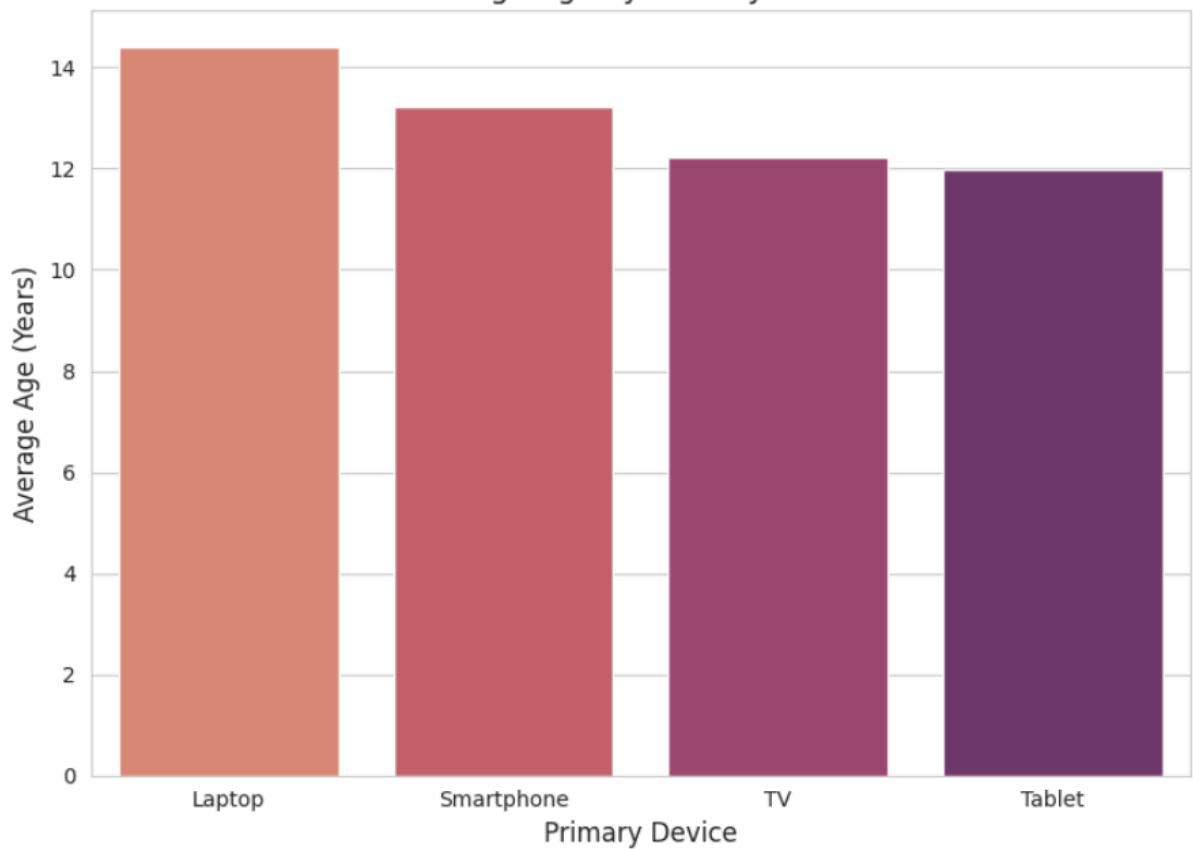
- Overall Average Daily Screen Time: 4.37 hours
- Percentage of kids exceeding recommended limit: 85.77%

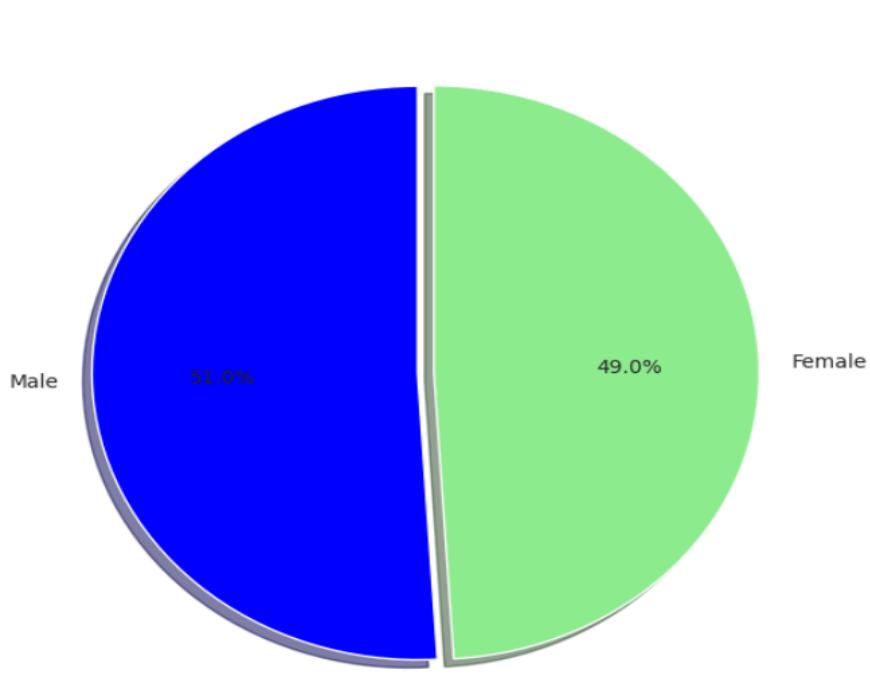
## 5. Basic Visualization

- Bar chart: *Health impacts by number of kids*.
- Line graph: *Average use of primary device by age*.
- Pie chart: *Urban vs Rural screen time*.
- Bar comparison: *Screen time by gender*.



Average Age by Primary Device





## 6. Working with Excel

- Added new calculated columns for further visualization.

I	J	K	L	M	N
Screen_Time_Category	Screen_Time_Usage	Screen_Usage_Impact	Educational_Use_Hr	Recreational_Use_Hr	Age_Category
Moderate	Entertainment Use	Moderate Impact	1.68	2.31	Teen
High	Entertainment Use	Low Impact	1.38	3.23	Kid
Moderate	Entertainment Use	Low Impact	1.19	2.54	Adult
Low	Entertainment Use	None	0.47	0.74	Teen
High	Entertainment Use	Moderate Impact	2.89	3.00	Kid
High	Entertainment Use	Low Impact	2.15	2.73	Teen
Moderate	Entertainment Use	None	1.43	1.54	Teen
Moderate	Educational Use	None	1.48	1.26	Kid
High	Entertainment Use	Moderate Impact	1.66	2.95	Teen
Moderate	Entertainment Use	Moderate Impact	1.56	1.68	Adult
Moderate	Entertainment Use	Low Impact	1.55	1.98	Adult
High	Entertainment Use	Low Impact	1.78	3.16	Teen
High	Entertainment Use	Low Impact	1.97	2.62	Kid
Very High	Entertainment Use	Low Impact	2.92	3.16	Kid
Very High	Entertainment Use	Low Impact	2.46	3.69	Teen
Moderate	Entertainment Use	Low Impact	1.25	2.44	Teen
Very High	Entertainment Use	Moderate Impact	3.05	4.05	Kid
Very High	Entertainment Use	Low Impact	3.35	3.63	Teen
Moderate	Entertainment Use	Low Impact	1.53	2.29	Kid
Moderate	Entertainment Use	Low Impact	1.89	1.96	Teen
High	Entertainment Use	Low Impact	1.34	2.71	Teen
Low	Educational Use	None	0.95	0.68	Kid