

SCREEN SENSE – KIDS’ SCREENTIME VISUALIZATION

WEEK 1 REPORT – PROJECT INITIALIZATION AND DATASET SETUP

1. OBJECTIVE

The main goal for Week 1 was to initialize the project environment, load the dataset, and explore its structure to understand data quality and trends. This stage lays the foundation for the next steps — preprocessing, feature creation, and visualization — aimed at analysing kids’ screentime patterns for insights beneficial to parents, educators, and policymakers.

2. IMPLEMENTATION

Step 1: Importing Required Libraries

Imported Python libraries such as pandas, NumPy, matplotlib, and seaborn for data handling and visualization. These tools will be essential throughout the project for analysis and reporting.

```
[2]: # Import Libraries

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

Step 2: Dataset Loading

Loaded the 'Indian Kids Screentime 2025' dataset and verified successful import with 9,712 rows and 8 columns.

```
[3]: # Load Dataset

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

```
[4]: # Basic overview

print(" Dataset Loaded Successfully!\n")
print("Total Rows:", df.shape[0])
print("Total Columns:", df.shape[1])
```

Dataset Loaded Successfully!

Total Rows: 9712
Total Columns: 8

Step 3: Schema and Summary Exploration

Used `df.info()` and `df.describe()` to review column data types, statistics, and data completeness. No missing or duplicate values were found.

```
[5]: # Explore Schema and Data Types

print("\n Dataset Information:\n")
df.info()

Dataset Information:

<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
```

Step 4: Unique Values and Data Types Check

Analyzed unique values per column to understand categorical distributions, confirming proper data formats for further preprocessing.

```
[10]: print("\n Duplicate Rows:", df.duplicated().sum())

Duplicate Rows: 44

[11]: # 5 Unique Values per Column

print("\n Unique Values per Column:\n")
for col in df.columns:
    print(f"{col}: {df[col].nunique()} unique values")

Unique Values per Column:

Age: 11 unique values
Gender: 2 unique values
Avg_Daily_Screen_Time_hr: 899 unique values
Primary_Device: 4 unique values
Exceeded_Recommended_Limit: 2 unique values
Educational_to_Recreational_Ratio: 31 unique values
Health_Impacts: 15 unique values
Urban_or_Rural: 2 unique values

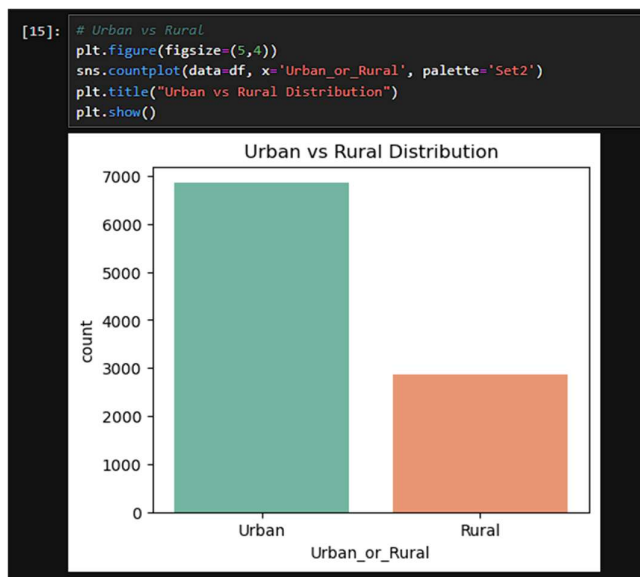
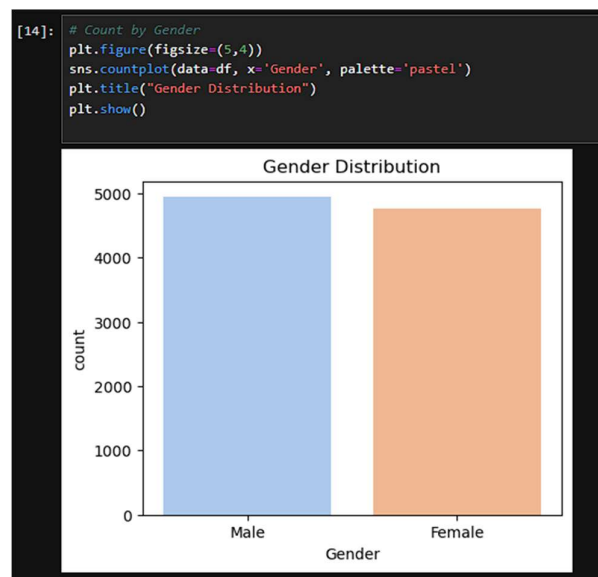
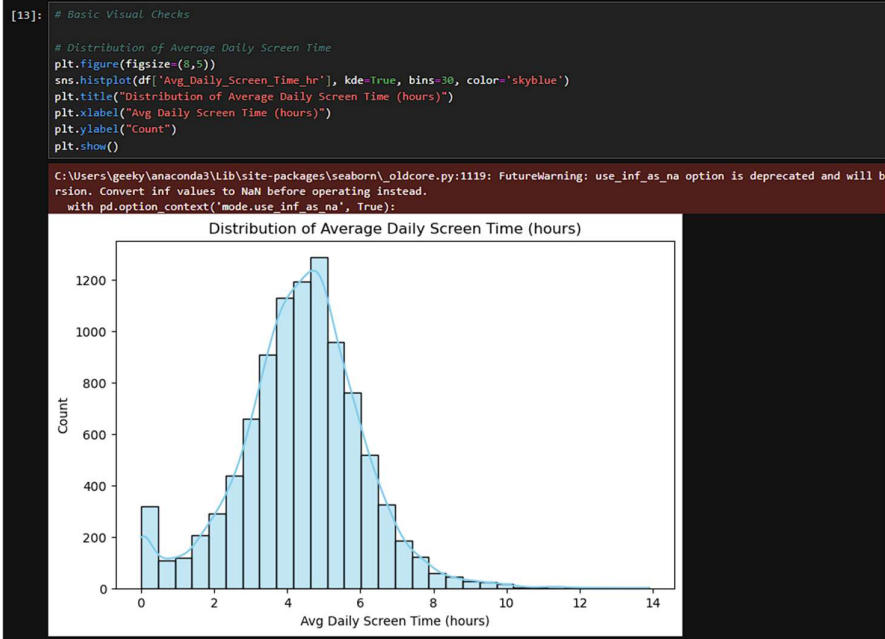
[12]: # Data Type Overview

print("\n Data Types Summary:")
print(df.dtypes)

Data Types Summary:
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
```

Step 5: Initial Visualizations

Plotted histograms for average screen time and bar charts for gender and location. These visualizations revealed that most children spend 3–6 hours on screens, with a slight male majority and mostly urban representation.



Step 6: Initial Data Quality & Assumptions:

1. No missing values or nulls detected — dataset is complete.
2. Age range: 8 to 18 years, which aligns with 'kids' definition.
3. Average daily screen time ranges from 0 to ~14 hrs.
4. Gender column: balanced but slightly more Male entries.
5. Primary devices: Smartphone is dominant (~45% of records).
6. Urban/Rural distribution: Mostly Urban (~70% of records).
7. 'Health Impacts' column contains multiple comma-separated entries.
8. No immediate formatting issues, but outlier detection will be needed in Week 2.

3. CONCLUSION

Week 1 successfully established a strong foundation for the Screen Sense project. The dataset was explored and validated, confirming its quality and readiness for further analysis. Key findings — such as high screen time concentration in the 3–6-hour range and urban dominance — will guide deeper feature engineering and visualization in Week 2. This ensures data readiness for generating meaningful insights aligned with the project's goal of understanding kids' screentime behaviour and its health implications.

Report By – Yash Verma