

ScreenSense: Kids' Screentime Visualization

WEEK 1 – PROJECT INITIALIZATION AND DATASET SETUP

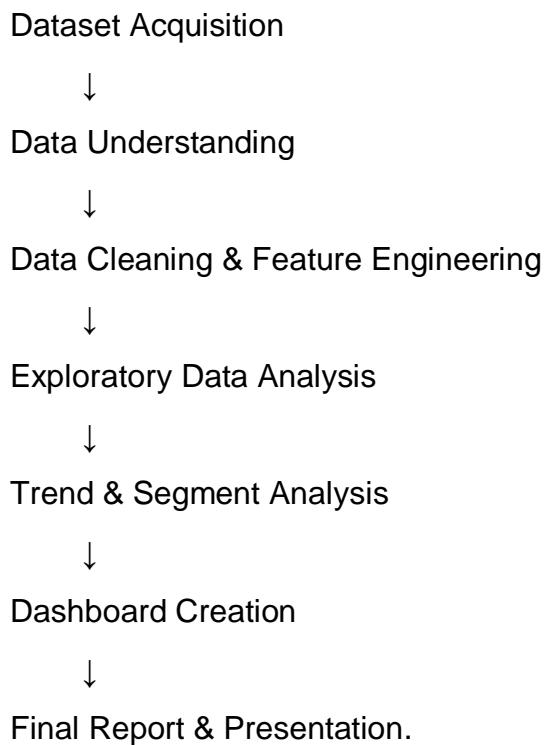
Objective:

The primary objective of Week 1 is to establish a strong foundation for the Screen Sense project by setting up the environment, loading the dataset, and performing an in-depth exploratory review of its structure, completeness, and quality. This phase ensures that subsequent stages of data preprocessing and visualization are based on clean, well-understood data.

1. Project Overview

The project 'Screen Sense' aims to explore and visualize children's screen time behavior in India using the Kaggle dataset 'Indian Kids Screentime 2025'. The goal is to uncover key behavioral trends across demographics such as age, gender, location type, and device or activity preferences. The insights from this study are intended to help parents, educators, and policymakers understand the digital exposure of children in a structured, data-driven way.

PROJECT WORK FLOW:



2. Week 1 Goals

- Define clear project goals and workflow.
- Load and explore the dataset for structure and integrity.
- Identify missing values, inconsistent entries, and potential data quality issues.
- Summarize findings in a structured data understanding report.
- Document the assumptions, dataset limitations, and preparation needs for Week 2 preprocessing.

3. Week 1 Activities and Deliverables

. Goal 1: Implementation: install and Importing Required Libraries

At Initially, we Install and Imported Required Python libraries such as pandas, NumPy, matplotlib, and seaborn which are useful for the data handling and visualization. These tools will be essential throughout the project for analysis and reporting.

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

.Goal 2: Load and explore the dataset for structure and integrity

import the CSV (Indian_Kids_Screen_Time.csv) using pandas.read_csv() and perform an initial scan:

- df.head() → shows the first few rows for a quick look at column names.
- df.info() → reveals data types, memory usage, and non-null counts.
- df.shape → gives total rows × columns.
- df.describe() → summarizes numeric columns (mean, median, std dev).

Goal 3 – Identify missing values, inconsistent entries, and potential data quality issues:

Once the dataset is loaded, data quality must be checked.

Typical techniques:

```
df.isnull().sum()  
df.duplicated().sum()  
df['DeviceType'].unique()
```

Checks reveal:

- **Nulls** in ActivityCategory or DeviceType.
- **Inconsistent spellings** (e.g., “mobile”, “Mobile Phone”, “MOBILE”).
- **Extreme values** in Daily_ScreenTime_Hours (> 15 h).

Goal 4 – Summarize findings in a structured data-understanding

- Table of columns, data types, sample values, missing %.
- Observations about patterns (e.g., higher usage in certain age ranges).

- Possible data corrections or derived fields needed.

4. Exploratory Observations

- After the initial inspection, the dataset revealed several important characteristics and patterns:
- Dataset Size: 9,712 records.
- Numerical Fields: Age, Daily Screen Time Hours, Educational to Recreational Ratio
- Categorical Fields: Gender, Location Type (Urban/Rural), Device Type, Health impacts.
- Potential Data Issues: Missing or blank entries in Device Type and Activity Category. Inconsistent text formats (e.g., 'mobile', 'Mobile Phone', 'Mobile').
- Possible outliers (Daily Screen Time Hours > 15 hours).

5. Tools and Technology Used

- Python Libraries: pandas, numpy for data handling and inspection.
- Visualization Libraries: matplotlib, seaborn for distribution and null-value plots.
- Development Environment: Jupyter Notebook for step-by-step documentation and testing.

6. Week 1 Deliverables

1. Dataset successfully loaded and inspected.
2. Structural summary (columns, data types, memory footprint).
3. Missing value report and categorical value validation.
4. Preliminary data observations and quality notes.
5. Jupyter Notebook: `Week1.ipynb` containing all commands and outputs.

Conclusion

The Week 1 phase of the *ScreenSense: Kids' Screenshot Visualization* project successfully established the foundation for an effective data-driven analysis. The dataset was collected, explored, and examined to understand its structure, contents, and overall quality. Through systematic inspection, the initial week ensured that all subsequent stages—cleaning, visualization will be based on reliable and well-organized data.

The project now stands ready for the preprocessing phase, where missing values and inconsistencies will be addressed, and new derived fields will be engineered. This structured approach ensures that the insights generated in later stages will be accurate, meaningful, and relevant for understanding the screen-time behaviors of Indian children.

Submitted By:

LAKSHMI REDDY KOTTAM