ScreenSense: Kids' Screentime Visualization

Milestone 1: Data Foundation and Cleaning

Week 1: Project Initialization and Dataset Setup

1. Define Goals and Workflow

Goal: Analyze kids' screentime to identify patterns based on:

- Age
- Gender
- Location type (Urban/Rural)
- Device type
- Weekday vs Weekend
- Activity category

Purpose: Provide actionable insights for parents, teachers, and policymakers to understand children's digital habits.

2. Load the Dataset

Import your raw dataset into your working environment (Python, Power BI, Excel, or Tableau).

Check whether the file loads correctly and understand what each column represents (e.g., age, gender, screentime).

File formats: CSV, Excel, JSON, etc.

Example (Python):

```
import pandas as pd
df = pd.read_csv('kids_screentime.csv')
df.head()
```

3. Explore Schema, Data Types, Size, and Nulls

Before cleaning, inspect the structure and quality of your data.

Column	Type	Description
Age	Numerical	Age of the child
Gender	Categorical	Male/Female
LocationType	Categorical	Urban/Rural
Device	Categorical	Phone, Tablet, TV, etc.
Activity	Categorical	Learning, Gaming, etc.
ScreenTime_Minutes	Numerical	Total screen time per day
DayOfWeek	Categorical	Monday–Sunday

Data Types:

Use **df.info()** to check column types and convert incorrect ones if necessary.

Size: Use **df.shape** to get total records and columns.

Nulls: Use **df.isnull().sum()** to check for missing values.

4. Capture Initial Notes on Data Quality and Assumptions

Document your initial observations and assumptions after exploring the dataset.

Example Notes:

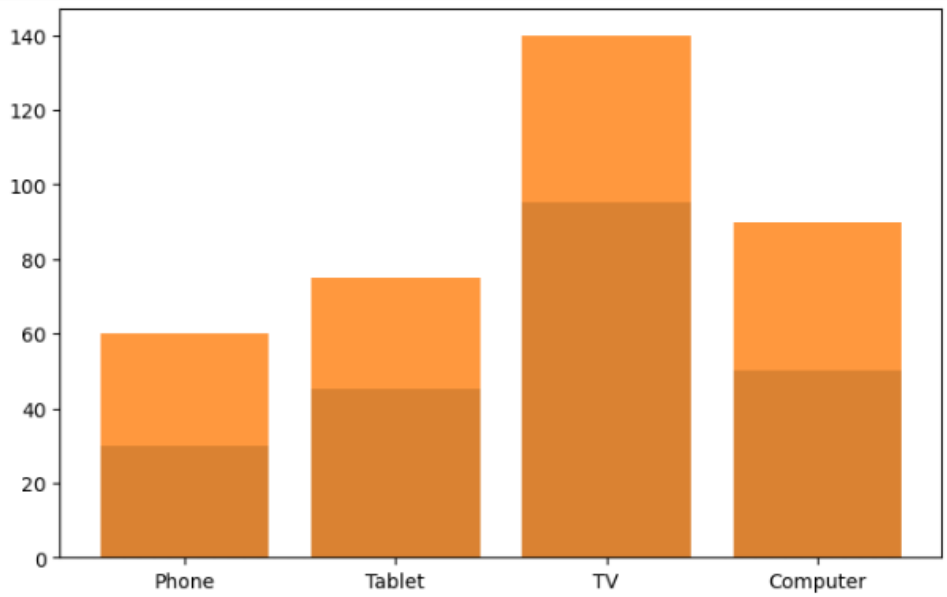
- Found 2% missing values in Activity → fill with “Unknown”.
- Age values range from 3–17 → looks valid.
- Some entries in Device are lowercase (“mobile”, “Mobile”) → standardize.
- Assume ScreenTime_Minutes = total daily screen time.
- Will group ages later (e.g., 5–8, 9–12, 13–17).

Why-- it’s important: These notes guide your Week 2: Data Cleaning phase and help track identified issues and planned fixes.

Program Execution :

```
import pandas as pd
import matplotlib.pyplot as plt
data = {
    'AgeGroup': ['5-8', '5-8', '5-8', '5-8', '9-12', '9-12', '9-12', '9-12'],
    'DeviceType': ['Phone', 'Tablet', 'TV', 'Computer', 'Phone', 'Tablet', 'TV', 'Computer'],
    'AvgScreenTime_Min': [30, 45, 95, 50, 60, 75, 140, 90]
}

df = pd.DataFrame(data)
print("Dataset Preview:")
print(df)
print("\n--- Data Info ---")
print(df.info())
print("\n--- Missing Values ---")
print(df.isnull().sum())

plt.figure(figsize=(8,5))
for age_group in df['AgeGroup'].unique():
    subset = df[df['AgeGroup'] == age_group]
    plt.bar(subset['DeviceType'], subset['AvgScreenTime_Min'], label=f'{age_group} years', alpha=0.8)

plt.figure(figsize=(6,6))
device_sum = df.groupby('DeviceType')['AvgScreenTime_Min'].sum()
plt.pie(device_sum, labels=device_sum.index, autopct='%1.1f%%', shadow=True, startangle=90)
plt.title('Overall Screen Time Share by Device Type')
plt.show()

plt.title('Average Screen Time by Device Type and Age Group')
plt.xlabel('Device Type')
plt.ylabel('Average Screen Time (minutes)')
plt.legend(title='Age Group')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

```
Dataset Preview:
  AgeGroup DeviceType  AvgScreenTime_Min
0      5-8      Phone                 30
1      5-8     Tablet                 45
2      5-8         TV                 95
3      5-8   Computer                 50
4     9-12      Phone                 60
5     9-12     Tablet                 75
6     9-12         TV                140
7     9-12   Computer                 90

--- Data Info ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 3 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   AgeGroup           8 non-null      object
 1   DeviceType         8 non-null      object
 2   AvgScreenTime_Min  8 non-null      int64 
dtypes: int64(1), object(2)
memory usage: 324.0+ bytes
None

--- Missing Values ---
AgeGroup            0
DeviceType          0
AvgScreenTime_Min   0
dtype: int64
```



Overall Screen Time Share by Device Type