

*Worksheet 6 - Oct 17, 2025*

**🎵 Describing Data with Statistics**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Dataset:** [spotify\_mini.csv](https://github.com/Sammyjoon/Intro-to-Data-Science/tree/main/Week%206)

**Goal:** Learn how to describe data with numbers — mean, median, mode — and explore simple correlations.

🧩 Part 1: Load & Explore the Data

1️. Download **spotify\_mini.csv** from GitHub (Week 6) and upload it to your Colab sidebar (Files ➜ Upload).  
2️. Run this code:

A close-up of a computer code

Description automatically generated

**📝 Answer:**

1. How many rows (songs) are in this dataset?
2. List three column names:  
    -  
    -  
    -

### 📊 Part 2: What Are Statistics?

Statistics means **summarizing data** with numbers.  
Example: Instead of looking at all 100 songs, we can describe the dataset with:

* **Mean** → average
* **Median** → middle value
* **Mode** → most common value

🧠 Let’s use these ideas to describe our Spotify data!

🎯 Part 3: Find Mean, Median, and Mode

We’ll start with **streams** (how many times each song was played).

A screenshot of a computer program

Description automatically generated

**📝 Questions:**

* What’s the average number of streams?
* What’s the median (middle value)?
* What’s the most common number of streams?
* Which measure (mean, median, or mode) do you think best describes the dataset — and why?

💡 Part 4: Compare Skip Rate

Let’s describe another column — **skip\_rate** (how often a song gets skipped).

A close-up of words

Description automatically generated

**📝 Questions:**

1. Is the average skip rate low or high?
2. What might cause some songs to have a higher skip rate?

🔍 Part 5: Correlation – How Two Things Relate

We’ll check if songs with more streams are skipped less often.

A screen shot of a computer code

Description automatically generated

**📝 Questions:**

1. What does the scatterplot show?
2. Does it look like there’s a relationship?
3. What is the correlation number (close to +1, 0, or -1)? What does it mean?

🧠 Hint:

* Positive (→ +1): when one goes up, the other goes up
* Negative (→ -1): when one goes up, the other goes down
* Near 0: no relationship

🧮 Part 6: Your Turn!

Now you decide what to explore!  
Pick **two columns** from the dataset and find a relationship between them.  
Some ideas:

* minutes\_listened vs. streams
* genre vs. skip\_rate
* day\_of\_week vs. streams

🧩 Write your code!

**📝 Question:**  
What story does your chart tell?

### 💬 Part 7: Reflection (5 min)

Answer in your own words:

1. Which statistic (mean, median, or mode) do you find easiest to understand?
2. What did you learn about your Spotify data?
3. Complete the sentence:

“Statistics help me understand data because…”