

Describing my Data Project Choice

By Michael Campoverde and
Dmytro Lukinchuk


A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.


Title and link to my project of choice.

The title of my Data project that we choice from pudding is called “Can an AI make a data-driven, visual story?”

The link (MLA format) - The Pudding.
“We Had an AI Attempt to Make a Data-Driven Story like We Do at the Pudding.” *The Pudding*, 2020, pudding.cool/2024/07/ai/. Accessed 30 Oct. 2024.

Brief summary of the digital project you are exploring.

►  We like the song idea, can you get more specific?

▼  Sure, let me clarify some details for you.

Great question! Let me clarify the details and address your concerns regarding the unit of emotional analysis and the constraints of the song corpus.

Unit of emotional analysis:

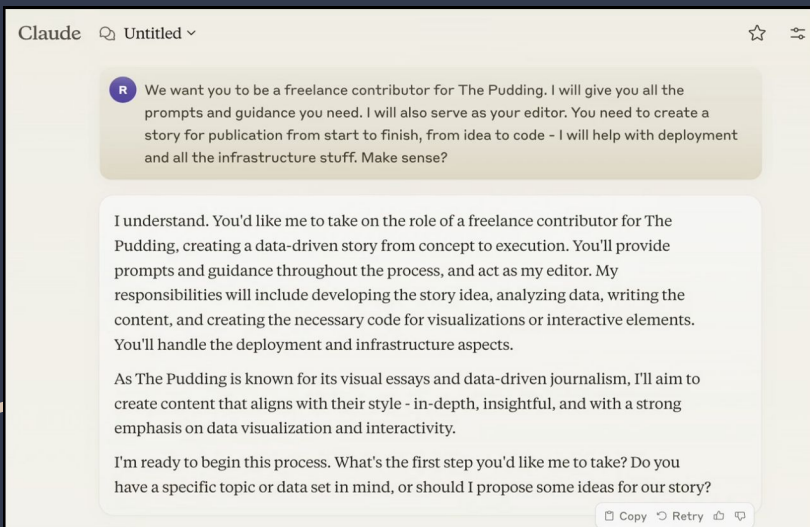
The primary unit of analysis will be the overall emotional sentiment of each song, determined by analyzing the lyrics in their entirety. This approach will give us a high-level understanding of the dominant emotions and sentiments expressed in the song.

However, to add depth to the analysis, we can also explore the emotional trajectory within each song by breaking down the lyrics into smaller units, such as verses and choruses. This

The whole idea behind this Data project was to see if AI could make a data-driven visual story. It was given prompts and specific instructions to see how well AI could gather all of this data and put it well formatted and how does it do it exactly. For example, they chose to talk about songs and music and they helped AI to structure and explain what they wanted.

What are some objectives of the authors/researchers/scholars?

The main objectives of the author was to grade and explore the thought process of AI when creating a data project. They did this by helping give AI (Claude) a list of prompts and guidance on steps and what to do in order to reach their end goal.



Describe the visualizations.

- ▶  The project has been greenlit. Write some scripts to collect the data.

- ▼  Absolutely! Here is a bunch of code.

```
import requests
import csv
from bs4 import BeautifulSoup

# Replace 'YOUR_ACCESS_TOKEN' with your actual Genius API access token
access_token = 'YOUR_ACCESS_TOKEN'

# Input file name
input_file = 'billboard_top_100.csv'

# Output file name
output_file = 'song_lyrics.csv'

# Write the data to a CSV file
with open(input_file, 'r', encoding='utf-8') as infile, open(output_file, 'w', encoding='utf-8') as outfile:
    reader = csv.reader(infile)
    writer = csv.writer(outfile)
```

Most of the visualization shown was the authors giving AI the prompts such as questions or tasks and AI explaining its thought process. They also show the different websites and coding that AI wrote in order to create a sufficient data project like showing graphs and charts.

```
In [4]: df = pd.read_csv('song_emotion_results.csv')
        print(df.head())
```

	chart_week	current_week	title	performer
0	2022-01-01	1	All I Want For Christmas Is You	Mariah Carey
1	2022-01-01	2	Rockin' Around The Christmas Tree	Brenda Lee
2	2022-01-01	3	Jingle Bell Rock	Bobby Helms
3	2022-01-01	4	A Holly Jolly Christmas	Burl Ives
4	2022-01-01	5	Easy On Me	Adele

	last_week	peak_pos	wks_on_chart	Sentiment
0	1.0	1	50	Positive
1	2.0	2	44	Positive
2	4.0	3	41	Positive
3	5.0	4	25	Positive
4	3.0	1	11	Positive


```
Emotions
0 {'anger': 2, 'anticipation': 6, 'disgust': 2, ...
1 {'anger': 7, 'anticipation': 11, 'disgust': 7, ...
2 {'anticipation': 14, 'joy': 1, 'trust': 2}
3 {'anticipation': 5, 'joy': 3, 'surprise': 3, ...
4 {'anticipation': 10, 'joy': 11, 'surprise': 6, ...
```

What strategies in visualization do they use to facilitate their objectives?

Our grade: B-

A too-literal but adequate data analyst

This is pretty impressive. Typically, coding is a practice of being very thorough and specific with how you give a computer instructions. But here, we simply asked Claude something like "can you analyze it" (extremely vague) and it produced workable scripts almost immediately. During this stage, we often found ourselves giddy with the fun of working with such a powerful tool.

Which didn't always last for long. If Claude ever got derailed, it was hard to get it back on track. We've found that when we're interacting with AI using our technical background, we're able to help it solve bugs quickly by helping it identify the issue. But since we acted as if we had no coding skills for this story, and simply reported errors and asked Claude to fix them, it often went in circles and recovered inefficiently from issues.

For complex coding tasks, we feel like these tools are only as good as their operator. You might get lucky and get flawless code, but chances are, it will falter eventually and struggle to correct itself.

Even when its scripts worked, a closer look revealed some sloppiness. At every step, Claude made a few mistakes or oversights. Unless prompted, Claude couldn't notice or account for things like considering songs with multiple artists. It did the bare minimum, strictly following its prompts, for better or worse. The effort it takes us humans to fix things like this and produce stories that cover all the bases can often take days on end. It involves questioning results, identifying edge cases, doing more research, and repeating. This stage of our projects is often the most deceptively time-consuming. So while Claude was faster, it is "satisfied" with whatever it generates. Critical thinking still falls to the user.

They use screenshots of the AI's ideas of creating a story as a guide to visualization. Along with this they also provided a grade at the end of each process. This help grade overall summarized what they liked or didn't like from Claude himself. They also explain what they were mainly looking for and what they were hoping for when giving Claude prompts.

Which was the most effective visualization?



Can you write a storyboard?

Thank you for the insights. We are ready to move on to the storyboard stage of this project. Here is what we are looking for: an outline of the actual story. There should be discrete sections that explain what a summary of the copy will be about, what data and chart types you might use. Keep in mind The Pudding's approach to storytelling, it should be a tight narrative with visuals that support the thesis. We will give you a round of feedback after your first draft before moving into production. Keep in mind you haven't seen the full data, but assume your insights (or at least their general concepts) will be interesting on the full data.



No problem, here is my outline.

Understood! Here's my proposed outline for the data story on the rise of negative sentiment in popular music:

Section 1: Introduction

- * Open with an anecdote about a recent hit song with surprisingly dark/negative lyrics

- * Thesis: Over the past decade, there has been a steady rise in the proportion of emotionally negative songs topping the Billboard charts, reflecting a broader shift in popular music's emotional landscape.

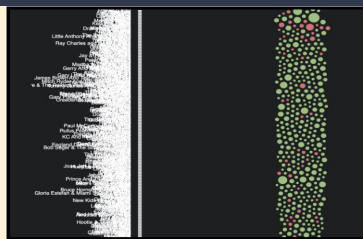
- * Preview key insights: the trend correlates with Hip hop's chart dominance, turbulent cultural events, and certain artists'

The most effective visualization was showing the chats between the AI and the authors of this article. The chats also helped gave us an idea of what they were trying to accomplish and give readers an idea of what was happening and their thought process

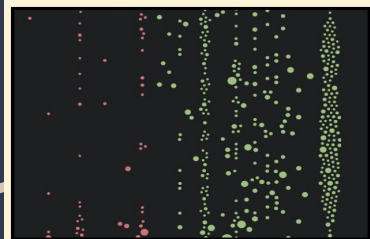
Which visualization would you make changes on?



"i should be clearer, a swarm chart uses d3 force to prevent overlap"



"it looks like the red and green colors are spread throughout, i expect positive to be at the top, and negative to be at bottom"



"it looks like negative/positive is horizontal, not vertical"



"very close - except the circles also go from left to right, they should all try to be in the center on the x axis"

These data charts are mainly what I would change as they are too clustered and too difficult to understand. As you can't see the names at all for the swarm charts. For the bubble charts at first a reader may misinterpret this chart and completely not understand it at all. Most of the data shown is too much and the reader won't be able to follow along with what the authors are trying to accomplish.

Level-up
component:



Explain why what you have chosen is not a good visualization

The reason why I chose this was because this visualization is mainly inappropriate as instead of charts or graphs it uses a neck of a person this makes it very distraction.

Another reason is it's mainly cluttered and small trying to look at the numbers. The last reason is I believe using a graph would better show and compare the values of his earnings such as a pie chart.

What improvements
will you make to
improve the
visualization and
why?

As I explained they should scrap the whole picture of the guy and the data being his neck as its very distracting and inappropriate. Instead, I would use pie charts or a bar graph in order to properly visualize the data and not making it as cluttered and small. I would also try to label each team and explain more what the data is showing as in the years and why. The reason why is because for it to be a proper visual data its important to be able to show the data clear and be able to show some comparison between them.