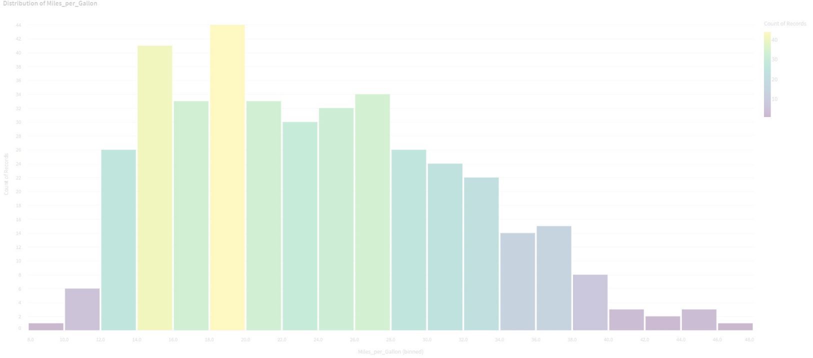


VisuaLing

Enhancing User Interaction and Aesthetics in LLM-Based Visualization Systems

Chandramita Dutta
Chirag Manghani

Problem Statement

Current Situation:

- Research on LLM-generated visualizations prioritize accuracy over aesthetics.
- Lack of emphasis on essential design elements.

Challenges:

- Missing elements like proper positioning, axes, and color schemes.
- Impact on overall aesthetic appeal and user engagement.

Critical Gap:

- Current visualizations may not meet user expectations for design and appeal.

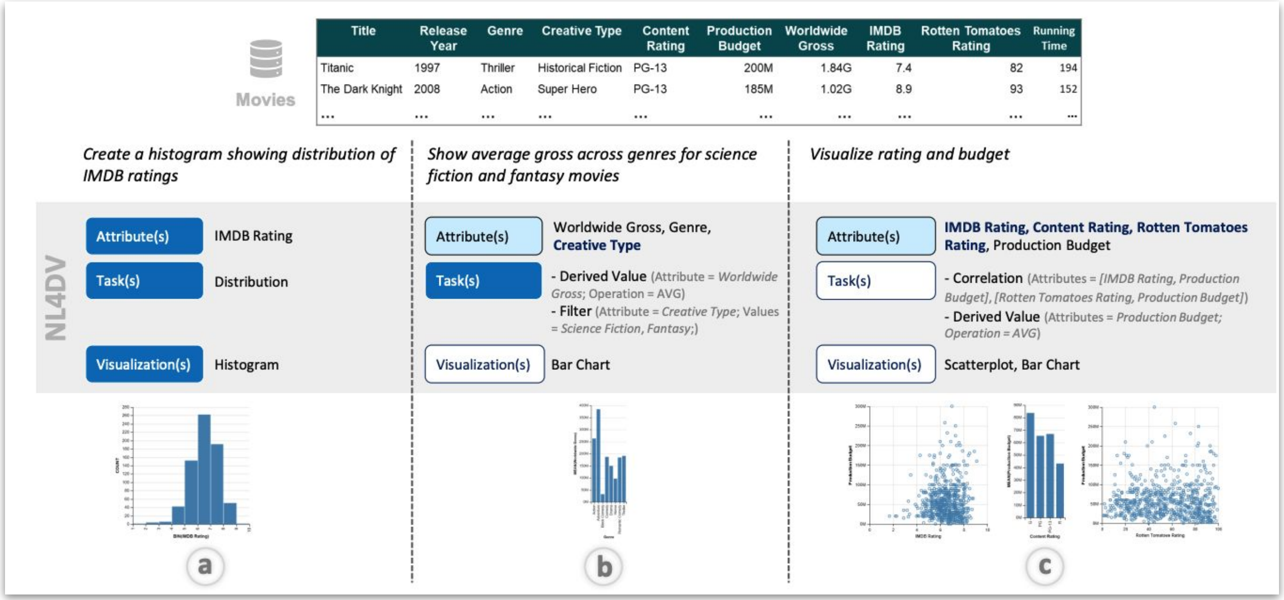
Research Objective:

- Empower users to confidently utilize GPT-backed tools for end-to-end visualization generation.
- Promote wider adoption and utilization across various domains.

Previous Work

NL4DV – Natural Language Toolkit for Data Visualization

- Python package
- Input: Table + Natural Language Query
- Output: analytic specification
- Modeled as: JSON object
    - Containing – data attributes, analytic tasks, and a list of Vega-Lite specifications relevant input query.
- Aids developers to create new visualization NLIs



[2020, NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries
Arpit Narechania*, Arjun Srinivasan*, and John Stasko]

NL4DV: Facilitating Conversational Interaction in Natural Language Interfaces for Visualization

..a step forward in the realm of Natural Language Interfaces (NLIs) for data visualization.

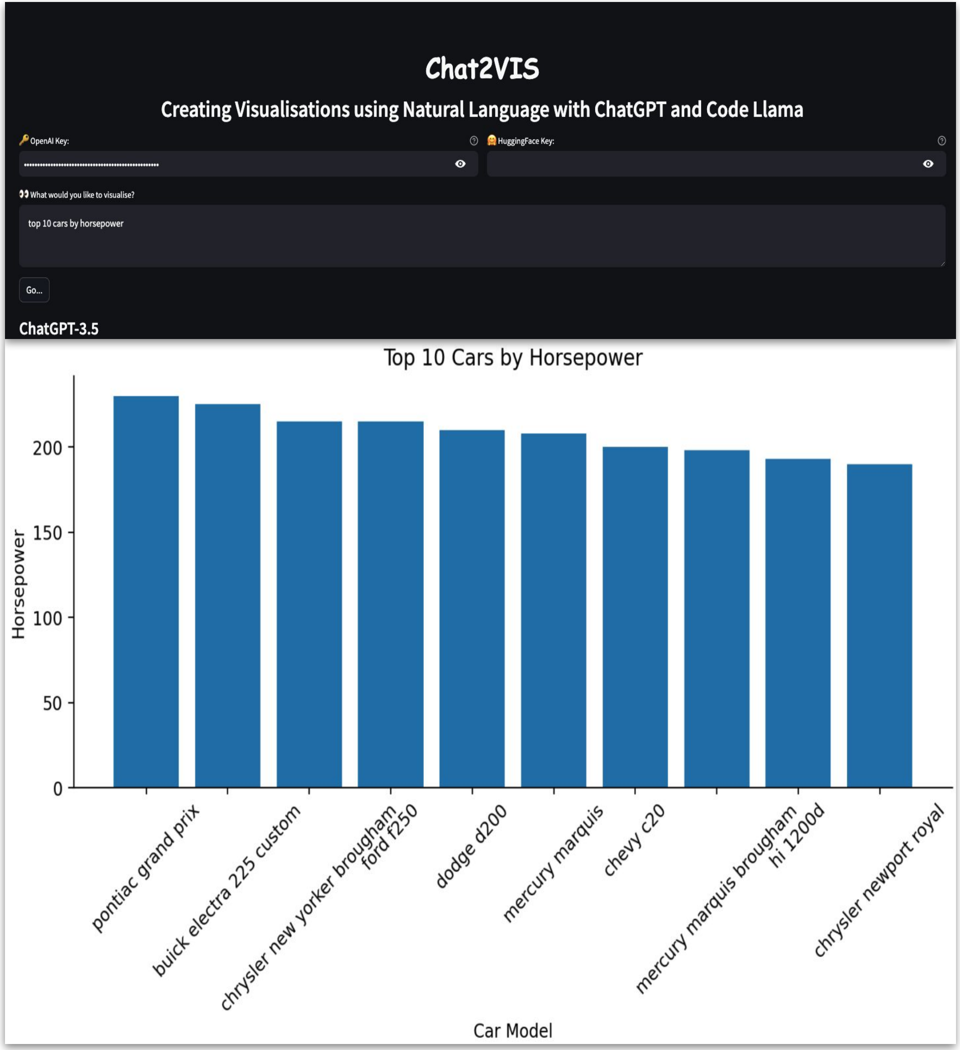
- Current focus on singleton queries
- **Emphasizing the importance of multi-turn dialogues**
- Addresses challenges of ambiguity and context
- Augments output JSON with conversational information

2022, Facilitating Conversational Interaction in Natural Language Interfaces for Visualization
Rishab Mitra, Arpit Narechania,Alex Endert,John Stasko

Chat2VIS:
Generating Data Visualisations via Natural Language using ChatGPT, Codex and GPT-3 Large Language Models

Using Existing Language Models for Visualization Code

- Utilizing pre-trained LLMs (ChatGPT, Codex, GPT-3)
- Efficiently handling complex, ambiguous queries
- Simplifies NLI development, cost-effective
- Ensures data privacy/security
- **LLMs with prompts outperform traditional NLP**



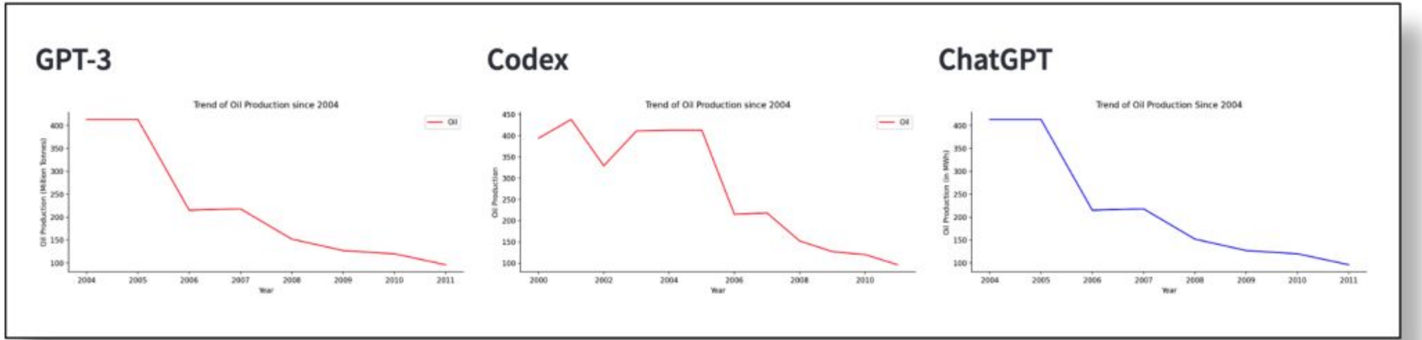
[2023, Chat2VIS: Generating Data Visualisations via Natural Language using ChatGPT, Codex and GPT-3 Large Language Models, Paula Maddigan et al.]

Comparison between LLM models and NL4DV

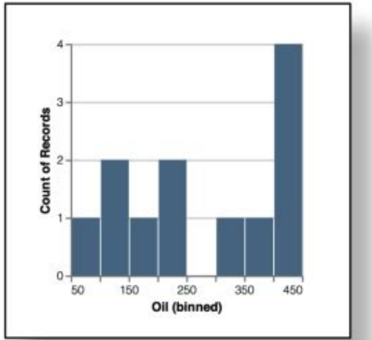
"What is the trend of oil production since 2004?"

Year	Population(M)	Coal	Oil	Gas	Nuclear
2000	282.1700	6968	394	2179	2672
2001	285.0800	6679	438	2274	2697
2002	287.8000	6717	329	2441	2710
2003	290.3300	6798	411	2292	2631

Energy Production Dataset



Chat2VIS



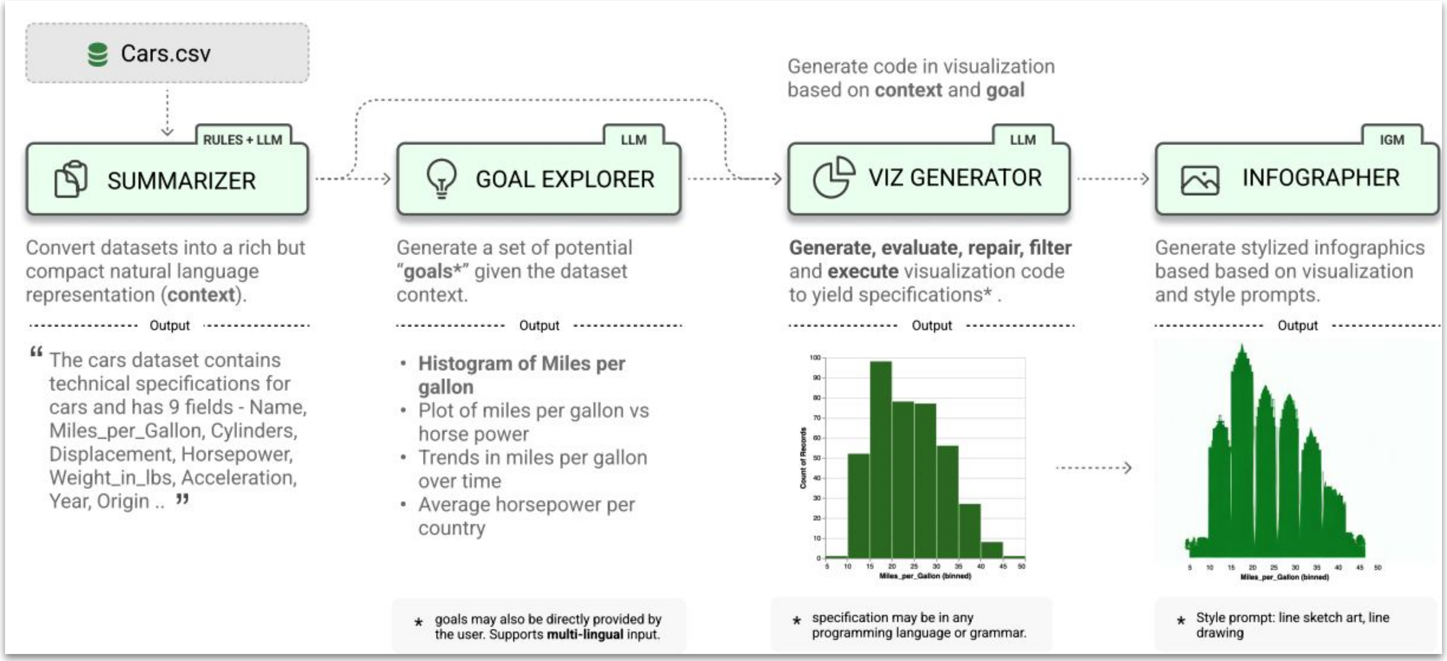
NL4DV

- GPT-3, ChatGPT: Correctly show data from 2004 – in a single plot line (most suitable for this query)
- Codex: Misses detail, depicts data from 2000
- NL4DV: Incorrect visualization, semantic parsing limits which lacks flexibility

[2023, Chat2VIS: Generating Data Visualisations via Natural Language using ChatGPT, Codex and GPT-3 Large Language Models, Paula Maddigan et al.]

LIDA

developed by Microsoft in July 2023



- Core Modules: SUMMARIZER, GOAL EXPLORER, VISGENERATOR, INFOGRAPHER
- Grammar-agnostic approach, addressing limitations of existing methods
- Goal explorer allows users to explore LLM suggested goals without having to query themselves, helping with cold start
- Open source – we built upon the existing capabilities of LIDA to generate enhanced visualizations

[July, 2023, LIDA: A Tool for Automatic Generation of Grammar-Agnostic Visualizations and Infographics using Large Language Models Victor Dibia]

Beyond Generating Code: Evaluating GPT on a Data Visualization Course

This paper evaluates GPT based models in Data Visualization tasks

- Evaluates data interpretation, visualization design, exploration, and insight communication.
- Methodology: Used OpenAI APIs for Harvard's CS171 data visualization course assignments.

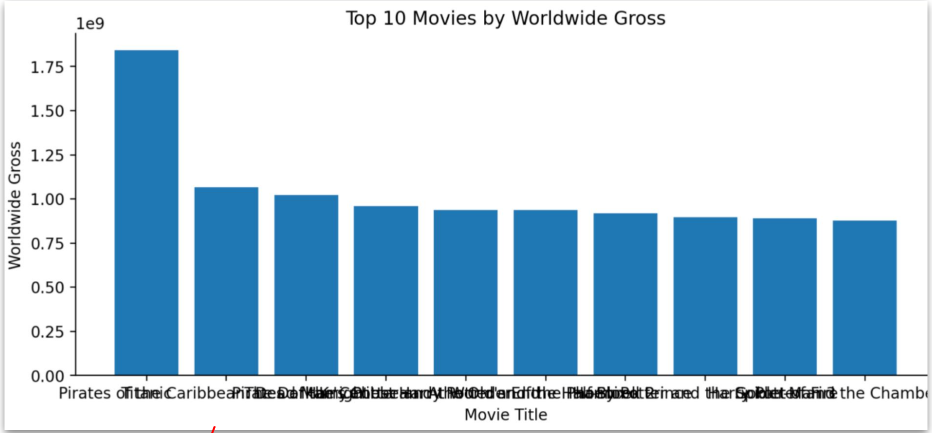
Key Findings:

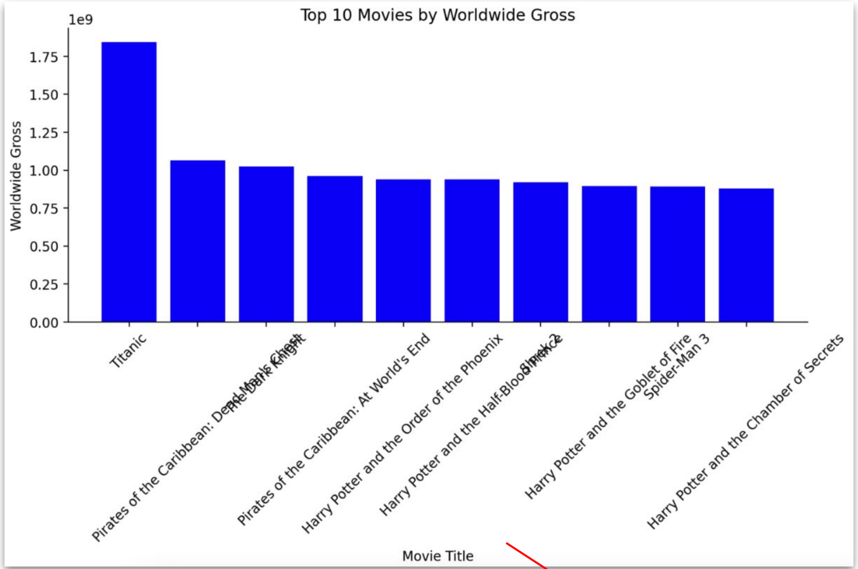
- GPT-4 proficient in visualization assignments, showcasing strengths and limitations.
- **Efficient prompt engineering may address some LLM limitations.**
- **LLMs struggle with aesthetic visualizations and positioning defects.**

Beyond Generating Code: Evaluating GPT on a Data Visualization Course Chen Zhu-Tian et al.

Our work holds importance as it focuses on improving aesthetics and interaction, key elements that directly impact user experience in both data visualization exploration and presentation.

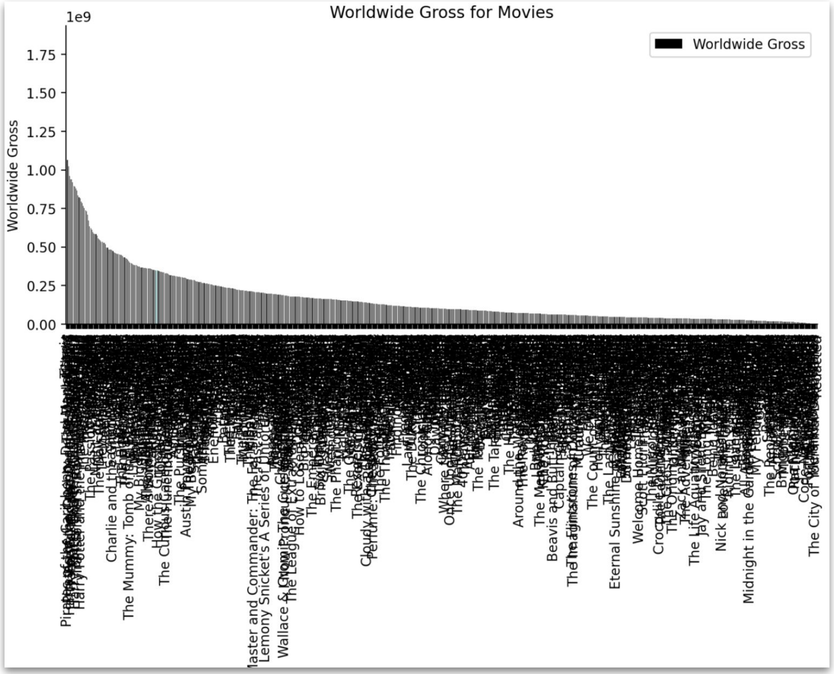
Examples from Chat2Viz and LIDA on Aesthetics Impacting Readability

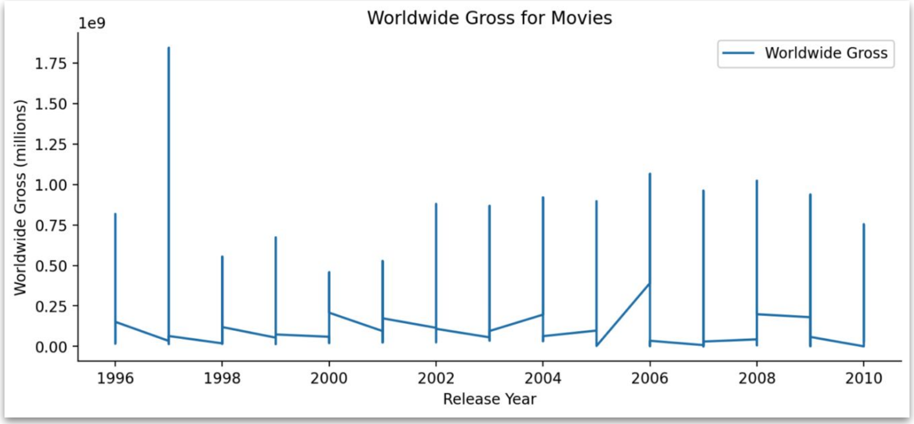




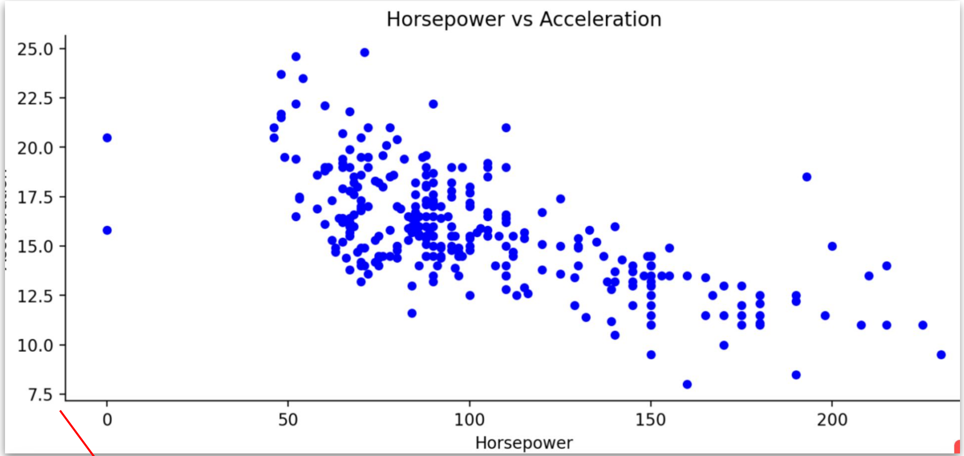
Too many axes labels overlapping on each other

Ambiguous labelling





Both charts are not readable because of ambiguous aesthetics



Y axis starts at 7.5

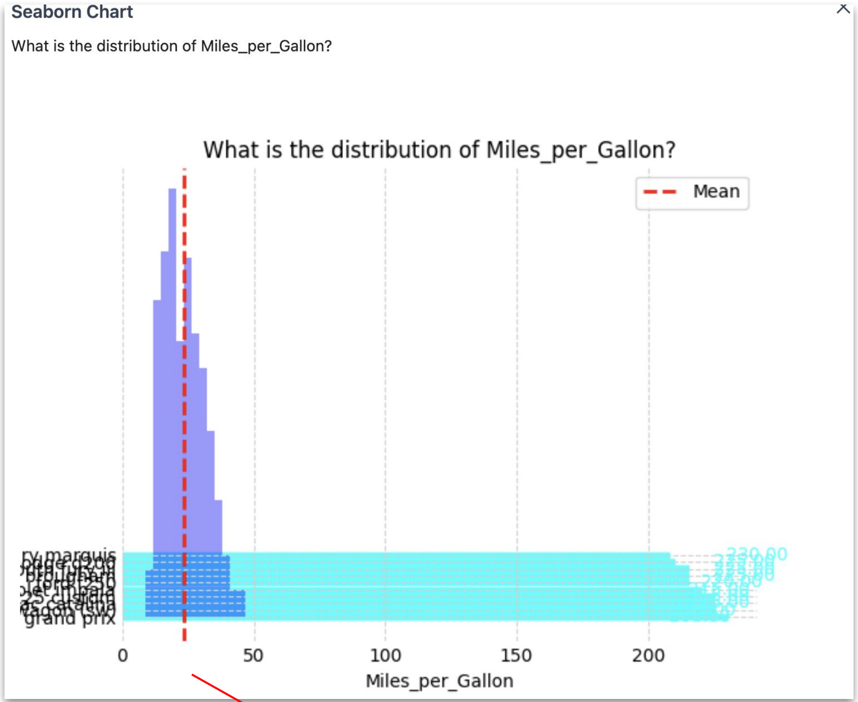


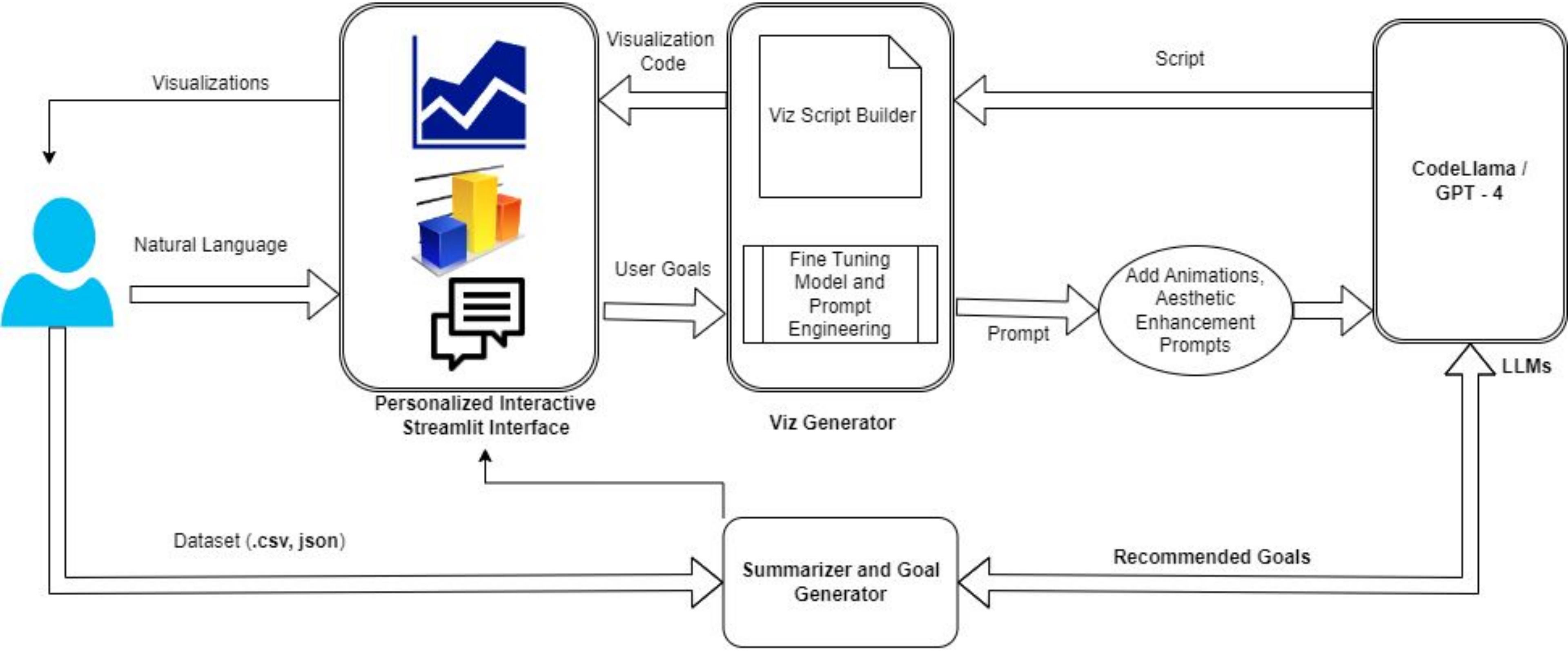
Chart is not readable because of ambiguous aesthetics

Key Limitations in Previous Systems

- LLMs are fairly new and produce non-deterministic responses.
- Lack human interaction and feedback loop. Mostly emphasize on building accurate graphs.
- Existing solutions use Seaborn and Matplotlib primarily generate static images, limiting the user's ability to interact with and explore the data visually
- Focus is on generating an image of the viz but not exporting as a SVG to embed / publish it

Introducing VisuaLing

Architecture



Method

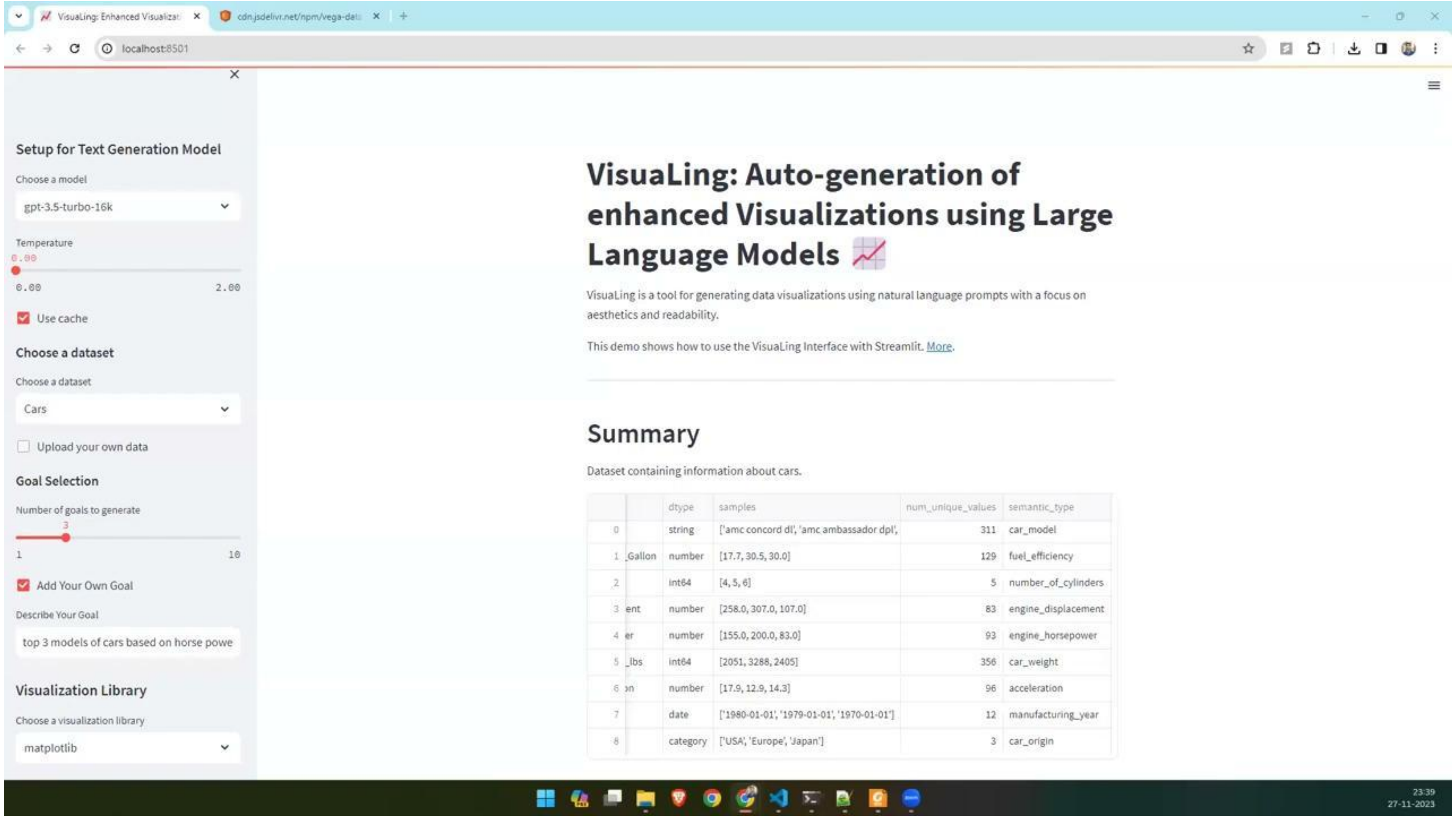
Building upon the existing LIDA architecture with an emphasis on interactivity and user readability.

Carefully worded context and prompts added a layer to focus on aesthetics

Understanding and fine tuning the model parameters (seed, temperature, logit_bias) to produce deterministic results

Caching the user parameters

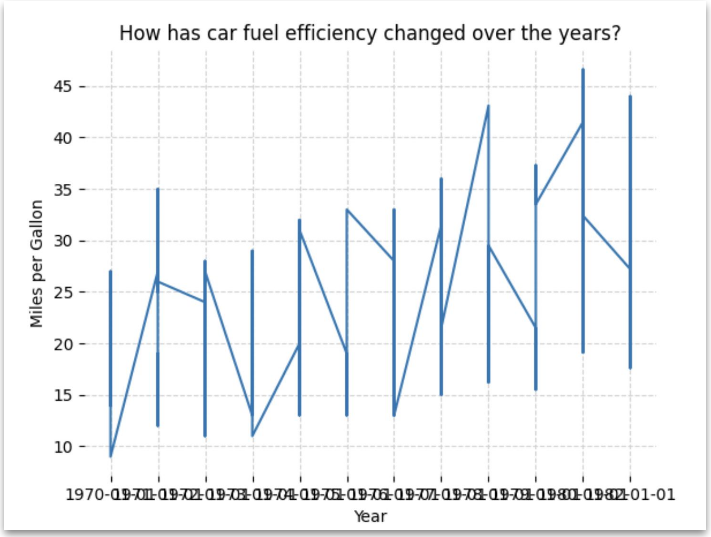
Interface – Demo

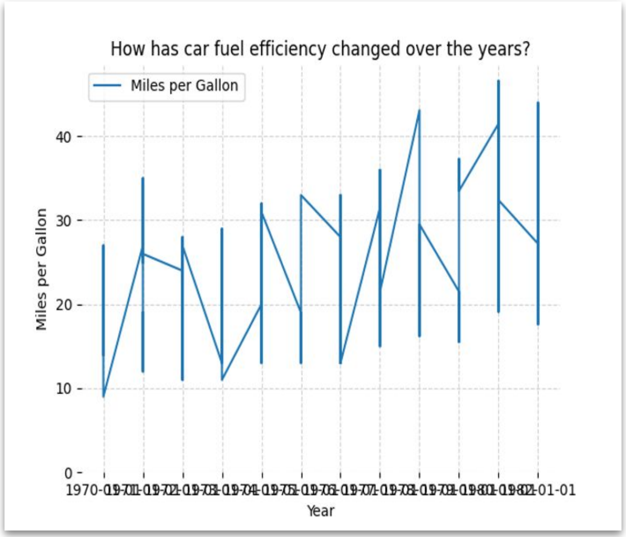


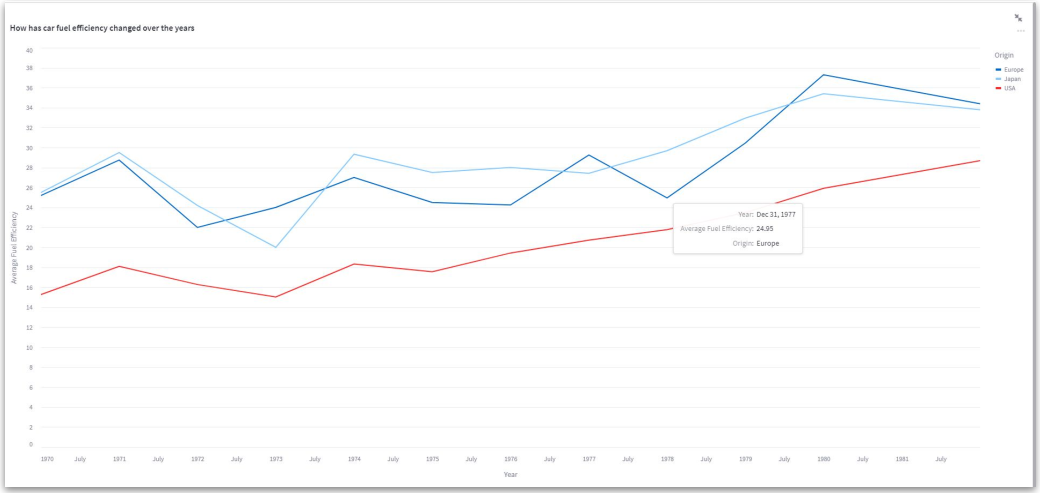
Evaluation – Case study

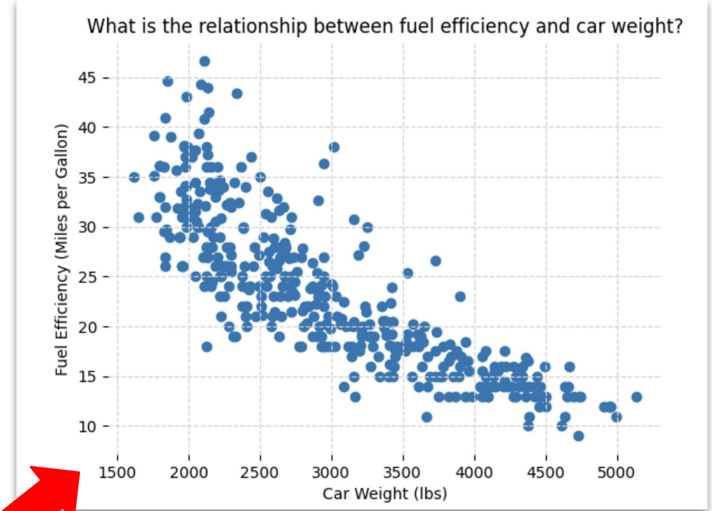
Visual comparison of charts generated by LIDA and VisuaLing

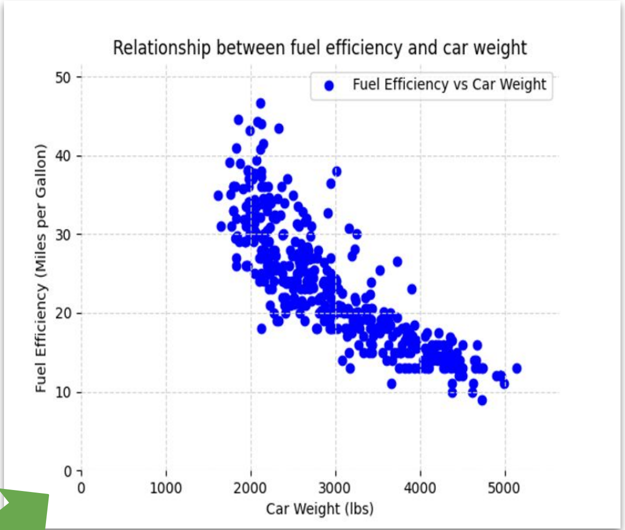
Model - GPT-3.5-turbo
All charts were created using the same dataset, same prompt and the same model temperature

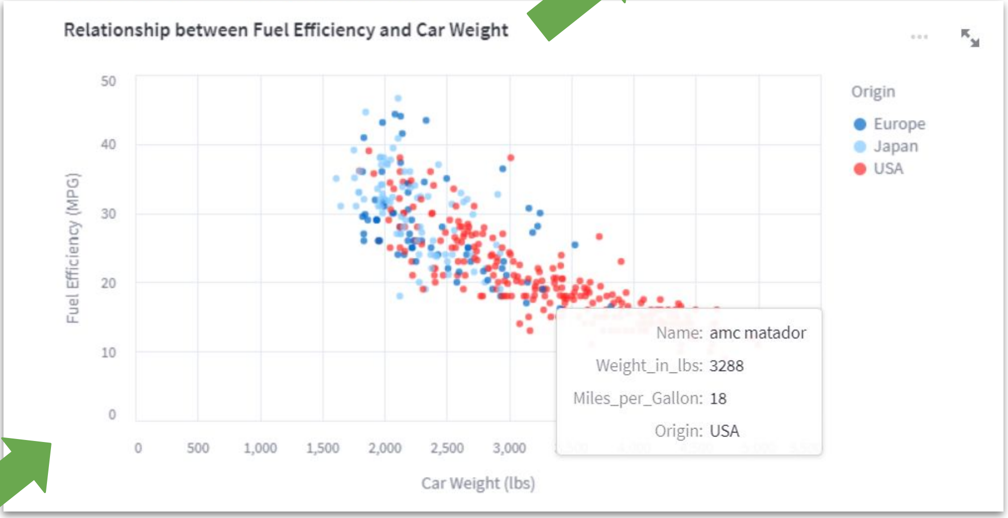

How has car fuel efficiency changed over the years?
Miles per Gallon
Year
45
40
35
30
25
20
15
10

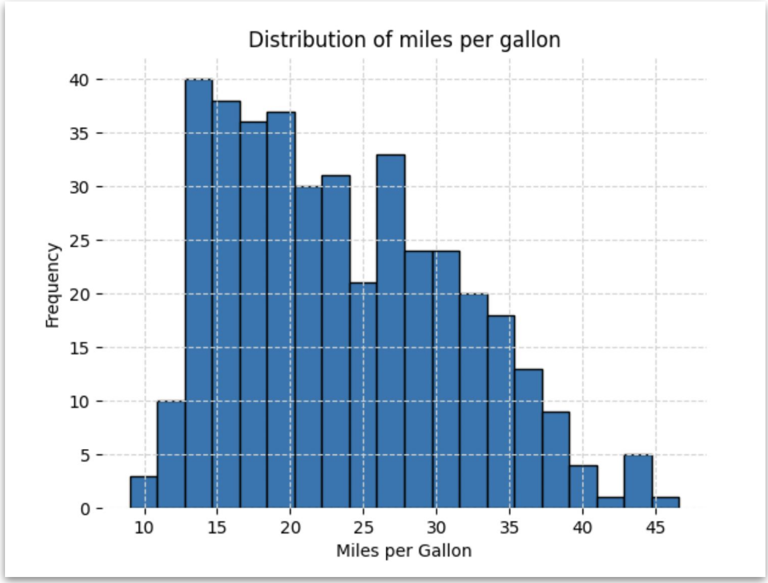

How has car fuel efficiency changed over the years?
Miles per Gallon
Miles per Gallon
Year
40
30
20
10
0

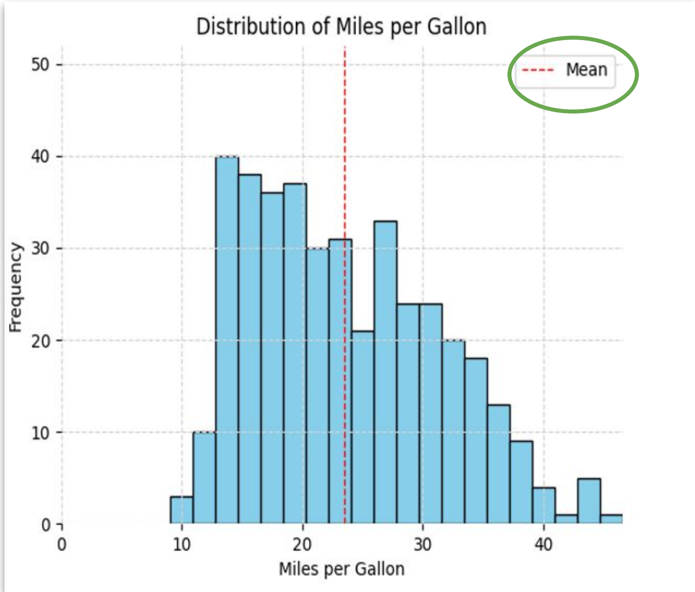

How has car fuel efficiency changed over the years
Origin
Europe
Japan
USA
Year: Dec 31, 1977
Average Fuel Efficiency: 24.95
Origin: Europe
Average Fuel Efficiency
Year
1970
July
1971
July
1972
July
1973
July
1974
July
1975
July
1976
July
1977
July
1978
July
1979
July
1980
July
1981
July


What is the relationship between fuel efficiency and car weight?
Fuel Efficiency (Miles per Gallon)
45
40
35
30
25
20
15
10
1500
2000
2500
3000
3500
4000
4500
5000
Car Weight (lbs)

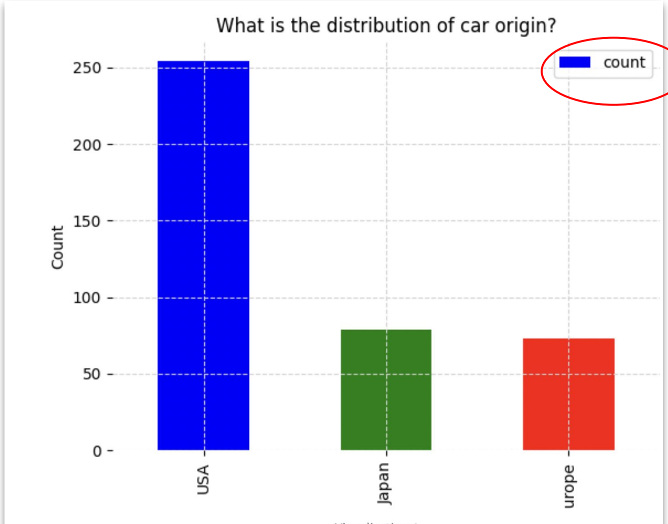

Relationship between fuel efficiency and car weight
Fuel Efficiency vs Car Weight
Fuel Efficiency (Miles per Gallon)
50
40
30
20
10
0
0
1000
2000
3000
4000
5000
Car Weight (lbs)


Relationship between Fuel Efficiency and Car Weight
Origin
Europe
Japan
USA
Fuel Efficiency (MPG)
50
40
30
20
10
0
0
500
1,000
1,500
2,000
2,500
3,000
Name: amc matador
Weight_in_lbs: 3288
Miles_per_Gallon: 18
Origin: USA
Car Weight (lbs)

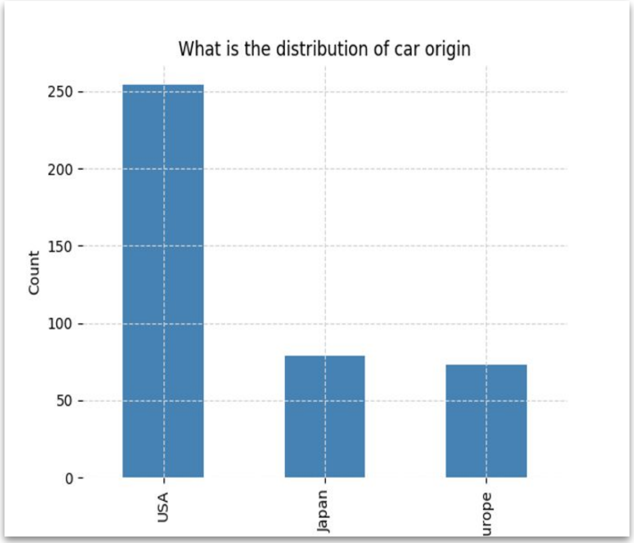

Distribution of miles per gallon
Frequency
0
5
10
15
20
25
30
35
40
10
15
20
25
30
35
40
45
Miles per Gallon

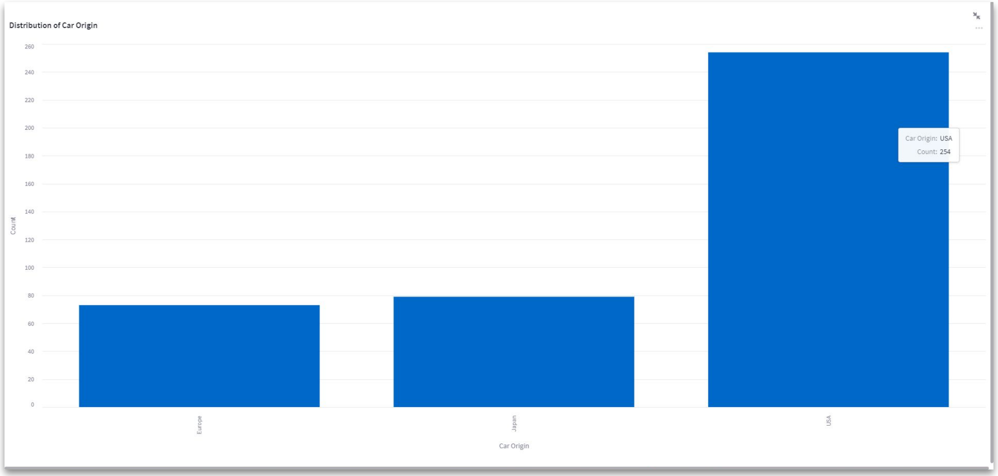

Distribution of Miles per Gallon
Mean
Frequency
0
10
20
30
40
50
0
10
20
30
40
Miles per Gallon

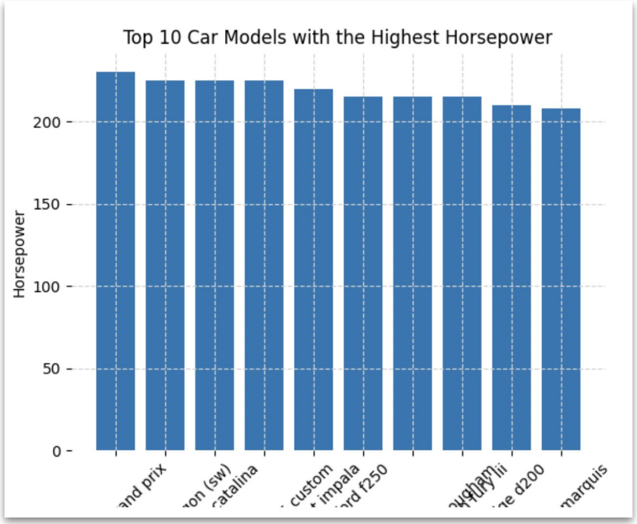

Distribution of Miles_per_Gallon
Count of Records
0
10
20
30
40
8
10
12
14
16
18
38
40
42
44
46
48
Miles_per_Gallon (bin… 18 – 20
Count of Records: 44
Miles_per_Gallon (binned)
Count of Records
44
1

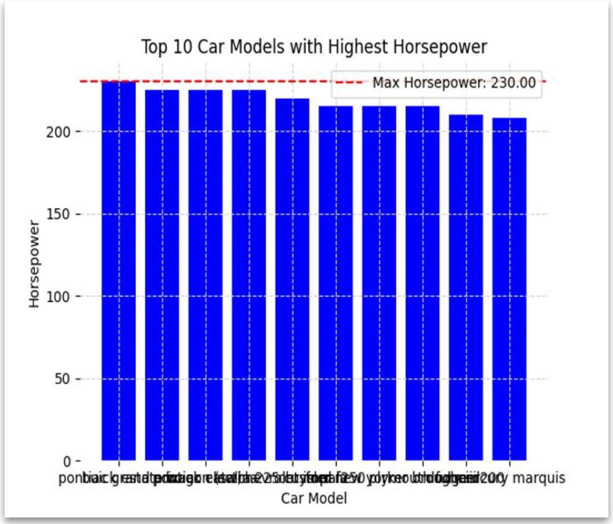


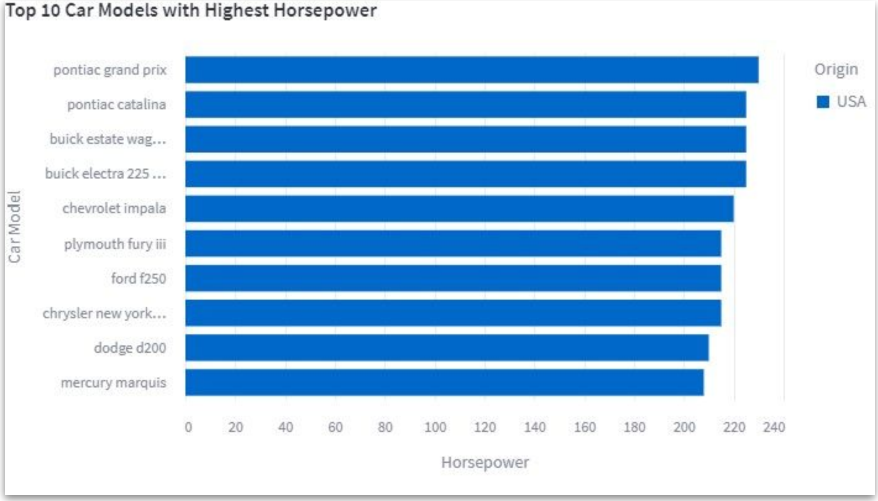
Visualization 1











The bars are flipped to accommodate the axis labelling

Findings

Easy to understand and interactive plots

Works best for small datasets (<2MB)

Direct references and simple charting to columns produce great results

Complex computations and large datasets don’t produce accurate results and sometimes completely hallucinated results.

This is a hard problem

The approach to improve visualisation aesthetics goes beyond prompt engineering. Because –

1) Different behaviour on different models
2) Each visualisation library produces different charts and interpretations of the same prompts
3) LLM systems are not truly human in the loop. Traditional selections can help alleviate some of the frustrations
4) There is no insight into why a particular model is prioritising a particular visualization over another. Similarly, why a certain visualisation library is showing the visualisation differently than another.
5) Performance issues for large dataset and computation
6) Errors are ambiguous

To-dos before final report

Implementation of visualization editor on the generated interactive chart

Comprehensive testing with more complex prompts and by incorporating different open source models (Llama and Mistral)

Thanks for listening!