

DATA MINING (2238-CSE-5334-002)
ASSIGNMENT-1
LLM, AI REPORT

Team - 1

Sai Sumana Adurugatla: 1002073766

MIR MUSTAFA ALI: 1002117402

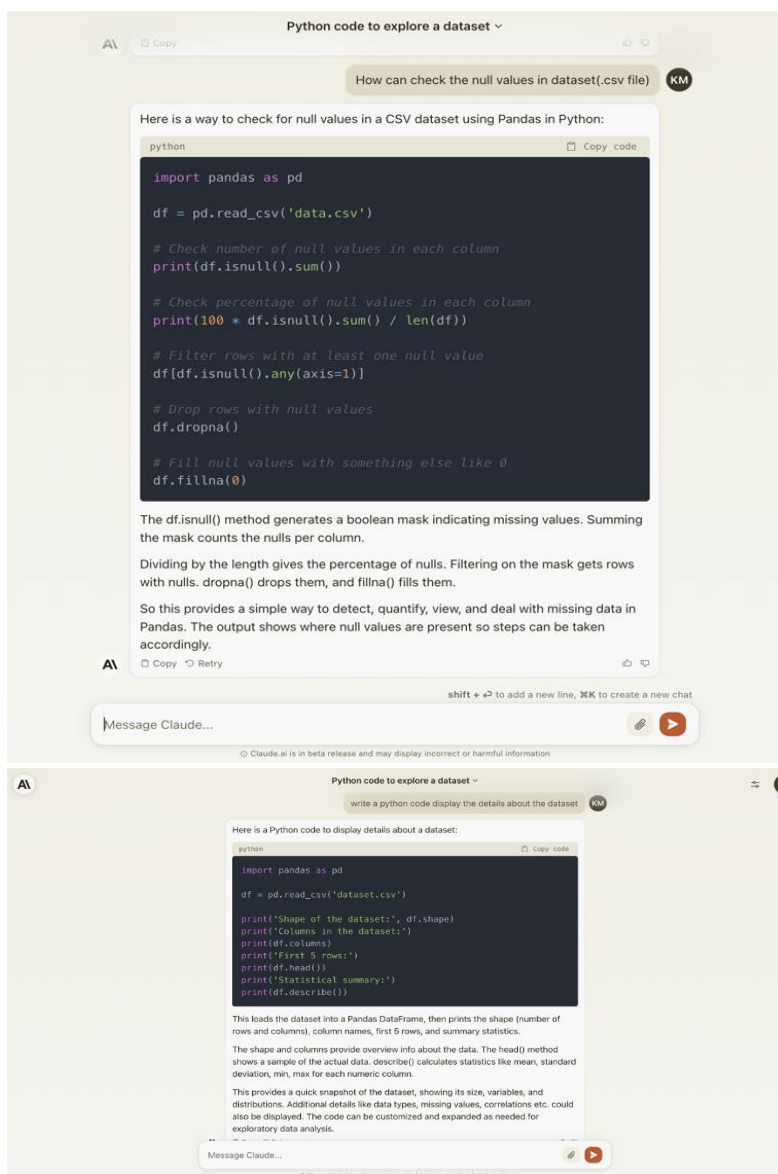
Kiran Mai Akaram: 1002115618

AI Prompts and Responses:

We have Different AI tool for various purposes. Few of them are Claude.ai, gemini.google.com, Perplexity.ai, ChatGPT, Bing AI, Blackbox.ai

1) Claude.ai


Claude is an artificial intelligence (AI) assistant created by Anthropic. With the Claude 1.3 language model at its core, it was initially made available in March 2023. But in July 2023, a revised version that was driven by the Claude 2 language model was made available. On certain benchmarks, Claude outperforms ChatGPT, but not on others. Although ChatGPT has a better premium membership with more information and capabilities than Claude, Claude's free version is still superior than ChatGPT's.



2) gemini.google.com

Similar to OpenAI's GPT, Google Gemini is a family of AI models. The primary distinction is that, in addition to being able to comprehend and produce text like other LLMs, Gemini is also capable of natively processing, combining, and understanding other types of data, including images, audio, video, and code.

After used both platforms, I have to conclude that Gemini performs somewhat better than ChatGPT at conducting web searches and incorporating the results into its responses.

 how to change data format in R language

Show drafts



Unfortunately, changing data formats in R depends on the specific type of data you have and what you want to achieve. To guide you better, I need some more information:

1. What type of data do you want to format?

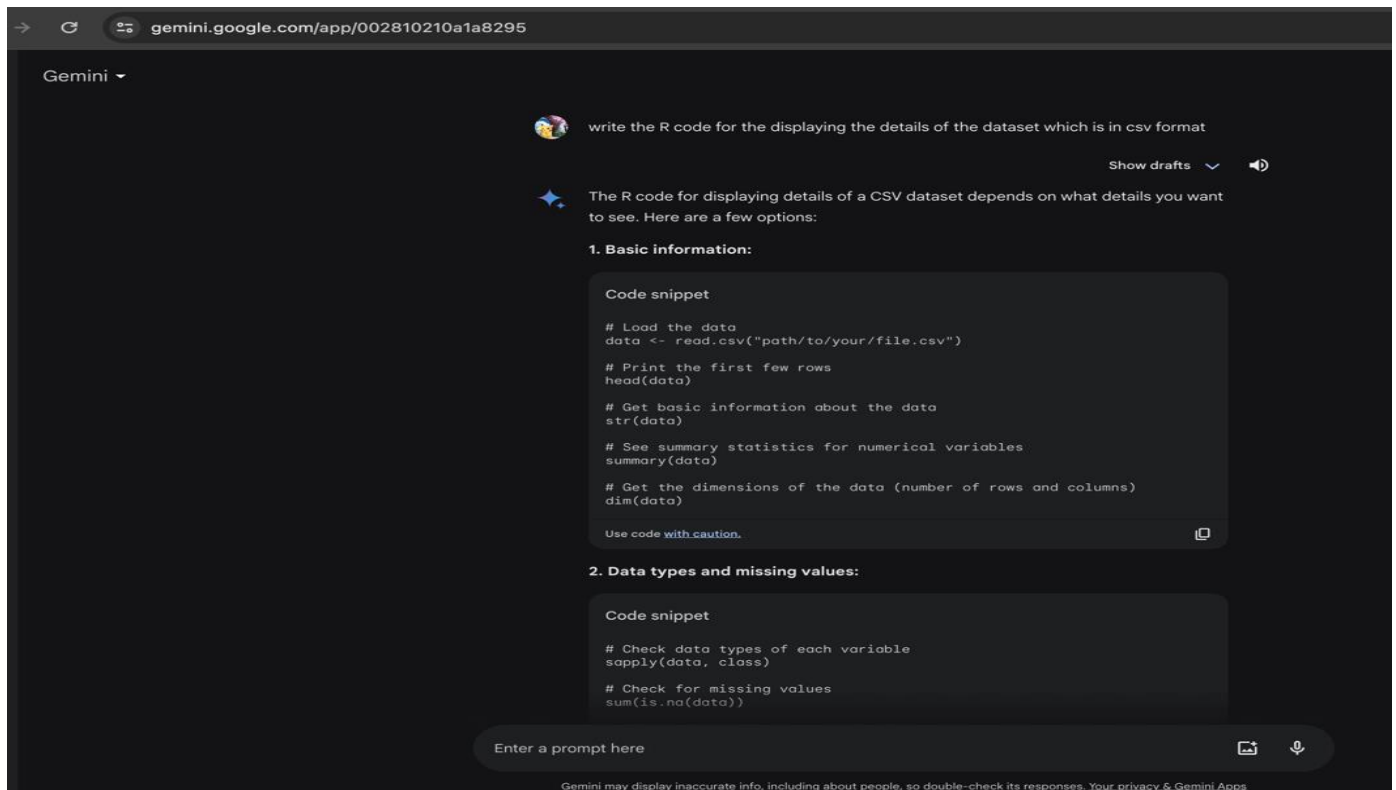
- Is it numerical data (like age, weight), dates, text, or something else?

2. What's the current format of your data?

- For example, are dates stored as "yyyy-mm-dd" or "dd/mm/yyyy"? Is text in uppercase or lowercase?

3. What is your desired format?

- Do you want to change decimal places for numbers. change date format.



The screenshot shows the Gemini AI web interface. At the top, the URL is `gemini.google.com/app/002810210a1a8295`. The user's prompt is: "write the R code for the displaying the details of the dataset which is in csv format". The AI's response is: "The R code for displaying details of a CSV dataset depends on what details you want to see. Here are a few options: 1. Basic information: [Code snippet] 2. Data types and missing values: [Code snippet]". The code snippets are as follows:

```
# Load the data
data <- read.csv("path/to/your/file.csv")

# Print the first few rows
head(data)

# Get basic information about the data
str(data)

# See summary statistics for numerical variables
summary(data)

# Get the dimensions of the data (number of rows and columns)
dim(data)
```

```
# Check data types of each variable
sapply(data, class)

# Check for missing values
sum(is.na(data))
```

At the bottom, there is a text input field labeled "Enter a prompt here" and a small disclaimer: "Gemini may display inaccurate info, including about people, so double-check its responses. Your privacy & Gemini Apps".

3)Perplexity.ai

Perplexity is a measure of how well a probability distribution or probability model predicts a sample in artificial intelligence. It is employed to assess and compare the prediction capacity of language models.

Both ChatGPT and Perplexity are incredibly strong artificial intelligence technologies with special features to assist create content and respond to queries. While ChatGPT is more appropriate for content creation and everyday use, Perplexity AI is superior at offering search engine functionality and sourcing skills.

The screenshot shows the Perplexity AI interface. The search query is "In R how can we print top 10 values by taking csv as a dataset?". The results are categorized under "Sources" and "Answer".

Sources:

- How to print the head (first 10 rows) with only... - stackoverflow - 1
- Select Top N Highest Values by Group in R - - geeksforgeeks - 2
- Chapter 6 Extracting Values from Data Frame... - michaelgastner - 3
- View 2 more

Answer:

To print the top 10 values from a CSV dataset in R, you can use the `read.csv()` function to read the CSV file into a data frame, and then use the `head()` function to display the first 10 rows of the data frame. Here's the code to achieve this:

```
R
# Load the CSV file into a data frame
data <- read.csv("your_dataset.csv")

# Print the top 10 rows of the data frame
print(head(data, 10))
```

This code first reads the CSV file into a data frame using the `read.csv()` function, and then uses the `head()` function to display the first 10 rows of the data frame.

If you want to select specific columns while printing the top 10 rows, you can do so by subsetting the data frame before using the `head()` function. For example, if you want to select only the first three columns, you can use:

```
R
# Load the CSV file into a data frame
data <- read.csv("your_dataset.csv")

# Print the top 10 rows of specific columns
print(head(data[, 1:3], 10))
```

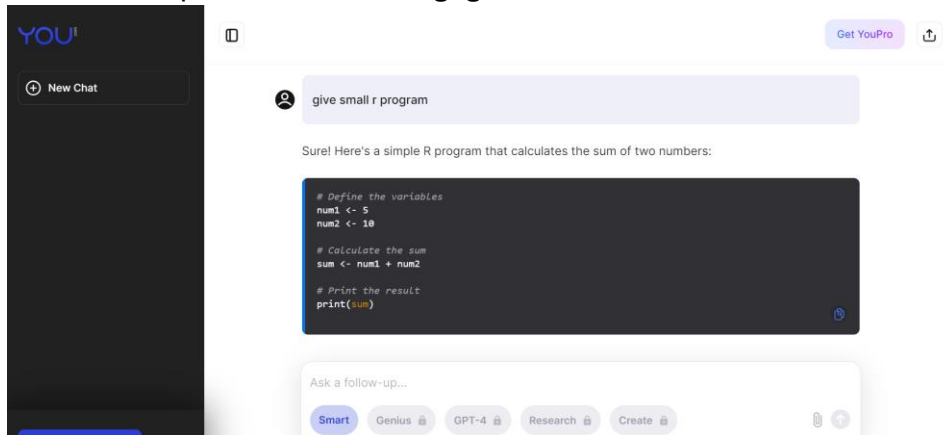
function is used to display the first 10 rows of a data frame.

The `read.csv()` function is part of base R and is used to read comma-separated files.

4) YOU ai

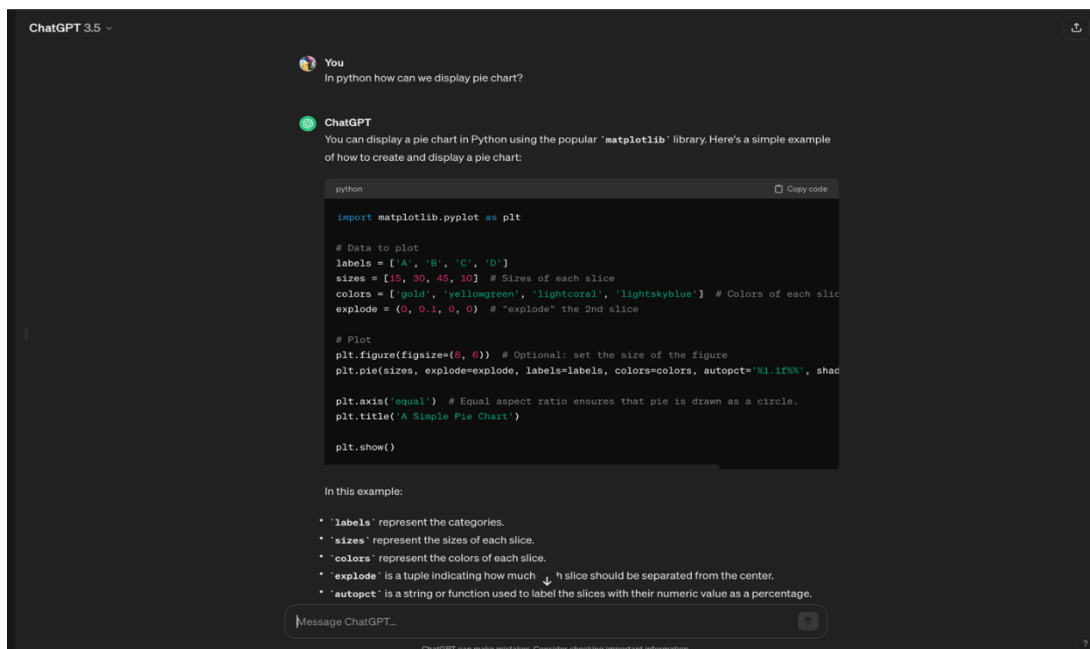
AI-driven, YouWrite is an AI writing tool intended to assist writers, novelists, and bloggers in overcoming writer's block and enhancing their talents. YouAI is intended to be a coaching and learning platform made for course developers. It can support programming in some way.

Use You.com's AI chatbot to get personalized suggestions & a personal AI search assistant. Uncover the power of AI and engage in natural conversation.



5) Chat GPT

ChatGPT is an AI-powered natural language processing tool, you may converse with the chatbot in a manner akin to that of a human and accomplish a lot more. The language model may help you with things like writing emails, essays, and programming, as well as provide answers to your inquiries. At the moment, anyone can use it for free.



6)Bing.ai

Microsoft introduced Bing Chat to enhance search outcomes and facilitate natural language query response. One way that Bing Chat sets itself apart from rivals is through the creation of both written and visual material inside the Chat. Furthermore, Bing Chat added multilingual support for their Image Creator.

The screenshot shows the Bing AI interface. At the top, there's a navigation bar with "Microsoft Bing", "SEARCH", and "COPILOT" tabs. On the right, there's a user profile "kxa5618@ma...", a notification bell with "10", and a "Mobile" button. Below the navigation bar, the user's query is "In Python how can we write to display side by side graphs?". The Copilot response starts with "Certainly! To display side-by-side graphs in Python, you can use libraries like **Matplotlib** or **Plotly**. I'll provide examples for both:" followed by a numbered list item: "1. **Matplotlib**: You can create subplots using `plt.subplot()` and adjust the layout to display graphs side by side. Here's an example:". Below this, there's a code block labeled "Python" containing the following code:

```
import numpy as np
import matplotlib.pyplot as plt

x1 = np.linspace(0.0, 5.0)
x2 = np.linspace(0.0, 2.0)
y1 = np.cos(2 * np.pi * x1) * np.exp(-x1)
y2 = np.cos(2 * np.pi * x2)

plt.subplot(1, 2, 1) # 1 row, 2 columns, first subplot
plt.plot(x1, y1, 'ko-')
plt.title('Damped oscillation')
plt.ylabel('Damped oscillation')

plt.subplot(1, 2, 2) # 1 row, 2 columns, second subplot
plt.plot(x2, y2, 'r.-')
plt.xlabel('time (s)')
plt.ylabel('Undamped')

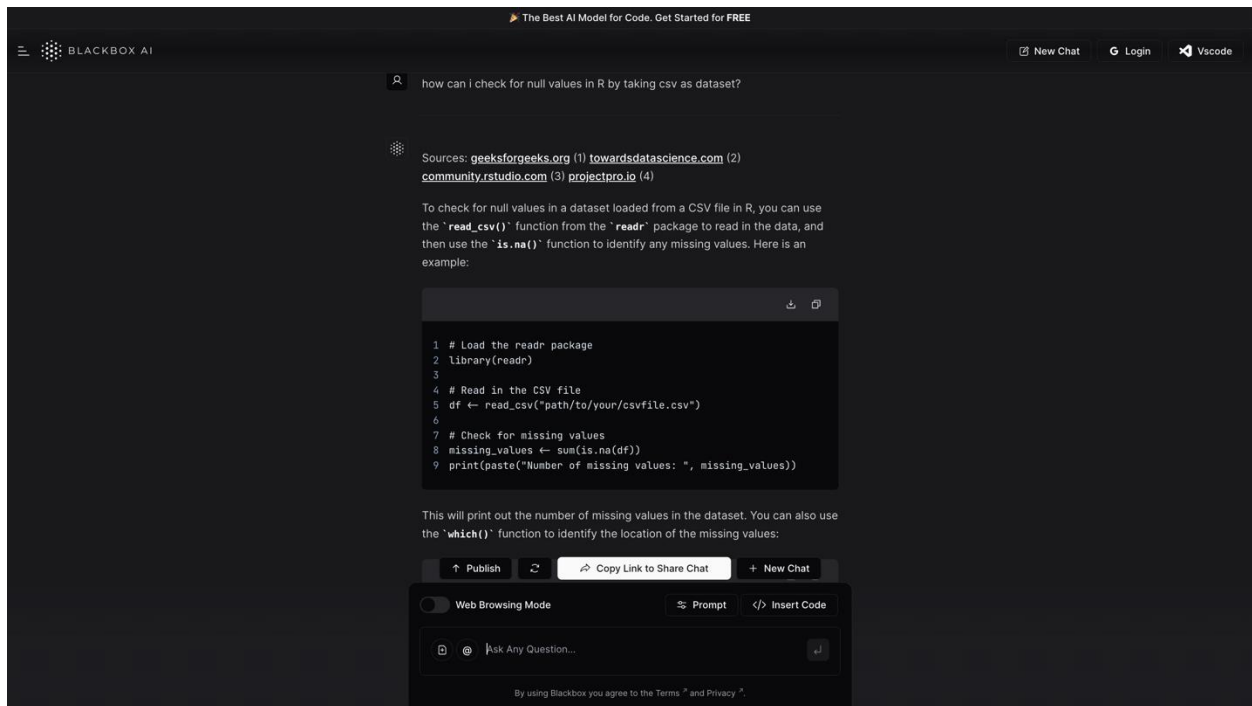
plt.tight_layout() # Minimize subplot overlap
plt.show()
```

Below the code block, there's a small text: "AI-generated code. Review and use carefully. [More info on FAQ](#)." At the bottom, there's a chat input area with a microphone icon, the text "Ask me anything...", a character count "0/2000", and a "Feedback" button.

7)Blackbox AI

Blackbox AI is commonly employed in sophisticated applications like picture identification, speech recognition, and natural language processing, where the inputs and outputs are too complex for humans to understand. The potential for extreme accuracy is one of blackbox AI's greatest benefits.

Reviewers believed that Blackbox more closely satisfies their company's needs than ChatGPT. In terms of continuing product support quality, reviewers thought Blackbox was the better choice.



LLM Report

Task 3: LLM

1. For this task you will be using the "CarPrice.csv" data file.
 2. Load the data and perform the analysis asked for each question.
 3. Load and Display the clear screenshot for each question analysis respectively.
- Note: Target variable is 'price'

```
from lida import Manager, TextGenerationConfig, llm
2.1s

lida = Manager(text_gen = llm("openai", api_key="sk-gx0ecp396ucKz1YMbF2T3B1bkFJVVRDspDsORwghLzPtec")) # !! api key
textgen_config = TextGenerationConfig(n=1, temperature=0.5, model="gpt-3.5-turbo-0301", use_cache=True)

summary = lida.summarize("C:\\Users\\suman\\Documents\\Data Mining\\SPRING2024_Section3\\SPRING2024_Section3\\CarPrice.csv",
goals = lida.goals(summary, n=2, textgen_config=textgen_config)

for goal in goals:
    display(goal)
8.3s
```

Goal 0

Question: What is the relationship between car price and engine size?

Visualization: scatter plot of engine size vs price

Goal 1

Question: What is the most common carbody type among the dataset?

Visualization: bar chart of carbody type counts

Rationale: This visualization will give us insight into the most common carbody type among the dataset, which can help us understand the preferences of car buyers. Additionally, this can help manufacturers understand which carbody types are in high demand and should be produced more.


```
# goals can also be based on a persona
persona = "a mechanic who wants to buy a car that has low price"
personal_goals = lida.goals(summary, n=1, persona=persona, textgen_config=textgen_config)
for goal in personal_goals:
    display(goal)
```

✓ 3.7s

Python

Goal 0

Question: What is the relationship between car price and engine size?

Visualization: scatter plot of engine size vs price

Rationale: This visualization will help the mechanic identify if there is a correlation between the engine size and the price of the car. It will also help them identify any outliers or patterns in the data that may be useful in their decision-making process.

```
# goals can also be based on a persona
persona = "a mechanic who wants to buy a car that has low price"
personal_goals = lida.goals(summary, n=2, persona=persona, textgen_config=textgen_config)
for goal in personal_goals:
    display(goal)
```

✓ 8.3s

Python

Goal 0

Question: What is the relationship between engine size and price for cars with diesel fuel type?

Visualization: scatter plot of engine size vs price for cars with diesel fuel type

Rationale: This visualization will help the mechanic to understand the relationship between engine size and price for cars with diesel fuel type. The mechanic can use this information to identify cars with a larger engine size that may be available at a lower price point.

Goal 1

Question: What is the distribution of curb weight for cars with different car body types?

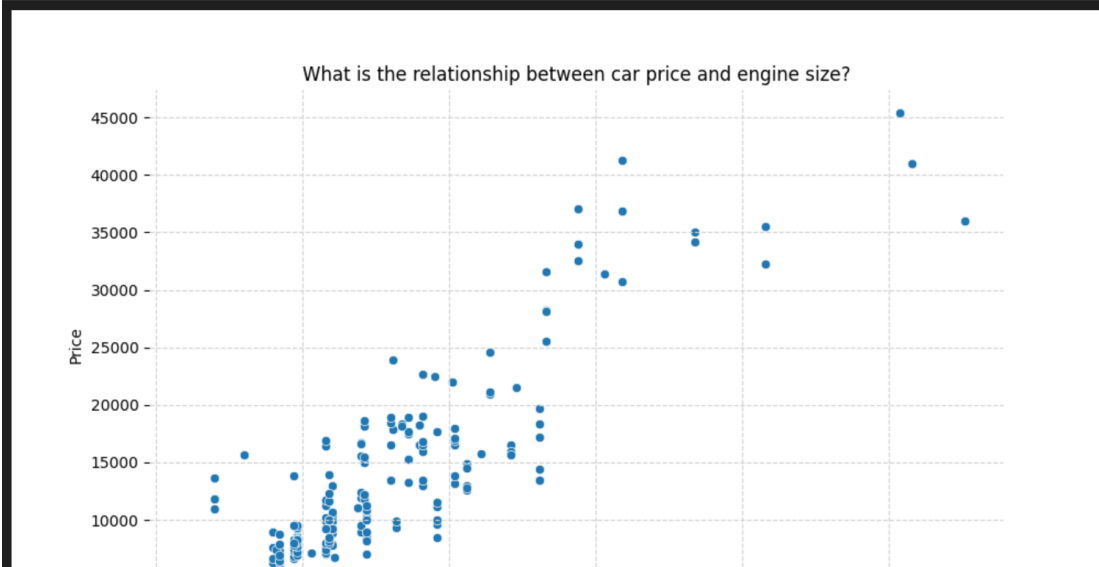
Visualization: box plot of curb weight for each car body type

Rationale: This visualization will help the mechanic to understand the distribution of curb weight for each car body type. The mechanic can use this information to identify car body types that tend to have a higher or lower curb weight, which may impact the car's performance and fuel efficiency.

Ln 2, Col 64 Spaces: 4 CRLF Cell 4 of 4 3 Spell

```
i = 0
library = "seaborn"
textgen_config = TextGenerationConfig(n=1, temperature=0.2, use_cache=True)
charts = lida.visualize(summary=summary, goal=goals[i], textgen_config=textgen_config, library=library)
charts[0]
```

✓ 11.4s

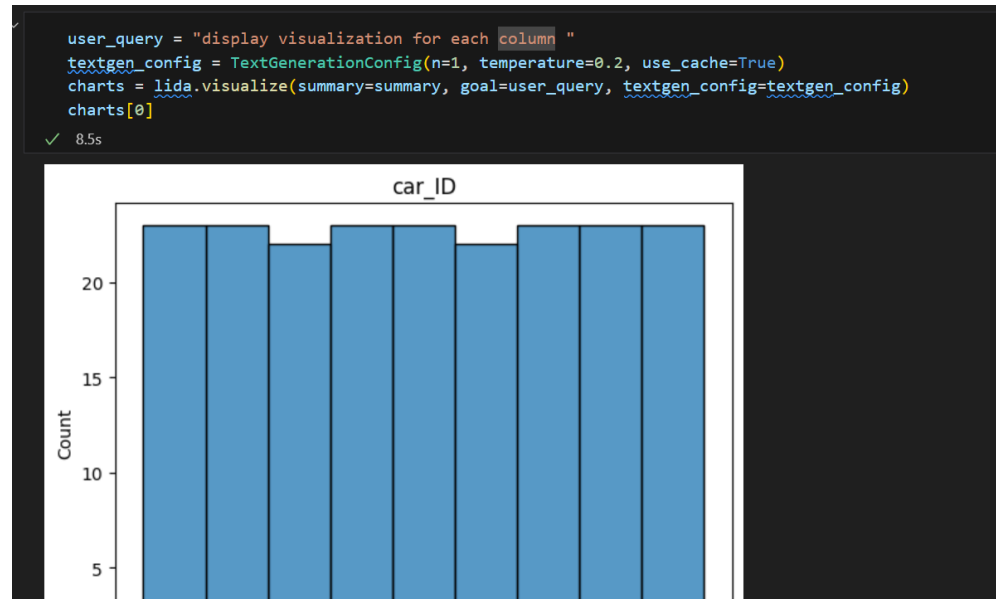


Ln 37 0 Sourcery

Ln 3, Col 41 Spaces: 4 CRLF Cell

Task 3A:

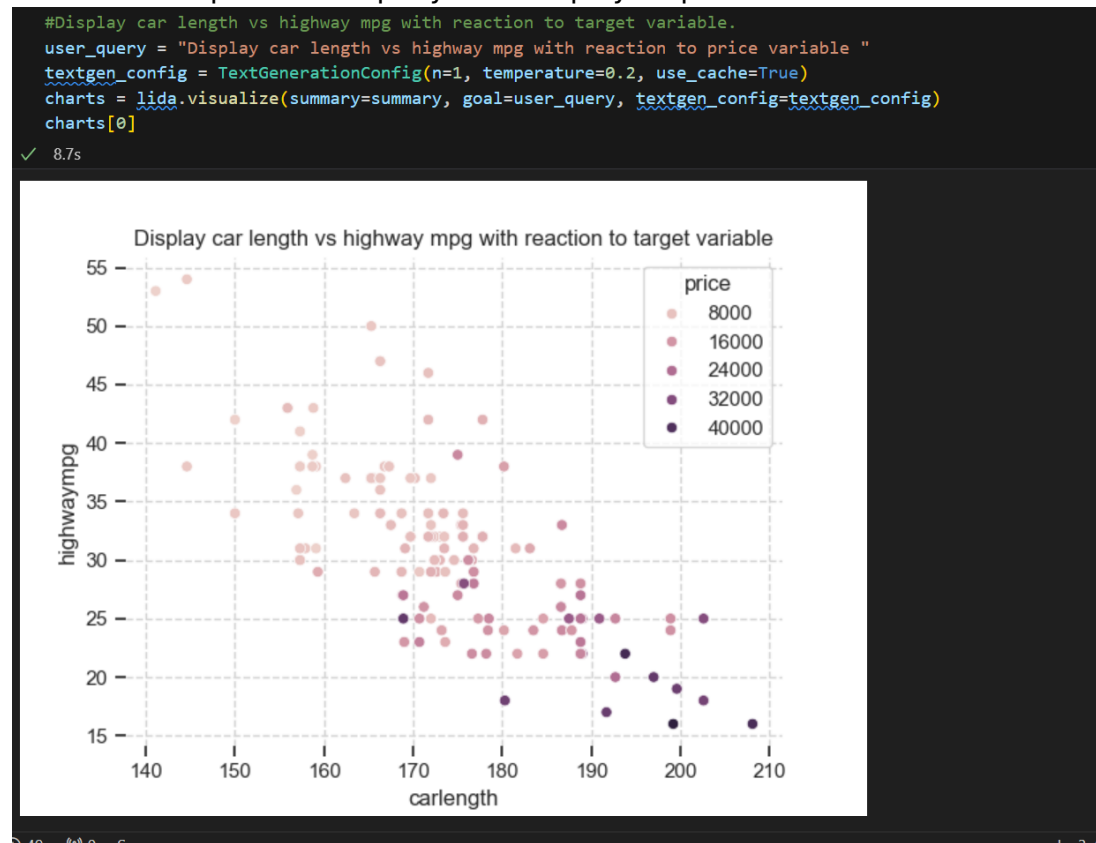
The visualization of each column in “CarPrice” dataset.



Task 3B:

Display car length vs highway mpg with reaction to target variable.

When we ask question as query it will display output we can visualize data



- The horizontal axis of the graph depicts car length, while the vertical axis represents highway miles per gallon (mpg).
- The variable of interest in this scenario is the price of the car.
- Analyzing the graph, it suggests that as the length of the car increases, the highway mpg tends to decrease.
- Within the range of 160 to 180 inches in car length, the highway mpg typically falls between 23 and 40 miles per gallon.
- Regarding the target variable, it appears that prices tend to be higher when either the car is longer, or the highway mpg is lower.

Task 3C:

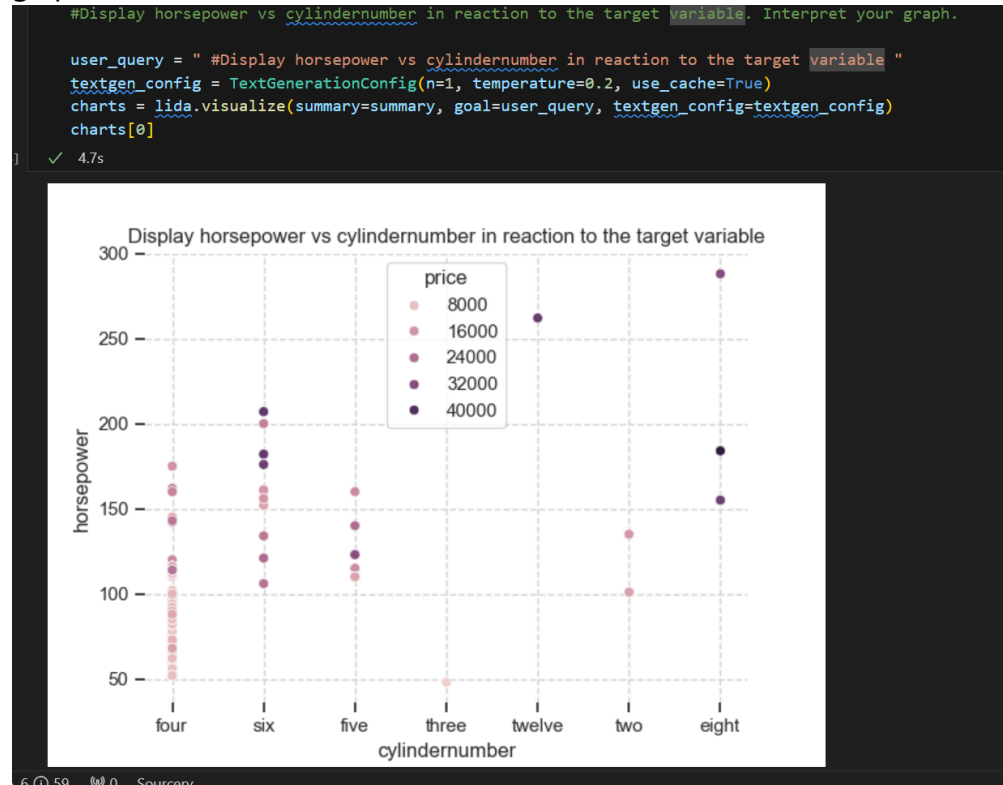
Display peak rpm vs horsepower with reaction to target variable. Interpret your graph.



- In the following graph, peak rpm is represented on the Y-axis and horsepower is on the X-axis.
- The price of the car is the target variable in this scenario. According to the graph's interpretation, horsepower tends to be moderate at peak rpm, and automobile prices are likewise within the average range.
- Neither low nor high rpm is favorable for efficient horsepower output.

Task 3D:

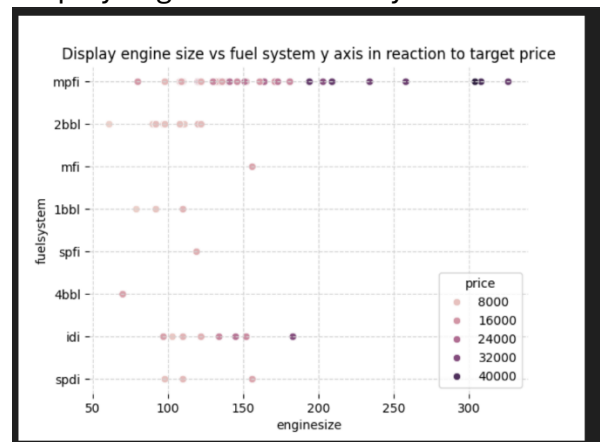
Display horsepower vs cylinder number in reaction to the target variable. Interpret your graph.



- In this picture, horsepower is indicated by the Y-axis, while cylinder number is indicated by the X-axis.
- In this instance, the car's price is the target variable.
- Typically, cars with fewer cylinders and moderate horsepower are less expensive.

Task 3E:

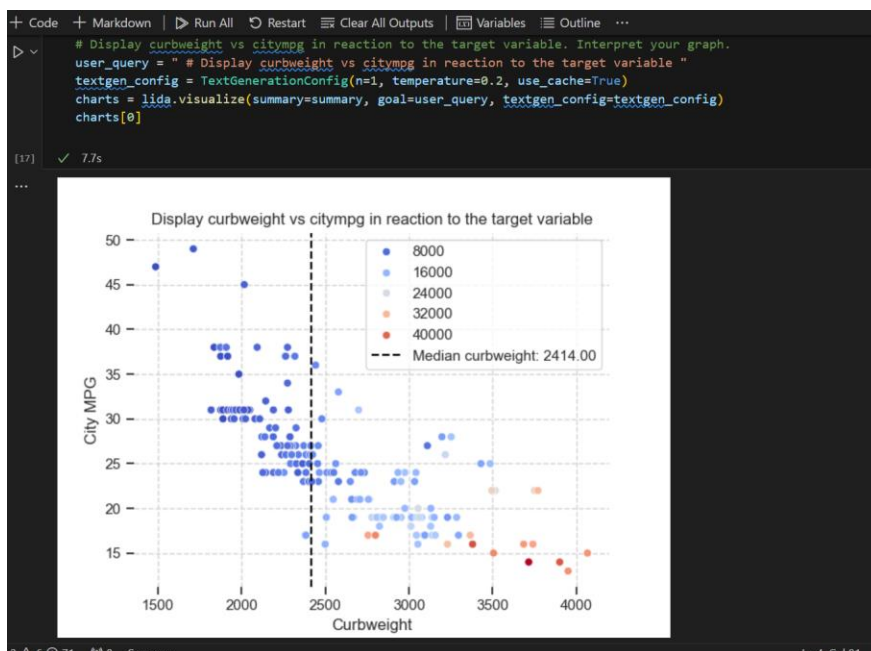
Display engine size vs fuel system in reaction to the target variable. Interpret your graph



- The fuel system is shown on the Y-axis in the following picture, while the engine size is indicated on the X-axis.
- In this instance, the car's price is the target variable.
- Mpci is the sole fuel system available for larger engines, which are expensive;
- Cars with smaller engines and mpci fuel systems are more popular than those with 2bbl or idi fuel systems.

Task 3F:

Display curbweight vs citympg in reaction to the target variable. Interpret your graph.



- In the figure below, the X-axis shows the weight of the curb, and the Y-axis reflects the city mpg.
- The target variable in this case is the car's price.
- Based on the preceding graph interpretation, we may conclude that with lower curb weights, city mpg is modest, and cars are less expensive.
- As curb weight grows, city mpg drops and prices rise.

Notebook links:

Python:

<https://colab.research.google.com/drive/1uaHDLvX656IKPk66GwNTDhdFFqRucmBL>

R:

<https://colab.research.google.com/drive/1Wl8XgFk4sUgHM65GPU01s-bjWQR2x10#scrollTo=M1FSOvHcGkpy>