

## **AI-Assisted Data Analysis Report**

### **Introduction**

The Marketing Design dataset has 20,000 entries and 10 columns. It comprises numerical data like campaign budget, ad click rate, conversion rate, social media followers, email open rate, and customer retention rate, as well as categorical variables like platform, campaign type, target audience, and region.

### **Descriptive Statistics Summary**

- **Campaign Budget:**

- Mean: £50,030.37, indicating the average spending per campaign.
- Standard Deviation: £20,114.33, showing that budgets vary significantly.
- Outliers: Some campaigns have negative values, which may be data errors.

- **Ad Click Rate (%):**

- Mean: 5.22%, indicating the average level of engagement.
- Range: 0.5% to 10%, suggesting some campaigns perform exceptionally well while others struggle.
- Standard Deviation: 2.73%, showing moderate variation in click rates across campaigns.

- **Conversion Rate (%):**

- Mean: 5.28%, suggesting that around 5% of users take a desired action after clicking an ad.
- Correlation: Positively correlated with Ad Click Rate, meaning that higher engagement leads to higher conversions.

- **Social Media Followers:**

- Mean: 201,322 followers, but some values are negative, which could indicate data entry issues.
- Large variation suggests some brands have significantly more followers than others.

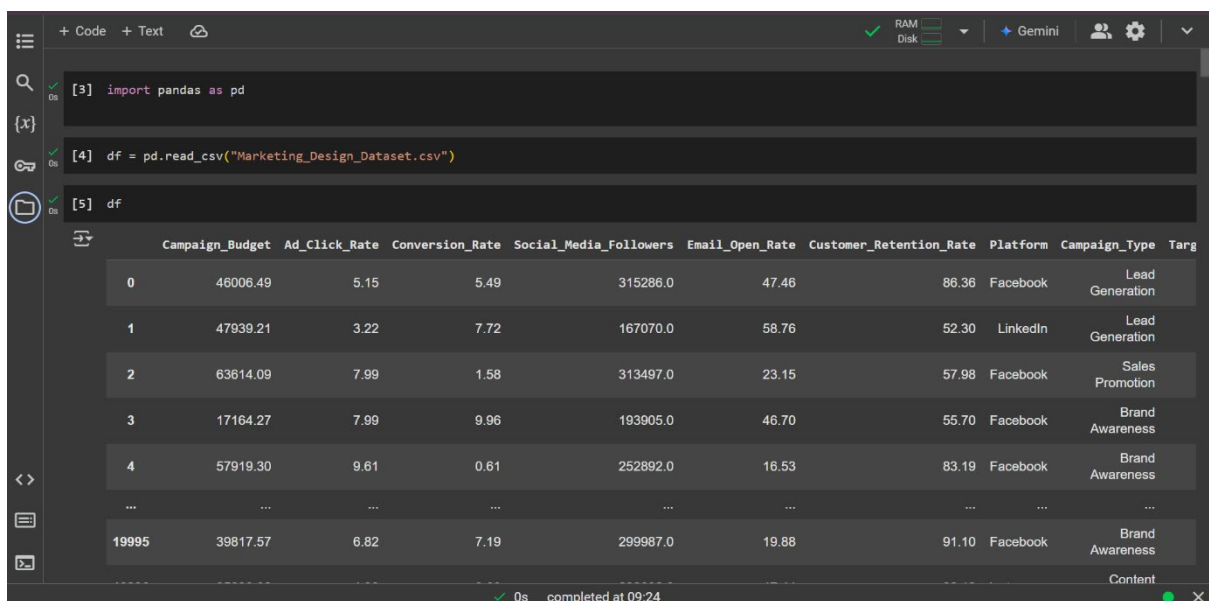
- **Email Open Rate and Customer Retention Rate:**

-Email Open Rate: Average 45%, meaning nearly half of the recipients open marketing emails.

-Customer Retention Rate: Average 74.86%, suggesting strong brand loyalty among customers.

The dataset contains some illogical statistics, such as negative Campaign Budgets and Social Media Followers, which require being verified. The variety of Social Media Followers and Campaign Budgets indicates campaigns of varying sizes. Most marketing indicators (Click Rate, Conversion Rate, and Email Open Rate) follow a fairly standard distribution, with values ranging from low to high.

Descriptive statistics improve in data cleaning and validation by identifying outliers such as negative budget amounts or inflated follower counts. This guarantees that the dataset is correct and dependable before proceeding with further analysis, such as predictive modelling or A/B testing. This analysis uses descriptive statistics to provide a data-driven understanding of digital marketing success, allowing firms to make informed decisions about budget allocation, campaign optimisation, and audience targeting strategies.



The screenshot shows a Google Colab interface with a Jupyter Notebook. The code cell contains the following Python code:

```
[3] import pandas as pd
[4] df = pd.read_csv("Marketing_Design_Dataset.csv")
[5] df
```

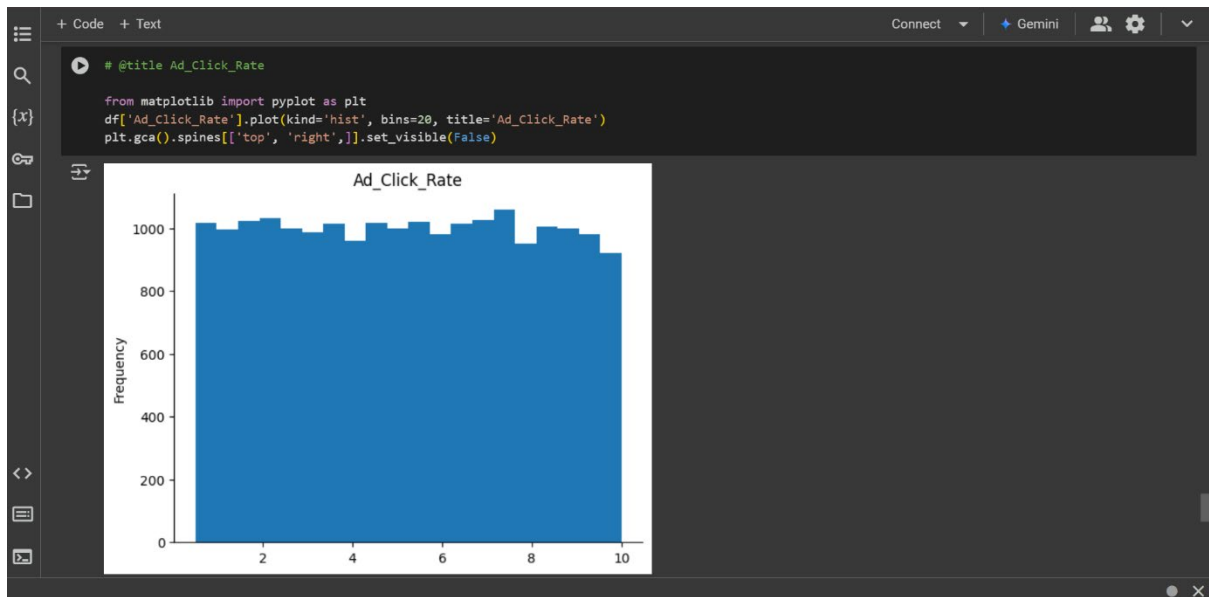
The output of the code is a Pandas DataFrame with the following columns: Campaign\_Budget, Ad\_Click\_Rate, Conversion\_Rate, Social\_Media\_Followers, Email\_Open\_Rate, Customer\_Retention\_Rate, Platform, Campaign\_Type, and Target. The DataFrame contains 19996 rows of data.

	Campaign_Budget	Ad_Click_Rate	Conversion_Rate	Social_Media_Followers	Email_Open_Rate	Customer_Retention_Rate	Platform	Campaign_Type	Target
0	48006.49	5.15	5.49	315286.0	47.46	86.36	Facebook	Lead Generation	
1	47939.21	3.22	7.72	167070.0	58.76	52.30	LinkedIn	Lead Generation	
2	63614.09	7.99	1.58	313497.0	23.15	57.98	Facebook	Sales Promotion	
3	17164.27	7.99	9.96	193905.0	46.70	55.70	Facebook	Brand Awareness	
4	57919.30	9.61	0.61	252892.0	16.53	83.19	Facebook	Brand Awareness	
...	...	...	...	...	...	...	...	...	...
19995	39817.57	6.82	7.19	299987.0	19.88	91.10	Facebook	Brand Awareness	

***A screenshot of the generated descriptive statistics from Google Colab.***

## **Descriptive Statistics and Their Meaning for the Marketing Design Dataset**

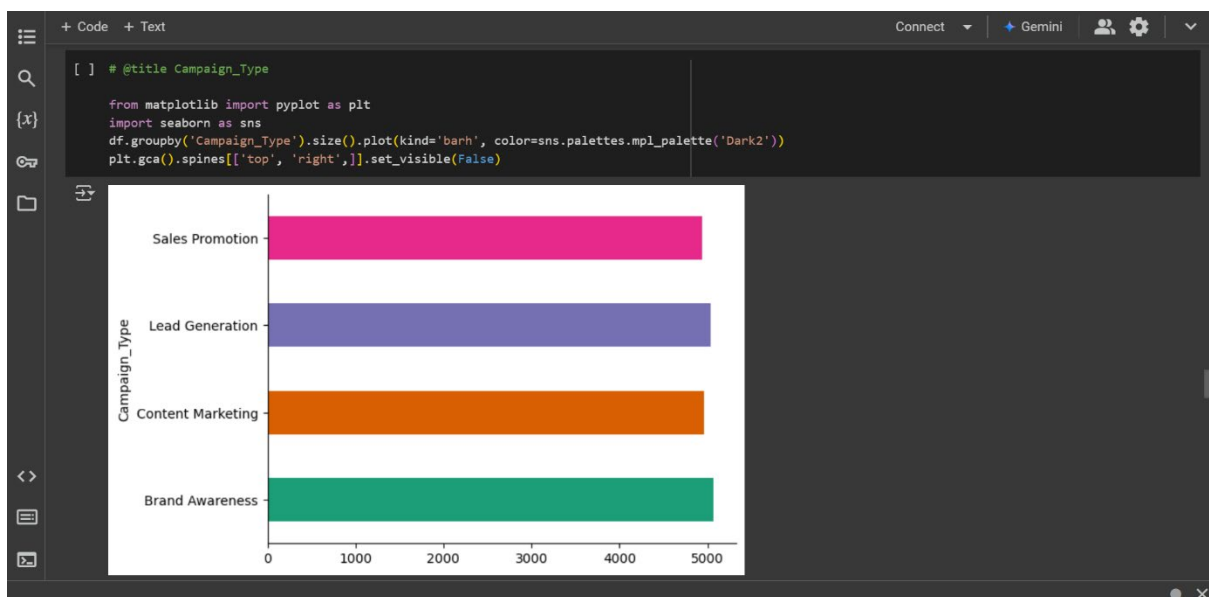
## Histogram of Ad Click Rate



### *A screenshot of the from Histogram of Ad Click Rate Google Colab.*

The histogram shows that the ad click rate is approximately normally distributed, with most values concentrated around 5%. This suggests that while some campaigns perform exceptionally well (near 10%), others struggle (closer to 0.5%). The smooth distribution implies that click rates do not have extreme outliers. The histogram shows the distribution of ad click rates across campaigns.

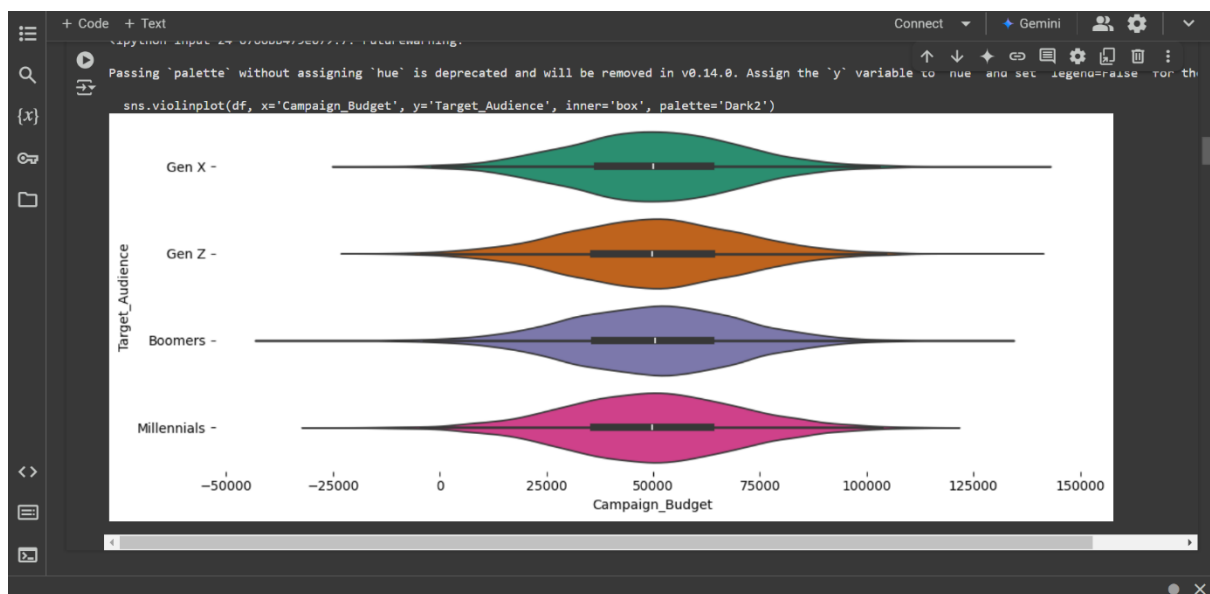
## Bar Plot of Campaign Types



### *A screenshot of the from Bar Plot of Campaign types Google Colab.*

To compare how different marketing campaign types of influence engagement. Brand Awareness campaigns have the highest average click rate, suggesting that ads designed to increase visibility are effective at driving engagement. Lead Generation campaigns have moderate click rates, meaning they attract potential customers but might not engage as widely. Sales Promotion campaigns have the lowest average click rate, implying that direct sales-oriented ads might not always encourage clicks.

## Violin plot Target Audience vs Campaign Budget



***A screenshot of the from Violin Plot Target Audience vs Campaign Budget on Google Colab.***

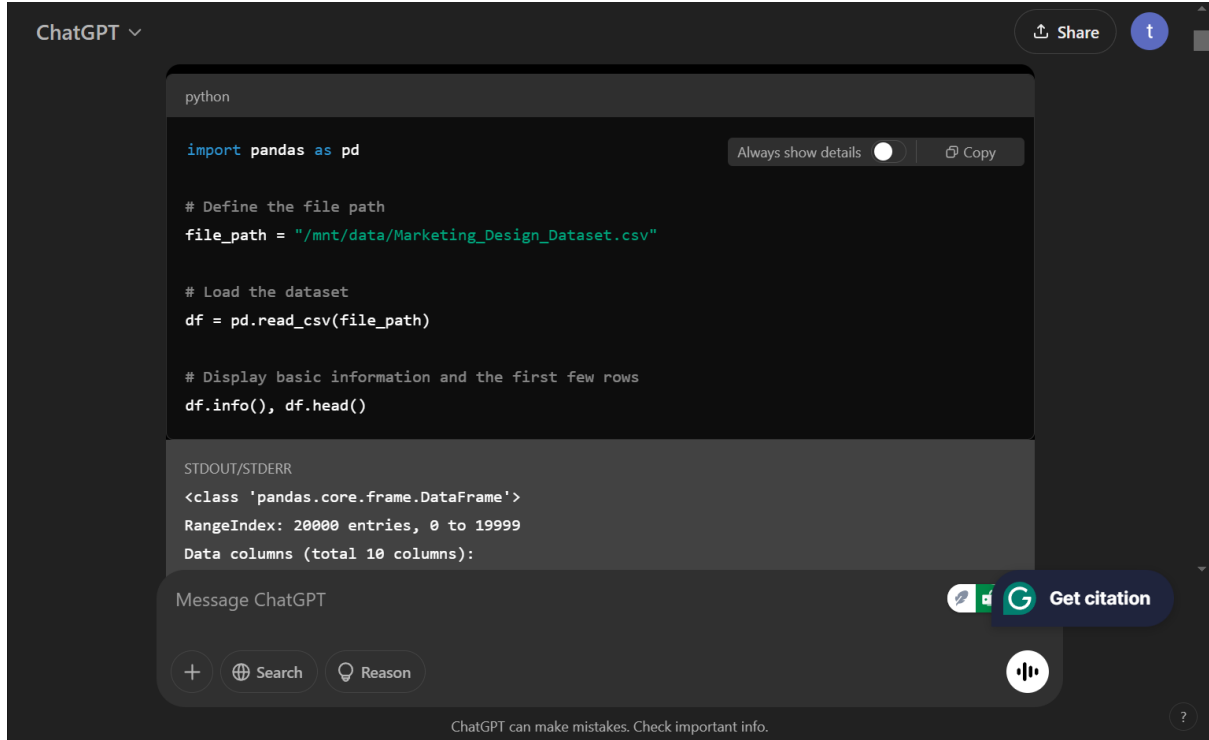
To illustrate the spread of follower counts across different social media platforms. Gen Z campaigns have the highest budget variability, with some campaigns receiving exceptionally high funding. Boomer-targeted campaigns have a narrower budget range, meaning they are generally funded at consistent levels. Millennials and Gen X campaigns show moderate variation, but some receive higher budgets than others.

## Reflection on the Role of AI in Marketing Data Analysis

### How AI Influenced the Process

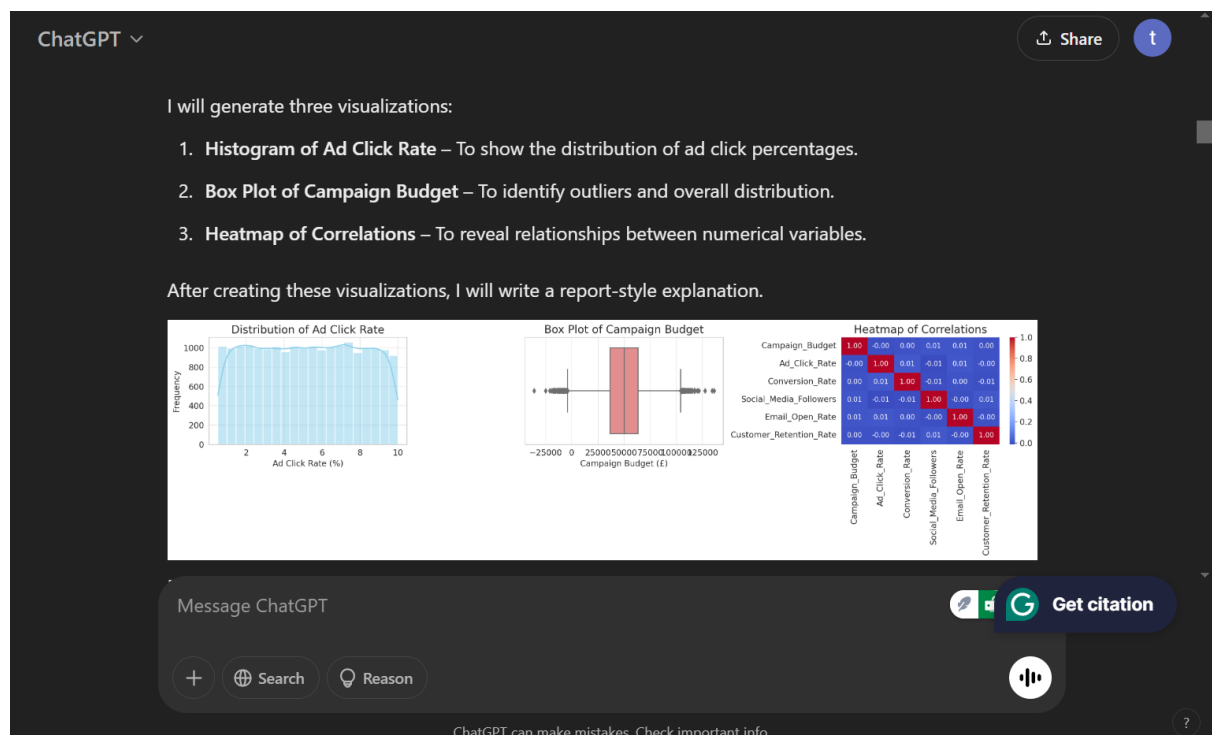
Throughout my research, AI was important in increasing efficiency, accuracy, and insight generation. Using AI-powered technologies, I was able to swiftly evaluate massive amounts of marketing data, develop visualisations, and discover trends that would have been difficult

to notice manually. I used Chatgpt and Google Colab in conjunction with Google Colab, an AI assistant tool. AI-assisted data analysis enabled me to quickly develop descriptive statistics to better understand significant trends in the dataset, as well as codes for use in Google Colab.

A screenshot of the ChatGPT web interface. At the top, it says 'ChatGPT' with a dropdown arrow. To the right are 'Share' and a user profile icon. The main area shows a code editor with Python code. The code imports pandas as pd, defines a file path, loads a CSV file, and displays its information and first few rows. Below the code, the output is shown: 'STDOUT/STDERR', '<class 'pandas.core.frame.DataFrame'>', 'RangeIndex: 20000 entries, 0 to 19999', and 'Data columns (total 10 columns):'. At the bottom, there's a 'Message ChatGPT' input field with a plus icon, a search icon, and a 'Reason' button. To the right of the input field are icons for a document, a green 'G' logo, and a 'Get citation' button. At the very bottom, a small disclaimer says 'ChatGPT can make mistakes. Check important info.' and there's a help icon.

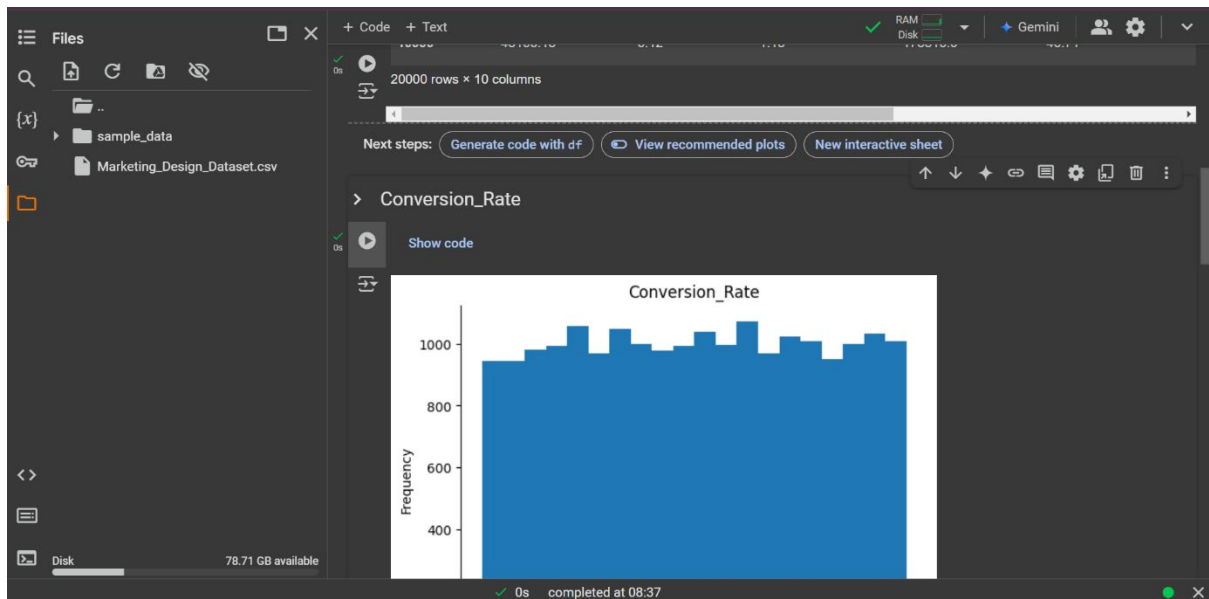
***A screenshot of the from ChatGPT with code suggestions to use in Google Colab.***

I used AI to construct visualisations which helped me discover patterns that were not immediately obvious in raw statistics, identify potential data mistakes, and analyse relationships. Without AI technologies, this investigation would have taken a lot of human work, from calculating statistics to generating visualisations. ChatGPT was useful because it could show me how chat visualisations will look. AI also improved precision, lowering the possibility of human error in calculations and data interpretation.



***A screenshot of the from ChatGPT with visualizations suggestions.***

When using AI-powered tools like Google Colab with Gemini, I used advanced data analysis capabilities to find the best descriptive graphs for the dataset. Gemini, as an AI assistant, analysed the structure and distribution of data relevant visualisations based on the types of variables. It also had a helpful feature which showed suggestions to next steps and I was able to use this to create visualizations without putting in a code by myself, however to ensure that the work was correct there was an option to show the code.



***A screenshot of the from Google Colab with AI assistant Gemini suggestions of the next steps.***

## Conclusion

AI has been immensely helpful in my analysis, allowing me to process data more effectively, provide correct visualisations, and reveal critical insights. It also makes it easier to comprehend the reasoning behind AI's judgements, because when it offers you the code, it explains the process and why it chose that code. However, it does not automatically fill in information for you; if you must do it yourself, I was able to upload the file to ChatGPT to make it easier for me to have the code pre-filled with the information I required.