**Benchmarking LLMs**

Prompt 1

Persona: You are a data visualization specialist. Context: You have been tasked with obtaining climate data about the city of Nairobi. Action: Access the Wikipedia page for the city of Nairobi. Find the climate data for the Jomo Kenyatta International Airport. Download/scrape the data. Plot a line graph with month on the x-axis and temperature in degrees celsius (C) on the y-axis. The line graph should contain three lines with values for the daily mean, mean daily maximum, and mean daily minimum. The style/theme for the plot should be NY Times. Write an R script that obtains, cleans, and visualizes the data. Tone: Professional. Output Format: Provide an R script.

**LLM Data Analysis**

1. LLM Prompt Template developed by DigitalSreeni ()

Prompt 1

Persona: You are a data analyst working on exploratory data analysis at a newly funded startup. Context: You have been tasked with working on the penguins dataset and developing new techniques for EDA. Action: Write an R script to analyze the penguins the dataset, clean and process the data with tidyverse, perform EDA with plots, and finally generate a pairplot. Tone: Professional. Output Format: Provide an R script.

Prompt 2

Persona: You are a GIS data analyst working on spatial mapping at a newly funded startup. Context: You have been tasked with plotting county populations in the USA. Action: Write an R script to load the required packages (usmap) and datasets (county population data from the usmaps package), as well as plot the map using the economist theme. Tone: Professional. Output Format: Provide an R script.

Prompt 3

Persona: You are a data scientist working on machine learning models at a newly funded startup. Context: You have been tasked with working on the penguins dataset and working on machine learning classification models. Action: Write an R script to analyze the penguins the dataset, clean and process the data with tidyverse, perform EDA with plots, and use tidymodels to design 5 different classification models including xgboost. Additionally, perform hyperparameter tuning and optimization to get the best model. Provide a table to compare the classification models. Use concepts of interpretable machine learning to evaluate the model performance. Additionally, test the code before output. Tone: Professional. Output Format: Provide an R script.