# Cheatsheet: Data Analytics and Generative AI

### Important Terms

| Term | Description |
|---|---|
| **Generative AI** | Is a category of AI that focuses on creating new, synthetic data. Unlike traditional AI models that predict or classify, generative models generate entirely new data points, opening a realm of possibilities for data analytics. |
| **Data Augmentation** | Is a powerful technique for improving the performance of machine learning models, especially when the training data sets of the models are small or imbalanced. |
| **Data Preparation** | Is crucial in the data analytics journey. To prepare raw data for analysis, it must be cleaned, transformed, and arranged in a format that makes it easy for analytical tools to use. |
| **Database Querying** | Is the process of working with a database to extract relevant details for analysis. It includes interacting with a database and retrieving data that satisfies certain criteria using query languages, most often SQL (Structured Query Language). |
| **Data Q&A** | Asking questions and getting answers about certain data sets or data analysis activities is called Q&A for data. With Q&A, you can accomplish data exploration, extract insights, and get a better comprehension of the underlying patterns and trends in the data. |

---

### Generative AI Platforms/Tools Used in this Module

| Task Performed | Generative AI Pltform/Tools |
|---|---|
| Data Augmentation and Generation | DataRobot<br>Colab<br>ChatGPT<br>Bard<br>MOSTLY.AI<br>universaldata |
| Data Preparation | ChatCSV<br>Tomat.AI |
| Data Querying | SQLthroughAI<br>dbsensei |
| Data Insights through QnA | Akkio |

---

### Some Generic Prompts

| Task | Prompt | Example |
|---|---|---|
| Create dataset in a particular domain | Create <…..> dataset for <…..> | Create *patient* data set for the *symptoms of diabetes* |
| Create dataset with specific attributes and format | Create a dataset with attributes as <…..> in a <…..> format | Create a dataset with attributes as *temperature in Fahrenheit, Temperature Category, Humidity in percentage, Rain and Snow as categorical type in yes or no categories, Month, Year* in a *CSV* format |
| Get the insights through Q&A:<br>Finding highest value within the data attribute | Identify the <…..> with the highest <…..> | Identify the products with the highest sales<br>What are the top-selling products? |
| Get the insights through Q&A:<br>See patterns of a data attribute over a period of time | How has <…..> changed over time? | How has the quantity ordered changed over time? |
| Identify missing data | Write a <…> code to Identify <…..> missing values.<br><br>Identify the attributes with missing data | Write a *python* code to Identify *the columns* with missing values<br>(ChatGPT)<br>Identify the attributes with missing data<br>(ChatCSV) |
| Handling missing values | Write a <…> code to replace missing values with <…..> in the dataset<br><br>Replace the missing values <…> in the <…> and save the updated dataset | Write a *python* code to replace missing values with *mean values* in the dataset<br>(ChatGPT)<br>Replace the missing values with *the mean value* in the *Screen_size_cm column* and save the updated dataset<br>(ChatCSV) |
| Join two tables | Write a SQL query to join <…..> with <…..> on the <…..> as a primary key<br><br>join <…..> with <…..> on <…..> as primary key | Write an SQL query to join *customer table* with the *product sales tables* on *product ID* as primary key<br>(ChatGPT)<br>Join *customer table* with the *product sales table* on *product ID* as primary key<br>(dbsensei) |
| Create database | Write a <…..> query to Create a database on <…..>, create a <…> and insert values in these tables | Write a *SQL* query to Create a database on *sales*, create a *customer table*, *a sales table*, *a product table* and insert values in these tables<br>(ChatGPT) |

| Task | Prompt | Example |
| --- | --- | --- |
| | Create a database on <….>, create tables, insert values | Create a database on *customers and sales*, create tables, insert values (dbsensei) |

## Author(s)

Dr. Pooja