

# Module 2 Cheatsheet: Use of Generative AI for Data Science

## Popular GenAI tools

Name of model	Usage	Link
Hal9	EDA tool to identify key insights on data	<a href="https://www.hal9.com/">https://www.hal9.com/</a>
Columns.ai	Data visualization tool to create useful charts	<a href="https://columns.ai/">https://columns.ai/</a>
Akkio	Data visualization tool to create data plots like regression plots, box plots, correlation heatmaps, and so on	<a href="https://www.akkio.com/">https://www.akkio.com/</a>

## Important prompts for generating data insights and visualizations

Task	Prompt
Generate a statistical description of data.	Write a Python code to generate the statistical description of all the features used in the data set. Include "object" data types as well.
Create regression plots between a target variable and a continuous valued source variable.	Write a Python code to generate a regression plot between a target variable and a source variable of a data frame.
Create box plots between a target and categorical source variable.	Write a Python code to generate a box plot between a target variable and a source variable of a data frame.
Evaluate parametric interdependence using correlation, p-value and pearson coefficient	Write a Python code to evaluate correlation, pearson coefficient, and p-values for all attributes of a data frame against the target attribute.
Group variables to create pivot tables. Create a p-color plot for the pivot table.	Write a Python code that performs the following actions: 1. Groups three attributes as available in a data frame df. 2. Creates a pivot table for this group, using a target attribute and aggregation function as mean. 3. Plots a pcolor plot for this pivot table.

## Important prompts for model development and refinement

Task	Prompt
------	--------

Task	Prompt
Linear regression between a single source attribute and target attribute and evaluate it	<p>Write a Python code that performs the following tasks:</p> <ol style="list-style-type: none"><li>1. Develops and trains a linear regression model that uses one attribute of a data frame as the source variable and another as a target variable.</li><li>2. Calculates and displays the MSE and R^2 values for the trained model.</li></ol>
Linear regression between multiple source attributes and target attributes and evaluate it	<p>Write a Python code that performs the following tasks:</p> <ol style="list-style-type: none"><li>1. Develops and trains a linear regression model that uses some attributes of a data frame as the source variables and one of the attributes as a target variable.</li><li>2. Calculates and displays the MSE and R^2 values for the trained model.</li></ol>
Polynomial regression model with single source and target variable	<p>Write a Python code that performs the following tasks:</p> <ol style="list-style-type: none"><li>1. Develops and trains multiple polynomial regression models, with orders 2, 3, and 5, that use one attribute of a data frame as the source variable and another as a target variable.</li><li>2. Calculates and displays the MSE and R^2 values for the trained models.</li><li>3. Compares the performance of the models.</li></ol>
Pipeline creation for scaling, polynomial feature creation, and linear regression	<p>Write a Python code that performs the following tasks:</p> <ol style="list-style-type: none"><li>1. Create a pipeline that performs parameter scaling, polynomial feature generation, and linear regression. Use the set of multiple features as before to create this pipeline.</li><li>2. Calculate and display the MSE and R^2 values for the trained model.</li></ol>
Grid search with ridge regression and cross validation	<p>Write a Python code that performs the following tasks:</p> <ol style="list-style-type: none"><li>1. Use polynomial features for some of the attributes of a data frame.</li><li>2. Perform a grid search on a ridge regression model for a set of values of hyperparameter alpha and polynomial features as input.</li><li>3. Use cross-validation in the grid search.</li><li>4. Evaluate the resulting model's MSE and R^2 values.</li></ol>