4/9/24, 7:59 p.m. about:blank

Module 2 Cheatsheet: Use of Generative AI for Data Science

Popular GenAI tools

Name of mode!LinkHal9EDA tool to identify key insights on datahttps://www.hal9.com/Columns.aiData visualization tool to create useful chartshttps://columns.ai/AkkioData visualization tool to create data plots like regression plots, box plots, correlation heatmaps, and so on https://www.akkio.com/

Important prompts for generating data insights and visualizations

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

Important prompts for model development and refinement

Task **Prompt** Write a Python code that performs the following tasks: Linear regression between a single source attribute and 1. Develops and trains a linear regression model that uses one attribute of a data frame as the source variable target attribute and evaluate it and another as a target variable. 2. Calculates and displays the MSE and R^2 values for the trained model. Write a Python code that performs the following tasks: Linear regression between multiple source attributes 1. Develops and trains a linear regression model that uses some attributes of a data frame as the source and target attributes and evaluate it variables and one of the attributes as a target variable. 2. Calculates and displays the MSE and R^2 values for the trained model. Write a Python code that performs the following tasks: 1. Develops and trains multiple polynomial regression models, with orders 2, 3, and 5, that use one attribute Polynomial regression model with single source and of a data frame as the source variable and another as a target variable. target variable 2. Calculates and displays the MSE and R^2 values for the trained models. Compares the performance of the models. Write a Python code that performs the following tasks: Pipeline creation for scaling, polynomial feature 1. Create a pipeline that performs parameter scaling, polynomial feature generation, and linear regression. creation, and linear regression Use the set of multiple features as before to create this pipeline. 2. Calculate and display the MSE and R^2 values for the trained model. Write a Python code that performs the following tasks: 1. Use polynomial features for some of the attributes of a data frame. 2. Perform a grid search on a ridge regression model for a set of values of hyperparameter alpha and Grid search with ridge regression and cross validation polynomial features as input. 3. Use cross-validation in the grid search.

4. Evaluate the resulting model's MSE and R^2 values.

Author(s)

Abhishek Gagneja



about:blank 1/1