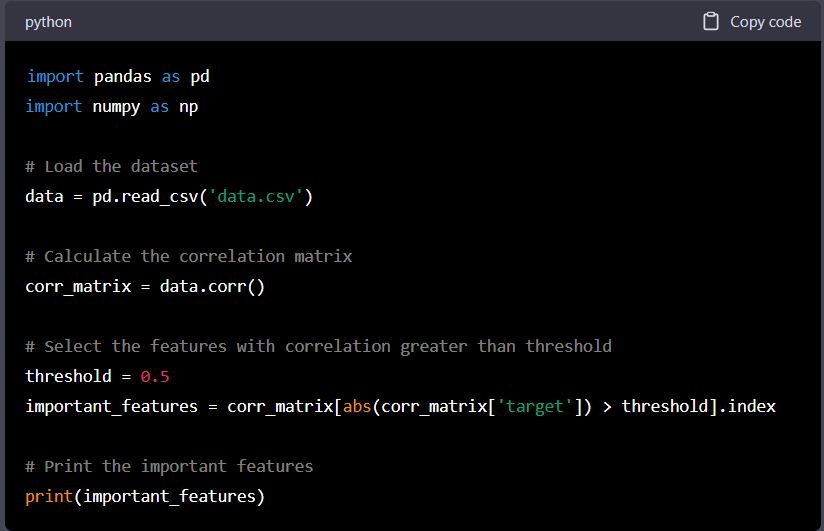
***Feature Selection***

1. ***Correlation Method:***

The correlation method of feature selection is a statistical technique that evaluates the linear relationship between features and the target variable. It measures the strength and direction of the relationship between the two variables, and features with high correlation to the target variable are considered important for the predictive model.

Here's an example code that demonstrates how to perform feature selection using correlation method:



In this code, we first load the dataset using pandas library. Then, we calculate the correlation matrix using the corr() method of pandas. The resulting correlation matrix is a table that shows the correlation coefficients between all pairs of features in the dataset. Next, we select the features with correlation greater than a threshold value (in this case, 0.5) using the abs() and index methods. The resulting important\_features variable contains a list of the features that are important for the predictive model.

Finally, we print the important features using the print() method.

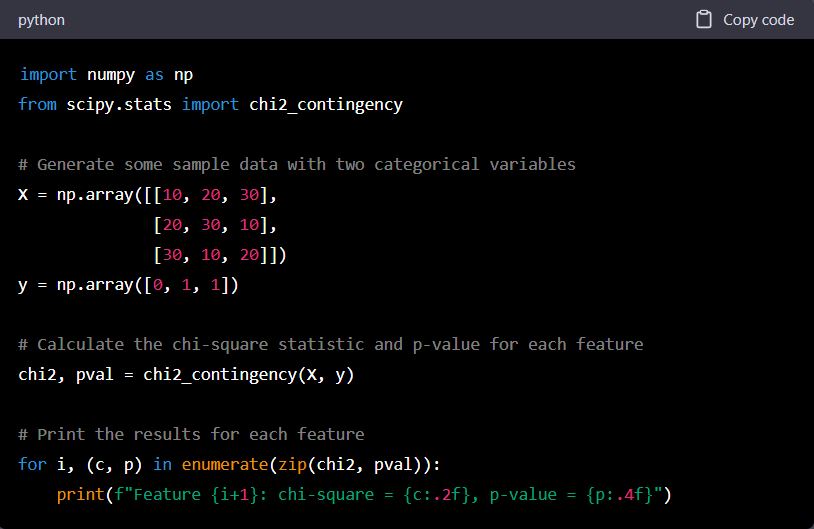
Note that this is a simple example, and there are several other factors to consider when performing feature selection, such as the distribution of the data, the presence of outliers, and the correlation between features. Additionally, correlation method only captures linear relationships, and it may not identify important nonlinear relationships between the features and target variable. Therefore, it is always important to carefully evaluate and test different feature selection techniques before selecting the final set of features for the model.

1. ***CHI-Squared Test:***

The chi-square test is a statistical method used to determine the relationship between two categorical variables. It is commonly used in feature selection to identify the most important features in a dataset by measuring the association between each feature and the target variable.

The chi-square test works by calculating the difference between the expected and observed frequencies of each category in the two variables, and then summing these differences across all categories. The resulting chi-square statistic can be compared to a critical value from a chi-square distribution with a certain number of degrees of freedom to determine whether the relationship between the two variables is significant.

Here is an example of how to perform a chi-square test for feature selection in Python using the chi2 function from the scipy.stats library:



***Wrapper Algorithm:***

The wrapper algorithm is a feature selection technique that selects subsets of features by using a predictive model to evaluate the performance of each subset. It involves creating many models, each with a different subset of features, and then selecting the subset that gives the best performance.

Here's an example code implementation of the wrapper algorithm for feature selection using the scikit-learn library in Python:

