LOADING AND PREPROCESSING THE DATASET FOR FUTURE SALES PREDICTION

**Loading and Preprocessing the dataset :-**

Importing the necessary libraries and the dataset, the data has been loaded and preprocessed by finding the missing data and removing the null columns.

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

data = pd.read\_csv('Sales.csv')

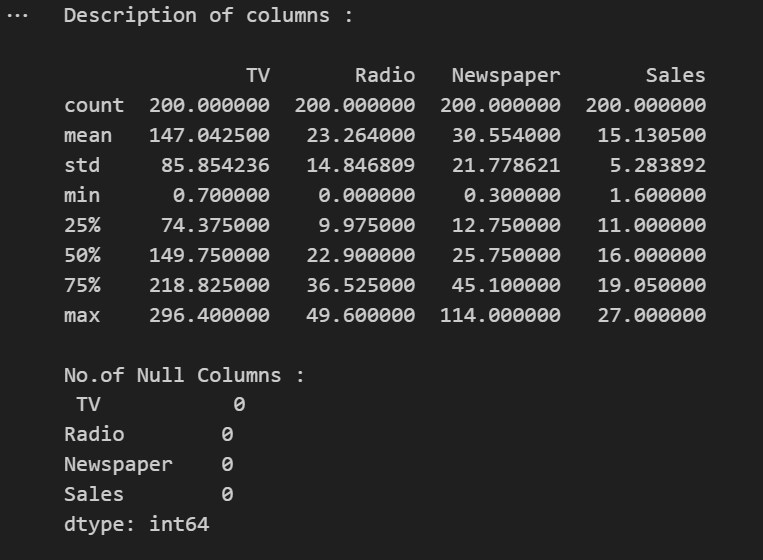
print("Description of columns : \n")

print(data.describe()) #decription of each column

print("\nNo.of Null Columns :\n",data.isnull().sum()) #count of null values in columns

data = data.dropna() #to remove null data

The description of the data like count,maximum,minimum using the following attributes is shown.



The null columns have been identified and it has been removed.

Finding highly related attributes :-

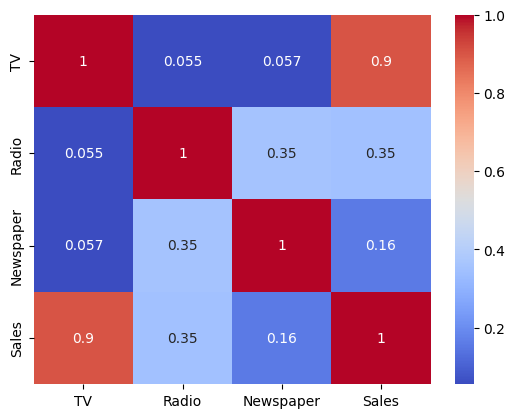
This part of the code shows the correlation between the dataset's features.

correlation=data.corr(method='pearson')

sns.heatmap(correlation, cmap="coolwarm",annot=True)

plt.show() #To show correlation between attributes

The visualized output is shown.



Positive correlation indicate that two attributes are highly correlated . For this dataset TV and Sales are highly related .