Applied Machine Learning Workshop (CSE 3193)

ASSIGNMENT-6: EXPLORATORY DATA ANALYSIS AND DATA AUGMENTATION

1. Write a Python code to rotate an image to generate 10 images. The angle used for rotation will be any random angle taken between 0 to 80 degrees.

2. Write a python code to flip an image from left to right and from top to bottom.

3. Write a Python code to crop an image using coordinates as left, right, top, and bottom. The bottom and right signify the height and width of the image. So the dimension of the cropped image should be 0, 0,200,250.

4. Write a python code to change the character of a sentence that show the new sentence consisting of optical illusions.

5. Write a Python code to replace words with synonyms of the given text.
   Text= " The quick brown fox jumps over the lazy dog"

6. Write a Python code to find the important words from a sentence based on the tfidf score and also print the tfidf score.

7. Write a Python code to upload the given dataset roller coaster from the given link https://www.kaggle.com/datasets/robikscube/rollercoaster-database and perform the following operations:
   (A) Load the dataset and print the 500 rows. List all the columns of the dataset.
   (B) Display the numeric and non-numeric features of the dataset.
   (C) Change the opening_date_clean, time, datatype date and time.
   (D) Find out the cardinality of the categorical features.
   (E) Create a data frame by selecting the following columns 'coaster_name', 'Location', 'Status', 'Manufacturer', 'year_introduced', 'latitude', 'longitude', 'Type_Main', 'opening_date_clean', 'speed_mph', 'height_ft', 'Inversions_clean', 'Gforce_clean'
   (F) Drop the column opening date.

8. Write Python code to find outliers in the given dataset. Also, use a box plot to show the outliers.
   Dataset=2, 19, 22, 27, 29, 30, 32, 35, 52, 59.

9. Write a Python code to find the score generated randomly around mean 60 and standard deviation 12. The number of samples to be generated is 200. Also, show the distribution of scores using Displot.

10. Write Python code to find the score for each data point. Also, find and remove the outliers using the z-score and IQR methods. Also, it shows the distribution of scores after removing outliers.

11. Write Python code to load the diabetes dataset from sklearn.dataset and perform the univariate analysis of it. Find the outliers in the feature bmi using IQR method and remove the outliers.