Data Collection and Preprocessing Phase

Date	28 june 2025
Student Name	Prathmesh Arvind Kumbhar
Project Title	Restaurant Recommendation System
Maximum Marks	6 Marks

Section	**Description**
Data Overview	The dataset contains restaurant information from Zomato, including name, reviews, ratings, cuisines, cost, and more. The data is cleaned, deduplicated, and preprocessed for building a content-based recommendation system.
Resizing	*Not applicable for text data.*
Normalization	Ratings are normalized to a 1-5 scale using MinMaxScaler. Text is lowercased and punctuation is removed.

Data Augmentation	Not applicable for text data.
Denoising	Text is cleaned by removing newline characters and punctuation.

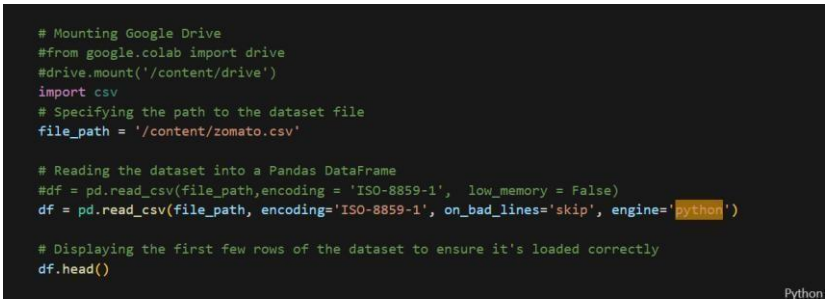
Edge Detection	Not applicable for text data.

Data Preprocessing

The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network training, ensuring robust and efficient performance across various computer vision tasks.

Color Space Conversion	Not applicable for text data.
Image Cropping	Not applicable for text data.
Batch Normalization	Not applicable for text data.

Data Preprocessing Code Screenshots

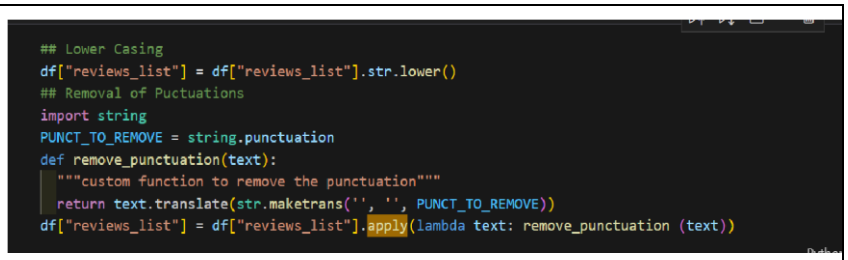
Loading Data	
Resizing	*Not applicable*

```python
# Mounting Google Drive
#from google.colab import drive
#drive.mount('/content/drive')
import csv
# Specifying the path to the dataset file
file_path = '/content/zomato.csv'

# Reading the dataset into a Pandas DataFrame
#df = pd.read_csv(file_path,encoding = 'ISO-8859-1',  low_memory = False)
df = pd.read_csv(file_path, encoding='ISO-8859-1', on_bad_lines='skip', engine='python')

# Displaying the first few rows of the dataset to ensure it's loaded correctly
df.head()
```

Normalization	
Data Augmentation	*Not applicable*

Denoising	
Edge Detection	*Not applicable*

Color Space Conversion	*Not applicable*
Image Cropping	*Not applicable*
Batch Normalization	*Not applicable*