**Numerical dataset**

**KNN & Linear Regression:**

* ***General Information on dataset:***

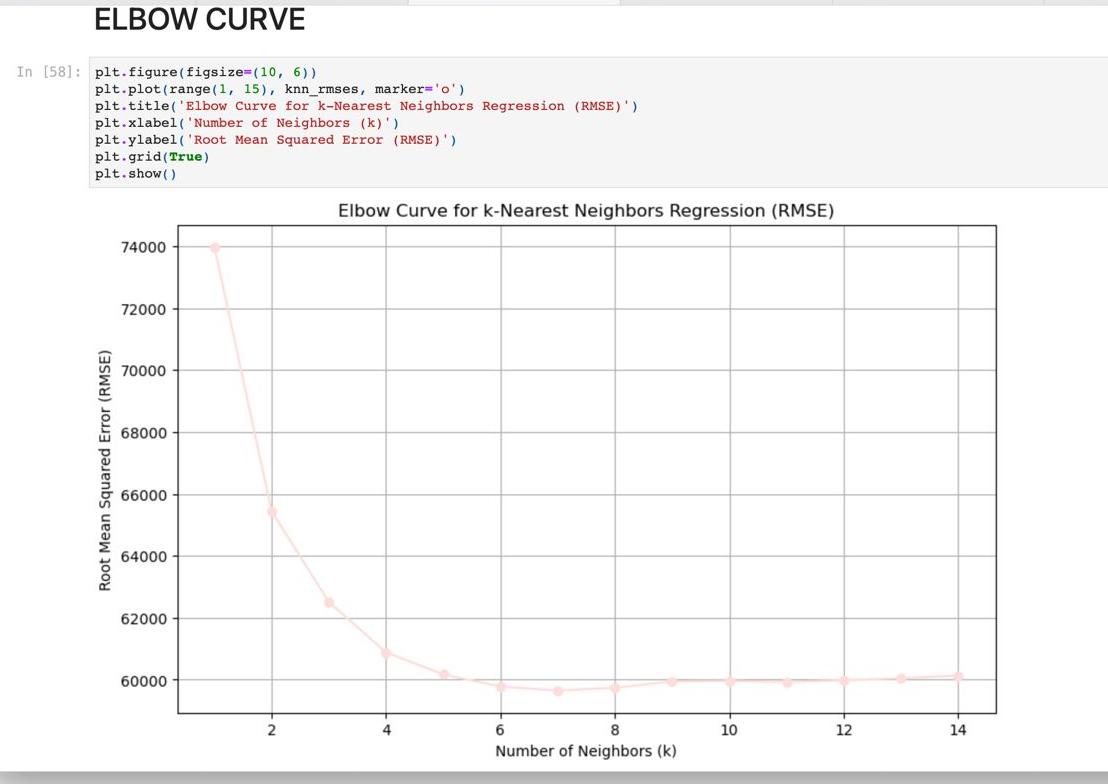
# Dataset Name: California Housing Data (1990)

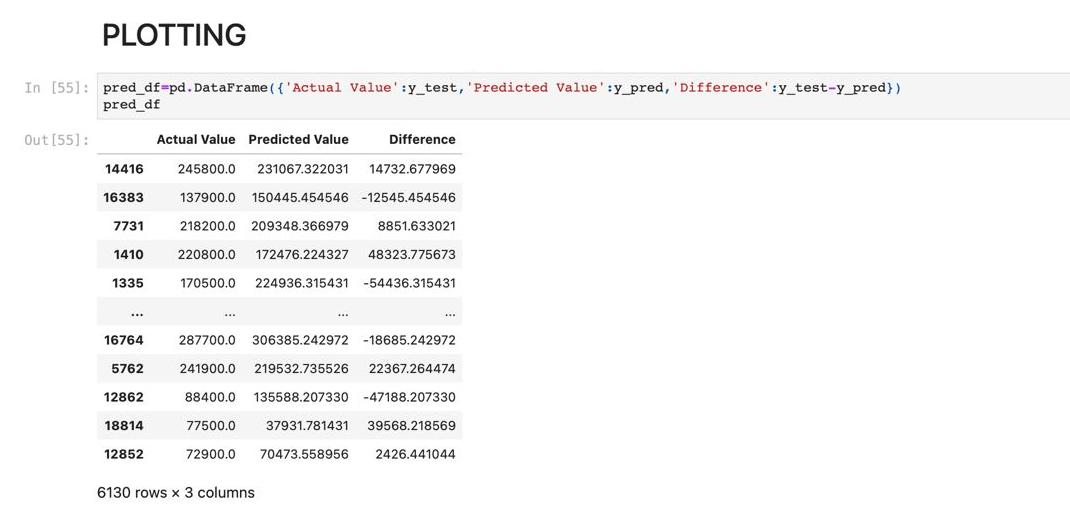
* ***Implementation details:***
* Our feature extraction is 16 features :

[longitude, latitude, housing\_median\_age, total\_rooms, total\_bedrooms, population, households, median\_income, median\_house\_value, ocean\_proximity, bedroom\_ratio, household\_rooms]

* Num of fold and ratio of training and testing :

(7 , 0.7 , 0.3)

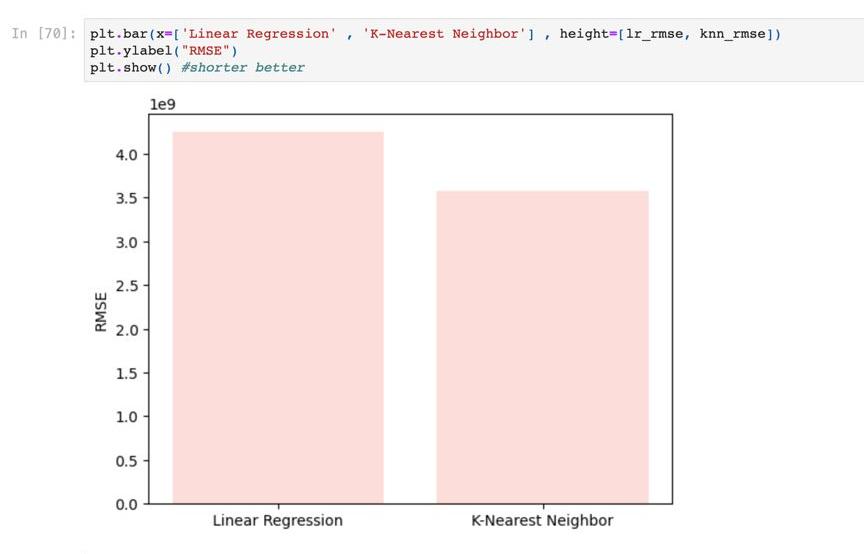
* Hyperparameters:
* [***regularization***  (lasso , ridge](https://www.google.com/search?client=avast-a-2&sca_esv=592805013&sxsrf=AM9HkKlHS-kI7fOxfYAu-g9J2-gO3tf4og:1703159001220&q=regularization+lasso+,+ridge&spell=1&sa=X&ved=2ahUKEwj20Y3auaCDAxUkhf0HHeGhCKwQkeECKAB6BAgHEAI))
* [***optimization*** (grid search , random](https://www.google.com/search?client=avast-a-2&sca_esv=592805013&sxsrf=AM9HkKkZTQizILP_IxMqANHyzCOYwZkLVA:1703159104594&q=optimization+grid+search+random&spell=1&sa=X&ved=2ahUKEwiv_bKLuqCDAxVYnf0HHVKWALoQkeECKAB6BAgJEAI) search)
* ***Results details:***
* KNN:
* ****- **Linear regression** :



- R2 between knn and linear regression:



-RMSE between knn and linear regression:



**Image dataset**

1. **Kmeans:**

* ***General Information on dataset:***

# Dataset Name: Food\_101

* + Number of classes: two classes
  + Labels :
* Cheesecake
* Pizza
  + total number of samples: 2000
  + Size of samples: 512×384
  + the number of samples used in training:
  + the number of samples used in validation :
  + the number of samples used in and testing:
* ***Implementation details:***
* Our feature extraction is Four features :
* Loading and Preprocessing Image:

converts them to RGB format using OpenCV, and resizes them to a fixed size (100x100 pixels in this case)

* Flattening Image Data:

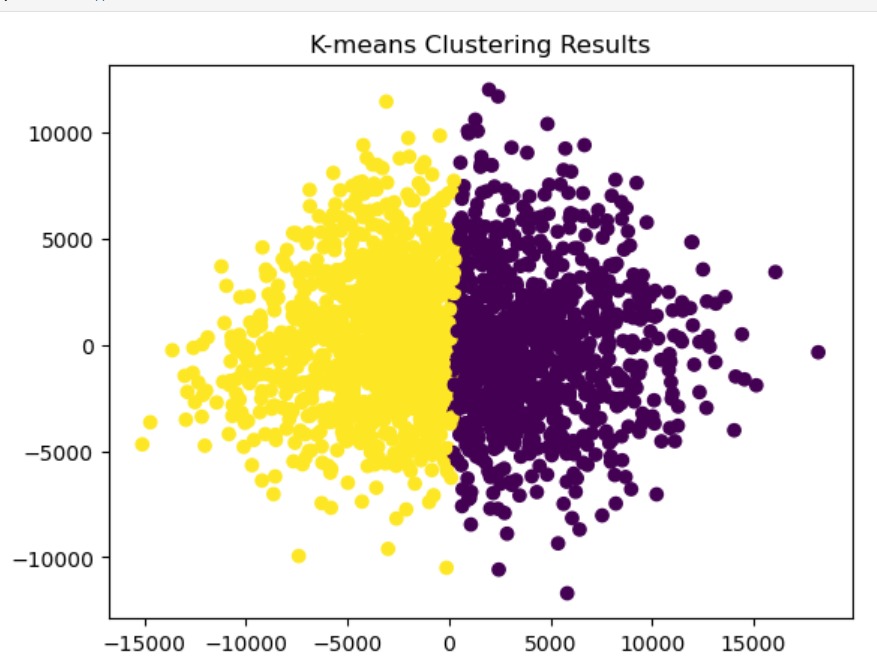
is reshaping the image data into a flat array. Each image, originally a 100x100 RGB image, is flattened into a 1D array of size 100 \* 100 \* 3

* Normalization (MinMax Scaling) :

The MinMaxScaler is applied to scale the flattened data. Scaling is a common preprocessing step, and it ensures that each feature (pixel value in this case) has a similar scale.

* Reshaping Scaled Data:

The scaled data is reshaped back to the original image shape (scaled\_image\_data) after scaling.

* There is no explicit use of cross-validation
* The hyperparameter in the K-means clustering model is the number of clusters (n\_clusters)
* ***Results details:***
* ****

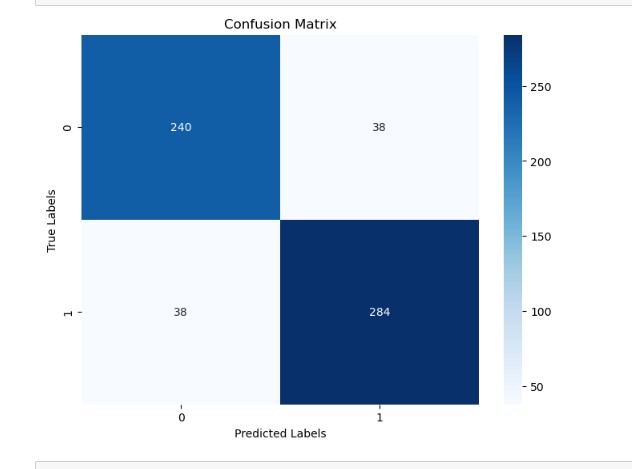
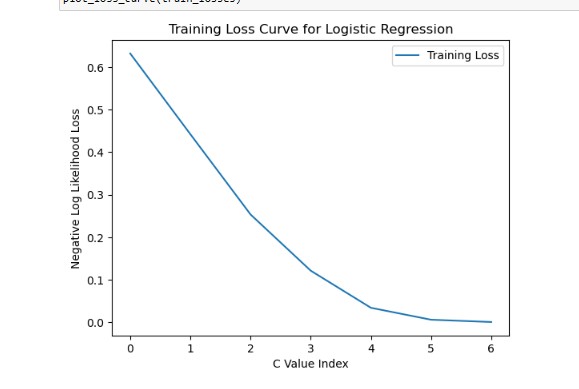
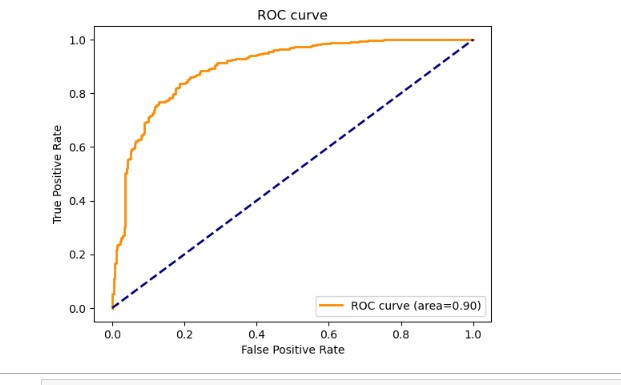
1. **Logistic regression:**

* ***General Information on dataset:***

# Dataset Name: Food\_101

* + Number of classes: two classes
  + Labels :
* Cheesecake
* Pizza
  + total number of samples: 2000
  + Size of samples: 512×384
  + the number of samples used in training: 700
  + the number of samples used in validation : We didn't not explicitly split the dataset into validation .
  + the number of samples used in and testing: 300
* ***Implementation details:***
* Our feature extraction is Four features :
* Function to extract HOG features :feature extraction using HOG 2916 features were extracted with dimensions (2916,).
* Loading and Preprocessing image data:
* resizing data to (64, 64, 3)
* Converts class labels to numeric values using LabelEncoder
* Splits the dataset into training and testing sets
* Logistic Regression Model Training:
* Applies standardization to the feature vectors using StandardScaler.
* Performs a grid search with cross-validation to find the best logistic regression model based on accuracy.
* Tracks and plots the training loss curve for different values of the regularization parameter C.
* Cross validation is used with Grid search and the number of folds is 4.
* Model Evaluation:
* Computes the confusion matrix for the test set predictions.
* Plots the confusion matrix as a heatmap.
* Calculates the accuracy score for model .

* Hyperparameters:
* the regularization parameter C
* The max\_iter parameter determines the maximum number of iterations taken for the solver to converge
* The random\_state parameter is used to seed the random number generator for reproducibility.
* ***Results details:***

****