

Welcome to Peer Review Assignments

DPhi Mini Bootcamp: Unsupervised Learning

A **Peer Review Assignment** is where you and your fellow learners have an opportunity to review and grade each other's work. You can provide both qualitative and quantitative feedback based on a rubric that we have created (available at the end of this document). Peer reviews also provide pedagogical benefits and can help learners develop evaluative judgement and self-regulation.

Expectations: Reviewers

While grading, be respectful, encouraging, and honest. Acknowledge what your fellow learner did well and offer specific suggestions on how they can improve. Scores should reflect the learner's grasp of the concepts and should not be deducted points for difficulties with language or differences of opinion.

Note: Reviewing peers assignment carries 5 marks. Your reviewed assignments' feedback and scores will be reviewed by the DPhi team. Any anomalies in the scores will carry a penalty. We request all the learners to review notebooks responsibly.

Scoring Rubric

You can look at the solution notebook here:

<https://colab.research.google.com/drive/1h2d5kfpTtgzcZsvwSLJpzRFe2P9IF3R1?usp=sharing>

Task 1: Data loading and Data Analysis

1. Load the data file into a dataframe object : df
 - The learner has correctly loaded the data into the dataframe - 1 mark
2. Display the first 5 observation of the dataset
 - The learner has correctly loaded the data into the dataframe - 1 mark
3. Display a concise summary of the provided data and list out 2 observations/inferences that you observe from the result. You can use the info() method for this.
 - The learner has correctly displayed the summary - 1 mark
 - The learner has listed the 2 observations/inferences - 1 mark
4. Perform EDA viz. Is there any missing values in each column of the provided dataset
 - The learner has checked if there are any missing values or not - 1 mark
5. Display all the unique status_type of the dataframe
 - The learner has correctly displayed all the unique status type - 1 mark
6. Convert categorical variable, i.e., status_type into numerical representation using a label encoder
 - The learner has correctly converted the status_type into numerical representation - 1 mark

Task 2: Feature Scaling

1. Perform normalization to scale numerical data prior to modeling. You may use Min-max scaling technique
 - The learner has scaled the data using any scaling technique - 1 mark
2. Use Elbow method to determine the number of cluster to be formed
 - The learner has used the elbow method to determine the number of clusters - 1 mark
3. Verify optimal number of clusters on the basis of silhouette scores
 - The learner has used silhouette_scores/plot to verify the no. of clusters - 1 mark
4. Assigning cluster label to each instance of the dataframe
 - The learner has assigned the cluster labels - 1 mark

5. Validating obtained clusters using external cluster validation approaches like `rand_score` and `adjusted_rand_score`
 - The learner has used `rand_score` - 1 mark
 - The learner has used `adjusted_rand_score` - 1 mark

Task 3: Perform PCA on Fruit Dataset

1. Scale the value. Each feature should be centered (zero mean) and with unit variance
 - The learner has scaled the value - 1 mark
2. Analysis - perform PCA with 4 components and print their explained variance_ratio
 - The learner has performed PCA with 4 components and printed the explained variance ratio - 1 mark