**Scoring guide (Rubric) - Unsupervised Learning Project Rubric**

| **Criteria** | **Points** |
| --- | --- |
| **Import the required libraries and load the data**   1. Load the required libraries and read the dataset. (1 point) 2. Check the first few samples, shape, info of the data and try to familiarize yourself with different features (2 point) | 3 |
| **Data cleansing and Exploratory data analysis**   1. Check if there are any duplicate records in the dataset? If any, drop them. (1) 2. Drop the columns which you think redundant for the analysis.(Hint: drop columns like ‘id’, ‘review’) (1) 3. Check the column 'weight', Is there any presence of string data? If yes, remove the string data and convert to float. (Hint: 'weight' has the suffix as lbs) (2) 4. Check the unique categories for the column 'rented for' and group 'party: cocktail' category with 'party'. (2) 5. The column 'height' is in feet with a quotation mark, Convert to inches with float datatype. (3) 6. Check for missing values in each column of the dataset? If it exists, impute them with appropriate methods. (3) 7. Check the statistical summary for the numerical and categorical columns and write your findings. (3) 8. Are there outliers present in the column age? If yes, treat them with the appropriate method. (3) 9. Check the distribution of the different categories in the column 'rented for' using appropriate plot. (2)   **Data Preparation for model building** | 20 |
| 12. Encode the categorical variables in the dataset. (1)  13. Standardize the data, so that the values are within a particular range. (1) | 2 |
| **Principal Component Analysis and Clustering**   1. Apply PCA on the above dataset and determine the number of PCA components to be used so that 90-95% of the variance in data is explained by the same. (7) 2. Apply K-means clustering and segment the data. (You may use original data or PCA transformed data) (8)    * + 1. Find the optimal K Value using elbow plot for K Means clustering.        2. Build a Kmeans clustering model using the obtained optimal K value from the elbow plot.        3. Compute silhouette score for evaluating the quality of the K Means clustering technique.   16. Apply Agglomerative clustering and segment the data. (You may use original data or PCA transformed data) (8)  Find the optimal K Value using dendrogram for Agglomerative clustering.  b. Build a Agglomerative clustering model using the obtained optimal K value observed from dendrogram.  Compute silhouette score for evaluating the quality of the Agglomerative clustering technique. (Hint: Take a sample of the dataset for agglomerative clustering to reduce the computational time) | 23 |
| **Conclusion**  17. Perform cluster analysis by doing bivariate analysis between cluster labels and different features and write your conclusion on the results. (2) | 2 |
| Points | 50 |