HW 3
(submit R code to eLearning by April 11, 11:59pm)

Marketing to Frequent Fliers

Download the data **AirlineLoyalty.csv** from eLearning. The data contain information on passengers who belong to an airline's frequent flyer program. The data include information on customers' mileage history and the different ways they accrued or spent miles in the last year. The primary objective is to cluster customers into different segments, analyze each group's characteristics, and design targeted marketing campaigns (i.e., different types of mileage offers). Detailed descriptions of the dataset can be found in the table below.

Feature	Description
ID	Unique ID for each passenger
Balance	Number of miles eligible for award travel
Qual_miles	Number of miles counted as qualifying for top flight status
cc1_miles	Number of miles earned with Frequent Flyer credit card in the past 12 months
cc2_miles	Number of miles earned with Rewards credit card in the past 12 months
cc3_miles	Number of miles earned with Small Business credit card in the past 12 months
Bonus_miles	Number of miles earned from non-flight bonus transactions in the past 12 months
Bonus_trans	Number of non-flight bonus transactions in the past 12 months
Flight_miles_12mo	Number of flight miles in the past 12 months
Flight_trans_12	Number of flight transactions in the past 12 months
Days_since_enroll	Number of days since enrollment date
Award	Dummy variable for the last award (1=not null, 0=null)

To guide you through the process, please proceed with the following steps to perform the clustering analysis.

1. (1 pt) Load the data and normalize all variables except ID.
2. (2 pt) Apply hierarchical clustering with Euclidean distance and Ward's method. Cut the dendrogram into two clusters. Then add a column as the cluster label to the data.
3. (2 pt) Using silhouette score to find the best k for $k \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Create the silhouette plot, with x-axis being the number of clusters k, y-axis being the average silhouette score .
4. (3 pt) Compare the resulting clusters from step 2 and the resulting clusters when applying k-means with k=2. What are the characteristics of customers in each group? Assume you work in the marketing department, what types of offers would you target to customers in each cluster?

For question 4, please briefly describe your findings in the form of comments in the R script.