	Complete each task and question by filling in the blanks ( ) with one or more lines of code or text. Tasks 1-11 and questions 1-5 are worth <b>0.5 points</b> each and questions 6-7 are worth <b>1 point</b> each (out of <b>10 points</b> ).  Submission:
	This assignment is due Monday, October 26, at 11:59PM (Central Time).  This assignment must be submitted on Gradescope as a PDF file containing the completed code for each task and the corresponding output. To save your Jupyter notebook as a PDF file, go to File > Export Notebook As > HTML or File > Download As > HTML, open the HTML file and print it as a PDF file. Additionally, this assignment has a single question on Gradescope and all pages of PDF file must be assigned to this question. A 0.5-point (5%) penalty will be applied to submissions that do not follow these guidelines. It more instructions on how to submit assignments on Gradescope, see this guide.
	Late submissions will be accepted within <b>0-12</b> hours after the deadline with a <b>0.5-point (5%) penalty</b> and within <b>12-24</b> hours after the deadline with a <b>2-point (20%) penalty</b> . No late submissions will be accepted more than 24 hours after the deadline. <b>This assignment is individual</b> . Offering or receiving any kind of unauthorized or unacknowledged assistance is a violation of the University's academic integrity policies, will result in a grade of zero for the assignment, and will be subject to disciplinary action.
[39]:	# Load libraries import pandas as pd import matplotlib.pyplot as plt import seaborn as sns from sklearn.preprocessing import StandardScaler from scipy.cluster.hierarchy import linkage, fcluster from sklearn.cluster import KMeans, DBSCAN
[40]: ====================================	# Load dataset and display the first five rows data = pd.read_csv('data_lab_05.csv') data.head()  Account International Voice Number Total
	No.   Plan   P
[41]:	Task 01 (of 11): Create a new numerical attribute named 'Total charge' that contains the sum of the attributes 'Total day charge', 'Total eve charge', and 'Total night charge'.  data['Total charge'] = data['Total day charge'] + data['Total eve charge'] + data['Total night charge']
[42]:	<pre># Partition the dataset into attributes and true clusters (churned/non-churned) # Consider only the following attributes: 'International plan', 'Total charge', and 'Customer service alls' X = data[['International plan', 'Total charge', 'Customer service calls']] Y = data['Churn']</pre> Task 02 (of 11): Standardize the attributes.
[43]:	<pre>scaler = StandardScaler() scaler.fit(X) X_scaled = scaler.transform(X)</pre> Part 2: Hierarchical Clustering
[44]:	Task 03 (of 11): Cluster the dataset using hierarchical clustering with single linkage method. Use single linkage as the method a Euclidean distance as the distance metric.  clustering = linkage(X_scaled, method='single', metric='euclidean') clusters = fcluster(clustering, 2, criterion = 'maxclust')
[45]:	<pre>Task 04 (of 11): Plot contingency matrix and compute evaluation metrics for hierarchical clustering with single linkage method.</pre> cont_matrix = metrics.cluster.contingency_matrix(data['Churn'], clusters) sns.heatmap(cont_matrix, annot = True, fmt = ".3f", square = True, cmap = plt.cm.Blues) plt.ylabel('Actual') plt.xlabel('Predicted') plt.title('Contingency matrix') plt.tight layout()
	Contingency matrix - 2500 - 2664.000 186.000 - 2000
	- 1500 - 1000 - 500 - 500
[46]:	<pre>adjusted_rand_index = metrics.adjusted_rand_score(data['Churn'], clusters) silhouette_coefficient = metrics.silhouette_score(X_scaled, clusters, metric='euclidean') print([adjusted_rand_index, silhouette_coefficient])  [0.201747491556069, 0.5588810706001527]</pre>
[ 1 / ] •	<pre># Plot clusters found using hierarchical clustering with single linkage method data['clusters'] = clusters ax = data.plot(kind = 'scatter', x = 'Total charge', y = 'International plan', c = 'clusters', colors = plt.cm.brg) ax = data.plot(kind = 'scatter', x = 'Total charge', y = 'Customer service calls', c = 'clusters', colors rmap = plt.cm.brg)</pre> 10 - Customer service calls', c = 'clusters', colors 10 - Customer service calls', c = 'clusters', colors 10 - Customer service calls', c = 'clusters', colors 11 - Customer service calls', c = 'clusters', colors 12 - Customer service calls', c = 'clusters', colors 12 - Customer service calls', c = 'clusters', colors 13 - Customer service calls', c = 'clusters', colors 14 - Customer service calls', c = 'clusters', colors 15 - Customer service calls', c = 'clusters', colors 16 - Customer service calls', c = 'clusters', colors 17 - Customer service calls', c = 'clusters', colors 18 - Customer service calls', c = 'clusters', colors 19 - Customer service calls', c = 'clusters', colors 10 - Customer service calls', c = 'clusters', colors 10 - Customer service calls', c = 'clusters', colors 10 - Customer service calls', c = 'clusters', colors 10 - Customer service calls', c = 'clusters', c = 'clus
	0.8 - 18 - 16 - 16 - 14 planting of the state of the stat
	0.2
	-16 stayson 2 -12
	Question 01 (of 07): What can you conclude about the clusters found using hierarchical clustering with single linkage method from the plot and the evaluation metrics?  Answer: There is a obviously separation that split international plan into two clusters. But for the Customer service calls, it seems two clusters mixed and no division of these clusters
	The adjusted rand index tells us that single linkage does not clustering the observation well, and sihuotte coefficient tells us there is about the average quality in terms of cohesion and seperation  Task 05 (of 11): Cluster the dataset using hierarchical clustering with complete linkage method. Use complete linkage as the method and Euclidean distance as the distance metric.
[48]: [49]:	clustering = linkage(X_scaled, method='complete', metric='euclidean') clusters = fcluster(clustering, 2, criterion = 'maxclust')  Task 06 (of 11): Plot contingency matrix and compute evaluation metrics for hierarchical clustering with complete linkage method  cont_matrix = metrics.cluster.contingency_matrix(data['Churn'], clusters) sns.heatmap(cont_matrix, annot = True, fmt = ".3f", square = True, cmap = plt.cm.Blues)
	<pre>plt.ylabel('Actual') plt.xlabel('Predicted') plt.title('Contingency matrix') plt.tight_layout()  Contingency matrix - 2500</pre>
	- 11.000 2839.000 - 2000 - 1500 - 100
[50]:	<pre>silhouette_coefficient = metrics.silhouette_score(X_scaled, clusters, metric='euclidean')</pre>
[51]:	<pre>print([adjusted_rand_index, silhouette_coefficient])  [0.017507021065270037, 0.5130278392459514]  # Plot clusters found using hierarchical clustering with complete linkage method data['clusters'] = clusters ax = data.plot(kind = 'scatter', x = 'Total charge', y = 'International plan', c = 'clusters', colors = plt.cm.brg) ax = data.plot(kind = 'scatter', x = 'Total charge', y = 'Customer service calls', c = 'clusters', colors = colors =</pre>
	rmap = plt.cm.brg)  10 - 20  0.8 - 18
	0.6 - 1.6 state   0.7 - 1.4   0.9 - 1.2   0.0 - 1.2   0.0 - 1.2   1.0
	8 - 18 - 18 - 16 s saying a service calls a service call a service c
	Question 02 (of 07): What can you conclude about the clusters found using hierarchical clustering with complete linkage methological clustering with clustering
	from the plot and the evaluation metrics?  Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation  The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and sihuotte coefficient tells us there is about the average quality in terms of cohesion and seperation
[52]:	Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation  The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and sihuotte coefficier tells us there is about the average quality in terms of cohesion and separation  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter random_state to 0.
[52]: [53]:	Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation  The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and sihuotte coefficier tells us there is about the average quality in terms of cohesion and seperation  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter random_state to 0.  clustering = KMeans(n_clusters=2, init='random', n_init=10, random_state=0).fit(X_scaled)  clusters = clustering.labels_  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  cont_matrix = metrics.cluster.contingency_matrix(data['Churn'], clusters) sns.heatmap(cont_matrix, annot = True, fmt = ".3f", square = True, cmap = plt.cm.Blues) plt.ylabel('Actual')
	Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation  The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and sihuotte coefficient tells us there is about the average quality in terms of cohesion and separation  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter random_state to 0.  clustering = KMeans (n_clusters=2, init='random', n_init=10, random_state=0).fit(X_scaled) clusters = clustering.labels_  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  cont_matrix = metrics.cluster.contingency_matrix(data['Churn'], clusters) sns.heatmap(cont_matrix, annot = True, fmt = ".3f", square = True, cmap = plt.cm.Blues) plt.ylabel('Actual') plt.xlabel('Perdicted') plt.title('Contingency matrix') plt.tight_layout()  Contingency matrix
	Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation  The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and sihuotte coefficient tells us there is about the average quality in terms of cohesion and seperation  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter random_state to 0.  clustering = KMeans (n_clusters=2, init='random', n_init=10, random_state=0).fit (X_scaled) clusters = clustering.labels_  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  cont_matrix = metrics.cluster.contingency_matrix(data['Churn'], clusters) sns.heatmap(cont_matrix, annot = True, fmt = ".3f", square = True, cmap = plt.cm.Blues) plt.ylabel('Actual') plt.xlabel('Predicted') plt.xlabel('Predicted') plt.tight_layout()  Contingency matrix  -zsoo
	Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being spilt well, because we can see there is a obviously separation  The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and sihuotte coefficient tells us there is about the average quality in terms of cohesion and separation  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter random_state to 0.  clustering = KMeans (_clusters=2, init='random', n_init=10, random_state=3).fit (X_scaled) clusters = clustering.libels_  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  cont_natrix = metrics.cluster.contingency_matrix(data['Churn'], clusters)
[53]:	Answer: There is only one cluster for international plan, so the complete inkage does not work well for international pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation. The adjusted rand index is very low, it tells us that complete linkage does not disserring the observation NOT well, and sihuotte coefficient tells us there is about the average quality in terms of cohesion and separation.  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter random state to 0.  clustering = thread (not clusters = 2, init="random", n_init=10, random_state=0).flt(X_scaled) clusters = observing.labols_  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  cont_matrix = metrics.cluster.contingency_matrix (data("churn"), clusters) ans.beatrag(cont_matrix, annot = True, fint = ".3f", square = True, cmap = plt.cm.slues) plt.ylabel("Acctual") plt.ylabel("Acctual") plt.ylabel("Acctual") plt.ylabel("Fredicted") plt.ylabel("Fredicted") plt.title("Contingency matrix)  contingency matrix  contingency matrix  contingency matrix  contingency matrix  adjusted_rand_index = metrics.adjusted_rand_score(data("churn"), clusters) allhousette_coefficient)  contingency matrix  contingency matrix  contingency matrix  print(ladjusted_rand_index, siteoutte_coefficient))  (0.201747491586069, 0.5588016706001527)  #### Plot coursers found using ## Reans clustering data("clusters") = clusters", well-coefficient)  (0.201747491586069, 0.5588016706001527)  ###################################
[53]:	Answer: There is only one cluster for international plan, so the complete linkage does not work well for International pan. But for the Customer service calls, it seems two clusters are being split well, because we can see there is a obviously separation. The adjusted rand index is very low, it tells us that complete linkage does not clustering the observation NOT well, and situotic coefficient tells us there is about the average quality in terms of cohesion and separation.  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter candom extacts to 0.  Clustering = %Pouncin clustorum=2, init="random", u_init=10, random_ptata=01.fit (X_sealod) clusters = clustering.lacels  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  cont_matrix = matrics.cluster.contingency_matrix(data("tburn"), clusters) when the computer of the contingency matrix (acta("tburn"), clusters) when the computer of the contingency matrix (acta("tburn"), clusters) when the contingency matrix ("burn") when the
[53]:	Answer There is only one cluster for international plan, so the complete inhage does not work well for international pan. But for the Customers enter cash, it seems to clusters are being spill well, because we can see there is a obviously separation.  The adjusted rand index is very low, it tells us that complete inhage does not clustering the observation NOT well, and shouldo coefficients to there is about the average quality in terms of cohesion and separation.  Part 3: KMeans Clustering  Task 07 (of 11): Cluster the dataset using KMeans clustering. Use random initialization of centroids, 10 iterations, and set parameter yeardom, state to 0.  clustering - "Owegate in clustering." Initiation of centroids, 10 iterations, and set parameter and understance, clusters, and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Contingency matrix and compute evaluation metrics for K-Means clustering.  Contingency matrix and compute evaluation metrics for K-Means clustering.  Contingency matrix   "Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Contingency matrix   "Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.")   plot (a contingency matrix   "Task 08 (of 11): Plot (a contingency matrix   "Task 08 (of
[53]:	Answer There is only one dustar for international plan, so the complete linkage does not work well for international pan. But for the Coustomer service calls, it seems two dusters are being spill well, because we can see there is a obviously separation. The adjusted most index is very low, it tells us that complete linkage does not clustering the observation NOT well, and situation coefficients there is about the average quality in forms of cohesion and separation.  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of controlds, 10 iterations, and set parameter random_state to 0.  **Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of controlds, 10 iterations, and set parameter random_state to 0.  **Task 07 (of 11): Plot contingency matrix and compute availation metrics for K-Means clustering.  **Count matrix = metricos.clusters.contingency matrix flates ("Chish"), related ("Chish")
[53]:	Answer: There is only one cluster for international plan, so the complote inhage does not work well for international plan. But for the Caustomer service calls, it seems two clusters are being soft two because we can see there is a dovolousy separation.  The adjusted mind make is very low. It falls us that complete inhage does not clustering the observation NOT well, and shoulds coefficient in should be average quality in terms of cohecian and appearation.  Part 3: K-Means Clustering  Task 07 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter candom_state to 0.  Slossborung = observation clusterers*, initia*/random*, n. initia*10, random state**(1): fit(1X state*0) clusterers = observation and act parameter candom_state to 0.  Slossborung = observation clusterers*, initia*/random*, n. initia*10, random state**(1): fit(1X state*0) clusterers = observation and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix and compute evaluation metrics for K-Means clustering.  Task 08 (of 11): Plot contingency matrix.  Task 08 (of 11): Plot contingency matrix.  Task 08 (of 12): Plot contingency matrix.  Task 08 (of 13): Plot contingency matrix.  Task 08 (of 13): Plot contingency matrix.  Task 08 (of 13): Plot contingency matrix.  Task 08 (of 14): Plot contingency matrix.  Task 08 (of 15): Plot contingency matrix.  Ta
[53]:	Answer: There is not you no cluster for international plan, so the complete linkage does not work well for international pan. But for the Calciforms service calls, if severe two clades are being spill well, because we can see there is a detinosal yeapardion.  The adjusted and index is very low. It tells us that complete linkage does not dusticing the observation NOT well, and silvation coefficients that is a detailed and index is very low. It tells us that complete linkage does not dusticing the observation NOT well, and silvation coefficients that is a detailed to the average quality in terms of contesson and seperation.  Part 3: K-Means Clustering  Task 67 (of 11): Cluster the dataset using K-Means clustering. Use random initialization of centroids, 10 iterations, and set parameter remoder, active to 0.0.    Clustering
[53]:	Accessed Two is not and the evaluation metebra?  Accessed Two is any one education for immulsional claims so the computed inhage does not use to all do immunistrations to the the Continuous and accessed and in a control control is a control with a control in a control control is a control with a control control in a control control is a control with a control control in a control control control in a control contr
[54]: [55]:	Names. There is not because the recording policy to the complete indeps does not work well as literational para. On for the Coscorer service and it, iterates to claster as in terms gall on the complete indeps does not work well as literational para para to the complete indeps does not be the substance and the complete indeps does not be the sub
[54]: [55]:	Assert Time to give a claim for international plan, as the conglate integral diserted value for international part. But for the Customer service cities, experts to substitute the conglate integral control costs which international part. But for the Customer service cities, experts to substitute the congress of the co
[54]: [55]:	Assert There is allow on state for international plan, as the complete integer does not sent and for international part. But for its Castoma service cells, ill seems to obtained are being all well because and on sever the relational programs and the corresponding of the state o
[54]: [55]:	Assert Prot in a Use and the contraction materials of any, will be applied that person of another the following and the first of the Contract vice can't became the discrete and any and the Contract vice can't became the discrete and any and the Contract vice can't became the discrete and any and the Contract vice can't became the discrete and any and the Contract vice can't became and any and the Contract vice can't be a several part of the contract the contract vice can't be a several part of the contract vice
[54]: [56]:	The Market Plane of the contribution control of the complex integer can an early adding management of the Charles and an extractional control of the complex integer can be an early and control of the plane of the complex of the com
[54]: [57]:	Advanced Tooks but the control for more and an analysis of the control for the control for the control of the control for the control of the control for the c
[54]: [57]:	More Than the option of the evaluation motions:  Account Than the process could be considered above the complete of the policy of the control
[54]: [57]:	Account To the Control of the Principles of the Control of the
[54]: [57]:	Contract the contraction in the inhabitation and the contraction of cathod at the receive of cathod at the Contract of cathod at the cathod at
[54]: [57]:	Security of the control of the contr
[54]: [57]:	Received in the reconstruction contents of the content of the cont
[54]: [57]:	Common form to the parties in the content of the co
[54]: [57]: [58]:	Constraint of the content of the con
[54]: [57]: [57]:	Account to the characteristic interface of the characteristic and the characteristic in
[54]: [57]: [57]:	More than set in a discontinuous control of the con
[54]: [57]: [57]:	Record Control