# Αναλυτική καταγραφή του περιβάλλοντος υλοποίησης (βιβλιοθήκες λογισμικού κτλ.) καθώς και τα βήματα που απαιτούνται για την εγκατάστασή του.

import pandas as pd

import os

import matplotlib.pyplot as plt

import seaborn as sns

from yellowbrick.cluster import KElbowVisualizer

from sklearn.preprocessing import StandardScaler

import pandas as pd

import os

import matplotlib.pyplot as plt

import seaborn as sns

   from sklearn.model\_selection import train\_test\_split

    from sklearn.metrics import accuracy\_score

    from sklearn.naive\_bayes import GaussianNB

    from sklearn.neural\_network import MLPClassifier

    from sklearn.ensemble import RandomForestClassifier

    from sklearn.cluster import KMeans

# o Σύντομη περιγραφή της διαδικασία υλοποίησης.

Ερώτημα 1

Ερώτημα 2

Ερώτημα 3

# o Σχολιασμό των τελικών αποτελεσμάτων

Ερώτημα 1

A screen shot of a black and white screen

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A graph of a bar graph

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a graph

Description automatically generated

1. Διαγώνιες τιμές: Οι διαγώνιες τιμές είναι 1 (τέλεια συσχέτιση), το οποίο είναι αναμενόμενο αφού κάθε στατιστικό στοιχείο συσχετίζεται τέλεια με τον εαυτό του.

2. Υψηλές συσχετίσεις: Ο μέσος όρος και το 25%, 50%, 75% συσχετίζονται σε μεγάλο βαθμό (συσχέτιση ≈ 0,99). Αυτό σημαίνει ότι οι κεντρικές τάσεις των δεδομένων (μέσος όρος και τεταρτημόρια) είναι στενά συνδεδεμένες, υποδεικνύοντας ότι η κατανομή των τιμών είναι αρκετά συνεπής. Η τυπική απόκλιση (std) και το 25% έχουν υψηλή συσχέτιση (~0,95), γεγονός που υποδηλώνει ότι η διασπορά των δεδομένων (που αντιπροσωπεύεται από την std) έχει ισχυρή επιρροή στο κατώτερο τεταρτημόριο (25%).

3. Η τυπική απόκλιση (std) και το 25% έχουν υψηλή συσχέτιση (~0,95). Μέτριες συσχετίσεις: Η συσχέτιση μεταξύ του μέσου όρου και του ελάχιστου είναι χαμηλότερη (~0,92). Αυτό υποδηλώνει ότι ενώ υπάρχει σχέση μεταξύ του μέσου όρου και των ελάχιστων τιμών, η επιρροή δεν είναι τόσο ισχυρή όσο είναι με τα άλλα εκατοστημόρια. Std και min παρουσιάζουν επίσης μέτρια συσχέτιση (~0,90), υποδηλώνοντας ότι η μεταβλητότητα επηρεάζει τις ελάχιστες τιμές σε κάποιο βαθμό, αλλά όχι τόσο πολύ όσο τα άλλα στατιστικά στοιχεία.

4. Η μεταβλητότητα επηρεάζει τις ελάχιστες τιμές σε κάποιο βαθμό, αλλά όχι τόσο πολύ όσο τα άλλα στατιστικά στοιχεία. Χαμηλές συσχετίσεις: Δεν υπάρχουν εξαιρετικά χαμηλές συσχετίσεις, καθώς οι περισσότερες τιμές παραμένουν αρκετά υψηλές, αλλά οι min και max έχουν ελαφρώς ασθενέστερη σχέση (~0,91) σε σύγκριση με άλλα ζεύγη, γεγονός που θα μπορούσε να υποδηλώνει μεγαλύτερο εύρος μεταβλητότητας στις ακραίες τιμές.

5. Ερμηνεία χρώματος: Σκούρα χρώματα (πιο κοντά στο 0,90): Πιο ανοιχτά χρώματα (πιο κοντά στο 1,00): Αυτά υποδεικνύουν ισχυρότερες συσχετίσεις, π.χ. μέσος όρος με 50%, 75% κ.λπ.

Συνολική εικόνα: Ο χάρτης θερμότητας δείχνει ότι ο μέσος όρος και τα τεταρτημόρια συσχετίζονται πολύ στενά, υποδεικνύοντας μια συνεπή κατανομή στα δεδομένα. Η τυπική απόκλιση έχει σημαντική επιρροή στα τεταρτημόρια και τις ελάχιστες τιμές, αλλά ελαφρώς μικρότερη επιρροή στις μέγιστες τιμές.

Αυτό μπορεί να είναι χρήσιμο για την κατανόηση της σχέσης μεταξύ διαφορετικών περιγραφικών στατιστικών και του τρόπου με τον οποίο αντικατοπτρίζουν την υποκείμενη κατανομή των δεδομένων.

Ερώτημα 2

A computer screen shot of a black background with white text

Description automatically generated

Ερώτημα 3