

Comparison of Annual Fresh Water Withdrawal to Population Growth Using Clustering Analysis



Name : SRAVANI GOKA
Student id : 22013825
Repository link: <https://github.com/Sravanigooka11/Ads-assignment3.git>

Abstract:

Cluster analysis is a technique used in data analysis to discover groups or clusters within a given data set. Unlike supervised learning methods, this technique does not require any predefined groups to be specified. The k-means cluster analysis algorithm is an example of this technique, which groups objects with similar characteristics together into clusters. The ultimate goal of cluster analysis is to create a set of distinct clusters where objects within each cluster are similar to one another, but different from those in other clusters. One popular way of grouping objects using this technique is scatter plot clustering, which creates clusters based on the similarity between objects within each group. A scatter plot can be used to visualize the clusters. The following report focuses on stimulating the comparison of annual fresh water withdrawals to population growth.

Introduction:

We will be using data on annual freshwater withdrawal and population growth from different countries across the world. We will analyze the data using the K-means clustering algorithm to identify clusters of countries that have similar freshwater withdrawal and population growth patterns. We will also explore the relationship between the identified clusters and other relevant variables

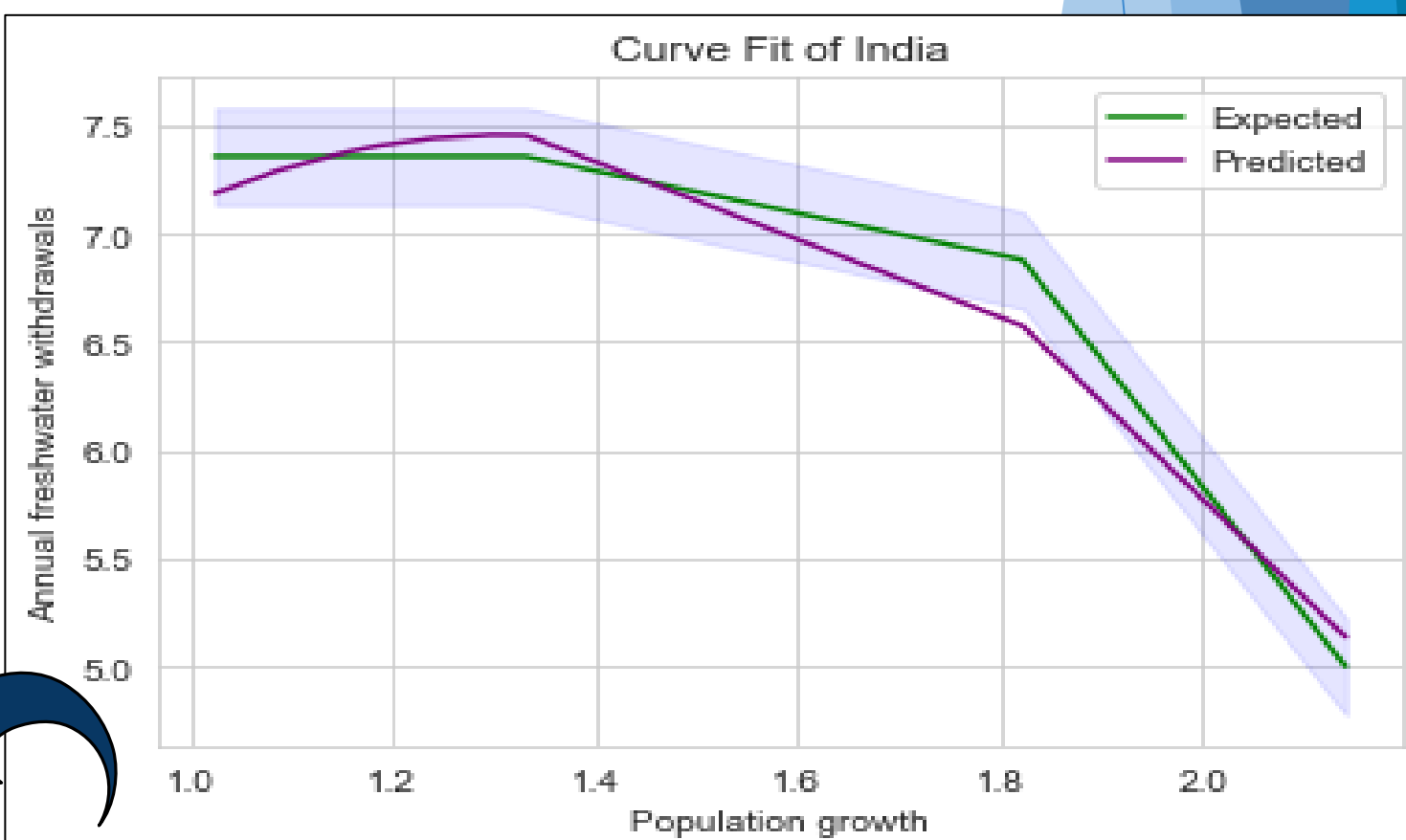
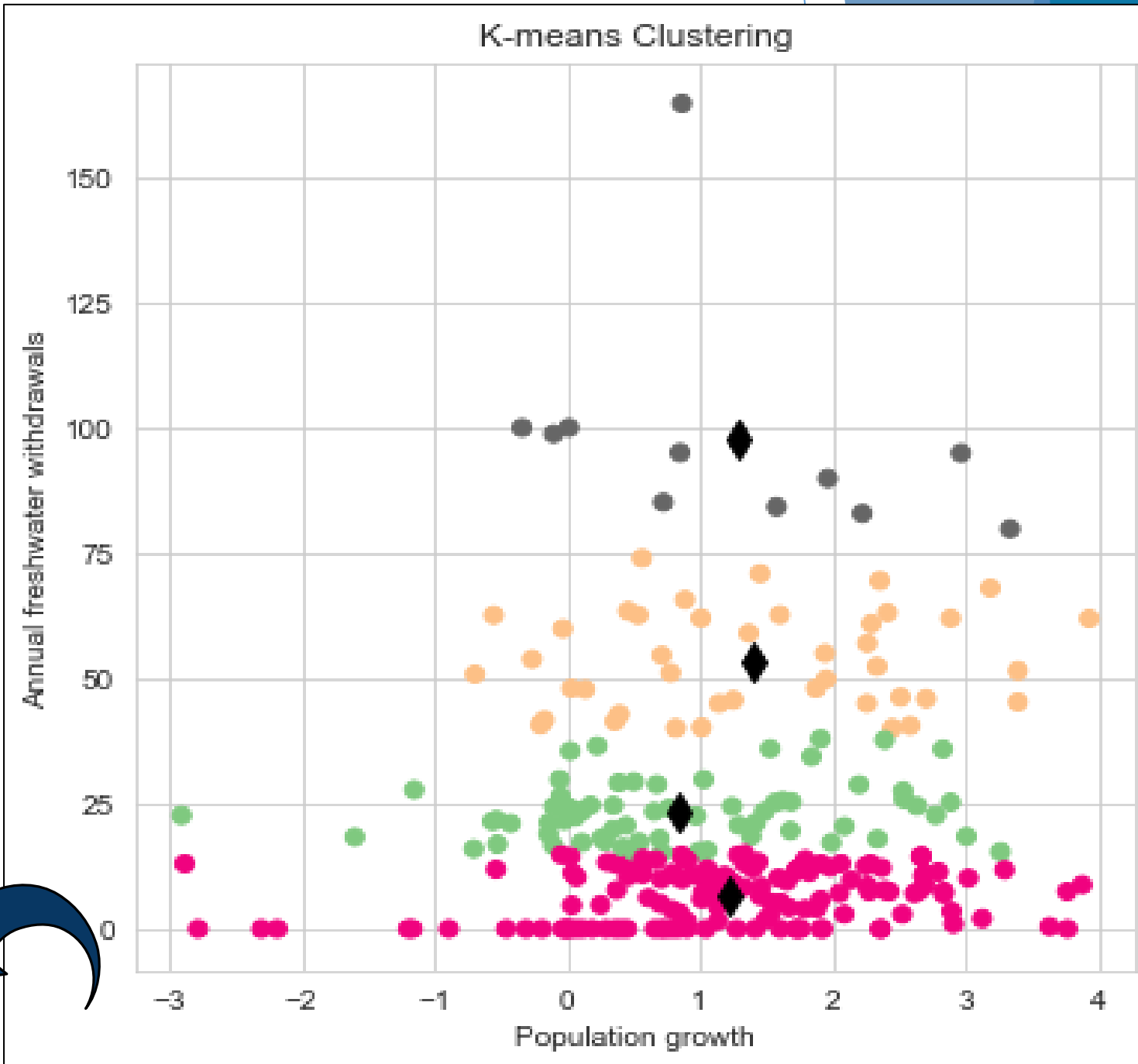
Steps to follow:

Code is used to create a Curve Fit and K Means Clustering for the dataset that contains information about the annual freshwater withdrawals (in cubic meters) by countries, along with their population.

- The **return df** function is used to read the CSV file and returns two data frames, one is of countries as columns and one as years as columns.
- The **convert to numbers** function is used to convert the non-numeric values to numeric values for a given column list of the data frame.
- The **objective** function is used to create a quadratic function that will be used to predict the respective Annual freshwater withdrawals for the population.
- The **curve Fit** function generates the curve fit for a given country using the **objective** function.

Expectations & Predictions:

- The study is expected to reveal separate clusters of countries that have comparable patterns of fresh water withdrawal and population increase.
- The study is expected to indicate that some countries have high rates of fresh water depletion and population increase, while others have low rates. Furthermore, certain countries may have high fresh water consumption rates but low population growth, or vice versa.



CONCLUSION :

In conclusion, the comparison of annual freshwater withdrawal to population growth using clustering analysis can provide valuable insights into the relationship between these two factors. Through this analysis, it is possible to identify different groups of countries with similar patterns of freshwater withdrawal and population growth. However, it is important to note that clustering analysis is just one method of exploring the relationship between freshwater withdrawal and population growth, and it should be complemented by other approaches to gain a more comprehensive understanding of the issue.