

**Department of CSE/IT**

**Jaypee Institute of Information Technology, Noida**

**INTRODUCTION TO DATA WAREHOUSE**

**AND DATA MINING**

**(16B1NCI438)**

**CLUSTERING METHOD**

**Submitted By: Submitted To:**

**Lakshya khera (9917103014) Mr. Avinash Pandey**

**Gaurav Parihar (9917103015)**

**Shubham Tak (9917103031)**

**PROBLEM STATEMENT**

The purpose of this project was to cluster the data set using k-means,

Agglomerative algorithm and Divisive algorithm

**DATASET USED**

Mall Customers dataset

CustomerID

Unique ID assigned to the customer

Gender

Gender of the customer

Age

Age of the customer

Annual Income (k$)

Annual Income of the customee

Spending Score (1-100)

Score assigned by the mall based on customer behavior and spending nature

**LIBRARIES USED:**

* **Pandas –** It is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](https://www.python.org/) programming language.
* **Sklearn –** It is on NumPy, SciPy and matplotlib, this library contains a lot of effiecient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction
* **Matplotlib -** It is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object – oriented API for embedding plots into applications.
* **Numpy –** It is a library for the Python programming language, adding support for large, muti-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

**RESULT**