PROBLEM STATEMENT

We know that information and data of all the Padma bhushan award recipients are in the form of rows and columns. We cannot determine the most popular fields of achievement among recipients, how the recipients' demographics have changed over time, and other significant patterns with such kind of data.

CASE STUDY

The Padma Bhushan awards are third of highest civilian honors in India, awarded annually by the Government of India. They are awarded to individuals for distinguished service of a higher order in any field, including service rendered by government servants. The awards are typically announced on the occasion of Republic Day, which is celebrated on January 26th each year. The awards are given in recognition of outstanding achievements and contributions in various fields such as art, literature, science, public affairs, and sports, among others.

So just by looking at the data set one cannot draw conclusions or analyze about different trends and aspects about particular attribute.

Existing System: The existing system is unprocessed and noisy data. Data contains ambiguity and lack integrity which is complex to draw conclusions and predict future.

Required Resources:

Data Visualization Tools: (e.g., Python libraries like Matplotlib).

Statistical Software: (e.g., Python for statistical analysis).

Report Writing Tools: (e.g., Microsoft Word)

Generate the dataset that includes the Padma Bhushan winners' details. To address errors, inconsistencies, and missing values, clean up the dataset. To understand the award distribution, perform exploratory data analysis (EDA).across numerous sectors, genders, age groups, states, and nations. To find out if there are significant variations in the distribution of prizes depending on different parameters, run statistical tests (such as hypothesis tests). To convey the analysis's conclusions, create photographs (such as bar charts, histograms).On the basis of the analysis, offer conclusions and suggestions for further thinking.
The deliverables of the analysis will include a report and data visualizations. The analysis is expected to be completed within a timeline of two months, with key milestones including data collection, analysis, and report writing. The analysis will be conducted using Python programming language and relevant data analysis libraries. Here is an over view of aspects of drawing conclusion.

1. Trends in Fields of Achievement
2. Demographic Trends:

3. Historical Trends:
4. Regional Representation
5. Impact of the Awards
6. Underrepresented Groups

ABSTRACT.

Perform a data analysis on the recipients of the Padma Bhushan award to gain insights into the trends, patterns, and demographics of the awardees. The data set has year of receiving the award, recipient name, from what state of the country they belong to, recipient age when receiving the reward, field of study and post effects of the awards

By analyzing the data in the Padma Bhushan award individuals we can understand the trends and patterns in Padma Bhushan awards recipients over the years. The analysis will be based on a dataset containing information about the recipients, including their name, year of award, field of achievement, and other relevant variables let's say their age, gender etc. The analysis will involve data cleaning, data preprocessing to identify the most common fields of achievement among recipients, how the demographics of recipients have changed over time, and other relevant trends. The expected outcome of the analysis is to gain insights into the impact of the Padma Bhushan awards on various fields and to highlight any underrepresented groups among the recipients.