Name: Rohit Sanjay Inamdar
SBU ID: 114504643

CSE 564: Visualization(Spring 2023) as Basic Project CSE 522 Assignment 2a Report

About Dataset:
The dataset I used for this assignment(same as assignment 1) consists of data pertaining to the requirements for a PC/Video game to function on Computer. The reason for using this dataset is because nowadays numerous high-end games are being developed and many remastered versions of older games are released. Hence this dataset contains information on what would be the optimal requirements(hardware specs) for smoothly running a particular game on a given system.
The dataset was obtained from Kaggle.
Link: https://www.kaggle.com/datasets/baraazaid/pc-video-game-requirements-v2
Number of columns in the actual dataset: 90
Number of records in the actual dataset: 10843
I've selected 1000 random records from the original dataset and 16 columns from the original dataset.

Column Names(categoricals encoded):
1. **Min_OS(encoded):** Required Operating System for the game to run
2. **Recom_OS(encoded):** Recommended Operating Operating System for the game
3. **Min_Direct_X(encoded):** Required Direct X setup version(important for running high-end games)
4. **Min_HDD_Space(encoded):** Required Hard Disk space since most games are nowadays have disk size >= 100GB
5. **Recom_CPU_Max_Temp:** Recommended CPU operating temperature to achieve optimal gaming performance(typically around 40 - 80 degree celsius)
6. **Recom_HDD_Space(encoded):** Recommended Hard Disk Space(useful for cost effectiveness)
7. **Min_CPU_Lithography:** Required CPU Lithography describes the distance between transistors embedded on the CPU (the lower the value(in nm), the better the CPU)
8. **Min_CPU_Physical_Cores:** Number of required cores in CPU(more cores means higher performance)
9. **Recom_CPU_CPU_Speed:** CPU performance speed(in GHz) crucial for running high end games
10. **Recom_CPU_Lithography:** Recommended CPU Lithography(usually lower than required)
11. **Min_GPU_Memory:** Required GPU VRAM memory(for achieving higher texture and frame processing)

12. **Min_GPU_Memory_Speed:** Required VRAM processing speed for faster processing of pixels, textures and frames
13. **Min_GPU_Pixel_Rate:** Required rate for processing a certain number of pixels per second
14. **Min_GPU_Texture_Rate:** Required texture rate(used for texture filtering) in GigaTexels/second
15. **Min_GPU_Memory_Bandwidth:** Required GPU bandwidth(higher value implies faster transfer of processed components such as pixels, textures and frames) in Gb/s.
16. **Min_GPU_Direct_X:** Required Direct X setup for GPU.

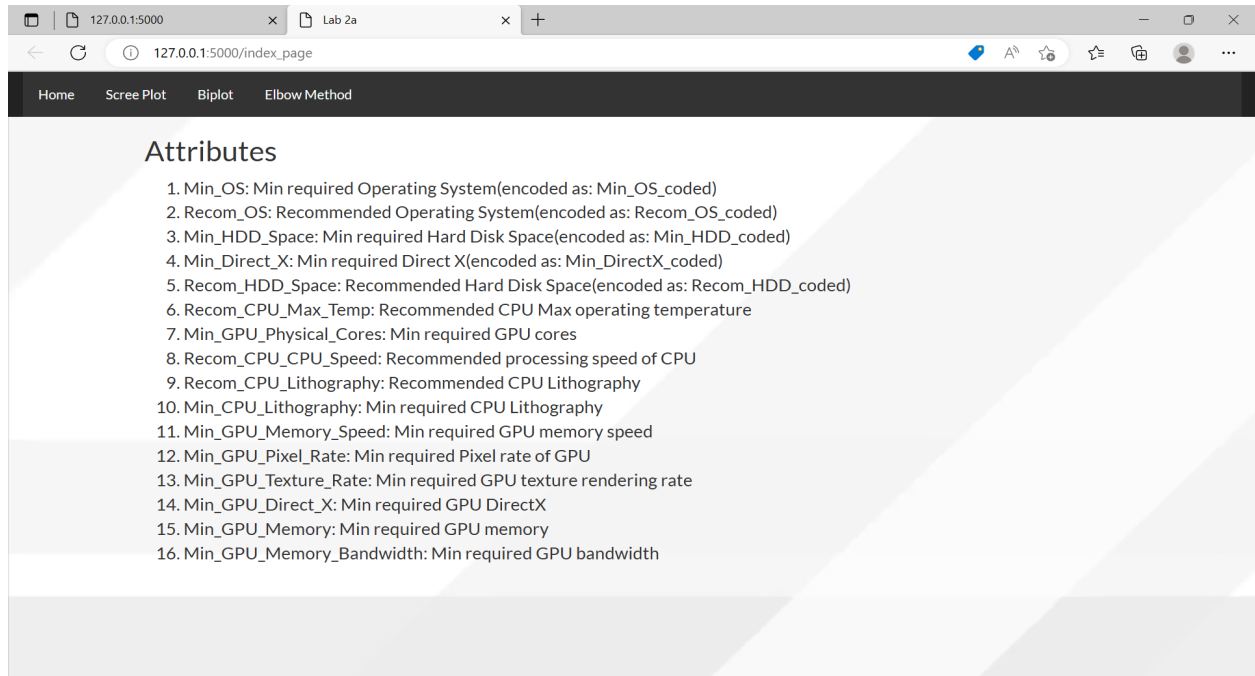**Tasks: Visualization tools in visual analytics**
Front page dashboard:
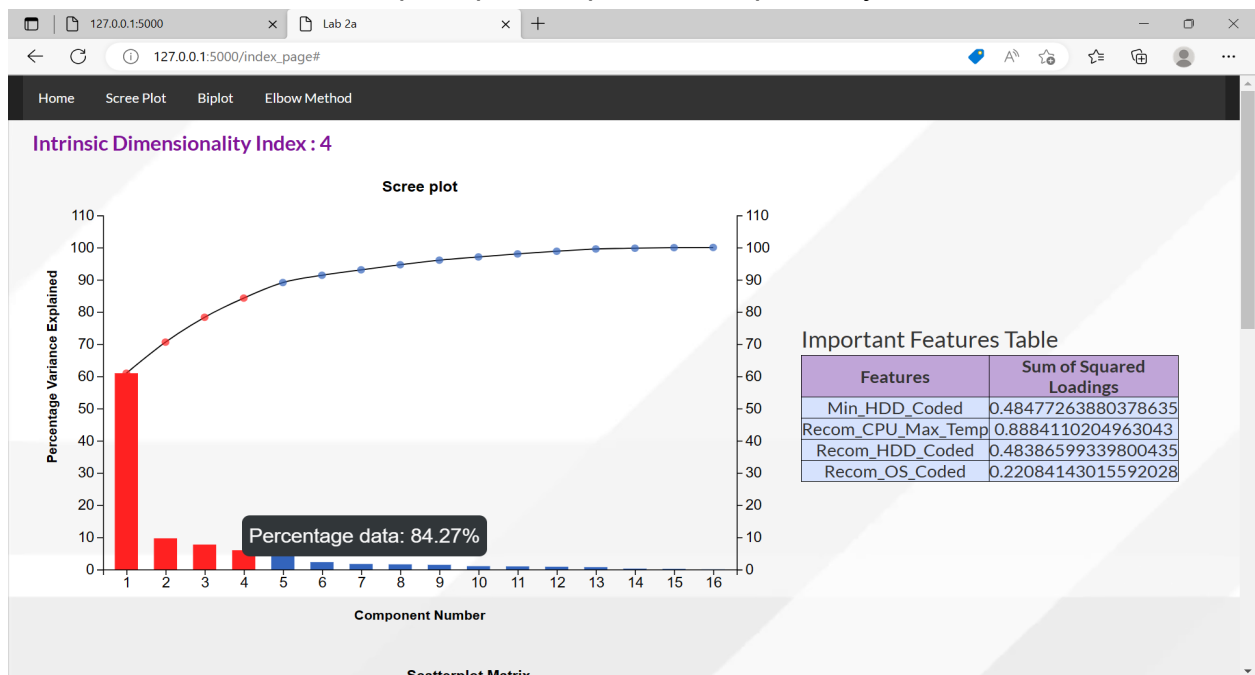The image below shows the front page of the visual interface. This is done using bootstrap
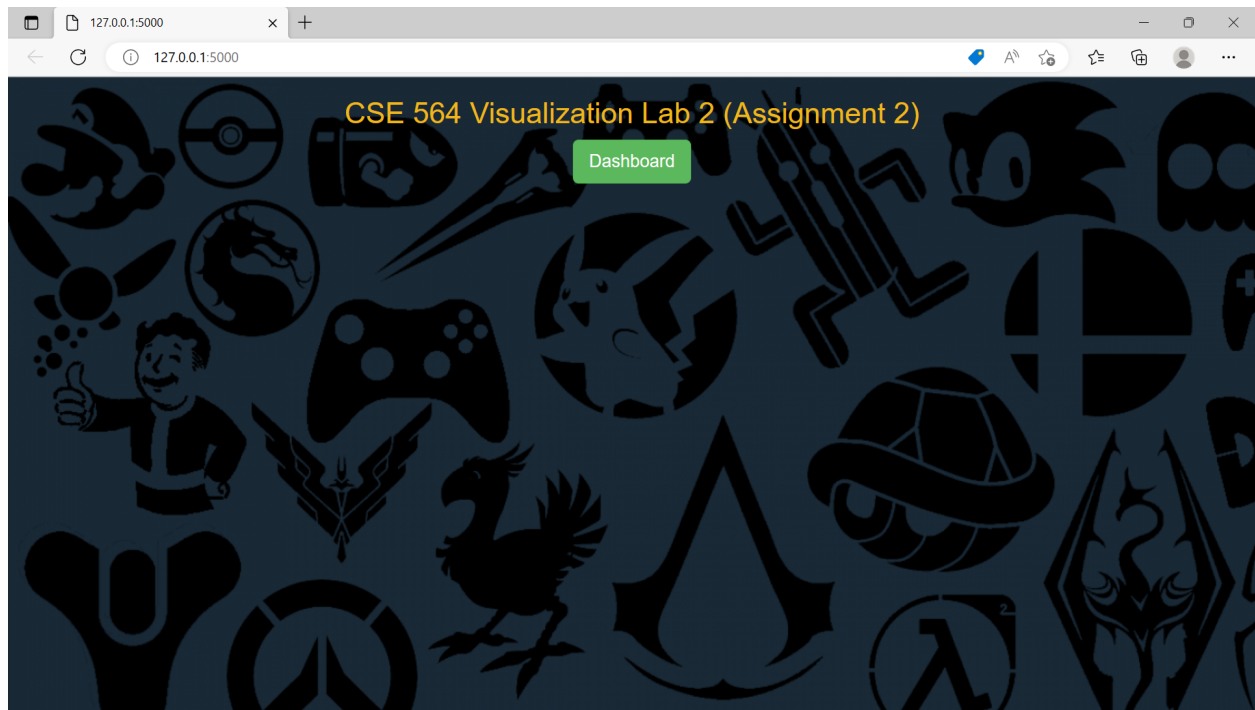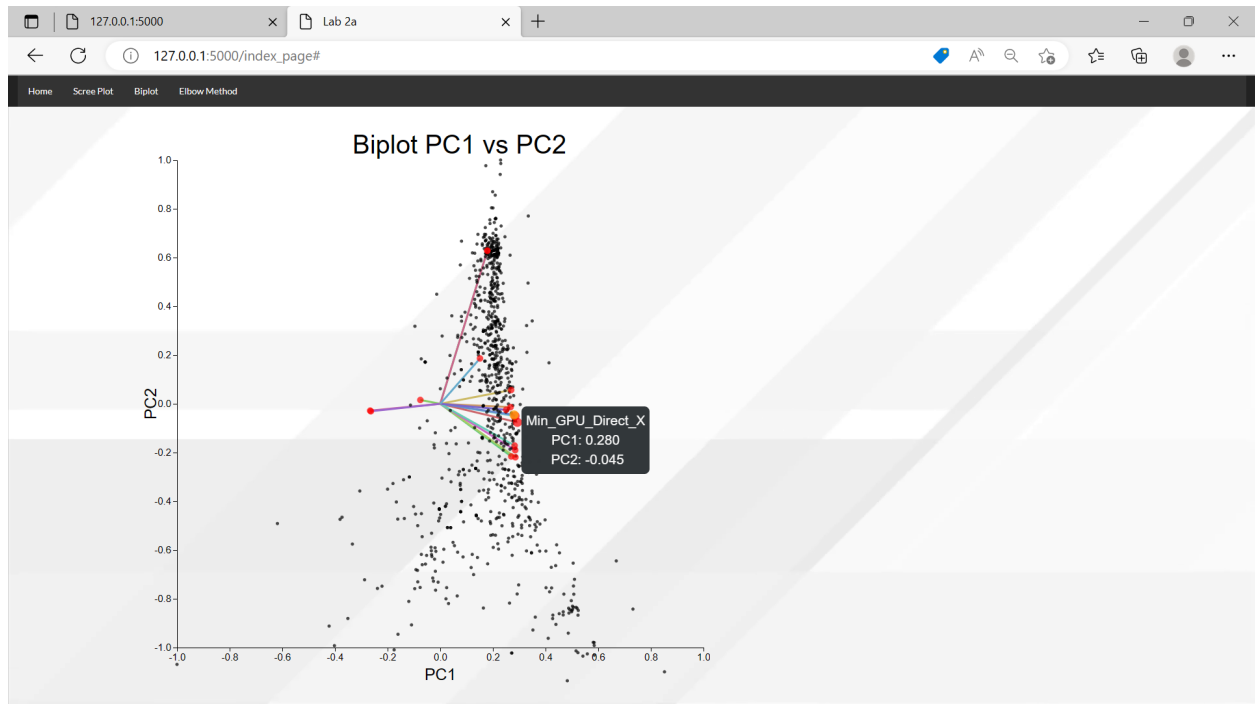

Menu Page:
The image below shows the menu page which gets loaded on clicking the "Dashboard" button in the previous page. This page displays the attributes that are used for the PCA analysis.

Attributes

1. Min_OS: Min required Operating System(encoded as: Min_OS_coded)
2. Recom_OS: Recommended Operating System(encoded as: Recom_OS_coded)
3. Min_HDD_Space: Min required Hard Disk Space(encoded as: Min_HDD_coded)
4. Min_Direct_X: Min required Direct X(encoded as: Min_DirectX_coded)
5. Recom_HDD_Space: Recommended Hard Disk Space(encoded as: Recom_HDD_coded)
6. Recom_CPU_Max_Temp: Recommended CPU Max operating temperature
7. Min_GPU_Physical_Cores: Min required GPU cores
8. Recom_CPU_CPU_Speed: Recommended processing speed of CPU
9. Recom_CPU_Lithography: Recommended CPU Lithography
10. Min_CPU_Lithography: Min required CPU Lithography
11. Min_GPU_Memory_Speed: Min required GPU memory speed
12. Min_GPU_Pixel_Rate: Min required Pixel rate of GPU
13. Min_GPU_Texture_Rate: Min required GPU texture rendering rate
14. Min_GPU_Direct_X: Min required GPU DirectX
15. Min_GPU_Memory: Min required GPU memory
16. Min_GPU_Memory_Bandwidth: Min required GPU bandwidth

1. Basic dimension reduction and visualization using PCA:
   In this task the PCA based scree plot is displayed in which the user can interact with the bars on the graph and select the required dimensionality indices i.e number of bars selected will set the index number and the biplot can be viewed as well. On moving the mouse cursor on the scree plot bar or the biplot attributes, we can view the percentage variance and the contribution of the attribute towards the principal components respectively

Intrinsic Dimensionality Index : 4

Scree plot

Percentage data: 84.27%

Important Features Table

| Features	Sum of Squared Loadings
Min_HDD_Coded	0.48477263880378635
Recom_CPU_Max_Temp	0.8884110204963043
Recom_HDD_Coded	0.48386599339800435
Recom_OS_Coded	0.22084143015592028

Scatterplot Matrix

** webpage zoomed out to avoid scrolling **

2. Data visualization using scatter plot matrix:
   In this task, the intrinsic dimensionality index selected in the scree plot will display the feature table containing the top 4 features and their sum of squared loading values. A scatter plot matrix representing these 4 attributes with their unique coloured clusters on the basis of k means will be displayed.
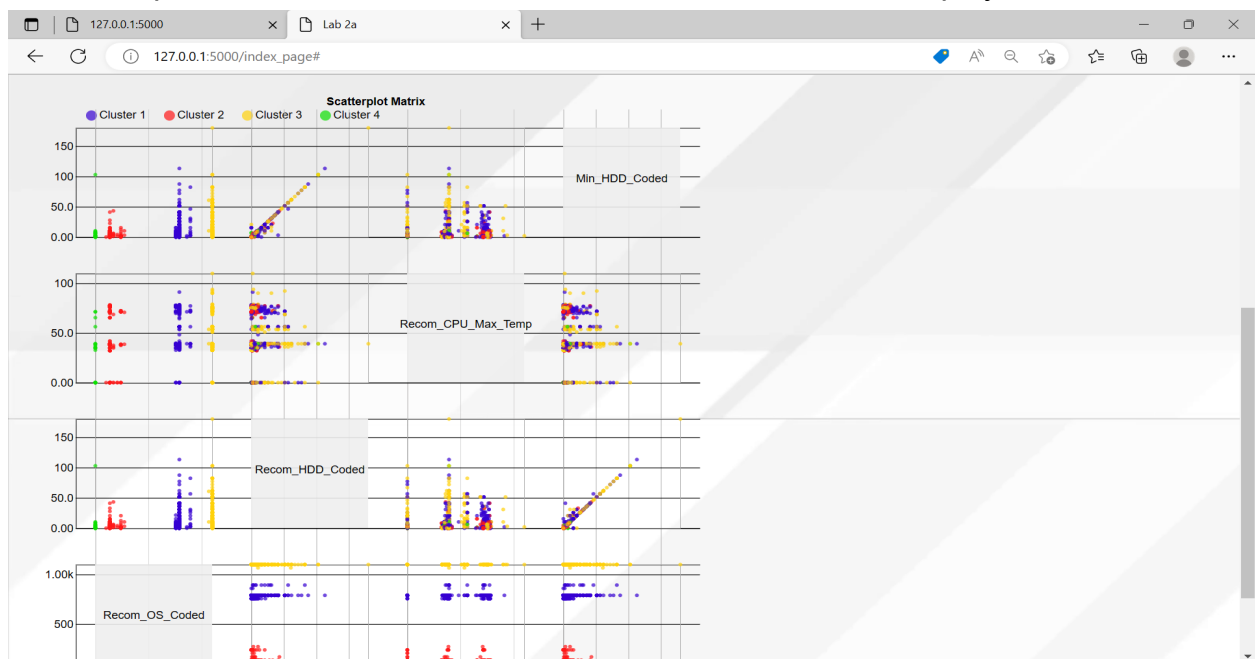
3. In this task the elbow plot describing the optimal value of k is displayed.