**DATA SCIENCE**

**PROJECT**

**MUNZIR KALIM AHMED(20P-0477)**

**&**

**NABEEL YASEEN(20P-0486)**

**BSSE-7A**

**ETL:**

In our data processing workflow, we employ the Extract, Transform, and Load (ETL) methodology to seamlessly transition from raw data to a refined state. The process begins with the extraction of raw data, followed by a transformative phase where we simplify its structure and introduce necessary changes. This ensures that the data becomes more manageable and relevant to our analytical objectives. Subsequently, the refined data is loaded into an Excel file, a structured format that facilitates easy accessibility and comprehension. By systematically executing these steps, we enhance the overall quality and utility of the data, paving the way for more effective analysis and interpretation.

**Extract:**

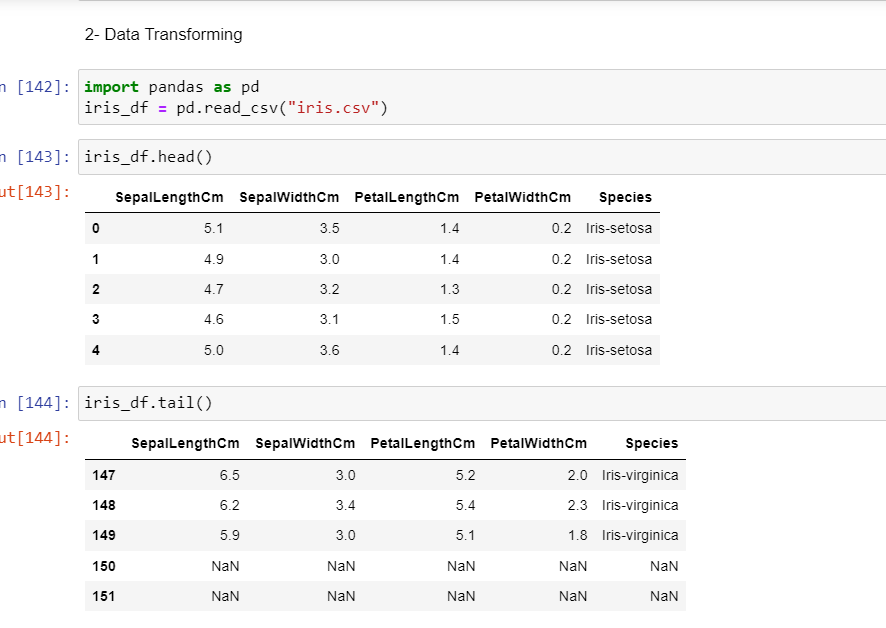
To commence our data processing task, we initiate the download of the "iris.data" dataset. Following the download, we embark on the extract process, where we meticulously extract the data from its original row-oriented format. Subsequently, we transform the extracted data into a more structured arrangement by organizing it into columns. This organized dataset is then systematically stored in a CSV (Comma-Separated Values) file. This structured file format ensures that the data is neatly organized, making it conducive to efficient storage, retrieval, and subsequent analysis. Through this methodical approach, we establish a streamlined dataset ready for further exploration and utilization.

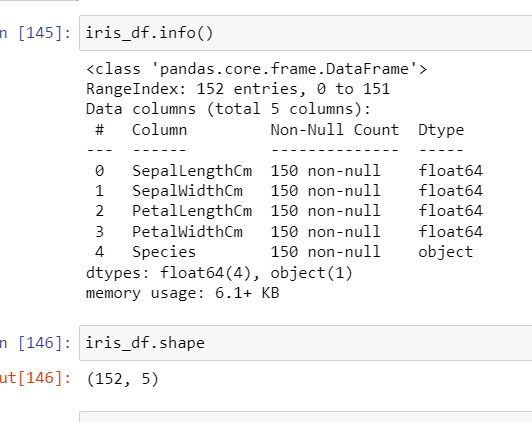


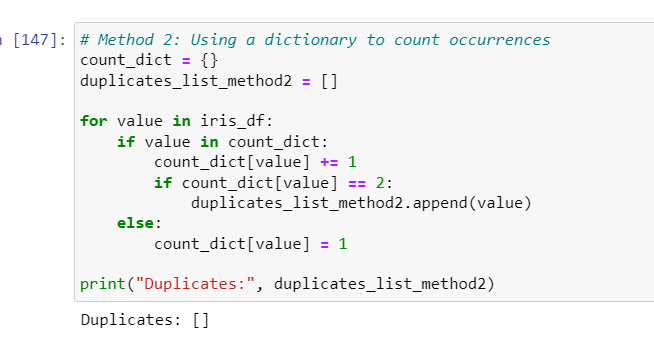
**Transform:**

In the course of our data processing endeavor, we engage in a transformative phase aimed at enhancing the quality and usability of the dataset. This involves a meticulous examination to identify and address missing values and duplicates, ensuring the integrity of the data. Additionally, we undertake a restructuring of columns by assigning them meaningful and descriptive names, thereby enhancing clarity and interpretability.

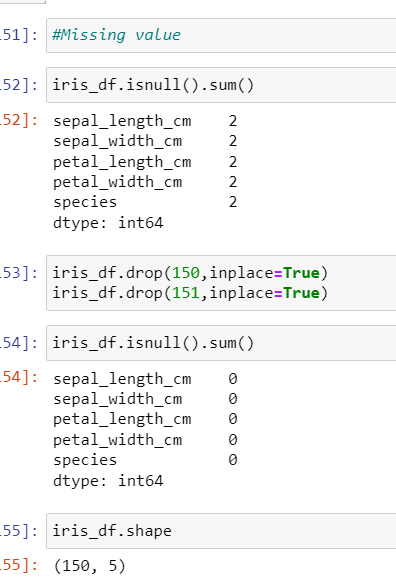
Furthermore, we delve into the exploration of interrelationships within the dataset by calculating correlations between each pair of columns. This insightful analysis unveils dependencies among the variables, shedding light on how they influence each other. By systematically conducting these transformations and correlations, we not only refine the dataset but also lay the groundwork for more informed and nuanced analyses.

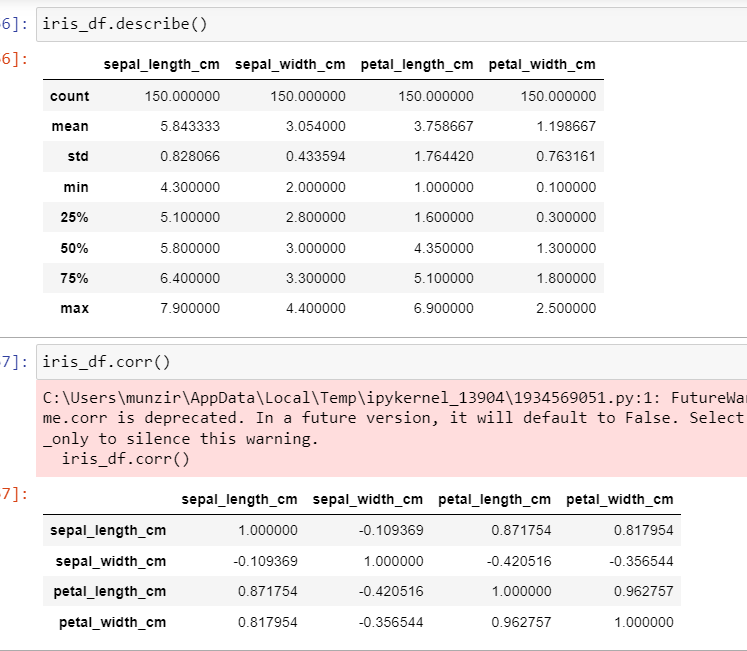


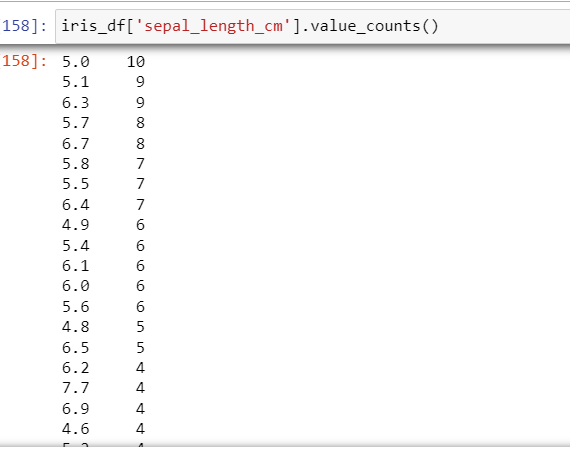


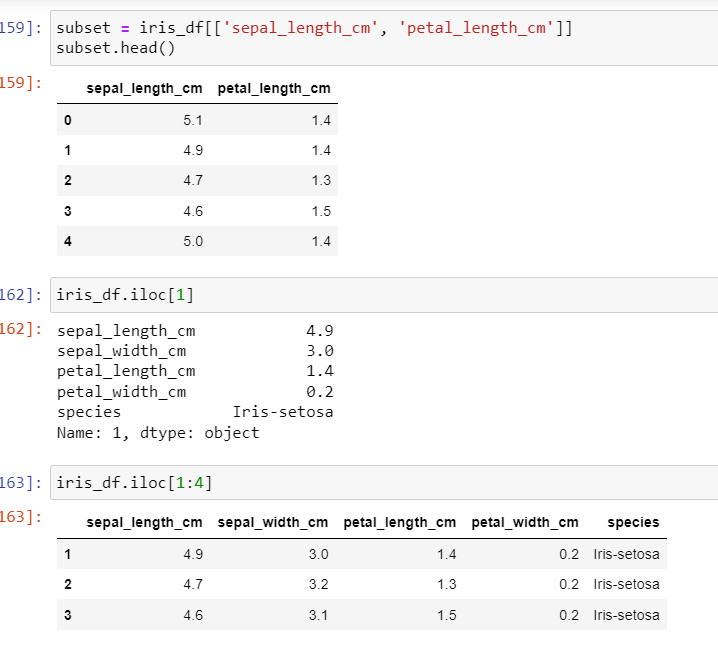


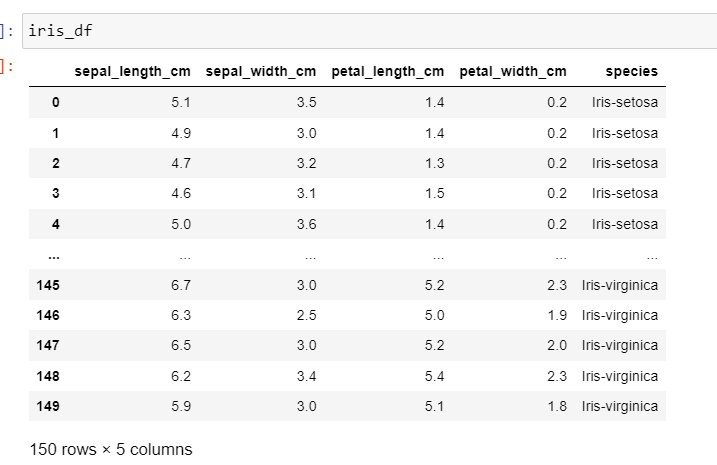












**Load:**

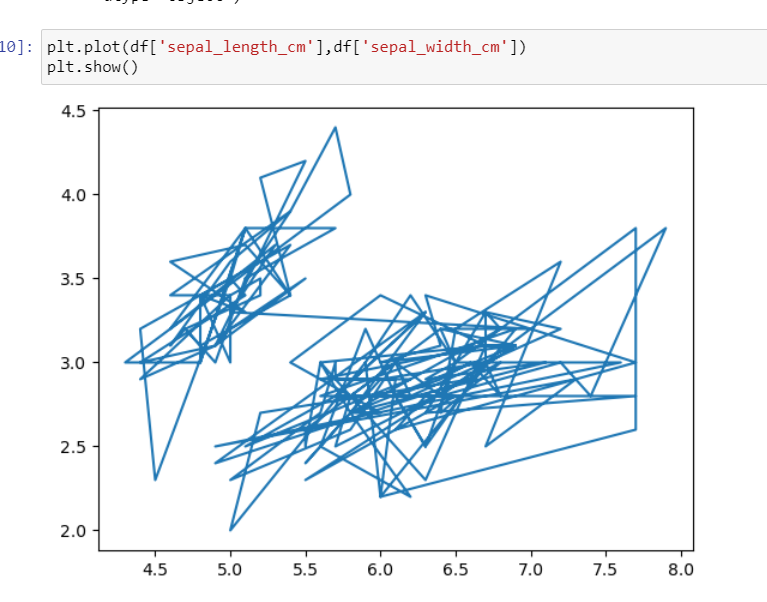
In the final stage of our data processing workflow, we execute the Load process by incorporating the changes made to our dataset into an Excel file. This involves systematically sending and recording the refined data, along with all the modifications and enhancements, into the structured format of an Excel file. By doing so, we create a comprehensive and organized record of the processed data. This Excel file serves as a repository of the refined information, providing a clear and accessible snapshot that reflects the culmination of our data extraction, transformation, and loading efforts.

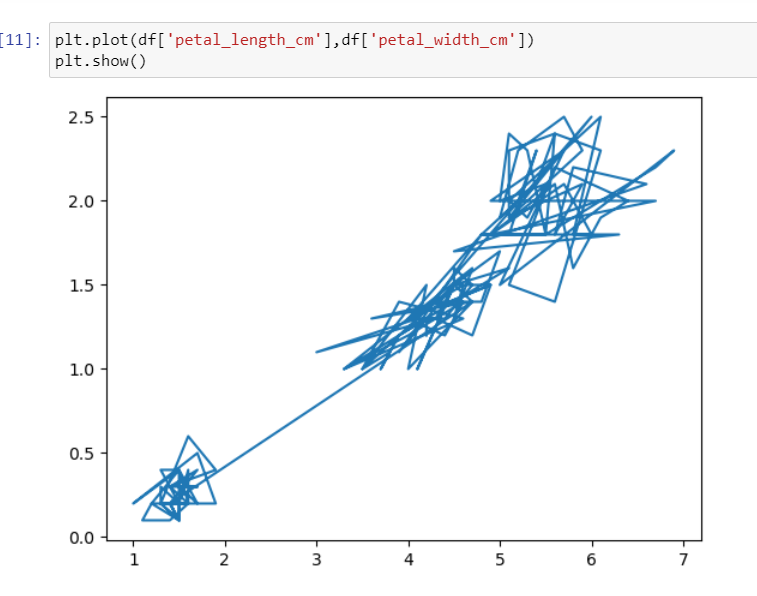


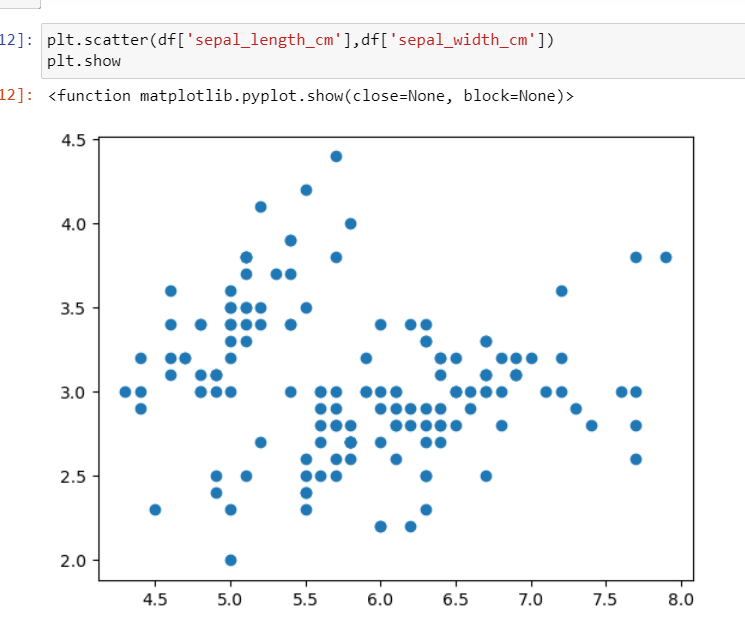
**Visualize:**

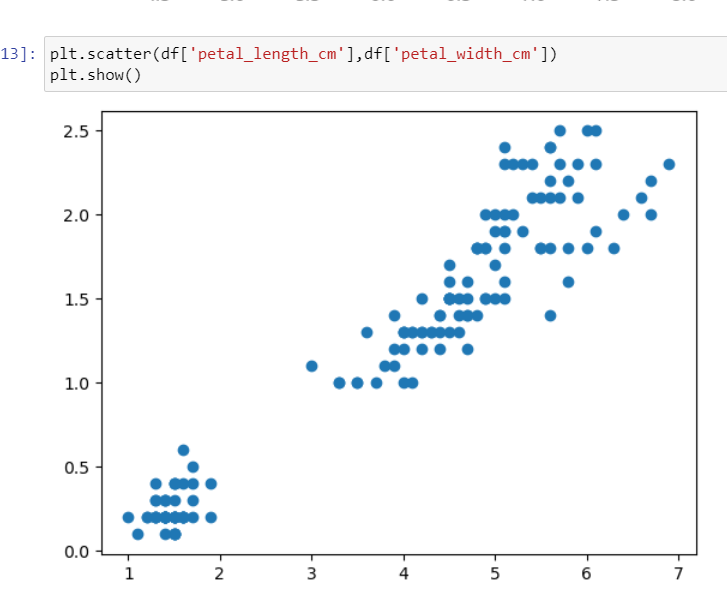
In the next phase of our data analysis, we initiate the visualization process to create informative graphs that offer a visual representation of our dataset. These visualizations serve as powerful tools to unveil patterns, trends, and insights that might not be immediately apparent in raw data. By leveraging various graph types, we enhance our ability to comprehend complex information and perform different analytical tasks. This visual exploration not only facilitates a deeper understanding of the dataset but also provides a visually intuitive platform for further analysis and decision-making

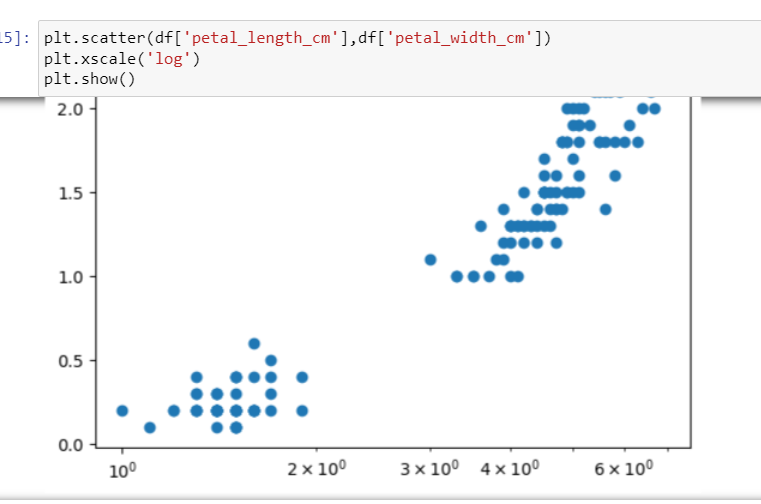


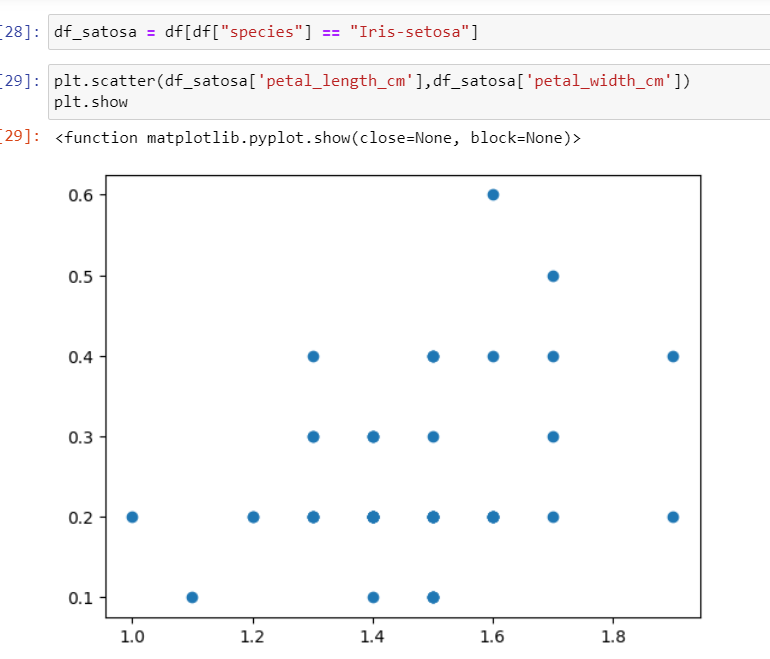


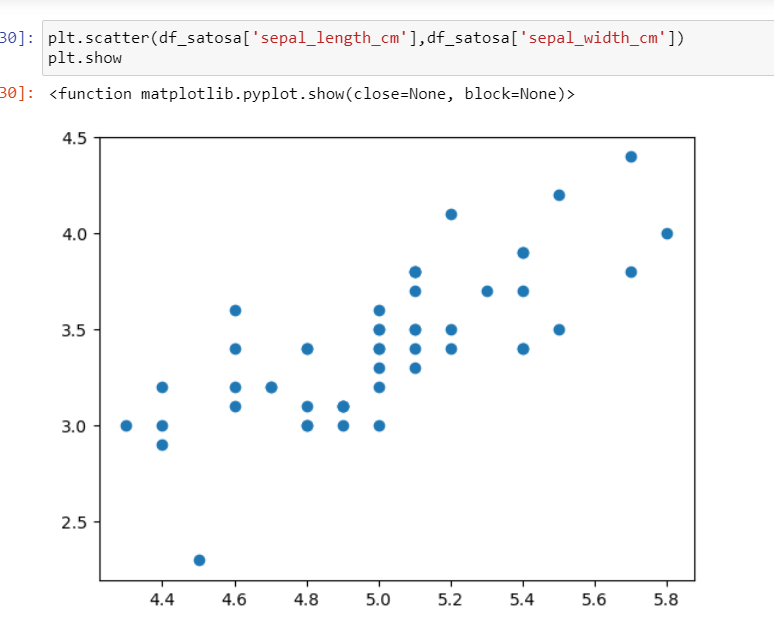


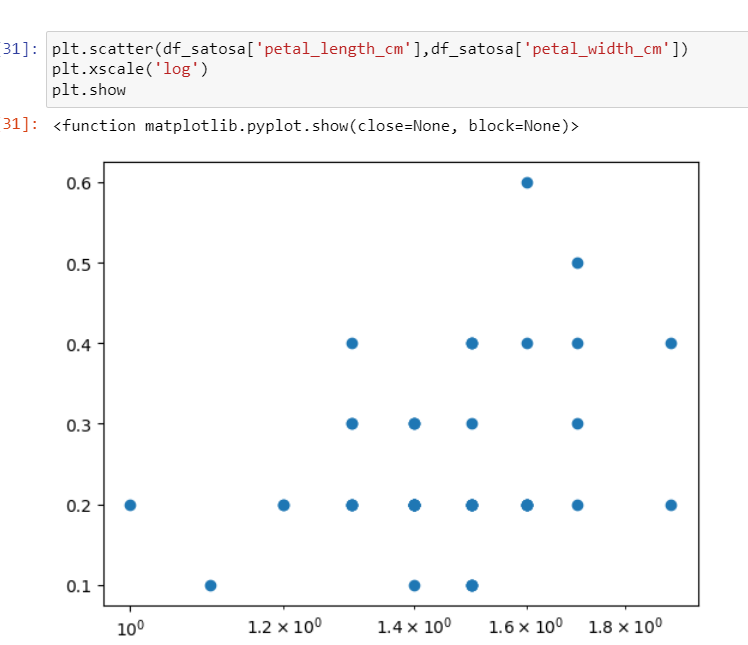


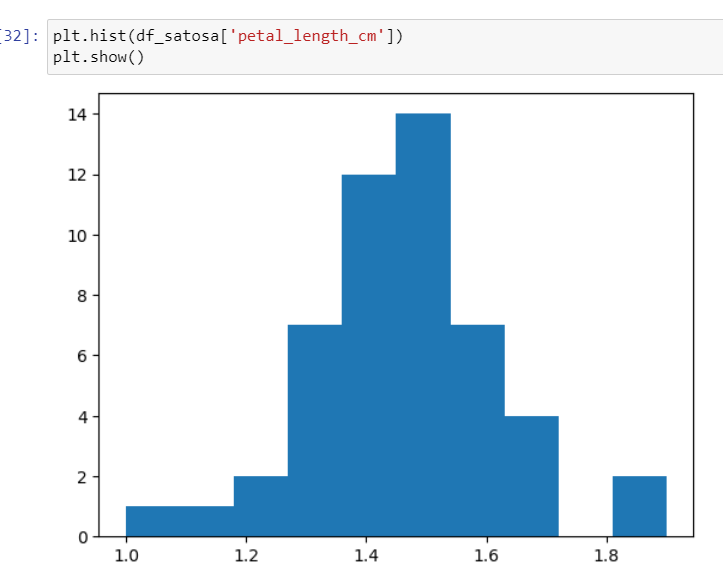


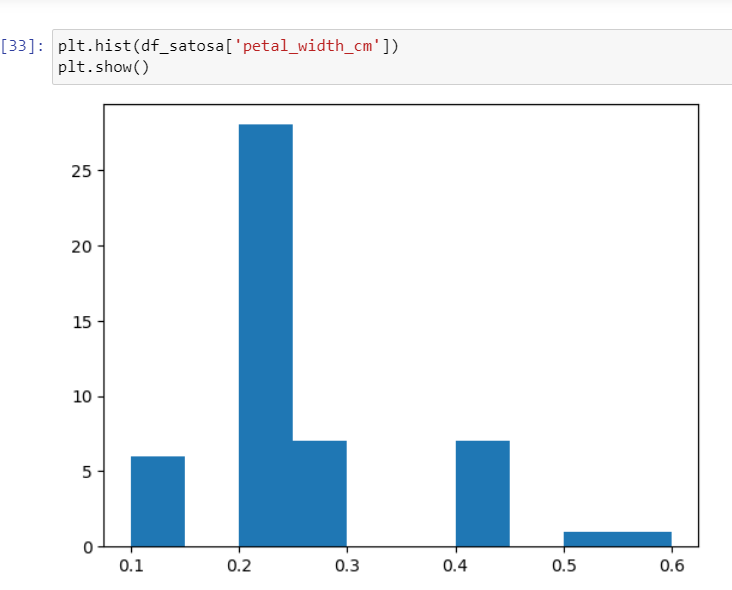


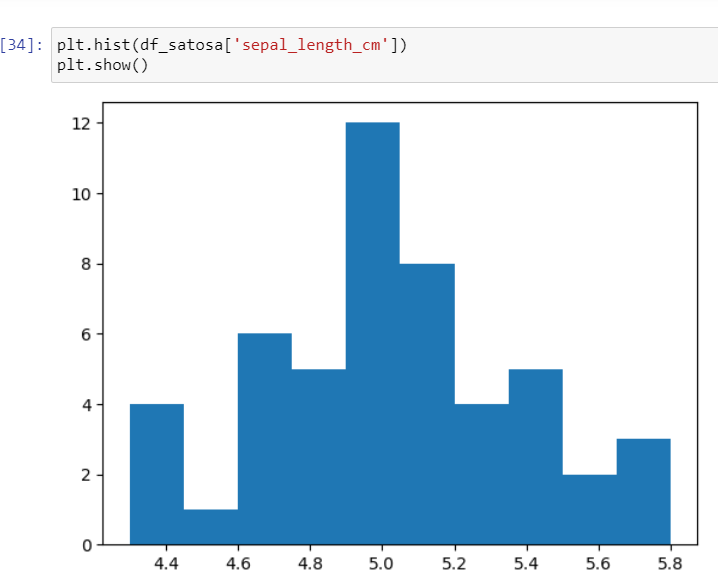


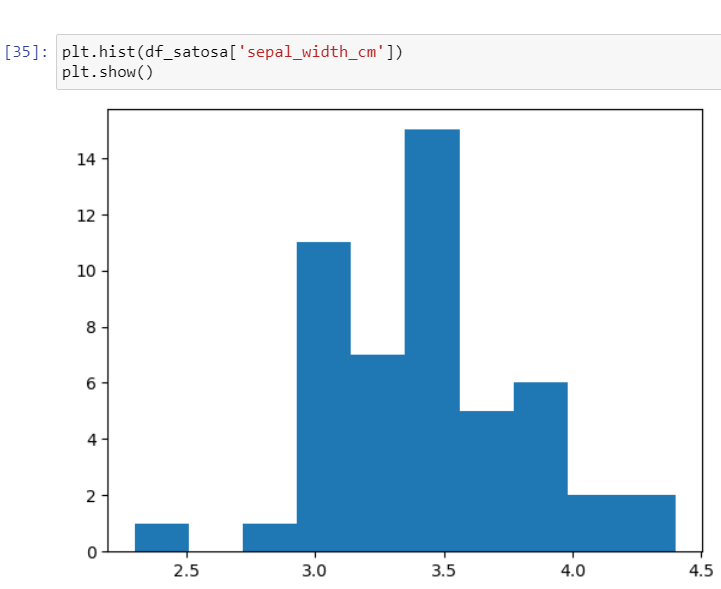


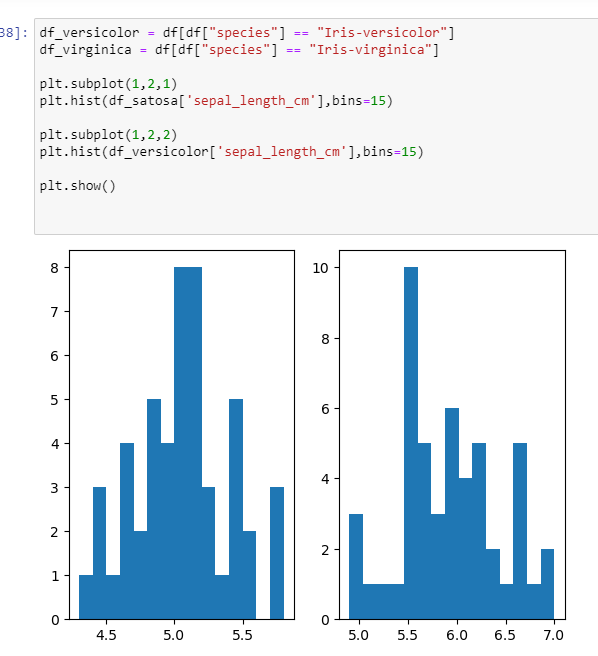


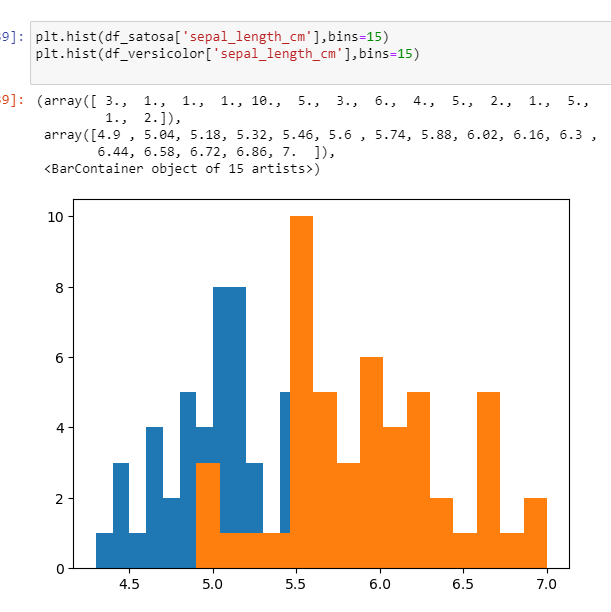


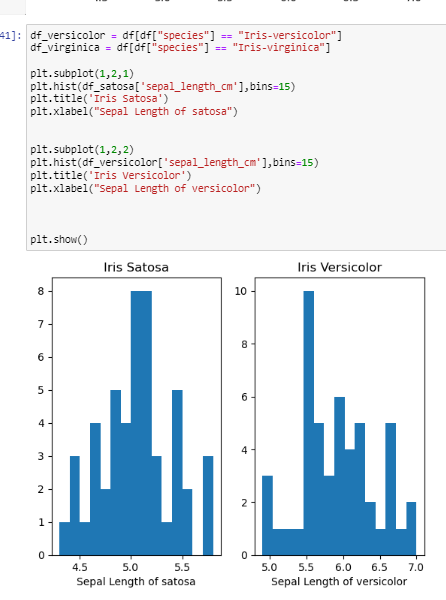






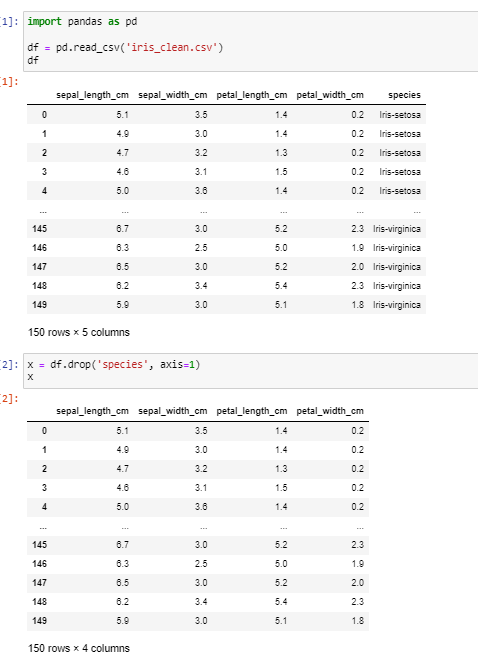


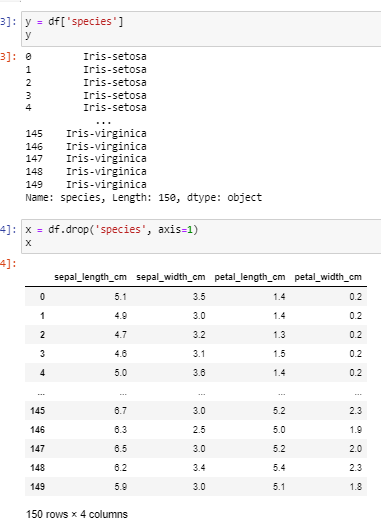


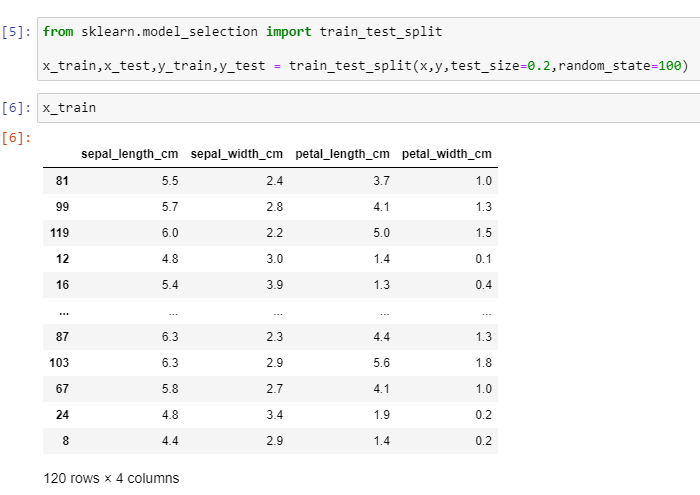


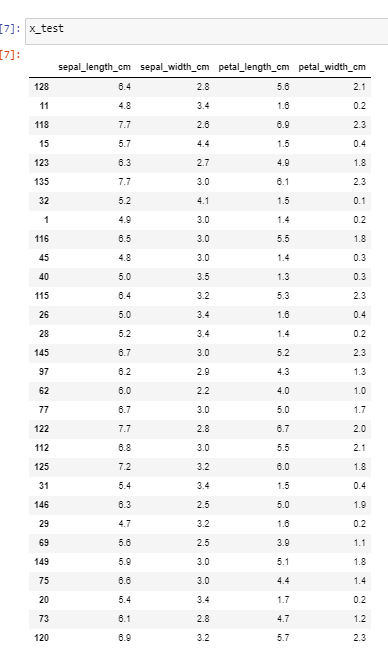
**Machine learning:**

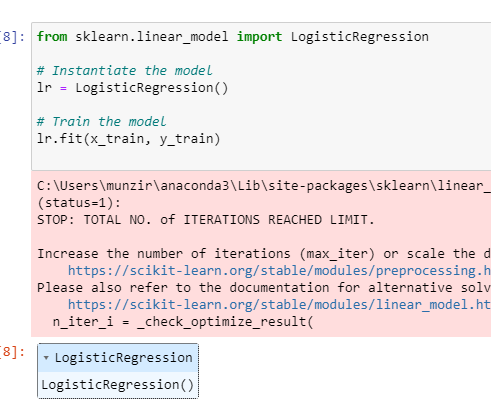
Our approach involves the selection of diverse models for training on the dataset, allowing us to thoroughly evaluate and compare their performance. This systematic process entails training each chosen model on the data, meticulously examining how well they grasp and interpret the underlying patterns. By rigorously assessing the performance of these various models, we gain valuable insights into their respective strengths and weaknesses. This methodical examination is crucial in determining the most suitable model for our specific dataset, ensuring that our final choice aligns optimally with the intricacies of the data at hand.

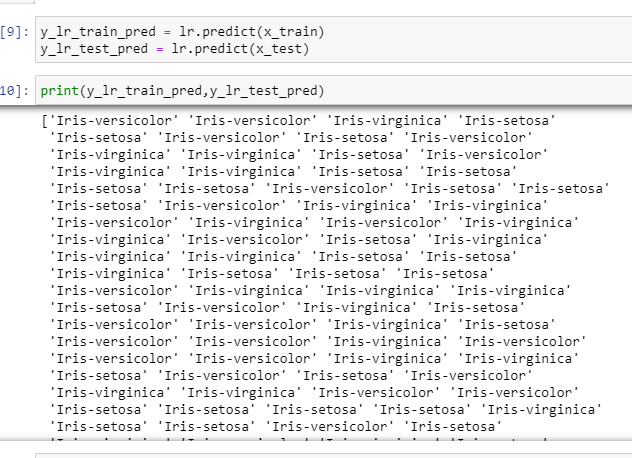


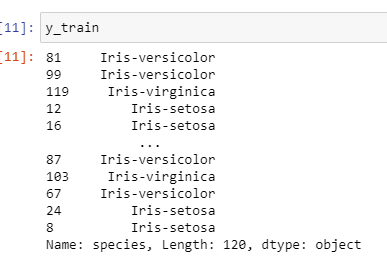


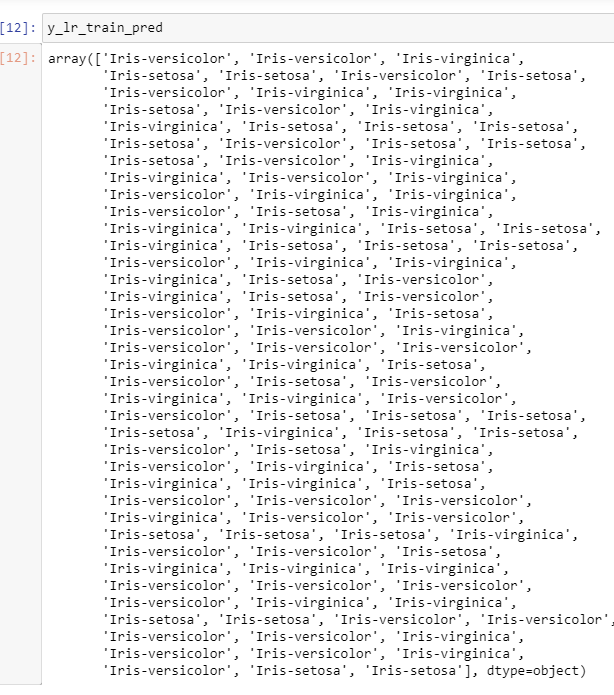




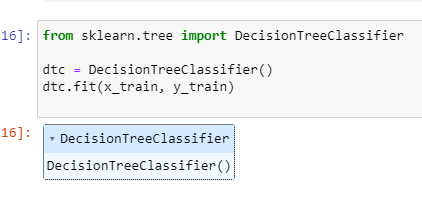


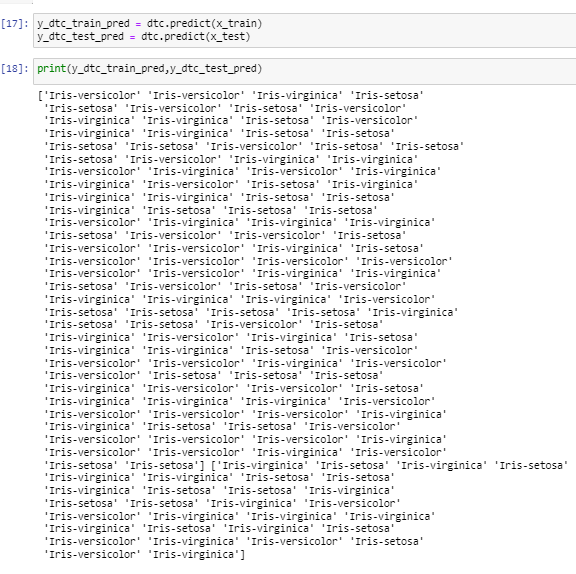


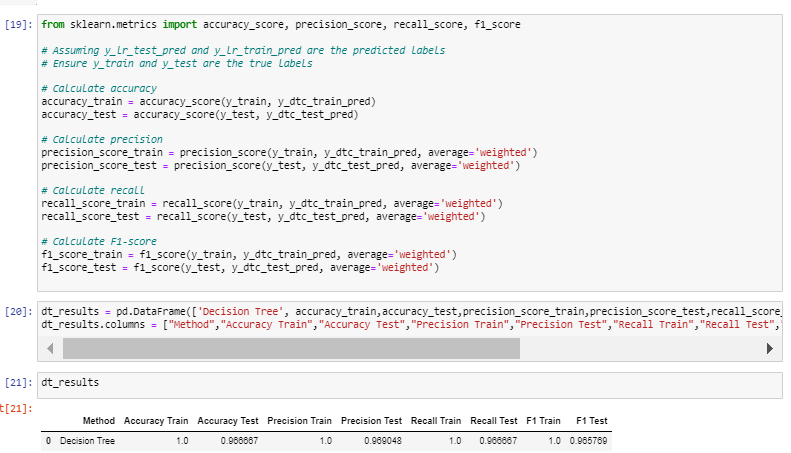


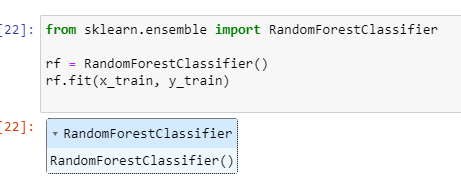


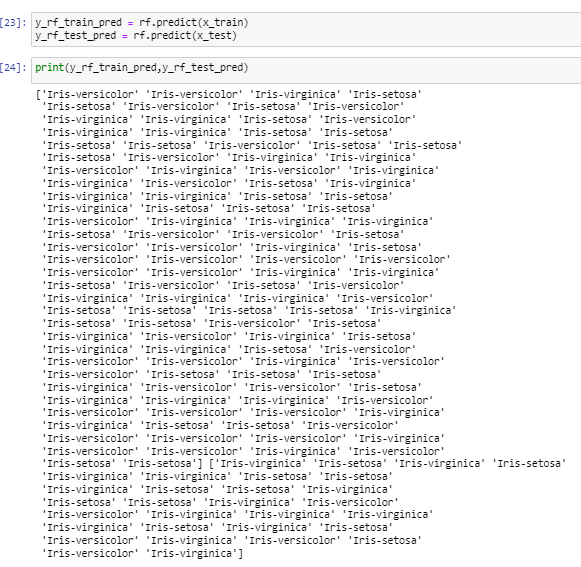




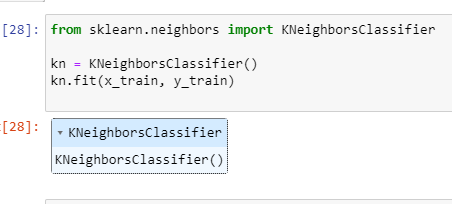


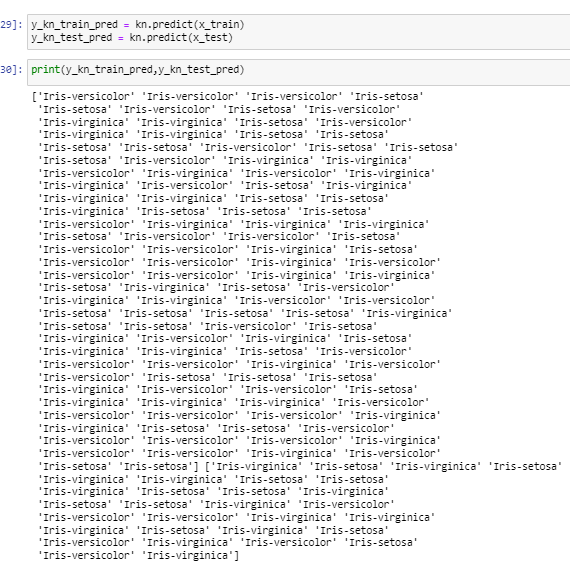




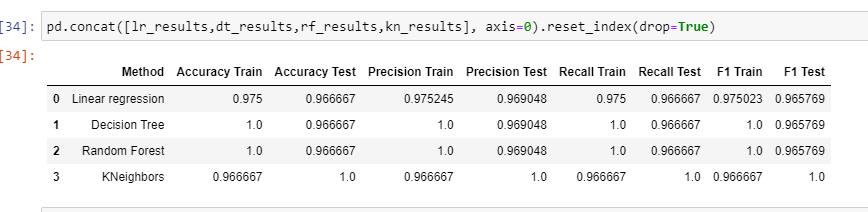












**Story telling:**

In the process of transforming raw data from the "iris.data" dataset, we engage in a comprehensive Extract, Transform, and Load (ETL) procedure. This involves extracting pertinent information, performing crucial preprocessing steps to enhance the meaningfulness of the data, and ultimately loading the refined data into a structured format, specifically an "iris.csv" file. Subsequently, we delve into the realm of visualization, employing various graph types to distill complex information into comprehensible insights. These visualizations serve as powerful tools to unveil nuances that might remain concealed in the raw data. Following the ETL and visualization stages, we embark on the machine learning phase. This entails training the dataset using different models to ascertain their accuracy and effectiveness in handling the underlying patterns within the data.