Automated Data Analysis Report

1. Dataset Overview

Columns: Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm, Species

Rows: 165

2. Basic Statistics

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	165.000000	161.000000	159.000000	159.000000	165.000000
mean	83.000000	7.510559	4.713208	4.022642	2.810909
std	47.775517	11.162995	13.468999	2.052505	12.097722
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	42.000000	5.100000	2.800000	1.600000	0.400000
50%	83.000000	5.800000	3.000000	4.500000	1.500000
75%	124.000000	6.500000	3.300000	5.150000	1.900000
max	165.000000	116.300000	130.000000	13.800000	111.100000

3. Missing Values

Id	0
SepalLengthCm	4
SepalWidthCm	6
PetalLengthCm	6
PetalWidthCm	0
Species	5

4. Key Insights

1. The accuracy of the classification model is 1.00.

2. The most important feature is 'SepalWidthCm' with a coefficient of 0.0698.

3.

The pivot table shows the average Id values for each species of Iris. The Iris-setosa species has the lowest average Id value of 25.52, indicating that the sepal width and petal width of this species are generally smaller compared to the other two species. The Iris-versicolor species has an average Id value of 75.29, which is higher than that of Iris-setosa but lower than Iris-virginica. The Iris-virginica species has the highest average Id value of 132.49, suggesting that the sepal width and petal width of this species are significantly larger than the other two species. This could potentially indicate differences in the growth and development of these species.

4.

The bar plot displays the average 'Id' for each species in the dataset. Among the three species, 'Iris-virginica' has the highest average 'Id' value of 132.49, indicating that this species has the most instances or observations in the dataset. On the other hand, 'Iris-setosa' has the lowest average 'Id' value of 78.86, suggesting that it has fewer instances compared to the other two species. 'Iris-versicolor' falls in between these two species with an average 'Id' of 112.47. These findings may imply that the distribution of instances or observations across the three species is not equal, with 'Iris-virginica' having the most instances and 'Iris-setosa' having the least.

5.

The bar plot illustrates the average Sepal Length in centimeters for each species of Iris flowers. Notably, the Iris-versicolor species has the highest average sepal length of 8.18 centimeters, which is significantly larger than the other species. Conversely, the Iris-setosa species has the smallest average sepal length of 5.01 centimeters. Lastly, the Iris-virginica species has an average sepal length of 6.59 centimeters, which falls between the other two species. These differences in sepal length could potentially be used to differentiate between the species of Iris flowers.

6.

The bar plot illustrates the average sepal width in centimeters for each species of iris flowers. Interestingly, the iris species 'Iris-setosa' has the highest average sepal width of 6.75 centimeters, which could suggest that these flowers generally have larger sepals compared to the other two species. On the other hand, 'Iris-versicolor' and 'Iris-virginica' have relatively lower average sepal widths of 5.94 and 6.00 centimeters, respectively. These findings might indicate that the 'Iris-setosa' flowers have distinct characteristics in terms of sepal size among the three species.

7.

The bar plot reveals that the average Petal Length in centimeters for the species of Iris varies among the three categories. The species 'Iris-virginica' has the highest average Petal Length of 5.52 centimeters, suggesting that the petals of this species tend to be larger compared to the other two species. On the other hand, 'Iris-setosa' has the lowest average Petal Length of 1.46 centimeters, indicating that the petals of this species are generally smaller. 'Iris-versicolor' falls in between the two with an average Petal Length of 4.26 centimeters. This suggests a noticeable difference in the size of petals among the three species of Iris.

8.

The bar plot shows the average petal width in centimeters for three different species of irises. Among these species, the Iris-versicolor has the highest average petal width of 3.87 centimeters. This suggests that the petals of Iris-versicolor are generally wider than those of the other two species. The average petal width for Iris-setosa is 1.46 centimeters, which is significantly smaller than that of Iris-versicolor, and the average petal width for Iris-virginica is 2.97 centimeters, which is still smaller than Iris-versicolor but larger than Iris-setosa. In summary, the Iris-versicolor has the largest petals, followed by Iris-virginica, and Iris-setosa has the smallest petals.

9.

The histogram of 'SepalLengthCm' indicates a right-skewed distribution, suggesting that there are more observations with smaller values of sepal length than larger ones. This could potentially be due to the presence of immature or smaller species in the dataset, or perhaps an outlier with a significantly smaller sepal length value. Further analysis and exploration of the data may help to better understand the reasons behind this skewed distribution.

10.

The histogram of 'SepalWidthCm' exhibits a skewed distribution, with a longer tail extending towards the right. This suggests that there is a higher concentration of data points with smaller values of sepal width, while there are fewer, but larger, sepal width measurements. This could be indicative of a species with a more uniform sepal width among most individuals, but a few exceptions with significantly wider sepal widths.

11.

The histogram of 'PetalLengthCm' indicates a right-skewed distribution. This means that the majority of the petal lengths are clustered towards the shorter end, while there is a long tail extending towards longer petal lengths. This suggests that most flowers in this dataset have shorter petals, but a few flowers have unusually long petals. This could be due to various factors such as genetic variation, environmental conditions, or measurement errors. It is essential to be aware of this skewness when performing statistical analysis on this data, as it could potentially impact the results.

12.

The histogram of 'PetalWidthCm' indicates a skewed distribution, leaning towards smaller values. This suggests that the majority of iris flowers in this dataset have relatively narrow petals compared to a smaller number of flowers with wider petals. This skewed distribution may be due to a variety of factors such as genetic variations, environmental conditions, or measurement errors, but further investigation would be required to confirm these hypotheses.

13.

The boxplot of 'SepalLengthCm' shows that the majority of the data points fall between 5.1 and 6.5 centimeters. This indicates that the length of sepal for most of the iris flowers is between these two values. There are some outliers present, as there are data points that are significantly higher than 6.5 centimeters, suggesting that there are a few iris flowers with unusually long sepals.

14.

The boxplot of 'SepalWidthCm' indicates that most of the data points, which are the middle 50% of the values, fall between 2.8 and 3.3 centimeters. This suggests that the majority of the sepal widths for these flowers are fairly consistent and concentrated in this range. However, there are some outliers present, as there are values that extend beyond the upper and lower limits of the boxplot, indicating that a few sepal widths are significantly larger or smaller than the typical range observed.

15.

The boxplot of 'PetalLengthCm' indicates that most of the petal lengths are between 1.6 and 5.15 centimeters. Approximately 50% of the petal lengths fall within this range. There are some outliers present, as there are a few petal lengths that are significantly larger than the majority, reaching up to about 7 centimeters. This suggests that while the majority of irises have relatively similar petal lengths, there are a few exceptions with unusually long petal lengths.

16.

The petal width of the iris flowers varies significantly, as shown in the boxplot. Most of the petal widths fall between 0.4 and 1.9 centimeters, with 50% of the observations falling within this range. The data also suggests that there are outliers present, as there are some petal widths that are significantly larger than the rest, exceeding 2.5 centimeters. These large petal widths are not common but do occur in some instances among the iris flowers.

17.

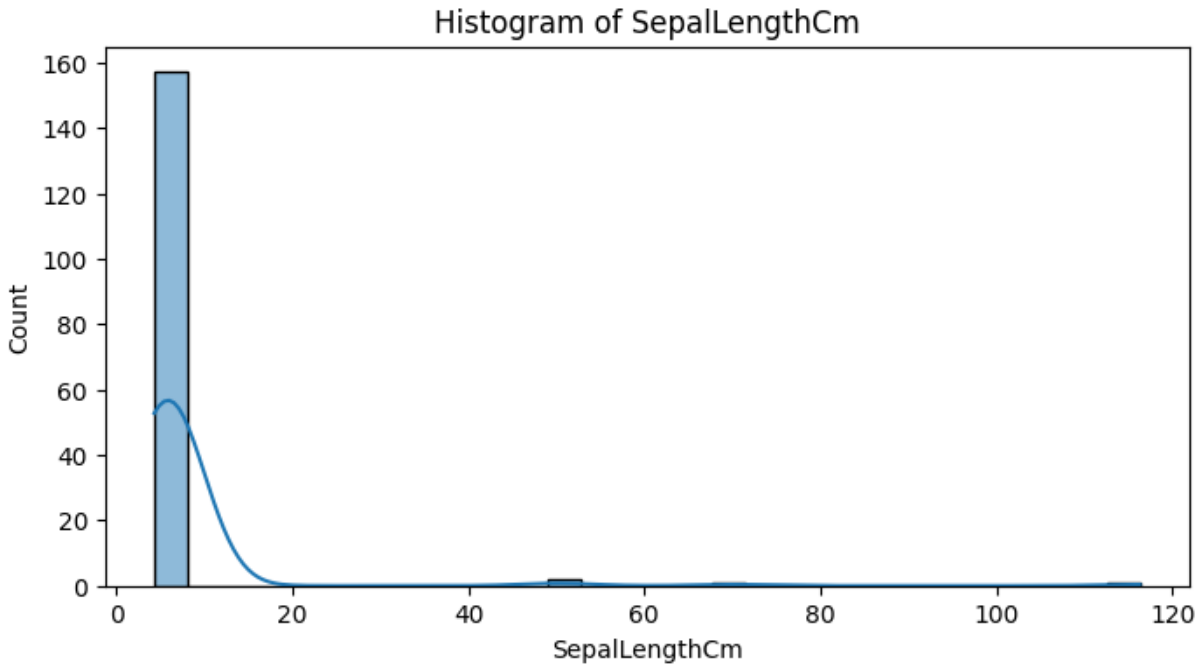
The majority of the data in this count plot represents Iris-virginica, making up 63 instances, which suggests that this species is the most common among the iris flowers in the dataset. This could indicate a higher population or prevalence of Iris-virginica in the area where these iris flowers were collected. However, it's important to remember that the count plot only provides a simple overview and further analysis would be needed to draw more definitive conclusions.

5. Classification Results

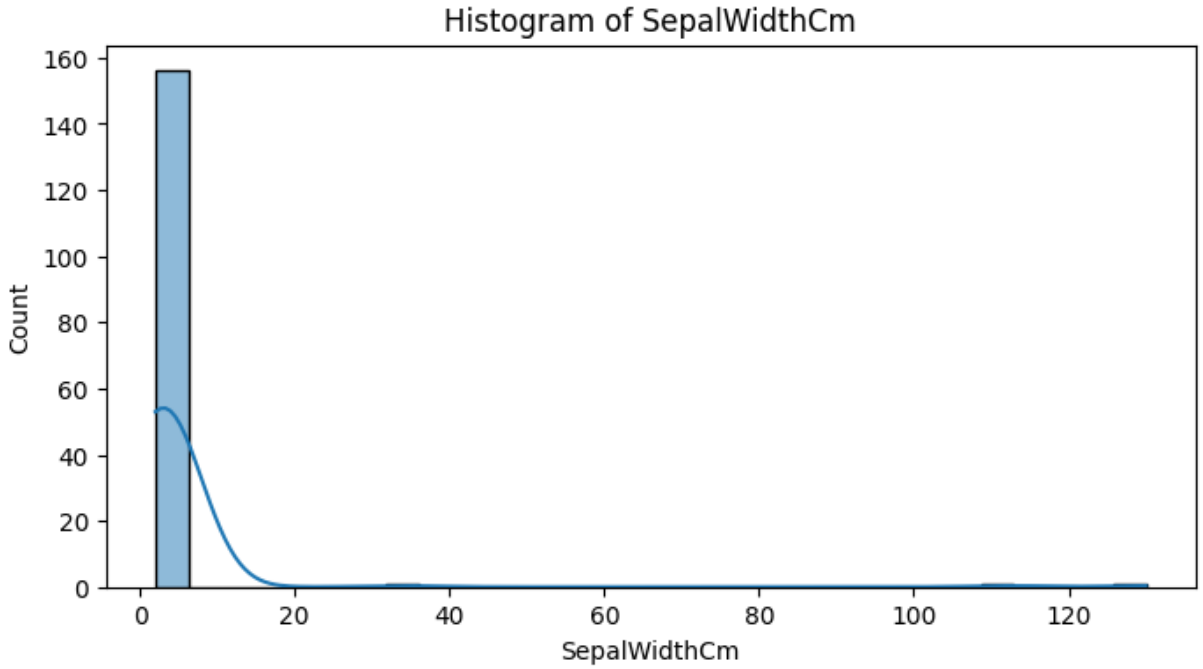
	precision	recall	f1-score	support
Iris-setosa	1.0	1.0	1.0	46.0
Iris-versicolor	1.0	1.0	1.0	46.0
Iris-virginica	1.0	1.0	1.0	55.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	147.0
weighted avg	1.0	1.0	1.0	147.0

6. Visualizations

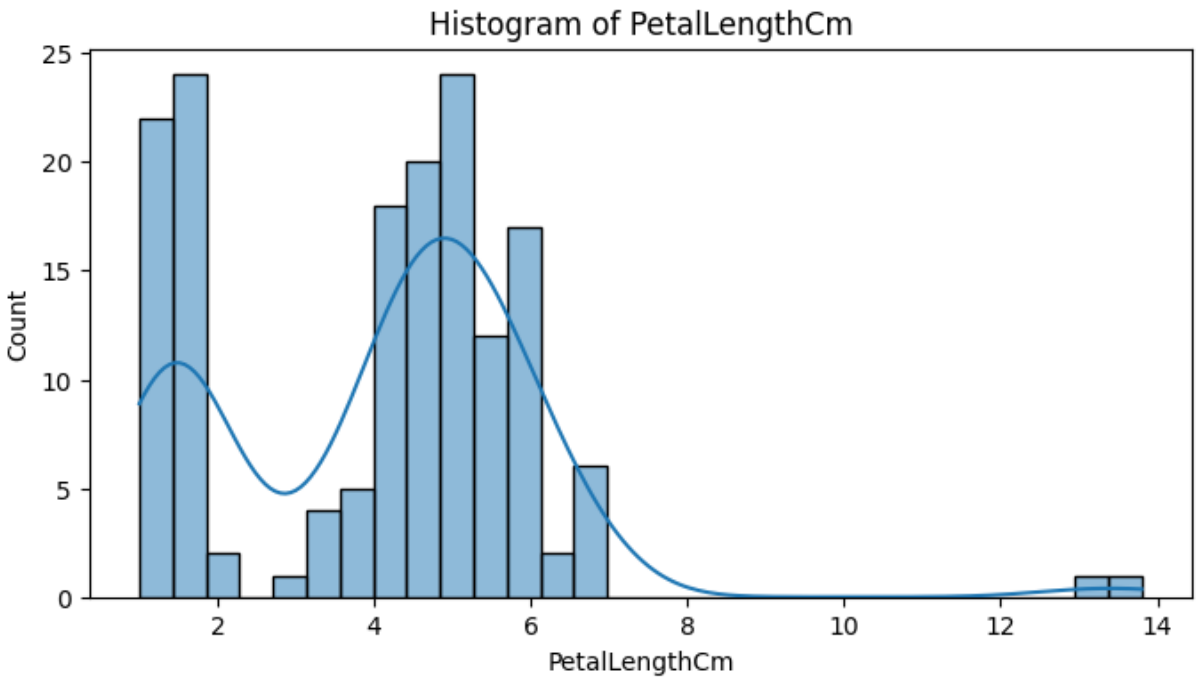
Visualization 1



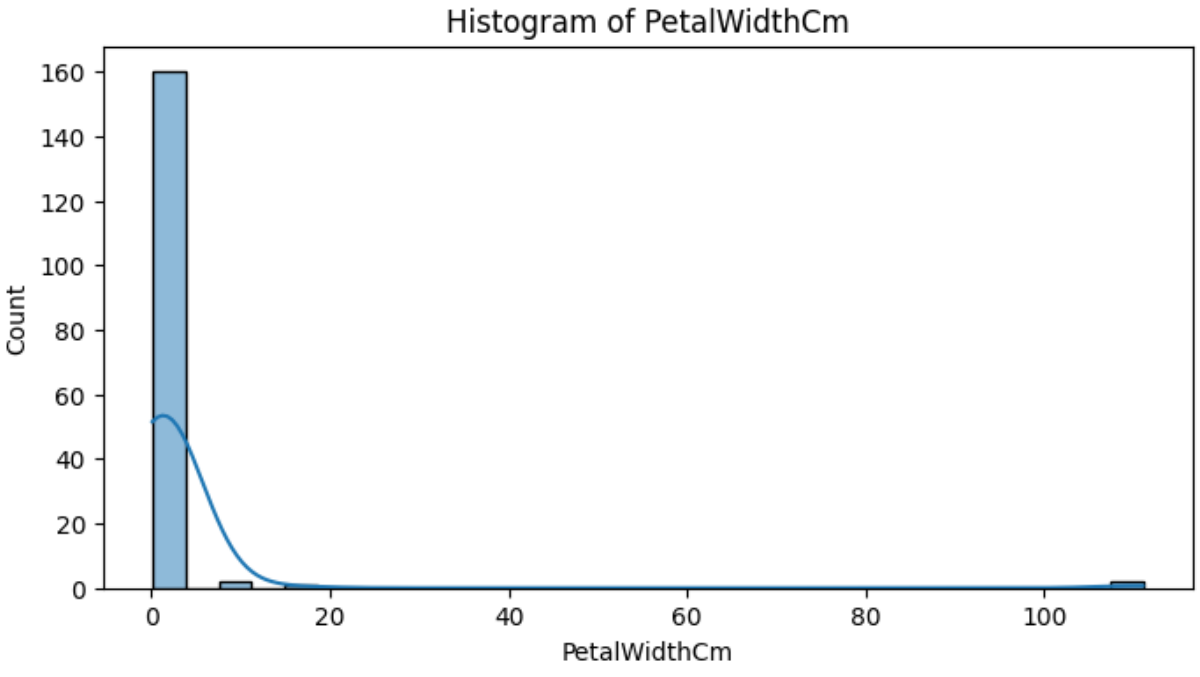
Visualization 2



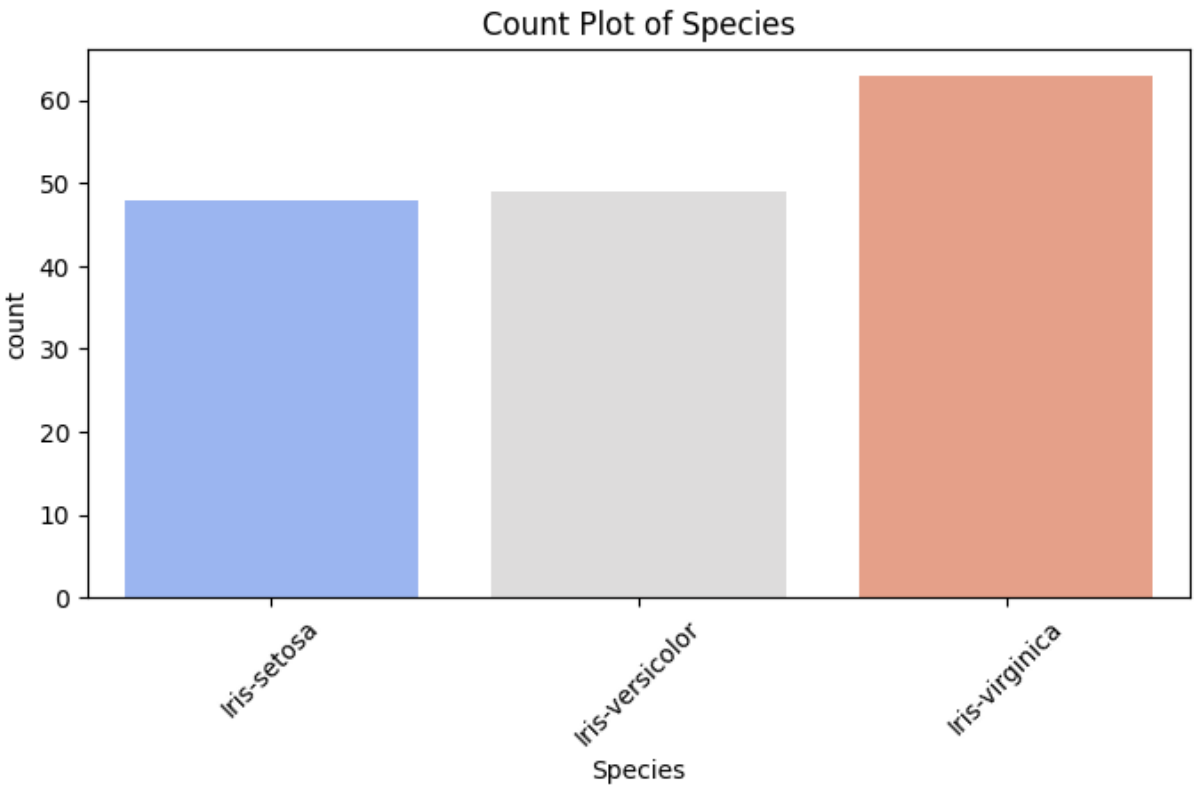
Visualization 3



Visualization 4



Visualization 5



Visualization 6