Submission Format: Please Submit one **ZIP FILE** that contains:

* **\*.py** files with necessary code and
* **\*.docx** file with IDs and students names, results and explanations.

The name of the ZIP FILE should be: lab<#>\_<IDnumber1>\_<IDnumber2>, where # is the lab number.

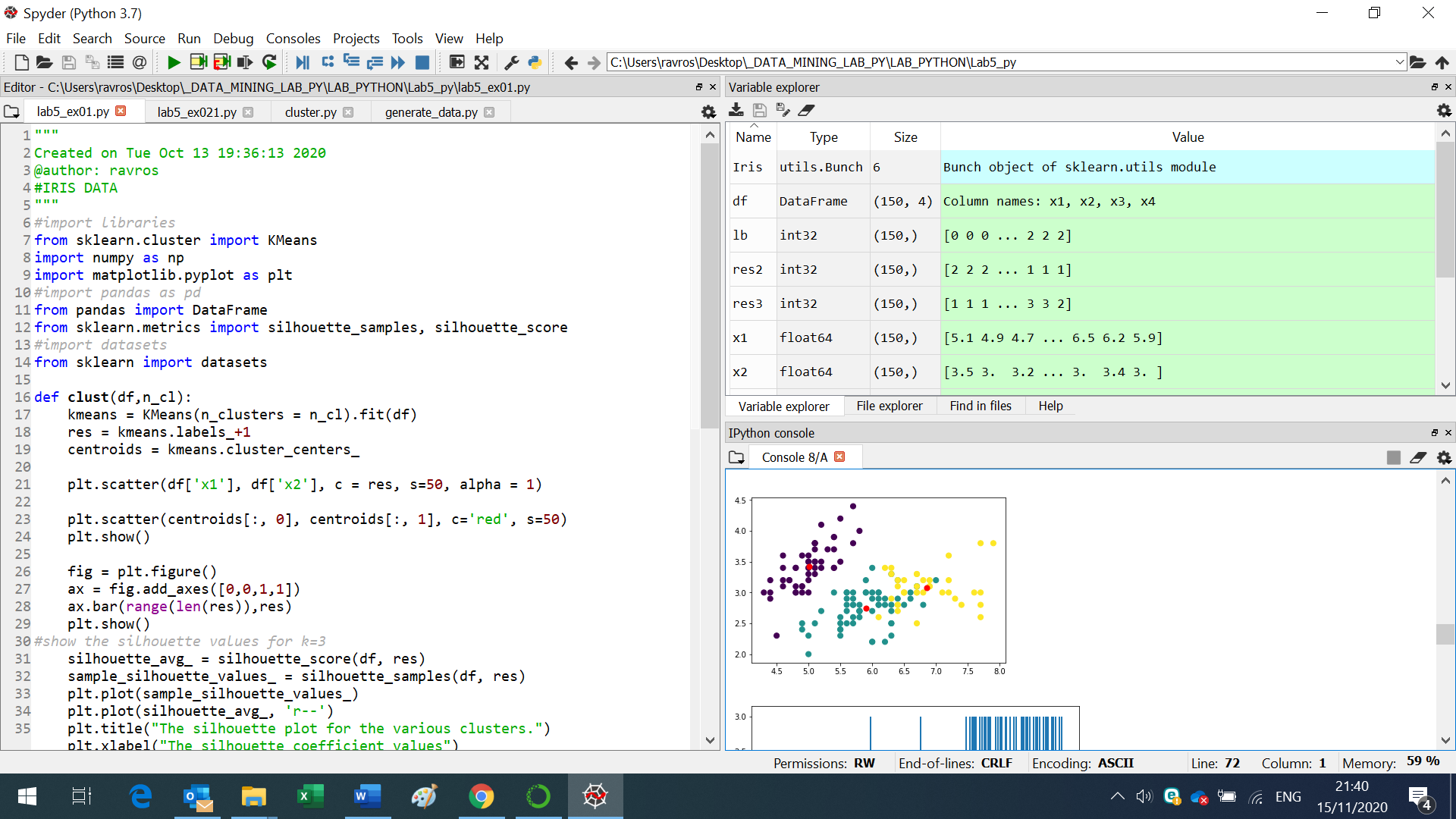
**Lab 5: Visualization of Clustering Data**

**Tasks to do:**

1. Open the file EXAM\_task1 and read the task solution.
2. Open the file lab5\_ex01.py and read the code.
3. Load the data Iris Dataset # Iris = datasets.load\_iris()

*Iris dataset consists of 4 measurements on the sepal length, sepal width, petal length, and petal width for 150 iris specimens. There are 50 specimens from each of three species.*

1. Cluster the Iris Database into 3 clusters using k-means.
2. Show the results of clustering on the plot as in the figure below: each cluster sign by different color with selected centroid.
3. Compare the clustering results with the true labeling.

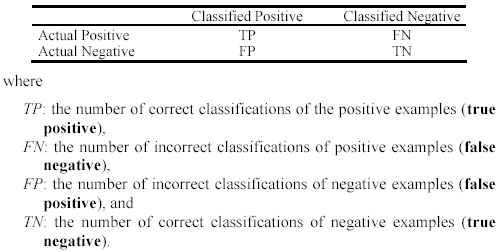


**Individual work:**

1. **Select from the Iris Data 100 points from 2 species. Cluster these points into 2 clusters.**
2. **Show the results of the clustering and the silhouette values on plots. Compare the clustering results with the true labeling.** **(TP, FN, FP and TN).**
3. **Calculate the True Positive rate, False Positive Rate, Accuracy and Precision.**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Open the file lab5\_02.py and read the code. Simulate 2 sets of points (in according with different values of cluster\_std, random\_state).**
2. **Cluster the received datasets into 2 clusters using k-means. Show the clustering results.**
3. **Calculate the True Positive rate, False Positive Rate, Accuracy and Precision.**



**True Positive Rate (TPR) or Sensitivity TPR= TP / (TP + FN)**

**False Positive Rate FPR = 1 - (TN / (TN + FP))**

**Accuracy = (TP + TN) / (TP + TN + FP + FN)**

**Precision = TP / (TP + FP)**