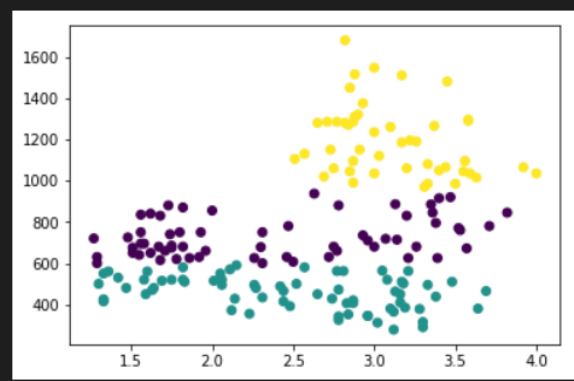


```
xs=wine.loc[:, 'OD280']  
ys=wine.loc[:, 'Proline']  
plt.scatter(xs,ys,c=labels)
```

[19] Python

<matplotlib.collections.PathCollection at 0x7f053b953828>



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

sp500\_companies.csv | unsupervised-learning.ipynb | comcast-telecom-consumer-complaints.ipynb

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M\*Clustering for dataset exploration > wine.var(axis=0)

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

labeled 2, has all the three class of Wines in it.

```
wine.var(axis=0)
```

[20] Python

Alcohol	0.659062
Malic_acid	1.248015
Ash	0.075265
Alcalinity_of_ash	11.152686
Magnesium	203.989335
Total_phenols	0.391690
Flavanoids	0.997719
Nonflavanoid_phenols	0.015489
Proanthocyanins	0.327595
Color_intensity	5.374449
Hue	0.052245
OD280	0.504086
Proline	99166.717355

dtype: float64

PROBLEMS 3 | OUTPUT | DEBUG CONSOLE | TERMINAL | JUPYTER

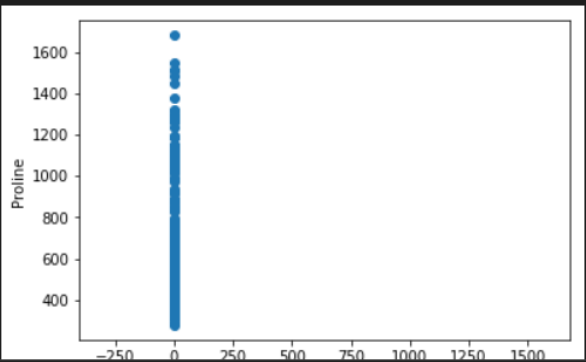
pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```
# Variance comparison between Proline and OD280
plt.scatter(wine['OD280'],wine['Proline'])
plt.xlim(-400,max(wine['Proline']))
plt.xlabel('OD280')
plt.ylabel('Proline')
plt.show()
```

[21]

Python

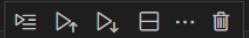


[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M Clustering for dataset exploration > from sklearn.preprocessing import StandardScaler

+ Code + Markdown | Run All Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)



```

from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline
scaler=StandardScaler()
kmeans=KMeans(n_clusters=3)
pipeline=make_pipeline(scaler,kmeans)
pipeline.fit(wine)
labels=pipeline.predict(wine)

```

[22] Python

```

df=pd.DataFrame({'labels':labels,'class':wine_class })
ct=pd.crosstab(df['labels'],df['class'])
ct
# After scaling we get tight clusters

```

[23] Python

```

...
class 1 2 3
labels

```

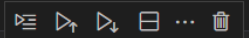
PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M Clustering for dataset exploration > df=pd.DataFrame({'labels':labels,'class':wine\_class })

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)



```
df=pd.DataFrame({'labels':labels,'class':wine_class })
ct=pd.crosstab(df['labels'],df['class'])
ct
# After scaling we get tight clusters
```

[23] Python

class	1	2	3
0	0	65	0
1	59	3	0
2	0	3	48

```
xs=wine.loc[:, 'OD280']
ys=wine.loc[:, 'Proline']
plt.scatter(xs,ys,c=labels)
```

[24] Python

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M Clustering for dataset exploration > xs=wine.loc[:, 'OD280']

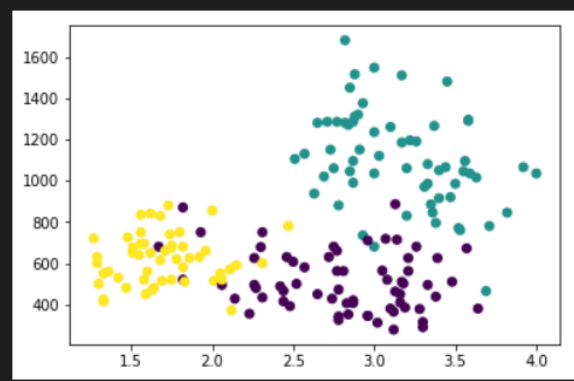
+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

Python

```
xs=wine.loc[:, 'OD280']
ys=wine.loc[:, 'Proline']
plt.scatter(xs,ys,c=labels)
```

[24] Python
<matplotlib.collections.PathCollection at 0x7f0537f9fcc0>



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C:\Users\Nadeem Farooqui\Downloads> unsupervised-learning.ipynb > M\*Clustering for dataset exploration > M\*Scaling without pipeline

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

### Scaling without pipeline

```

scaled_wine=scaler.fit_transform(wine)
scaled_wine.var(axis=0)

```

[25]

Python

... array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.])

```

scaled_wine=pd.DataFrame(scaled_wine)
scaled_wine.columns=['Alcohol','Malic_acid','Ash','Alcalinity_of_ash','Magnesium','Total_phenols','Flavanoids','Nonflavanoid_phenols','Proanthocyanins','Color_intensity','Hue','OD28

```

[26]

Python

```

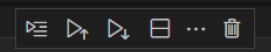
lables=KMeans(n_clusters=3).fit_predict(scaled_wine)
xs=scaled_wine.loc[:, 'OD280']
ys=scaled_wine.loc[:, 'Proline']
plt.scatter(xs,ys,c=lables)

```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

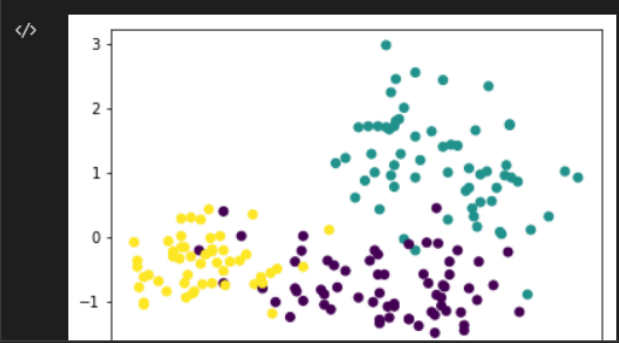
[Error - 11:00:48] Starting client failed  
 Launching server using command: c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



```
lables=KMeans(n_clusters=3).fit_predict(scaled_wine)
xs=scaled_wine.loc[:, 'OD280']
ys=scaled_wine.loc[:, 'Proline']
plt.scatter(xs,ys,c=lables)
```

[27] Python

... <matplotlib.collections.PathCollection at 0x7f0537f072b0>



[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > Visualization with hierarchical clustering and t-SNE > mergings=linkage(seeds,method='complete')

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

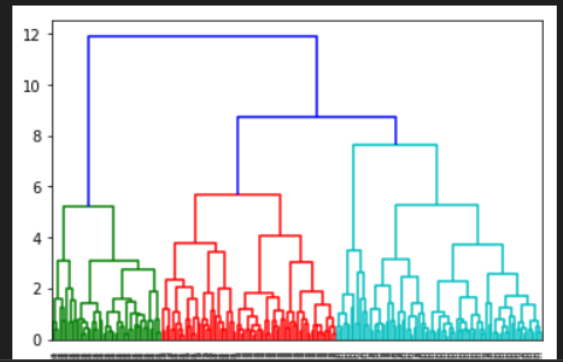
base (Python 3.8.8)

Run, Stop, Refresh, Close, More

```
mergings=linkage(seeds,method='complete')
plt.figure(figsize=(50,50))
dendrogram(mergings,labels=seed_list,leaf_rotation=90,leaf_font_size=6)
plt.show()
```

[29]

Python



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



```
df=pd.DataFrame({'labels':labels,
                 'seed_variety':seed_list})

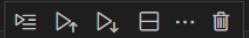
ct=pd.crosstab(df['labels'],df['seed_variety'])
ct
```

[31] Python

seed_variety	Canadian	Kama	Rosa
1	0	0	47
2	0	52	23
3	13	1	0
4	57	17	0

\*\* t-SNE for 2- dimentional maps \*\* t distributed stochastic neighbor embedding Maps samples to 2D or 3D space Map approximately preserves nearness of samples Great for inspecting datasets

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

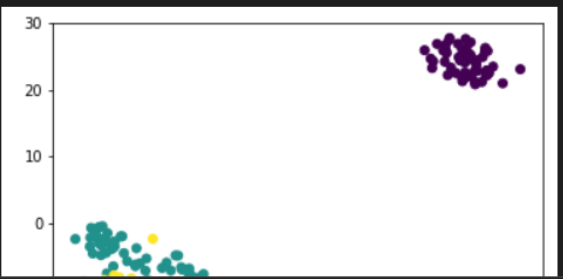


```

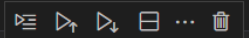
import seaborn as sns
model=TSNE(learning_rate=100)
transformed=model.fit_transform(data)
xs=transformed[:,0]
ys=transformed[:,1]
plt.scatter(xs,ys,c=species)
#plt.legend(species)
#sns.scatterplot(xs,ys,hue=species)
plt.show()

```

[33] Python



[Error - 11:00:48] Starting client failed  
 Launching server using command : c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



```

# TSNE on Seed Data

# Create a TSNE instance: model
model = TSNE(learning_rate=200)

# Apply fit_transform to samples: tsne_features
tsne_features = model.fit_transform(seeds)

# Select the 0th feature: xs
xs = tsne_features[:,0]

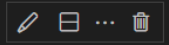
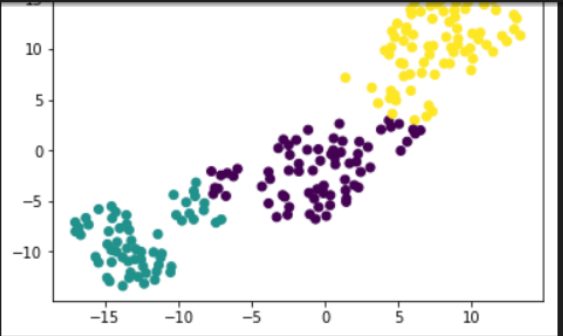
# Select the 1st feature: ys
ys = tsne_features[:,1]

# Scatter plot, coloring by variety_numbers
plt.scatter(xs,ys,c=seed_labels)
plt.show()

# the t-SNE visualization manages to separate the 3 varieties of grain samples.
# Annotate the points
# for x, y, seedlabel in zip(xs, ys, seed_labels):

```

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



## Decorrelating your data and dimension reduction

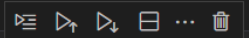
Dimension reduction summarizes a dataset using its common occurring patterns. In this section, you'll learn about the most fundamental of dimension reduction techniques, "Principal Component Analysis" ("PCA"). PCA is often used before supervised learning to improve model performance and generalization. It can also be useful for unsupervised learning. For example, you'll employ a variant of PCA will allow you to cluster Wikipedia articles by their content!

```
[Error - 11:00:48] Starting client failed
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.
```

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M+Decorrelating your data and dimension reduction > # WINE DATA

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)



```

# WINE DATA
# Samples contain two wine features
samples=wine[['Total_phenols','OD280']]
from sklearn.decomposition import PCA
model=PCA()
model.fit(samples)
transformed=model.transform(samples)
print(model.components_)
transformed[:10]

```

[36] Python

```

... [[-0.64116665 -0.76740167]
      [-0.76740167  0.64116665]]

array([[ -1.32771994,  0.45139607],
       [-0.83249607,  0.23309966],
       [-0.75216868, -0.02947892],
       [-1.64026613, -0.65572401],
       [-0.56799228, -0.18335891],
       [-0.80794847, -0.59533103],

```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```

# Seeds width length Data
grains=pd.read_csv('../input/seeds-width-length/seeds-width-vs-length.csv',header=None)

from scipy.stats import pearsonr

width = grains.iloc[:,0]
length = grains.iloc[:,1]

# Scatter plot width vs length
plt.scatter(width, length)
plt.axis('equal')
plt.show()

# Calculate the Pearson correlation
correlation, pvalue = pearsonr(width,length)

# Display the correlation
print(correlation)
'''
As you would expect, the width and length of the grain samples are highly correlated.
Pearson Correlation is 0.86
'''

```

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M+Decorrelating your data and dimension reduction > model = PCA()

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

```

model = PCA()
# Apply the fit_transform method of model to grains: pca_features
pca_features = model.fit_transform(grains)
# Assign 0th column of pca_features: xs
xs = pca_features[:,0]
# Assign 1st column of pca_features: ys
ys = pca_features[:,1]
# Scatter plot xs vs ys
plt.scatter(xs, ys)
plt.axis('equal')
plt.show()

# Calculate the Pearson correlation of xs and ys
correlation, pvalue = pearsonr(xs, ys)
# Display the correlation
print(correlation)

...

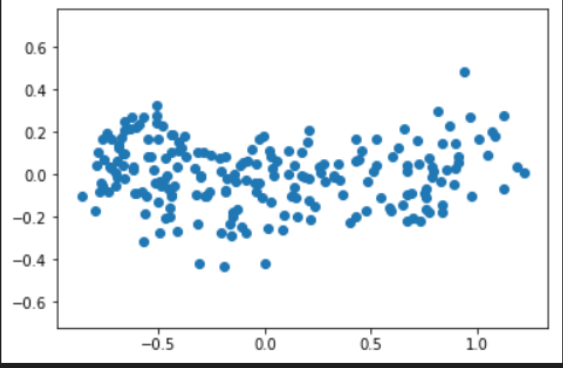
We've successfully decorrelated the grain measurements with PCA!
...

```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

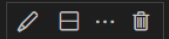
pvs-language-server

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



2.5478751053409354e-17

"\nWe've successfully decorrelated the grain measurements with PCA!\n"



**Intrinsic dimensions** Intrinsic dimension is the number of features needed to approximate the dataset It is the essential idea behind dimension reduction Intrinsic dimension = number of PCA Features with significant variance

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M Decorrelating your data and dimension reduction > # Make a scatter plot of the untransformed points

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline

Python

```

# Make a scatter plot of the untransformed points
plt.scatter(grains.iloc[:,0],grains.iloc[:,1])

model=PCA()
model.fit(grains)
# the coordinates of the mean of the data
mean=model.mean_
# Get the first principal component: first_pc Data varies the most along this direction
first_pc = model.components_[0]
# Plot first_pc as an arrow, starting at mean
plt.arrow(mean[0], mean[1], first_pc[0], first_pc[1], color='red', width=0.01)

# Keep axes on same scale
plt.axis('equal')

```

[39] ... (2.549993369175627, 4.113006630824374, 4.795023399558499, 6.778976600441501)



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > MDecorrelating your data and dimension reduction > # FISH DATA

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline

base (Python 3.8.8)

[40]

```
# FISH DATA
fish=pd.read_csv("../input/fishdata/fish.csv", header=None).drop([0],axis=1)
fish.head()
```

Python

	1	2	3	4	5	6
0	242.0	23.2	25.4	30.0	38.4	13.4
1	290.0	24.0	26.3	31.2	40.0	13.8
2	340.0	23.9	26.5	31.1	39.8	15.1
3	363.0	26.3	29.0	33.5	38.0	13.3
4	430.0	26.5	29.0	34.0	36.6	15.1

[41]

```
fish.var()
```

Python

```
1 145110.683403
2 127.451689
```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M+Decorrelating your data and dimension reduction > fish.var()

+ Code + Markdown | ▶ Run All | ☰ Clear Outputs of All Cells | ☰ Outline ...

base (Python 3.8.8)

⌵ ▶ ⏪ ⏩ ⌵ ... 🗑

fish.var()

[41] Python

```
... 1 145110.683403
2 127.451689
3 148.475328
4 174.168524
5 107.154440
6 4.483703
dtype: float64
```

```
scaler=StandardScaler()
pca=PCA()
pipeline=make_pipeline(scaler,pca)
pipeline.fit(fish)
# Plot the explained variances
features = range(pca.n_components_)
plt.bar(features, pca.explained_variance_)
```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server ⌵ 🔒 📄 ⬆ ⬇

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```

# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load in

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list the files in the input directory

import os
print(os.listdir("../input"))

# Any results you write to the current directory are saved as output.

```

[1] Python ... ['seed-from-uci', 'seeds-width-length', 'fishdata', 'wineuci', 'wikipediavectors']

# Clustering for dataset exploration

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > Clustering for dataset exploration

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

# Clustering for dataset exploration

## Definition

Unsupervised learning is all about finding patterns in the data.  
Ex: Clustering Customers by their purchases , Compressing the data using purchase pattern (dimentionality reduction)

## Difference between Supervised and Unsupervised Learning

Supervised learning finds the pattern for a prediction task , unsupervised learning find patterns but without specific prediction in mind.

## Datasets used Iris

*Empty markdown cell, double click or press enter to edit.*

## K Means Clustering Finds clusters of samples

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

```
[Error - 11:00:48] Starting client failed
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.
```

**K Means Clustering** Finds clusters of samples  
Number of clusters must be specified in advance.

New samples can be assigned to existing cluster  
K means remembers the mean of each cluster (Cluster Centroids)  
Finds the nearest centroid for each new sample

```
[2] from sklearn.datasets import load_iris
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

Python

```
iris=load_iris()
samples=iris.data
model=KMeans(n_clusters=3)
model.fit(samples)
labels=model.predict(samples)
```

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



sp500\_companies.csv x unsupervised-learning.ipynb x comcast-telecom-consumer-complaints.ipynb

C:\Users\Nadeem Farooqui\Downloads\unsupervised-learning.ipynb > M Clustering for dataset exploration > iris=load\_iris()

Code Markdown Run All Clear Outputs of All Cells Outline

```
iris=load_iris()
samples=iris.data
model=KMeans(n_clusters=3)
model.fit(samples)
labels=model.predict(samples)
print(labels)
```

[3]

Python

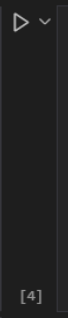
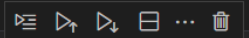
```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 1 2 2 2
2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2
2 1]
```

```
# plotting sepal length and petal length
xs=samples[:,0]
ys=samples[:,2]
plt.scatter(xs,ys,c=labels)
```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

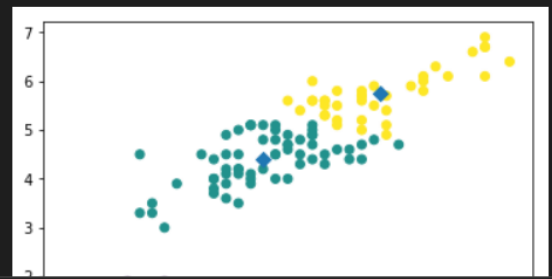
pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command : c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nlDSL-mac-0.5.0\PythonInterpreter\Windows\python failed.

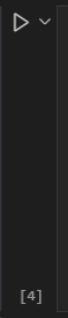
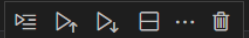


```
# plotting sepal length and petal length
xs=samples[:,0]
ys=samples[:,2]
plt.scatter(xs,ys,c=labels)
centroids=model.cluster_centers_
centroids_x=centroids[:,0]
centroids_y=centroids[:,2]
plt.scatter(centroids_x,centroids_y,marker='D',s=50)
plt.show()
```

Python

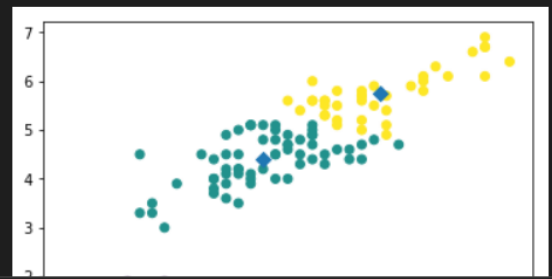


[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



```
# plotting sepal length and petal length
xs=samples[:,0]
ys=samples[:,2]
plt.scatter(xs,ys,c=labels)
centroids=model.cluster_centers_
centroids_x=centroids[:,0]
centroids_y=centroids[:,2]
plt.scatter(centroids_x,centroids_y,marker='D',s=50)
plt.show()
```

Python

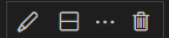


[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M\*Clustering for dataset exploration > M\*Analysing seeds data from UCI

+ Code + Markdown | ▶ Run All | ☰ Clear Outputs of All Cells | ☰ Outline ...

base (Python 3.8.8)



### Analysing seeds data from UCI

```
seeds=pd.read_csv('../input/seed-from-uci/Seed_Data.csv')
seeds.head()
```

[5]

Python

	A	P	C	LK	WK	A_Coef	LKG	target
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220	0
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956	0
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825	0
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805	0
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175	0

```
seed_dict={
    0: 'Kama',
    1: 'Rosa',
```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```
seed_dict={
  0:'Kama',
  1:'Rosa',
  2:'Canadian'
}
seed_list=seeds['target'].map(seed_dict).tolist()
```

[6] Python

```
# Preparing seeds for clustering by dropping the target column
seeds=seeds.drop(['target'],axis=1)
seeds.head()
```

[7] Python

	A	P	C	LK	WK	A_Coef	LKG
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```
# Preparing seeds for clustering by dropping the target column

seeds=seeds.drop(['target'],axis=1)
seeds.head()
```

[7] Python

	A	P	C	LK	WK	A_Coef	LKG
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175

```
ks=range(1,6)
inertias=[]

for k in ks:
```

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

0 3 Jupyter Server: local Cell 12 of 76

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M Clustering for dataset exploration > # Preparing seeds for clustering by dropping the target column

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

Python

```
# Preparing seeds for clustering by dropping the target column

seeds=seeds.drop(['target'],axis=1)
seeds.head()
```

	A	P	C	LK	WK	A_Coef	LKG
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175

```
ks=range(1,6)
inertias=[]

for k in ks:
```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M\*Clustering for dataset exploration > ks=range(1,6)

+ Code + Markdown | ▶ Run All | ☰ Clear Outputs of All Cells | ☰ Outline ...

base (Python 3.8.8)

⌵ ▶ ⏪ ⏩ ⌵ ... 🗑️

```

ks=range(1,6)
inertias=[]

for k in ks:
    model=KMeans(n_clusters=k)
    model.fit(seeds)
    inertias.append(model.inertia_)

plt.plot(ks,inertias,'-o')
plt.xlabel('Number of Clusters , k')
plt.ylabel('Inertia')
plt.xticks(ks)
plt.show()

# Inertia decreases from 3 to 4 very slowly , so 3 can be a good choice

```

[8] Python



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

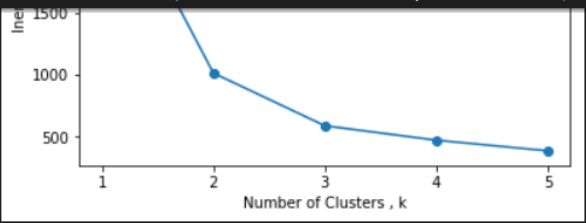
[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M+Clustering for dataset exploration > seeds.head()

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)



Run, Stop, Refresh, Close, Delete icons

seeds.head()

[9] Python

	A	P	C	LK	WK	A_Coef	LKG
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C:\Users\Nadeem Farooqui\Downloads> cd unsupervised-learning.ipynb > M+Clustering for dataset exploration > model=KMeans(n\_clusters=3)

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

```

model=KMeans(n_clusters=3)
seed_labels=model.fit_predict(seeds)
centroids=model.cluster_centers_
centroids

```

```

[10]
... array([[14.64847222, 14.46041667,  0.87916667,  5.56377778,  3.27790278,
           2.64893333,  5.19231944],
          [18.72180328, 16.29737705,  0.88508689,  6.20893443,  3.72267213,
           3.60359016,  6.06609836],
          [11.96441558, 13.27480519,  0.8522    ,  5.22928571,  2.87292208,
           4.75974026,  5.08851948]])

```

```

# A vs LKG
xs_A=seeds.iloc[:,0]
ys_LKG=seeds.iloc[:,-1]

```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

```

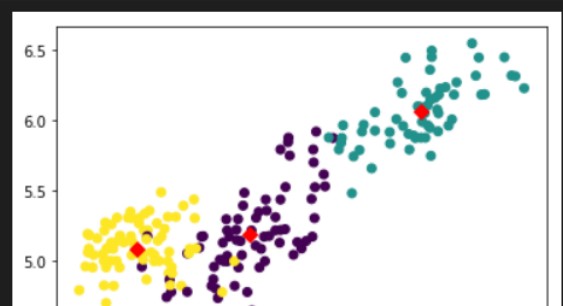
[Error - 11:00:48] Starting client failed
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```

```
# A vs LKG
xs_A=seeds.iloc[:,0]
ys_LKG=seeds.iloc[:,-1]
centroids_xs_A=centroids[:,0]
centroids_ys_LKG=centroids[:,-1]
plt.scatter(xs_A,ys_LKG,c=seed_labels)
plt.scatter(centroids_xs_A,centroids_ys_LKG,marker='D',s=50,c='red')
plt.show()
```

[11]

Python

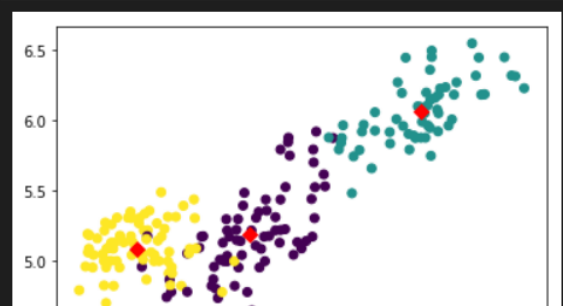


[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

```
# A vs LKG
xs_A=seeds.iloc[:,0]
ys_LKG=seeds.iloc[:,-1]
centroids_xs_A=centroids[:,0]
centroids_ys_LKG=centroids[:,-1]
plt.scatter(xs_A,ys_LKG,c=seed_labels)
plt.scatter(centroids_xs_A,centroids_ys_LKG,marker='D',s=50,c='red')
plt.show()
```

[11]

Python



[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M+ Clustering for dataset exploration > # P vs C

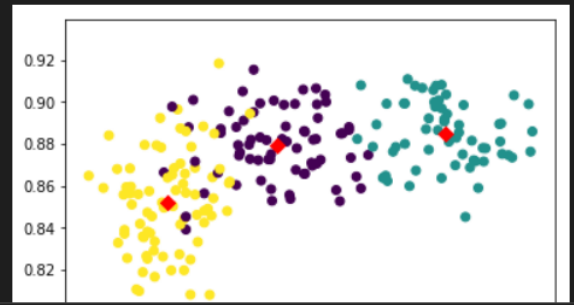
+ Code + Markdown | ▶ Run All | ☰ Clear Outputs of All Cells | ☰ Outline ...

base (Python 3.8.8)

⌵ ▶ ⏪ ⏩ ⏹ ... 🗑

```
# P vs C
xs_A=seeds.iloc[:,1]
ys_LKG=seeds.iloc[:,2]
centroids_xs_A=centroids[:,1]
centroids_ys_LKG=centroids[:,2]
plt.scatter(xs_A,ys_LKG,c=seed_labels)
plt.scatter(centroids_xs_A,centroids_ys_LKG,marker='D',s=50,c='red')
plt.show()
```

[12] Python



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M Clustering for dataset exploration > # WK vs A\_Coeff

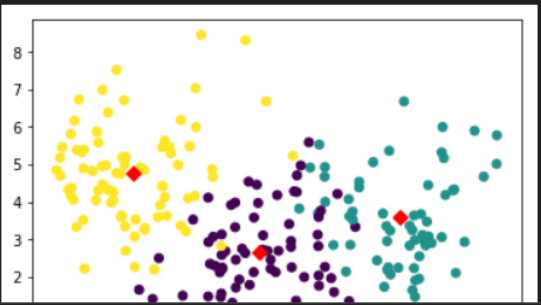
+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

⌵ ⏪ ⏩ ⏹ ... 🗑

```
# WK vs A_Coeff
xs_A=seeds.iloc[:,4]
ys_LKG=seeds.iloc[:,5]
centroids_xs_A=centroids[:,4]
centroids_ys_LKG=centroids[:,5]
plt.scatter(xs_A,ys_LKG,c=seed_labels)
plt.scatter(centroids_xs_A,centroids_ys_LKG,marker='D',s=50,c='red')
plt.show()
```

[13] Python



PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server ⌵ 🔒 📄 ⏪ ⏩

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

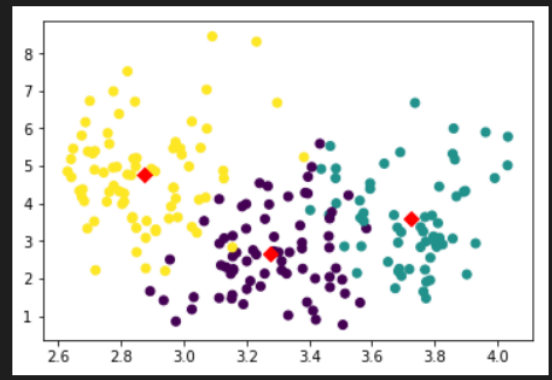
C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M+Clustering for dataset exploration > M+Transforming Features for Better Clustering

+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)

```
plt.scatter(xs_A,ys_LKG,c=seed_labels)  
plt.scatter(centroids_xs_A,centroids_ys_LKG,marker='D',s=50,c='red')  
plt.show()
```

[13] Python



### Transforming Features for Better Clustering

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

pvs-language-server

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

### Piedmont wine Dataset

```
wine=pd.read_csv('../input/wineuci/Wine.csv',header=None)
wine.head()
```

[14]

Python

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735

```
wine.columns=['Class','Alcohol','Malic_acid','Ash','Alcalinity_of_ash','Magnesium','Total_phenols','Flavanoids','Nonflavanoid_phenols','Proanthocyanins','Color_intensity','Hue','OD2
```

[15]

Python

[Error - 11:00:48] Starting client failed  
Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.



C:\Users\Nadeem Farooqui\Downloads> unsupervised-learning.ipynb > M Clustering for dataset exploration > wine\_class=wine['Class']

Code | Markdown | Run All | Clear Outputs of All Cells | Outline

```
wine.columns=['Class','Alcohol','Malic_acid','Ash','Alcalinity_of_ash','Magnesium','Total_phenols','Flavanoids','Nonflavanoid_phenols','Proanthocyanins','Color_intensity','Hue','OD280','Proline']
```

```
wine.head()
```

	Class	Alcohol	Malic_acid	Ash	Alcalinity_of_ash	Magnesium	Total_phenols	Flavanoids	Nonflavanoid_phenols	Proanthocyanins	Color_intensity	Hue	OD280	Proline
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735

```
wine_class=wine['Class']
wine.drop('Class',axis=1,inplace=True)
```

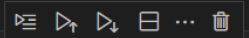
PROBLEMS 3 | OUTPUT | DEBUG CONSOLE | TERMINAL | JUPYTER

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.

C: > Users > Nadeem Farooqui > Downloads > unsupervised-learning.ipynb > M\*Clustering for dataset exploration > model=KMeans(n\_clusters=3)

+ Code + Markdown | Run All Clear Outputs of All Cells | Outline ...

base (Python 3.8.8)



```

model=KMeans(n_clusters=3)
labels=model.fit_predict(wine)

df=pd.DataFrame({'labels':labels , 'class':wine_class})
ct=pd.crosstab(df['labels'],df['class'])
ct

```

[18] Python

class	1	2	3
0	13	20	29
1	0	50	19
2	46	1	0

```

xs=wine.loc[:, 'OD280']
ys=wine.loc[:, 'Proline']
plt.scatter(xs,ys,c=labels)

```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER pvs-language-server

[Error - 11:00:48] Starting client failed  
 Launching server using command c:\Users\Nadeem Farooqui\.vscode\extensions\pvs-ifi-heidelberg-university-germany.vscode-nldsl-mac-0.5.0\PythonInterpreter\Windows\python failed.