

```
[2]: #restart the kernel after installation
!pip install numpy pandas-profiling matplotlib seaborn --quiet

[5]: !pip install jovian opendatasets graphviz scikit-learn plotly --upgrade --quiet

[6]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('darkgrid')
%matplotlib inline
```

IRIS FLOWER DATASET:

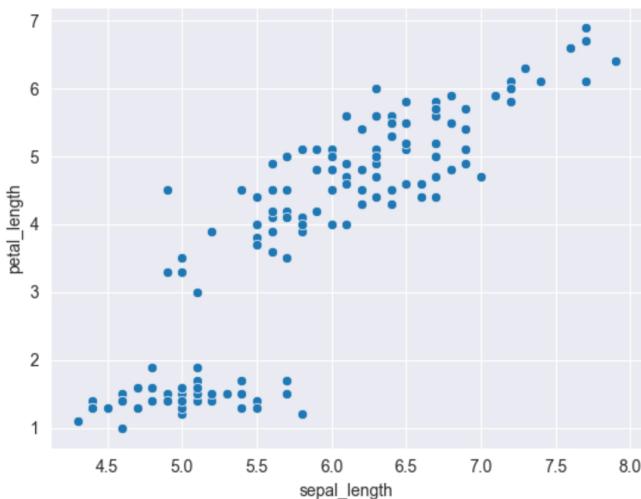
```
[7]: iris_df = sns.load_dataset('iris')
```

```
[8]: iris_df
```

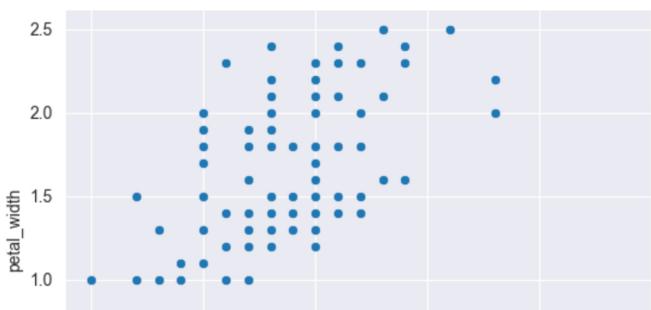
```
[8]:   sepal_length  sepal_width  petal_length  petal_width  species
  0          5.1         3.5         1.4         0.2    setosa
  1          4.9         3.0         1.4         0.2    setosa
  2          4.7         3.2         1.3         0.2    setosa
  3          4.6         3.1         1.5         0.2    setosa
  4          5.0         3.6         1.4         0.2    setosa
  ...
  145        6.7         3.0         5.2         2.3  virginica
  146        6.3         2.5         5.0         1.9  virginica
  147        6.5         3.0         5.2         2.0  virginica
  148        6.2         3.4         5.4         2.3  virginica
  149        5.9         3.0         5.1         1.8  virginica
```

150 rows × 5 columns

```
[14]: sns.scatterplot(data = iris_df, x = 'sepal_length', y = 'petal_length');
```



```
[13]: sns.scatterplot(data = iris_df, x = 'sepal_width', y = 'petal_width');
```





```
[15]: numeric_cols = ["sepal_length", "sepal_width", "petal_length", "petal_width"]
```

```
[16]: X = iris_df[numeric_cols]
```

[17]: X

[17]:	sepal_length	sepal_width	petal_length	petal_width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

KMEANS CLUSTERING

```
[18]: from sklearn.cluster import KMeans
```

```
[19]: model = KMeans(n_clusters = 3, random_state = 42)
```

```
[20]: model.fit(X)
```

```
[20]: KMeans
```

```
[21]: model.cluster_centers
```

```
[21]: array([[6.85384615, 3.07692308, 5.71538462, 2.05384615],  
           [5.006     , 3.428     , 1.462     , 0.246    ],  
           [5.88360656, 2.74098361, 4.38852459, 1.43442623]])
```

[22]: X

[22]:	sepal_length	sepal_width	petal_length	petal_width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

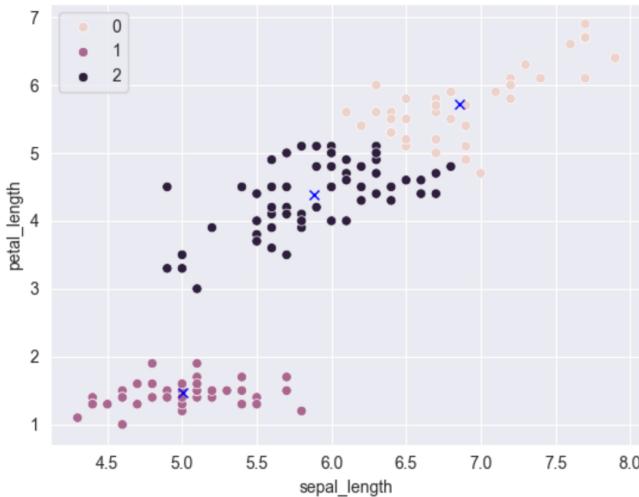
150 rows × 4 columns

```
[23]: preds = model.predict(X)  
preds
```

```
0, 0, 0, 2, 2, 0, 0, 0, 0, 2, 0, 2, 0, 2, 0, 0, 2, 2, 0, 0, 0, 0, 0,  
0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 2]
```

```
[24]: sns.scatterplot(data = X, x = 'sepal_length', y = 'petal_length', hue = preds);  
centers_x,centers_y = model.cluster_centers_[:,0], model.cluster_centers_[:,2]  
plt.plot(centers_x, centers_y, 'xb')
```

```
[24]: [<matplotlib.lines.Line2D at 0x2579a93f2d0>]
```



```
[25]: model.inertia_
```

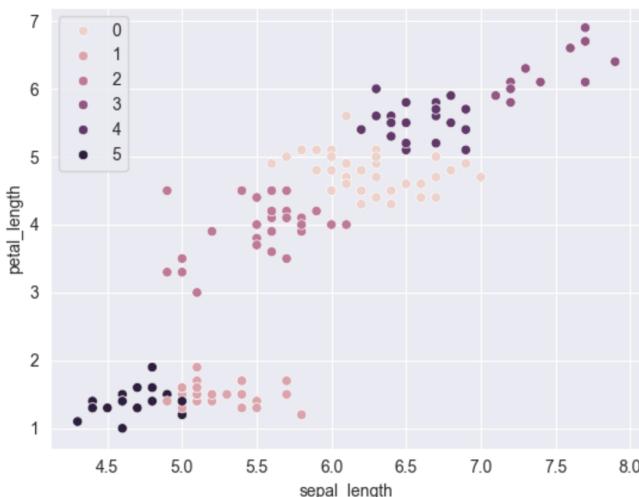
```
[25]: 78.8556658259773
```

```
[26]: model = KMeans(n_clusters = 6, random_state = 42).fit(X)
```

```
[27]: preds = model.predict(X)  
preds
```

```
[27]: array([1, 5, 5, 5, 1, 1, 5, 1, 5, 5, 1, 5, 5, 1, 1, 1, 1, 1, 1, 1,  
5, 1, 5, 5, 1, 1, 1, 5, 5, 1, 1, 1, 5, 5, 1, 1, 5, 1, 1, 5, 5, 1,  
1, 5, 1, 5, 1, 5, 0, 0, 0, 2, 0, 2, 0, 2, 0, 2, 2, 2, 0, 2, 0, 0, 0,  
2, 2, 0, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
2, 2, 2, 0, 2, 2, 2, 2, 0, 2, 2, 4, 0, 3, 4, 4, 3, 2, 3, 4, 3,  
4, 4, 4, 0, 4, 4, 4, 3, 3, 0, 4, 0, 3, 0, 4, 3, 0, 4, 3, 3, 3,  
4, 0, 0, 3, 4, 4, 0, 4, 4, 4, 0, 4, 4, 4, 0, 4, 4, 0])
```

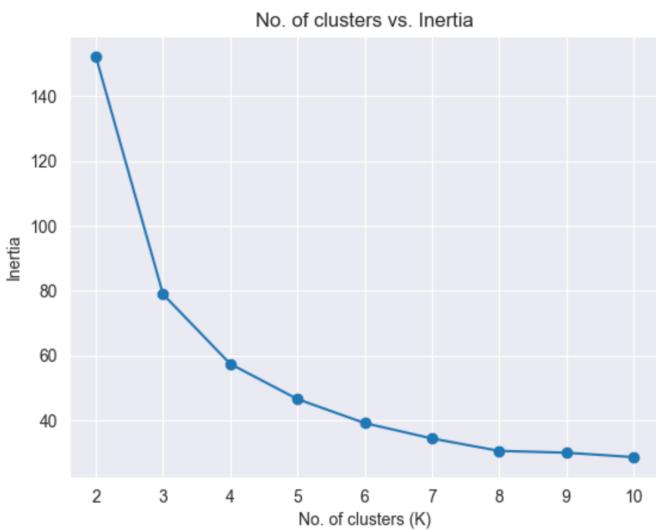
```
[29]: sns.scatterplot(data = X, x = 'sepal_length', y = 'petal_length', hue = preds );
```



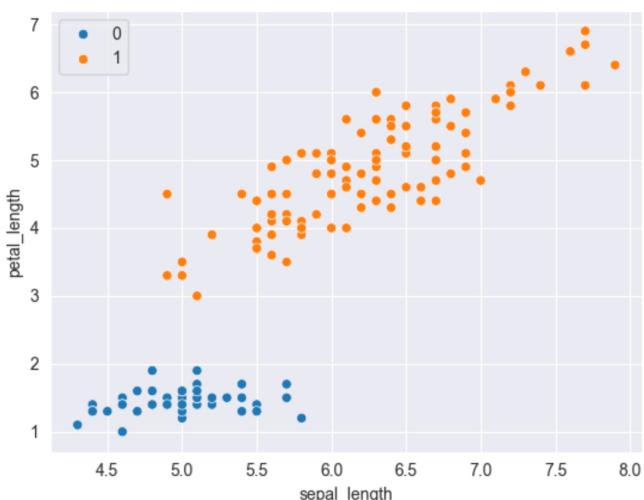
```
[30]: model.inertia_
```

```
[30]: 39.0660353535355
```

```
[31]: options = range(2,11)  
inertias = []  
  
for n_clusters in options:  
    model = KMeans(n_clusters, random_state = 42).fit(X)  
    inertias.append(model.inertia_)  
  
plt.title("No. of clusters vs. Inertia")  
plt.plot(options, inertias, '-o')  
plt.xlabel('No. of clusters (K)')  
plt.ylabel('Inertia');
```



DBSCAN:



```
[38]: iris_ut = sns.load_dataset('iris')
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

```
[39]: numeric_cols
```

```
[39]: ['sepal_length', 'sepal_width', 'petal_length', 'petal_width']
```

PRINCIPAL COMPONENT ANALYSIS (PCA)

```
[40]: from sklearn.decomposition import PCA
```

```
[41]: pca = PCA(n_components = 2)
```

```
[42]: pca.fit(iris_df[numeric_cols])
```

```
[42]: PCA(n_components=2)
```

```
[47]: pca.components_
```

```
[47]: array([[ 0.36138659, -0.08452251,  0.85667061,  0.3582892 ],  
           [ 0.65658877,  0.73016143, -0.17337266, -0.07548102]])
```

```
[48]: transformed = pca.transform(iris_df[numeric_cols])
```

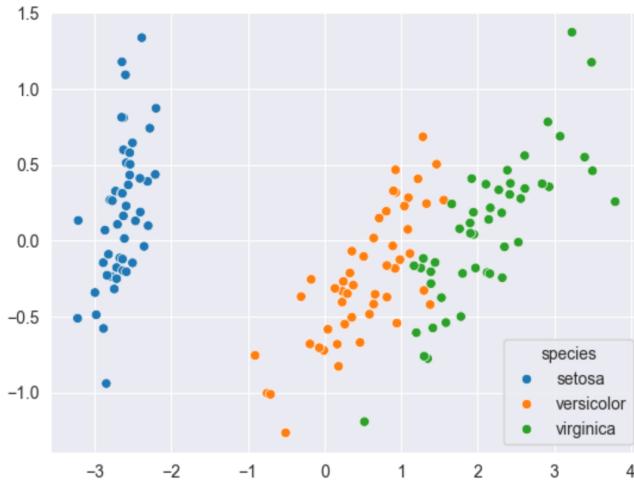
```
[49]: transformed
```

```
[49]: array([[-2.68412563,  0.31939725],  
           [-2.71414169, -0.17700123],  
           [-2.88899057, -0.14494943],  
           [-2.74534286, -0.31829898],  
           [-2.72871654,  0.32675451],  
           [-2.28085963,  0.74133045],  
           [-2.82053775, -0.08946138],  
           [-2.62614497,  0.16338496],  
           [-2.88638273, -0.57831175],  
           [-2.6727558 , -0.11377425],  
           [-2.50694709,  0.6450689 ],  
           [-2.61275523,  0.01472994],  
           [-2.78610927, -0.235112 ],  
           [-3.223380374, -0.51139459],  
           [-2.64475039,  1.17876464],  
           [-2.38603903,  1.33806233],  
           [-2.62352788,  0.81067951],  
           [-2.64829671,  0.31184914],  
           [-2.19982032,  0.87283904],  
           [-2.5879864 ,  0.51356031],  
           [-2.31025622,  0.39134594],  
           [-2.54370523,  0.43299606],  
           [-3.21593942,  0.13346807],  
           [-2.30273318,  0.09870885],  
           [-2.35575405, -0.03728186],  
           [-2.50666891, -0.14601688],  
           [-2.46882007,  0.13095149],  
           [-2.56231991,  0.36771886],  
           [-2.63953472,  0.31203998],  
           [-2.63198939, -0.19696122],  
           [-2.58739848, -0.20431849],  
           [-2.4099325 ,  0.41092426],  
           [-2.64886233,  0.81336382],  
           [-2.59873675,  1.09314576],  
           [-2.63692688, -0.12132235],  
           [-2.86624165,  0.06936447],  
           [-2.62523805,  0.59937002],  
           [-2.80068412,  0.26864374],  
           [-2.98050204, -0.48795834],  
           [-2.5900631,  0.22904384],  
           [-2.77010243,  0.26352753],  
           [-2.84936871, -0.94096057],
```

[-2.99740655, -0.34192606],
[-2.40561449, 0.18887143],
[-2.20948924, 0.43666314],
[-2.71445143, -0.2502082],
[-2.53814826, 0.50377114],
[-2.83946217, -0.22794557],
[-2.54308575, 0.57941002],
[-2.70335978, 0.10770608],
[1.28482569, 0.68516047],
[0.93248853, 0.31833364],
[1.46430232, 0.50426282],
[0.18331772, -0.82795901],
[1.08810326, 0.07459068],
[0.64166908, -0.41824687],
[1.09506066, 0.28346827],
[-0.74912267, -1.00489096],
[1.04413183, 0.2283619],
[-0.0087454 , -0.72308191],
[-0.50784088, -1.26597119],
[0.51169856, -0.10398124],
[0.26497651, -0.55003646],
[0.98493451, -0.12481785],
[-0.17392537, -0.25485421],
[0.92786078, 0.46717949],
[0.66028376, -0.35296967],
[0.23610499, -0.33361077],
[0.94473373, -0.54314555],
[0.04522698, -0.58383438],
[1.11628318, -0.08461685],
[0.35788842, -0.06892503],
[1.29818388, -0.32778731],
[0.92172892, -0.18273779],
[0.71485333, 0.14905594],
[0.90017437, 0.32850447],
[1.33202444, 0.24444088],
[1.55780216, 0.26749545],
[0.81329065, -0.1633503],
[-0.30558378, -0.36826219],
[-0.06812649, -0.70517213],
[-0.18962247, -0.68028676],
[0.13642871, -0.31403244],
[1.38002644, -0.42095429],
[0.58800644, -0.48428742],
[0.80685831, 0.19418231],
[1.22069088, 0.40761959],
[0.81509524, -0.37203706],
[0.24595768, -0.2685244],
[0.16641322, -0.68192672],
[0.46480029, -0.67071154],
[0.8908152 , -0.03446444],
[0.23054802, -0.40438585],
[-0.70453176, -1.01224823],
[0.35698149, -0.50491009],
[0.33193448, -0.21265468],
[0.37621565, -0.29321893],
[0.64257601, 0.01773819],
[-0.90646986, -0.75609337],
[0.29900084, -0.34889781],
[2.53119273, -0.00984911],
[1.41523588, -0.57491635],
[2.61667602, 0.34390315],
[1.97153105, -0.1797279],
[2.35000592, -0.04026095],
[3.39703874, 0.55083667],
[0.52123224, -1.19275873],
[2.93258707, 0.3555],
[2.32122882, -0.2438315],
[2.91675097, 0.78279195],
[1.66177415, 0.24222841],
[1.80340195, -0.21563762],
[2.1655918 , 0.21627559],
[1.34616358, -0.77681835],
[1.58592822, -0.53964071],
[1.90445637, 0.11925069],
[1.94968906, 0.04194326],
[3.48705536, 1.17573933],
[3.79564542, 0.25732297],
[1.30079171, -0.76114964],
[2.42781791, 0.37819601],
[1.19900111, -0.60609153],
[3.49992004, 0.4606741],
[1.38876613, -0.20439933],
[2.2754305 , 0.33499061],
[2.61409047, 0.56090136],
[1.25850816, -0.17970479],
[1.29113206, -0.11666865],
[2.12360872, -0.20972948],
[2.38800302, 0.4646398],
[2.84167278, 0.37526917],
[3.23067366, 1.37416509],
[2.15943764, -0.21727758],
[1.44416124, -0.14341341],
[1.78129481, -0.49990168],
[3.07649993, 0.68808568],
[2.14424331, 0.1400642],
[1.90509815, 0.04930053],
[1.16932634, -0.16499026],
[2.10761114, 0.37228787],
[2.31415471, 0.18365128],

```
[ 1.9222678 ,  0.40920347],
[ 1.41523588, -0.57491635],
[ 2.56301338,  0.2778626 ],
[ 2.41874618,  0.3047982 ],
[ 1.94410979,  0.1875323 ],
[ 1.52716661, -0.37531698],
[ 1.76434572,  0.07885885],
[ 1.90094161,  0.11662796],
[ 1.39018886, -0.28266094]])
```

```
[50]: sns.scatterplot(x=transformed[:,0], y=transformed[:,1], hue=iris_df['species']);
```



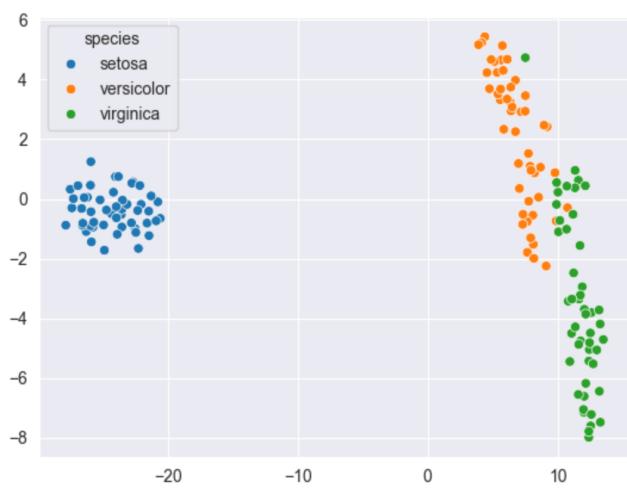
t-Distributed Stochastic Neighbor Embedding (t-SNE):

```
[51]: from sklearn.manifold import TSNE
[52]: tsne = TSNE(n_components = 2)
[54]: transformed = tsne.fit_transform(iris_df[numerical_cols])
[56]: transformed
[56]: array([[-2.35809307e+01, -5.22060156e-01],
       [-2.62917042e+01, -1.09369588e+00],
       [-2.61933346e+01,  5.30571826e-02],
       [-2.66382561e+01, -3.21696430e-01],
       [-2.35450134e+01, -9.41376507e-01],
       [-2.16082954e+01, -8.10961246e-01],
       [-2.59608002e+01,  4.58464950e-01],
       [-2.43470783e+01, -4.96076971e-01],
       [-2.73912773e+01, -3.01681638e-01],
       [-2.57687702e+01, -9.45293069e-01],
       [-2.20654278e+01, -1.82509467e-01],
       [-2.49495239e+01, -8.73362005e-01],
       [-2.65746098e+01, -8.87391508e-01],
       [-2.75171375e+01,  3.22735608e-01],
       [-2.07749718e+01, -9.77787822e-02],
       [-2.05931244e+01, -6.52583599e-01],
       [-2.14729099e+01, -4.12662476e-01],
       [-2.35401649e+01, -3.40562731e-01],
       [-2.12914257e+01,  9.89724994e-02],
       [-2.25217514e+01, -9.96302843e-01],
       [-2.26839333e+01,  5.70334435e-01],
       [-2.28101254e+01, -8.04687738e-01],
       [-2.59318562e+01,  1.24782252e+00],
       [-2.40347214e+01,  7.45504856e-01],
       [-2.48908157e+01, -1.71913588e+00],
       [-2.58891830e+01, -1.44035780e+00],
       [-2.41925945e+01,  2.23819539e-01],
       [-2.31403904e+01, -1.81380153e-01],
       [-2.34945908e+01, -4.12371419e-02],
       [-2.59228287e+01, -4.30750549e-01],
       [-2.59456367e+01, -9.00707841e-01],
       [-2.27738113e+01,  5.28339386e-01],
       [-2.14538536e+01, -1.23106670e+00],
       [-2.09043598e+01, -7.36066878e-01],
       [-2.57058487e+01, -7.81319261e-01],
       [-2.52224865e+01, -3.24066244e-02],
       [-2.21609688e+01,  4.49673384e-01],
       [-2.39009266e+01, -1.18546867e+00],
       [-2.72753696e+01,  8.64682905e-03],
       [-2.39778614e+01, -2.59965837e-01],
       [-2.39615498e+01, -6.34557307e-01],
       [-2.78628788e+01, -8.81407559e-01],
       [-2.69254074e+01,  4.44037199e-01],
       [-2.38071499e+01,  7.48348713e-01],
       [-2.22719498e+01, -1.66061199e+00],
       [-2.65504227e+01, -8.15608799e-01],
       [-2.24669056e+01, -1.12465620e+00],
       [-2.64441872e+01,  4.36418355e-02],
       [-2.22281784e+01, -2.87805161e-01]]
```

L 2.4294001e+01, -3.0e+00],
[-2.47328072e+01, -3.72594863e-01],
[7.67057943e+00, -1.79082298e+00],
[7.78293991e+00, -7.89327249e-02],
[8.17553139e+00, -1.99173152e+00],
[5.65694475e+00, 4.65532684e+00],
[8.09762669e+00, -5.45519173e-01],
[7.18582630e+00, 2.91833282e+00],
[8.53758335e+00, 5.73346242e-02],
[4.09063196e+00, 5.25111008e+00],
[7.66601610e+00, -7.60379612e-01],
[6.12745523e+00, 4.68019819e+00],
[4.39984322e+00, 5.43702555e+00],
[6.75730658e+00, 2.25450516e+00],
[5.74789000e+00, 5.13668919e+00],
[8.21599770e+00, 8.72165978e-01],
[4.75599289e+00, 3.69317007e+00],
[7.31080484e+00, -8.54035378e-01],
[7.50655508e+00, 2.93865752e+00],
[5.58163977e+00, 3.32526064e+00],
[9.19466877e+00, 2.41996694e+00],
[5.35062504e+00, 4.23819065e+00],
[9.79706669e+00, 8.76046538e-01],
[5.85655069e+00, 2.33508992e+00],
[9.88593483e+00, -7.37212539e-01],
[7.88447762e+00, 1.09638894e+00],
[7.06188011e+00, 3.49349171e-01],
[7.33840895e+00, -5.15883148e-01],
[8.12901306e+00, -1.52181315e+00],
[9.11799145e+00, -2.24973226e+00],
[7.75346851e+00, 1.52166224e+00],
[4.56025648e+00, 4.23269272e+00],
[5.12946518e+00, 4.60401773e+00],
[4.88813591e+00, 4.67032290e+00],
[5.41178560e+00, 3.51656890e+00],
[1.07649126e+01, -2.94311285e-01],
[7.52379322e+00, 3.46076846e+00],
[8.68296051e+00, 1.06104612e+00],
[7.92673540e+00, -1.30351341e+00],
[8.95549011e+00, 2.46904421e+00],
[6.35707092e+00, 3.24137592e+00],
[5.81201315e+00, 4.30562115e+00],
[6.75271559e+00, 3.97774482e+00],
[7.95118666e+00, 9.59324777e-01],
[5.61352873e+00, 3.68298769e+00],
[4.14842415e+00, 5.23422527e+00],
[6.394744440e+00, 3.74552751e+00],
[6.41322899e+00, 2.94826150e+00],
[6.49610233e+00, 3.08218336e+00],
[6.99272688e+00, 1.18641365e+00],
[3.91202879e+00, 5.16871214e+00],
[6.11202860e+00, 3.34574914e+00],
[1.35206690e+01, -4.71305466e+00],
[1.13344374e+01, 3.73205036e-01],
[1.21688089e+01, -6.18319464e+00],
[1.18957453e+01, -2.94585228e+00],
[1.25215912e+01, -4.49549818e+00],
[1.25458508e+01, -7.60910845e+00],
[7.51970863e+00, 4.73458862e+00],
[1.20268364e+01, -7.15669775e+00],
[1.09437685e+01, -5.45042753e+00],
[1.32177639e+01, -6.44768333e+00],
[1.08205328e+01, -3.42842245e+00],
[1.12339220e+01, -2.48026347e+00],
[1.17774429e+01, -4.75986338e+00],
[1.16014652e+01, 6.23467624e-01],
[1.21346455e+01, 4.41442817e-01],
[1.25989037e+01, -3.81008172e+00],
[1.16449280e+01, -3.35743356e+00],
[1.33026037e+01, -7.49872351e+00],
[1.24047041e+01, -7.99148417e+00],
[1.11872721e+01, -5.15895724e-01],
[1.24107971e+01, -5.43478537e+00],
[1.13483229e+01, 9.55082119e-01],
[1.24065790e+01, -7.78985596e+00],
[1.01736193e+01, -7.24107385e-01],
[1.24363422e+01, -5.04861593e+00],
[1.20451689e+01, -6.61607552e+00],
[9.91461277e+00, -1.83136404e-01],
[1.00286541e+01, 2.25165367e-01],
[1.20567379e+01, -3.70011282e+00],
[1.15929918e+01, -6.55592346e+00],
[1.19902811e+01, -7.05997705e+00],
[1.32899113e+01, -7.48531771e+00],
[1.21674728e+01, -3.86144447e+00],
[1.00674219e+01, -1.10187650e+00],
[1.17257872e+01, -1.56192005e+00],
[1.25974159e+01, -7.22445393e+00],
[1.32728443e+01, -4.19393826e+00],
[1.17538919e+01, -3.22705865e+00],
[9.92642212e+00, 5.48063874e-01],
[1.16302233e+01, -4.87757254e+00],
[1.24635401e+01, -4.81949997e+00],
[1.10740538e+01, -4.50459385e+00],
[1.13344374e+01, 3.73205036e-01],
[1.27405863e+01, -5.52893686e+00],
[1.30165920e+01, -5.06265783e+00],
[1.13630991e+01, -4.28907728e+00],
[1.06912041e+01, -1.01544368e+00],
[1.10070810e+01, -3.25719117e+00]

```
[ 1.10270040e+01, -5.55710417e+00],  
[ 1.31959906e+01, -3.72454739e+00],  
[ 1.07124100e+01,  4.27633196e-01]], dtype=float32)
```

```
[55]: sns.scatterplot(x=transformed[:,0], y=transformed[:,1], hue=iris_df['species']);
```



```
[ ]:
```