KMC (student clustering)

December 18, 2022

```python
[18]: import pandas as pd
      import numpy as np
```

```python
[19]: df = pd.read_csv('student_clustering.csv')
      print("The shape of data is",df.shape)
      df.head()
```
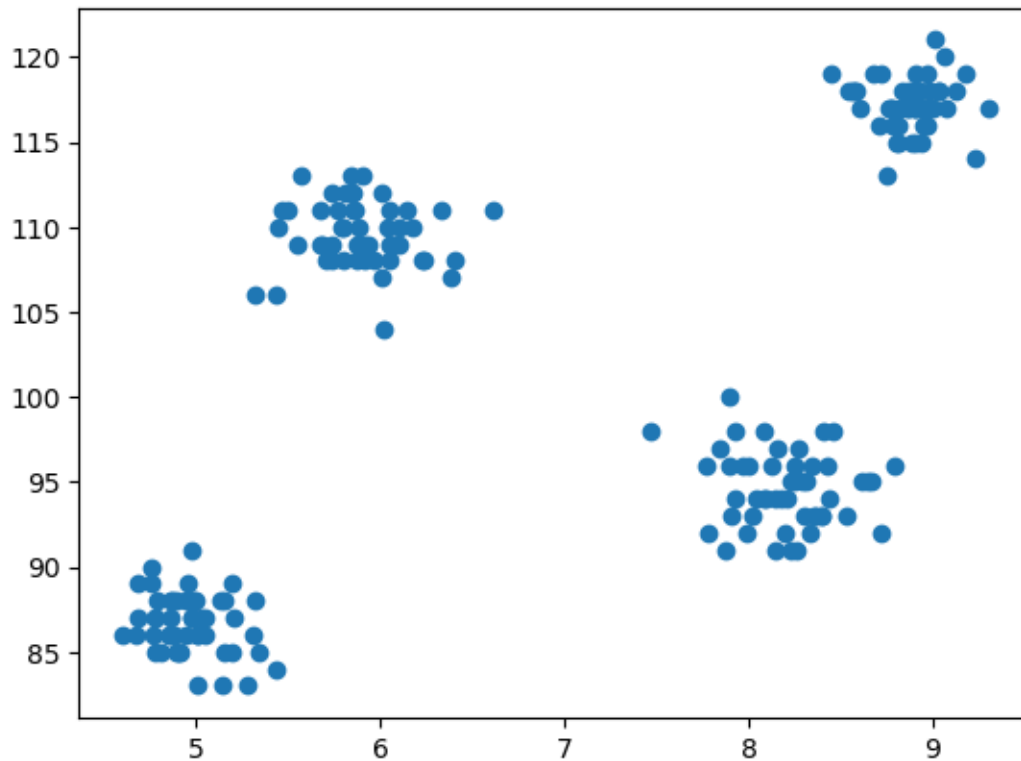
```
The shape of data is (200, 2)
```

```
[19]:    cgpa   iq
     0  5.13   88
     1  5.90  113
     2  8.36   93
     3  8.27   97
     4  5.45  110
```

```python
[20]: from matplotlib import pyplot as plt
      %matplotlib inline
```

```python
[21]: plt.scatter(df['cgpa'],df['iq'])
```

```
[21]: <matplotlib.collections.PathCollection at 0x1ec40036850>
```

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

```
[23]: wcss = []

      for i in range (1,11):
          km = KMeans(n_clusters=i)
          km.fit_predict(df)
          wcss.append(km.inertia_)
```

C:\Users\Deepak\ana-conda-3\lib\site-packages\sklearn\cluster_kmeans.py:1036:
UserWarning:

KMeans is known to have a memory leak on Windows with MKL, when there are less
chunks than available threads. You can avoid it by setting the environment
variable OMP_NUM_THREADS=1.
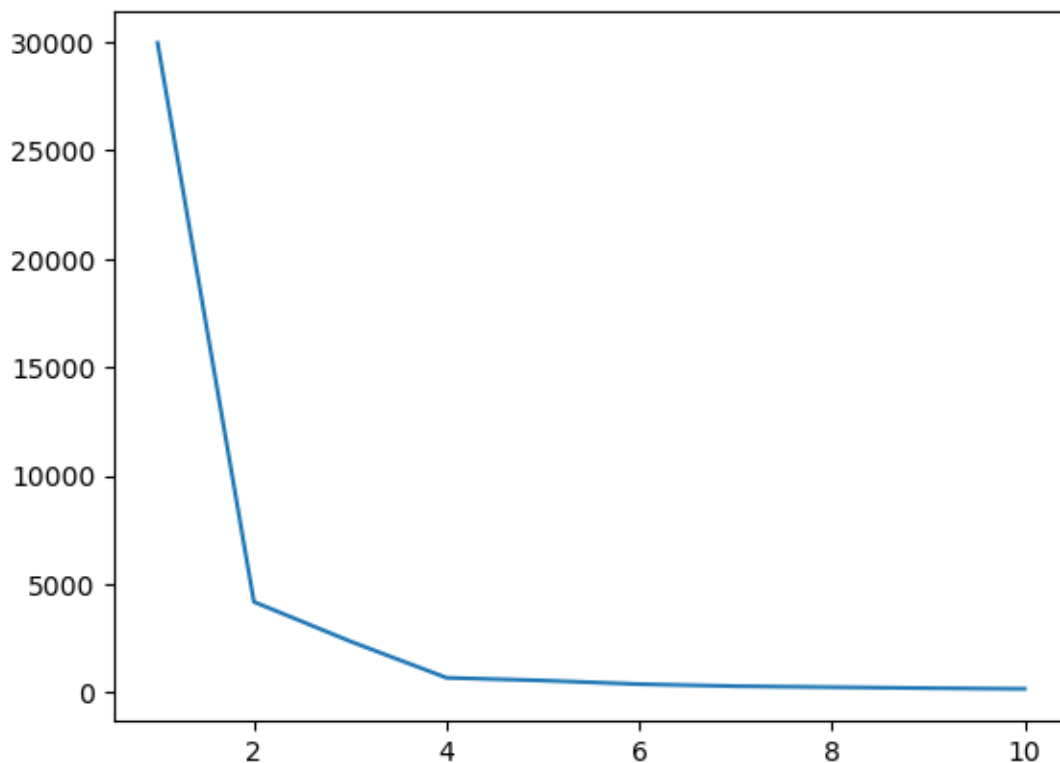
```
[24]: wcss
```

```
[24]: [29957.898287999997,
       4184.14127,
       2364.005583420083,
```

```
      681.96966,
      556.6603823704864,
      388.8524026875981,
      295.4391895943192,
      249.63018358945305,
      203.82503394280948,
      171.405907521685]
```

[25]: `plt.plot(range(1,11),wcss)`

[25]: `[<matplotlib.lines.Line2D at 0x1ec3fccfe50>]`

[26]:
```python
x = df.iloc[:,:].values
km = KMeans(n_clusters=4)
y_means = km.fit_predict(x)
```

[27]: `y_means`

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

```
       0, 2, 0, 0, 3, 1, 3, 3, 0, 3, 1, 0, 2, 1, 1, 0, 1, 1, 3, 1, 0, 0,
       3, 0, 2, 2, 3, 0, 3, 2, 0, 1, 1, 2, 3, 0, 3, 1, 3, 2, 1, 3, 3, 2,
       1, 1, 2, 0, 2, 1, 3, 3, 3, 1, 2, 1, 1, 0, 1, 0, 2, 1, 0, 1, 0, 0,
       1, 3, 2, 0, 2, 3, 1, 0, 2, 3, 0, 1, 2, 1, 1, 0, 0, 2, 0, 1, 1, 3,
       0, 2, 1, 0, 0, 2, 2, 2, 3, 1, 3, 3, 0, 2, 3, 3, 1, 1, 3, 1, 0, 2,
       2, 0])
```

[28]:
```
x[y_means ==3,1]
```

[28]:
```
array([ 93.,  97.,  98.,  94.,  97.,  95.,  91.,  98.,  92.,  98.,  94.,
        96.,  96.,  96.,  93.,  94.,  96.,  96.,  95.,  93.,  95.,  94.,
        92.,  91.,  92.,  95.,  94.,  95.,  92.,  94.,  91.,  95.,  93.,
        97.,  98.,  96.,  93., 100.,  96.,  94.,  95.,  93.,  92.,  98.,
        96.,  93.,  91.,  93.,  94.,  96.])
```

[29]:
```
plt.scatter(x[y_means == 0,0],x[y_means == 0,1],color='blue')
plt.scatter(x[y_means == 1,0],x[y_means == 1,1],color='red')
plt.scatter(x[y_means == 2,0],x[y_means == 2,1],color='green')
plt.scatter(x[y_means == 3,0],x[y_means == 3,1],color='yellow')
```

[29]: <matplotlib.collections.PathCollection at 0x1ec3fd0c910>
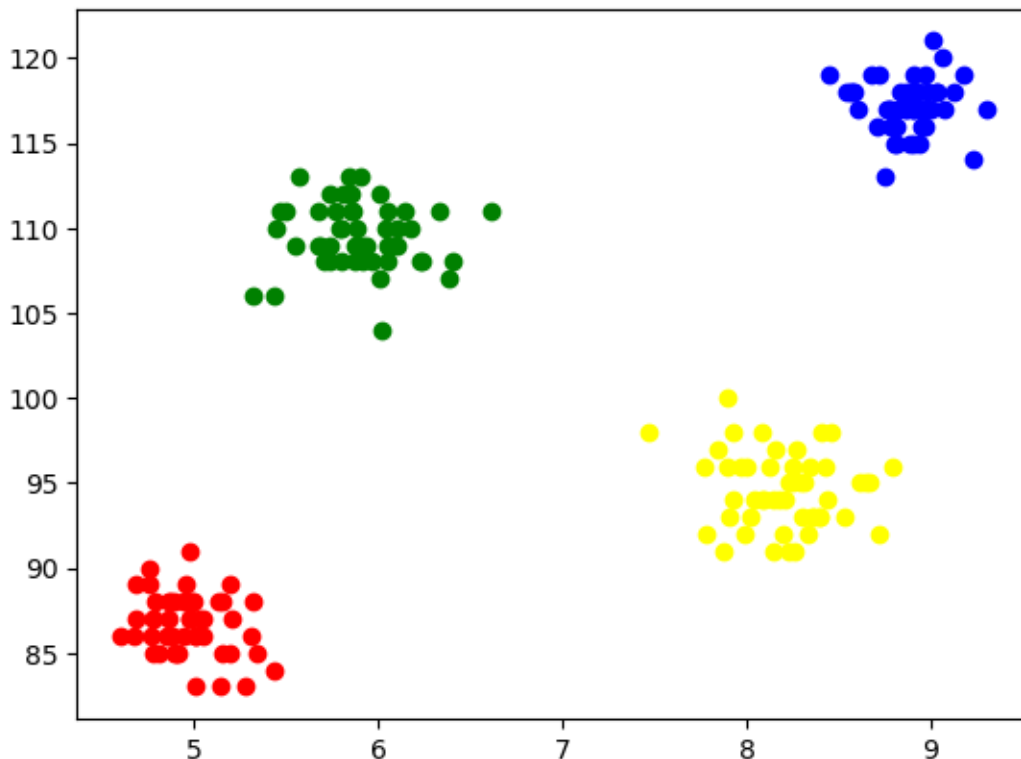
```
[30]: from sklearn.datasets import make_blobs

      centroids = [(-5,-5,5),(5,5,-5),(3.5,-2.5,4),(-2.5,2.5,-4)]
      cluster_std = [1,1,1,1]

      x,y =␣
       ↪make_blobs(n_samples=200,cluster_std=cluster_std,centers=centroids,n_features=3,random_stat
```

```
[31]: x
```

```
[31]: array([[ 4.33424548,  3.32580419, -4.17497018],
             [-3.32246719,  3.22171129, -4.625342  ],
             [-6.07296862, -4.13459237,  2.6984613 ],
             [ 6.90465871,  6.1110567 , -4.3409502 ],
             [-2.60839207,  2.95015551, -2.2346649 ],
             [ 5.88490881,  4.12271848, -5.86778722],
             [-4.68484061, -4.15383935,  4.14048406],
             [-1.82542929,  3.96089238, -3.4075272 ],
             [-5.34385368, -4.95640314,  4.37999916],
             [ 4.91549197,  4.70263812, -4.582698  ],
             [-3.80108212, -4.81484358,  4.62471505],
             [ 4.6735005 ,  3.65732421, -3.88561702],
             [-6.23005814, -4.4494625 ,  5.79280687],
             [-3.90232915,  2.95112294, -4.6949209 ],
             [ 3.72744124,  5.31354772, -4.49681519],
             [-3.3088472 ,  3.05743945, -3.81896126],
             [ 2.70273021, -2.21732429,  3.17390257],
             [ 4.06438286, -0.36217193,  3.214466  ],
             [ 4.69268607, -2.73794194,  5.15528789],
             [ 4.1210827 , -1.5438783 ,  3.29415949],
             [-6.61577235, -3.87858229,  5.40890054],
             [ 3.05777072, -2.17647265,  3.89000851],
             [-1.48617753,  0.27288737, -5.6993336 ],
             [-5.3224172 , -5.38405435,  6.13376944],
             [-5.26621851, -4.96738545,  3.62688268],
             [ 5.20183018,  5.66102029, -3.20784179],
             [-2.9189379 ,  2.02081508, -5.95210529],
             [ 3.30977897, -2.94873803,  3.32755196],
             [ 5.12910158,  6.6169496 , -4.49725912],
             [-2.46505641,  3.95391758, -3.33831892],
             [ 1.46279877, -4.44258918,  1.49355935],
             [ 3.87798127,  4.48290554, -5.99702683],
             [ 4.10944442,  3.8808846 , -3.0439211 ],
             [-6.09989127, -5.17242821,  4.12214158],
             [-3.03223402,  3.6181334 , -3.3256039 ],
             [ 7.44936865,  4.45422583, -5.19883786],
             [-4.47053468, -4.86229879,  5.07782113],
```

```
[-1.46701622,  2.27758597, -2.52983966],
[ 3.0208429 , -2.14983284,  4.01716473],
[ 3.82427424, -2.47813716,  3.53132618],
[-5.74715829, -3.3075454 ,  5.05080775],
[-1.51364782,  2.03384514, -2.61500866],
[-4.80170028, -4.88099135,  4.32933771],
[ 6.55880554,  5.1094027 , -6.2197444 ],
[-1.48879294,  1.02343734, -4.14319575],
[ 4.30884436, -0.71024532,  4.45128402],
[ 3.58646441, -4.64246673,  3.16983114],
[ 3.37256166,  5.60231928, -4.5797178 ],
[-1.39282455,  3.94287693, -4.53968156],
[-4.64945402, -6.31228341,  4.96130449],
[ 3.88352998,  5.0809271 , -5.18657899],
[ 3.32454103, -3.43391466,  3.46697967],
[ 3.45029742, -2.03335673,  5.03368687],
[-2.95994283,  3.14435367, -3.62832971],
[-3.03289825, -6.85798186,  6.23616403],
[-4.13665468, -5.1809203 ,  4.39607937],
[-3.6134361 ,  2.43258998, -2.83856002],
[ 2.07344458, -0.73204005,  3.52462712],
[ 4.11798553, -2.68417633,  3.88401481],
[ 3.60337958,  4.13868364, -4.32528847],
[-5.84520564, -5.67124613,  4.9873354 ],
[-2.41031359,  1.8988432 , -3.44392649],
[-2.75898285,  2.6892932 , -4.56378873],
[-2.442879  ,  1.70045251, -4.2915946 ],
[ 3.9611641 , -3.67598267,  5.01012718],
[-7.02220122, -5.30620401,  5.82797464],
[ 2.90019547, -1.37658784,  4.30526704],
[ 5.81095167,  6.04444209, -5.40087819],
[-5.75439794, -3.74713184,  5.51292982],
[-2.77584606,  3.72895559, -2.69029409],
[ 3.07085772, -1.29154367,  5.1157018 ],
[ 2.206915  ,  6.93752881, -4.63366799],
[ 4.2996015 ,  4.79660555, -4.75733056],
[ 4.86355526,  4.88094581, -4.98259059],
[-4.38161974, -4.76750544,  5.68255141],
[ 5.42952614,  4.3930016 , -4.89377728],
[ 3.69427308,  4.65501279, -5.23083974],
[ 5.90148689,  7.52832571, -5.24863478],
[-4.87984105, -4.38279689,  5.30017032],
[ 3.93816635, -1.37767168,  3.0029802 ],
[-3.32862798,  3.02887975, -6.23708651],
[-4.76990526, -4.23798882,  4.77767186],
[-2.12754315,  2.3515102 , -4.1834002 ],
[-0.64699051,  2.64225137, -3.48649452],
```

```
[-5.63699565, -4.80908452,  7.10025514],
[-1.86341659,  3.90925339, -2.37908771],
[ 4.82529684,  5.98633519, -4.7864661 ],
[-5.24937038, -3.53789206,  2.93985929],
[-4.59650836, -4.40642148,  3.90508815],
[-3.66400797,  3.19336623, -4.75806733],
[ 6.29322588,  4.88955297, -5.61736206],
[-2.85340998,  0.71208711, -3.63815268],
[-2.35835946, -0.01630386, -4.59566788],
[ 5.61060505, -3.80653407,  4.07638048],
[-1.78695095,  3.80620607, -4.60460297],
[-6.11731035, -4.7655843 ,  6.65980218],
[-5.63873041, -4.57650565,  5.07734007],
[ 5.62336218,  4.56504332, -3.59246   ],
[-3.37234925, -4.6619883 ,  3.80073197],
[-5.69166075, -5.39675353,  4.3128273 ],
[ 7.19069973,  3.10363908, -5.64691669],
[-3.86837061, -3.48018318,  7.18557541],
[-4.62243621, -4.87817873,  6.12948391],
[ 5.21112476,  5.01652757, -4.82281228],
[-2.61877117,  2.30100182, -2.13352862],
[-2.92449279,  1.76846902, -5.56573815],
[-2.80912132,  3.01093777, -2.28933816],
[ 4.35328122, -2.91302931,  5.83471763],
[ 2.79865557, -3.03722302,  4.15626385],
[-3.65498263,  2.3223678 , -5.51045638],
[ 4.8887794 , -3.16134424,  7.03085711],
[ 4.94317552,  5.49233656, -5.68067814],
[ 3.97761018, -3.52188594,  4.79452824],
[-3.41844004,  2.39465529, -3.36980433],
[ 3.50854895, -2.66819884,  3.82581966],
[-2.63971173,  3.88631426, -3.45187042],
[-3.37565464, -5.61175641,  4.47182825],
[-2.37162301,  4.26041518, -3.03346075],
[ 1.81594001, -3.6601701 ,  5.35010682],
[ 5.04366899,  4.77368576, -3.66854289],
[-4.19813897, -4.9534327 ,  4.81343023],
[ 5.1340482 ,  6.20205486, -4.71525189],
[ 3.39320601, -1.04857074,  3.38196315],
[ 4.34086156, -2.60288722,  5.14690038],
[-0.80619089,  2.69686978, -3.83013074],
[-5.62353073, -4.47942366,  3.85565861],
[ 5.56578332, -3.97115693,  3.1698281 ],
[ 4.41347606,  3.76314662, -4.12416107],
[ 4.01507361, -5.28253447,  4.58464661],
[-5.02461696, -5.77516162,  6.27375593],
[ 5.55635552, -0.73975077,  3.93934751],
```

```
[-5.20075807, -4.81343861,  5.41005165],
[-2.52752939,  4.24643509, -4.77507029],
[-3.85527629, -4.09840928,  5.50249434],
[ 5.78477065,  4.04457474, -4.41408957],
[ 1.74407436, -1.7852104 ,  4.85270406],
[ 3.27123417, -0.88663863,  3.62519531],
[ 7.18697965,  5.44136444, -5.10015523],
[-2.78899734,  2.10818376, -3.31599867],
[-3.37000822,  2.86919047, -3.14671781],
[-4.30196797, -5.44712856,  6.2245077 ],
[ 3.95541062,  7.05117344, -4.414338  ],
[ 3.55912398,  6.23225307, -5.25417987],
[-3.09384307,  2.15609929, -5.00016919],
[-5.93576943, -5.26788808,  5.53035547],
[ 5.83600472,  6.54335911, -4.24119434],
[ 4.68988323,  2.56516224, -3.9611754 ],
[-5.29809284, -4.51148185,  4.92442829],
[-1.30216916,  4.20459417, -2.95991085],
[ 4.9268873 ,  6.16033857, -4.63050728],
[-3.30618482,  2.24832579, -3.61728483],
[ 4.50178644,  4.68901502, -5.00189148],
[ 3.86723181, -1.26710081,  3.57714304],
[ 4.32458463, -1.84541985,  3.94881155],
[ 4.87953543,  3.76687926, -6.18231813],
[ 3.51335268, -3.1946936 ,  4.6218035 ],
[-4.83061757, -4.25944355,  4.0462994 ],
[-1.6290302 ,  1.99154287, -3.22258079],
[ 1.62683902, -1.57938488,  3.96463208],
[ 6.39984394,  4.21808832, -5.43750898],
[ 5.82400562,  4.43769457, -3.04512192],
[-3.25518824, -5.7612069 ,  5.3190391 ],
[-4.95778625, -4.41718479,  3.89938082],
[ 2.75003038, -0.4453759 ,  4.05340954],
[ 3.85249436, -2.73643695,  4.7278135 ],
[-5.10174587, -4.13111384,  5.75041164],
[-4.83996293, -4.12383108,  5.31563495],
[ 1.086497  , -4.27756638,  3.22214117],
[ 4.61584111, -2.18972771,  1.90575218],
[-4.25795584, -5.19183555,  4.11237104],
[ 5.09542509,  5.92145007, -4.9392498 ],
[-6.39649634, -6.44411381,  4.49553414],
[ 5.26246745,  5.2764993 , -5.7332716 ],
[ 3.5353601 , -4.03879325,  3.55210482],
[ 5.24879916,  4.70335885, -4.50478868],
[ 5.61853913,  4.55682807, -3.18946509],
[-2.39265671,  1.10118718, -3.91823218],
[ 3.16871683, -2.11346085,  3.14854434],
```

```
       [ 3.95161595, -1.39582567,  3.71826373],
       [-4.09914405, -5.68372786,  4.87710977],
       [-1.9845862 ,  1.38512895, -4.76730983],
       [-1.45500559,  3.1085147 , -4.0693287 ],
       [ 2.94250528, -1.56083126,  2.05667659],
       [ 2.77440288, -3.36776868,  3.86402267],
       [ 4.50088142, -2.88483225,  5.45810824],
       [-5.35224985, -6.1425182 ,  4.65065728],
       [-2.9148469 ,  2.95194604, -5.57915629],
       [-4.06889792, -4.71441267,  5.88514116],
       [ 3.47431968,  5.79502609, -5.37443832],
       [ 3.66804833,  3.23931144, -6.65072127],
       [-3.22239191,  3.59899633, -4.90163449],
       [-3.6077125 ,  2.48228168, -5.71939447],
       [ 5.5627611 ,  5.24073709, -4.71933492],
       [ 1.38583608, -2.91163916,  5.27852808],
       [ 4.42001793, -2.69505734,  4.80539342],
       [ 4.71269214,  5.68006984, -5.3198016 ],
       [-4.13744959,  6.4586027 , -3.35135636],
       [-5.20889423, -4.41337681,  5.83898341],
       [ 2.6194224 , -2.77909772,  5.62284909],
       [-1.3989998 ,  3.28002714, -4.6294416 ]])
```

[32]: `!pip install plotly.express`

Requirement already satisfied: plotly.express in c:\users\deepak\ana-
conda-3\lib\site-packages (0.4.1)
Requirement already satisfied: statsmodels>=0.9.0 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly.express) (0.13.2)
Requirement already satisfied: patsy>=0.5 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly.express) (0.5.2)
Requirement already satisfied: numpy>=1.11 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly.express) (1.21.5)
Requirement already satisfied: scipy>=0.18 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly.express) (1.9.1)
Requirement already satisfied: plotly>=4.1.0 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly.express) (5.9.0)
Requirement already satisfied: pandas>=0.20.0 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly.express) (1.4.4)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\deepak\ana-
conda-3\lib\site-packages (from pandas>=0.20.0->plotly.express) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\deepak\ana-
conda-3\lib\site-packages (from pandas>=0.20.0->plotly.express) (2022.1)
Requirement already satisfied: six in c:\users\deepak\ana-conda-3\lib\site-
packages (from patsy>=0.5->plotly.express) (1.16.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\deepak\ana-
conda-3\lib\site-packages (from plotly>=4.1.0->plotly.express) (8.0.1)
Requirement already satisfied: packaging>=21.3 in c:\users\deepak\ana-

conda-3\lib\site-packages (from statsmodels>=0.9.0->plotly.express) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\deepak\ana-
conda-3\lib\site-packages (from
packaging>=21.3->statsmodels>=0.9.0->plotly.express) (3.0.9)

```python
import plotly.express as px
fig = px.scatter_3d(x=x[:,0],y=x[:,1],z=x[:,2])
fig.show()
```

[ ]:

[ ]:

[ ]:

[ ]: