MATPLOTLIB:

Matplotlib is easy to use and an amazing visualizing library in Python. It is built on NumPy arrays and designed to work with the broader SciPy stack and consists of several plots like line, bar, scatter, histogram, etc.

Key Features of Matplotlib:

1.**Versatility**: Matplotlib can generate a wide range of plots, including line plots, scatter plots, bar plots, histograms, pie charts, and more.

2.**Customization**: It offers extensive customization options to control every aspect of the plot, such as line styles, colors, markers, labels, and annotations.

3.**Integration with NumPy**: Matplotlib integrates seamlessly with NumPy, making it easy to plot data arrays directly.

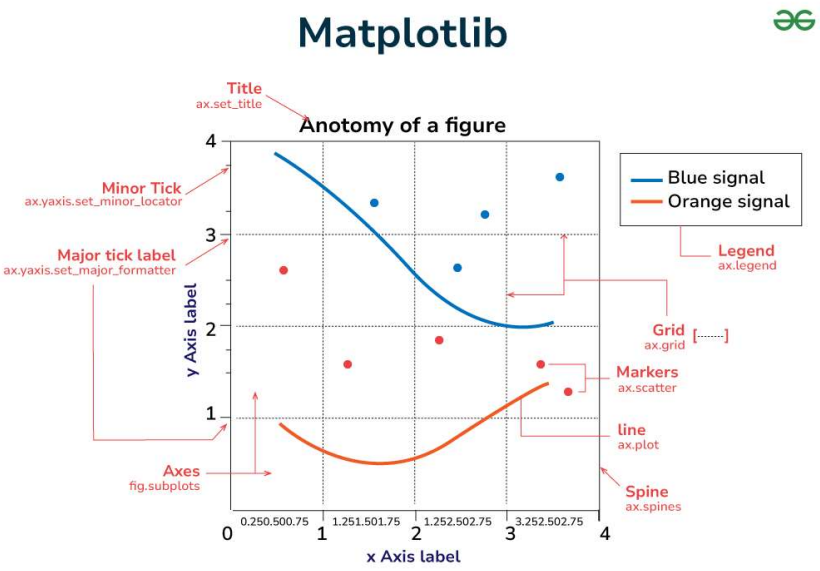
4.**Extensible**: Matplotlib is highly extensible, with a large ecosystem of add-on toolkits and extensions like Seaborn, Pandas plotting functions, and Basemap for geographical plotting.5.**Cross-**

4.5.**Platform**: It is platform-independent and can run on various operating systems, including Windows, macOS, and Linux.

6.**Interactive Plots**: Matplotlib supports interactive plotting through the use of widgets and event handling, enabling users to explore data dynamically.

What is a Matplotlib Figure?

In Matplotlib, a figure is the top-level container that holds all the elements of a plot. It represents the entire window or page where the plot is drawn.



The graphs in Matplotlib are classified based on the data we are trying to visualise This results in the 5 main types of data that matplotlib can handle:

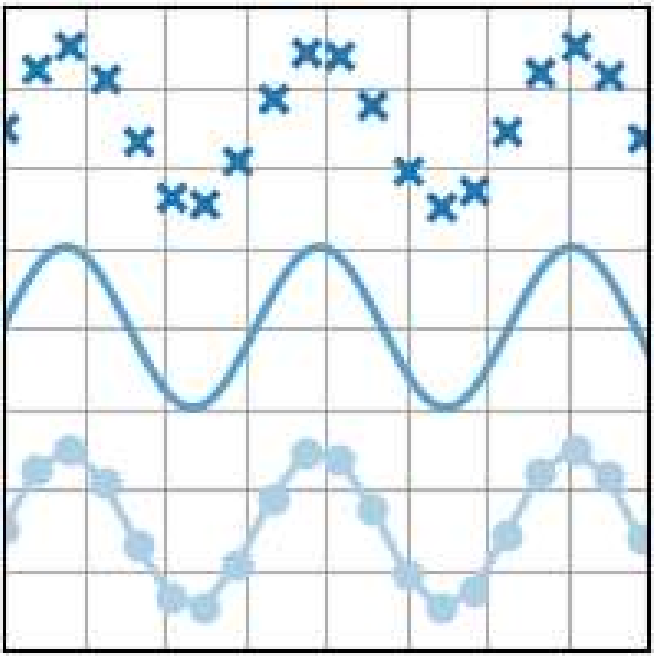
1)Pairwise Data:

Allows us to see both the distribution of a single variable and the relationship between two variables. Here are the types of graphs:

plot(x,y):

This helps us project a basic plot using given x andy coordinate values.

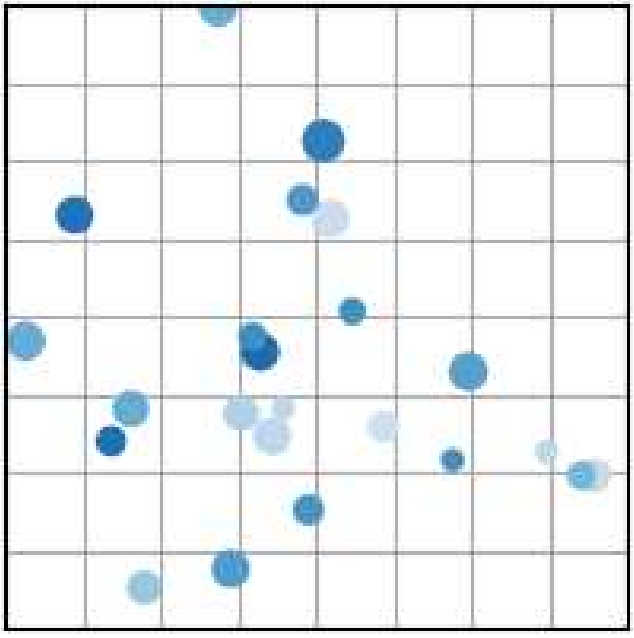
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
x = np.linspace(0, 10, 100)
y = 4 + 1 * np.sin(2 * x)
x2 = np.linspace(0, 10, 25)
y2 = 4 + 1 * np.sin(2 * x2)

# plot
fig, ax = plt.subplots()

ax.plot(x2, y2 + 2.5, 'x', markeredgewidth=2)
ax.plot(x, y, linewidth=2.0)
ax.plot(x2, y2 - 2.5, 'o-', linewidth=2)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



scatter(x,y):

This helps us create a scatter plot for pairwise data.

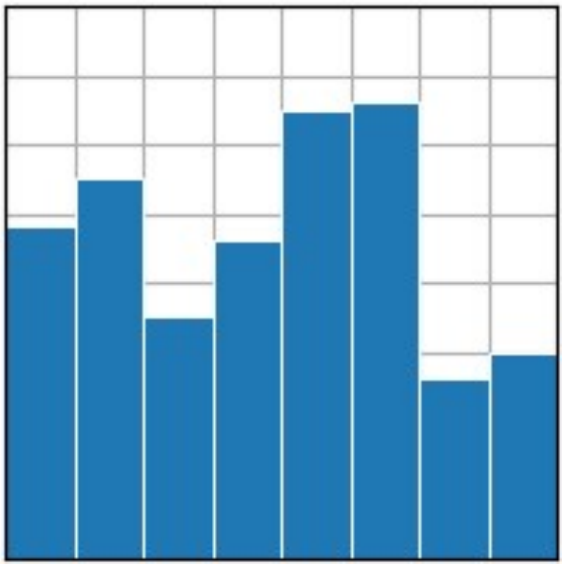
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make the data
np.random.seed(3)
x = 4 + np.random.normal(0, 2, 24)
y = 4 + np.random.normal(0, 2, len(x))
# size and color:
sizes = np.random.uniform(15, 80, len(x))
colors = np.random.uniform(15, 80, len(x))

# plot
fig, ax = plt.subplots()

ax.scatter(x, y, s=sizes, c=colors, vmin=0, vmax=100)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



bar(x,height):

Helps you create a bar plot for a given batch of variables.

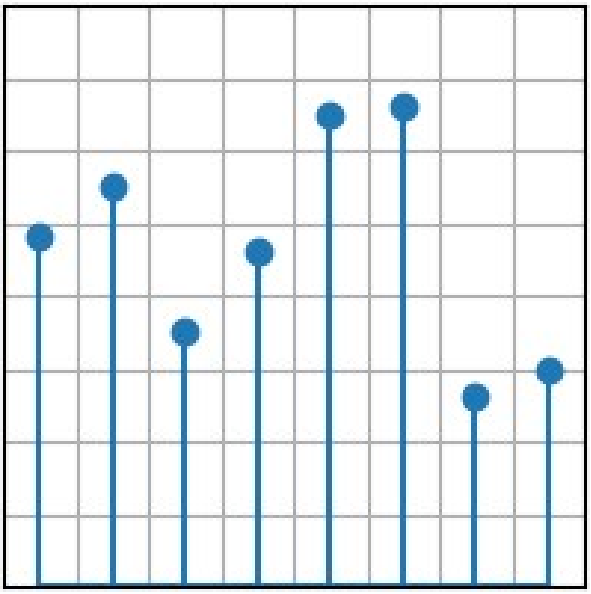
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data:
x = 0.5 + np.arange(8)
y = [4.8, 5.5, 3.5, 4.6, 6.5, 6.6, 2.6, 3.0]

# plot
fig, ax = plt.subplots()

ax.bar(x, y, width=1, edgecolor="white", linewidth=0.7)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



stem(x,y):

Helps classify either discrete or continuous variables.

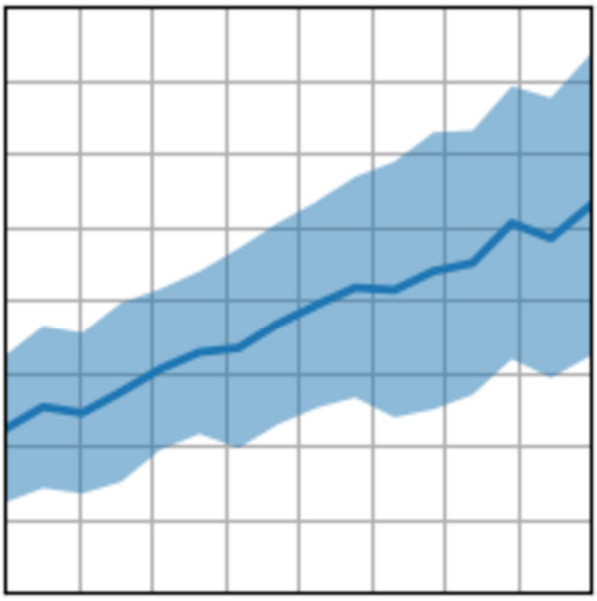
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
x = 0.5 + np.arange(8)
y = [4.8, 5.5, 3.5, 4.6, 6.5, 6.6, 2.6, 3.0]

# plot
fig, ax = plt.subplots()

ax.stem(x, y)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



fill_between(x,y1,y2):

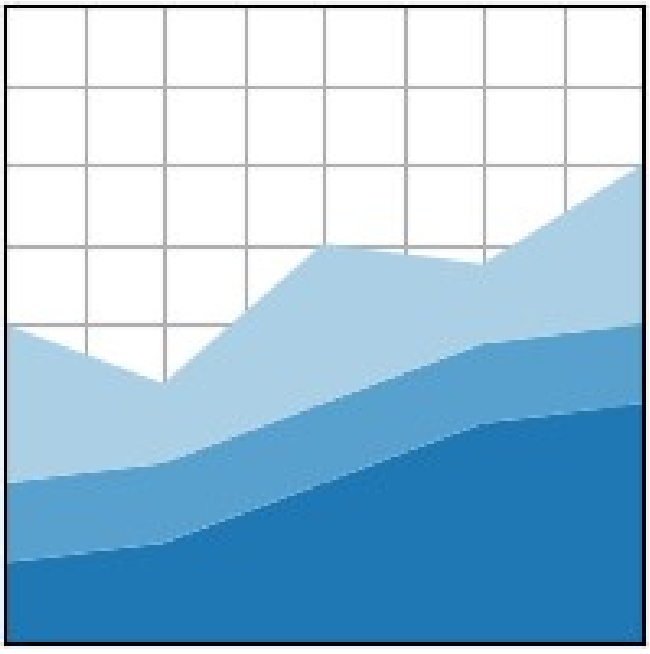
Fills the graph between a given range of values.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
np.random.seed(1)
x = np.linspace(0, 8, 16)
y1 = 3 + 4*x/8 + np.random.uniform(0.0, 0.5, len(x))
y2 = 1 + 2*x/8 + np.random.uniform(0.0, 0.5, len(x))

# plot
fig, ax = plt.subplots()

ax.fill_between(x, y1, y2, alpha=.5, linewidth=0)
ax.plot(x, (y1 + y2)/2, linewidth=2)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



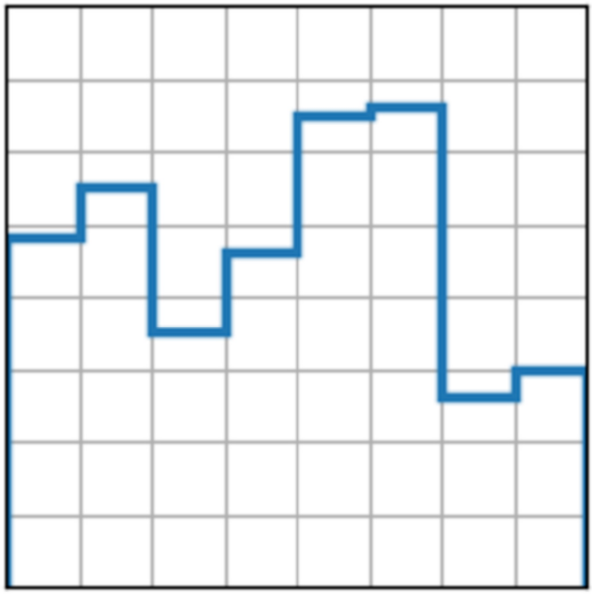
stackplot(x,y):

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
x = np.arange(0, 10, 2)
ay = [1, 1.25, 2, 2.75, 3]
by = [1, 1, 1, 1, 1]
cy = [2, 1, 2, 1, 2]
y = np.vstack([ay, by, cy])

# plot
fig, ax = plt.subplots()

ax.stackplot(x, y)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



stairs(values):

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
y = [4.8, 5.5, 3.5, 4.6, 6.5, 6.6, 2.6, 3.0]

# plot
fig, ax = plt.subplots()

ax.stairs(y, linewidth=2.5)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



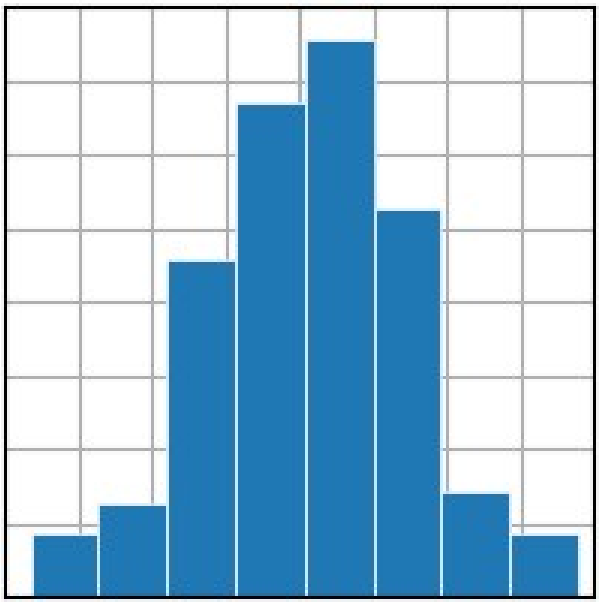
2)Statistical Distributions:

It helps map data in a probabilistic format.Which help predict certain outcomes.

hist(x):

Helps plot distribution of numeric values as series of bars.

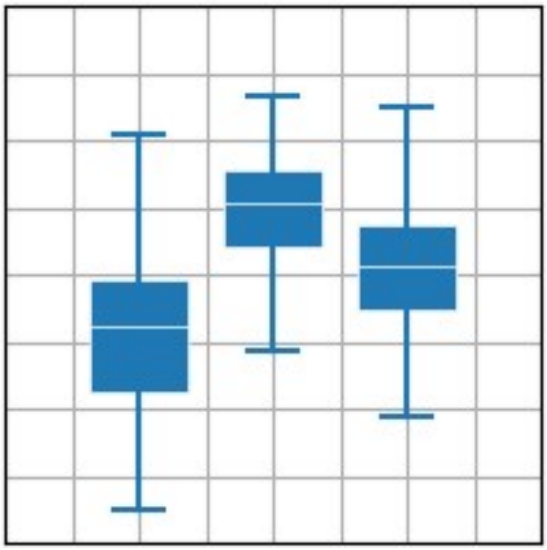
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
np.random.seed(1)
x = 4 + np.random.normal(0, 1.5, 200)

# plot:
fig, ax = plt.subplots()

ax.hist(x, bins=8, linewidth=0.5, edgecolor="white")

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 56), yticks=np.linspace(0, 56, 9))

plt.show()
```



boxplot(X):

A graphical method to visualize data distribution for gaining insights and making informed decisions.

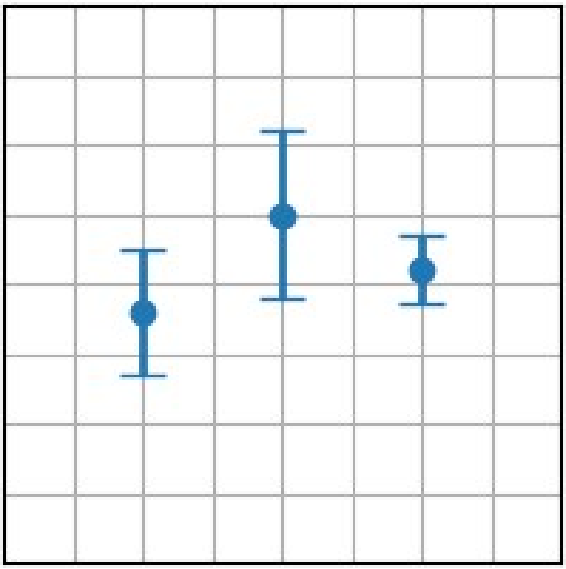
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data:
np.random.seed(10)
D = np.random.normal((3, 5, 4), (1.25, 1.00, 1.25), (100, 3))

# plot
fig, ax = plt.subplots()
VP = ax.boxplot(D, positions=[2, 4, 6], widths=1.5, patch_artist=True,
                showmeans=False, showfliers=False,
                medianprops={"color": "white", "linewidth": 0.5},
                boxprops={"facecolor": "C0", "edgecolor": "white",
                          "linewidth": 0.5},
                whiskerprops={"color": "C0", "linewidth": 1.5},
                capprops={"color": "C0", "linewidth": 1.5})

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



errorbar(x,y,yerr,xerr):

A graphical representation of the variability of data in a chart, and is used to indicate the uncertainty or error in a reported measurement.

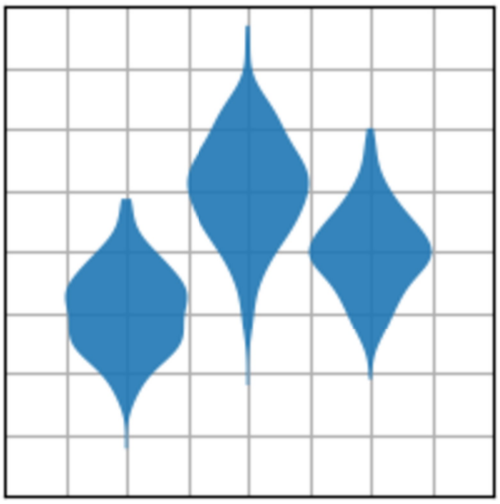
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data:
np.random.seed(1)
x = [2, 4, 6]
y = [3.6, 5, 4.2]
yerr = [0.9, 1.2, 0.5]

# plot:
fig, ax = plt.subplots()

ax.errorbar(x, y, yerr, fmt='o', linewidth=2, capsize=6)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



violinplot(D):

Depicts distributions of numeric data for one or more groups using density curves.

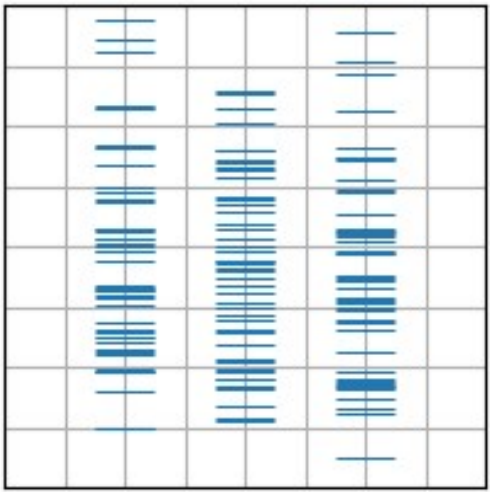
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data:
np.random.seed(10)
D = np.random.normal((3, 5, 4), (0.75, 1.00, 0.75), (200, 3))

# plot:
fig, ax = plt.subplots()

vp = ax.violinplot(D, [2, 4, 6], widths=2,
                   showmeans=False, showmedians=False, showextrema=False)
# styling:
for body in vp['bodies']:
    body.set_alpha(0.9)
ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



eventplot(D):

The sequence of events in a story, where each event is connected to the next by cause and effect.

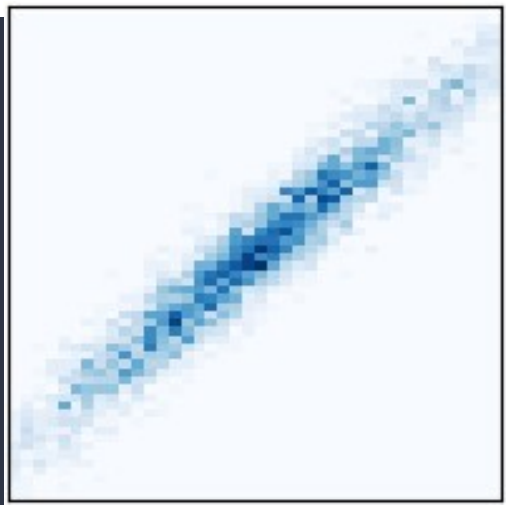
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data:
np.random.seed(1)
x = [2, 4, 6]
D = np.random.gamma(4, size=(3, 50))

# plot:
fig, ax = plt.subplots()

ax.eventplot(D, orientation="vertical", lineoffsets=x, linewidth=0.75)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



hist2d(x,y):

A 2D Histogram plot.

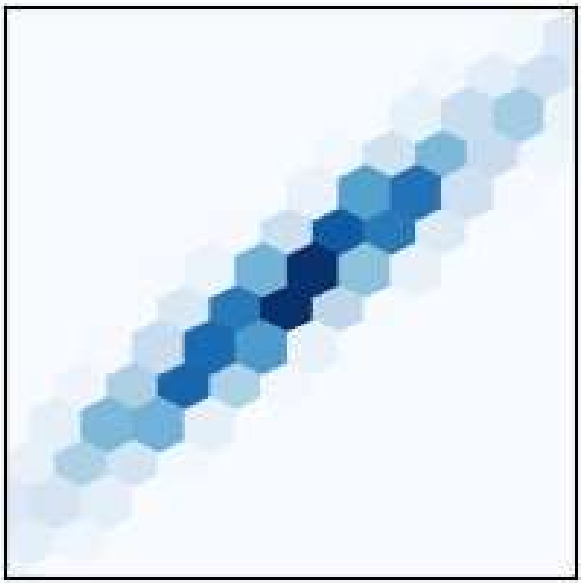
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data: correlated + noise
np.random.seed(1)
x = np.random.randn(5000)
y = 1.2 * x + np.random.randn(5000) / 3

# plot:
fig, ax = plt.subplots()

ax.hist2d(x, y, bins=(np.arange(-3, 3, 0.1), np.arange(-3, 3, 0.1)))

ax.set(xlim=(-2, 2), ylim=(-3, 3))

plt.show()
```



hexbin(x,y,c):

Makes a hexagonal binning plot of points of x,y.

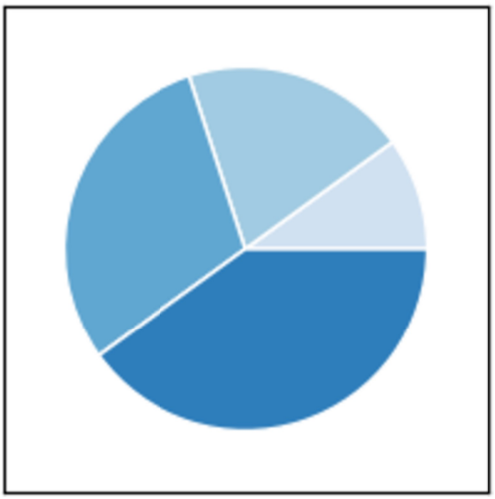
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data: correlated + noise
np.random.seed(1)
x = np.random.randn(5000)
y = 1.2 * x + np.random.randn(5000) / 3

# plot:
fig, ax = plt.subplots()

ax.hexbin(x, y, gridsize=20)

ax.set(xlim=(-2, 2), ylim=(-3, 3))

plt.show()
```



pie(x):

A way of summarizing a set of nominal data or displaying the different values of a given variable.

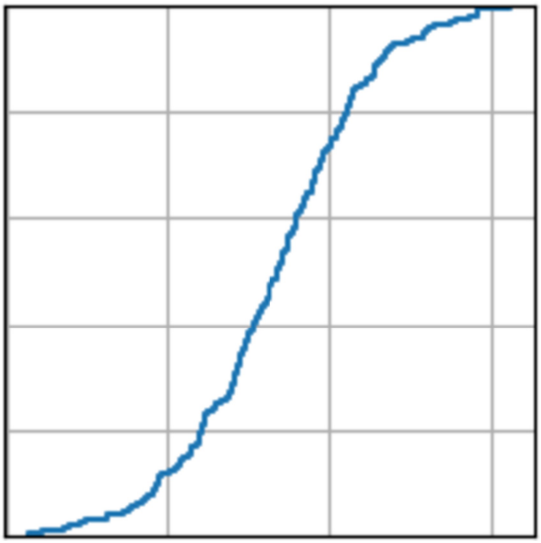
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')


# make data
x = [1, 2, 3, 4]
colors = plt.get_cmap('Blues')(np.linspace(0.2, 0.7, len(x)))

# plot
fig, ax = plt.subplots()
ax.pie(x, colors=colors, radius=3, center=(4, 4),
       wedgeprops={"linewidth": 1, "edgecolor": "white"}, frame=True)

ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
       ylim=(0, 8), yticks=np.arange(1, 8))

plt.show()
```



ecdf(x):

Compute and plot the empirical cumulative distribution function of x.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# make data
np.random.seed(1)
x = 4 + np.random.normal(0, 1.5, 200)

# plot:
fig, ax = plt.subplots()
ax.ecdf(x)
plt.show()
```



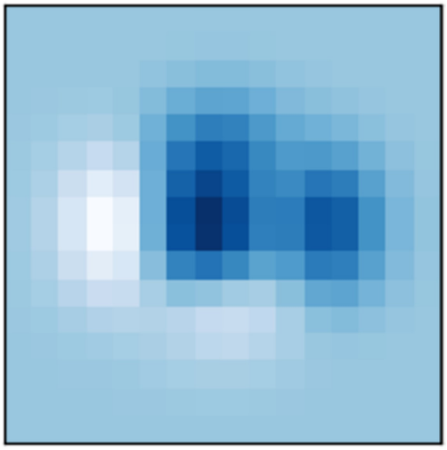
3)Gridded Data:

A collection of values or measurements that are organized in a grid at regular intervals.

imshow(Z):

Display data as an image, i.e., on a 2D regular raster.

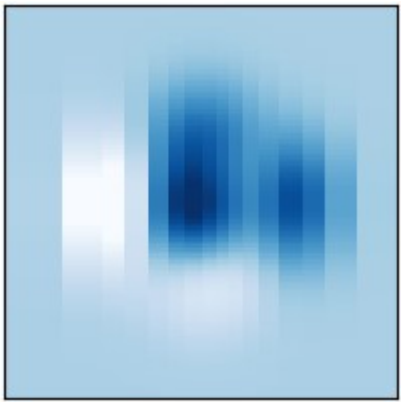
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data
X, Y = np.meshgrid(np.linspace(-3, 3, 16), np.linspace(-3, 3, 16))
Z = (1 - X/2 + X**5 + Y**3) * np.exp(-X**2 - Y**2)

# plot
fig, ax = plt.subplots()

ax.imshow(Z, origin='lower')

plt.show()
```



pcolormesh(X,Y,Z):

Create a pseudocolor plot with a non-regular rectangular grid.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data with uneven sampling in x
x = [-3, -2, -1.6, -1.2, -.8, -.5, -.2, .1, .3, .5, .8, 1.1, 1.5, 1.9, 2.3, 3]
X, Y = np.meshgrid(x, np.linspace(-3, 3, 128))
Z = (1 - X/2 + X**5 + Y**3) * np.exp(-X**2 - Y**2)

# plot
fig, ax = plt.subplots()

ax.pcolormesh(X, Y, Z, vmin=-0.5, vmax=1.0)

plt.show()
```



contour(X,Y,Z):

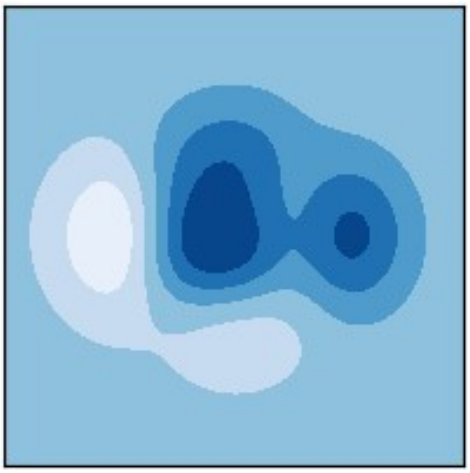
Plots Contour Lines.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data
X, Y = np.meshgrid(np.linspace(-3, 3, 256), np.linspace(-3, 3, 256))
Z = (1 - X/2 + X**5 + Y**3) * np.exp(-X**2 - Y**2)
levels = np.linspace(np.min(Z), np.max(Z), 7)

# plot
fig, ax = plt.subplots()

ax.contour(X, Y, Z, levels=levels)

plt.show()
```

contourf(X,Y,Z):

Plot filled contours.

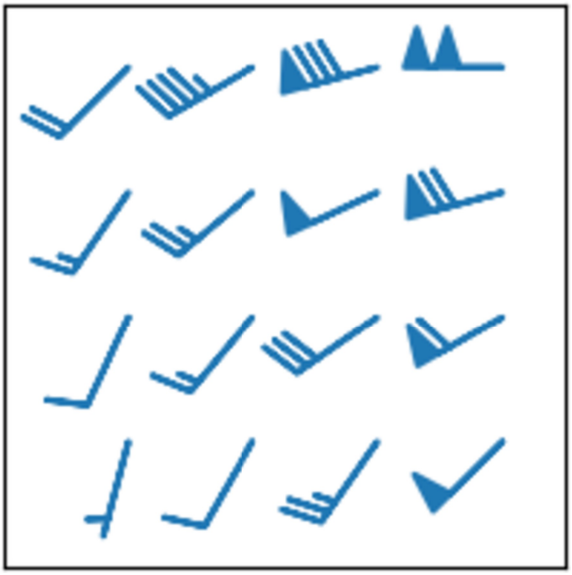
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data
X, Y = np.meshgrid(np.linspace(-3, 3, 256), np.linspace(-3, 3, 256))
Z = (1 - X/2 + X**5 + Y**3) * np.exp(-X**2 - Y**2)
levels = np.linspace(Z.min(), Z.max(), 7)

# plot
fig, ax = plt.subplots()

ax.contourf(X, Y, Z, levels=levels)

plt.show()
```



barbs(X,Y,U,V):

Plot a 2D field of wind barbs.

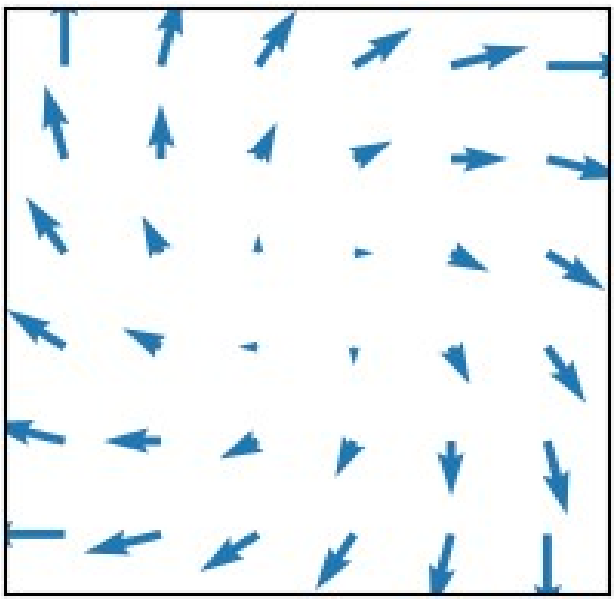
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data:
X, Y = np.meshgrid([1, 2, 3, 4], [1, 2, 3, 4])
angle = np.pi / 180 * np.array([[15., 30, 35, 45],
                                [25., 40, 55, 60],
                                [35., 50, 65, 75],
                                [45., 60, 75, 90]])
amplitude = np.array([[5, 10, 25, 50],
                      [10, 15, 30, 60],
                      [15, 26, 50, 70],
                      [20, 45, 80, 100]])
U = amplitude * np.sin(angle)
V = amplitude * np.cos(angle)

# plot:
fig, ax = plt.subplots()

ax.barbs(X, Y, U, V, barbcolor='C0', flagcolor='C0', length=7, linewidth=1.5)

ax.set(xlim=(0, 4.5), ylim=(0, 4.5))

plt.show()
```



quiver(X,Y,U,V):

A field of arrows.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data
x = np.linspace(-4, 4, 6)
y = np.linspace(-4, 4, 6)
X, Y = np.meshgrid(x, y)
U = X + Y
V = Y - X

# plot
fig, ax = plt.subplots()

ax.quiver(X, Y, U, V, color="C0", angles='xy',
          scale_units='xy', scale=5, width=.015)

ax.set(xlim=(-5, 5), ylim=(-5, 5))

plt.show()
```



steamplot(X,Y,U.V):

Draw Streamlines of a vector flow.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make a stream function:
X, Y = np.meshgrid(np.linspace(-3, 3, 256), np.linspace(-3, 3, 256))
Z = (1 - X/2 + X**5 + Y**3) * np.exp(-X**2 - Y**2)
# make U and V out of the streamfunction:
V = np.diff(Z[1:, :], axis=1)
U = -np.diff(Z[:, 1:], axis=0)

# plot:
fig, ax = plt.subplots()

ax.streamplot(X[1:, 1:], Y[1:, 1:], U, V)

plt.show()
```

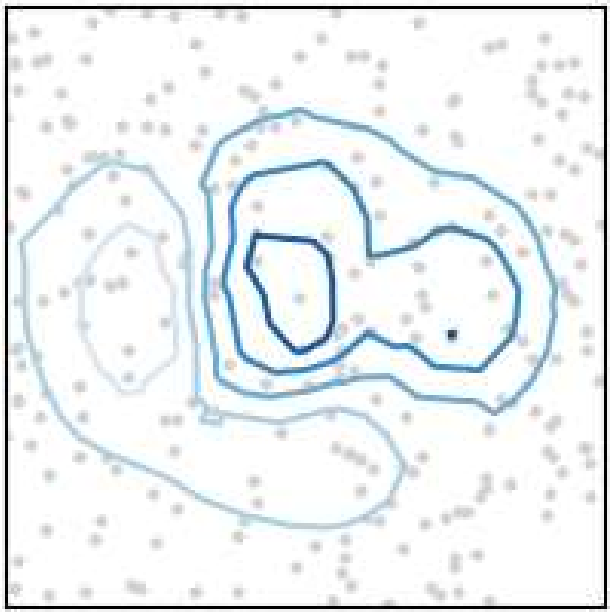
4)Irregularly Gridded Data:

An unstructured grid or irregular grid is a tessellation of a part of the Euclidean plane or Euclidean space by simple shapes, such as triangles or tetrahedra, in an irregular pattern.

tricontour(x,y,z):

Drawing contour lines on an unstructured triangular grid.

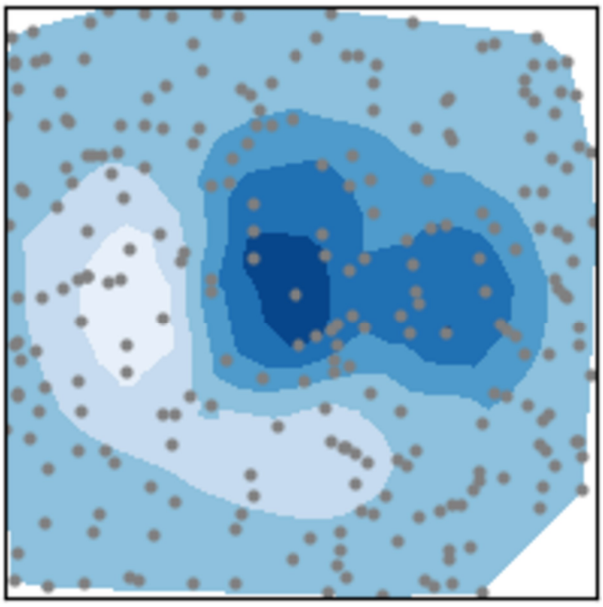
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data:
np.random.seed(1)
x = np.random.uniform(-3, 3, 256)
y = np.random.uniform(-3, 3, 256)
z = (1 - x/2 + x**5 + y**3) * np.exp(-x**2 - y**2)
levels = np.linspace(z.min(), z.max(), 7)

# plot:
fig, ax = plt.subplots()

ax.plot(x, y, 'o', markersize=2, color='lightgrey')
ax.tricontour(x, y, z, levels=levels)

ax.set(xlim=(-3, 3), ylim=(-3, 3))

plt.show()
```



tricontourf(x,y,z):

Drawing filled contours on an unstructured triangular grid.

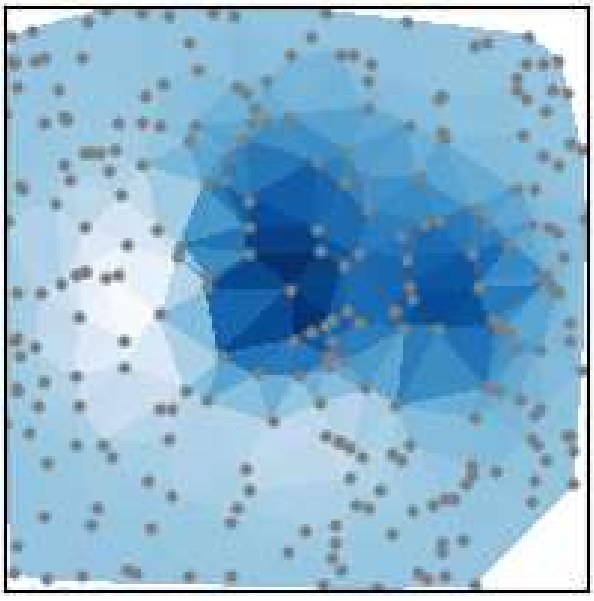
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data:
np.random.seed(1)
x = np.random.uniform(-3, 3, 256)
y = np.random.uniform(-3, 3, 256)
z = (1 - x/2 + x**5 + y**3) * np.exp(-x**2 - y**2)
levels = np.linspace(z.min(), z.max(), 7)

# plot:
fig, ax = plt.subplots()

ax.plot(x, y, 'o', markersize=2, color='grey')
ax.tricontourf(x, y, z, levels=levels)

ax.set(xlim=(-3, 3), ylim=(-3, 3))

plt.show()
```



tripcolor(x,y,z):

Create a pseudocolor plot of an unstructured triangular grid.

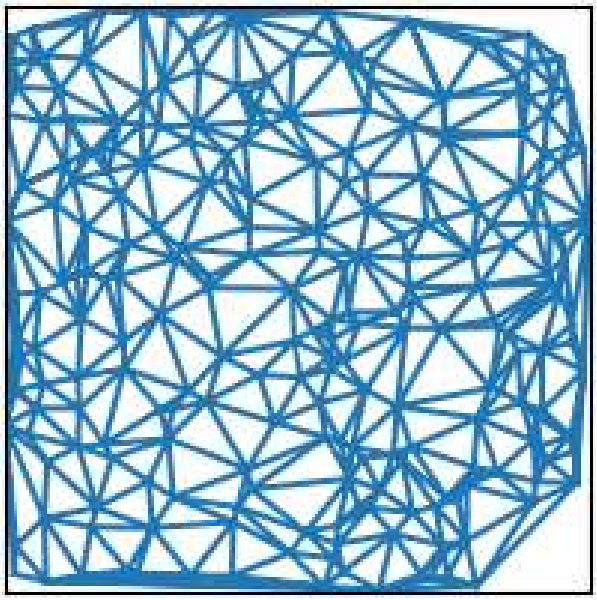
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data:
np.random.seed(1)
x = np.random.uniform(-3, 3, 256)
y = np.random.uniform(-3, 3, 256)
z = (1 - x/2 + x**5 + y**3) * np.exp(-x**2 - y**2)

# plot:
fig, ax = plt.subplots()

ax.plot(x, y, 'o', markersize=2, color='grey')
ax.tripcolor(x, y, z)

ax.set(xlim=(-3, 3), ylim=(-3, 3))

plt.show()
```



triplot(x.y):

Draw an unstructured triangular grid as lines and/or markers.

```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery-nogrid')

# make data:
np.random.seed(1)
x = np.random.uniform(-3, 3, 256)
y = np.random.uniform(-3, 3, 256)
z = (1 - x/2 + x**5 + y**3) * np.exp(-x**2 - y**2)

# plot:
fig, ax = plt.subplots()

ax.triplot(x, y)

ax.set(xlim=(-3, 3), ylim=(-3, 3))

plt.show()
```



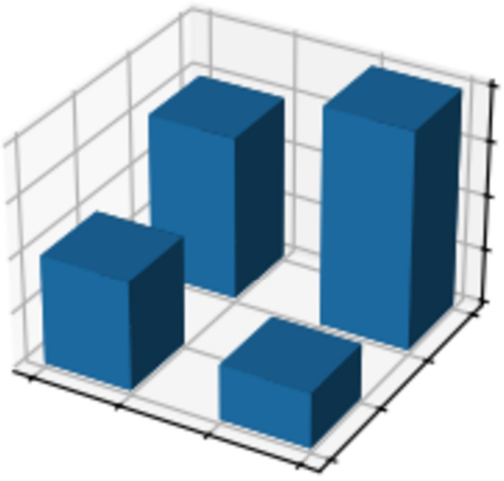
5)3D and Volumetric Data:

a collection of samples in 3D space, where each sample has coordinates and a value that represents a property like color, density, or pressure.

barplot3d(x,y,z,dx,dy,dz):

This method creates three-dimensional barplot where the width, depth, height, and color of the bars can all be uniquely set.

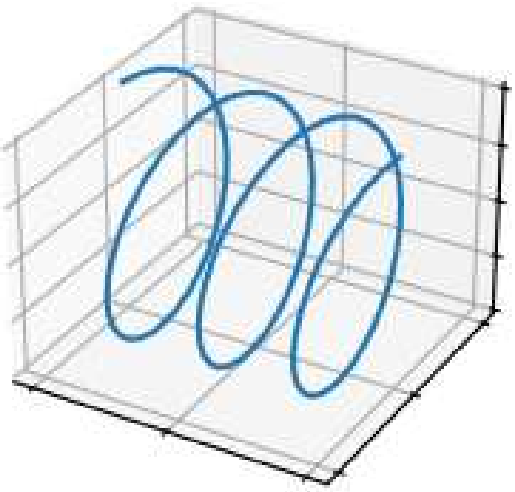
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# Make data
x = [1, 1, 2, 2]
y = [1, 2, 1, 2]
z = [0, 0, 0, 0]
dx = np.ones_like(x)*0.5
dy = np.ones_like(x)*0.5
dz = [2, 3, 1, 4]

# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.bar3d(x, y, z, dx, dy, dz)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



plot(xs,ys,zs):

Draws the basic plot in a 3d axis.

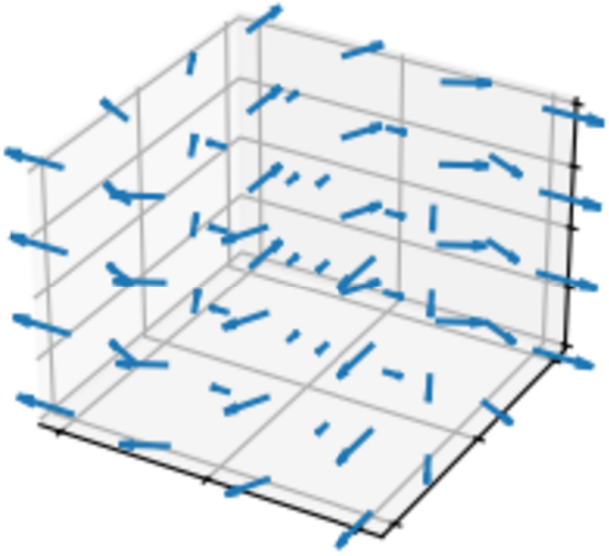
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# Make data
n = 100
xs = np.linspace(0, 1, n)
ys = np.sin(xs * 6 * np.pi)
zs = np.cos(xs * 6 * np.pi)

# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.plot(xs, ys, zs)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



quiver(X,Y,Z,U,V,W):

Plots quivers in a 3d field.

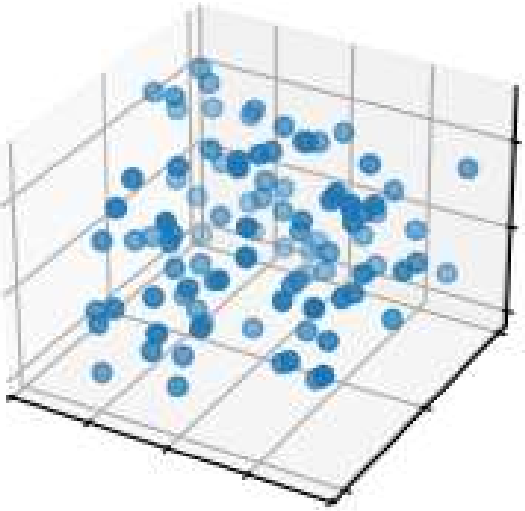
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# Make data
n = 4
x = np.linspace(-1, 1, n)
y = np.linspace(-1, 1, n)
z = np.linspace(-1, 1, n)
X, Y, Z = np.meshgrid(x, y, z)
U = (X + Y)/5
V = (Y - X)/5
W = Z*0


# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.quiver(X, Y, Z, U, V, W)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



scatter(xs,ys.zs):

Plots a scatter plot in a 3d field.

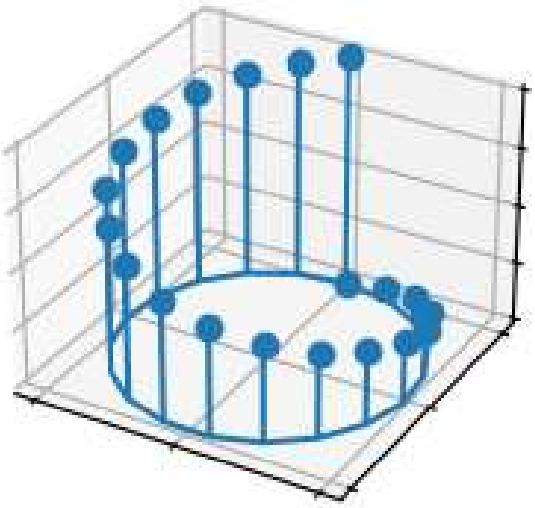
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# Make data
np.random.seed(19680801)
n = 100
rng = np.random.default_rng()
xs = rng.uniform(23, 32, n)
ys = rng.uniform(0, 100, n)
zs = rng.uniform(-50, -25, n)

# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.scatter(xs, ys, zs)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



stem(x,y,z):

Plots a stem graph in a 3d field.

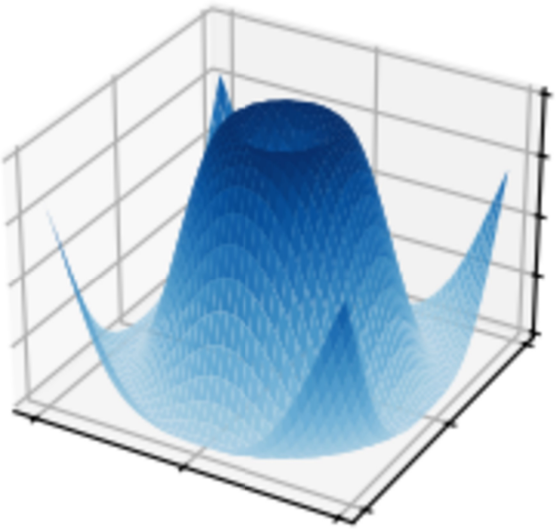
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# Make data
n = 20
x = np.sin(np.linspace(0, 2*np.pi, n))
y = np.cos(np.linspace(0, 2*np.pi, n))
z = np.linspace(0, 1, n)

# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.stem(x, y, z)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



plot_surface(X.Y,Z):

Plots a 3D plane in a 3D field.

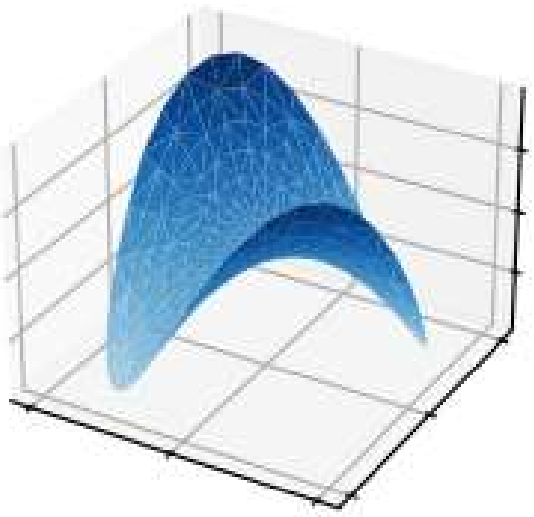
```
import matplotlib.pyplot as plt
import numpy as np

from matplotlib import cm

plt.style.use('_mpl-gallery')

# Make data
X = np.arange(-5, 5, 0.25)
Y = np.arange(-5, 5, 0.25)
X, Y = np.meshgrid(X, Y)
R = np.sqrt(X**2 + Y**2)
Z = np.sin(R)

# Plot the surface
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.plot_surface(X, Y, Z, vmin=Z.min() * 2, cmap=cm.Blues)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



plot_trisurf(x,y,z):

plots a triangulated surface in a 3D plane:

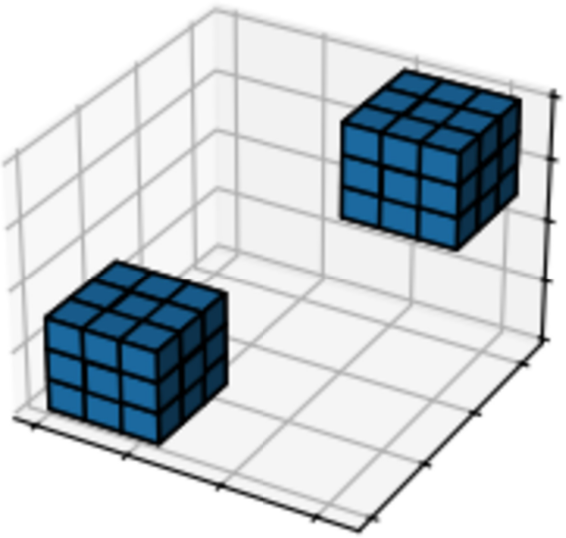
```
import matplotlib.pyplot as plt
import numpy as np

from matplotlib import cm

plt.style.use('_mpl-gallery')

n_radii = 8
n_angles = 36

# Make radii and angles spaces
radii = np.linspace(0.125, 1.0, n_radii)
angles = np.linspace(0, 2*np.pi, n_angles, endpoint=False)[..., np.newaxis]

# Convert polar (radii, angles) coords to cartesian (x, y) coords.
x = np.append(0, (radii*np.cos(angles)).flatten())
y = np.append(0, (radii*np.sin(angles)).flatten())
z = np.sin(-x*y)

# Plot
fig, ax = plt.subplots(subplot_kw={'projection': '3d'})
ax.plot_trisurf(x, y, z, vmin=z.min() * 2, cmap=cm.Blues)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



voxels([x,y,z],filled):

A voxel is a three-dimensional counterpart to a pixel . It prints said voxels in a 3D field.

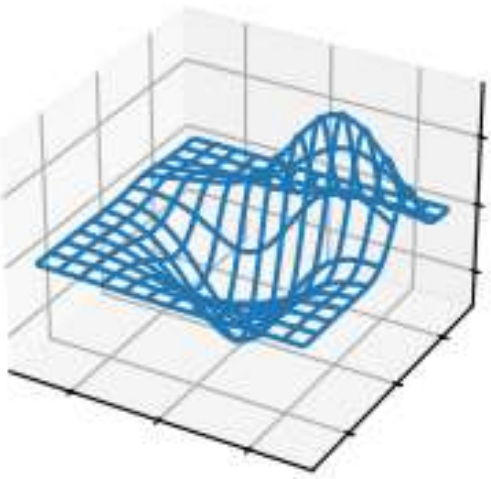
```
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('_mpl-gallery')

# Prepare some coordinates
x, y, z = np.indices((8, 8, 8))

# Draw cuboids in the top left and bottom right corners
cube1 = (x < 3) & (y < 3) & (z < 3)
cube2 = (x >= 5) & (y >= 5) & (z >= 5)

# Combine the objects into a single boolean array
voxelarray = cube1 | cube2

# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.voxels(voxelarray, edgecolor='k')

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



plot_wireframe(X,Y,Z):

Plots the wireframe of a specified surface using grid lines.

```
import matplotlib.pyplot as plt

from mpl_toolkits.mplot3d import axes3d

plt.style.use('_mpl-gallery')

# Make data
X, Y, Z = axes3d.get_test_data(0.05)

# Plot
fig, ax = plt.subplots(subplot_kw={"projection": "3d"})
ax.plot_wireframe(X, Y, Z, rstride=10, cstride=10)

ax.set(xticklabels=[],
       yticklabels=[],
       zticklabels=[])

plt.show()
```



PANDAS:

Pandas is a powerful open-source data analysis and manipulation library for Python. It provides data structures and functions needed to work with structured data seamlessly. The library is particularly well-suited for handling labeled data, such as tables with rows and columns, making it a staple in the data science community.

Key Features for Data Visualization with Pandas:

Pandas offers several features that make it a great choice for data visualization:

1) **Variety of Plot Types**: Pandas supports various plot types, including line plots, bar plots, histograms, box plots, and scatter plots, catering to different visualization needs.

2) **Customization**: Users can customize plots by adding titles, labels, and styling, enhancing the readability and aesthetics of the visualizations.

3) **Handling of Missing Data**: Pandas efficiently handles missing data, ensuring that visualizations accurately represent the dataset without errors.

4) **Integration with Matplotlib**: Pandas seamlessly integrates with Matplotlib, allowing users to create a wide range of static, animated, and interactive plots.

To perform basic plotting with Pandas, **we can leverage the built-in plot () method**, which is a wrapper around Matplotlib's plotting functions. You can also just call df.plot(kind='hist') or replace that kind argument with any of the key terms shown in the list above (e.g. 'box', 'barh', etc). Let us take a look at.

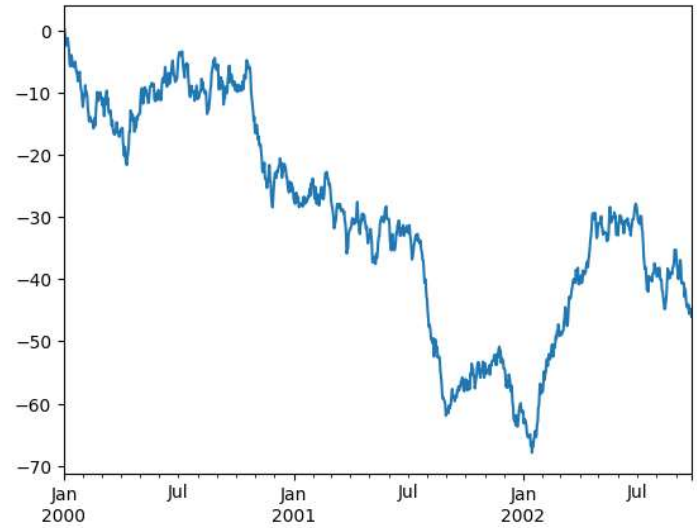
There are a total of 11 plots in pandas let us take a look at these plots.

PLOT:

The “plot” method on Series and DataFrame is just a simple wrapper around “plt.
Let us look at a simple line plot in pandas:

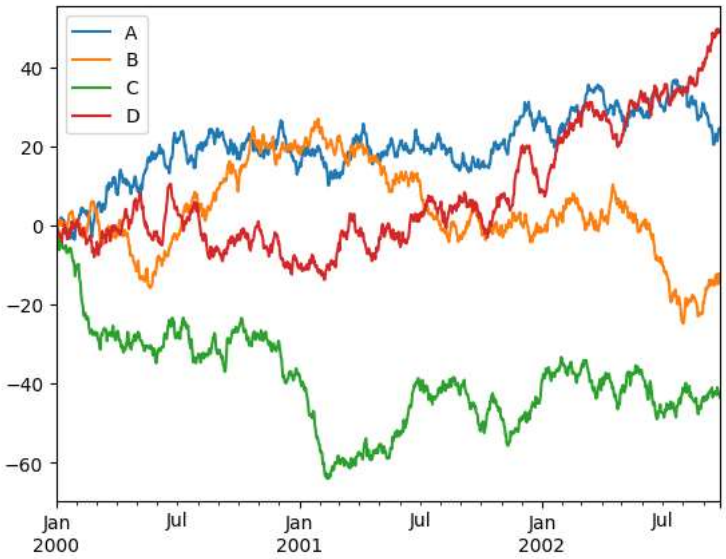
SINGLE LINE:

```
import pandas as pd
import numpy as np
np.random.seed(123456)
ts = pd.Series(np.random.randn(1000), index=pd.date_range("1/1/2000", periods=1000))
ts = ts.cumsum()
ts.plot();
```



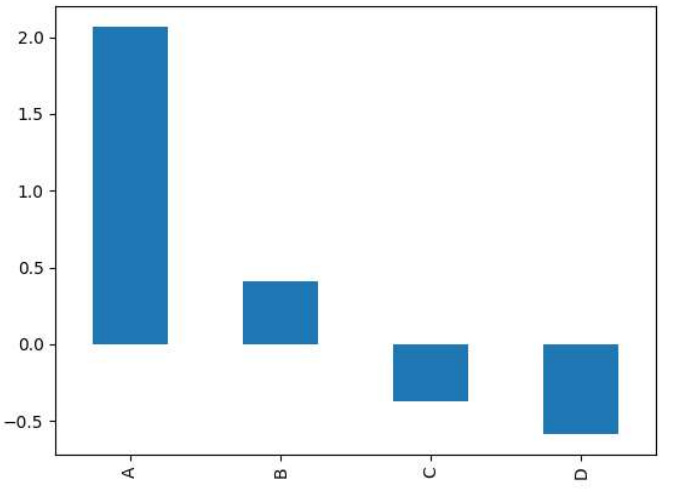
MULTI-LINE:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
df = pd.DataFrame(np.random.randn(1000, 4), index=ts.index, columns=list("ABCD"))
df = df.cumsum()
plt.figure();
df.plot();
```



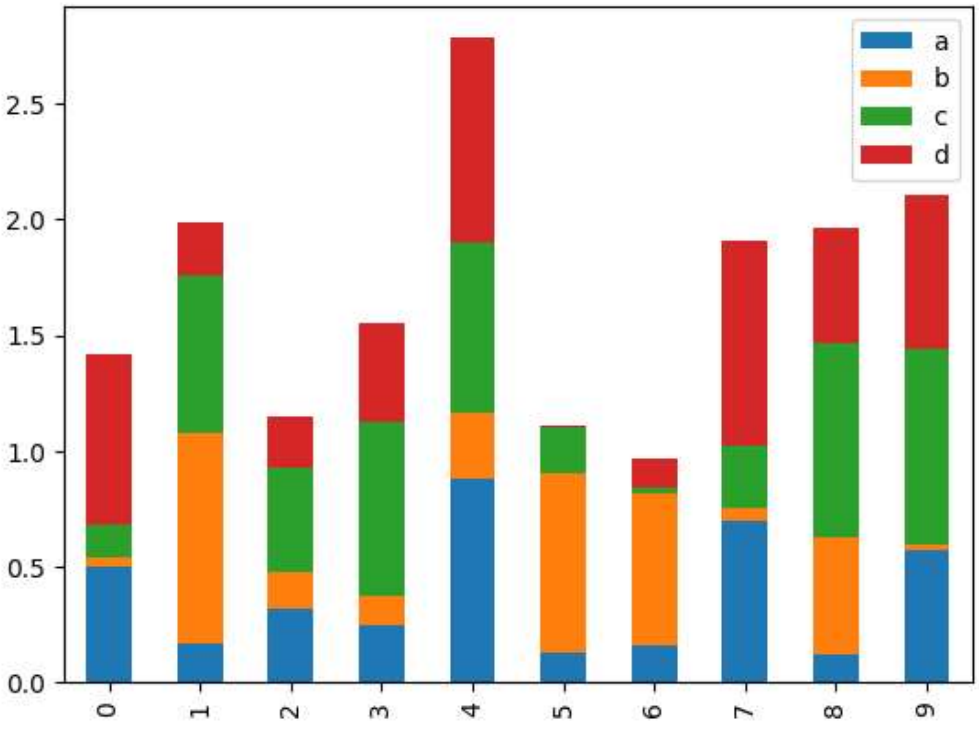
Barplot:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
df = pd.DataFrame(np.random.randn(1000, 4), index=ts.index, columns=list("ABCD"))
plt.figure();
df.iloc[5].plot(kind="bar");
```



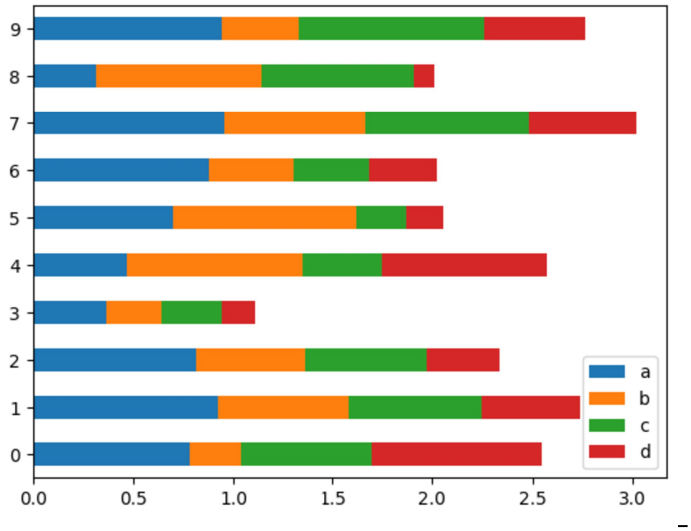
Stacked barplot:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
df2 = pd.DataFrame(np.random.rand(10, 4), columns=["a", "b", "c", "d"])
df2.plot.bar(stacked=True);
```



Stacked Horizontal barplot:

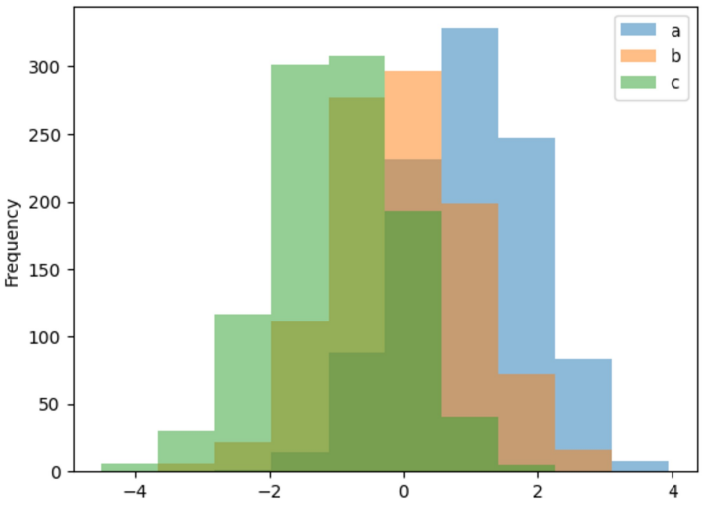
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
df2 = pd.DataFrame(np.random.rand(10, 4), columns=["a", "b", "c", "d"])
df2.plot.barh(stacked=True);
```



Histogram:

A histogram is a type of chart that shows the frequency distribution of data points across a continuous range of numerical values. The values are grouped into bin or buckets that are arranged in consecutive order along the horizontal x-axis at the bottom of the chart. Each bin is represented by a vertical bar that sits on the x-axis and extends upward to indicate the number of data points within that bin.

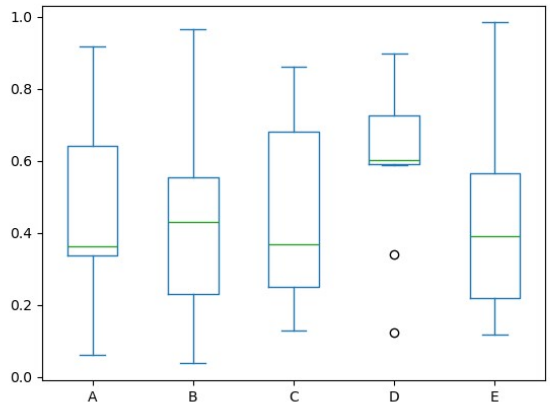
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
df4 = pd.DataFrame(
    {
        "a": np.random.randn(1000) + 1,
        "b": np.random.randn(1000),
        "c": np.random.randn(1000) - 1,
    },
    columns=["a", "b", "c"],
)
plt.figure();
df4.plot.hist(alpha=0.5);
```



Box plot:

Box Plot is a graphical method to visualize data distribution for gaining insights and making informed decisions. Box plot is a type of chart that depicts a group of numerical data through their quartiles.

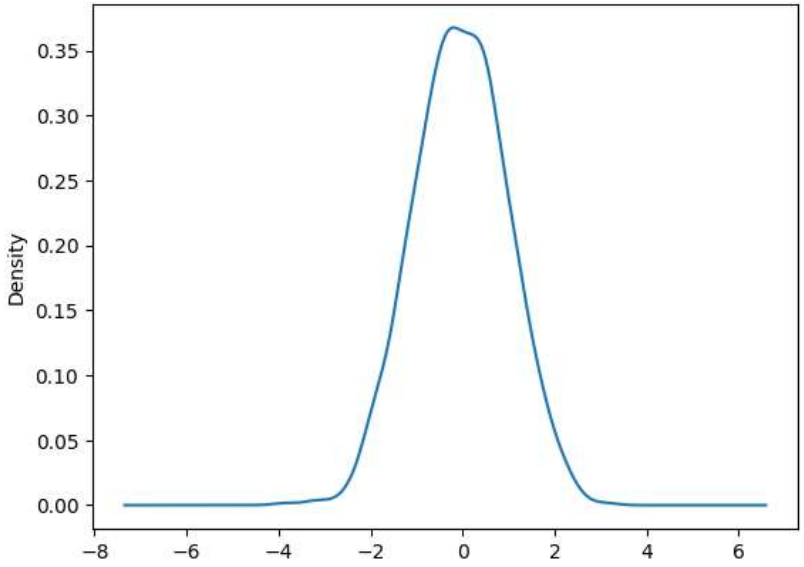
```
import pandas as pd
import numpy as np
df = pd.DataFrame(np.random.rand(10, 5), columns=["A", "B", "C", "D", "E"])
df.plot.box();
```



kde or density plot:

A density plot is a representation of the distribution of a numeric variable. It uses a kernel density estimate to show the probability density function of the variable (see more). It is a smoothed version of the histogram and is used in the same concept.

```
import pandas as pd
import numpy as np
ser = pd.Series(np.random.randn(1000))
ser.plot.kde();
```

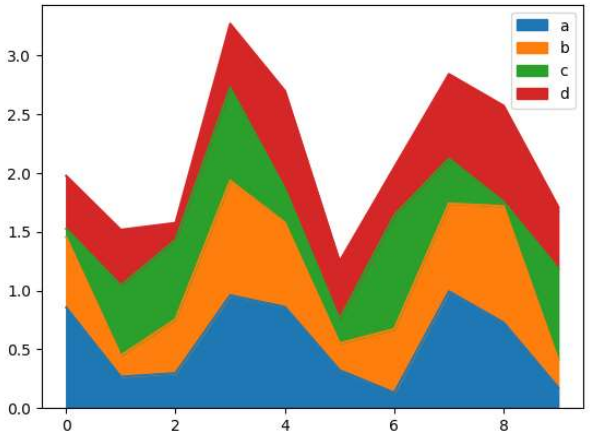


Area Chart:

An area chart or area graph displays graphically quantitative data. It is based on the line chart.

The area between axis and line are commonly emphasized with colors, textures and hatchings. Commonly one compares two or more quantities with an area chart.

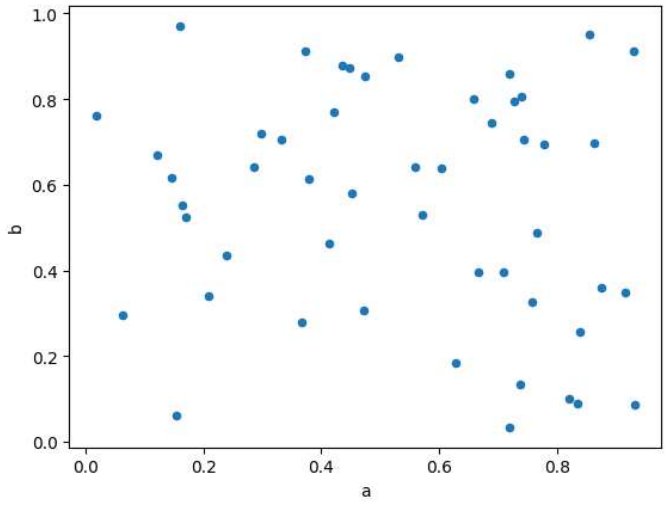
```
import pandas as pd
import numpy as np
df = pd.DataFrame(np.random.rand(10, 4), columns=["a", "b", "c", "d"])
df.plot.area();
```



Scatter Plot:

A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

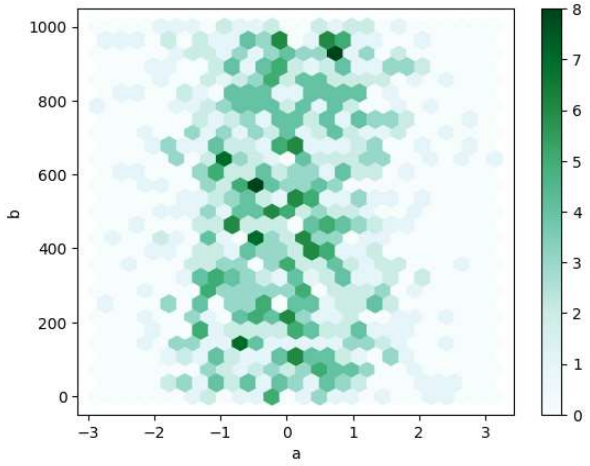
```
import pandas as pd
import numpy as np
df = pd.DataFrame(np.random.rand(50, 4), columns=["a", "b", "c", "d"])
df["species"] = pd.Categorical(
    ["setosa"] * 20 + ["versicolor"] * 20 + ["virginica"] * 10
)
df.plot.scatter(x="a", y="b");
```



Hexagonal Bin Plot:

A hexagonal bin plot is a way to visualize data by grouping points into hexagonal bins and coloring the bins based on the number of points in each bin.

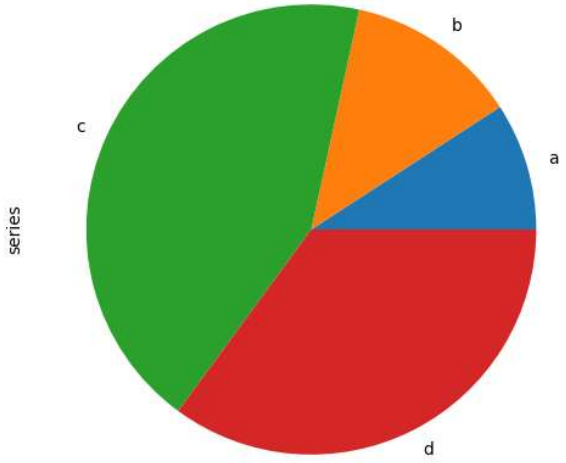
```
import pandas as pd
import numpy as np
df = pd.DataFrame(np.random.rand(50, 4), columns=["a", "b", "c", "d"])
df = pd.DataFrame(np.random.randn(1000, 2), columns=["a", "b"])
df["b"] = df["b"] + np.arange(1000)
df.plot.hexbin(x="a", y="b", gridsize=25);
```



Pie Chart:

A pie chart is a type of graph representing data in a circular form, with each slice of the circle representing a fraction or proportionate part of the whole. All slices of the pie add up to make the whole equaling 100 percent and 360 degrees.

```
import pandas as pd
import numpy as np
series = pd.Series(3 * np.random.rand(4), index=["a", "b", "c", "d"], name="series")
series.plot.pie(figsize=(6, 6));
```



MATPLOTLIB: Advantages and Disadvantages

Pros	Cons
Extensive Customization: Highly customizable, allowing detailed control over visual elements (colors, shapes, fonts, etc.).	**Complex Syntax:** Has a steep learning curve, especially for beginners, due to a complex syntax and many configurations.
Wide Usage & Support: Well-documented, widely used, and supported by a large community with many tutorials available.	**Not as High-Level:** Requires more code for basic plots compared to libraries like Seaborn and Plotly.
Static & Publication-Quality Plots: Produces high-quality, static images suitable for publications and presentations.	**Limited Interactivity:** Lacks built-in interactivity; requires additional libraries (e.g., Plotly) for interactive visualizations.
Versatile & Comprehensive: Supports a wide range of plot types, from basic charts to complex subplots and 3D graphics.	**Performance Limitations:** Can be slow when handling large datasets or complex plots.
Integrates Well with Other Libraries: Easily integrates with other Python libraries like Pandas, NumPy, and SciPy, making it versatile for various data science tasks.	**Outdated Aesthetic:** Default styles may look less modern, although styles can be customized with additional effort.
Cross-Platform Support: Works well across different operating systems and is compatible with Jupyter Notebooks.	**Overlapping Elements:** Requires manual adjustments to prevent overlapping elements, such as tick labels and legends.

PANDAS: Advantages and Disadvantages

Pros	Cons
Easy Data Handling: Simplifies data manipulation with powerful data structures (DataFrames, Series).	**Memory Intensive:** Consumes significant memory, which can limit handling of very large datasets.
Data Cleaning Tools: Offers robust tools for data cleaning, filtering, and aggregation.	**Performance Limitations with Large Data:** Not optimized for very large datasets, impacting speed and efficiency.
Integration with Other Libraries: Works seamlessly with libraries like NumPy, Matplotlib, and SciPy for data analysis and visualization.	**Learning Curve for Complex Operations:** Advanced operations and method chaining can be challenging for beginners.
Rich I/O Support: Supports reading and writing to multiple file formats (CSV, Excel, SQL, JSON, etc.).	**Limited Parallelism:** Lacks optimized multi-threaded processing, which can reduce efficiency for data-heavy tasks.
Built-In Time Series Support: Provides excellent functionality for time-based indexing and analysis.	**Performance Bottlenecks:** Certain operations can be inefficient, especially with row-wise operations.

Comparison Between Pandas and Matplotlib:

Feature	Pandas	Matplotlib
Primary Purpose	Data manipulation and analysis	Data visualization and plotting
Data Structure	Uses DataFrames and Series for data handling	Uses plots and figures for visual representation
Ease of Use	Simple for basic data tasks; complex operations have a steeper learning curve	Moderate learning curve for customization; extensive configuration options
Integration	Integrates well with libraries like NumPy and Matplotlib for visualization	Integrates with Pandas for plotting DataFrames directly
Output	Primarily tabular or text-based data representations	Creates static, publication-quality visuals suitable for presentations

Pandas Applications:

1. **Data Cleaning and Preprocessing**: Handling missing values, duplications, and transforming data for analysis.
2. **Data Aggregation and Grouping**: Summarizing data through grouping, filtering, and aggregation functions.
3. **Time Series Analysis**: Processing and analyzing time-stamped data for trend analysis and forecasting.
4. **Data Merging and Joining**: Combining datasets from multiple sources using merging and joining techniques.
5. **Exploratory Data Analysis (EDA)**: Generating summary statistics and quick insights from data before detailed analysis.

Matplotlib Applications:

1. **Data Visualization**: Creating a wide range of charts (line, bar, histogram, scatter plots) to represent data visually.
2. **Trend Analysis**: Plotting data over time to identify patterns, trends, and seasonal variations.
3. **Statistical Analysis**: Visualizing statistical distributions and relationships in data (box plots, histograms).
4. **Presentation Graphics**: Generating high-quality, publication-ready visuals for research papers, presentations, and reports.
5. **Geospatial Visualization**: Creating basic geographic maps by plotting data points with Matplotlib's `basemap` or using coordinates.

Semantic analysis of Twitter data

OBJECTIVES:

- Understanding data preprocessing of text data
- Application of data vectorization using TfidfVectorizer
- Training model based of vectorized text
- Savingt the data of the model in a .sav file using pickle

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from collections import Counter
import numpy as np
from sklearn.model_selection import train_test_split
```

```
import nltk
nltk.download('stopwords')
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourse
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
#Loading data from the csv file using pandas
X_data=pd.read_csv('/content/X data.csv',encoding='ISO-8859-1')
```

```
#Checking number of rows and columns
X_data.shape
```

```
(162980, 2)
```

```
#Printing firt 5 rows of data
X_data.head()
```

	clean_text	category
0	when modi promised â□□minimum government maxim...	-1.0
1	talk all the nonsense and continue all the dra...	0.0
2	what did just say vote for modi welcome bjp t...	1.0
3	asking his supporters prefix chowkidar their n...	1.0
4	answer who among these the most powerful world...	1.0

```
#Naming columns and reading data sets again
COL_NAMES=['TEXT','FLAG']
X_data=pd.read_csv('/content/X data.csv',encoding='ISO-8859-1',names=COL_NAMES)
X_data.head()
```

	TEXT	FLAG
0	clean_text	category
1	when modi promised â□□minimum government maxim...	-1
2	talk all the nonsense and continue all the dra...	0
3	what did just say vote for modi welcome bjp t...	1
4	asking his supporters prefix chowkidar their n...	1

```
#counting number of missing values in X_data
X_data.isnull().sum()
```

	0
TEXT	4
FLAG	7

dtype: int64

```
#Filling null values with -1
X_data=X_data.fillna(-1)
X_data['TEXT'].replace(to_replace="-1",value="Modi will win")
```

	TEXT
0	clean_text
1	when modi promised â□□minimum government maxim...
2	talk all the nonsense and continue all the dra...
3	what did just say vote for modi welcome bjp t...
4	asking his supporters prefix chowkidar their n...
...	...
162976	why these 456 crores paid neerav modi not reco...
162977	dear rss terrorist payal gawar what about modi...
162978	did you cover her interaction forum where she ...
162979	there big project came into india modi dream p...
162980	have you ever listen about like gurukul where ...

162981 rows × 1 columns

dtype: object

```
#checking the distributions of target column
X_data['TEXT'].value_counts()
```

	count
TEXT	
-1	4
2019	2
clean_text	1
should vote modi for cpas after years	1
lok sabha election 2019 live modi has ignored his own constituency varanasi says priyanka gandhi	1
...	...
modi destroying india for personal benefit	1
sreeniwho announced buddha laughing nuclear testjust for sake dont frowl before knowing the achievementthis not achievement modiits nations achievementmake your heart bit enlarged	1
modi thunders indiaâ□□ entry into indiaâ□□ space club raga calls â□□happy theatreÂ dayâ□□	1
back basics jobs\nfarmers\nsmall businesses\ngst reform\neducation\nhealth\nenvironment\nwater\ninfrastructure\ninvestment revival national security indiaâ□□ armed forces are handling that donâ□□ worry space narendra modi thanks	1
have you ever listen about like gurukul where discipline are maintained even narendra modi rss only maintaining the culture indian more attack politics but someone attack hinduism rss will take action that proud for	1

162977 rows × 1 columns

dtype: int64

```
X_data['FLAG'].value_counts()
```

FLAG	count
1	72250
0	55213
-1	35510
-1	7
category	1

dtype: int64

0 -->neutral tweet 1 -->positive tweet -1 --> negative tweet

Stemming is the process of reducing a word to its root form i.e Swimming to swim. We do this using the porter stemmer function

```
port_stem=PorterStemmer()

def stemming(content):
    # Convert content to string to handle non-string values
    content = str(content)
    stemmed_content=re.sub('[^a-zA-Z]',' ',content)
    stemmed_content=stemmed_content.lower()
    stemmed_content=stemmed_content.split()
    # Correct variable name from stemmed_conted to stemmed_content
    stemmed_content=[port_stem.stem(word) for word in stemmed_content if not word in stopwords.words('english')]
    stemmed_content=' '.join(stemmed_content)

    return stemmed_content


X_data['STEM_TEXT']=X_data['TEXT'].apply(stemming) #about 7mins to complete this execution!!! SOOOOO LOOOONG!!!!!


#Viewing new data
X_data.head()
```

	TEXT	FLAG	STEM_TEXT
0	clean_text	category	clean text
1	when modi promised â□□minimum government maxim...	-1	modi promis minimum govern maximum govern expe...
2	talk all the nonsense and continue all the dra...	0	talk nonsens continu drama vote modi
3	what did just say vote for modi welcome bjp t...	1	say vote modi welcom bjp told rahul main campa...
4	asking his supporters prefix chowkidar their n...	1	ask support prefix chowkidar name modi great s...

```
print(X_data['FLAG'])
```

```
0         category
1               -1
2                0
3                1
4                1
              ...
162976          -1
162977          -1
162978           0
162979           0
162980           1
Name: FLAG, Length: 162981, dtype: object
```

```
#seperating data and label
X=X_data['STEM_TEXT'].values
Y=X_data['FLAG'].values.astype(str)


print(X)
```

```
['clean text'
 'modi promis minimum govern maximum govern expect begin difficult job reform state take year get justic state busi exit psu templ'
 'talk nonsens continu drama vote modi' ... 'cover interact forum left'
 'big project came india modi dream project happen realiti'
 'ever listen like gurukul disciplin maintain even narendra modi rss maintain cultur indian attack polit someon attack hinduism rss
```

```
print(Y)
```

```
['category' '-1' '0' ... '0' '0' '1']
```

```
# Calculate class distribution
class_distribution = Counter(Y)

# Find classes with only one sample
classes_to_remove = [cls for cls, count in class_distribution.items() if count < 2]

# Remove samples belonging to the under-represented classes
mask = ~np.isin(Y, classes_to_remove)
X = X[mask]
Y = Y[mask]


#SPLITTING DATA TO TRAIN AND TEST
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,stratify=Y,random_state=2)


print(X.shape,Y_train.shape,X_test.shape)
```

```
(162980,) (130384,) (32596,)
```

```
print(X_train)
```

```
['kaha tha nachoron sarkarscamist parti hai modi ruin countri dont forget vote'
 'chines citizen power right speak'
 'missil defenc india strengthen modi govern' ...
 'pm narasimha rao communist lack vision like communist includ shastri moraji desai today due vajpaye narendra modi other lack visic
 'smita prakash modi fangirl would modi without ani would ani without modi'
 'lok sabha elect campaign live make sure rahul defeat say prakash karat financialxpress']
```

```
print(X_test)
```

```
['modi anti nation lie creat commun divid mislead manipul data'
 'jaya pradha join bjp convass goten shorten amount also popular bjp big rich leader bythat famou cine field enjoy modi also'
 'power popular leadership modi one tweet whole countri goe desper' ... ''
 'modi address massiv ralli kurnool andhra pradesh via namo app'
 'watch video voic whatsapp section volunt modul narendra modi app']
```

```
#Converting textual Data to numerical data
vectorizer=TfidfVectorizer(lowercase=False)
# Fit and transform the original text data 'X_text' (assuming this is your original text data variable)
X = vectorizer.fit_transform(X_text)
# Now transform the training and testing sets using the fitted vectorizer
X_train=vectorizer.transform(X_train)
X_test=vectorizer.transform(X_test)



print(X_train)
```

```
  (0, 5877)     0.23211504791694398
  (0, 7538)     0.23548945446948935
  (0, 9418)     0.354204268832559
  (0, 10840)    0.2753084460429206
  (0, 13783)    0.47083934854041315
  (0, 16942)    0.06500689091180703
  (0, 19727)    0.25053561795816454
  (0, 22855)    0.4318397312580794
  (0, 26449)    0.4087484798215461
  (0, 28611)    0.2171804773161407
  (1, 4756)     0.5918469474568181
  (1, 5044)     0.4551164112905416
  (1, 20743)    0.3366327029443095
  (1, 22610)    0.3759564931415472
  (1, 25003)    0.4334989983070124
  (2, 6582)     0.5300020214201105
  (2, 10358)    0.35074894432798903
  (2, 12299)    0.23366050593559784
  (2, 16809)    0.43637908746264226
  (2, 16942)    0.0916516983467352
  (2, 25434)    0.5853623335377406
  (3, 405)      0.2523482267005551
  (3, 3567)     0.35357562788310454
  (3, 8468)     0.2191756447841152
  (3, 8638)     0.5081332140365633
  :     :
  (130381, 21875)       0.23022080815445803
  (130381, 24083)       0.25638451452365046
  (130381, 26956)       0.13746263513427678
  (130381, 28173)       0.20083845408518594
  (130381, 28539)       0.3585143380350224
```

```
(130382, 1140)        0.5902580758686817
(130382, 8875)        0.37522459976450834
(130382, 16942)       0.12281166280105922
(130382, 20820)       0.31199308145928956
(130382, 24710)       0.3647377075919171
(130382, 29193)       0.3930048745543684
(130382, 29344)       0.33300614024675346
(130383, 4076)        0.22549246814283908
(130383, 6574)        0.2608135213898228
(130383, 8051)        0.16215926088637278
(130383, 9165)        0.43666985999467206
(130383, 13931)       0.4285314472717363
(130383, 15275)       0.21640635333884672
(130383, 15360)       0.25961397199213104
(130383, 15839)       0.18627600953805795
(130383, 20820)       0.38204335439728315
(130383, 21659)       0.18595018555306403
(130383, 22961)       0.2584458163033459
(130383, 23449)       0.16610043871510108
(130383, 25794)       0.23804619772332206
```

```
print(X_test)
```

```
  (0, 1230)     0.26958717265727083
  (0, 5360)     0.3121892578044157
  (0, 6004)     0.2877293106675747
  (0, 6386)     0.3445650741673561
  (0, 7407)     0.3619912275820138
  (0, 15163)    0.2895043909183832
  (0, 16023)    0.44650548079773944
  (0, 16789)    0.41099085054222745
  (0, 16942)    0.05972087114921265
  (0, 18019)    0.20408265440005308
  (1, 909)      0.24233964199864133
  (1, 1018)     0.1940119208092101
  (1, 3142)     0.14963154522697322
  (1, 3273)     0.19966985781677263
  (1, 4013)     0.3035493266924947
  (1, 5018)     0.3035493266924947
  (1, 5740)     0.3035493266924947
  (1, 8257)     0.18786060786814415
  (1, 8859)     0.2158496436901553
  (1, 9107)     0.2003429948755047
  (1, 10342)    0.3035493266924947
  (1, 13257)    0.24283093573712491
  (1, 13578)    0.16215342950663475
  (1, 14945)    0.1276272236408723
  (1, 16942)    0.03186232085438606
  :     :
  (32592, 23569)        0.39236097060969066
  (32592, 23918)        0.39236097060969066
  (32592, 24964)        0.16872467473369665
  (32592, 25723)        0.22384292133048414
  (32592, 26234)        0.21329268036979332
  (32594, 405)  0.2572984313499474
  (32594, 1106) 0.3839118348072669
  (32594, 1396) 0.31007248344097077
  (32594, 14649)        0.45649845172633596
  (32594, 16203)        0.3686756673788957
  (32594, 16942)        0.05947675840780109
  (32594, 17905)        0.28432703479309335
  (32594, 20803)        0.3451404640437686
  (32594, 21798)        0.28523238717232124
  (32594, 28371)        0.24716975759466478
  (32595, 1396) 0.27895174324173294
  (32595, 16942)        0.053507313051794896
  (32595, 17232)        0.4767332578224748
  (32595, 17969)        0.1762845855602194
  (32595, 23700)        0.3469172423358947
  (32595, 28397)        0.27296612439328555
  (32595, 28589)        0.3263140040976818
  (32595, 28601)        0.39772489655740606
  (32595, 28827)        0.24696891876614663
  (32595, 29002)        0.37476711518967865
```

```
#Training the Machine learning model
model=LogisticRegression()
model.fit(X_train,Y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:469: ConvergenceWarning: lbfgs failed to converge (status=
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

```
#Model Evaluation accurace on traininf data
X_train_prediction  = model.predict(X_train)
training_data_accuracy=accuracy_score(Y_train,X_train_prediction)
```

```
print('Accuracy score on the training Data:',training_data_accuracy)
```

```
Accuracy score on the training Data: 0.8684807951895939
```

Testing Data Accuracy = 86.84%

```
#Model Evaluation accurace on training data
X_test_prediction  = model.predict(X_test)
test_data_accuracy=accuracy_score(Y_test,X_test_prediction)
```

```
print('Accuracy score on the training Data:',test_data_accuracy)
```

```
Accuracy score on the training Data: 0.8392747576389741
```

Model Accuracy= 83.92%

We are saving this model to be used for later to use model directily without having to train multiple times! using the pickle library!!!

```
import pickle
```

```
#Saving the pretrained model to a .sav file using Pickle moduld
filename='trained_model.sav'
pickle.dump(model,open(filename,'wb'))
```

Sales data Analysis

OBJECTIVES:

- importing zipdata and creating a dataframe
- figuring out optimal data visualization formats to visualize data
- Applying Data Visualization
- Creating interactive dashboard or tables and charts to better explain dataset

```
import pandas as pd
import matplotlib.pyplot as plt  # Corrected import statement
import numpy as np
import seaborn as sns

Sales_data = pd.read_csv('/content/Sales data/retail_sales_dataset.csv')
```

```
Sales_data.head()
```

	Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount
0	1	2023-11-24	CUST001	Male	34	Beauty	3	50	150
1	2	2023-02-27	CUST002	Female	26	Clothing	2	500	1000
2	3	2023-01-13	CUST003	Male	50	Electronics	1	30	30
3	4	2023-05-21	CUST004	Male	37	Clothing	1	500	500
4	5	2023-05-06	CUST005	Male	30	Beauty	2	50	100

Next steps: Generate code with `Sales_data` View recommended plots New interactive sheet

Checking for null values

```
Sales_data.isna().sum()
```

	0
Transaction ID	0
Date	0
Customer ID	0
Gender	0
Age	0
Product Category	0
Quantity	0
Price per Unit	0
Total Amount	0

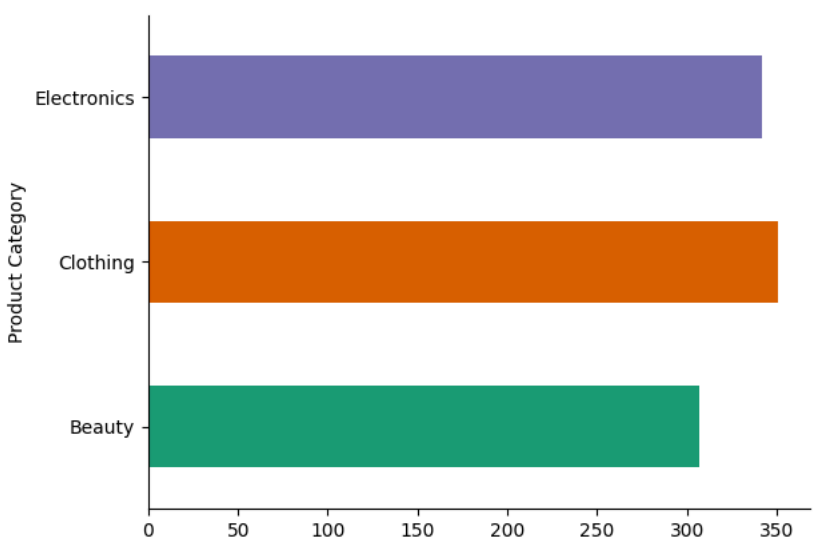
dtype: int64

```
Table has no null values so we can move ahead to data visualization

**Checking what category of product sells the most**
```

Table has no null values so we can move ahead to data visualization

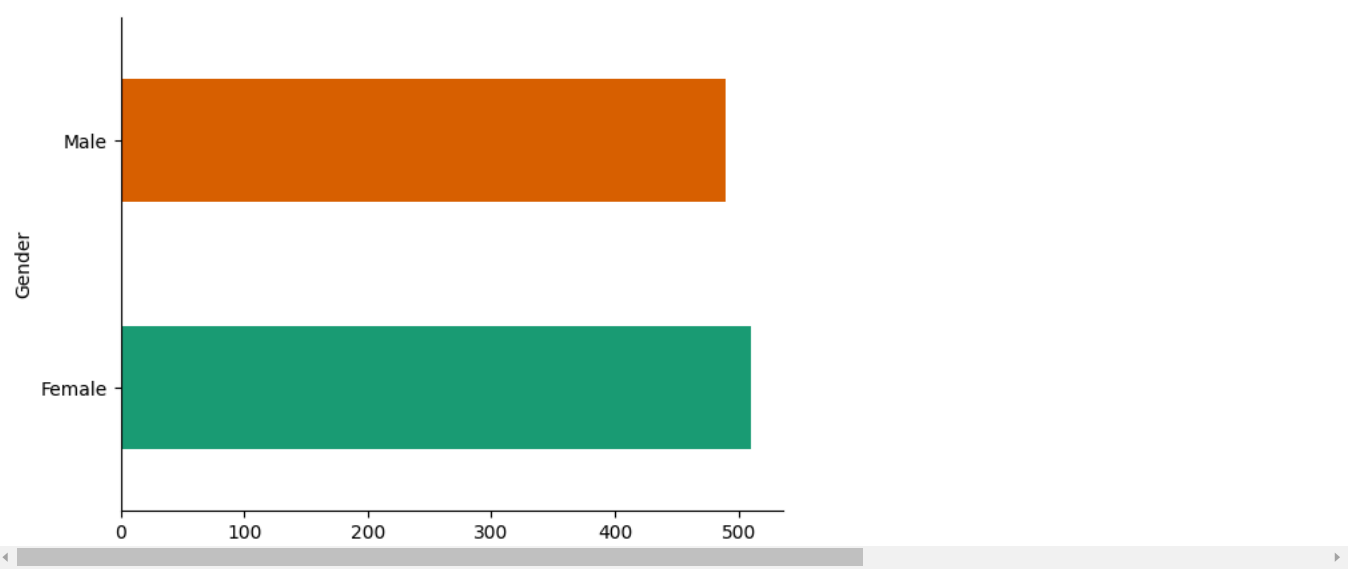
Checking what category of product sells the most

```
Sales_data.groupby('Product Category').size().plot(kind='barh', color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right',]].set_visible(False)
```



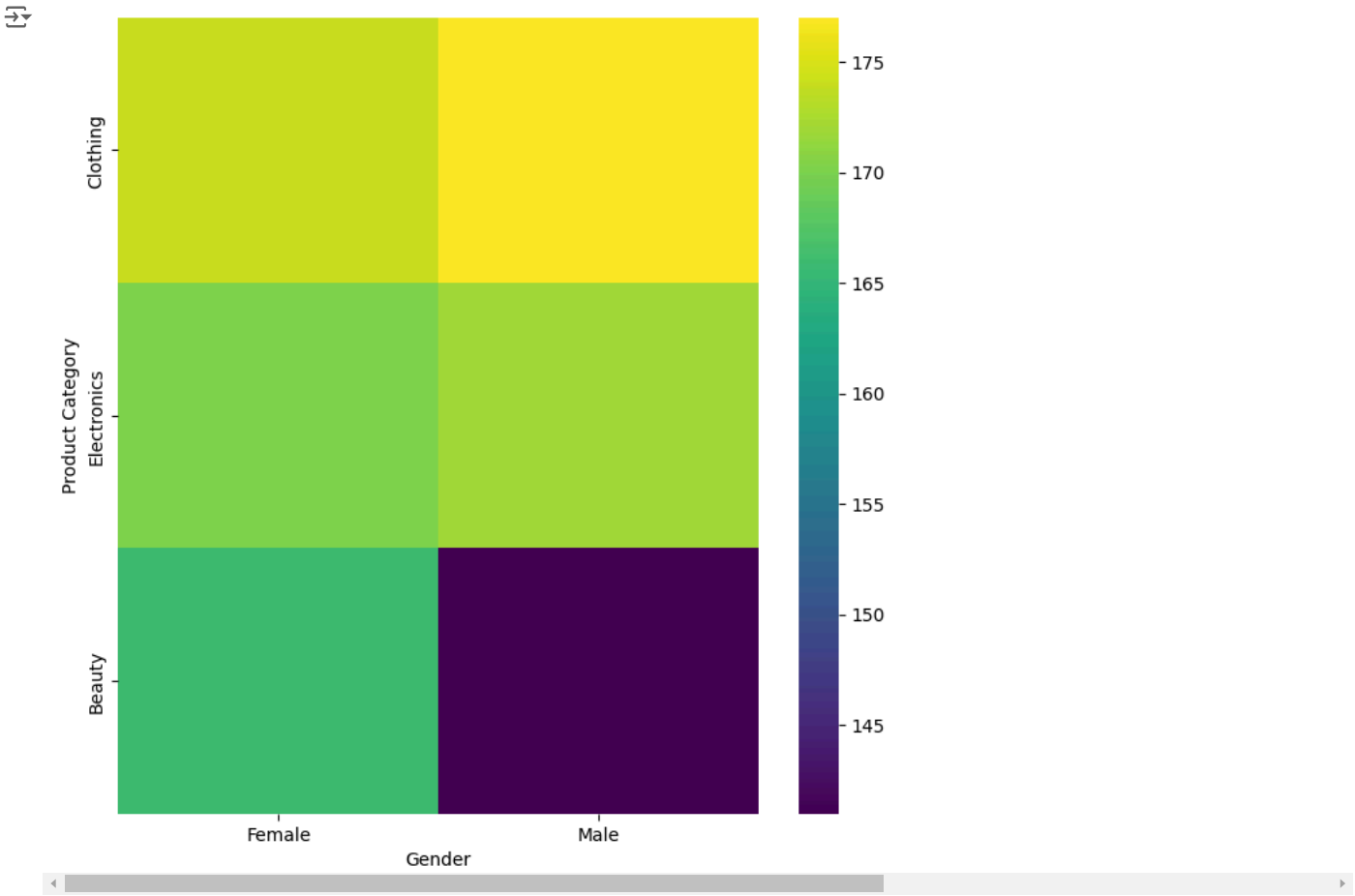
Checking Which Gender Buys the most products

```
Sales_data.groupby('Gender').size().plot(kind='barh', color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right',]].set_visible(False)
```



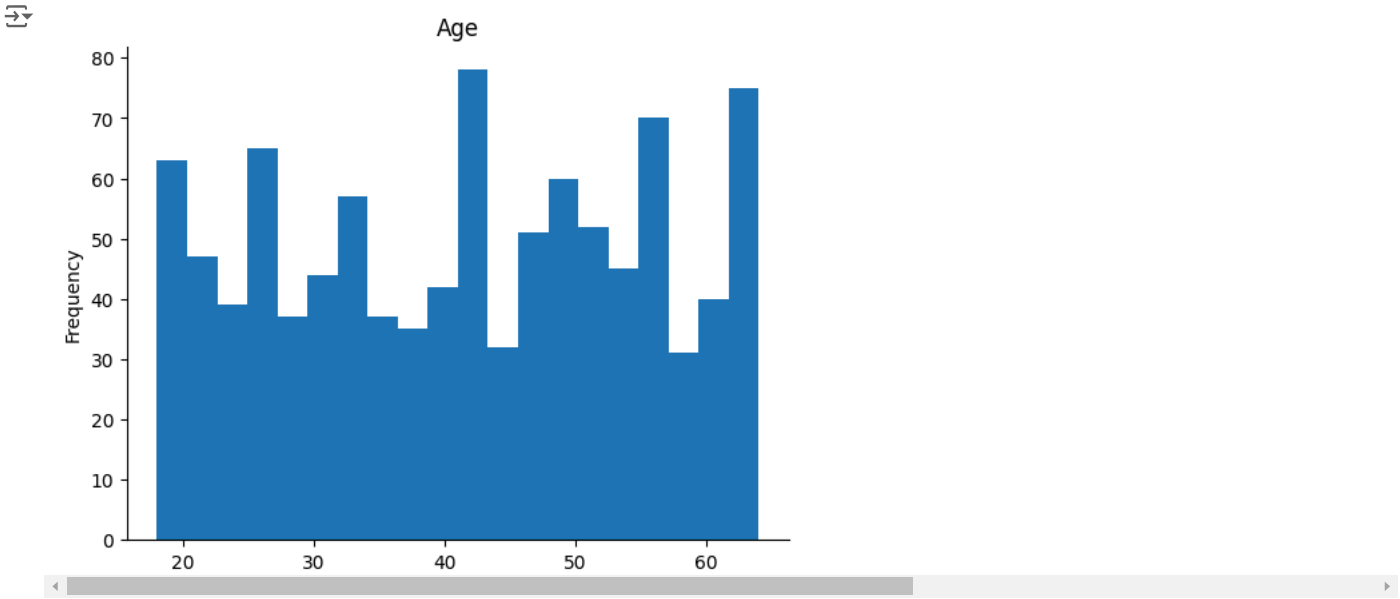
Mapping Sales product category for each Gender

```
plt.subplots(figsize=(8, 8))
df_2dhist = pd.DataFrame({
    x_label: grp['Product Category'].value_counts()
    for x_label, grp in Sales_data.groupby('Gender')
})
sns.heatmap(df_2dhist, cmap='viridis')
plt.xlabel('Gender')
_ = plt.ylabel('Product Category')
```



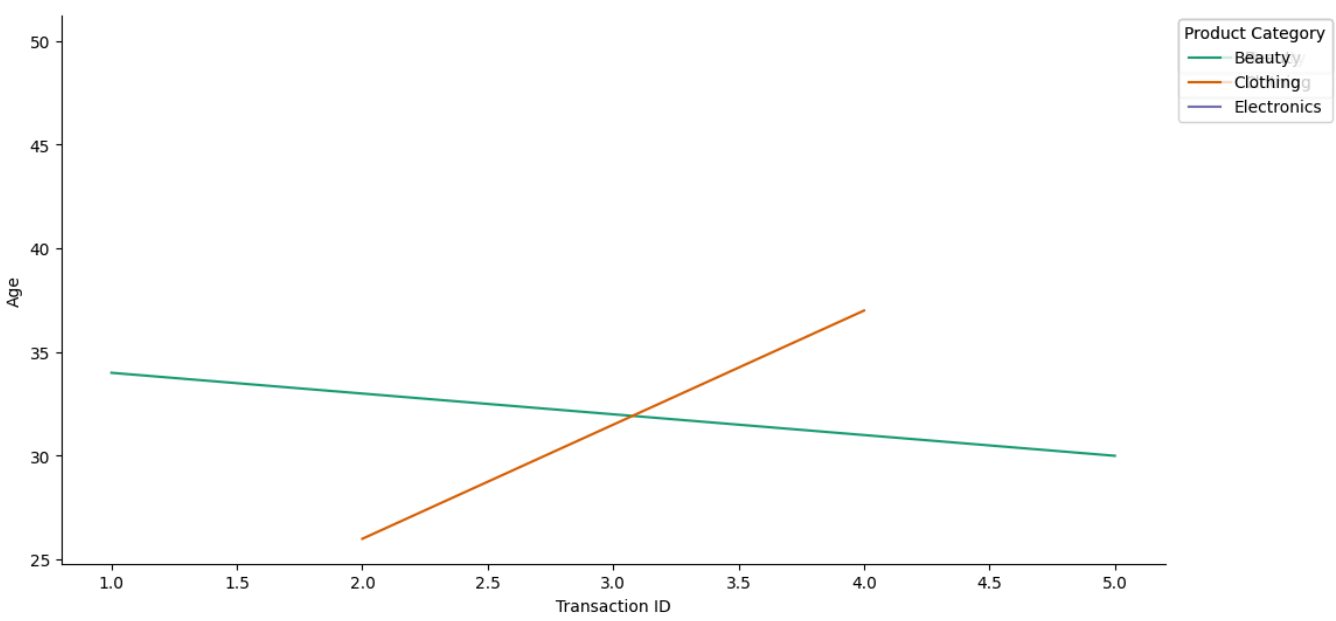
Age vs Buying Frequency

```
Sales_data['Age'].plot(kind='hist', bins=20, title='Age')
plt.gca().spines[['top', 'right',]].set_visible(False)
```



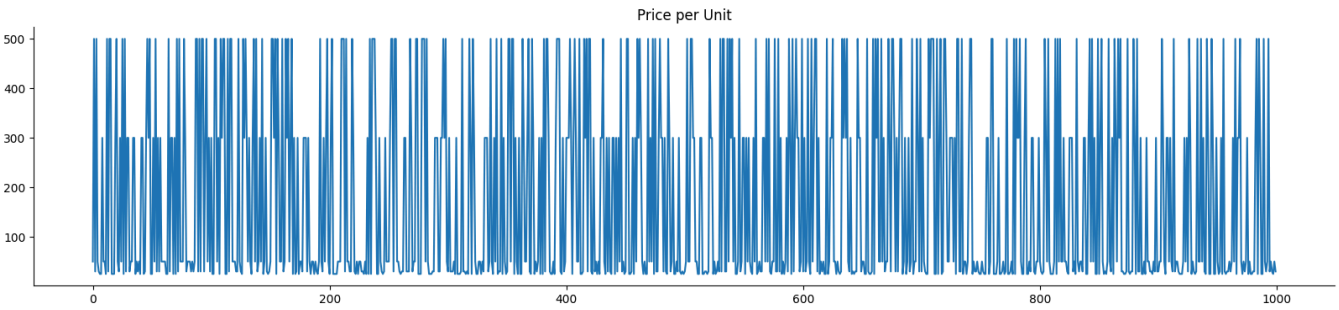

```
def _plot_series(series, series_name, series_index=0):
  palette = list(sns.palettes.mpl_palette('Dark2'))
  xs = series['Transaction ID']
  ys = series['Age']

  plt.plot(xs, ys, label=series_name, color=palette[series_index % len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = _df_42.sort_values('Transaction ID', ascending=True)
for i, (series_name, series) in enumerate(df_sorted.groupby('Product Category')):
  _plot_series(series, series_name, i)
  fig.legend(title='Product Category', bbox_to_anchor=(1, 1), loc='upper left')
sns.despine(fig=fig, ax=ax)
```

```
plt.xlabel('Transaction ID')
_ = plt.ylabel('Age')
```



Price per Unit of Products

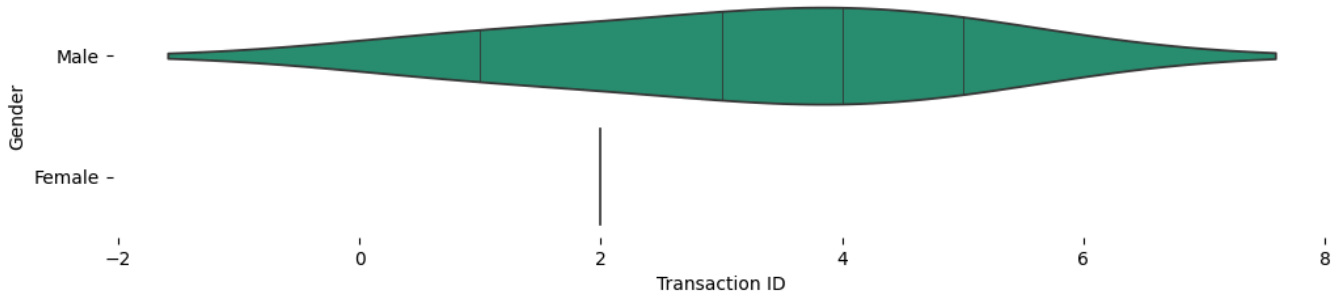
```
Sales_data['Price per Unit'].plot(kind='line', figsize=(20, 4), title='Price per Unit')
plt.gca().spines[['top', 'right']].set_visible(False)
```



Gender Vs Transaction ID

```
figsize = (12, 1.2 * len(_df_52['Gender'].unique()))
plt.figure(figsize=figsize)
sns.violinplot(_df_52, x='Transaction ID', y='Gender', inner='stick', palette='Dark2')
sns.despine(top=True, right=True, bottom=True, left=True)
```

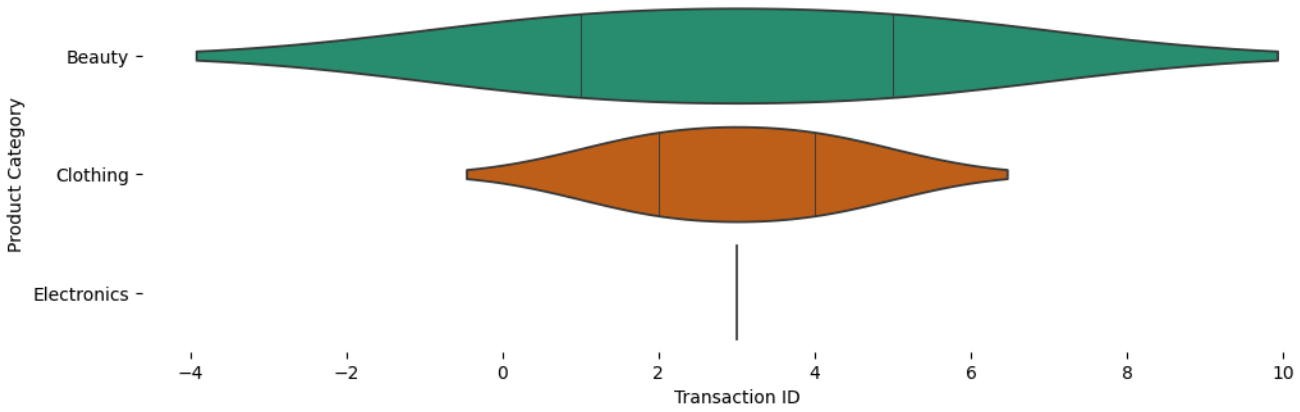
```
<ipython-input-155-25025927d693>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `le

  sns.violinplot(_df_52, x='Transaction ID', y='Gender', inner='stick', palette='Dark2')
```



Transaction ID vs Product Category

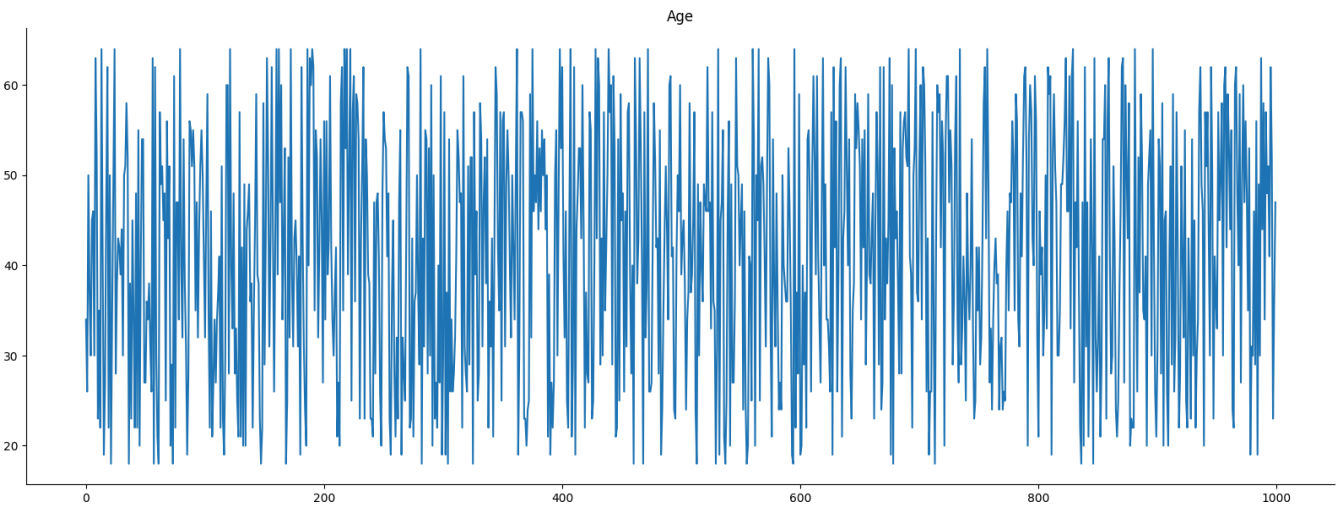
```
figsize = (12, 1.2 * len(_df_53['Product Category'].unique()))
plt.figure(figsize=figsize)
sns.violinplot(_df_53, x='Transaction ID', y='Product Category', inner='stick', palette='Dark2')
sns.despine(top=True, right=True, bottom=True, left=True)
```

```
<ipython-input-156-89f27e2f743b>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `le

  sns.violinplot(_df_53, x='Transaction ID', y='Product Category', inner='stick', palette='Dark2')
```



Age vs Quantity Plot

```
from matplotlib import pyplot as plt
Sales_data['Age'].plot(kind='line', figsize=(20, 7), title='Age')
plt.gca().spines[['top', 'right']].set_visible(False)
```



TRANSACTION ID VS AGE

Transaction ID vs Age

```
# @title Transaction ID vs Age

from matplotlib import pyplot as plt
import seaborn as sns
def _plot_series(series, series_name, series_index=0):
  palette = list(sns.palettes.mpl_palette('Dark2'))
  xs = series['Transaction ID']
  ys = series['Age']

  plt.plot(xs, ys, label=series_name, color=palette[series_index % len(palette)])
```

```
  plt.plot(xs, ys, label=series_name, color=palette[series_index % len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = Sales_data.sort_values('Transaction ID', ascending=True)
for i, (series_name, series) in enumerate(df_sorted.groupby('Gender')):
  _plot_series(series, series_name, i)
  fig.legend(title='Gender', bbox_to_anchor=(1, 1), loc='upper left')
sns.despine(fig=fig, ax=ax)
plt.xlabel('Transaction ID')
_ = plt.ylabel('Age')
```