```python
import pandas as pd
data = pd.read_csv('data.csv')
   print(data.info())
   print(data.shape)
     (11.2)
     print(data.size)
       22
     print(data.ndim)
       2
```

Graph plot:
```python
   import matplotlib.pyplot as plt.
   plt.plot(data['year'],data['No. of bag sold'])
   plt.show()
```

Bar Plot:
```python
   import matplotlib.pyplot as plt.
   plt.bar(data['year'], data['No. of bag sold'])
   plt.show()
```

Scatter plot:
```python
   import matplotlib.pyplot as plt.
   plt.scatter(data['year'], data['No. of bag sold'])
   plt.show()
```

Lab1: Warm Up – Familiarity with Data types and Visualization

Objectives
In this lab, you will get familiarity with downloading, reading, printing properties and visualizing datasets. Also, you will work on various Notebooks such as Google Colab and Azure notebooks.

Learning Outcomes
After completing this lab, you will be able to
- Understand various data formats
- Understand various file formats
- Visualize various kinds of data such as text, images, video and audios
- Get familiarity with Google Colab and Azure notebooks

Step 1: Download the dataset files that belong to the following data formats from internet. The files may belong to any dataset available online.

Step 2: Read these files inside the python code. Some of the file formats cannot be read using default python packages. In this case, explore the python packages suitable for reading the files.

Step 3: Print the properties of the data files such as size, shape, dimensions, etc.

Step 4: Visualize each of these data files using graphs, diagrams, etc.
- Table data visualization: line graph, bar graph, histogram chart, pie chart, scatter plot
- Image visualization: image plot, 3d plot
- Video visualization: video player
- Audio visualization: audio player, spectrogram
- Text visualization: Word cloud, bubble cloud (some more in http://vallandingham.me/textvis-talk/)

1. Tabular, Spreadsheet and Interchange Data Formats
- "Table" — generic tabular data (.dat), "CSV" — comma-separated values (.csv), "TSV" — tab-separated values (.tsv), "ARFF" - Attribute-Relation File Format (.arff) – Read and visualize the data
- "XLS" — Excel spreadsheet (.xls), "XLSX" — Excel 2007 format (.xlsx), "ODS" — OpenDocument spreadsheet (.ods), "SXC" — OpenOffice 1.0 spreadsheet file (.sxc), "DIF" — VisiCalc data interchange format (.dif) – Read and visualize the data
- "JSON" — JavaScript Object Notation (.json), "UBJSON" — Universal Binary JSON (.ubj), "HTML" – Hypertext Markup Language (.html), "XML" - eXtensible Markup Language (.xml) - Read and Parse the data

2. Data File Formats
- PKL – Pickle format, HDF5, Zip, SQL, MAT, NPY, NPZ – Read and display the data

3. Image Data Formats
- JPG, PNG, BMP, TIFF – Read and display the image
- 3D medical Images: DICOM, MHA – Read and display the image

4. Video Data Formats
- MP4, AVI, MPEG – Read and play the video

5. Audio Data Formats
- MP3, MIDI, WAV – Read and play the audio

histograms:-

```
import matplotlib.pyplot as plt.
plt.hist (data ['No. of bag sold']
plt.show ()
```

Pie chart:-

```
import matplotlib.pyplot as plt.
import pandas as pd.
df = pd.read_csv ("data.csv")
month_data = df ["Year"]
acc_data = df ["No. of bag sold"]
colors = ["#1f7764", "#bb7bce", "#2ca02c",
          "#d62728", "#8c564b", "#1f7764", "#bf7be"]
explode = (0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1)
plt.pie (acc_data, labels= month_data, explode.
    = explode, colours = colors, cutpit = "%1.1f"
    "%", shadow = True, startangle = 145)
plt.title ("No. of bag sold | n" "t" "in 2019")
plt.show()
```

6. Text Data Formats
- TXT, PDF, DOC – Read and parse the data

Familiarity: Get familiarity with **Google Colab Notebook** and Microsoft **Azure Notebooks**.

Image Visualization:

```
% matplotlib inline
import image io
import matplotlib.pyplot as plt.
import matplotlib. chook
(-0.5, 699.5, 499.5, -0.5)
%. matplotlib notebook
import matplotlib. pyplot as plt.
from mpl-toolkits.mplot 3d import Axes 3D
import numpy as np.
from scipy. stats import multivariate-normal
X = np. linspace (-5,5,50)
Y = np. linspace (-5,5,50)
X, Y = np. mesgrid (X, Y)
```

```python
X_mean = 0; Y_mean = 0
X_var = 5; Y_var = 8
pos = np.empty(x.shape + (2,))
pos[:, :, 0] = x
pos[:, :, 1] = y
rv = multivariate_normal([X_mean, Y_mean],
        [[X_var, 0], [0, X_var]])

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(X, Y, rv.pdf(pos), cmap="plasma")
plt.show()

from Ipython.display import HTML

HTML("""<video width="320", height="240"
controls><Source src="jeevamshamayi.mp4" type=
"video/mp4">
</video>
""")
```

NOTES

```
import Ipython.display as ipd.
import librosa
import librosa.display
import matplotlib.pyplot as plt.
ipd. Audio ("Jeevamshamayi.m4a")
filename1 = ("Jeevamshamayi.m4a")
plt.figure (figsize ('5,4))
data1, sample_rate1 = librosa.load (filename1,
    sr = 22050, mono=true, offset =0.0, duration
librosa.display.waveplot (data1, sr = sample_rate1,
    max_points = 50000.0, x-axis = 'time'
```