1. Consider Salaries dataset from Kaggle and write the code to find the following

1. BasePay
2. TotalPay
3. Benefits average (mean) BasePay of all employees per year
4. No. of unique job titles
5. top 5 most common jobs
6. How many people have the word Chief in their job title

Ans:

import numpy as np  
import pandas as pd  
sal = pd.read\_csv('../input/Salaries.csv', na\_values='Not Provided')

1. Sal.BasePay()
2. Sal.TotalPay()
3. sal.groupby('Year').mean()['BasePay']
4. sal['JobTitle'].nunique()
5. sal['JobTitle'].value\_counts().head()
6. sal['JobTitle'].apply(lambda str:('chief' in str.lower())).sum()

2. Consider Salaries dataset from Kaggle and write the code to find the following

1. What is the average Purchase Price
2. How many people have English 'en' as their Language of choice on the website
3. How many people have the job title of "Lawyer"
4. What are the 5 most common Job Titles?
5. What is the email of the person with the following Credit Card Number: 4926535242672853
6. How many people have a credit card that expires in 2025

Ans:

import pandas as pd  
ecom=pd.read\_csv('../input/ecommerce-purchases-csv/Ecommerce Purchases.csv')

1. ecom['Purchase Price'].mean()
2. len(ecom[ecom['Language']=='en'])
3. len(ecom[ecom['Job']=='Lawyer'].index)
4. ecom['Job'].value\_counts().head(5)
5. ecom[ecom['Credit Card']==4926535242672853]['Email']
6. sum(ecom['CC Exp Date'].apply(lambda exp: exp[3:]=='25'))

3. Write the code for the following

1. Print a continuous signal x = (0.85) \*\* t
2. Print the Discrete signal x = (0.85) \*\* n
3. Draw a unit impulse signal with a line space of (0, 10, 100) and frequency 50Hz.

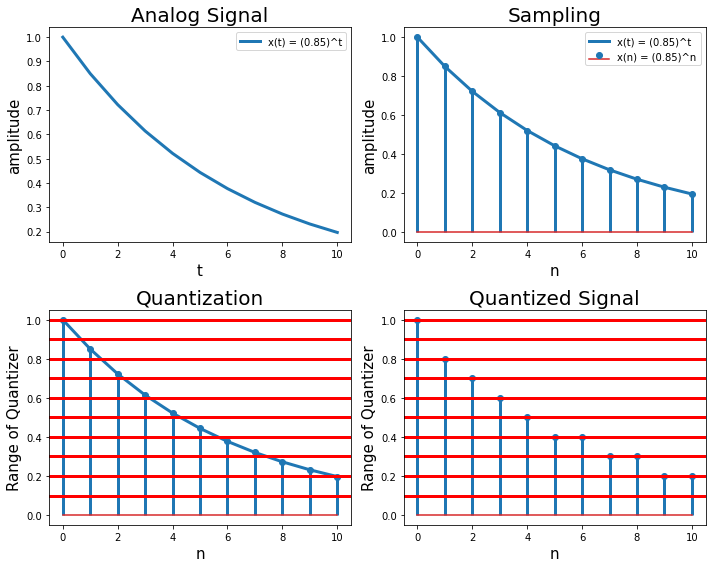
Ans:

import numpy as np  
import matplotlib.pyplot as plt  
  
Code:

import scipy  
from scipy import signal

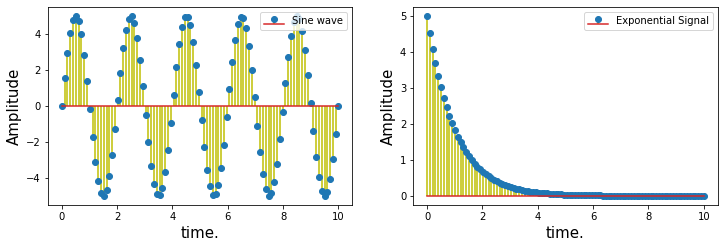
#continoues signal

t = np.arange(0, 11)  
x = (0.85) \*\* t  
plt.figure(figsize = (10,8)) # set the size of figure  
plt.subplot(2, 2, 1)  
plt.title('Analog Signal', fontsize=20)



#Discrete signal

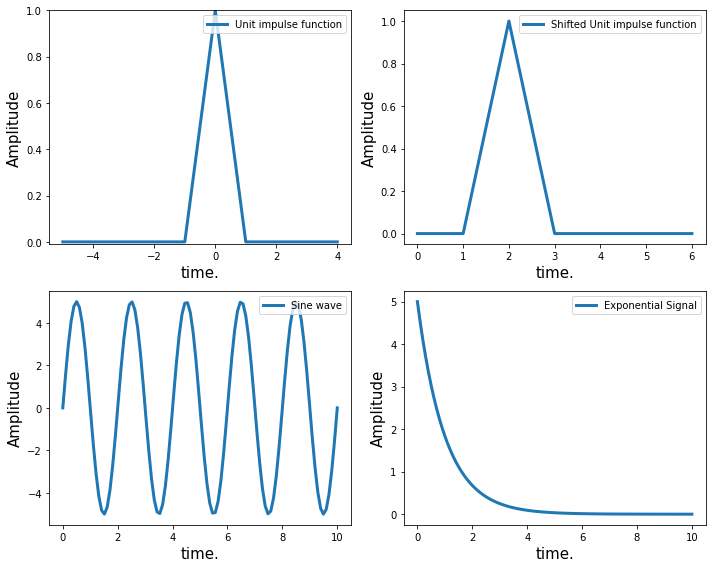
n = np.linspace(0, 10, 100)  
amp = 5 # Amplitude  
f = 50s  
x = amp \* np.sin(2 \* np.pi \* f \* n)  
plt.figure(figsize=(12, 8))  
plt.subplot(2, 2, 1)  
plt.stem(n, x, 'yo', label='Sine wave')  
plt.xlabel('time.', fontsize=15)  
plt.ylabel('Amplitude', fontsize=15)  
plt.legend(fontsize=10, loc='upper right')



#Unit impulse signal

impulse = signal.unit\_impulse(10, 'mid')  
shifted\_impulse = signal.unit\_impulse(7, 2)  
# Sine wave  
t = np.linspace(0, 10, 100)  
amp = 5 # Amplitude  
f = 50  
x = amp \* np.sin(2 \* np.pi \* f \* t)  
# Exponential Signal  
x\_ = amp \* np.exp(-t)

plt.figure(figsize=(10, 8))  
plt.subplot(2, 2, 1)  
plt.plot(np.arange(-5, 5), impulse, linewidth=3, label='Unit impulse function')  
plt.ylim(-0.01,1)  
plt.xlabel('time.', fontsize=15)  
plt.ylabel('Amplitude', fontsize=15)  
plt.legend(fontsize=10, loc='upper right')



4.Answer the following:

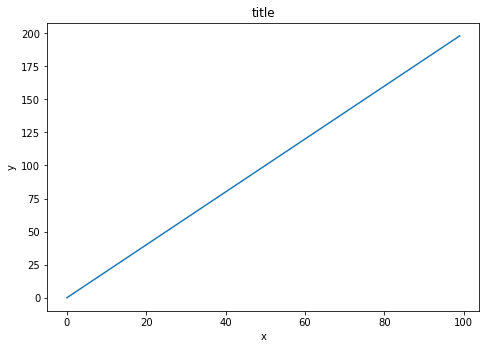
1. Create a figure object called fig using plt.figure()
2. Use add\_axes to add an axis to the figure canvas at [0,0,1,1]. Call this new axis ax.
3. Plot (x,y) on that axes and set the labels and titles

Ans:

import matplotlib.pyplot as plt  
%matplotlib inline

fig = plt.figure()  
ax = fig.add\_axes([0,0,1,1])  
ax.plot(x,y)  
ax.set\_xlabel('x')  
ax.set\_ylabel('y')  
ax.set\_title('title')

Text(0.5, 1.0, 'title')



5. Explain in detail about data transformation and its methods?

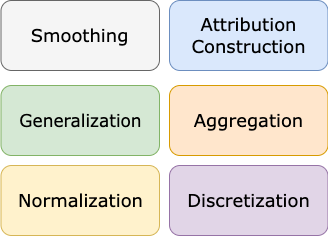
Data transformation is **the process of converting, cleansing, and structuring data into a usable format that can be analyzed to support decision making processes, and to propel the growth of an organization**. Data transformation is used when data needs to be converted to match that of the destination system.

The process of data transformation can be handled manually, automated or a combination of both.

Transformation is an essential step in many processes, such as data integration, migration, warehousing and wrangling. The process of data transformation can be:

* Constructive, where data is added, copied or replicated
* Destructive, where records and fields are deleted
* Aesthetic, where certain values are standardized, or
* Structural, which includes columns being renamed, moved and combined

## Different types of data transformation techniques

Datascience assignment

There are 6 basic data transformation techniques that you can use in your analysis project or data pipeline:

* Data Smoothing
* Attribution Construction
* Data Generalization
* Data Aggregation
* Data Discretization
* Data Normalization