Lab Exercise – 5: Data Transformation & Synthetic Data Generation

Note:

* Prepare a PDF document and name the file as “Lab5_RegisterNo.pdf”.
* PDF file should consist Question No, Code, and Result for each Question.
* File Should be headed with your Register number, Slot number, Lab Exercise number.

* * *

1. a. Download the bank.csv file and do the following transformations:

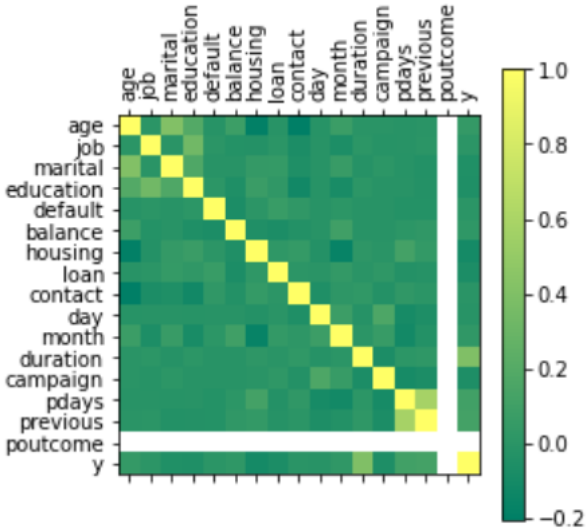
Attribute	**Transformation**	
	From	**To**
marital	single	0
	Other	1
housing	no	0
	yes	1
loan	no	0
	yes	1
job	'unknown'	np.nan
	'management'	0
	'technician'	1
	'entrepreneur'	2
	'blue-collar'	3
	'retired'	4
	'admin.'	5
	'services'	6
	'self-employed'	7
	'unemployed'	8
	'housemaid'	9
	'student'	10
education	'unknown'	np.nan
	'tertiary'	0
	'secondary'	1
	'primary'	2
default	no	0
	yes	1
contact	unknown	np.nan
	telephone	0
	cellular	1
month	jan-dec	1-12
poutcome	'unknown'	np.nan
	'failure'	0
	'other'	1
	'success'	2
y	no	0
	yes	1

1. b. Develop user defined functions for min-max normalization to [0,1] and Z-Score normalization for the attributes, duration, pdays and balance.

During min-max normalization, observe the data distribution before and after with a plot, and in Z-Score normalization identify data points which are far from three standard deviations.

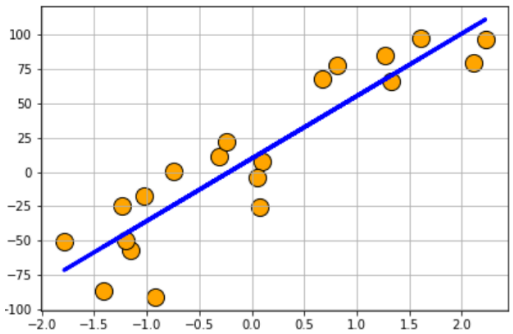
After data transformation, save the data into preprocessed_bank.csv file.

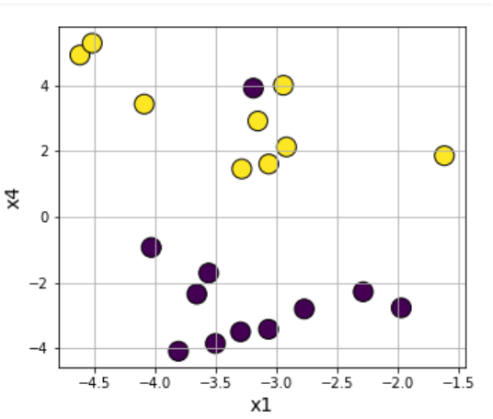
2. Reload the preprocessed_bank.csv file and calculate covariance and correlation matrices. Display the correlation matrix with a diagram as:

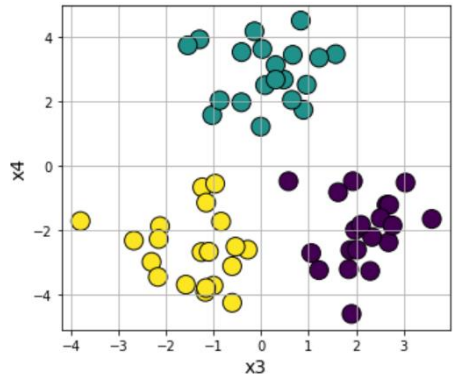


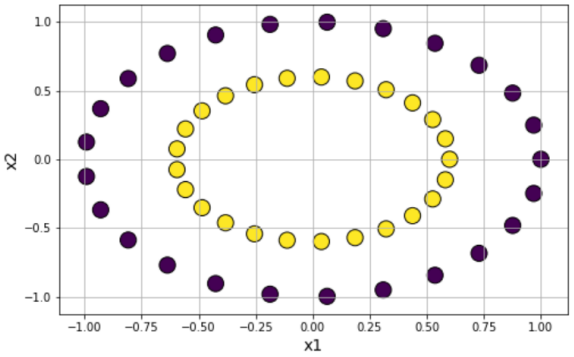
Sort Correlation matrix values of attribute “y” in descending order.

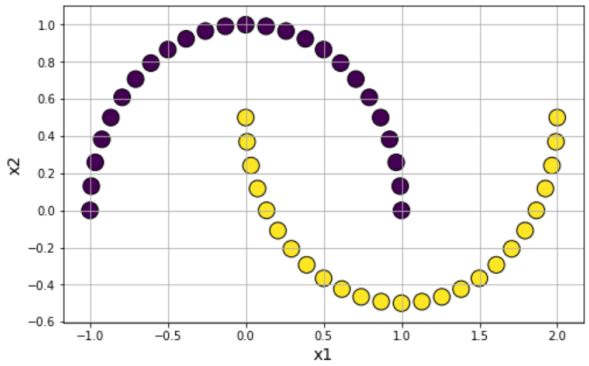
3. Generate synthetic data to visualize following kind of images. (Use sklean package)











4. Create Fake Customer data of super market with the following fields and store it in a CSV file. (Using Faker package and random package)

- Customer id, Customer name, Customer address, date of birth, gender, blood group, email, phone no.
- Customer id, Transaction ID, Purchase amount
- Transaction ID, Product purchased, # of items for each product
- Product Id, Product Name, cost of single product.