


National University of Computer and Emerging Sciences, Lahore Campus

	Course Name:	Data Science	Course Code:	CS-4048
	Degree Program:	BS(Computer Science)	Semester:	Fall 2023
	Exam Duration:	60 Minutes	Total Marks:	30
	Paper Date:	30-Sept-2023	Weight	15%
	Sections:	A,B	No of Page(s):	2
	Exam Type:	Midterm-I		

Student : Name: _____ **Roll No.** _____ **Section:** _____

Instruction/Notes: Attempt all question in sequence. Provide your answers in the answer booklet.
Exam is open book. Use of internet and electronic devices is not allowed.

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Payment
75-8169	A	Yangon	Member	Male	Sports and travel	15.81	10.0	7.905	\$166.00	3/6/2019	Credit card
853-23-2453	B	Mandalay	Member	Male	Health and beauty	75.74	4.0	15.148	\$318.11	2/14/2019	Cash
877-22-3308	A	Yangon	Member	Male	Health and beauty	15.87	10.0	7.935	\$166.63	3/13/2019	Cash
838-78-4295	C	Naypyitaw	Normal	Female	Health and beauty	33.47	2.0	3.347	\$70.29	2/10/2019	Ewallet
109-28-2512	B	Mandalay	Member	Female	Fashion accessories	97.61	6.0	29.283	\$614.94	1/7/2019	Ewallet

Problem#1 (CLO-1)

10 Marks

This dataset represents transaction records from a supermarket, capturing various aspects of each transaction. It includes details such as mentioned in the above given sample. This dataset is valuable for performing data analysis, identifying purchasing patterns, and gaining insights into customer behavior within the supermarket environment. Load the csv file into a data frame.

- Are there any missing values in the dataset? Write a command show the missing values for each column.

```
df.isnull.sum()
```

- Write a command to drop those columns which have more than 5% missing data.

```
threshold = len(df) * 0.95 # 5% threshold
```

```
df_cleaned = df.dropna(axis=1, thresh=threshold)
```

- Write a command to convert the 'Total' column into numeric.

```
df.Total = df.Total.str.replace('$', '').astype('float')
```

- How would you encode the 'Payment' and 'Customer type' columns? Provide code.

```
df['Customer type'] = df['Customer type'].astype('category').cat.codes
```

```
pd.get_dummies(data=df, columns=['Payment'])
```

5. Write code to extract and create new columns for specific date components like day, month, and year from the 'Date' column.

```
df['Date'] = pd.to_datetime(df['Date'])
# Extract day, month, and year into new columns
df['Day'] = df['Date'].dt.day
df['Month'] = df['Date'].dt.month
df['Year'] = df['Date'].dt.year
```

Problem#2 (CLO-2)

10 Marks

You need to reload the original dataset into a new data frame. Make sure the 'Total' column contains numeric values. Provide code and output both for the following problems.

1. What is the most common "Payment" method used by customers in each city?

```
df.groupby('City')['Payment'].agg(lambda x: x.mode().iloc[0])
```

OR

```
pd.crosstab(df['City'], df['Payment']).idxmax(axis=1)
```

2. Find the city where most customers use a credit card as their payment method.

```
df[df['Payment'] == 'Credit card']['City'].mode().iloc[0]
```

OR

```
pd.crosstab(df['City'], df['Payment'])['Credit card'].idxmax()
```

3. What is the average Total sales for each month in the dataset?

```
df.groupby('Month')['Total'].mean()
```

4. What is the most common customer type in each branch, and how many customers of that type are there?

```
pd.crosstab(df['Branch'], df['Customer type']).idxmax(axis=1)
```

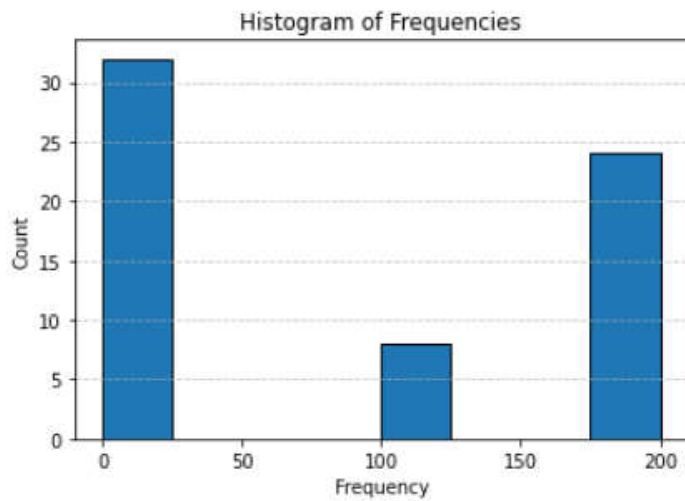
Problem#3 (CLO-1)

10 Marks

```
[ [ 0 0 0 0 0 0 0 0 ]
  [ 0 0 100 100 100 100 0 0 ]
  [ 0 200 200 200 200 200 200 0 ]
  [ 0 200 200 200 200 200 200 0 ]
  [ 0 200 200 200 200 200 200 0 ]
  [ 0 200 200 200 200 200 200 0 ]
  [ 0 0 100 100 100 100 0 0 ]
  [ 0 0 0 0 0 0 0 0 ] ]
```

Answer the following question with respect to the given 8x8 grayscale image represented as a matrix.

1. Create the histogram of frequencies.



2. Manually apply a horizontal flip to the image i.e. using array manipulation.

```
h_flip = img[:, :: -1]
```

3. Provide two reasons for applying image augmentation to the dataset of images.

To increase data size, generalized learning

4. Calculate the size of the image in bits.

$8 \times 8 \times 8 \times 1 = 512$ bits