Here’s a detailed explanation and code for each of the questions in your assignment:

**Q1: List any five functions of the pandas library with execution.**

1. head(): Displays the first few rows of the DataFrame.
2. tail(): Displays the last few rows of the DataFrame.
3. describe(): Generates summary statistics for numeric columns.
4. groupby(): Groups data based on a column.
5. dropna(): Removes rows or columns with missing values.

import pandas as pd

# Sample DataFrame

data = {'A': [1, 2, 3, None], 'B': [5, 6, None, 8]}

df = pd.DataFrame(data)

print("1. Head of DataFrame:")

print(df.head())

print("\n2. Tail of DataFrame:")

print(df.tail())

print("\n3. Describe DataFrame:")

print(df.describe())

print("\n4. Group by column A and count values:")

print(df.groupby('A').count())

print("\n5. Drop rows with missing values:")

print(df.dropna())

**Q2: Re-index the DataFrame with a new index starting from 1 and incrementing by 2.**

def reindex\_dataframe(df):

new\_index = range(1, 2 \* len(df) + 1, 2)

df.index = new\_index

return df

# Example DataFrame

df = pd.DataFrame({'A': [1, 2, 3], 'B': [4, 5, 6], 'C': [7, 8, 9]})

print(reindex\_dataframe(df))

**Q3: Calculate the sum of the first three values in the 'Values' column.**

def sum\_first\_three(df):

result = df['Values'][:3].sum()

print(f"Sum of the first three values: {result}")

# Example DataFrame

df = pd.DataFrame({'Values': [10, 20, 30, 40, 50]})

sum\_first\_three(df)

**Q4: Create a new column 'Word\_Count' with the word count of 'Text' column.**

def add\_word\_count(df):

df['Word\_Count'] = df['Text'].apply(lambda x: len(str(x).split()))

return df

# Example DataFrame

df = pd.DataFrame({'Text': ['Hello world', 'Data Science Masters', 'Pandas library']})

print(add\_word\_count(df))

**Q5: Difference between DataFrame.size and DataFrame.shape.**

* DataFrame.size: Returns the total number of elements in the DataFrame (rows × columns).
* DataFrame.shape: Returns the tuple (number of rows, number of columns).

**Q6: Function to read an Excel file.**

pd.read\_excel('filename.xlsx')

**Q7: Extract 'Username' from 'Email' and create a new column.**

def extract\_username(df):

df['Username'] = df['Email'].apply(lambda x: x.split('@')[0])

return df

# Example DataFrame

df = pd.DataFrame({'Email': ['john.doe@example.com', 'jane.doe@test.com']})

print(extract\_username(df))

**Q8: Select rows where column 'A' > 5 and column 'B' < 10.**

def select\_rows(df):

return df[(df['A'] > 5) & (df['B'] < 10)]

# Example DataFrame

df = pd.DataFrame({'A': [3, 8, 6, 2, 9], 'B': [5, 2, 9, 3, 1], 'C': [1, 7, 4, 5, 2]})

print(select\_rows(df))

**Q9: Calculate mean, median, and standard deviation of 'Values'.**

def calculate\_stats(df):

mean = df['Values'].mean()

median = df['Values'].median()

std\_dev = df['Values'].std()

return mean, median, std\_dev

# Example DataFrame

df = pd.DataFrame({'Values': [10, 20, 30, 40, 50]})

print(calculate\_stats(df))

**Q10: Create a 'MovingAverage' column with a 7-day window.**

def add\_moving\_average(df):

df['MovingAverage'] = df['Sales'].rolling(window=7, min\_periods=1).mean()

return df

# Example DataFrame

df = pd.DataFrame({'Date': pd.date\_range(start='2023-01-01', periods=10),

'Sales': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]})

print(add\_moving\_average(df))

**Q11: Add a 'Weekday' column with the weekday name.**

def add\_weekday(df):

df['Weekday'] = pd.to\_datetime(df['Date']).dt.day\_name()

return df

# Example DataFrame

df = pd.DataFrame({'Date': pd.date\_range(start='2023-01-01', periods=5)})

print(add\_weekday(df))

**Q12: Select rows where the date is between '2023-01-01' and '2023-01-31'.**

def filter\_date\_range(df):

return df[(df['Date'] >= '2023-01-01') & (df['Date'] <= '2023-01-31')]

# Example DataFrame

df = pd.DataFrame({'Date': pd.date\_range(start='2023-01-01', periods=10)})

print(filter\_date\_range(df))

**Q13: Necessary library for pandas.**

import pandas as pd