

Name: Vinayak V Thayil  
Roll No:AM.EN.U4CSE21161

```
import numpy as np
import pandas as pd
# Set a random seed for reproducibility
np.random.seed(42)
# Generate random data for three features: Age, Income, and Score
num_samples = 100 # Number of data points
age = np.random.randint(18, 65, size=num_samples)
income = np.random.normal(50000, 10000, size=num_samples)
score = np.random.uniform(0, 100, size=num_samples)
# Create a DataFrame to store the data
data = pd.DataFrame({
    'Age': age,
    'Income': income,
    'Score': score
})
# Print the first few rows of the dataset
print(data.head())
```

	Age	Income	Score
0	56	56363.051083	17.495493
1	46	40932.793314	98.216834
2	32	54760.425874	51.663589
3	60	63036.612684	26.082917
4	25	52115.870123	99.625370

```
import faker
import datetime
import random
np.random.seed(42)
def random_dates(start_date, end_date, n=10):
    date_list = [start_date + datetime.timedelta(days=random.randint(0, (end_date - start_date).days)) for _ in range(n)]
    return date_list
fake = Faker()
num_records = 100
employee_ids = list(range(1, num_records + 1))
employee_names = [fake.name() for _ in range(num_records)]
departments = ['Sales', 'Marketing', 'Engineering', 'Finance', 'HR']
employee_departments = [random.choice(departments) for _ in range(num_records)]
employee_salaries = [random.randint(50000, 100000) if random.random() > 0.2 else np.nan for _ in range(num_records)]
start_date = datetime.date(2010, 1, 1)
end_date = datetime.date(2023, 1, 1)
employee_joining_dates = random_dates(start_date, end_date, num_records)
data = {
    'Employee ID': employee_ids,
    'Employee Name': employee_names,
    'Department': employee_departments,
    'Salary': employee_salaries,
    'Joining Date': employee_joining_dates
}
employee_df = pd.DataFrame(data)
print(employee_df.head())
```

	Employee ID	Employee Name	Department	Salary	Joining Date
0	1	Carlos Bentley	Marketing	82418.0	2022-07-16
1	2	Brian Morse	Marketing	82785.0	2017-12-13
2	3	James Jimenez	Sales	79352.0	2011-06-15
3	4	Amy Moss	Marketing	60858.0	2020-01-02
4	5	David Martinez	HR	84281.0	2022-02-28

```
missing_count = employee_df.isnull().sum()
missing_percentage = (missing_count / len(employee_df)) * 100
missing_info = pd.DataFrame({
    'Missing Count': missing_count,
    'Missing Percentage': missing_percentage
})
print(missing_info)
```

	Missing Count	Missing Percentage
Employee ID	0	0.0
Employee Name	0	0.0

Department	0	0.0
Salary	20	20.0
Joining Date	0	0.0