Data Anonymization

1. Data Masking

```
import pandas as pd

data = {
    'name': ['Nandu', 'Parthiba', 'Adithya'],
    'email': ['nkv@duk.com', 'prathibaprathi@duk.com',
'itsmeadhii@duk.com'],
    'phone': ['8495568678', '9667844777', '9876538796'],
}

df = pd.DataFrame(data)

df["name"] = df["name"].apply(lambda x: x[:3] + "******" + x[9:])
df["email"] = df["email"].apply(lambda x: x[:3] + "******" +
x[9:])
df["phone"] = df["phone"].apply(lambda x: x[:3] + "******" +
x[9:])

print(df)
```

```
       name                email      phone
0  Nan******          nkv******om  849******8
1  Par******  pra******rathi@duk.com  966******7
2  Adi******      its******i@duk.com  987******6
```

Submitted by ***Nanda Krishnan V, Msc Cyber Security***

2. Pseudonymization

```
import pandas as pd
import random
import string

data = {
    'name': ['Nandu', 'Parthiba', 'Adithya'],
    'email': ['nkv@duk.com', 'prathibaprathi@duk.com',
'itsmeadhii@duk.com'],
    'phone': ['8495568678', '9667844777', '9876538796'],
}

df = pd.DataFrame(data)

def generate_random_string(length=8):
    return ''.join(random.choices(string.ascii_lowercase +
string.digits, k=length))

df["name"] = [f"User{i+1}" for i in range(len(df))]
df["email"] = [f"user{i+1}@example.com" for i in range(len(df))]
df["phone"] = [generate_random_string(3) +
''.join(random.choices(string.digits, k=7)) for _ in
range(len(df))]

print(df)
```

```
    name              email       phone
0  User1  user1@example.com  z5n2798593
1  User2  user2@example.com  0kt0899011
2  User3  user3@example.com  cct2424440
```

3. Suppression

```
import pandas as pd
data = pd.DataFrame({
 'name': ['Nandu', 'Parthiba', 'Adithya'],
    'email': ['nkv@duk.com', 'prathibaprathi@duk.com',
'itsmeadhii@duk.com'],
    'phone': ['8495568678', '9667844777', '9876538796'],
})

data["phone"] = None
data["email"] = "NA"
print("\nAnonymized Dataset:")
print(data)
```

```
Anonymized Dataset:
       name email phone
0     Nandu    NA  None
1  Parthiba    NA  None
2   Adithya    NA  None
```