

## Phase: 03

### Public Health Awareness Campaign

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import geopandas as gpd

#data cleansing
data = pd.read_csv('survey.CSV')
data = data.drop_duplicates()
data['AgeID'] = data['GenderID'].int('desired_data_type')
data = data[(data['CountryName'] >= lower_threshold) &
            (data['NumberOfDays'] <= upper_threshold)]
from sklearn.preprocessing import StandardScaler, MinMaxScaler
scaler = StandardScaler()
data['CountryID'] = scaler.fit_transform(data[['AgeID']])
data['Family_History'] = data['NumberOfDays'].str.strip()
data['CountryName'] = data['CountryName'].str.upper()
data.to_csv('cleaned_survey.CSV', index=False)

#data visualization
AgeID = [55,28,12,57,89]
Country ID = ['US','CANADA','FRANCE', 'RUSSIA', 'SWITZERLAND']
plt.plot(AgeID, Country ID)
```

```
plt.AgeIDlabel("AgeID")
plt.CountryIDlabel("Country ID")
plt.show()

#geospatial analysis
gdf = gpd.read_file('survey.CSV')
print(gdf.head())
print(gdf.crs)
gdf.plot()
plt.title('Geospatial Data Visualization')
plt.show()
```

```
#data mining
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

data = load_iris()
AgeID = data.data
CountryID = data.target

AgeID_2014, AgeID_test, CountryID_Canada, CountryID_test =
train_test_split(AgeID, CountryID , test_size=0.3, random_state=42)

classifier = DecisionTreeClassifier()
classifier.fit(AgeID_2014, CountryID_Canada)

CountryID_pred = classifier.predict(AgeID_test)

accuracy = accuracy_score(CountryID_test, CountryID_pred)

print("Accuracy:", accuracy)
```