NUST

NATIONAL UNIVERSITY
OF SCIENCES & TECHNOLOGY

PCA Assignment

Muhammad Anser Sohaib

367628

Code Snippets and One Liner Explanations:

(1) Data Loading:

```
data = {
    'Category': [
        'Alcoholic drinks', 'Beverages', 'Carcase meat', 'Cereals', 'Cheese',
        'Confectionery', 'Fats and oils', 'Fish', 'Fresh fruit', 'Fresh potatoes',
        'Fresh veg', 'Other meat', 'Other veg', 'Processed potatoes', 'Processed veg',
        'Soft drinks', 'Sugars'
    ],
    'England': [375, 57, 245, 1472, 105, 54, 193, 147, 1102, 720, 253, 685, 488, 198, 360, 1374, 156],
    'N Ireland': [135, 47, 267, 1494, 66, 41, 209, 93, 674, 1033, 143, 586, 355, 187, 334, 1506, 139],
    'Scotland': [458, 53, 242, 1462, 103, 62, 184, 122, 957, 566, 171, 750, 418, 220, 337, 1572, 147],
    'Wales': [475, 73, 227, 1582, 103, 64, 235, 160, 1137, 874, 265, 803, 570, 203, 365, 1256, 175]

}
```

(2) PCA Function:

***

Performs PCA using Singular Value Decomposition (svd)

Arguments:

data: pandas data frame

k: number of principal components to find

Returns:

transformed_data: The data projected onto the principal components.
explained_variance_ratio: The percentage of variance explained by each PC.

***

```
def pca_svd(data, k):

  mean = np.mean(data, axis=0)
  centered_data = data - mean

  U, S, Vt = np.linalg.svd(centered_data, full_matrices=False)

  singular_values = S[:k]
  principal_components = Vt[:k]  # V transpose already gives principal component

  transformed_data = centered_data.dot(principal_components.T)

  explained_variance = (singular_values ** 2) / (data.shape[0] - 1)
  total_variance = np.sum(explained_variance)
  explained_variance_ratio = explained_variance / total_variance

  return transformed_data, explained_variance_ratio
```

(3) First two principal components:

Shows variance explained by each country in two principal components.

```
transformed_data
```

	0	1
England	-144.993152	2.532999
N Ireland	477.391639	58.901862
Scotland	-91.869339	-286.081786
Wales	-240.529148	224.646925

(4) Variance explained by first two PCs:

Using this formula (k=2) $\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{m} S_{ii}}$

```
explained_variance_ratio
```

```
array([0.69892817, 0.30107183])
```

(5) Plot [PC1]: