Exploratory Data Analysis
Recipients of Social Protection Schemes

Prepared by : Alya Almanqour

1.Abstract

The data in my project show Quarterly data of recipients on selected social protection schemes by county from Ireland. Since the given data is a single unit drawn from huge information, certain actions must be taken. As the organization considers it a valuable asset because of the information stored in business and operations.
Therefore, in this project we will analyze the data given to handle, in the case of a new classification system for number of recipients , the impact on society could be a better suggestion , depending on their needs.

2.Datasets Description

This data was provided by Department of Social Protection of Welfare by virtue of the Central Statistical Office in Ireland .Period of time covered between 2014-01-01 and 2021-09-30. Dataset shows 10 columns for 16,736 entries.

Structure of the data:

id : Unique number to identify each row .

period: represents time of year

scheme: Subunit of scheme_ description

scheme_ description: shows the type of benefits

Basis: Subunit of scheme_ description

county_code: County name abbreviations

county: county name

recipients: the number of recipients of social assistance.

```
RangeIndex: 16736 entries, 0 to 16735
Data columns (total 9 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   period              16736 non-null  object
 1   programme           16736 non-null  object
 2   scheme              16736 non-null  object
 3   scheme_description  16736 non-null  object
 4   basis               16736 non-null  object
 5   UGI                 16736 non-null  object
 6   county_code         16736 non-null  object
 7   county              16736 non-null  object
 8   recipients          16736 non-null  int64
dtypes: int64(1), object(8)
memory usage: 1.1+ MB
```

3.Goals

The main goal of the project is to develop, integrate, and share knowledge using new technologies to improve the production and outputs (products and
services) of official statistics. After the recipients are classified by county we must answer some questions:
1. Which county have the largest number of recipients?

2. Is there any relationship between the total number of recipients in some county with the type of service provided (scheme).

3. Is the number of recipients increasing or decreasing over the years?

4. Is the service make differences over the year?

4.Tools

There are several tools which are offered for data science project. However, tools chosen from open source platforms could be preferred. In my Project I will use
Technologies : Python, Jupyter notebook, SQL.
Libraries: Pandas, Matplotlib, NumPy.

5.Expected Conclusion

Managing and communicating dataset to answering the above questions that transform data into actionable information .