Text

Description automatically generated

Jader Leonardo Gomes da Rosa – **sba22199**

**sba22199@student.cct.ie**

**Production and Population**

**Continuous Assessment 2**

**Dublin**

**December 30,2022**

**TABLE OF CONTENT**

|  |  |
| --- | --- |
| INTRODUCTION | 3 |
| Data Manipulation | 4 |
| Exploratory Data Analysis | 5 |
| Statistics and appropriate plots | 9 |
| Boxplot payment per year | 9 |
| Add a calculated column on Data Frame | 11 |
| Created new Data Frame | 12 |
| Report the payment | 12 |
| Boxplot number of spaces per year | 15 |
| Number of spaces per amount in 2006 | 17 |
| Distribution (Binomial) | 18 |
| Decision Trees and Random Forest | 19 |
| Decision TreeRandom forest | 20 |
| Explain about SEMMA | 23 |
| Explain about CRISP-DM | 23 |
| Problems with Testing Machine Learning Models | 25 |
| Principles & Best Practices | 25 |
| Reproducibility | 26 |
| How to Test Machine Learning Models | 26 |
| Invariance Test | 27 |
| Matplotlib pyplot: Pair plot | 28 |
| Columns description and details | 31 |
| CONCLUSION | 35 |
| REFERENCES | 31 |

**INTRODUCTION**

This project is based on two types of databases, the amount of population of the European Union countries between 2015 and 2020 and production of Barley, Meat Cattle, Meat Pig, and Raw Milk for the same period and countries.

Data bases were collected from Food and Agriculture Organization of the United Nations (FAO).

All information contained in this document was based on instructions obtained in the classroom and all complements are correct referenced.

All additions and changes were made by code in the Jupyter notebook and are added in this document.

All of the documents and creations made in this document are on GitHub and saved directly in the CA2\_Project repository

FAO allows filters and extraction of report in CSV format, the following parameters were applied to population, filtered by:

Population data bases: European Union countries, total female population, total male population, total rural population, and total urban population for the years 2015 to 2020.

Production data bases: European Union countries, total female population, total male population, total rural population, and total urban population for the years 2015 to 2020.

after the data were extracted, manipulations and collisions were made.

**Data Manipulation**

Jupyter Notebook, Population\_Eurpean:

1 – Import 4 csv databases. (1. Population\_Female, 2. Population\_Male, 3. Population\_Rural and 4. Population\_Urban).

2 – Renamed value columns according to respect reference (Female\_K, Male\_K, Rural\_K and Urban\_K).

3 - The future period was identified where the data was run to the correct year. (07/06/2006 instead of 07/06/2026).

4 – The row with no data was removed.

Some “location\_of\_spaces” in the column “E” were adjusted because we identified some similar ones with different abbreviations or double space.

1 – The “146 Parnel St. Dublin 1”, “146 Parnell Street” and “146 Parnell Street. Dn 1” were changed to “146 Parnel Street Dublin 1”

2 – The “17, Shops, New Cabra Road, Cabra Dublin 7”, “17 Shops New Cabra Road Dublin 7” and 17 Shops, New Cabra Road were changed to “17, Shops, New Cabra Road, Cabra Dublin 7”

3 – The “Adelaide Road (in front of No. 24)”, “Adelaide Road, Dublin 2” and “Adelaide Road” were changed to “Adelaide Road, Dublin 2”

4 – The “Amiens Street, Dublin 1”, “Amien Street (Opposite Connelly Station)”, “Amiens Street”, “Amien Street”, “Amiens Street opposite Luas” were changed to “Amiens Street, Dublin 1”

5 – The “Anglesea Road (Outside No. 7)”, “Anglesea Road”, “Anglesea Road (outside the Herbert Park Hotel)” were changed to “Anglesea Road”

6 – The “Aungier Street, Redmond's Hill” and “Aungier Street, Redmonds Hill” were changed to “Aungier Street, Redmond's Hill”

7 – The “Baggott Street Upper” and “Baggot Street Upper” were changed to “Baggot Street Upper”

8 – The “Blackhall Place” and “Blackhall Place” were changed to “Blackhall Place”

9 – The “Castle Street” and “Castle Street” were changed to “Castle Street”

10 – The “Dame Court” and “Dame Court, Dublin 2” were changed to “Dame Court, Dublin 2”

11 – The “Georges Quay” and George's Quay” were changed to “George's Quay”

12 – The “Glengarriff Parade” and “Glengariff Parade” were changed to “Glengariff Parade”

13 – The “Grenville Street” and “Greenville Street” were changed to “Greenville Street”

**Exploratory Data Analysis**

**CONCLUSION**

**REFERENCES**

<https://github.com/jaderleonardo/CA2_Project>

https://www.fao.org/faostat/en/#data/OA