**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | *Programming for DA*  *Statistics for Data Analytics*  *Machine Learning for Data Analysis*  *Data Preparation & Visualisation* |
| **Assessment Title:** | *MSC\_DA\_CA1* |
| **Lecturer Name:** | *Sam Weiss/* *David Gonzalez*  *Bharathi Chakravarthi/* *Marina Iantorno*  *Muhammad Iqbal*  *David McQuaid* |
| **Student Full Name:** | Jose Mario da Cruz Costa |
| **Student Number:** | 2023422 |
| **Assessment Due Date:** | 10/11/2023 |
| **Date of Submission:** | 10/11/2023 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

**Exploring the Link Between Population Trends and Renting Prices**

CONTENTS

ABSTRACT

1 Data preparation and Visualization

* 1. **Early Data Analysis/Exploratory Data Analysis (EDA)**
  2. Merging all Datasets into only solid

ABSTRACT

The aim of this project is to explore the link between population trends and renting prices in Ireland. To achieve this goal, I will be investigating how population growth, immigration, migration, and other population factors affect the cost of renting a home in Ireland.

For this purpose, we need to answer the following questions:

How has the population of Ireland changed in recent years?

How have renting prices changed in recent years?

Is there a relationship between population growth and renting prices in Ireland?

Population trends can be used to predict future rent prices in Ireland?

To answer these questions, I have collected and analysed data on population trends and renting prices in Ireland.

The main sources of these dataset are: Central Statistics Office (CSO) and the Residential Tenancies Board (RTB).

Once you have collected the data, you can use a variety of statistical methods to analyze it. To choose the best way to predict something, we are exploring and comparing different regression models. The models we are considering are multiple linear regression, lasso and ridge regularization, and decision tree regression.

**As d**ifferent regression models make different assumptions about the data. Comparing different models, I will find the one that produces the most accurate predictions for the dataset.

I also will use regression analysis to model the relationship between population growth and renting prices.

The datasets show Estimated Immigration into Ireland since 1996, Estimated Population in Ireland since 2016 and Average renting price since 2007, To improve the accuracy of my results, I will merge the three datasets into a unified dataset, retaining only data from 2016 to 2023.

Data preparation and Visualization

**Early Data Analysis/Exploratory Data Analysis (EDA)**

Before getting any result from the data, it is necessary perform the correct EDA process, this process provide insights into its characteristics, relationships, and patterns.

**EDA method for the dataset**

The dataset contains three variables:

* Estimated Immigration into Ireland since 1996
* Estimated Population in Ireland since 2016
* Average renting price since 2007

**As we have three different datasets, the best approach is the multivariate analysis, what is** exploring the relationships between three or more variables, so I will look at the relationship between immigration, population, and renting prices.

I chose these methods because they will allow me to gain insights over the overall trends in immigration, population, and renting prices over time.

**4. Insights gained from EDA**

The insights gained from EDA will help me to choose the appropriate prediction method for the dataset and to develop a more accurate and reliable model. For example, if I find that immigration is a strong predictor of population growth, then I can include immigration in my prediction model.

**Here are some specific examples of insights that I might gain from EDA:**

* I might find that immigration has been increasing steadily over time, while population growth has been slowing down. This could suggest that immigration is becoming a more important driver of population growth in Ireland.
* I might find that there is a strong positive correlation between immigration and renting prices. This could suggest that an increase in immigration leads to an increase in demand for housing, which drives up renting prices.
* I might find that there is a negative correlation between population growth and renting prices. This could suggest that an increase in population leads to an increase in the supply of housing, which drives down renting prices.

By identifying these insights, I can develop a better understanding of the data and choose the appropriate prediction method.