**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Programming for DA,  Statistics for Data Analytics,  Machine Learning for Data Analysis,  Data Preparation & Visualisation |
| **Assessment Title:** | CA2 |
| **Lecturer Name:** | Marina Iantorno  Muhammad Iqbal  David McQuaid |
| **Student Full Name:** | Rita Raher |
| **Student Number:** | sbs22115 |
| **Assessment Due Date:** | ?? |
| **Date of Submission:** | ?? |

Icon

Description automatically generated with medium confidence

**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.

# Abstract

Some text

# Introduction

Due to the conflict in Ukraine, there has been increased pressure on Irish farmers due to rising input costs (Teagasc, 2022).

the outlook in 2022 for Irish farmers

Github repo: <https://github.com/RitRa/Msc_CA2>

Link to dashboard: <https://share.streamlit.io/ritra/msc_ca2/notebooks/dashboard.py>

# Methodology

Some text

The project was tracked using Github account <https://github.com/RitRa/Msc_CA2>

# Data Collection

Using the [Central Statistics Office](https://www.cso.ie/en/statistics/), CSO, a dataset on fertiliser prices was sourced with data dating back as far as 1980 to 2022.

According to x, fertiliser is made up from potassium, nitrogen and phosphate. These are readily available on [www.indexmundi.com](http://www.indexmundi.com).

Data for potassium chloride was collected from

|  |  |
| --- | --- |
| Datasets |  |
| Fertiliser | [source](https://data.cso.ie/table/AJM05) |
| Potassium | [source](https://www.indexmundi.com/commodities/?commodity=potassium-chloride&months=240&currency=eur) |
| Nitrogen |  |
| Phosphate | [source](https://www.indexmundi.com/commodities/?commodity=rock-phosphate&months=360&currency=eur) |

Fertilizer production require large amounts of gas

Find gas

## Twitter

Part of this project is to analysis sentiment related to agriculture. One of the best places to find out what the general population are thinking is Twitter. One way to gain access to tweets, is by using a developer account. This will give you api key, api key secret, access token and access token secret. To get elevated access to twitter you need to explain what you are doing with the tweets. This gives you access to 2M Tweets per month / Project. There is a python library called tweepy which allows for easy access to the twitter api. Unfortunately, Twitter only allows api access up to 5-7 days of tweets or 1,500 tweets in total so the total number of tweets returned was only 9, not enough for this project. Snscrape is a python web scraper for social networking services which allows you to add search criteria and gives back tweets without the restrictions of others like Tweepy. Specific keywords were targeted for focus on agriculture, they included: #farmers, Fertiliser, #farming, #agriculture, farm and each search criteria included “near: Dublin” to try and collected tweets in Ireland only. After appending all the datasets together, it resulted in ~1000 tweets ready for sentiment analysis. [jupyter]

# Data understanding

Some text

# Data preparation

## Inferential statistics

Inferential statistics allows you to make inferences or estimations about the population based on the sample data (B. S. Everitt, 2010) . There are a number of ways of sampling methods that can be used, some examples are random sampling and stratified sampling. Random sampling

Stratified sampling is a way of obtaining samples that best represent the population as it divides the data into subgroups and takes a sample from each subgroup (Weiss, 2017).

Using this stratified sampling approach, the dataset was grouped by fertiliser\_type to sample each subgroup of fertiliser type, taking 5 from each [jupyter]. In order to determine whether the sample data are normally distributed, a probability plot was graphed and a Shapiro-Wilk test was performed. The Shapiro-Wilk test revealed that the sample data was not normally distributed as it had a p-value of 0.00001329.

### T-Test

For this project, it will be assumed that it is normally distributed. A t-test is performed on the sample dataset to check whether the sample mean price of fertiliser is equal to 288 yielding a p-value of 0.00097. The null hypothesis is rejected and the alternative hypothesis is accepted. The difference between the mean of the sample data and 288 is statistically significant.

### Anova

A sample size of x was selected

To ensure that the sample is

Using a T-test

ANOVA - Analysis of variance

As stated in

## Comparing countries against Ireland

### T-Test

Parametric test (Brownlee, 2021)

The Independent variable is geo, IE and PL (Ireland and Poland) and a t-test can be used to see if there is a difference between the fertiliser consumption volume of each group or if it is the same.

A sample of 15 is randomly selected from each dataset [jupyter] and then a function is applied to perform the t-test. The test output yields a t-statistic of 3.843 and a p-value is 0.001, as the p-value is less than alpha (0.05) and the null hypothesis is rejected. The difference between the means is statistically significant.

### ANOVA

Analysis of variance can be used to compare groups of more than two.

The Independent variable is geo, IE, PL and DE (Ireland, Poland, Germany).

A random sample of 15 from each country was selected for the test.

The ANOVA test produced an f-statistics of 0.0002, this is less than 0.05 and therefore the null hypothesis is rejected. Th difference between the 3 countries is statistically significant.

# Modelling

Some text

Sentiment analysis

Added in stopwords for farm, farms, amp as the topics

Decided to apply the year to

Timeseries analysis

## **ARIMA Model**

<https://siebert-julien.github.io/time-series-analysis-python/overview.html>

Using facebooks prophet for time series analysis, we can predict the future price of fertiliser

With autoregressive model

Time series

# Evaluation

text

# Conclusion

text

Add bib

Article

<https://www.irishtimes.com/news/ireland/irish-news/price-of-fertiliser-doubles-for-farmers-with-knock-on-anticipated-in-food-prices-1.4827498#:~:text=Office%20(CSO).-,Fertiliser%20prices%20rose%20by%20127.2%20per%20cent%20since%20January%202021,output%20and%20input%20price%20indices>.

<https://www.irishexaminer.com/farming/arid-40829501.html>

**Fertiliser prices rose by 127.2 per cent since January 2021**

What is causing the increase?

<https://asmith.ucdavis.edu/news/story-rising-fertilizer-prices>

is it caused by Russia invasion?

Check the dates of when it jumped?

<https://www.tfi.org/the-feed/fertilizer-101-big-3-nitrogen-phosphorus-and-potassium>

<https://www.cso.ie/en/statistics/>

<https://www.agriland.ie/farming-news/fertiliser-prices-impact-farmers-and-consumers-mcguinness/#:~:text=The%20spiraling%20cost%20of%20fertiliser,(CAN)%20and%20urea%20fertilisers>.

Milk prices

Diesel

Grain

Ideas for sentiment

positive and negative sentiment

<https://ec.europa.eu/eurostat/databrowser/view/T2020_RN310/bookmark/table?lang=en&bookmarkId=0c6ee6ae-b496-4703-a0dd-66b2b2fc8184>