Group ID - MSc in Data Analytics

Author: Olena Pleshan
e-mail: sbs24043@student.cct.ie
Student ID: sbs24043
GitHub Link: https://github.com/CCT-Dublin/term-1-ca-2-sbs24043.git

Abstract

This study provides a comprehensive comparison of crop yield and production across various EU countries, with a primary focus on Ireland and its key agricultural crops. Utilizing data from multiple sources, including FAO's FAOSTAT and Reddit, the study makes heavy use of public data, and open source libraries to turn that data into insights. The analysis integrates exploratory data analysis, statistics, and machine learning, using open-source libraries developed by big tech companies, notably Meta's Prophet model for time series forecasting, and also BERT topic analysis model developed by Cornell University based on research done by Alphabet company. Data preprocessing involved advanced techniques to ensure integrity and suitability for machine learning, while visualizations were created using Plotly and Dash to enhance accessibility for the interested persons. The study demonstrates the value of combining robust statistical methods with open-source tools to provide actionable insights for policymakers, farmers, and businesses.

Introduction

The study compares various EU countries in terms of crop yield and production. It seeks to identify patterns in crop output changes across the EU and to forecast future yields and outputs. The study also aims to create dashboards featuring exploratory data analysis, statistics, and machine learning, which can be utilized by policymakers, farmers, and businesses. While the study looks at different EU countries, it had a primary focus on Ireland and top crops of interest for the Irish agricultural industry.

Materials and Methods

Descriptive analysis was used for ordering, manipulating, and interpreting raw data from various sources. Data visualization techniques were used to provide insights into the datasets.

Exploratory and statistical analysis was used, as the dataset contains a lot of entities that have no relationship but which can benefit from comparison using statistical tests. It helped to find connections and generate graphs to identify suitable methods for data preprocessing, and further hypotheses validation.

To ensure data integrity and suitability for machine learning analysis, various preprocessing techniques such as imputation of missing values, scaling, and encoding were implemented (Feature Engineering).

API queries were performed for development of datasets.

Results / Discussion

Programming

In this study, data collection involved various sources, each demanding distinct processing tools, with Python Pandas adeptly handling tabular formats like CSV, TSV, Excel, JSON upon setting appropriate flags. This was easy to achieve with Pandas, contrasting with the basic functionality offered by the built-in csv module, which provides lower-level control over the underlying data structures.

A significant portion of the study focused on JSON data obtained through direct calls to the Reddit API, which required prettification for clarity and traversal as dictionaries to retrieve relevant information, prompting the preference for specialized libraries like PRAW to abstract away complexities.

Assembly of data from FAOSTAT, lacking an API, involved manual segmentation and output into multiple CSV files, concatenated via a for-loop, with additional steps like unstacking and melting the DataFrame to accommodate different operations' data representations for different purposes.

Due to issues with the PRAW API, the raw Reddit API was utilized to gather posts based on combined queries, while complexities with Reddit JSON and API data structures prompted reverting to PRAW for gathering post comments.

Profiling was used to determine resource usage using in-cell magic functions such as %time. Profiling during API calls and model training included attention to complexity of nested for loops, with in-advance counting of Reddit API calls to remove unnecessary ones. Storing gathered data as CSV files optimized resource usage, while certain machine learning libraries enabled explicit GPU usage (such as BERT), enhancing performance and freeing up CPU resources during model training (Maarten, 2022). Also, saving ML models into JSON files eliminated the need to retrain the models every time.

Statistics

To determine suitable statistical methods, it's essential to assess data for normality. Although aggregate data from all countries fails normality tests (see Fig 1), data from individual countries approximates a normal distribution (see Fig 2).

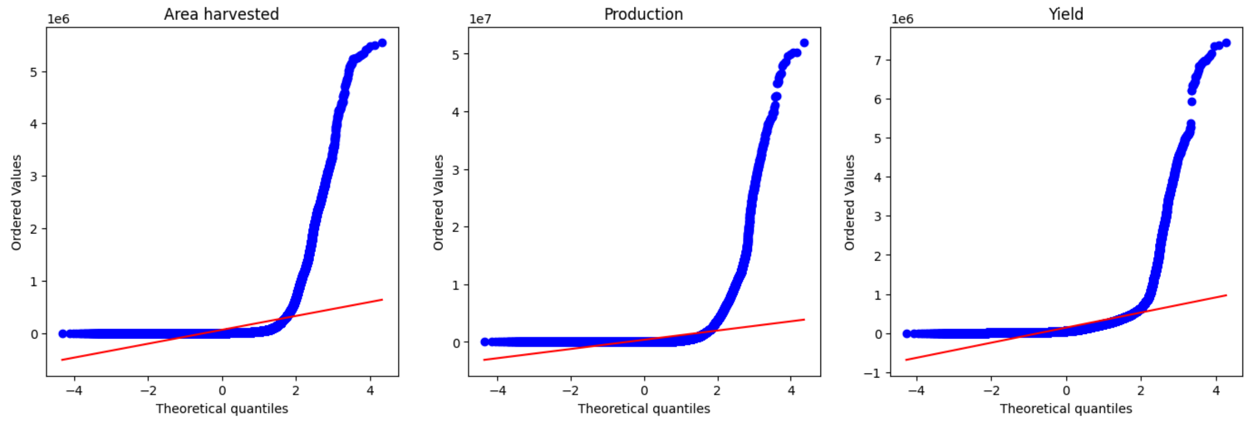


Figure 1: QQ plot for aggregated metric distributions

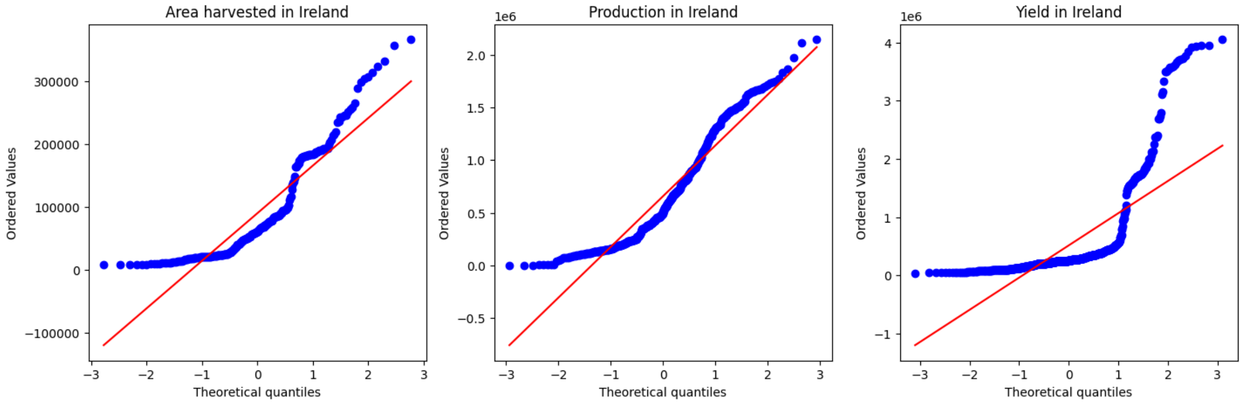


Figure 2: QQ plot for per country data distribution approaches normality

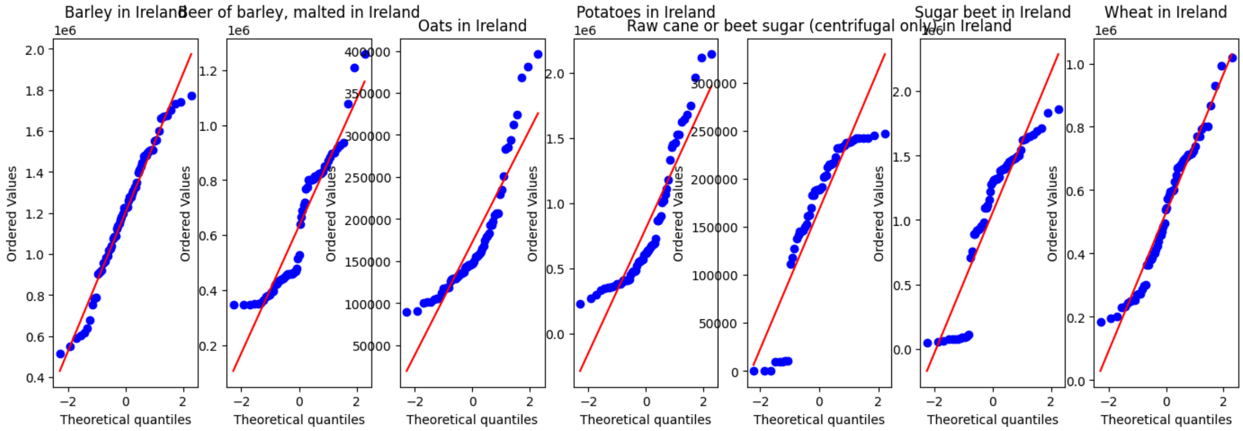
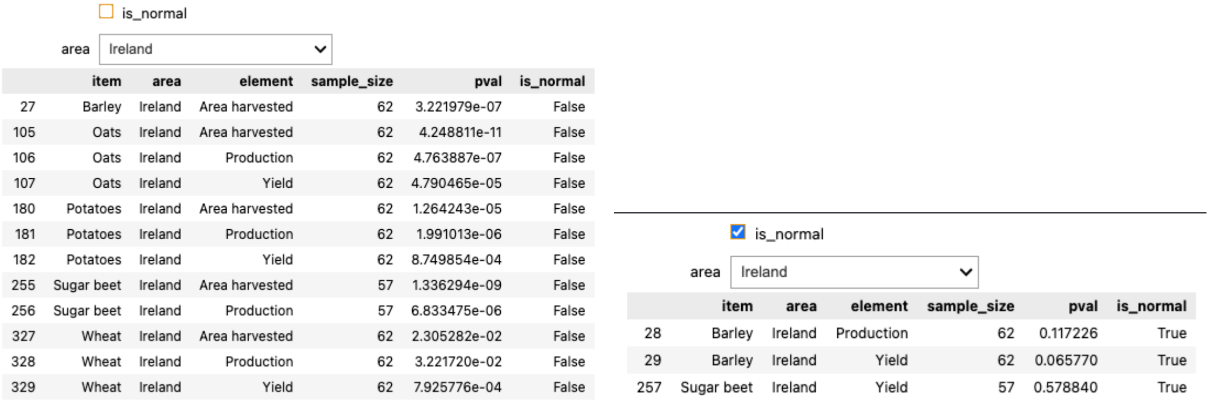


Figure 3: QQ plot at the individual crop level, the QQ plot results are showing normality

Shapiro normality tests were performed on all countries for each three metrics, and resulted were output into an interactive chart:



□ is_normal

area: Ireland

	item	area	element	sample_size	pval	is_normal
27	Barley	Ireland	Area harvested	62	3.221979e-07	False
105	Oats	Ireland	Area harvested	62	4.248811e-11	False
106	Oats	Ireland	Production	62	4.763887e-07	False
107	Oats	Ireland	Yield	62	4.790465e-05	False
180	Potatoes	Ireland	Area harvested	62	1.264243e-05	False
181	Potatoes	Ireland	Production	62	1.991013e-06	False
182	Potatoes	Ireland	Yield	62	8.749854e-04	False
255	Sugar beet	Ireland	Area harvested	57	1.336294e-09	False
256	Sugar beet	Ireland	Production	57	6.833475e-06	False
327	Wheat	Ireland	Area harvested	62	2.305282e-02	False
328	Wheat	Ireland	Production	62	3.221720e-02	False
329	Wheat	Ireland	Yield	62	7.925776e-04	False

☑ is_normal

area: Ireland

	item	area	element	sample_size	pval	is_normal
28	Barley	Ireland	Production	62	0.117226	True
29	Barley	Ireland	Yield	62	0.065770	True
257	Sugar beet	Ireland	Yield	57	0.578840	True

Figure 4: Shapiro Normality Tests Interactive Output. Most of the data is not normally distributed (most likely due to exponential distribution in case of Production output)

Unfortunately, only a few countries, crops and metrics follow the normal distribution, so parametric tests could be applied. It was found with 5% significance that, for example, Sweden and Ireland exhibit the same barley crop production post year 2006.

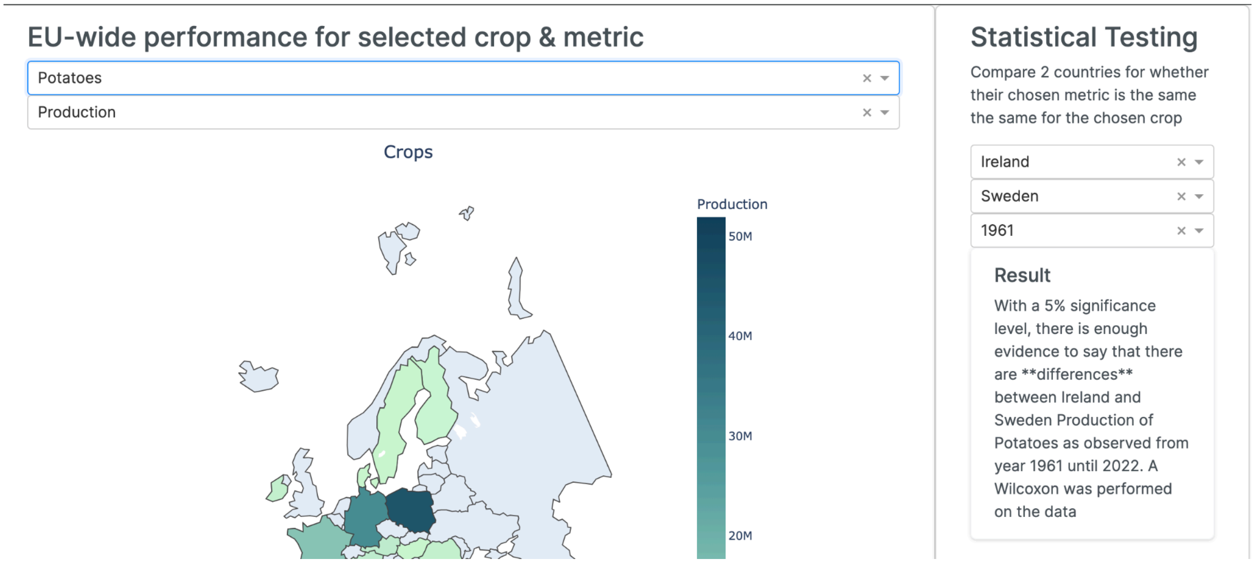
ANOVA test performed on Oat Yield across Ireland, Germany, France, Belgium, Denmark, Netherlands showed that At 5% significance level there is NOT enough evidence to suggest that the average yields are the same across the countries.

However, for comparison between the rest of the countries, nonparametric tests would need to be used, such as Wilcoxon for comparing 2 groups and Kruskall-Wallis for comparison of multiple groups (Moore, 2017).

In order to scale the use of statistics for this study, the decision was made to incorporate the statistical testing into the Dashboard. The dashboard performs a normality check on the two countries provided, and either runs a t-test (when H0 for normality of both county's observations can be accepted), or a Wilcoxon test (please see the update_stats function in dashboards.ipynb). An extension to this dashboard would be to allow for comparison between multiple countries, and adding Levene testing for having equal variances to either run ANOVA test, or Kruskal-Wallis in case the data is non-homogenous.

Based on the findings presented in the Jupyter notebook, most of the tests would appear to be Wilcoxon and Kruskall-Wallis, as most of the observations are either non-normally distributed or do not pass Levene test.

The results below demonstrate how the dashboard can be used for statistical comparisons:



EU-wide performance for selected crop & metric

Potatoes

Production

Statistical Testing

Compare 2 countries for whether their chosen metric is the same the same for the chosen crop

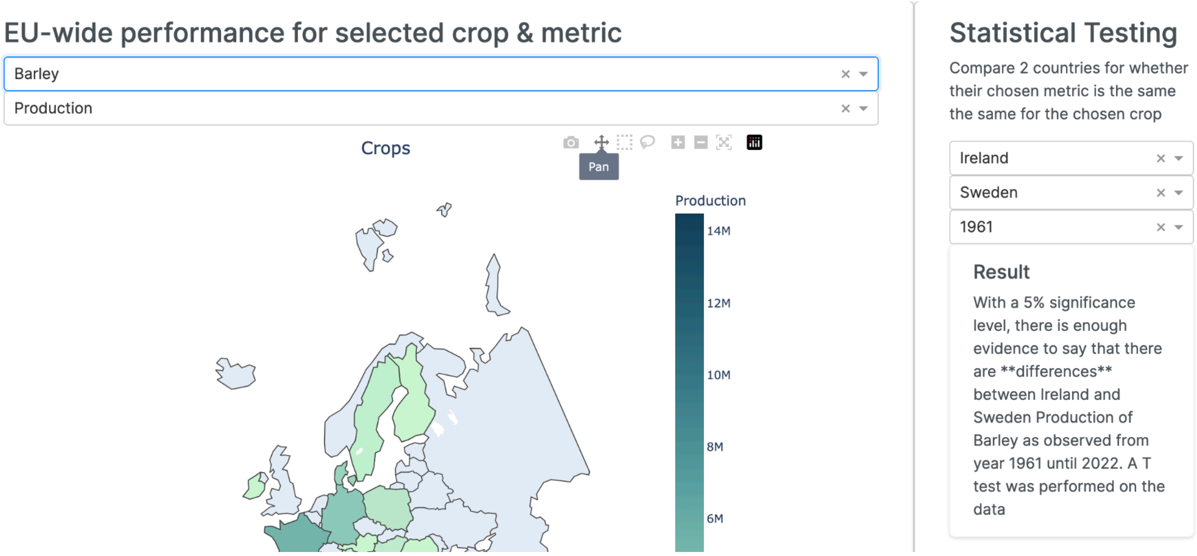
Ireland

Sweden

1961

Result

With a 5% significance level, there is enough evidence to say that there are **differences** between Ireland and Sweden Production of Potatoes as observed from year 1961 until 2022. A Wilcoxon was performed on the data



EU-wide performance for selected crop & metric

Barley

Production

Statistical Testing

Compare 2 countries for whether their chosen metric is the same the same for the chosen crop

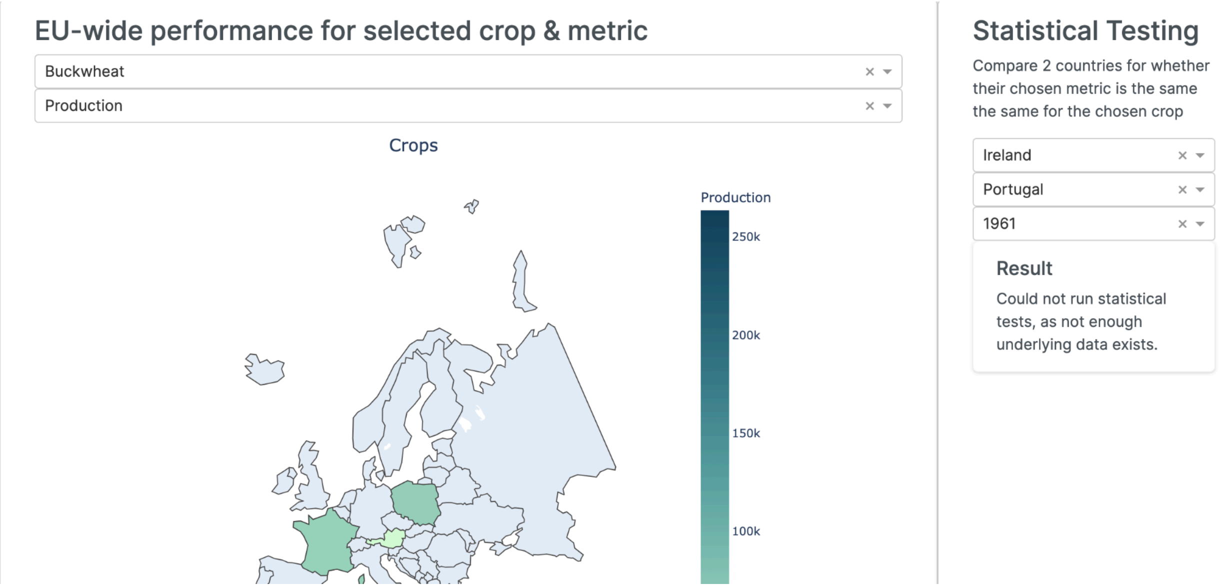
Ireland

Sweden

1961

Result

With a 5% significance level, there is enough evidence to say that there are **differences** between Ireland and Sweden Production of Barley as observed from year 1961 until 2022. A T test was performed on the data



EU-wide performance for selected crop & metric

Buckwheat

Production

Statistical Testing

Compare 2 countries for whether their chosen metric is the same the same for the chosen crop

Ireland

Portugal

1961

Result

Could not run statistical tests, as not enough underlying data exists.

These findings show the significance of relying on statistical methods to draw accurate conclusions, as visual appearances on graphs can appear to be deceptive. Incorporating statistical tests into data analysis tools enables users to make unbiased and reliable interpretations. Thus, elements of statistical analysis have been integrated into the final Dashboard to enhance its utility and credibility.

Additionally, since the ML part of this study relies on time series analysis, the importance of data being stationary vs non-stationary when choosing a time series machine learning algorithm, as most methods assume time series can be rendered approximately stationary (Box, 2015). As respective library documentation suggests, many time series models, like ARIMA, assume the underlying data are stationary. Non-stationary data will violate the model assumptions and often result in unreliable predictions (ming-zhao.github.io).

So, Adfuller tests were conducted, confirming the data's non-stationarity, thereby limiting the suitability of many time series models for analysis. For this study, thus, Meta Prophet ML model was chosen, based on this statistical test, since Prophet doesn't require stationary time series: a trend component is generated natively (Taylor, 2017).

Finally, confidence intervals were used to determine, for example, oat production in Ireland from 2006 onwards. These ascertain the range of the true population parameter, which for example in Ireland, with 95% confidence interval are at (159225, 191414) tonnes per year for Oat production. These confidence levels feed into the understanding for the Time Series models created by Meta Prophet, as they provide confidence levels for their predictions.

Data Preparation & Visualisation

Data Acquisition

The data for this study was sourced from two primary sources:

1. FAO's FAOSTAT platform (https://www.fao.org/faostat/en/), subject to licensing detailed at https://www.fao.org/contact-us/terms/db-terms-of-use/en/#:~:text=LICENSES.

   The CC BY-NC-SA 3.0 IGO license mandates proper attribution to FAO for any work produced or data re-disseminated, with citation format specified.

2. Data from Reddit (https://www.reddit.com/) was obtained adhering to Reddit's User Agreement (www.redditinc.com, n.d.), which prohibits commercial use without permission and obligates anonymization of personally identifiable information for user protection and ethical research. So, minimal Post and Comment attributes were saved for this study to mitigate the risk of storing Personally Identifiable Information.

Open-source models used in this study require citations, including Meta Prophet, under the MIT license.

Additionally, the BERT model used for sentiment analysis (bertweet-base-sentiment-analysis) is available for non-commercial use and scientific research purposes only. Since the latter is trained with third-party datasets and are subject to their respective licenses, with many models on Huggingface released under open-source licenses such as MIT, Apache 2.0, or Creative Commons, permitting academic and commercial use with specified conditions, including proper attribution and limited liability for the authors.

Obtaining certain data posed challenges, particularly for custom queries on Reddit or FAOSTAT, which seemed to lack an API and necessitated manual retrieval.

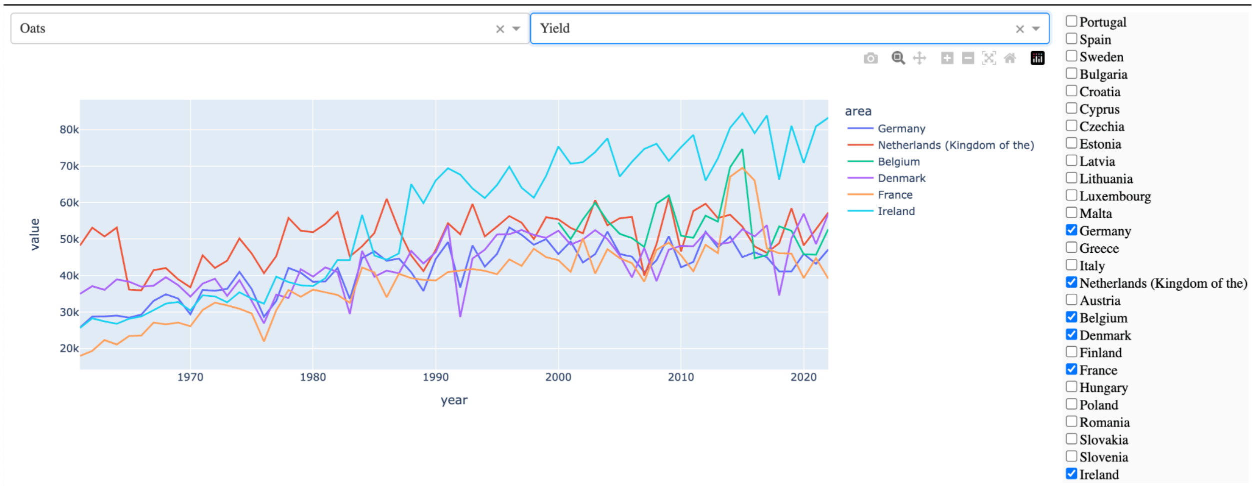
Data Preparation

Data Cleaning: This process involves removing or correcting erroneous data, removing duplicates and irrelevant observations, and dealing with missing values. In Python, libraries like pandas are often used for data cleaning. The most common issue with FAOSTAT data appeared to be data structure, naming conventions, and missing values.

Data Visualization: Seaborn was used for EDA and preliminary analysis, however, it appeared ot be quite limiting when analysis data across multiple countries and hundreds of crops. Therefore, Plotly was the primary data visualization library for this study, which allowed for data filtering based on different criteria and providing annotations for each data point.

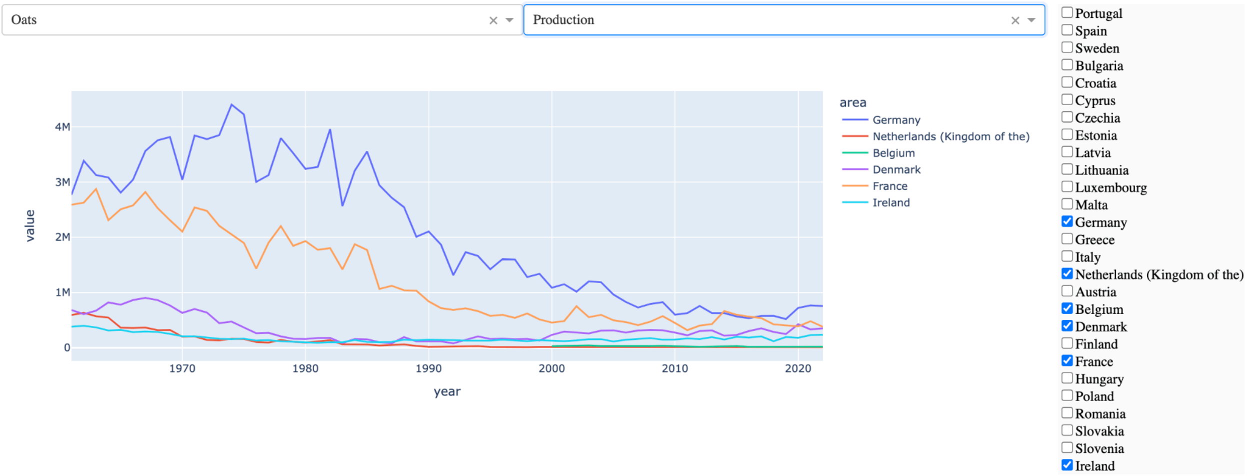
Exploratory Data Analysis

As part of exploratory data analysis, the need arose to break data down into percentiles as this was necessary to partition data (Moore, 2017), denoise it and allow for understanding the distribution of data and identifying outliers.


Oats

Yield

☐ Portugal
☐ Spain
☐ Sweden
☐ Bulgaria
☐ Croatia
☐ Cyprus
☐ Czechia
☐ Estonia
☐ Latvia
☐ Lithuania
☐ Luxembourg
☐ Malta
☑ Germany
☐ Greece
☐ Italy
☑ Netherlands (Kingdom of the)
☐ Austria
☑ Belgium
☑ Denmark
☐ Finland
☑ France
☐ Hungary
☐ Poland
☐ Romania
☐ Slovakia
☐ Slovenia
☑ Ireland



Oats

Production

☐ Portugal
☐ Spain
☐ Sweden
☐ Bulgaria
☐ Croatia
☐ Cyprus
☐ Czechia
☐ Estonia
☐ Latvia
☐ Lithuania
☐ Luxembourg
☐ Malta
☑ Germany
☐ Greece
☐ Italy
☑ Netherlands (Kingdom of the)
☐ Austria
☑ Belgium
☑ Denmark
☐ Finland
☑ France
☐ Hungary
☐ Poland
☐ Romania
☐ Slovakia
☐ Slovenia
☑ Ireland

Figure 5: Dynamically explore the dataset using Plotly & Dash. Allows to Compare any number of countries for a given crop and metric

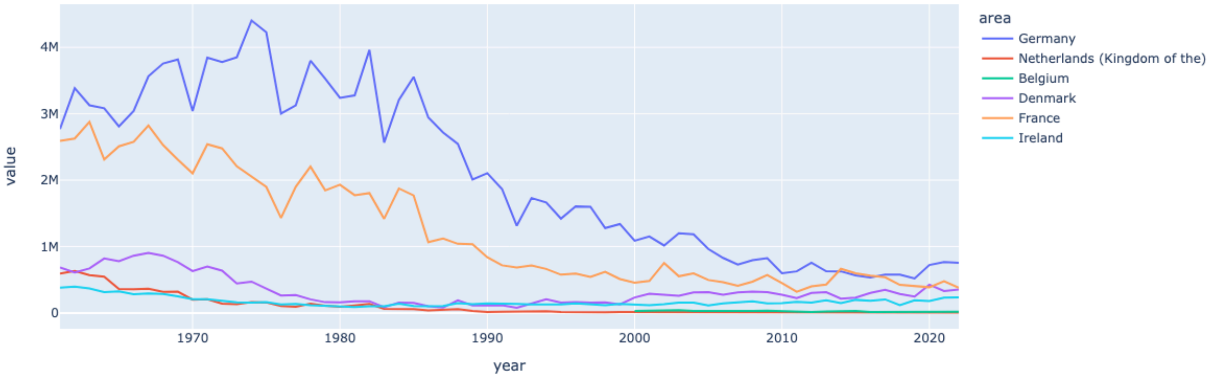


Figure 6: Results of preliminary graphic can give visual hints as to the countries and timelines which could be explored for statistical comparisons

One of Tufte's principles advocates for Annotations and Labels, emphasizing the direct labeling of key points, trends (Iversen, 1988), and outliers on the graph whenever feasible, instead of relying on legends that necessitate viewer cross-referencing, ensuring labels are inconspicuous yet easily legible.

Moreover, maintaining consistent scales and formats across multiple graphs enhances intuitive comparisons.

This is where the Plotly library stands-out. Plotly is advantageous due to its great interactivity, allowing users to explore data through zooming, panning, and hovering to see detailed information. Its annotation features enable clear, customizable labeling of data points and trends directly on the graph (Dabbas, 2021). Additionally, Plotly supports integration with various programming languages and platforms, making it a versatile tool for creating dynamic, web-based visualizations, and for this study Dash was used in conjunction with Plotly.

In order to prepare the data for consumption by machine learning algorithms, Ordinal Encoding was used, and the data was scaled as part of the testing. The final results (in particular, when using the Prophet Forecasting model) did not require data preparation apart from imputing.

Exploratory data analysis also helped find correlations between features to make judgment on which feature are best to use for the machine learning models.

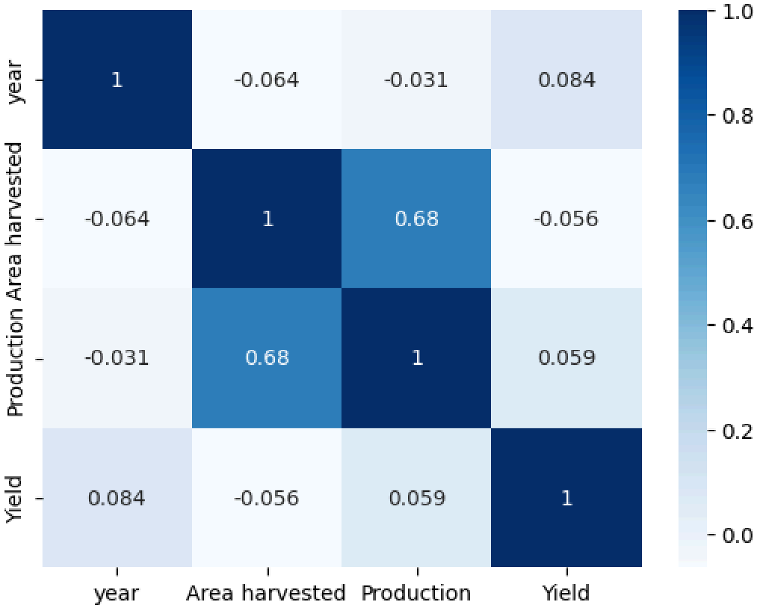


Figure 7: Correlation Matrix can help understand which features show the most variance in the underlying data and perform feature selection for the ML models.

Machine Learning

Time Series Analysis

Model Choices: Due to limited underlying data, a univariate time series analysis has been used. In order to achieve a multivariate analysis, more features would need to be added such as population growth data, farming structure in different countries, environmental changes and impact on the growth of crops, and different economic factors. The challenge with this approach would that to make any meaningful future predictions, all of this data would need to be supplied for any prediction to be made (Bishop, 2006), which would not be feasible as it would require the user of that model to supply all the data for years not yet observed and would defeat the usefulness of the deliverables.

Multiple approaches were explored for the univariate analysis:

	Random Forest	Linear Regression	Meta Prophet
Measurement	R^2 of 0.98	R^2 of -1.68	RMSE of 20924 *different for different models trained
Challenges	Can only go one step ahead and requires custom optimizations / data transformations	Inaccurate, unless used with Polynomial Features	N / A
Fine-tuning Techniques	GridSearch can be performed	Polynomial features could improve performance	Requires little fine tuning unless the data is seasonal

Below are the graphs that demonstrated challenges faced when using algorithms that are not specific to time series analysis and which would require a lot of data preparation or fine tuning to work.

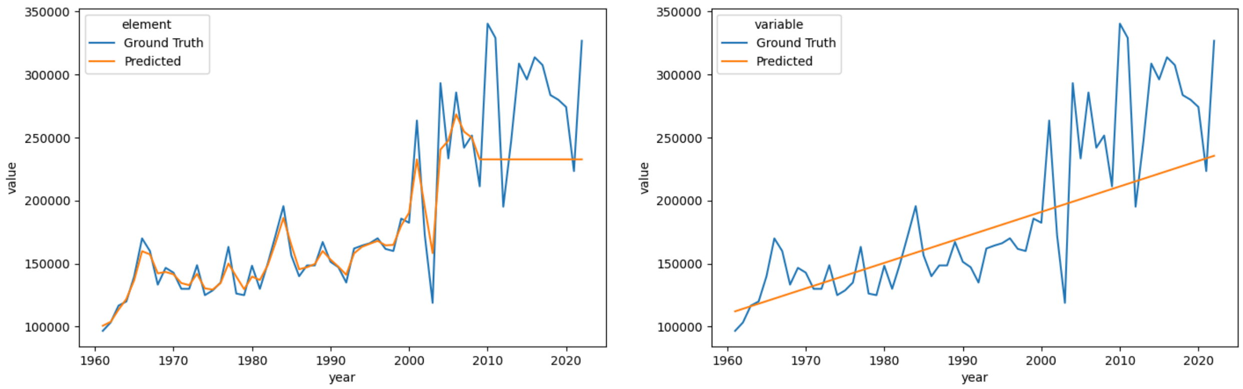


Figure 8: Random Forest and Linear Regression produce poor results even after fine tuning

In order to solve these challenges, a mode that is specific to the task should bemused. The Prophet model was trained on the different combinations of Crops & EU Countries and the results were output into the `prophet_models` folder. This prevents from retraining the model every time the model needs to be used, and the models can be loaded in from the folder using the custom load_in_prophet_models function.

While random forest and linear regression were evaluated using the R^2, the Prophet model evaluation relies on the RMSE - a single number to judge a model’s performance, whether it be during training, cross-validation, or monitoring after deployment (James, 2017). Root mean square error is one of the most widely used measures for this. It is a proper scoring rule that is intuitive to understand and compatible with some of the most common statistical assumptions (Gneiting, 2007).

Below are the examples of the predictions made by the Prophet model for different countries and crops, as well as EU totals. The advantage of prophet is that is provides an uncertainty interval that allows the users of these predictions to make more rational and weighted decisions based on the recommendations coming from the model.

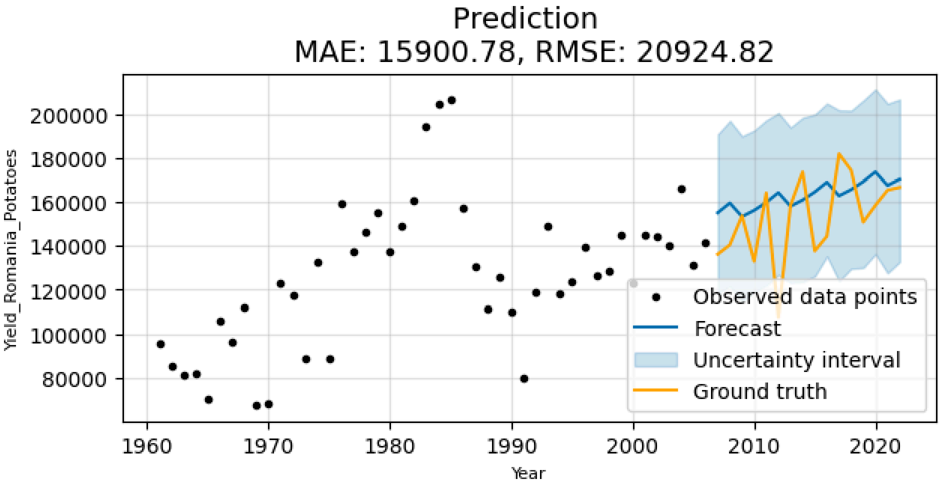


Figure 9: Potation Yield in Romania Meta Prophet Model Training Results

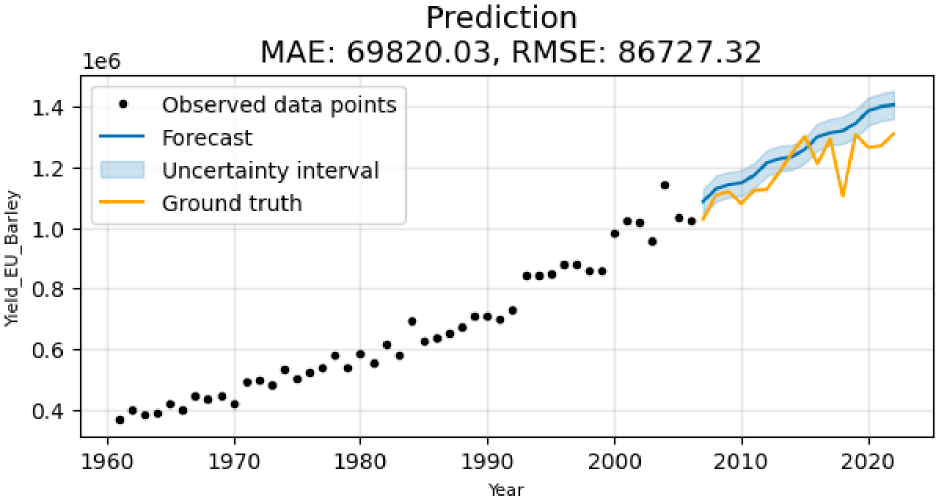


Figure 10: Barley Yield in EU Meta Prophet Model Training Results

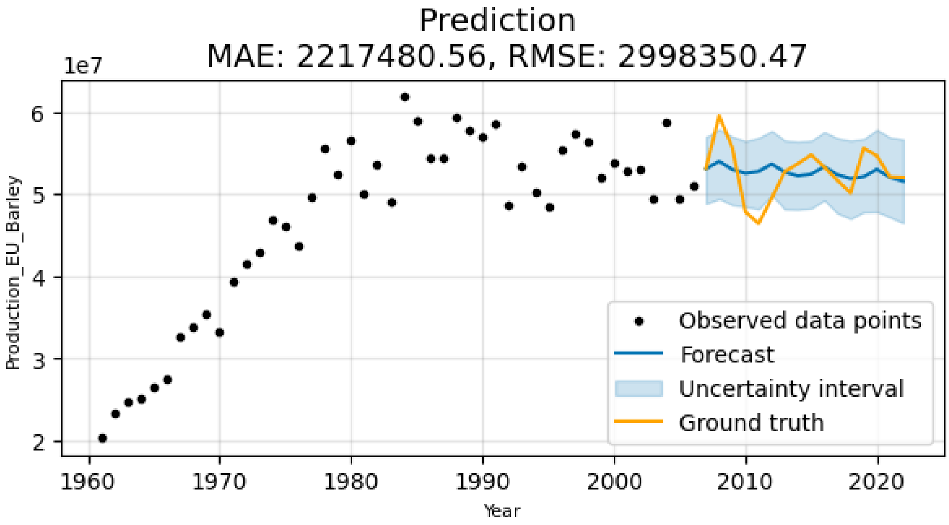


Figure 11: Barley Production in EU Meta Prophet Model Training Results

EU total seems to have best performance across the different crops, based on the testing performed for this study.

The pre-trained models were saved and can be reused, and / or presented in the Dashboard as demonstrated below.

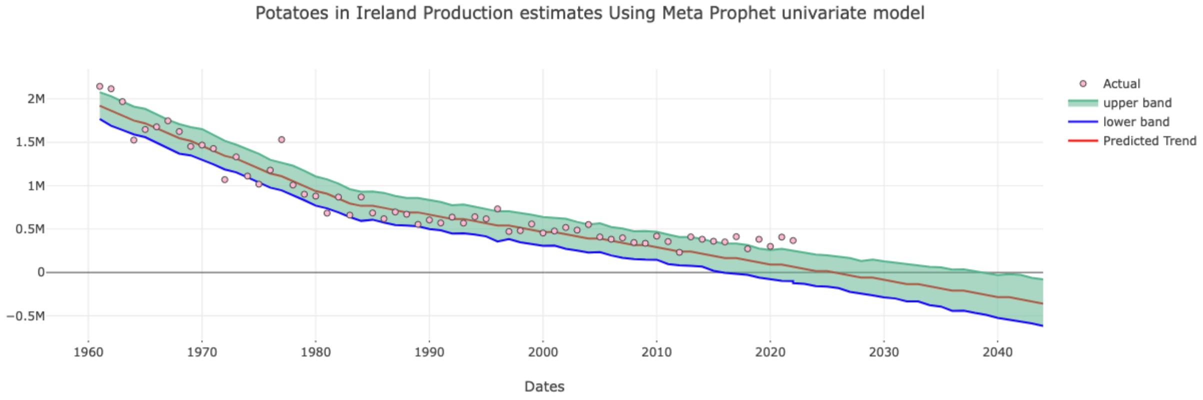


Figure 12: Prototype of predictions to be used in the final Dashboard

Algorithm cross validation was performed using rolling window. This type of validation uses a fixed-size window that moves over the time series data. At each step, a new model is trained on the data within the window, and its performance is evaluated on the subsequent data points (Box, 2015), as implemented for this study in the graph below:

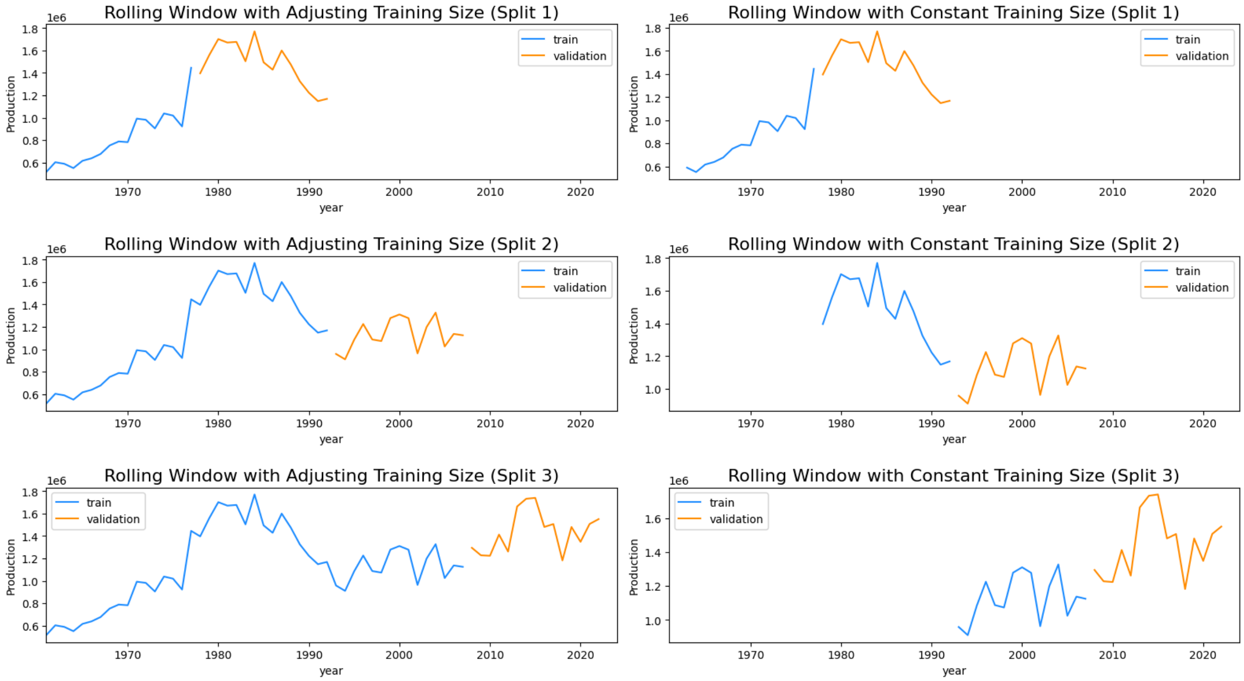


Figure 13: Rolling window validation that can be used for Time series modeling with Linear Regression or Meta Prophet

Sentiment Analysis & Topic Modelling

In order to perform sentiment analysis, a dataset had to be developed. Developing a dataset for sentiment analysis using data from the Reddit API involved extracting comments and posts from certain subreddits for specified queries (Tiago Rocha-Silva, 2023). This process requires setting up API calls to fetch the data, preprocessing the text for analysis, and possibly using automated or manual methods to annotate the sentiment of each entry.

After the data was gathered it was stored in CSV files to eliminate the need to making new calls to the Reddit API every time the notebook needs to be restarted / re-run.

After that, the data was loaded into dataframes, and the text entries were labeled based on their emotional tone or sentiment. To perform the labelling, the NLTK vader_lexicon package was used providing scores for positive, negative, neutral, and compound sentiment (Hutto, 2014). Different thresholds were tested out for the compound score to receive the sentiment labels. Following a detailed study of Perkins, 2011, lemmetization and data clean up had to be performed to ensure more authentic results for both sentiment, and topic analysis.

To cross-validate its results, the best approach would be to compare sentiment scores generated by VADER with human-labeled sentiment data, calculating metrics such as accuracy, precision, recall, and F1-score to assess its performance (Hutto, 2014). However, the approach used for this study was to use a pre-trained BERT sentiment analysis model (Pérez, 2021) on a sample of 1500 Reddit posts for cross-validation, yielding a 0.67% overlap between the two models.

Finally, The bag of words model was used to represent text data by converting it into a matrix of token counts. This approach disregards grammar and word order, focusing on the frequency of words. It is related to MultinomialNB (Multinomial Naive Bayes) because this classifier is commonly used with bag of words representations to predict categories based on the distribution of word frequencies in the text data. So, in this study the MultinomialNB was thus trained on the data labelled by Vader and achieved a **0.61** accuracy score even after performing a GridSearchCV.

Additionally, in order to gain more insights into the gathered data and what the above mentioned sentiment relates to, a topic analysis was conducted using BERTopic pre-trained model (Maarten Grootendorst).

The key results of the topics are visualized in Figure 14 below:

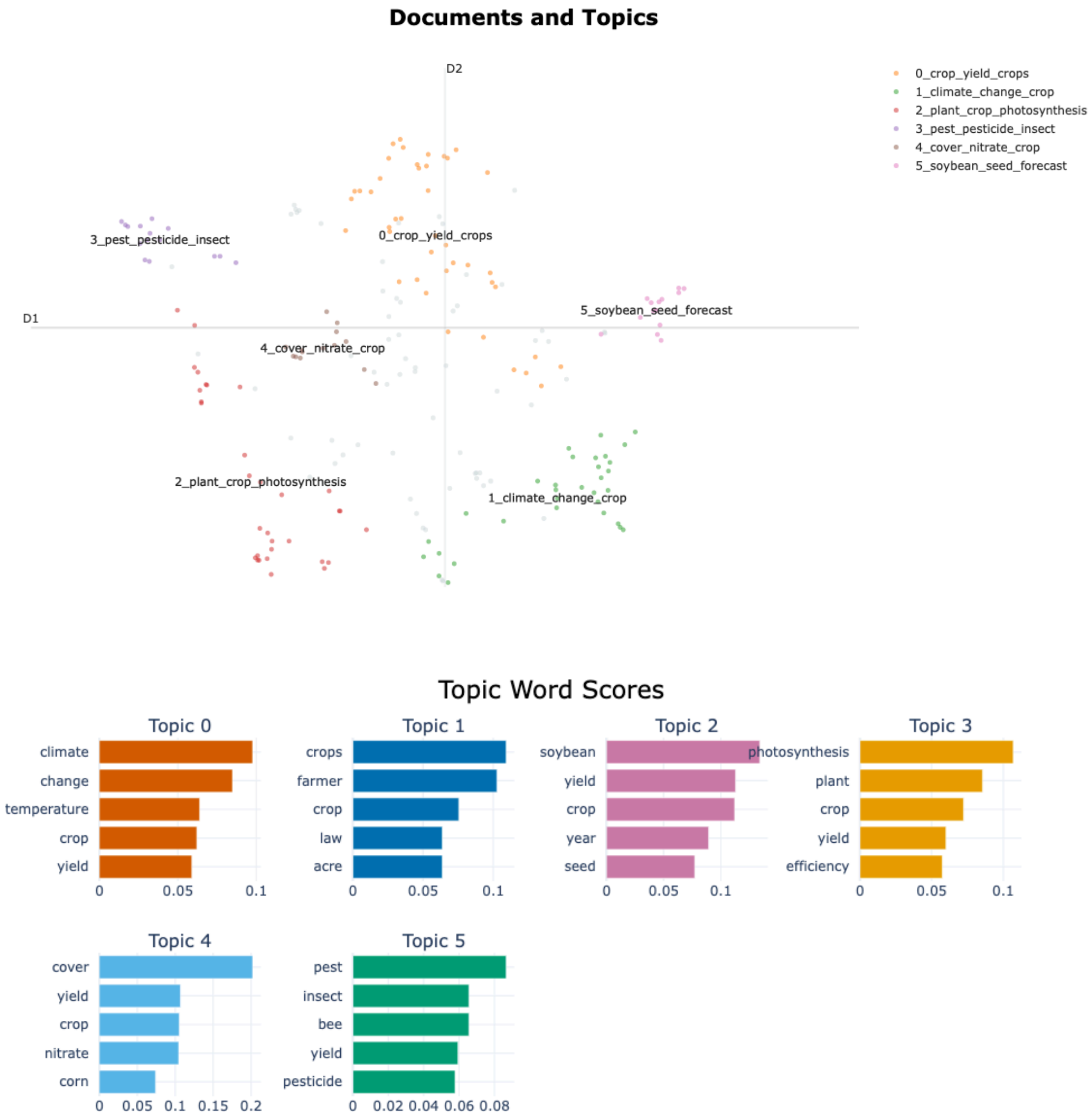


Figure 14: Visualizing common reddit topics related to crop farming in Ireland

Topic analysis surfaces some of the major concerns by the farmers, in particular around the impact of climate change on crop yields, as well as issues with pollinators caused by the use of pesticides. These can be used to inform public policies in Ireland and worldwide.

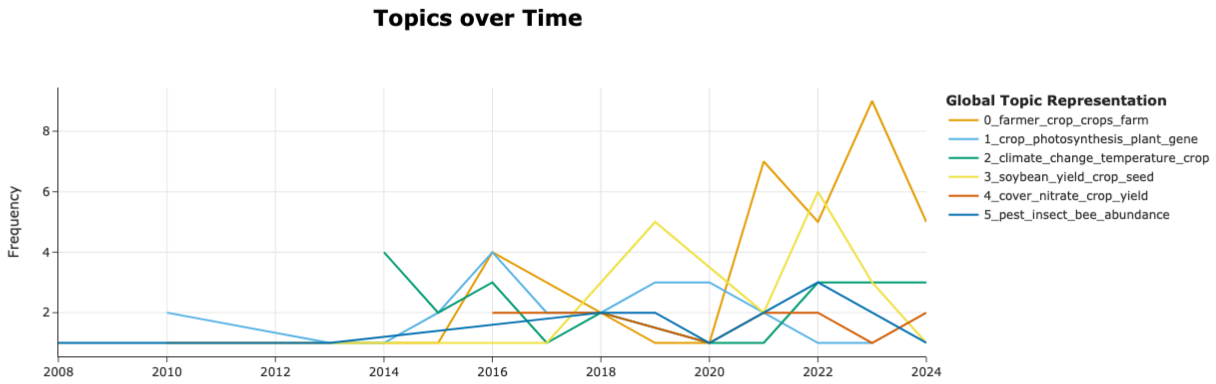


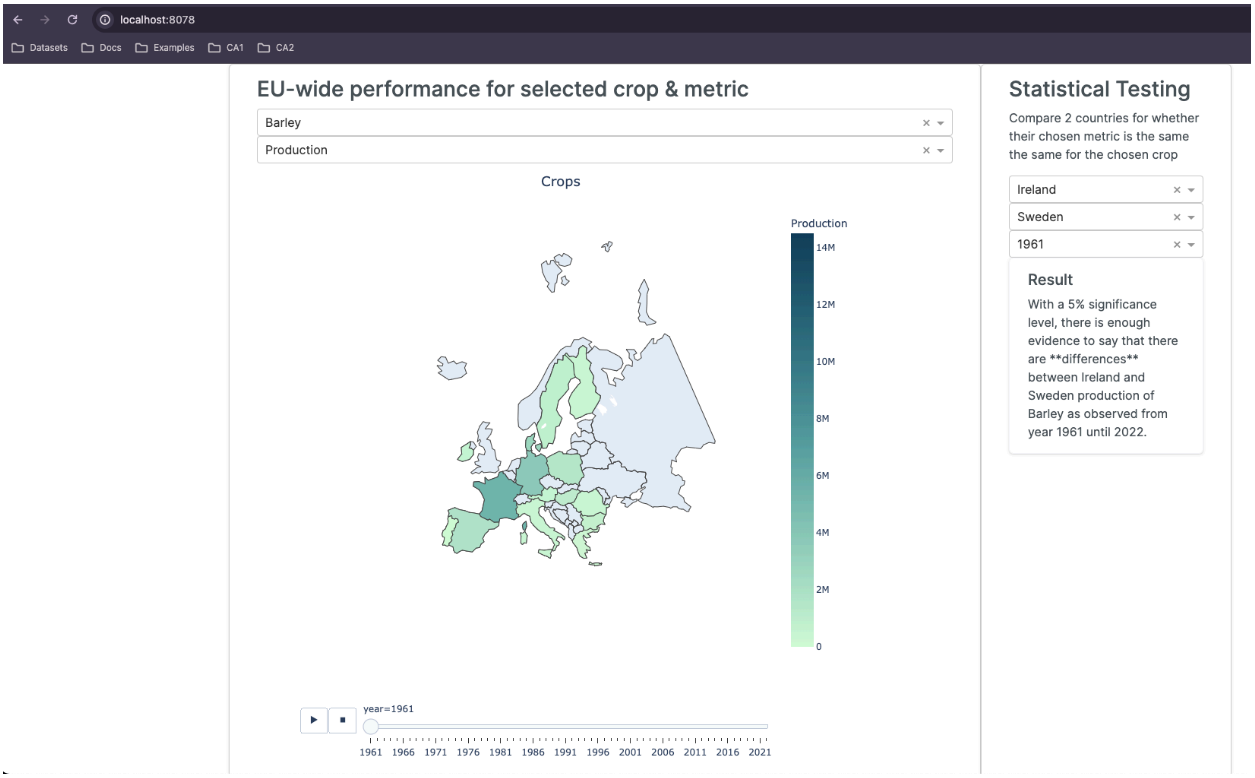
Figure 15: The timeseries combines with topic analysis shows an increased emphasis on farmer crop issues and an assumed connection to climate change, at least as observed by the public

Dashboarding Results of The Study

In order to deliver results of this study to a potential end-user (farmer, business, policy maker, etc), the data needs to be visualised and easily accessible and interpretable.

Dash and Plotly were used to visualize and deliver the results of the Data Preparation, feature engineering, model training using Meta Prophet, described above.

The resulting processed dataframes were saved into files, including the ML models were serialised into json files and read from the files in the Dashboard notebook. Below are the screenshots of the dashboard. It provides for interactivity and follows Tuftes principles (Iversen, 2018).



EU-wide performance for selected crop & metric

Barley

Production

Statistical Testing

Compare 2 countries for whether their chosen metric is the same the same for the chosen crop

Ireland

Sweden

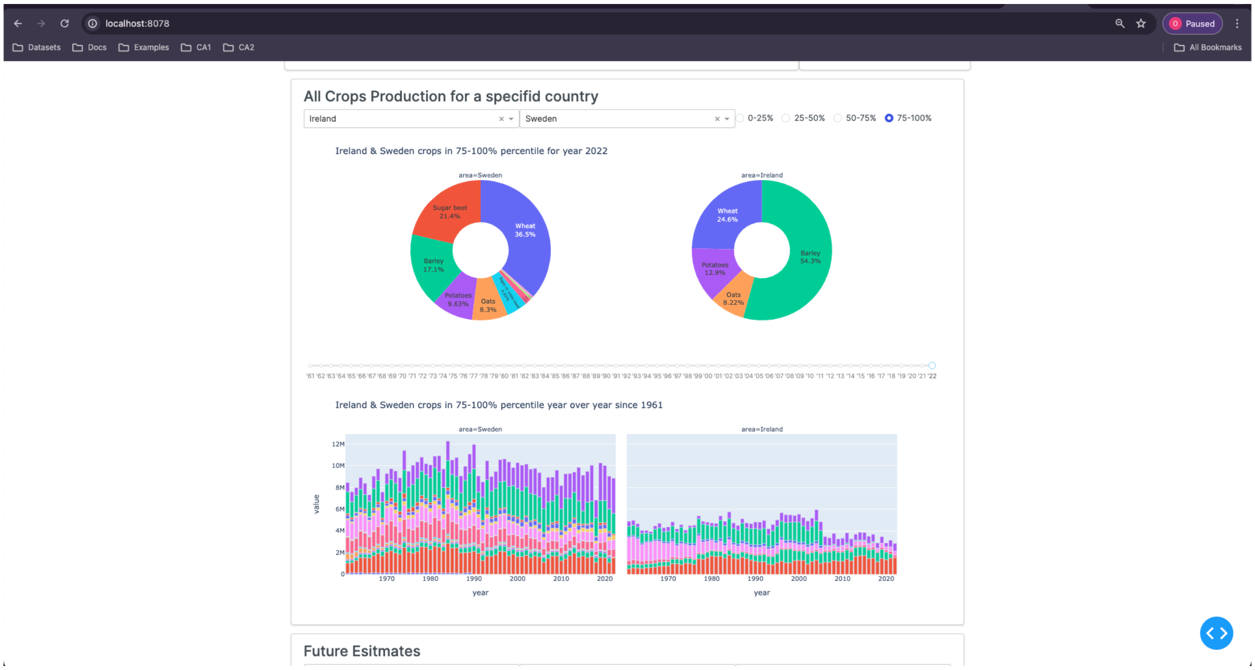
1961

Result

With a 5% significance level, there is enough evidence to say that there are **differences** between Ireland and Sweden production of Barley as observed from year 1961 until 2022.

Figure 16: Geoplot of crop Production / Yield / Area Harvested for a selected country to the left, with Statistical testing to the right. Statistical testing provides a country comparison providing information on differences / similarities in the selected metrics for 2 countries, as observed from the year selected by the user until 2022.

Use consistent scales and formats across multiple graphs to make comparisons intuitive, which is one of Tufte's principles (Iversen, 2018), hence min() and max() scale is derived for all years.



All Crops Production for a specifid country

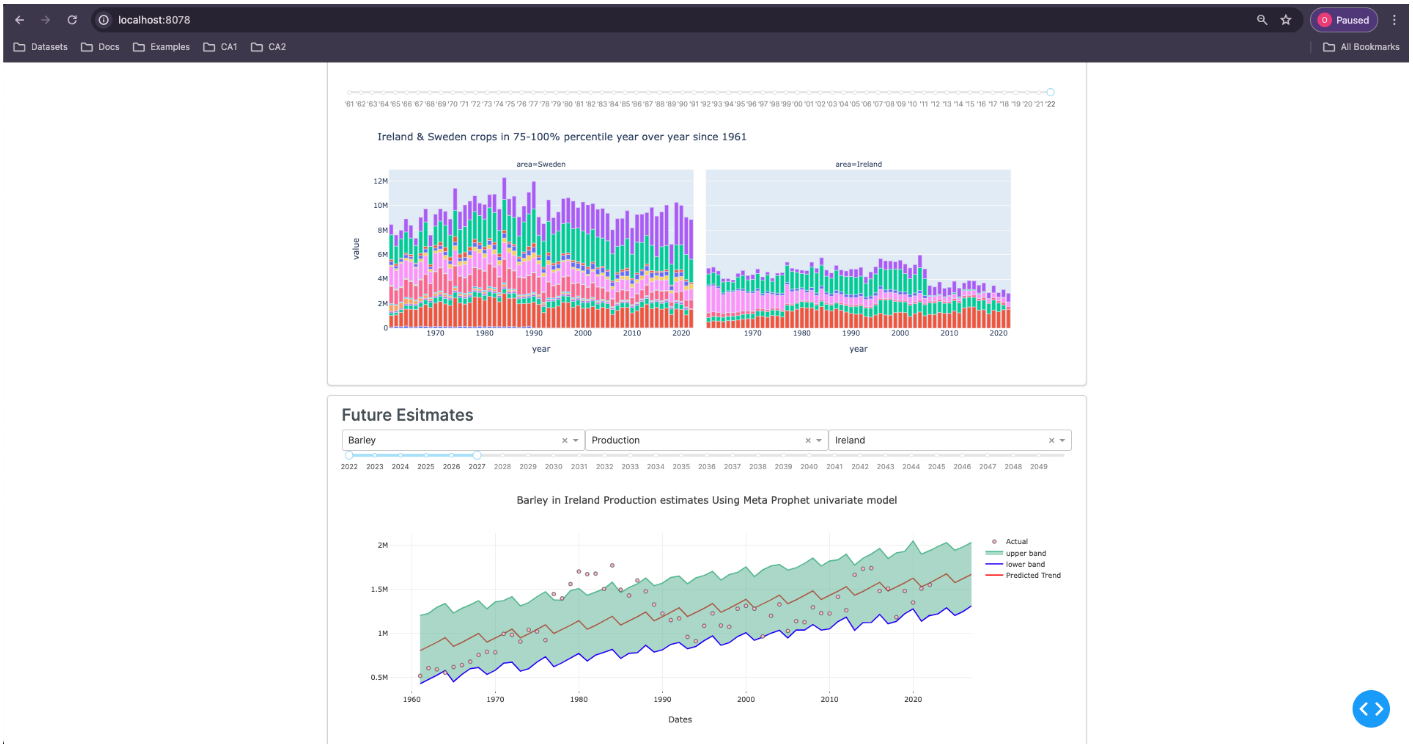
Ireland × ▾ Sweden × ▾ ○ 0-25% ○ 25-50% ○ 50-75% ● 75-100%

Ireland & Sweden crops in 75-100% percentile for year 2022

Ireland & Sweden crops in 75-100% percentile year over year since 1961

Future Esitmates

Figure 17: Allows to compare production of top / bottom crops in 2 selected EU countries. Provides pie charts with crops proportions for a selected year and a timeline for production of the crops over the given timescale (from 1961 to 2022). Choosing the percentiles allows to view the data at an appropriate scale and removes the noise crops which are produced at different scales



Ireland & Sweden crops in 75-100% percentile year over year since 1961

Future Esitmates

Barley × ▾ Production × ▾ Ireland × ▾

Barley in Ireland Production estimates Using Meta Prophet univariate model

Figure 18: Using Meta Prophet model, allows to forecast future production / yield for a particular crop in a selected country until year 2050.

Conclusion

The study successfully compares crop yield and production across various EU countries, with a specific focus on Ireland and its key agricultural crops. Utilizing descriptive and exploratory data analysis, the study identifies patterns in crop output and predicts future yields. Machine learning models, particularly Meta Prophet, were employed to enhance the accuracy of these forecasts. Data visualization techniques, such as interactive dashboards, were integrated to facilitate the use of this information by policymakers, farmers, and businesses. The findings underscore the importance of using robust statistical methods and advanced forecasting tools to make informed decisions. Overall, this study provides valuable insights and practical tools for stakeholders in the agricultural sector.

References

Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer.

Box, G.E.P. and Al, E. (2015). Time series analysis : forecasting and control. Hoboken, New Jersey: John Wiley & Sons.

Hutto, C. and Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), pp.216–225. doi:https://doi.org/10.1609/icwsm.v8i1.14550.

Dabbas, E. (2021). Interactive Dashboards and Data Apps with Plotly and Dash. Packt Publishing Ltd.

Iversen, I.H. (1988). TACTICS OF GRAPHIC DESIGN: A REVIEW OF TUFTE'S THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION. Journal of the Experimental Analysis of Behavior, 49(1), pp.171–189. doi:https://doi.org/10.1901/jeab.1988.49-171.

Franzke, Aline Shakti, Bechmann, Anja, Zimmer, Michael, Ess, Charles and the Association of Internet Researchers (2020). Internet Research: Ethical Guidelines 3.0. https://aoir.org/reports/ethics3.pdf

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning. New York, Ny Springer New York.

Gneiting, T. and Raftery, A.E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association, 102(477), pp.359–378. doi:https://doi.org/10.1198/016214506000001437.

Maarten Grootendorst (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.2203.05794.

ming-zhao.github.io. (n.d.). ARIMA Models — Business Analytics 1.0 documentation. [online] Available at: https://ming-zhao.github.io/Business-Analytics/html/docs/time_series/arima.html.

Moore, D.S., Mccabe, G.P. and Craig, B.A. (2017). Introduction to the practice of statistics. New York, Ny: W. H. Freeman And Company.

Pérez, J.M., Rajngewerc, M., Giudici, J.C., Furman, D.A., Luque, F., Alemany, L.A. and Martínez, María Vanina (2021). pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.2106.09462.

Perkins, J. (2011). Python Text Processing with Nltk 2.0 Cookbook. Packt Publishing Ltd.

Taylor, S.J. and Letham, B. (2017). Forecasting at scale. [online] doi:https://doi.org/10.7287/peerj.preprints.3190v2.

Tiago Rocha-Silva, Nogueira, C. and Rodrigues, L. (2023). Passive data collection on Reddit: a practical approach. Research Ethics. doi:https://doi.org/10.1177/17470161231210542.

FAO.[Crops and livestock products]. License: CC BY-NC-SA 3.0 IGO. Extracted from: [https://www.fao.org/faostat/en/#data/QCL]. Data of Access: 25-04-2024.

Raschka, S. and Vahid Mirjalili (2017). Python machine learning : machine learning and deep learning with Python, scikit-learn, and TensorFlow. Birmingham (Uk): Packt Publishing.

www.redditinc.com. (n.d.). User Agreement - September 25, 2023 - Reddit. [online] Available at: https://www.redditinc.com/policies/user-agreement-september-25-2023.

Index of Figures