WFP Supply Chain Planning & Optimization Unit Technical Assessment Report

Kalogeropoulos Dionisis

1 Introduction

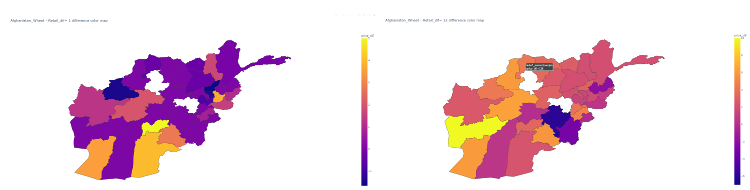
For the purpose of this study an analysis on monthly food prices will be produced by examining an old repository that consists of global nationwide datapoints. Each row of this dataset provides a datapoint that consists of multiple variables that can be used to extract insights on food allocation per country per region.

The dataset consists of a 215MB file, with 2050638 rows and 18 columns / features. This study focuses on a specific region and commodity, namely Region : Afghanistan Commodity : Wheat − Retail The focus produces a reduced dataset that consists of 2312 rows and 18 columns / features.

2 Exploration Phase

2.1 Geography Analysis

In order to get a grasp of the dataset and the insights that can provide a first step would be to chart the information in a map presenting heatmaps of change. The charts provided are on a difference scale of 1 month and yearly.



(a) 1 month dif. (b) 12 month dif.

Figure 1: Afghanistan wheat price difference heatmap.

Here we can see that the past month the south regions increased their price while most of the rest of the regions decreased. Moreover for the 1 year period a trend of increase is being noticed except for some regions in the east where the price was decreased.

A note to be taken is that the last date/month that the differences are being performed is July which is one of the months that the value is close to its highest value.

2.2 Timeseries Analysis

A second step would be to look the prices as timeseries and dive into the statistical properties of these signals. The focus here is Afghanistan and all its regions on average. The following represents the dataset:

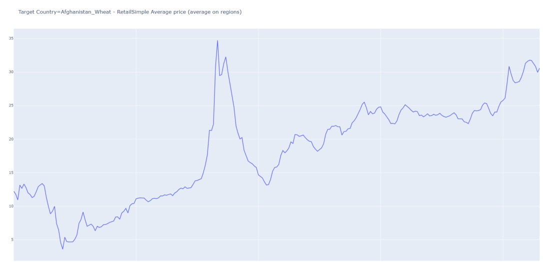
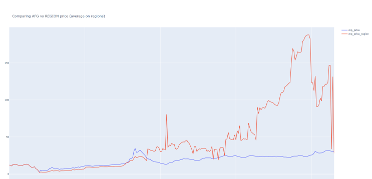
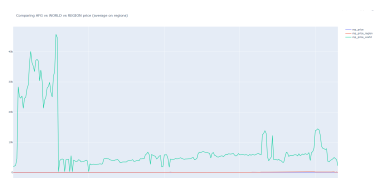


Figure 2: Afghanistan Wheat-Retail price.

Here an upwards trend is evident and also a seasonality factor. Let's compare this timeseries versus the Asia region and the rest of the world as well.



(a) Afghanistan vs Asia region

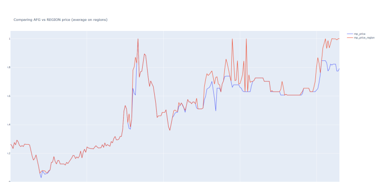


(b) Afghanistan vs World

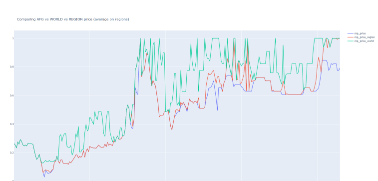
Figure 3: Wheat price on different scales.

An immediate insight is that prices in Asia skyrocket after 2010 compared to Afghanistan region. Moreover prices in a global scale seem very volatile as for example prices in Somalia can reach 10k.

And with normalizing via MinMax scaling after hashing the countries. This scaling will collapse the data range to the 0-1 range and will provide the following plots:



(a) Afghanistan vs Asia region



(b) Afghanistan vs World

Figure 4: Normalized wheat price on different scales.

In order to increase the depth of the analysis a seasonal decomposition is being performed. The chart below shows the decomposition into trend, seasonality and residuals.

Figure 5: Afghanistan Wheat-Retail price seasonal decomposition.

It is evident from the previous plot that there is a seasonality effect with a period of 1 year. This is something we would expect as the study is on wheat commodity and validates the methodology.

3 Analysis and Prediction

3.1 Sourcing the commodity

A goal of this study would be to decide where would be ideal to source the commodity and thus a geo plot of recent prices around the region is presented and some hypotheses as well.

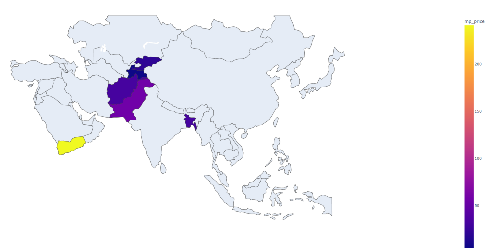


Figure 6: Asia region Wheat-Retail price and availability.

Due to incomplete data / unavailability of the commodity from all the neighboring countries, sourcing from **Tajikistan** would be a good recommendation as the price is 1/5 of the average of the country of Afghanistan.

When considering the sourcing of wheat for Afghanistan, several factors need to be taken into account and thus just studying the price will not be sufficient for a fair and optimal approach. These include geographical proximity, trade agreements, transportation infrastructure, cost, quality, and reliability of suppliers.

An alternative would be sourcing the commodity from **Kyrgyzstan** but the price difference (20/31) might not be sufficient to label this as an efficient trade due to the aforementioned factors.

Additionally, considerations such as political stability, diplomatic relations, and import/export regulations would influence the choice of a sourcing country. Therefore, further research and analysis would be necessary to determine the most suitable sourcing option of wheat for Afghanistan.

3.2 Feature Engineering & Forecasting price

The task of predicting time series is in general a non trivial task. In order to provide a basis an autoregressive model could be used (e.g. sarima) that uses the historical data to give a forecast and confidence intervals. There are numerous other options in the machine learning literature (supervised and unsupervised depending the data structure and transformations) that can give more accurate predictions but using only the price as feature and reducing the problem to 1 dimension increases the uncertainty of the outcome and reduces the predictive power of the model. To predict wheat prices for the next 12 months, in addition to price, several other parameters are essential. These parameters include:

Inventory levels: Monitoring the level of wheat reserves or stockpiles is important. High inventories can lead to lower prices as it indicates an abundance of supply, whereas low inventories may indicate potential price increases due to scarcity.

Climate and weather patterns: Droughts, floods, or extreme temperatures can affect crop yields and production. Adverse weather conditions often lead to higher prices due to reduced supply.

Economic indicators & Exchange rates: Macroeconomic factors such as inflation, interest rates, and GDP growth can influence wheat prices. Strong economic growth may lead to increased demand, while economic downturns can decrease demand.

Beside these factors that are ready to be transformed numerically and applied to any machine learning forecasting algorithm we have to mention some other factors that are crucial in creating a robust prediction system and need to be treated as well as **Trade agreements**, **Government policies** and **Market speculation**.

3.3 Forecasting uncertainty & Confidence Intervals

There are various sources of uncertainty that can impact the accuracy of a forecast. Some common sources of uncertainty include:

- **Model selection & Assumptions**: Forecasts of a specific model are based on a set of assumptions about future conditions. If these assumptions turn out to be incorrect, it can introduce uncertainty into the forecast.
- **Data quality**: The accuracy of a forecast is heavily dependent on the quality and reliability of the data used for analysis.
- **Parameter estimation**: Forecasting models require estimating parameters and the accuracy of these parameter estimates affects the accuracy of the forecast. Uncertainty arises if different parameter estimates lead to different forecasts.

To estimate the uncertainty of a forecast, one common approach is to calculate a confidence interval. To calculate a confidence interval, you can use methods such as bootstrapping, Monte Carlo simulation, or analytical approaches depending on the nature of the data and the forecasting model used. These methods involve generating multiple alternative forecasts based on different plausible scenarios or assumptions, and then calculating the distribution of forecast outcomes. It is important to note that many forecasting algorithms have pre-built in functions that calculate these confidence intervals from the training dataset (e.g. sarimax).

A basic prediction plot for the next 12 months trained via a SARIMAX (2,1,0)x(2,1,0)(12) model with confidence intervals:

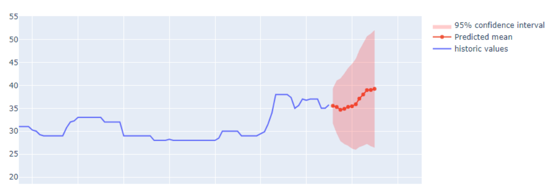


Figure 7: Afghanistan Wheat-Retail price prediction for the next 12 months.

4 Summary

This report is being presented as a basis and is not a complete solution to the problem. Through basic exploration and analysis we are able to decide how to navigate into sourcing the selected commodity and what are the important parts that will affect the analytic decision. The seasonality and trend will give us power over time decision processes and a region predictive modeling will be able to define a strategy over future decisions. Hopefully with more data, feature engineering and robust forecasting methods we will be able to achieve a noble cause and distribute resources fairly and efficient.