A REPORT ON ANALYSIS OF WEATHER AND FOOD PRICE

Prepared for
Invitation Only
Data Science Cohort

Prepared by
Rakibuzzaman Rahat

December 2, 2021

Introduction

As a final project for the "Invitation Only Data Science Cohort" course, we've been given 4 (four) datasets to choose from. These are as followings:

1. Weather and Food Prices Dataset
2. Google Playstore Dataset
3. Crunchbase Bangladesh Dataset
4. Growth Index of Bangladesh Dataset

This report focuses on the "Weather and Food Prices Dataset" of Bangladesh. This dataset contains different categories of goods (food), prices of foods, time and regional details with weather related to those regions and timeframe. As for a data analysis project, the first step is to define the goal of the project and collect data accordingly. Here we are given the prepared data, so after conducting a few brief analyses the objective of the project was defined.

Objective

To find patterns and relations in goods prices and sales in accordance with weather components and in respect to regions if any exists.

Dataset Analysis

This dataset contains 3 (three) categories of goods with sub-categories, these are as follows:

1. Cereals and tubers
    a. Rice (coarse, BR-8/ 11/, Guti Sharna) - Wholesale & Retail
    b. Wheat - Wholesale & Retail
    c. Wheat flour - Wholesale & Retail
    d. Rice (coarse, Guti Sharna) - Wholesale & Retail
    e. Rice (coarse) - Retail
    f. Rice (medium grain) - Wholesale & Retail
2. Oil and fats
    a. Oil (palm) - Retail
3. Pulses and nuts
    a. Lentils (masur) - Retail

Along with these goods, the above mentioned dataset contains goods prices, timestamps, locational data (Dhaka, Barisal, Khulna, Sylhet, Rajshahi), weather components. It's been divided into 3 (three) groups for further analysis:

1. Goods related (G-1)
2. Weather related (G-2)
3. Time & locations (G-3)

Serial No.	Goods related (G-1)	Weather Related (G-2)	Time & locations (G-3)
1	Name of goods	Rainfall	Regions
2	Category of goods	Relative humidity	Time (Year and month)
3	Price of goods	Cloud coverage	Locational data (Lat/Long)
4	Goods sales amount	Bright sunshine	Stations number (Regions)
5	Unit price of goods	Wind speed	Altitude
6		Max & Min Temperature	X & Y coordinates
Remarks	**Dependent variables**		**Independent variables**

Table-1: Categorization of dataset

From the G-3, few data points have been omitted which are:

1. Stations number - Refers to individual regions
2. X & Y coordinates

As we can see the dataset contains the price of goods over a period (years and months), Inflation rate [1] can be a major factor for price changes. Also, to compare Inflation rate with goods prices new calculated fields have been introduced i.e. Consumer Price Index (CPI) to calculate goods basis inflation rate namely "Individual Inflation Rate". The GDP growth rate [2] of Bangladesh has also been added to the dataset. These measures have been taken to enrich the dataset. So, the modified dataset table is as below:

Serial No.	Goods related (G-1)	Weather Related (G-2)	Time, locations & others (G-3)
1	Name of goods	Rainfall	Regions
2	Category of goods	Relative humidity	Time (Year and month)
3	Price of goods	Cloud coverage	Locational data (Lat/Long)
4	Goods sales amount	Bright sunshine	Inflation rate
5	Unit price of goods	Wind speed	Altitude
6	Individual inflation rate	Max & Min Temperature	GDP growth rate
Remarks	**Dependent variables**		**Independent variables**

Table-2: Categorization of dataset (modified)

Methodology

For conducting further analysis, the above mentioned 3 (three) groups are used. Remarks section of the table indicates that G-1 and G-2 contain dependent variables and G-3 contains independent variables in context to the dataset. In relation to G-3, G-1 can change. The price of goods in a single month can be different in different regions of the country which can be seen from the dataset and will be discussed in the analysis section. Also G-2 can change in relation to G-3, rainfall in Dhaka in a certain month is different from the amount of rainfall in Sylhet. As the objective of the project suggests, we conduct analysis to find patterns and relations between G-1 and G-2 varying G-3.

Instead of using all features/variables from G-2 to directly correlate to G-1, weather components are compared in between to find the best probable feature which can be compared to G-1. Here we've opted to use Rainfall as the primary feature from G-2. Correlation conducted in Orange Data Mining Tool suggests the following characteristics:

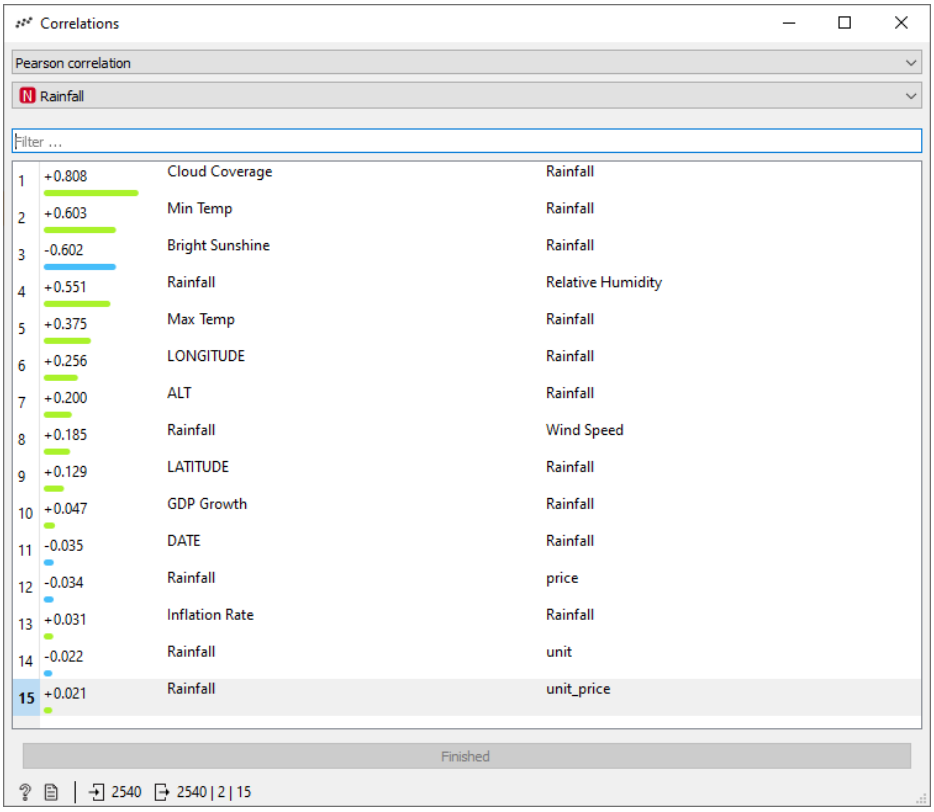

Fig-1: Correlations of weather components to Rainfall

However in some cases in spite of having higher correlations wasn't directly observed, rather indirectly i.e. the relationship between Rainfall and Bright Sunshine (-0.602) wasn't plotted against rather Cloud Coverage vs Bright Sunshine was used (-0.652).

Fig-2: Correlation between Cloud Coverage and Bright Sunshine

For G-1, analysis was conducted using G-3 and other growth rates were compared. Finally G-1 and G-2 were compared in between.

Analysis

Price of goods over the years

At first EDA was conducted on G-1. Prices of goods were plotted against time.

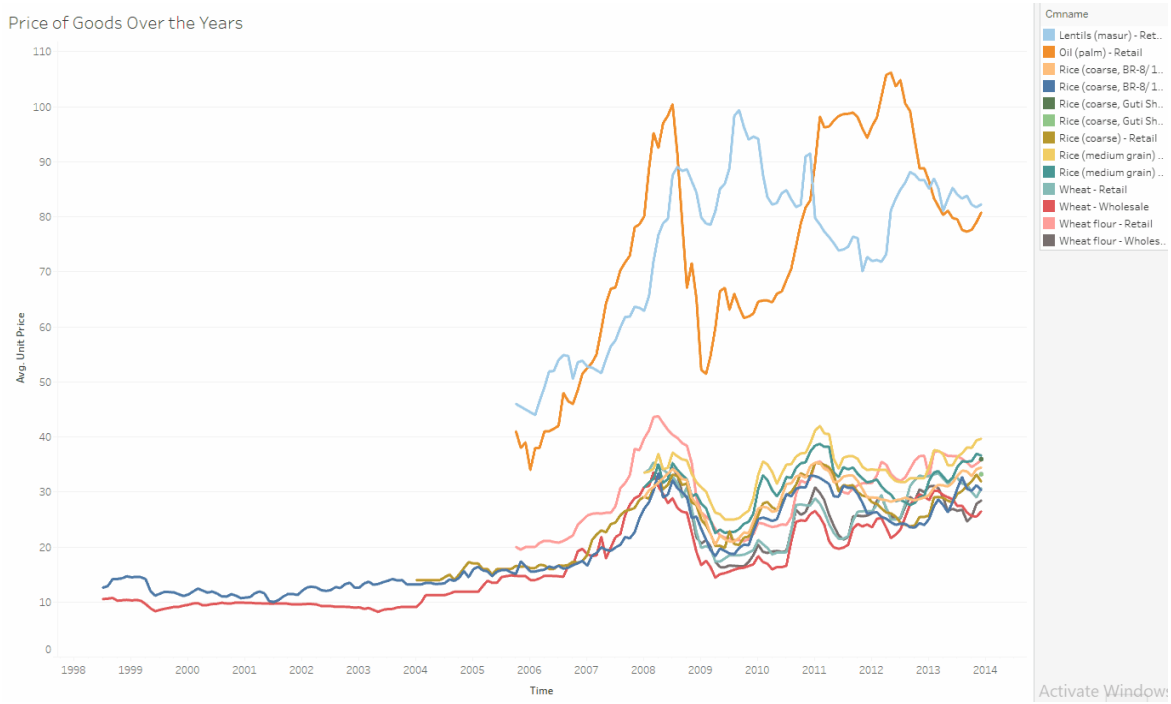

Fig-3: Price of goods over the years (Individual)

From this time series chart of price of goods, few interesting findings have been made.

1. For **Rice (coarse, BR-8/ 11/, Guti Sharna) - Wholesale** and **Wheat - Wholesale,** from 1998-2004(ish) changes in price were more or less in a margin. But from 2004 prices followed an ever increasing trend.

2. There's a huge spike in 2007 and continued till the first and second quarter of 2008 and price decreased in first quarter of 2009.

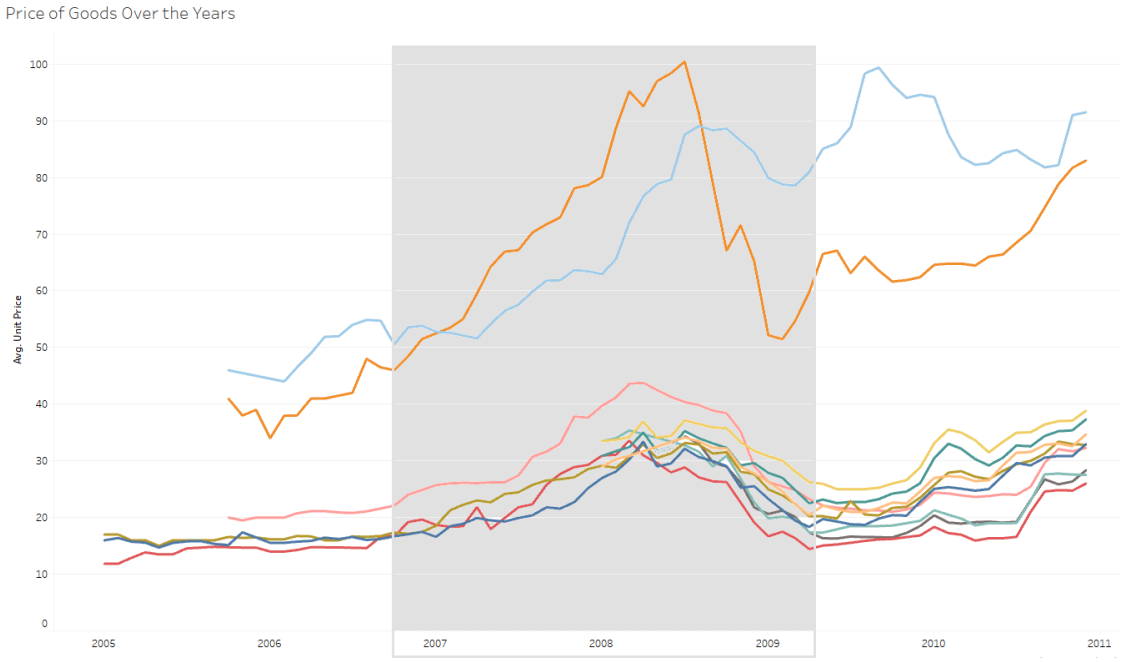

Fig-6: Price hike in 2007-08

This matches the global food price hikes during 2007–08 resulted in a sharp rise in staple food prices in Bangladesh [6].

Also inflation rate Bangladesh was at a peak of 9.107 in 2007 and 8.902 in 2008 [1] which were the all time highest and second highest respectively during 1998-2013. Also GDP growth rate was the highest in 2007 at 7.059% [2]. The high growth rate of GDP contributed to a demand-pull inflation and the price increased [6].

At the political front, Bangladesh was run by an unelected and undemocratic 'civil' caretaker government backed by the military [6]. It is not surprising that the election manifestos of all political parties keep inflation control high on their economic agenda [5].

3. Also 2010-12 world food crisis [4] is evident here. Prices of foods started to increase in late 2009 and 2010 reached new heights in 2011-12. Although not all of the goods' prices increased simultaneously but followed the trend.

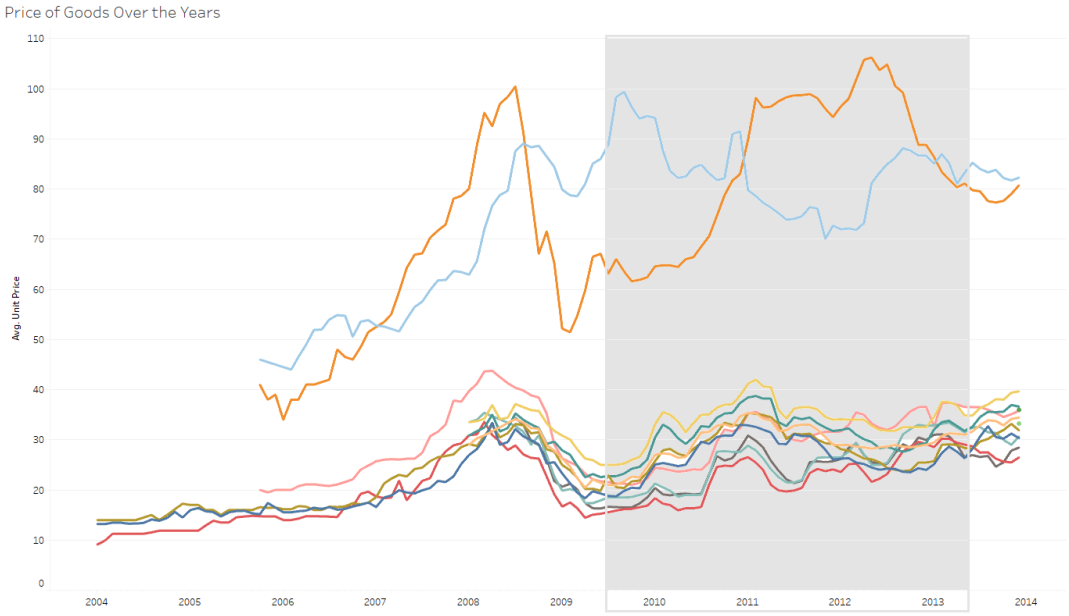

Fig-6: Price hike in 2011-12

4. “Individual Inflation Rate” of goods was plotted against the overall inflation rate on a year by year basis and showed similar trends.

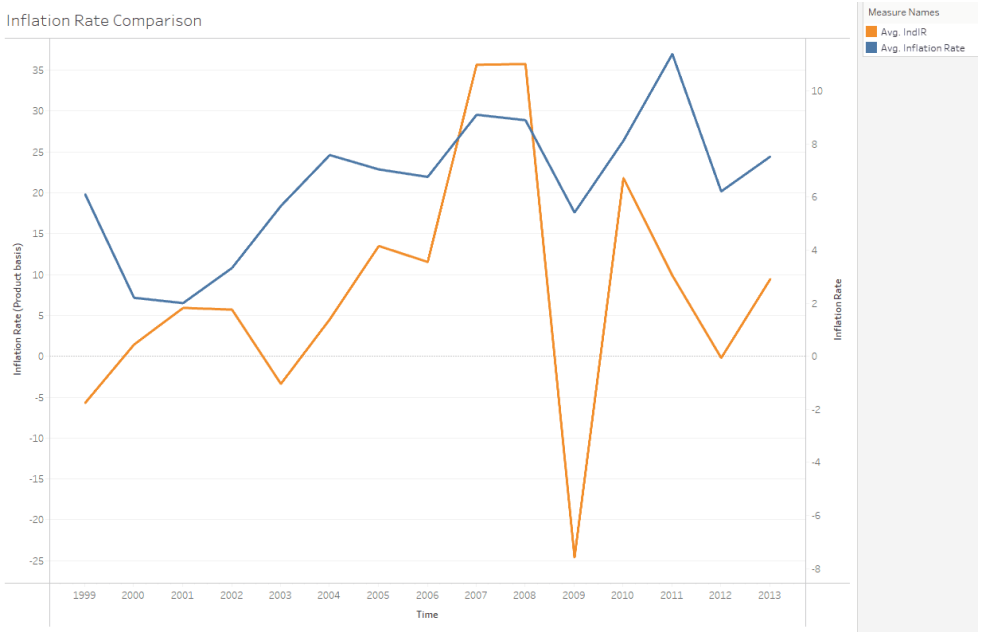

Fig-7: Individual inflation rate and overall inflation rate comparison

Goods Sales Distribution in Regions and Average Prices

Here, we've plotted total sales of individual goods over the years in different regions against their average prices.

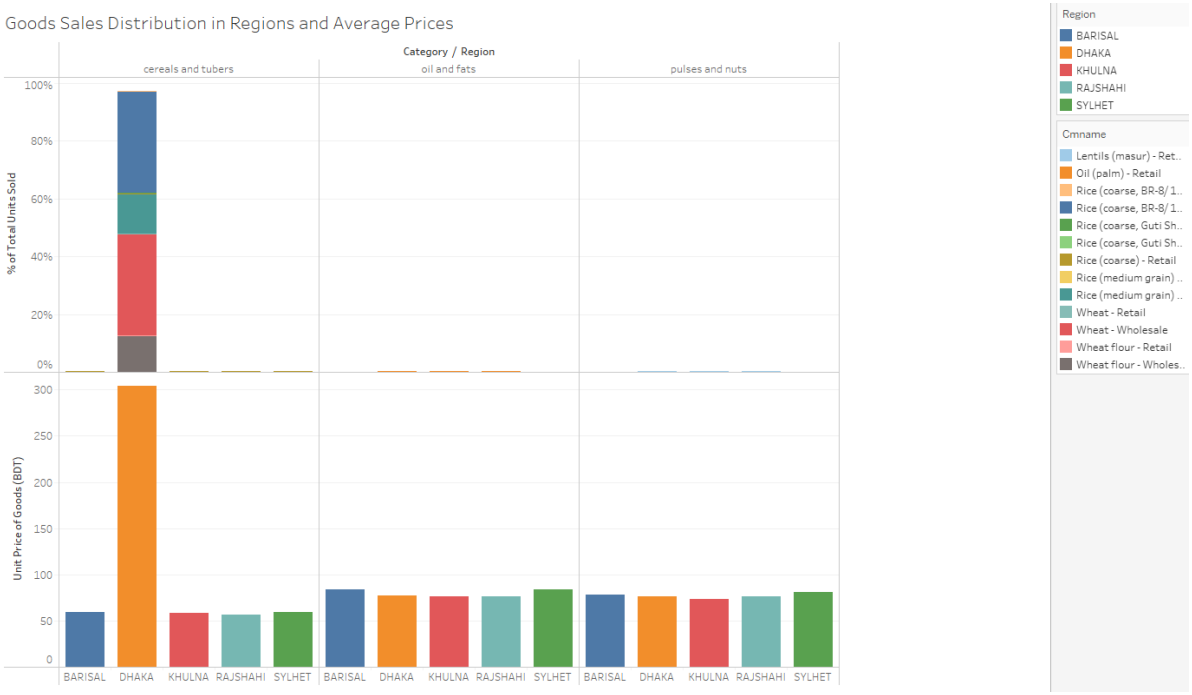

Fig-8: Goods Sales Distribution in Regions and Average Prices (for all goods)

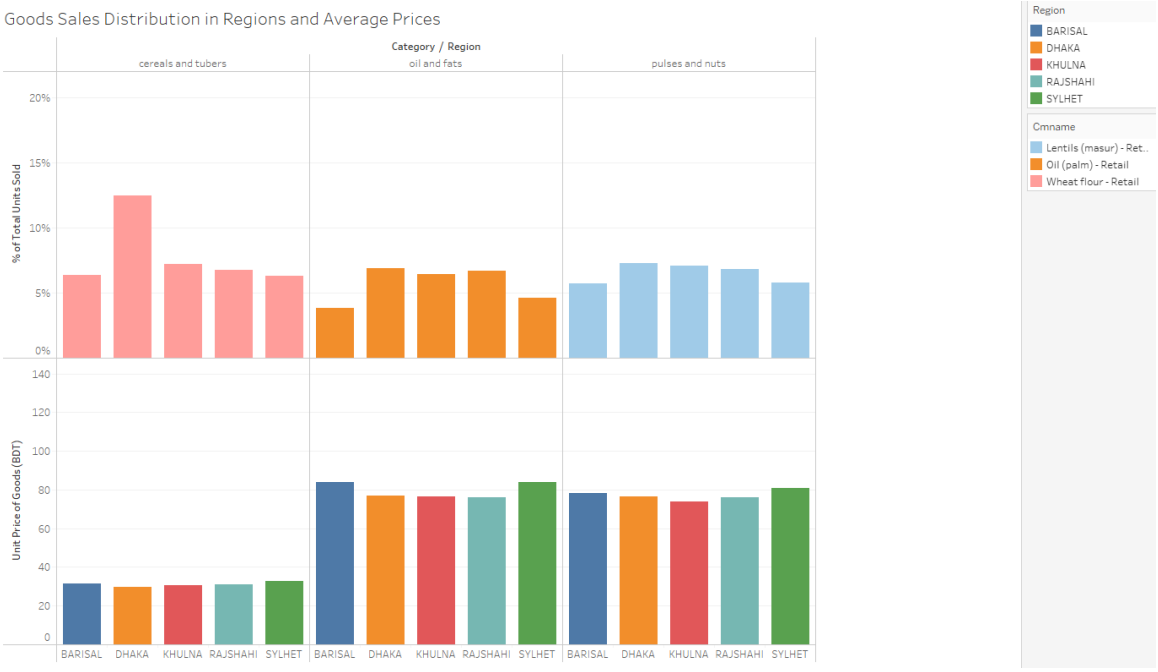

Fig-9: Goods Sales Distribution in Regions and Average Prices (Lentils, Oil-palm, Wheat flour)

Fig-8 represents the total account of the above mentioned plot where Fig-9 shows filtered data. From Fig-8, we can see that Cereals and tubers have the most demand in Dhaka region. Reasons behind this demand could be due to the huge population living in this region, day to day meal and bakery needs create the highest demand. Also for all of the goods, Dhaka has the highest demand which can be seen from the above plot.

Total Goods Sold and Goods Unit Price Comparison

This graph presents the total units sold of goods and their respective unit prices.

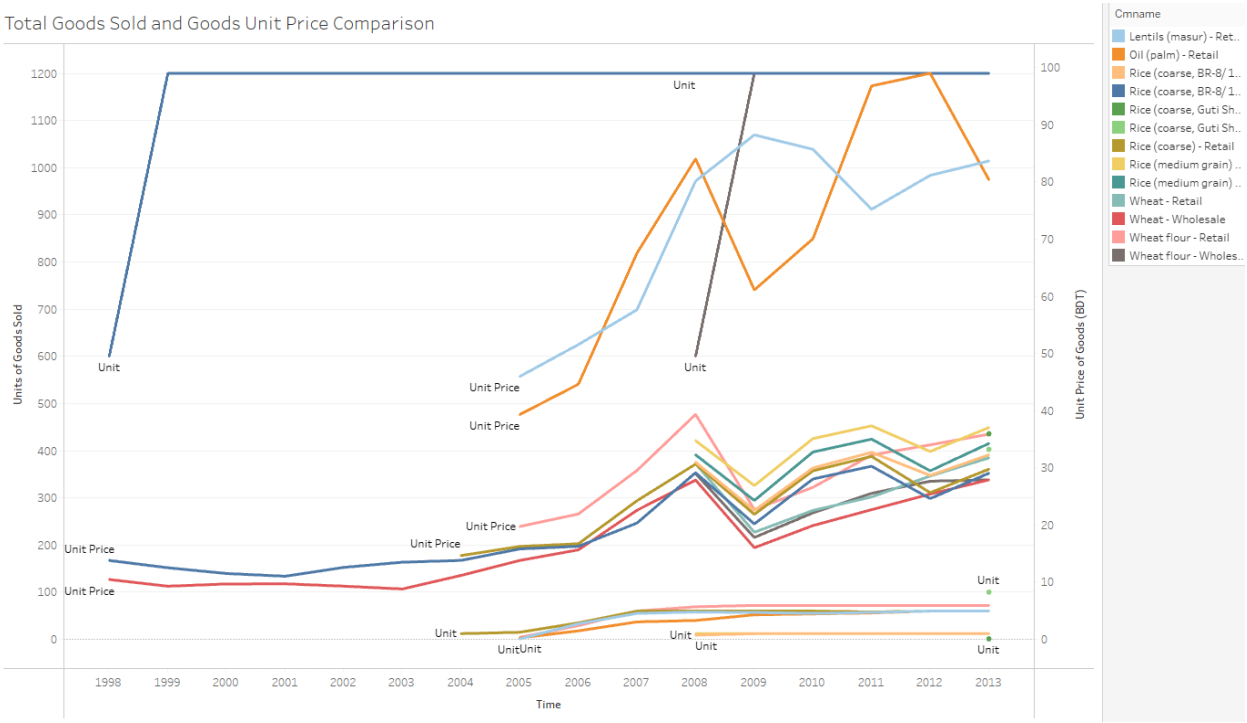

Fig-10: Total Goods Sold and Goods Unit Price

For all of the goods after a certain period, units sold per month and per year become constant. In reality with changes in price sales changes. The price change is evident here but the unit of goods is constant. So this comparison should not be taken into account.

Analysis for G-1 in relation to G-3 concludes here. We'll be conducting analysis between G-2 and G-3 features.

Rainfall and Relative Humidity Comparison

As previously seen Rainfall and Relative humidity showed well enough correlation [Fig-1].

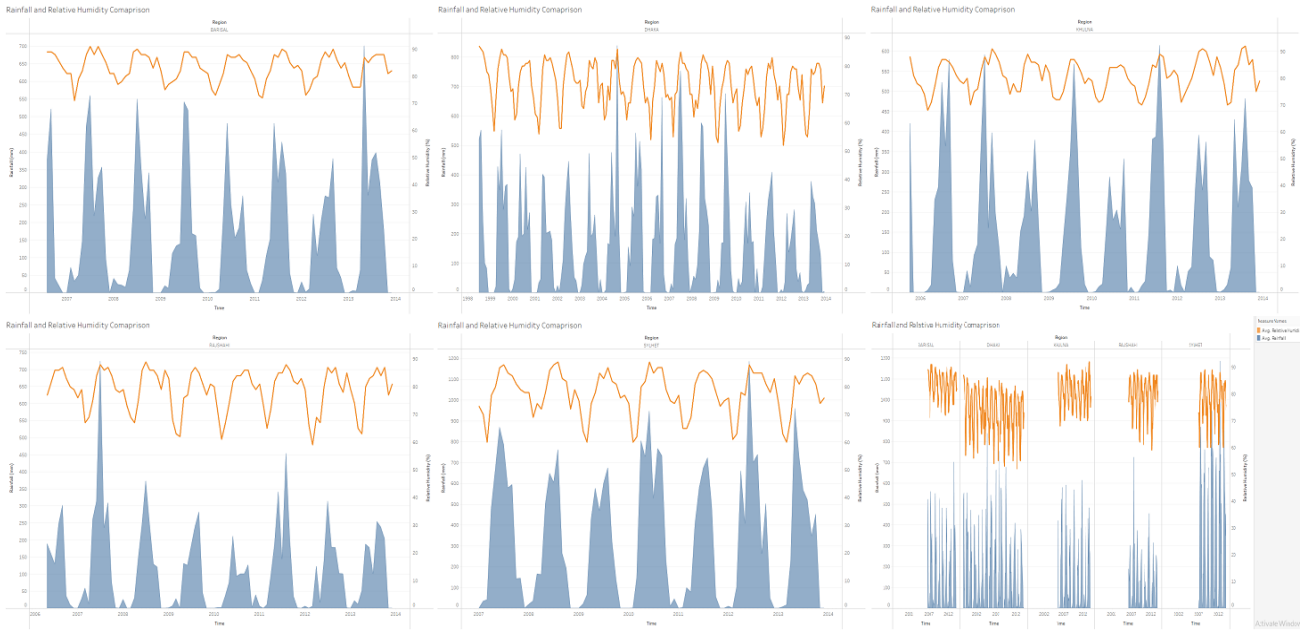

Fig-11: Rainfall and Relative humidity comparison

Fig-11 shows a cyclical pattern for rainfall and relative humidity. Except Dhaka, other 4 (four) regions possess almost the same amount of relative humidity. This could be due to their geological positions. When it comes to rainfall, Sylhet has the most amount out of the 5 (five) regions. This is also due to its geological position. For Khulna and Rajshahi, rainfall isn't consistent on a year by year basis in context to this dataset.

Relative humidity decreases late December to mid February (during winter) and increases during monsoon, peaks in the month of July.

Rainfall and Cloud Coverage Comparison

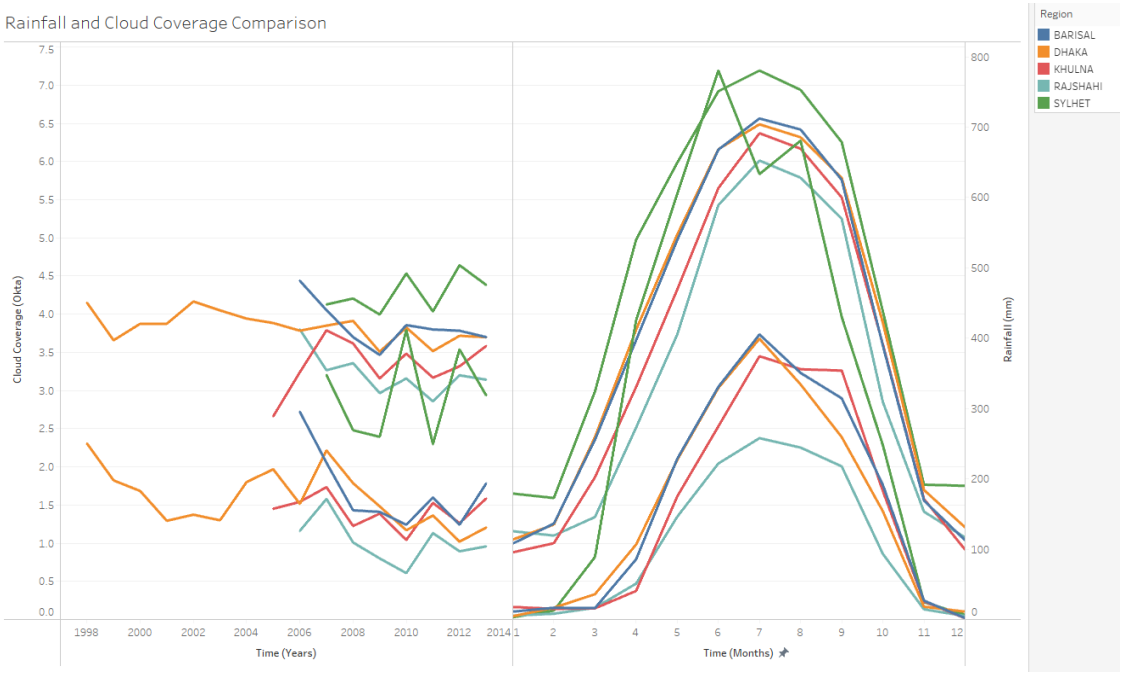

Fig-12: Rainfall and Cloud Coverage Comparison

This graph was plotted to show over the years relation between rainfall and cloud coverage as well on monthly average. It shows a strong correlation between cloud coverage and rainfall as there must be clouds for rainfall and the other way around is not always true. So this data reproves the common understanding of rainfall mechanism.

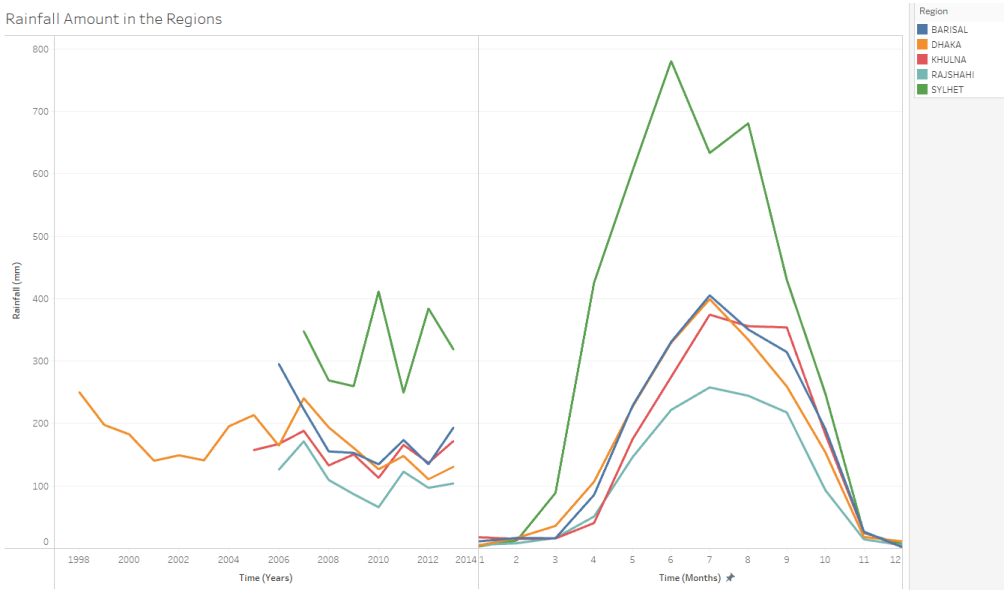

Fig-13: Rainfall amounts in different regions

But it also points out that, unlike the rest of the 4 (four) regions, rainfall patterns on a monthly basis are different in Sylhet [Fig-13]. Where in other regions the peak of average rainfall occurs in July, for Sylhet it occurs in June, decreases in July and increases in August.

Cloud Coverage and Bright Sunshine Comparison

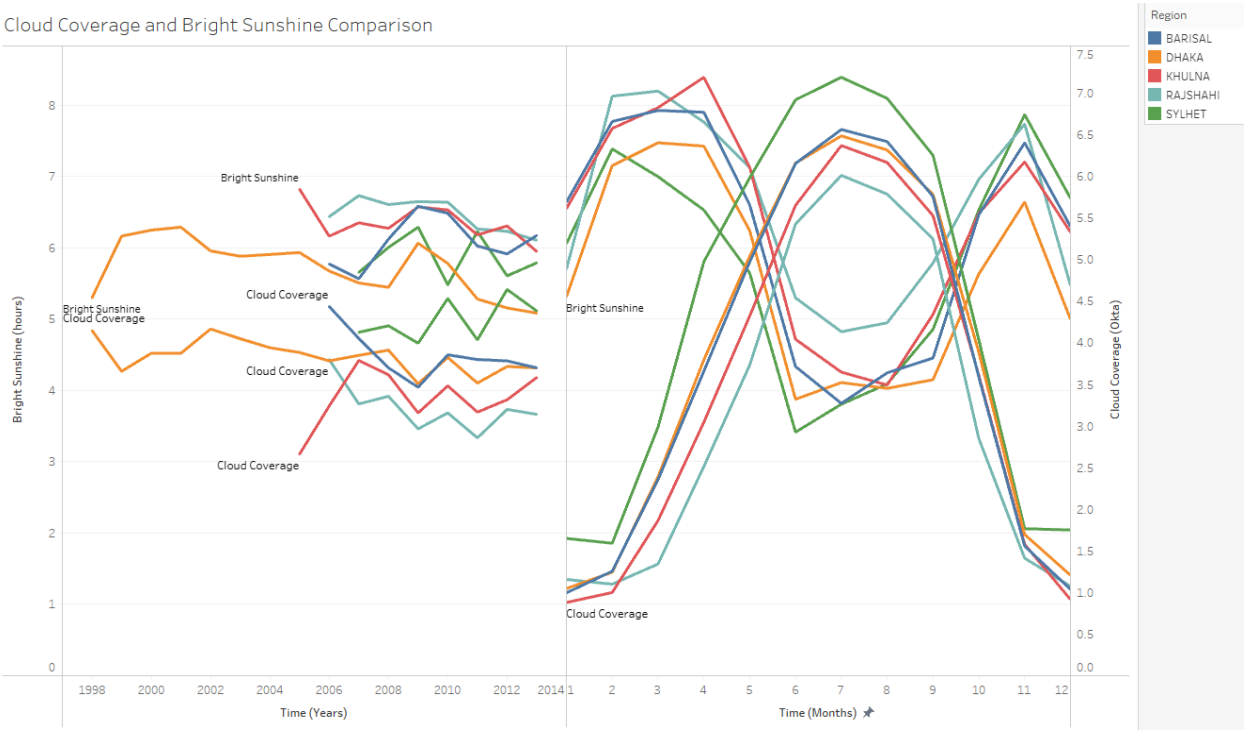

Fig-14: Cloud Coverage and Bright Sunshine Comparison

From this graph we can see that cloud coverage and bright sunshine shows an inversely proportional relation. As more clouds cover the sky less sunshine hence less bright sunshine is possible.

Another interesting finding is for Sylhet, for the month of June the Bright sunshine is at its lowest point, this is also the month when Sylhet gets the most amount of average rainfall during a year [Fig-13].

So, from rainfall to cloud coverage and from cloud coverage to bright sunshine relation we can deduce that rainfall and bright sunshine have an inversely proportional relation which can be seen in nature as well.

Cloud Coverage and Wind Speed Comparison

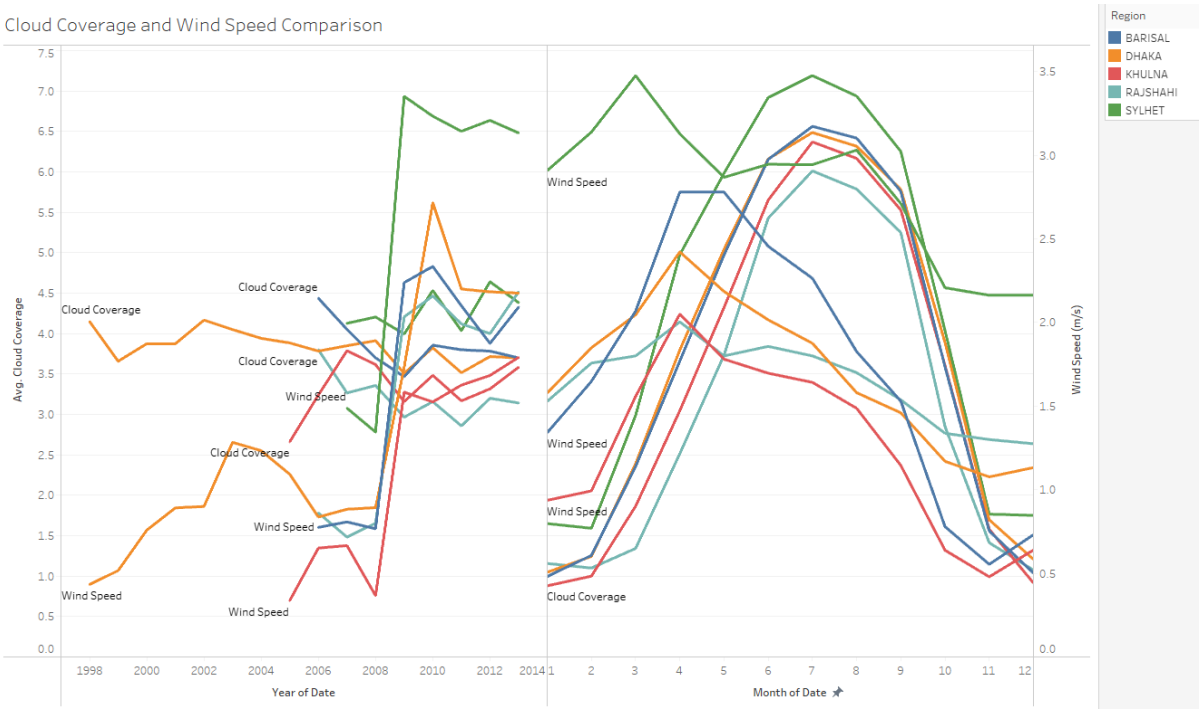

Fig-15: Cloud Coverage and Wind Speed Comparison

As the wind speed increases so does the cloud cover. But a decrease in wind speed doesn't simultaneously decrease cloud cover. Generally wind speed and cloud cover has an inverse relation. More data points and factors should be used to deduce further.

Relative Humidity and Temperature Comparison

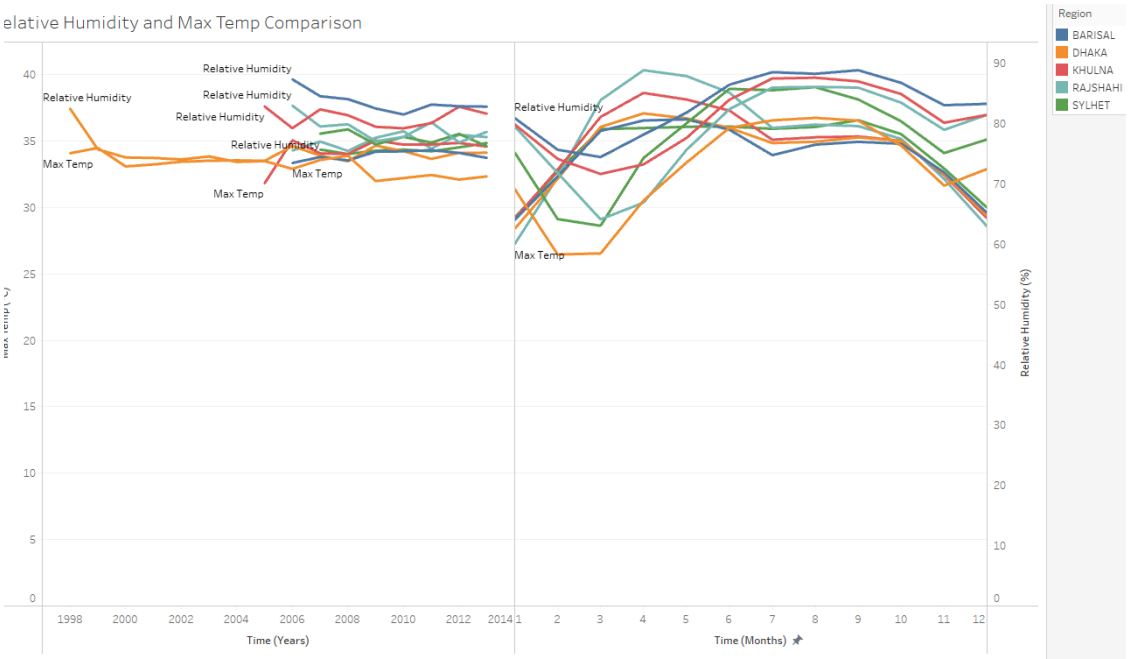

Fig-16: Relative Humidity and Max Temp Comparison

Relative humidity and Max temperature follows a cyclic pattern over the months of years. With increase in max temperature relative humidity decreases because warm air can hold more water vapour so with increasing temperature relative humidity falls if no water vapour is added. Also with decrease in max temperature relative humidity increases as the air gets saturated with previously contained water vapour.

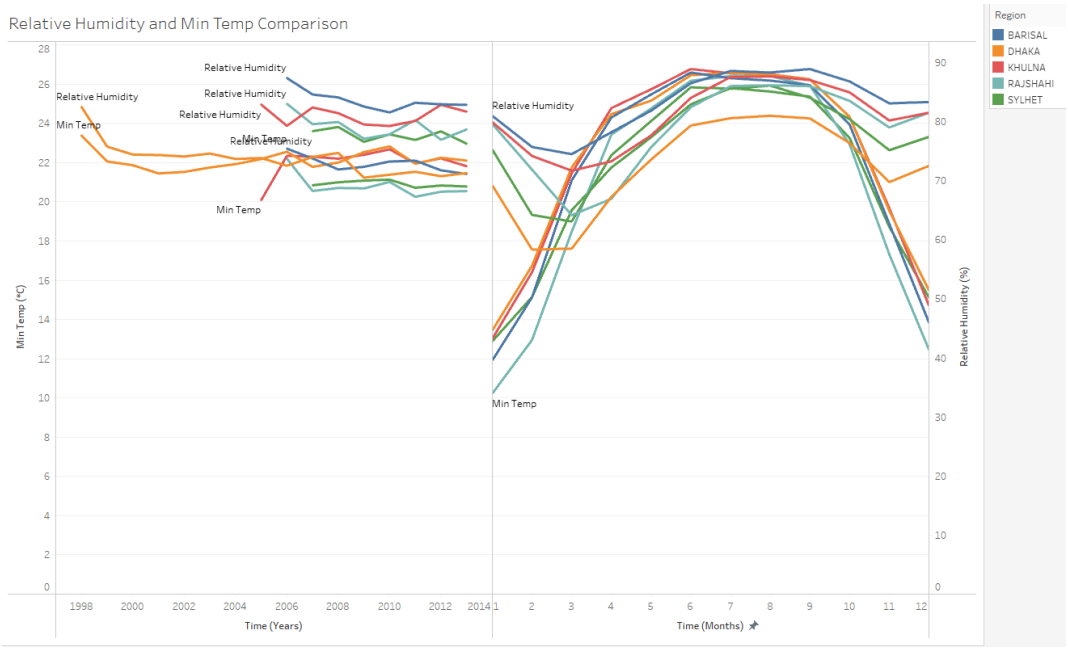

Fig-17: Relative Humidity and Min Temp Comparison

In case of min temperature and relative humidity it follows a cyclical pattern. With increasing min temperature relative humidity decreases then increases again with increasing min temperature and mirrors this pattern for the other half of the year.

Analysis for G-2 in relation to G-3 concludes here. Now we'll compare between rainfall (G-2) and unit price of goods (G-1) to identify patterns if any exists.

Rainfall and Unit Price of Goods Comparison

Fig-18: Rainfall and Unit Price of Goods Comparison

After comparing features to intra and inter groups (G-1& G-3, G-2 & G-3), we are comparing features between G-1 (unit price of goods) and G-2 (rainfall). Overall goods unit price on a region basis doesn't exactly correlate to rainfall. 5 (five) regions of our dataset behave differently to rainfall.

Discussion

As stated earlier, categorizing the dataset in 3 (three) parts gave us a different way to look at the problem. We were able to conduct intra-group to inter-group feature comparison. We've seen price surge in no relation to weather (or to any weather features). There is a factor of import goods, most of these goods are imported in high volume. So local weather changes except natural and transportational calamities won't affect the price directly or that much in general. Also man made constraints like syndicate or maintaining personal gain could affect the price changes.

Further analysis

We used Orange Data mining tool to predict Rainfall and Unit price of goods. Used three machine learning models:

1. Linear Regression
2. Decision Tree
3. Random Forest

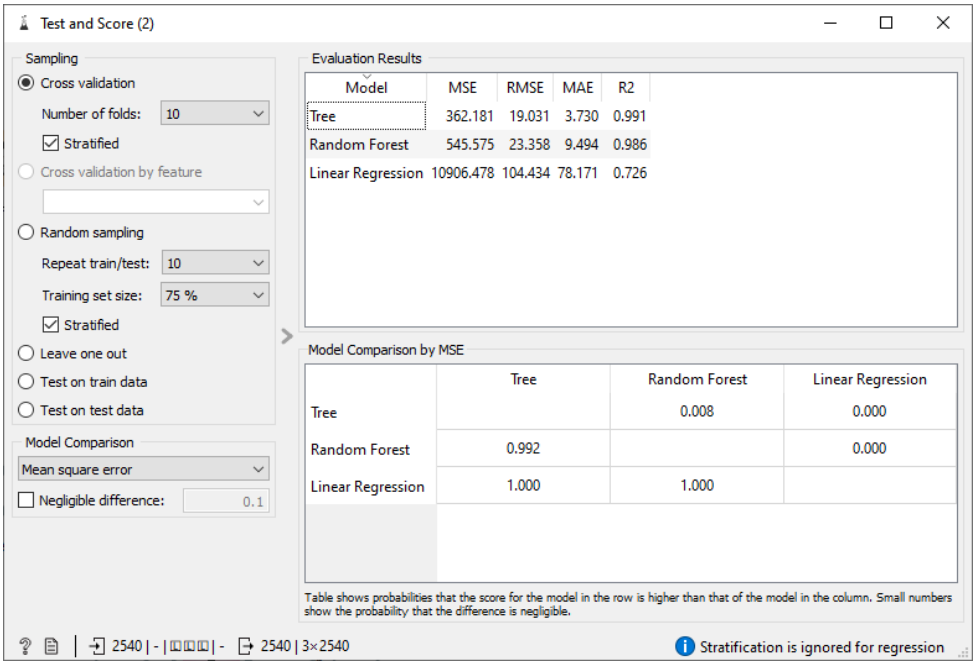

Fig-19: Rainfall prediction using machine learning models

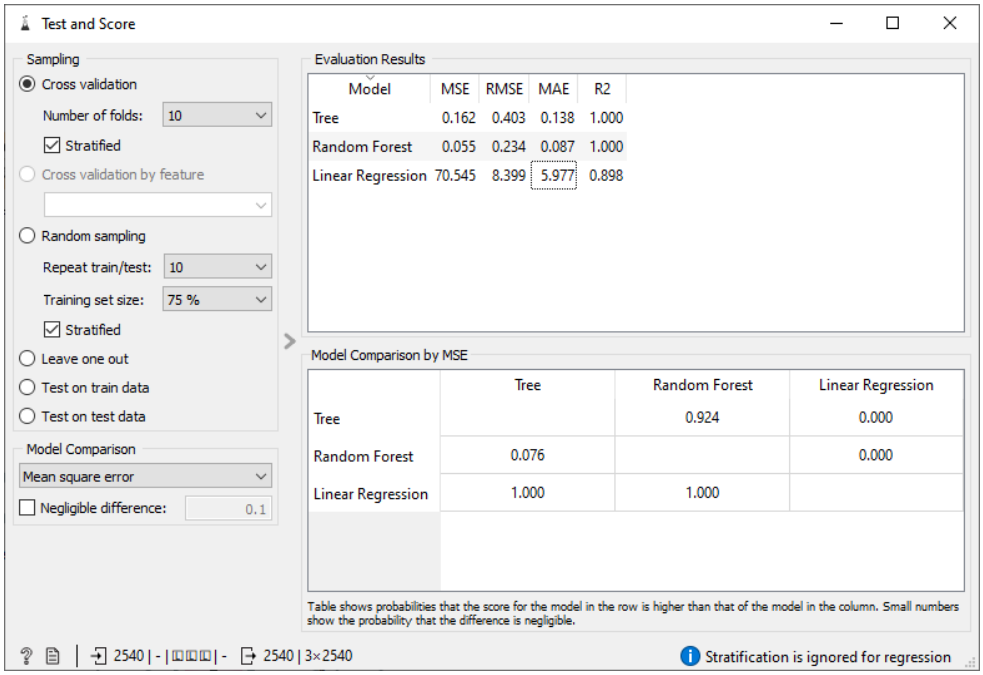

Fig-20: Unit price of goods prediction using machine learning models

Models did very well in predicting Unit price of goods in respect to predicting Rainfall amount. Decision Tree and Random Forest have a R-Squared value of 1 (predicted data accurately).

Conclusion

To summarize the whole process, we categorized the dataset in 3 (three) different categories to conduct different levels of feature analysis. Weather data correlates to rainfall in seemingly manner whereas prices of goods seem not to rely on weather rather on other socio-economic factors both global and local.

References

1. Inflation Data - https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?end=2013&locations=BD&start=1998
2. GDP Grwoth - https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2013&locations=BD&start=1998
3. 2007–2008 world food price crisis - https://en.wikipedia.org/wiki/2007%E2%80%932008_world_food_price_crisis
4. 2010–2012 world food price crisis - https://en.wikipedia.org/wiki/2010%E2%80%932012_world_food_price_crisis
5. Managing rice prices - https://www.thedailystar.net/news-detail-68574
6. The Political Economy of Food Price Policy in Bangladesh - https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198718574.001.0001/acprof-9780198718574-chapter-11