Results and Analysis

Kapil Thakkar and Reshma Kumari

June 10, 2016

Chapter 1

Results and Analysis

We executed our library functions on the onion data. This data consists of Wholesale Price, Retail Price and Arrival since 1st January 2006 to 6th July 2015. In this chapter, we will show results produced by our system and will analyse these results along with each method.

1.1 Results

We have performed 4 types of analysis and result for each of this method is as follows. Note that these are primary results. Data for 2 centers are considered over here, Mumbai and Delhi.

Here are some results related with Mumbai Center. Table 1.1 stats the result of anomalies reported by our system, with details about anomalies reported by each method. So here First 5 columns corresponds to each method. Column 6 is union of results of first 3 methods and column 7 is union of result of method 4 and 5, as described in table. Column 8 is intersection of results of column 6 and column 7, which is final result of our system.

Table 1.2 stats the result of number of articles matched with the dates reported by our system as anomaly, for each method. So here First 5 columns corresponds to each method. Column 6 is union of results of first 3 methods and column 7 is union of result of method 4 and 5, as described in table. Column 8 is intersection of results of column 6 and column 7, which is final result of our system.

Note that total number of articles present for center Mumbai is **143**. Note one thing that articles are present from 2010 onwards. Apart from Graph Based Anomaly method, all methods are producing results from 2006 onwards as input data is from that time.

Now, we present detailed analysis for each of the different type of time-series. Fist type of such analysis is in table 1.3. This table shows distribution of news articles present year-wise for each method when retail price time series is compared with average retail price time series. Fist type of such analysis is in table 1.4. This table shows distribution of news articles present year-wise for each method when retail price time series is compared with arrival data of onion time series. Third type of such analysis is in table 1.5. This table shows distribution of news articles present year-wise for each method when retail price time series is compared with wholesale price time series. Fourth type of such analysis is in table 1.6. This table shows distribution of news articles present year-wise for each method when wholesale price time series is compared with arrival data of onion time series.

Distribution of anomalies present year-wise, for each method is also shown in table. Result is described in different tables for different analysis. Result for various analysis is described in tables 1.7, 1.8, 1.9 and 1.10.

Such results of different cities can also be calculated.

1.2 Analysis of Each Method

In this section, we try to analyse each method, what is limitation of each method and where it is performing good. So, we will describe each method one by one and study them. Note that we have articles from 2010 onwards, so we will be focussing on anomalies reported after 2010 and comparing with them news articles which we have.

Methods	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
Retail Vs Average	742	1245	353	100	177	1871	192	136
Retail Vs Arrival	420	120	353	100	167	810	267	173
Retail Vs Wholesale	658	1230	310	100	167	1819	229	132
Wholesale Vs Arrival	448	525	282	100	186	1165	286	217

Table 1.1: Anomalies Reported

Methods	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
Retail Vs Average	742	1245	353	100	177	1871	192	136
Retail Vs Arrival	420	120	353	100	167	810	267	173
Retail Vs Wholesale	658	1230	310	100	167	1819	229	132
Wholesale Vs Arrival	448	525	282	100	186	1165	286	217

Table 1.2: Number of news articles matched with system

Distribution of All Articles	Articles Present	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2010	7	0	5	6	0	0	6	0	0
2011	7	0	3	0	2	0	2	2	2
2012	2	1	1	0	0	0	1	0	0
2013	77	46	31	17	30	52	50	52	50
2014	37	6	18	14	0	1	20	1	1
2015	13	0	12	0	0	0	12	0	0
Total	143	53	70	37	32	53	91	55	53

Table 1.3: Retail Price VS Average Retail Price

Distribution of All Articles	Articles Present	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2010	7	0	5	6	0	0	6	0	0
2011	7	0	3	0	2	0	2	2	2
2012	2	1	1	0	0	0	1	0	0
2013	77	46	31	17	30	52	50	52	50
2014	37	6	18	14	0	1	20	1	1
2015	13	0	12	0	0	0	12	0	0
Total	143	53	70	37	32	53	91	55	53

Table 1.4: Retail Price VS Arrival data of onion

Distribution of All Articles	Articles Present	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2010	7	0	5	6	0	0	6	0	0
2011	7	0	3	0	2	0	2	2	2
2012	2	1	1	0	0	0	1	0	0
2013	77	46	31	17	30	52	50	52	50
2014	37	6	18	14	0	1	20	1	1
2015	13	0	12	0	0	0	12	0	0
Total	143	53	70	37	32	53	91	55	53

Table 1.5: Retail Price VS Wholesale Price

Distribution of All Articles	Articles Present	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2010	7	0	5	6	0	0	6	0	0
2011	7	0	3	0	2	0	2	2	2
2012	2	1	1	0	0	0	1	0	0
2013	77	46	31	17	30	52	50	52	50
2014	37	6	18	14	0	1	20	1	1
2015	13	0	12	0	0	0	12	0	0
Total	143	53	70	37	32	53	91	55	53

Table 1.6: Wholesale Price VS Arrival data of onion

Distribution of All Articles	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2006	35	30	0	78	0	58	78	7
2007	63	0	0	22	0	63	22	0
2008	28	0	0	0	0	28	0	0
2009	28	15	0	0	0	43	0	0
2010	48	15	46	0	0	82	0	0
2011	36	30	40	0	0	105	0	0
2012	77	0	0	0	0	77	0	0
2013	59	0	168	0	161	203	161	161
2014	46	30	99	0	6	151	6	5
2015	0	0	0	0	0	0	0	0
Total	420	120	353	100	167	810	267	173

Table 1.7: Distribution of Anomalies reported by system for Retail Price VS Average Retail Price

Distribution of All Articles	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2006	35	30	0	78	0	58	78	7
2007	63	0	0	22	0	63	22	0
2008	28	0	0	0	0	28	0	0
2009	28	15	0	0	0	43	0	0
2010	48	15	46	0	0	82	0	0
2011	36	30	40	0	0	105	0	0
2012	77	0	0	0	0	77	0	0
2013	59	0	168	0	161	203	161	161
2014	46	30	99	0	6	151	6	5
2015	0	0	0	0	0	0	0	0
Total	420	120	353	100	167	810	267	173

Table 1.8: Distribution of Anomalies reported by system for Retail Price VS Arrival data of onion

Distribution of All Articles	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2006	35	30	0	78	0	58	78	7
2007	63	0	0	22	0	63	22	0
2008	28	0	0	0	0	28	0	0
2009	28	15	0	0	0	43	0	0
2010	48	15	46	0	0	82	0	0
2011	36	30	40	0	0	105	0	0
2012	77	0	0	0	0	77	0	0
2013	59	0	168	0	161	203	161	161
2014	46	30	99	0	6	151	6	5
2015	0	0	0	0	0	0	0	0
Total	420	120	353	100	167	810	267	173

Table 1.9: Distribution of Anomalies reported by system for Retail Price VS Wholesale Price

Distribution of All Articles	Slope Based (1)	Window Correlation (2)	Linear Regression (3)	Graph Based (4)	Multivariate (5)	1 U 2 U 3 (6)	4 U 5 (7)	6 ∩ 7
2006	35	30	0	78	0	58	78	7
2007	63	0	0	22	0	63	22	0
2008	28	0	0	0	0	28	0	0
2009	28	15	0	0	0	43	0	0
2010	48	15	46	0	0	82	0	0
2011	36	30	40	0	0	105	0	0
2012	77	0	0	0	0	77	0	0
2013	59	0	168	0	161	203	161	161
2014	46	30	99	0	6	151	6	5
2015	0	0	0	0	0	0	0	0
Total	420	120	353	100	167	810	267	173

Table 1.10: Distribution of Anomalies reported by system for Wholesale Price VS Arrival data of onion

1.2.1 Slope Based Anomaly Detection

The main functionality of this method is to find change in one variable with respect to other. Given two time-series, here we try to find, between two points in time series, how much dependent variable changed corresponding to independent variable. If this change is huge, than it is reported as anomaly.

We have four type of analysis which are as follows:

1. **Retail Price vs Average of Retail Price**: Here, we first take average of retail price at all centres and than compare change in retail price with change in average of retail price for different time window.

2. **Retail Price vs Arrival of Onion**: Here, we try to find change in retail price with respect to change in arrival of onion for different windows.

3. **Retail Price vs Wholesale Price**: Here, retail price is dependent on wholesale price and here we try to find change in retail price with respect to change wholesale price for different windows in this method.

4. **Wholesale Price vs Arrival of Onion**: Here, we try to find change in wholesale price with respect to change in arrival of onion for different window size.

So, in each of the case, we try to find change with respect to another, and if this change is huge, crossing threshold than it is reported as anomaly. Now, not that in analysis 1 and 3 stated above, both the time series are directly proportional to each other and in the analysis 2 and 4 both the time series are inversely proportional to each other. So, limitations faced by this method for analysis 1 and 3 will be similar and for analysis 2 and 4 will be similar. While describing this method, each analysis will be referenced by its corresponding number.

First we will start with analysis 1 and 3. Here, we have few observations as follows:

- Dates are reported as anomalies, if retail price at centre is increasing more as compared to average retail price for analysis 1 or if retail price at centre is increasing more as compared to wholesale price for analysis 3. Such cases are reported for the following tenure by this method:

  - *Analysis 1*: June 2010, August 2010, May 2011, June 2011, Nov 2012, Apr 2013, May 2013 (Prices went too high as compared to average)

  - *Analysis 3*:

  So, even though we do not have articles for these anomalies, but method is behaving as it should be.

- This observation is limitation of this method. There exist few cases where, drop in retail price for one center is quite huge as compared to drop in price for average retail price (in case of analysis 1) or wholesale price (in case of analysis 3). This is good thing for centres, and should bot be anomaly. But in this case, slope value goes high and that's why our method reports that tenure as anomaly as well. Such cases are reported for the following tenure by this method:

  - *Analysis 1*: January 2010

– *Analysis 3*:

Also, note one thing that, this method reports anomaly as whole window. So, because of that too, we have more anomaly dates.