Data Warehousing and Business Intelligence Project

on

Smart Agriculture and Management of Resources

Kaushik Rajan
17165849

MSc Data Analytics – 2018/9

Submitted to: Dr Horacio Gonzlez-Velez

National College of Ireland
Project Submission Sheet – 2017/2018
School of Computing

Student Name:	Kaushik Rajan
Student ID:	17165849
Programme:	MSc Data Analytics
Year:	2018/9
Module:	Data Warehousing and Business Intelligence
Lecturer:	Dr Horacio Gonzlez-Velez
Submission Due Date:	26/11/2018
Project Title:	Agriculture Crop Production

I hereby certify that the information contained in this (my submission) is information pertaining to my own individual work that I conducted for this project. All information other than my own contribution is fully and appropriately referenced and listed in the relevant bibliography section. I assert that I have not referred to any work(s) other than those listed. I also include my TurnItIn report with this submission.

ALL materials used must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is an act of plagiarism and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	November 26, 2018

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table 1: Mark sheet – do not edit

Criteria	Mark Awarded	Comment(s)
Objectives	of 5	
Related Work	of 10	
Data	of 25	
ETL	of 20	
Application	of 30	
Video	of 10	
Presentation	of 10	
Total	of 100	

Project Check List

This section capture the core requirements that the project entails represented as a check list for convenience.

- ☒ Used LaTeX template
- ☒ Three Business Requirements listed in introduction
- ☒ At least one structured data source
- ☒ At least one unstructured data source
- ☒ At least three sources of data
- ☒ Described all sources of data
- ☒ All sources of data are less than one year old, i.e. released after 17/09/2017
- ☒ Inserted and discussed star schema
- ☒ Completed logical data map
- ☒ Discussed the high level ETL strategy
- ☒ Provided 3 BI queries
- ☒ Detailed the sources of data used in each query
- ☒ Discussed the implications of results in each query
- ☒ Reviewed at least 5-10 appropriate papers on topic of your DWBI project

Data Warehousing and Business Intelligence Project on Smart Agriculture and Management of Resources

Kaushik Rajan
17165849

November 26, 2018

Abstract

With advancement in technology, the agriculture sector has also started to use data analytics, automation and machine learning. Also, Vast data are generated in the agriculture field which can be used for crop management, resource management, predicting the weather which in turn help in crop cultivation. The project focuses on building a data warehouse model with an intention to improve the agriculture production by providing better resources such as agricultural land and better water supply to improve the crop production. The core idea of this project is to show the historical agricultural data from different sources which are in different format in a simplified way so that it can be useful in visualizing multiple Business Intelligence queries such as comparing the total agriculture land (Source 1) with the crop produced per year (Source 2), to check if there is increase of pesticide use (Source 3) when there is increase in crop production (Source 2) and to check if provided water resource (Source 4) were sufficient enough for crop production (Source 2). To help framing the above-mentioned BI queries, datasets of Crops produced per year, Total Agricultural land, Pesticide use and water resources were used. Tools used: RStudio for cleaning the datasets extracted from different sources, SQL Server Management Studio for creating and maintaining the Database, SQL Server Integration Services for ETL process and finally Tableau for visualizing the Business Intelligence Queries.

1 Introduction

Agriculture has been on decline in the UK due to multiple reasons, such as land encroachment, growing population, growth of MNCs, exploitation of resources, extensive use of pesticides and less availability of natural resources such as water and fertile soil. The agriculture sector is the main user of resources and has a complicated relationship with the environment (OECD, 2017). Two of the challenges that is confronting agriculture in the Europe are climate change (EEA, 2017c) and land take, i.e. the conversion of land to settlements and infrastructure (EEA, 2017a). With the advancement in IT and communication service, historical data can be collected and stored, using which, agriculture can be improved. Data warehouse has been an integral part of Information technology (IT) strategy for huge multi-national companies of different sectors. As per author Inmon (2005), a data warehouse is a repository of data taken from operational systems, aggregated and summarized to provide decision support. Data warehouses are subject-oriented, integrated, time-variant, and non-volatile. The main motive of this project is to combine the ever-growing technology with agriculture for better decision making. Data warehouse model for agriculture is built by collecting historical data from multiple sources using which multiple Business queries can be visualized. The Business queries that can be visualized using this Data Warehouse model are given below.

1) Total Agricultural land vs Crops produced.

2) Crops produced vs Pesticide use.

3) Total Agricultural land vs Available water resource.

Building such a data warehouse model which is suitable for analytics is the main step which can be used in decision making. The benefits here include improvement of crop production, lesser usage of chemicals and better land management to save for the worlds future needs.

2 Data Sources

Source	Type	Brief Summary
Statista	Structured	Extracted to get the information about the total agricultural land
Eurostat	Structured	Extracted the production information of multiple crops
FAO	Structured	Extracted the information about the amount of pesticide use.
FAO	Unstructured	Scraped data from image which contains the information about the water percentage region wise.

Table 2: Summary of sources of data used in the project

2.1 Source 1: Statista

The dataset containing the information on the total agricultural land in UK was downloaded from the link : `https://www.statista.com/statistics/315937/total-agricultural-land-area-in-the-united-kingdom-uk/` which contains the information for the years 2003 to 2017. The dataset was published online in May 2018 [Figure 1] and this data source is used to address the following business query mentioned in Section 1

- Total agricultural land vs crop production of different countries

Figure 1: Source 1

2.2 Source 2: Eurostat

The dataset contains information on the crop production for different countries in Europe which were downloaded from the link: `https://ec.europa.eu/eurostat/data/database`. The website contains data of different crops such as Cereals, Wheat, Rye, Barley, Oats, Grain maize, Dry pulses, Rape, Green Maize, Root Crops, Fresh vegetables, Permanent crops, Grapes and olives, which were cleaned and combined using RStudio to form a dataset and the R code that was used has been given below in the appendix section. The datasets were published on 17-08-2018[Figure 3] and are used in the following business queries.

- Total Agricultural land vs Crops produced per year country wise.
- Crops produced vs the amount of pesticides used

The datasets which were downloaded initially contained 24 columns which are the years - 2007 to 2018, and 55 rows, which are the different countries. (A Screenshot of the sample dataset has been provided)

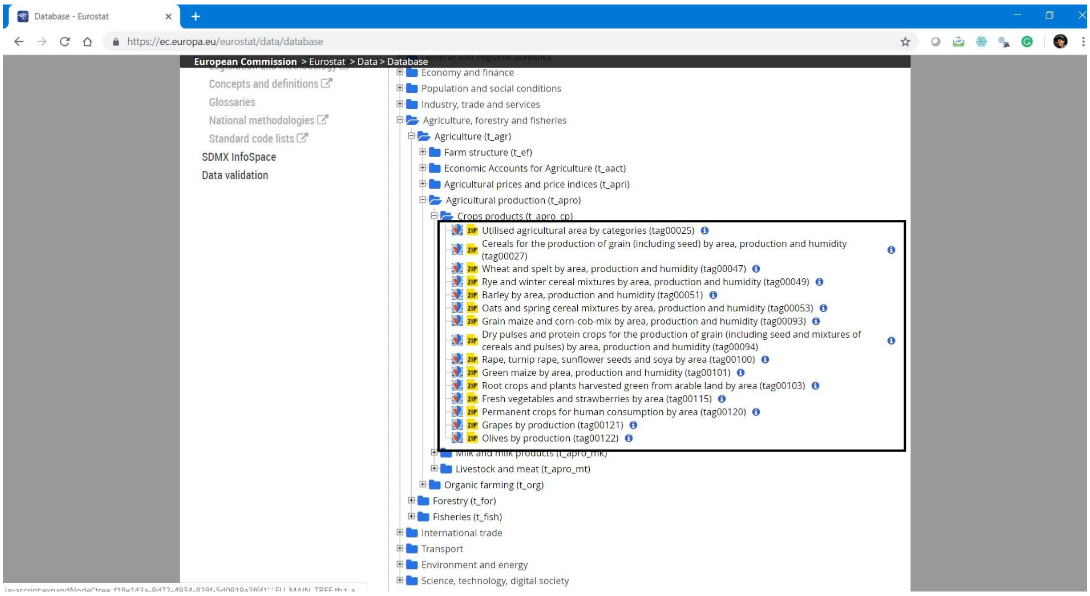
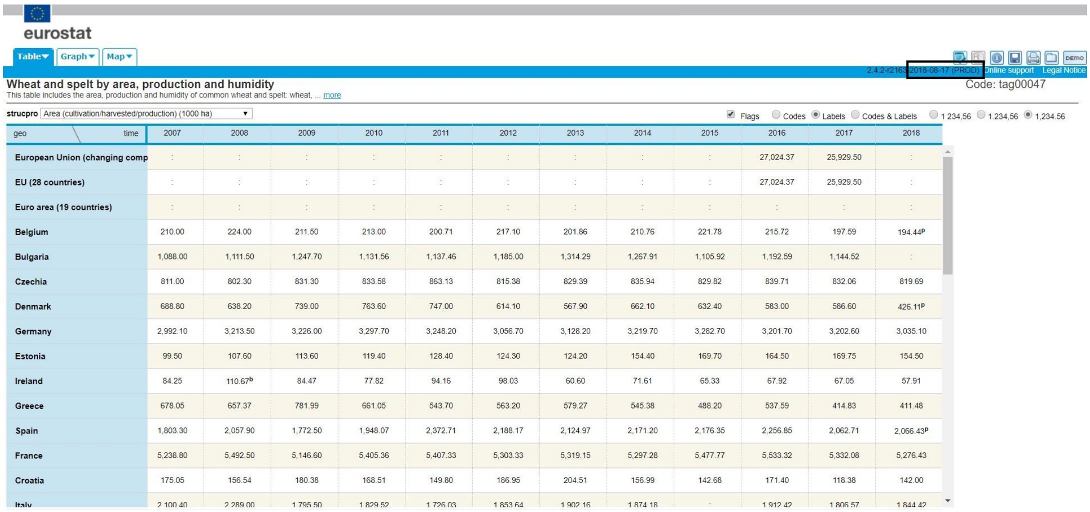


Figure 2: Source 2



geo \ time	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
European Union (changing comp	:	:	:	:	:	:	:	:	:	27,024.37	25,929.50	:
EU (28 countries)	:	:	:	:	:	:	:	:	:	27,024.37	25,929.50	:
Euro area (19 countries)	:	:	:	:	:	:	:	:	:	:	:	:
Belgium	210.00	224.00	211.50	213.00	200.71	217.10	201.86	210.76	221.78	215.72	197.59	194.44 p
Bulgaria	1,088.00	1,111.50	1,247.70	1,131.56	1,137.46	1,185.00	1,314.29	1,267.91	1,105.92	1,192.59	1,144.52	:
Czechia	811.00	802.30	831.30	833.58	863.13	815.38	829.39	835.94	829.82	839.71	832.06	819.69
Denmark	688.80	638.20	739.00	763.60	747.00	614.10	567.90	662.10	632.40	583.00	586.60	426.11 p
Germany	2,992.10	3,213.50	3,226.00	3,297.70	3,248.20	3,056.70	3,128.20	3,219.70	3,282.70	3,201.70	3,202.60	3,035.10
Estonia	99.50	107.60	113.60	119.40	128.40	124.30	124.20	154.40	169.70	164.50	169.75	154.50
Ireland	84.25	110.67 b	84.47	77.82	94.16	98.03	60.60	71.61	65.33	67.92	67.05	57.91
Greece	678.05	657.37	781.99	661.05	543.70	563.20	579.27	545.38	488.20	537.59	414.83	411.48
Spain	1,803.30	2,057.90	1,772.50	1,948.07	2,372.71	2,188.17	2,124.97	2,171.20	2,176.35	2,256.85	2,062.71	2,066.43 p
France	5,238.80	5,492.50	5,146.60	5,405.36	5,407.33	5,303.33	5,319.15	5,297.28	5,477.77	5,533.32	5,332.08	5,276.43
Croatia	175.05	156.54	180.38	168.51	149.80	186.95	204.51	156.99	142.68	171.40	118.38	142.00
Italy	2,100.40	2,289.00	1,795.50	1,829.52	1,726.03	1,853.64	1,902.16	1,874.18	:	1,912.42	1,806.57	1,844.42

Figure 3: Source 2

2.3 Source 3: FAO

The dataset contains information on the pesticide use for different countries in Europe were downloaded from the link: `http://www.fao.org/faostat/en/#data/RP`. The Datasets contains information for the years 2000 to 2016 and was published on 10-08-2018. Initially, the datasets taken from the website were region wise, which was then joined and created into a single datasets having the attributes Area, Item, Year, Value_Pestisides, region and country code. This dataset is used address the following business query mentioned in Section 1

- Crops produced per year vs Pesticide use

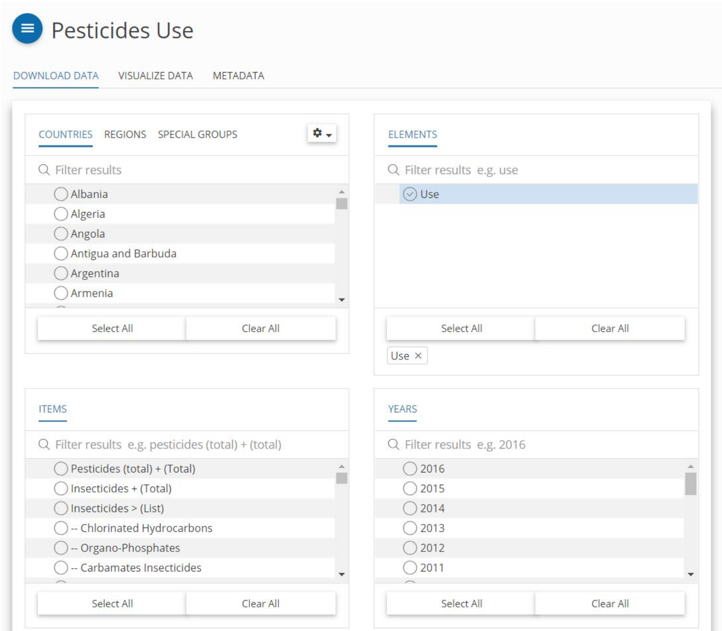


Figure 4: Source 3

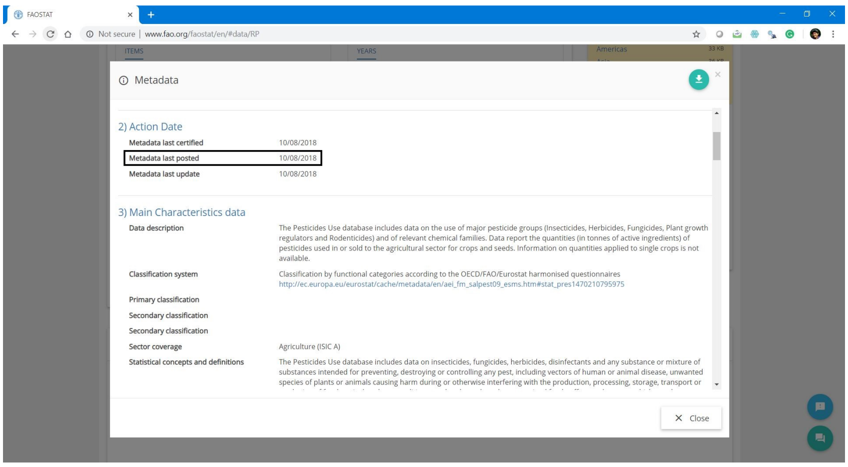


Figure 5: Source 3

2.4 Source 4: FAO

Image from the link: `http://www.fao.org/docrep/005/Y4473E/y4473e0r.gif` has been scraped to get the data of water resource available in each region. The scraping was done using RStudio and the code used for scraping has been provided below in the appendix part. The dataset after cleaning contains the attributes Region Water_Resource_in_Percentage. This dataset is used address the following business query mentioned in Section 1

- Crops produced vs Available water resource.

Water resources regions	Water resources subregions	Total area (km^2) (FAOSTAT, 1999)	Total population (inh.) (FAOSTAT, 2000)	Internal resources: total (km^3/year)	External resources: actual (km^3/year)		Total resources: actual (km^3/year)	% of internal water resources of the region	IRWR/inhab. (m^3/year)	TRWR (actual)/inhab. (m^3/year)
4 Western and Central Europe	Central Europe	1 123 550	115 802 000	284.5	87.9	(1)	372.4	13.11	2 457.0	3 216.0
	Mediterranean Europe	1 095 300	124 408 000	422.8	30.0	(2)	452.8	19.48	3 398.7	3 639.9
	Northern Europe	1 258 080	24 082 000	864.1	0.0	(3)	864.1	39.81	35 881.6	35 881.6
	Weslern Europe	1 421 486	246 492 000	598.9	14.7	(4)	613.6	27.59	2 429.7	2 489.3
	Western and Central Europe	**4 898 416**	**510 784 000**	**2 170.4**	**10.9**	**(5)**	**2 181.3**	**100.00**	**4 249.1**	**4 270.4**
	World	**133 845 436**	**6 052 577 900**	**43 764.3**	**0.0**		**43 764.3**		**7 230.7**	**7 230.7**
	Western and Central Europe as % of world	3.7	8.4	5.0			5.0			

Notes:

(1) 77 km^3/year from another European subregion and 10.9 km^3/year from Eastern Europe.
(2) 29.95 km^3/year from another European subregion.
(3) No exchanges.
(4) 14.7 km^3/year from another European subregion.
(5) From Eastern Europe.

Figure 6: Source 4

3 Related Work

In recent times, many papers have been published, that emphasis on usage of analytics and big data, in the agriculture sector and its importance with regards to crop production.

In Ngo et al. (2018), the author has used different Big data storage systems in order to capacitate the amount of data generated in the agriculture field.

The ideology of the author is to create a precision agriculture to create agriculture Intelligence which will help in decision making in usage of resources. The author has used continent level dataset for this project. The author answers for the business queries such as the suitable crops for a particular field, how to improve crop yields, and where can the crops be sold for the highest price. The author here has suggested to use IOT for collection of data and to use databases such as Hadoop, MongoDB.

In Liakos et al. (2018), the author, upon researching, has found out that there are more than 5 machine learning algorithms which already exist and which are used for crop management. As per the author, the usage of machine learning algorithms in the field of agriculture have been on rise. Machine learning algorithms can be used to mine data, which can be automated. In this paper, the author has taken a survey to find which algorithm has been the preferred one. Machine learning and artificial intelligence have been used to predict the crop yield, crop quality, water management, soil management. Such predictions are used for decision making.

In Cojocariu & Stanciu (2009), the author, has presented a paper on agricultural intelligence where the author emphasis on using artificial intelligence to predict the future outcomes such the crop production and required resources.

In Janssen et al. (2017), the author has presented paper on how information technology can help the growth of agriculture. The paper concentrates on the path to gather information for decision making. The flow chat starts with collecting information from various sources such as social media, satellites, experimental data, sensors, citizen observations, Statistics, the data collected is then converted into information using models, and then converted to knowledge, which is useful for decision making.

In Lokers et al. (2016), the author has presented his ideas for using Big-data technologies for data mining and collecting and storing data in different big-data environment. The authors compares different databases by testing range of Big Data characteristics.

4 Data Model

Kimballs bottom-up approach has been used here in which the dimensions are first created, which are eventually integrated together to form the data warehouse. [Kimball et al. (2013)]

Key advantages of Kimballs approach are as follows,

- Quick to set up
- Easily understood by business users
- Easy for reporting

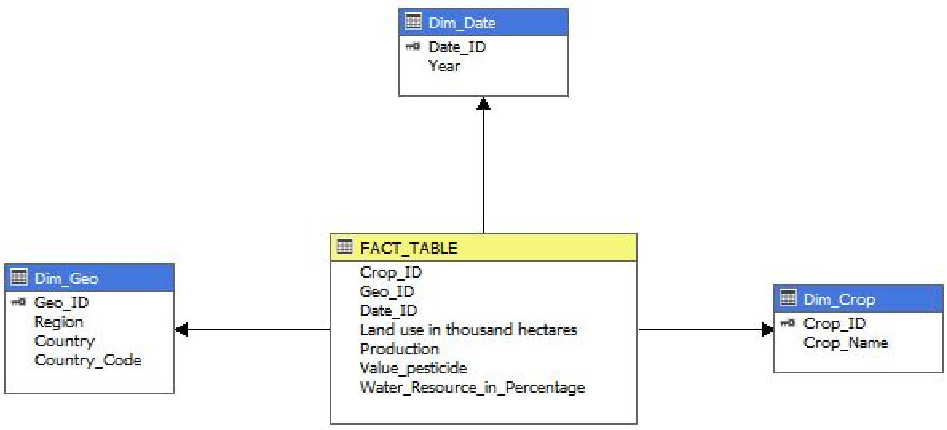


Figure 7: Star Schema

Star Schema[Figure 7] has been used as the model for this data warehouse project. Star schema is simple and more efficient. Star schema consists of one fact table, which is surrounded by multiple dimension tables. Dimensions provide labelling information to the numeric values that are in the fact table. The Dimensions used in this project are Dimension Crop (Dim_Crop), Dimension Geo (Dim_Geo) and Dimension Date (Dim_Date). Each dimension has a primary key, which has been included in the fact table as a foreign key. Composition of each dimensions and the attributes present in each are given in detail, below

4.1 Dim_Geo

Dim_Geo consists of the following columns,
a) Geo_ID
b) Region
c) Country
d) Country Code
Geo_ID is the primary key in this dimension and is unique for 33 different countries of Europe. The countries have been split into different regions for drill down purpose. Country : Different countries of use. Country code : assigned to different countries.

4.2 Dim_Crop

Dim_Crop consists of the following columns,
a) Crop_ID
b) Crop_Name
Crop_ID is the primary key in this dimension and is unique for different crops. 14 types of crops are using for this project. Crop_Name is the name of each crop.

4.3 Dim_Date

Dim_Date consists of the following columns,
a) Date_ID
b) Year
Date_ID is the primary key in this dimension and is unique for different crops. The data has been gathered for the years 2003 to 2017

4.4 Fact Table

The Fact table contains crop_ID, Geo_ID, Date_ID and the measures Land use in thousand hectares, Production, Value_pesticide, Water_Resource_in_Percentage. The Crop_ID, Geo_ID and Date_ID are referenced here as the foreign key which were the primary keys pulled from the dimension tables Dim_Crop, Dim_Geo and Dim_Date respectively.

5 Logical Data Map

Table 3: Logical Data Map describing all transformations, sources and destinations for all components of the data model

Source	Column	Destination	Column	Type	Transformation
1	Total agricultural land area in the United Kingdom (UK) 2003-2017	Dim_Date	Year	Dimension	Converted the column name to year using R
1	Land use in thousand hectares	FactTable	Land use in thousand hectares	Fact	Removed the comma from all the values using R
2	Region	Dim_Geo	Region	Dimension	Moved from Dimension table to fact table. Region is used for drill down purpose. Region - country
2	Country	Dim_Geo	Country	Dimension	Moved from Dimension table to fact table. Country is used for drill down purpose. Region - Country
2	Year	Dim_Date	Year	Dimension	Year has been moved from dimension table to fact table. Year can be used for drill down purpose. Contains the years 2007 to 2018
2	Production	FactTable	Production	Fact	Contains the production value measure of each crop which was rounded off to remove the decimal points using RStudio.

Continued on next page

Table 3 – *Continued from previous page*

Source	Column	Destination	Column	Type	Transformation
2	Crop_name	Dim_Crop	Crop_Name	Dimension	Crop names from 13 different datasets downloaded from the source were combined to form the data for 13 different crops. Cleaning and combining the datasets were done using RStudio
3	Area	Dim_Geo	Country	Dimension	Moved from dimension table to fact table as country
3	Year	Dim_Date	Year	Dimension	Moved from Dimension table to fact table
3	Value_pesticide	FactTable	Value_pesticide	Fact	Contains the pesticide use value measure which was rounded off to remove the decimal points using RStudio
3	Region	Dim_Geo	Region	Dimension	Moved from Dimension table to fact table
3	Country_code	Dim_Geo	Country_code	Dimension	Generated using RStudio
4	Region	Dim_Geo	Region	Dimension	Moved from Dimension table to fact table
4	Water_Resource_in_Percentage	FactTable	Water_Resource_in_Percentage	Fact	Contains the information on available water resource region wise value measure which was rounded off to remove the decimal points using RStudio

6 ETL Process

ETL is the process of Extracting data from several sources which are in different format, cleaning them and converting them as per our need and inserting them into the data warehouse. For this project SSIS is used for ETL process.

Extraction:

The initial stage in the ETL process is to extract the desired data from different sources. There are 3 sources of data from where 19 different datasets have been extracted and transformed. Out of the 18 different datasets, 17 are structured (1 from Statista,13 from Eurostat and 4 from FAO) and 1 is unstructured. The Structured data were extracted in CSV format.

Transformation and Loading:

After extracting the desired datasets from the different sources, they were combined and cleaned as per the project requirement. The dataset that was extracted from Statista had the columns with headers as Total agricultural land area in the United Kingdom (UK) 2003-2017 and Land use in thousand hectares out of which the header of the first column was changed to Year. The 13 structured datasets extracted from Eurostat were cleaned and combined into a single dataset which contains the information on different crops which are organized region wise, country wise and year wise. The 4 structured datasets extracted from FAO were combined to form region wise data of pesticide use. The final dataset was extracted from image taken from FAO, from which the value of water resource was extracted.

The above cleaning process was entirely done using RStudio and the code used is provided below in the appendix. After cleaning the datasets, the datasets are uploaded into the staging table. All the tables are initially truncated using the execute Sql task in SSIS. The cleaning part using RStudio is automated inside the SSIS using the Execute Process Task. Each file is loaded in the data flow task using the flat file source, which is then connected to SSMS using the OLE DB destination. The columns in the flat file source are correctly linked to the destination database. Once this the done, the data gets loaded in the database.

After loading all the datasets into the database, Dimension tables are created using Execute SQL task, where SQL code is used to move the required column from the raw table to the dimension table. 3 Dimension tables were created, namely Dim_Crop, Dim_Geo, Dim_Date. The Dimension tables contain dimensions and their respective primary key. After the dimensions are created, fact table is generated, which contains all the measures and the primary keys of all the dimensions. After executing the fact table, the cube is automated using the sequence generator, where the dimensions are processed first then the cube is executed. After all the connections are the made, the control flow is executed. The entire process including the cleaning part and the cube generation part have been automated. The Cube deployment is done is SQL Server Analysis Services.

	Year	Agriculture_Land
1	2003	18465
2	2004	18431
3	2005	18486
4	2006	18770
5	2007	18692
6	2008	18697
7	2009	18296
8	2010	18282
9	2011	18263
10	2012	18349

	Region	Country	Year	Production	crop_name	Country_Code
1	Western_Europe	Belgium	2007	210	wheat	BEL
2	Eastern_Europe	Bulgaria	2007	1088	wheat	BGR
3	Eastern_Europe	Czechia	2007	811	wheat	CZE
4	Northern_Europe	Denmark	2007	689	wheat	DNK
5	Western_Europe	Germany	2007	2992	wheat	DEU
6	Northern_Europe	Estonia	2007	100	wheat	EST
7	Northern_Europe	Ireland	2007	84	wheat	IRL
8	Southern_Europe	Greece	2007	678	wheat	GRC
9	Southern_Europe	Spain	2007	1803	wheat	ESP
10	Western_Europe	France	2007	5239	wheat	FRA
11	Southern_Europe	Croatia	2007	175	wheat	HRV
12	Southern_Europe	Italy	2007	2100	wheat	ITA

	Area	Item	Year	Region	Country_Code	Value_pesticide
1	Belarus	Pesticides (total)	2000	Eastern_Europe	BLR	8306
2	Belarus	Pesticides (total)	2001	Eastern_Europe	BLR	8306
3	Belarus	Pesticides (total)	2002	Eastern_Europe	BLR	8306
4	Belarus	Pesticides (total)	2003	Eastern_Europe	BLR	8306
5	Belarus	Pesticides (total)	2004	Eastern_Europe	BLR	8306
6	Belarus	Pesticides (total)	2005	Eastern_Europe	BLR	8306
7	Belarus	Pesticides (total)	2006	Eastern_Europe	BLR	8306
8	Belarus	Pesticides (total)	2007	Eastern_Europe	BLR	8306
9	Belarus	Pesticides (total)	2008	Eastern_Europe	BLR	8306
10	Belarus	Pesticides (total)	2009	Eastern_Europe	BLR	8306
11	Belarus	Pesticides (total)	2010	Eastern_Europe	BLR	8306
12	Belarus	Pesticides (total)	2011	Eastern_Europe	BLR	8306

	Region	Water_Resource_in_Percentage
1	Southern_Europe	13.110
2	Eastern_Europe	19.480
3	Northern_Europe	30.810
4	Western_Europe	27.590
5	Southern_Europe	13.110
6	Eastern_Europe	19.480
7	Northern_Europe	30.810
8	Western_Europe	27.590
9	Southern_Europe	13.110
10	Eastern_Europe	19.480
11	Northern_Europe	30.810
12	Western_Europe	27.590
13	Southern_Europe	13.110

Figure 8: Database

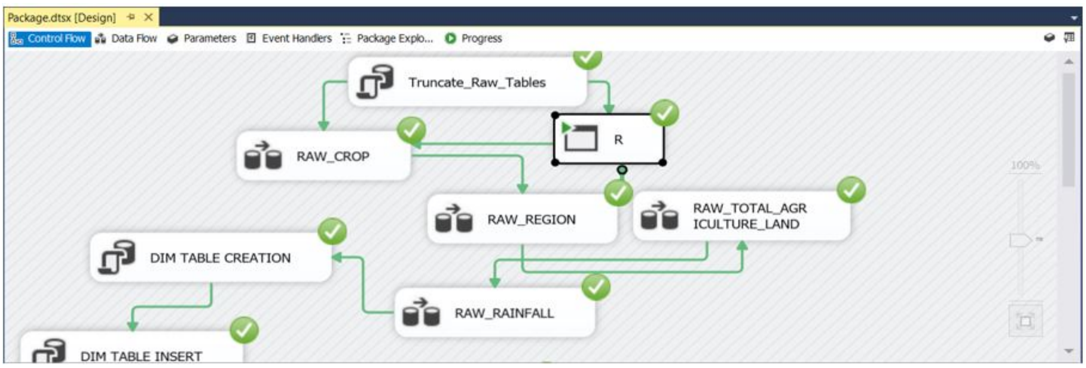


Figure 9: Control Flow

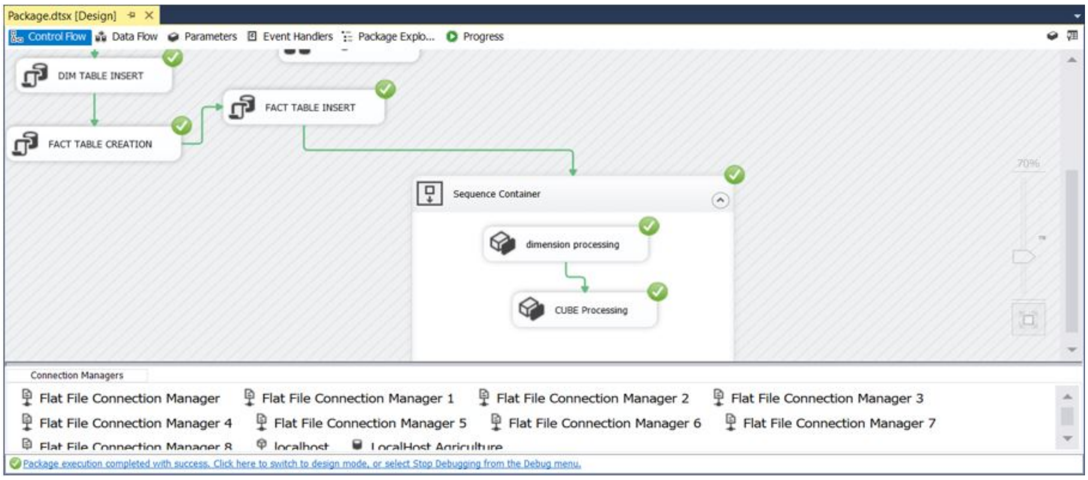


Figure 10: Control Flow

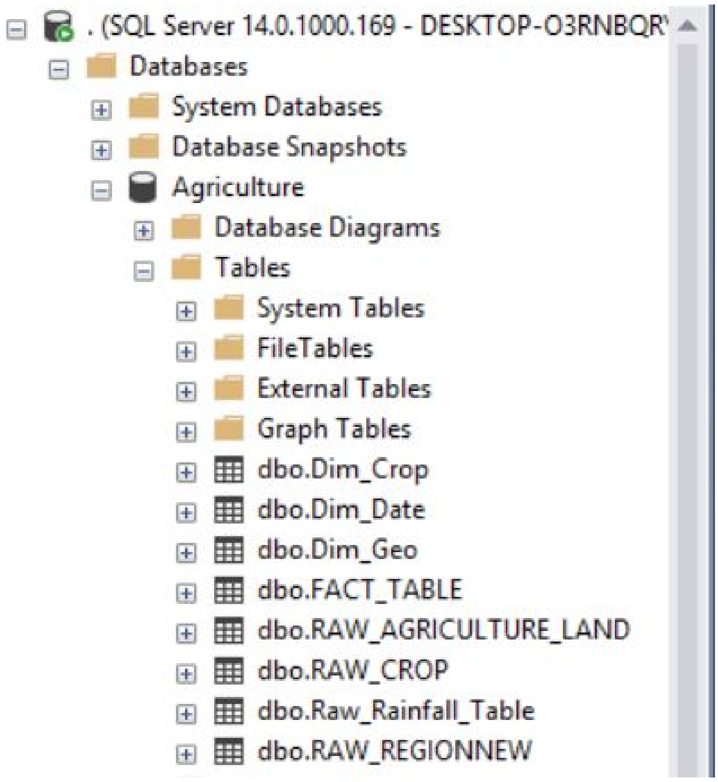


Figure 11: Database

Cube Deployment: Once the Dimension tables and the Fact table are loaded, the cube is processed using SQL Server Analysis Services. The cube contists of dimensions and a fact table. Once the cube is loaded, the star schema is created. The fact table creation is automated using the analysis services automation task in SSIS

Deployment.JPG Deployment.JPG

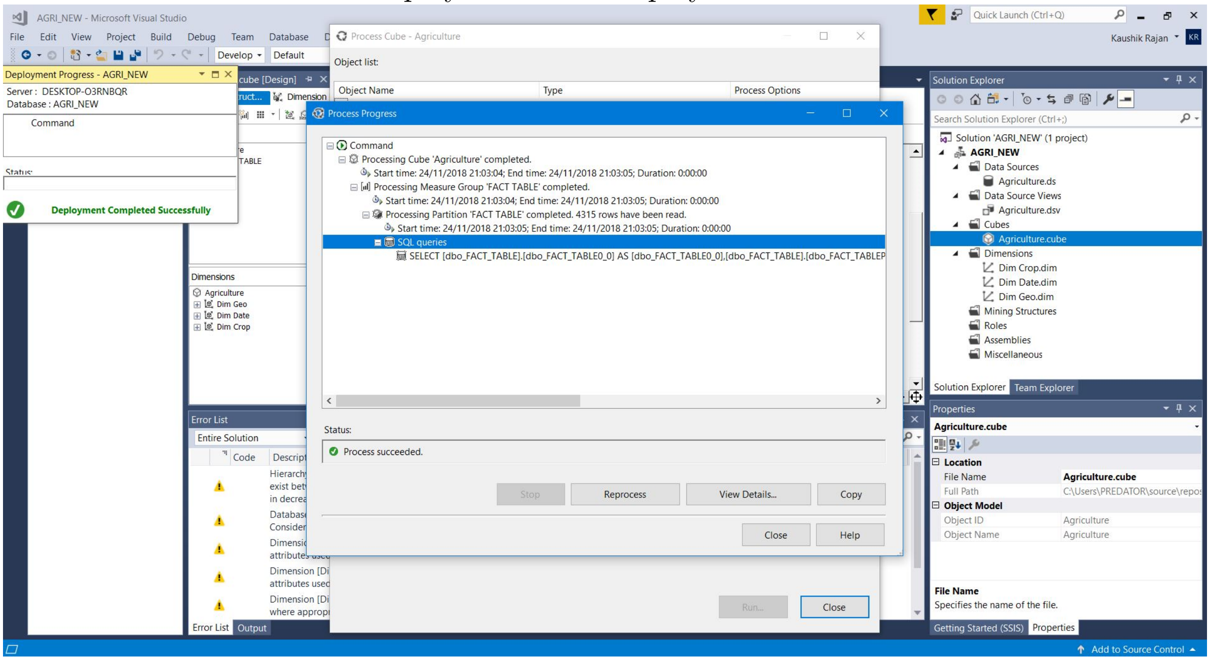


Figure 12: Cube Processing

7 Application

7.1 BI Query 1: Total Agricultural land vs Crops produced

For this business query, the contributing data sources are Statista and Eurostat. This query compares the total agricultural land used and the amount of crops produced.

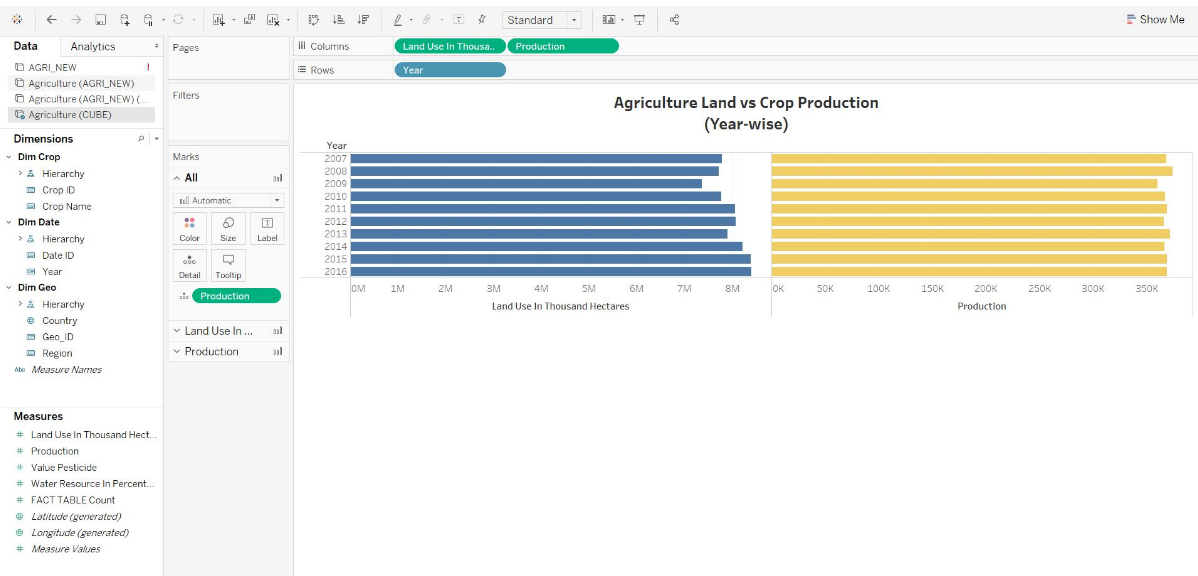


Figure 13: Results for BI Query 1

From the (Figure 13) it can be inferred that, the available land for agriculture have been on decline from 2007 till 2011 and from 2012 there have been recovery of sorts due the conservation efforts made by the government. While the available land has gone up and down and up again, the production of crops has been steady from the year 2007 till 2016. as illustrated in

7.2 BI Query 2: Crops produced vs Pesticide use

For this query, the contributing sources of data are: Eurostat and FAO This query compares the amount of crop produced vs the amount of pesticide used.

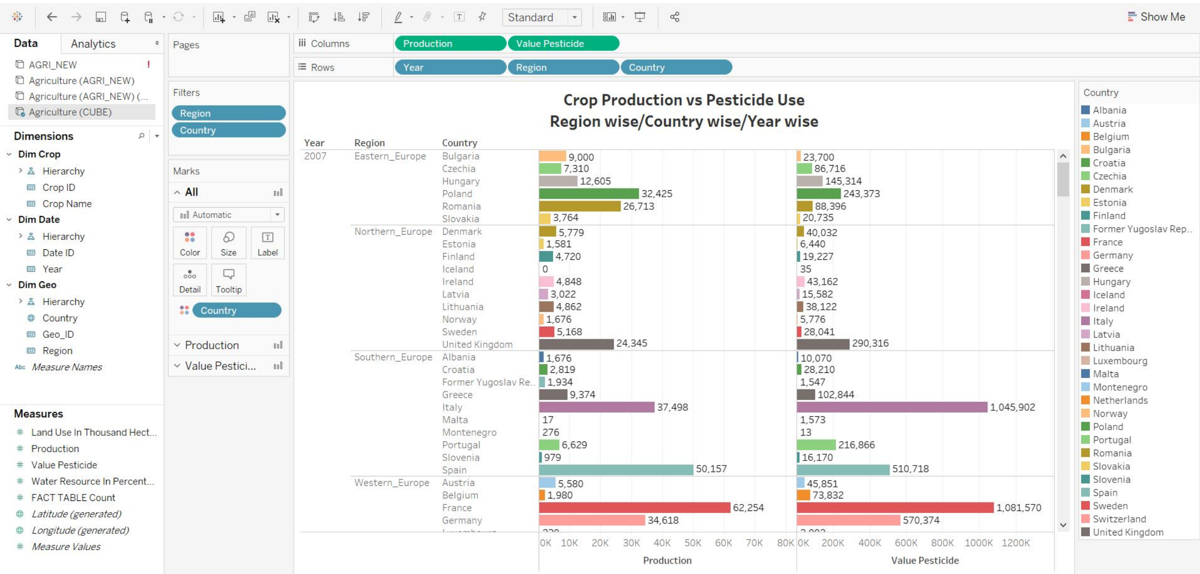


Figure 14: Results for BI Query 2

The graph (Figure 14) depicts the correlation between the crop production and pesticide use. More the amount of crops produced, more the amount of pesticide is used. The graph can further be drilled down by selecting a particular year or country or region.

7.3 BI Query 3: Total Agricultural land vs Available water resource

For this Business query, the contributing data sources are Statista and FAO. This query compares the Total Agricultural land used with the available water resource in each region.

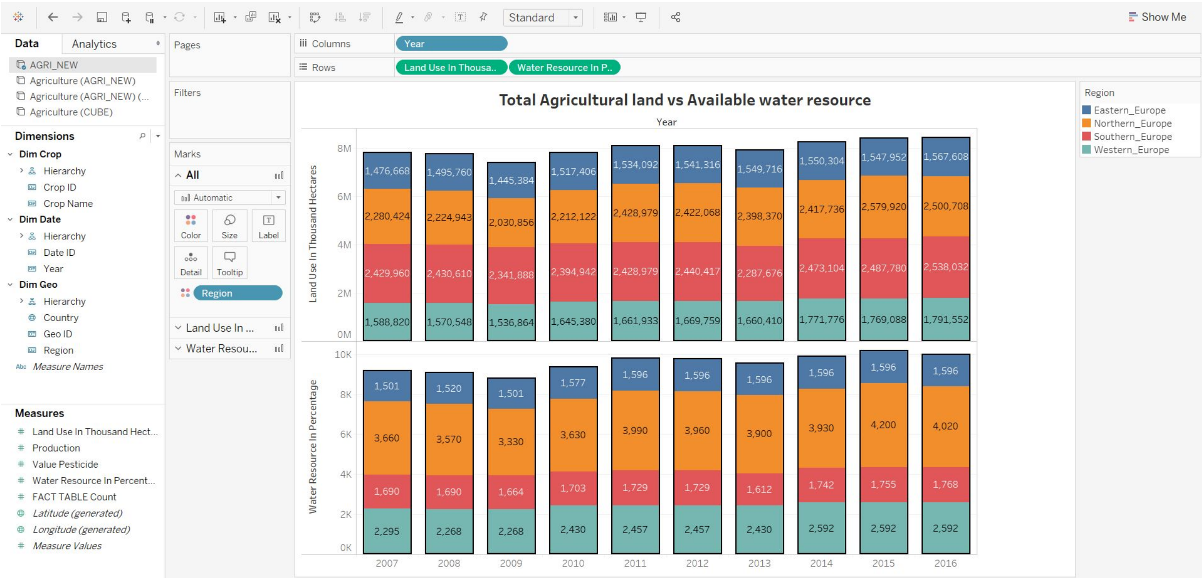


Figure 15: Results for BI Query 3

The Graph(Figure 15) shows the division of land used for agriculture in each region of Europe and also the percentage of water resource available in each region.

7.4 Discussion

From the 3 business queries, it can be summarized that the agriculture production depends on multiple factors such as the available land, available water resource, amount of pesticide use.

From Query 1, it can be understood that the land usage for agriculture have been on decline from 2007 till 2011 due to the conservation efforts made by the government, the available land for agriculture have been increasing. Even though, the land usage for agriculture was highly fluctuating, the production of crops had steadily increased from 2007 till 2016. From Query 2, it can be understood that there is a correlation between the crop production and pesticide use. When more crops are produced, more pesticides are used. The query 3 shows the quantity of water resource available and the amount of land used for agriculture, region wise

In the paper Ngo et al. (2018) the author, has used continent level dataset for his project, using which the author has answered the BI queries such as crops suitable for each field, crop sale price and how to improve crop yields. Since the datasets used are vast, the author has used Big data-NoSql to store the datasets. The main difference between the author's paper and this project are the datasets (For this project datasets of Europe were only considered), database (SSMS is used as the database for this project)

8 Conclusion and Future Work

The project was developed with motive to improve agricultural production by researching on the factors that directly affect agriculture. For this project, factors such as land management, pesticide use, available natural resources were considered. For this, data

of land usage, pesticide usage and available natural resources were downloaded from different resources and were accumulated into the data warehouse model. The model answers successfully answers the Business queries for which the model was built for. The Business Intelligence Queries were successfully visualized using tableau. The project can be used by the government for improving the agriculture yield by checking the different factors which are responsible for the decline of agriculture. The model is built in the such a way that it can handle changes. The model can be updated with new data at any point of time.
In the future, various other factors that affect the agriculture crop production, such as Soil quality, weather, immigration, climate change, pollution, etc, can also be added to the model. In the current model, the data extraction part was done manually, but in the future, with the arrival of new machine learning techniques, the extraction part can also be fully automated[Liakos et al. (2018)]. In near future, Big data technologies can also be used to handle the storage of high volume of data generated. There are loads of other ways to mine data in the agriculture, e.g. sensors can be used to collected and transferred to a Big data storage system using Internet of things[Ngo et al. (2018)].

References

Cojocariu, A. & Stanciu, C. O. (2009), 'Data warehouse and agricultural intelligence.', *Agricultural Management / Lucrari Stiintifice Seria I, Management Agricol* **11**(2), 1 – 6.
URL: *http://search.ebscohost.com/login.aspx?direct=true AuthType=ip,cookie,shibdb=bthAN=72447758site=eds-live scope=sitecustid=ncirlib*

Inmon, W. H. (2005), *Building the data warehouse : [electronic book].*, Indianapolis, Ind. : Wiley, c2005.
URL: *http://search.ebscohost.com/login.aspx?direct=true AuthType=ip,cookie,shibdb=cat05743aAN=nci.30574site=eds-livescope=sitecustid=ncirlib*

Janssen, S. J., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W. & Antle, J. M. (2017), 'Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology', *Agricultural Systems* **155**, 200 – 212.
URL: *http://www.sciencedirect.com/science/article/pii/S0308521X16305637*

Kimball, R., Ross, M. & Kimball, R. (2013), *The data warehouse toolkit : the definitive guide to dimensional modeling : [electronic book].*, Indianapolis, Ind. : John Wiley, 2013.
URL: *http://search.ebscohost.com/login.aspx?direct=true AuthType=ip,cookie,shibdb=cat05743aAN=nci.31479site=eds-livescope=sitecustid=ncirlib*

Liakos, K. G., Busato, P., Moshou, D., Pearson, S. & Bochtis, D. (2018), 'Machine learning in agriculture: A review', *Sensors* **18**(8).
URL: *http://www.mdpi.com/1424-8220/18/8/2674*

Lokers, R., Knapen, R., Janssen, S., van Randen, Y. & Jansen, J. (2016), 'Analysis of big data technologies for use in agro-environmental science', *Environmental Modelling*

Software **84**, 494 – 504.
URL: *http://www.sciencedirect.com/science/article/pii/S1364815216304194*

Ngo, V. M., Le-Khac, N.-A. & Kechadi, M. T. (2018), 'An efficient data warehouse for crop yield prediction', *CoRR* **abs/1807.00035**.

Appendix

R code example

```
setwd("D:/Education/IRELAND/NCI/Moodle␣Documents/SEM␣1/DWBI/
DataSets/Agriculture")
// Cleaning Datasets from FAO

library(countrycode)
library(tidyr)
library(sqldf)

Eastern_Europe<-read.csv("Eastern␣Europe.csv")
Western_Europe<-read.csv("Western␣Europe.csv")
Southern_Europe<-read.csv("Southern␣Europe.csv")
Northern_Europe<-read.csv("Northern␣Europe.csv")

Eastern_Europe$Region<- "Eastern_Europe"
Western_Europe$Region<- "Western_Europe"
Southern_Europe$Region<- "Southern_Europe"
Northern_Europe$Region<- "Northern_Europe"

Region<-rbind(Eastern_Europe,Western_Europe,Southern_Europe,Northern_Europe)

library(countrycode)
Region<-Region[c(4,8,10,12,15)]
Region$Country_Code <- countrycode(Region$Area,'country.name','iso3c')

//write.csv(Region,file="Region.csv")

//Importing data

Barley<- read.csv("Barley␣by␣area,␣production␣and␣humidity.csv")
Cereals<-read.csv("Cereals␣for␣the␣production␣of␣grain.csv")
Dry_pulses<-read.csv("Dry␣pulses␣and␣protein␣crops.csv")
Fresh_vegetables<-read.csv("Fresh␣vegetables␣and␣strawberries␣by␣area.csv")
Grain_maize<-read.csv("Grain␣maize␣and␣corn-cob-mix␣by␣area,
production␣and␣humidity.csv")
Root_Crops<-read.csv("Root␣crops␣and␣plants␣harvested␣green
from␣arable␣land␣by␣area.csv")
Grapes<-read.csv("Grapes␣by␣production.csv")
green_maize<-read.csv("Grapes␣by␣production.csv")
oats<-read.csv("Oats␣and␣spring␣cereal␣mixtures␣by␣area,
```

```
production␣and␣humidity.csv")
Olives<-read.csv("Olives␣by␣production.csv")
Permanent_crops<-read.csv("Permanent␣crops␣for␣human␣consumption
by␣area.csv")
turnip_rape<-read.csv("Rape,␣turnip␣rape,␣sunflower␣seeds␣and
soya␣by␣area.csv")
Rye<-read.csv("Rye␣and␣winter␣cereal␣mixtures␣by␣area,
production␣and␣humidity.csv")
Utilised_agricultural_area<-read.csv("Utilised␣agricultural
area␣by␣categories.csv")
Wheat<-read.csv("Wheat.csv")
total_agriculture_land_statista<-read.csv("statistic_id315937_
total-agricultural-land-area-in-the-united-kingdom--uk--2003-2017.csv")

//Cleaning

Barley<-Barley[c(2,6:44),]
Cereals<-Cereals[c(2,6:44),]
Dry_pulses<-Dry_pulses[c(2,6:44),]
Fresh_vegetables<-Fresh_vegetables[c(3,7:45),]
Grain_maize<-Grain_maize[c(2,6:44),]
Grapes<-Grapes[c(3,7:45),]
green_maize<-green_maize[c(3,7:45),]
oats<-oats[c(2,6:44),]
Olives<-Olives[c(3,7:45),]
Permanent_crops<-Permanent_crops[c(3,7:45),]
Root_Crops<-Root_Crops[c(3,7:45),]
Rye<-Rye[c(2,6:44),]
total_agriculture_land_statista<-total_agriculture_land_statista[c(2:17),]
turnip_rape<-turnip_rape[c(3,7:45),]
Utilised_agricultural_area<-Utilised_agricultural_area[c(3,7:45),]
Wheat<-Wheat[c(2,6:44),]

// Wheat Cleaning

Wheat<-read.csv("Wheat.csv")
Wheat<-Wheat[c(2,6:44),]
colnames(Wheat) <- as.character(unlist(Wheat[1,]))
Wheat = Wheat[-1, ]
rownames(Wheat) <- c(1:nrow(Wheat))
Wheat <- Wheat[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Wheat <- Wheat %>% gather(Year,Production,`2007`,`2008`,`2009`,`2010`
,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Wheat) <- c("Country","Year","Production")


library(sqldf)
Wheat<- sqldf('Select␣*␣from␣Wheat␣Where␣Production␣!=␣":"')
Wheat$crop_name <- "wheat"
Wheat$Country_Code <- countrycode(Wheat$Country,'country.name','iso3c')
```

```
Wheat<- na.omit(Wheat)

// Barley cleaning

colnames(Barley)<-as.character(unlist(Barley[1,]))
Barley = Barley[-1,]
rownames(Barley)<-c(1:nrow(Barley))
Barley <- Barley[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Barley <- Barley %>% gather(Year,Production,`2007`,`2008`,`2009`,`2010`,
`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Barley) <- c("Country","Year","Production")

Barley<- sqldf('Select * from Barley Where Production != ":"')
Barley$crop_name <- "Barley"
Barley$Country_Code <- countrycode(Barley$Country,'country.name','iso3c')

Barley<- na.omit(Barley)

// Cereals cleaning
colnames(Cereals)<-as.character(unlist(Cereals[1,]))
Cereals = Cereals[-1,]
rownames(Cereals)<-c(1:nrow(Cereals))
Cereals <- Cereals[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Cereals <- Cereals %>% gather(Year,Production,`2007`,`2008`,`2009`,`2010`,
`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Cereals) <- c("Country","Year","Production")

Cereals<- sqldf('Select * from Cereals Where Production != ":"')
Cereals$crop_name <- "Cereals"
Cereals$Country_Code <- countrycode(Cereals$Country,'country.name','iso3c')

Cereals<- na.omit(Cereals)

// Dry_pulses Cleaning

colnames(Dry_pulses)<-as.character(unlist(Dry_pulses[1,]))
Dry_pulses = Dry_pulses[-1,]
rownames(Dry_pulses)<-c(1:nrow(Dry_pulses))
Dry_pulses <- Dry_pulses[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Dry_pulses <- Dry_pulses %>% gather(Year,Production,`2007`,`2008`,`2009`,
`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Dry_pulses) <- c("Country","Year","Production")

Dry_pulses<- sqldf('Select * from Dry_pulses Where Production != ":"')
Dry_pulses$crop_name <- "Dry_pulses"
Dry_pulses$Country_Code <- countrycode(Dry_pulses$Country,
'country.name','iso3c')

Dry_pulses<- na.omit(Cereals)

// Fresh_vegetables Cleaning
```

```
colnames(Fresh_vegetables)<-as.character(unlist(Fresh_vegetables[1,]))
Fresh_vegetables = Fresh_vegetables[-1,]
rownames(Fresh_vegetables)<-c(1:nrow(Fresh_vegetables))
Fresh_vegetables <- Fresh_vegetables[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,
-21,-23,-25)]
Fresh_vegetables <- Fresh_vegetables %>% gather(Year,Production,`2007`
,`2008`,`2009`,`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,
`2017`,`2018`)
colnames(Fresh_vegetables) <- c("Country","Year","Production")

Fresh_vegetables<- sqldf('Select * from Fresh_vegetables Where
Production != ":"')
Fresh_vegetables$crop_name <- "Fresh_vegetables"
Fresh_vegetables$Country_Code <- countrycode(Fresh_vegetables$Country,
'country.name','iso3c')

Fresh_vegetables<- na.omit(Fresh_vegetables)

// Grain_maize Cleaning

colnames(Grain_maize)<-as.character(unlist(Grain_maize[1,]))
Grain_maize = Grain_maize[-1,]
rownames(Grain_maize)<-c(1:nrow(Grain_maize))
Grain_maize <- Grain_maize[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Grain_maize <- Grain_maize %>% gather(Year,Production,`2007`,
`2008`,`2009`,`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`
colnames(Grain_maize) <- c("Country","Year","Production")

Grain_maize<- sqldf('Select * from Grain_maize Where Production != ":"')
Grain_maize$crop_name <- "Grain_maize"
Grain_maize$Country_Code <- countrycode(
Grain_maize$Country,'country.name','iso3c')

Grain_maize<- na.omit(Grain_maize)

// Grapes Cleaning

colnames(Grapes)<-as.character(unlist(Grapes[1,]))
Grapes = Grapes[-1,]
rownames(Grapes)<-c(1:nrow(Grapes))
Grapes <- Grapes[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Grapes <- Grapes %>% gather(Year,Production,`2007`,`2008`,`2009`,`2010`,
`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Grapes) <- c("Country","Year","Production")

Grapes<- sqldf('Select * from Grapes Where Production != ":"')
Grapes$crop_name <- "Grapes"
Grapes$Country_Code <- countrycode(Grapes$Country,'country.name','iso3c')

Grapes<- na.omit(Grapes)
```

```
// green_maize Cleaning

colnames(green_maize)<-as.character(unlist(green_maize[1,]))
green_maize = green_maize[-1,]
rownames(green_maize)<-c(1:nrow(green_maize))
green_maize <- green_maize[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
green_maize <- green_maize %>% gather(Year,Production,`2007`,`2008`,`2009`,`
`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(green_maize) <- c("Country","Year","Production")

green_maize<- sqldf('Select * from green_maize Where Production != ":"')
green_maize$crop_name <- "green_maize"
green_maize$Country_Code <- countrycode(green_maize$Country,'country.name','

green_maize<- na.omit(green_maize)

// oats Cleaning

colnames(oats)<-as.character(unlist(oats[1,]))
oats = oats[-1,]
rownames(oats)<-c(1:nrow(oats))
oats <- oats[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
oats <- oats %>% gather(Year,Production,`2007`,`2008`,`2009`
,`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(oats) <- c("Country","Year","Production")

oats<- sqldf('Select * from oats Where Production != ":"')
oats$crop_name <- "oats"
oats$Country_Code <- countrycode(oats$Country,'country.name','iso3c')

oats<- na.omit(oats)

//Olives Cleaning

colnames(Olives)<-as.character(unlist(Olives[1,]))
Olives = Olives[-1,]
rownames(Olives)<-c(1:nrow(Olives))
Olives <- Olives[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Olives <- Olives %>% gather(Year,Production,`2007`,`2008`,`2009`,`2010`,
`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Olives) <- c("Country","Year","Production")

Olives<- sqldf('Select * from Olives Where Production != ":"')
Olives$crop_name <- "Olives"
Olives$Country_Code <- countrycode(Olives$Country,'country.name','iso3c')

Olives<- na.omit(Olives)

// Permanent_crops Cleaning
```

```
colnames(Permanent_crops)<-as.character(unlist(Permanent_crops[1,]))
Permanent_crops = Permanent_crops[-1,]
rownames(Permanent_crops)<-c(1:nrow(Permanent_crops))
Permanent_crops <- Permanent_crops[,c(-3,-5,-7,-9,-11,-13,-15,-17,
-19,-21,-23,-25)]
Permanent_crops <- Permanent_crops %>% gather(Year,Production,`2007`,
`2008`,`2009`,`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,
`2017`,`2018`)
colnames(Permanent_crops) <- c("Country","Year","Production")

Permanent_crops<- sqldf('Select * from Permanent_crops Where
Production != ":"')
Permanent_crops$crop_name <- "Permanent_crops"
Permanent_crops$Country_Code <- countrycode(Permanent_crops$Country,
'country.name','iso3c')

Permanent_crops<- na.omit(Permanent_crops)

// Root_Crops Cleaning

colnames(Root_Crops)<-as.character(unlist(Root_Crops[1,]))
Root_Crops = Root_Crops[-1,]
rownames(Root_Crops)<-c(1:nrow(Root_Crops))
Root_Crops <- Root_Crops[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Root_Crops <- Root_Crops %>% gather(Year,Production,`2007`,`2008`,
`2009`,`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,
`2017`,`2018`)
colnames(Root_Crops) <- c("Country","Year","Production")

Root_Crops<- sqldf('Select * from Root_Crops Where Production != ":"')
Root_Crops$crop_name <- "Root_Crops"
Root_Crops$Country_Code <- countrycode(Root_Crops$Country,
'country.name','iso3c')

Root_Crops<- na.omit(Root_Crops)

// Rye Cleaning

colnames(Rye)<-as.character(unlist(Rye[1,]))
Rye = Rye[-1,]
rownames(Rye)<-c(1:nrow(Rye))
Rye <- Rye[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
Rye <- Rye %>% gather(Year,Production,`2007`,`2008`,
`2009`,`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Rye) <- c("Country","Year","Production")

Rye<- sqldf('Select * from Rye Where Production != ":"')
Rye$crop_name <- "Rye"
Rye$Country_Code <- countrycode(Rye$Country,
'country.name','iso3c')
```

```
Rye<- na.omit(Rye)

// turnip_rape Cleaning

colnames(turnip_rape)<-as.character(unlist(turnip_rape[1,]))
turnip_rape = turnip_rape[-1,]
rownames(turnip_rape)<-c(1:nrow(turnip_rape))
turnip_rape <- turnip_rape[,c(-3,-5,-7,-9,-11,-13,-15,-17,-19,-21,-23,-25)]
turnip_rape <- turnip_rape %>% gather(Year,Production,`2007`,`2008`,`2009`,
`2010`,`2011`,`2012`,`2013`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(turnip_rape) <- c("Country","Year","Production")

turnip_rape<- sqldf('Select␣*␣from␣turnip_rape␣Where␣Production␣!=␣":"')
turnip_rape$crop_name <- "turnip_rape"
turnip_rape$Country_Code <- countrycode(turnip_rape$Country,
'country.name','iso3c')

turnip_rape<- na.omit(turnip_rape)

// Utilised_agricultural_area

Utilised_agricultural_area<-read.csv("Utilised␣agricultural
area␣by␣categories.csv")
Utilised_agricultural_area<-Utilised_agricultural_area[c(3,7:45),]
Utilised_agricultural_area <- Utilised_agricultural_area[,c(-3,-5,-7,-9,
-11,-13,-15,-17,-19,-21,-23,-25)]
colnames(Utilised_agricultural_area)<-as.character(unlist(Utilised
_agricultural_area[1,]))
Utilised_agricultural_area = Utilised_agricultural_area[-1,]
rownames(Utilised_agricultural_area)<-c(1:nrow(Utilised_agricultural_area))

Utilised_agricultural_area <- Utilised_agricultural_area %>% gather(Year,
Production,`2007`,`2008`,`2009`,`2010`,`2011`,`2012`,`2013
`,`2014`,`2015`,`2016`,`2017`,`2018`)
colnames(Utilised_agricultural_area) <- c("Country","Year","Production")

Utilised_agricultural_area<- sqldf('Select␣*␣from
Utilised_agricultural_area␣Where␣Production␣!=␣":"')
Utilised_agricultural_area$crop_name <- "Utilised_agricultural_area"
Utilised_agricultural_area$Country_Code <- countrycode(
Utilised_agricultural_area$Country,'country.name','iso3c')

Utilised_agricultural_area<- na.omit(Utilised_agricultural_area)


crop <- rbind(Wheat,Barley,Cereals,Dry_pulses,Fresh_vegetables,
Grain_maize,Grapes,green_maize,oats,Olives,Permanent_crops,
Root_Crops,Rye,turnip_rape,Utilised_agricultural_area)
crop <- sqldf('select␣distinct␣b.Region,a.
*␣from␣crop␣a␣join␣Region␣b␣on␣a.Country_Code␣=␣b.Country_Code')
colnames(Region)[4] <- "Value_pesticide"
```

```
// Statista

colnames(total_agriculture_land_statista)<-as.character(unlist(
total_agriculture_land_statista[1,]))
total_agriculture_land_statista = total_agriculture_land_statista[-1,]
colnames(total_agriculture_land_statista) <- c("Year","Land use
in thousand hectares")
library(stringr)


crop$Production<-as.numeric(crop$Production)
crop$Production<-round(crop$Production)
Region$Value_pesticide<-round(Region$Value_pesticide)

write.csv(crop, file = "crop.csv",row.names = FALSE)

write.csv(Region,file="Region.csv", row.names = FALSE)

colnames(total_agriculture_land_statista)[2] <- "Agriculture_Land"
total_agriculture_land_statista$Agriculture_Land<- str_replace_all(
total_agriculture_land_statista$Agriculture_Land, "[^[:alnum:]]", "")


write.csv(total_agriculture_land_statista,file="total_
agriculture_land_Statista.csv",row.names = FALSE)

// SQL
// DIM Table Creation
USE [Agriculture]

GO

DROP table FACT_TABLE
DROP table Dim_Crop
DROP table Dim_Date
DROP table Dim_Geo


CREATE TABLE [dbo].[Dim_Crop](
        [Crop_ID] [int] IDENTITY(1,1) NOT NULL PRIMARY KEY,
        [Crop_Name] [varchar](50) NULL,
        )


CREATE TABLE [dbo].[Dim_Geo](
        [Geo_ID] [numeric](18, 0) IDENTITY(1,1) NOT NULL PRIMARY KEY,
        [Region] [varchar](50) NULL,
        [Country] [varchar](50) NULL,
        [Country_Code] [varchar](50) NULL
```

```
        )

CREATE TABLE [dbo].[Dim_Date](
        [Date_ID] [numeric](18, 0) IDENTITY(1,1) NOT NULL PRIMARY KEY,
        [Year] [varchar](50) NULL
        )

GO
// DIM Table Insertion

USE [Agriculture]
GO

TRUNCATE TABLE [dbo].[Dim_Crop]
INSERT INTO [dbo].[Dim_Crop]
           ([Crop_Name])
SELECT DISTINCT crop_name
from RAW_CROP

TRUNCATE TABLE [dbo].[Dim_Geo]
INSERT INTO [dbo].[Dim_Geo]
           ([Region]
           ,[Country]
           ,[Country_Code])
SELECT DISTINCT Region,
        Country,
        Country_Code
from [dbo].[RAW_CROP]

TRUNCATE TABLE [dbo].[Dim_Date]
INSERT INTO [dbo].[Dim_Date]
           ([Year])
SELECT DISTINCT Year
        from [Agriculture].[dbo].[RAW_AGRICULTURE_LAND]
GO

// Fact Table creation
USE [Agriculture]
GO

IF OBJECT_ID('FACT_TABLE', 'U') IS NOT NULL
  drop TABLE dbo.FACT_TABLE;
GO

CREATE TABLE [dbo].[FACT_TABLE](
        [Crop_ID] int,
        [Geo_ID] [numeric](18, 0),
        [Date_ID] [numeric](18, 0),
        [Land use in thousand hectares] int NULL,
        [Production] int NULL,
        [Value_pesticide] int NULL,
```

```
        [Water_Resource_in_Percentage] int NULL
) ON [PRIMARY]
GO

ALTER TABLE [FACT_TABLE]
ADD
CONSTRAINT  fk_Crop FOREIGN KEY (Crop_ID) REFERENCES [Dim_Crop](Crop_ID),
CONSTRAINT  fk_Location FOREIGN KEY (Geo_ID) REFERENCES [Dim_Geo](Geo_ID),
CONSTRAINT  fk_Year FOREIGN KEY (Date_ID) REFERENCES [Dim_Date](Date_ID)
go
GO

// Fact table Insertion

INSERT INTO [dbo].[FACT_TABLE]
           ([Crop_ID]
           ,[Geo_ID]
           ,[Date_ID]
           ,[Land use in thousand hectares]
           ,[Production]
           ,[Value_pesticide]
                   ,[Water_Resource_in_Percentage])
select distinct
                        [Crop_ID]
           ,[Geo_ID]
           ,[Date_ID]
                   ,cast(a.[Agriculture_Land] as numeric(18,0)) as
                   Agriculture_Land
                   ,cast(b.[Production] as numeric(18,0))
as Production
                   ,CAST(c.[Value_pesticide] as numeric(18,0)) as
                   Value_pesticide
                   ,[Water_Resource_in_Percentage]
                   from [dbo].[RAW_AGRICULTURE_LAND] a
                   join [dbo].[RAW_CROP] b
                   on a.Year = b.Year
                   join [dbo].RAW_REGIONNEW c
                   on b.Year = c.Year  and b.Country_Code = c.
                   Country_Code
                   join [dbo].Raw_Rainfall_Table d
                   on b.Region = d.Region
                   join [dbo].Dim_Crop e
                   on b.crop_name = e.Crop_Name
                   join [dbo].Dim_Date f
                   on a.Year = f.Year
                   join [dbo].Dim_Geo g
                   on b.Country_Code = g.Country_Code

GO
```