National College of Ireland

Data Warehousing and Business Intelligence Project

on

Agriculture Imapct On Indonesia

Anurag Abhay Singh
x18104053

MSc/PGDip Data Analytics – 2018/9

Submitted to: Dr. Simon Caton

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

Student Name:	Anurag Abhay Singh
Student ID:	x18104053
Programme:	MSc Data Analytics
Year:	2018/9
Module:	Data Warehousing and Business Intelligence
Lecturer:	Dr. Simon Caton
Submission Due Date:	26/11/2018
Project Title:	Agriculture Imapct On Indonesia

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

Agriculture Imapct On Indonesia

Anurag Abhay Singh
x18104053

November 26, 2018

Abstract

This Project will analyse existing key relationship agriculture with government and FMCG industry. Project will also demonstrate about approach and platform used to deliver this project. This project will provide insight to Agriculture Impact on Indonesia which helped its economy in many ways, different dataset used for purpose of generating knowledge in respect of Agriculture production, export and FMCG company stock analysis with respect to my topic with the help of different tools. Project shows general trend among people adoption to regarding agriculture industry and how implementing the factors in project with the help of government various companies and people of that country can be benefited. If this method is adopted it can surely build relationship to other industry which can make agriculture as centre of attraction to all industry.

1 Introduction

1 Introduction Food Industry is increasing at exponential rate so is the data increasing with FMCG industry, effecting both people and economy of country as whole, so in this project of Data Warehouse tried to fetch out some useful knowledge from data available on internet using some datasets, web scraping and using API's in general Data warehouse project is shaped to give insight or confidence in decision making with relevant all-round information present with Country/Industry. It is an optimized version of operational database provide only relevant information for and also provide fast and reliable access to Data, Data Warehouse is method oriented, Integrated, Time-variant, Non-Volatile Platform This report will provide convenient architecture used and overall implementation approach followed for this data warehouse project. Report consists of following parameters given below: 1. Methodologies and Architecture used to build Data Warehouse. 2. Data Modelling – Properly documenting the schema, usage of Datasets and Drill down approach usage in the model. 3. Extract, Transform and Load(ETL) – Information about complexity of ETL, usage of Emerging Technologies in ETL, Automated ETL, describing methodologies used for ETL. 4.Business Intelligence – Number of Business Queries will be explained in this document with methodologies used and how all datasets have been used in the process of building, critical evaluation of Business Queries using appropriate academics.

For this project Ralph Kimball methodologies is used because it suits the requirement of the small scale project, here I will be discussing what made me choose Ralph Kimball approach, one of the main reason behind is for choosing for Ralph Kimball because

Source	Type	Brief Summary
Statista	Structured	Used because it was mandatory to get find one data set from here and this dataset connect to all my dataset with year and country same in it.
OECD	Structured	Have Suitable form of data in excel of all country which was cleaned to get details about specific country.
Investing.com	Unstructured	This data obtained because it is part of relationship which is required to connect my one dataset to another and make reliable analysing.
World Bank Data	Structured	The data obtained from here have micro level information about each aspect needed for my project.

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

its suitability to Small Scale Projects, Small Business Area, User based, small budget with Bottom-up approach through informative bus architecture, Multi-Dimensional and Structure which can be easily understood by actual user. My business requirement for this project which generated after Data Warehouse was implemented.

(Req-1) 1. How people employment can bring out more productivity in terms of agriculture and Industria production?

(Req-2) 2.Fertilizer Industry growth with use of agriculture land which can lead to more productivity?

(Req-3) 3. FMCG companies growth with increase urban population?

2 Data Sources

Here you should present and formally describe your sources of data used in the project.

2.1 Source 1:Statista

The STATISTA dataset downloaded from: `https://www.statista.com/statistics/603387/indonesia-urban-population/` is structured dataset provides 14 columns of information on urban population from year 2005-2018. These is relevant to this project because this dataset addresses the business requirements listed in Section 1 by providing how can FMCG industry can build business model.

2.2 Source 2:OECD

The OECD dataset downloaded from: `https://data.oecd.org/agroutput/crop-production.htm/` is structured dataset provides columns of information on crop production from year 2009-2017.

The OECD dataset downloaded from: https://data.oecd.org/gdp/investment-by-asset.htm#indicator-chartis structured dataset provides columns of information on cultivated asset from year 2009-2017. These is relevant to this project because this dataset addresses the business requirements listed in Section 1 by providing how can agriculture industry can help in industrial production.

2.3 Source 3:World Bank Data

The WORLD BANK DATA dataset downloaded from:http://databank.worldbank.org/data/source/indonesia-database-for-policy-and-economic-research is structured dataset provides columns of information on agriculture food export, gdp, agriculture land used, fertilizer production from year 2009-2017. These is relevant to this project because this dataset addresses the government attitude in focussing on its agriculture strength to help its economy.

2.4 Source :Investing.com

BISI stock data is downloaded from stock chart from year from May 2007- Nov 2018 https://www.investing.com/common/modules/js_instrument_chart/api/data.php?pair_id=101353&pair_id_for_news=101353&chart_type=area&pair_interval=month&candle_count=120&events=yes&volume_series=yes&period=max

UNVR stock data is downloaded from stock chart from year Jan 2004 – Nov 2018 https://www.investing.com/common/modules/js_instrument_chart/api/data.php?pair_id=101622&pair_id_for_news=101622&chart_type=area&pair_interval=month&candle_count=120&events=yes&volume_series=yes&period=max

INDF stock data is downloaded from stock chart from year from Jan 2008- Nov 2018 https://www.investing.com/common/modules/js_instrument_chart/api/data.php?pair_id=101443&pair_id_for_news=101443&chart_type=area&pair_interval=month&candle_count=120&events=yes&volume_series=yes&period=max this all dataset is unstructured dataset provides columns of information of stock value of FMCG stock listed company in Indonesia from year 2009-2017. These is relevant to this project because this dataset addresses how FMCG companies are performing in Indonesia.

3 Related Work

Sustainable agriculture in Indonesia: Facts and challenges to keep growing in harmony with environment M. Faiz Syuaib has been the journal which had previously worked on somewhat related to this topic. The dataset used in this literature review preparation is agriculture land measurement, crop production, fertilizer usage and urban population demographic to show the rapid pace of agricultural development in the last four decades as well as the commercialization, industrialization and urbanization which has led to significant changes in agricultural production systems. In some regions up gradation of agricultural technologies has increased production to keep pace with the population growth in urban areas. This literature review is providing support to my project that if listed factors in dataset is worked properly it can benefit and create a new scope to our business requirement by developing suitable business intelligence. Agriculture has never been priority of any countries government because they don't consider it's a driving force for its economy, but consider just as a part of system which we need in our daily lifestyle.

There has been always attempt to boost industrial production on verge of declining agriculture production by replacing agriculture land with industrial areas. Many people who work as farmer have lost their bread because of government policies which is making economic corridor for industrial companies. Due to which there is huge migration from rural areas to urban areas which has created a huge imbalance in urban development model. This whole scenario turns into vicious circle which can has long term effect in future by degrading life in urban areas. The project can give results how can agriculture development and providing up gradation connection with industrialization to produce FMCG goods can create a proper balance of socio-economic development in country, by helping government prioritising its duties to help farmers of its country which can never be replaced. Implementing records data analytics in agriculture can boost its productivity by creating awareness among people in these occupation by making realizing them future scope.

4 Data Model

So, far in this project I have used bottom up Approach of Ralph Kimball, according to Kimball Approach we have to start with Star schema when we have fact tables and some dimension tables, fact will give you all measures whereas dimensions contains information, dimensions contains a primary which is used by foreign key in fact table.

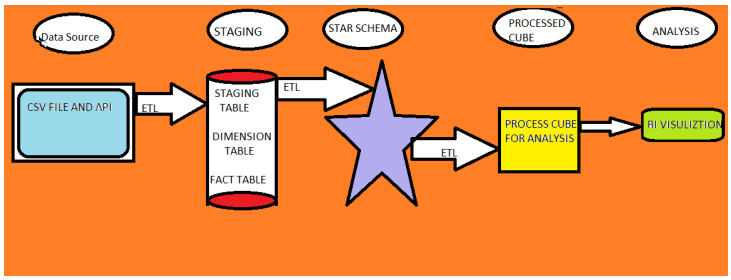


Figure 1: Data Warehouse Architecture by Ralph Kimball

Four Step of Data Modelling used in this project:

4.1 Define Business Process: - Business process used for this project is to provide efficiency in agriculture production and what impact it will have on industrial production by relationship with other factors. Finding stock price of FMCG company if going up, how production is going high, how employment is going high, according change the working style and improve in lacking area. Main motive of this analysis is to analyse agriculture data on yearly basis comparing it with gdp, export of food, employment, stock prices improve as required. Today Agriculture Sector required to be on active in terms of business relationship and satisfaction of customer as whole. So, it is required to be efficient in terms of providing quantitative as well qualitative as whole and analysing the services given on time using data available from various countries on their site. So, I tried to provide a general predictive Analysis Engine which can be used by all countries in general to analyse the fact on yearly basis.

4.2 Defining grain of the Data Warehouse: Analysis will be done on yearly basis, so it will provide us information on yearly basis and analysis can be done on granular level and take effective measure instantly.

4.3 Create Dimensions: Identify the attributes from Data Tables and create separate dimension table for each of them. Will try to provide information about what and where about business process, in this project we are using two-dimension tables (Year, Country). Year Dimension will be a Role-Playing Dimension used for many year.

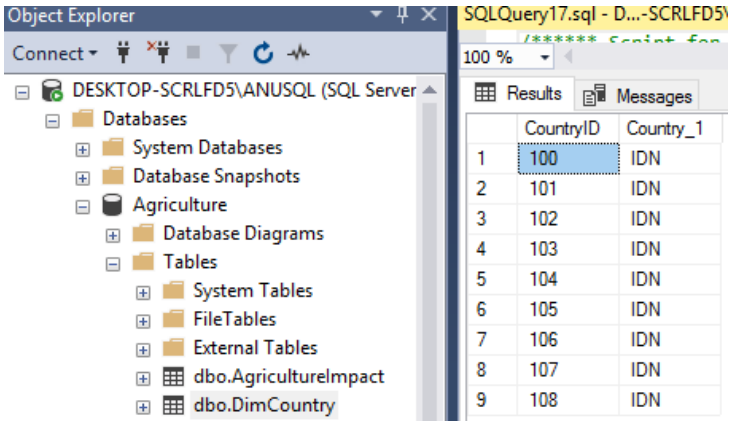


Figure 2: Figure 2: Dimension table of Country

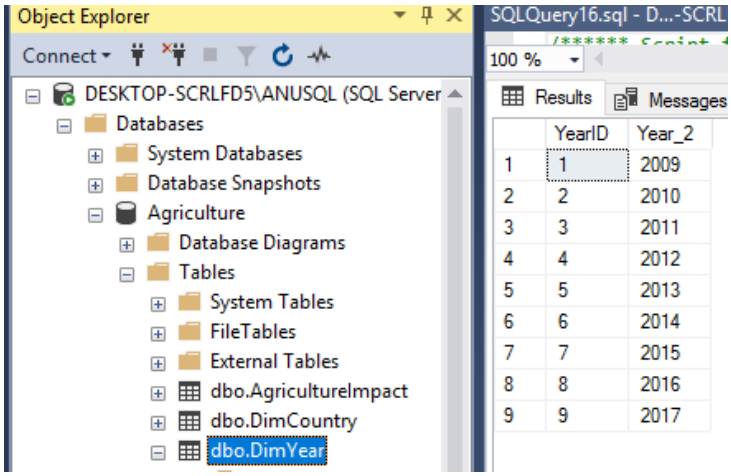
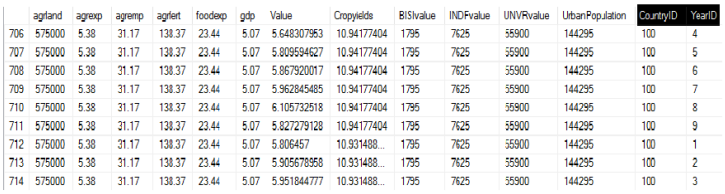


Figure 3: Figure 3: Dimension table of Year

Primary Key Used by dimension table to uniquely identify the row in the table. YearId, CountryID are primary keys for dimension tables shown in above tables.

4.4 Creating Fact table: After creating dimension create a fact table with all measures in it, these measures further be used for purpose of analysis.

Fact Table: This table store all processed granulated measurement of business process, in this project review our analysis will be on yearly basis



	agrland	agrexp	agremp	agrfert	foodexp	gdp	Value	Cropyields	BISIvalue	INDFvalue	UNVRvalue	UrbanPopulation	CountryID	YearID
706	575000	5.38	31.17	138.37	23.44	5.07	5.648307953	10.94177404	1795	7625	55900	144295	100	4
707	575000	5.38	31.17	138.37	23.44	5.07	5.809594627	10.94177404	1795	7625	55900	144295	100	5
708	575000	5.38	31.17	138.37	23.44	5.07	5.867920017	10.94177404	1795	7625	55900	144295	100	6
709	575000	5.38	31.17	138.37	23.44	5.07	5.962845485	10.94177404	1795	7625	55900	144295	100	7
710	575000	5.38	31.17	138.37	23.44	5.07	6.105732518	10.94177404	1795	7625	55900	144295	100	8
711	575000	5.38	31.17	138.37	23.44	5.07	5.827279128	10.94177404	1795	7625	55900	144295	100	9
712	575000	5.38	31.17	138.37	23.44	5.07	5.806457	10.931488...	1795	7625	55900	144295	100	1
713	575000	5.38	31.17	138.37	23.44	5.07	5.905678958	10.931488...	1795	7625	55900	144295	100	2
714	575000	5.38	31.17	138.37	23.44	5.07	5.951844777	10.931488...	1795	7625	55900	144295	100	3

Figure 4: Figure 4:Fact Table

4.5 Star Schema Design Star Schema has a single table for each dimension, each table contains all attributes for that dimension, particularly a demoralized form.

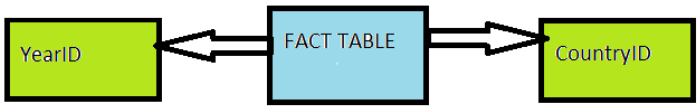


Figure 5: Raw Star Schema representation.

Star Join in Dimensional modelling is used to join both fact and Dimension tables, in start join facts are contained in fact table like agriculture land area, cultivated asset, stock value etc and information in Dimension table like Year dimension, Country dimension. Star join use both fact and dimension tables to build relationship and answer any query related to dimension and facts and up to most granular level of join.

5 Logical Data Map

Table 3: Logical Data Map describing all transformations, sources and destinations for all components of the data model illustrated in Figure 10

Source	**Column**	**Destination**	**Column**	**Transformation**
3	CountryID	DimCountry	Dimension	Converted into Integer
3	agrland	AgricultureImpact	Fact	Converted into Integer
3	agrexp	AgricultureImpact	Fact	Converted into percentage
3	agremp	AgricultureImpact	Fact	Converted into percentage
3	foodexp	AgricultureImpact	Fact	Converted into percentage
3	agrfert	AgricultureImpact	Fact	Converted into percentage
3	gdp	AgricultureImpact	Fact	Converted into percentage
2	Value	AgricultureImpact	Fact	Converted into percentage
2	CropYields	AgricultureImpact	Fact	Converted into Integer
3	YearID	DimYear	Dimension	Converted into Integer
4	BISIvalue	AgricultureImpact	Fact	Converted into currency form
4	UNVRvalue	AgricultureImpact	Fact	Converted into currency form
4	INDFvalue	AgricultureImpact	Fact	Converted into currency form
3	UrbanPopulation	AgricultureImpact	Fact	Converted in Million form

6 ETL Process

ETL is process of Extracting Data from raw sources, transforming it into measurable form and then loading it into Data Warehouse. 6.1 Extracting Data Sources which are used in Data Warehouse, Sources can be any type Structured, Semi structured, Unstructured. In this project different sources of Data being used like Structured Data (CSV Files) downloaded from Websites, some data was scraped from website and some unstructured data is extracted from stock line graph using R Code which then converted to csv. Two Types of Extractions methods used:

6.1.1 Logical: Full Extraction is extraction of data one time and no keys required in this extraction. Incremental Extraction is used when only changed data being extracted. 6.1.2 Physical: Online Extraction is done directly from source. Offline Extraction is done from Flat File, Dump File. In this project online Extraction done through R Code offline Extraction of Datasets Done using R Code. • R Code used: given in Appendices I • Data extracted is stored in SQL Server 2017 Database – Dimension Database for Dimensions and Facts, Stage Dimension is used as staging Area.

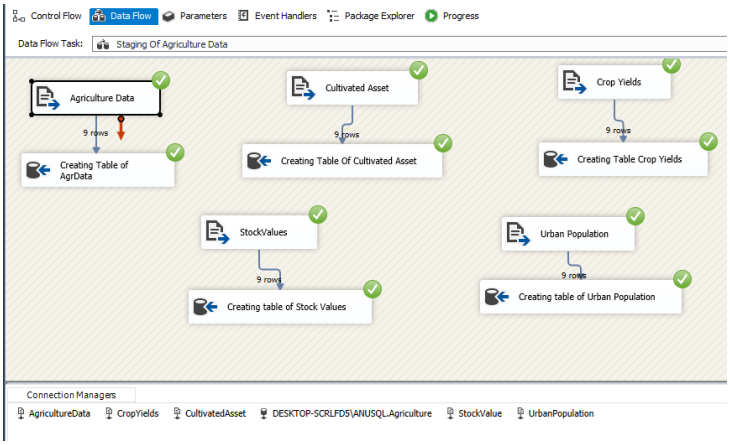


Figure 6: Screenshots of extractions used in this project.

We can see that flat file is loaded through connection manager in flat file source and then through SSIS pipeline into OLEDB Destination to create table in database of SQL server 2017 which is connected to SSIS through service analysis.

6.2 Transformation: It is one of most difficult in terms of processing time, here we do simple data conversion to complex data aggregation, merging, editing. Number of transformations techniques are present in SQL Server Integration Services, but in these project minimum transformations is used as required for the project. Some Transformation done before staging Area and some before loading Data into dimensions and fact table. Different Transformations used are: • Multistage Data transfer. • Pipelined Transfer. • Create Table using SQL. • Use of Truncate, Lookup • Multistage insert of Data.

Truncate is used by using SQL Command so that no values of table is repeated it contains all tables. It is used before staging.

6.3 Loading: Loading Data into Dimensions and Fact: • Using SQL, used SQL code to Populate YearDimension and CountryDimension (Code Taken online) • Using pipelined multistage insert of Data

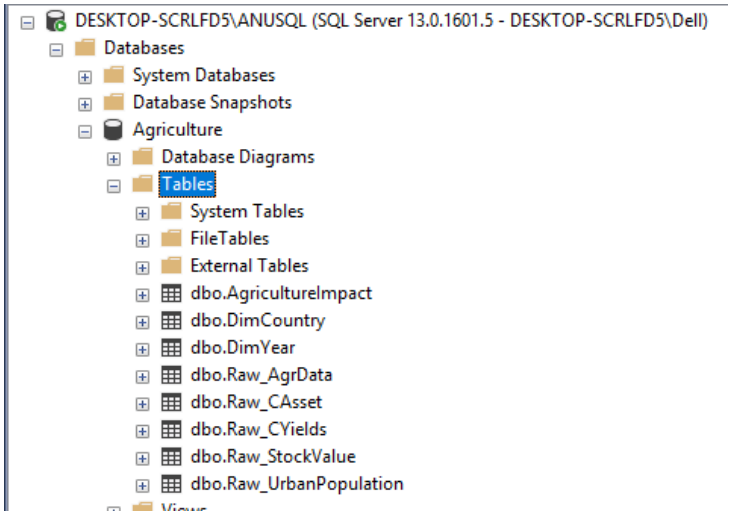


Figure 7: Screenshots of Showing Table Creation automatically after SSIS Execution

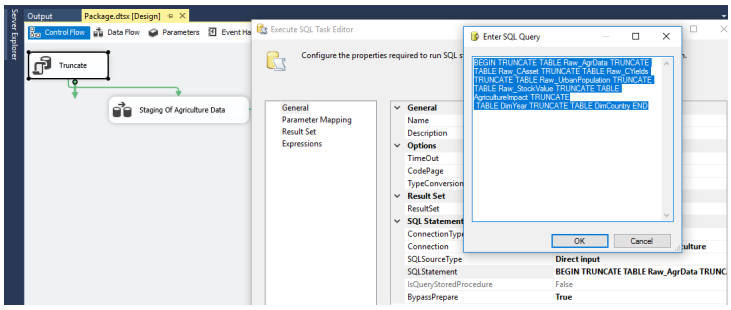


Figure 8: Screenshots Showing Truncation using SQL Command

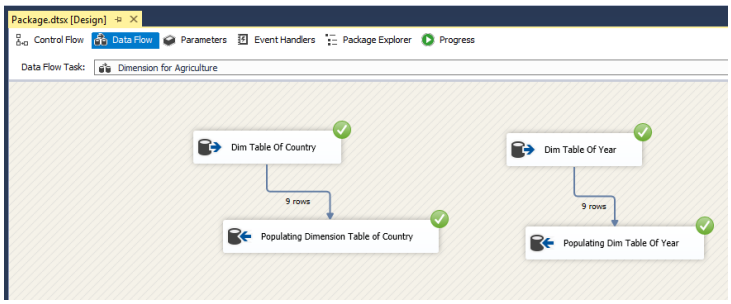


Figure 9: Screenshots Showing Creation and Population of Dimension table using SQL Command

• Using SQL, used SQL code to Populate Fact Table AgricultureImpact (Code Taken online) • Using pipelined multistage insert of Data. • Using lookup so dimension can be matched and there can be connection between them using primary key. • Foreign key is given between fact table and dimension so proper dimension can be created.

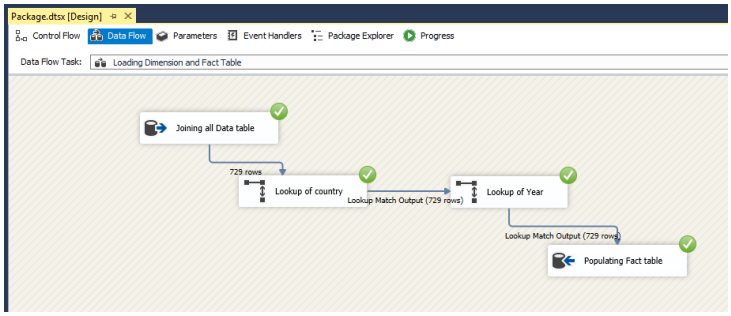


Figure 10: Screenshots Showing Creation and Population of Fact table using SQL Command and joining dimension to fact table by lookup

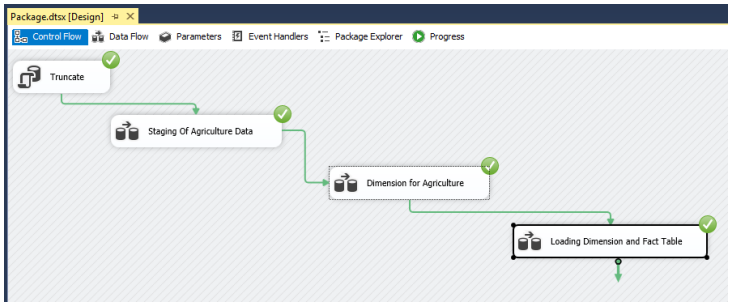


Figure 11: Screenshots Showing Creation and Population of all table and there flow from staging to Fact table

• AgricultureImpact is Fact table which is connected to dimension year and dimension country by primary key connection from dimension key to fact table

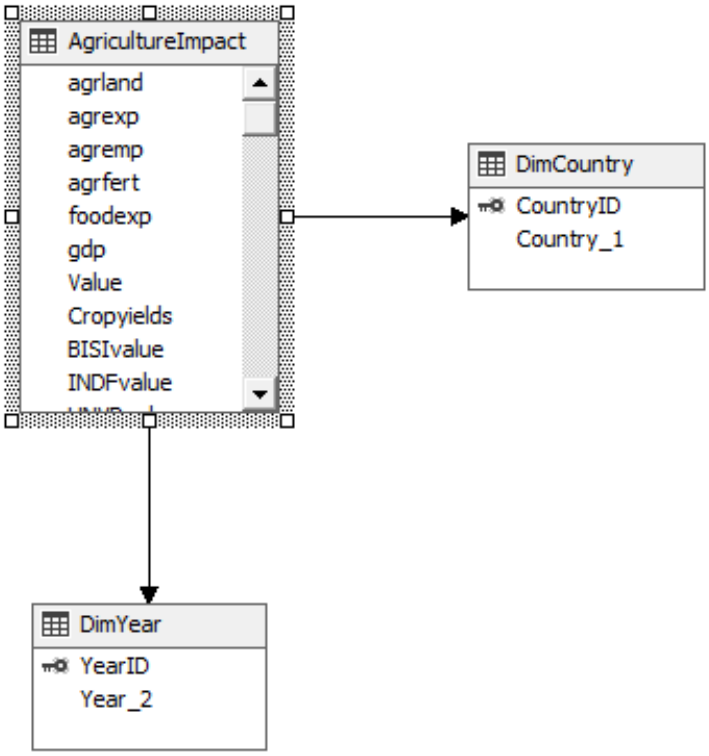


Figure 12: Star Schema Generation with primary key connected to fact table.

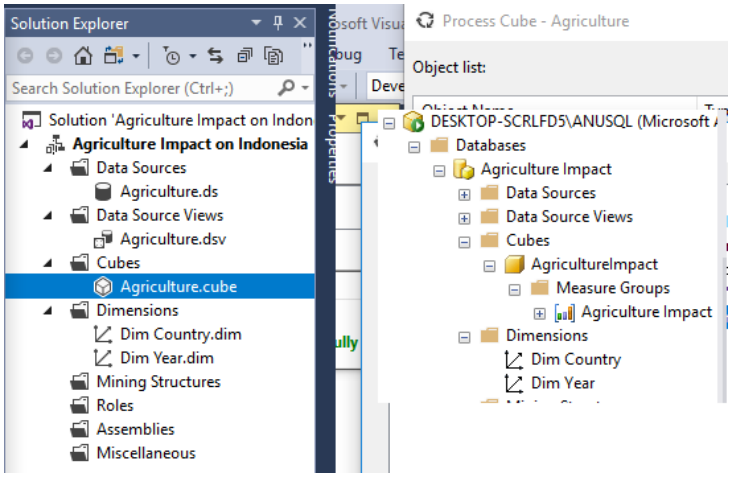


Figure 13: View of Cube in SQL Server Analysis Service and Analysis Server in SQL Server Management Studio

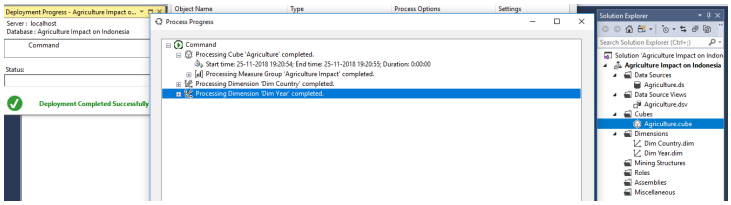


Figure 14: Successful deployment of Cube

7 Application

This project is designed to deliver proper ideas to government and FMCG companies where they should look for opportunity for their growth Cube was placed in essential column to get desired result and find answer to different Business queries, that's why in the project I'll be representing three main Business Queries which I found most impactful and provide insight in terms of choosing agriculture as base for answering queries on basis of yearly and historical data.Using Data we can a business intelligence report.

7.1 BI Query 1: Do less people employment in agriculture means more productivity?

Data source which contributed tho this business query was from OECD and Statista.Food export, crop yields, Gdp of Indonesia is not affected by people employment in agriculture so government can transform this workforce in industrial areas to gain maximum result in other areas and there might be reason that government is automating the field of agriculture production which shows less people employment in agriculture sector which in future create less dependency and continuous productivity. The general findings are that as illustrated in

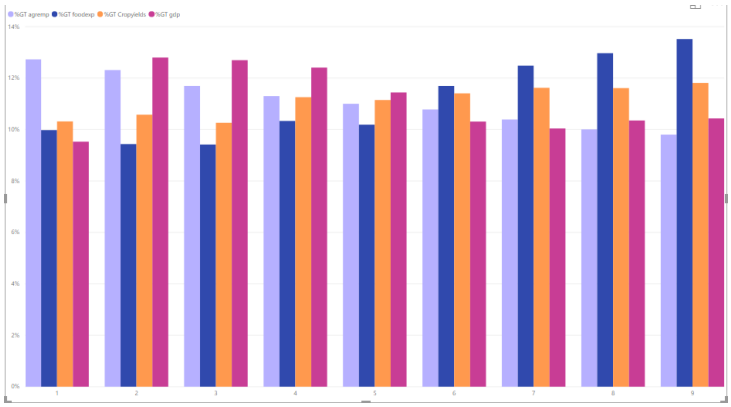


Figure 15: Results Of BI query1 from Power BI

7.2 BI Query 2:Business Query 2: Do fertilizer companies need to setup businesses in Indonesia?

Data source which contributed tho this business query was from Data world bank,OECD and Statista. Government of Indonesia is seriously considering taking agriculture as major field for its economic growth, since agriculture land usage in Indonesia has increased with time and use of fertilizer also shows that there is seriousness about productivity which can help to get more quantitative and qualitative production, since urban population is also increasing so there might be more requirement in coming years. Fertilizer Industry can benefit from it by getting great consumer base.

The general findings are that ...as illustrated in

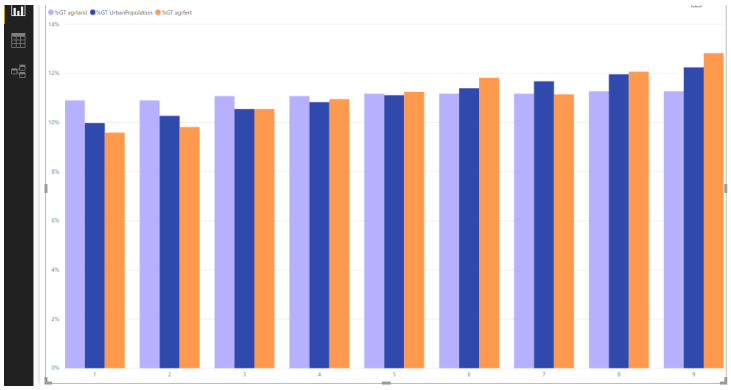
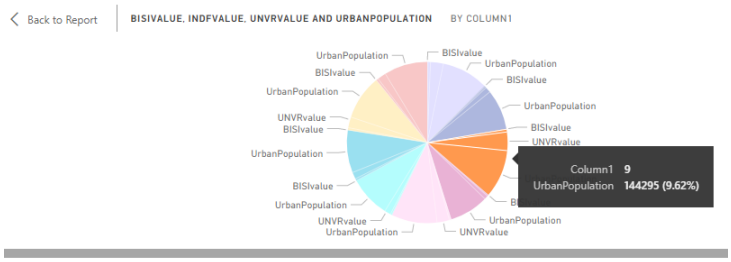


Figure 16: Results Of BI query2 from Power BI

7.3 BI Query 3: Do FMCG companies need to setup companies in Indonesia?

Data source which contributed tho this business query was from investing.com and Statista.Urban Population in Indonesia has increased so has demand for food also increased with time productivity which can help FMCG companies to get great consumer base in Indonesia ,because there has been rise in stock values of existing FMCG companies in Indonesia in recent years.

The general findings are that as illustrated in



Column1	BISIvalue	INDFvalue	UNVRvalue	UrbanPopulation
8	1900	7925	38800	140973
2	1870	4875	16500	121053
9	1795	7625	55900	144295
1	1350	3550	11050	117597
7	1350	5175	37000	137635
3	910	4600	18800	124315
4	790	5850	20850	127618
6	790	6750	32300	134288
5	560	6600	26000	130944

Figure 17: Results Of BI query3 from Power BI

8 Conclusion and Future Work

This project can be further modified in improved way by including rural areas, industrial companies producing agriculture equipment in project, but there is no data available from government, only two dimension are included in these project because complexity of agriculture can only be understand with respect to country and year. Running R Code and connecting it with SSIS to check monthly analysis of FMCG in particular city. We can also further automated process to find review analysis score for all food deliver app data for most influential cities in terms of food and using this monthly analysis Engine with comparing it with historical Data.

9 References

• Sustainable agriculture in Indonesia: Facts and challenges to keep growing in harmony with environment Syuaib, "Sustainable agriculture in Indonesia: Facts and challenges to keep growing in harmony with environment", Cigrjournal.org, 2018. [Online]. Available: http://www.cigrjournal.org/index.php/Ejounral/article/view/3747. [Accessed: 26- Nov- 2018].

- Youtube
- Stackover flow

Appendix

9.1 sql query code

```
DESKTOP-SCRLFD5\ANUSQL
SELECT        A.agrland, A.agrexp, A.agremp, A.agrfert, A.foodexp, A.gdp, A.Country_1, A.Year_1, C.[Value], C.Country_2, C.Year_2,
              D.Country_3, D.Year_3, D.Cropyields, S.Year_5, S.BISIvalue, S.INDFvalue, S.UNVRvalue,
              U.Country_4, U.Year_4, U.UrbanPopulation
FROM            Raw_AgrData AS A INNER JOIN
                         Raw_CAsset AS C ON C.Country_2 = A.Country_1 INNER JOIN
                         Raw_CYields AS D ON D.Country_3 = A.Country_1 INNER JOIN
                         Raw_StockValue AS S ON S.Year_5 = A.Year_1 INNER JOIN
                         Raw_UrbanPopulation AS U ON U.Year_4 = A.Year_1

CREATE TABLE [AgricultureImpact] (
    [agrland] int,
    [agrexp] float,
    [agremp] float,
    [agrfert] float,
    [foodexp] float,
    [gdp] float,
    [Value] float,
    [Cropyields] float,
    [BISIvalue] int,
    [INDFvalue] int,
    [UNVRvalue] int,
    [UrbanPopulation] int,
    [CountryID] int,
    [YearID] int
)
```

Figure 18: SQL QUERY CODE

sql query code

```
Dimension Table
CREATE TABLE [DimCountry] (
    [CountryID] int IDENTITY (100,1) PRIMARY KEY NOT NULL,
[Country_1] varchar(50)
)
CREATE TABLE [DimYear] (
[YearID] int IDENTITY (1,1) PRIMARY KEY NOT NULL,
    [Year_2] varchar(50)
)

TRUNCATE
BEGIN TRUNCATE TABLE Raw_AgrData TRUNCATE TABLE Raw_CAsset TRUNCATE TABLE Raw_CYields
TRUNCATE TABLE Raw_UrbanPopulation TRUNCATE TABLE Raw_StockValue TRUNCATE TABLE DimCountry TRUNCATE TABLE DimYear END
```

Figure 19: SQL QUERY CODE

getwd()

Importing Raw Agriculture Data of Indonesia agr.data<-read.csv ("C:/Users/Dell/Documents/Indo
Data.csv",header=TRUE) agr.data

Clean Columns rows of Agriculture Data Of Indonesia agr.data1<-agr.data[-(7:13),c(2,5:13)]
agr.data1

changing row to Year of agriculture Data names(agr.data1)[2:10] <-c(2009,2010,2011,2012,2013,2014 agr.data1

creating a new column of id of Agriculture dataset $\text{Agr}_id < -c(\text{"}agrland\text{"}, \text{"}agrexp\text{"}, \text{"}agremp\text{"}, \text{"}agr$

Adding that column to dataset agr.data2=cbind($\text{Agr}_id, agr.data1)agr.data2$

creating Matrix from dataframe and transposing it as.matrix(agr.data2) agr.data2.t=t(agr.data2) agr.data2.t

Transforming matrix to dataframe as.data.frame(agr.data2.t) agr.data3<-agr.data2.t agr.data3

Doing Modification in column and row by renaming it and deleting NA agrdata4<-agr.data3[-c(1:2),] agrdata4 colnames(agrdata4)[1]<-"agrland" colnames(agrdata4)[2]<-"agrexp" colnames(agrdata4)[3]<-"agremp" colnames(agrdata4)[4]<-"agrfert" colnames(agrdata4)[5]<-"foodexp" colnames(agrdata4)[6]<-"gdp" agrdata4 rownames(agrdata4)<-c(1:9) agrdata4 agr5<-agrdata4[-(10:49),] agr5

Creating previous column of country and year $\text{Year}_1 < -c(2009, 2010, 2011, 2012, 2013, 2014, 2015, 20$ $-c(\text{"}IDN\text{"})Country_1$

Adding that column to dataset agr6=cbind($\text{Country}_1, Year_1, agr5)agr6$

Importing Data from data frame to csv setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(agr6,"Agrdata.csv") write.csv(agr6)

Importing Raw Crop Yields of Indonesia cyd <-read.csv("C:/Users/Dell/Documents/Indonesia/DW yields.csv",header=TRUE) cyd

Clean Columns of crop yields Of Indonesia cyd1<-cyd[,-c(2,5,8)] cyd1

Creating Dataframe for matrix df=cyd1[(1:9),5] df df1=cyd1[(10:18),5] df1 df2=cyd1[(19:27),5] df2 df3=cyd1[(28:36),5] df3 Creating Matrix of dataframe and adding year values

as.matrix(df) as.matrix(df1) as.matrix(df2) as.matrix(df3)

Cropyields=df+df1+df2+df3 Cropyields

again transforming matrix into dataframe as.data.frame(Cropyields)

Clean renaming Columns rows of Crop Yields Of Indonesia cy2<-cyd1[-(10:36),c(1:4)] cy2 names(cy2)[1]<-c("Country$_3$")$names(cy2)[4] < -c("Year_3")$

cy2[,2]<-c("RWMR") cy2 Adding calculated dataframe column to dataset cy3=cbind(cy2,Cropyields cy3

Importing Data from data frame to csv setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(cy3,"Cyields.csv") write.csv(cy3)

Importing Raw Cultivated Asset of Indonesia ca<-read.csv("C:/Users/Dell/Documents/Indonesia/D Asset.csv",header=TRUE) ca

Clean Columns of cultivated asset Of Indonesia ca1<-ca[,-c(2,5)] ca1 names(ca1)[1]<-c("Country$_2$")$names(ca1)[4] < -c("Year_2")ca1ImportingDatafromdataframetocsvsetwd("C:/Users/Dell/Documents/Indonesia/DWBI")write.csv(ca1,"Casset.csv")write.csv(ca1)$

Importing Raw Urban Population of Indonesia up<-read.csv("C:/Users/Dell/Documents/Indonesia/ Population.csv",header=TRUE) up Clean Columns renaming of urban population Of Indonesia up1<-up[-c(1:6,16),] up1 names(up1)[1]<-c("Year$_4$")$names(up1)[2] < -c("UrbanPopulation")$ $-c(1:9)up1Country_4 < -c("IDN")Country_4Addingthatcolumntodatasetup2 = cbind(Country_4, up1)$ $/Users/Dell/Documents/Indonesia/DWBI")write.csv(up2,"urbpop.csv")write.csv(up2)$

READING BISI STOCK UNSTRUCTURED DATA FROM TEXT BISI1 <- read.delim(" /Indonesia/DWBI/BISI.txt", header=FALSE, row.names=1, quote="", stringsAsFactors=FALSE) View(BISI1)

writing it to csv format setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(BISI1,"BISI write.csv(BISI1)

reading csv file converting it to data frame B1 <-read.csv("C:/Users/Dell/Documents/Indonesia/DWBI/BISI.CSV",header = TRUE) B1 selecting stock value column B2<-B1[3] B2 BISIvalue<-B2[-(101),] BISIvalue giving it month name so data of december month can be taken into account month<-c("MAY","JUN","JUL","AUG","SEPT", "OCT","NOV","DEC","JAN","FEB","MAR","APR") month

modification in row and column data after identification of year B4=cbind(month,BISIvalue) B4 B5<-B4[-c(1:31,33:43,45:55,57:67, 69:79,81:91,93:103,105:115,117:127,129:140),] B5

adding year column to make data frame of BISI stock value Year$_5$ $< -c(2009, 2010, 2011, 2012, 2013, 2$

B6=cbind(Year$_5$, $B5)B6$

READING INDF STOCK UNSTRUCTURED DATA FROM TEXT INDF1 <- read.delim(" /Indonesia/DWBI/INDF.txt", header=FALSE, row.names=1, quote="", stringsAsFactors=FALSE) View(INDF1) writing it to csv format setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(INDF1,"INDF.csv") write.csv(INDF1)

reading csv file converting it to data frame I1<-read.csv("C:/Users/Dell/Documents/Indonesia/DWI = TRUE) I1 modifying row columns of INDF excel I2<-I1[-(1:8),] I2 I3<-I2[2] I3 INDFvalue<-I3[-101 ,] INDFvalue

giving it month name so data of december month can be taken into account month1<-c("DEC","JAN","FEB","MAR","APR","MAY","JUN","JUL","AUG","SEPT","OCT","NOV") month1

making final data frame of INDF stock I5=cbind(month1,INDFvalue) I5 I6<-I5[-c(1:72,74:84,86:96,98:108,110:120,122:132,134:144,146:156,158:168,170:180),] I6

READING UNVR STOCK UNSTRUCTURED DATA FROM TEXT UNVR1<-read.delim(" /Indonesia/DWBI/UNVR.txt", header=FALSE, row.names=1, quote="", stringsAsFactors=FALSE) View(UNVR1) /newpage writing it to csv format setwd("C:/Users/Dell/Doc write.csv(UNVR1,"UNVR.csv") write.csv(UNVR1)

reading csv file converting it to data frame U1<-read.csv("C:/Users/Dell/Documents/Indonesia/DW = TRUE) U1

modifying row columns of uNVR excel U2<-U1[-(1:8),] U2 U3<-U2[2] U3 UNVRvalue<-U3[-101 ,] UNVRvalue giving it month name so data of december month can be taken into account month2<-c("JAN","FEB","MAR","APR","MAY","JUN","JUL","AUG","SEPT","OCT","NC month2 making final data frame of UNVR stock U5=cbind(month2,UNVRvalue) U5 U6<-U5[-c(1:71,73:83,85:95,97:107,109:119,121:131,133:143,145:155,157:167,169:179),] U6 combining all stock values SV=cbind(B6,I6,U6) SV STOCKVALUE<-SV[,c(1,3,5,7)] STOCKVALUE setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(STOCKVALUE,"STC write.csv(STOCKVALUE)

R code example

getwd()

Importing Raw Agriculture Data of Indonesia agr.data<-read.csv ("C:/Users/Dell/Documents/Indo Data.csv",header=TRUE) agr.data

Clean Columns rows of Agriculture Data Of Indonesia agr.data1<-agr.data[-(7:13),c(2,5:13)] agr.data1

changing row to Year of agriculture Data names(agr.data1)[2:10] <-c(2009,2010,2011,2012,2013,2014 agr.data1

creating a new column of id of Agriculture dataset $\mathrm{Agr}_i d < -c("agrland","agrexp","agremp","agr$

Adding that column to dataset agr.data2=cbind($\mathrm{Agr}_i d, agr.data1) agr.data2$

creating Matrix from dataframe and transposing it as.matrix(agr.data2) agr.data2.t=t(agr.data2) agr.data2.t

Transforming matrix to dataframe as.data.frame(agr.data2.t) agr.data3<-agr.data2.t agr.data3

Doing Modification in column and row by renaming it and deleting NA agrdata4<-agr.data3[-c(1:2),] agrdata4 colnames(agrdata4)[1]<-"agrland" colnames(agrdata4)[2]<-"agrexp" colnames(agrdata4)[3]<-"agremp" colnames(agrdata4)[4]<-"agrfert" colnames(agrdata4)[5]<-"foodexp" colnames(agrdata4)[6]<-"gdp" agrdata4 rownames(agrdata4)<-c(1:9) agrdata4 agr5<-agrdata4[-(10:49),] agr5

Creating previous column of country and year $\text{Year}_1 < -c(2009, 2010, 2011, 2012, 2013, 2014, 2015, 20$
$-c(\text{"}IDN\text{"})Country_1$

Adding that column to dataset agr6=cbind($\text{Country}_1, Year_1, agr5)agr6$

Importing Data from data frame to csv setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(agr6,"Agrdata.csv") write.csv(agr6)

Importing Raw Crop Yields of Indonesia cyd <-read.csv("C:/Users/Dell/Documents/Indonesia/DW yields.csv",header=TRUE) cyd

Clean Columns of crop yields Of Indonesia cyd1<-cyd[,-c(2,5,8)] cyd1

Creating Dataframe for matrix df=cyd1[(1:9),5] df df1=cyd1[(10:18),5] df1 df2=cyd1[(19:27),5] df2 df3=cyd1[(28:36),5] df3 Creating Matrix of dataframe and adding year values

as.matrix(df) as.matrix(df1) as.matrix(df2) as.matrix(df3)

Cropyields=df+df1+df2+df3 Cropyields

again transforming matrix into dataframe as.data.frame(Cropyields)

Clean renaming Columns rows of Crop Yields Of Indonesia cy2<-cyd1[-(10:36),c(1:4)] cy2 names(cy2)[1]<-c("Country$_3$")$names(cy2)[4] < -c("Year_3")$

cy2[,2]<-c("RWMR") cy2 Adding calculated dataframe column to dataset cy3=cbind(cy2,Cropyields cy3

Importing Data from data frame to csv setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(cy3,"Cyields.csv") write.csv(cy3)

Importing Raw Cultivated Asset of Indonesia ca<-read.csv("C:/Users/Dell/Documents/Indonesia/D Asset.csv",header=TRUE) ca

Clean Columns of cultivated asset Of Indonesia ca1<-ca[,-c(2,5)] ca1 names(ca1)[1]<-c("Country$_2$")$names(ca1)[4] < -c("Year_2")ca1ImportingDatafromdataframetocsvsetwd("C:/Users/Dell/Documents/Indonesia/DWBI")write.csv(ca1,"Casset.csv")write.csv(ca1)$

Importing Raw Urban Population of Indonesia up<-read.csv("C:/Users/Dell/Documents/Indonesia/ Population.csv",header=TRUE) up Clean Columns renaming of urban population Of Indonesia up1<-up[-c(1:6,16),] up1 names(up1)[1]<-c("Year$_4$")$names(up1)[2] < -c("UrbanPopulation")$ $-c(1:9)up1Country_4 < -c("IDN")Country_4Addingthatcolumntodatasetup2 = cbind(Country_4, up1)$ $/Users/Dell/Documents/Indonesia/DWBI")write.csv(up2,"urbpop.csv")write.csv(up2)$

READING BISI STOCK UNSTRUCTURED DATA FROM TEXT BISI1 <- read.delim(" /Indonesia/DWBI/BISI.txt", header=FALSE, row.names=1, quote="", stringsAsFactors=FALSE) View(BISI1)

writing it to csv format setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(BISI1,"BISI write.csv(BISI1)

reading csv file converting it to data frame B1 <-read.csv("C:/Users/Dell/Documents/Indonesia/DWBI/BISI.CSV",header = TRUE) B1 selecting stock value column B2<-B1[3] B2 BISIvalue<-B2[-(101),] BISIvalue giving it month name so data of december month can be taken into account month<-c("MAY","JUN","JUL","AUG","SEPT", "OCT","NOV","DEC","JAN","FEB","MAR","APR") month

modification in row and column data after identification of year B4=cbind(month,BISIvalue) B4 B5<-B4[-c(1:31,33:43,45:55,57:67, 69:79,81:91,93:103,105:115,117:127,129:140),] B5

adding year column to make data frame of BISI stock value Year$_5$ $< -c(2009, 2010, 2011, 2012, 2013, 2$

B6=cbind(Year$_5$, $B5)B6$

READING INDF STOCK UNSTRUCTURED DATA FROM TEXT INDF1 <- read.delim(" /Indonesia/DWBI/INDF.txt", header=FALSE, row.names=1, quote="", stringsAsFactors=FALSE) View(INDF1) writing it to csv format setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(INDF1,"INDF.csv") write.csv(INDF1)

reading csv file converting it to data frame I1<-read.csv("C:/Users/Dell/Documents/Indonesia/DWI = TRUE) I1 modifying row columns of INDF excel I2<-I1[-(1:8),] I2 I3<-I2[2] I3 INDFvalue<-I3[-101 ,] INDFvalue

giving it month name so data of december month can be taken into account month1<-c("DEC","JAN","FEB","MAR","APR","MAY","JUN","JUL","AUG","SEPT","OCT","NOV") month1

making final data frame of INDF stock I5=cbind(month1,INDFvalue) I5 I6<-I5[-c(1:72,74:84,86:96,98:108,110:120,122:132,134:144,146:156,158:168,170:180),] I6

READING UNVR STOCK UNSTRUCTURED DATA FROM TEXT UNVR1<-read.delim(" /Indonesia/DWBI/UNVR.txt", header=FALSE, row.names=1, quote="", stringsAsFactors=FALSE) View(UNVR1) /newpage writing it to csv format setwd("C:/Users/Dell/Doc write.csv(UNVR1,"UNVR.csv") write.csv(UNVR1)

reading csv file converting it to data frame U1<-read.csv("C:/Users/Dell/Documents/Indonesia/DW = TRUE) U1

modifying row columns of uNVR excel U2<-U1[-(1:8),] U2 U3<-U2[2] U3 UNVRvalue<-U3[-101 ,] UNVRvalue giving it month name so data of december month can be taken into account month2<-c("JAN","FEB","MAR","APR","MAY","JUN","JUL","AUG","SEPT","OCT","NO month2 making final data frame of UNVR stock U5=cbind(month2,UNVRvalue) U5 U6<-U5[-c(1:71,73:83,85:95,97:107,109:119,121:131,133:143,145:155,157:167,169:179),] U6 combining all stock values SV=cbind(B6,I6,U6) SV STOCKVALUE<-SV[,c(1,3,5,7)] STOCKVALUE setwd("C:/Users/Dell/Documents/Indonesia/DWBI") write.csv(STOCKVALUE,"ST write.csv(STOCKVALUE)