National College of Ireland

Data Warehousing and Business Intelligence Project

on

Implementation of Data Warehouse on Properties in Ireland

Kanak Kaushik
X18136966

MSc Data Analytics – 2019/20

Submitted to: Prof. Sean Heeney

National College of Ireland
Project Submission Sheet – 2019-2020
School of Computing

Student Name:	Kanak Kaushik
Student ID:	X18136966
Programme:	MSc Data Analytics
Year:	2019/20
Module:	Data Warehousing and Business Intelligence
Lecturer:	Prof. Sean Heeney
Submission Due Date:	12/04/2019
Project Title:	Implementation of Data Warehouse on Properties in Ireland

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:	Kanak Kaushik
Date:	April 12, 2019

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

Implementation of Data Warehouse on Properties in Ireland

Kanak Kaushik
X18136966

April 12, 2019

Abstract

Buying a residential property is always a critical point in any persons life. A potential buyer will have a lot more options to choose varying from the location of the property, the value of the property, and neighborhood of the property. Especially, in such an ever-changing market of property in Dublin, Ireland. The author has tried to analyze various factors which will help a person to take the decision of buying a property. The author analyzed datasets to get to the findings which will help to decide on a property. The objective for the research was to find out different business decision which will create new values or increase the efficiency of investing in the right way of the business. The business queries answer the questions about where to purchase when to purchase, or which to purchase, which are discussed in the later part of the report.

1 Introduction

Ireland is continuing to emerge as one of the strongest economies in the EU region by the continuous rise in its GDP. Ireland is giving competition to the giants like Germany and France with respect to the GDP. Even the unemployment rate is decreasing while the jobs growth rose. Due to a strong economy, there is a rise in investment of properties across the country. Rapid job growth, heavy investments and increase in the disposable economy led to the high markets and low availability of the properties. Mccartney (2019) Dublin, the capital city of Ireland, is famous for its rich heritage, historic buildings, and emerging as one of the technical hub cities in Europe. Due to its popularity, there is a price rise in the residential property market. The average prices were increased by 5.9 in the first quarter of the year. In Dublin, the inflation rate is well kept under in comparison to the other cities. *An analysis of recent trends in the Irish residential sales market for 2019 Q1* (2019) Any person who wants to buy a residential property in Dublin generally uses the internet to look up for properties locations, prices and other details of it which results in buying a property or not. Therefore, a detailed analysis on the properties in County Dublin is carried out in this project, which will make it easier for the buyer to make their decision based on the current market scenario and to make it worth investing in the property.

Source	Type	Brief Summary
Statista	Structured	Data about house price index of 2016-2018
Property.ie	Unstructured	Data about the properties in Dublin
Daft.ie	Unstructured	Data about the properties in Dublin
CSO	Structures	Data about house price index of different location as well as types of property

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

2 Data Sources

The data sources used for building up the data warehouse on the properties in Dublin are:

2.1 Source 1: Statista

Statista dataset downloaded from: `https://www.statista.com/statistics/329735/house-price-index-in-ireland/` provides us house price index from 1st quarter of 2016 to the 2nd quarter of 2018. It is a structured data source. Details like the House price index of specific quarters are given. House price index consists of all price changes of the residential properties purchased by the owners, both new and old properties.

2.2 Source 2: Property.ie

Properties of all the location in County Dublin are hosted on Property.ie website. Link for this source: `https://www.property.ie/property-for-sale/dublin/` There are parameters like the property name, property price, area of the property is given on the website. This information is scraped through R. Specific data and cleaning of the data is also done by R. This information is used to answer business queries.

2.3 Source 3: Daft.ie

Properties of all the area in County Dublin are facilitated on Daft.ie site. Link for this source: `https://www.daft.ie/dublin/property-for-sale/` There are parameters like the property name, property value, territory of the property is given on the site. This data is scratched through R. Explicit information and cleaning of the information is additionally done by R. This data is utilized to answer business inquiries.

2.4 Source 4: CSO

Residential property index of different types of houses in different areas are taken from Central Statistics Office. Link for this source: `https://www.cso.ie/px/pxeirestat/Statire/SelectVarVal/Define.asp?maintable=HPM06&PLanguage=0` There are columns of different types of location as well as types of houses which contains the house price index values in different rows containing quarters of years.

3 Related Work

There are several techniques which are used to calculate the impact of a number of factors on the pricing of the houses. One of the most commonly used is Hedonic regression. This technique is used to make house price index of a certain property. Hedonic factors, like the location of the property, type of the property, and the construction plays a far more important role in deciding its price. This technique is proved to be better than any other techniques. The demand for the residential property is on the grow because of growth in the disposable income and rising of the employment rate which results in a shift in the demand curve. Due to this, there is an escalation of prices in the market. Berry et al. (2003) Prices also change with the common cycle in the housing market and the short-lived presence of the price bubbles. The prices distribution fluctuates between different period. It also depends on the hedonic indexes, which affects it. There will always be a difference in the listing price and the selling price, which could make a relationship between them. Houses are multidimensional items depending on both the consumption and the investment in the market. Due to this, the house price index varies according to the market and various factors. McGreal et al. (2010) Consumer behavior also subjects into buying a residential property. Investigate the characteristics that impact the estimation of private property from the purchasers' point of view, look at the manner in which that subjective variables sway on the estimation of private property from the purchasers' perspective, and explore the real valuation strategies utilized by valuers to decide regardless of whether they represent purchaser conduct precisely. Thus, it does not tell every one of the variables that impact the estimation of property, however, qualitative factors do impact estimation of residential property from the customers perspective. Daly et al. (2003) Recovery concerns the physical and financial restoration of areas with advancement and interest in property a principal part of the procedure and item. Be that as it may, the property is impacted by demand- supply side connections that mirror the energy of the market. McGreal et al. (2004) Understanding the advancement of residential property is a subject of impressive premium both from a hypothetical and connected point of view. Housing is a vital piece of the development speculation, GDP, and the wealth of a nation. A key component disregarded in hedonic regression is the way that the areas and time of abiding exchanges are random. That is, the selling price is demonstrated restrictively on area and time; randomness is presented just in the regression given highlights noted above at the area and time. Understanding the example of sales movement over space and time is an alternate however a conceivably valuable issue. A similarity can be drawn with natural demonstrating of species. Paci et al. (2017)

The model discussed was based on the space-time point patterns of the transactions of the properties. Poisson process and log-Gaussian Cox process were used to carry out the research model. It shows the hotspots of the sale of the properties using the mentioned methods. The author used the location and the price of the properties to carry out the project. From this project, the author expects to contribute a new way to get information about the property with using the mentioned method. It will tell the trendy area of the offered residential property and how the house price index will affect the price in the near future.

4 Data Model

Kimball and Inmon are the two prevalent data warehouse structures. The design utilized relies upon the necessity of the data warehousing framework. Inmon is a top-bottom methodology that implies there is a single point data warehouse distribution center which contains all the data of the association at one spot. After that distinctive data dimension can be added to query from central warehouse. While Kimball deciphers this idea in all respects in an unexpected way, as per Kimball, based upon the business inquiries, a data warehouse is built. That implies in Kimball's methodology right off the bat the dimension and the fact tables are built according to the business inquiries and choose a schema. Mor (2014) There are two schemata: Star schema and Snowflake schema. As per the name suggest, Star schema contains a single fact table connected with multiple dimension tables. All the primary keys of the dimension tables are stored as foreign keys in the fact table. And as for the Snowflake schema, it contains single fact connected with multiple dimension tables, even the dimension table are then branched out to the different dimension table. Kimball & Ross (2013)

In this project, the author has used star schema which is constructed from the fact and dimension tables for the business queries, which is then used to populate the warehouse. The author has used the Kimball approach of building the data warehouse from bottom up.

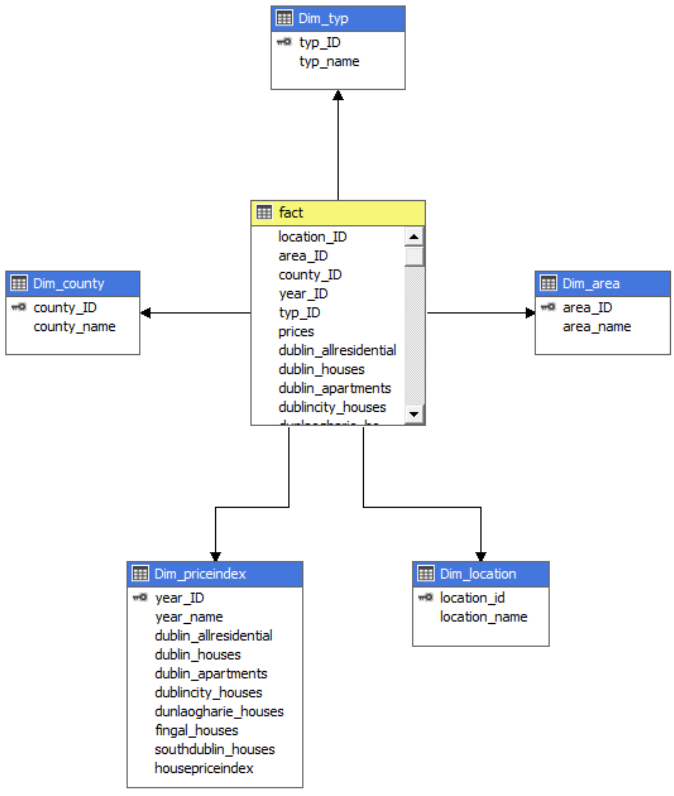

Figure 1: Star schema

This data warehouse project contains one fact table and five dimensions table.(Fig.1) The fact table contains all the foreign keys of the dimension table and the measures which will be required to get the solution of the business queries. The four-dimension

table contains all the information about the properties. The one-dimension table contains the house price index of the properties with respect to year. All these dimension tables are connected to fact table by having relationships between them. All primary keys of the dimension tables are kept as a foreign key in the fact table. To get the results of the business queries, measures from the fact table are taken to answer them.

Fact Table:

Name: fact Fields: `location_ID`,`area_ID`, `county_ID`, `year_ID`, `typ_ID`, `prices`, `dublin_allresident`
`dublin_houses`, `dublin_apartments`, `dublincity_houses`, `dunlaogharie_houses`,`fingal_house`
`southdublin_houses`, housepriceindex

Dimension Table:

Dimension table 1:
Name: `Dim_location`
Fields: `location_ID, location_name`
Dimension table 2:
Name: `Dim_area`
Fields: `area_ID, area_name`

Dimension table 3:
Name: `Dim_county`
Fields: `county_ID, county_name`

Dimension table 4:
Name: `Dim_typ`
Fields: `typ_ID, typ_name`

Dimension table 5:
Name: `Dim_priceindex`
Fields: `year_ID, year_name, dublin_allresidential, dublin_houses, dublin_apartments, dublincity_houses, dunlaogharie_houses, fingal_houses, southdublin_houses, housepriceindex`

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 1

Source	Column	Destination	Column	Type	Transformation
2	allloc1	Raw_daft	allloc1	Dimension	All location were removed from symbol ','
2	typ	Raw_daft	typ	Dimension	Separated location and then looped to get the type of the property
2	county	Raw_daft	county	Dimension	Separated location and then looped to get the county of the property
2	area	Raw_daft	area	Dimension	Separated and then looped to get the area
2	apr	fact	prices	Fact	Removed the euro symbol and then converted to numeric by functions
3	allloc1	Raw_property	allloc1	Dimension	All location were removed from symbol ','
3	typ	Raw_property	typ	Dimension	Separated location and then looped to get the type of the property
3	county	Raw_property	county	Dimension	Separated location and then looped to get the county of the property
3	area	Raw_property	area	Dimension	Separated and then looped to get the area
3	apr	fact	prices	Fact	Removed the euro symbol and then converted to numeric by functions
1	year_name	Raw_csostats	year_name	Dimension	Merged two files by the same column $
4	year_name	Raw_csostats	year_name	Dimension	Merged two files by the same column $
4	dublin_allre	fact	dublin_allre	Fact	Tranformed into float
4	dublin_house	fact	dublin_house	Fact	Tranformed into float
4	dublin_apart	fact	dublin_apart	Fact	Tranformed into float
4	dublincity_h	fact	dublincity_	Fact	Tranformed into float

Continued on next page

Table 3 – *Continued from previous page*

Source	Column	Destination	Column	Type	Transformation
4	dunlaogharie	fact	dunlaogharie	Fact	Tranformed into float
4	southdublin_	fact	southdublin_	Fact	Tranformed into float
1	housepricein	fact	housepricein	Fact	Tranformed into float
2	location_ID	Dim_location	location_ID	Dimension	Transformed into int
3	location_ID	Dim_location	location_ID	Dimension	Transformed into int
2	location_name	Dim_location	location_name	Dimension	Transformed into NVARCHAR
3	location_name	Dim_location	location_name	Dimension	Transformed into NVARCHAR
2	typ_ID	Dim_typ	typ_ID	Dimension	Transformed into int
3	typ_ID	Dim_typ	typ_ID	Dimension	Transformed into int
2	typ_name	Dim_typ	typ_name	Dimension	Transformed into NVARCHAR
3	typ_name	Dim_typ	typ_name	Dimension	Transformed into NVARCHAR
2	county_ID	Dim_county	county_ID	Dimension	Transformed into int
3	county_ID	Dim_county	county_ID	Dimension	Transformed into int
2	county_name	Dim_county	county_name	Dimension	Transformed into NVARCHAR
3	county_name	Dim_county	county_name	Dimension	Transformed into NVARCHAR
2	area_ID	Dim_area	area_ID	Dimension	Transformed into int
3	area_ID	Dim_area	area_ID	Dimension	Transformed into int
2	area_name	Dim_area	area_name	Dimension	Transformed into NVARCHAR
3	area_name	Dim_area	area_name	Dimension	Transformed into NVARCHAR

*Some of the column names are incomplete because there was not enough space to write.

6 ETL Process

A Database is made in the SSMS, named test which contains all the Raw tables, Execute process tasks, Dimension and fact tables of this warehouse. A project is created in Visual Studio 2017 by using the Integration Services template. In this project, the database was created under the name of the test and was connected to the MOLAP server. The ETL (Extract, Transform and Load) is the base of any given data warehouse. A well-designed ETL system extracts the data from the source, make sure that the data is consistent, and its quality is maintained, then do transformation as required, and finally delivers the data in a presentable format which will be used by developers to make decisions.Kimball & Ross (2013) The project flow is shown in SSIS (Fig. 2).

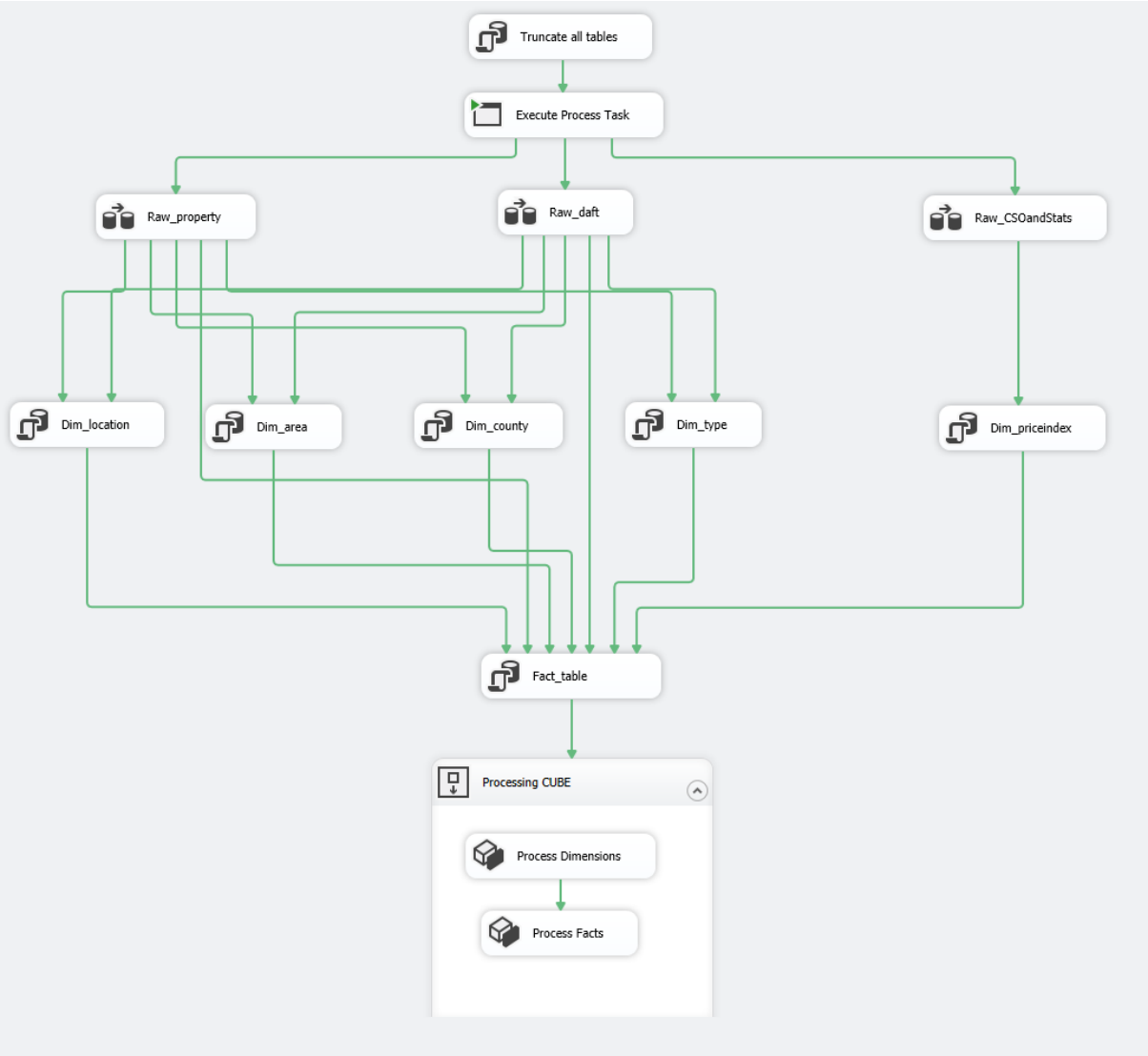

Figure 2: SSIS Workflow

The first thing that happens is there is a SQL statement written to truncate all the tables in the SSMS by executing SQL script. This is done to make the ETL process Re-runnable.

Then there is Execute process task, which contains the R script to scrape the data, combine the data from our structured and unstructured source. It also cleans the data

and transforms it into the required format.

Then comes the staging area, in which the there is an injection of flat file into OLE DB server which contains our database. It forms our Raw tables from which we will be creating our Dimension and Fact tables. It populates all the Raw tables by correctly mapping the columns. In this step, there are 3 Raw tables created.

After the execution of the Raw tables, the Dimension tables are created by it. This is done by executing SQL scripts which then load the data from the raw tables to the dimension tables.

Then Fact table is created by the help of executing SQL script in order to load it from the raw tables and the dimension table. All the dimension tables are connected to the fact table by having primary keys as foreign keys in the fact table.

In the end, there is cube deployment by the help task sequencer which contains the cube deployment functions. As you see from the diagram, Dimensions are processed first then the Facts are processed. Cube deployment is automated in this project. It is connected to the SSMS server, which helps to get all the dimension tables and the fact table.

Cleaning of Unstructured data

The scraped data from the sources are cleaned and then transform it to the data warehouse. The process of cleaning of data is explained below: While scraping the property location from the websites, the format of the location was a single string containing the location. But in order to get the area and the county of the property, it was separated by the comma and then stored into different vectors. Then it was looped to get the area and the county of the property. This was done in R and then cleaned and then loaded to the SSMS. Price also was scraped from the website, but it was coming in a string format. It should be in numeric format to calculate on it. But there was Euro symbol in front of the price, and it was to be removed, so to remove it the Encoding of the prices were changed as it was not recognizing the symbol. Then the pattern was matched and removed from the column. There were also many text values which was then substituted. Then it was transformed to numeric field. All codes is shown in the appendix.

7 Application

The cube which was created in SSAS is deployed in Tableau to analyze and visualize the data.

7.1 BI Query 1: Did Brexit had any affect on the property market ?

For this query, the contributing sources of data are: Statista and CSO This case study tries to find correlation between the Brexit and the prices of the property.This is analyzed in Tableau using multiple bar graphs.

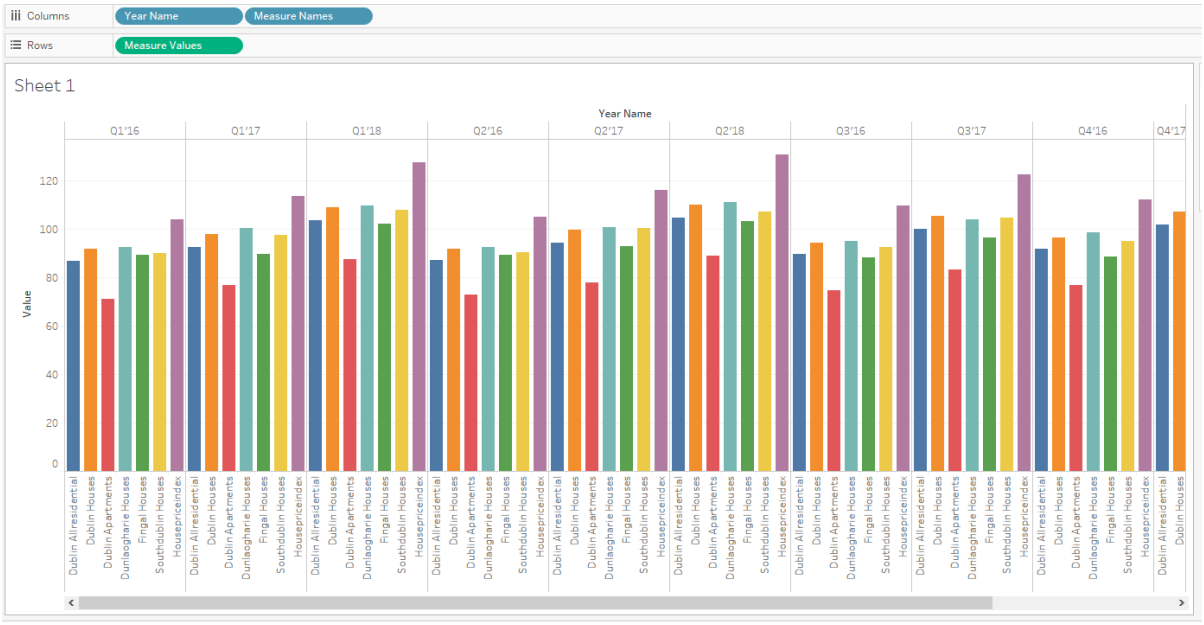

Figure 3: Results for BI Query 1

Analysis: From the bar graph, It is clear there is a significant rise in the house price index after the announcement of Brexit in Q2'2016 from the comparison with Q2'2018. There is rise in house price index of houses that are located near Dun Laogharie houses and South Dublin houses. For a new buyer, it is looking good for investment in property as from the result the house price index will get higher as near the brexit date gets. Figure 3.

7.2 BI Query 2: Expensive properties offered in area

For this query, the contributing sources of data are: Property.ie and Daft.ie This case study tries to analyze among all the areas which is most expensive to buy.

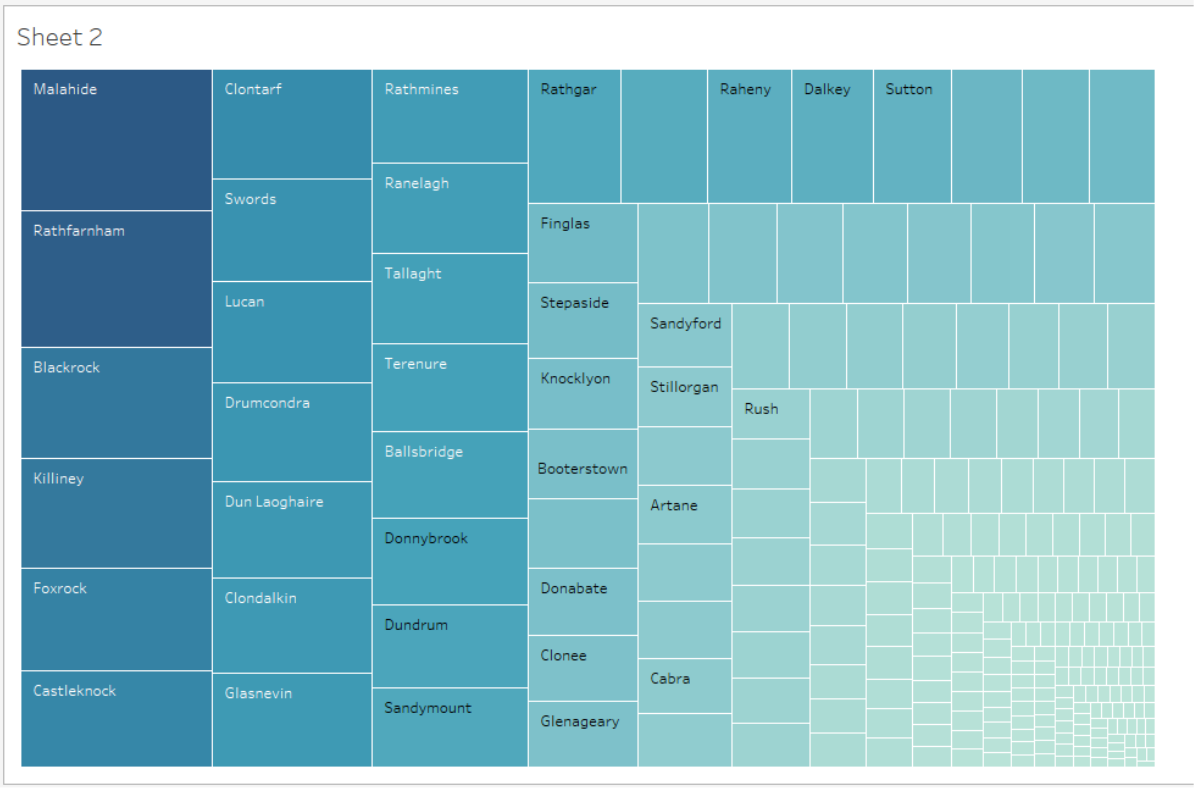

Figure 4: Results for BI Query 2

Analysis: From this, one can study that Malahide area is the most popular area and expensive area to buy a residence property. Then comes the Rathfarnham, which is on the second position. As the price will go higher in these areas, Its a nice investment. Also Blackrock and Killiney aren't that far away. Figure 4.

7.3 BI Query 3: What types of properties are preferred ?

For this query, the contributing sources of data are: Property.ie and Daft.ie This case study tells that what type of property are in the market right now and which area has the most popular type of property.

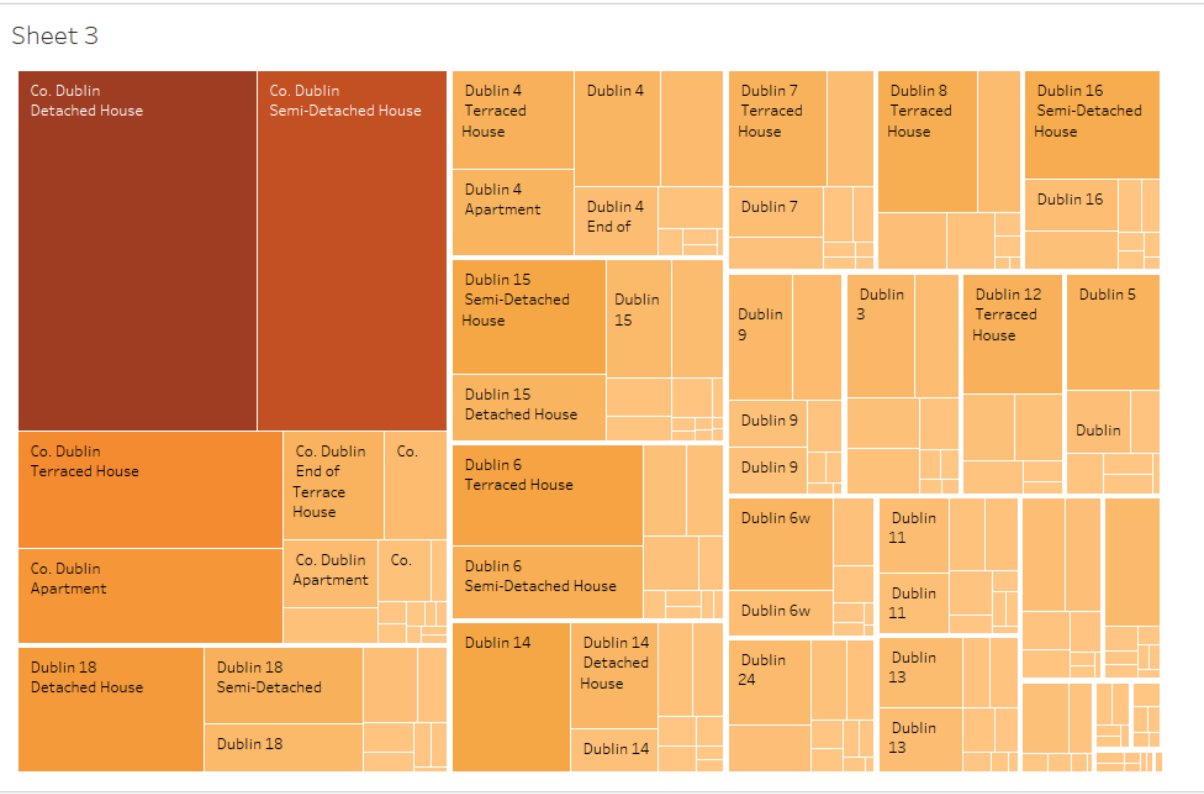

Figure 5: Results for BI Query 3

Analysis: From this, one can study that Detached and Semi-Detached houses are the most popular type of property which are currently in the market. After that Terraced house and End of terraced houses are preferred by the people. An investor can build the type of houses that people prefers more than the other. Figure 5.

8 Conclusion and Future Work

The data warehouse is successfully deployed by the bottom-up approach.Kimball & Ross (2013) The data warehouse seems to have significant factor in the estimating the value of the properties based on the area and the type of the property. It can answer many BI queries regarding the property. This used a location-price system which gave the results to have an understanding of the current property market.As against the use of space-time pattern for the estimation of the market.Paci et al. (2017) Therefore, the benefit of this analysis is to provide insight into the property market, which will then help make the decision better while buying a residence property.

Video link for this project : `https://youtu.be/XHOqUiWsl8c`

As for the future work, this model can be extended to use the data about the entire country and continent. Energy rating of the property can also be included to see the efficiency of the certain properties.

References

An analysis of recent trends in the Irish residential sales market for 2019 Q1 (2019), Technical report, Daft.ie.
URL: *https://www.daft.ie/report/2019-Q1-houseprice-daftreport.pdf*

Berry, J., McGreal, S., Stevenson, S., Young, J. & Webb, J. R. (2003), 'Estimation of Apartment Submarkets in Dublin, Ireland'.
URL: *https://www.jstor.org/stable/24887656*

Daly, J., Gronow, S., Jenkins, D. & Plimmer, F. (2003), 'Consumer behaviour in the valuation of residential property', *Property Management* **21**(5), 295–314.
URL: *https://www.emeraldinsight.com/doi/10.1108/02637470310508653*

Kimball, R. & Ross, M. (2013), *The Data Warehouse Toolkit*, John Wiley & Sons, Inc., Indianapolis, Indiana.
URL: *http://www.essai.rnu.tn/Ebook/Informatique/The Data Warehouse Toolkit, 3rd Edition.pdf*

Mccartney, J. (2019), Savills World Research-Ireland Investment, Technical report.
URL: *https://www.ntma.ie/business-areas/funding-and-debt-management/investor-relations/credit-ratings*

McGreal, S., Berry, J., McParland, C. & Turner, B. (2004), 'Urban regeneration, property performance and office markets in Dublin', *Journal of Property Investment & Finance* **22**(2), 162–172.
URL: *http://www.emeraldinsight.com/doi/10.1108/14635780410536160*

McGreal, S., Brown, L. & Adair, A. (2010), 'List price and sale price variation across the housing market cycle', *International Journal of Housing Markets and Analysis* **3**(2), 89–99.
URL: *http://www.emeraldinsight.com/doi/10.1108/17538271011049731*

Mor, Y. (2014), 'Inmon vs. Kimball - The Big Data Warehouse Duel · Xplenty'.
URL: *https://www.xplenty.com/blog/inmon-vs-kimball-the-big-data-warehouse-duel/*

Paci, L., Asunción Beamonte, M., Gelfand, A. E., Gargallo, P. & Salvador, M. (2017), 'Analysis of residential property sales using space-time point patterns', *Spatial Statistics* **21**, 149–165.
URL: *http://dx.doi.org/10.1016/j.spasta.2017.06.0072211-6753/*

9 Appendix

R Code

```
library(tidyverse)
library(rvest)
library(tidyr)

allloc <- NULL
apr<- NULL
typ<- NULL

urls <- 'https://www.property.ie/property-for-sale/dublin/'
urll <- read_html(urls)
loc <- urll %>% html_nodes(xpath = '//div[contains(@class,"sresult_address")
]//h2//a') %>% html_text()
pr <- urll %>% html_nodes(xpath = '//div[contains(@class,"sresult_descriptio
n")]//h3') %>% html_text()
ty <- urll %>% html_nodes(xpath = '//div[contains(@class,"sresult_descriptio
n")]//h4') %>% html_text()

allloc <- append(allloc,loc)
apr <- append(apr,pr)
typ <- append(typ,ty)
urls <- 'https://www.property.ie/property-for-sale/dublin/p_'
for (i in 2: 100) {
  newurl<- paste0(urls,i)
  cat(newurl)
  linkss<- read_html(newurl)
  loc <- linkss %>% html_nodes(xpath = '//div[contains(@class,"sresult_addre
␣␣ss")]//h2//a') %>% html_text()
  pr <- linkss %>% html_nodes(xpath = '//div[contains(@class,"sresult_descri
␣␣ption")]//h3') %>% html_text()
  ty <- linkss %>% html_nodes(xpath = '//div[contains(@class,"sresult_descri
␣␣ption")]//h4') %>% html_text()
  allloc <- append(allloc,loc)
  apr <- append(apr,pr)
  typ <- append(typ,ty)
  cat(i)
}

allloc1 <- gsub('\n|,','',allloc)
typ <- gsub('\n|For␣Sale','',typ)
ar <- strsplit(allloc,",")
tya <- strsplit(typ,",")
county<- NULL
area<- NULL
types <- NULL
for (i in 1: (length(ar))) {
  count <- length(ar[[i]])
  county <- append(county,ar[[i]][count])
```

```
  area <- append(area,ar[[i]][count-1])
}

for (i in 1: (length(tya))) {
  count <- length(tya[[i]])
  types <- append(types,tya[[i]][count])
}

Encoding(apr)<- "UTF-8-BOM"
apr <- gsub("   |       ","",apr)
types <- gsub('For␣Sale|\n','',types)
df <- data.frame(allloc1,types,apr,area,county)
df$apr <- gsub(",","",df$apr)
df$apr <- gsub("\n|Region|From|Asking␣Price|Excess|Reserve␣not␣to␣exceed
|AMV|not␣to␣exceed|Reserve|.Auction.|.Tender.|(␣)","",df$apr)
df$apr <- gsub("Price␣on␣Application|PriceonApplication|(␣)","350000"
,df$apr)
df$apr <- as.numeric(df$apr)
df$types <- gsub('for␣sale|\n','',df$types)
df$types <- gsub('Site␣For␣Sale','Apartment',df$types)
df$apr <- format(df$apr, scientific=FALSE)
write.csv(df,file = 'b.csv',row.names=FALSE)

library(tidyverse)
library(rvest)
library(tidyr)

allloc <- NULL
apr<- NULL
typ<- NULL

urls <- 'https://www.daft.ie/dublin/property-for-sale'
urll <- read_html(urls)
loc <- urll %>% html_nodes(xpath = '//a[contains(@class,"PropertyInformation
CommonStyles__addressCopy--link")]') %>% html_text()
pr <- urll %>% html_nodes(xpath = '//a[contains(@class,"PropertyInformation
CommonStyles__propertyPrice--link")]//strong') %>% html_text()
ty <- urll %>% html_nodes(xpath = '//div[contains(@class,"QuickPropertyDeta
ils__propertyType")]') %>% html_text()
allloc <- append(allloc,loc)
apr <- append(apr,pr)
typ <- append(typ,ty)
a<-20
xa <- 'https://www.daft.ie/dublin/property-for-sale/?offset='
for (i in 1: 20) {
  newurl<- paste0(xa,a)
  cat(newurl)
  a<-a+20
  linkss<- read_html(newurl)
  loc <- linkss %>% html_nodes(xpath = '//a[contains(@class,"PropertyInfor
␣␣mationCommonStyles__addressCopy--link")]') %>% html_text()
```

```
  pr <- linkss %>% html_nodes(xpath = '//a[contains(@class,"PropertyInforma
␣␣tionCommonStyles__propertyPrice--link")]//strong') %>% html_text()
  ty <- linkss %>% html_nodes(xpath = '//div[contains(@class,"QuickProperty
␣␣Details__propertyType")]') %>% html_text()
  allloc <- append(allloc,loc)
  apr <- append(apr,pr)
  typ <- append(typ,ty)
  cat(i)
}
allloc1 <- gsub('\n|,','',allloc)
ar <- strsplit(allloc,",")
finalarea<- NULL
county<- NULL
area<- NULL
for (i in 1: (length(ar))) {
  count <- length(ar[[i]])
  county <- append(county,ar[[i]][count])
  area <- append(area,ar[[i]][count-1])

}
Encoding(apr)<- "UTF-8-BOM"
apr <- gsub("   |       ","",apr)
apr <- gsub("Price␣On␣Application","350000",apr)
apr <- gsub("AMV:␣","",apr)
apr <- gsub(",","",apr)
apr <- as.numeric(apr)
typ <- gsub('for␣sale|\n','',typ)
typ <- gsub('Site␣For␣Sale','Apartment',typ)
apr <- format(apr, scientific=FALSE)
df1 <- data.frame(allloc1,typ,apr,area,county)
write.csv(df1,file ='a.csv',row.names=FALSE )
r <- read.csv("cso.csv")
r1 <- read.csv("stats.csv")
a <- merge(r,r1,by='year')
write.csv(df,file = 'c.csv',row.names=FALSE)
```