Sri Lankan Institute of Information Technology



Data Warehousing and Business Intelligence (IT3021)

1.6 million UK traffic accidents Assignment 2

Submitted to
Sri Lanka Institute of Information Technology

Bachelor of Science Special Honors Degree in Data Science

IT19176666 Aththanayake A.B.P.S 25.06.2021

STEP 01

Data Source (Data Warehouse Introduction)

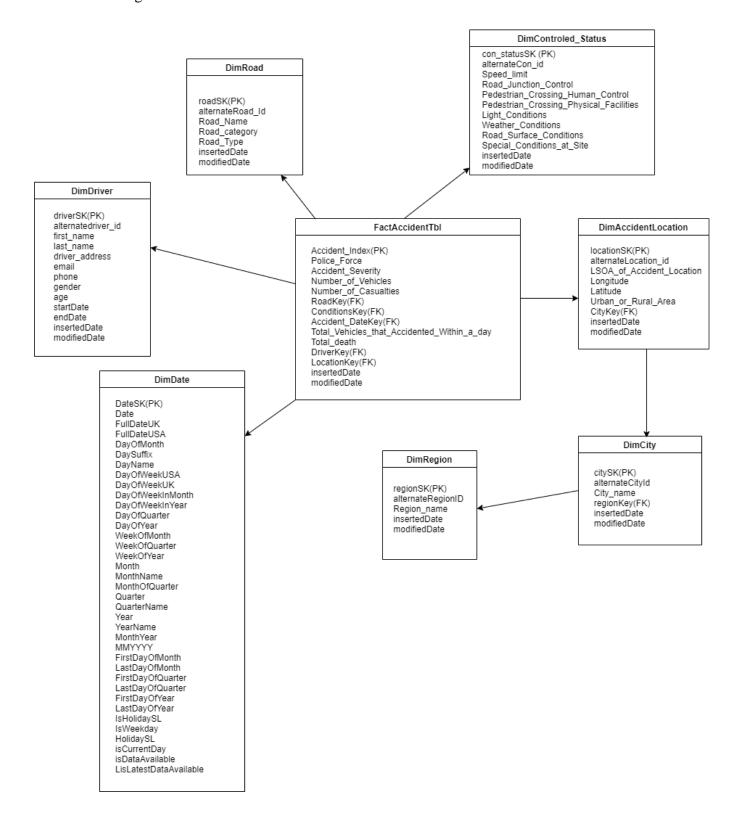
In here used data source is the 'Accident_Warehouse' which is the database which I have created as assignment one. I have chosen **1.6 million UK traffic accidents** data set of UK government amassed traffic data from 2000-2016. But here I have inly used data from 2012 – 2014, because of complexity. As well as I have done some modifications accordingly to the data set derived from source for assignment one.

Therefore, I have divided my data set into seven tables and eight-dimension tables (Include fact table) using my scenario. As I have mentioned before my tables was, FactAccidentTbl, DimRoad, DimControlled_Status, DimDriver, DimAccidentLocation, DimCity, DimRegion.

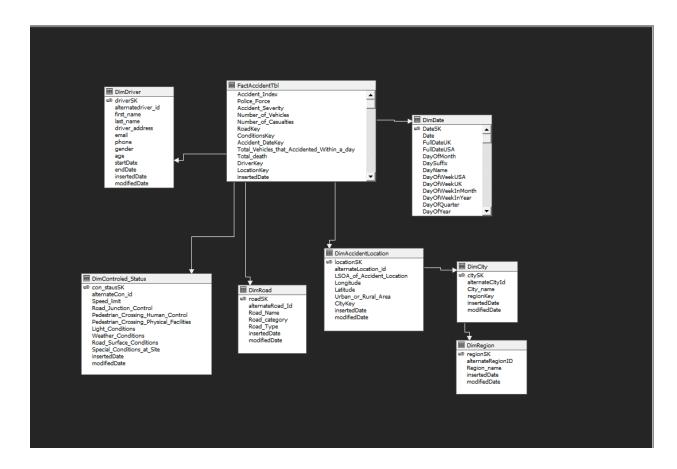
I. DATA SOURCE: Accident_Warehouse database

☐ Accident_Warehouse
☐ Database Diagrams
☐ Tables
☐ System Tables
☐ FileTables
☐ External Tables
☐ dbo.DimAccidentLocation
☐ dbo.DimCity
☐ dbo.DimControled_Status
☐ dbo.DimDate
☐ dbo.DimDriver
☐ dbo.DimRegion
☐ dbo.DimRoad
☐ dbo.FactAccidentTbl

II. ER -Diagram for Data warehouse



III. Snowflake Schema used



STEP 02

Cube Implementation

OLAP cube is a method for storing data in a multi-dimensional form. This multi-dimensional cube is used for analytical purpose. In this process I have used SQL Server Data Tools. Dimensions and measures are main constituents of a cube. In addition to two main facts are,

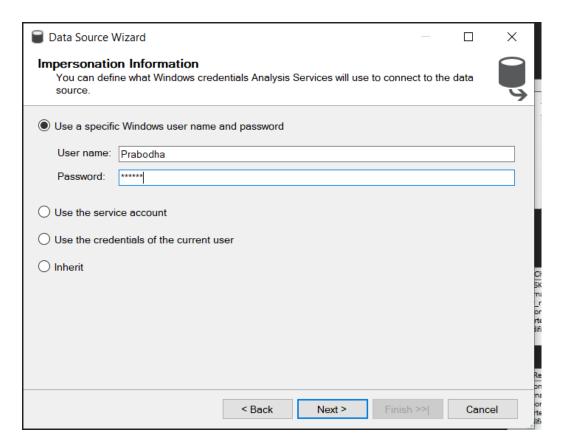
- Measures: Provide aggregated numerical values of interest to the end user.
- Dimensions: Define the structure of the cube that you use to slice and dice over.

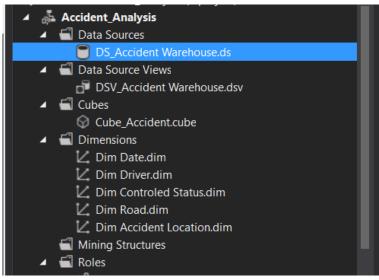
2.1

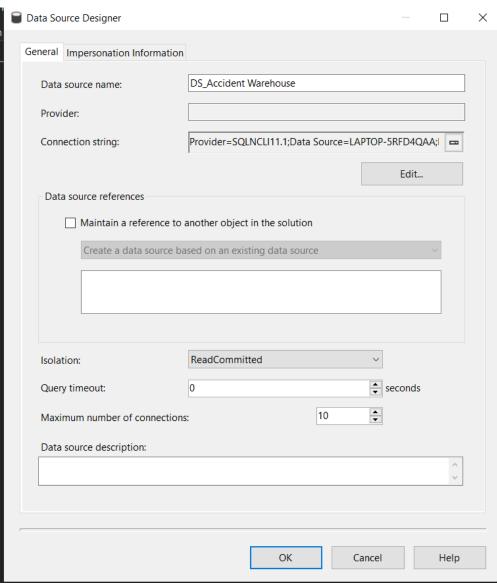
STEP 01: Data Source Creation

When creating data source, it was connected to the 'Accident_Warehouse' database via SQL Server Management Studio using windows username and password for connecting to the SQL Server Management Studio.

I. Data Source creation

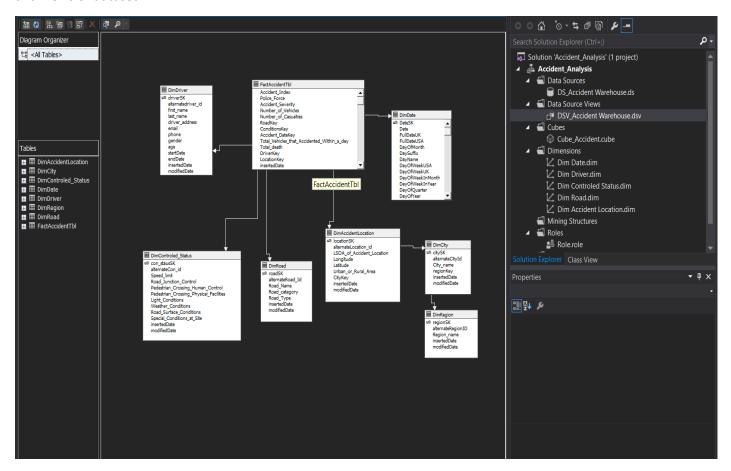






STEP 02: Data Source Views Creation

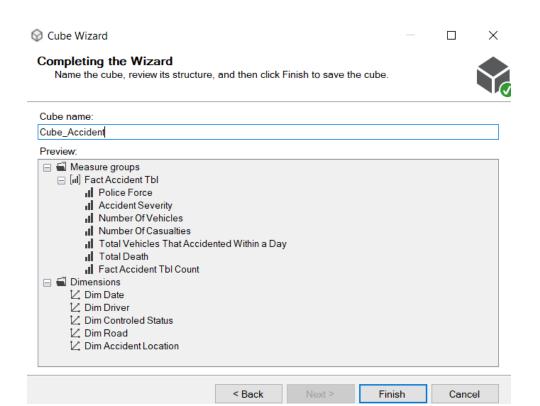
After creating data sources my next step was creating data source view. The step describes the relations and views of dataset.



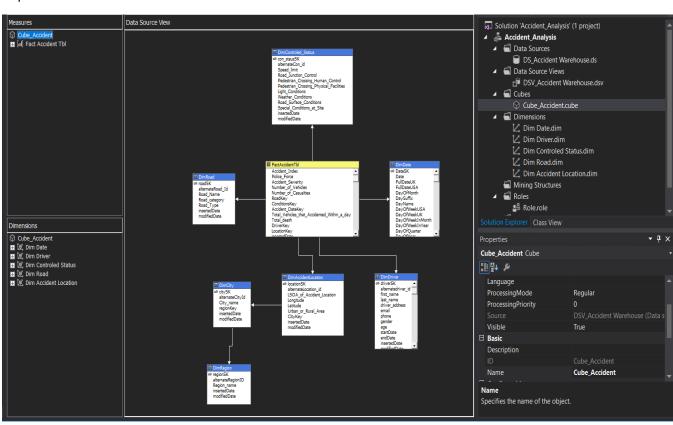
2.3

STEP 03: The Cube Creation

In here, the cube was created using data source view. We can use the existing data source to create the cube. As well as FactAccidentTbl is fact table and it contains all included measures required for the analysis process. In measure group there is FactAccidentTbl and in dimensions there is DimDate, DimCity, DimRegion, DimAccidentLocation, DimDriver, DimRoad, DimControledStatus.



Implemented cube is below.



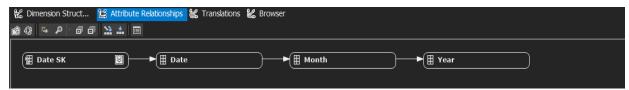
STEP 04: Hierarchies Creation

Hierarchies are used reduce complexity and normalized tables. This will use to drill down behavior.

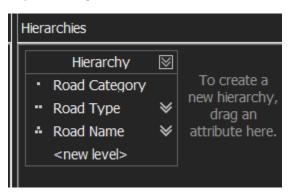
I have created three hierarchies for Date, Location and Road.

> Implementing hierarchies to the DimDate



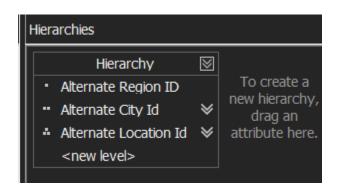


Implementing hierarchies to the DimAccidentLocation





> Implementing hierarchies to the DimRoad

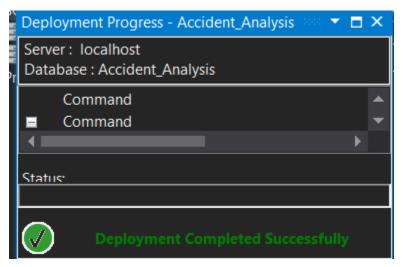




2.5

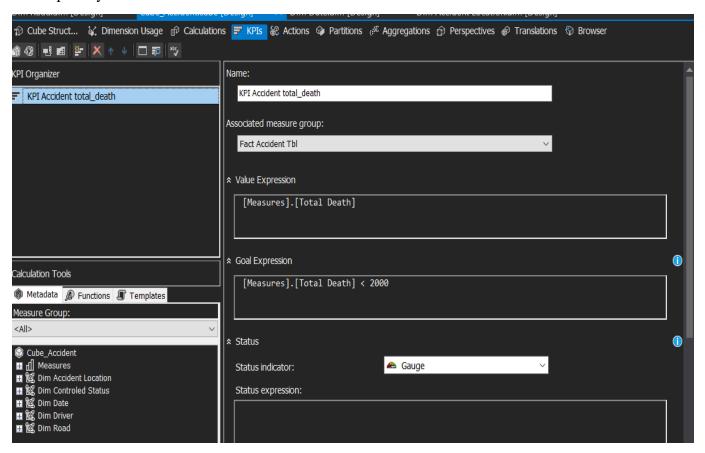
STEP 5: Deploy the Cube

The cube must be deployed in accident to be used for analysis. Having proper connection details provided will take effect in this stage when you try to deploy. Once deployed, SSAS Cube will be available for analysis under SSAS database accessible via SSMS.



STEP 6: KPIs Creation

KPI's are created based on the business requirements. KPIs depend on what the organization want to monitor and measure. According to my scenario I wanted to measure Accident Total deaths per day was less than 2000.



STEP 03

Demonstration of OLAP Operations

Online Analytical Processing (OLAP) is the technology behind many Business Intelligence (BI) Applications. As well as this will provide powerful capabilities for data mining and trend analysis.

OLAP technology has neem defined as the ability to archive "Fast access to share multidimensional information".

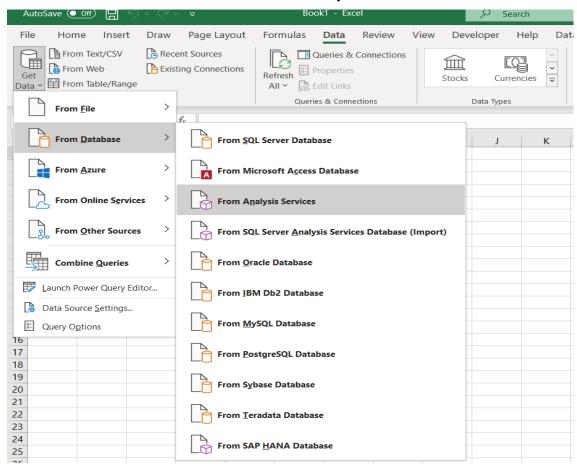
MDX: Multi-Dimensional Expressions is used for retrieving data from SSAS cubes.

According to my scenario I did not used MDX and I have used data tab in the ribbon.

3.1

STEP 01: SSAS Cube Connection

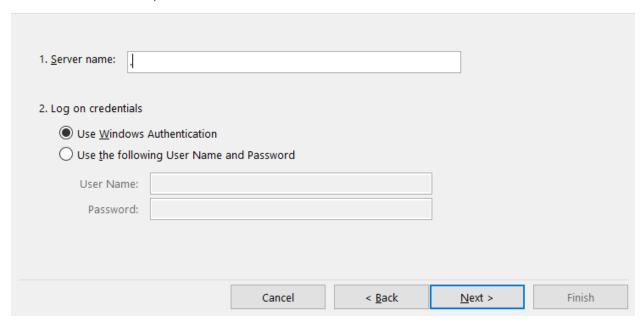
Data → Get Data → From Database → From Analysis Services



Data Connection Wizard ? X

Connect to Database Server

Enter the information required to connect to the database server.



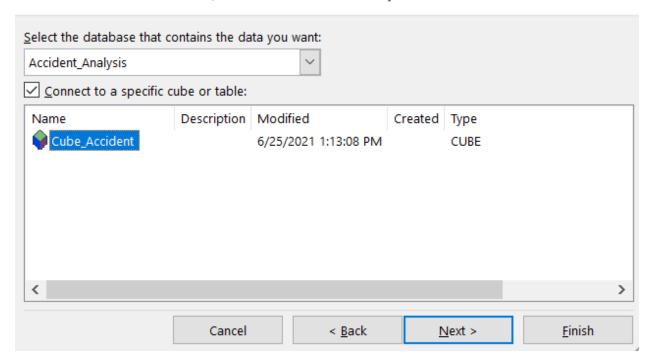
?

Х

Data Connection Wizard

Select Database and Table

Select the Database and Table/Cube which contains the data you want.



STEP 02: Demonstration of Excel Report for OLAP Operations

There is five reports for showing OLAP operations. Those are,

Report 1: Roll UpReport 2: Drill Down

❖ Report 3: Slice❖ Report 4: Dice❖ Report 5: Pivot

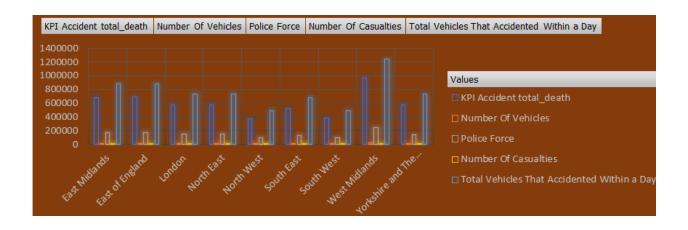
Report 01

Roll Up

Climbing up a hierarchy of a dimension or by dimension reduction to aggregate data. This is also known as consolidation or aggregation.

According to my scenario I have used total incidents of accidents in different regions and cities.

TOTAL INCIDENTS OF ACCIDENTS										
Row Labels	KPI Accident total_death	Number Of Vehicles	Police Force 1	Number Of Casualties	Total Vehicles That Accidented Within a Day					
± East Midlands	682641	17893	173128	13025	87850					
± East of England	687914	18029	174186	13125	88481					
± London	570706	15152	148185	10906	73538					
± North East	573450	15088	147647	10979	73885					
± North West	378053	9835	95596	7166	48601					
± South East	527185	13752	134135	10069	67776					
± South West	383962	10058	99797	7342	49374					
± West Midlands	968625	25571	249124	18564	124774					
+ Yorkshire and The Humbe	er 575134	14912	145613	10914	73867					
Grand Total	5347670	140290	1367411	102090	688149					



Drill Down

Stepping down a hierarchy of a dimension allowing navigation through details. Data is divided into smaller parts. It is the opposite of the roll up process. According to my scenario I used drill down for check total Number of Casualties, Number of Vehicles, Total Deaths within a day and total vehicles that accident within a day by region, cities and year wise.

Row Labels	Number Of Casualties Nur	mber Of Vehicles P	olice Force K	PI Accident total_deatl Total Vehic	les That Accidented Within a Da
⊞ East Midlands	13025	17893	173128	682641	878501
■ East of England					
2012	531	743	7267	28261	36341
2013	610	839	8705	31487	40567
2014	89	118	1036	4148	5368
□ Derbyshire					
2012	547	711	6640	26694	34454
2013	553	769	7298	29165	37445
2014	94	127	1196	4708	6088
⊕ East Sussex	1161	1621	14999	60723	78363
⊕ Gloucestershire	1150	1582	15396	61667	79147
⊕Greater Manchester	1168	1653	15522	61931	79791
⊞Isle of Wight	1235	1717	16625	67627	86467
⊕ Kent	1142	1588	15462	59721	77281
⊕ Rutland	1241	1678	17197	63970	82290
⊕Slough	1208	1615	14630	62000	79760
⊕Tyne and Wear	1186	1613	15568	61600	79200
⊕ Wiltshire	1210	1655	16645	64212	82252
⊕ London	10906	15152	148185	570706	735386
■ North East	10979	15088	147647	573450	738850
■ North West	7166	9835	95596	378053	486013
⊞South East	10069	13752	134135	527185	677765
■ South West	7342	10058	99797	383962	493742
■ West Midlands	18564	25571	249124	968625	1247745
■ Yorkshire and The Humber	10914	14912	145613	575134	738674
Grand Total	102090	140290	1367411	5347670	6881490

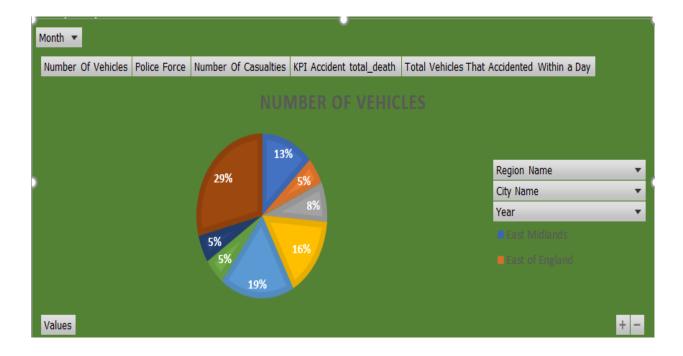


Slice

A slice representation two or more dimensional views of an OLAP data cube. A slice function much like a report or a query. Then we can request data what we need to see.

According to my scenario I have used slice for check total Number of Casualties, Number of Vehicles, Total Deaths within a day and total vehicles that accident within a day in specific road name by region, city and year wise.

Row Labels	Number Of Vehicles	Police Force	Number Of Casualties	KPI Accident total_death	Total Vehicles That Accidented Within a Day	Road Name 👙 💍	١×
⊕ East Midlands	5	41	3	228	288	A1031	٨
± East of England	2	42	2	144	184		
⊕ North East	3	34	2	152	192	A1033	
⊕ North West	6	50	6	197	257	A1033	
⊕ South East	7	6	4	222	282	14000	
South West						A1033	
⊟ Bedfordshire						A1033	
2013	2	. 1	1	64	84	44022	
☐ Lincolnshire						A1033	
2012	2	20	1	46	66	A1033	
⊕ West Midlands	11	115	5	218	298	A1033	
<u> </u>	·		-	·	1651	ATOSS	٧

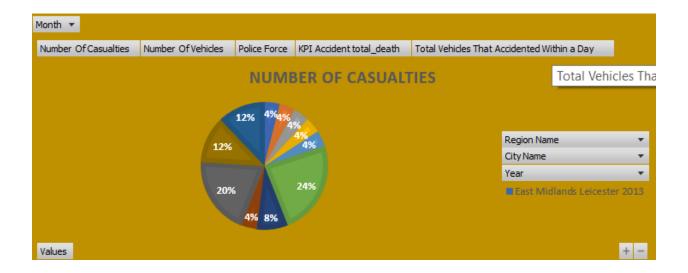


Dice

Select two or more dimensions from s given cube and provides a new sub-cube by selecting specific values on those selected dimensions. Dice operation is similar to a slice.

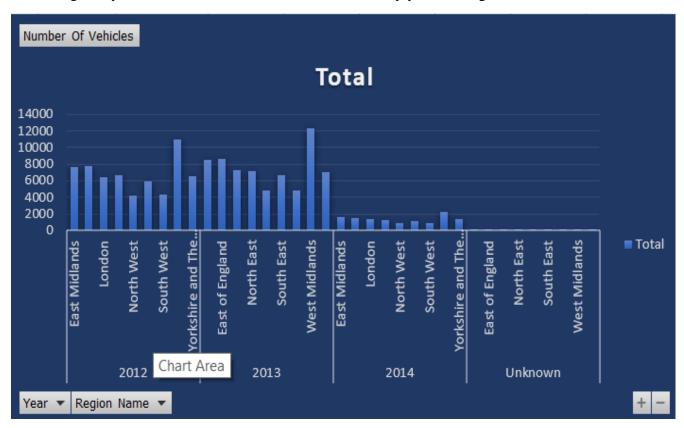
According to my scenario I have used slice for check total Number of Casualties, Number of Vehicles, Total Deaths within a day and total vehicles that accident within a day. in specific road name and speed limit by region, city and year wise.

Month	All	w										
Row Labels	Number Of Casual	ties Nur	mber Of Vehicles P	olice Force I	KPI Accident total_deatl	Total Vehicles That Accidented Within a Da						
■ East Midlands												
∃ Leicester							Speed	limit	% ∃ %	Road N	ame	% ≡ 5
2013		1	2	20	41	61	- Speed		V- IX			
■ East of England							30			A58		
∃Slough							20			A64		
2014		1	2	20	15	35						
∃London							40	40			A64M	
∃ Derby							50			A38		
2013		1	1	30	71	91						
∃Norfolk							70			A38M		
2012		1	2	1	126	146	Unkno	wn		A58		
■ West Sussex										45414		
2014		1	2	31	93	113				A64M		
■ North East		6	8	43	243	323				A38		
■ North West		2	4	23	212	252						
■ South East		1	2	32	91	111						
■ South West		5	8	56	221	301						
		3	3	6	128	148						
■ Yorkshire and The Humb	iei	3	3	44	159	199						
Grand Total		25	37	306	1400	1780						



Pivot Chart

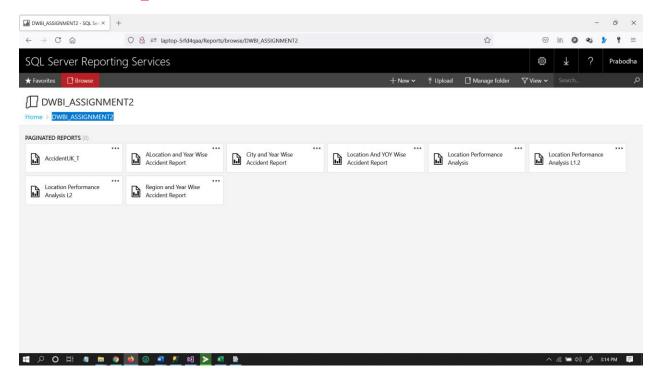
According to my scenario I have used Number of vehicles by year and regions.



STEP 04

SQL Service Reporting Service Report

(SSRS Report)



The **SQL Server Reporting Service** (**SSRS**) is a reporting application that provides you to generate formatted tables reports in the form of data, pictures, diagrams and charts. The reports are hosted on a server which may be performed with user-defined settings any time. A reporting services server's web portal is a web-based experience. The report, mobile report, KPI and element in your report server instance may be viewed on the portal. You may also manage a single report server instance using the web portal.

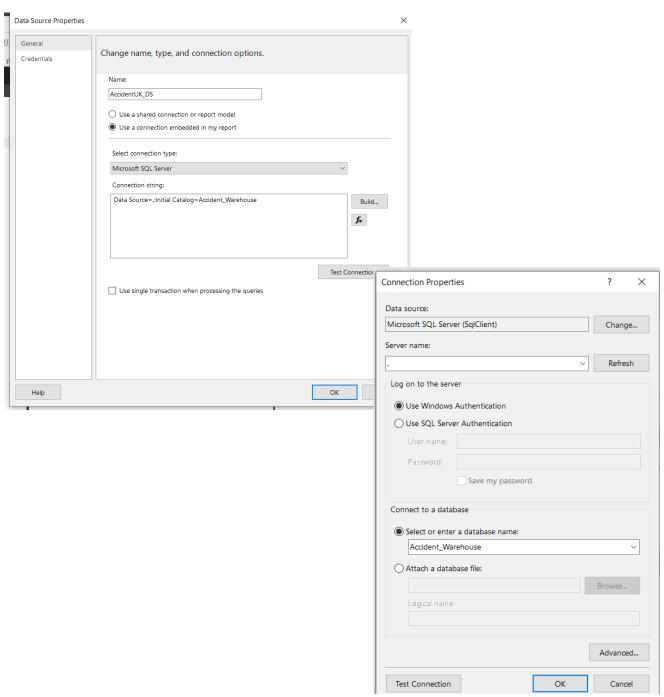
I have used report builder application for this. The **Report Builder** is an independent software that you or an administrator installs on your computer. It may be installed through the Microsoft Download Center, the 2016 reporting services SQL Servers or later (SSRS) or from a reporting services SharePoint site.

4.1

Basic Steps of Report Builder

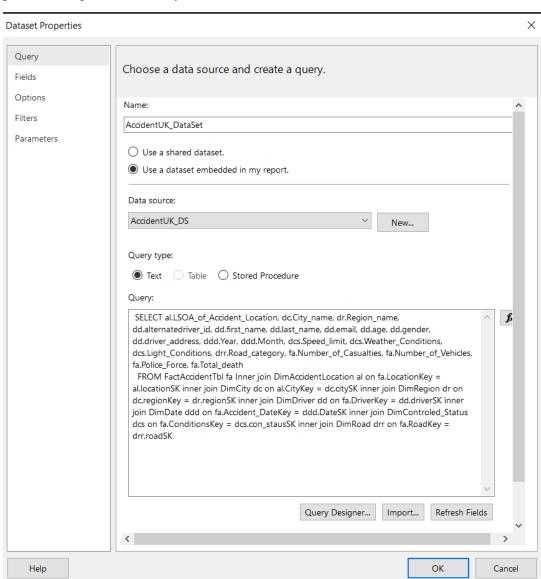
Step 01: Data Source Creation

Throughout this stage we will also be going to connect to our previously created data warehouse data source 'Accident_Warehouse'.



Step 02: Dataset Creation

In this step a data set is imported into our report builder with the help of a query which is manual process assigned as actually needed.



Step 3: Table creation

4.2

Report Demonstrations



Report with a matrix

Matrix is similar to a table in SSRS, but it is configured to display data categorized by rows and columns, with statistical data at the crossing. In a spreadsheet, this is equivalent to using a pivot table.

Query for AccidentUK_Dataset:

SELECT al.LSOA_of_Accident_Location, dc.City_name, dr.Region_name, dd.alternatedriver_id, dd.first_name, dd.last_name, dd.email, dd.age, dd.gender, dd.driver_address, ddd.Year, ddd.Month, dcs.Speed_limit, dcs.Weather_Conditions, dcs.Light_Conditions, drr.Road_category, fa.Number_of_Casualties, fa.Number_of_Vehicles, fa.Police_Force, fa.Total_death

FROM FactAccidentTbl fa

Inner join DimAccidentLocation al on fa.LocationKey = al.locationSK

inner join DimCity dc on al.CityKey = dc.citySK

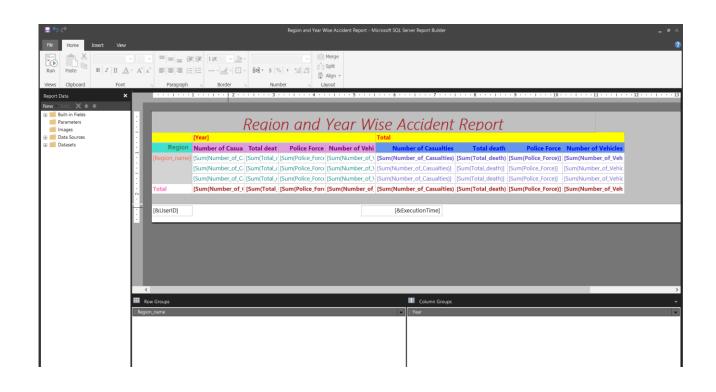
inner join DimRegion dr on dc.regionKey = dr.regionSK

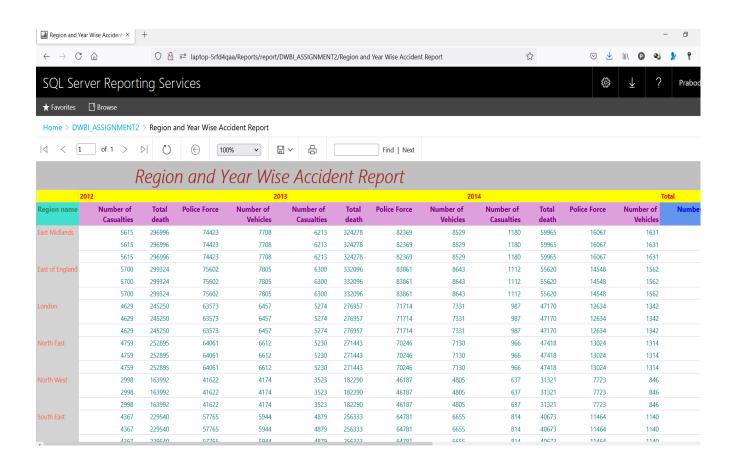
inner join DimDriver dd on fa.DriverKey = dd.driverSK

inner join DimDate ddd on fa.Accident_DateKey = ddd.DateSK

inner join DimControled_Status dcs on fa.ConditionsKey = dcs.con_stausSK

inner join DimRoad drr on fa.RoadKey = drr.roadSK





Report with more than one parameter

Applying Multiparameter-values in SSRS enables us to transfer one or more than the input data to the report. This one has a "Select All" choice which always allows you to select all parameter values.

Query for AccidentUK_Dataset:

This data set contains the UK accident data which will be displayed in the report through the matrix.

SELECT al.LSOA_of_Accident_Location, dc.City_name, dr.Region_name, dd.alternatedriver_id, dd.first_name, dd.last_name, dd.email, dd.age, dd.gender, dd.driver_address, ddd.Year, ddd.Month, dcs.Speed_limit, dcs.Weather_Conditions, dcs.Light_Conditions, drr.Road_category, fa.Number_of_Casualties, fa.Number_of_Vehicles, fa.Police_Force, fa.Total_death

FROM FactAccidentTbl fa

Inner join DimAccidentLocation al on fa.LocationKey = al.locationSK

inner join DimCity dc on al.CityKey = dc.citySK

inner join DimRegion dr on dc.regionKey = dr.regionSK

inner join DimDriver dd on fa.DriverKey = dd.driverSK

inner join DimDate ddd on fa.Accident_DateKey = ddd.DateSK

inner join DimControled_Status dcs on fa.ConditionsKey = dcs.con_stausSK

inner join DimRoad drr on fa.RoadKey = drr.roadSK

Where dc.City_name in (@City)

Query for Regions:

1st parameter. This dataset describes regions details.

select regionSK, alternateRegionID, Region_name

from DimRegion

Query for City:

2nd parameter. This dataset describes city details.

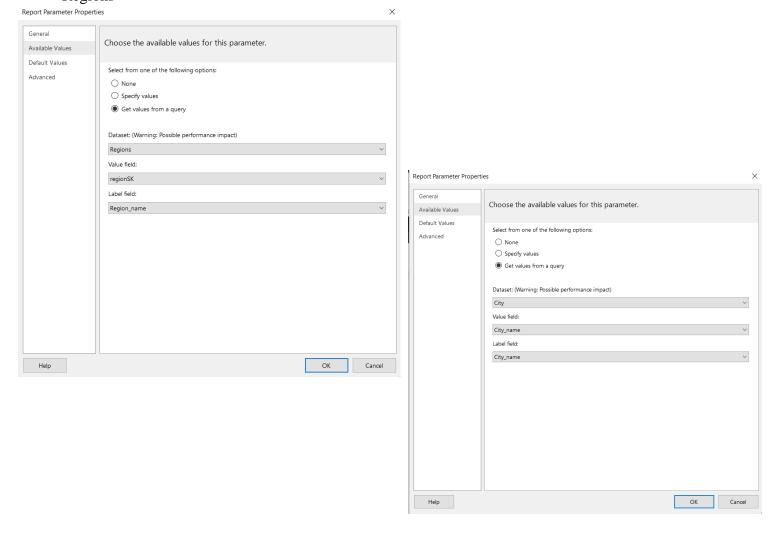
select citySK, alternateCityId, City_name, regionKey

from DimCity

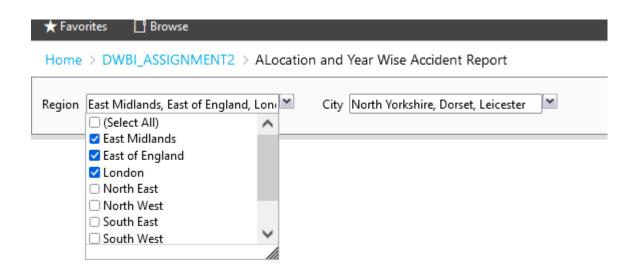
where regionKey in (@Region)

parameters

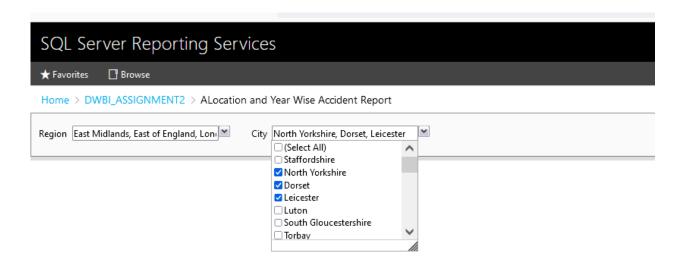
Regions

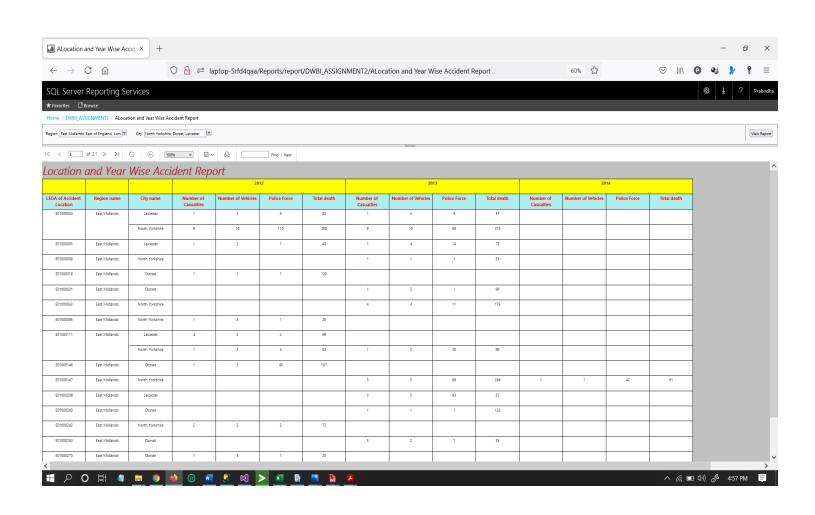


1st parameter - Region



2nd parameter -City





SSRS drill-down report

Using Drill Down in SSRS simply allows users to display or cover up column records by presenting plus and minus signs on such a text field (In short, providing interactivity to the user.).

Query for AccidentUK_Dataset:

SELECT al.LSOA_of_Accident_Location, dc.City_name, dr.Region_name, dd.alternatedriver_id, dd.first_name, dd.last_name, dd.email, dd.age, dd.gender, dd.driver_address, ddd.Year, ddd.Month, dcs.Speed_limit, dcs.Weather_Conditions, dcs.Light_Conditions, drr.Road_category, fa.Number_of_Casualties, fa.Number of Vehicles, fa.Police Force, fa.Total death

FROM FactAccidentTbl fa

Inner join DimAccidentLocation al on fa.LocationKey = al.locationSK

inner join DimCity dc on al.CityKey = dc.citySK

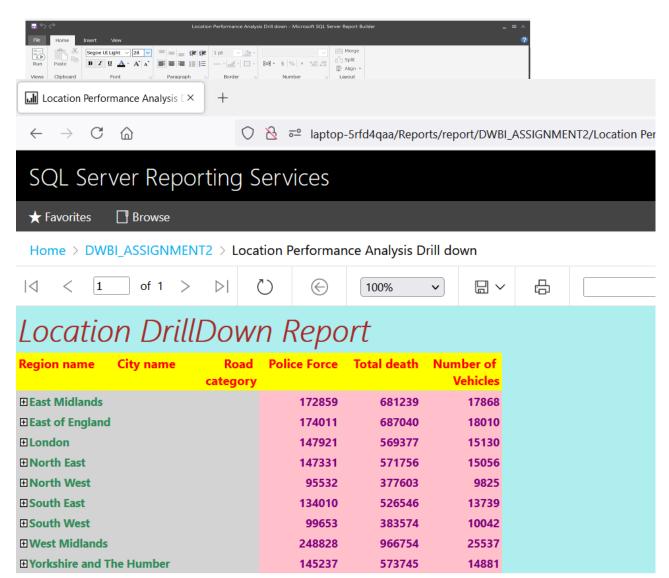
inner join DimRegion dr on dc.regionKey = dr.regionSK

inner join DimDriver dd on fa.DriverKey = dd.driverSK

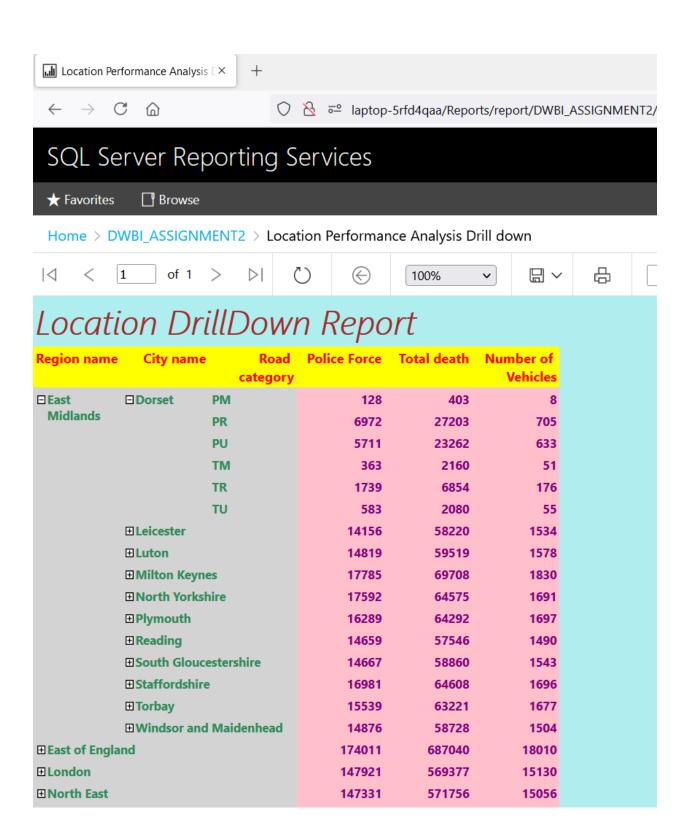
inner join DimDate ddd on fa.Accident_DateKey = ddd.DateSK

inner join DimControled_Status dcs on fa.ConditionsKey = dcs.con_stausSK

inner join DimRoad drr on fa.RoadKey = drr.roadSK



LAPTOP-5RFD4QAA\Pra bodha 6/25/2021 6:50:44 PM



SSRS drill-through report

Main Query:

SELECT al.LSOA_of_Accident_Location, dc.City_name, dr.Region_name, dd.alternatedriver_id, dd.first_name, dd.last_name, dd.email, dd.age, dd.gender, dd.driver_address, ddd.Year, ddd.Month, dcs.Speed_limit, dcs.Weather_Conditions, dcs.Light_Conditions, drr.Road_category, fa.Number_of_Casualties, fa.Number of Vehicles, fa.Police Force, fa.Total death

FROM FactAccidentTbl fa

Inner join DimAccidentLocation al on fa.LocationKey = al.locationSK

inner join DimCity dc on al.CityKey = dc.citySK

inner join DimRegion dr on dc.regionKey = dr.regionSK

inner join DimDriver dd on fa.DriverKey = dd.driverSK

inner join DimDate ddd on fa.Accident_DateKey = ddd.DateSK

inner join DimControled_Status dcs on fa.ConditionsKey = dcs.con_stausSK

inner join DimRoad drr on fa.RoadKey = drr.roadSK

Sub Query:

SELECT al.LSOA_of_Accident_Location, dc.City_name, dr.Region_name, dd.alternatedriver_id, dd.first_name, dd.last_name, dd.email, dd.age, dd.gender, dd.driver_address, ddd.Year, ddd.Month, dcs.Speed_limit, dcs.Weather_Conditions, dcs.Light_Conditions, drr.Road_category, fa.Number_of_Casualties, fa.Number of Vehicles, fa.Police Force, fa.Total death

FROM FactAccidentTbl fa

Inner join DimAccidentLocation al on fa.LocationKey = al.locationSK

inner join DimCity dc on al.CityKey = dc.citySK

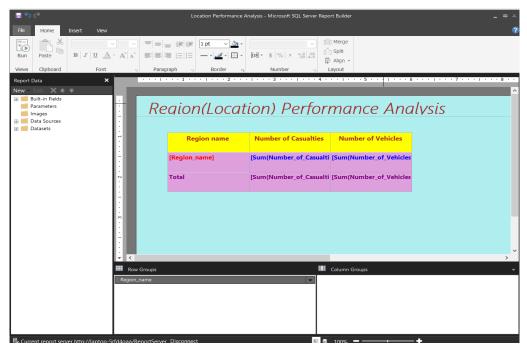
inner join DimRegion dr on dc.regionKey = dr.regionSK

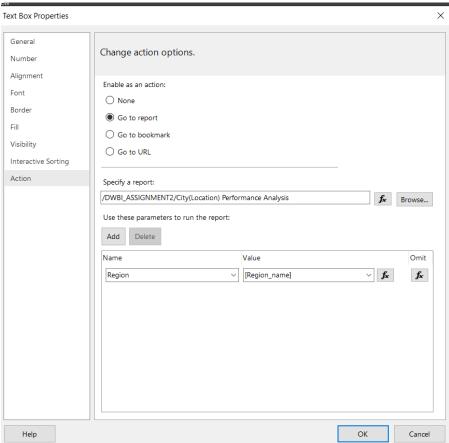
inner join DimDriver dd on fa.DriverKey = dd.driverSK

inner join DimDate ddd on fa.Accident_DateKey = ddd.DateSK

inner join DimControled_Status dcs on fa.ConditionsKey = dcs.con_stausSK inner join DimRoad drr on fa.RoadKey = drr.roadSK

Region (Location) Performance Analysis





City (Location) Performance Analysis

