MSIS 5663
Data Warehousing

TEAM 8 - FINAL DOCUMENT

Anveshi Atul, A20040500
Praisy Mary Abraham, A20110475
Saumitra Joshi, A20094307

OKLAHOMA STATE UNIVERSITY

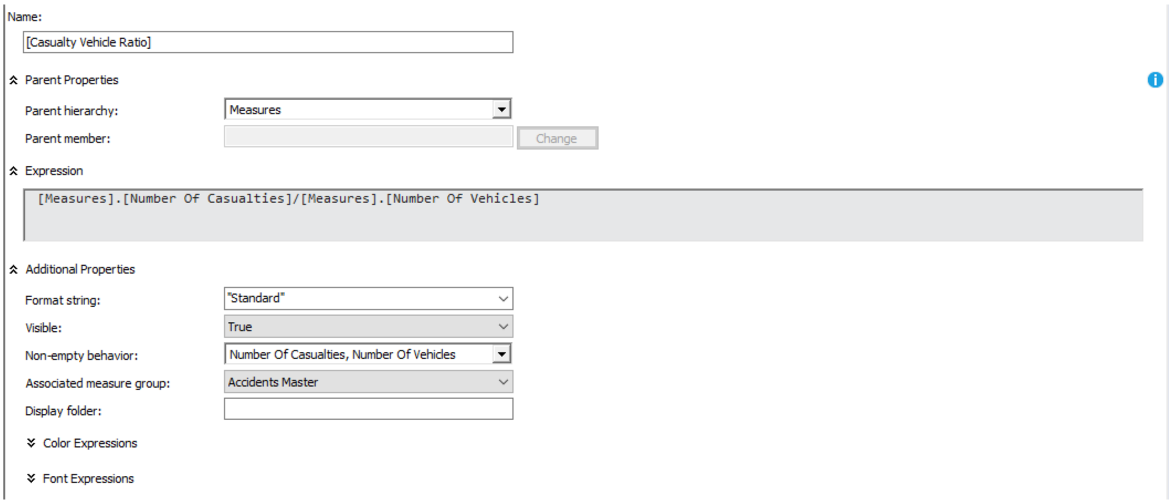
Table of Contents

CALCULATED MEASURES

The calculated measures were created using the existing measure groups and here is there specification as follows:

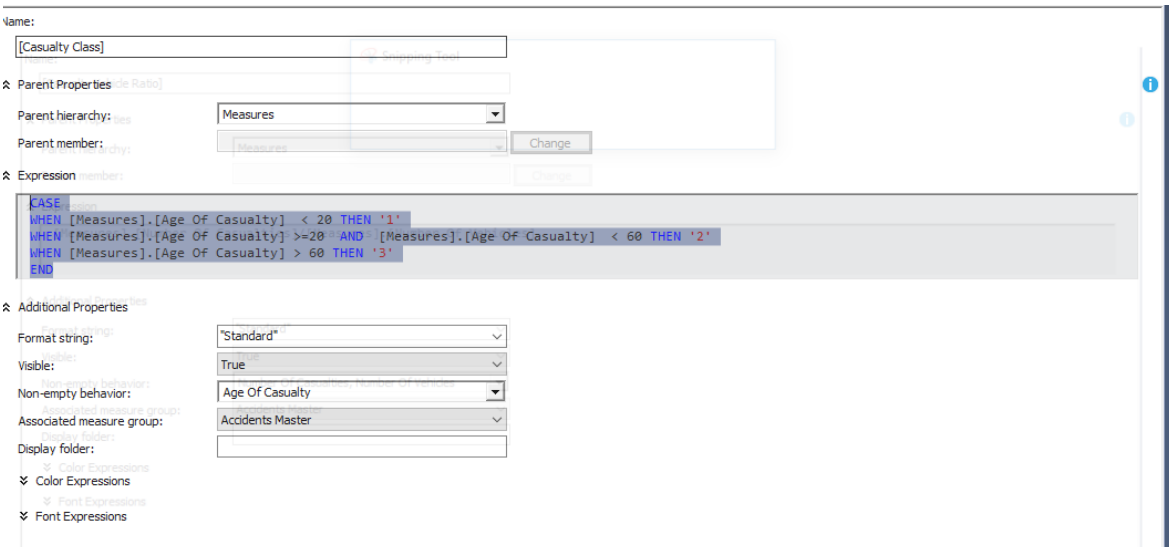
Casualty Vehicle Ratio:

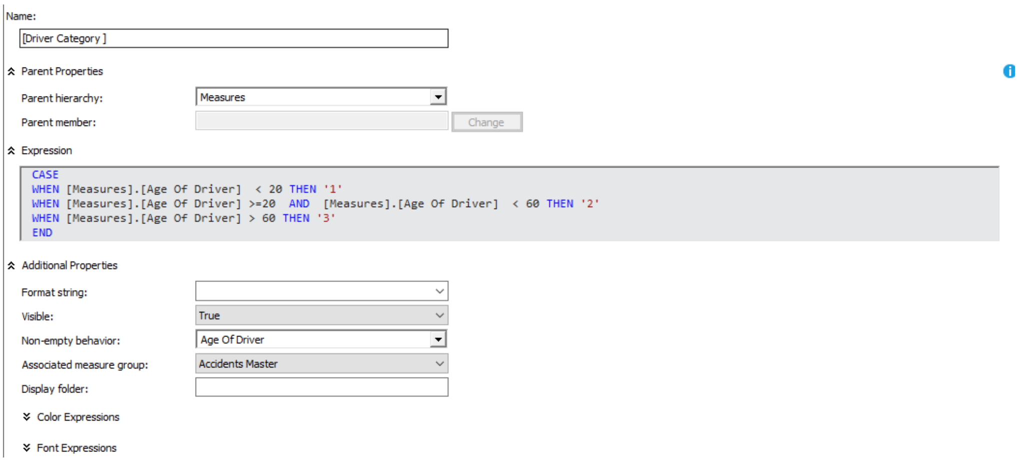
Signifies the ratio of people dead to the number of casualties in the accident.



Casualty Class and Driver Category:

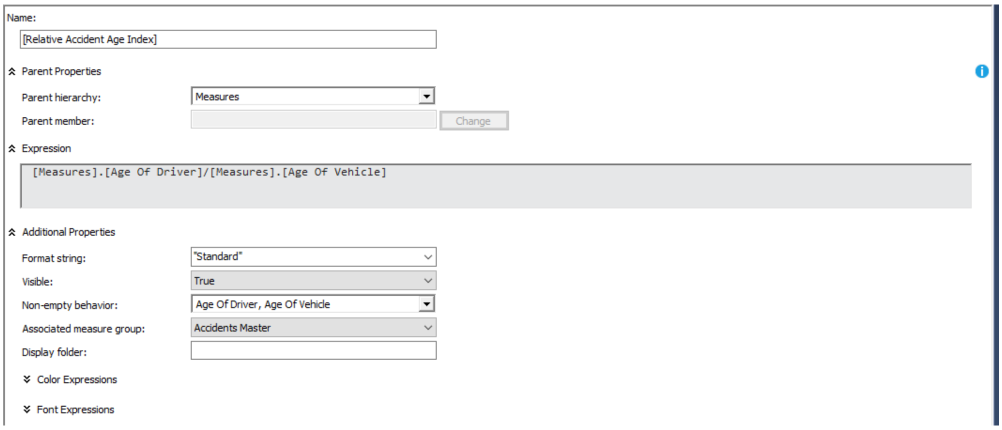
Based on age of casualty and driver is classified as Young, Middle Aged and Old.



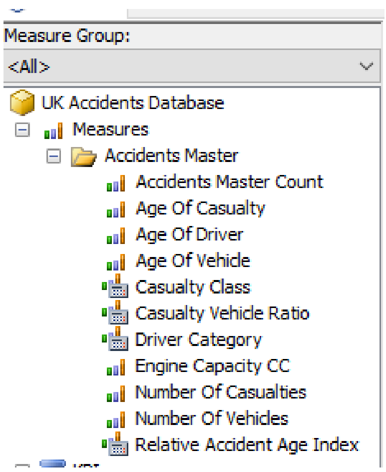


Relative Age Index:

It captures the ratio of the age of the driver to vehicle which can be used in calculating probability of accidents.



Deployed:



HIERARCIES

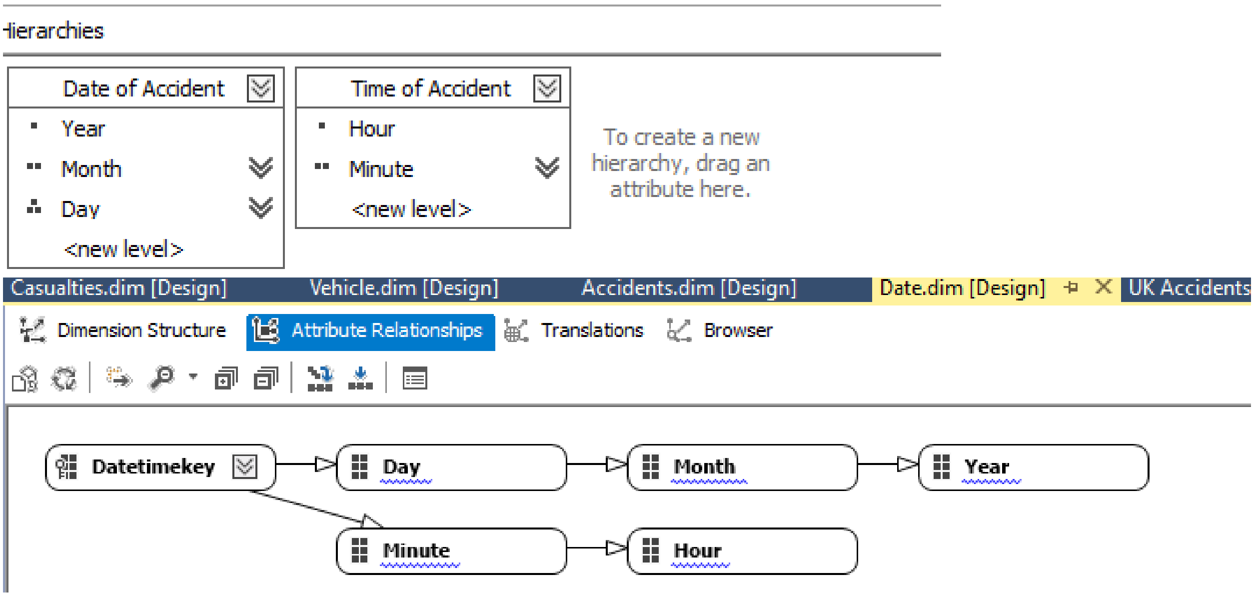
The date dimension hierarchies created pertain to the day and time of accident.

Day of Accident:

In this hierarchy you can drill down to day of week starting from Year to Month to Day.

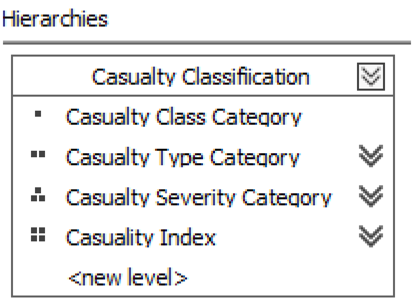
Time of Accident:

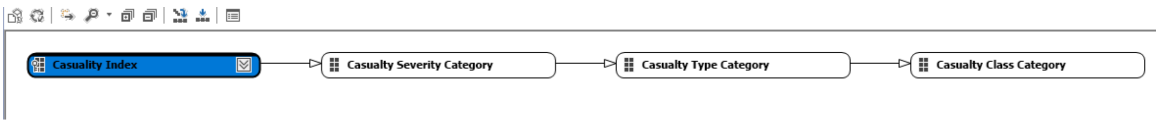
In this hierarchy you can drill down to minute from hour at which the accident occurred.



Casualty dimension hierarchies focus on the casualty class and then drill down to the type of casualty before further drilling down to the casualty severity of the particular victim.

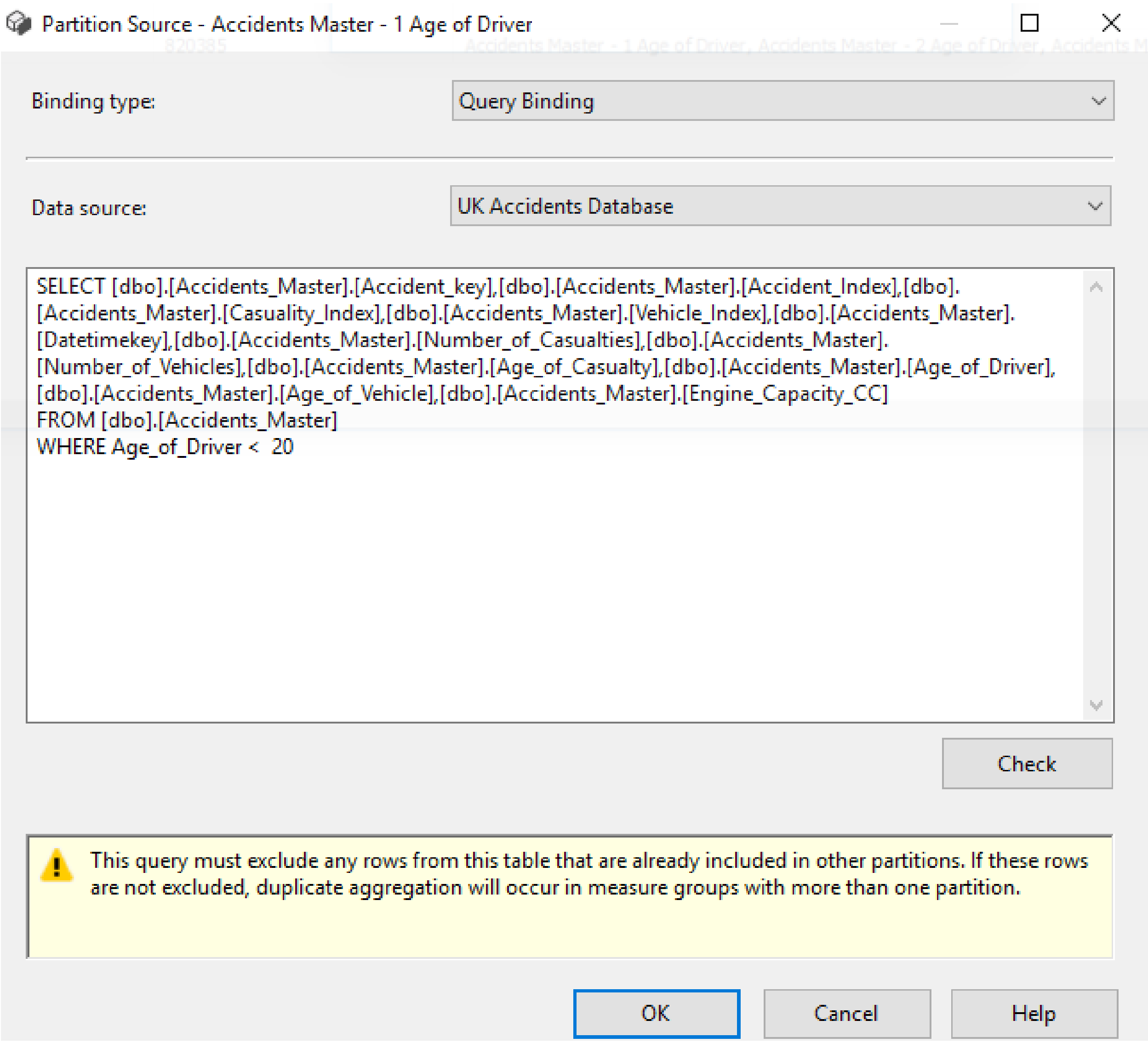
Casualty Classification Hierarchy:

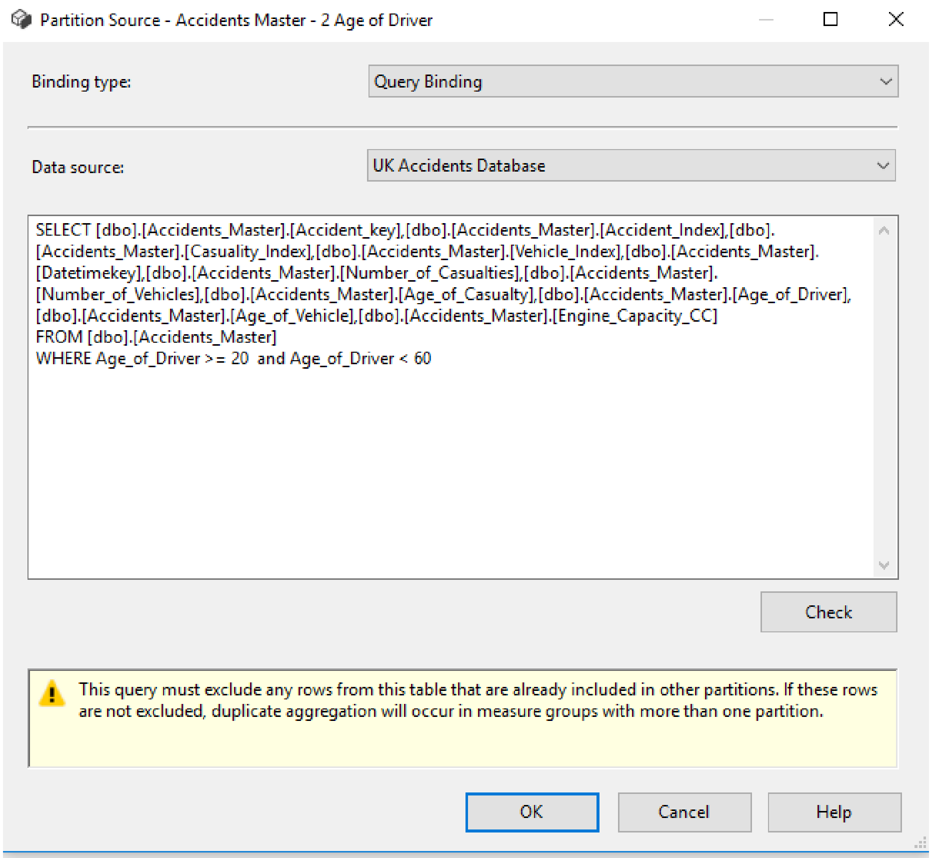


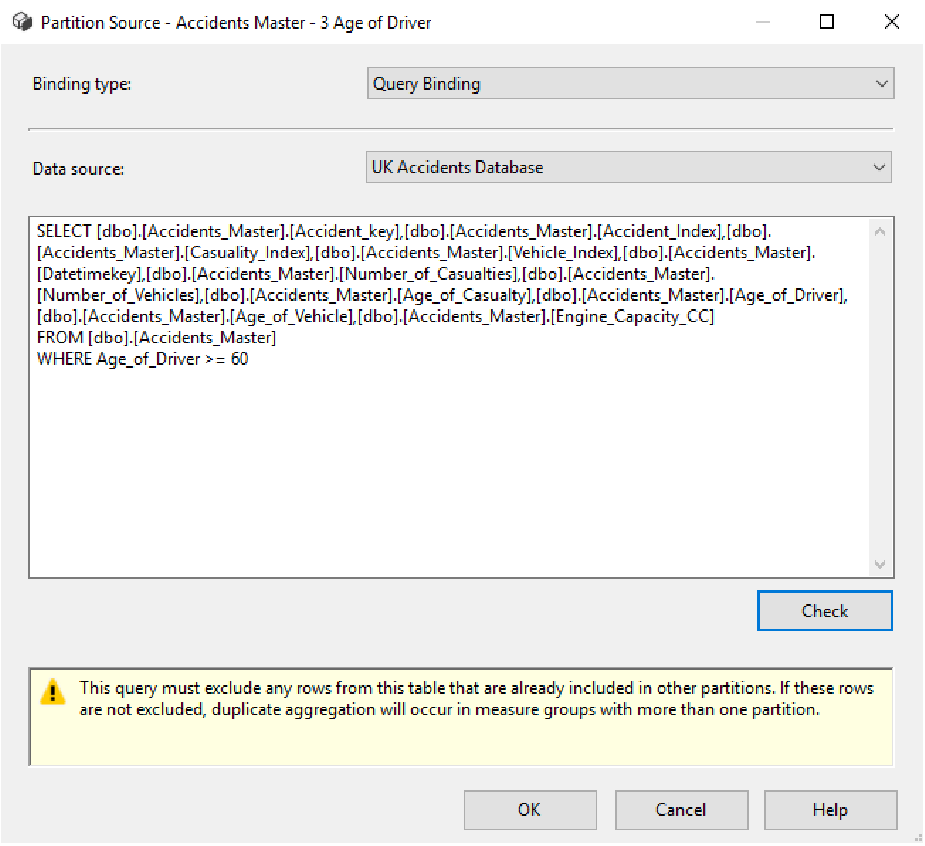


PARTITIONS AND AGGREGATIONS

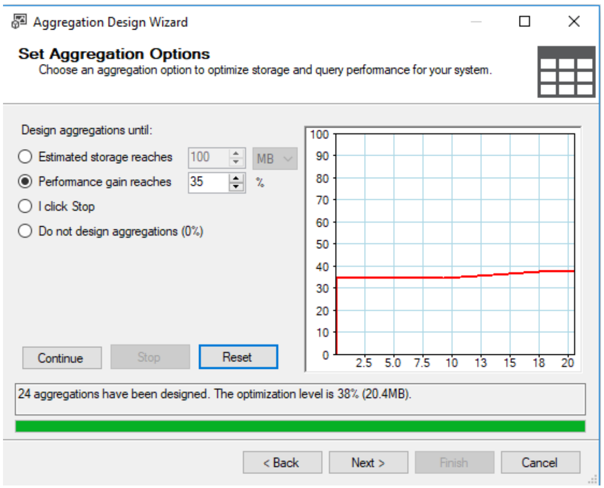
The partitions are created on the basis of the age of the driver involved in the accident and the partition specifications are as shown below:

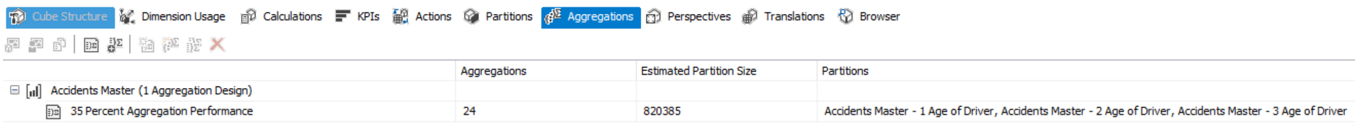



Partition Source - Accidents Master - 2 Age of Driver
Binding type:
Query Binding
Data source:
UK Accidents Database
SELECT [dbo].[Accidents_Master].[Accident_key],[dbo].[Accidents_Master].[Accident_Index],[dbo].
[Accidents_Master].[Casuality_Index],[dbo].[Accidents_Master].[Vehicle_Index],[dbo].[Accidents_Master].
[Datetimekey],[dbo].[Accidents_Master].[Number_of_Casualties],[dbo].[Accidents_Master].
[Number_of_Vehicles],[dbo].[Accidents_Master].[Age_of_Casualty],[dbo].[Accidents_Master].[Age_of_Driver],
[dbo].[Accidents_Master].[Age_of_Vehicle],[dbo].[Accidents_Master].[Engine_Capacity_CC]
FROM [dbo].[Accidents_Master]
WHERE Age_of_Driver >= 20 and Age_of_Driver < 60
Check
This query must exclude any rows from this table that are already included in other partitions. If these rows are not excluded, duplicate aggregation will occur in measure groups with more than one partition.
OK
Cancel
Help

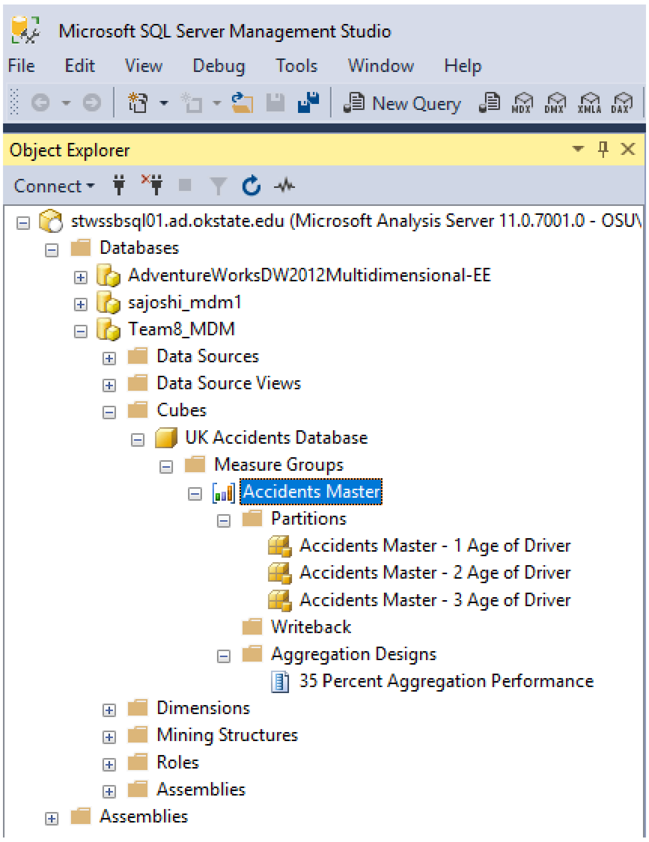

Partition Source - Accidents Master - 3 Age of Driver
Binding type:
Query Binding
Data source:
UK Accidents Database
SELECT [dbo].[Accidents_Master].[Accident_key],[dbo].[Accidents_Master].[Accident_Index],[dbo].
[Accidents_Master].[Casuality_Index],[dbo].[Accidents_Master].[Vehicle_Index],[dbo].[Accidents_Master].
[Datetimekey],[dbo].[Accidents_Master].[Number_of_Casualties],[dbo].[Accidents_Master].
[Number_of_Vehicles],[dbo].[Accidents_Master].[Age_of_Casualty],[dbo].[Accidents_Master].[Age_of_Driver],
[dbo].[Accidents_Master].[Age_of_Vehicle],[dbo].[Accidents_Master].[Engine_Capacity_CC]
FROM [dbo].[Accidents_Master]
WHERE Age_of_Driver >= 60
Check
This query must exclude any rows from this table that are already included in other partitions. If these rows are not excluded, duplicate aggregation will occur in measure groups with more than one partition.
OK
Cancel
Help

The aggregations were created for 35% performance gain in query execution in order to enhance query performance and the results are shown below:





Partitions and aggregations deployed:



MDX QUERIES

- Display Number of Vehicles involved in accidents every year.

```
SELECT [Measures].[Number Of Vehicles] ON 0,
[Date].[Year].CHILDREN ON 1
FROM [UK Accidents Database]
```

100 %

Messages | Results

	Number Of Vehicles
2005	1200680
2006	1138336
2007	1104397
2008	1003139
2009	969672
2010	905458
2011	965219
2012	847437
2013	1107294
2014	851339
2015	868988

- Comparing Casualty Vehicle Ratio over Years using **UNION**

```
SELECT Union
([Date].[Date of Accident].[Year].Members,
{[Date].[Date of Accident].[Year].&[2005]},
{[Date].[Date of Accident].[Year].&[2006]}
) ON 0,
[Measures].[Casualty Vehicle Ratio] ON 1
FROM [UK Accidents Database]
```

100 %

Messages | Results

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Casualty Vehicle Ratio	0.88	0.88	0.90	0.90	0.88	0.90	0.97	0.89	0.92	0.91	0.88

- Displaying number of vehicles involved in the October month of year 2015 using **ANCESTORS.**

```
SELECT {[Measures].[Number Of Vehicles]} ON 0,
{
(ANCESTORS([Date].[Date of Accident].[Month].&[2015]&[10],0)),
(ANCESTORS([Date].[Date of Accident].[Month].&[2015]&[10],1))
} ON ROWS
FROM [UK Accidents Database]
```

100 %

Messages | Results

	Number Of Vehicles
10	72788
2015	868988

- Display Number of Casualties for different Vehicle Type in each year with the 3 highest Number of Casualties using **GENERATE** and **TOPCOUNT**

```
SELECT {[Measures].[Number Of Casualties]} ON 0,
GENERATE (
[Date].[Year].[Year].MEMBERS,
TOPCOUNT
({[Date].[Year].CURRENTMEMBER * [Vehicle].[Vehicle Type Desc].[Vehicle Type Desc]}, 3, [Measures].[Number Of Casualties])
) ON 1
FROM [UK Accidents Database]
```

100 %

Messages | Results

		Number Of Casualties
2005	Car	843572
2005	Van / Goods 3.5 tonnes mgw or under	47076
2005	Bus or coach (17 or more pass seats)	40681
2006	Car	805090
2006	Van / Goods 3.5 tonnes mgw or under	44751
2006	Bus or coach (17 or more pass seats)	35109
2007	Car	775591
2007	Van / Goods 3.5 tonnes mgw or under	45861
2007	Bus or coach (17 or more pass seats)	39393
2008	Car	709019
2008	Bus or coach (17 or more pass seats)	41629
2008	Van / Goods 3.5 tonnes mgw or under	37148
2009	Car	691154
2009	Van / Goods 3.5 tonnes mgw or under	37148
2009	Bus or coach (17 or more pass seats)	28321
2010	Car	645215
2010	Bus or coach (17 or more pass seats)	36118
2010	Van / Goods 3.5 tonnes mgw or under	34941
2011	Car	698806
2011	Van / Goods 3.5 tonnes mgw or under	64051
2011	Bus or coach (17 or more pass seats)	50166
2012	Car	590989
2012	Van / Goods 3.5 tonnes mgw or under	37532
2012	Bus or coach (17 or more pass seats)	25027
2013	Car	770545
2013	Van / Goods 3.5 tonnes mgw or under	107512
2013	Bus or coach (17 or more pass seats)	24927
2014	Car	584338
2014	Bus or coach (17 or more pass seats)	49554
2014	Van / Goods 3.5 tonnes mgw or under	38868
2015	Car	599644
2015	Van / Goods 3.5 tonnes mgw or under	40387
2015	Bus or coach (17 or more pass seats)	25116

- Display Top 5 Casualty Type Category for each day and month of the year 2005 using **DESCENDANTS.**

```
WITH SET [Top_5_Casualty_Type] AS
TOPCOUNT(([Casualties].[Casualty Classifiication].[Casualty Type Category]), 5)
SELECT [Top_5_Casualty_Type] ON 0,
DESCENDANTS
(
[Date].[Date of Accident].[Year].&[2005]
) ON ROWS
FROM [UK Accidents Database]
```

100 %

Messages | Results

	Cyclist	Motorcycle 50cc and under rider or passenger	Motorcycle 125cc and under rider or passenger	Motorcycle over 125cc and up to 500cc rider or passenger	Motorcycle over 500cc rider or passenger
2005	35082	10020	13996	5257	24179
1	2105	749	1098	404	1138
1	12	3	14	2	8
2	49	10	16	7	25
3	14	21	38	8	17
4	85	18	39	9	26
5	70	40	37	17	35
6	86	16	27	14	46
7	62	20	34	5	14
8	36	40	31	9	20
9	28	10	17	9	17
10	93	23	26	18	34
11	84	21	40	16	40
12	81	23	37	21	42
13	122	23	64	11	77
14	86	43	39	11	47
15	16	17	50	10	24
16	42	21	33	18	19
17	90	35	30	17	50
18	112	29	44	11	33
19	87	47	52	10	55
20	82	15	30	26	46
21	90	38	50	11	44
22	42	36	25	12	68
23	53	19	25	7	40
24	79	14	36	14	23
25	83	20	40	14	41
26	97	18	30	20	57
27	77	34	35	27	44
28	84	30	54	17	30
29	51	13	26	16	46
30	28	20	29	9	38
31	84	32	50	8	32
2	1972	648	850	329	971
1	102	21	51	15	35

- Display the Number of Casualities over the years using **Extract function**.

osoft SQL Server Management Studio

Generic Debugger

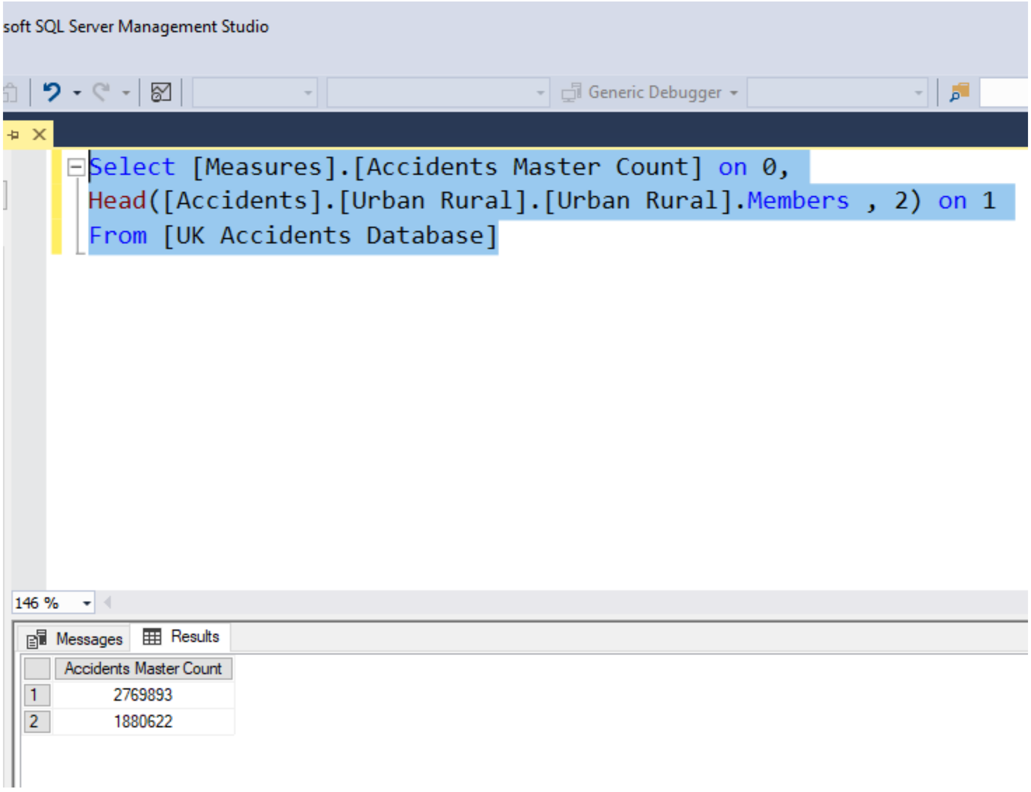
```
SELECT [Measures].[Number Of Casualties] ON 0,
Extract (([Date].[Year].[Year].Members , [Accidents].[Speed Limit].[Speed Limit].Members), [Date].[Year]) on 1
From [UK Accidents Database]
```

100 %

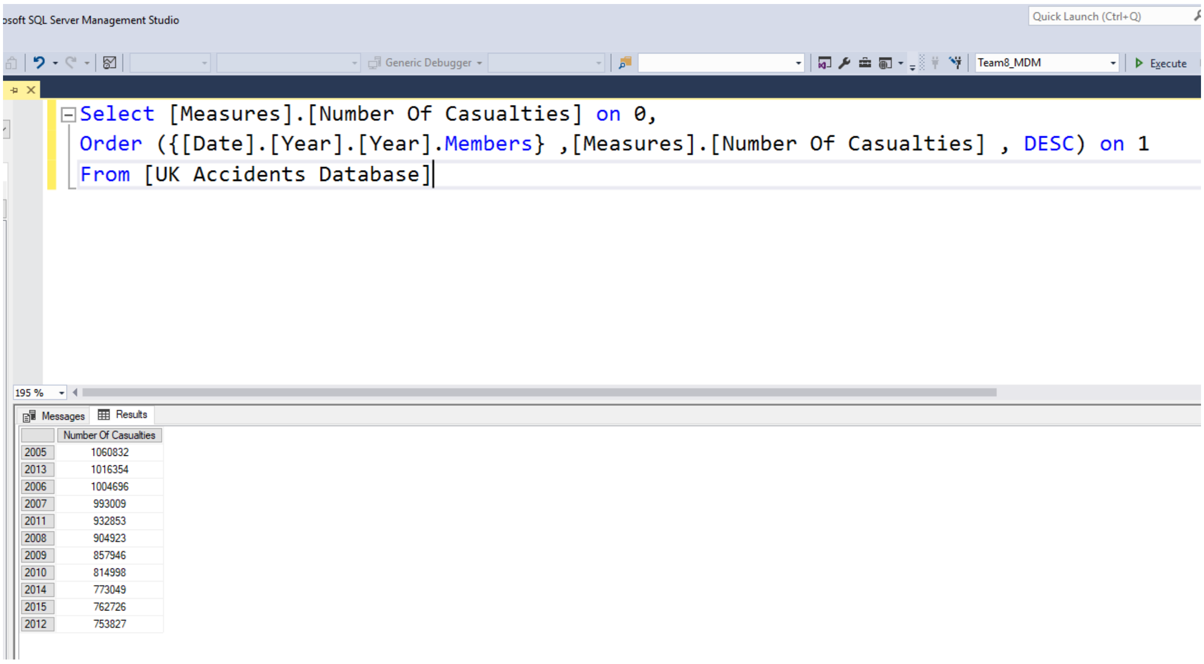
Messages | Results

	Number Of Casualties
2005	1060832
2006	1004696
2007	993009
2008	904923
2009	857946
2010	814998
2011	932853
2012	753827
2013	1016354
2014	773049
2015	762726

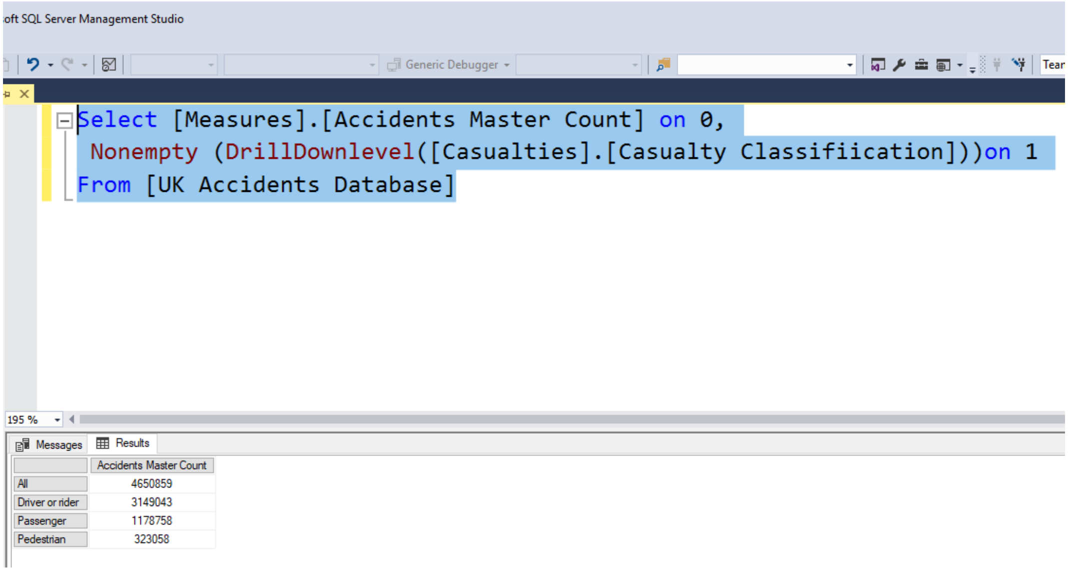
- Display first 2 members of Urban/Rural along with the total count of accidents occurred using **Head** function.



- Show the ordered set according to Number of Casualties in a descending order which shows the year having maximum number of causalities in the beginning and so on using **ORDER**.



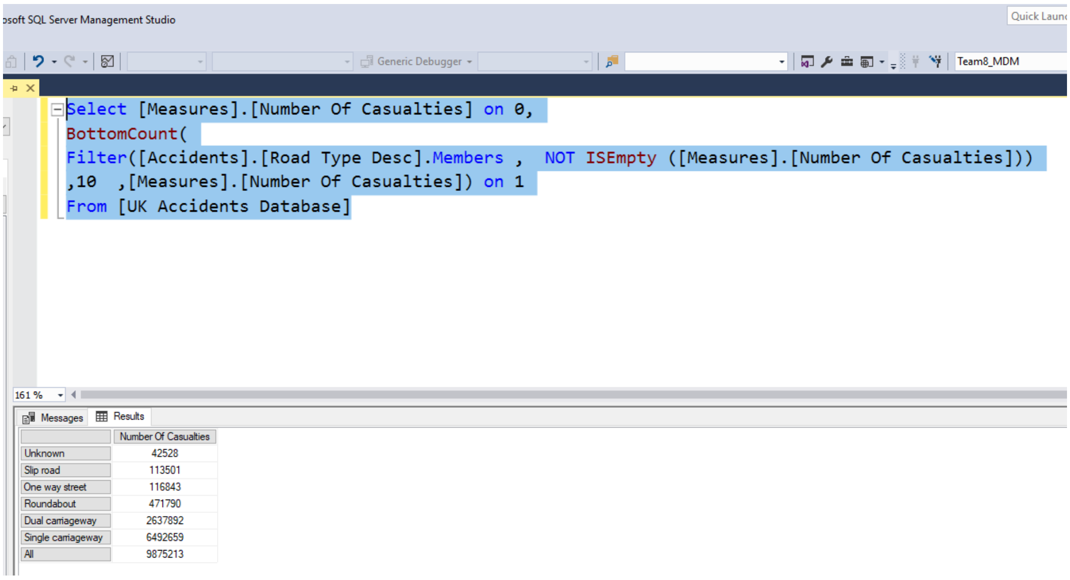
- Show the Accident occurrence count in the Causality Classification Hierarchy using **Drilldown** function.



```
Select [Measures].[Accidents Master Count] on 0,
 Nonempty (DrillDownlevel([Casualties].[Casualty Classifiication]))on 1
From [UK Accidents Database]
```

	Accidents Master Count
All	4650859
Driver or rider	3149043
Passenger	1178758
Pedestrian	323058

- This query filters accident based on the road type's description of the 10 lowest causalities using **Bottom Count and Filter** functions



```
Select [Measures].[Number Of Casualties] on 0,
BottomCount(
Filter([Accidents].[Road Type Desc].Members ,  NOT ISEmpty ([Measures].[Number Of Casualties]))
,10  ,[Measures].[Number Of Casualties]) on 1
From [UK Accidents Database]
```

	Number Of Casualties
Unknown	42528
Slip road	113501
One way street	116843
Roundabout	471790
Dual carriageway	2637892
Single carriageway	6492659
All	9875213

DMX MODELS

We used the **DimAccidents** table to predict accident severity based on various factors like road conditions, light conditions, weather conditions, and speed limits. To do so we built and implement mining structure and mining models using Data Mining Extensions. Our target variable is **Accident_Severity.**

Following are the DMX Queries for the Mining Structure and Mining Models:

DMX QUERIES:

Creating Mining Structure UK ACCIDENTS DMX:

```
CREATE MINING STRUCTURE [UK ACCIDENTS DMX]
(
   [Accident_Index] LONG KEY,
   [Accident_Severity] LONG DISCRETE,
   [First_Road_Class] LONG DISCRETE,
   [Road_Type]  LONG DISCRETE,
   [Speed_limit] LONG CONTINUOUS,
   [Junction_Detail] LONG DISCRETE,
   [Junction_Control] LONG DISCRETE,
   [Second_Road_Class] LONG DISCRETE,
   [Ped_Cross_Human] LONG DISCRETE,
   [Ped_Cross_Physical] LONG DISCRETE,
   [Light_Conditions] LONG DISCRETE,
   [Weather_Conditions]  LONG DISCRETE,
   [Road_Surface_Conditions] LONG DISCRETE,
   [Special_Conditions_at_Site] LONG DISCRETE,
   [Carriageway_Hazards] LONG DISCRETE,
   [Urban_Rural] LONG DISCRETE


)
WITH HOLDOUT (40 PERCENT or 1000 CASES)
```

Adding the decision tree, neural network, logistic regression and clustering Mining Models to the mining structure created above:

Logistic Regression:

ALTER MINING STRUCTURE [UK ACCIDENTS DMX]

```
ADD MINING MODEL [Logistic Regression]
(
   [Accident_Index],
   [Accident_Severity] PREDICT,
   [First_Road_Class] ,
   [Road_Type]  ,
   [Speed_limit] ,
   [Junction_Detail] ,
   [Junction_Control] ,
   [Second_Road_Class] ,
   [Ped_Cross_Human] ,
   [Ped_Cross_Physical] ,
   [Light_Conditions] ,
   [Weather_Conditions] ,
   [Road_Surface_Conditions] ,
   [Special_Conditions_at_Site] ,
   [Carriageway_Hazards] ,
   [Urban_Rural]

)
USING Microsoft_Logistic_Regression
```

Decision Tree:

```
ALTER MINING STRUCTURE [UK ACCIDENTS DMX]
ADD MINING MODEL [Decision Tree DMX]
(
   [Accident_Index],
   [Accident_Severity] PREDICT,
   [First_Road_Class] ,
   [Road_Type]  ,
   [Speed_limit] ,
   [Junction_Detail] ,
   [Junction_Control] ,
   [Second_Road_Class] ,
   [Ped_Cross_Human] ,
 [Ped_Cross_Physical] ,
    [Light_Conditions] ,
   [Weather_Conditions] ,
   [Road_Surface_Conditions] ,
   [Special_Conditions_at_Site] ,
   [Carriageway_Hazards] ,
   [Urban_Rural]

)
USING Microsoft_Decision_Trees
```

Neural Network:

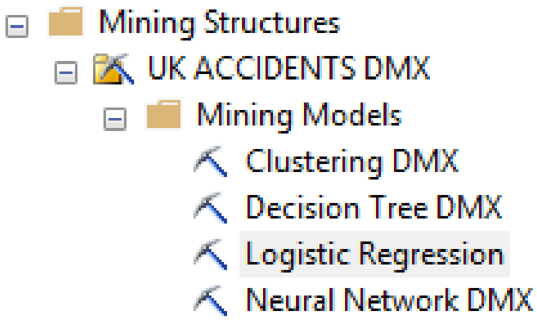
```
ALTER MINING STRUCTURE [UK ACCIDENTS DMX]
ADD MINING MODEL [Neural Network DMX]
(
  [Accident_Index],
  [Accident_Severity] PREDICT,
  [First_Road_Class] ,
  [Road_Type]  ,
  [Speed_limit] ,
  [Junction_Detail] ,
  [Junction_Control] ,
  [Second_Road_Class] ,
  [Ped_Cross_Human] ,
  [Ped_Cross_Physical] ,
  [Light_Conditions] ,
  [Weather_Conditions] ,
  [Road_Surface_Conditions] ,
  [Special_Conditions_at_Site] ,
  [Carriageway_Hazards] ,
  [Urban_Rural]

)
USING Microsoft_Neural_Network
```

Clustering:

```
ALTER MINING STRUCTURE [UK ACCIDENTS DMX]
ADD MINING MODEL [Clustering DMX]
(
  [Accident_Index],
  [Accident_Severity] PREDICT,
  [First_Road_Class] ,
  [Road_Type]  ,
  [Speed_limit] ,
  [Junction_Detail] ,
  [Junction_Control] ,
  [Second_Road_Class] ,
  [Ped_Cross_Human] ,
  [Ped_Cross_Physical] ,
  [Light_Conditions] ,
  [Weather_Conditions] ,
  [Road_Surface_Conditions] ,
  [Special_Conditions_at_Site] ,
  [Carriageway_Hazards] ,
  [Urban_Rural]
)
USING Microsoft_Clustering
```

The deployed mining models look as shown below:



Processing of the model is done by using the following query:

```
INSERT INTO MINING STRUCTURE [UK ACCIDENTS DMX]
(

 [Accident_Index] ,
  [Accident_Severity],
  [First_Road_Class] ,
  [Road_Type] ,
  [Speed_limit] ,
  [Junction_Detail] ,
  [Junction_Control],
  [Second_Road_Class],
  [Ped_Cross_Human] ,
  [Ped_Cross_Physical],
  [Light_Conditions],
  [Weather_Conditions],
  [Road_Surface_Conditions],
  [Special_Conditions_at_Site],
  [Carriageway_Hazards],
  [Urban_Rural]


)
OPENQUERY([UK Accidents Database],
'SELECT top 1000000 Accident_Index,Accident_Severity,First_Road_Class,
Road_Type,Speed_limit,Junction_Detail,Junction_Control,
Second_Road_Class,Ped_Cross_Human,Ped_Cross_Physical,
Light_Conditions,Weather_Conditions,Road_Surface_Conditions,
Special_Conditions_at_Site,Carriageway_Hazards,Urban_Rural
FROM dbo.DimAccidents
order by Accident_Severity')
```

COMPARISON OF ALGORITHMS

Now we assess the performance of the 4 algorithms to decide the best performing algorithm:

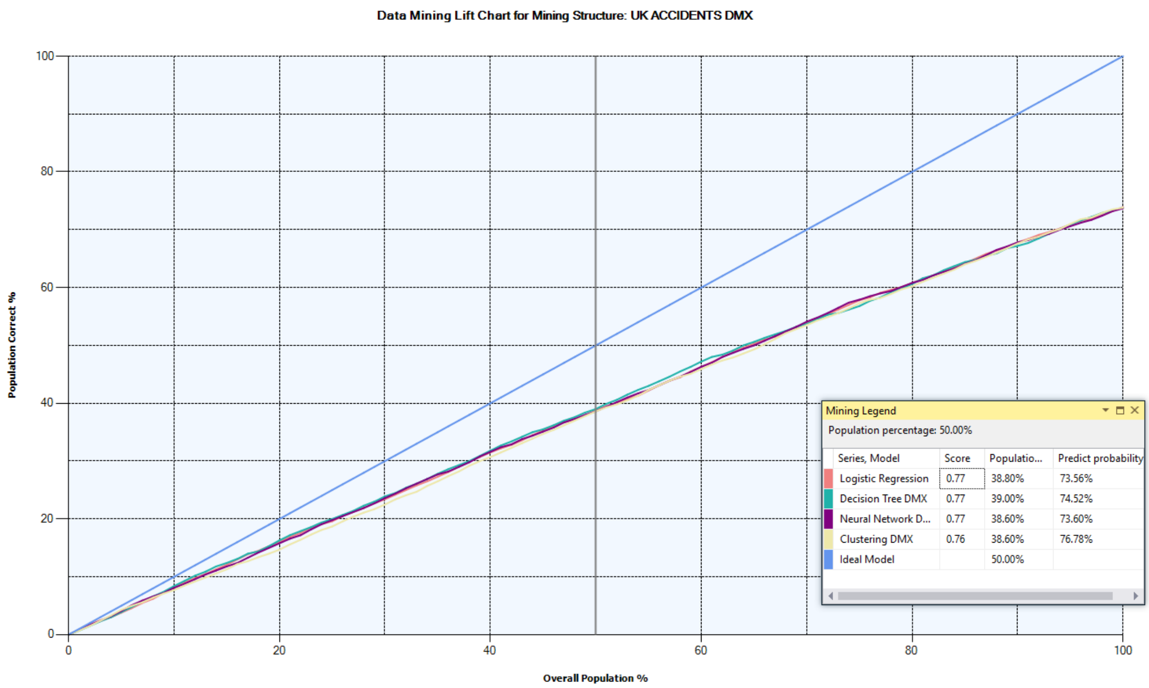
The following query signifies that the data is highly skewed and the models are biased towards the majority class:

```sql
SELECT COUNT(*) AS Accident_Count_By_Severity_Category FROM DimAccidents
GROUP BY Accident_Severity
```

	Accident_Count_By_Severity_Category
1	22998
2	242080
3	1515575

We use lift charts and classification matrices to assess the performance of the algorithms and they are as shown below:

Lift Chart for Accident_Severity:



The Neural Network, Decision Tree, and Logistic Regression performed at par which is signified by the Lift score. However, as seen in the confusion matrix and the data being skewed led to models being heavily biased towards category 3.

Confusion Matrix:

Counts for Logistic Regression on Accident_Severity				
	Predicted	1 (Actual)	2 (Actual)	3 (Actual)
	1	0	0	0
	2	0	1	0
	3	26	235	738
Counts for Decision Tree DMX on Accident_Severity				
	Predicted	1 (Actual)	2 (Actual)	3 (Actual)
	1	0	0	0
	2	0	0	0
	3	26	236	738
Counts for Neural Network DMX on Accident_Severity				
	Predicted	1 (Actual)	2 (Actual)	3 (Actual)
	1	0	0	0
	2	0	1	2
	3	26	235	736
Counts for Clustering DMX on Accident_Severity				
	Predicted	1 (Actual)	2 (Actual)	3 (Actual)
	1	0	0	0
	2	0	0	0
	3	26	236	738

In order to drill down and understand the effects of the predictor variables in predicting **Accident_Severity = 1 (Fatal),** we created a lift chart for the same category for all mining models.

Mining Legend

Population percentage: 50.00%

Series, Model	Score	Targe...	Predi...
Logistic Regression	0.67	65.38...	4.49%
Decision Tree DMX	0.68	73.08...	1.20%
Neural Network D...	0.62	69.23...	3.36%
Clustering DMX	0.61	65.38...	1.66%
Random Guess M...		50.00...	
Ideal Model for: L...		100.0...	

Fig: *Lift Score for Accident_Severity = 1 (Fatal)*

From the above lift scores, we can see that the Decision Tree is the best model in predicting Accident_Severity = 1 (Fatal).

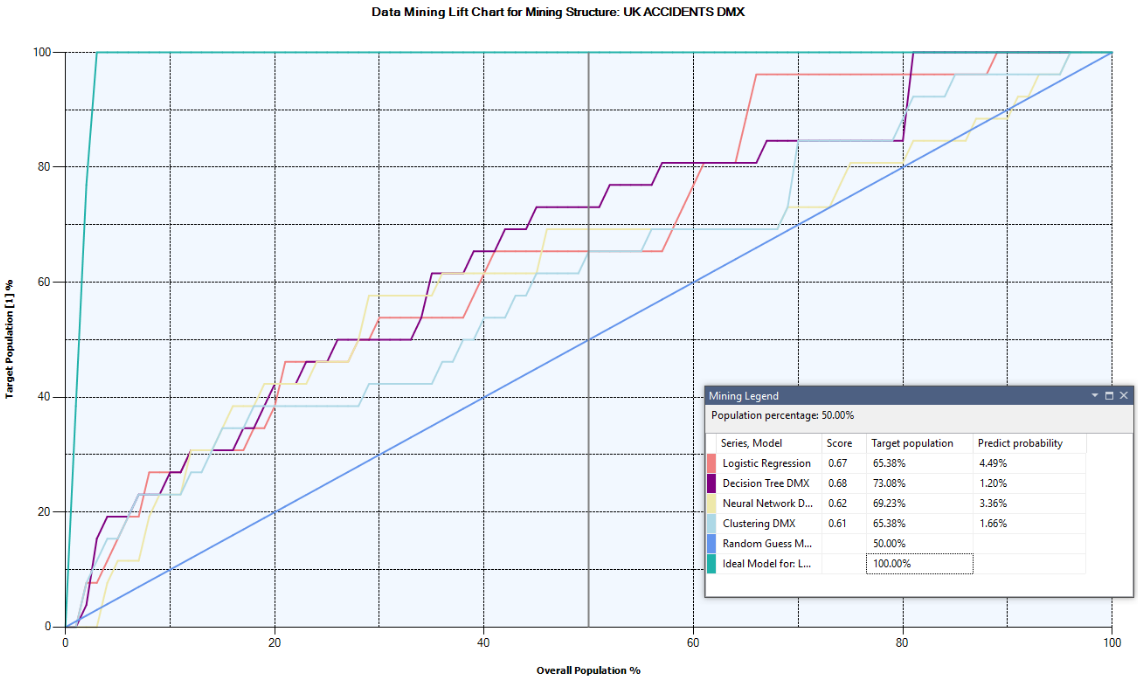


Fig: *Lift Chart for Accident_Severity = 1 (Fatal)*

For category Accident_Severity = 1, there were 22998 records in the data set which consisted of 1760853 total records that is about 1.3% of the population. The Decision Tree prediction model will correctly identify 73.08% (0.7308 * 22998 = 16,807) of all fatal accidents in the population.

Cross Validation:

Fold Count: 5 | Max Cases: 5000 | Get Results

Target Attribute: Accident_Severity | Target State: 1 | Target Threshold:

Logistic Regression

Partition Index	Partition Size	Test	Measure	Value
1	1000	Classification	True Positive	14
2	1000	Classification	True Positive	17
3	1000	Classification	True Positive	12
4	1000	Classification	True Positive	20
5	1000	Classification	True Positive	21
			Average	16.8
			Standard Deviation	3.4293
1	1000	Classification	False Positive	506
2	1000	Classification	False Positive	522
3	1000	Classification	False Positive	685
4	1000	Classification	False Positive	511
5	1000	Classification	False Positive	633
			Average	571.4
			Standard Deviation	73.5734
1	1000	Classification	True Negative	469
2	1000	Classification	True Negative	453
3	1000	Classification	True Negative	290
4	1000	Classification	True Negative	464
5	1000	Classification	True Negative	342
			Average	403.6
			Standard Deviation	73.5734
1	1000	Classification	False Negative	11
2	1000	Classification	False Negative	8
3	1000	Classification	False Negative	13
4	1000	Classification	False Negative	5
5	1000	Classification	False Negative	4
			Average	8.2
			Standard Deviation	3.4293
1	1000	Likelihood	Log Score	-2.4461
2	1000	Likelihood	Log Score	-3.9908
3	1000	Likelihood	Log Score	-3.3755
4	1000	Likelihood	Log Score	-2.8265
5	1000	Likelihood	Log Score	-3.1467
			Average	-3.1571

			Standard Deviation	0.5211
1	1000	Likelihood	Lift	-1.7942
2	1000	Likelihood	Lift	-3.3389
3	1000	Likelihood	Lift	-2.7224
4	1000	Likelihood	Lift	-2.1746
5	1000	Likelihood	Lift	-2.4948
			Average	-2.505
			Standard Deviation	0.521
1	1000	Likelihood	Root Mean Square Error	0.3197
2	1000	Likelihood	Root Mean Square Error	0.2454
3	1000	Likelihood	Root Mean Square Error	0.2592
4	1000	Likelihood	Root Mean Square Error	0.2786
5	1000	Likelihood	Root Mean Square Error	0.281
			Average	0.2768
			Standard Deviation	0.0251

Decision Tree DMX

Partition Index	**Partition Size**	**Test**	**Measure**	**Value**
1	1000	Classification	True Positive	0.000e+000
2	1000	Classification	True Positive	0.000e+000
3	1000	Classification	True Positive	0.000e+000
4	1000	Classification	True Positive	0.000e+000
5	1000	Classification	True Positive	0.000e+000
			Average	0.000e+000
			Standard Deviation	0.000e+000
1	1000	Classification	False Positive	0.000e+000
2	1000	Classification	False Positive	0.000e+000
3	1000	Classification	False Positive	0.000e+000
4	1000	Classification	False Positive	0.000e+000
5	1000	Classification	False Positive	0.000e+000
			Average	0.000e+000
			Standard Deviation	0.000e+000
1	1000	Classification	True Negative	975
2	1000	Classification	True Negative	975
3	1000	Classification	True Negative	975

4	1000	Classification	True Negative	975
5	1000	Classification	True Negative	975
			Average	975
			Standard Deviation	0.000e+000
1	1000	Classification	False Negative	25
2	1000	Classification	False Negative	25
3	1000	Classification	False Negative	25
4	1000	Classification	False Negative	25
5	1000	Classification	False Negative	25
			Average	25
			Standard Deviation	0.000e+000
1	1000	Likelihood	Log Score	-0.6503
2	1000	Likelihood	Log Score	-0.6382
3	1000	Likelihood	Log Score	-0.6445
4	1000	Likelihood	Log Score	-0.6421
5	1000	Likelihood	Log Score	-0.6398
			Average	-0.643
			Standard Deviation	0.0042
1	1000	Likelihood	Lift	0.0016
2	1000	Likelihood	Lift	0.0136
3	1000	Likelihood	Lift	0.0085
4	1000	Likelihood	Lift	0.0098
5	1000	Likelihood	Lift	0.0121
			Average	0.0091
			Standard Deviation	0.0042
1	1000	Likelihood	Root Mean Square Error	0.2705
2	1000	Likelihood	Root Mean Square Error	0.2673
3	1000	Likelihood	Root Mean Square Error	0.2643
4	1000	Likelihood	Root Mean Square Error	0.262
5	1000	Likelihood	Root Mean Square Error	0.2664
			Average	0.2661
			Standard Deviation	0.0029

Neural Network DMX				
Partition Index	**Partition Size**	**Test**	**Measure**	**Value**
1	1000	Classification	True Positive	23

2	1000	Classification	True Positive	25
3	1000	Classification	True Positive	23
4	1000	Classification	True Positive	24
5	1000	Classification	True Positive	24
			Average	23.8
			Standard Deviation	0.7483
1	1000	Classification	False Positive	884
2	1000	Classification	False Positive	956
3	1000	Classification	False Positive	937
4	1000	Classification	False Positive	876
5	1000	Classification	False Positive	836
			Average	897.8
			Standard Deviation	43.3793
1	1000	Classification	True Negative	91
2	1000	Classification	True Negative	19
3	1000	Classification	True Negative	38
4	1000	Classification	True Negative	99
5	1000	Classification	True Negative	139
			Average	77.2
			Standard Deviation	43.3793
1	1000	Classification	False Negative	2
2	1000	Classification	False Negative	0.000e+000
3	1000	Classification	False Negative	2
4	1000	Classification	False Negative	1
5	1000	Classification	False Negative	1
			Average	1.2
			Standard Deviation	0.7483
1	1000	Likelihood	Log Score	-6.7925
2	1000	Likelihood	Log Score	-45.3055
3	1000	Likelihood	Log Score	-9.6432
4	1000	Likelihood	Log Score	-9.7721
5	1000	Likelihood	Log Score	-7.4093
			Average	-15.7845
			Standard Deviation	14.8078
1	1000	Likelihood	Lift	-6.1406
2	1000	Likelihood	Lift	-44.6536
3	1000	Likelihood	Lift	-8.9901
4	1000	Likelihood	Lift	-9.1202
5	1000	Likelihood	Lift	-6.7574
			Average	-15.1324

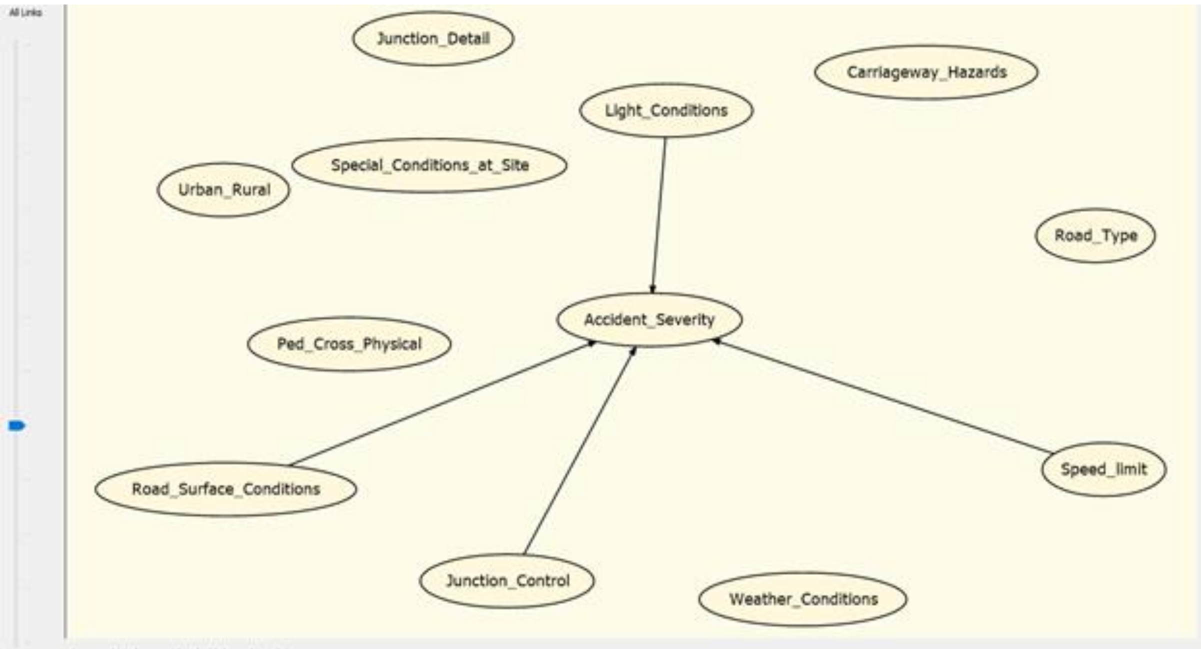
			Standard Deviation	14.8079
1	1000	Likelihood	Root Mean Square Error	0.131
2	1000	Likelihood	Root Mean Square Error	0.0243
3	1000	Likelihood	Root Mean Square Error	0.0987
4	1000	Likelihood	Root Mean Square Error	0.1209
5	1000	Likelihood	Root Mean Square Error	0.1805
			Average	0.1111
			Standard Deviation	0.051

Clustering DMX

Partition Index	**Partition Size**	**Test**	**Measure**	**Value**
1	1000	Classification	True Positive	0.000e+000
2	1000	Classification	True Positive	0.000e+000
3	1000	Classification	True Positive	0.000e+000
4	1000	Classification	True Positive	0.000e+000
5	1000	Classification	True Positive	0.000e+000
			Average	0.000e+000
			Standard Deviation	0.000e+000
1	1000	Classification	False Positive	0.000e+000
2	1000	Classification	False Positive	0.000e+000
3	1000	Classification	False Positive	0.000e+000
4	1000	Classification	False Positive	0.000e+000
5	1000	Classification	False Positive	0.000e+000
			Average	0.000e+000
			Standard Deviation	0.000e+000
1	1000	Classification	True Negative	975
2	1000	Classification	True Negative	975
3	1000	Classification	True Negative	975
4	1000	Classification	True Negative	975
5	1000	Classification	True Negative	975
			Average	975
			Standard Deviation	0.000e+000
1	1000	Classification	False Negative	25
2	1000	Classification	False Negative	25
3	1000	Classification	False Negative	25
4	1000	Classification	False Negative	25

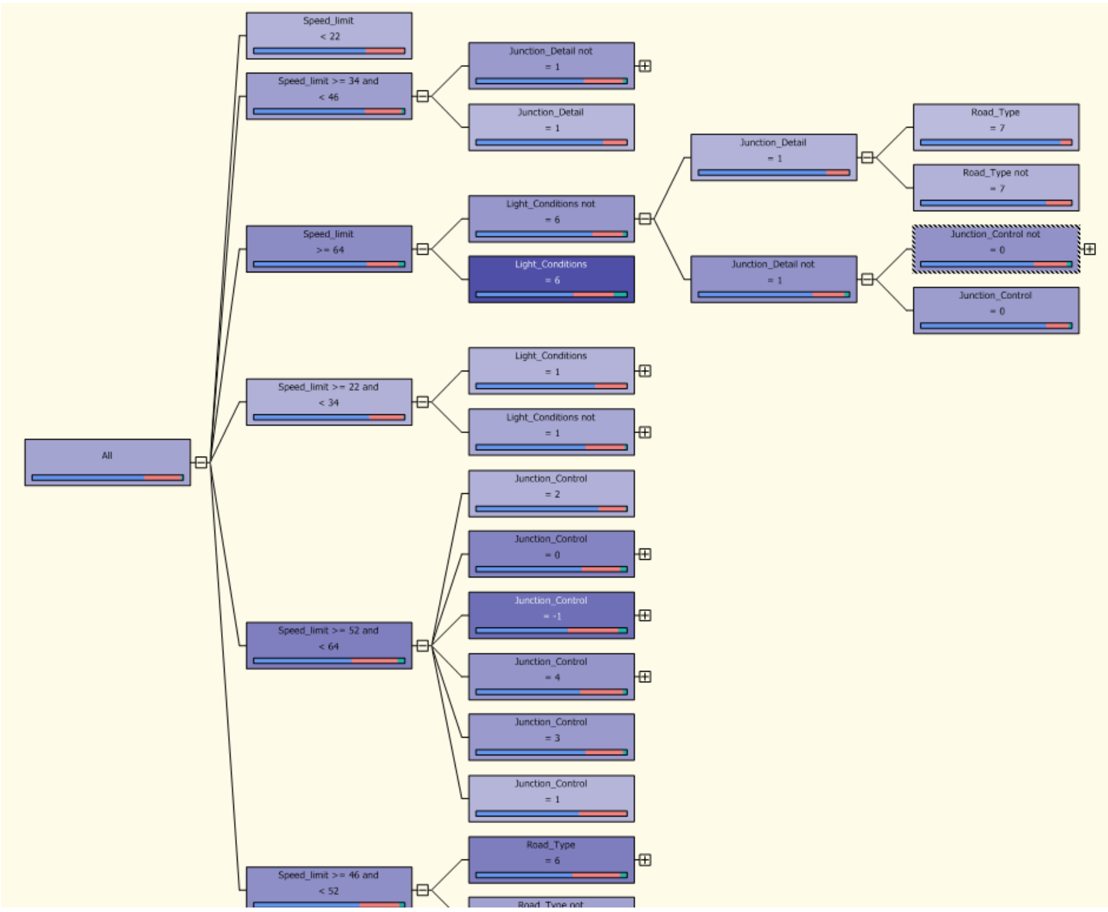
5	1000	Classification	False Negative	25
			Average	25
			Standard Deviation	0.000e+000
1	1000	Likelihood	Log Score	-0.6436
2	1000	Likelihood	Log Score	-0.6366
3	1000	Likelihood	Log Score	-0.6433
4	1000	Likelihood	Log Score	-0.6314
5	1000	Likelihood	Log Score	-0.6412
			Average	-0.6392
			Standard Deviation	0.0046
1	1000	Likelihood	Lift	0.0083
2	1000	Likelihood	Lift	0.0153
3	1000	Likelihood	Lift	0.0098
4	1000	Likelihood	Lift	0.0204
5	1000	Likelihood	Lift	0.0107
			Average	0.0129
			Standard Deviation	0.0044
1	1000	Likelihood	Root Mean Square Error	0.2763
2	1000	Likelihood	Root Mean Square Error	0.2665
3	1000	Likelihood	Root Mean Square Error	0.2637
4	1000	Likelihood	Root Mean Square Error	0.2668
5	1000	Likelihood	Root Mean Square Error	0.2669
			Average	0.268
			Standard Deviation	0.0043

SUMMARY OF THE FINDINGS



From the dependency network, we can see that the most important causing fatal injuries are

– Speed_limit, Light_Conditions, Junction_Control, and Road_Surface_Conditions.



- Speed_limit is greater than or equal to 64 which means vehicles are travelling at high speed, and there is no lighting then chances of Accident_Severity being Fatal are high.
- If Lighting was present and the junction was roundabout irrespective of road type, severity will be fatal.
- If the junction is not roundabout, however, the vehicle was not at junction or within 20 meters of the junction, then severity will be fatal.
- If junction is within 20 meters, and road surface conditions are dry in rural areas, then severity can be fatal.

RECOMMENDATIONS

- Speed limits should vary depending on light conditions.
- Proper junction control at roundabouts is required.
- Lighting should be present at every junction.
- There should be stop sign boards within 20 meters of the junction.
- Rural areas road should be properly maintained by the government to avoid fatal accidents.