Sri Lanka Institute of Information Technology

Data Warehousing and Business Intelligence

(IT3021)

Assignment 2

Hotel Reviews

Submitted by :

IT19374666

Dissanayake Adikaramge D.M.

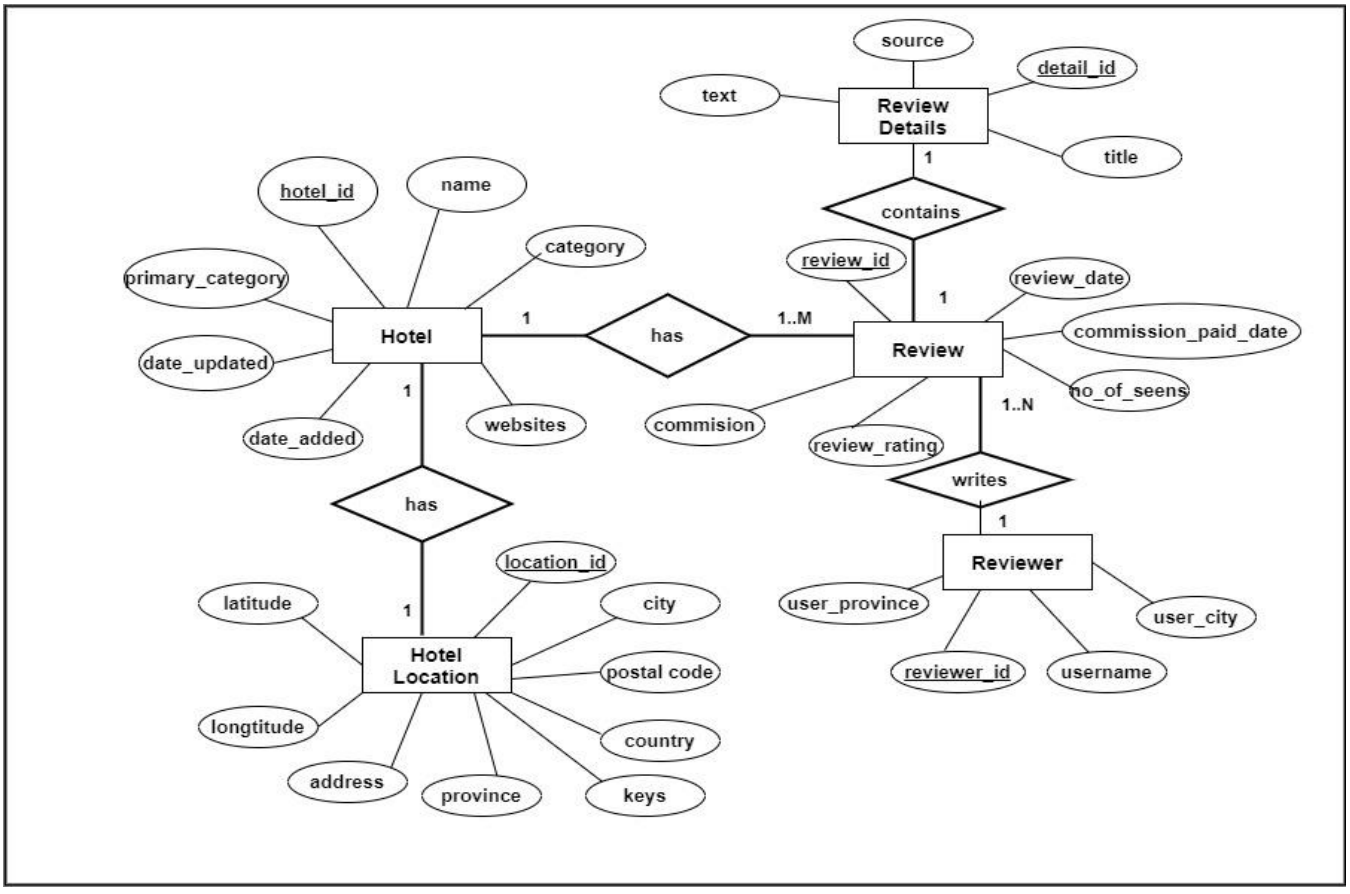
1. Data Source

I selected a hotel review data set for the assignment which includes the ratings by the reviewers, number of reviews seen of a review by the reviewees, the commission paid to the reviewer, the relevant hotel details, and the reviewer details.

This hotel review data set comprises of 10,000 records of hotel reviews over 16 years from 2002 – 2018.

The data set has hierarchies in hotel location such as country-> province -> city and in reviewer entity has user_province -> user_city hierarchy.

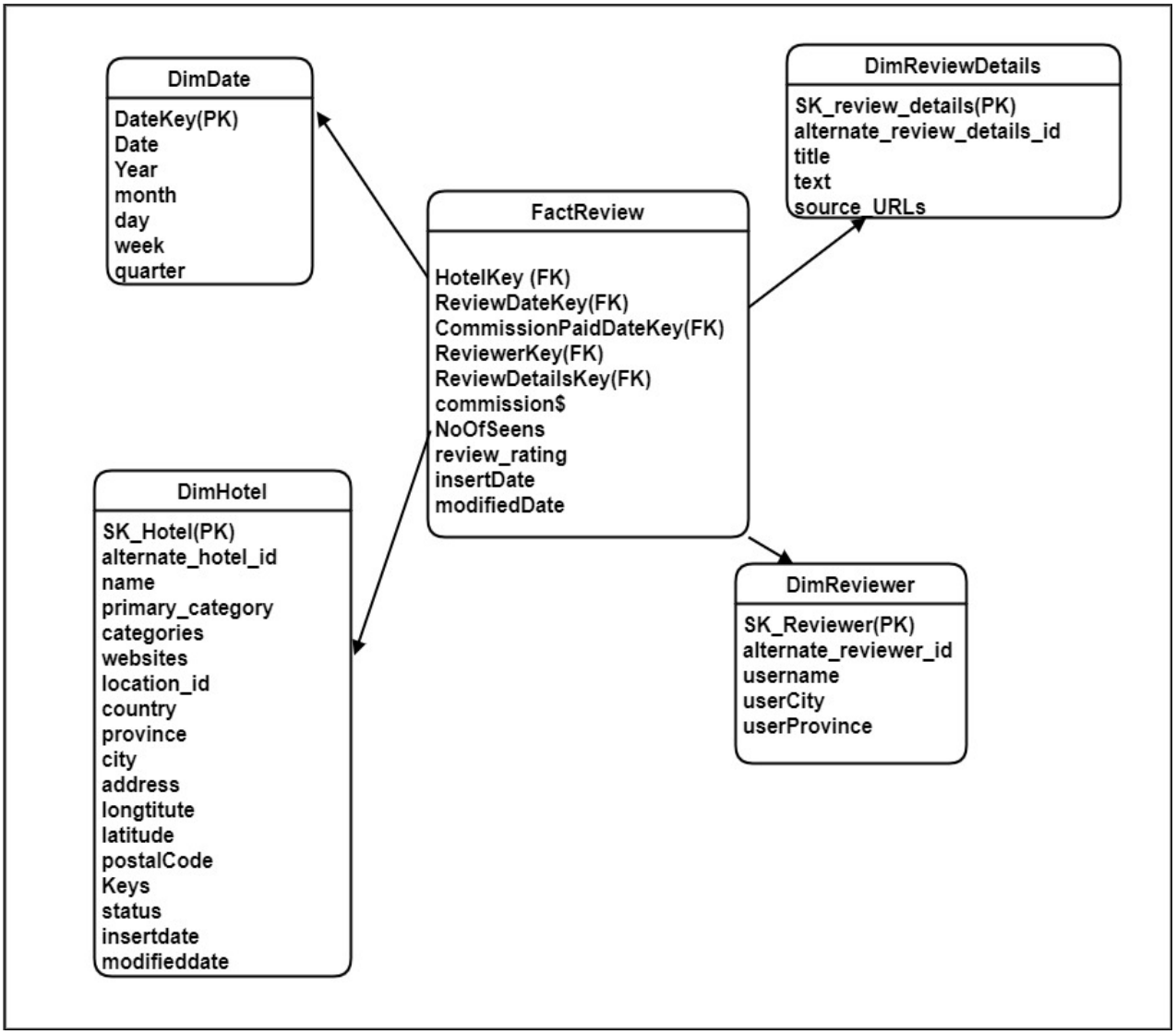
Following is the ER Diagram for the chosen data set.



Data set was downloaded from the following link :

https://www.kaggle.com/datafiniti/hotel-reviews

Dimensional Model



For the data warehouse of the review data set ,I implemented a star schema. DimHotel,DimReviewer , DimReviewDetails and DimDate are dimensions and Review is the fact table in the data warehouse.

I merged the hotel and hotel location tables as shown in the relational model above together to create a single dimension table named DimHotel as shown in the above dimensional model to develop it to a star schema.

Further I implemented the DimHotel dimension as a **slowly changing dimension**.

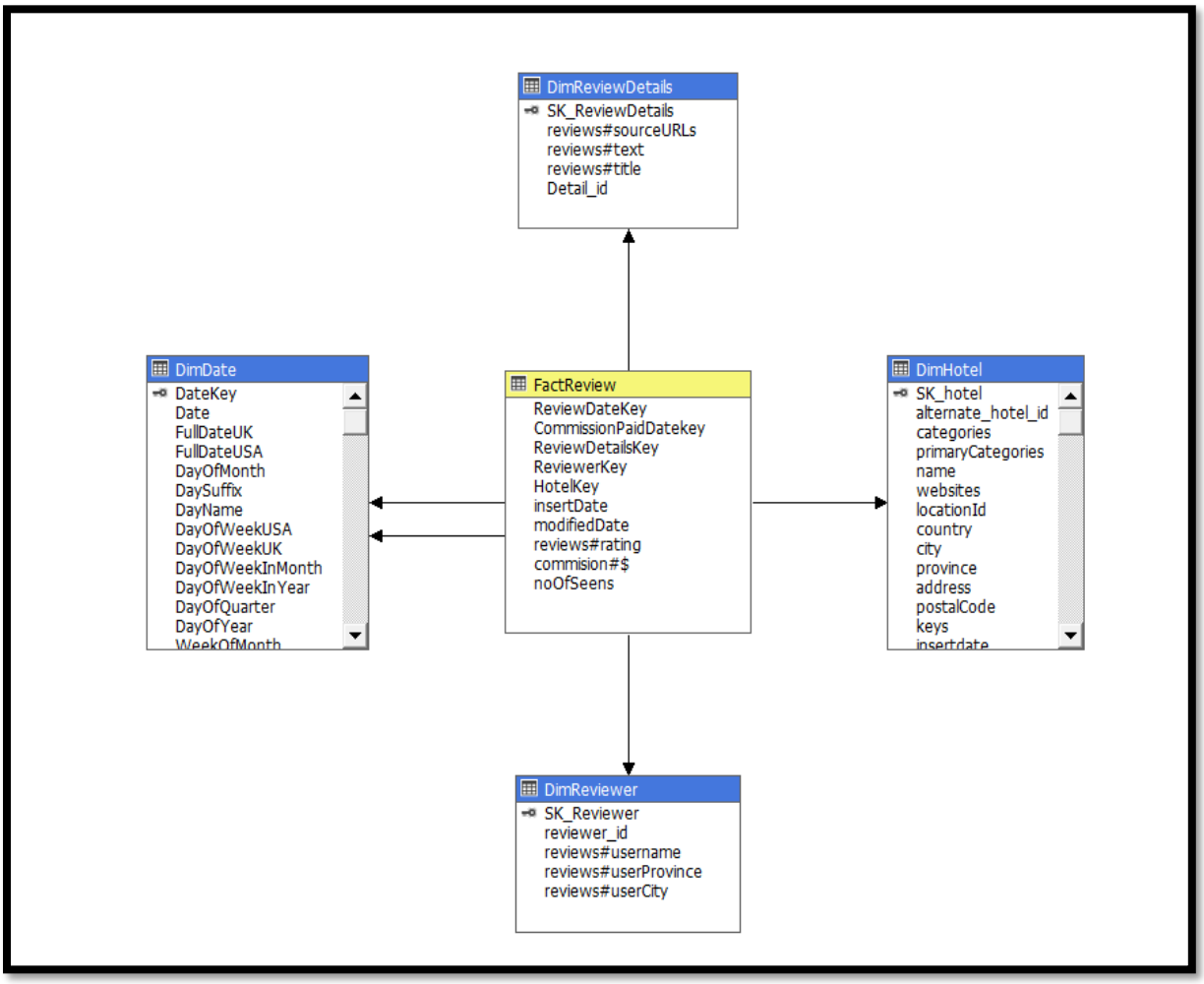
Grain : Details of a review on a specific hotel by a reviewer .

Assumptions:

I decided hotel dimension is a slowly changing dimension assuming hotel name, categories ,primary categories and location details can be changed over time.

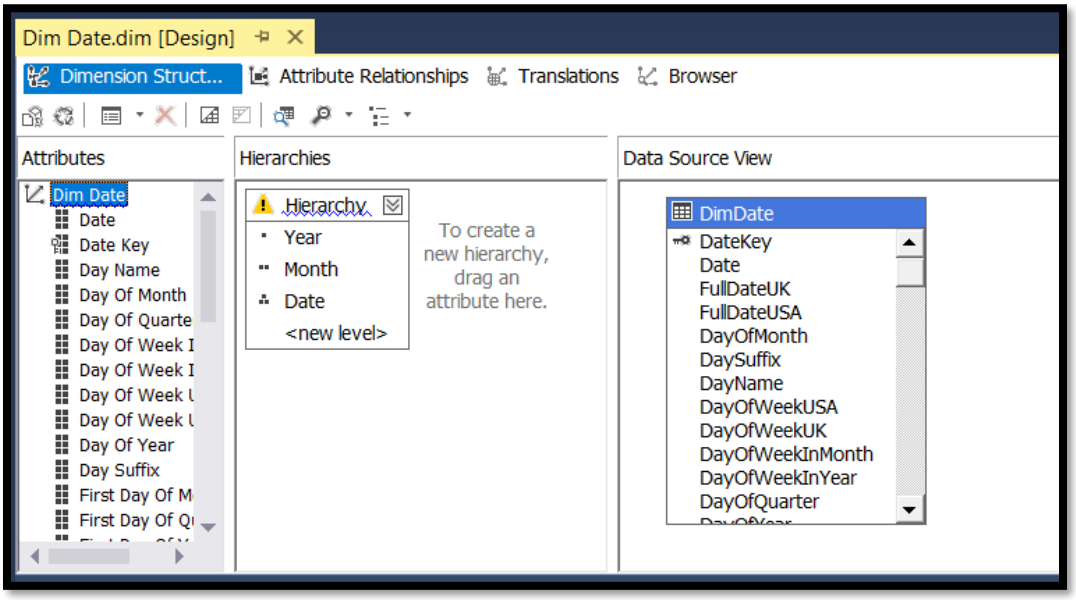
2. SSAS Cube implementation

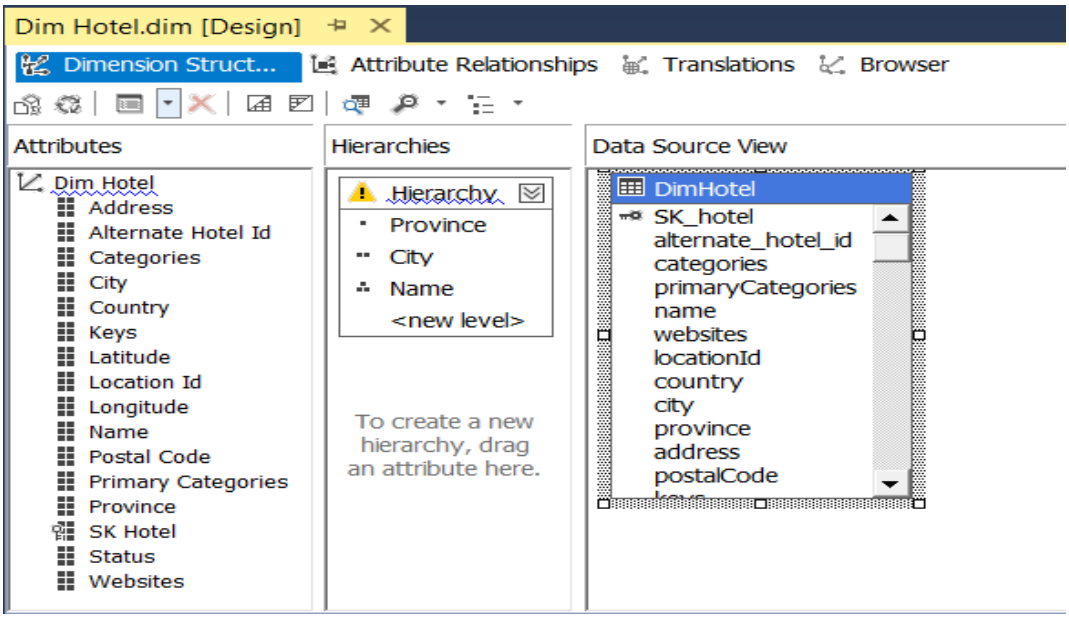
1. Initially I added the 'DataWarehouse_Hotel_Reviews' database as the data source in the SSIS analysis service.
2. Then I created the data source view as follows;



3. I created a cube for the data source using all the dimensions (DimDate, DimReviewer, DimHotel and DimReviewDetails) and FactReview as the measure group table with all the measures in the Fact table as measures.

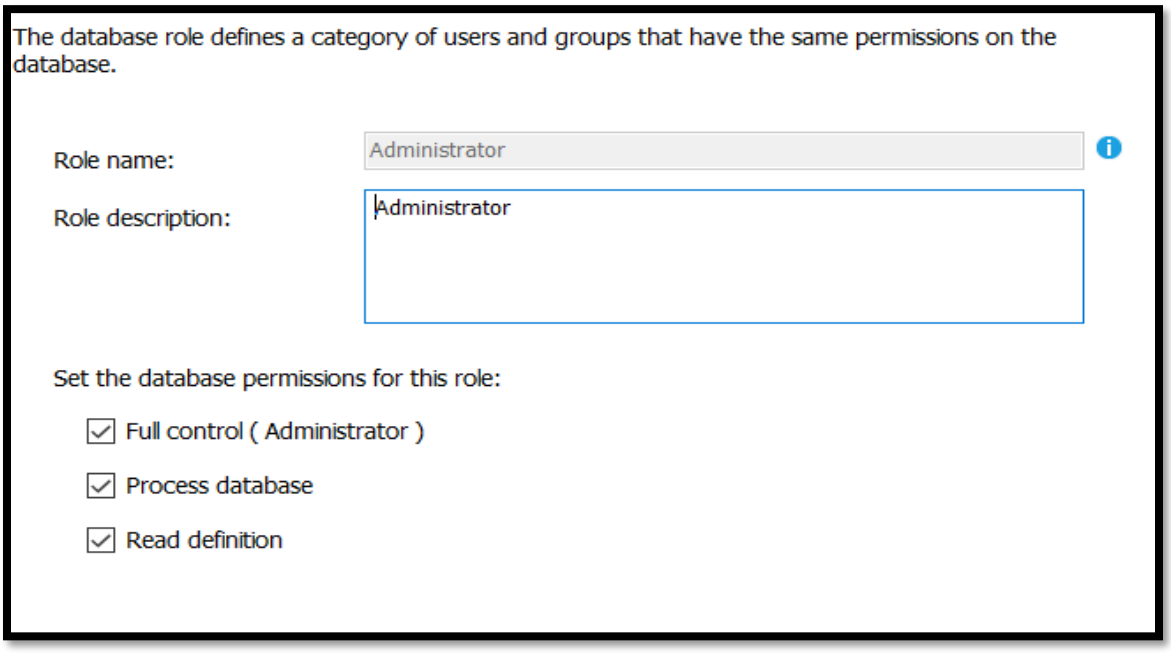
4. Then I added a hierarchy for the DimDate as Year>Month>Date and a hierarchy for the DimHotel as Province > City >Name.



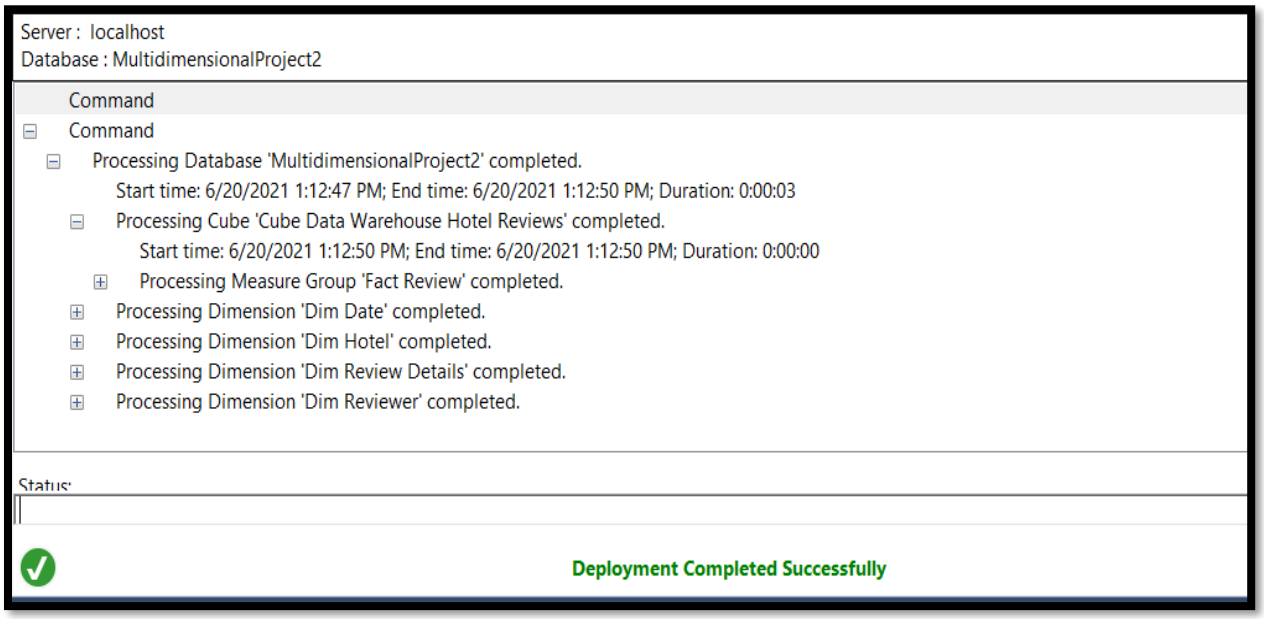


5. I created a KPI to measure the review counts to be greater than 15.

6. I created a role ‘Administrator’ giving all the access and permission.

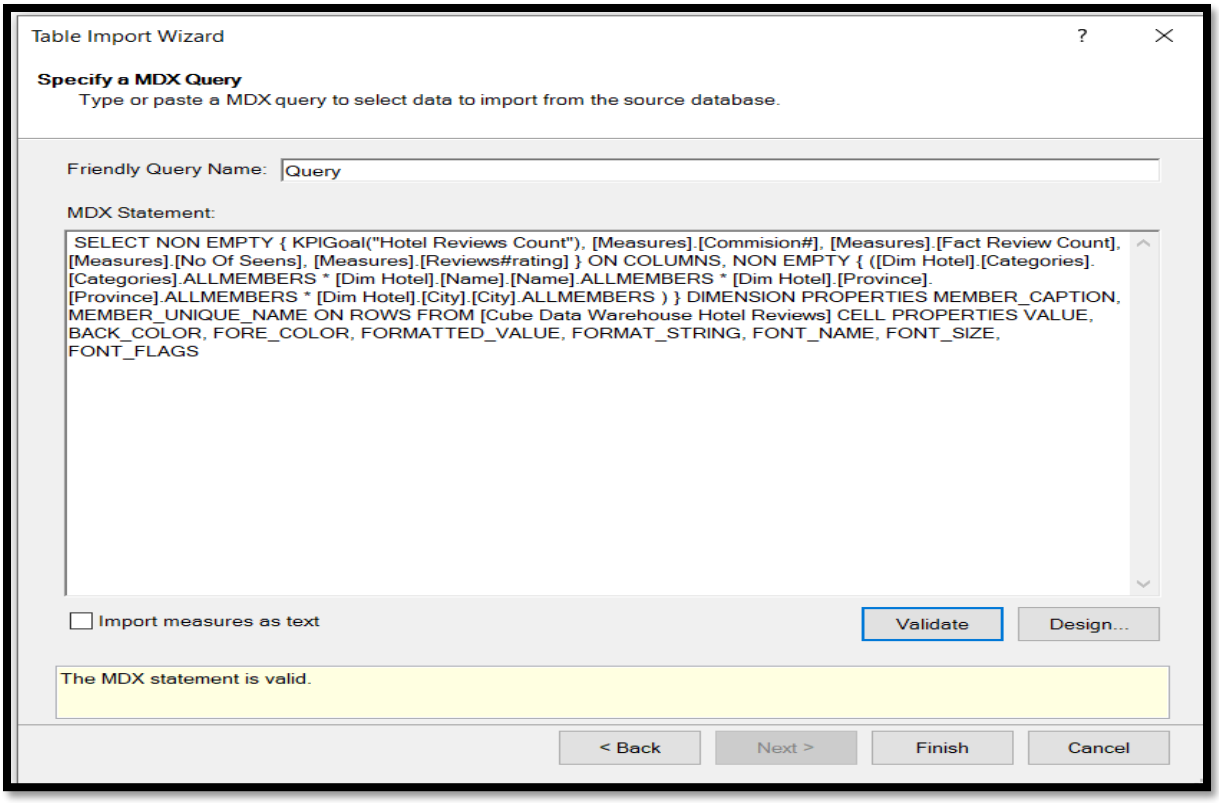


7. I deployed the cube as a SSAS database.

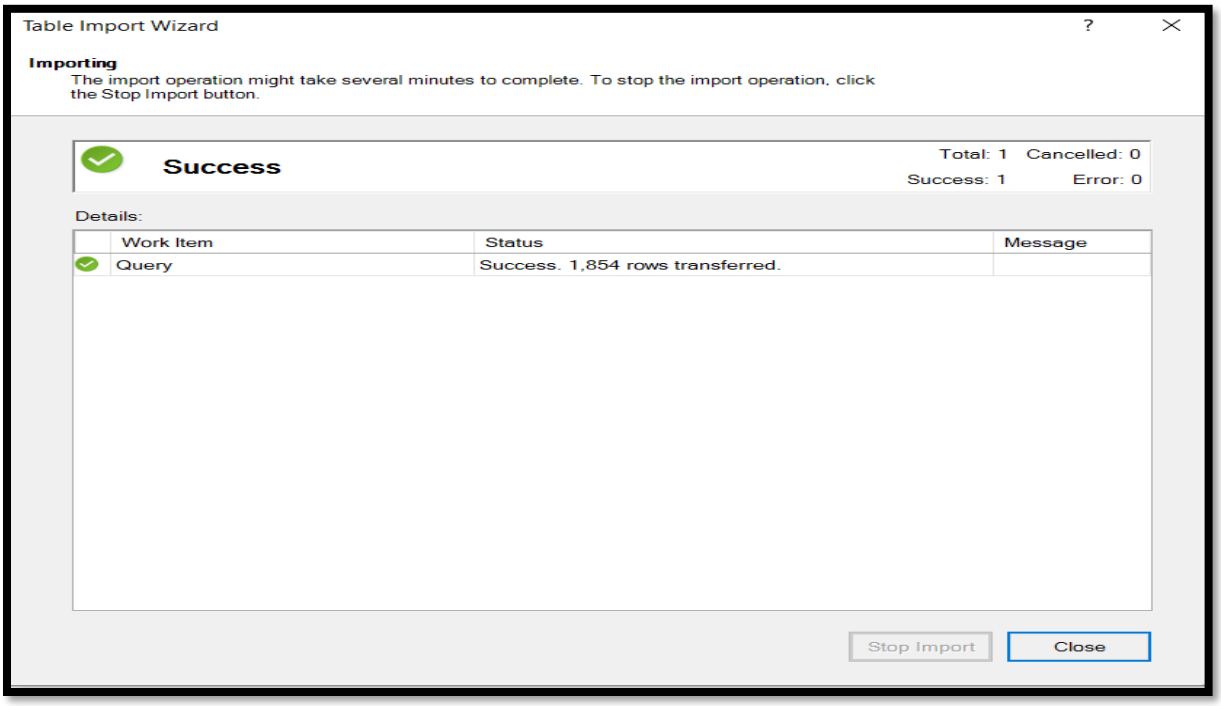


3. Demonstration of OLAP operations

1. Initially I connected Excel PowerPivot to the cube using MDX query.
   I used the following MDX query for the OLAP drill down and roll up operations.

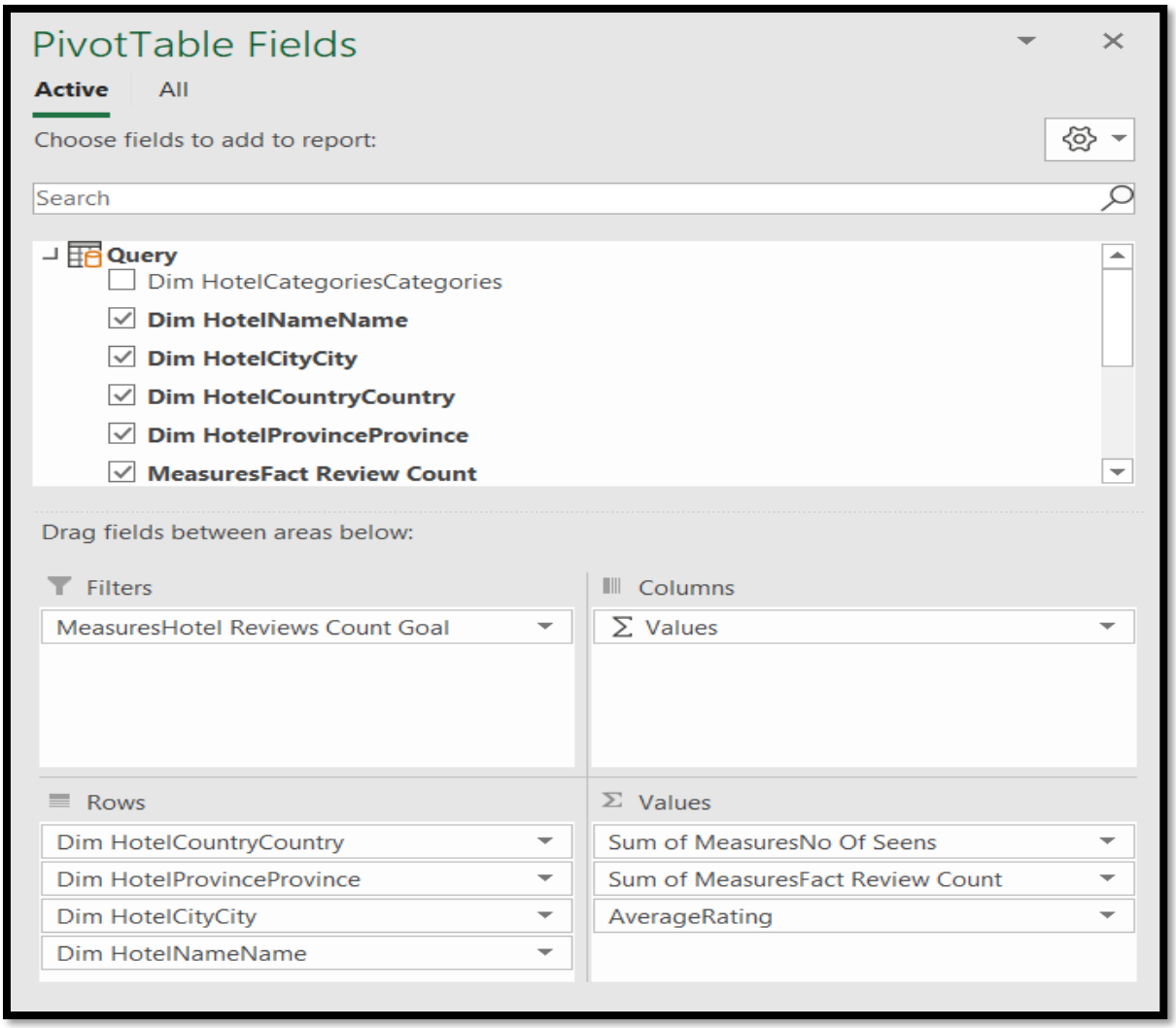


Then I successfully loaded the query data to the PowerPivot backend.



Drill down

I created an pivot Table to analyze count of views, count of reviews and average rating for Hotel where we can drill down from the Country to Province to City to Hotel.



Following is the table that created for the drill down operation.

Row Labels	Sum of MeasuresNo Of Seens	Sum of MeasuresFact Review Count	AverageRating
US			
AZ			
Phoenix	650	80	3.8
Sedona			
Best Western Plus Arroyo Roble Hotel & Creekside Villas	144	19	4.052631579
CA			
Anaheim	270	38	4.078947368
Camarillo	175	24	4.416666667
Long Beach			
Hotel Maya - a DoubleTree by Hilton Hotel	171	25	4.24
Los Angeles			
Miyako Hotel Los Angeles	133	20	4.4675
Napa			
Best Western Plus Inn At The Vines	573	74	4.148648649
Hampton Inn and Suites Napa	216	25	4.32
Pismo Beach			
Seacrest Oceanfront Hotel	291	41	4.512195122
San Diego			
Best Western Seven Seas	129	19	3.526315789
Fairmont Grand Del Mar	115	16	4.875
San Francisco			
Hotel Zetta San Francisco	524	63	4.350793651
InterContinental San Francisco	241	30	4.633333333
Santa Monica			
Shore Hotel	292	32	4.0625
South Lake Tahoe			
Inn by the Lake	189	28	3.928571429
West Hollywood			

Roll up.

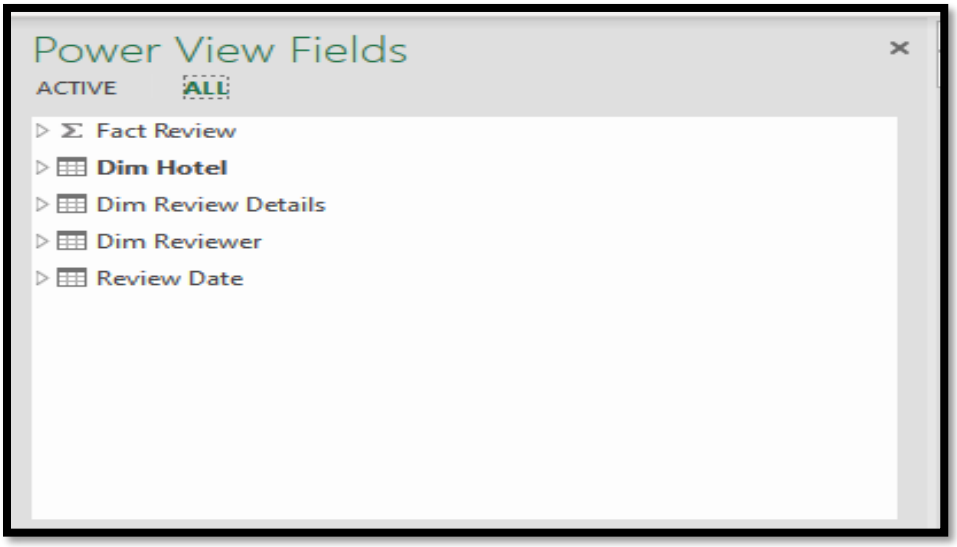
I used the above created table to perform the roll up operation for the hotel review details analysis.

Row Labels	Sum of MeasuresNo Of Seens	Sum of MeasuresFact Review Count	AverageRating
⊟US			
⊞AZ	794	99	3.848484848
⊞CA	3532	464	4.270366379
⊞DE	221	31	3.193548387
⊞FL	2898	401	4.196134663
⊞GA	701	97	4.144329897
⊞HI	1537	196	4.397959184
⊞ID	147	18	4.5
⊞IL	2376	301	4.346843854
⊞KY	176	22	3.954545455
⊞LA	933	122	4.418032787
⊞MA	1664	227	4.256167401
⊞MD	6314	834	3.881354916
⊞ME	137	18	4.444444444
⊞MO	329	45	4.044444444
⊞NC	180	25	4.76
⊞NJ	421	53	3.301886792
⊞NV	3590	499	4.158116232
⊞NY	384	51	4.529411765
⊞OR	182	23	4.347826087
⊞PA	563	73	4.506849315
⊞SC	1191	151	4.40397351
⊞TN	275	36	4.083333333
⊞TX	156	18	4.111111111
⊞VA	5734	788	3.981535533
⊞WA	667	93	4.279569892
Grand Total	35102	4685	4.127865528

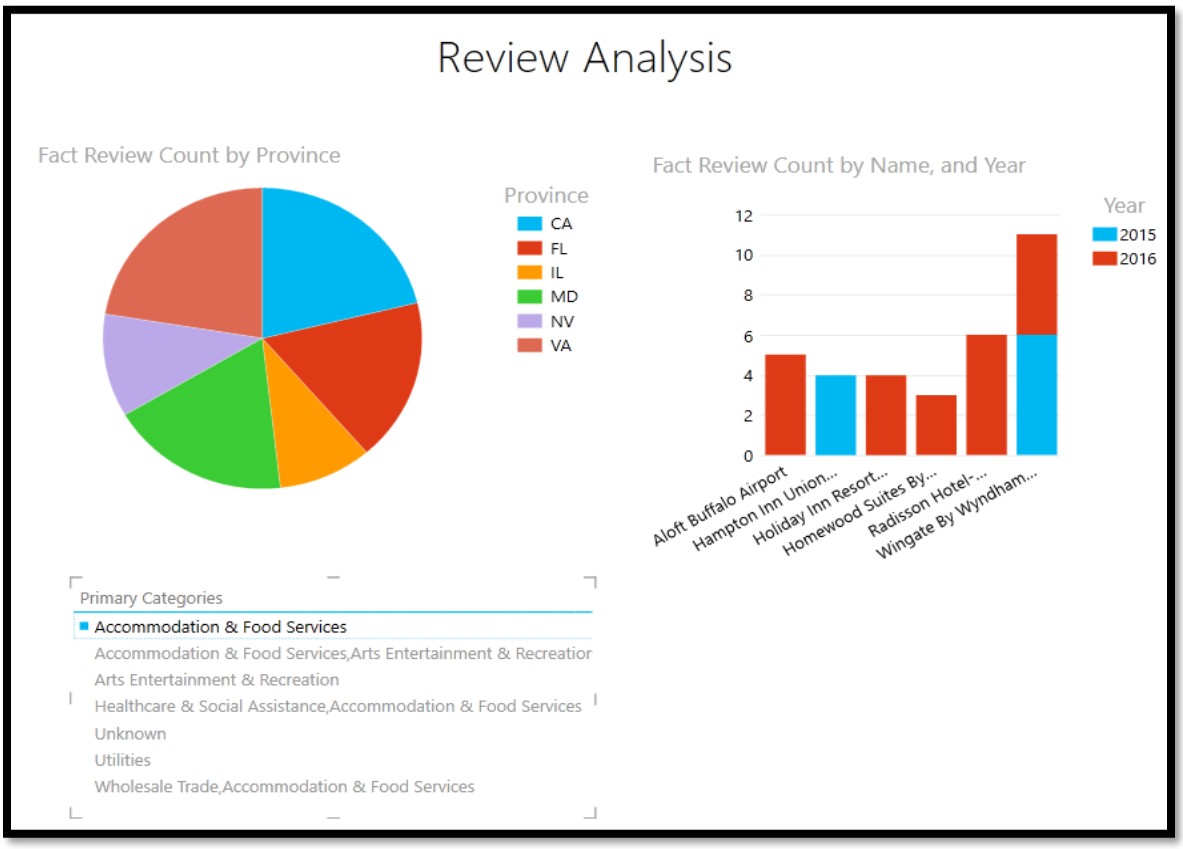
We can analyze province wise review details for hotels using above table.

Slice

1. I initially connected Excel Power View to the cube.

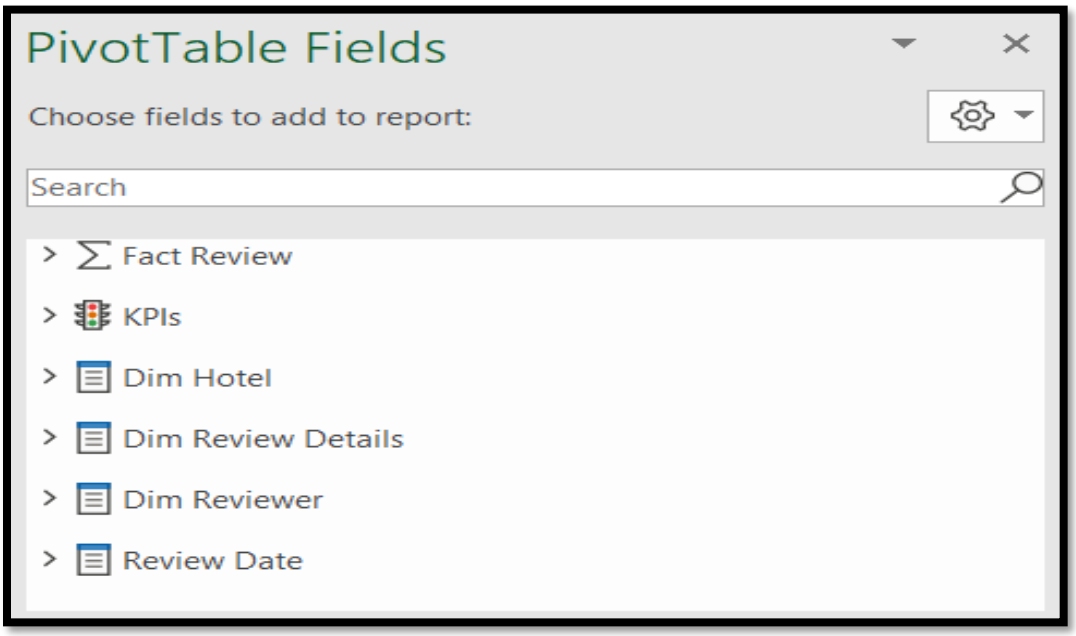


2. Then I added a pie chart to analyze province wise review counts and stacked bar chart to analyze hotel wise review counts for 2015 and 2016.
3. Finally, I added a slicer to analyze primary category wise review counts.



Dice

1. I initially connected PowerPivot to the implemented cube.



1. I used Hotel Dimension to analyze review count, view count based on the selected provinces as below and Date Dimension to analyze the same measures from year 2013-2018.

	Colum													
	Fact Review Count						No Of Seens						Total Fact Review Count	Total No Of Seens
Row Labels	2013	2014	2015	2016	2017	2018	2013	2014	2015	2016	2017	2018		
AK	3		8	8	1		37		55	57	5		20	154
AR	1	5	10	11	5		1	27	70	115	35		32	248
AZ	8	10	46	107	25	9	71	56	388	839	198	107	205	1659
CA	29	69	359	431	66	41	231	554	2668	3383	497	278	995	7611
CO	2	6	48	69	6		15	49	320	527	38		131	949
CT		1	13	15				1	85	133			29	219
DE	8	6	11	18			60	48	64	141			43	313
FL	8	14	261	464	91	34	60	108	1865	3396	670	288	872	6387
GA	7	8	106	131	13	6	51	79	792	970	82	28	271	2002
Grand Total	**66**	**119**	**862**	**1254**	**207**	**90**	**526**	**922**	**6307**	**9561**	**1525**	**701**	**2598**	**19542**

Pivot

1. From the connected cube I implemented the following pivot table.
   It depicts the review count and view counts based on year and province.

Row Labels	Fact Review Count	No Of Seens
2013		
AK	3	37
AR	1	1
AZ	8	71
CA	29	231
CO	2	15
DE	8	60
FL	8	60
GA	7	51
HI	13	99
ID	1	4
IL	12	98
KS	2	6
MA	1	1
MD	126	949
ME	1	3
MI	3	24
MO	6	43
NC	3	34
ND	1	1
NH	3	12
NJ	6	64
NV	8	55
NY	2	15
OH	3	30
OK	1	2
OR	3	13
PA	2	14
SC	22	162
SD	1	5
TX	2	8
VA	104	746
WA	2	26
WI	2	19
2014		
AR	5	27
AZ	10	56
CA	69	554
CO	6	49
CT	1	1

2. I changed the year row to column for better analysis of the measures.

Row Labels	Column Fact Review Count						No Of Seens						Total Fact Review Count	Total No Of Seens
	2013	2014	2015	2016	2017	2018	2013	2014	2015	2016	2017	2018		
AK	3		8	8	1		37		55	57	5		20	154
AR	1	5	10	11	5		1	27	70	115	35		32	248
AZ	8	10	46	107	25	9	71	56	388	839	198	107	205	1659
CA	29	69	359	431	66	41	231	554	2668	3383	497	278	995	7611
CO	2	6	48	69	6		15	49	320	527	38		131	949
CT		1	13	15				1	85	133			29	219
DE	8	6	11	18			60	48	64	141			43	313
FL	8	14	261	464	91	34	60	108	1865	3396	670	288	872	6387
GA	7	8	106	131	13	6	51	79	792	970	82	28	271	2002
HI	13	21	83	41	1	1	99	174	690	297	15	15	160	1290
IA		1	18	23	7			10	144	176	36		49	366
ID	1	2	18	30	4	1	4	19	148	182	43	14	56	410
IL	12	16	109	175	93	33	98	111	914	1377	722	250	438	3472
IN		8	38	42	8			51	237	300	62		96	650
KS	2	3	22	27	4		6	20	172	208	25		58	431
KY			33	83	7	1			246	696	65	10	124	1017
LA		3	32	115	50	35		34	241	839	386	265	235	1765
MA	1	10	81	124	83	35	1	81	570	960	552	242	334	2406
MD	126	126	207	139	85	18	949	971	1581	1089	600	152	701	5342
ME	1	6	38	27	2		3	36	289	237	15		74	580
MI	3	8	33	52	1		24	52	249	411	10		97	746
MN		1	15	24	1			11	98	206	9		41	324
MO	6	13	50	65	6		43	82	348	482	47		140	1002
MS			8	10	2				42	75	17		20	134
MT		1	35	17				12	283	105			53	400
NC	3	6	30	73			34	48	191	521			112	794
ND	1	2	7	6	1	1	1	12	66	51	8	6	18	144
NE		3	12	29	1			23	74	230	9		45	336
NH	3	4	15	16			12	27	140	80			38	259
NJ	6	8	39	51	6	2	64	74	258	351	46	12	112	805
NM		1	19	46	3			1	144	363	23		69	531
NV	8	23	38	227	191	31	55	142	311	1676	1367	252	518	3803
NY	2	15	54	111	11	17	15	98	423	873	81	137	210	1627

4. SSRS Reports

Report with a matrix

1. In report builder ,I created a data source connection to 'HotelReviewDataWarehouse' which is embedded to the report.

2. Then I created a data set using the following SQL query.

```
SELECT
  DimDate.[Date]
  ,DimDate.[Month]
  ,DimDate.[Year]
  ,DimHotel.city
  ,DimHotel.country
  ,DimHotel.province
  ,DimHotel.name
  ,DimHotel.categories
  ,DimHotel.primaryCategories
  ,DimReviewer.reviews#userProvince
  ,DimReviewer.reviews#userCity
  ,FactReview.reviews#rating
  ,FactReview.[commision#$]
  ,FactReview.noOfSeens
FROM
  DimDate
  INNER JOIN FactReview
    ON DimDate.DateKey = FactReview.ReviewDateKey
  INNER JOIN DimHotel
    ON DimHotel.SK_hotel = FactReview.HotelKey
  INNER JOIN DimReviewer
    ON DimReviewer.SK_Reviewer = FactReview.ReviewerKey
```

The following matrix report demonstrates province wise average review rating and view count for years(2015-2018); year demonstrates in columns and province in rows.

Province wise Hotel Review Details between 2015 -2018

	2015		2016		2017		2018		Total	
province	reviews rating	no Of Seens	reviews rating	no Of Seens	reviews rating	no Of Seens	reviews rating	no Of Seens	reviews rating	no Of Seens
AK	3.875	55	2.875	57	3	5			3.3333333333333	157
AR	4.6	70	4.0909090909090	115	3	35			3.7878787878787	251
AZ	3.8043478260869	388	3.8878504672897	839	4.48	198	4.5555555555556	107	3.9339622641509	1725
CA	4	2668	4.0013921113689	3383	4.1795454545454	497	4.2914634146341	278	4.0243589743589	8008
CO	4.0833333333333	320	4.1159420289855	527	4.6666666666667	38			4.0746268656716	965
CT	4.0769230769230	85	4.5333333333333	133					4.3	226
DE	3.6363636363636	64	3.7222222222222	141					3.3214285714285	398
FL	3.9670498084291	1865	4.1051724137931	3396	4.2098901098901	670	3.9661764705882	288	4.0619726339794	6424
GA	3.7924528301886	792	4.0992366412213	970	4.0769230769230	82	2.3333333333333	28	3.8612099644128	2053
HI	4.3734939759036	690	4.1219512195122	297	5	15	4	15	4.3584905660377	1657
IA	3.6666666666667	144	3.7391304347826	176	4	36			3.7450980392156	383
ID	3.9444444444444	148	4.7666666666666	182	3.5	43	5	14	4.3392857142857	410

Paramiterized report

1. Initially I created a new data set from following SQL query.

```
SELECT
  DimDate.[Date]
  ,DimDate.[Month]
  ,DimDate.[Year]
  ,DimHotel.city
  ,DimHotel.country
  ,DimHotel.province
  ,DimHotel.name
  ,DimHotel.categories
  ,DimHotel.primaryCategories
  ,DimReviewer.reviews#userProvince
  ,DimReviewer.reviews#userCity
  ,FactReview.reviews#rating
  ,FactReview.[commision#$]
  ,FactReview.noOfSeens
FROM
  DimDate
  INNER JOIN FactReview
    ON DimDate.DateKey = FactReview.ReviewDateKey
  INNER JOIN DimHotel
    ON DimHotel.SK_hotel = FactReview.HotelKey
  INNER JOIN DimReviewer
    ON DimReviewer.SK_Reviewer = FactReview.ReviewerKey

where DimHotel.name in (@Hotel)
```

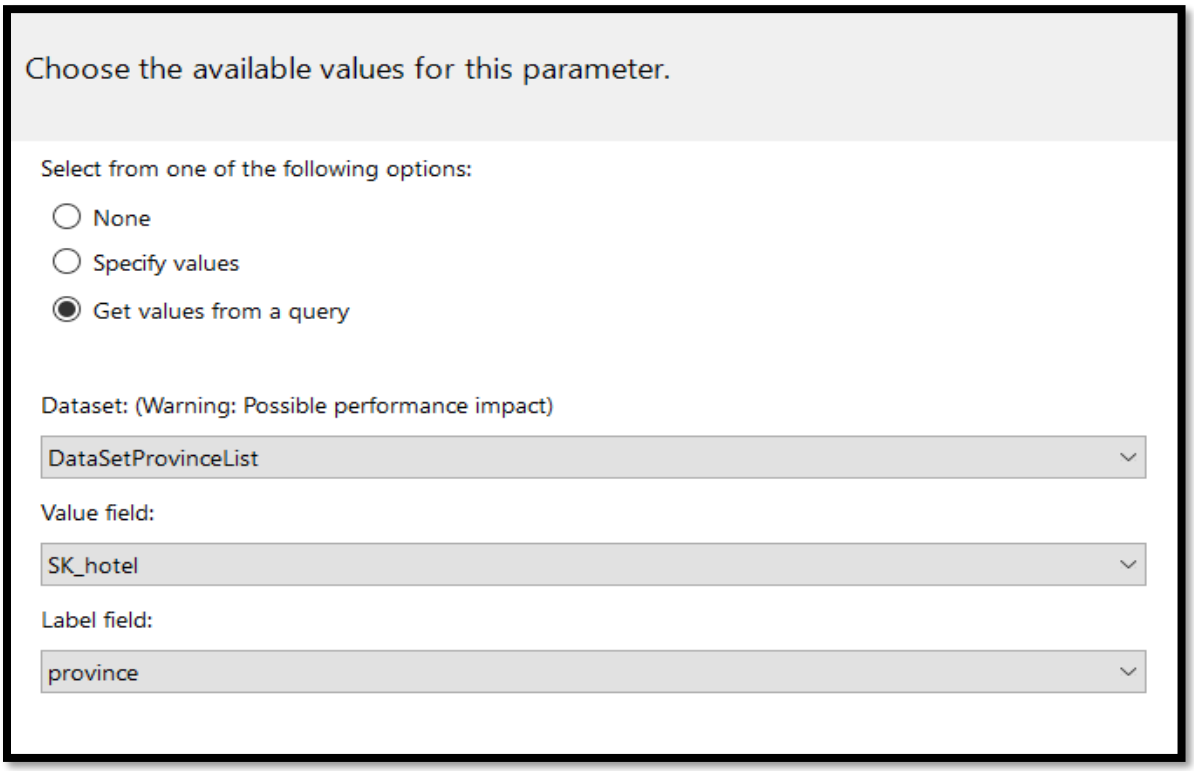
2. Then I added a new data set for Province details.

```
SELECT
  DimHotel.SK_hotel
  ,DimHotel.alternate_hotel_id
  ,DimHotel.province
FROM
  DimHotel
```

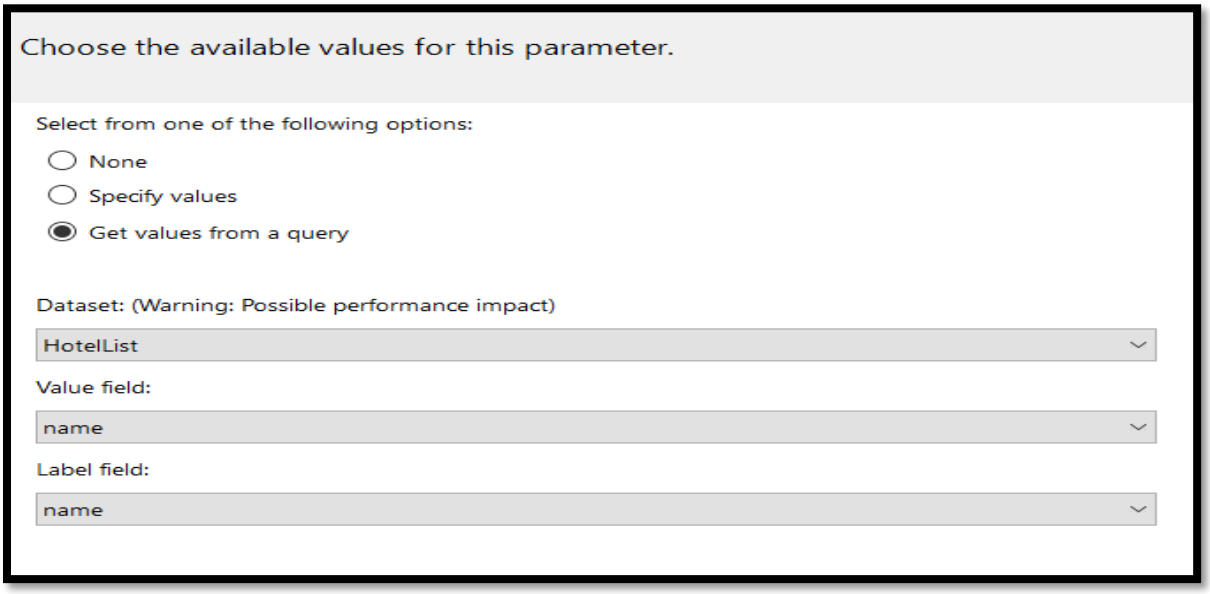
3. Next I added a data set for Hotel details.

```
SELECT
  DimHotel.SK_hotel
  ,DimHotel.alternate_hotel_id
  ,DimHotel.province
  ,DimHotel.name
FROM
  DimHotel
where  DimHotel.SK_hotel in (@Province)
```

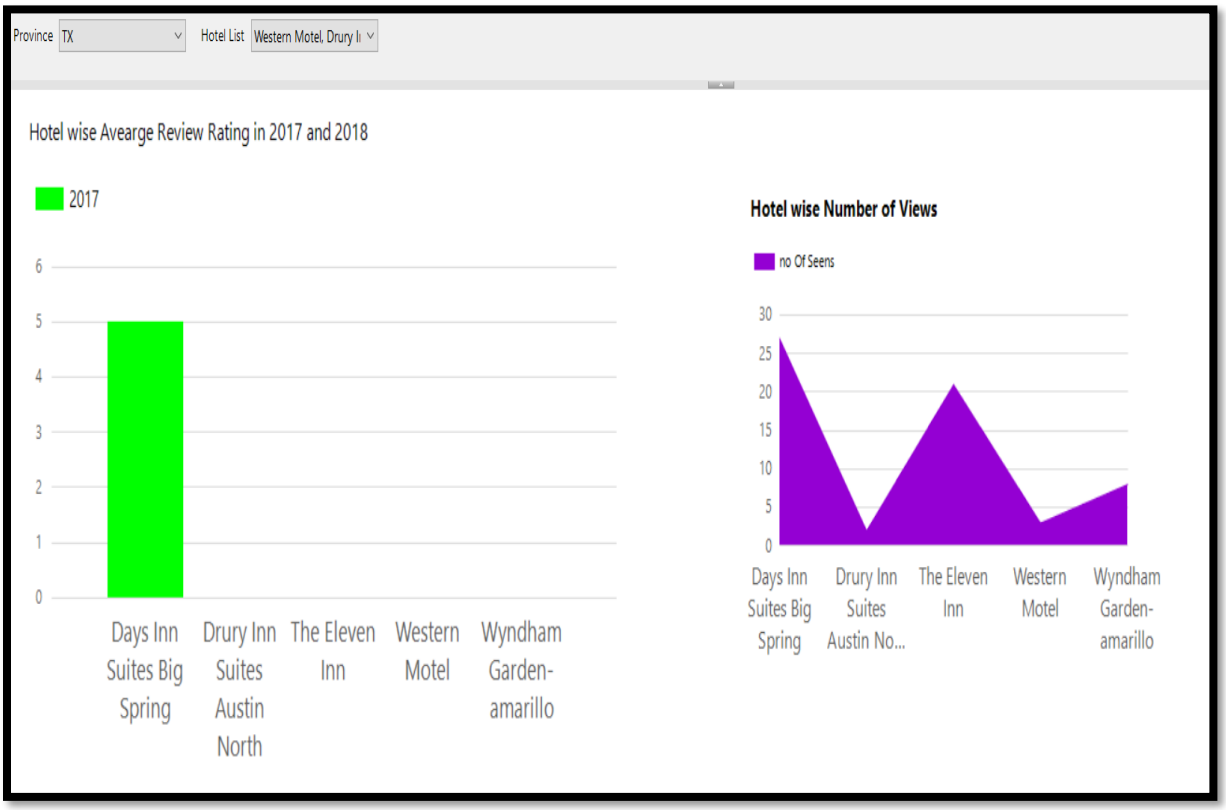
4. I added a parameter for Province list as follows to accept multiple values.



5. Then I added a parameter for Hotel list as follows to accept multiple values.



Following report consists of a bar char for hotel wise average rating for 2017 and 2018 and an area graph for hotel wise views when province and hotel names are passed as parameters.



Drill-down report.

1. Initially I created a new data set from following SQL query.

```
SELECT
  DimDate.[Date]
  ,DimDate.[Month]
  ,DimDate.[Year]
  ,DimHotel.city
  ,DimHotel.country
  ,DimHotel.province
  ,DimHotel.name
  ,DimHotel.categories
  ,DimHotel.primaryCategories
  ,DimReviewer.reviews#userProvince
  ,DimReviewer.reviews#userCity
  ,FactReview.reviews#rating
  ,FactReview.[commision#$]
  ,FactReview.noOfSeens
FROM
  DimDate
  INNER JOIN FactReview
    ON DimDate.DateKey = FactReview.ReviewDateKey
  INNER JOIN DimHotel
    ON DimHotel.SK_hotel = FactReview.HotelKey
  INNER JOIN DimReviewer
    ON DimReviewer.SK_Reviewer = FactReview.ReviewerKey
```

Following report depicts hotel wise average review rating and view count which is drilled down from province for 2015-2018.

Province wise Hotel Review Details between 2015 -2018

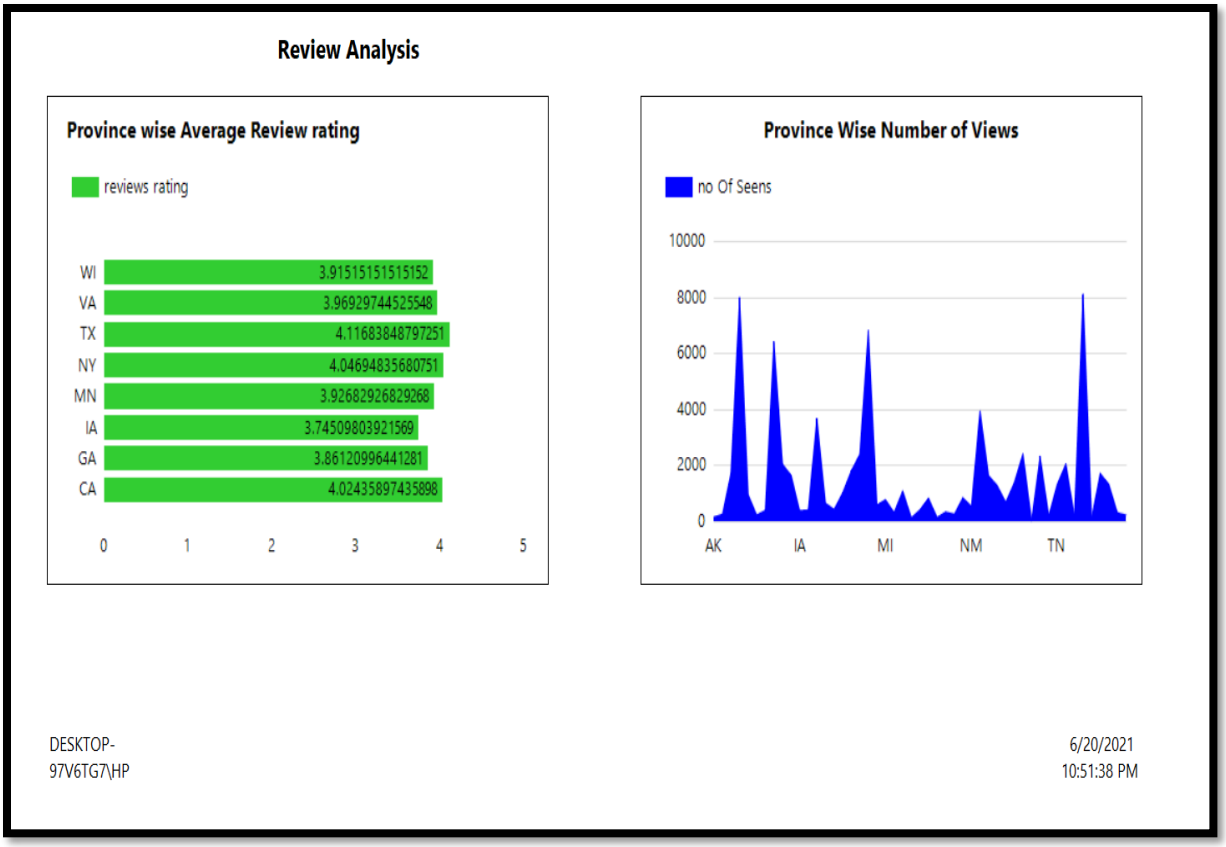
		2015		2016		2017		2018		Total	
province	name	reviews rating	no Of Seens	reviews rating	no Of Seens	reviews rating	no Of Seens	reviews rating	no Of Seens	reviews rating	no Of Seens
⊟AK	Alicia's Eagle Rock Lodge									3	3
	Americas Best Value Inn - Executive Suite Airport	3.2	37	2	34					2.6	71
	Aviator Hotel Anchorage					3	5			3	5
	Best Western Grandma's Feather Bed	5	15	5	16					5	31
	Golden North Motel			3	7					3	7
	Royal Suite Lodge									1	14
	Sunrise Inn									4	12
	Waldo Arms Hotel									5	11
	Yakutat Lodge	5	3							5	3
	Total	3.875	55	2.875	57	3	5			3.3333333333333	157
⊞AR	Total	4.6	70	4.0909090909090	115	3	35			3.7878787878787	251
⊞AZ	Total	3.8043478260869	388	3.8878504672897	839	4.48	198	4.5555555555556	107	3.9339622641509	1725
⊞CA	Total	4	2668	4.0013921113689	3383	4.1795454545454	497	4.2914634146341	278	4.0243589743589	8008
⊞CO	Total	4.0833333333333	320	4.1159420289855	527	4.6666666666667	38			4.0746268656716	965
⊞CT	Total	4.0769230769230	85	4.5333333333333	133					4.3	226
⊞DE	Total	3.6363636363636	64	3.7222222222222	141					3.3214285714285	398
⊞FL	Total	3.9670498084291	1865	4.1051724137931	3396	4.2098901098901	670	3.9661764705882	288	4.0619726339794	6424
⊞GA	Total	3.7924528301886	792	4.0992366412213	970	4.0769230769230	82	2.3333333333333	28	3.8612099644128	2053

Drill-through report.

1. Initially I created a new data set from following SQL query for the level 1 report.

```
SELECT
  DimDate.[Date]
  ,DimDate.[Year]
  ,DimDate.[Month]
  ,DimHotel.SK_hotel
  ,DimHotel.alternate_hotel_id
  ,DimHotel.categories
  ,DimHotel.primaryCategories
  ,DimHotel.name
  ,DimHotel.city
  ,DimHotel.country
  ,DimHotel.province
  ,DimReviewDetails.reviews#title
  ,DimReviewDetails.reviews#text
  ,DimReviewer.reviews#username
  ,DimReviewer.reviews#userProvince
  ,DimReviewer.reviews#userCity
  ,FactReview.noOfSeens
  ,FactReview.[commision#$]
  ,FactReview.reviews#rating
FROM
  DimDate
  INNER JOIN FactReview
    ON DimDate.DateKey = FactReview.ReviewDateKey
  INNER JOIN DimHotel
    ON DimHotel.SK_hotel = FactReview.HotelKey
  INNER JOIN DimReviewDetails
    ON DimReviewDetails.SK_ReviewDetails = FactReview.ReviewDetailsKey
  INNER JOIN DimReviewer
    ON DimReviewer.SK_Reviewer = FactReview.ReviewerKey
```

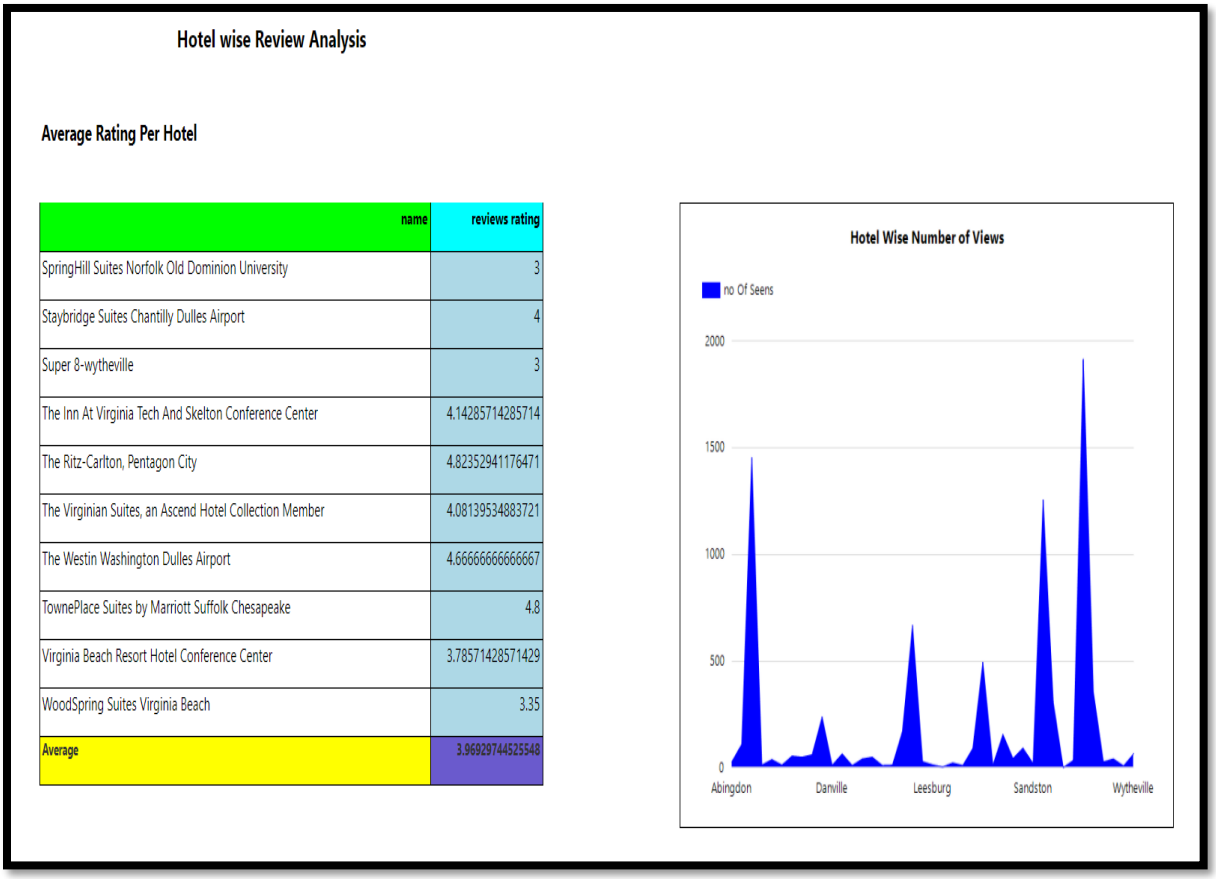
2. Then I created the following report with a bar chart to depict province wise average rating and an area graph to depict province wise view count.



3. I created a data set from following SQL query for the level 2 report.

```
SELECT
  DimDate.[Date]
  ,DimDate.[Year]
  ,DimDate.[Month]
  ,DimHotel.SK_hotel
  ,DimHotel.alternate_hotel_id
  ,DimHotel.categories
  ,DimHotel.primaryCategories
  ,DimHotel.name
  ,DimHotel.city
  ,DimHotel.country
  ,DimHotel.province
  ,DimReviewDetails.reviews#title
  ,DimReviewDetails.reviews#text
  ,DimReviewer.reviews#username
  ,DimReviewer.reviews#userProvince
  ,DimReviewer.reviews#userCity
  ,FactReview.noOfSeens
  ,FactReview.[commision#$]
  ,FactReview.reviews#rating
FROM
  DimDate
  INNER JOIN FactReview
    ON DimDate.DateKey = FactReview.ReviewDateKey
  INNER JOIN DimHotel
    ON DimHotel.SK_hotel = FactReview.HotelKey
  INNER JOIN DimReviewDetails
    ON DimReviewDetails.SK_ReviewDetails = FactReview.ReviewDetailsKey
  INNER JOIN DimReviewer
    ON DimReviewer.SK_Reviewer = FactReview.ReviewerKey
where @ProvinceList = DimHotel.province
```

The following level 2 report demonstrates the hotel wise average rating and hotel wise view counts in an area graph.



Hotel wise Review Analysis

Average Rating Per Hotel

name	reviews rating
SpringHill Suites Norfolk Old Dominion University	3
Staybridge Suites Chantilly Dulles Airport	4
Super 8-wytheville	3
The Inn At Virginia Tech And Skelton Conference Center	4.14285714285714
The Ritz-Carlton, Pentagon City	4.82352941176471
The Virginian Suites, an Ascend Hotel Collection Member	4.08139534883721
The Westin Washington Dulles Airport	4.66666666666667
TownePlace Suites by Marriott Suffolk Chesapeake	4.8
Virginia Beach Resort Hotel Conference Center	3.785714285714 29
WoodSpring Suites Virginia Beach	3.35
Average	3.96929744525548

4. Then I created a parameter for the provinces.

5. Finally, I added an action in the level 1 report to navigate to the level 2 report as follows;