

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ
ТАРАСА ШЕВЧЕНКА
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

Звіт до лабораторної роботи №1

Студента групи ТТП-41
Маркова Максима Юрійовича

викладач: Панченко Тарас Володимирович

1. domain short description

Набір даних, використаний в роботі, містить інформацію про дорожні будівельні та закриття події, що охоплюють 49 штатів США. Будівельні події в цьому наборі даних можуть включати будь-які роботи на дорозі, починаючи від ремонту покриттів до масштабних проєктів, які можуть тривати місяцями. Дані збиралися з січня 2016 року по грудень 2021 року за допомогою кількох API, які надають поточні дані про інциденти на дорогах. Ці API передають дані про рух, які захоплюються різними організаціями, такими як міністерства транспорту США та штатів, правоохоронні органи, камери спостереження за дорожнім рухом та датчики руху в дорожніх мережах. На даний момент у цьому наборі даних налічується близько 6,2 мільйона записів про будівельні та закриті події.

2. original dataset link + short description

<https://www.kaggle.com/datasets/sobhanmoosavi/us-road-construction-and-closures>

US_Constructions_Dec21.csv:

ID

This is a unique identifier of construction record. Severity Shows the severity of the construction, an integer between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the construction) and 4 indicates a significant impact on traffic (i.e., long delay).

Start_Time

Shows start time of construction in local time zone.

End_Time

Shows end time of construction in local time zone. Start_Lat Shows latitude in GPS coordinate of the start point.

Start_Lng

Shows longitude in GPS coordinate of the start point.

End_Lat

Shows latitude in GPS coordinate of the end point.

End_Lng

Shows longitude in GPS coordinate of the end point. Distance(mi) The length of the road extent affected by the construction.

Description

Shows a human provided description of the event.

Number

Shows the street number in address field.

Street

Shows the street name in address field.

Side

Shows the relative side of the street (Right/Left) in address field.

City

Shows the city in address field.

County

Shows the county in address field.

State

Shows the state in address field.

Zipcode

Shows the zipcode in address field.

Country

Shows the country in address field.

Timezone

Shows timezone based on the location of the construction event (eastern, central, etc.).

Airport_Code

Denotes an airport-based weather station which is the closest one to location of the construction event.

Weather_Timestamp

Shows the time-stamp of weather observation record (in local time).

Temperature(F)

Shows the temperature (in Fahrenheit).

Wind_Chill(F)

Shows the wind chill (in Fahrenheit).

Humidity(%)

Shows the humidity (in percentage).

Pressure(in)

Shows the air pressure (in inches).

Visibility(mi)

Shows visibility (in miles).

Wind_Direction

Shows wind direction.

Wind_Speed(mph)

Shows wind speed (in miles per hour).

Precipitation(in)

Shows precipitation amount in inches, if there is any.

Weather_Condition

Shows the weather condition (rain, snow, thunderstorm, fog, etc.)

Amenity

A POI (point-of-interest) annotation which indicates presence of amenity in a nearby location.

Bump

A POI annotation which indicates presence of speed bump or hump in a nearby location.

Crossing

A POI annotation which indicates presence of crossing in a nearby location.

Give_Way

A POI annotation which indicates presence of give_way in a nearby location.

Junction

A POI annotation which indicates presence of junction in a nearby location.

No_Exit

A POI annotation which indicates presence of no_exit in a nearby location.

Railway

A POI annotation which indicates presence of railway in a nearby location.

Roundabout

A POI annotation which indicates presence of roundabout in a nearby location.

Station

A POI annotation which indicates presence of station in a nearby location.

Stop

A POI annotation which indicates presence of stop sign in a nearby location.

Traffic_Calming

A POI annotation which indicates presence of traffic_calming in a nearby location.

Traffic_Signal

A POI annotation which indicates presence of traffic_signal in a nearby location.

Turning_Loop

A POI annotation which indicates presence of turning_loop in a nearby location.

Sunrise_Sunset

Shows the period of day (i.e. day or night) based on sunrise/sunset. This is based on start time of the event.

Civil_Twilight

Shows the period of day (i.e. day or night) based on civil twilight. This is based on start time of the event.

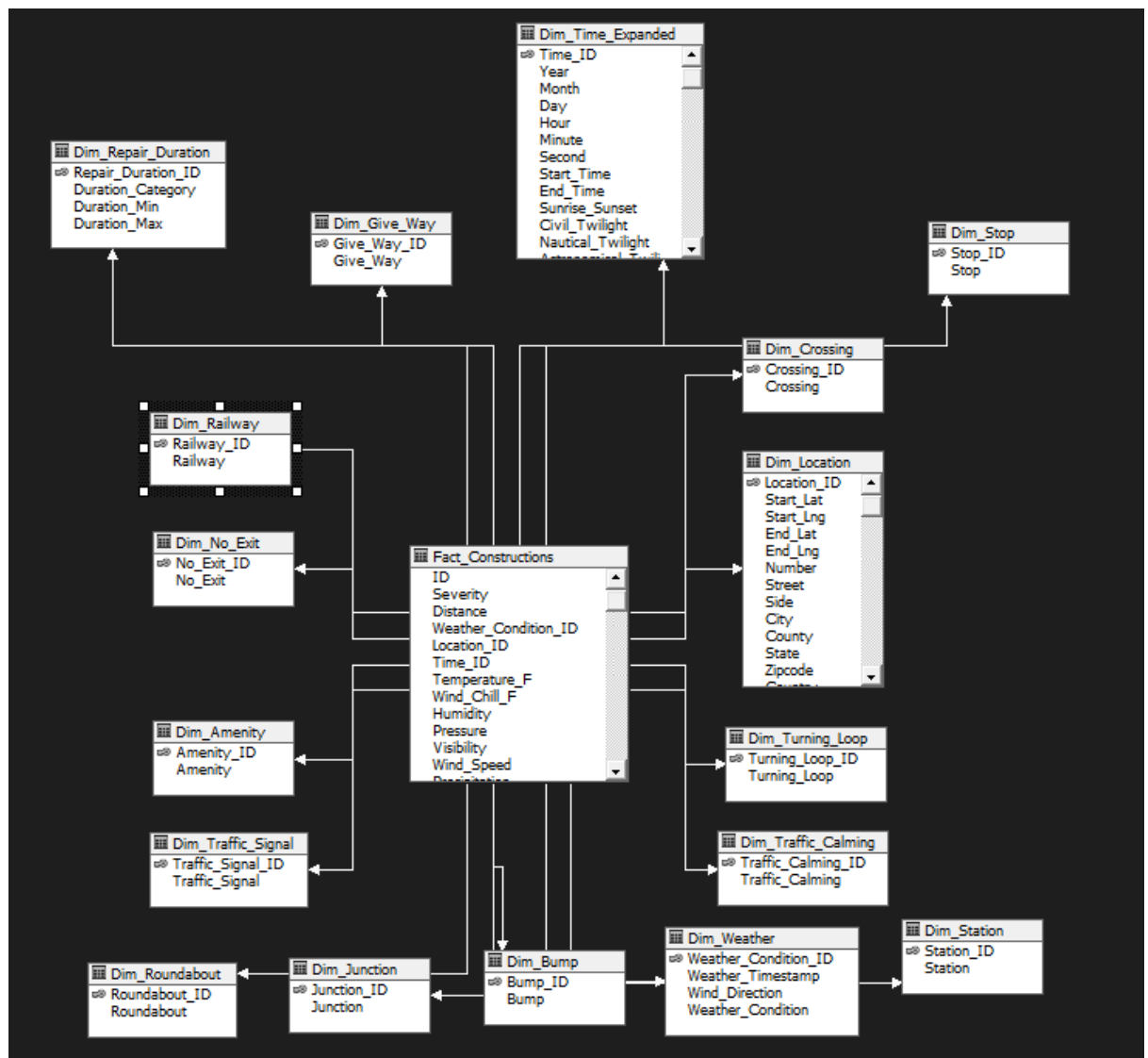
Nautical_Twilight

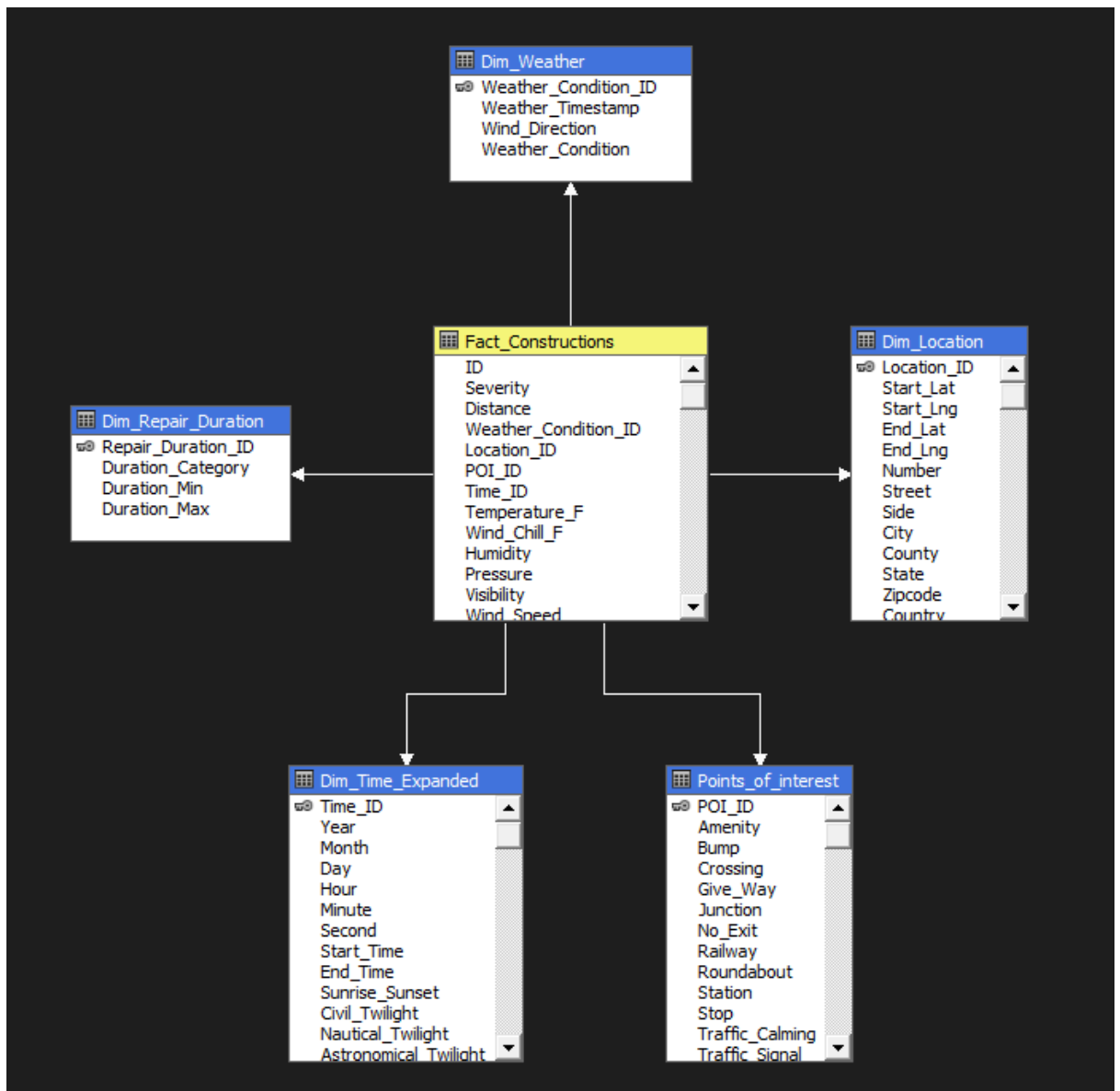
Shows the period of day (i.e. day or night) based on nautical twilight. This is based on start time of the event.

Astronomical_Twilight

Shows the period of day (i.e. day or night) based on astronomical twilight. This is based on start time of the event.

3. star scheme for OLAP warehouse
























4. ETL description

Дані подаються у вигляді csv файлу. У реляційному сховищі даних створюємо таблицю з відповідною схемою та завантажуюмо дані:

ID	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance_mi	Description	Number	Street	Side	City
1	C-1	2019-04-05 16:00:00.000	2020-09-29 11:53:57.000	32.83836	-93.152378	32.85074	-93.164388	1.10349651559	Construction on LA-534 WB near EDMONDS LOOP Road cl...	4200.0	Highway 534	R	Haynesville
2	C-1000	2021-03-05 06:49:30.000	2021-03-05 11:29:26.000	25.887873	-80.210708	25.889155	-80.210251	0.0930213610061	Stationary traffic on US-441 N from NW 124th St to I-95 S du...	12353.0	NW 7th Ave	L	North Miami
3	C-1000001	2021-10-22 16:03:30.000	2021-10-24 02:58:55.000	39.207054	-81.537915	39.205742	-81.536833	0.107579171151	Roadwork on Blair Ave from Sharon St to WW-14.	6749.0	Pike St	R	Mineral Wells
4	C-1000002	2021-01-08 12:28:00.000	2021-01-09 02:30:28.000	27.997469	-82.582765	27.996669	-82.581225	0.109005496084	Slow traffic on FL-580 - County Hwy-589 from Mohr Rd to St...	5416.0	Sheldon Rd	L	Tampa
5	C-1000004	2020-05-21 00:34:15.000	2020-05-21 05:33:55.000	40.81374	-74.10911	40.81513	-74.1083	0.104965366933	Closed at Van Riper Ave - Closed due to roadwork.	2.0	Orient Way	L	Lyndhurst
6	C-1000008	2020-09-30 13:02:00.000	2020-10-01 03:31:28.000	40.609677	-73.965219	40.608351	-73.962121	0.188548836178	Queueing traffic from Ocean Pkwy to Kings Hwy due to roadw...	798.0	Avenue P	R	Brooklyn
7	C-100001	2021-01-19 08:30:00.000	2021-03-02 04:00:00.000	33.483887	-112.073702	33.486328	-112.073689	0.168658700862	MONTEREY WAY - COLUMBUS AVE - Construction - MANH...	2.0	E Earll Dr	L	Phoenix
8	C-1000011	2021-02-13 05:53:30.000	2021-02-14 02:00:00.000	45.417947	-122.404674	45.428846	-122.378234	1.48896882857	Slow traffic on OR-212 from SE 242nd Ave (OR-212) to SE 2...	14499.0	SE Holly View Ter	L	Damascus
9	C-1000012	2019-11-07 13:49:07.000	2020-08-24 20:37:01.000	41.40104	-81.99027	41.38878	-82.02379	1.93290421451	Between Bradley Rd and OH-301/OH-57JFK Mem Pky - Ro...	33742.0	Center Ridge Rd	R	North Ridgevi
10	C-1000014	2021-07-10 10:19:00.000	2021-07-10 12:33:26.000	42.196402	-88.110586	42.204723	-88.111457	0.576652244371	Stationary traffic on US-12 W from IL-22W Main St (S Rand ...	62.0	S Rand Rd	R	Lake Zurich
11	C-1000015	2021-12-19 18:54:30.000	2021-12-19 19:14:30.000	45.311425	-121.801182	45.309605	-121.838852	1.83473480503	Stationary traffic on US-26 E - US-26 W from Government C...	86464.0	Mount Hood Hwy	R	Government C
12	C-1000017	2019-03-04 00:42:24.000	2019-03-08 00:41:56.000	40.71862	-74.00251	40.71831	-74.00183	0.0415552674493	At Cortlandt Aly - Construction work.	68.0	Walker St	L	New York
13	C-1000019	2021-11-24 16:18:19.000	2021-11-24 16:41:59.000	39.000714	-77.345924	39.000195	-77.344886	0.0662750059617	Roadwork on Sugarland Rd (VA-7) from Reston Pkwy (VA-7)...	11567.0	Leesburg Pike	L	Herndon
14	C-1000021	2021-05-26 17:48:58.000	2021-12-02 12:57:42.000	25.939399	-80.261341	25.940338	-80.262775	0.110216626949	Slow traffic on FL-953 - FL-860 from NW 181st St to NW 38...	18201.0	NW 37th Ave	R	Miami Gardel
15	C-1000023	2021-08-16 23:00:00.000	2021-08-16 23:26:00.000	29.964902	-90.032283	29.964149	-90.028833	0.212963178249	Stationary traffic on LA-46 from Franklin Ave (LA-46) to Caffin...	4328.0	Saint Claude Ave	R	New Orleans
16	C-1000024	2021-11-11 10:00:00.000	2021-11-11 10:00:00.000	44.044243	-122.007030	44.044189	-122.007014	0.00000117554195	Closed from W 14th Ave to W 15th Ave due to roadwork.	1000.0	Highway 534	R	Haynesville

Виділимо основні виміри: Time, Weather, Location, Repair_Duration (ієрархія залежна від часу ремонтних робіт) та виміри, що відповідають POI.

Для другої star схеми, у якій всі POI зв'язані одним FK у таблиці фактів, за допомогою Microsoft SQL Server Analysis Services можемо створити виміри для кожного атрибуту таблиці Points_of_interests, які будуть базуватись на цій таблиці:

Measure Groups 	
Dimensions 	 Fact Constructions
 Dim Location 2 (Dim Location)	Location ID
 Dim Weather 2 (Dim Weather)	Weather Condition ID
 Dim Repair Duration 1 (Dim Repair Durati...	Repair Duration ID
 Dim Time Expanded 2 (Dim Time Expand...	Time ID
 Amenity1 (Amenity)	POI ID
 Bump1 (Bump)	POI ID
 Crossing1 (Crossing)	POI ID
 Give Way 1 (Give Way)	POI ID
 Junction1 (Junction)	POI ID
 No Exit 1 (No Exit)	POI ID
 Railway	POI ID
 Roundabout1 (Roundabout)	POI ID
 Station1 (Station)	POI ID
 Stop1 (Stop)	POI ID
 Traffic Calming 1 (Traffic Calming)	POI ID
 Traffic Signal 1 (Traffic Signal)	POI ID

	Column Name	Data Type	Allow Nulls
▶	ID	nvarchar(50)	<input type="checkbox"/>
	Severity	int	<input checked="" type="checkbox"/>
	Distance	float	<input checked="" type="checkbox"/>
	Weather_Condition_ID	int	<input checked="" type="checkbox"/>
	Location_ID	int	<input checked="" type="checkbox"/>
	POI_ID	int	<input checked="" type="checkbox"/>
	Time_ID	int	<input checked="" type="checkbox"/>
	Temperature_F	float	<input checked="" type="checkbox"/>
	Wind_Chill_F	float	<input checked="" type="checkbox"/>
	Humidity	float	<input checked="" type="checkbox"/>
	Pressure	float	<input checked="" type="checkbox"/>
	Visibility	float	<input checked="" type="checkbox"/>
	Wind_Speed	float	<input checked="" type="checkbox"/>
	Precipitation	float	<input checked="" type="checkbox"/>
	Repair_Duration_Seconds	int	<input checked="" type="checkbox"/>
	Repair_Duration_ID	int	<input checked="" type="checkbox"/>
	Amenity_ID	int	<input checked="" type="checkbox"/>
	Bump_ID	int	<input checked="" type="checkbox"/>
	Crossing_ID	int	<input checked="" type="checkbox"/>
	Give_Way_ID	int	<input checked="" type="checkbox"/>
	Junction_ID	int	<input checked="" type="checkbox"/>
	No_Exit_ID	int	<input checked="" type="checkbox"/>
	Railway_ID	int	<input checked="" type="checkbox"/>
	Roundabout_ID	int	<input checked="" type="checkbox"/>
	Station_ID	int	<input checked="" type="checkbox"/>
	Stop_ID	int	<input checked="" type="checkbox"/>
	Traffic_Calming_ID	int	<input checked="" type="checkbox"/>
	Traffic_Signal_ID	int	<input checked="" type="checkbox"/>

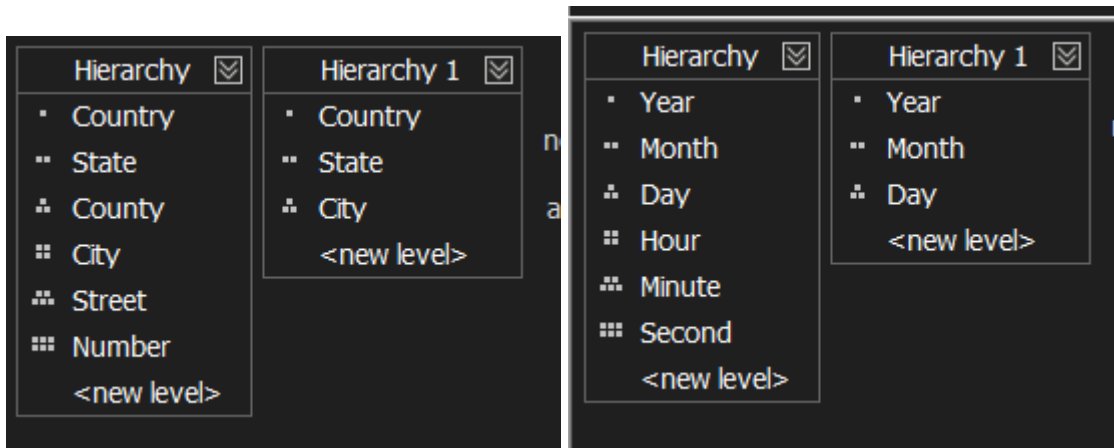
Для створення ієрархічного виміру по часу виконання ремонтних робіт створимо таблицю виміру та саму ієрархію:

	Column Name	Data Type	Allow Nulls
🔑	Repair_Duration_ID	int	<input type="checkbox"/>
	Duration_Category	nvarchar(50)	<input checked="" type="checkbox"/>
	Duration_Min	int	<input checked="" type="checkbox"/>
	Duration_Max	int	<input checked="" type="checkbox"/>

Розділяємо дані та зв'язуємо їх за допомогою FK в таблиці фактів.
Далі, за допомогою Microsoft SQL Server Analysis Services створюємо OLAP куб з вказаними Dimensions and Measures
Додаємо Calculated Measures (переважно агрегаційні функції AVG для KPI):

```
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_Severity]
AS [Measures].[Severity_Sum] / [Measures].[Severity_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_Distance]
AS [Measures].[Distance_Sum] / [Measures].[Distance_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_Humidity]
AS [Measures].[Humidity_Sum] / [Measures].[Humidity_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_Pressure]
AS [Measures].[Pressure_Sum] / [Measures].[Pressure_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_Visibility]
AS [Measures].[Visibility_Sum] / [Measures].[Visibility_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_WindSpeed]
AS [Measures].[WindSpeed_Sum] / [Measures].[WindSpeed_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
CREATE MEMBER CURRENTCUBE.[Measures].[Avg_Temperature]
AS [Measures].[Temp_Sum] / [Measures].[Temp_Count],
VISIBLE = 1 , ASSOCIATED_MEASURE_GROUP = 'Fact Constructions';
```

Додаємо ієрархії для Location та Time:



Деплоїмо куб та підключаємо його до Power BI у live режимі:

SQL Server Analysis Services database

Server

DESKTOP-JG8P8MC

Database (optional)

US_Roads2

☐ Import
☒ Connect live

MDX or DAX query (optional)

OK Cancel

Pivot Table Screenshots:

Year

2020

Country

Fact

Constructions

Count

Distance_Sum

Avg_Distance

Avg_Severity

Avg_Humidity

Avg_Pressure

Avg_Temperature

Avg_Visibility

Avg_WindSpeed

US

729446

524

653,74

0,72

2,29

59,72

29,13

65,62

9,47

7,68

AL

5735

6

516,07

1,14

2,20

66,69

29,55

70,66

9,55

5,59

AR

8353

5

408,73

0,65

2,09

63,98

29,65

65,61

9,55

6,29

AZ

45594

10

276,35

0,23

2,10

18,89

28,50

83,51

9,98

5,72

CA

55960

84

559,75

1,51

2,36

56,08

29,32

65,91

9,09

5,53

CO

27835

20

278,28

0,73

2,22

46,45

24,24

52,74

9,86

7,20

CT

2354

809,77

0,34

2,82

67,09

29,86

57,09

9,09

7,50

DC

2638

628,23

0,24

2,91

69,63

29,99

62,59

9,60

7,11

DE

6479

3

496,13

0,54

2,25

64,87

30,00

64,59

9,50

8,79

FL

95337

31

171,73

0,33

2,12

66,00

29,99

80,21

9,72

10,68

GA

10333

8

010,86

0,78

2,69

66,71

29,23

66,07

9,46

5,77

IA

7463

16

757,12

2,25

2,64

66,29

28,98

62,18

9,32

9,59

ID

18247

12

374,22

0,68

2,04

47,23

27,25

57,42

9,29

7,38

IL

32788

15

667,40

0,48

2,29

58,51

29,30

61,01

9,44

10,73

IN

20751

18

533,37

0,89

2,29

64,67

29,20

61,68

9,44

8,65

KS

6137

7

613,94

1,24

2,38

60,28

28,62

64,91

9,52

9,75

KY

4189

3

076,02

0,73

2,37

68,05

29,06

59,33

9,41

7,50

LA

13521

8

449,32

0,62

2,19

63,79

30,00

72,13

9,55

7,77

MA

3454

2

118,19

0,61

2,89

62,91

29,89

58,54

9,35

9,09

MD

10152

6

183,38

0,61

2,25

62,61

29,79

64,07

9,47

6,46

Total

729446

524

653,74

0,72

2,29

59,72

29,13

65,62

9,47

7,68

2021

Fact

Constructions

Count

Distance_Sum

Year, Month, Day, Hour, Minute, Second

2016

2017

2018

2019

2020

2021

Country, State, County, City, Street, Number

US

AL

AR

AZ

CA

CO

CT

Вимір Repair Duration:

Duration	Category	Long-term jobs		Middle-term jobs		Short-term jobs		Total	Distance	Fact Constructions Count
		Distance	Fact Constructions Count	Distance	Fact Constructions Count	Distance	Fact Constructions Count			
☐	US	188 206,90	282115	302 190,70	404603	1 094 736,73	2153528	1 585 134,33	2840246	
☐	AL	1 573,36	1899	1 527,20	2200	11 908,87	12945	15 009,44	17044	
☐	AR	1 927,98	2975	1 275,41	1545	17 977,93	28734	21 181,32	33254	
☐	AZ	6 584,91	33342	4 466,86	21989	51 134,81	250242	62 186,57	305573	
☐	CA	8 129,71	9680	26 782,54	41076	131 342,81	107671	166 255,06	158427	
☐	CO	8 039,29	10465	8 668,70	15183	47 922,08	88531	64 630,08	114179	
☐	CT	283,51	1066	764,58	3373	2 108,59	5536	3 156,69	9975	
☐	DC	108,87	595	1 170,79	3096	848,94	4303	2 128,59	7994	
☐	DE	885,02	2012	3 594,58	3221	4 941,25	11294	9 420,85	16527	
☐	FL	18 526,62	46655	20 489,49	34462	119 636,51	469914	158 652,61	551031	
☐	GA	1 451,80	3351	5 693,53	13519	20 994,56	34167	28 139,89	51037	
☐	IA	5 350,59	3639	18 847,37	5056	31 471,70	18837	55 669,66	27532	
☐	ID	2 663,82	2457	10 096,50	3256	16 857,05	34184	29 617,37	39897	
☐	IL	6 053,56	14251	11 830,34	11659	30 816,29	122054	48 700,20	147964	
☐	IN	9 529,52	9378	11 177,02	7837	35 276,75	57310	55 983,28	74525	
☐	KS	3 721,25	2655	2 762,83	2270	14 894,16	15166	21 378,24	20091	
☐	KY	857,32	2303	1 618,91	3219	6 637,63	10939	9 113,86	16461	
☐	LA	6 094,11	7169	4 703,14	5037	20 041,56	38369	30 838,81	50575	
	Total	188 206,90	282115	302 190,70	404603	1 094 736,73	2153528	1 585 134,33	2840246	

Year, Month, Day, Hour, Minute, Second

▼ ☐ 2016

▼ 2017

2018

2019

2020

▼ ☐ 2021

☐ ELET

Country, State, County, City, Street, Number

^ ☐ US

✓ ☐ AL

AR

▼ ☐ AZ

CA

CO

✓ ☐ CT

✓ ☐ DC

1