



Báo cáo

PHÂN TÍCH KDL VỀ CÁC CUỘC BIỂU TÌNH

GVHD: Đỗ Thị Minh Phụng

Nội dung







Thành Viên Nhóm



Trần Nhật Khuê 19521713



Lê Ngọc Minh Thư 19522305







Nội Dung Mở đầu 01 Giới thiệu đề tài, XX kho dữ liêu 02 Quá trình SSIS Tích hợp dữ liệu vào kho 03 Quá trình SSAS Phân tích dữ liêu trong kho Quá trình SSRS 04 Tạo các báo biểu Quá trình 05 **Data Mining**



O1 KHO DỮ LIỆU VỀ CÁC CUỘC BIỂU TÌNH







HARVARD

Dataverse

Mass Mobilization Data Project Dataverse (Binghamton University)

Harvard Dataverse > Mass Mobilization Data Project Dataverse >

Mass Mobilization Protest Data

Version 5.0



Clark, David; Regan, Patrick, 2016, "Mass Mobilization Protest Data", https://doi.org/10.7910/DVN/HTTWYL, Harvard Dataverse, V5, UNF:6:F/k8KUqKpCa5UssBbL/gzg== [fileUNF]

Cite Dataset ▼

Learn about Data Citation Standards.

Access Dataset ▼									
Contact Owner	Share								
Dataset Metrics ②									
19,353 Downloads ②									

Description @

Mass Mobillization Data project protests against governments, all countries, 1990-2020. Visit the project page at http://www.binghamton.edu/massmobilization/ (2019-02-07)

Add Data ▼

Search ▼

About

Subject @

Social Sciences

Keyword 🕣

protests, demonstrations

Notes 😱

version 4.0 - 1990-March 31, 2020 Update includes formerly missing notes for 2700+ cases.









1990 - 2020 09/01/2021

[1] [7] [4] [6]

Dòng dữ liệu

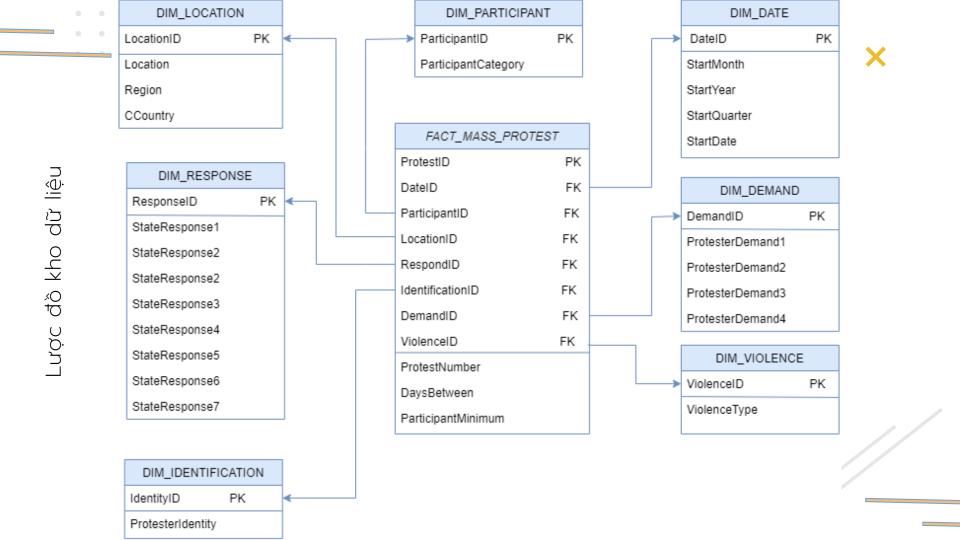


Thuộc tính

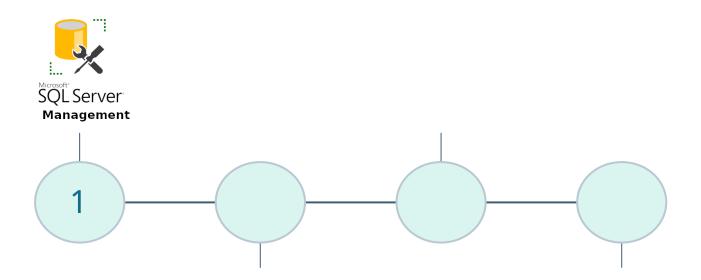


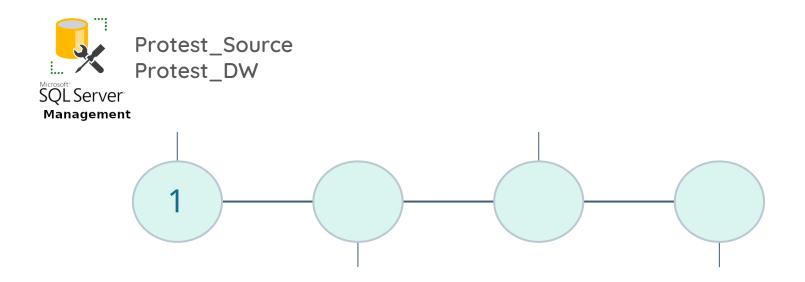
Đất nước

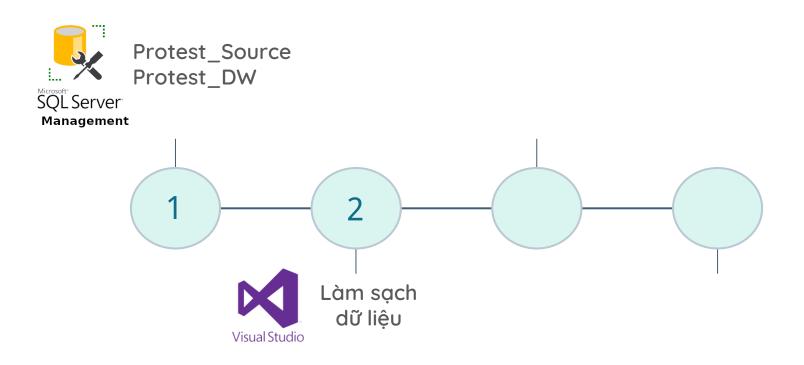


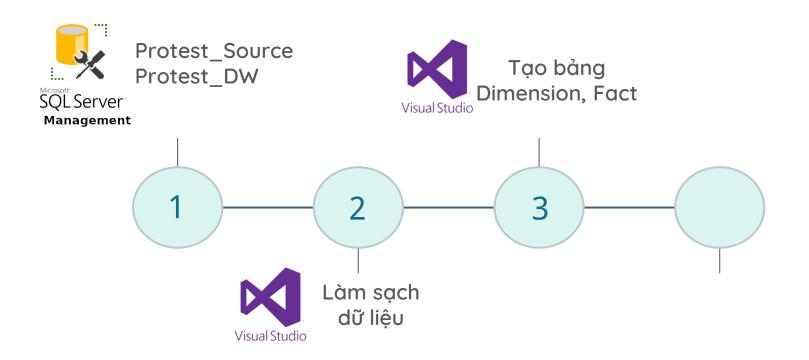


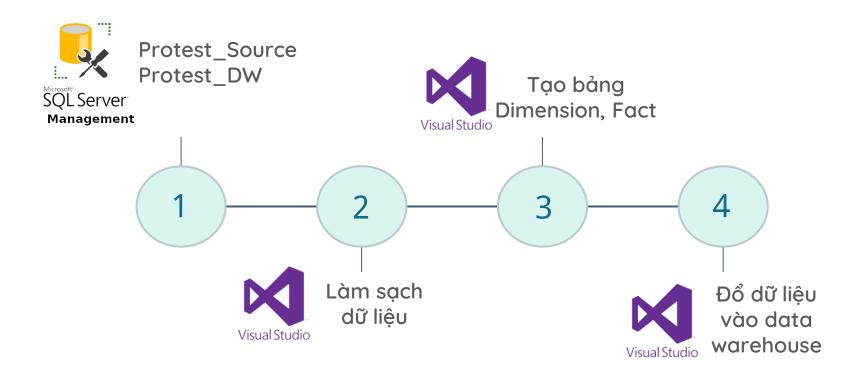
01











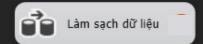
ccode	year	region	protest	I	protestnun s	tartday	startmonth	startyear	endday	endmonth	endyear	protesterv	ilocation	participant	participant	protes
20	199	0 North Ame	9	1	1	15	1	1990	15	1	1990	0	national		1000s	unspe
20	199	0 North Ame	•	1	2	25	6	1990	25	6	1990	0	Montreal,	Quebec	1000	unspe
20	199	0 North Ame	•	1	3	1	7	1990	1	7	1990	0	Montreal,	Quebec	500	separa
20	199	0 North Ame	•	1	4	12	7	1990	6	9	1990	1	Montreal,	Quebec	100s	mohav
20	199	0 North Ame		1	5	14	8	1990	15	8	1990	1	Montreal,	Quebec	950	local r
20	199	0 North Ame		1	6	19	9	1990	19	9	1990	C	Kahnawa	ke Reservat	i 200	moha\
20	199	1 North Ame		1	1	10	9	1991	17	9	1991	0	national		110000	public
20	199	1 North Ame		1	2	28	9	1991	2	10	1991	0	national		110000	public
20	199	2 North Ame		1	1	4	5	1992	5	5	1992	1	Toronto		1000	gangs
20	199	3 North Ame		1	1	16	5	1993	16	5	1993	0	Ottawa		10000s	unspe
20	199	3 North Ame		1	2	1	7	1993	31	8	1993	0	Clayoquo	t Sound, Brit	t 100s	enviro
20	199	4 North Ame		1	1	1	9	1994	1	9	1994	0	Ottawa		10000	gun o
20	199	4 North Ame		1	2	18	11	1994	18	11	1994	0	Ottawa		6000	studer
20	199	5 North Ame		1	1	20	2	1995	20	2	1995	C	Pickering	, Ontario	3000	unspe
20	199	5 North Ame		1	2	8	9	1995	8	9	1995	1	Ipperwas	h Provincial	100s	chippe
20	199	6 North Ame		1	1	26	10	1996	26	10	1996	0	Toronto,	Ontario	1000s	labor u
20	199	7 North Ame		1	1	28	10	1997	9	11	1997	0	Ontario		126000	ontario
20	199	7 North Ame		1	2	19	11	1997	4	12	1997	0	national		45000	canad
20	199	8 North Ame		0	0											
20	199	9 North Ame		0	0											
20	200	0 North Ame		1	1	22	2	2000	23	2	2000	1	Quebec (City, Quebec	2000	french
20	200	0 North Ame		1	2	26	2	2000	26	2	2000	C	Saskatoo	n, Saskatch	400	unspe
20	200	0 North Ame	,	1	3	9	5	2000	9	5	2000	0	Edmonto	n. Alberta	500	unspe

startday	startmonth	startyear	endday	endmonth	endyear
15	1	1990	15	1	1990
25	6	1990	25	6	1990
1	7	1990	1	7	1990
12	7	1990	6	9	1990
14	8	1990	15	8	1990
19	9	1990	19	9	1990
10	9	1991	17	9	1991
28	9	1991	2	10	1991
4	5	1992	5	5	1992
16	5	1993	16	5	1993
1	7	1993	31	8	1993
1	9	1994	1	9	1994
18	11	1994	18	11	1994
20	2	1995	20	2	1995
Я	q	1995	Q	q	1995

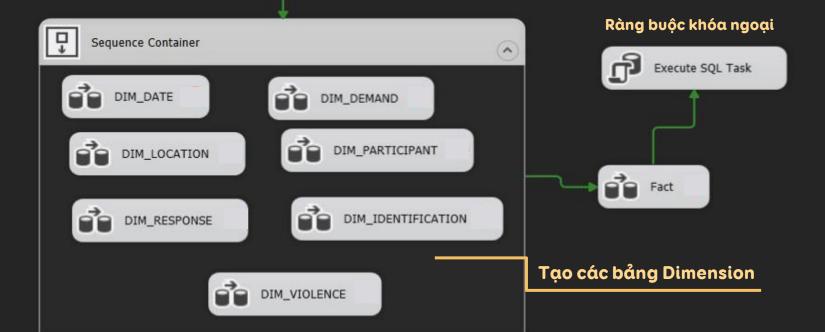
startday	startmonth	startyear	endday	endmonth	endyear
15	1	1990	15	1	1990
25	6	1990	25	6	1990
1	7	1990	1	7	1990
12	7	1990	6	9	1990
14	8	1990	15	8	1990
19	9	1990	19	9	1990
10	9	1991	17	9	1991
28	9	1991	2	10	1991
4	5	1992	5	5	1992
16	5	1993	16	5	1993
1	7	1993	31	8	1993
1	9	1994	1	9	1994
18	11	1994	18	11	1994
20	2	1995	20	2	1995
R	Q	1995	R	Q	1995

StartDate EndDate

DaysBetween=
EndDate-StartDate



Đổ dữ liệu, lọc và sắp xếp

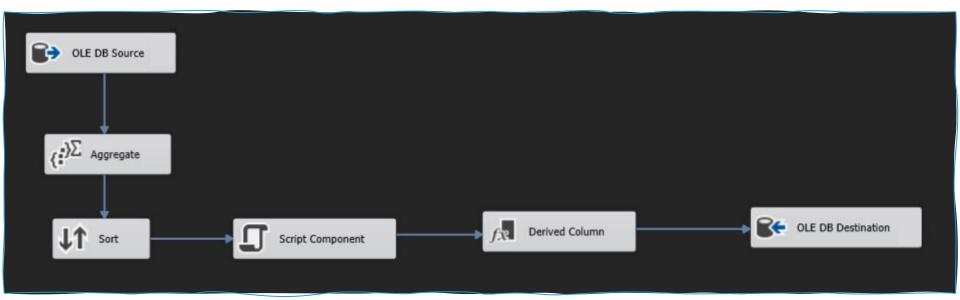


■ Tạo Bảng Dim_Date

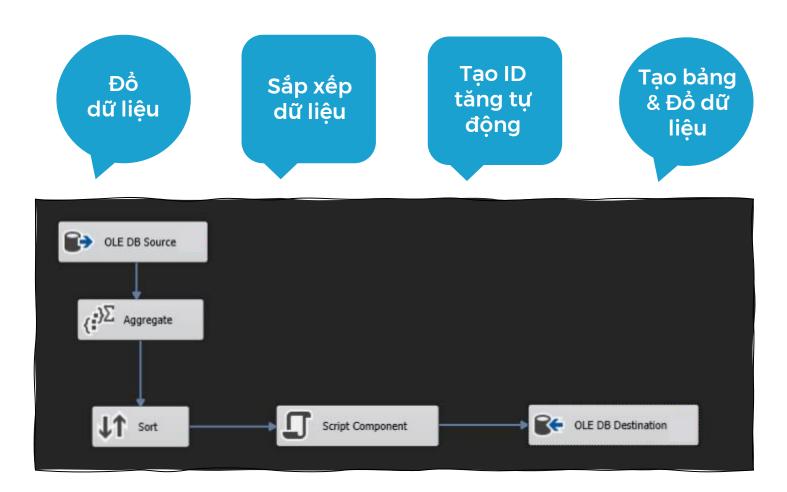


Tạo ID tăng tự động Tạo thuộc tính 'quý'

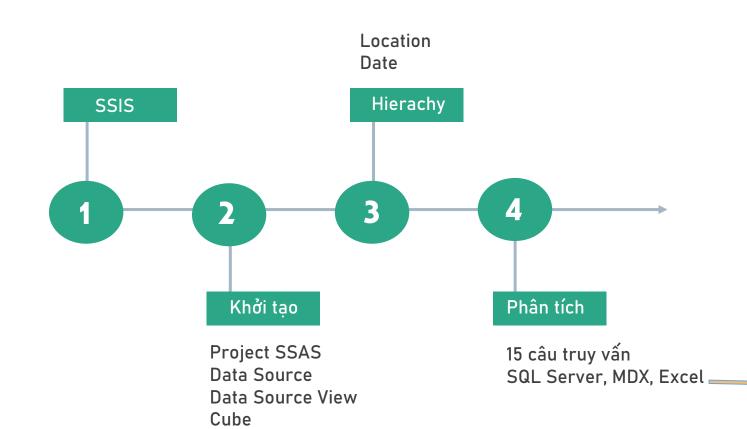
Tạo bảng & Đổ dữ liệu



Tạo bảng Dimension còn lại (Location, Identification, Demand, Response, Participant, Violence)



02



```
----Câu 2: Cho biết thông tin cuộc biểu tình, nước xảy ra biểu tình có tổng số ----lượng người ước tính tham gia biểu tình cao nhất trong quý 3/1990

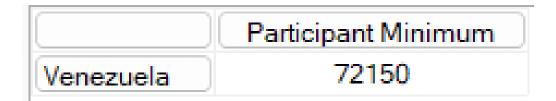
Select NON EMPTY {[Measures].[Participant Minimum]} on Columns ,

{TOPSUM([Location].[Country].CHILDREN,1,[Measures].[Participant Minimum])} on Rows

From [Protest Dw v9]

Where [Date].[Y_Q].[Start Quarter].&[3]&[1990];
```

```
----Câu 2: Cho biết thông tin cuộc biểu tình, nước xảy ra biểu tình có tổng số ---lượng người ước tính tham gia biểu tình cao nhất trong quý 3/1990
Select NON EMPTY {[Measures].[Participant Minimum]} on Columns ,
{TOPSUM([Location].[Country].CHILDREN,1,[Measures].[Participant Minimum])} on Rows
From [Protest Dw v9]
Where [Date].[Y_Q].[Start Quarter].&[3]&[1990];
```



```
---- Câu 4:Theo từng tháng, quý, năm liệt kê số lần diễn ra biểu tình tại các nước ở châu á
|Select Non empty{[Location].[Country].Children*[Measures].[Protestnumber]} on Columns,
| {DrillDownLevel(
| DrillDownLevel(
| DrillDownLevel([Date].[Y_Q_M])
| )
| )
| on Rows
```

79988											
} 0	n Rows										
Fro	m [Protes	t Dw v9]									
who	re {[Loca	tion] [P	Cl [Pog	ion I O [Ac	iall.						
WITE	e ([Loca	CION].[K	_c].[reg.	LOII] . a[As	stall,						
	Afghanistan	Bangladesh	Bhutan	Cambodia	China	India	Indonesia	Japan	Kazakhstan	Kuranatan	Laos
-	Protestnumber	Kyrgyzstan Protestnumber	Protestnumber								
			1								
All	108	4978	2	142	3841	1328	764	169	266	2014	2
1990	(null)	55	1	(null)	6	55	1	15	1	1	(null)
1	(null)	(null)	(null)	(null)	1	10	1	1	(null)	(null)	(null)
1	(null)	(null)	(null)	(null)	1	3	1	(null)	(null)	(null)	(null)
2	(null)	(null)	(null)	(null)	(null)	3	(null)	(null)	(null)	(null)	(null)
3	(null)	(null)	(null)	(null)	(null)	4	(null)	1	(null)	(null)	(null)
2	(null)	3	(null)	(null)	5	5	(null)	(null)	(null)	1	(null)
4	(null)	(null)	(null)	(null)	2	(null)	(null)	(null)	(null)	(null)	(null)
5	(null)	(null)	(null)	(null)	(null)	5	(null)	(null)	(null)	(null)	(null)
6	(null)	3	(null)	(null)	3	(null)	(null)	(null)	(null)	1	(null)
3	(null)	3	1	(null)	(null)	40	(null)	(null)	1	(null)	(null)
7	(null)	(null)									

```
--- câu 5: Cho biết thông tin nhóm người biểu tình và tổng số cuộc biểu tình của cuộc biểu tình đó lớn hơn 15
----được đặt theo từng quốc gia
Select {[Measures].[Protestnumber]} on Columns,
{Generate(
    [Location].[Country].Children,
    [Location].[Country].CurrentMember *
    Filter(
        [Identification].[Protesteridentity].CHILDREN,[Measures].[Protestnumber]>15)
        )
    } on Rows
```

From [Protest Dw v9];

```
--- câu 5: Cho biết thông tin nhóm người biểu tình và tổng số cuộc biểu tình của cuộc biểu tình đó lớn hơn 15
----được đặt theo từng quốc gia
Select {[Measures].[Protestnumber]} on Columns,
{Generate(
     [Location].[Country].Children,
     [Location].[Country].CurrentMember *
     Filter(
          [Identification].[Protesteridentity].CHILDREN,[Measures].[Protestnumber]>15)
     )
} on Rows
```

		D(
At	ghanistan	hazaras	

From [Protest Dw v9];

Afghanistan	hazaras	16
Afghanistan	protesters	31
Albania	albanians	88
Albania	opposition	162
Albania	protesters	28
Albania	students	63
Algeria	judicial activists	58
The Section II	TI Consequence	20

Protestnumber

	All	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Africa	5120	125	157	99	115	106	69	84	167	356	72	139	122	129	98	171
Asia	9346	277	82	152	65	126	131	137	275	228	143	458	434	233	175	305
Central America	1159	14	8	13	20	29	79	59	48	19	34	25	20	18	28	24
MENA	4453	20	39	40	34	37	23	16	9	26	44	19	28	33	21	31
North America	1952	139	81	85	161	32	47	30	56	19	39	20	9	13	39	14
Oceania	121	4	43	(null)	1	(null)	1	1	3	1	(null)	2	18	(null)	(null)	2
South America	7209	138	347	205	203	223	192	327	362	272	342	173	146	279	233	236

	Fact Count
Europe	4994
Oceania	38

03

Quá trình lập báo biểu

Quá Trình SSRS

Quá Trình SSRS



Thống kê số cuộc biểu tình, số ngày biểu tình, số người tham gia biểu tình tại Nga từ 1992 đến 1999

Ngày lập báo cáo: 5/13/2022 10:02:13 PM

Năm	Quý	Số ngày biểu tình	Số người tham gia	Số cuộc biểu tình
	1	6	33000	21
	2	3	30000	24
1992	3	1	100	10
	4	1	6 33000 3 30000 1 100	11
	Tổng cộng	11	64100	66
	1	2	20000	3
1993	2	3	20100	12
1993	4	2	12000	13
	Tổng cộng	7	52100	28
	2	1		1
1994	4	4	30100	14
	Tổng cộng	5	40100	15
	2	3	20100	6
1995	3	1	50	4
1885	4		10000	5
	Tổng cộng	5	3 20100 1 50 1 10000 5 30150 4 13000 2 2100	15
	1	4		10
1996	4			11
	Tổng cộng	6		21
1997	1	1	10000	1
1001	Tổng cộng	1	10000	1
	2	4		10
1998	3	1		5
1550	4			13
	Tổng cộng	7		28
	1	1		1
1999	4			2
	Tổng cộng			3
Tổng cộng		44	282550	177

Quá Trình SSRS Report Visual Studio

Thống kê số cuộc biểu tình, số ngày biểu tình, số người tham gia biểu tình tại Nga từ 1992 đến 1999

Thống kê số người tham gia biểu tình tại Nam Mĩ và Bắc Mĩ trong 2019

Ngày lập báo cáo: 5/13/2022 10:12:34 PM

Quá Trình SSRS Report Visual Studio

Châu lục	Quốc gia	Số người tham gia
	Canada	2250
Canada Cuba	200	
North America	Haiti	59300
	Mexico	2400
	Tổng cộng	64150
	Argentina	100
	Bolivia	90250
	Brazil	10150
	Chile	90100
South America	Colombia	50000
South America	Ecuador	65250
	Paraguay	2000
	Peru	4000
	Venezuela	32450
	Tổng cộng	344300
Tổng cộng		408450

Thống kê người tham gia biểu tình tại Nam Mĩ và Bắc Mĩ trong năm 2019

Quá Trình SSRS Report Visual Studio

Thống kê số người tham gia biểu tình ở Central America trong năm 2019 - 2020 (theo tháng quý năm)

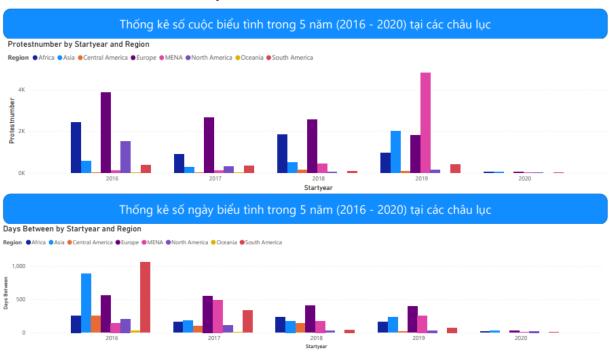
Ngày lập báo cáo: 5/21/2022 8:38:17 PM

			Central America						
Năm	Quý	Tháng	Guatemala	Honduras	Nicaragua	Tổng			
	1	1	100			100			
		2			50	50			
		3			150	150			
		Tổng	100		200	300			
	2	4		15000	200	15200			
		5		10000		10000			
2019		6		10100		10100			
2019		Tổng		35100	200	35300			
	3	7	150			150			
		8	2000			2000			
		9			100	100			
		Tổng	2150		100	2250			
	4	10		6000		6000			
		Tổng		6000		6000			
Tổng			2250	41100	500	43850			

Thống kê số người tham gia biểu tình ở Central America trong năm 2019 - 2020 (theo tháng quý năm)

Quá Trình SSRS

Repertisent Studio



Thống kê số cuộc biểu tình, số ngày biểu tình trong 5 năm (2016 - 2020) tại các Châu lục

Quá Trình SSRS

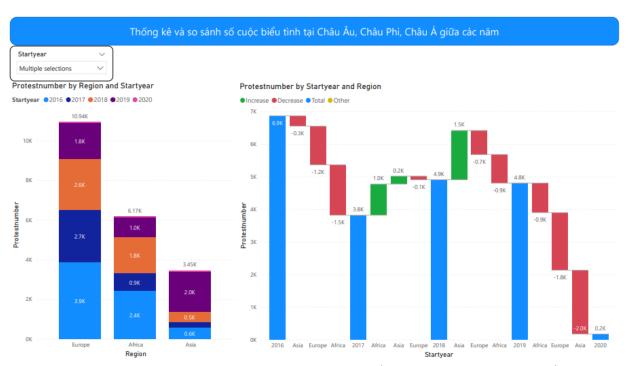
Report Power BI



Thống kê số người tham gia các cuộc biểu tình ôn hoà tại MENA

Quá Trình SSRS

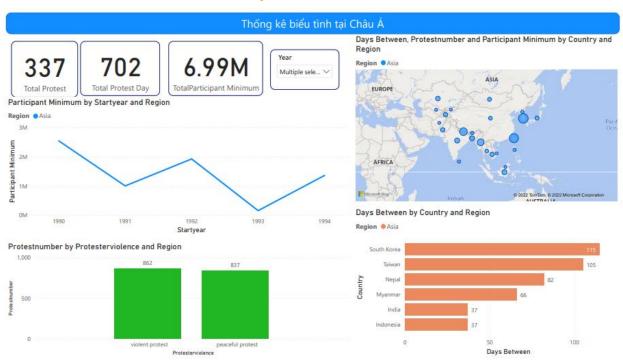
Report Power BI



Thống kê và so sánh số cuộc biểu tình tại Châu Âu, Châu Phi, Châu Á giữa các năm

Quá Trình SSRS

Report Power BI



Thống kê biểu tình tại Châu Á

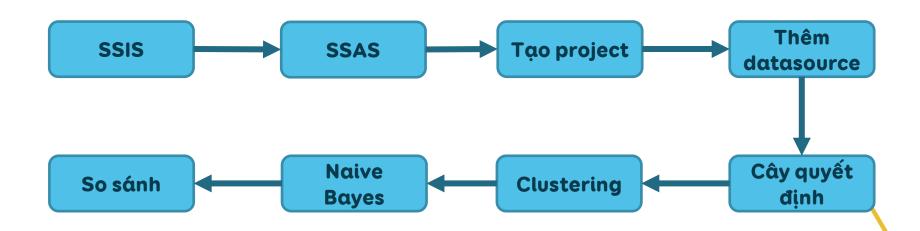
04

Quá trình Data Mining

Phân tích các yếu tố gây ra bệnh tim

Quá trình Data Mining

Phân tích các yếu tố gây ra bệnh tim



Quá trình Data Mining





New Notebook





:

Personal Key Indicators of Heart Disease

2020 annual CDC survey data of 400k adults related to their health status



Data

Code (98)

Discussion (9)

Metadata

About Dataset

Key Indicators of Heart Disease

2020 annual CDC survey data of 400k adults related to their health status

Usability ①

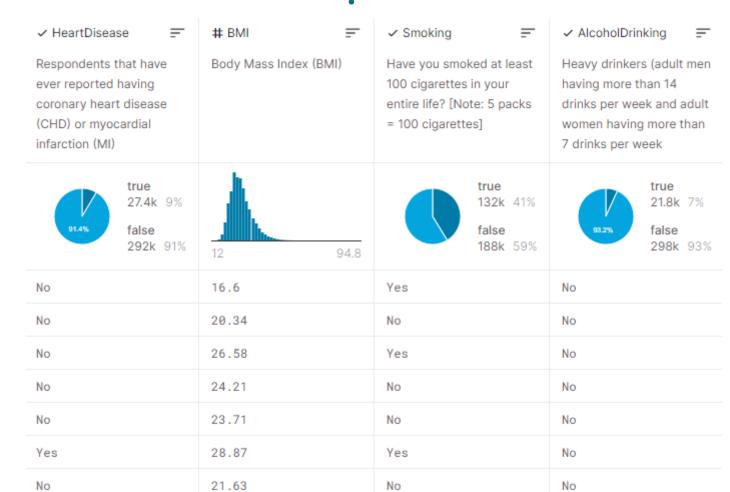
10.00

License

CC0: Public Domain

Expected update frequency

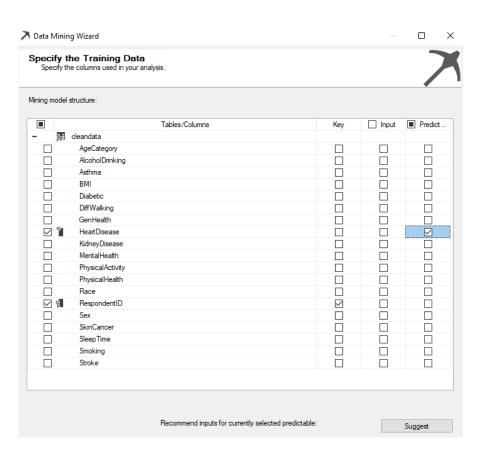
Giới thiệu Dataset



Giới thiệu Dataset

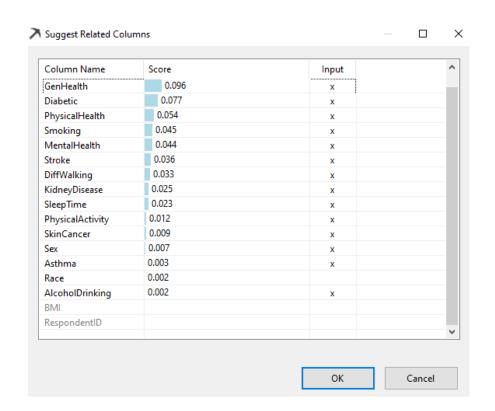
AlcoholE	Orir Stroke	PhysicalHe	MentalHea	DiffWalking	Sex	AgeCatego	Race	Diabetic	PhysicalAd	GenHealth	SleepTime Asthma	KidneyDise	SkinCance
No	No	3	30	No	Female	55-59	White	Yes	Yes	Very good	5 Yes	No	Yes
No	Yes	0	0	No	Female	80 or older	White	No	Yes	Very good	7 No	No	No
No	No	20	30	No	Male	65-69	White	Yes	Yes	Fair	8 Yes	No	No
No	No	0	0	No	Female	75-79	White	No	No	Good	6 No	No	Yes
No	No	28	0	Yes	Female	40-44	White	No	Yes	Very good	8 No	No	No
No	No	6	0	Yes	Female	75-79	Black	No	No	Fair	12 No	No	No
No	No	15	0	No	Female	70-74	White	No	Yes	Fair	4 Yes	No	Yes
No	No	5	0	Yes	Female	80 or older	White	Yes	No	Good	9 Yes	No	No
No	No	0	0	No	Female	80 or older	White	No, border	No	Fair	5 No	Yes	No
No	No	0	0	Yes	Male	65-69	White	No	Yes	Good	10 No	No	No
No	No	30	0	Yes	Male	60-64	White	Yes	No	Poor	15 Yes	No	No
No	No	0	0	No	Female	55-59	White	No	Yes	Very good	5 No	No	No
No	No	0	0	Yes	Male	75-79	White	Yes	Yes	Very good	8 No	No	No
No	No	7	0	Yes	Female	80 or older	White	No	No	Good	7 No	No	No
No	No	0	30	Yes	Female	60-64	White	No	No	Good	5 No	No	No
No	No	1	0	No	Female	50-54	White	No	Yes	Very good	6 No	No	No
No	No	5	2	No	Female	70-74	White	No	No	Very good	10 No	No	No
No	No	0	30	Yes	Male	70-74	White	No, border	Yes	Good	8 No	No	No
No	No	0	0	Yes	Female	75-79	Black	Yes	No	Fair	5 No	Yes	No
No	No	0	0	No	Male	80 or older	White	No	Yes	Excellent	8 No	No	Yes
No	No	0	0	No	Female	80 or older	Black	No, border	No	Good	8 No	No	No
No	No	2	5	No	Female	70-74	White	Yes	No	Very good	7 No	No	No
No	No	0	15	No	Female	45-49	White	No	Yes	Very good	6 No	No	No

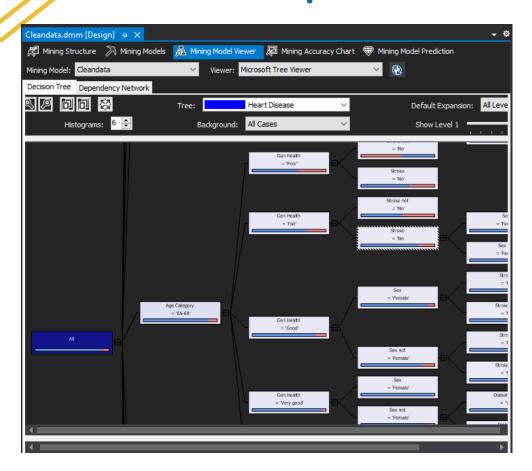
Quá trình thực hiện

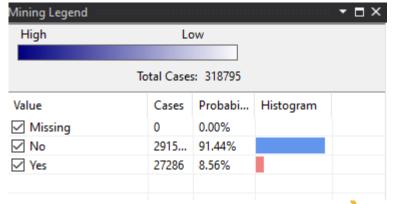


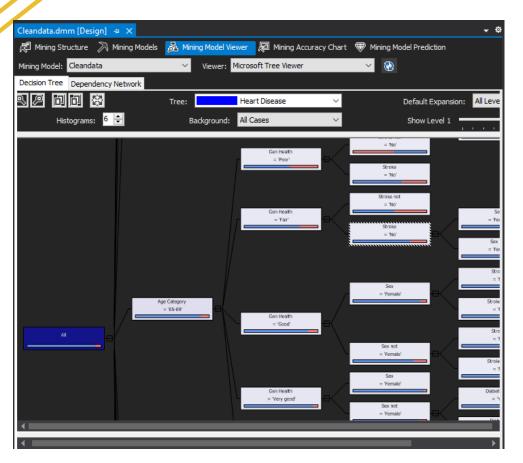
Quá trình thực hiện

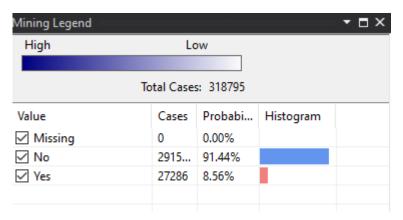








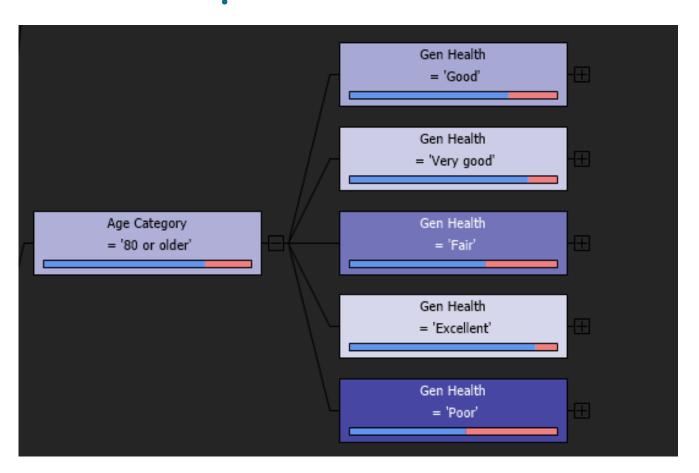


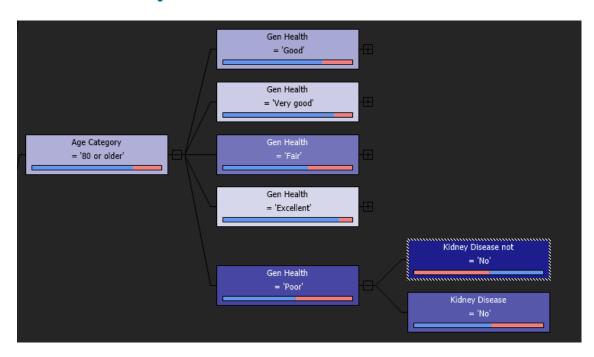


Với thuộc tính dự đoán là HeartDisease.

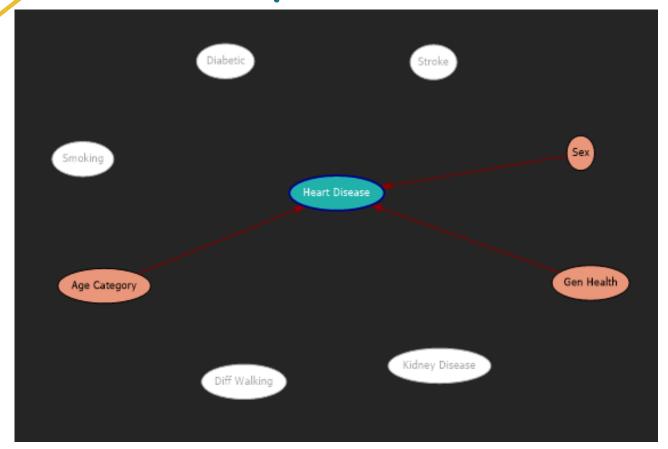
Tỷ lệ bị bệnh tim (Yes) là 8.56%

Tỷ lệ không bị bệnh tim (No) là 91.44%



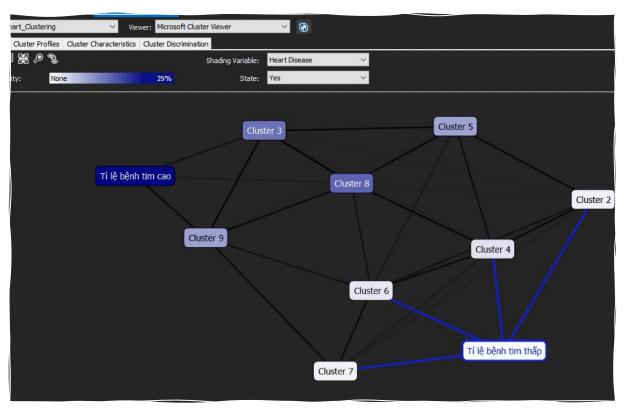


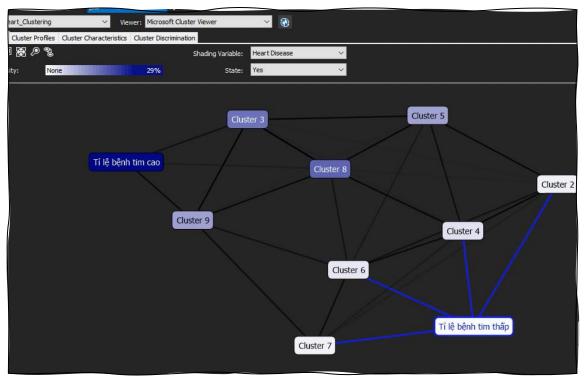
⇒ Rút ra tập luật: Những người thực hiện khảo sát tại Mỹ, nếu là người từ 80 tuổi trở lên (Age Category = '80 or older'), sức khỏe yếu (Gen Health = 'Poor') và mắc bệnh về thận (Kidney disease = 'Yes') thì có khả năng mắc bệnh về tim (nhồi máu cơ tim hoặc bệnh về động mạch vành) cao hơn các nhóm tuổi còn lại.



Age Category Gen Health Sex Stroke Diabetic Smoking Diff Walking

Dependency Network



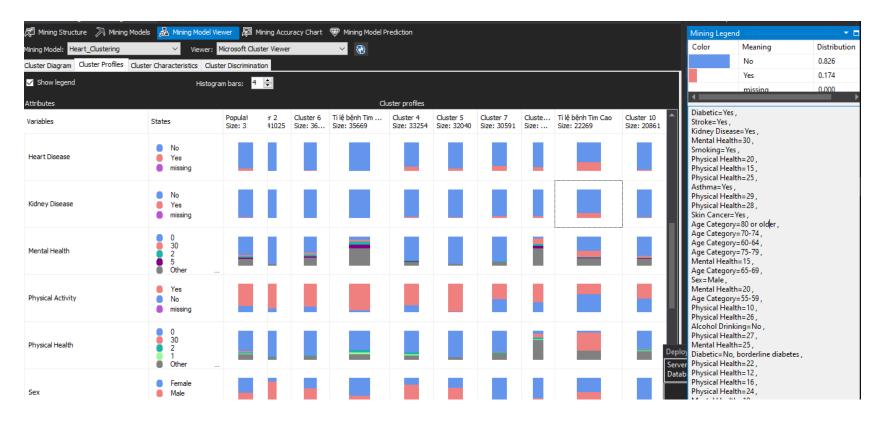


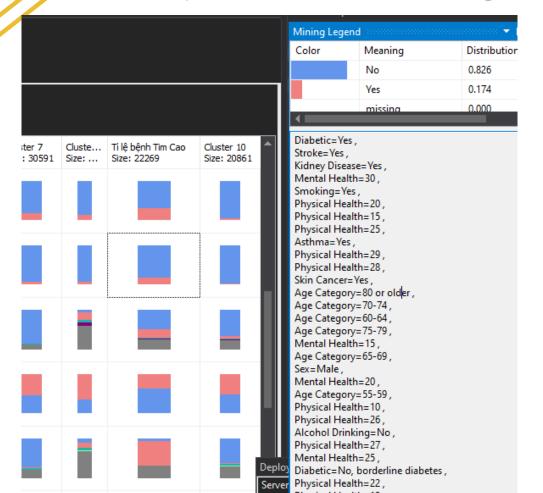
Cluster Diagram của thuật toán với:

Shading Variable là '*HeartDisease*', value là '*yes*' (mắc bệnh tim).

Cluster đậm màu nhất:

- -> Khả năng bị bệnh tim cao nhất Cluster nhạt màu nhất:
- -> Khả năng bị bệnh tim thấp nhất





Với *Cluster Profiles*, Cluster tên 'Tỉ lệ bệnh tim Cao' cho ta thấy tập lục của những người được phỏng vấn bị bệnh tim chiếm 17.4%

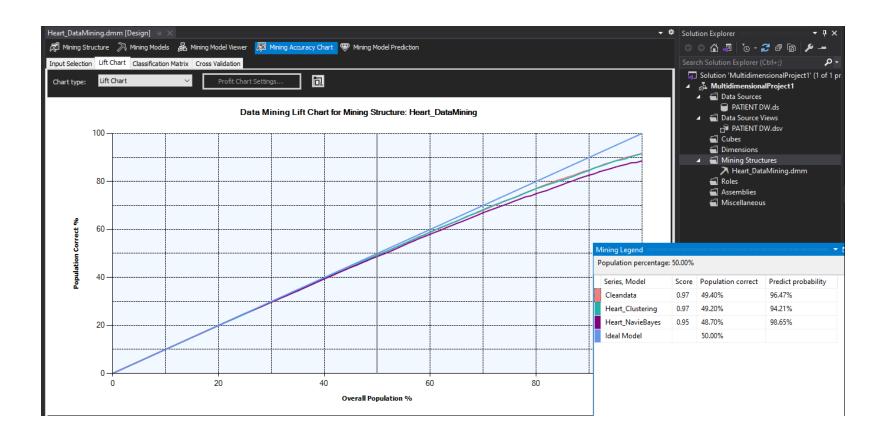
Tập lục: Những người có độ tuổi từ 60 tuổi trở lên, giới tính nam (sex), mắc bệnh tiểu đường (Diabetic), bị đột quy (stroke), mắc bệnh về thận (kidney disease), cảm thấy không khỏe về mặt thể chất từ 15 ngày trở lên, mắc bệnh về da (Skin cancer) có tỉ lê mắc bênh tim cao.

Thuật toán Naive Bayes



So sánh điểm giữa hai cluster 'tỉ lệ bệnh tim cao' và 'tỉ lệ bệnh tim thấp'

So sánh thuật toán với biểu đồ Lift



THANK —YOU