Data Model and DAX

Venkata Reddy Konasani

Contents

- Introduction to DAX
- DAX formulas
- Important DAX formulas and how to work with them

Stroke Case study

Step4: Model the data and create derived columns

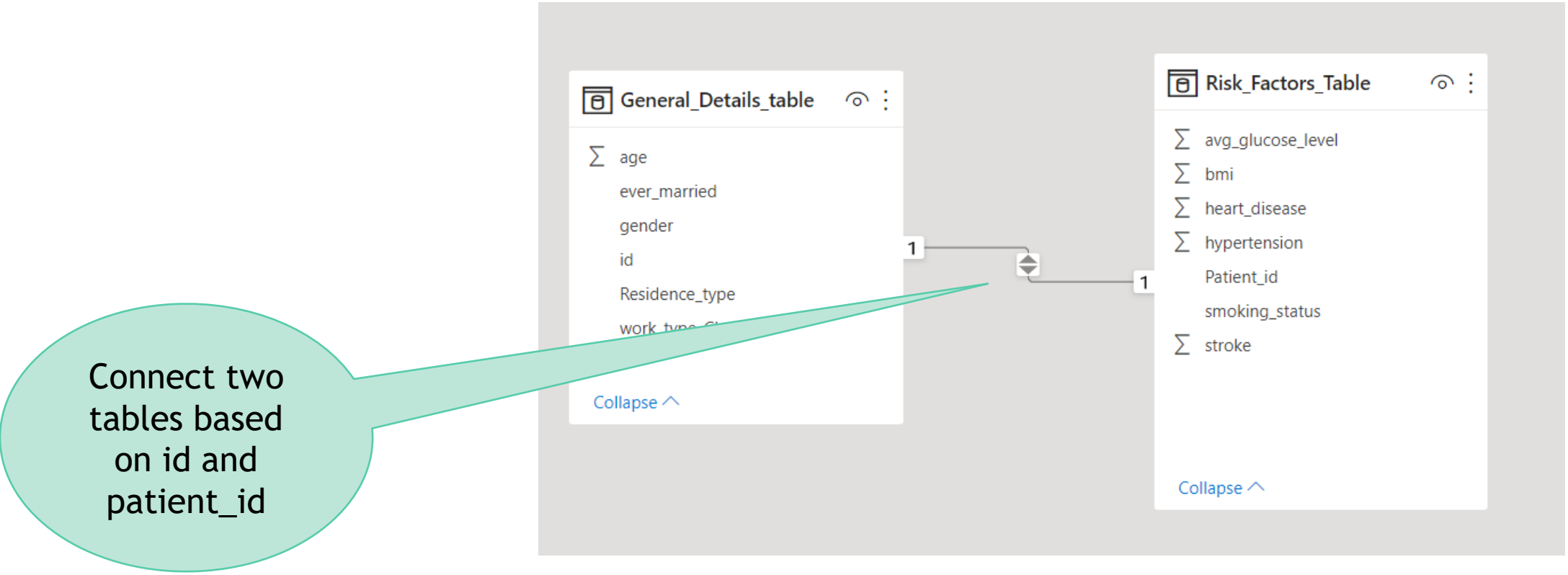
Data Model

- Find out fact table and Lookup table.
- Connect them based on the common ids.

- More details on data model are coming-up in our course.

Join the datasets

- Drag and drop the relevant keys to connect the tables.



DAX

- Data Analysis Expressions **(DAX)** is a programming language that is used throughout Microsoft Power BI **for creating calculated columns**, **measures**, and custom tables.
- **It is a collection of functions,** operators, and constants that can be used in a formula, or expression, to calculate and return one or more values.
- You can use DAX **to solve** a number of calculations and **data analysis** problems, which can help you create new information from data that is already in your model.

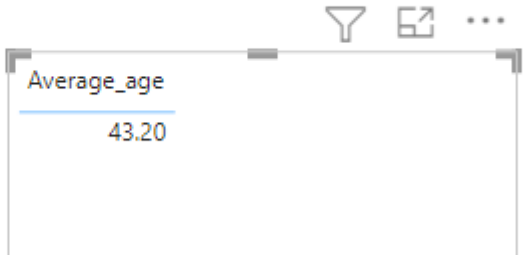
What is a measure?

- Measure are similar Aggregation functions in SQL. Sum(), Avg(), count(), min() and max()
- You can create measures yourself, they appear in the fields list with a **calculator icon.**
- You can name measures whatever you want, and add them to a new or existing visualization just like any other field.
- **Remember** measures are calculated on the whole data or a column. Measures are calculated using DAX formulas.
- DAX formulas look **similar to excel formulas**, but excel formulas are applied on each **cell**, measures are applied on **full table**

Calculate the average age

- Create a new measure average age.
- Take help of intelligence, if it is not loading then submit the partial formula with errors.
- Display the value in a table

```
1  Average_age = AVERAGE(General_Details_table[age])
```



Context

- We have calculated the average age for the overall data.
- Can we change the context and find the average age in Male and Female?
- Can we change the context and find the average age in Male and Female? – Or do we need to write a new DAX formula.
- Can we change the context and find the average age by marital status?
- Can we find the average age in every job category.

- We need NOT write multiple formulas. With one measure, DAX automatically calculates the average based on the context.

Context based measure calculations

- Since DAX automatically takes care of the context. **Its is suggested to use DAX measures as much as possible** and avoid creating new columns in the data transformation phase.

Average_age
43.20

gender	Average_age
Female	43.70
Male	42.46
NA	52.57
Other	26.00
Total	**43.20**

Single formula used in multiple contexts

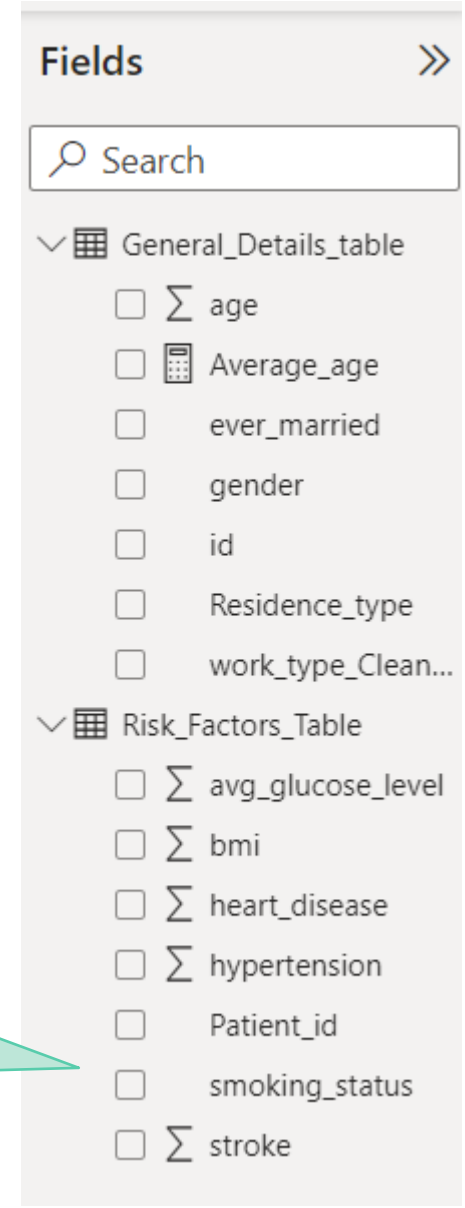
work_type_Cleaned1	Average_age
children	6.84
Govt_job	50.88
NA	43.20
Never_worked	16.18
Private	45.48
Self-employed	60.19
Total	**43.20**

Residence_type	Average_age
NA	43.20
Rural	42.88
Urban	43.52
Total	**43.20**

Context in connected tables

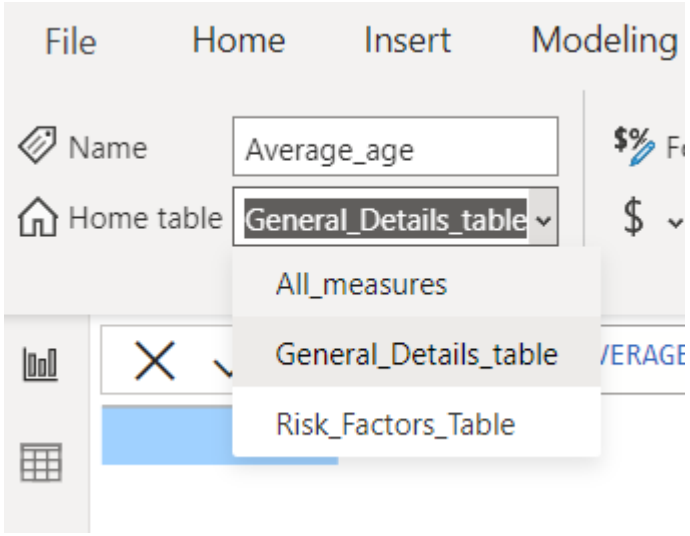
smoking_status	Average_age
formerly smoked	54.97
never smoked	46.73
smokes	47.04
Unknown	30.28
Total	**43.20**

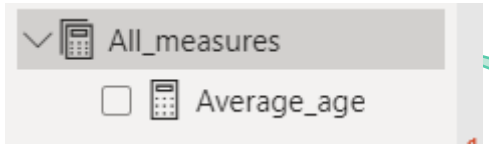
- Smoking status is in the second table, the average age is in the first table
- Both the tables are connected to each other, automatically the calculations are adjusted



Keep all the measures in a new table

- Home >>Enter Data >> Create a new table
- Click on Average Age measure >> Measure tools>> Home table >> All Measures





Drag and drop may not work

Identify the target variable

- Here our target variable is “Stroke”. A person was effected with stroke or not.
- Lets create a measure to calculate the overall stroke rate(stroke percentage). It can be used in the analysis later on.

```
1  Stroke_rate = SUM(Risk_Factors_Table[stroke])/count(Risk_Factors_Table[stroke])
```

Stroke_rate
4.87%

Overall stroke rate in this data is 4.87%

High risk Factor-1

- According to research people above age 55 are at higher risk in USA. Lets create a new column Age_high_risk
- Note- Here we can create a measure. But, measures work on the full data or full column. We have to fill the values of Age_high_risk row-wise
- For this operation, we have to create a new column, NOT a measure

```
Age_high_risk = if(General_Details_table[age]>55, "High Risk Age", "Low Risk Age")
```

To crate a frequency table

- Click on empty space on canvas >> Click on Table >> Select Age_high_risk >> Select id >> Convert id to count or count (distinct)
- To format the table
  - Click on the table >> Select format >> Style Presets >>Contrasting alternating rows

Age_high_risk	Count of id
High Risk Age	1692
Low Risk Age	3418
Total	**5110**

We will later verify whether Age is really an influencing factor on Stroke or not – later on

High Risk Factor-2

- **Gender** - Men are at a higher risk of stroke than women. Stoke probability is high for patients with **cardiovascular diseases.**
- Lets create a new column to capture patients containing both the characteristics.
- Gender is in the **first table**, Heart disease is in the **second table.** You have to use RELATED() function

High Risk Factor-2

- Access values from the second table using RELATED().

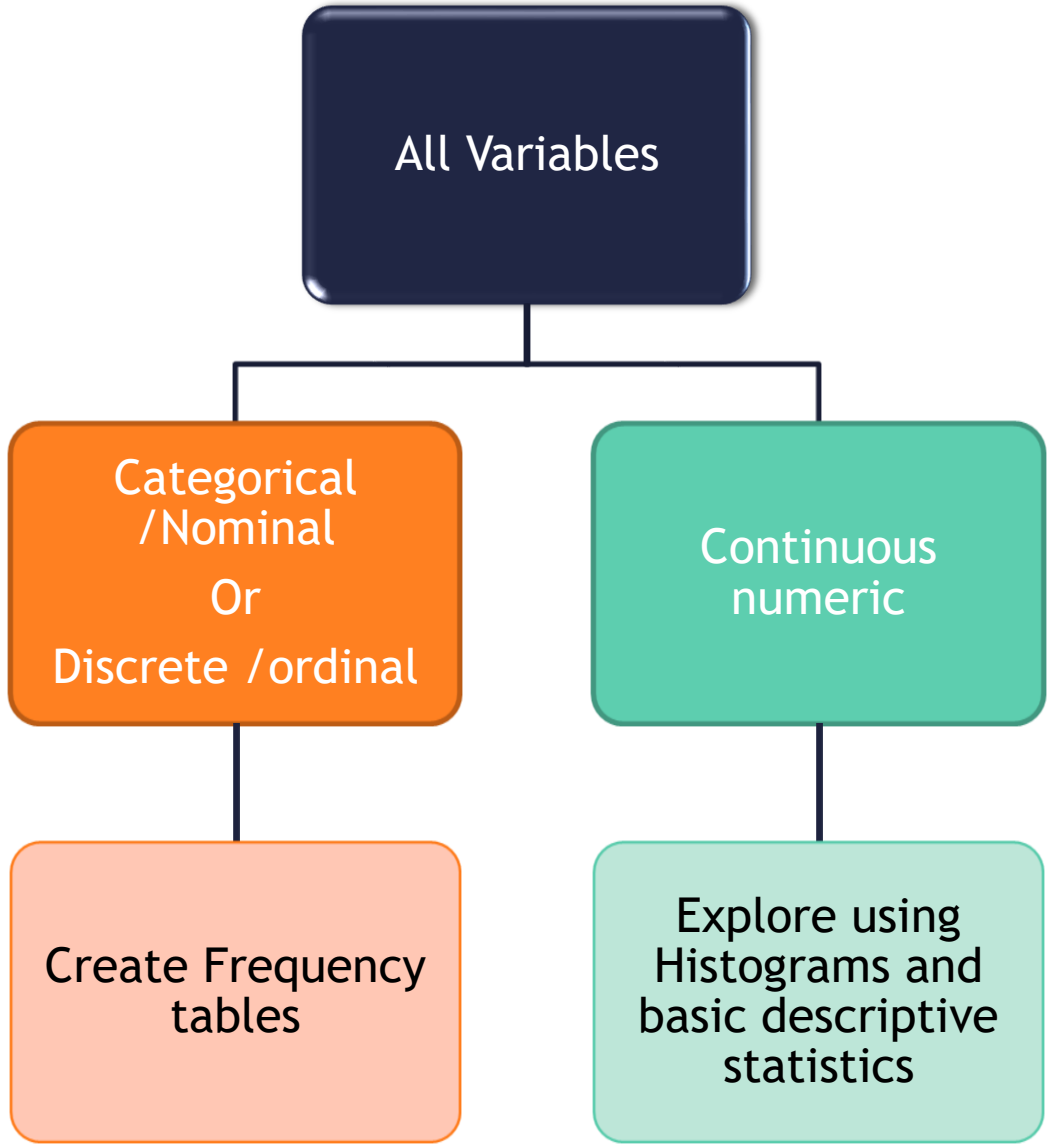
```
1  High_Risk_factor2 =
2  if(
3      (General_Details_table[gender]="Male" && RELATED(Risk_Factors_Table[heart_disease])=1),
4       "High Risk2",
5       "Low Risk2"
6     )
```

High_Risk_factor2	Count of id
High Risk2	161
Low Risk2	4949
Total	**5110**

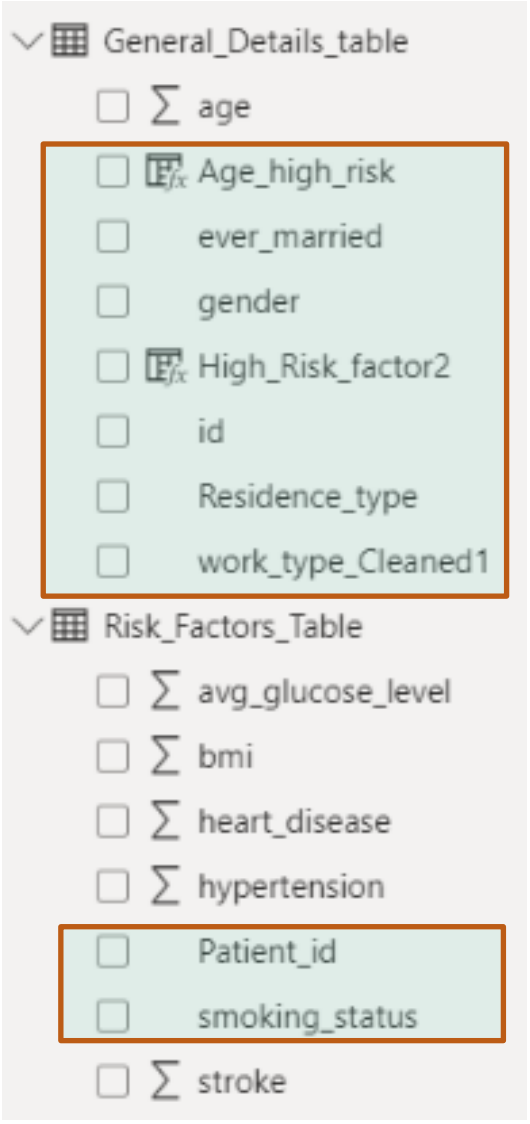
Data Exploration –Step by Step Process

1. Univariate Analysis – Explore individual variables
2. Bi-variate Analysis – Every variable vs. Target variable (Stroke)
3. Multi-variable analysis. – Multiple factors vs. Target variable (Stroke)

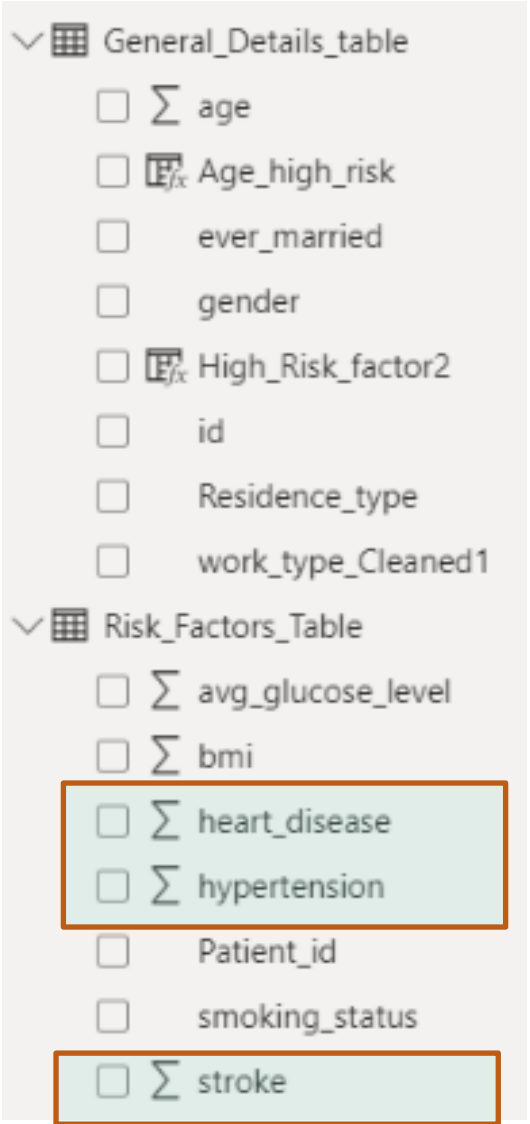
Univariate Analysis



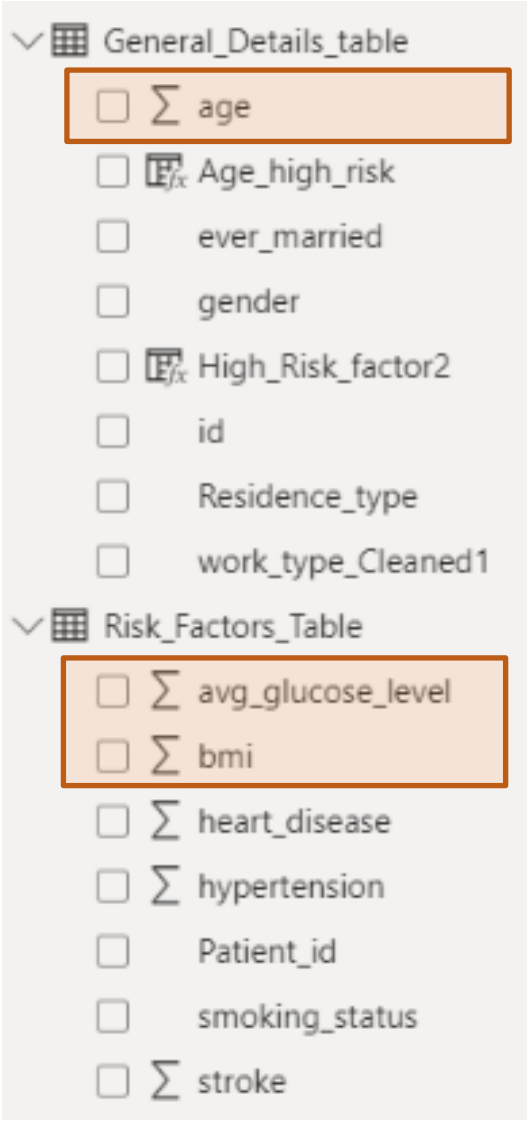
Categorical Variables



Discrete Variables



Continuous Variables



Next Step –
Step 5 - Univariate Analysis