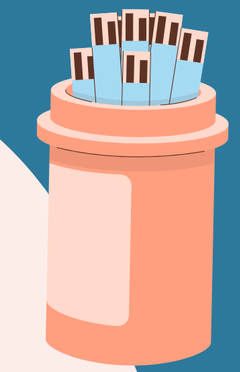
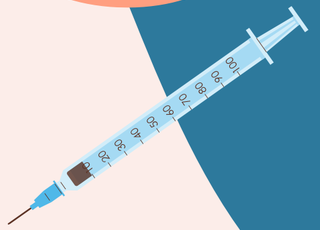
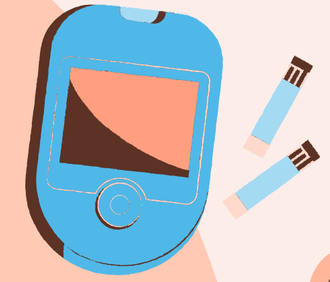




DIABETES PREDICTION ANALYSIS

Created by: PRATHAM SRIVASTAVA



DATASET

EmployeeName	Patient_id	gender	D.O.B	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes	Age	BMICategory
ARY JIMENEZ	PT102	Female	11 November 1992	0	0	No Info	27.32	6.6	80	0	32	Overweight
NGERINE BRIGHAM	PT388	Female	28 September 1997	0	0	No Info	27.32	6.6	140	0	27	Overweight
MES HARRIGAN	PT399	Female	02 October 1997	0	0	No Info	27.32	6.6	85	0	27	Overweight
SLIE COGAN	PT1369	Female	10 March 1999	0	0	No Info	27.32	6.6	145	0	25	Overweight
AETANO CALTAGIRONE	PT1604	Female	25 March 1999	0	0	No Info	27.32	6.6	160	0	25	Overweight
LACY BOES	PT1769	Female	02 April 1999	0	0	No Info	27.32	6.6	100	0	25	Overweight
ROOK MEBRAHTU	PT1832	Female	04 April 1999	0	0	No Info	27.32	6.6	85	0	25	Overweight
NOTA MARTINEZ	PT2056	Female	11 April 1999	0	0	No Info	27.32	6.6	126	0	25	Overweight
ARTER ROHAN	PT2061	Female	11 April 1999	0	0	No Info	27.32	6.6	140	0	25	Overweight
HN UPDIKE	PT2130	Female	12 April 1999	0	0	No Info	27.32	6.6	159	0	25	Overweight
YNTHIA MANUEL	PT2516	Female	22 April 1999	0	0	No Info	27.32	6.6	90	0	25	Overweight
MES KELLY	PT2677	Female	27 April 1999	0	0	No Info	27.32	6.6	145	0	25	Overweight
RRY DAVIS	PT2842	Female	01 May 1999	0	0	No Info	27.32	6.6	200	0	25	Overweight
JSAN KEARNEY	PT2937	Female	03 May 1999	0	0	No Info	27.32	6.6	200	0	25	Overweight
N GEE	PT2982	Female	04 May 1999	0	0	No Info	27.32	6.6	159	0	25	Overweight
AYNE CHAN	PT3362	Female	17 May 1999	0	0	No Info	27.32	6.6	100	0	25	Overweight
HOMAS FONG	PT3455	Female	19 May 1999	0	0	No Info	27.32	6.6	159	0	25	Overweight
IRA DE BERNARDI	PT3672	Female	23 May 1999	0	0	No Info	27.32	6.6	80	0	25	Overweight
YMOND DRISCOLL	PT3888	Female	29 May 1999	0	0	No Info	27.32	6.6	159	0	25	Overweight
ARTIN COVARRUBIAS	PT3895	Female	29 May 1999	0	0	No Info	27.32	6.6	145	0	25	Overweight
VENA HOLMES	PT4030	Female	01 June 1999	0	0	No Info	27.32	6.6	159	0	25	Overweight
AVID SUMMERHILL	PT4080	Female	02 June 1999	0	0	No Info	27.32	6.6	145	0	25	Overweight
IL HOPPER	PT4082	Female	02 June 1999	0	0	No Info	27.32	6.6	155	0	25	Overweight
AYNE CHEW	PT4207	Female	04 June 1999	0	0	No Info	27.32	6.6	200	0	25	Overweight
JDY URIBE	PT4848	Female	17 June 1999	0	0	No Info	27.32	6.6	159	0	25	Overweight
SS O'REILLY	PT5060	Female	22 June 1999	0	0	No Info	27.32	6.6	80	0	25	Overweight
HARTI MUNI	PT5219	Female	24 June 1999	0	0	No Info	27.32	6.6	100	0	25	Overweight
HANDRA JOHNSON	PT5455	Female	28 June 1999	0	0	No Info	27.32	6.6	160	0	25	Overweight
MIE DWYER	PT5530	Female	30 June 1999	0	0	No Info	27.32	6.6	145	0	25	Overweight

betesData (1,00,000 rows)

Data

Search

- DiabetesData
  - Age
  - AgeGroup
  - Average Gluco
  - AverageAge
  - AverageBMI
  - AverageHbA1
  - blood_glucose
  - BloodGlucoseS
  - bmi
  - BMICategory
  - CustomRiskFa
  - D.O.B
  - diabetes
  - DiabetesProba
  - DiabetesRiskS
  - DiabeticPercer
  - EmployeeNam
  - gender
  - HbA1c_level
  - heart_disease
  - hypertension
  - HypertensionS
  - Patient_id
  - PatientCount

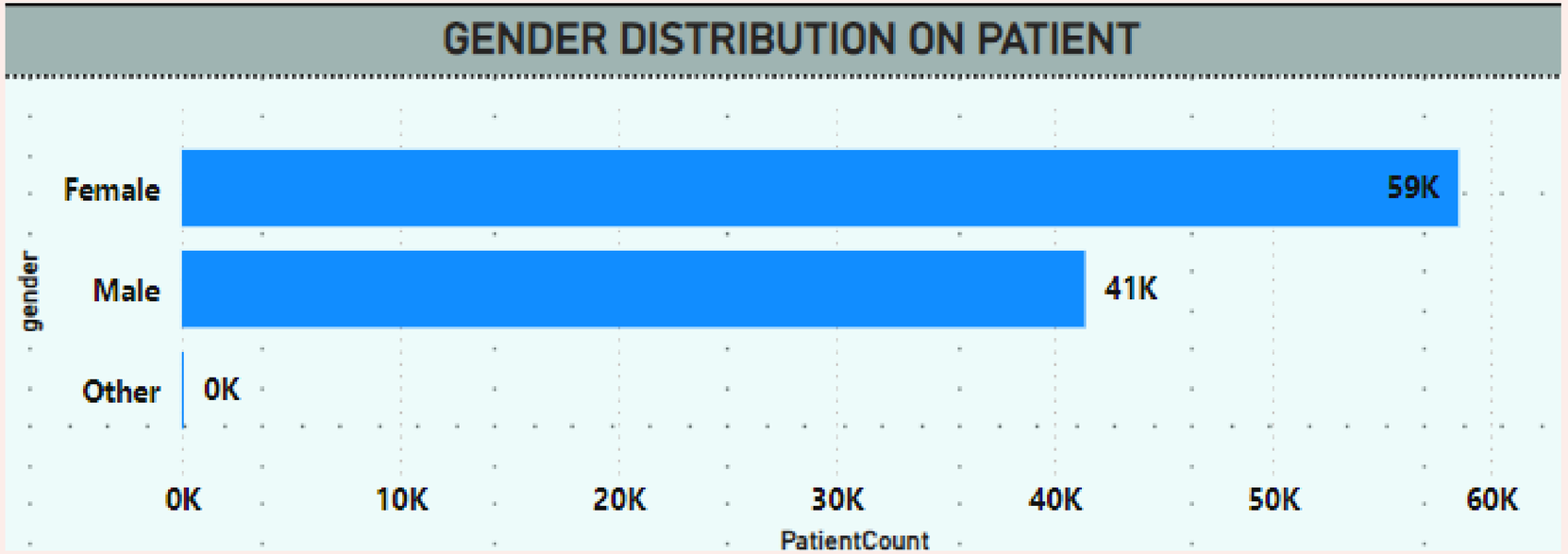
DATA CLEANSING: WHAT STEPS WOULD YOU TAKE TO CLEAN THE DATA, SUCH AS HANDLING MISSING VALUES OR OUTLIERS?

Data Cleansing

Steps to clean the data:

- Handling missing values: Identify columns with missing values using functions like isnull() in Python or conditional formatting in Excel. Handle missing data by removing records, filling with mean/median values, or using interpolation methods.
- Outliers: Identify outliers using visualization methods like box plots. Outliers can be handled by capping or removing extreme values, or transforming the data using log transformation.

BASIC VISUALIZATION: CREATE A SIMPLE BAR CHART TO SHOW THE DISTRIBUTION OF GENDERS IN THE DATASET.



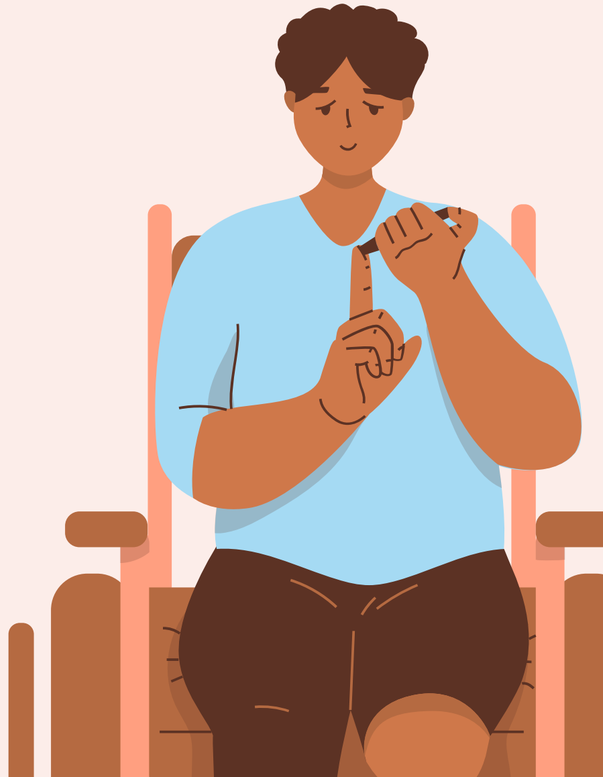
DAX INTRODUCTION: WHAT IS DAX, AND WHY IS IT USED IN POWER BI?

DAX Introduction:

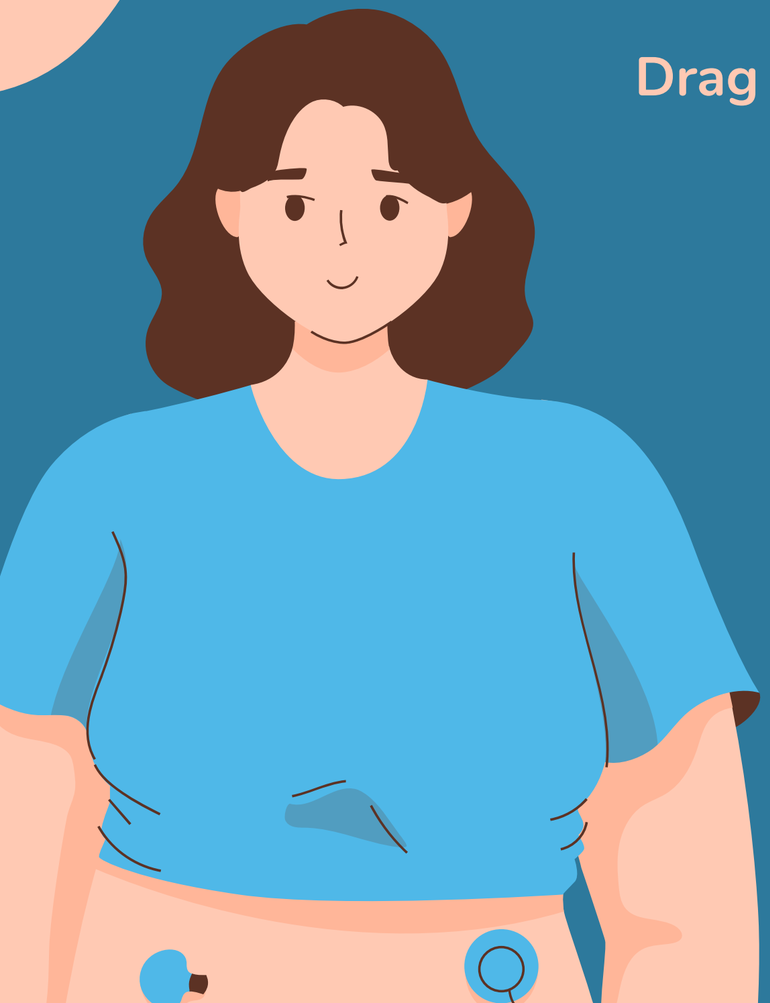
DAX (Data Analysis Expressions) is a collection of functions, operators, and constants used in Power BI to perform advanced calculations and create customized reports. DAX is essential for creating measures, calculated columns, and performing complex data aggregations in Power BI.

Calculated Columns: How can you create a calculated column in Power BI, and why might you need one?

- Calculated Columns in Power BI
- To create a calculated column in Power BI:
- Navigate to the Data View, and select New Column.
- Write a DAX expression, e.g., AgeGroup = IF([Age] >= 30 && [Age] <= 40, "30-40", "Other"). Calculated columns are used when you need new columns based on other fields in the dataset, often for categorization or grouping.
-

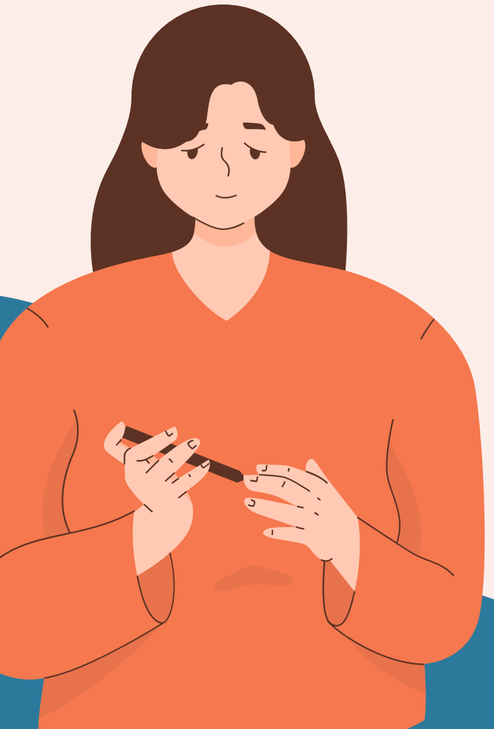


FILTERING DATA: EXPLAIN HOW TO FILTER DATA TO DISPLAY INFORMATION FOR EMPLOYEES AGED BETWEEN 30 AND 40.



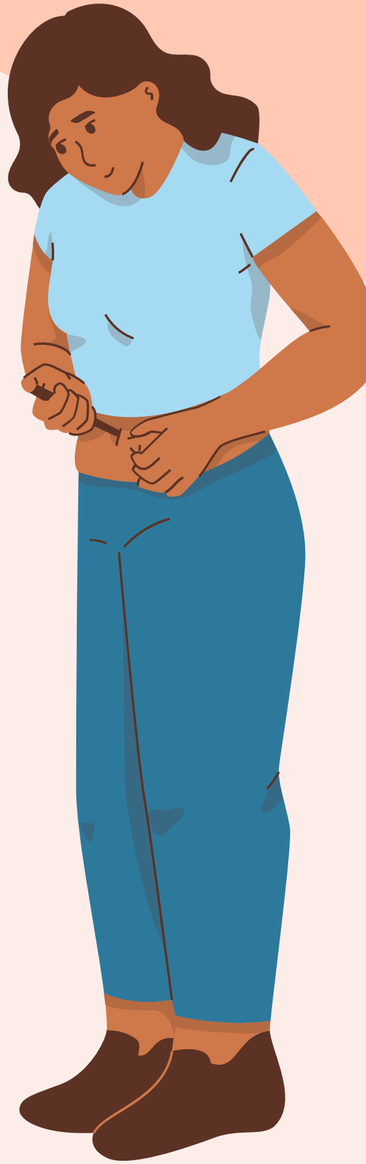
In Power BI, you can filter data using the Filter Pane:
Drag the Age field into the filter area and set the filter condition to show records where age is between 30 and 40.

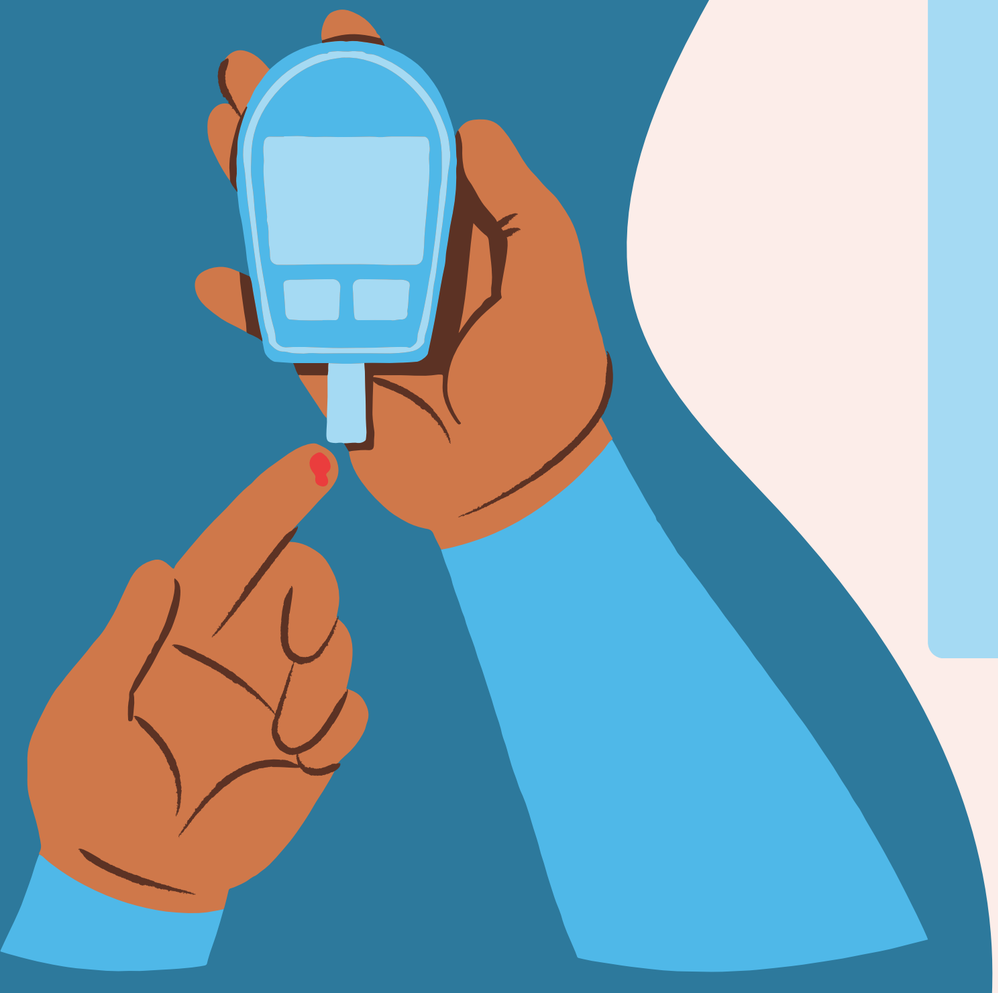
Joins: What is the difference between inner join and left join, and when would you use each in Power BI?



Inner Join vs. Left Join

- Inner Join: Returns only the rows where both tables have matching values.
- Left Join: Returns all rows from the left table and the matching rows from the right table. Unmatched rows will have NULL for columns from the right table. Use Inner Join when you only need matching data, and Left Join when you want to keep all data from one table.





MEASURES IN DAX: CREATE A DAX MEASURE TO CALCULATE THE AVERAGE BMI FOR EMPLOYEES WITH HYPERTENSION

```
Avg_BMI_Hypertension =
CALCULATE(AVERAGE(DiabetesData[BMI]),
DiabetesData[Hypertension] = 1)
.
```

TIME INTELLIGENCE: EXPLAIN HOW TO USE DAX TO CALCULATE THE YEAR-TO-DATE TOTAL FOR BLOOD GLUCOSE LEVELS

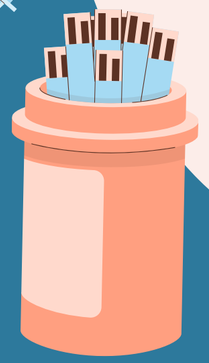
```
YTD_Blood_Glucose =
TOTALYTD(SUM(DiabetesData[BloodGlucose]),
DiabetesData[Date])
```

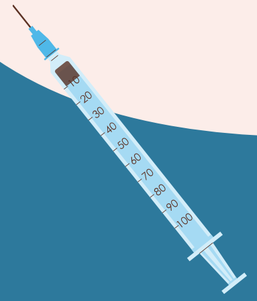
DATA AGGREGATION: WHAT IS THE PURPOSE OF SUMMARIZE IN DAX, AND HOW CAN YOU USE IT TO AGGREGATE DATA?

The SUMMARIZE function is used to aggregate data by grouping columns:

```
SummarizedTable = SUMMARIZE(DiabetesData,
DiabetesData[Age], "AverageBMI",
AVERAGE(DiabetesData[BMI]))
```

ROW-LEVEL SECURITY: HOW CAN YOU IMPLEMENT ROW-LEVEL SECURITY IN POWER BI TO RESTRICT ACCESS TO SENSITIVE EMPLOYEE DATA?





Row-Level Security

Implement Row-Level Security (RLS) by defining roles with specific filters on the data in Power BI. Use the Manage Roles feature to restrict data access based on conditions like Employee[Department] = "HR"

ADVANCED DAX: WRITE DAX CODE TO CALCULATE THE ROLLING AVERAGE OF BLOOD GLUCOSE LEVELS OVER A 3-MONTH PERIOD.

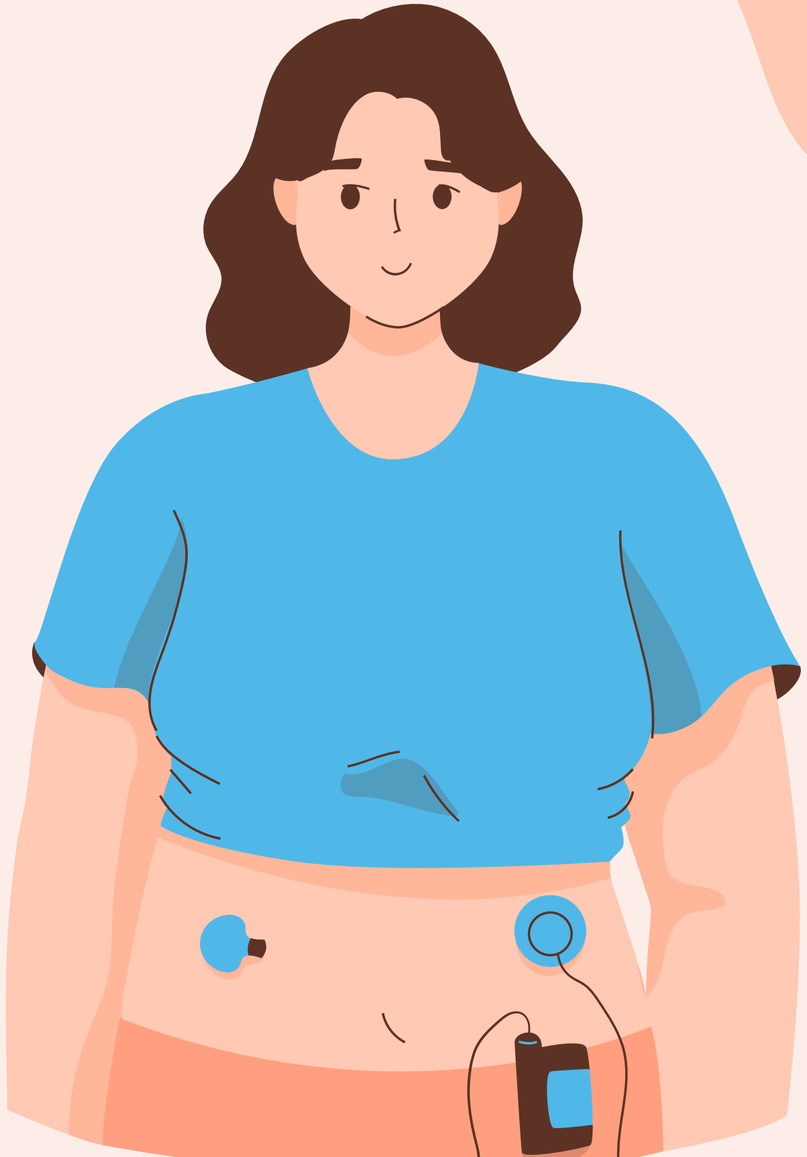
```
Rolling_Avg_Blood_Glucose =
CALCULATE(AVERAGE(DiabetesData[BloodGlucose]),
DATESINPERIOD(DiabetesData[Date],
LASTDATE(DiabetesData[Date]), -3, MONTH))
.
```

ADVANCED CALCULATIONS: CREATE A DAX MEASURE THAT CALCULATES THE PROBABILITY OF AN EMPLOYEE HAVING DIABETES BASED ON THEIR AGE, GENDER, AND OTHER FACTORS

```
Diabetes_Probability = IF([Age] >= 45 && [BMI] > 30 &&
[Gender] = "Male", 0.7, 0.2)
```

DASHBOARD

DIABETES PREDICTION ANALYSIS



DIABETES PREDICTION ANALYSIS

smoking_history: All

Age: All

AVERAGE GULUCOSE LEVEL

138.06

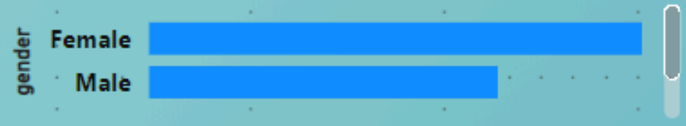
TOTAL PATIENT

100K

HYPERTENSION COUNT

7485

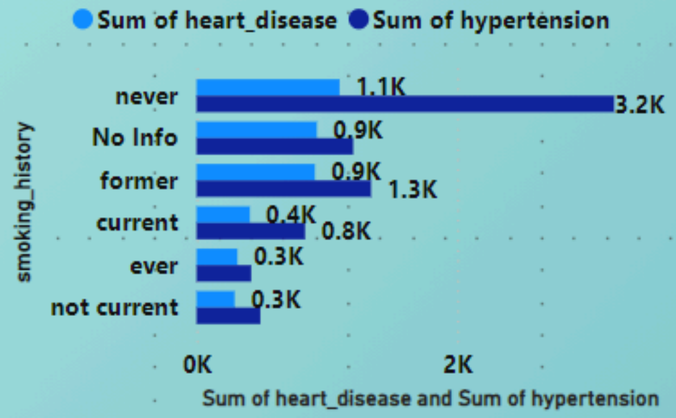
DIABETES RISK SCORE



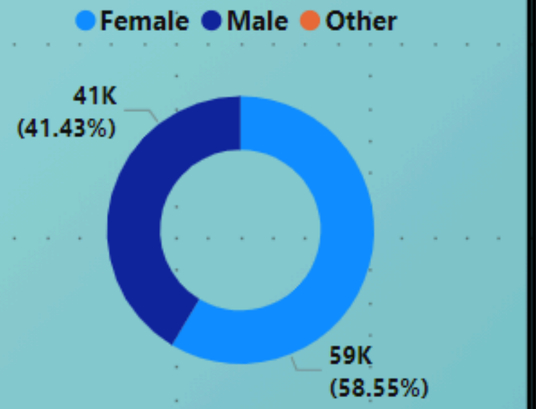
AGE VS BMI

AgeGroup	BMICategory	PatientCount
20-29	Overweight	45258
20-29	Obese	23484
20-29	Normal	21737
20-29	Underweight	8475
20-29	Unknown	810
60+	Overweight	81
60+	Obese	43
60+	Normal	41
Total		100000

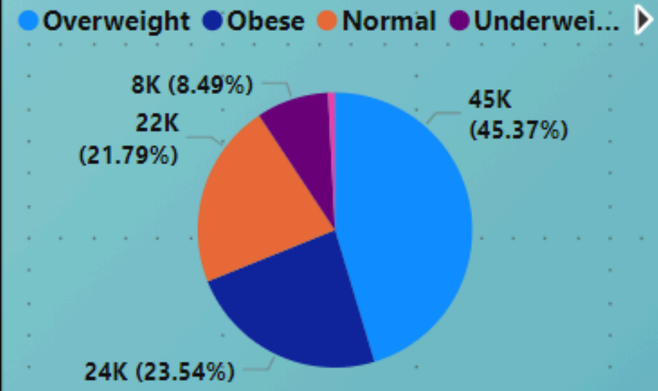
HYPERTENSION COUNT



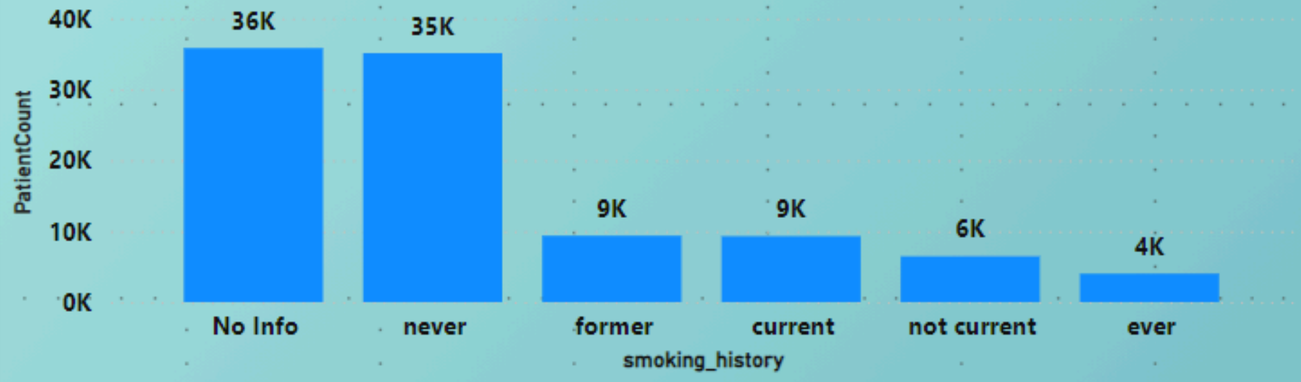
GENDER DISTRIBUTION



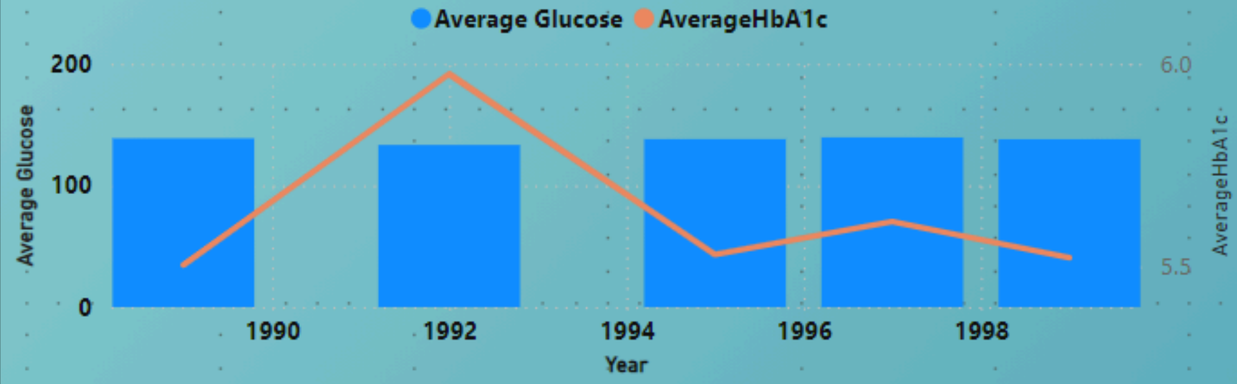
BMI CATEGORY DISTRIBUTION



PATIENT BY SMOKING HISTORY



HbA1c and Glucose Over Time



THANK YOU!