

# Title: Diabetes and Possible Interventions

Project 8 Group Members:

1. S2164604/1 ELAINE LI
2. S2174087/1 JUNHAO XU
3. S2175919/1 HAOJIE CHU
4. S2180377/1 YING MING TANG
5. S2157792/1 PENGWEI LI
6. S2174898/1 ZHENG WANG

## Objective:

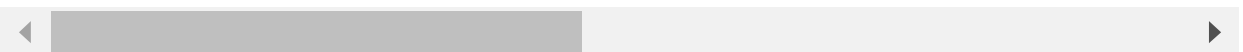
1. To predict the risk of having diabetes based on lifestyle.
2. To predict the level of diabetes (Type 1 diabetes, Type 2 diabetes, Prediabetes, Gestational Diabetes) based on lifestyle.
3. To predict the chance of having other diseases due to diabetes.

## Dataset:

Source:

[https://figshare.com/articles/dataset/EDDMQoL\\_Dataset\\_for\\_the\\_Study\\_on\\_Diabetes\\_Related\\_Distress/12345678](https://figshare.com/articles/dataset/EDDMQoL_Dataset_for_the_Study_on_Diabetes_Related_Distress/12345678)

700 adults of the three main ethnicity in Malaysia on regular follow-ups and medical care at three public health clinics. Participants were at least 30 years of age, known T2DM for more than one year at the time of participating in the study in 2013.



```
In [1]: import pandas as pd
```

```
In [2]: path='C:\\Users\\USER\\Documents\\WQD7003 Data Analytics\\Group Project\\Diabetes Data\\Diabetes Data.csv'
data=pd.read_csv(path)
type(data)
```

```
Out[2]: pandas.core.frame.DataFrame
```

```
In [3]: data.head()
```

Out[3]:

	CodeCentre	Dengkil1	CodeNumber	Age	AgeGroups	DiabetesDuration	DiabDuration3Cat	Gender
0	1	2	275	68	3	5	2	
1	1	2	112	65	3	33	3	
2	1	2	141	56	2	9	2	
3	1	2	295	61	3	5	2	
4	1	2	5	58	2	20	3	

5 rows × 106 columns

In [4]:

```
print(data.shape)
data.describe()
```

(700, 106)

Out[4]:

	CodeCentre	Dengkil1	CodeNumber	Age	AgeGroups	Gender	DDS2	Gender
count	700.000000	700.000000	700.000000	700.000000	700.000000	700.000000	700.000000	700.000000
mean	2.184286	2.328571	166.702857	59.614286	4.927143	4.751429	6.701429	7.964286
std	0.889638	0.756875	123.066867	51.337419	53.255226	65.277319	65.160071	75.190476
min	1.000000	1.000000	1.000000	31.000000	1.000000	0.000000	1.000000	1.000000
25%	1.000000	2.000000	66.750000	50.000000	1.000000	0.000000	1.000000	1.000000
50%	3.000000	3.000000	138.500000	57.000000	2.000000	0.000000	2.000000	2.000000
75%	3.000000	3.000000	283.000000	64.000000	3.000000	1.000000	3.000000	3.000000
max	3.000000	3.000000	434.000000	999.000000	999.000000	999.000000	999.000000	999.000000

8 rows × 54 columns