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:

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

In [2]:
```
path='C:\\Users\\USER\\Documents\\WQD7003 Data Analytics\\Group Project\\Diabetes Data
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	Gend
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	D
count	700.000000	700.000000	700.000000	700.000000	700.000000	700.000000	700.000000	700.00
mean	2.184286	2.328571	166.702857	59.614286	4.927143	4.751429	6.701429	7.96
std	0.889638	0.756875	123.066867	51.337419	53.255226	65.277319	65.160071	75.19
min	1.000000	1.000000	1.000000	31.000000	1.000000	0.000000	1.000000	1.00
25%	1.000000	2.000000	66.750000	50.000000	1.000000	0.000000	1.000000	1.00
50%	3.000000	3.000000	138.500000	57.000000	2.000000	0.000000	2.000000	2.00
75%	3.000000	3.000000	283.000000	64.000000	3.000000	1.000000	3.000000	3.00
max	3.000000	3.000000	434.000000	999.000000	999.000000	999.000000	999.000000	999.00

8 rows × 54 columns