

In [36]:

```
print("Name : ")
print("We will be cleaning the big data and make a comparison to show who has a healthier heart smokers OR non smokers, using a line graph")
print("Also we will be deriving which age group has the high chances of coronary heart disease in 10 years")
```

Name :

We will be cleaning the big data and make a comparison to show who has a healthier heart smokers OR non smokers, using a line graph

Also we will be deriving which age group has the high chances of coronary heart disease in 10 years

## Task 1 - Plot a line graph to show the difference between heart rate of smokers and non smokers

In [37]:

```
#Import Libraries
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
#read the csv
df = pd.read_csv("framingham.csv")
df
```

Out[37]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHypertension
0	1	39	4.0	0	0.0	0.0	0	0
1	0	46	2.0	0	0.0	0.0	0	0
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	0
4	0	46	3.0	1	23.0	0.0	0	0
...	...	...	...	...	...	...	...	...
4233	1	50	1.0	1	1.0	0.0	0	0
4234	1	51	3.0	1	43.0	0.0	0	0
4235	0	48	2.0	1	20.0	NaN	0	0
4236	0	44	1.0	1	15.0	0.0	0	0
4237	0	52	2.0	0	0.0	0.0	0	0

4238 rows × 16 columns

In [38]:

```
#Filter and make a new dataframe for non smokers
non_smoker = df.loc[df["currentSmoker"] == 0]
non_smoker
```

Out[38]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalent
0	1	39	4.0	0	0.0	0.0	0	
1	0	46	2.0	0	0.0	0.0	0	
5	0	43	2.0	0	0.0	0.0	0	
6	0	63	1.0	0	0.0	0.0	0	
8	1	52	1.0	0	0.0	0.0	0	
...	...	...	...	...	...	...	...	
4226	1	58	1.0	0	0.0	0.0	0	
4228	0	50	1.0	0	0.0	0.0	0	
4231	1	58	3.0	0	0.0	0.0	0	
4232	1	68	1.0	0	0.0	0.0	0	
4237	0	52	2.0	0	0.0	0.0	0	

2144 rows × 16 columns



In [39]:

```
#Group by age column and find average heart rate at different age  
age_heart = df.groupby("age")["heartRate"].mean().reset_index()  
age_heart
```

Out[39]:

	age	heartRate
0	32	80.000000
1	33	75.600000
2	34	73.555556
3	35	73.785714
4	36	74.761905
5	37	75.043478
6	38	76.631944
7	39	75.958580
8	40	76.518325
9	41	75.977011
10	42	75.522222
11	43	77.603774
12	44	74.867470
13	45	76.166667
14	46	78.027473
15	47	76.184397
16	48	76.028902
17	49	75.363636
18	50	76.457143
19	51	75.821918
20	52	76.255034
21	53	77.273381
22	54	75.598485
23	55	74.965517
24	56	73.902439
25	57	75.276423
26	58	76.675214
27	59	74.705882
28	60	76.099099
29	61	74.681818
30	62	74.454545
31	63	74.281818
32	64	76.706522
33	65	74.473684
34	66	78.868421
35	67	76.266667
36	68	80.277778

	age	heartRate
37	69	77.000000
38	70	64.000000

```
#Filter and make a new dataframe for smokers
smoker = df.loc[df["currentSmoker"] == 1]
smoker
```

Out[40]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHypertension
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	0
4	0	46	3.0	1	23.0	0.0	0	0
7	0	45	2.0	1	20.0	0.0	0	0
9	1	43	1.0	1	30.0	0.0	0	0
...	...	...	...	...	...	...	...	...
4230	0	56	1.0	1	3.0	0.0	0	0
4233	1	50	1.0	1	1.0	0.0	0	0
4234	1	51	3.0	1	43.0	0.0	0	0
4235	0	48	2.0	1	20.0	NaN	0	0
4236	0	44	1.0	1	15.0	0.0	0	0

2094 rows × 16 columns



In [41]:

```
#Group by age column and find average heart rate at different age  
age_heart = df.groupby("age")["heartRate"].mean().reset_index()  
age_heart
```

**Out[41]:**

	<b>age</b>	<b>heartRate</b>
<b>0</b>	32	80.000000
<b>1</b>	33	75.600000
<b>2</b>	34	73.555556
<b>3</b>	35	73.785714
<b>4</b>	36	74.761905
<b>5</b>	37	75.043478
<b>6</b>	38	76.631944
<b>7</b>	39	75.958580
<b>8</b>	40	76.518325
<b>9</b>	41	75.977011
<b>10</b>	42	75.522222
<b>11</b>	43	77.603774
<b>12</b>	44	74.867470
<b>13</b>	45	76.166667
<b>14</b>	46	78.027473
<b>15</b>	47	76.184397
<b>16</b>	48	76.028902
<b>17</b>	49	75.363636
<b>18</b>	50	76.457143
<b>19</b>	51	75.821918
<b>20</b>	52	76.255034
<b>21</b>	53	77.273381
<b>22</b>	54	75.598485
<b>23</b>	55	74.965517
<b>24</b>	56	73.902439
<b>25</b>	57	75.276423
<b>26</b>	58	76.675214
<b>27</b>	59	74.705882
<b>28</b>	60	76.099099
<b>29</b>	61	74.681818
<b>30</b>	62	74.454545
<b>31</b>	63	74.281818
<b>32</b>	64	76.706522
<b>33</b>	65	74.473684
<b>34</b>	66	78.868421
<b>35</b>	67	76.266667
<b>36</b>	68	80.277778

	age	heartRate
37	69	77.000000
38	70	64.000000

Conslusion -

## Task 2 - Which age group have high chances of having coronary heart disease in 10 years

In [42]:

```
#Read the csv
csv = pd.read_csv("framingham.csv")
#Filter and make a new dataframe for those who has chances of having coronary heart dise
filter_in_ten_yrs = csv.loc[csv["TenYearCHD"] == 1]
filter_in_ten_yrs
```

Out[42]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalen
3	0	61	3.0	1	30.0	0.0	0	
6	0	63	1.0	0	0.0	0.0	0	
15	0	38	2.0	1	20.0	0.0	0	
17	0	46	2.0	1	20.0	0.0	0	
25	1	47	4.0	1	20.0	0.0	0	
...	...	...	...	...	...	...	...	
4221	1	50	1.0	0	0.0	0.0	0	
4223	1	56	4.0	0	0.0	1.0	0	
4226	1	58	1.0	0	0.0	0.0	0	
4232	1	68	1.0	0	0.0	0.0	0	
4233	1	50	1.0	1	1.0	0.0	0	

644 rows × 16 columns





In [43]:

```
#Group by age column and count the rows of TenYearCHD column  
group_by_age_Ten_Year_CHD = csv.groupby("age")["TenYearCHD"].count().reset_index()  
group_by_age_Ten_Year_CHD
```

Out[43]:

	age	TenYearCHD
0	32	1
1	33	5
2	34	18
3	35	42
4	36	84
5	37	92
6	38	144
7	39	169
8	40	191
9	41	174
10	42	180
11	43	159
12	44	166
13	45	162
14	46	182
15	47	141
16	48	173
17	49	132
18	50	140
19	51	146
20	52	149
21	53	139
22	54	132
23	55	145
24	56	123
25	57	123
26	58	117
27	59	119
28	60	111
29	61	110
30	62	99
31	63	110
32	64	93
33	65	57
34	66	38
35	67	45
36	68	18

	age	TenYearCHD
37	69	7
In [44]:	70	2

```
#Plot a line graph to show total number of people having a chance of coronary heart disease
label = group_by_age_Ten_Year_CHD["age"]
value = group_by_age_Ten_Year_CHD["TenYearCHD"]

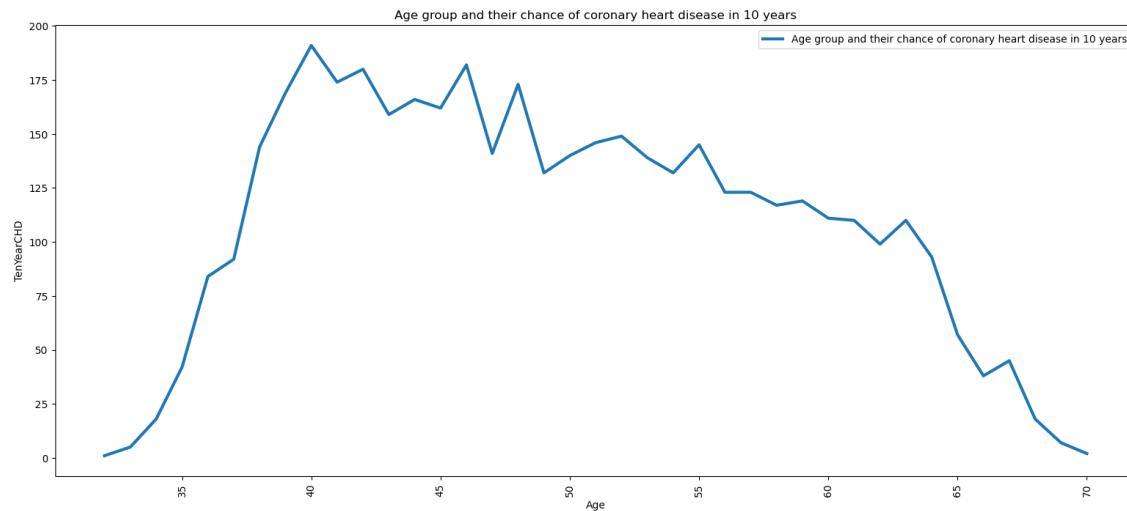
plt.subplots(figsize=(19,8))

plt.plot(label, value, label="Age group and their chance of coronary heart disease in 10 years")

plt.xlabel("Age")
plt.xticks(rotation="vertical")
plt.ylabel("TenYearCHD")
plt.title("Age group and their chance of coronary heart disease in 10 years")
plt.legend()
```

Out[44]:

<matplotlib.legend.Legend at 0x1bc478e8160>



Conslusion -

In [ ]: