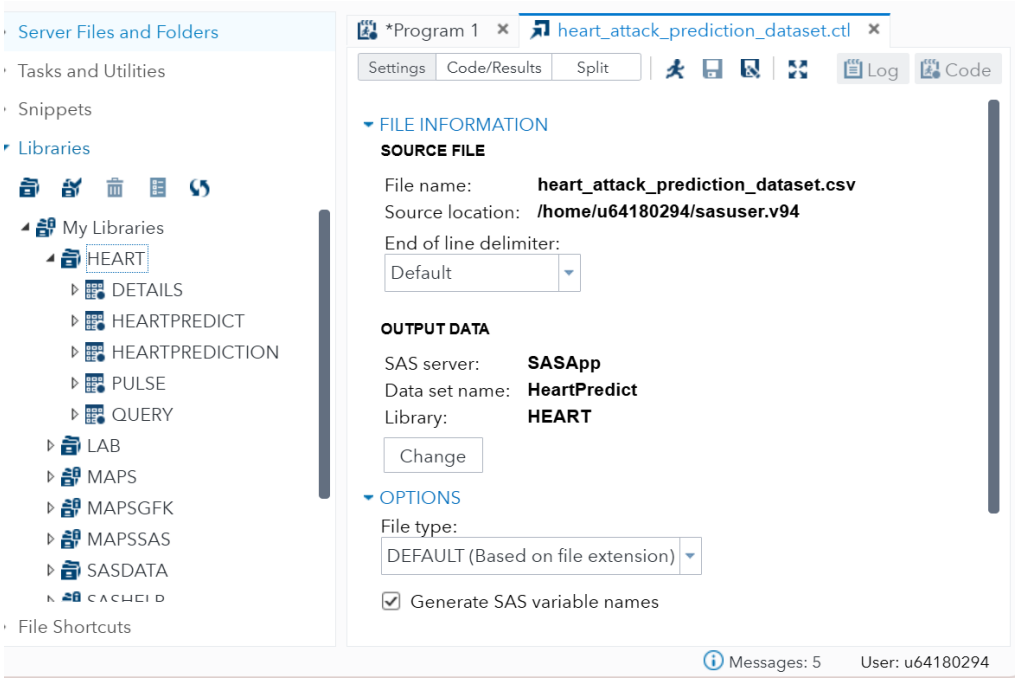
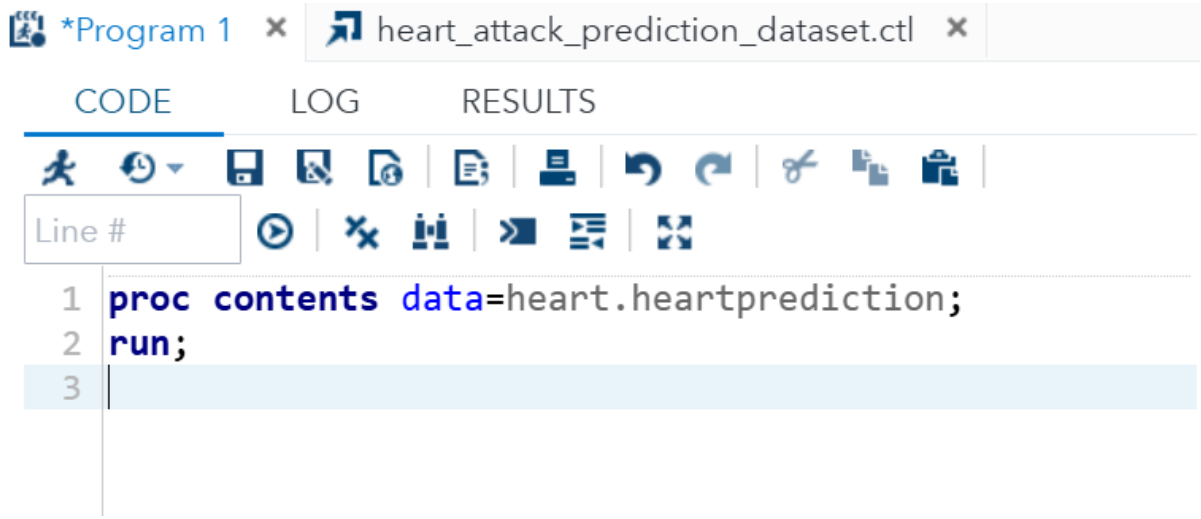
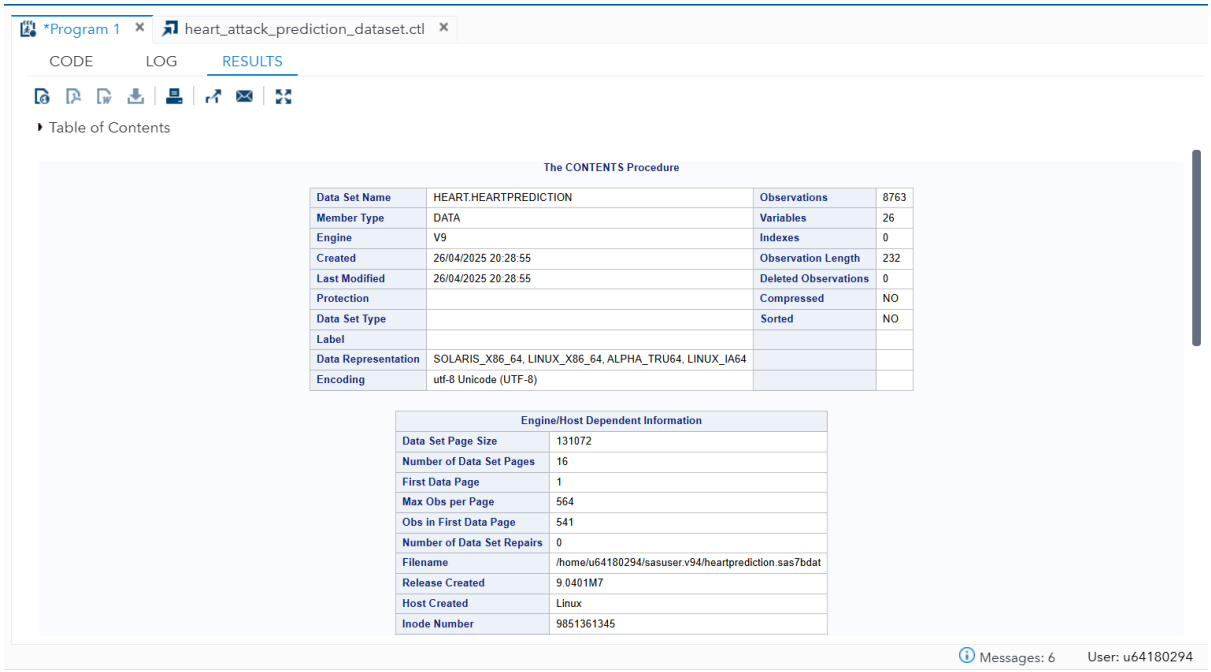
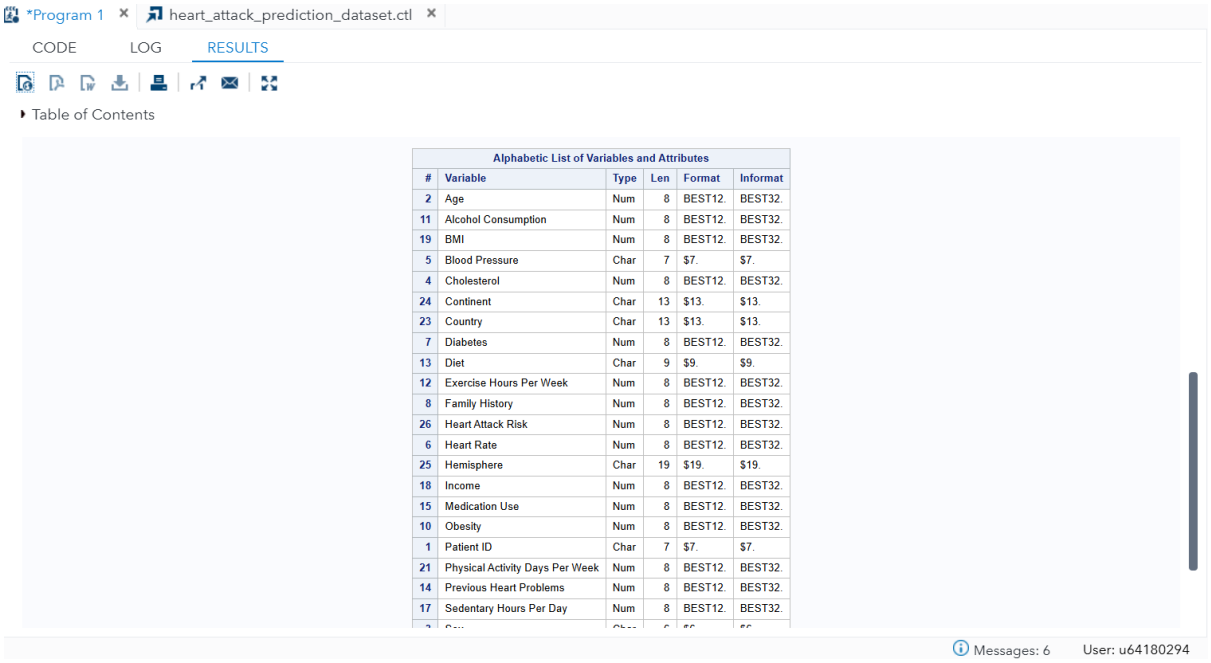
The first step before starting the research is to upload our dataset to our SAS program, then create a library for it so data and work get saved, and not discarded upon shutting down, then create name the dataset and choose the corresponding library to save it in.



As a starting point we must get a general idea on the dataset we're working on, so we start with displaying the variables names, types, and labels of the dataset:







Then the first 10 rows of the dataset to familiarize ourselves with the content and format:

```
4
5  proc print data=heart.heartprediction (obs=10);
6  run;
7
8
```

Obs	Patient ID	Age	Sex	Cholesterol	Blood Pressure	Heart Rate	Diabetes	Family History	Smoking	Obesity	Alcohol Consumption	Exercise Hours Per Week	Diet	Previous Heart Problems	Medication Use	Stress Level	Sedentary Hours Per Day	Income	
1	BMW7812	67	Male	208	158/88	72	0	0	1	0	0	4.1681888354	Average	0	0	9	6.6150014529	261404	31.2512
2	CZE1114	21	Male	389	165/93	98	1	1	1	1	1	1.8132416179	Unhealthy	1	0	1	4.9634588398	285768	27.1949
3	BNI9906	21	Female	324	174/99	72	1	0	0	0	0	2.0783529861	Healthy	1	1	9	9.463425838	235282	28.1765
4	JLN3497	84	Male	383	163/100	73	1	1	1	0	1	9.8281295935	Average	1	0	9	7.6489808245	125640	36.4647
5	GFO8847	66	Male	318	91/88	93	1	1	1	1	0	5.8042988203	Unhealthy	1	0	6	1.5148209264	160555	21.8091
6	ZOO7941	54	Female	297	172/86	48	1	1	1	0	1	0.6250080237	Unhealthy	1	1	2	7.7987524086	241339	20.1468
7	WYV0966	90	Male	358	102/73	84	0	0	1	0	1	4.098177091	Healthy	0	0	7	0.627356001	190450	28.8858
8	XXM0972	84	Male	220	131/68	107	0	0	1	1	1	3.4279287543	Average	0	1	4	10.543780239	122093	22.2218
9	XCQ5937	20	Male	145	144/105	68	1	0	1	1	0	16.868302239	Average	0	0	5	11.348786873	25086	35.8099
10	FTJ5456	43	Female	248	160/70	55	0	1	1	1	1	0.1945150606	Unhealthy	0	0	4	4.0551147818	209703	22.5589

Messages: 7 User: u64180294

*Task2.sas heart_attack_prediction_dataset.ctl

CODE LOG RESULTS

Table of Contents

y	Smoking	Obesity	Alcohol Consumption	Exercise Hours Per Week	Diet	Previous Heart Problems	Medication Use	Stress Level	Sedentary Hours Per Day	Income	BMI	Triglycerides	Physical Activity Days Per Week	Sleep Hours Per Day	Country	Continent	Hemisphere	Heart Attack Risk
0	1	0	0	4.1681888354	Average	0	0	9	6.6150014529	261404	31.251232725	286	0	6	Argentina	South America	Southern Hemisphere	0
1	1	1	1	1.8132416179	Unhealthy	1	0	1	4.9634588398	285768	27.194973352	235	1	7	Canada	North America	Northern Hemisphere	0
0	0	0	0	2.0783529861	Healthy	1	1	9	9.463425838	235282	28.176570684	587	4	4	France	Europe	Northern Hemisphere	0
1	1	0	1	9.8281295935	Average	1	0	9	7.6489808245	125640	36.464704293	378	3	4	Canada	North America	Northern Hemisphere	0
1	1	1	0	5.8042988203	Unhealthy	1	0	6	1.5148209264	160555	21.809144181	231	1	5	Thailand	Asia	Northern Hemisphere	0
1	1	0	1	0.6250080237	Unhealthy	1	1	2	7.7987524086	241339	20.146839503	795	5	10	Germany	Europe	Northern Hemisphere	1
0	1	0	1	4.098177091	Healthy	0	0	7	0.627356001	190450	28.885810607	284	4	10	Canada	North America	Northern Hemisphere	1
0	1	1	1	3.4279287543	Average	0	1	4	10.543780239	122093	22.221861739	370	6	7	Japan	Asia	Northern Hemisphere	1
0	1	1	0	16.868302239	Average	0	0	5	11.348786873	25086	35.809901319	790	7	4	Brazil	South America	Southern Hemisphere	0
1	1	1	1	0.1945150606	Unhealthy	0	0	4	4.0551147818	209703	22.558916752	232	7	7	Japan	Asia	Northern Hemisphere	0

Messages: 7 User: u64180294

Then to perform descriptive statistics where we use “proc means” which is a summarization tool to compute the descriptive statistics across all observations, our goal is to identify median, mean, standard deviation along with minimum and maximum value of over variables to understand what we’re working with, understand variable type and ranges to analyze and identify patterns within:

```
8  proc means data=heart.heartprediction mean median std min max;
9  var Age BMI Cholesterol 'Heart Rate'n Diabetes Obesity Cholesterol Triglycerides;
10 run;
11
```

The MEANS Procedure

Variable	Mean	Median	Std Dev	Minimum	Maximum
Age	53.7079767	54.0000000	21.2495088	18.0000000	90.0000000
BMI	28.8914459	28.7689994	6.3191813	18.0023366	39.9972108
Cholesterol	259.8772110	259.0000000	80.8632761	120.0000000	400.0000000
Heart Rate	75.0216821	75.0000000	20.5509479	40.0000000	110.0000000
Diabetes	0.6522880	1.0000000	0.4762712	0	1.0000000
Obesity	0.5014265	1.0000000	0.5000265	0	1.0000000
Triglycerides	417.6770512	417.0000000	223.7481368	30.0000000	800.0000000

The output shows the mean/average, for instance average age from dataset population is 54 with ages ranging from 18 (minimum field) to 90 (maximum field), indicating majority of people in this sample population are middle-aged or older, a 21.2 standard deviation indicates a relatively spread-out data distribution from the mean

Additionally, diabetes mean suggests that 65.2%. of population has diabetes Similarly, obesity is 0.501, indicating that 50% or half the overall population is obese, which raises an important question, are those two play as key factors for a high risk of a heart attack or heart attack related conditions.

Furthermore, the Triglycerides which represent the level of fat in blood stream mean 417 which is alarming considering the normal level is150 mg/dL.

Next step is using “Proc Freq” to summarize categorial variables, we use that procedure to display frequency or the number of times a certain value occurred (“freq” in the query), percent, cumulative frequency, total of current frequency along with previous ones, well as cumulative percent, which is tracking the percentage total that’s been accumulated so far as we progress through the data.

Then the presentation format is chosen, here we chose table, followed by the variable which we want to display, as “blood pressure” is the only variable in our current query that has a space we use ‘’n to indicate it’s a name literal indicating the actual name contains a space:

```
12 proc freq data=heart.heartprediction;
13 table Sex 'Blood Pressure'n Diet Country Continent Hemisphere;
14 run;
15
16
```

The FREQ Procedure

Sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	2652	30.26	2652	30.26
Male	6111	69.74	8763	100.00

.

Blood_Pressure	Frequency	Percent	Cumulative Frequency	Cumulative Percent
100/100	2	0.02	2	0.02
100/102	1	0.01	3	0.03
100/103	4	0.05	7	0.08
100/104	4	0.05	11	0.13
100/105	4	0.05	15	0.17
100/106	3	0.03	18	0.21
100/107	1	0.01	19	0.22
100/108	2	0.02	21	0.24
100/109	2	0.02	23	0.26
100/110	2	0.02	25	0.29
100/60	4	0.05	29	0.33
100/61	1	0.01	30	0.34
100/63	3	0.03	33	0.38
100/64	1	0.01	34	0.39
100/65	3	0.03	37	0.42

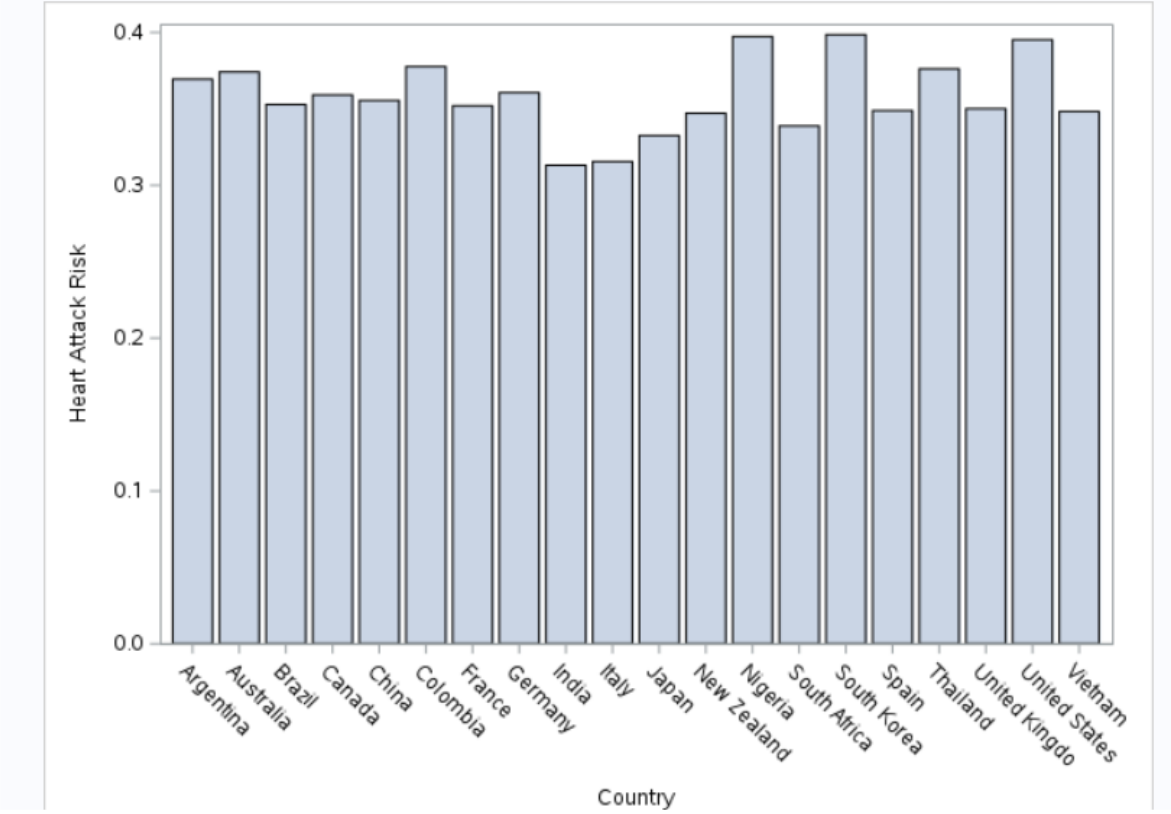
Here we notice the number of males in the dataset population is almost three times higher than the females, it's followed by a table displaying blood pressure values, how frequent is it, the percentage of the population that falls within followed by cumulative frequency and cumulative percentage

Diet	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Average	2912	33.23	2912	33.23
Healthy	2960	33.78	5872	67.01
Unhealthy	2891	32.99	8763	100.00

We then have a table with the same attributes yet to measure the diet, categorizing it into records of average, healthy, and unhealthy. And we observe that they are all relatively even.

```
proc means data=heart.heartprediction mean;
class Country;
var 'Heart Attack Risk'n;
run;
```

```
proc sgplot data=heart.heartprediction;
vbar Country / response= 'Heart Attack Risk'n stat=mean;
    xaxis label= "Country";
    yaxis label= "Heart Attack Risk";
run;
```



The MEANS Procedure

Analysis Variable : Heart Attack Risk		
Country	N Obs	Mean
Argentina	471	0.3694268
Australia	449	0.3741648
Brazil	462	0.3528139
Canada	440	0.3590909
China	436	0.3555046
Colombia	429	0.3776224
France	446	0.3520179
Germany	477	0.3605870
India	412	0.3131068
Italy	431	0.3155452
Japan	433	0.3325635
New Zealand	435	0.3471264
Nigeria	448	0.3973214
South Africa	425	0.3388235
South Korea	409	0.3985330
Spain	430	0.3488372
Thailand	428	0.3761682
United Kingdo	457	0.3501094
United States	420	0.3952381
Vietnam	425	0.3482353

We then observe the countries, Argentina is the most reoccurring value, closely followed by Australia, then Brazil. However, the countries with the highest are South Korea, Nigeria, and the United States.

Continent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Africa	873	9.96	873	9.96
Asia	2543	29.02	3416	38.98
Australia	884	10.09	4300	49.07
Europe	2241	25.57	6541	74.64
North America	860	9.81	7401	84.46
South America	1362	15.54	8763	100.00

Hemisphere	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Northern Hemisphere	5660	64.59	5660	64.59
Southern Hemisphere	3103	35.41	8763	100.00

Here we notice that the most frequent continents, or majority of the population reside in countries located in Asia, followed by Europe by a mere 302 and then South America. Additionally, the majority are in the Northern Hemisphere by 64.59%.

As number of males in the population is almost 3 times the female population, so we compare portion mean rather than total mean to draw a fairer theory:

```
17 proc means data=heart.heartprediction mean;
18 class Sex;
19 var Diabetes Obesity Cholesterol Triglycerides 'Heart Attack Risk'n;
20 run;
21
22
23
```

The MEANS Procedure

Sex	N Obs	Variable	Mean
Female	2652	Diabetes	0.6496983
		Obesity	0.4996229
		Cholesterol	258.9426848
		Triglycerides	416.6809955
		Heart Attack Risk	0.3559578
Male	6111	Diabetes	0.6534119
		Obesity	0.5022091
		Cholesterol	260.2827688
		Triglycerides	418.1093111
		Heart Attack Risk	0.3591883

The query output indicates that both means are relatively similar in all aspects, diabetes mean is evenly distributed within both as well as obesity, even though the male population is slightly higher, yet the difference isn't big enough to make an effect here, both cholesterol levels are similar and both are higher than the normal which is 200, and the same thing is observed in the triglycerides levels both are similar yet dangerously higher from the average of 150 hence both genders in the population are at 35% risk of a heart attack.

We run the following query to group population by income to identify whether low income is a factor indirectly resulting in a heart attack risk:

```
proc sql;
    select
        case when Income < 25000 the 'Low Income'
        when Income between 25000 and 74999 then 'Mid Income'
        else 'High Income' end as Income_Group,
    mean('Heart Attack Risk'n) as Avg_Heart_Attackk_Rate
from heart.heartprediction
group by Income_Group;
quit;
```

Income_Group	Avg_Heart_Attackk_Rate
High Income	0.359424
Low Income	0.327778
Mid Income	0.356278

The output table suggests that the average risk of heart attack is highest individuals with high income 35.9%, with Mid income falling close behind with 35.6%, while individuals who fall within the low-income group have a lower percentage with average risk of 32.8%. which indicates income might not be a major or direct effect of inducing a heart attack, for instance by affecting healthcare access and diet nutrients an individual intake.

As income doesn't have a major effect, we then check whether family history highly affects the possibility of a heart attack

```
proc Sql;
    select
        'Family History'n as Family_History,
        mean('Heart Attack Risk'n)*100 as Heart_Attack_Risk_Risk_Percent
        from heart.heartprediction
        group by 'Family History'n;
quit;
```

Family_History	Heart_Attack_Risk_Risk_Percent
0	35.89917
1	35.74074

The table suggests that family history of heart attacks doesn't majorly affect the risk of having a heart attack as those without heart attack family history have a 35.8% chance, while those who do are at 35.7% risk of it.

Additionally, we assess the risk of heart attacks caused by the diet for people in the population

```
proc means data=heart.heartprediction mean;
    class Diet;
    var 'Heart Attack Risk'n;
run;
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk		
Diet	N Obs	Mean
Average	2912	0.3523352
Healthy	2960	0.3645270
Unhealthy	2891	0.3576617

Evidently, the diet the population follows doesn't solely contribute to the risk of a heart attack as the group at highest risk is the one following a healthy diet with a risk percentage of 36.4% followed by the unhealthy diet group with a 35.7% risk then the average diet group with a 35.2% risk.

Next, we observe if the heart attack risk percentage is affected by smoking

```
40 proc means data=heart.heartprediction mean;
41     class Smoking;
42     variable 'Heart Attack Risk'n;
43 run;
44
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk		
Smoking	N Obs	Mean
0	904	0.3639381
1	7859	0.3575519

The observed result indicates that the population who smoke are slightly less heart attack risk dropping to 35.7% than those within the population who don't who are at 36.3%

and to confirm lifestyle isn't independently a major factor, we check alcohol consumption with the risk of a heart attack

```
45 proc means data=heart.heartprediction mean;
46     class 'Alcohol Consumption'n;
47     var 'Heart Attack Risk'n;
48 run;
49
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk		
Alcohol Consumption	N Obs	Mean
0	3522	0.3662692
1	5241	0.3527953

The observed result indicates that the population who smoke have a slightly lower heart attack risk falling about 35.7% than those within the population who don't are at 36.3%.

Results were relatively similar to smoking percentage with alcohol consumers aa 35.2% while the population sample the doesn't is at 36.6 % risk of a heart attack so consuming alcohol or smoking on its own isn't enough to increase the risk of a heart attack

And checking the percentage of risk for the population that both drink alcohol and smoke

```
50 proc means data=heart.heartprediction mean;
51     class 'Alcohol Consumption'n Smoking;
52     var 'Heart Attack Risk'n;
53 run;
54
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk			
Alcohol Consumption	Smoking	N Obs	Mean
0	0	380	0.3894737
	1	3142	0.3634628
1	0	524	0.3454198
	1	4717	0.3536146

This confirms that lifestyle isn't a necessity for increasing the risk as the highest risk percentage is for the sample of the population that doesn't consume alcohol nor smoke with 38.9%. hence countering the assumption that lifestyle factors such as drinking alcohol and smoking alone aren't key factors to determining the risk of a heart attack, but possibly a cumulative effect of lifestyle. Diet, stress and genetic may greatly affect in increasing a heart attack risk.

As a final confirmation to all the previously proposed assumptions we extract the number of people who smoke, consume alcohol, have family history and have an unhealthy diet

```
60 proc sql;
61     select
62         count(*) as N,
63         mean('Heart Attack Risk'n) as Risk_Percentage
64     from heart.heartprediction
65     where Smoking = 1
66       and 'Alcohol Consumption'n = 1
67       and Diet = "Unhealthy"
68       and 'Family History'n = 1;
69 quit;
70
```

N	Risk_Percentage
782	35.29%

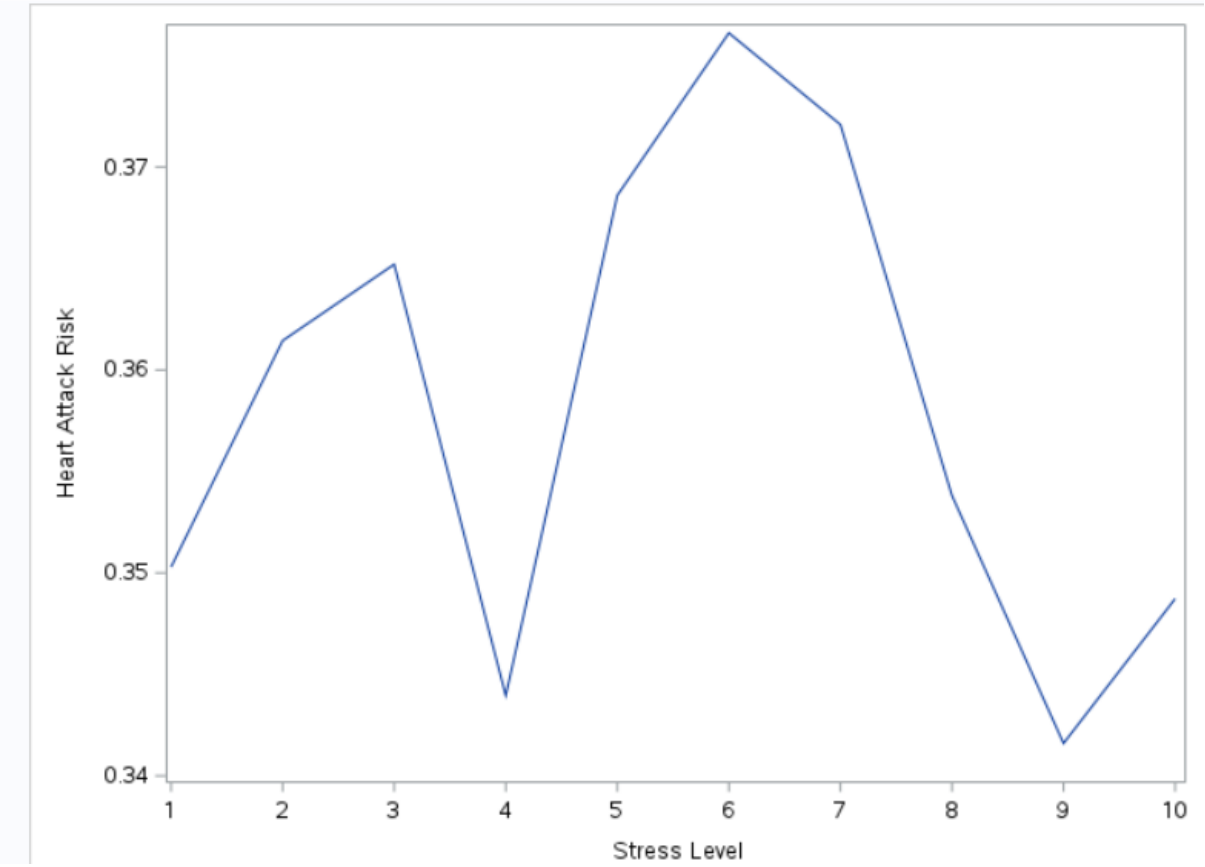
And results prove us with a sample of 782 people, with a sample mean of 35.29% heart attack risk percentage proving that the lifestyle on its own along with family history isn't a grand factor in increasing the risk of heart attacks yet other factors might come into play such as the overall environment, sleep hours and stress levels hence we check them next.

```
proc means data=heart.heartprediction;
class 'Stress level'n;
var 'Heart Attack Risk'n;
run;

proc sgplot data=heart.heartprediction;
    vline 'Stress Level'n / response='Heart Attack Risk'n stat=mean;
    xaxis label= "Stress Level";
    yaxis label= "Heart Attack Risk";
run;
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk						
Stress Level	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	865	865	0.3502890	0.4773364	0	1.0000000
2	913	913	0.3614458	0.4806826	0	1.0000000
3	868	868	0.3652074	0.4817659	0	1.0000000
4	910	910	0.3439560	0.4752878	0	1.0000000
5	860	860	0.3686047	0.4827072	0	1.0000000
6	855	855	0.3766082	0.4848189	0	1.0000000
7	903	903	0.3720930	0.4836309	0	1.0000000
8	879	879	0.3538111	0.4784237	0	1.0000000
9	887	887	0.3416009	0.4745140	0	1.0000000
10	823	823	0.3487242	0.4768563	0	1.0000000



The group with the highest risk percentage is the one with stress level of 6 with a 37.6% risk percent followed by stress level 7 with 37.2% risk chance of a heart attack.

Then we assess the risk percentage of the population that uses medication and of those with pervious heart problems

```
77 proc means data=heart.heartprediction;
78 class 'Medication Use'n;
79 var 'Heart Attack Risk'n;
80 run;
81 proc means data=heart.heartprediction;
82 class 'previous Heart Problems'n;
83 var 'Heart Attack Risk'n;
84 run;
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk						
Medication Use	N Obs	N	Mean	Std Dev	Minimum	Maximum
0	4396	4396	0.3571429	0.4792119	0	1.0000000
1	4367	4367	0.3592856	0.4798460	0	1.0000000

The MEANS Procedure

Analysis Variable : Heart Attack Risk						
Previous Heart Problems	N Obs	N	Mean	Std Dev	Minimum	Maximum
0	4418	4418	0.3580806	0.4794903	0	1.0000000
1	4345	4345	0.3583429	0.4795688	0	1.0000000

Heart attack risk for the population that uses medication is slightly higher, by a mere 0.21% as the sample that uses medication is at 35.92 risk while the sample that doesn't has a 35.71% risk. Furthermore, the effect pervious heart problems have on the risk is very minimal, only a 0.026% difference, where the sample who had experienced heart problems has a 35.83% risk, while those who didn't have a 35.80% risk.

The physical activity effect is evaluated next

```
86 proc means data=heart.heartprediction;
87 class 'Physical Activity Days Per Week'n;
88 var 'Heart Attack Risk'n;
89 run;
90
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk						
Physical Activity Days Per Week	N Obs	N	Mean	Std Dev	Minimum	Maximum
0	1065	1065	0.3887324	0.4876913	0	1.0000000
1	1121	1121	0.3461195	0.4759442	0	1.0000000
2	1109	1109	0.3543733	0.4785388	0	1.0000000
3	1143	1143	0.3508311	0.4774391	0	1.0000000
4	1077	1077	0.3574745	0.4794788	0	1.0000000
5	1079	1079	0.3419833	0.4745940	0	1.0000000
6	1074	1074	0.3528864	0.4780904	0	1.0000000
7	1095	1095	0.3744292	0.4841963	0	1.0000000

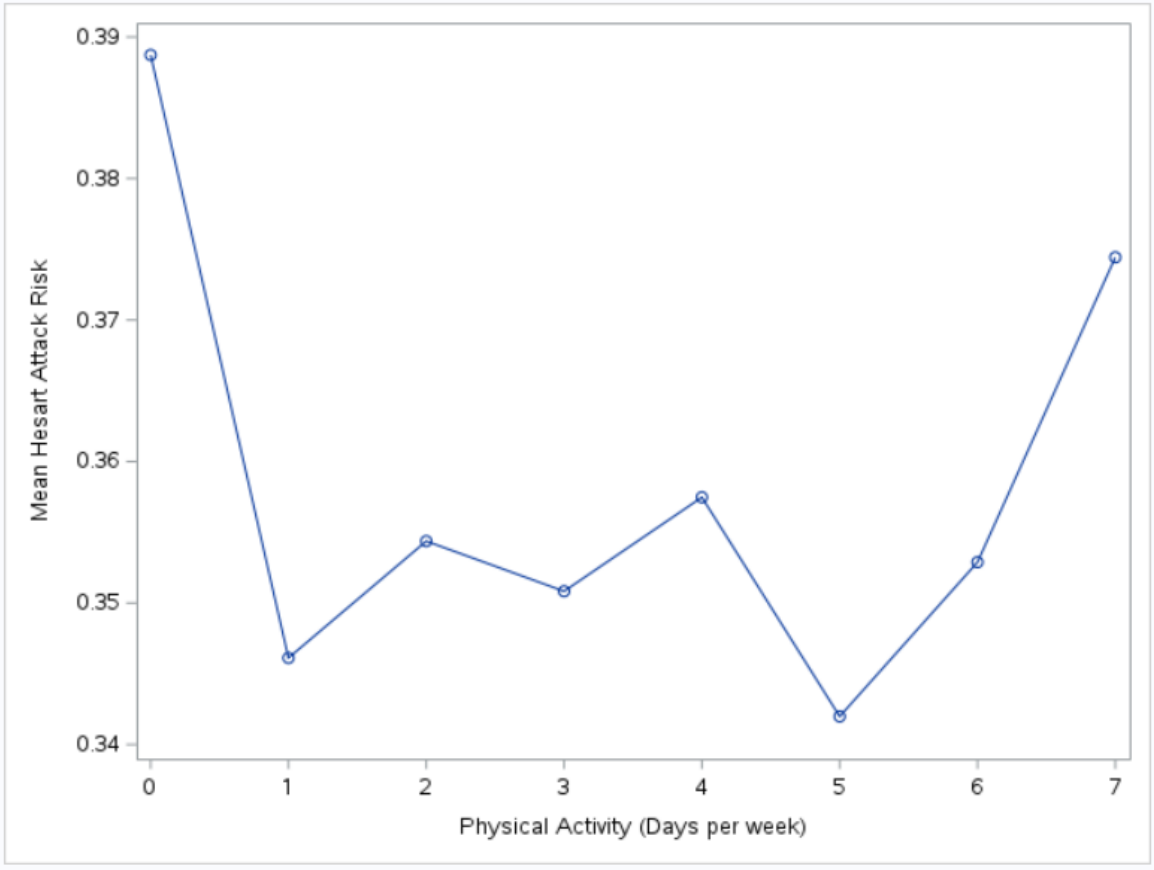


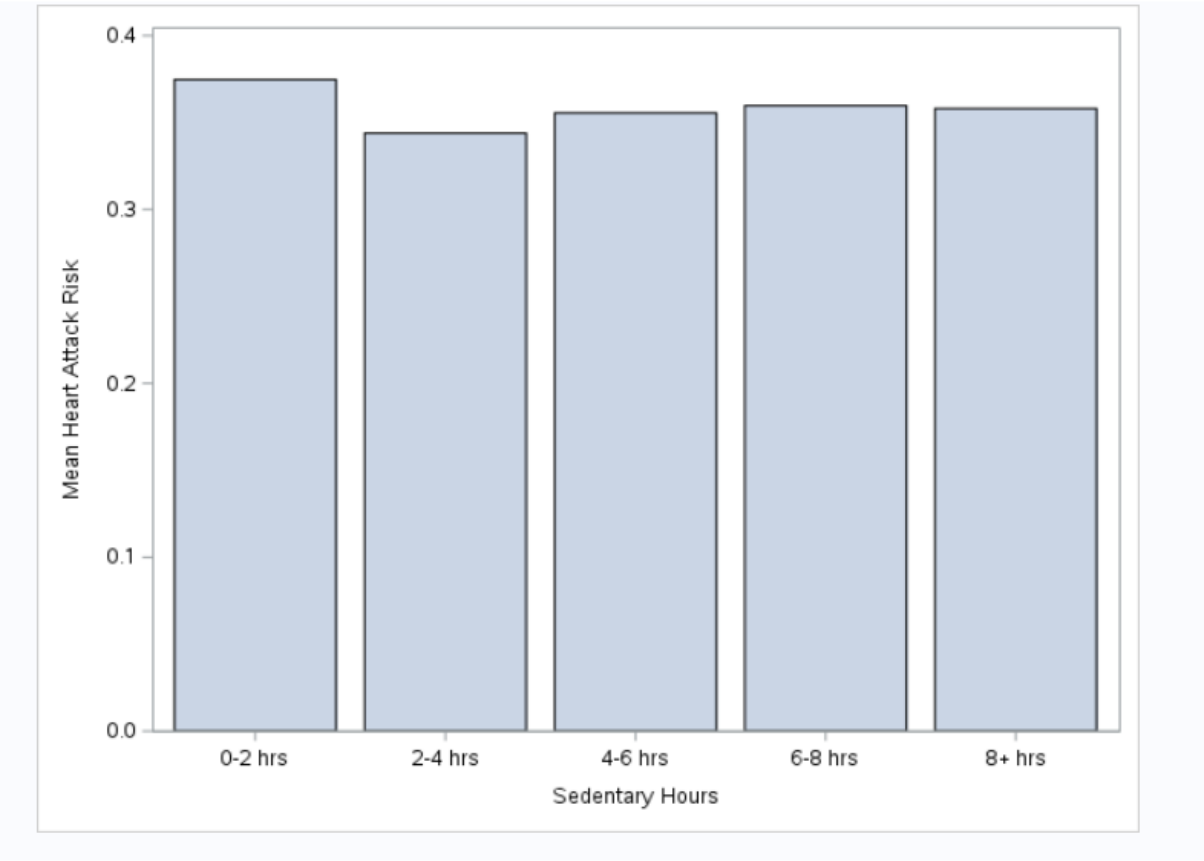
Table and chart suggest that the group with the highest risk percentage is the one that doesn't exercise at all with a 38.87% risk percent followed by sample who exercise 7 times a week with a 37.44% risk chance of a heart attack.

On the other hand, with sedentary hours per day

```
91  data heart.binned;
92      set heart.heartprediction;
93      if 'Sedentary Hours Per Day'n < 2 then SedentaryGroup = '0-2 hrs';
94      else if 'Sedentary Hours Per Day'n < 4 then SedentaryGroup = '2-4 hrs';
95      else if 'Sedentary Hours Per Day'n < 6 then SedentaryGroup = '4-6 hrs';
96      else if 'Sedentary Hours Per Day'n < 8 then SedentaryGroup = '6-8 hrs';
97      else SedentaryGroup = '8+ hrs';
98  run;
99
100 proc means data=heart.binned mean maxdec=3;
101     class SedentaryGroup;
102     var 'Heart Attack Risk'n;
103 run;
104
105 proc sgplot data=heart.binned;
106     vbar SedentaryGroup / response='Heart Attack Risk'n stat=mean;
107     yaxis label="Mean Heart Attack Risk";
108     xaxis label="Sedentary Hours (Binned)";
109 run;
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk		
SedentaryGroup	N Obs	Mean
0-2 hrs	1433	0.375
2-4 hrs	1486	0.344
4-6 hrs	1519	0.355
6-8 hrs	1429	0.360
8+ hrs	2896	0.358



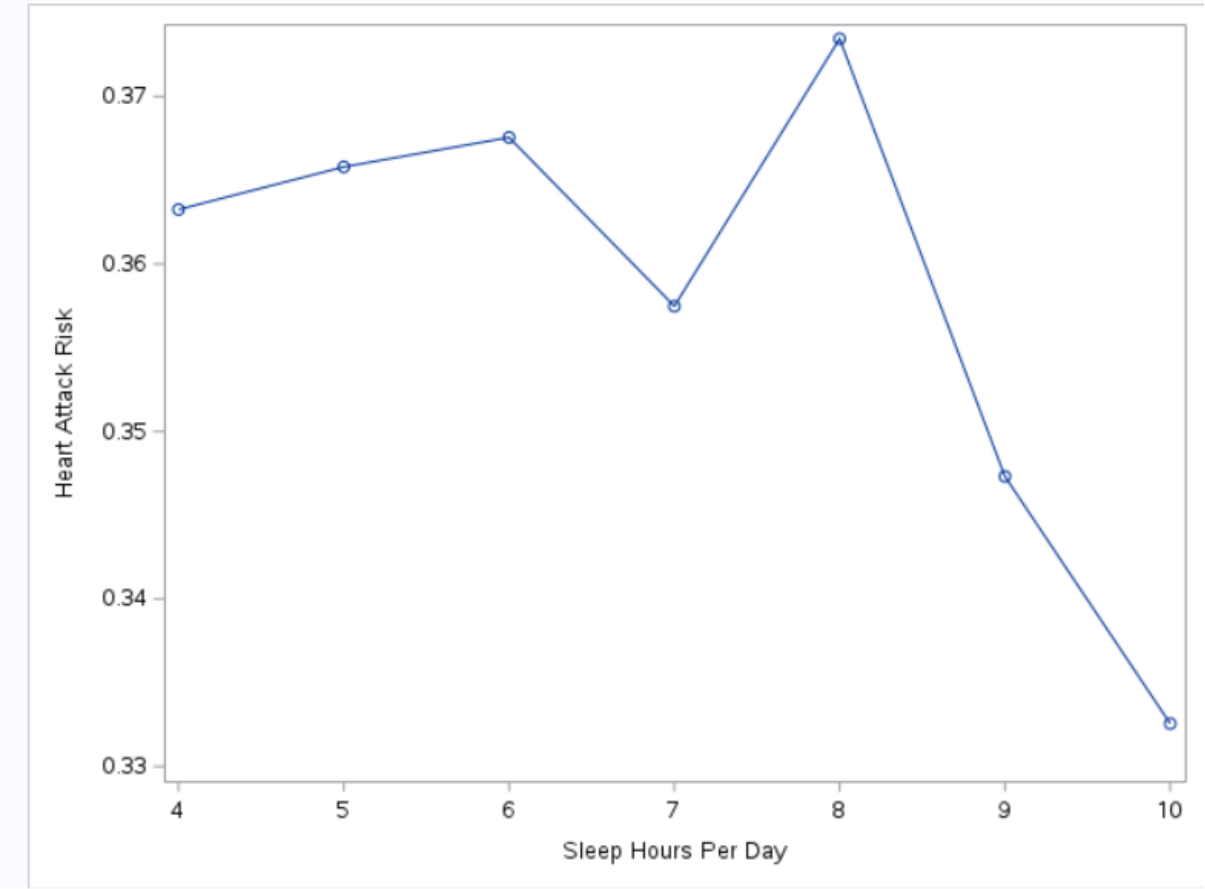
Bar chart indicates that the difference is relatively small between them all yet the out with the most risk is the group with 0-2 hours a day of sedentary hours with a 37.5% risk, followed by 36.0% risk for the 6-8 hours sedentary group.

Both charts indicate that both extremes can lead to increasing the risk of a heart attack whether it’s high sedentary hours or exercising 7 days a week.

```
proc sgplot data=heart.heartprediction;
vline 'Sleep Hours Per Day'n / response= 'Heart Attack Risk'n stat=mean markers;
yaxis label='Heart Attack Risk';
xaxis label='Sleep Hours Per Day';
run;
```

The MEANS Procedure

Analysis Variable : Heart Attack Risk						
Sleep Hours Per Day	N Obs	N	Mean	Std Dev	Minimum	Maximum
4	1181	1181	0.3632515	0.4811402	0	1.0000000
5	1263	1263	0.3657957	0.4818434	0	1.0000000
6	1276	1276	0.3675549	0.4823283	0	1.0000000
7	1270	1270	0.3574803	0.4794467	0	1.0000000
8	1288	1288	0.3734472	0.4839072	0	1.0000000
9	1192	1192	0.3473154	0.4763169	0	1.0000000
10	1293	1293	0.3325599	0.4713127	0	1.0000000



As for sleep's effect on the heart attack risk, risk percentage is at its highest within the sample that sleeps for 8 hours a day with a 37.34%, followed by the group that sleeps for 6 hours with a 36.75%, it's notable to mention that as sleep hours increase beyond 8, the risk decrease

```
proc sql;
    select
        case when Cholesterol <101 then 'Healthy cholesterol level'
        else 'unhealthy' end as Cholesterol_Level,
    mean("Heart Attack Risk"n) as Avg_Heart_Attack_Risk
from heart.heartprediction
group by Cholesterol_Level;
run;
```

Cholesterol_Level	Avg_Heart_Attack_Risk
unhealthy	0.358211

All the population has high cholesterol, and which makes their risk 35.82%

Now we assess all the previously attained factors, to gauge the most affecting ones contributing to a heart attack

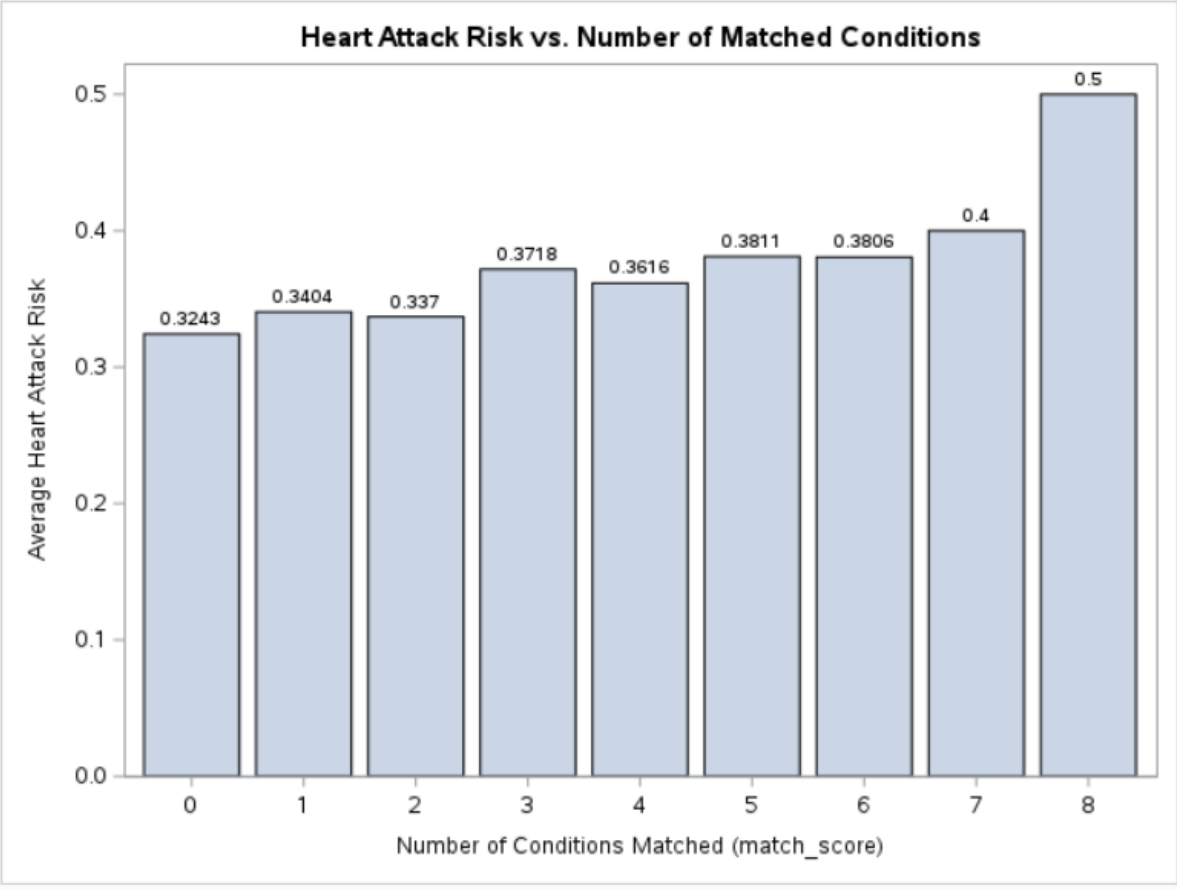
```
proc sql;
    create table scored as
    select *,
        ( (Smoking = 0) +
          ('Alcohol Consumption'n = 0) +
          (Diet = "Healthy") +
          ('Family History'n = 0) +
          (Sex = "Male") +
          (Country = "South Korea") +
          ('Sleep Hours Per Day'n = 8) +
          ('Physical Activity Days Per Week'n = 0) +
          ('Stress Level'n = 6) +
          ('Previous Heart Problems'n = 1)
        ) as match_score
    from heart.heartprediction;
quit;

proc sql;
    create table match_summary as
    select match_score,
           count(*) as N,
           mean('Heart Attack Risk'n) as Avg_Risk
    from scored
    group by match_score
    order by match_score;
quit;

proc sgplot data=match_summary;
    vbar match_score / response=Avg_Risk stat=mean datalabel;
    xaxis label="Number of Conditions Matched (match_score)";
    yaxis label="Average Heart Attack Risk";
    title "Heart Attack Risk vs. Number of Matched Conditions";
run;
```

Total rows: 9 Total columns: 3 Rows 1-9

	match_score	N	Avg_Risk
1	0	111	0.3243243243
2	1	896	0.3404017857
3	2	2190	0.3369863014
4	3	2792	0.3717765043
5	4	1831	0.361551065
6	5	761	0.3810775296
7	6	155	0.3806451613
8	7	25	0.4
9	8	2	0.5



2 observations satisfy 8 of the conditions, resulting in a 50% risk, higher than the generated average of 35% for each of the previous factors individually

Finally, we identify the 8 factors

```
proc sql;
    create table matched_8 as
    select *,
        /* Flag each condition with 1 (met) or 0 (not met) */
        (Smoking = 0)                              as match_smoking,
        ('Alcohol Consumption'n = 0)               as match_alcohol,
        (Diet = "Healthy")                         as match_diet,
        ('Family History'n = 0)                    as match_family_history,
        (Sex = "Male")                             as match_sex,
        (Country = "South Korea")                  as match_country,
        ('Sleep Hours Per Day'n = 8)               as match_sleep,
        ('Physical Activity Days Per Week'n = 0)   as match_activity,
        ('Stress Level'n = 6)                      as match_stress,
        ('Previous Heart Problems'n = 1)           as match_heart_history,

        /* Total match score */
        calculated match_smoking +
        calculated match_alcohol +
        calculated match_diet +
        calculated match_family_history +
        calculated match_sex +
        calculated match_country +
        calculated match_sleep +
        calculated match_activity +
        calculated match_stress +
        calculated match_heart_history as match_score
    from heart.heartprediction
    having match_score = 8;
quit;

proc print data=matched_8 noobs;
    var match_smoking match_alcohol match_diet match_family_history
        match_sex match_country match_sleep match_activity
        match_stress match_heart_history match_score 'Heart Attack Risk'n;
    title "Observations Matching 8 of 10 Conditions and Their Heart Attack Risk";
run;
```

Observations Matching 8 of 10 Conditions and Their Heart Attack Risk

match_smoking	match_alcohol	match_diet	match_family_history	match_sex	match_country	match_sleep	match_activity	match_stress	match_heart_history	match_score	Heart Attack Risk
0	1	1	1	1	0	1	1	1	1	8	0
0	1	1	1	1	1	0	1	1	1	8	1

Hence the sample that matches our query with the highest heart attack risk is males, located in South Korea, with a healthy diet, consumes alcohol, doesn't smoke or have a family history with heart attacks yet had previous heart problems, sleeps 8 hours a day but doesn't exercise, and has a relatively high stress level.