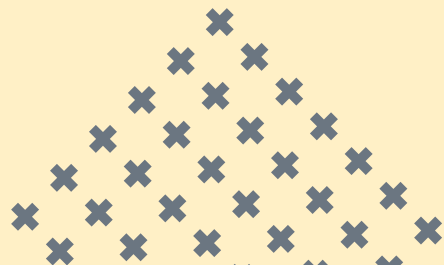
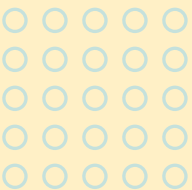
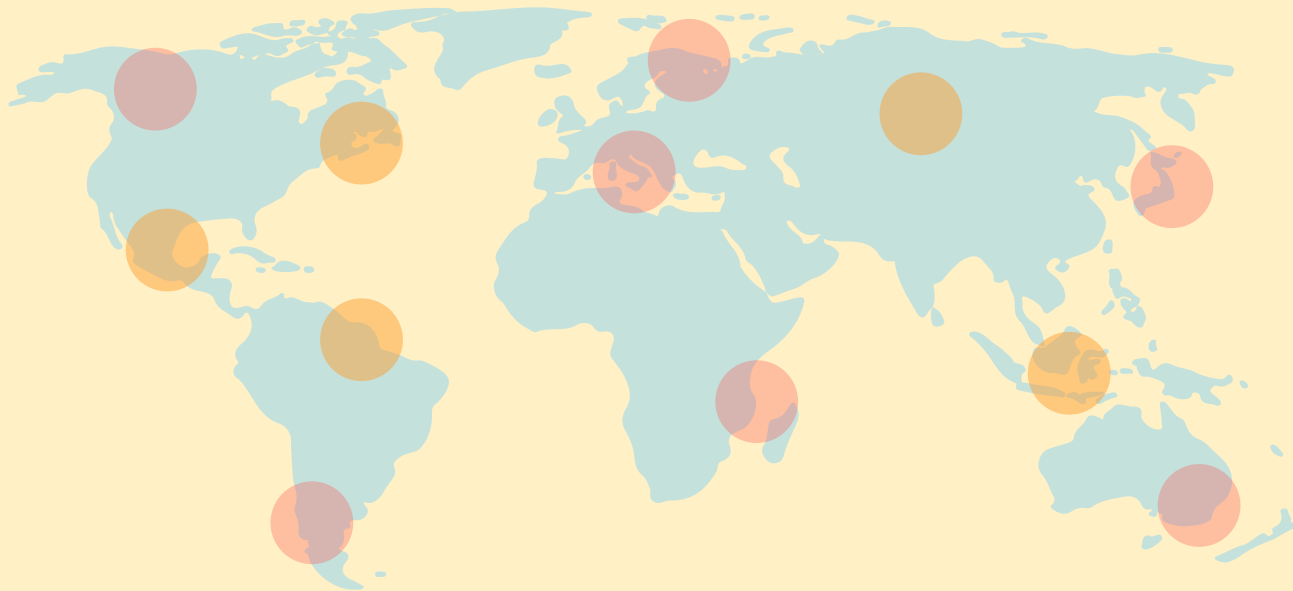




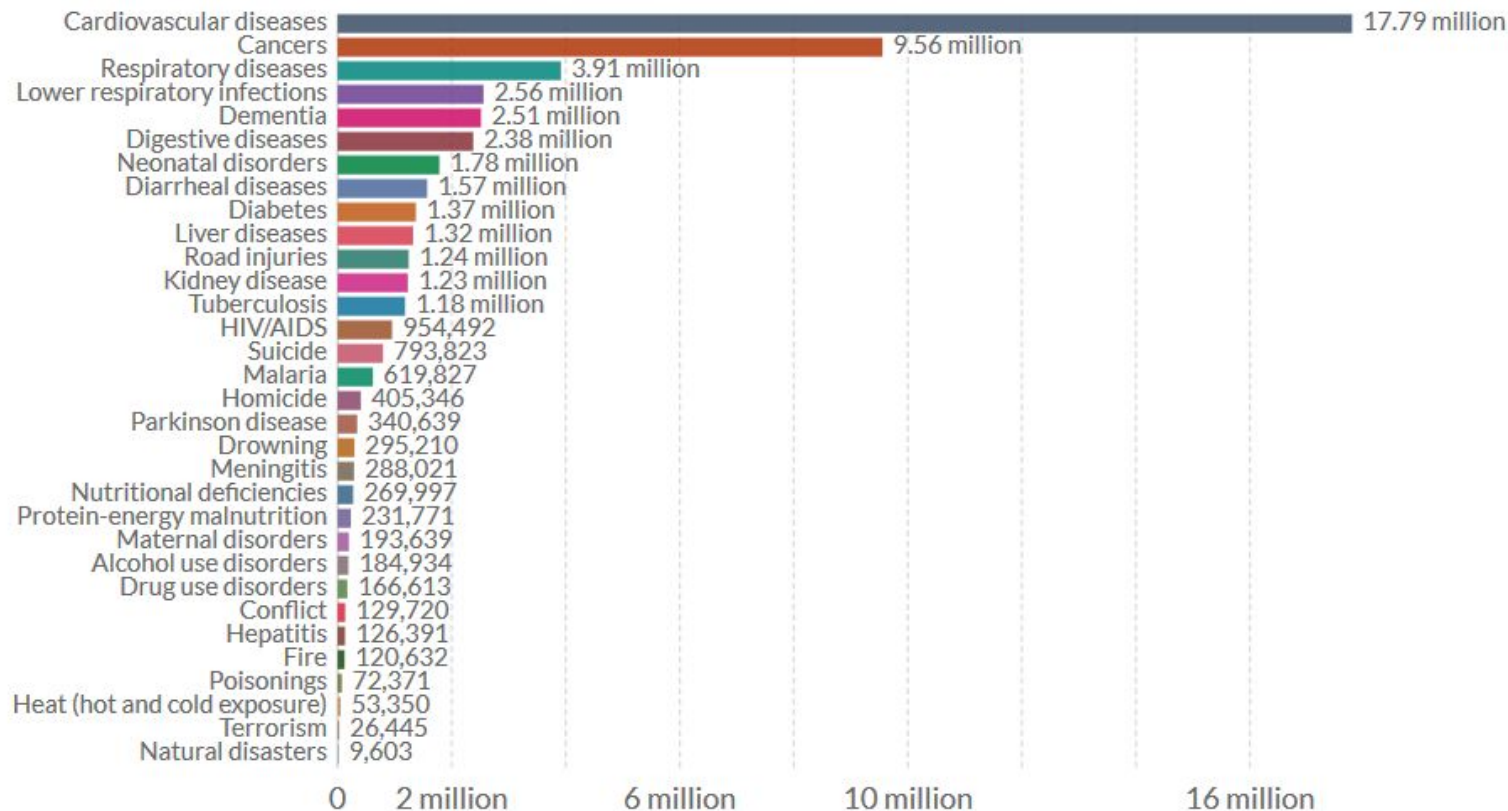
# PREDICTING HEART DISEASES

Miguel Estepa Polonio  
Data Analytics, March 2020





# Number of deaths by cause, World, 2017





**CAN WE  
PREDICT  
THEM?**



The diagram features a central text 'FRAMINGHAM SCORE' with a large pink arrow pointing down to 'HEART DISEASE RISK'. Surrounding the central text are six colored circles, each containing a risk factor: 'SMOKING' (light blue), 'BLOOD PRESSURE' (orange), 'GENDER' (light orange), 'AGE' (orange), 'CHOLESTEROL' (light orange), and 'DIABETES' (light blue). The background is a solid light yellow. Decorative elements include a cluster of red 'x' marks at the top left, a vertical line of light blue diamonds on the right, and a vertical line of light blue diamonds on the left. At the bottom right, there are three light blue chevrons pointing upwards.

**SMOKING**

**BLOOD  
PRESSURE**

**GENDER**

**AGE**

# **FRAMINGHAM SCORE**

**CHOLESTEROL**


**DIABETES**



# **HEART DISEASE RISK**

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	50.0	2	168	62.0	110	80	1	1	0	0	1	0
1	1	55.0	1	156	85.0	140	90	3	1	0	0	1	1
2	2	52.0	1	165	64.0	130	70	3	1	0	0	0	1
3	3	48.0	2	169	82.0	150	100	1	1	0	0	1	1
4	4	48.0	1	156	56.0	100	60	1	1	0	0	0	0

**def framingham\_score\_risk(kaggle\_dataset)**



	age	gender	BMI	gluc	smoke	alco	active	risk_group
0	50.0	2	22.0	1	0	0	1	1
1	55.0	1	35.0	1	0	0	1	3
2	52.0	1	24.0	1	0	0	0	2
3	48.0	2	29.0	1	0	0	1	2
4	48.0	1	23.0	1	0	0	0	1

# ML

# KNN

# 61%

```
X = df.drop(["risk_group"], axis=1)
y = df["risk_group"]
```

```
scaler = MinMaxScaler()
X = scaler.fit_transform(X)
```

```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2,
    random_state=158)
```

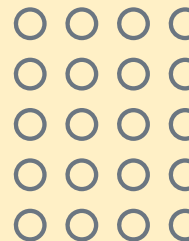
```
model = KNeighborsClassifier(algorithm='brute', leaf_size= 1, n_jobs= -1,
                             n_neighbors= 25, weights= 'uniform')
clf = model.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

```
print("Accuracy:", metrics.accuracy_score(y_pred, y_test))
print("Confusion Metrix:\n", metrics.confusion_matrix(y_pred, y_test))
```

Accuracy: 0.6114285714285714

Confusion Metrix:

```
[[2070  387   82   18    0]
 [ 363 1830  887  234   28]
 [   49   530 2777 1246  208]
 [    1    69  609 1208  325]
 [    2     4   89  309  675]]
```

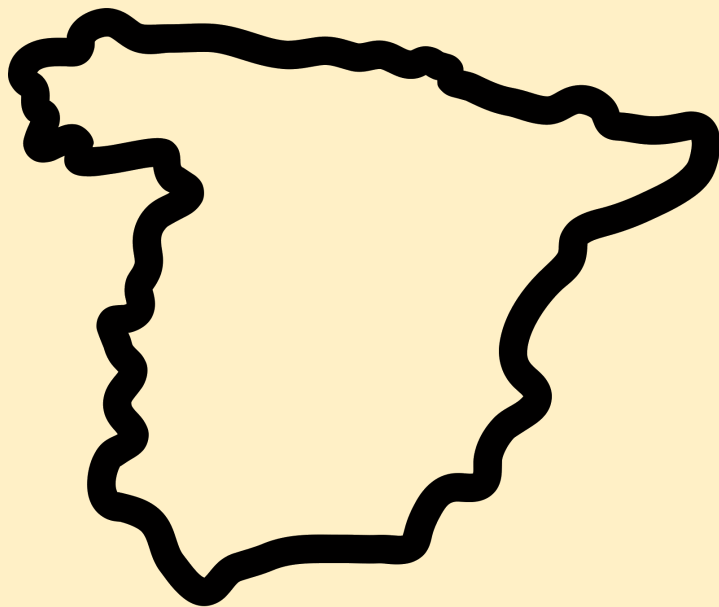




**SPAIN**

**30M PEOPLE  
OVER 30 YEARS**





**AGE**

**GENDER**

**BMI**

**SMOKING**

**ALCOHOL**

**DIABETES**

**SPORT**



# 1.000.000

**Big numbers are great.**

**I modeled a dataset with**

**1.000.000 imaginary Spanish**

**people over 30 using data from INE**



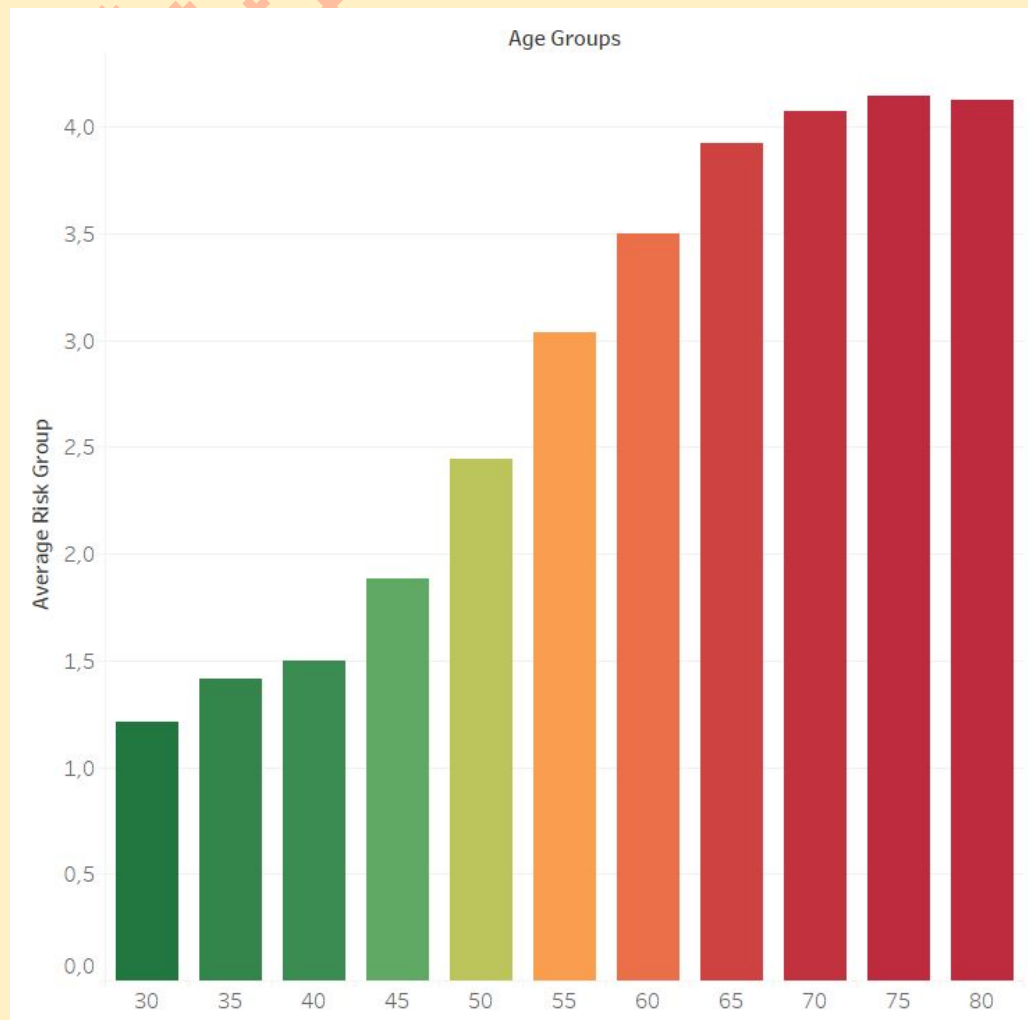
# KNN

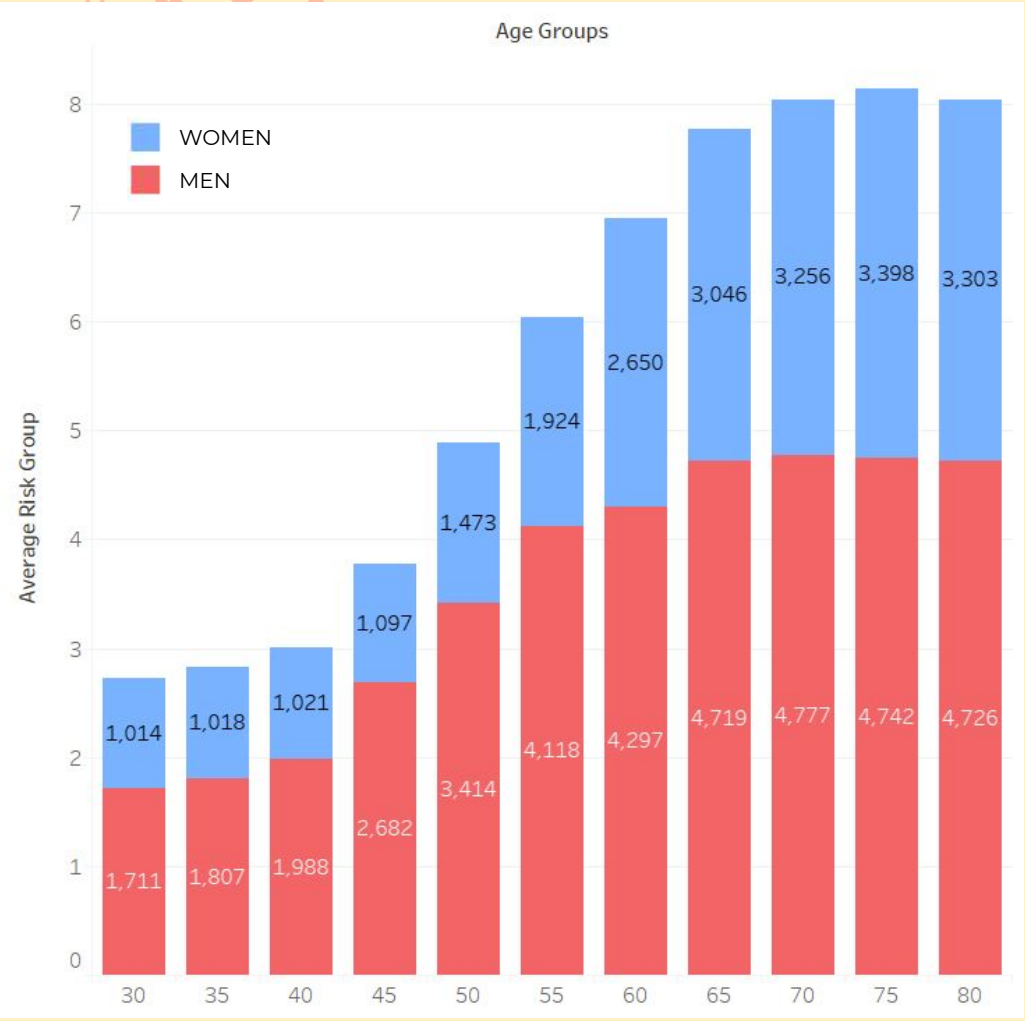
	age	gender	BMI	gluc	smoke	alco	active	risk_group
828492	59	2	25.404160	1	1	1	1	0
530876	34	2	17.356998	1	1	0	0	0
690731	47	2	25.404160	1	1	1	1	0
15879	31	1	17.524374	1	0	1	0	0
434752	70	1	17.524374	1	0	0	1	0

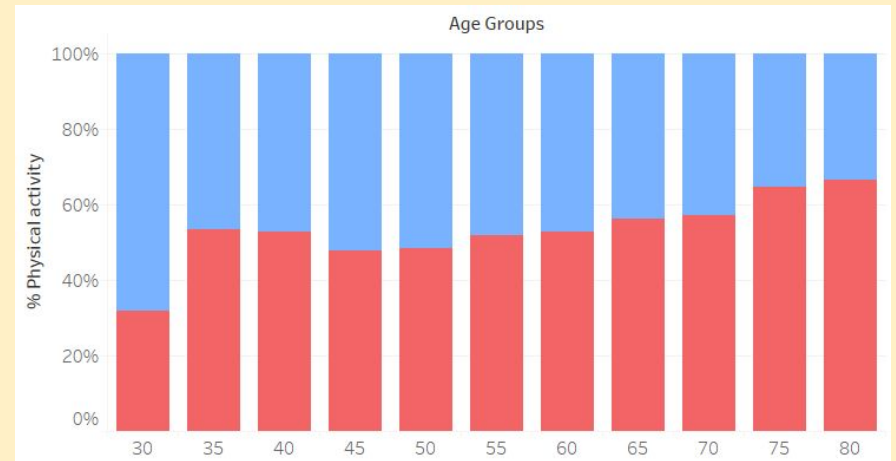
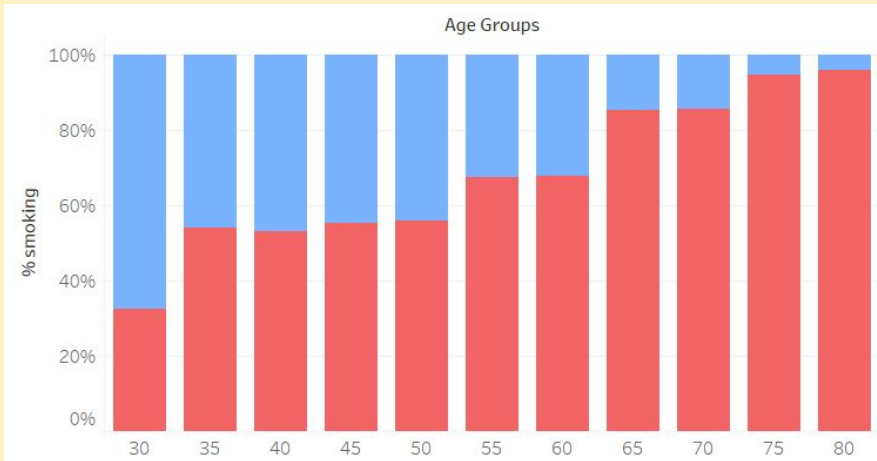
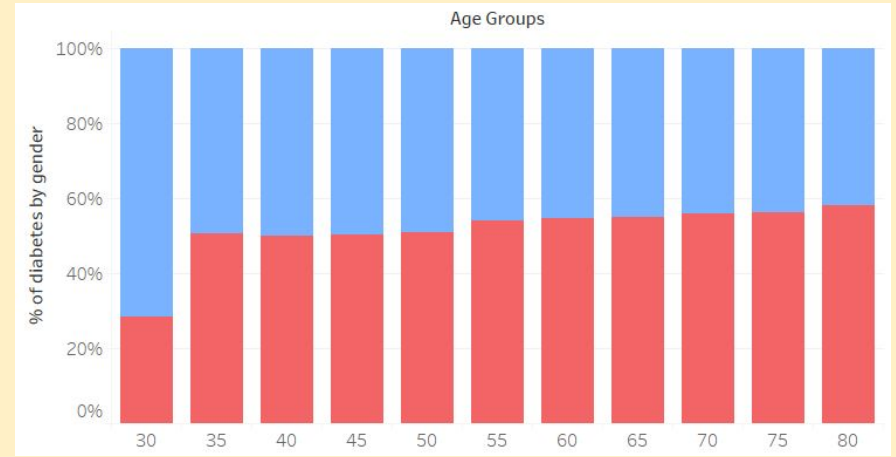
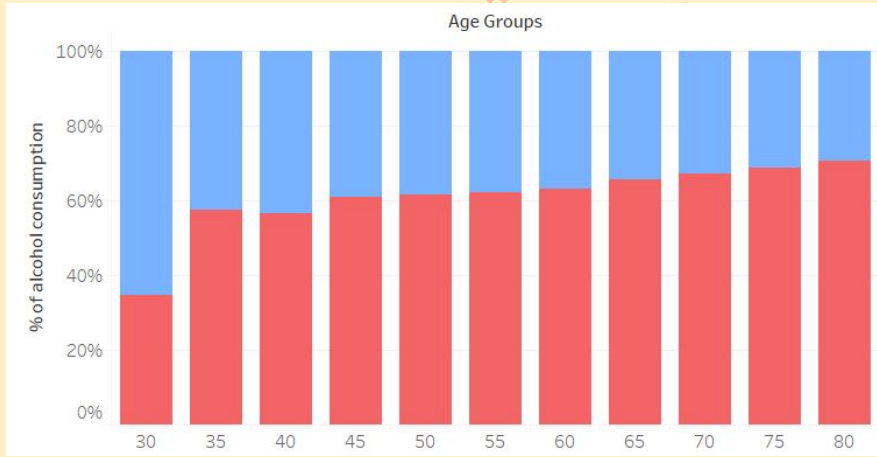
	age	gender	BMI	gluc	smoke	alco	active	risk_group
459037	74	1	17.524374	1	0	0	1	3
755086	52	2	25.404160	1	1	1	0	4
107097	40	1	17.524374	1	1	0	1	1
863370	62	2	17.356998	1	0	0	0	2
255058	51	1	17.524374	1	1	0	0	1



# EXPLORATION



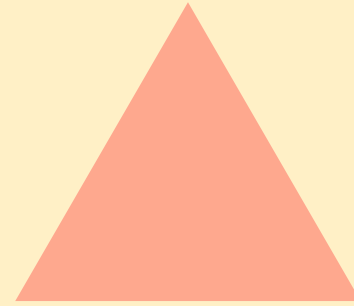
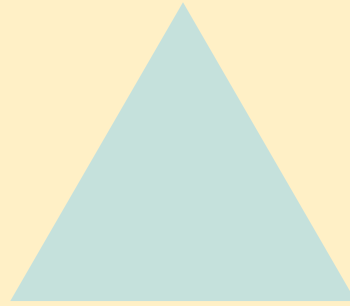






# CONCLUSIONS





## **40 YEARS OLD IS THE CRITICAL AGE**

At that age, the probability of suffering a heart disease increases dramatically.

## **BAD HABITS KILL**

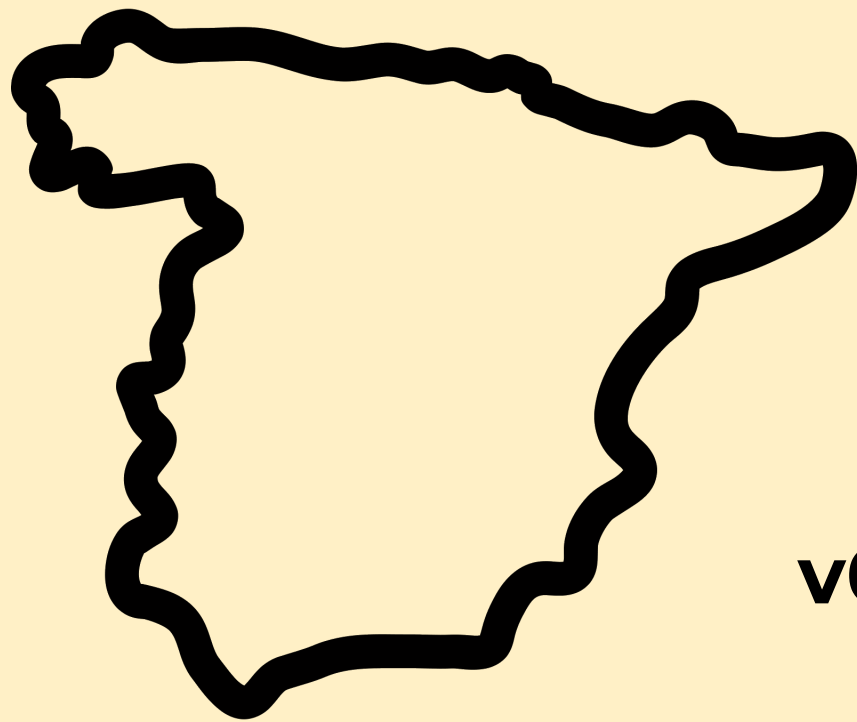
Even when women have more natural probabilities, due to bad habits men die more

## **CONSUMPTION HABITS ARE CHANGING**

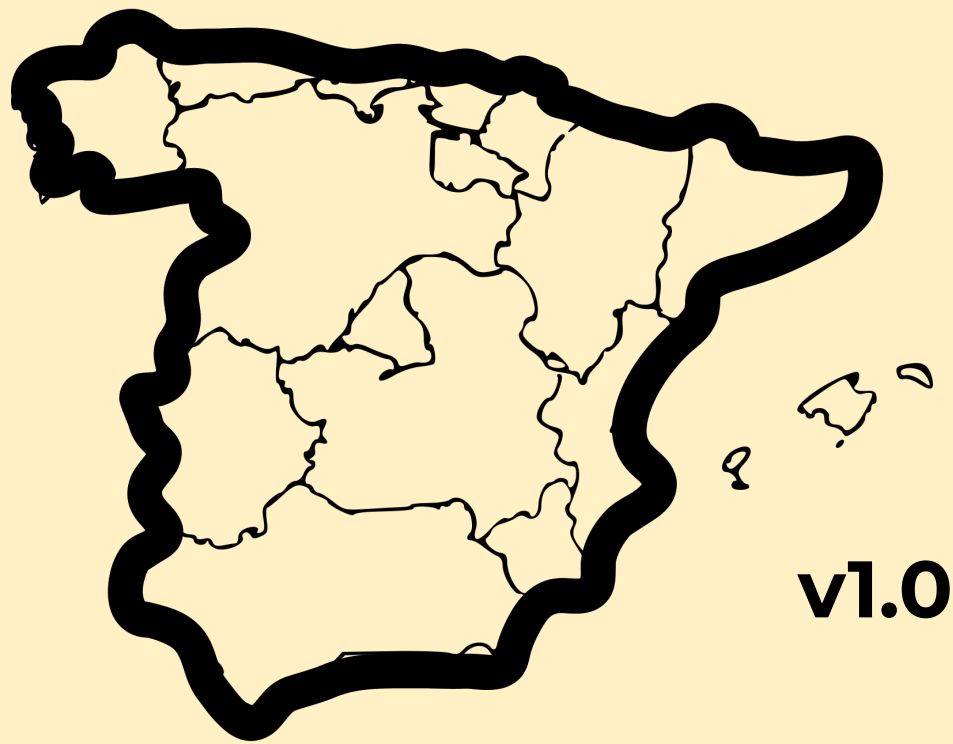
The percentage of smoking and drinking alcohol in woman is increasing in younger generations



# **FUTURE**



**v0.1**



**v1.0**



# THANKS!

QUESTIONS?

