

**Nombre y Apellidos: Javier Díaz de Rábago Pemán**

**Github con notebook:** <https://github.com/javierdrp/audio-effects-simulator/tree/main/FINAL>

*Nota: Por favor, seguir esta estructura para el documento*

## 1. Resumen Ejecutivo

Máximo 2 páginas

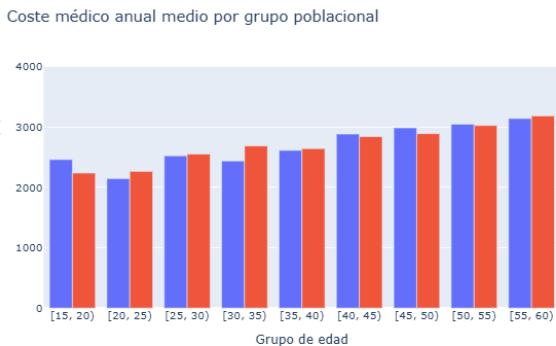
No se observa gran diferencia entre los datos de diferentes regiones. Independientemente de la región, hay algunos factores muy claros que aumentan el riesgo de la persona asegurada, y como consecuencia el coste médico anual y las reclamaciones pagadas.

- El grupo de edad. A partir de los 35 años, el gasto médico empieza a crecer.
- Si el cliente es o ha sido fumador. Hay una diferencia de hasta un 50% en el gasto medio de clientes fumadores respecto a no fumadores.
- La composición corporal. Las personas con sobrepeso empiezan a considerarse de riesgo entre cinco y diez años antes que las personas con un peso saludable.
- Las enfermedades crónicas. Este es el factor que más dispara el gasto médico en una persona. Las personas con enfermedades crónicas cuestan casi el doble que las personas sanas.

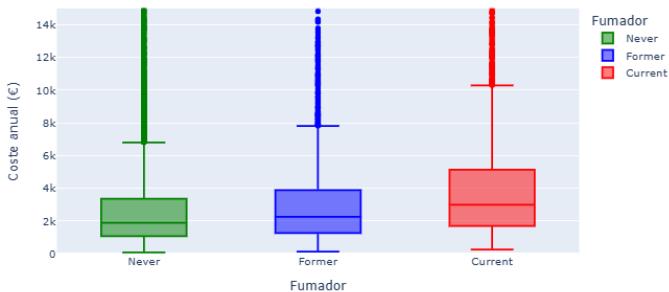
De cara a reducir costes, sería clave realizar un estudio exhaustivo de estas variables en los posibles clientes.

### Análisis de Costes Médicos

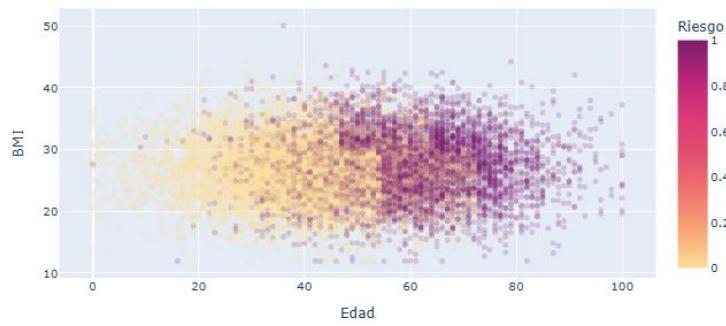
North



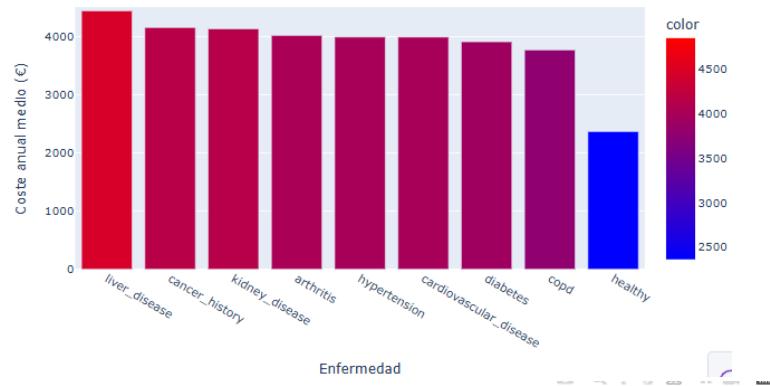
Coste médico anual según hábitos de vida (fumador/no fumador)



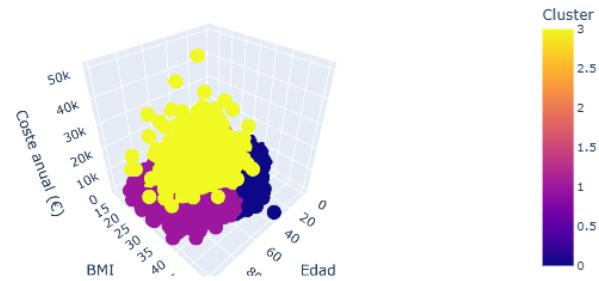
Pacientes de riesgo según edad y composición corporal



Coste medio anual por enfermedad

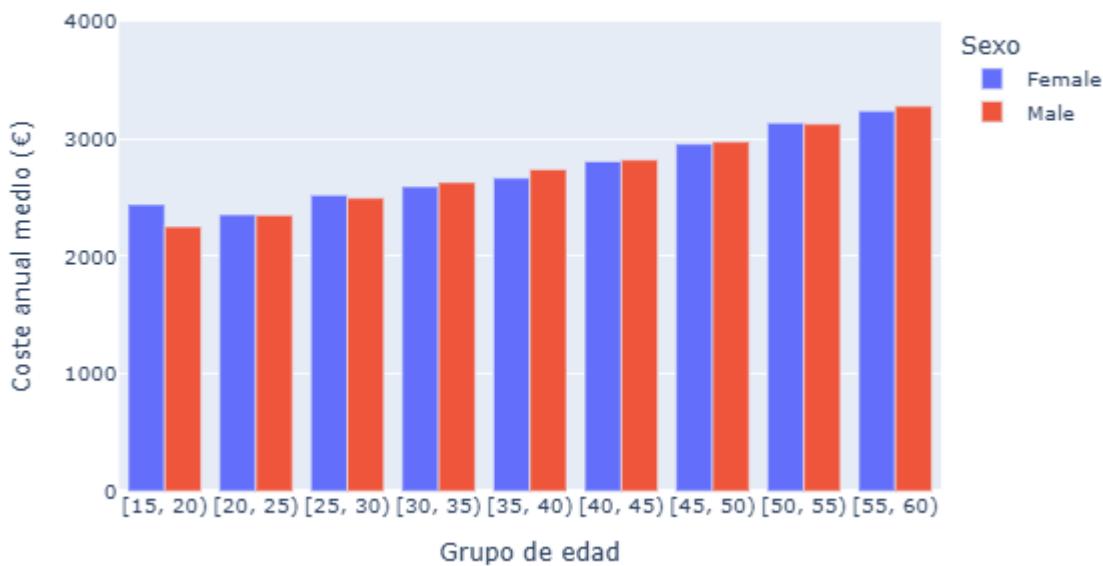


Clustering de pacientes según edad, BMI y coste médico anual



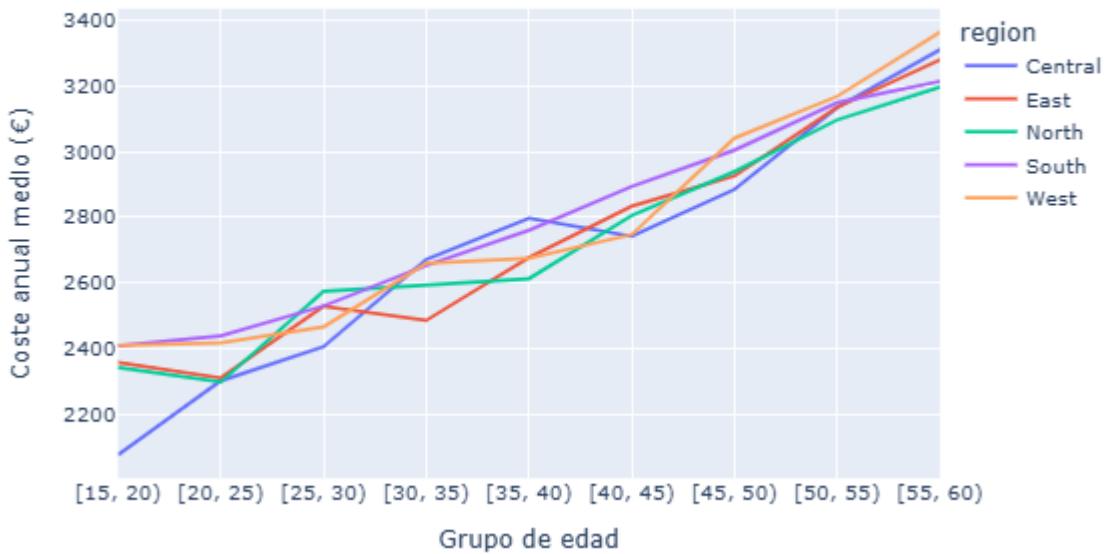
## 2. Gráficas del análisis exploratorio y breve explicación de cada una

Coste médico anual medio por grupo poblacional



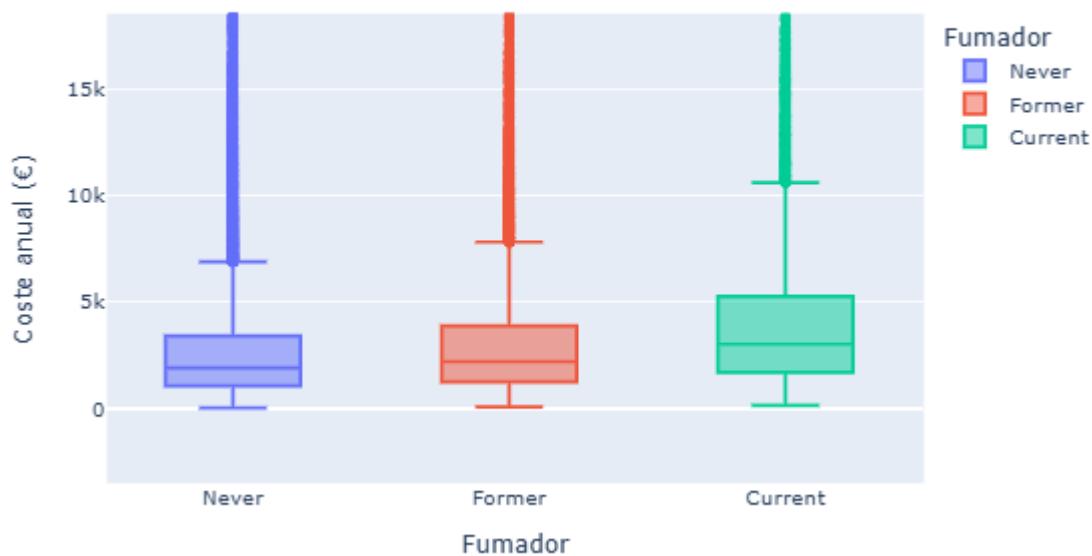
Se puede ver claramente cómo el gasto anual aumenta con la edad. No parece depender casi del sexo del paciente.

Coste médico anual medio por grupo poblacional



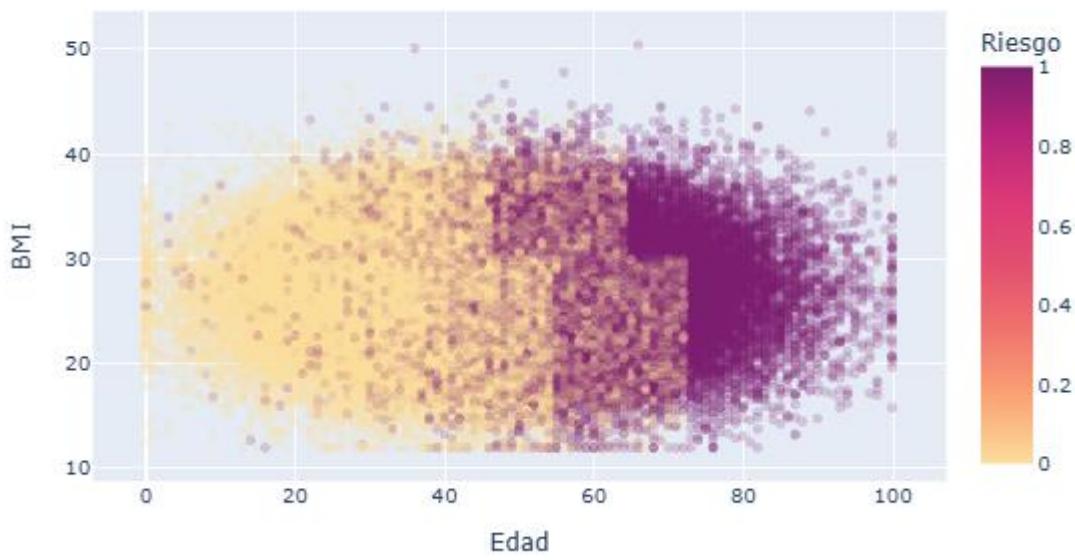
Aquí se puede ver cómo las regiones central y sur tienen más coste médico en edades intermedias (30 a 50) años, pero a partir de los 50 es el oeste donde más crece el gasto.

### Coste médico anual según hábitos de vida (fumador/no fumador)



Fumar es claramente un factor que influye en el coste anual en salud de cada persona. Todas las métricas (cuartiles, media, mediana, outliers) son hasta un 50% más altas en el caso de las personas fumadoras.

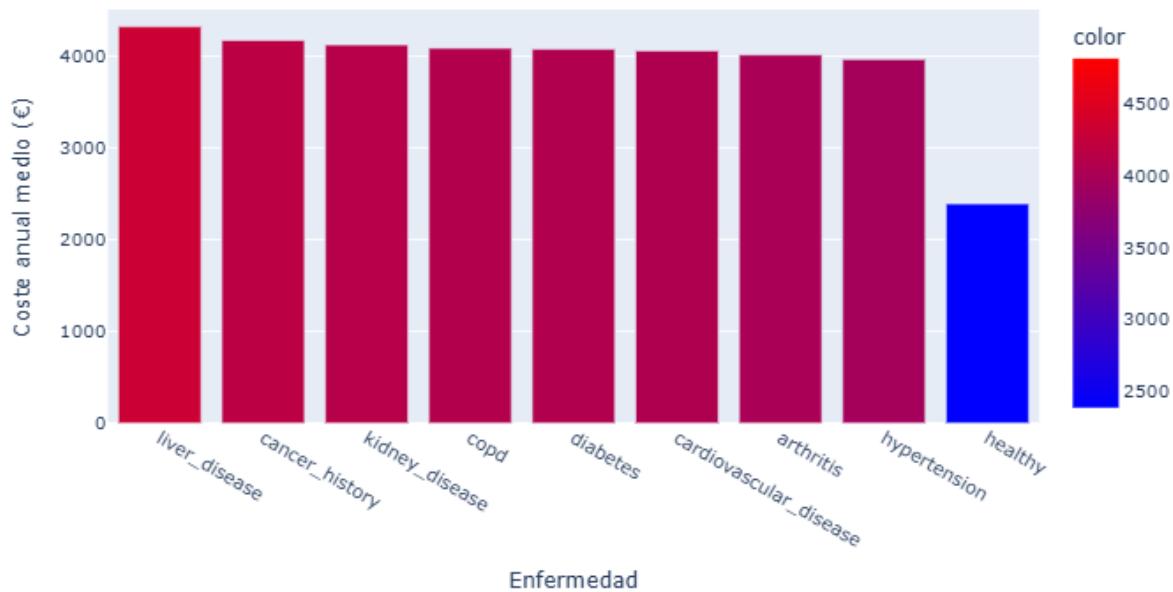
### Pacientes de riesgo según edad y composición corporal



Aprovechando la transparencia, se pueden distinguir tres áreas claras: sin riesgo, riesgo medio y alto riesgo. A partir de un BMI de 30, se ve cómo claramente se reduce la edad en la que empiezan a considerarse pacientes de riesgo. Es consistente con el consenso de que el BMI se considera saludable entre 18.5 y 25. Curiosamente, este aumento de riesgo no se

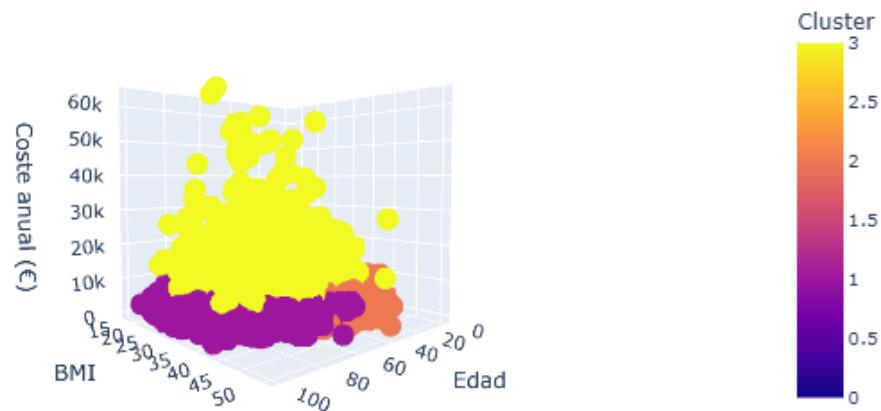
da con BMIs por debajo del límite inferior. Podría ser por falta de datos (hay pocos por debajo de BMI 18).

Coste medio anual por enfermedad

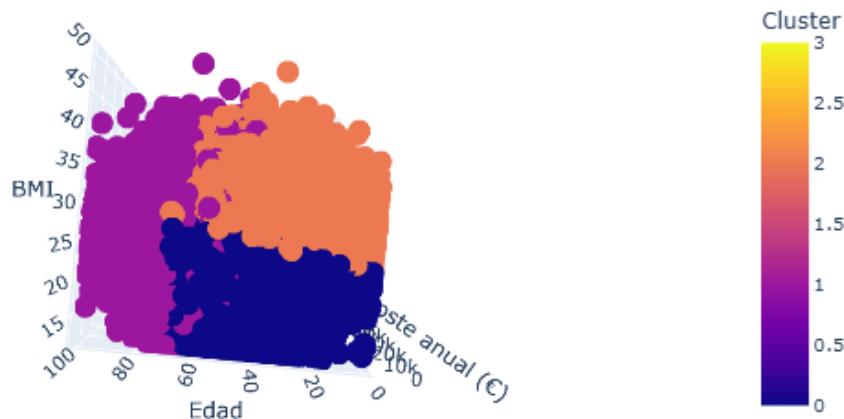


Hay una clara diferencia entre personas con enfermedades crónicas y personas sanas. El gasto anual promedio se dispara desde unos 2400€ hasta más de 4000€, siendo las enfermedades del hígado las más problemáticas.

Clustering de pacientes según edad, BMI y coste médico anual



Clustering de pacientes según edad, BMI y coste médico anual



Al realizar un clustering tras normalizar, para obtener diferentes grupos de personas, el algoritmo separa claramente varios grupos:

- Personas con más gasto en salud (amarillo)
- Personas de edad avanzada (morado)
- Personas con sobrepeso (naranja)
- Personas en edad y BMI saludable (azul)

### 3. Modelo predictivo explicado y con tablas

Calculando una clasificación de 'is\_high\_risk' con las siguientes variables: 'age', 'income', 'household\_size', 'dependents', 'bmi', 'chronic\_count', 'risk\_score', 'visits\_last\_year', 'provider\_quality'

Obtenemos:

Classification report					
	precision	recall	f1-score	support	
Sí	0.88	0.89	0.88	13228	
No	0.80	0.79	0.80	7748	
accuracy			0.85	20976	
macro avg	0.84	0.84	0.84	20976	
weighted avg	0.85	0.85	0.85	20976	
Confusion matrix					
[[11718 1510]					
[ 1634 6114]]					

Vemos que se consigue clasificar correctamente más del 80% de los casos. Se prueba a reducir el número de variables explicativas, dejando: 'age', 'income', 'household\_size', 'bmi', 'chronic\_count', 'provider\_quality'

Classification report				
	precision	recall	f1-score	support
Sí	0.87	0.88	0.87	13228
No	0.79	0.77	0.78	7748
accuracy			0.84	20976
macro avg		0.83	0.82	20976
weighted avg		0.84	0.84	20976
Confusion matrix				
[[11641 1587] [ 1802 5946]]				

Este modelo sigue teniendo una buena precisión en la clasificación, requiriendo menos variables de entrada.