

Nombre y Apellidos: José Guardo

Github con notebook:

1. Resumen Ejecutivo

Máximo 2 páginas

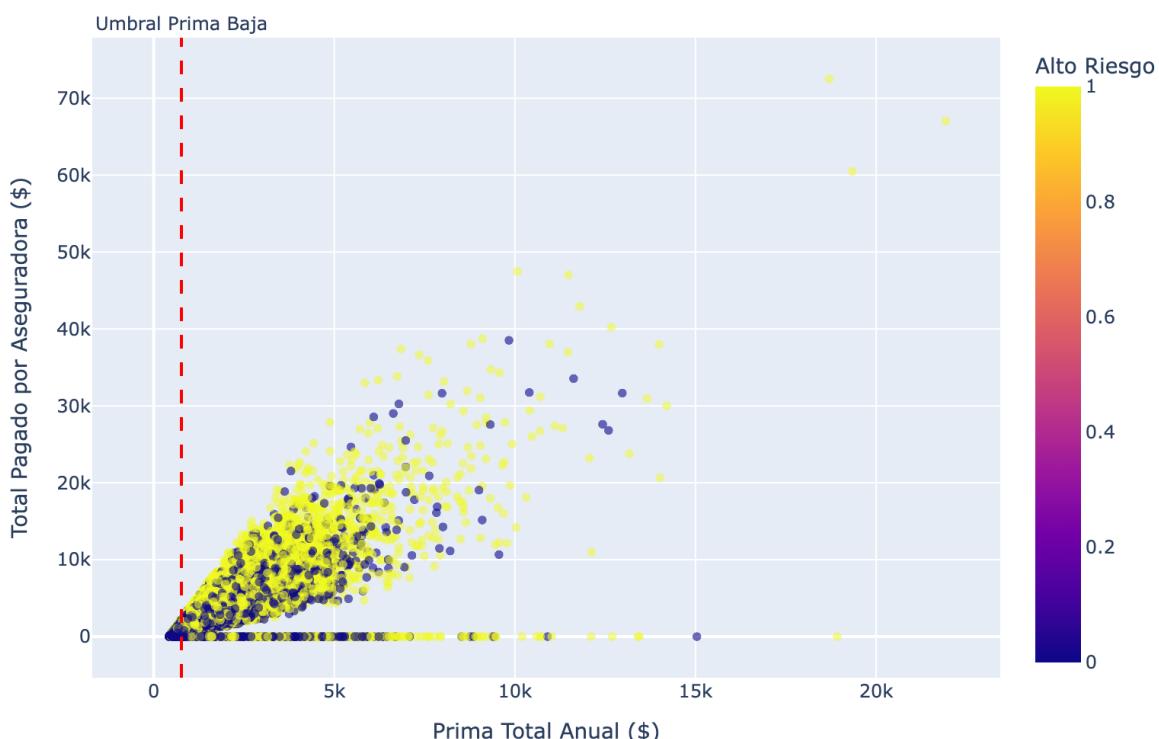
El objetivo de este trabajo consiste en hacer una evaluación de los clientes de la aseguradora X. Esta compañía tiene una línea de negocio en la vertical sanitaria y está buscando optimizar la forma en la que determinan el perfil de riesgo de un cliente y al mismo tiempo cómo reducir los costes promedios por cliente para poder aumentar el margen de beneficios.

Este trabajo tiene como objetivo darle una propuesta a la compañía realizando un análisis explorativo de sus clientes. Se realizará una evaluación en detalle que permita diseñar una estrategia para la aseguradora.

Se parte de la base de que el cliente ideal de la aseguradora es aquel que tiene una prima total (prima mensual + prima anual) elevada y que el coste que supone para la compañía sea lo menor posible.

Por el contrario, se interpreta que el peor cliente posible para la compañía es el que paga una póliza reducida, pero supone un coste elevado para la compañía.

Relación entre Prima Pagada y Reclamaciones: Identificación de Perfiles Problemáticos



Esta relación se puede observar en la gráfica. Aquellos pacientes que están más pegados al eje y (Total Pagado por Aseguradora) y más elevados son los peores clientes para la compañía mientras que aquellos que están más cercanos al eje x (Prima Total Anual) son los

que preferidos por la misma. Esta añadimos como perfil de cliente que alto potencial de pérdidas a aquellos que tienen un perfil de riesgo alto y están pagando una cuota baja.

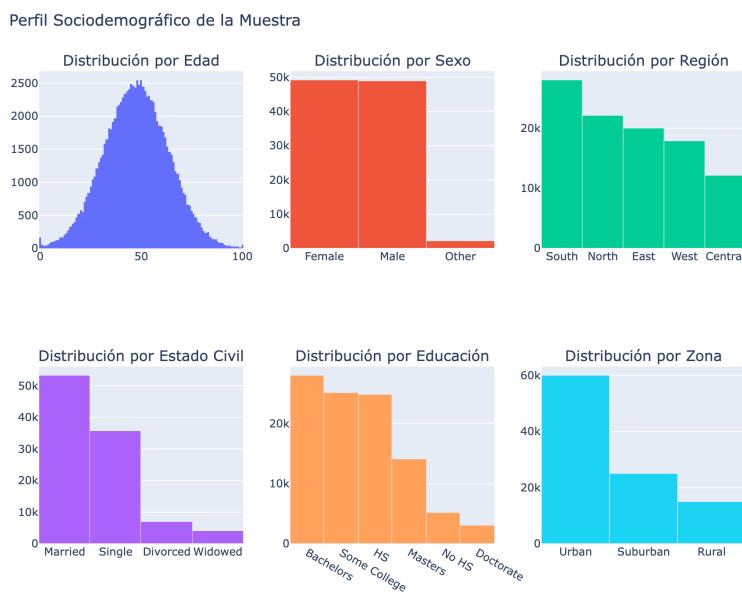
Total, Clientes de Alto Riesgo	Clientes de alto Riesgo con Prima Baja	Porcentaje
36781	7067	19.21%

Tras realizar una serie de evaluaciones y exposiciones se centra el estudio en la evaluación de perfiles de grupo 2 (riesgo alto + cuota baja). Debido a la falta de inteligencia extractable de las visualizaciones generales se procede a entrenar un modelo Random Forest para predecir cuál de los factores permite predecir el perfil de riesgo del cliente. Esto lleva a la edad como principal indicativo de riesgo alto. En particular, los clientes de tercera edad suponen la gran parte de los clientes de perfil 2 lo que era de esperar, no obstante, incluso estos clientes que tienen baja tendencia por cambiar su póliza, la compañía no es capaz de entender el riesgo y persiste las pérdidas asociadas a un estado de salud peor por la edad.

Se le aconseja a la empresa a revisar la política de renovación de pólizas para clientes de alto riesgo así realizar una proyección correcta para aquellas pólizas que terminen pronto para personas de la tercera edad.

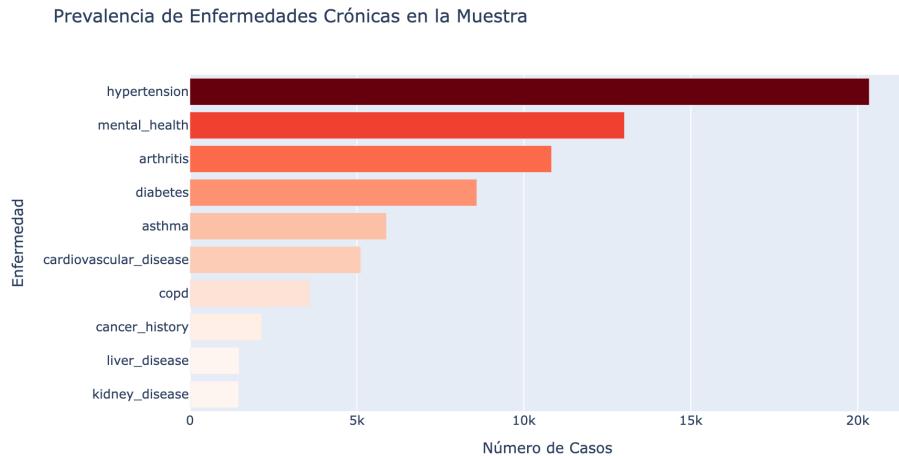
2. Gráficas del análisis exploratorio y breve explicación de cada una

2.1 Perfil Sociodemográfico del Cliente



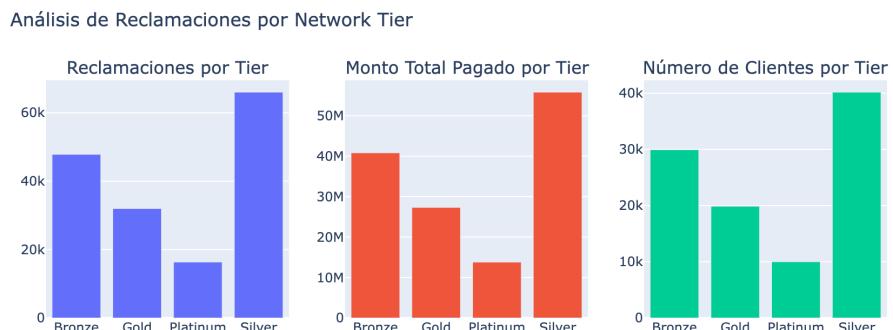
A continuación, se realiza una descripción de cómo está distribuida la muestra. Se puede observar que la edad se asemeja a una distribución normal centrada en 50 y una distribución equitativa entre mujeres y hombres. En cuanto a la distribución por Estado Civil, así como la Distribución por Educación se puede observar claramente como casados y solteros, así como personas que realizaron el instituto y la carrera (completada o no) constituyen una porción mayoritaria de la muestra.

2.2 Prevalencia de Enfermedades Crónicas en la Muestra



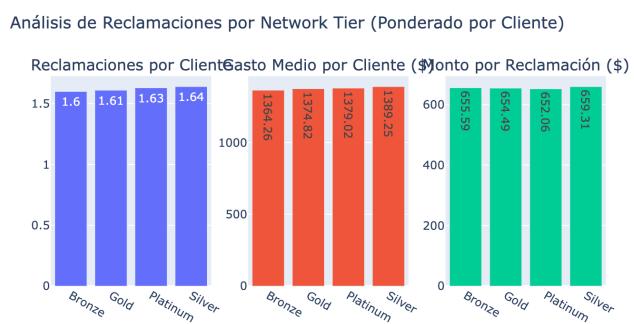
En este caso se puede observar claramente como existe una clara prevalencia de las enfermedades más comunes como la hipertensión y enfermedades relacionadas a la salud mental.

2.3 Análisis de Reclamaciones por Network Tier



En línea con la tesis inicial la distribución de reclamaciones por Network Tier muestra una señal indicativa de que los clientes que más reclamaciones realizan y que suponen un mayor coste son aquellos que forman parte del Tier plata.

No obstante, si analizamos la media ponderada de cuanto se paga por cada Tier se puede observar que no existe una clara diferencia en cuanto a donde se puede encontrar el cliente que representan un riesgo para la compañía.



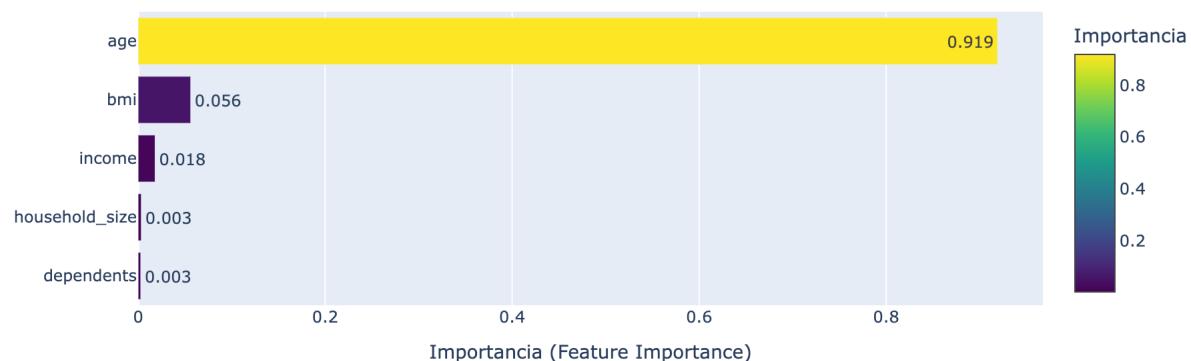
3. Modelo predictivo explicado y con tablas

Frente a la dificultad de entender cuáles son los factores que hacen ver cuáles son los perfiles que suponen un mayor potencial de pérdida para la empresa se procede a entrenar un modelo para predecir cuál de las variables numéricas relacionadas con el perfil demográfico del cliente son las más significativas para predecir su nivel de riesgo.

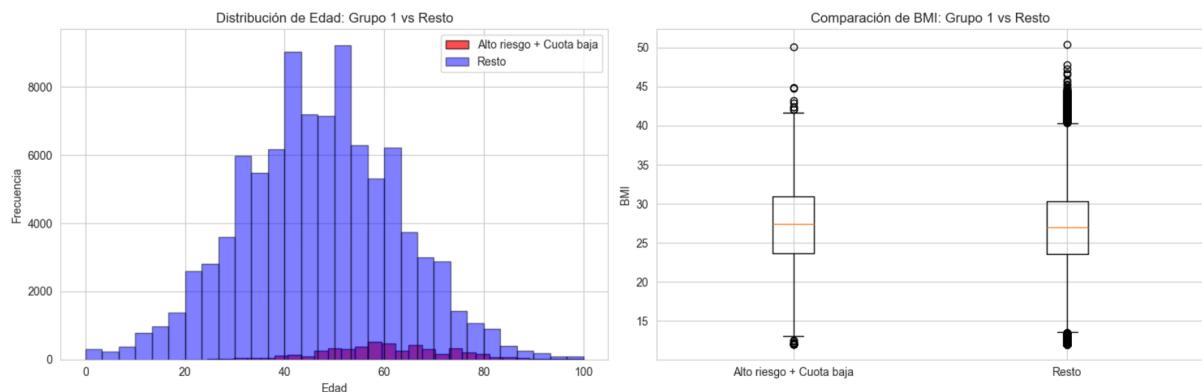
Se entrena para este caso un modelo *Random Forest Classifier* con el que se consigue el siguiente rendimiento:

	Precisión	Recall	F1-score	Support
Bajo Riesgo	0,85	0,83	0,84	12644
Alto Riesgo	0,71	0,74	0,73	7356

Importancia de Variables en la Predicción de Alto Riesgo



Cómo puede ser de esperar el factor que indica con mayor precisión el perfil de riesgo del cliente es la edad. Esta vista especialmente interesante para exponer los perfiles con alto potencial de pérdida que corresponden al perfil **Alto Riesgo + Cuota Baja**.



Para poder entender mejor que políticas se pueden llevar a cabo con las personas que cumplen este perfil se procede a estudiar cómo es el comportamiento de estos clientes frente a la posibilidad de acometer cambios en su póliza en los dos últimos años.

Las conclusiones de este caso son que únicamente 258 clientes de un total de 7067 clientes de este grupo realizan cambios en su póliza. Además, no solo conservan pólizas que no implican mejoras para la compañía dado que la pérdida anual promedia se mantiene igual y no capturan la transición del perfil de riesgo debido a su edad.

	Clientes Con Cambios (258)	Clientes Sin Cambios (6809)
Duración Poliza	5,4 años	5,56 años
Edad Promedia	60,7 años	60,9
Risk Score	0,75	0,75
Costo Médico Anual	879,88\$	881,24
Prima Mensual	25,73\$	25,83\$
Pérdida Anual Promedio	571,18\$	571,23\$