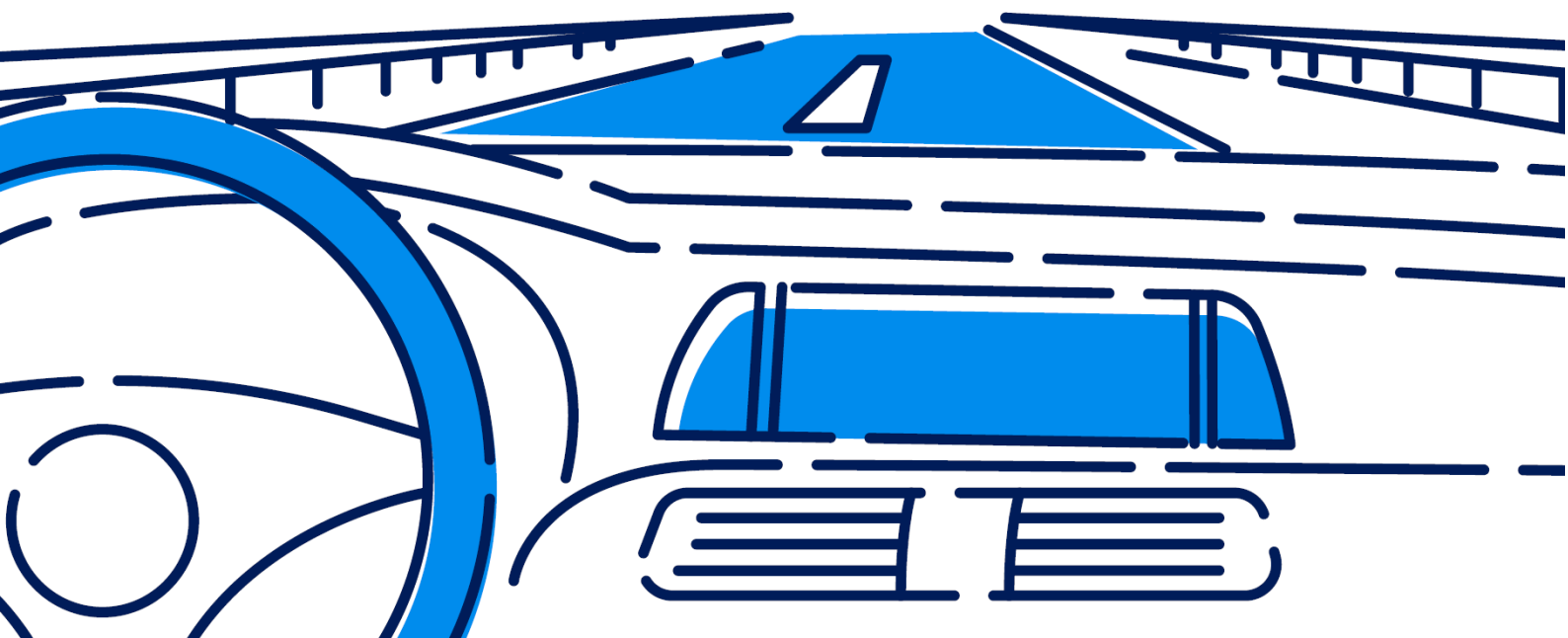


Análisis Exploratorio de Datos “Evolución y Tendencias - Entorno Insurtech”



Indice

1. Introducción

1.1 Contexto

1.2 Hipótesis

2. Exploración de los Datos

2.1 Análisis Preliminar

2.2 Calidad de los Datos

3. Análisis Global de los Datos

3.1 Análisis Univariante

3.2 Análisis Bivariante

3.3 Análisis Multivariante

4. Análisis Específico

4.1 Tendencias en la Contratación

4.2 Perfil de los Clientes por Provincia

5. Conclusiones

1. Introducción

A. Contexto

El mercado de los seguros ha evolucionado notablemente en los últimos años, impulsado por la tecnología. El sector **insurtech** ha permitido transformar la forma en que las aseguradoras ofrecen sus productos, facilitando la digitalización, la personalización de servicios y la creación de nuevos modelos de negocio. Dentro de este contexto, Hello Auto se presenta como una plataforma innovadora que ofrece pólizas inteligentes, utilizando grandes volúmenes de datos para mejorar la experiencia del cliente y optimizar la oferta de seguros.

El análisis de datos juega un papel crucial en el negocio de Hello Auto, ya que la empresa utiliza estos datos para ajustar sus estrategias comerciales, identificar patrones de comportamiento de los clientes, y mejorar la toma de decisiones en tiempo real. El objetivo de este informe es explorar los datos proporcionados sobre las pólizas contratadas y los perfiles de los clientes, con el fin de obtener información valiosa que ayude a mejorar las operaciones de Hello Auto y aumentar la satisfacción de los clientes.

B. Hipótesis

Con base en los datos disponibles y el contexto del sector **insurtech**, se plantea la siguiente hipótesis:

1. **Perfil de los clientes:** Existen diferencias significativas en los perfiles de los clientes que contratan pólizas Smart y Flex, especialmente en variables como la edad, tipo de vehículo y la provincia de residencia.
2. **Tendencias de contratación:** Las pólizas de tipo Smart y Flex presentan tendencias diferenciadas a lo largo del tiempo, lo que podría implicar estacionalidades o cambios en las preferencias de los consumidores debido a factores externos (como la economía o eventos específicos del sector).
3. **Impacto de los siniestros:** La frecuencia de siniestros está relacionada con ciertas características del cliente o del vehículo, como la edad del vehículo y el tipo de combustible.

2. Exploración de los Datos

En esta sección se realiza una exploración preliminar de los datos proporcionados, correspondientes al periodo 2020-2023. El análisis se centrará en las siguientes columnas relevantes:

- **policyId**: Identificador de la póliza.
- **policyStatus**: Estado de la póliza (activa o cancelada).
- **tipoPoliza**: Tipo de póliza (Smart o Flex).
- **customerId**: Identificación del cliente.
- **age**: Edad del cliente al momento de contratar la póliza.
- **vehicleAge**: Edad del vehículo.
- **fuel**: Tipo de combustible del vehículo.
- **province**: Provincia de residencia del cliente.
- **FAclaims**: Número de siniestros registrados para la póliza.

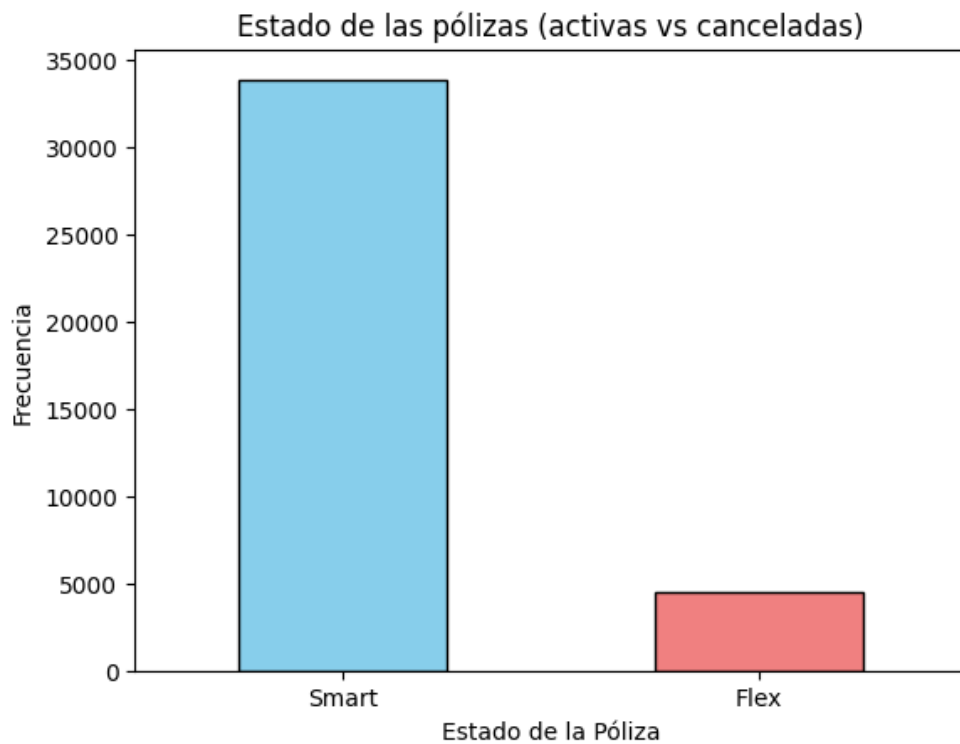
Además, se analiza la calidad de los datos, identificando la presencia de valores nulos o erróneos que podrían impactar el análisis posterior.

3. Análisis Global de los Datos

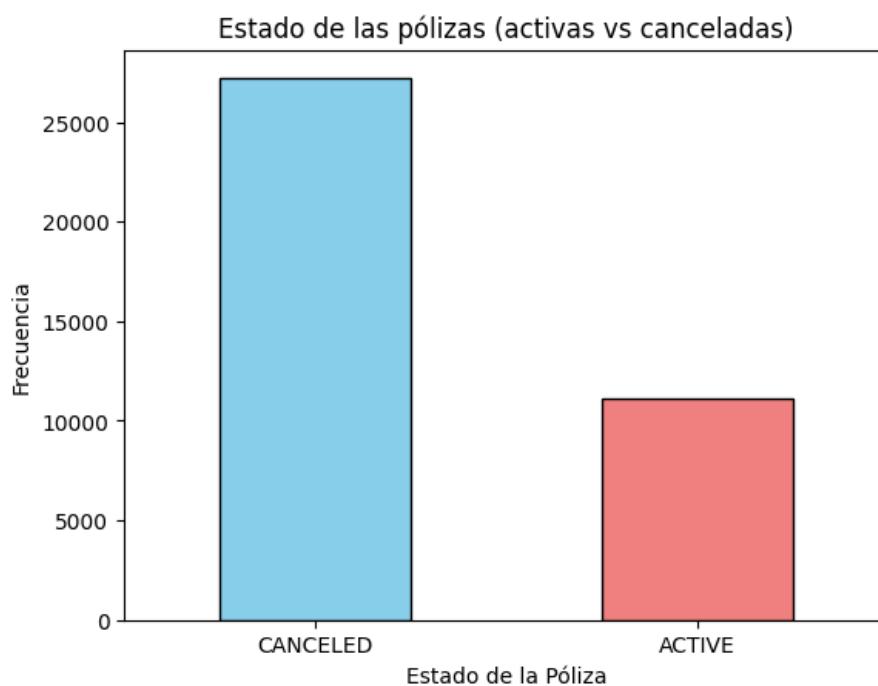
A. Análisis Univariante

El análisis univariante se enfoca en entender la distribución de cada variable de forma individual. En esta etapa, se generan gráficos y estadísticos descriptivos como:

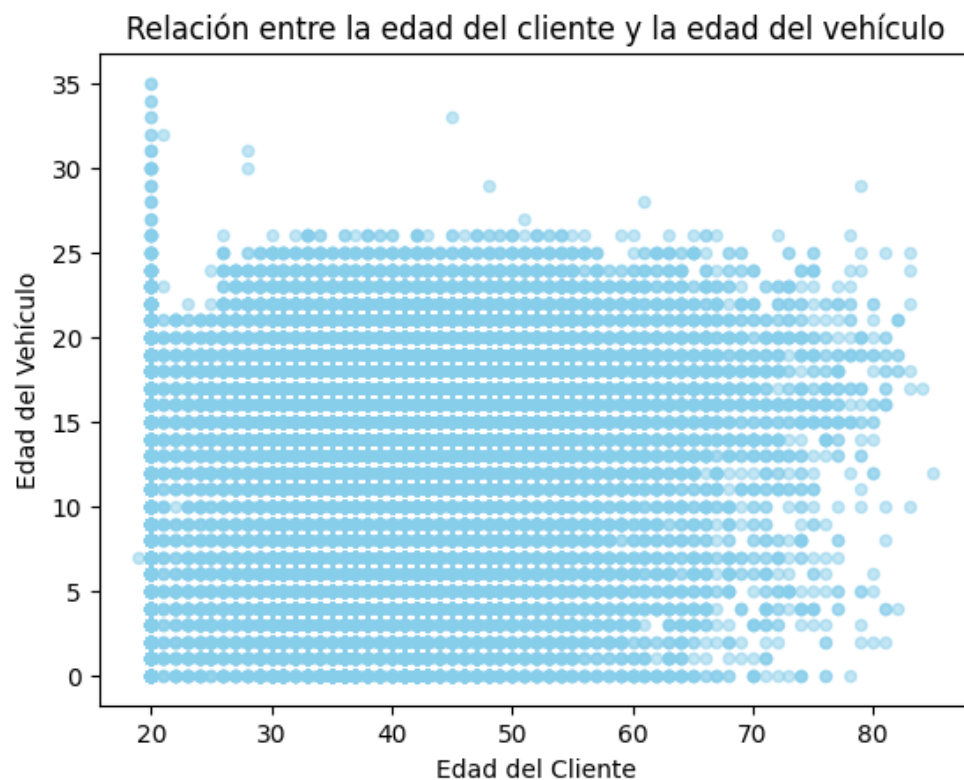
- **Distribución de las pólizas por tipo (Smart y Flex).**



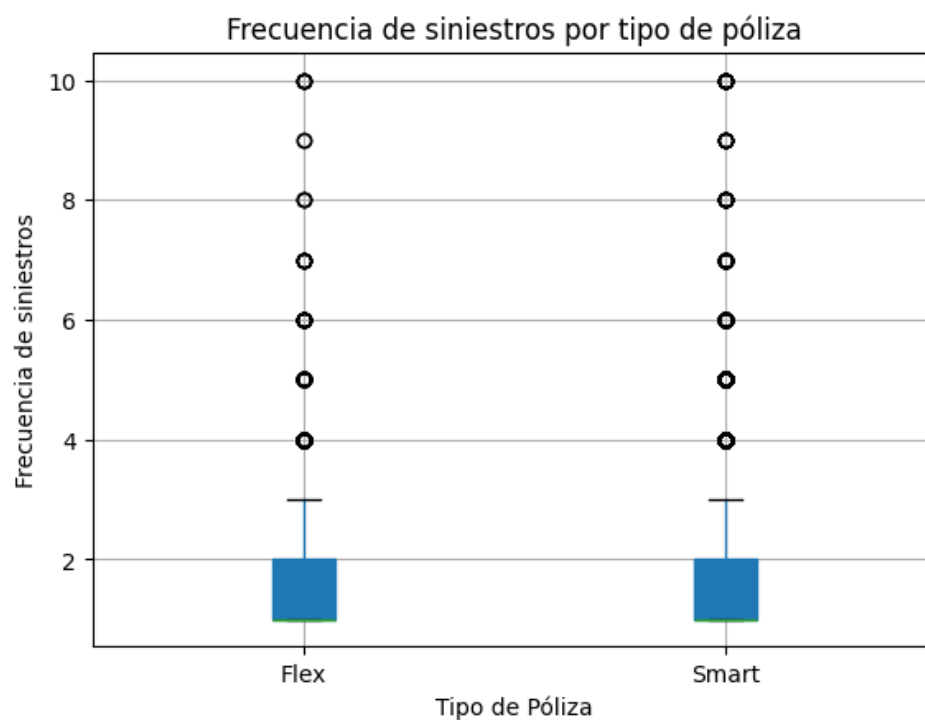
- **Distribución de las pólizas por estado (activas o canceladas).**



- **Edad del cliente y antigüedad del vehículo:** Análisis de las características demográficas y de los vehículos asociados a las pólizas.



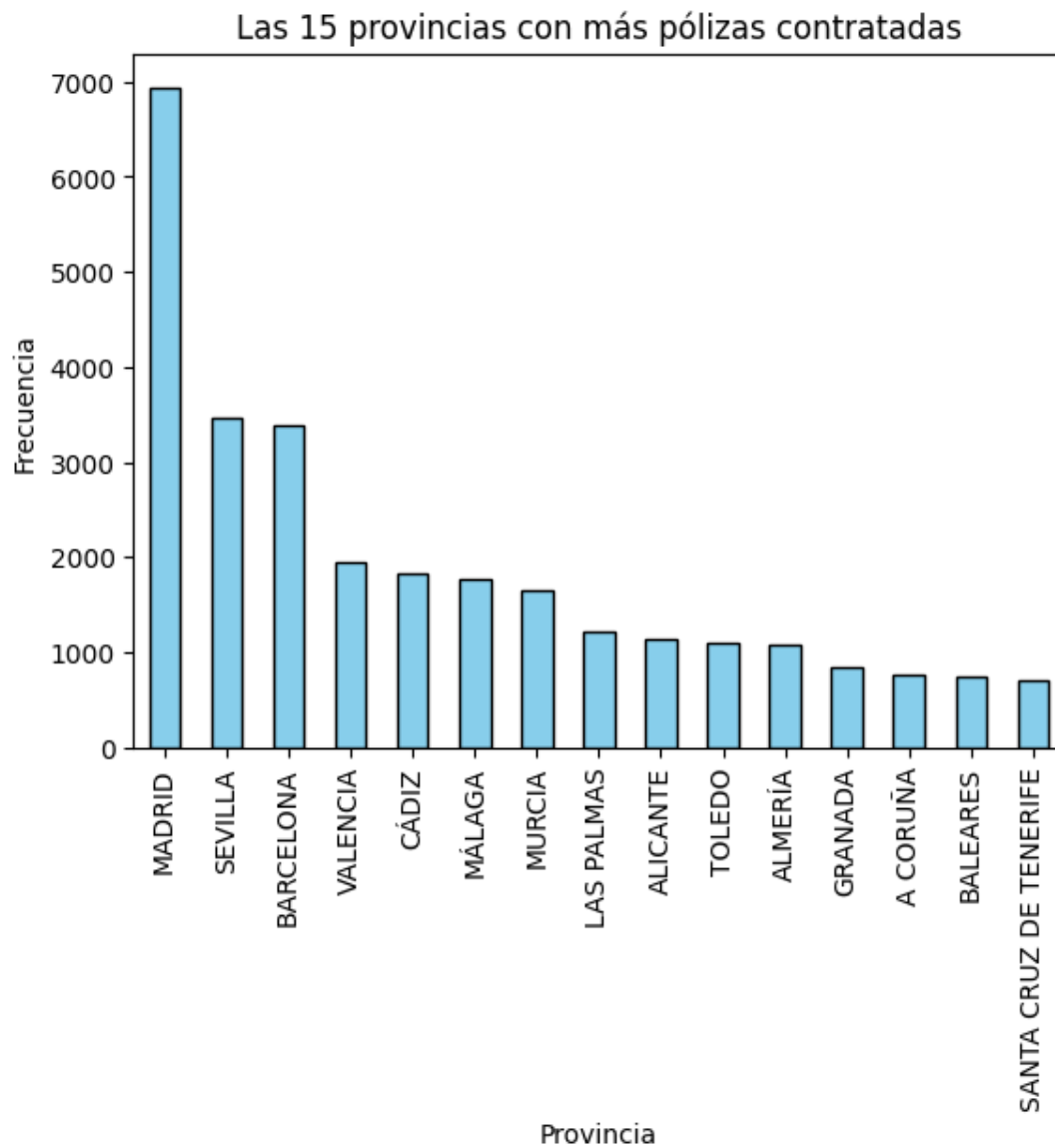
- **Frecuencia de siniestros por tipo de póliza:** Relación de la variable FAclaims con el tipo de póliza.

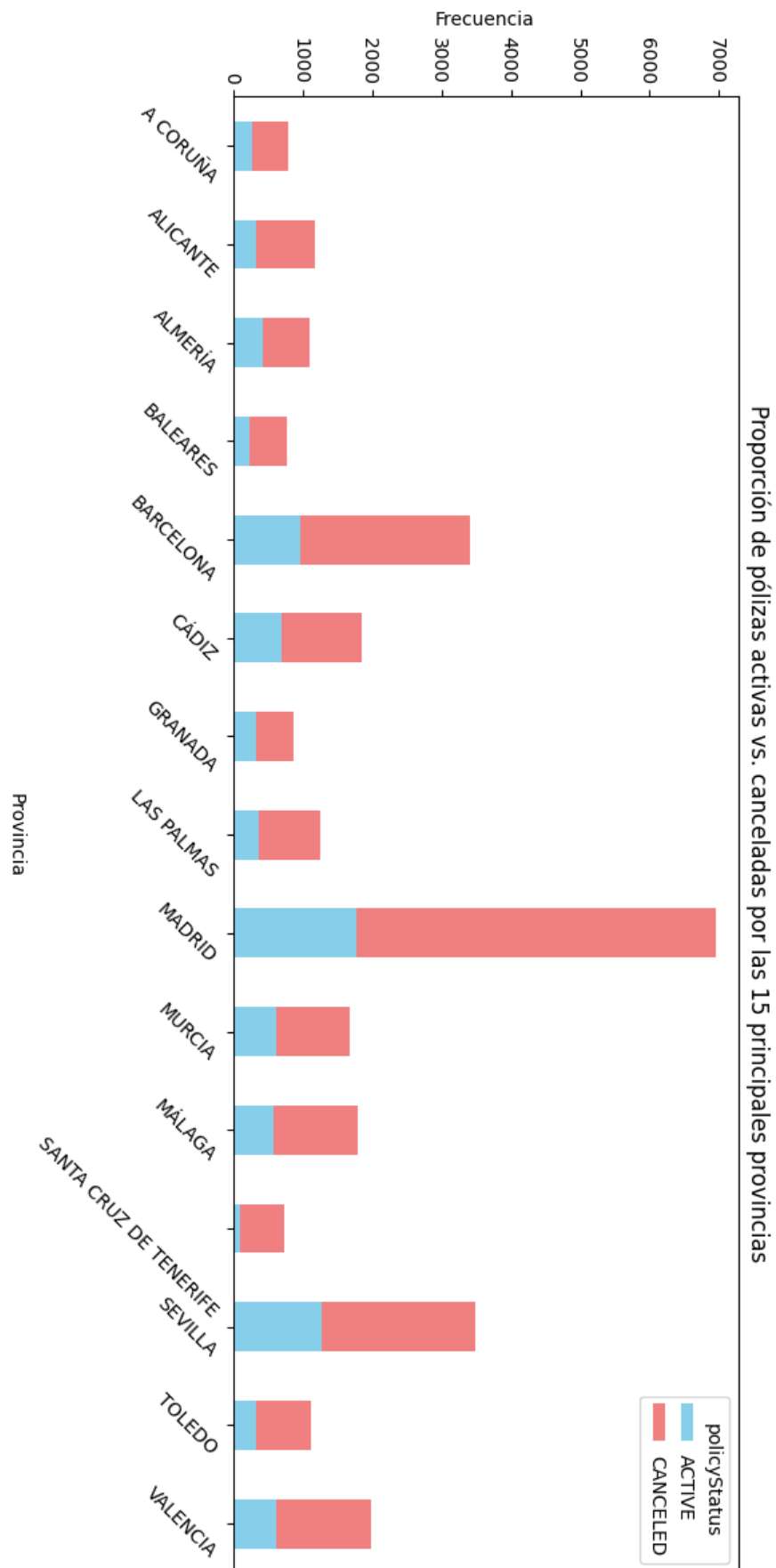


B. Análisis Bivariante

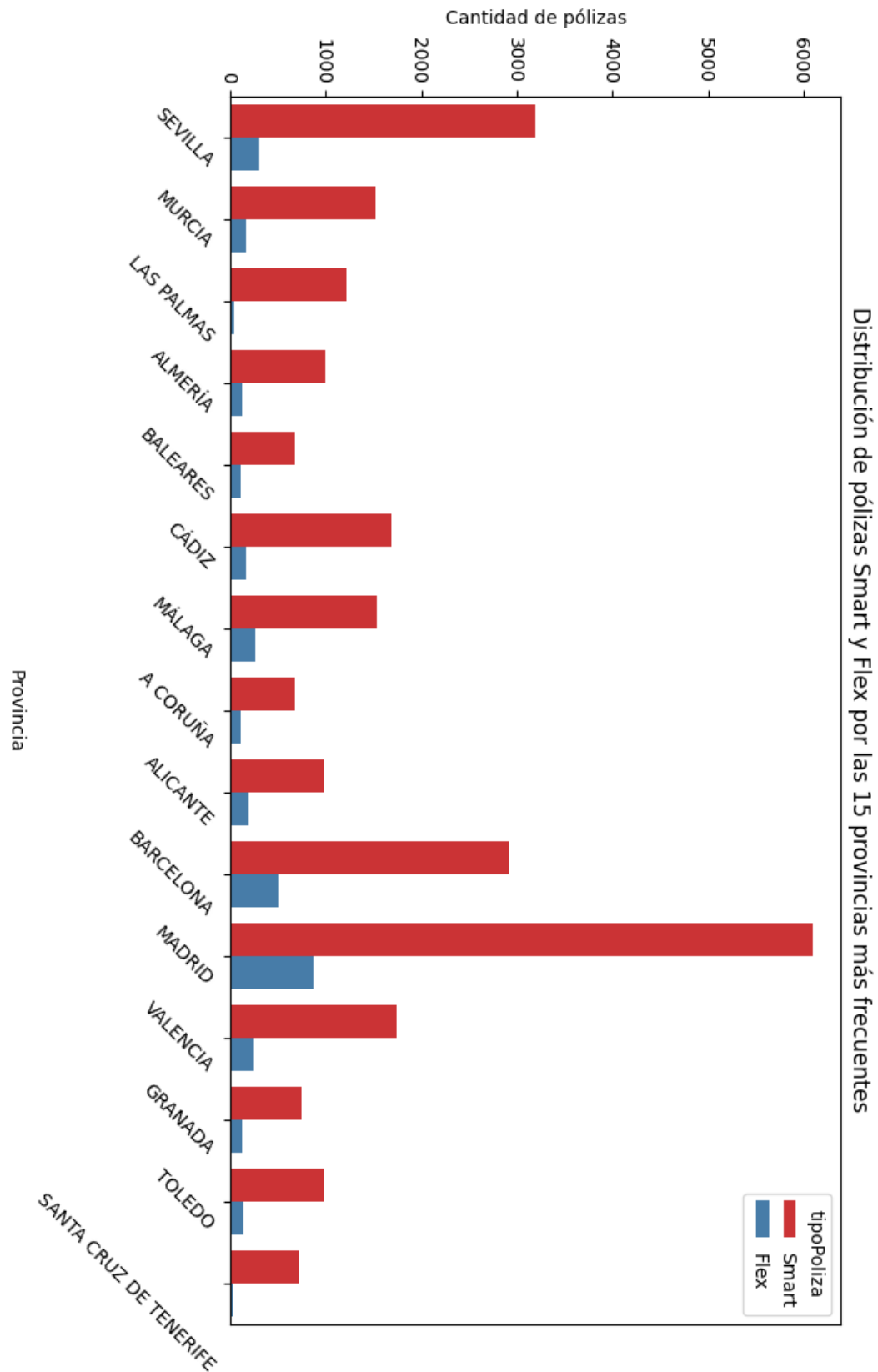
En el análisis bivariante, se investigan las relaciones entre dos variables. Algunos ejemplos de análisis que se realizarán:

- **Proporción de pólizas activas y canceladas por provincia:** Análisis de la distribución de las pólizas activas y canceladas en las provincias. Segmentamos los datos en función de las 15 provincias con más pólizas contratadas, que representan el 90% de la muestra.

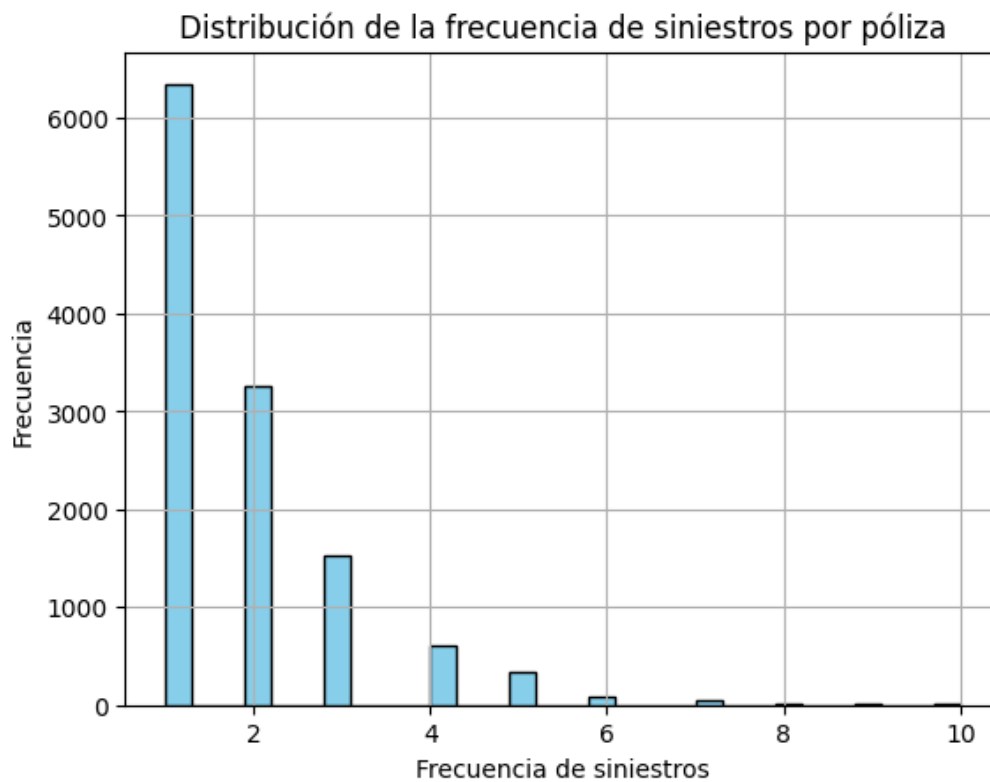
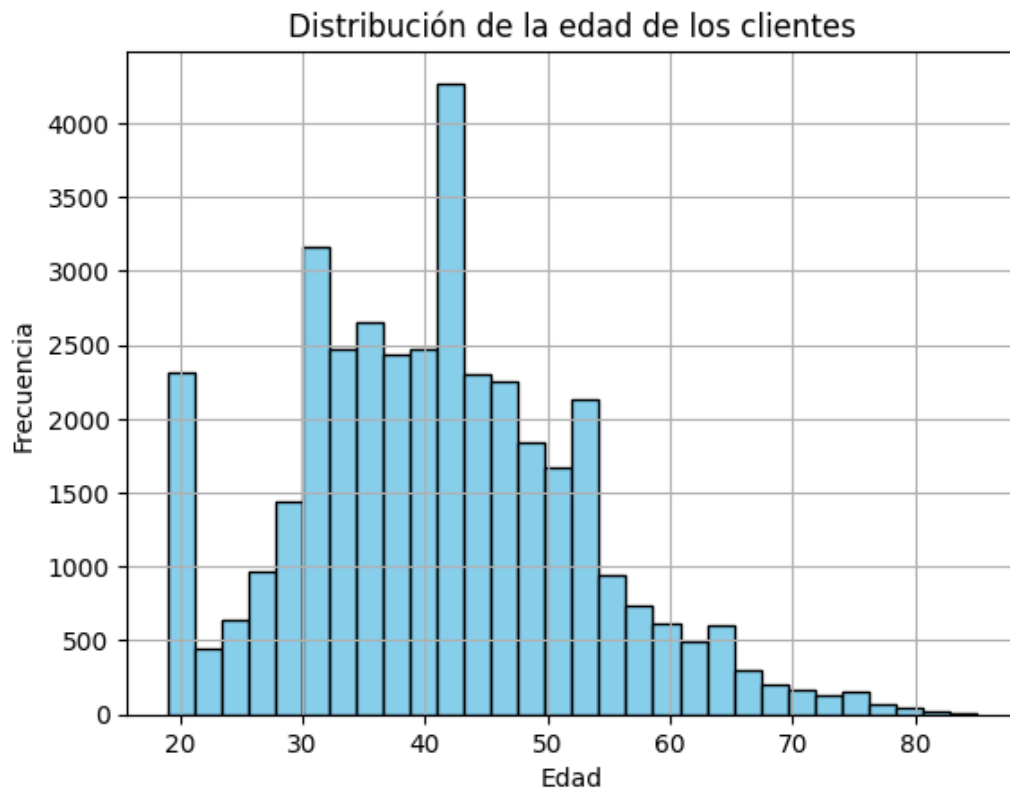


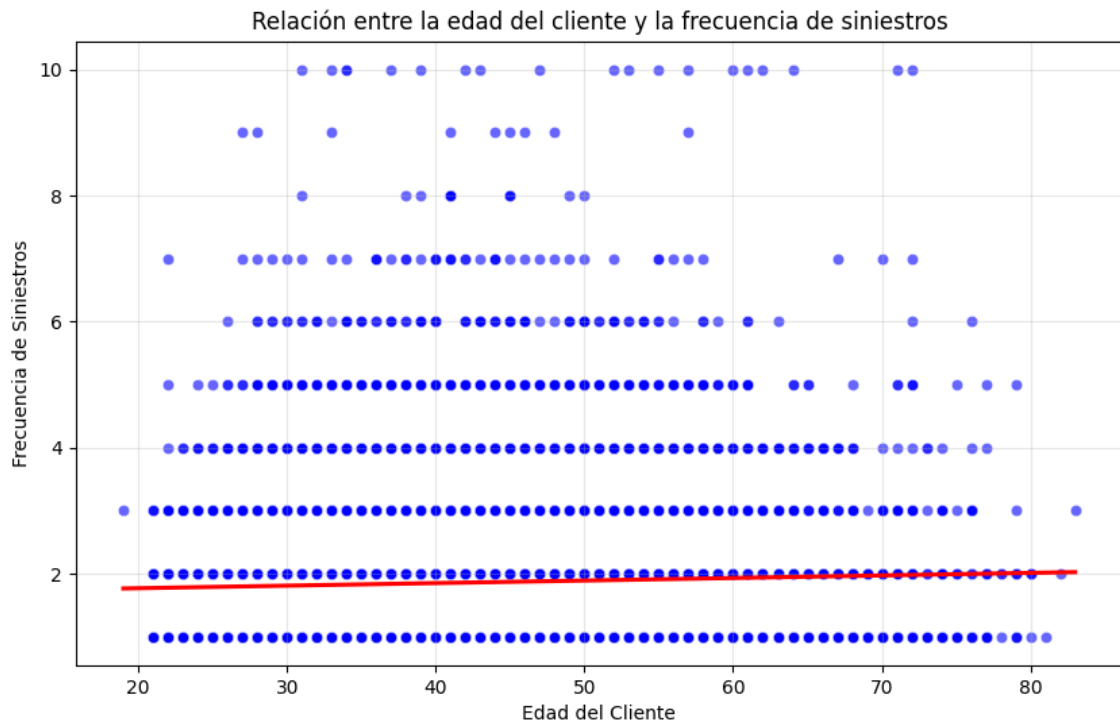


- **Proporción de pólizas Smart y Flex por provincia:** Análisis de la distribución de las pólizas Smart y Flex en las provincias. Segmentamos los datos en función de las 15 provincias con más pólizas contratadas, que representan el 90% de la muestra.

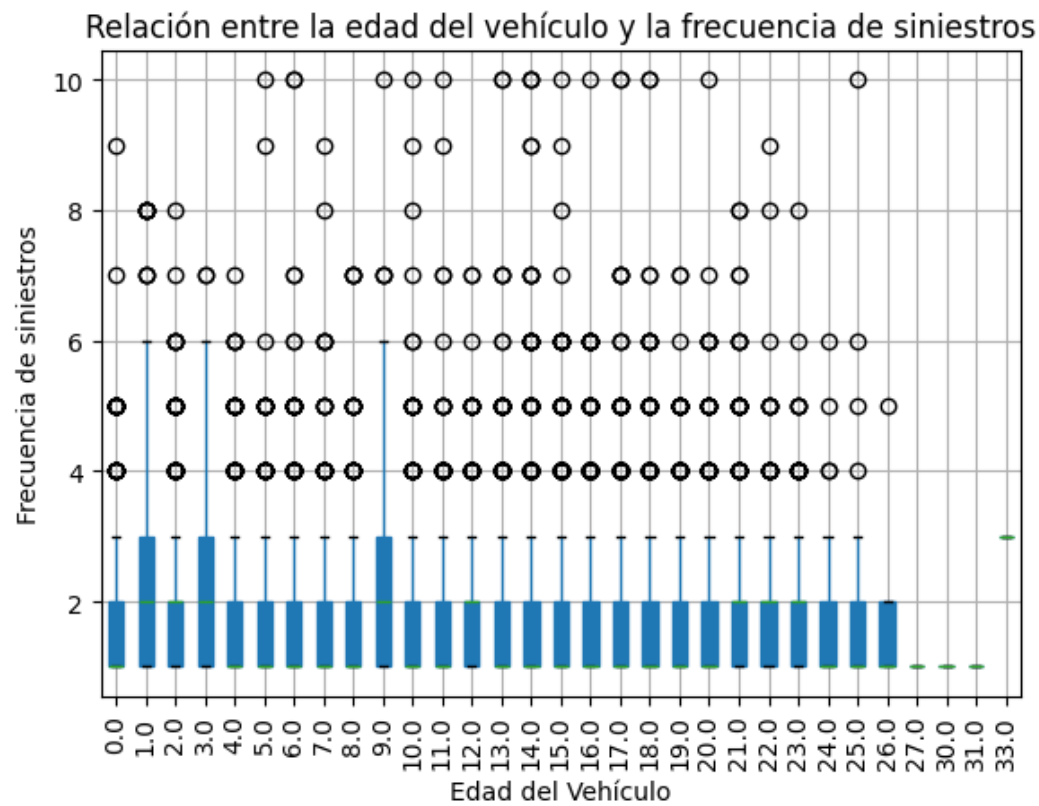


- **Relación entre la edad del cliente y la frecuencia de siniestros:** Se explorará si los clientes de mayor edad presentan más siniestros, o si esta relación es significativa.





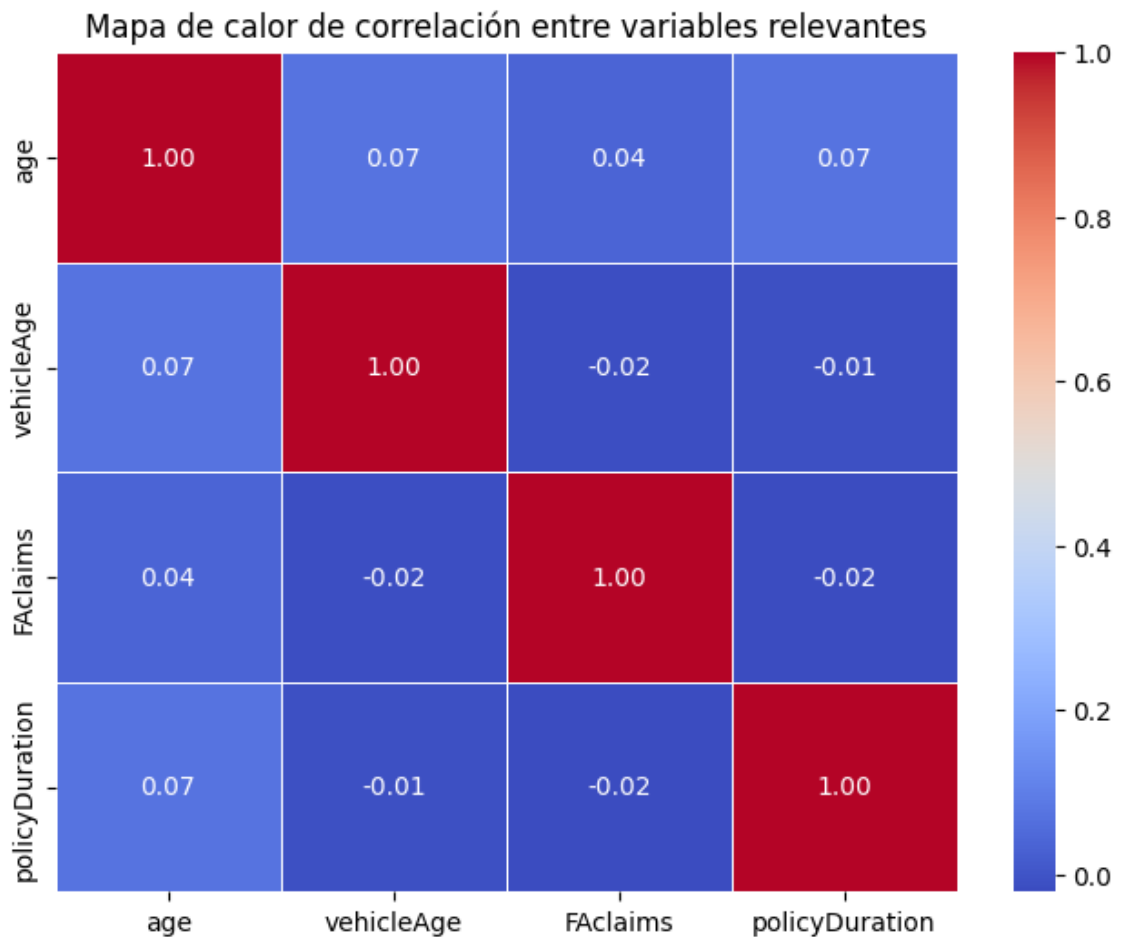
- **Comparación entre la edad del vehículo y la frecuencia de siniestros:** Investigaremos si los vehículos más antiguos tienen una mayor probabilidad de generar siniestros.



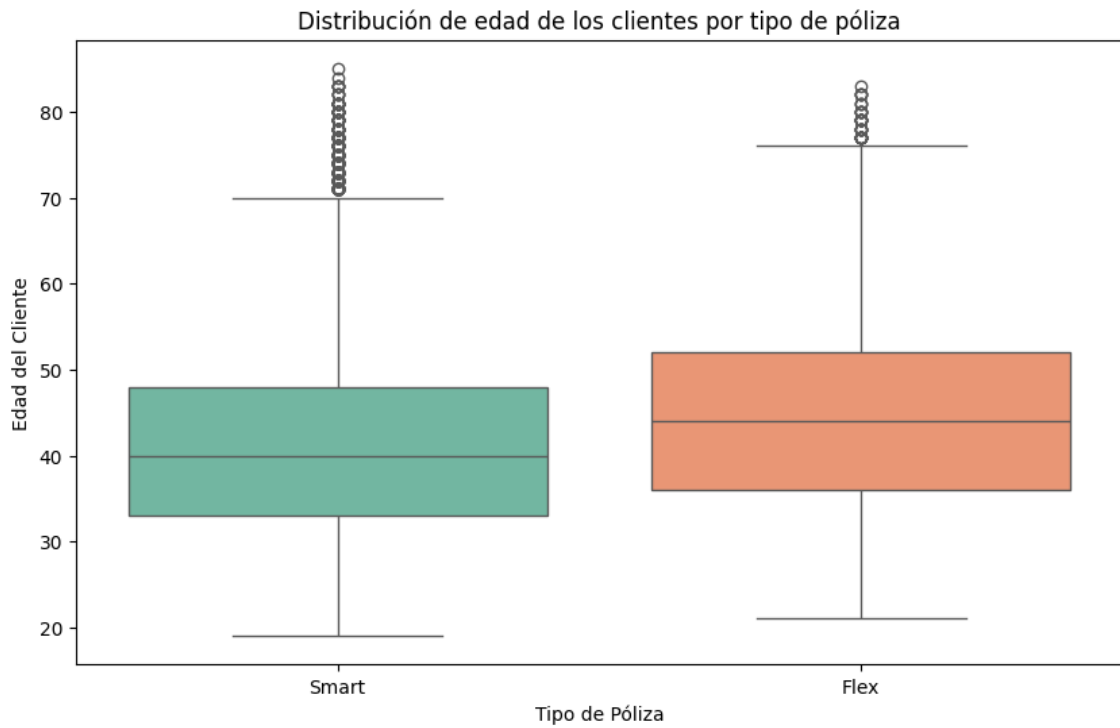
C. Análisis Multivariante

En el análisis multivariante, se explorarán las interacciones entre varias variables para obtener una comprensión más profunda de los factores que afectan la contratación de pólizas. Algunos análisis multivariantes incluyen:

- **Búsqueda de correlación entre variables:** Se realiza un mapa de calor en búsqueda de correlación entre las variables numéricas.



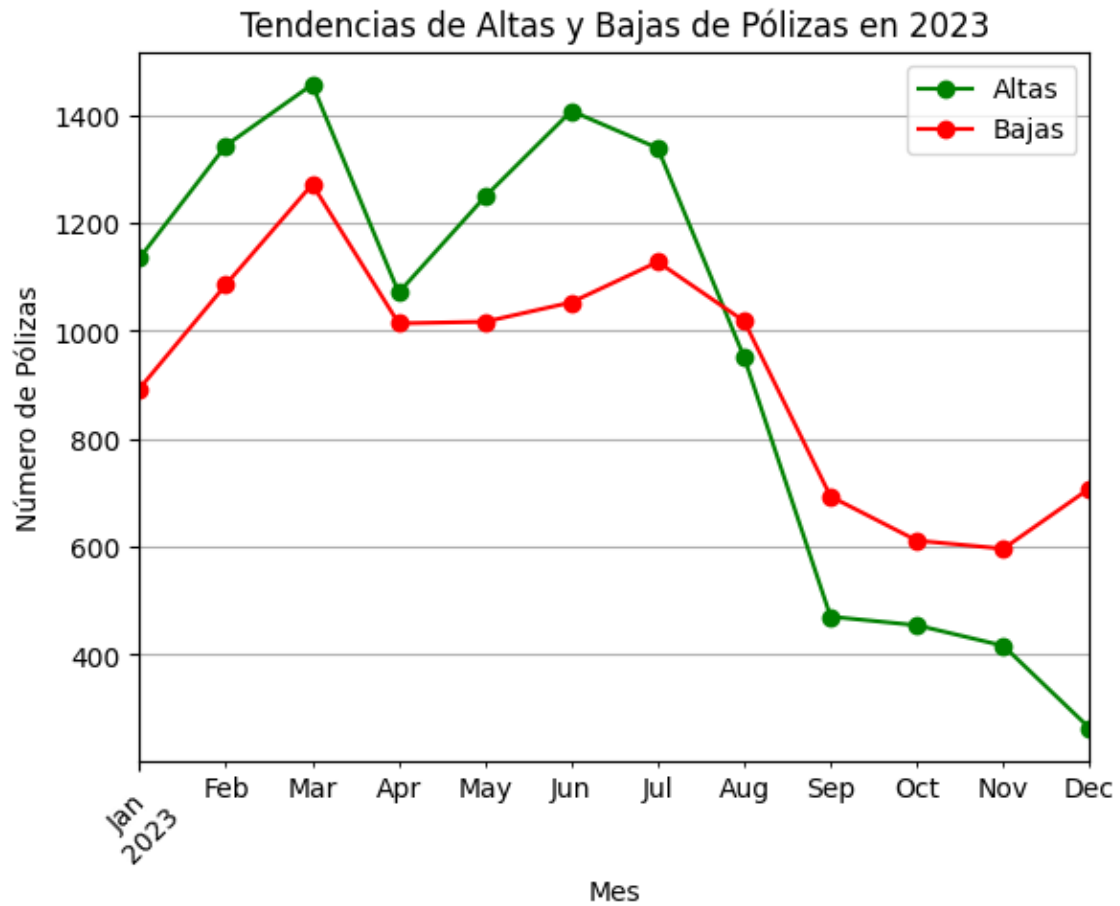
- **Segmentación de clientes:** Agrupación de clientes según múltiples características (edad, tipo de póliza) para identificar patrones de comportamiento.



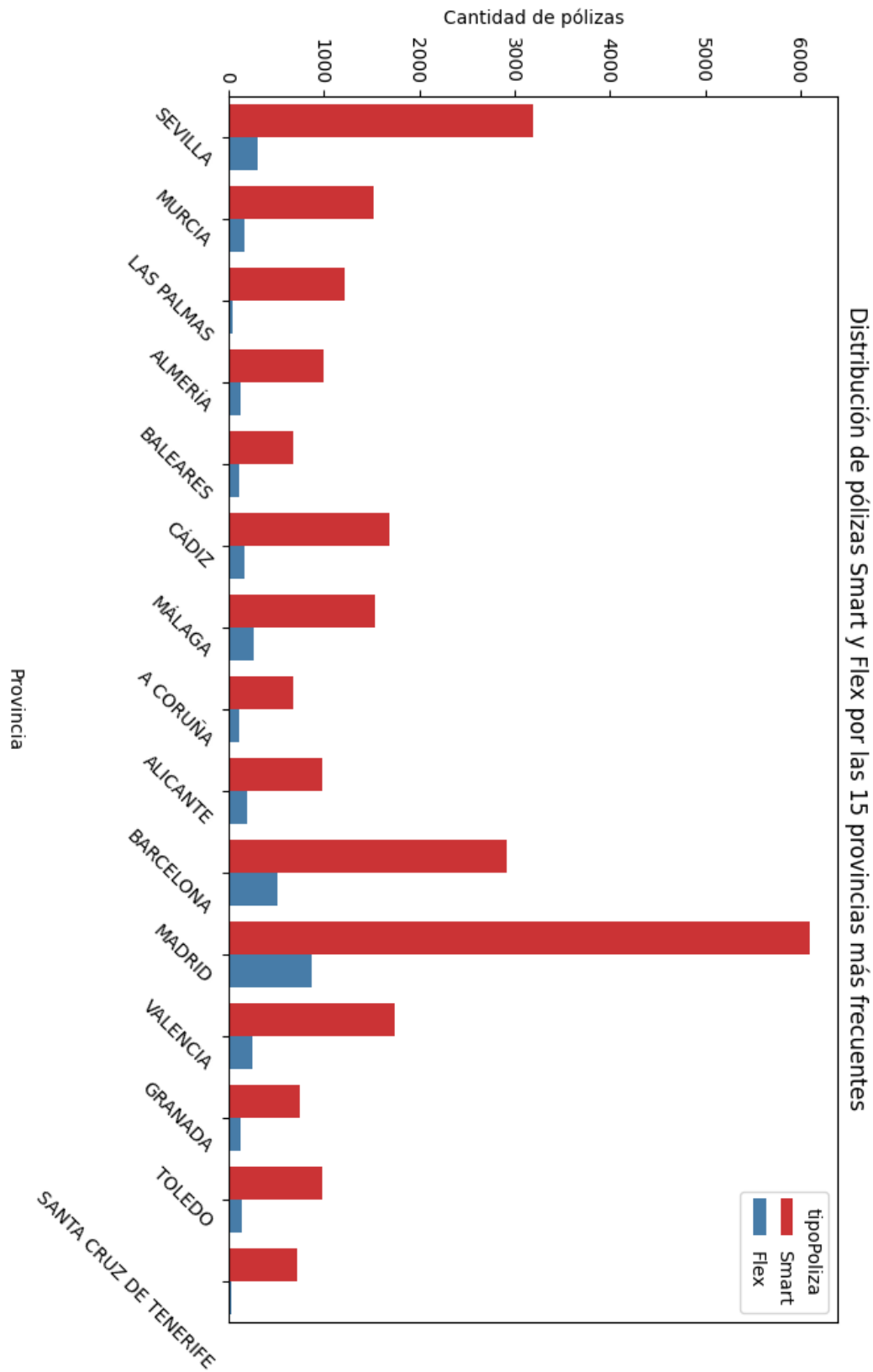
4. Análisis Específico

Aquí se abordan algunos puntos específicos que podrían ser de interés particular para Hello Auto, como:

- **Tendencias en la contratación de pólizas:** Analizar si las pólizas Smart o Flex han tenido un crecimiento o disminución en determinados periodos.



- **Perfil de los clientes por provincia:** Analizar cómo se distribuyen los tipos de pólizas (Smart y Flex) en las diferentes provincias, y si hay diferencias significativas entre regiones.



5. Conclusiones

En esta sección se resumen los hallazgos más importantes del análisis. Algunas posibles conclusiones podrían ser:

1. Segmentación de Clientes:

- Las pólizas Flex son ligeramente más populares entre los clientes de mayor edad.
- El rango de edad de los clientes se sitúa mayormente entre 33 y 48 años, según los percentiles 25% y 75%.
- Los clientes menores de 30 años prefieren mayoritariamente las pólizas Smart debido a su flexibilidad y características tecnológicas avanzadas. En cambio, los mayores de 50 años muestran una preferencia hacia las Flex por su enfoque más tradicional.

2. Tendencias y Estacionalidad:

- Durante 2023, se observó una tendencia descendente en la contratación de pólizas en julio, similar a las bajas, aunque estas últimas disminuyen con menor intensidad.
- A partir de agosto, las bajas superaron a las altas, alcanzando su punto máximo en noviembre.
- Los meses de verano (junio a agosto) presentan una mayor contratación de pólizas Smart, posiblemente debido a viajes y vacaciones.
- Las cancelaciones aumentan notablemente en los últimos meses del año, lo que podría estar relacionado con el fin de ciclos anuales de contratación o ajustes económicos personales.

3. Siniestros:

- El promedio de siniestros es de dos por cliente, sin una relación significativa con la edad.
- La antigüedad del vehículo está positivamente correlacionada con la frecuencia de siniestros, especialmente para vehículos con más de 10 años. Esto sugiere que los vehículos más antiguos son más propensos a incidentes.
- Los clientes más jóvenes, aunque menos propensos a incidentes graves, presentan una mayor frecuencia de pequeños siniestros.

4. Diferencias Regionales:

- Existen disparidades significativas en la distribución de pólizas Smart y Flex entre provincias. Las provincias con mayor densidad urbana tienden a contratar más pólizas Smart, mientras que en regiones rurales o con menos densidad poblacional predominan las Flex.
- Las provincias con mayor dinamismo económico tienden a tener más pólizas activas, lo que sugiere una relación directa entre el contexto económico y la estabilidad de las pólizas.

5. Eficiencia Operativa:

- La mayoría de las pólizas activas se mantienen durante más de 2 años, lo que indica una alta retención y satisfacción del cliente en general.
- Sin embargo, las cancelaciones en los primeros seis meses de contratación representan un área de mejora para ajustar las expectativas del cliente en el proceso inicial.

6. Otros Hallazgos Relevantes:

- El análisis de correlaciones no mostró relaciones significativas entre las variables numéricas estudiadas.

Este análisis proporcionará a Hello Auto una visión más clara de su base de clientes y de las dinámicas que influyen en las decisiones de contratación de pólizas.