**1. Business Case & Data Collection**

**Business understanding**:  
Queremos analizar los datos relacionados con las pólizas de seguros para tomar decisiones estratégicas que mejoren la comprensión del comportamiento de los clientes y su relación con los productos ofrecidos. Los estudios específicos son:

* **Evolución de las pólizas**: Identificar tendencias en altas y bajas de pólizas a lo largo del tiempo.
* **Relación entre clientes y tipos de pólizas**: Identificar qué perfiles de clientes contratan pólizas *Smart* o *Flex*.
* **Siniestros y características de clientes/vehículos**: Determinar qué tipo de cliente es más propenso a tener siniestros y si esto depende de las características del vehículo.

**Hipótesis**:

1. **Evolución de pólizas**: Las altas de pólizas tienen picos en determinadas épocas del año (por ejemplo, inicio de año o verano). Las bajas pueden estar asociadas a falta de renovación o insatisfacción.
2. **Clientes y pólizas**: Las pólizas *Smart* son más populares entre clientes más jóvenes y con vehículos compactos, mientras que las *Flex* se asocian con perfiles familiares y vehículos grandes.
3. **Siniestros**: Clientes jóvenes o con vehículos de gama alta podrían ser más propensos a reportar siniestros.

**Plan de acción**:

* Si detectamos tendencias de bajas, se desarrollarán estrategias de retención, como descuentos o campañas personalizadas.
* La información sobre clientes y pólizas permitirá ajustar estrategias de marketing para mejorar la oferta según el perfil de cliente.
* Los hallazgos sobre siniestros guiarán la implementación de políticas de riesgo, precios personalizados o campañas educativas para reducir siniestros.

**Requerimientos de los datos**:

1. **Evolución de pólizas**: Fechas de alta y baja de pólizas, tipo de póliza, y tiempo de permanencia.
2. **Clientes y pólizas**: Perfil del cliente (edad, género, estado civil, localización), tipo de póliza contratada (*Smart* o *Flex*).
3. **Siniestros y vehículos**: Información del cliente, características del vehículo (marca, modelo, año), y registros de siniestros.

**Disponibilidad**:  
Los datos requeridos están disponibles en la base de datos interna de la compañía. Confirmar que incluyen las fechas de alta/baja, tipos de pólizas, datos de los clientes y registros de siniestros.

**Adquisición de datos**:  
Extraer información de los sistemas internos de la compañía mediante consultas SQL o mediante archivos previamente recopilados.

**Calidad**:  
Verificar que los datos no tienen duplicados, valores inconsistentes, ni faltantes críticos. Revisar la completitud de los registros de clientes y vehículos.

**Revisión de hipótesis**:  
Si falta información clave (por ejemplo, datos incompletos de vehículos), replantear la hipótesis sobre siniestros y limitar el análisis a lo disponible.

**2. Data Understanding**

**Exploratorio inicial**:

* Revisión del número de tablas, columnas y filas.
* Comprobar naturaleza de las variables: fechas, categóricos (*Smart/Flex*), y numéricos (edad, número de siniestros).
* Inspeccionar proporciones de missings en cada columna.

**Tabla de variables**:  
Crear un inventario con información como:

* **Variable**: Nombre de la columna.
* **Tipo**: Categórica, numérica, fecha.
* **Descripción**: Qué representa la variable.
* **Missings**: Porcentaje de valores faltantes.

**3. Data Cleaning**

1. **Juntar tablas**:  
   Unir información de pólizas, clientes, vehículos y siniestros en una tabla maestra.
2. **Eliminar columnas**:  
   Descartar identificadores únicos irrelevantes y otras columnas no relacionadas con los análisis.
3. **Formato de datos**:  
   Convertir fechas, categorías y numéricos al formato adecuado para evitar errores posteriores.
4. **Missings**:

* Investigar la razón detrás de valores faltantes.
* Si los missings son escasos, eliminar filas/columnas; si son relevantes, imputar valores adecuados.

1. **Outliers**:  
   Identificar outliers en variables críticas como edad, siniestros y antigüedad del vehículo, destacando su impacto.
2. **Duplicados**:  
   Eliminar registros duplicados para evitar sesgos en los resultados.
3. **Transformaciones**:

* Crear variables como tiempo activo de póliza (fecha baja - fecha alta).
* Agrupar por rangos de edad y categorías del vehículo (compacto, SUV, lujo).

**4. Analysis**

**Análisis univariante**:

* Evolución temporal de altas y bajas: gráficas de líneas para observar tendencias.
* Distribución de tipos de pólizas: histogramas o diagramas de barras para visualizar proporciones de *Smart* y *Flex*.
* Características de clientes: histogramas de edad, gráficos de torta para género y estado civil.

**Análisis bivariante**:

* Clientes y pólizas: cruces entre perfil del cliente (edad, género) y tipo de póliza.
* Siniestros y vehículos: scatter plots entre edad del cliente, antigüedad del vehículo y número de siniestros.
* Relación entre siniestros y tipo de póliza.

**Análisis multivariante**:

* Heatmap de correlación entre variables relevantes como edad, antigüedad del vehículo, número de siniestros, y tiempo activo de póliza.

**5. Resultados**

1. **Contraste con hipótesis**:

* Identificar si las tendencias de altas/bajas se confirman.
* Validar si el tipo de cliente coincide con el tipo de póliza contratada.
* Evaluar si los clientes más propensos a siniestros tienen características comunes.

1. **Plan de acción**:

* Diseñar estrategias de retención para épocas con más bajas.
* Optimizar campañas de marketing para los segmentos más interesados en *Smart* o *Flex*.
* Ajustar políticas de precios y evaluación de riesgos basadas en siniestros.

1. **Productivización y comunicación**:

* Crear un dashboard para visualizar resultados clave.
* Presentar conclusiones con gráficos y recomendaciones prácticas al equipo directivo.