

## 1 Introducción:

Cuando las actividades de un proveedor de asistencia sanitaria son desviadas de los estándares aceptados por la comunidad médica, causando con ello alguna lesión o daño al paciente, es considerado como negligencia médica o acto realizado de forma incompetente, ya sea por actos no apropiados o por no ser suficientemente diligente respecto al caso en curso. En las pólizas relacionadas, el paciente lesionado o afectado presenta derecho a una indemnización, cuya reclamación puede variar según el ámbito en el que se haya cometido la reclamación, es decir, si está relacionado a sanidad del sector privado o del sector público y el tipo o grado de responsabilidad atribuido al profesional de la salud, ya sea administrativo, civil o penal.

Siendo así este problema de negligencia médica un problema crítico y complejo para las aseguradoras, debido a que en la estimación de provisiones y en la gestión del riesgo, existen factores que dificultan estas estimaciones, lo cual pueden generar considerables desafíos y desequilibrios en la estabilidad económica de las compañías. Dentro de ellos se pueden destacar los siguientes:

- La variabilidad que caracteriza a estas reclamaciones tanto en severidad, esto es, la magnitud del fallo médico, como en frecuencia u ocurrencia, sumado a la probabilidad de que el paciente presente la demanda, el período de tiempo entre el siniestro y la reclamación y el criterio judicial que define el resultado, sea compensación o exención.
- La diversidad y complejidad de los casos que pueden originar o implicar una mala praxis médica, casos específicos que requieren un estudio y una valoración pericial singulares, así como la adecuada y obligatoria clasificación del riesgo según el tipo de servicio sanitario, la institución médica y el rol o función del profesional, entre otros factores.
- Gran repercusión social y mediática, que puede derivar en un aumento de la demanda y la conciencia de los afectados, así como en una mayor responsabilidad y presión sobre las aseguradoras y los médicos.
- La accesibilidad y calidad de los datos históricos sobre la siniestralidad, que son imprescindibles para aplicar métodos estadísticos o actuariales que permitan estimar la frecuencia y la severidad de las demandas futuras.
- La atención a los factores externos que pueden incidir en la evolución de las demandas, como los cambios legislativos, jurisprudenciales, sociales o tecnológicos, que pueden alterar las expectativas y el comportamiento de los agentes involucrados.
- La medición de la incertidumbre asociada a las estimaciones, que supone considerar los posibles escenarios alternativos y sus probabilidades, así como fijar márgenes de prudencia o solvencia que aseguren la suficiencia de las reservas.

Nuestro enfoque se centra en las aseguradoras que operan en Estados Unidos en la línea de accidentes de propiedad, donde se presenta un aspecto crítico y complejo: las reclamaciones por negligencia médica. Estas reclamaciones afectan tanto al ámbito personal como al comercial, lo que aumenta la necesidad de un análisis preciso y sofisticado para calcular las provisiones correspondientes.

Los criterios de éxito del modelo se basan en validar las estimaciones futuras que genera el modelo con los valores reales que se requieren para cubrir las provisiones en esos momentos. De esta forma, se obtiene un nivel de confianza o de funcionalidad del proyecto.

## 2 Objetivos y Alcance:

El siguiente documento tiene como propósito presentar el proyecto de desarrollo de un modelo de pronóstico para estimar la reserva óptima para las reclamaciones por negligencia médica en los próximos períodos, considerando las características y complejidades de este tipo de reclamaciones, que pueden tener un alto impacto financiero y reputacional para la compañía.

El proyecto se basa en el uso de un método estadístico adecuado para el problema y los datos, que permita predecir el comportamiento y las tendencias de las reclamaciones por negligencia médica, utilizando variables históricas y futuras. El proyecto busca responder a la pregunta de cuál es el nivel óptimo de reserva que se debe mantener para cubrir las reclamaciones por negligencia médica, sin incurrir en una subestimación que ponga en riesgo la solvencia de la compañía, ni en una sobreestimación que implique una pérdida de oportunidades de inversión.

El proyecto se alinea con los objetivos del negocio, ya que aporta beneficios o valor para la compañía, tales como:

- Mejorar la eficiencia y efectividad de la asignación de recursos para las reservas, optimizando el uso del capital y evitando posibles sanciones o pérdidas por insuficiencia o exceso de reservas.
- Aumentar la confianza y credibilidad de la compañía ante los clientes, los accionistas, los reguladores y el mercado, demostrando una gestión responsable y transparente de las reclamaciones por negligencia médica.
- Generar conocimiento e información valiosa sobre el comportamiento y las tendencias de las reclamaciones por negligencia médica, que permitan diseñar estrategias preventivas y correctivas para reducir su frecuencia y severidad.

El proyecto se desarrollará en colaboración con diversas áreas clave dentro de la organización, que proveerán información o datos relevantes para el análisis y el modelado. Estas áreas son:

1. Área de Reclamaciones, que es la encargada de recibir, tramitar y pagar las reclamaciones por negligencia médica, y que tiene un conocimiento profundo y detallado de las características y circunstancias de cada caso.

Proveerá el histórico detallado de las reclamaciones ejecutadas, permitiendo la identificación de patrones y tendencias.

Facilitará la información sobre los años de ocurrencia de las reclamaciones y los tiempos de respuesta.

2. Área Financiera y Contable, que es la responsable de registrar, controlar y reportar los movimientos financieros y contables relacionados con las reclamaciones por negligencia médica, y que tiene un conocimiento amplio y actualizado de los estados financieros y contables de la compañía.

Proporcionará datos cruciales relacionados con las pérdidas acumuladas pagadas, las primas ganadas y la prima ganada cedida.

Ofrecerá información acerca de la reserva actual destinada a las reclamaciones en cuestión.

3. Área de Reservas, que es la responsable de calcular, revisar y ajustar las reservas para las reclamaciones por negligencia médica, y que tiene un conocimiento técnico y metodológico de los criterios y modelos utilizados para estimar las reservas.

Suministrará datos históricos y proyecciones futuras en relación con las reservas establecidas previamente para enfrentar las reclamaciones.

El proyecto se realizará siguiendo los estándares de calidad, es decir, asegurando que el modelo sea preciso, confiable y útil para el negocio, utilizando medidas e indicadores apropiados para evaluarlo. El proyecto se entregará en el tiempo acordado con el cliente o patrocinador del proyecto. Se respetarán las leyes y regulaciones que aplican al sector asegurador y al tratamiento de datos personales.

El proyecto se enfrentará a algunas restricciones o riesgos durante su desarrollo, que se tratarán de mitigar mediante acciones preventivas o correctivas. Estos son:

La disponibilidad o calidad de los datos, que puede afectar al alcance temporal o geográfico del proyecto, así como a la precisión o confiabilidad del modelo. Para mitigar este riesgo se realizará una revisión exhaustiva de los datos disponibles, verificando su completitud, consistencia, precisión y relevancia. También se aplicarán técnicas de limpieza o transformación de datos para corregir o eliminar los valores erróneos o faltantes.

La complejidad o incertidumbre del problema, que puede dificultar la elección o aplicación del método estadístico más adecuado para el problema y los datos. Para mitigar este riesgo se realizará una investigación previa sobre los métodos estadísticos existentes para este tipo de problemas, evaluando sus ventajas e inconvenientes. También se realizarán pruebas piloto con diferentes métodos para comparar sus resultados y elegir el más apropiado.

La resistencia al cambio o la falta de apoyo por parte de algunas áreas o personas involucradas en el proyecto, que puede afectar a la colaboración o coordinación necesaria para el éxito del proyecto. Para mitigar este riesgo se realizará una comunicación constante y efectiva con todas las partes interesadas, explicando los objetivos, beneficios y expectativas del proyecto. También se solicitará y se incorporará el feedback y las sugerencias de las áreas o personas involucradas, buscando su compromiso y participación activa en el proyecto.

La ejecución exitosa de este modelo reviste una gran relevancia para la toma de decisiones en el ámbito financiero y contable, así como para la gestión eficaz de las reservas de la compañía. Adicionalmente, esta iniciativa se alinea con los objetivos del equipo comercial, ofreciendo una ventaja competitiva al abordar con mayor precisión y estrategia las complejidades asociadas con las reclamaciones por negligencia médica.

### **3 Comprensión del Dominio:**

Las reclamaciones por negligencia médica en los Estados Unidos están reguladas por las leyes estatales y federales, así como por las entidades que supervisan el sector de la salud y el sector asegurador. Las leyes estatales establecen los requisitos y procedimientos para presentar una demanda por negligencia médica, así como los plazos de prescripción, los límites de responsabilidad y las normas de prueba. Las leyes federales se aplican cuando la reclamación involucra a programas o entidades del gobierno federal, como Medicare o el Departamento de Asuntos de Veteranos. Las entidades que regulan el sector de la salud y el sector asegurador son, entre otras, la Administración de Alimentos y Medicamentos (FDA), la Oficina del Inspector General (OIG), la Asociación Nacional de Comisionados de Seguros (NAIC) y la Oficina Federal de Protección Financiera del Consumidor (CFPB).

Las partes involucradas en una reclamación por negligencia médica son el demandante, que es la persona que alega haber sufrido un daño por la actuación o la omisión de un profesional de la salud; el demandado, que es el profesional de la salud o la entidad que presta el servicio médico; y la aseguradora, que es la compañía que cubre los riesgos del demandado. Las partes tienen derechos y obligaciones recíprocos, como el derecho a ser informados, asesorados y representados legalmente; el derecho a presentar pruebas, argumentos y testigos; el derecho a impugnar o apelar las decisiones judiciales; y la obligación de actuar con buena fe, veracidad y diligencia.

Los criterios que se usan para determinar la responsabilidad y el monto de las indemnizaciones son los siguientes:

- La existencia de un deber de cuidado por parte del demandado hacia el demandante, es decir, que haya una relación profesional entre ellos que implique una obligación de actuar con competencia y diligencia.
- La violación del deber de cuidado por parte del demandado, es decir, que haya actuado con negligencia, imprudencia o impericia, incumpliendo los estándares profesionales aceptados.
- La causalidad entre la violación del deber de cuidado y el daño sufrido por el demandante, es decir, que haya una relación directa y previsible entre la conducta del demandado y el perjuicio del demandante.
- La existencia y cuantificación del daño sufrido por el demandante, es decir, que haya una lesión física, psicológica o patrimonial que pueda ser valorada económicamente.

El análisis del mercado y la competencia para las aseguradoras que operan en la línea de accidentes de propiedad en Estados Unidos debe considerar los siguientes aspectos:

- El tamaño y las características del mercado. Según datos de la NAIC, el mercado de seguros de accidentes de propiedad en Estados Unidos representó \$ 697 mil millones en primas directas escritas en 2020, lo que supone un 3.1% más que en 2019. Este mercado se compone principalmente por los segmentos de seguros personales (como automóviles y viviendas) y seguros comerciales (como responsabilidad civil y propiedad comercial). El mercado se caracteriza por tener una alta penetración (el ratio entre primas directas escritas y PIB fue del 3.2% en 2020), una alta concentración (las diez primeras compañías aca-pararon el 51.5% del mercado en 2020) y una alta competencia (se registraron más de 2.600 compañías activas en 2020).
- Las ventajas o desventajas de las aseguradoras frente a sus competidores. Las aseguradoras pueden tener ventajas o desventajas competitivas según su capacidad para ofrecer productos diferenciados, precios competitivos, calidad en el servicio al cliente, eficiencia operativa, innovación tecnológica, solvencia financiera y reputación corporativa. Estos factores pueden influir en la captación y retención de clientes, así como en la rentabilidad y sostenibilidad del negocio.
- Las oportunidades o amenazas que se presentan para este sector. El sector de seguros de accidentes de propiedad en Estados Unidos enfrenta oportunidades y amenazas derivadas del entorno económico, social, tecnológico, regulatorio y competitivo. Algunas de estas son las siguientes:
  - Oportunidades:
    - \* El aumento de la demanda de seguros por parte de los consumidores y las empresas, impulsado por la recuperación económica tras la pandemia del COVID-19, el crecimiento demográfico, el aumento de la riqueza y el patrimonio, y la mayor conciencia sobre los riesgos.
    - \* El desarrollo de nuevos productos y servicios que respondan a las necesidades y expectativas de los clientes, aprovechando las oportunidades que brindan las nuevas tecnologías, los nuevos canales de distribución, los nuevos modelos de negocio y los nuevos segmentos de mercado.
    - \* La mejora de la eficiencia y la rentabilidad del negocio, mediante la optimización de los procesos operativos, la reducción de los costos, el uso de herramientas analíticas y predictivas, la gestión activa del capital y el riesgo, y la diversificación de las fuentes de ingresos.
  - Amenazas:
    - \* El aumento de la siniestralidad y la volatilidad del negocio, debido al incremento de la frecuencia y severidad de los eventos catastróficos naturales o provocados por el hombre, como huracanes, incendios, terremotos, ciberataques o actos terroristas.
    - \* El endurecimiento de la regulación y la supervisión del sector, que puede implicar mayores requisitos de solvencia, transparencia, protección al consumidor, cumplimiento normativo y responsabilidad social y ambiental.
    - \* La intensificación de la competencia y la presión sobre los márgenes, originada por la entrada de nuevos competidores (como insurtechs o gigantes tecnológicos), la consolidación del sector (mediante fusiones o adquisiciones), la guerra de precios (por la mayor sensibilidad al precio de los clientes) y la disrupción del modelo tradicional (por el cambio en las preferencias y comportamientos de los clientes).

El estudio de las necesidades y expectativas de los clientes o usuarios de las aseguradoras debe considerar los siguientes aspectos:

- El tipo de cobertura o indemnización que buscan o demandan los clientes. Los clientes buscan o demandan una cobertura o indemnización que les proteja adecuadamente frente a los riesgos que puedan afectar a su vida, salud o patrimonio. Esto implica que la cobertura o indemnización sea suficiente para cubrir el daño sufrido, sea accesible en términos de precio y condiciones, sea flexible para adaptarse a las circunstancias personales o profesionales, y sea ágil para tramitarse y pagarse.

- El nivel de satisfacción o insatisfacción que tienen con el servicio o la atención de las aseguradoras. Los clientes tienen un nivel de satisfacción o insatisfacción con el servicio o la atención de las aseguradoras que depende de varios factores, como la calidad del producto ofrecido, el trato recibido por parte del personal o del canal de contacto, la facilidad o dificultad para contratar o cancelar el seguro, la rapidez o lentitud para resolver las consultas o reclamaciones, la transparencia o confusión en la información proporcionada, y la percepción de valor o desvalor que tienen del seguro contratado.
- Los factores que influyen en su lealtad o deserción. Los clientes pueden ser leales o desertores a una aseguradora según su grado de satisfacción o insatisfacción con el servicio o la atención recibidos, así como según otros factores que pueden influir en su decisión de mantener o cambiar de aseguradora. Algunos de estos factores son: el nivel de confianza o desconfianza que tienen en la aseguradora, el grado de recomendación o crítica que hacen a otros potenciales clientes, la existencia o ausencia de incentivos o penalizaciones para permanecer o salir del seguro y la disponibilidad o escasez.

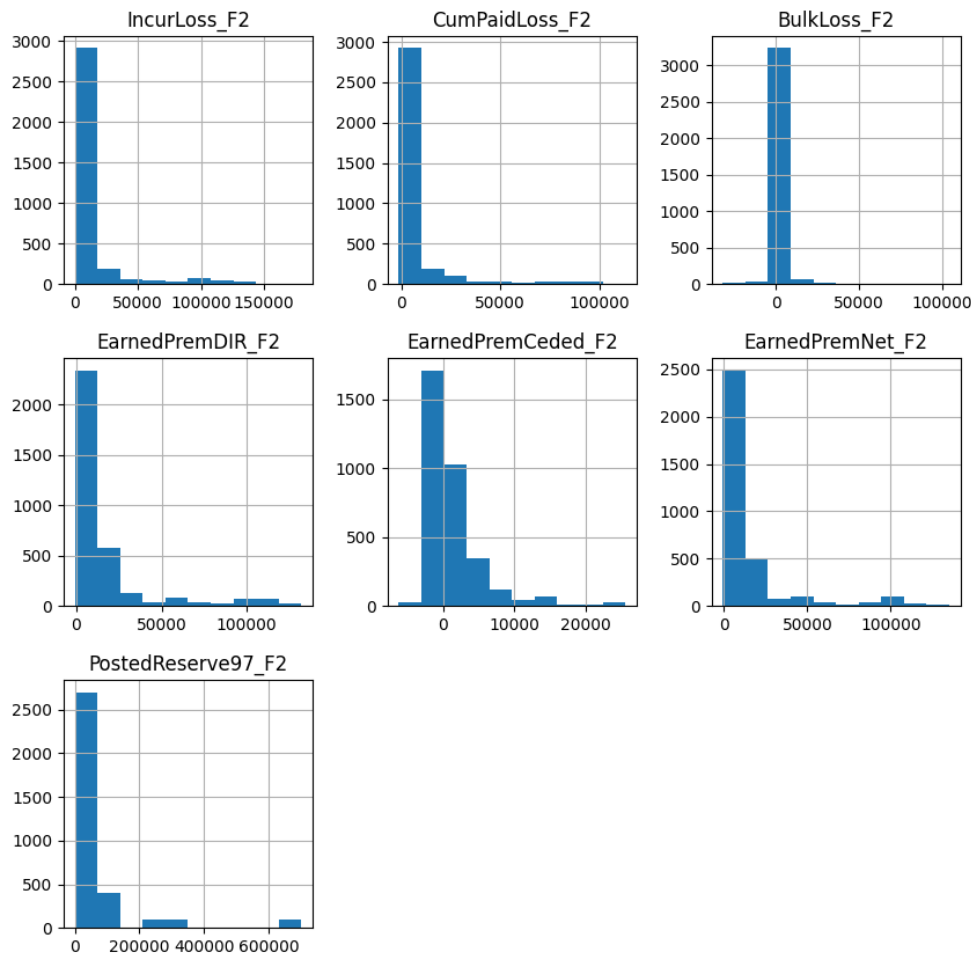
## 4 Desarrollo del problema:

Este modelo se basa en las compañías aseguradoras de accidentes de propiedad en Estados Unidos que se especializan en seguros de responsabilidad civil profesional que incluyen seguros de mala praxis médica. Los datos se obtuvieron del Anexo P - Análisis de pérdidas y gastos de pérdida, que forma parte de la base de datos de la Asociación Nacional de Comisionados de Seguros (NAIC por sus siglas en inglés). Por lo tanto, se establece que el proceso de recolección e integración de datos ya está hecho. Las variables de la base de datos son:

- **Pérdidas Incurridas (IncurLoss):** Esta variable representa las pérdidas totales que una compañía de seguros ha experimentado hasta la fecha, incluyendo las pérdidas que ya se han pagado y las que aún están pendientes de pago. Es una variable fundamental, ya que es la base para calcular las reservas futuras. Los valores negativos pueden hacer referencia a un exceso de liquidez después de pagar las reclamaciones con las reservas destinadas para ello, o que la compañía haya recibido reembolsos o recuperaciones por parte de terceros, como reaseguradores o subrogados, que superen el monto de las pérdidas pagadas o estimadas.
- **Cumulative Paid Losses (CumPaidLoss):** Estas son las pérdidas acumuladas que ya se han pagado a los asegurados. Es importante para evaluar el historial de pagos de pérdidas y cómo ha evolucionado con el tiempo. Los valores negativos pueden hacer referencia a un exceso de liquidez después de pagar las reclamaciones con las reservas destinadas para ello, o que la compañía haya recibido reembolsos o recuperaciones por parte de terceros, como reaseguradores o subrogados, que superen el monto de las pérdidas pagadas. Esto puede ocurrir, por ejemplo, cuando se resuelve una disputa legal a favor de la compañía o cuando se recupera parte del daño causado por un siniestro.
- **Reservas a Granel (BulkLoss):** Esta variable se refiere a las reservas establecidas para cubrir pérdidas futuras que se espera que surjan de eventos pasados pero que aún no se han notificado. Es esencial para estimar la cantidad de fondos que la compañía debe mantener para hacer frente a reclamaciones futuras. Los valores negativos pueden hacer referencia a una necesidad de liquidez en la reserva, la cual para esa fecha estaba ausente, o que la compañía haya sobreestimado el monto de las reservas necesarias en el pasado y haya realizado ajustes a la baja en el presente. Esto puede ocurrir, por ejemplo, cuando se resuelven reclamaciones con un costo menor al esperado o cuando se reduce la frecuencia o la severidad de los siniestros.
- **Primas Ganadas (EarnedPremDIR, EarnedPremCeded, EarnedPremNet):** Estas variables representan las primas ganadas en el año de incurrimento, desglosadas en primas directas, cedidas y netas. Las primas ganadas son cruciales para determinar la relación entre las primas generadas y las pérdidas incurridas. Los valores negativos pueden hacer referencia a que la compañía haya devuelto parte de las primas cobradas a los asegurados, por ejemplo, por cancelación anticipada de la póliza, por reducción del riesgo asegurado o por aplicación de bonificaciones. Esto puede ocurrir, por ejemplo, cuando se produce una disminución de la siniestralidad o una mejora de las condiciones del mercado.
- **Posted Reserves in Year 1997 (PostedReserve97):** Estas son las reservas publicadas en el año 1997, tomadas del Anexo de Suscripción e Inversiones - Parte 2A, incluyendo pérdidas netas no pagadas y gastos no pagados de ajuste de pérdidas. Pueden proporcionar una referencia importante para la evolución de las reservas a lo largo del tiempo. Los valores negativos pueden hacer referencia a que

la compañía haya realizado revisiones a la baja de sus estimaciones de reservas, debido a una mejor información sobre la frecuencia y la severidad de los siniestros, o a cambios en las normas contables o regulatorias. Esto puede ocurrir, por ejemplo, cuando se cierran reclamaciones con un costo menor al previsto o cuando se aplican criterios más estrictos para reconocer las reservas.

- Entidad Individual o Grupo (Single): Esta variable indica si se trata de una entidad individual (1) o una aseguradora de grupo (0). Esta distinción puede ser relevante para comprender las diferencias en la gestión de reservas entre compañías individuales y grupos de aseguradoras. Por lo tanto, se realizará un análisis general sin tomar en cuenta esta variable; y además, un análisis discriminado por esta variable, con el fin de identificar y describir adecuadamente la información y si estos valores cambian mucho para cada variable.



Con el uso del histograma y de las medidas de tendencia central, se logra identificar que las variables *IncurLossF2*, *CumPaidLossF2*, tienen sesgo a la derecha con cola pesada porque sus medianas son 645 y 187 respectivamente mientras que su máximo valor está en 179,425.0 y 113,189.0 respectivamente. Mientras que *BulkLossF2* presenta un sesgo a los dos lados siendo más pesada a la derecha porque su mediana es 0, es decir, presenta mayor rango de variabilidad a la derecha. Para las variables *EarnedPremDIRF2*, *EarnedPremNetF2*, *PostedReserve97F2*, todas presentan un sesgo a la derecha, cuyas medianas están dadas por 1,500.0, 1,302.0 y 5,875.0 respectivamente mientras que sus valores máximos están dados por 131,948.0, 25,553.0, 135,318.0 y 702,246.0. Finalmente la variable *EarnedPremCededF2* presenta sesgo a la derecha, pero muchos valores negativos cercanos a cero o con muy poco sesgo a la izquierda.

En la variable *IncurLossF2*, no es normal que tenga valores tan altos y tan dispersos. Esto puede indicar que la compañía ha tenido muchos siniestros graves o costosos, o que ha subestimado el riesgo de sus pólizas. También puede indicar que la compañía ha tenido dificultades para recuperar parte de las pérdidas por parte de terceros, como reaseguradores o subrogados. Siendo así, las compañías deberían revisar su política de tarificación y suscripción, así como su gestión de reservas, para evitar tener pérdidas tan elevadas en el futuro.

En la variable *CumPaidLossF2*, no es normal que tenga valores negativos. Esto puede indicar que la compañía ha recibido más reembolsos o recuperaciones por parte de terceros, como reaseguradores o subrogados, que el monto de las pérdidas pagadas. Esto puede ocurrir, por ejemplo, cuando se resuelve una disputa legal a favor

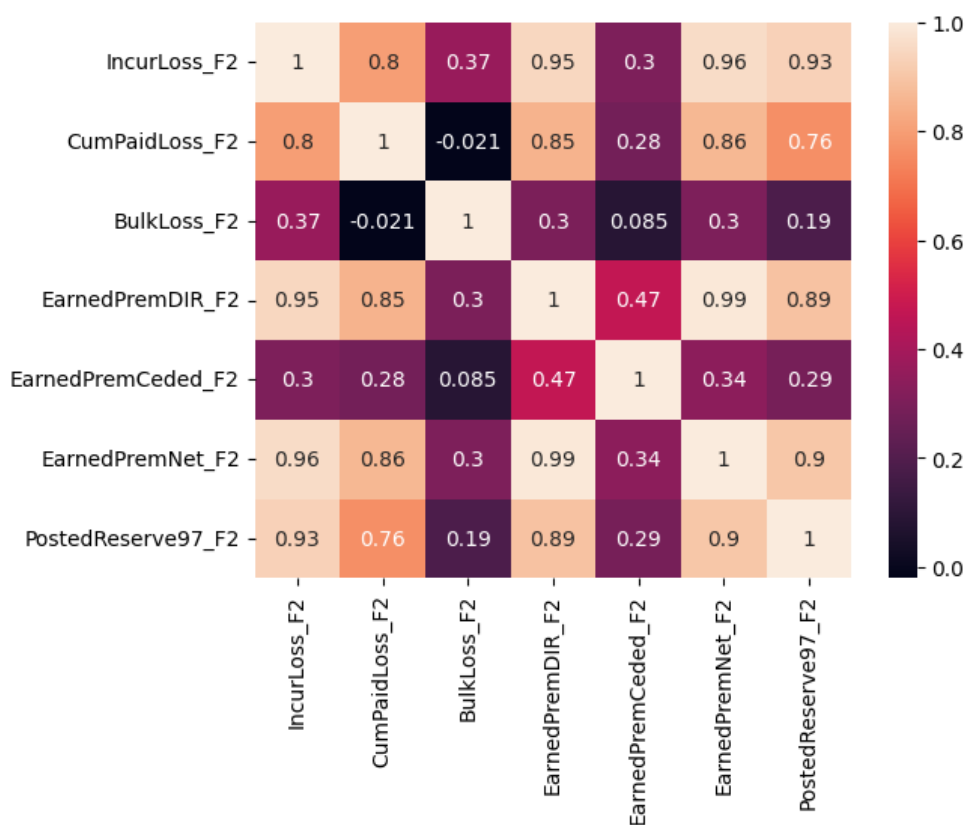
de la compañía o cuando se recupera parte del daño causado por un siniestro. La compañía debería revisar su política de reclamaciones y ajuste de pérdidas, así como su relación con los terceros involucrados, para evitar tener valores negativos en esta variable.

Para la variable BulkLossF2, no es normal que tenga valores negativos. Esto puede indicar que la compañía ha sobreestimado el monto de las reservas necesarias en el pasado y ha realizado ajustes a la baja en el presente. Esto puede ocurrir, por ejemplo, cuando se resuelven reclamaciones con un costo menor al esperado o cuando se reduce la frecuencia o la severidad de los siniestros. Aquí, la compañía debería revisar su política de estimación y reconocimiento de reservas, así como su metodología y criterios para hacerlo, para evitar tener valores negativos en esta variable.

Para las variables EarnedPremDIRF2, EarnedPremNetF2 y PostedReserve97F2, no es normal que tengan valores tan altos y tan dispersos. Esto puede indicar que la compañía ha cobrado primas muy elevadas a sus asegurados, o que ha establecido reservas muy conservadoras para hacer frente a las reclamaciones futuras. Esto puede ocurrir, por ejemplo, cuando la compañía tiene una posición dominante en el mercado o cuando tiene una aversión al riesgo muy alta. Aquí, que la compañía debería revisar su política de tarificación y suscripción, así como su política de estimación y reconocimiento de reservas, para evitar tener valores tan elevados en estas variables.

Para la variable EarnedPremCededF2, no es normal que tenga valores negativos, ya que representa las primas cedidas a los reaseguradores. Esto puede indicar que la compañía ha recibido más primas de los reaseguradores que las que les ha pagado. Esto puede ocurrir, por ejemplo, cuando la compañía cancela parte de sus contratos de reaseguro o cuando recibe bonificaciones por su buen desempeño.

Estos valores de las métricas de tendencia central, así como la forma y la desviación de la distribución, podrían variar y mejorar si se analizan por separado las aseguradoras individuales y las de grupo. Esto se debe a que parece que las aseguradoras de grupo son las que están sesgando tanto estas variables.

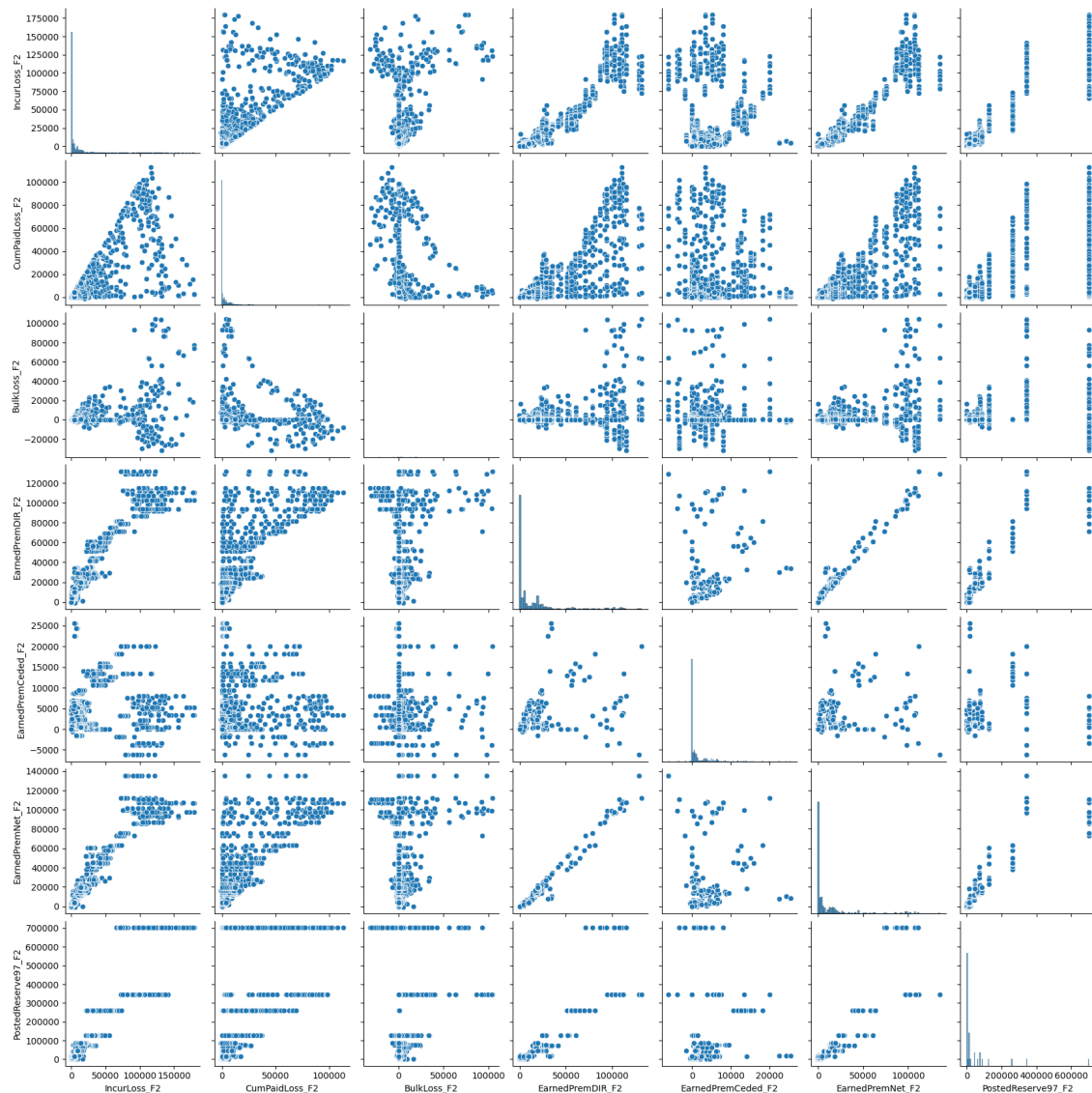


Por lo cual se puede deducir que las variables que tienen correlación directa son:

- IncurLossF2 con CumPaidLossF2, EarnedPremDIRF2, EarnedPremNetF2 y PostedReserve97F2.
- CumPaidLossF2 con EarnedPremDIRF2, EarnedPremNetF2 y PostedReserve97F2.
- EarnedPremDIRF2 con EarnedPremNetF2 y PostedReserve97F2.
- EarnedPremNetF2 con PostedReserve97F2.

Recordando así que la correlación es una relación simétrica.

Note que BulkLossF2 y EarnedPremCededF2 presentan niveles de correlación cercanos a cero, esto implica que no tienen ningún comportamiento lineal respecto del resto de variables. Viendo estas correlaciones gráficamente se tiene:



Los gráficos parecen un poco dispersos y sin una forma particular, sin embargo, tal vez al clasificar por aseguradoras individuales y de grupo, esta tendencia sea un poco más clara.

Sin embargo, se observa que hay algunas variables que tienen correlación lineal casi igual a 1, por ejemplo, EarnedPremNetF2 vs. EarnedPremDIRF2, esto porque no tiene puntos muy desviados de la línea de tendencia, como si fueran la misma variable.

Éste fenómeno se debe a que hay una relación matemática directa entre estas variables, diferente a la línea de tendencia de esa correlación, dada por:

$$\text{EarnedPremNetF2} = \text{EarnedPremDIRF2} - \text{EarnedPremCededF2} \quad (1)$$

Esto es porque sus definiciones están dadas por:

**Prima directa:** Corresponde a los ingresos por la venta de seguros efectuada por el total de las compañías, rebajadas las anulaciones e incluyendo su participación en coaseguro.

**Prima cedida:** Corresponde a la parte de la prima directa o prima aceptada, que las compañías traspasan al reasegurador a través de contratos de reaseguro proporcional.



**Prima neta o retenida:** Corresponde a la prima directa, más la prima aceptada y menos la prima cedida. Es el valor neto de las primas, retenido por la aseguradora.

por lo tanto se puede decir que son la misma variable salvo una constante de subida o bajada que es la variable `EarnedPremCededF2`.

El resto de variables que tienen una correlación positiva cercana a 1 están relacionadas con el tipo de dato y su significado. Por ejemplo, la prima, el riesgo, la pérdida y la reclamación son variables que se afectan mutuamente: a mayor prima, mayor riesgo; a mayor riesgo, mayor pérdida; y a mayor pérdida, mayor reclamación.

- **IncurLossF2:** muestra la situación de las pérdidas de las aseguradoras. De ellas se tiene:

2 aseguradoras tienen pérdidas nulas en todas las entradas y una de ellas, pérdidas muy bajas en el año más actual de la muestra, lo que significa que no registraron pérdidas en este periodo. Esto se explica porque iniciaron sus actividades en 1994 y 1995, respectivamente, y aún no se han consolidado en el mercado de seguros ni han recibido reclamaciones en este breve lapso de 3 y 2 años.

3 aseguradoras no reportan pérdidas en los años más recientes de la muestra. Esto podría deberse a que son compañías con mucha trayectoria en el mercado, fundadas en 1688, 1967 y 1980, respectivamente. Al tener una buena posición económica, pueden asumir deducibles altos y no registrar pérdidas menores. Además, pueden hacer mejores estimaciones de las pérdidas y contar con coseguros para cubrirlas, lo que reduce el impacto de los siniestros.

14 aseguradoras registran pérdidas nulas en los primeros años de la muestra. Esto puede obedecer a dos situaciones: que sean aseguradoras nuevas que tardan en recibir reclamaciones o que sean aseguradoras con gran capacidad de mercado que no consideran pérdidas menores. No obstante, puede haber cambios en las políticas internas de las compañías que hagan que las pérdidas pasen a ser positivas en algún momento. Por ejemplo, al evaluar el valor de las pérdidas y compararlo con los posibles coseguros, puede resultar más conveniente reportar las pérdidas.

El resto de las aseguradoras reportan pérdidas positivas. Esto indica que han tenido reclamaciones que han cubierto con sus reservas y su solvencia económica, sin importar su antigüedad o su posición en el mercado asegurador.

Solo una aseguradora, Homestead Ins Co, reporta pérdidas negativas en el año 1993 y en los rezagos de 7 a 10 años. Esto podría deberse a que se modificó el valor de las pérdidas por alguna razón legal o a que hubo un excedente de liquidez. Con el tiempo, estas pérdidas se ajustaron al valor previamente calculado.

**Matriz media y de desviación estándar:** La media y desviación estándar de esta variable por cada entrada de las matrices muestran una tendencia creciente desde la esquina superior derecha hasta la inferior izquierda. Esto se explica por dos factores: el primero es el periodo de los rezagos, ya que las compañías suelen pagar más en los años cercanos a la reclamación; el segundo es el tiempo de reclamación, que implica que las reclamaciones se encarecen por efectos económicos como la inflación, la devaluación y otros cambios en los mercados.

**Matriz de correlación:** Se observa que cada posición en las matrices tiene una correlación con el resto de las entradas. Esta correlación es más fuerte para los años más próximos al año estudiado. Se analizan 3 años (1988, 1992 y 1997) y se confirma que las correlaciones más altas se dan entre los años cercanos. Sin embargo, la correlación entre los otros años tampoco es muy baja, lo que sugiere que hay una relación entre todos los años. Con esta tendencia de correlación, se puede implementar el modelo de predicción para los años siguientes.

**Nota:** Por ahora se considerarán todos los elementos para el cálculo de estas métricas, pero en principio, teniendo en cuenta que hay aseguradoras que no estaban presentes en el mercado durante un intervalo de tiempo, tener en cuenta estos valores de cero sería sesgar la muestra. Para tener un análisis específico del monto pagado y reportado como pérdida, se deberían excluir todos los valores de 0, en los tiempos tales que las aseguradoras no existen.

- **CumPaidLossF2:** Los argumentos para esta variable son análogos a los anteriores por ser variables de pérdidas, solo que en la variable IncurLossF2 es la pérdida en el tiempo  $t$  con un rezago de  $x$  y para la variable CumPaidLossF2 es el acumulado, en el mismo tiempo  $t$  y un rezago de  $x$ .
- **BulkLossF2:** Esta variable muestra la situación de las reservas establecidas para cubrir pérdidas futuras que se espera que surjan de eventos pasados pero que aún no se han notificado. De ellas se tiene:

3 aseguradoras tienen reservas nulas en todas las entradas, incluyendo las dos aseguradoras de las variables anteriores que tienen ceros en todas sus componentes por ser tan nuevas en el mercado. La aseguradora adicional es Great Amer Grp, la cual inició actividades en 1971. Esto implica que estas aseguradoras no han tenido siniestros con demora en el aviso, otra posible razón es que han estimado correctamente el costo total de los siniestros ocurridos en el pasado y no tenga ninguna reserva adicional para cubrir siniestros no avisados. Esto podría deberse a que la compañía tenga una buena gestión de riesgos, un control eficiente de los procesos de reclamación y una base de datos confiable y actualizada, por otro lado, podría suceder, que la compañía haya utilizado una metodología de cálculo de la reserva de siniestros ocurridos no avisados que sea muy conservadora y que haya liberado el exceso de reserva en el periodo actual

6 aseguradoras registran reservas nulas en casi toda la muestra, lo que sugiere que estas aseguradoras han tenido una baja frecuencia o severidad de los siniestros o que han ajustado sus políticas de suscripción y tarificación para reducir la incertidumbre sobre el costo final de los siniestros.

7 aseguradoras no reportan reservas en los años más recientes de la muestra, lo que puede deberse a un cambio en la metodología de cálculo o a una falta de información actualizada sobre los siniestros ocurridos en el pasado.

10 aseguradoras registran reservas nulas en los primeros años de la muestra, lo que puede reflejar una mayor incertidumbre sobre el costo final de los siniestros o una mayor demora en el aviso en esos periodos.

La existencia o no de reservas de siniestros ocurridos no avisados tiene implicaciones para la solvencia y rentabilidad de las aseguradoras, pero también puede ser un riesgo si se produce un cambio adverso en la frecuencia o severidad de los siniestros, o si se presentan reclamaciones inesperadas por eventos pasados.

Los valores negativos, presentes en 14 aseguradoras, de las cuales 5 presentan este patrón negativo por largos periodos de tiempo, pueden hacer referencia a una necesidad de liquidez en la reserva, la cual para esa fecha estaba ausente, o que la compañía haya sobreestimado el monto de las reservas necesarias en el pasado y haya realizado ajustes a la baja en el presente. Esto puede ocurrir, por ejemplo, cuando se resuelven reclamaciones con un costo menor al esperado o cuando se reduce la frecuencia o la severidad de los siniestros.

**Matriz media:** La media de esta variable por cada entrada de las matrices muestra una tendencia decreciente desde la esquina superior izquierda hasta la inferior derecha, exceptuando algunos casos. Es decir, la media de esta variable es mayor para los rezagos más cortos y los tiempos de reclamación más largos, ya que esto implica que hay más siniestros ocurridos en el pasado que aún no han sido notificados y que requieren una reserva y además que a partir de la historia las aseguradoras han mejorado su modelo de estimación de reservas. Por el contrario, la media sea menor para los rezagos más largos y los tiempos de reclamación más cortos, ya que esto implica que hay menos siniestros ocurridos en el pasado que aún no han sido notificados y que requieren una reserva.

**Desviación estándar:** La desviación estándar se mantiene casi constante sobre el año de rezago, pero disminuye a medida que éste rezago aumenta, y con ello la variabilidad en ese rezago, respecto a los años de reclamación, va aumentando. Esto se debe a que los siniestros con rezagos más largos son menos frecuentes y más predecibles que los siniestros con rezagos más cortos, lo que reduce la incertidumbre sobre el costo final de los mismos. Por el contrario, los siniestros con rezagos más cortos son más frecuentes y menos predecibles que los siniestros con rezagos más largos, lo que aumenta la incertidumbre sobre el costo final de los mismos.

Matriz de correlación: En este caso, se observa que los valores cercanos tanto en tiempo de reclamación como de rezago tienen una correlación alta cercana a 1, pero a medida que se alargan los años, este coeficiente disminuye tan rápido hasta llegar a valores negativos. Esto se debe a que los siniestros con tiempos de reclamación y rezagos similares tienen características comunes que afectan el costo final de los mismos, como la frecuencia, la severidad, la inflación, la normatividad, etc. Por el contrario, los siniestros con tiempos de reclamación y rezagos muy diferentes tienen características distintas que afectan el costo final de los mismos, lo que genera una divergencia o una inversión en la tendencia de los valores de la variable.

- **PRIMAS GANADAS:** Los valores de estas tres variables, (EarnedPremDIRF2, EarnedPremCededF2, EarnedPremNetF2), están relacionados bajo la ecuación 1. Los triángulos asociados muestran las primas que la aseguradora ha reconocido como ingresos en el año de incurrimiento de los siniestros, es decir, el año en que se producen los eventos que dan lugar a las reclamaciones. Estas primas se mantienen constantes en todas las filas, ya que no dependen del año de desarrollo de los siniestros, es decir, el año en que se notifican o se pagan las reclamaciones. Las primas ganadas solo cambian cuando se produce una modificación en el contrato de seguro o en la cartera de clientes, lo que puede generar valores negativos en algunos casos. Por lo tanto, los triángulos de primas ganadas reflejan la exposición al riesgo de la aseguradora en el año de incurrimiento, independientemente del tiempo que tarde en conocer el costo final de los siniestros.

Los valores nulos en estos triángulos pueden tener diferentes interpretaciones según el contexto. Por ejemplo:

Si los valores nulos corresponden a años anteriores a la constitución de la aseguradora, significa que no hay datos disponibles sobre las primas ganadas en esos años, ya que la aseguradora no tenía actividad en el mercado.

Si los valores nulos corresponden a años posteriores al cierre de la muestra, como si la aseguradora hubiera sido vendida o fusionada o hubiera cambiado su razón social o su régimen contable, significa que la aseguradora ha dejado de reportar las primas ganadas en esos años, ya que ha cesado su operación o ha modificado su estructura jurídica o contable.

En cualquier caso, los valores nulos en estos triángulos implican una ausencia de datos que puede dificultar el análisis de la relación entre las primas y las pérdidas. Por lo tanto, es importante identificar la causa de los valores nulos y tratar de obtener o estimar los datos faltantes si es posible. Sin embargo, dado que no se pueden recolectar esos datos de ninguna manera, la opción más viable es limpiar la muestra, eliminando algunos años de algunas aseguradoras o eliminando algunas aseguradoras por completo. Esto será fundamental para describir exactamente los riesgos que asume la aseguradora y no incluir valores que no son reales, que sesgarían la muestra hacia abajo (cerca al cero) y no permitirían percibir patrones y tendencias relevantes para el mejor desarrollo de un modelo predictivo de provisiones.

Se observa que la media aumenta para estas variables y esto se debe a efectos de inflación, devaluación, crecimiento del riesgo de las primas aseguradas, etc. Estos factores hacen que las primas ganadas sean mayores en los años más recientes que en los años más antiguos, lo que refleja un incremento en el nivel de ingresos de la aseguradora.

Para las variables EarnedPremDIRF2 y EarnedPremNetF2 la desviación se mantiene casi que constante durante los años de la muestra y la correlación entre todos los años es bastante alta, lo que indica que los valores de estas variables tienen una distribución homogénea y una relación lineal positiva fuerte. Esto significa que las primas directas y netas tienden a moverse en la misma dirección y proporción, lo que implica una estabilidad en el nivel de riesgo asumido por la aseguradora.

Sin embargo, para EarnedPremCededF2 la desviación tiende a aumentar al pasar los años y la correlación tiende a llegar a niveles más bajos para estimar una correlación entre variables, ya sea positiva o negativa. Esto indica que los valores de esta variable tienen una distribución heterogénea y una relación lineal débil o nula. Esto significa que las primas cedidas tienden a moverse en direcciones o proporciones diferentes, lo que implica una variabilidad en el nivel de riesgo transferido por la aseguradora.

- **PostedReserve97F2:** Las reservas publicadas son iguales en todas las filas y columnas porque corresponden al valor reportado por la aseguradora en el año 1997, que es el año base de la muestra. Este valor

no cambia según el año de desarrollo de los siniestros, es decir, el año en que se notifican o se pagan las reclamaciones. Las reservas publicadas solo cambian cuando se produce una revisión de las estimaciones de reservas, lo que puede generar valores negativos en algunos casos. Por lo tanto, los triángulos de reservas publicadas reflejan el nivel de pasivo de la aseguradora en el año 1997, independientemente del tiempo que tarde en conocer el costo final de los siniestros.

En esta muestra se tiene 5 aseguradoras con valores nulos en su matriz, donde 2 de ellas son las aseguradoras más actuales del mercado. Por lo tanto, los valores nulos en estos triángulos tienen diferentes interpretaciones según el contexto:

Si los valores nulos corresponden a aseguradoras nuevas, se debe a la demora en la constitución de la aseguradora, y por ende no hay datos disponibles sobre las reservas publicadas, ya que la aseguradora no tenía actividad en el mercado.

Si los valores nulos corresponden a aseguradoras muy antiguas, como si la aseguradora hubiera sido vendida o fusionada o hubiera cambiado su razón social o su régimen contable, significa que la aseguradora ha dejado de reportar las reservas publicadas en esos años, ya que ha cesado su operación o ha modificado su estructura jurídica o contable.

Esto puede afectar el análisis de la evolución de las reservas y su relación con los siniestros, ya que se desconoce el nivel de pasivo final de la aseguradora.

Luego de un exhaustivo análisis de las variables en el contexto de nuestro proyecto de triángulos de runoff de compañías aseguradoras, se ha llegado a las siguientes conclusiones:

Hay dos aseguradoras en particular, a saber, Overseas Partners Us Reins Co y Texas Medical Ins Co. Estas compañías presentan un alto número de valores nulos en las variables estudiadas debido a su reciente ingreso al mercado en el momento en que se tomó la muestra. Su falta de suficiente cobertura para reportar pérdidas se refleja en el valor nulo o extremadamente bajo de las primas netas durante los años considerados.

En contraste, algunas otras aseguradoras, como Controlled Risk Ins Co Of VT Inc y Community Blood Cntr Exch RRG, también son relativamente nuevas en el mercado. Sin embargo, estas compañías tienen una posición lo suficientemente establecida en el mercado como para reportar pérdidas (incluso negativas) durante los períodos de cobertura incluidos en la muestra.

En cuanto a las demás aseguradoras que presentan valores nulos en todas las variables. Estos valores reflejan información real en los períodos existentes de las compañías y pueden representar situaciones como la liquidación de la compañía, cambios en la razón social o períodos inexistentes al comienzo del período considerado en la muestra. Modificar estos ceros mediante imputación o eliminación podría distorsionar la integridad de los datos y llevar a conclusiones erróneas.

Se diseñará un modelo que incluya todas las aseguradoras disponibles, y luego se hará un análisis de sensibilidad o de escenarios para ver cómo cambian los resultados excluyendo algunas aseguradoras.

Recordando que podría haber alguna relación entre el tamaño, el tipo, la ubicación o el segmento de mercado de las aseguradoras y sus matrices de runoff, variables para las cuales, en primera instancia no se tiene información y que en realidad no se necesitan por ser un proyecto de predicción y no de inferencia o causalidad.

Estas acciones, permitirán un análisis más preciso de la relación entre las primas y las pérdidas en el contexto del proyecto. Además, estas decisiones mejorarán la calidad de nuestro análisis y la confiabilidad de los resultados obtenidos.

Relativo a las variables a considerar, se tiene:

Se usará IncurLoss como la variable principal para el cálculo de las reservas, ya que representa las pérdidas totales que la aseguradora ha experimentado hasta la fecha.

Como, CumPaidLoss representa las pérdidas que ya se han pagado a los asegurados, se usará la diferencia entre IncurLoss y CumPaidLoss como una medida de las reservas pendientes de pago, y se comparará con la variable BulkLoss. Esta comparación ayuda a verificar si hay alguna diferencia significativa o alguna tendencia

en el tiempo. Es decir, identificar posibles subestimaciones o sobreestimaciones de las reservas.

Se dejará solo la variable `EarnedPremNet`, ya que representa las primas ganadas netas de la cesión al reaseguro, y que son la base para determinar la relación entre las primas y las pérdidas. Se eliminará `EarnedPremDIR` y `EarnedPremCeded`, ya que son variables redundantes que se pueden deducir a partir de `EarnedPremNet`. Además, se calculará el ratio de siniestralidad, que es la relación entre `IncurLoss` y `EarnedPremNet`, ya que es un indicador clave de la rentabilidad de la aseguradora.

se usará la variable `BulkLoss` como una variable auxiliar para el cálculo de las reservas, ya que representa las reservas establecidas para cubrir pérdidas futuras que se espera que surjan de eventos pasados pero que aún no se han notificado.

La variable `PostedReserve97` se usará como una variable de referencia para el cálculo de las reservas, ya que representa las reservas publicadas en el año 1997, que es el año base de la muestra. Se puede comparar esta variable con la diferencia entre `IncurLoss` y `CumPaidLoss`, para ver si hay alguna diferencia significativa o alguna tendencia en el tiempo. Esto podría ayudar a identificar posibles revisiones al alza o a la baja de las reservas.

Eventualmente se sacarán del modelo de predicción las variables `EarnedPremDIRF2`, `EarnedPremCededF2` por ser variables redundantes, `Single` por ser una variable no contemplada debido a que ya se hizo una previa clasificación con esta variable y no se notó relativa diferencia en las distribuciones de estos conjuntos de datos, y `PostedReserve97F2` por ser una variable constante en todas las entradas de la matriz, por ser la reserva publicada de 1997.

Por otro lado, se generarán nuevas variables derivadas a partir de las existentes, dadas por el ratio de siniestralidad ( $\text{IncurLossF2} / \text{EarnedPremNetF2}$ ) y la diferencia entre `IncurLoss` y `CumPaidLoss`.

Ratio de siniestralidad: Note que esta variable está bien definida solo si  $\text{EarnedPremNetF2} \neq 0$ . Si  $\text{EarnedPremNetF2} = 0$  hay dos posibles casos:

1.  $\text{IncurLossF2} = 0$ , lo cual sería natural, un año o una época en la que no hay reclamos y tampoco hay primas es porque no hay cobertura de algún tipo y por ende no hay riesgo, este es el caso trivial. Para este caso el ratio de siniestralidad se define como 0.

2.  $\text{IncurLossF2} \neq 0$ , esto se puede dar por varios factores:

La póliza de seguro es nueva y aún no se ha cobrado ninguna prima, pero ya se ha producido alguna reclamación.

La póliza de seguro ha sido cancelada o rescindida, y se ha devuelto la prima al asegurado, pero aún hay reclamaciones pendientes de pago.

La póliza de seguro tiene una cláusula de participación en beneficios o pérdidas, y la aseguradora ha devuelto parte o toda la prima al asegurado en función del resultado del negocio.

La póliza de seguro tiene una estructura de pago diferido o variable, y la prima se ajusta en función del rendimiento del negocio o de algún índice externo.

En estos casos, el ratio de siniestralidad será indefinido. Por lo tanto, la medida de la media no será descriptiva en esta variable, pero se podrían considerar la mediana o los cuartiles, que proporcionarán información relevante sobre la relación.

Es fundamental destacar la variable `IncurLossPremNetratio`, que muestra que en el percentil 75%, su valor asciende a un 87.57%. Este porcentaje representa la proporción de la prima neta destinada a cubrir pérdidas. Por lo tanto, en ciertos casos, resulta crucial llevar a cabo una evaluación más detallada de los riesgos con el fin de evitar la presencia de un porcentaje tan elevado en este percentil. Un indicador más favorable se encuentra en el percentil 50%, donde el 50% de la prima se destina al pago de reclamaciones. Este enfoque tiende a ser menos impactante y más sostenible en términos de gestión de riesgos.

En el percentil 25%, se presenta un caso completamente favorable en el que el 0% de la prima se destina al pago de reclamaciones. No obstante, también existen casos extremos, como el valor infinito mencionado anteriormente y valores negativos. Los valores negativos pueden originarse por dos razones: una prima

negativa o una reclamación negativa. En el primer caso, si la prima es negativa y la reclamación es positiva, se generan pérdidas en ambas áreas. En el segundo caso, un valor negativo de reclamación representa un ingreso para la compañía de seguros, al igual que una prima positiva. Este último escenario resulta altamente favorable para la compañía.

En el caso de la variable `diffIncurPaidLoss`, se observa que en su mayoría los valores son positivos. Sin embargo, es importante notar que los valores por debajo del percentil 25% son negativos. Esto implica que `CumPaidLossF2` es mayor que `IncurLossF2` en estos casos, es decir, las pérdidas pagadas superan la suma de las pérdidas pagadas y las pendientes de pago. En otras palabras, en estos casos, las pendientes de pago son negativas, lo que resulta en un ingreso en lugar de un egreso.

#### Data Modeling - Chain Ladder

El Chain Ladder es un método determinístico que se utiliza para estimar las reservas de siniestros en compañías de seguros.

El método se llama “Chain Ladder” porque utiliza una cadena de factores de desarrollo para proyectar las reservas futuras. El método es simple y fácil de entender, lo que lo hace popular entre las compañías de seguros.

El modelo utiliza una matriz acumulada triangular superior izquierda para estimar los siniestros. Cada celda de la diagonal y superior a la diagonal principal de la matriz representa el valor de la variable a estimar que ocurrió en un periodo de tiempo pasado y relacionado a una fecha de siniestro determinados, es decir, son valores conocidos. Las celdas por debajo de la diagonal principal representan los valores a estimar con el modelo.

El modelo se basa en la suposición de que la tasa de siniestralidad (número de siniestros por unidad de exposición) es constante a lo largo del tiempo. Esta suposición permite al modelo utilizar los datos históricos para predecir los siniestros futuros. Sin embargo, esta es una limitación, ya que no siempre se cumple en la realidad y puede haber cambios en las tendencias del mercado que no se reflejen en los datos históricos.

La fórmula para el Chain Ladder determinístico se puede expresar de la siguiente manera:

$$\hat{C}_{i,j} = C_{i,j-1} \frac{\sum_{i=1}^{10-j} C_{i,j}}{\sum_{i=1}^{10-j} C_{i,j-1}}$$

Donde:

$\hat{C}_{i,j}$  es la estimación de las reservas para el año de siniestro  $i$  y el desarrollo  $j$ .

$C_{i,j}$  es la cantidad de siniestros reportados para el año de siniestro  $i$  y el desarrollo  $j$ .

En el resultado anterior, se implementó la métrica MAPE para evaluar la precisión de un modelo. MAPE significa error porcentual absoluto medio y se calcula utilizando la siguiente fórmula matemática:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} 100$$

Es importante tener en cuenta que esta fórmula solo se aplica cuando el valor real es diferente de cero. En caso contrario, se toma como resultado porcentual el mismo valor del estimado pero en términos porcentuales.

En conclusión, tomando la medida MAPE sobre todas las estimaciones sobre cada variable, se observa que en la mayoría se presentan valores de la medida entre 20% y 30%, excepto la variable `bulkloss` que presenta una diferencia o MAPE del 60% aproximadamente entre la estimación y el valor real. Estos porcentajes de diferencia bastante elevados reflejan un desempeño medio del funcionamiento del Chain Ladder determinístico en estas variables y desempeño bajo en la variable `bulkloss`.

En adelante se implementarán otros modelos con el objetivo de compararlos y elegir el que mejor aproxime los valores reales.

Separar los conjuntos para las etapas de entrenamiento y pruebas Se tomará la aseguradora `SScpie Indemnity Co` para la etapa de pruebas y para la etapa de entrenamiento se iterará sobre las aseguradoras restantes realizando un cross validation sobre el modelo.

## Regresión Lineal

La regresión lineal Ridge y Lasso son técnicas de regularización que se utilizan para evitar el sobreajuste en modelos de regresión lineal. La regresión Ridge agrega una penalización L2 a la función de costo, mientras que la regresión Lasso agrega una penalización L1. La penalización reduce la magnitud de los coeficientes del modelo, lo que a su vez reduce la complejidad del modelo y evita el sobreajuste.

La fórmula para la regresión lineal Ridge es:

$$\hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T y$$

Donde:

-  $\hat{\beta}^{ridge}$  es el vector de coeficientes estimados. -  $X$  es la matriz de características. -  $y$  es el vector objetivo. -  $\lambda$  es el hiperparámetro que controla la fuerza de la penalización. -  $I$  es la matriz identidad.

La fórmula para la regresión lineal Lasso es:

$$\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|^2 + \alpha \|\beta\|_1 \right\}$$

Donde:

-  $\hat{\beta}^{lasso}$  es el vector de coeficientes estimados.  
-  $X$  es la matriz de características.  
-  $y$  es el vector objetivo.  
-  $\alpha$  es el hiperparámetro que controla la fuerza de la penalización.

Ahora, tomando el logaritmo de los triángulos acumulados, únicamente donde el logaritmo tiene sentido, es decir, para los valores mayores a cero, y además definiendo los valores para el vector  $Y$  y la matriz  $X$  para ejecutar la regresión.

## 5 Resultados:

Al evaluar la aseguradora del conjunto de pruebas, para la variable IncurLoss, se obtiene:

- Chain Ladder determinístico: MAPE = 31.39%.
- Regresión Ridge: MAPE = 25.2%.
- Regresión Lasso: MAPE = 45%.

El modelo de Regresión Ridge presenta un MAPE significativamente más bajo que Lasso, lo que indica una mayor precisión en las predicciones y por lo tanto una base más confiable para la toma de decisiones.

En el contexto de los triángulos de run-off, cada predictor es relevante para comprender la dinámica de las reclamaciones a lo largo del tiempo. Lasso, al poder eliminar predictores, podría descartar información crucial si los años de delay y desarrollo están correlacionados con otros predictores. Ridge, sin embargo, conserva toda la información al reducir proporcionalmente la importancia de los predictores sin eliminarlos, lo cual es crucial en este contexto.

Aunque Lasso puede simplificar la interpretación del modelo al realizar una selección de características, esto podría no ser deseable en el análisis actuarial. La regresión Ridge, al penalizar todos los coeficientes pero sin llevarlos a cero, mantiene la complejidad necesaria para un modelo que debe considerar la totalidad de los datos históricos.

Ridge tiende a ser más estable en presencia de multicolinealidad, una situación común con los datos de triángulos de run-off, donde los distintos años pueden estar altamente interrelacionados. Esto significa que Ridge es probablemente más robusto y consistente a través de diferentes conjuntos de datos o subconjuntos de datos.

La elección de Ridge sugiere un mejor equilibrio del compromiso de sesgo-varianza, especialmente importante en la modelización actuarial donde los pronósticos precisos son esenciales.

## 6 Bibliografía:

- <https://www.msdm anuals.com/es-mx/hogar/fundamentos/asuntos-legales-y-%C3%A9ticos/negligencia-m%C3%A9dica>
- <https://archivos.juridicas.unam.mx/www/bjv/libros/10/4734/13.pdf>
- <https://www.iberley.es/temas/reclamacion-sobre-negligencias-medicas-jurisdccion-civil-65312>
- <https://www.conceptosjuridicos.com/cl/negligencia-medica/>
- <https://carossoabogados.com/2023/06/30/presentar-reclamacion-negligencia-medica/>