

# Provisiones en Seguros

2023-2

Universidad Nacional  
de Colombia

Facultad de  
matemáticas

Aura Alaguna  
(aalaguna@unal.edu.co)

Este proyecto se centra en desarrollar un enfoque integral y preciso para la estimación de provisiones. Las provisiones son reservas financieras esenciales que una compañía de seguros debe establecer para garantizar que tenga los fondos adecuados disponibles para cubrir reclamaciones futuras de los asegurados. Nuestra meta principal es crear un marco de trabajo que combine técnicas actuales de modelado actuarial, análisis de datos y estadísticas para estimar con precisión las provisiones y gestionar los riesgos asociados. Este proyecto aborda los desafíos inherentes a la variabilidad en los tiempos de demora en la notificación y liquidación de reclamaciones, así como a la incertidumbre en la naturaleza y cantidad de futuras reclamaciones.

# Índice general

<b>1. Entendimiento empresarial</b>	<b>2</b>
1.1. Objetivos empresariales . . . . .	2
1.2. Estructura organizacional . . . . .	3
1.3. Evaluación de la Situación . . . . .	4
1.3.1. Inventario de recursos . . . . .	5
1.3.2. Requisitos, supuestos y limitaciones . . . . .	7
1.3.3. Riesgos y contingencia . . . . .	8
1.3.4. Análisis costos y beneficios . . . . .	10
1.4. Determinar los objetivos de la minería de datos . . . . .	10
1.4.1. Plan proyecto . . . . .	13
<b>2. Entendimiento de los datos</b>	<b>14</b>
2.1. Recolectando datos . . . . .	14
2.2. Descripción de los datos . . . . .	15
2.2.1. Exploración de los datos . . . . .	16

# Capítulo 1

## Entendimiento empresarial

### 1.1. Objetivos empresariales

Basado en la información proporcionada sobre el desarrollo de un siniestro en un seguro de P&C y los eventos asociados, se puede determinar el siguiente objetivo general el cual se basa en asegurar la estabilidad financiera y la sostenibilidad de la compañía de seguros.

En seguida podemos ver algunos objetivos empresariales:

1. **Gestión de Reservas y Capital:** El objetivo fundamental es garantizar que la compañía de seguros tenga suficiente capital reservado para cubrir las reclamaciones futuras. Dado que las reclamaciones pueden tener retrasos en la notificación y en la liquidación, es crucial establecer reservas adecuadas para asegurar que la compañía esté financieramente preparada para hacer frente a los pagos futuros. Esto implica calcular y mantener reservas precisas y suficientes para asegurar que la empresa cumpla con sus obligaciones hacia los asegurados.
2. **Predicción de Reclamaciones y Costos:** Un objetivo importante es desarrollar modelos y herramientas analíticas que permitan predecir con precisión las reclamaciones futuras y los costos asociados. Esto implica utilizar métodos estadísticos y matemáticos para estimar la probabilidad y el monto de las reclamaciones pendientes, teniendo en cuenta los diferentes tiempos de demora y liquidación. Estas predicciones ayudan a la compañía a planificar sus reservas y recursos financieros de manera eficiente.
3. **Optimización de la Política de Reservas:** La compañía debe determinar la política adecuada para establecer reservas en función de factores como el tipo de reclamación, el historial de reclamaciones y el riesgo asociado. El objetivo aquí es encontrar el equilibrio entre mantener suficientes reservas para la solvencia financiera y evitar la sobreestimación que podría afectar negativamente la rentabilidad.
4. **Gestión de Riesgos y Solvencia:** La gestión adecuada de riesgos es crucial. La compañía debe asegurarse de que las reservas sean suficientes para enfrentar escenarios adversos, como aumentos inesperados en las reclamaciones. El objetivo es mantener la solvencia de la compañía y garantizar que pueda cumplir con sus obligaciones incluso en situaciones de estrés financiero.

5. **Eficiencia Operativa:** La comprensión de los tiempos de demora en la notificación y liquidación de reclamaciones permite a la compañía optimizar sus procesos operativos. Esto puede involucrar la identificación de áreas donde se puede reducir la demora y mejorar la eficiencia en la liquidación de reclamaciones.
6. **Desarrollo de Productos y Precios:** Comprender los patrones de desarrollo de siniestros y los tiempos de demora puede influir en el desarrollo de nuevos productos y en la fijación de precios adecuados. Los datos sobre retrasos y liquidaciones pueden ser utilizados para establecer tarifas que reflejen con precisión el riesgo y el perfil de la cartera de seguros.

En conjunto, los objetivos empresariales giran en torno a la gestión financiera sólida, la predicción precisa de riesgos y costos, la optimización de las reservas y la eficiencia operativa en el contexto de los seguros de propiedad y accidentes. Estos objetivos trabajan en conjunto para asegurar que la compañía de seguros pueda mantener su estabilidad financiera, cumplir con sus obligaciones y operar de manera rentable en el mercado.

## 1.2. Estructura organizacional

Dentro del contexto del problema de estimación de provisiones en seguros de propiedad y accidentes, es importante determinar la estructura organizativa para asegurarse de que todas las partes relevantes estén involucradas y se puedan tomar decisiones informadas. Aquí está la estructura organizativa sugerida:

1. **Dirección Ejecutiva y Alta Gerencia:** - Director Ejecutivo (CEO): Responsable de la dirección estratégica de la empresa y de tomar decisiones clave. - Vicepresidentes y Directores Generales: Supervisan áreas específicas de la compañía, como finanzas, operaciones y seguros.
2. **Área de Seguros y Actuarial:**
  - Director de Seguros: Encargado de la gestión global de las operaciones de seguros de propiedad y accidentes.
  - Actuarios Principales: Lideran el equipo de actuarios y son responsables de la estimación de provisiones y la gestión de riesgos.
  - Analistas Actuariales: Trabajan en la recopilación y análisis de datos, modelado y cálculo de provisiones.
3. **Comité de Riesgos y Finanzas:**
  - Comité formado por miembros de la alta gerencia, actuarios y representantes financieros.
  - Supervisa y toma decisiones sobre cuestiones relacionadas con la estimación de provisiones, solvencia y riesgos.
4. **Unidades de Negocio Afectadas:**

- Gerentes de Área: Responsables de áreas específicas de la empresa que serán impactadas por las estimaciones de provisiones.
- Ejecutivos de Ventas: Informan sobre el volumen de ventas y nuevas pólizas, lo que afecta la generación de datos para las estimaciones.

#### 5. Tecnología de la Información:

- Director de TI: Encargado de la infraestructura tecnológica necesaria para la recopilación y el análisis de datos.
- Equipo de Desarrollo: Implementa herramientas y sistemas para gestionar y procesar datos relevantes.

#### 6. Departamento legal:

- Abogados especializados en regulaciones de seguros y compliance.
- Asesoran sobre las implicaciones legales y regulatorias de las estimaciones de provisiones.

#### 7. Comunicaciones y Marketing Interno:

- Responsables de comunicar los objetivos y beneficios del proyecto de estimación de provisiones a todo el personal.
- Garantizan la colaboración y comprensión en toda la organización.

#### 8. Recursos Humanos:

- Equipo de Recursos Humanos: Apoya en la gestión de personal, formación y contratación de expertos actuariales y analistas.

#### 9. Auditoría Interna:

- Asegura la integridad y precisión de los datos y los procesos de estimación de provisiones.

La estructura organizativa deberá ser adaptable y flexible para permitir la colaboración y el flujo de información entre todas estas áreas. Un enfoque interdisciplinario y una comunicación efectiva serán esenciales para abordar de manera exitosa el desafío de las provisiones en seguros de propiedad y accidentes.

## 1.3. Evaluación de la Situación

Un Escenario de Minería de Datos para Estimación de Provisiones Utilizando CRISP-DM

La compañía de seguros se adentra por primera vez en la minería de datos para estimación de provisiones y ha decidido consultar a un especialista en minería de datos para recibir orientación. Una de las primeras tareas que enfrenta el consultor es evaluar los recursos disponibles para llevar a cabo la estimación de provisiones.

**Personal:** La empresa cuenta con expertos internos en el área de seguros y actuaría que manejan las reclamaciones y datos históricos, pero se reconoce que se necesita más experiencia en modelado estadístico y análisis actuarial. Por lo tanto, también se podría considerar la consulta con analistas de datos y actuarios especializados. Dado que se espera que los resultados impacten la toma de decisiones continua, la administración debe evaluar si los roles creados para este proyecto serán permanentes.

**Datos:** Dado que la compañía tiene un historial extenso, hay una gran cantidad de datos históricos de reclamaciones y siniestros. Para este proyecto inicial, la atención se centrará en los seguros. Si los resultados son exitosos, se podría considerar expandir el análisis a otros tipos de seguros.

**Riesgos:** Además de los costos financieros asociados con la consultoría y el tiempo invertido por el personal, no se observa un riesgo inmediato significativo. Sin embargo, dada la importancia del tiempo, el proyecto se ha programado para un período de evaluación inicial de un trimestre financiero. Además, dada la situación financiera actual, es esencial que el proyecto se ajuste al presupuesto asignado. En caso de que cualquiera de estos objetivos esté en riesgo, la administración de la empresa sugiere la posibilidad de reducir el alcance del proyecto.

### 1.3.1. Inventario de recursos

#### Buscar recursos de hardware:

Para respaldar la estimación de provisiones en seguros de propiedad y accidentes, se requiere un hardware adecuado que sea capaz de manejar el procesamiento de datos complejos y el cálculo de modelos actuariales. Esto podría incluir:

- **Servidores Potentes:** Para procesar grandes volúmenes de datos históricos de reclamaciones y siniestros de manera eficiente.
- **Almacenamiento de Datos:** Capacidad suficiente para guardar y acceder a los datos históricos y modelos generados.
- **Procesadores de Alto Rendimiento:** Para realizar cálculos actuariales y análisis estadísticos complejos en tiempo razonable.
- **Memoria RAM Suficiente:** Para cargar y manipular grandes conjuntos de datos de manera efectiva.
- **Hardware Gráfico (si es necesario):** Para visualización de datos y modelos.
- **Conexión a Redes y Bases de Datos:** Para acceder a las fuentes de datos y sistemas necesarios.
- **Seguridad y Respaldo:** Implementación de medidas de seguridad para proteger los datos confidenciales y asegurar respaldos regulares.

Es fundamental contar con hardware que cumpla con los requisitos de procesamiento y almacenamiento, ya que la estimación de provisiones involucra el análisis de datos históricos y la ejecución de modelos complejos para predecir futuros eventos en el ámbito de seguros de propiedad y accidentes.

## Identificar las fuentes de datos y los almacenes de conocimientos

### Fuentes de Datos Disponibles

Para la estimación de provisiones en seguros de propiedad y accidentes, es crucial conocer las fuentes de datos disponibles. Esto incluye:

- **Registros de Reclamaciones y Siniestros:** Datos históricos de reclamaciones y siniestros, que proporcionan información sobre tipos de eventos, montos de reclamaciones, tiempos de demora, entre otros.
- **Información de Pólizas y Clientes:** Datos sobre los clientes asegurados, detalles de pólizas y coberturas contratadas.
- **Datos Demográficos y Socioeconómicos:** Información externa como datos demográficos, estadísticas económicas y sociodemográficas que podrían impactar las reclamaciones y siniestros.
- **Estadísticas del Sector:** Datos y métricas relacionadas con la industria de seguros en general, que podrían aportar contexto.
- **Datos de Regulación:** Información sobre regulaciones y requisitos legales que podrían influir en las provisiones.

### Almacenamiento de Datos

Los datos se almacenan de diversas formas, como:

- **Bases de Datos Internas:** Bases de datos específicas de la empresa que almacenan información sobre pólizas, reclamaciones y siniestros.
- **Almacenes de Datos:** Repositorios centralizados de datos históricos para análisis, que podrían ser utilizados en tiempo real o en lotes.
- **Fuentes Externas:** Posiblemente, se acceda a fuentes externas para adquirir datos demográficos y estadísticas sectoriales.

Es importante determinar si tienes acceso en tiempo real a almacenes de datos o bases de datos operativas, ya que esto afectará la velocidad y frecuencia de actualización de los análisis.

### Adquisición de Datos Externos

Se debe considerar si se planea adquirir datos externos, como información demográfica o estadísticas de reclamaciones, para enriquecer el análisis de provisiones. Estos datos pueden ser valiosos para comprender mejor los factores que influyen en las reclamaciones y mejorar la precisión de las estimaciones.

## **Identificar al personal con recursos**

### **Acceso a Expertos en Negocios y Datos**

Es esencial contar con acceso a expertos en negocios y datos para llevar a cabo con éxito la estimación de provisiones en seguros de propiedad y accidentes. Estos expertos pueden aportar conocimientos sobre la industria aseguradora, los productos de seguros y los detalles específicos de las reclamaciones y siniestros. Su experiencia contribuirá a la interpretación precisa de los datos y la formulación de modelos actuariales efectivos.

### **Identificación de Administradores de Bases de Datos y Personal de Soporte**

Es importante identificar a los administradores de bases de datos y otro personal de soporte que puedan ser necesarios para respaldar el proceso de estimación de provisiones. Los administradores de bases de datos desempeñan un papel crucial en la gestión y organización de los datos requeridos para el análisis. Además, se debe considerar la posibilidad de contar con personal de soporte técnico y analistas de datos para asegurar un flujo de trabajo fluido y una interpretación precisa de los resultados.

## **1.3.2. Requisitos, supuestos y limitaciones**

### **Requisitos**

#### **Existencia de Restricciones de Seguridad o Legales**

Es fundamental determinar si existen restricciones de seguridad o legales que puedan afectar los datos o resultados del proyecto de estimación de provisiones en seguros de propiedad y accidentes. Estas restricciones podrían incluir regulaciones de privacidad de datos, acuerdos de confidencialidad o requisitos de cumplimiento. Es esencial tener en cuenta estas restricciones para garantizar que el manejo de los datos y los resultados cumpla con todas las normativas aplicables.

#### **Alineación con Requisitos de Programación del Proyecto**

Es importante confirmar si todos los involucrados en el proyecto están alineados con los requisitos de programación. Esto abarca desde el equipo de desarrollo y análisis hasta los patrocinadores y partes interesadas. Asegurarse de que todos comprendan y estén comprometidos con los plazos y hitos del proyecto es esencial para una ejecución exitosa.

#### **Existencia de Requisitos sobre la Implementación de Resultados**

Es relevante determinar si existen requisitos específicos sobre cómo se implementarán y presentarán los resultados del proyecto. Por ejemplo, si se planea publicar los resultados en la web o almacenarlos en una base de datos, es importante identificar estos requisitos desde el principio. Esto ayudará a planificar la fase de implementación y asegurará que los resultados sean entregados de manera efectiva según las necesidades de las partes interesadas.

### **Supuestos**

#### **Existencia de Factores Económicos que Puedan Afectar al Proyecto**

Es relevante evaluar si existen factores económicos que puedan influir en el proyecto de estimación de provisiones en seguros de propiedad y accidentes. Esto podría incluir costos asociados a consultoría, adquisición de datos externos, o la competencia



de productos similares en el mercado. Considerar estos factores permitirá anticipar posibles impactos financieros en el proyecto.

#### **Supuestos sobre la Calidad de los Datos**

Es esencial identificar si hay supuestos sobre la calidad de los datos utilizados en el proyecto. Estos supuestos podrían abarcar la integridad, precisión y consistencia de los datos históricos de reclamaciones y siniestros. Reconocer y aclarar estos supuestos ayudará a evitar posibles sesgos o imprecisiones en los resultados del análisis.

#### **Expectativas del Patrocinador/Equipo Directivo sobre la Visualización de Resultados**

Es importante comprender cómo el patrocinador del proyecto y el equipo directivo esperan ver los resultados. ¿Desean tener un entendimiento detallado del modelo actuarial y estadístico utilizado en el proceso? ¿O simplemente desean ver los resultados finales de las estimaciones de provisiones? Aclarar estas expectativas será crucial para definir la presentación y nivel de detalle de los resultados entregados.

### **Limitaciones**

#### **Acceso a Contraseñas Necesarias**

Es esencial asegurarse de tener todas las contraseñas necesarias para acceder a los datos requeridos para la estimación de provisiones en seguros de propiedad y accidentes. Esto incluye acceso a bases de datos, sistemas de almacenamiento y cualquier recurso necesario para realizar análisis y modelado. La falta de acceso podría retrasar o impedir el progreso del proyecto.

#### **Verificación de Restricciones Legales sobre el Uso de los Datos**

Es crucial verificar y cumplir todas las restricciones legales que afecten el uso de los datos en la estimación de provisiones. Esto podría incluir regulaciones de privacidad, acuerdos de uso de datos y consideraciones legales que dictan cómo se pueden utilizar los datos y los resultados generados. Cumplir con estas restricciones es esencial para evitar problemas legales y asegurar la integridad del proyecto.

#### **Consideración de Restricciones Financieras en el Presupuesto del Proyecto**

Es importante garantizar que todas las restricciones financieras relevantes estén contempladas en el presupuesto del proyecto de estimación de provisiones. Esto podría incluir los costos asociados con la adquisición de datos externos, honorarios de consultoría, licencias de software y otros gastos relacionados. Asegurarse de que el presupuesto sea realista y completo es fundamental para evitar sorpresas financieras a lo largo del proyecto.

### **1.3.3. Riesgos y contingencia**

- **Programación (¿Y si el proyecto dura más de lo previsto?):** Riesgo: Retraso en la programación podría afectar los plazos y la entrega del proyecto. Plan de Contingencia: Mantener un seguimiento constante de los hitos y el progreso del proyecto. Si se identifican retrasos, evaluar la posibilidad de asignar más recursos o redefinir prioridades para acelerar el avance.
- **Financieros (¿Y si el patrocinador del proyecto tiene problemas presupuestarios?):** Riesgo: Limitaciones financieras podrían afectar la ejecución completa

del proyecto. Plan de Contingencia: Establecer una reserva financiera en el presupuesto para manejar imprevistos. En caso de limitaciones presupuestarias, priorizar tareas críticas y explorar oportunidades de financiamiento adicional.

- **Datos (¿Y si los datos son de mala calidad o cobertura?):** Riesgo: Datos insuficientes o de baja calidad podrían afectar la precisión de las estimaciones. Plan de Contingencia: Realizar un análisis exhaustivo de calidad de datos al inicio del proyecto. Si se detectan problemas, considerar la posibilidad de mejorar la calidad de los datos o ajustar los modelos para trabajar con limitaciones.
- **Resultados (¿Y si los resultados iniciales son menos espectaculares de lo esperado?):** Riesgo: Resultados iniciales por debajo de las expectativas podrían desalentar a los interesados. Plan de Contingencia: Mantener una comunicación transparente con los interesados. Enfocarse en resaltar los aspectos positivos y explicar cualquier diferencia en las expectativas iniciales. Identificar áreas de mejora y trabajar en ajustes.

### Documentación de Posibles Riesgos

Es esencial identificar y documentar cada posible riesgo asociado al proyecto de estimación de provisiones en seguros de propiedad y accidentes. Algunos ejemplos de posibles riesgos podrían incluir:

- Cambios en las regulaciones de seguros que afecten los cálculos de provisiones.
- Problemas de calidad de datos que afecten la precisión de las estimaciones.
- Falta de alineación entre los equipos de negocios y técnicos.
- Limitaciones tecnológicas que dificulten el procesamiento de grandes conjuntos de datos.
- Retrasos en la adquisición de datos externos.

### Plan de Contingencia para Cada Riesgo

Para cada riesgo identificado, es necesario elaborar un plan de contingencia que describa cómo se abordará el riesgo si llega a materializarse. Por ejemplo:

- Para el riesgo de cambios en las regulaciones, se establecerá un equipo de monitoreo que se mantenga al tanto de los cambios regulatorios y adapte los cálculos de provisiones en consecuencia.
- En caso de problemas de calidad de datos, se implementarán procedimientos de limpieza y validación de datos antes de realizar cualquier estimación.
- Para mitigar la falta de alineación entre equipos, se establecerán reuniones regulares de coordinación y se definirán claramente los roles y responsabilidades.
- Frente a limitaciones tecnológicas, se considerará la posibilidad de mejorar la infraestructura tecnológica o adoptar soluciones alternativas.
- En caso de retrasos en la adquisición de datos externos, se planificará un calendario flexible y se evaluarán fuentes alternativas.

El desarrollo de planes de contingencia contribuirá a mitigar los riesgos y a mantener la ejecución exitosa del proyecto.

### 1.3.4. Análisis costos y beneficios

Se debe considerar los costes estimados de:

- **Recogida de Datos y Datos Externos:** Calcular los costes asociados con la recopilación de datos internos y, si aplicable, la adquisición de datos externos como información demográfica o estadísticas sectoriales.
- **Despliegue de Resultados:** Evaluar los costes relacionados con la implementación de los resultados del proyecto. Esto puede incluir la creación de informes, visualizaciones y cualquier otra forma de presentación de los resultados.
- **Costes Operativos:** Considerar los costes operativos asociados con la ejecución continua del modelo y la actualización de los datos. Esto podría incluir costes de mantenimiento de hardware y software, así como costes de personal.

Es esencial tener en cuenta tanto los costes directos como los costes indirectos que puedan surgir a lo largo del ciclo de vida del proyecto.

Este es el texto traducido a LaTeX para responder a las preguntas sobre el análisis de costes y beneficios en el contexto de la estimación de provisiones en seguros de propiedad y accidentes.

#### Beneficios del Proyecto de Estimación de Provisiones

- **Cumplimiento del Objetivo Principal:** Lograr el objetivo principal del proyecto de estimación de provisiones en seguros de propiedad y accidentes, que es mejorar la precisión en la estimación de reservas. Esto puede conducir a una gestión más efectiva de riesgos y una toma de decisiones informada.
- **Conocimientos Adicionales Generados por la Exploración de Datos:** La exploración de datos durante el proyecto puede generar conocimientos valiosos sobre patrones y tendencias en las reclamaciones y siniestros. Estos conocimientos pueden ser utilizados para identificar oportunidades de mejora y optimización en la gestión de seguros.
- **Posibles Beneficios de una Mejor Comprensión de los Datos:** Al mejorar la comprensión de los datos relacionados con reclamaciones y siniestros, la compañía de seguros puede tomar decisiones más informadas. Esto incluye identificar áreas de riesgo potencial, ajustar estrategias de precios y optimizar la asignación de recursos.

Los beneficios mencionados pueden contribuir a la solidez financiera de la compañía, la satisfacción del cliente y la capacidad de adaptarse eficazmente a las dinámicas del mercado.

## 1.4. Determinar los objetivos de la minería de datos

#### Objetivos de Minería de Datos para el Problema de Provisiones

Los objetivos de minería de datos para el problema de provisiones en seguros de propiedad y accidentes son los siguientes:

1. **Identificación de Clientes de Alto Valor:** Utilizando la base de datos de compras recientes, el objetivo es identificar a los clientes de alto valor. Esto implica analizar los patrones de compra y comportamiento de los clientes para determinar quiénes generan ingresos significativos para la compañía.
2. **Construcción de un Modelo de Probabilidad de Rotación:** Mediante los datos disponibles de los clientes, se busca construir un modelo que prediga la probabilidad de que cada cliente abandone la compañía (rotación). Este modelo permitirá anticipar y abordar la pérdida de clientes antes de que ocurra.
3. **Asignación de Rangos a los Clientes:** Se busca asignar a cada cliente un rango basado en dos factores clave: su propensión a la pérdida (probabilidad de rotación) y el valor del cliente. Esta asignación de rangos permitirá identificar a los clientes más valiosos en términos de retención y ganancias.

## Objetivos de la minería de datos

### Descripción del Tipo de Problema de Minería de Datos

El problema de provisiones involucra principalmente la predicción. Se busca anticipar futuros eventos, como el monto de las reclamaciones, con el fin de estimar las provisiones financieras requeridas. Además, también puede involucrar elementos de análisis de riesgos y optimización de carteras.

### Objetivos Técnicos con Unidades de Tiempo Específicas

Los objetivos técnicos para el proyecto de estimación de provisiones se plantean con una validez de tres meses. Estos objetivos incluyen:

1. Construir un modelo de predicción de reclamaciones que sea válido para los próximos tres meses.
2. Desarrollar un enfoque estocástico para determinar la reserva pendiente con modelos lineales generalizados (GLMs) y obtener una distribución predictiva de la reserva pendiente en tres meses a través de la simulación.

### Resultados Deseados con Números Reales

Se espera lograr resultados concretos en términos de retención y gestión de riesgos. Ejemplos de resultados deseados con números reales incluyen:

1. Producir puntuaciones de churn para el 80 % de los clientes existentes, lo que permitirá identificar a los clientes con mayor probabilidad de abandonar la compañía en los próximos tres meses.
2. Alcanzar una reducción del 15 % en la volatilidad de las estimaciones de provisiones, mejorando así la estabilidad financiera de la compañía.

Estos resultados cuantificables reflejarán el éxito en la aplicación de técnicas de minería de datos para el problema de provisiones en seguros.

## Criterios de éxito de la minería de datos

### Métodos de Evaluación del Modelo

Para evaluar la efectividad del modelo de estimación de provisiones en seguros de propiedad y accidentes, se utilizarán métodos de evaluación como:

- **Precisión y Rendimiento:** Se medirá la precisión del modelo al comparar las predicciones con los resultados reales de las reservas y reclamaciones. También se evaluará el rendimiento general del modelo en términos de tiempo de respuesta y capacidad de manejar grandes conjuntos de datos.
- **Validación Cruzada:** Se realizará validación cruzada utilizando diferentes conjuntos de datos para asegurar que el modelo sea robusto y generalice bien a datos no vistos.
- **Métricas de Error:** Se utilizarán métricas como el error cuadrático medio (MSE) o el error absoluto medio (MAE) para medir la diferencia entre las predicciones y los valores reales.

### **Definición de Puntos de Referencia**

Para evaluar el éxito del proyecto, se definirán puntos de referencia con cifras concretas. Por ejemplo, se puede establecer como objetivo reducir el error de estimación en un cierto porcentaje en comparación con los métodos anteriores.

### **Medidas Subjetivas y Árbitro del Éxito**

Las medidas subjetivas podrían incluir la satisfacción de los actuarios y analistas con la calidad de las predicciones y la facilidad de uso del modelo. El árbitro del éxito podría ser el comité directivo que evaluará si los resultados cumplen con los objetivos establecidos y si el modelo mejora la precisión de las provisiones.

### **Despliegue de Resultados y Éxito**

El despliegue exitoso de los resultados del modelo es parte integral del éxito de la minería de datos. Si los resultados son implementados de manera efectiva y utilizados para mejorar la toma de decisiones y la gestión de riesgos, el proyecto se considerará exitoso.

### **Planificación del Despliegue**

La planificación del despliegue debe comenzar temprano. Esto incluye identificar las herramientas y plataformas para presentar los resultados, diseñar informes claros y accesibles, y capacitar al personal en la interpretación y uso del modelo.

**1.4.1. Plan proyecto**

Fase	Tiempo (días)	Recursos	Riesgos
Definición del Alcance	14	Equipo de proyecto, patrocinador interno	Cambios en los objetivos
Recopilación de Datos	28	Analistas de datos, base de datos	Datos incompletos o inconsistentes
Análisis Exploratorio	21	Analistas de datos, expertos en seguros	Problemas de calidad de datos
Construcción de Modelos	42	Analistas de datos, actuarios	Selección inadecuada de algoritmos
Validación y Ajuste	14	Analistas de datos, equipo técnico	Overfitting del modelo
Evaluación y Métricas	7	Analistas de datos, equipo técnico	Interpretación incorrecta de resultados
Despliegue y Capacitación	14	Equipo técnico, usuarios finales	Resistencia al cambio
Monitoreo y Mantenimiento	Continuo	Equipo técnico, actuarios	Cambios en patrones de reclamaciones

# Capítulo 2

## Entendimiento de los datos

### 2.1. Recolectando datos

1. **¿Cuáles atributos (columnas) de la base de datos parecen más prometedores?**

Los atributos que parecen más prometedores son aquellos que tienen un rango de valores amplio, contienen información relevante y presentan una variabilidad significativa entre las observaciones. Por ejemplo, `AccidentYear`, `DevelopmentYear`, `IncurLoss_D`, `CumPaidLoss_D`, `EarnedPremDIR_D` y `EarnedPremNet_D` parecen ser atributos prometedores debido a la diversidad de sus valores y su relación con el proceso de seguros y reclamos.

2. **¿Cuáles atributos parecen irrelevantes y se pueden excluir?**

Los atributos que parecen irrelevantes podrían ser aquellos que tienen valores constantes o repetitivos, como `Single` si solo tiene un valor predominante. También, si `GRNAME` solo contiene nombres de grupos y no aporta información significativa para el análisis, podría considerarse menos relevante.

3. **¿Hay suficientes datos para sacar conclusiones generalizables o hacer predicciones precisas?**

Para determinar si hay suficientes datos, es importante considerar la cantidad total de observaciones y la diversidad de los valores en cada atributo. Si hay un gran número de observaciones y una variabilidad sustancial en los valores de los atributos clave, es más probable que se puedan extraer conclusiones generalizables y realizar predicciones precisas.

4. **¿Hay demasiados atributos para el método de modelado elegido?**

Si hay demasiados atributos en comparación con la cantidad de observaciones, podría haber problemas de sobreajuste y dificultades para obtener modelos confiables. En este caso, es recomendable realizar selección de características o técnicas de reducción de dimensionalidad para manejar la complejidad.

5. **¿Estás fusionando varias fuentes de datos? Si es así, ¿hay áreas que podrían ser problemáticas al fusionar?**

Si estás fusionando datos de múltiples fuentes, es importante asegurarse de que las claves de fusión sean consistentes entre las fuentes. Las diferencias en las

representaciones de las claves podrían dificultar la fusión y requerir un proceso de limpieza y estandarización.

**6. ¿Has considerado cómo se manejan los valores faltantes en cada una de tus fuentes de datos?**

Los valores faltantes pueden afectar la calidad de los análisis y modelos. Es importante comprender cómo se manejan los valores faltantes en cada fuente de datos. Puedes utilizar métodos como la imputación, eliminación o análisis de patrones de valores faltantes para abordar este problema de manera adecuada.

## 2.2. Descripción de los datos

### Cantidad de los datos

La base de datos se distingue por su extensión sustancial, comprendiendo un total de 13,200 observaciones distribuidas en una matriz de múltiples columnas, cada una representando diversas características. La disposición tabular de los datos, donde cada fila denota una observación específica y cada columna encapsula una característica particular, posibilita un abordaje eficiente y analítico de los mismos mediante utilidades como hojas de cálculo y bibliotecas de programación.

Respecto al método empleado para adquirir los datos, no se proporcionan detalles específicos en la información disponible. Las metodologías pueden variar desde el ingreso manual en hojas de cálculo hasta sistemas automatizados de recolección de datos, incluyendo la importación desde fuentes externas mediante protocolos como ODBC (Open Database Connectivity) y otros mecanismos de acceso a bases de datos. En términos dimensionales, la base de datos alberga 13,200 filas, representando las observaciones, y múltiples columnas, correspondientes a los atributos. La cantidad exacta de columnas no se encuentra especificada. Dada la amplitud de observaciones, se potencia la capacidad para llevar a cabo análisis sustanciales y construir modelos más precisos, siempre y cuando se realice una selección y tratamiento adecuado de los atributos involucrados.

### Calidad de los datos

En relación a la calidad de los datos, se puede afirmar que estos contienen características directamente pertinentes a la cuestión comercial planteada. Entre los tipos de datos presentes, se encuentran tanto simbólicos como numéricos, abarcando una gama diversa de información. Además, se realizaron cálculos de estadísticas básicas para los atributos clave, lo que permitió obtener una comprensión más profunda sobre la distribución y las tendencias asociadas a la pregunta de negocios en cuestión. En cuanto a la priorización de atributos relevantes, se dispone de la capacidad para llevarla a cabo; no obstante, en caso de necesidad, los analistas comerciales se encuentran disponibles para aportar mayor perspicacia y contexto a este proceso.

Los datos exhiben una pertinencia inherente a la pregunta empresarial, con una diversidad de tipos de datos que abarcan lo simbólico y lo numérico. La evaluación de estadísticas fundamentales para los atributos clave ha proporcionado información



esencial para entender cómo estos se relacionan con la pregunta de negocios. La capacidad para priorizar atributos pertinentes está presente, pero en situaciones más complejas, los analistas empresariales pueden ofrecer una visión adicional y específica.

### 2.2.1. Exploración de los datos

Variable	Descripción	Mín.	1er Cuartil	Mediana	Media	3er Cuartil	Máx.
GRCODE	Código	86	8526	14110	17153	26983	44300
GRNAME	Nombre						
AccidentYear	Año Accidente	1988	1990	1992	1992	1995	1997
DevelopmentYear	Año Desarrollo	1988	1994	1997	1997	2000	2006
DevelopmentLag	Retraso	1.0	3.0	5.5	5.5	8.0	10.0
IncurLoss_D	Pérdida Incurrida	-59	0	544	11532	6526	367404
CumPaidLoss_D	Pérdida Pagada Acum.	-338.0	0.0	351.5	8215.7	4565.0	325322.0
BulkLoss_D	Pérdida Global	-4621.0	0.0	5.0	1570.1	259.2	145296.0
EarnedPremDIR_D	Prima Devengada Directa	-6518	0	1419	18438	11354	421223
EarnedPremCeded_D	Prima Devengada Cedida	-3522.0	0.0	144.5	1812.3	1141.0	78730.0
EarnedPremNet_D	Prima Devengada Neta	-9731	0	827	16626	9180	418755
Single	Único	0.0000	0.0000	1.0000	0.7273	1.0000	1.0000
PostedReserve97_D	Reserva Publicada en 1997	0	411	2732	39714	19266	1090093

Cuadro 2.1: Descripción de Datos

Los datos representan información relacionada con seguros y reclamos. A continuación, se detallan las estadísticas de diferentes variables presentes en el conjunto de datos:

- **GRCODE:** Los valores de esta variable representan códigos únicos de grupos. Se encuentran 10 grupos por cada código, es decir hay 132 códigos únicos. El valor mínimo es 86, el primer cuartil (25 %) es 8526 y el valor máximo es 44300.
- **GRNAME:** Esta variable contiene nombres de grupos en formato de caracteres.
- **AccidentYear:** Indica el año en que ocurrió el accidente. Los datos abarcan desde el año mínimo de 1988 hasta el máximo de 1997.
- **DevelopmentYear:** Muestra el año de desarrollo. Los valores oscilan entre 1988 y 2006.
- **DevelopmentLag:** Representa el lapso de tiempo en años entre el año del accidente y el año de desarrollo. La media es 5.5 años.
- **IncurLoss\_D:** Esta variable refleja pérdidas incurridas. El valor mínimo es -59 y el valor máximo es 367404.
- **CumPaidLoss\_D:** Muestra la suma acumulada de pérdidas pagadas. Los valores varían desde -338 hasta 325322.
- **BulkLoss\_D:** Indica la suma de pérdidas a granel. Los datos van desde -4621 hasta 145296.
- **EarnedPremDIR\_D:** Representa la prima devengada directa. Los valores van desde -6518 hasta 421223.

- **EarnedPremCeded\_D:** Indica la prima devengada cedida. Los valores oscilan entre -3522 y 78730.
- **EarnedPremNet\_D:** Muestra la prima devengada neta. La media es 16626.
- **Single:** Esta variable toma valores binarios (0 o 1) y representa si la entidad es un único reclamo. El 72.73 % de los casos tienen valor 1.
- **PostedReserve97\_D:** Muestra las reservas publicadas en el año 1997. Los valores oscilan desde 0 hasta 1090093.

En resumen, el conjunto de datos proporciona información sobre grupos de seguros y sus características asociadas, como años de accidente y desarrollo, pérdidas incurridas, primas devengadas y otros aspectos relevantes para el análisis de seguros y reclamos.