



Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas

MINERÍA DE DATOS

Maestra: MAYRA CRISTINA BERRONES REYES

Integrantes:

Romero Galván Asiel 1937895
Christian Servando Garza Gonzalez 1505813
Jorge Noe Zuñiga Gomez 1565813

Grupo: 003

Equipo: 003

1) **Título de la base de datos**

- a. Health Insurance data
- b. <https://www.kaggle.com/omartronco/health-insurance-data>

2) **Descripción de los datos**

- a. La información en la base de datos está representada por **tablas**.
- b. Dado que tenemos diferentes tablas, explicaremos las columnas por secciones:
 - i. **Coverage modifications per group:** Aquí manejaremos 4 columnas:
 - 1. Grupo: Aquí se clasifican el total de asegurados en 5 grupos de acuerdo con el tipo de seguro con el que cuentan.
 - 2. Deducible: De acuerdo con el seguro con el que cuentan (grupo), es la cantidad de deducible que pagara el asegurado. Ya hay un monto preestablecido.
 - 3. Coaseguro: De acuerdo con el seguro con el que cuentan (grupo), es la cantidad de coaseguro que pagara el asegurado. Aquí se aplica un porcentaje de acuerdo con el siniestro.
 - 4. Límite de pago: Es la cantidad máxima que pagara la asegurada según el grupo al que pertenece el asegurado.
 - ii. **Diseases per person:** Se encuentran 2 columnas:
 - 1. Subject: ID del asegurado.
 - 2. Número de reclamaciones: Son las reclamaciones que ha realizado un asegurado es un periodo de un año.
 - iii. **Payments 2015-2018:** Para las tablas de Payments 2015 a Payments 2018, se usan las mismas 4 columnas:
 - 1. Año: Hace referencia al año en el que se hizo la reclamación.
 - 2. Tipo: Es el tipo de la enfermedad/accidente de la reclamación.
 - 3. Grupo: A qué grupo pertenece el asegurado de acuerdo con la tabla "Coverage modifications per group".
 - 4. Pago: Es el monto que pago la aseguradora según el tipo de seguro.

3) **Justificación del uso de datos**

- a. Lo que más nos llamó la atención de esta base de datos es que utiliza información real de Aseguradora lo que nos servirá para fortalecer nuestra formación profesional como Actuarios.
- b. El mayor beneficio que encontramos al trabajar con estos datos es que tiene una aplicación en el mundo Actuarial, principalmente en el ámbito de seguros. Será una Buena práctica para aplicar nuestros conocimientos, e incluso realizar una auditoria real de una aseguradora.

4) **Planteamiento del problema**

- a. Desde el punto de vista de una empresa, es necesario justificar los datos en base a lo establecido en el lineamiento de la circular única de seguros y fianzas y así cumplir con el marco legal.

5) **Objetivo final**

- a. Utilizando datos previos, establecer como se comportarán los datos en un futuro, es decir, poder definir entre que rango se encontrará la severidad, y predecir qué porcentaje tendrá cada grupo de enfermedad todo esto de acuerdo con la Auditoria

según el anexo 6.3.2 para seguros no catastróficos de danos y seguros de accidentes y enfermedades. Nuestro objetivo es determinar la frecuencia del número de siniestros por medio del criterio 1,2,3,4 a través de la SESA del mismo modo la severidad de pago de siniestros individuales a través de los criterios 5,6,7,8. Van a incorporar año con año la nueva experiencia de los SESA, instituciones que no tenía suficiente información en cuanto al número de siniestros observados. Para utilizar sus propios parámetros de ajuste de panza y cola, esta podría presentar ya suficiente información para dejar de utilizar la información del mercado y utilizar su propia experiencia de factores de ajuste.

- b. Como objetivo secundario, determinaremos el esperado de la prima pura en el año posterior.

6) Planeación de la herramienta a utilizar

- a. El tipo de técnica que se desea utilizar es visualización ya que, por medio de esta, podremos confirmar la distribución de nuestros datos, si hay datos atípicos y/o si siguen alguna tendencia. Sin embargo, nuestros datos dejan la puerta abierta para poder implementar regresión lineal o proyección.
- b. La herramienta que tenemos pensado implementar es Python o R ya que, por medio de estas dos herramientas, podemos utilizar de manera eficiente la visualización con un gran número de datos.