



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**  
**SISTEMAS DE INFORMACION**  
**PROYECTO FINAL 1**  
**ANALISIS DE DATOS**

FC-FISC-1-8-2016)



Facilitador(a): Ing. Carmen Ortega Asignatura: GESTION DE LA INFORMACION

Fecha: \_\_\_\_\_ Grupo: 1SF141, 1SF142

Estudiantes: \_\_\_\_\_

### **PROCEDIMIENTO O ENUNCIADO DE LA EXPERIENCIA:**

- Elabore el informe de análisis de datos empleando las herramientas Python, MySQL, SQLite3, Anaconda y sus paquetes básicos (NumPy, SciPy, Pandas, matplotlib, statsmodels), Pytorch Keras, Tensorflow, Jupyter, Scikit –learn, Tableau, Weka
- Elabore un estudio de análisis de datos de cada Data Set, proponga su modelo, sustente y visualice los resultados del estudio.
- Ponga en práctica todos los conceptos dados a la fecha: preparación de los datos, filtros, algoritmos supervisados, algoritmos no supervisados con las herramientas usadas en el curso.

#### **Estructura del documento**

1. Introducción
2. Descripción del dataset
3. Preparación de los datos
4. Aplicación de Filtros
5. Modelado de datos y Análisis de los datos
6. Visualización de sus resultados

#### **Descripción del Proyecto Final 1**

### **Reto Desafío de Analítica de Datos**

#### **Tarificación de los Seguros de Automóviles**

Uno de los ejercicios fundamentales de una compañía de seguros es determinar el precio de una cobertura de seguro. Cada seguro es al final de cuentas una apuesta de la aseguradora hacia el cliente que las cosas van a salir bien. La pregunta entonces es: ¿Cuánto cobrar por la apuesta? Si se pone el precio del seguro muy bajo, no habrá suficiente dinero para pagar los reclamos a las coberturas. Si se pone el precio del seguro muy alto, los clientes se irán donde un competidor. El objetivo de la tarificación de seguros es entonces asignarles a los seguros el precio correcto: ni muy alto, ni muy bajo. Esta asignación se hace en base una estimación de futuros reclamos en base a los patrones que otros asegurados similares han tenido en el pasado. Este ejercicio es de



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**  
**SISTEMAS DE INFORMACION**  
**PROYECTO FINAL 1**  
**ANÁLISIS DE DATOS**

FC-FISC-1-8-2016)



regresión: se debe predecir un número real que puede ser la prima directamente o una tasa de prima en base a la suma asegurada del vehículo.

Este ejercicio es clave en el éxito en los seguros. Una empresa que puede estimar mejor el costo de un seguro podrá quedarse con clientes al precio correcto y dejarle a su competencia, que no estima bien los costos, los clientes con más siniestros.

Las variables “X” corresponden a las características del automóvil y su conductor: marca, modelo, año modelo del vehículo, sexo y edad del conductor. Las variables “Y” son los montos de dinero reclamados por dichas coberturas. En el Anexo A, está la descripción de cada variable y la representación de cada punto de datos.

### Características de los Datos

Ruido: es muy común que existan grupos en los que solamente un registro tiene reclamos. Por esto es por lo que se trata de predecir el “promedio” de primas de un grupo, aunque el valor reclamado más común sea cero.

Colinealidad: los asegurados y sus vehículos. Es muy común que ciertos diseños de marcas y modelos “atraigan” cierto grupo demográfico de asegurados. Por ejemplo: automóviles diseñados para atraer consumidores jóvenes, o automóviles diseñados para transportar familias. Un sub-reto de este ejercicio puede ser identificar esos “clusters” de automóviles de gente joven o “automóviles de mujeres”. En cualquier caso, el modelo predictivo de los reclamos debe tener en cuenta esta y cualquier otra colinealidad que exista en los datos.

### ¿Qué se ha hecho en el pasado?

Tradicionalmente estos problemas se han analizado utilizando Modelos Lineales Generalizados (GLM’s). Siendo los GLM’S el “state-of-the-art” en cuanto a modelos de tarificación de seguros de automóviles, se le recomienda al estudiante ejecutar este tipo de modelos para por lo menos como modelo base o modelo de referencia.

### Métricas de éxito de un modelo

¿Qué tan bien puede un modelo predictivo estimar el reclamo promedio y por consiguiente la base de la prima? Sugerencias: métricas como el Mean Square Error o Mean Absolute Error o definir creativamente otra métrica. El estudiante deberá establecer una o varias de estas métricas y explicar por qué una es mejor que otra.

### Herramientas

Para este proyecto se permitirá cualquier herramienta tecnológica que sea necesaria para el análisis y creación de modelos, enfocándonos en software de código libre (Free and Open Source Software).

En caso de que el estudiante requiera un paquete adicional, no debe dudar en solicitarlo.



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**  
**SISTEMAS DE INFORMACION**  
**PROYECTO FINAL 1**  
**ANALISIS DE DATOS**

FC-FISC-1-8-2016)



## Anexo A

### Renglón por renglón

Lo que se busca representar en cada renglón es un “momento de exposición a riesgo” de un vehículo. Por lo general esto significa que cada renglón representa un vehículo asegurado con el mismo dueño durante un año.

Si un vehículo estuvo asegurado durante “n” años, debe aparecer en “n” renglones, correspondiente a cada año que estuvo asegurado. Entre cada renglón debe tener la suma asegurada que le correspondía cada año (la suma asegurada tiende a disminuir cada año por depreciación).

### Descripción de Columnas

NUM_SERIEV	Número de serie del vehículo o VIN. Por motivos de privacidad este campo está serializado.
NUM_POL	Número de póliza del año
NUMUNIDADVP	Número de unidad del vehículo dentro de una póliza.
FECINIVIGENTE	Fecha de inicio de vigencia
FEC_FIN	Fecha original de inicio de vigencia
FEC_ANU	Fecha de cancelación (en caso de haber cancelación) de la póliza
GENERO	Sexo del contratante de la póliza
EDAD	Edad del contratante de la póliza
ANOMVEH	Año de modelo del vehículo
ACREEDOR	Si la póliza tiene acreedor hipotecario
COD_MARCAV	Código de la marca del vehículo
COD_MODELOV	Código del modelo del vehículo dentro de una marca. El modelo 1, es el primer modelo de la marca n. Por ejemplo: si la marca es 30 es Toyota, el modelo 1 es Avensis, y el modelo 10 es Yaris. En la marca 20, Hyundai, el modelo 1 es Accent y el modelo 10 es Tucson. La conclusión importante es: el número de modelo no nos dice nada si no se sabe de cuál marca es.
MCLIM_LC	Monto cubierto para LC (Lesiones Corporales)



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**  
**SISTEMAS DE INFORMACION**  
**PROYECTO FINAL 1**  
**ANALISIS DE DATOS**

FC-FISC-1-8-2016)



MCLIM_SO_LC	Monto cubierto para LC (Lesiones Corporales), Seguro Obligatorio (SO). Esta cobertura entró en vigor en 2014
MCLIM_DPA	Monto de cobertura para daños a la propiedad ajena (DP)
MCLIM_SO_DPA	Monto de cobertura para daños a la propiedad ajena (DP), Seguro Obligatorio (SO). Esta cobertura entró en vigor en 2014.
MCLIM_AM	Monto de cobertura para Asistencia Médica (AM)
MCLIM_CP	Monto de cobertura para Comprensivo (CP)
MCLIM_CO	Monto de cobertura para Colisión (CP)
MRECT_LC	Monto reclamado total en el año para Lesiones Corporales (LC)
MRECT_SO_LC	Monto reclamado total en el año para Lesiones Corporales (LC) Seguro Obligatorio SOAT
MRECT_DPA	Monto reclamado total en el año para Daños a la Propiedad (DPA)
MRECT_SO_DPA	Monto reclamado total en el año para Daños a la Propiedad (DPA), Seguro Obligatorio (SO)
MRECT_AM	Monto reclamado total en el año para Asistencia Médica (AM)
MRECT_CP	Monto reclamado total en el año para Comprensivo (CP)
MRECT_CO	Monto reclamado total en el año para Colisión (CO)
CANTAC_LC	Cantidad de accidentes durante el año que afectan LC
CANTAC_SO_LC	Cantidad de accidentes durante el año que afectan LC, Seguro Obligatorio (SO)
CANTAC_DP	Cantidad de accidentes durante el año que afectan DPA
CANTAC_SO_DP	Cantidad de accidentes durante el año que afectan DPA, Seguro Obligatorio (SO)
CANTAC_AM	Cantidad de accidentes durante el año que afectan AM
CANTAC_CP	Cantidad de accidentes durante el año que afectan CP
CANTAC_CO	Cantidad de accidentes durante el año que afectan CO



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**  
**SISTEMAS DE INFORMACION**  
**PROYECTO FINAL 1**  
**ANALISIS DE DATOS**

FC-FISC-1-8-2016)



## Abreviaturas

### **LC: Lesiones Corporales**

Cubre la responsabilidad civil hacia terceros por lesiones corporales (incluyendo muerte). Por ejemplo, si un asegurado nuestro atropella a un peatón. Antes de 2014 esta cobertura era obligatoria, a partir de 2014 pasó a ser opcional en exceso de LC-SOAT.

### **LC-SO: Lesiones Corporales SO**

Cubre la responsabilidad civil hacia terceros por lesiones corporales (incluyendo muerte) de Seguro Obligatorio de Accidentes de Tránsito (SO). Esta cobertura empezó a otorgarse en 2014. Esta cobertura es obligatoria.

### **DPA: Daños a la Propiedad Ajena**

Cubre la responsabilidad civil del asegurado con respecto a daños a la propiedad de terceros. Por ejemplo: si nuestro asegurado choca otro vehículo y es culpable, la aseguradora le paga los daños al vehículo del tercero que nuestro asegurado chocó.

### **DPA-SO: Daños a la Propiedad Ajena - SO**

Daños a la Propiedad Ajena. Cubre la responsabilidad civil del asegurado con respecto a daños a la propiedad de terceros, de Seguro Obligatorio de Accidentes de Tránsito (SO). Esta cobertura empezó a otorgarse en 2014. Esta cobertura es obligatoria.

### **AM: Asistencia Médica**

Cubre gastos médicos incurridos a raíz de un accidente dentro del vehículo asegurado para el conductor y los pasajeros. En otras palabras, si después de un accidente tienes que ir al hospital para recibir atención médica, el seguro paga esos gastos del hospital. Esta cobertura es opcional.

### **CP: Comprensivo**

Cubre gastos por daños al vehículo asegurado mientras está estacionado. Algunos riesgos cubiertos bajo esta cobertura son: inundación, robo, terremoto, caída de objetos sobre el vehículo.

### **CO: Colisión**

Cubre los daños al vehículo asegurado en caso de colisión contra otro vehículo, contra objeto fijo o por vuelco del vehículo asegurado.

## Bibliografía

Anderson et al. *"A practitioner's guide to Generalized Linear Models, a foundation for theory, interpretation and application"*. Casualty Actuarial Society Discussion Paper Program (2004).

Christmann, A. *"On a Strategy to Develop Robust and Simple Tariffs from Motor Vehicle Insurance Data"*. Acta Mathematicae Applicatae Sinica, English Series 21, 193–208 (2005).



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**  
**SISTEMAS DE INFORMACION**  
**PROYECTO FINAL 1**  
**ANALISIS DE DATOS**

FC-FISC-1-8-2016)



De Jong, P. & Heller, G. *“Generalized Linear Models for Insurance Data”*. Cambridge University Press. (2008)

Dugas, Bengio et al. *“Statistical learning Algorithms Applied to Automobile Insurance Ratemaking”* Intelligent and other Computational Techniques in Insurance Theory and Practice (2003)

McCullagh, P. & Nelder J.A. *“Generalized Linear Models”*. Springer (1983).

Hadidi, N. *“Classification Ratemaking Using Decision Trees”* Casualty Actuarial Society Publications (2001)

**RÚBRICAS:**

<i>Detalle - Evaluar</i>	<i>Valor de la respuesta</i>
<i>Estructura del documento</i>	<i>50 puntos</i>
<i>Presentación de su trabajo</i>	<i>50 puntos</i>