**ANÁLISIS DE DATOS**

**Rúbrica de Evaluación**

1. Calidad del Código (40%): El código debe ejecutarse correctamente. Uso eficiente y claro del lenguaje R, con código bien organizado y comentado.
2. Análisis y Comprensión (30%): Capacidad para interpretar los resultados y proporcionar insights relevantes.
3. Claridad en la Presentación de Resultados (30%): Presentación clara y eficaz de los resultados, incluyendo gráficos y tablas donde sea necesario.

**Para los siguientes ejercicios considera.**

* Crear una narrativa que describa los resultados obtenidos, utilizando gráficas y resaltando eventos clave.
* Generar un informe reproducible utilizando R Markdown que incluya todo el proceso de análisis desde la carga de datos hasta las visualizaciones, aplicando estrategias de manejo de datos.

**EJERCICIOS**

1. Preprocesamiento de Datos y Manejo de Calidad: Evaluar la calidad de los datos en campos como 'fecha\_constitucion', 'prima\_anual', y 'suma\_asegurada'. Identificar y corregir inconsistencias o valores atípicos.
2. Análisis Exploratorio en Relación a la Sucursal y el Ramo Comercial: Investigar si existe alguna correlación entre la 'sucursal' y el 'ramo\_comercial' en términos de 'prima\_emitida'. Esto implicaría agrupar los datos por sucursal y ramo comercial, y luego realizar análisis estadísticos y gráficos para identificar patrones.
3. Evaluación de la Evolución Temporal de las Pólizas: Analizar cómo ha evolucionado la 'prima\_anual' a lo largo del tiempo (usando 'fecha\_emision'). Crear visualizaciones de series temporales para mostrar tendencias o cambios significativos a lo largo de los años.
4. Segmentación de Clientes por Tipo de Póliza y Monto Asegurado: Utilizar métodos de clasificación o agrupamiento (como k-means) para segmentar las pólizas en grupos basados en 'tipo\_persona', 'tipo\_agente', y 'suma\_asegurada'. Esto podría revelar patrones interesantes en términos de qué tipos de clientes prefieren ciertos tipos de pólizas y montos asegurados.

**Entregables**

Se debe entregar un documento en formato acrobat (pdf) con los resultados solicitados y una presentación en un formato dinámico (power point, html, etc). Se debe entregar, también el código desarrollado para obtener los resultados (R o Python). Se prefiere que los entregables se realicen en el software estadístico R utilizando Rmarkdown.

Primera Diapositiva

Introducción:

* Introducción
  + Las pólizas de seguro son esenciales para proteger a individuos y empresas contra riesgos financieros imprevistos, ofreciendo estabilidad económica. Analizar estos datos ayuda a las aseguradoras a comprender mejor los riesgos y adaptar sus productos a las necesidades específicas de los clientes.
* ¿Que son las polizas de seguro?
  + Las pólizas de seguro son contratos entre una entidad aseguradora y una persona o empresa, en los cuales la aseguradora se compromete a compensar al asegurado por pérdidas específicas a cambio de una prima periódica. Estos contratos proporcionan protección financiera cubriendo diversos riesgos, como accidentes, enfermedades, daños a la propiedad, o responsabilidad legal.
* Importancia de las polizas de seguro.
  + Proteccion y seguridad financiera.
    - Las pólizas de seguro son clave para mitigar riesgos financieros, permitiendo que asegurados transfieran el riesgo de eventos imprevistos a la aseguradora a cambio de una prima. Esto facilita una recuperación más rápida y mantiene la estabilidad económica y personal tras adversidades.
  + Faciliatacion del riesgo empresarial.
    - Para las empresas, las pólizas de seguro son cruciales para gestionar riesgos que podrían obstaculizar su operación y crecimiento. Desde la cobertura de responsabilidad civil hasta el seguro de propiedad, las pólizas permiten que las empresas tomen riesgos calculados, lo cual es esencial para la innovación y la expansión.
  + Requerimiento Legal y Contractual
    - En muchos casos, tener ciertos tipos de seguro es un requisito legal o contractual. Por ejemplo, el seguro de vehículos es obligatorio en muchos países, y el seguro de responsabilidad civil puede ser necesario para contratos comerciales. Estas pólizas garantizan que tanto las partes involucradas como terceros estén protegidos en caso de un siniestro.

Entendimiento de los datos:

* Importación de librerías necesarias
* Importación de los datos
* Visualización de las primeras 5 filas de los datos
* Tipo de dato por columna y cantidad de valores no nulos

Preprocesamiento (EDA: Exploratory data analysis)

* Transformar columnas relacionados a fechas en tipo datetime
* Analizar outliers en valores numéricos
  + Solventar el problemas de los valores atipicos
    - Limitación (Winsorización)
    - Media, Mediana o Moda
    - KNN (K-Nearest Neighbors)
  + Gráficamente
    - Gráfico de dispersión
    - Gráfico de cajas
    - DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Descripción del dataset:

* El dataset contiene información sobre pólizas de seguro, destacando detalles como el giro de actividad, tipo de agente, sucursal, y ramo comercial. También incluye datos financieros y temporales como la prima anual, la suma asegurada, y las fechas de emisión y vigencia de las pólizas. Este conjunto de datos es crucial para análisis y gestión de seguros en sectores de producción y comercialización.
* Interpretación de las columnas
  + **suma\_aseg:** La columna suma\_aseg en el dataset representa la "suma asegurada", que es el monto total máximo que la compañía de seguros se compromete a pagar en caso de que ocurra un siniestro cubierto bajo los términos de la póliza de seguro. Esta cifra es crucial porque define la cobertura financiera del riesgo asegurado.
  + **giro\_actividad:** Indica el sector de actividad de la empresa asegurada, como producción o comercialización/distribución.
  + **fecha\_constitucion:** Fecha en la que se constituyó la póliza, expresada en formato de número de serie de fecha de Excel.
  + **id\_poliza:** Identificador único para cada póliza de seguro.
  + **anio\_mes:** Fecha correspondiente al mes y año de la póliza, también en formato de número de serie de fecha de Excel.
  + **tipo\_agente:** Tipo de agente que gestionó la póliza, que puede ser un productor, una unidad de producción o directo.
  + **sucursal:** Ubicación de la sucursal donde se gestionó la póliza, como Quito o Guayaquil.
  + **ramo\_comercial:** Tipo de seguro contratado, que puede ser, por ejemplo, "Buen uso de anticipo" o "Fiel cumplimiento de contrato".
  + **tipo\_persona:** Clasificación del asegurado, en este caso todos como "JURIDICO", indicando personas jurídicas (empresas).
  + **prima\_anual:** Monto anual que el asegurado paga por la cobertura de la póliza.
  + **prima\_emitida:** Parte de la prima anual que ha sido emitida efectivamente en el período registrado.
  + **fecha\_emision:** Fecha en que se emitió la póliza, en formato de número de serie de fecha de Excel.
  + **fecha\_vigencia\_desde:** Fecha de inicio de la vigencia de la póliza.
  + **fecha\_vigencia\_hasta:** Fecha de fin de la vigencia de la póliza.
  + **forma\_pago:** Método de pago utilizado para pagar la póliza, como efectivo.

Analizar la correlación entre la sucursal y el ramo comercial: