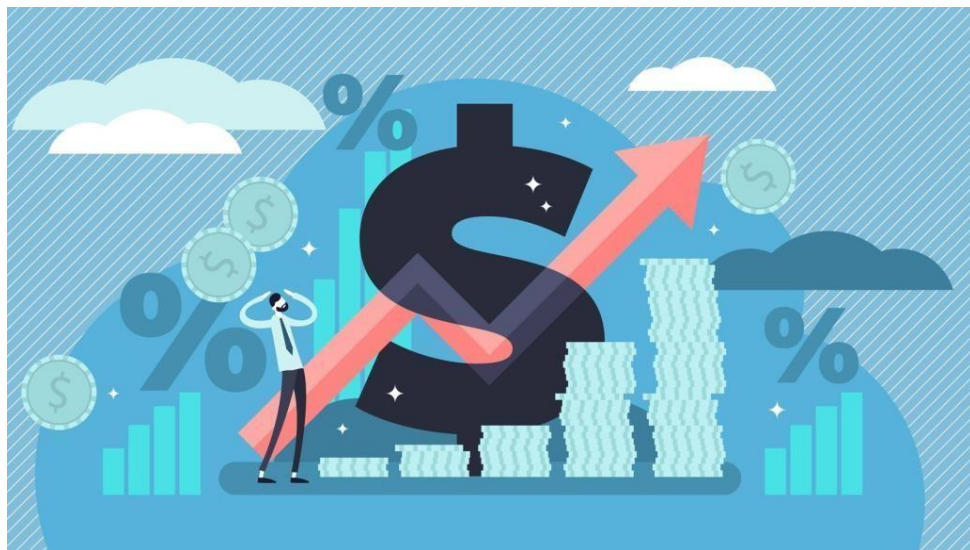


## TRABAJO FINAL

Se entrega el siguiente set de datos de clientes de una empresa del sector financiero, específicamente hablando encargada del otorgamiento de pólizas de seguros a los peruanos, DataInferencialIngresos.xlsx, en el cual tenemos los datos de sus clientes en el periodo del 2019 respecto a distintas variables de su negocio. En el último tiempo los niveles de aceptación y efectividad de sus campañas se han visto afectadas posiblemente porque sus ofertas no están siendo contextuales o asertivas con la realidad y necesidad de los clientes.

Debido a que las ofertas en las pólizas no están teniendo buenas aceptaciones, el área de analytics ha decidido hacer un análisis y reveló que las tasas a comparación de las ofertadas a la competencia eran muy superiores en unos casos y en otras muy inferiores respecto a los principales competidores.



El principal objetivo de esta problemática es entender si de una u otra manera la predicción correcta del ingreso de las personas está influyendo o determinando la aceptación de las pólizas de seguros, y si esto ocurre, cuantificar construir un estimador de ingresos adecuado. Debido a que usted está llevando un programa de especialización en Data Science requieren que los ayude a predecir o estimar los ingresos de los usuarios.



Las variables que se disponibilizan son:

Variables	Descripción de las Variables
edad	Edad del cliente.
AniosDireccior	Años de experiencia en el trabajo actual.
Gastocoche	Gasto medio en su vehículo principal.
Aniosempleo	Años de experiencia en total.
Aniosresiden	Años de residencia o de trabajo en el Perú.
ingres	Ingreso medio en el último año.

- El objetivo es realizar un Análisis Exploratorio de Datos y modelos predictivos para la Regresión
- Se puede presentar entrega en el jupyter Notebook
- Se entregará como fecha máxima 24 de Marzo 2022 (fecha límite)

## Entregables:

1. Definir el **problema** y los **objetivos de negocio**.
2. ¿Qué **tipo de variables** se utilizan en el problema de negocio?
3. Hacer un **resumen global del total de variables de estudio**, principalmente a nivel **univariado**, entendiendo globalmente la problemática en la cual estoy inmerso.
4. De acuerdo al objetivo de negocio planteado, realizar un **análisis de asociación o correlación** comentando y documentando sus posibles hallazgos.
5. Realizar un análisis de detección de posibles valores atípicos en las principales variables. Si es que detecta estos posibles valores atípicos, ¿Crees que **sería más adecuado retirar esos registros atípicos, o retirar las variables que tengan más valores atípicos?**
6. Aplicar el modelo regresión lineal múltiple, para ello como el dataset no es una información con temporalidad por ello se puede aplicar la partición train\_test\_split (testear con un 20%) y luego interpretar los coeficientes de dichas variables, por último, ver el resultado de el test la métrica de RSME e interpretarlo.
7. Aplicar 3 diferentes tipos de modelos regresión que se hizo en clase de acuerdo a su criterio, luego la partición use la misma de la pregunta 6. Y por último compare cual de ellos tienes el RSME más bajo para determinar cual fue el mejor modelo con el que se quedó.