

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO**  
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, IIMAS  
*Ciencia de Datos*  
**Visualización de la Información**

**Venta de Casas en Ames-Iowa**  
**TAREA-11**

### **Actividad**

Utilice el software **Qlik Sense** para comparar los diversos atributos relacionados con la venta de casas en Ames Iowa y hacer una visualización más informativa. Para ello utilice el conjunto de datos [train.csv](#)

**Limpieza de datos.** Elimine las variables con más de un 15% de datos faltantes (por ejemplo, 'PoolQC', 'MiscFeature', 'Alley', etc.); no parecen aspectos importantes a considerar al comprar una casa. Elimine también 'BsmtX'. En cuanto a las variables 'MasVnrArea' y 'MasVnrType', se puede decir que no son esenciales.

- Elabore una gráfica donde se visualicen los costos de las casas (**SalePrice**). Agrupe los valores en cubos para crear rangos de valores y utilice esta nueva variable (En cubos).
- Elabore una gráfica de barras donde se visualice cuándo las casas fueron construidas (**YearBuilt**). Considere un título para la grafica, etiquetas, colores y otros factores que sean convenientes.
- Para determinar si la época del año en que se vende una casa tiene algún efecto en el precio de venta, muestre una gráfica que nos permita visualizar cuándo las casas fueron vendidas (**YrSold**, **MoSold**). Para ello considere el año de venta y el mes de venta.
- Uno de los factores cuando se compra una casa es la **ubicación**, ya que esta relacionada con el precio de venta. Muestre una gráfica que indique en qué vecindarios están las casas (**Neighborhood**). Ponga un título a la grafica, etiquetas y modifique los colores por cada dimension.

Una característica del conjunto de datos es que presenta valores atípicos (ventas parciales que probablemente no representan los valores de mercado reales) como pueden ser ventas inusuales (casas muy grandes con un precio relativamente apropiado). Para determinar ello, muestre una gráfica que indique las dimensiones de las casas (**GrLivArea**). Agrupe los valores en cubos para crear rangos de valores y utilice esta nueva variable (En cubos).