**Reto Parcial 1**

De acuerdo con la base de datos NDA\_Data del Capitulo 13 del Libro Guía (Analítica de Negocios – Comunicación con Datos - Jaggia), implementar tres modelos que permitan determinar si el número de juegos jugados por jugador de la NBA esta por encima del promedio (games\_played). Para la implementación de los modelos, se deben tener en cuenta las siguientes recomendaciones:

* Las variables de entrada para la construcción del modelo son las siguientes: Salary, Age, Height, Weight, games\_started, minutes\_per\_game, assits, blocks, steals.
* Para aplicar la línea de código XDB.dropna(), es necesario primero seleccionar las variables de entrada y de salida.
* Determinar cual de los modelos tuvo un mejor comportamiento frente al pronóstico de la variable games\_played (por encima (1) – por debajo (0) del promedio de la variable. )
* Para la construcción del árbol, es necesario considerar solo 4 niveles, mientras que para el modelo de Clusterización, es importante tener en cuenta 5 grupos de datos.
* Para la obtención de la gráfica de los clusters, es importante tener en cuenta las variables que tienen una mayor correlación con respecto a la variable de salida, para luego mostrar las gráficas de correlación.
* Determinar la categoría a la que pertenece una persona con las siguientes características mediante la utilización del modelo Naive Bayes y Clustering:

[5000000,37,82,245,45,17.8,0.5,1.4,0.4]

* De acuerdo con la estructura del árbol de decisión, determinar la regla que cumple el deportista y cual es la probabilidad de pertenencia.