**Proyecto de análisis de datos:**

Este proyecto busca poder abordar una problemática de análisis de datos con condiciones reales, tal como ocurriría en cualquier empresa o entidad.

Elija alguno de los datasets que se encuentran en las páginas de datos abiertos de USA (<https://data.gov/>) o Colombia (<https://www.datos.gov.co/>). Asuma que usted fue contratado por dicha empresa o entidad para realizar un proyecto de analítica que resuelva alguna problemática que pueda ser abordada con el dataset elegido. La problemática debe ser relacionada con la temática de **modelamiento de series de tiempo.**

Tenga en cuenta que el dataset elegido debe tener una volumetría no muy pequeña (no puede ser un conjunto de datos muy pequeño).

De esta forma, siga todas **las etapas y subetapas** de la metodología CRISP-DM de principio a fin para dar solución a esta problemática. Se pueden guiar de este link: <https://pdf4pro.com/amp/cdn/crisp-dm-1-the-modeling-agency-com-73581e.pdf>

**El proyecto puede ser realizado en equipos de máximo 3 personas.**

Los entregables del proyecto serían los siguientes:

1. Informe en Word que explique cómo se ejecutaron y cuáles fueron los resultados de cada una de las **etapas** y **subetapas** de este proyecto, siguiendo la estructura propuesta por CRISP-DM en <https://pdf4pro.com/amp/cdn/crisp-dm-1-the-modeling-agency-com-73581e.pdf>. Se deben guiar del documento para las secciones y subsecciones que debe contener este informe (además de la portada y la introducción al documento, en total serían 6 secciones según lo dispuesto en CRISP-DM, cada una con múltiples subsecciones).
2. Se deben entregar también los notebooks utilizados para todo el desarrollo del proyecto, incluyendo los códigos ejecutados y las salidas de cada uno. Se recomienda entregar al menos 4 notebooks: uno para todo el análisis exploratorio y la preparación de datos, otro para el modelamiento, otro para la evaluación de resultados del modelo escogido y un último para el uso del modelo en producción (este último debe permitir generar nuevas predicciones a futuro bajo demanda del usuario, teniendo en cuenta que es posible que este usuario le suministre una versión actualizada a la fecha de los datos).

**Fecha de entrega del informe escrito: sábado 29 de octubre (20% nota materia)**

Recuerde que , adicionalmente, el **miércoles 26 de octubre (10% nota materia)** se debe realizar la exposición de dicho proyecto durante la clase (incluye también la creación de diapositivas y la **entrega de estas**). Para esta exposición, se **debe asumir que el público objetivo es el gerente general de la empresa o de la entidad a la que le interesa el proyecto** (esto implica tener que usar un lenguaje y explicaciones claras para un gerente general, el cual conoce bien de su área de negocio, pero no sería un experto en analítica).