

## IDENTIFICACIÓN

**DIVISIÓN/ VUAD:** Ciencias Económicas y Administrativas**FACULTAD/ DEPARTAMENTO/ INSTITUTO:** Facultad de Estadística**PROGRAMA ACADÉMICO:** Estadística**NOMBRE DEL DOCENTE:** Mónica Tatiana Gutierrez Ballen**DENOMINACIÓN DEL ESPACIO ACADÉMICO**

Métodos de minería de datos en Python

## DESARROLLO DEL PROYECTO

El objetivo es aplicar el proceso KDD para un modelo de minería de datos, utilizando una base de datos recopilados en el mundo real. Para ello los estudiantes deben elegir un data set que provenga de la sección recursos, de los cuales NO se podrá elegir data sets que estén en el top de los más populares y estén con el debido permiso de confidencialidad, en el cuál deben desarrollar los siguientes puntos:

Entrega 1:

- Breve explicación de la elección del dataset: características, número de atributos y variables, área al que pertenecen los datos.
- Posibles escenarios de solución en minería de datos ¿Es un problema de aprendizaje supervisado o no supervisado? ¿Es clasificación o regresión? ¿Es clustering o reducción de dimensionalidad? Justifique su elección.
- Haga un análisis exploratorio de datos y haga una pre-selección de variables basado en este análisis. Utilice al menos una visualización de datos por integrante de grupo. Saque sus primeras conclusiones de los datos.

El taller debe ser desarrollado en grupo de 2 personas en un notebook de Jupyter o notebook de Google Colab utilizando Python para el desarrollo y Markdown para la documentación, citando la fuente de dónde fueron extraídos los datos. Además, preparar una presentación (usando ppt) de 5 minutos resumiendo los anteriores puntos que serán presentados en hora de clase frente a los demás grupos.

Debe ser subido a la plataforma oficial en un '.zip' dónde el archivo debe nombrarse así: 'grupo(# de su grupo)\_proyecto1.zip'. Además, el taller debe ser subido a Github de todos los integrantes del grupo para su revisión. Tener en cuenta el plazo de entrega, modificaciones realizadas después de la hora acordada, no se tendrán en cuenta. Seguir las instrucciones garantiza que la calificación sea sobre el puntaje máximo.

## RECURSOS

<https://archive.ics.uci.edu/ml/index.php>  
<https://datasetsearch.research.google.com/>  
<https://datahub.io/search>