



Proyecto final

Como conclusión del programa, deberá realizar un proyecto final en el que aplique una técnica de *data science* para extraer información relativa a un problema. Este proyecto está diseñado para simular la experiencia real de trabajar en el ámbito de *data science*. Consta de tres partes: dos entregas y la presentación del proyecto completo.

Deberá elaborar una presentación de PowerPoint en la que describa su proyecto. Para su elaboración, debe tener en cuenta que toda persona que abra la presentación desconocerá el contexto del problema, por lo que debe ser capaz de presentar y sintetizar toda la información del proyecto con claridad y precisión. Para ello, debe poder transmitir el propósito principal de su proyecto, ofrecer una breve descripción del problema que va a abordar, definir sus objetivos y exponer las conclusiones a las que ha llegado.

A lo largo del programa, deberá entregar dos versiones en proceso de su presentación, al final de las semanas 3 y 6, que se calificarán como "completo" o "incompleto". Las primeras entregas del proyecto son parte de un proceso iterativo. No es necesario que el resultado sea perfecto, pero debe demostrar que ha comenzado a trabajar en el proyecto y seguir mejorándolo a medida que avance el programa. Es de esperar que introduzca cambios en las diapositivas entre la primera y la segunda entrega y entre la segunda y la presentación. La fecha de entrega de la presentación final del proyecto y el código será el final de la semana 9.

Siga este calendario y asegúrese de incluir los elementos que se detallan a continuación en cada entrega:

Primera entrega – Final de la semana 3:

- Definir el problema de forma clara
- Describir el conjunto de datos que se va a utilizar para el análisis
- Detallar los procedimientos de limpieza de datos (**debe adjuntar un archivo con los datos limpios**)
- Sugerir un plan relativo a los métodos de análisis que se van a utilizar
- **Entrega:** Presentación de entre 3 y 5 diapositivas y muestra de datos limpios

Segunda entrega – Final de la semana 6:

- Calcular las estadísticas descriptivas
- Trazar y visualizar los datos
- Resultados de modelos relevantes de clusterización o reducción de dimensiones

- Resumen de los hallazgos obtenidos a partir del análisis de los datos y sugerencias para realizar otros análisis mediante modelado supervisado
- **Entrega:** Presentación que incluye las diapositivas de la primera entrega y entre 3 y 5 diapositivas originales

Presentación del proyecto – Final de la semana 9:

- Resultados de los modelos lineales, de análisis de la varianza o de regresión logística relevantes
- Resultados de cualquier otro modelo relevante
- Conclusiones e información útil relativa al problema obtenidas del análisis
- Sugerencias de medidas prácticas para abordar el problema basadas en su análisis
- Ideas sobre las posibles mejoras y los siguientes pasos necesarios para el proyecto
- **Entrega:** Presentación completa y archive de código.

Calificación

El proyecto final representa un 20 % de la calificación total del programa. Los criterios de evaluación son los siguientes:

Tema	Porcentaje de la calificación
Primera entrega	
Entrega de la actividad correspondiente a la semana 3	10 %
Segunda entrega	
Entrega de la actividad correspondiente a la semana 6	10 %
Presentación del proyecto final	
Definición del problema y presentación de los datos	10 %
Presentación de los métodos y resultados del análisis no supervisado	20 %
Presentación de los métodos y resultados del análisis supervisado	20 %
Interpretación, conclusiones y capacidad general de comprender y evaluar modelos	30 %
TOTAL	100 %

Datos

Para este proyecto, puede utilizar cualquier conjunto de datos que le resulte interesante. Hay muchos conjuntos de datos integrados disponibles en R. La web

<https://www.kaggle.com/datasets> es un excelente recurso para encontrar datos que modelar. Puede utilizar sus propios datos o cualquier conjunto que le interese. Aunque recomendamos buscar conjuntos de datos de cierta complejidad, tenga en cuenta que, al terminar la tercera semana del programa, debería contar con un conjunto de datos limpio y bien estructurado con el que trabajar.

Código

Además de la presentación final, deberá entregar un archivo que contenga el código que haya utilizado para crear los modelos y ver los resultados. Puede ser un archivo .R, un archivo .Rmd o un archivo R Markdown generado con la opción "knit" en Word o PDF. Aunque este archivo no se calificará, los proyectos que no incluyan un código correcto que pueda generar los resultados presentados se considerarán incompletos y no recibirán la calificación correspondiente.