

A collection of approximately 15 squares in light blue, medium blue, and grey, arranged in a sparse, abstract pattern across the top half of the slide.

# MUBD

Màster Universitari en Enginyeria de Dades Massives (Big Data)

Estadística

# Estadística

## Ejercicio final

- Accede al dataset de “Census Income” de UCI-MLR.
- Realiza un análisis exploratorio detallado de los datos
- Plantea los aspectos necesarios de limpieza de datos
- Aplica diferentes algoritmos de clustering y comenta los resultados
- Realiza un modelo baseline y evalúa su performance
- Realiza varios modelos predictivos para mejorar el anterior en función de las características del dataset, evalúa los modelos como mejor consideres
- Obtén conclusiones de los resultados
- Plantea una visualización más avanzada de los resultados
- ¿Cómo podemos reducir el número de variables? Plantéalo como solución alternativa
- Crea un informe con todos los puntos anteriores y envíalo a [jordi.escayola@upc.edu](mailto:jordi.escayola@upc.edu)

A collection of approximately 15 squares in light blue, medium blue, and grey, arranged in a sparse, abstract pattern across the top half of the slide.

# MUBD

Màster Universitari en Enginyeria de Dades Massives (Big Data)

Estadística