1. Realizar un análisis exploratorio de los datos. Entre otras cosas, deben analizar la cantidad de datos, cantidad y tipos de atributos, cantidad de clases de la variable de interés (letras en lenguaje de señas) y otras características que consideren relevantes. Además se espera que con su análisis puedan responder las siguientes preguntas:

a. ¿Cuáles parecen ser atributos relevantes para predecir la letra a la que corresponde la seña? ¿Cuáles no? ¿Creen que se pueden descartar atributos?

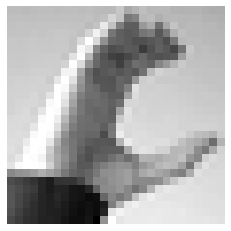
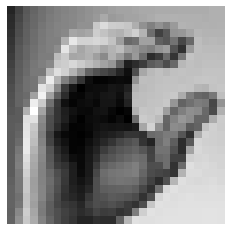
El contraste de la mano con el fondo, la forma de los dedos, dónde se encuentra la punta de los dedos. La rotación de la muñeca y la extensión de los dedos. No es relevante el fondo y las mangas de los -modelos-. Si se descarta el fondo, no podemos detectar el contraste con los dedos.

b. ¿Hay señas que son parecidas entre sí? Por ejemplo, ¿Qué es más fácil de diferenciar: la seña de la E de la seña de la L o la seña de la E de la seña de la M?

Sí, las letras más parecidas son las de puño cerrado, que son la A, E, M, N, S (que en particular la A y la S son casi indistinguibles). La L en cambio se hace con un ángulo recto entre el pulgar y el índice, que son muy distintas.

c. Tomen una de las clases, por ejemplo la seña correspondiente a la C, ¿Son todas las imágenes muy similares entre sí?

No mucho, si tomamos estos dos ejemplos, la primera es mucho más oscura en el centro y la separación entre los dedos es más pequeña, mientras que la segunda tiene tonos más claros y el índice con el pulgar están mucho más distanciados.



d. Este dataset está compuesto por imágenes, esto plantea una diferencia frente a los datos que utilizamos en las clases (por ejemplo, el dataset de Titanic). ¿Creen que esto complica la exploración de los datos?

Ahora lo que tenemos son 784 atributos numéricos en lugar de los 7 atributos string que tenía el de Titanic. Quizás vuelve más costosa a la exploración de datos por el hecho de que entrenar un modelo (por ejemplo con un árbol de decisión) puede necesitar más iteraciones (profundidad del árbol) para decidir correctamente de qué letra de trata. Fuera de eso no parece haber una complicación tan grande.

Importante: las respuestas correspondientes a los puntos 1.a, 1.b y 1.c deben ser justificadas en base a gráficos de distinto tipo.