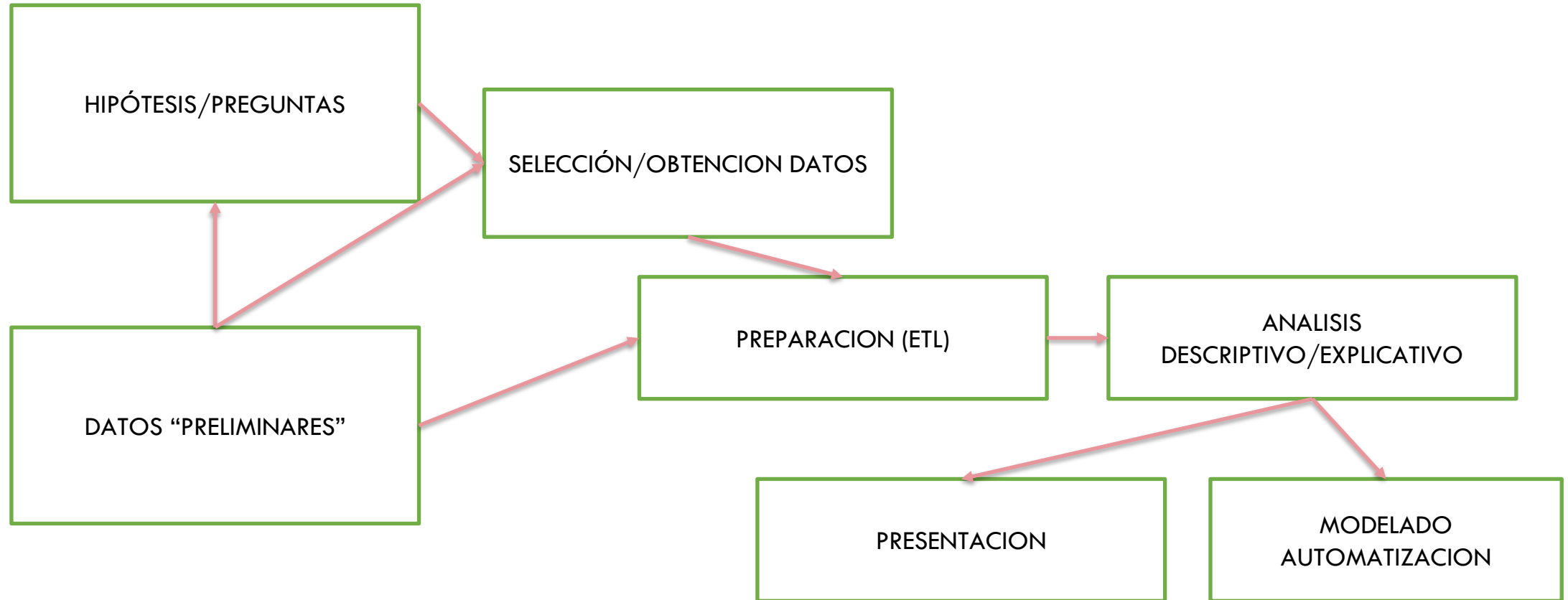




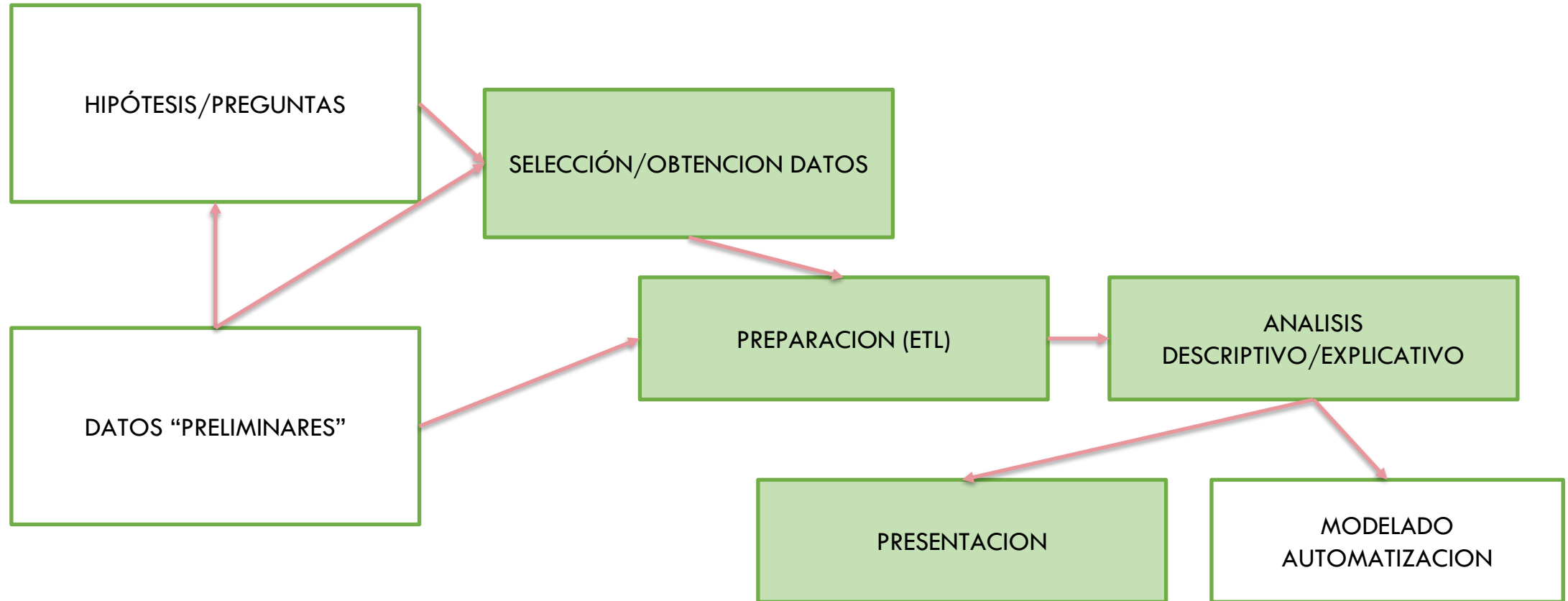
Análisis de Datos

Introducción al EDA

EDA, versión J.



EDA, versión J.



EDA, desde otra perspectiva



Hipótesis

Cuestiones de negocio como puntos débiles, oportunidades



Estructura

Cómo son tus datos, de qué tipo son, calidad del dato



Limpieza

Datos duplicados, missings, features correlacionadas.



Outliers

Técnicas de detección e imputación de outliers



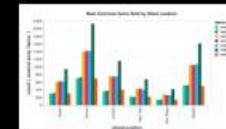
Feature Engineering

Transformación de variables y creación de nuevas a partir de las originales



Relación entre datos

Qué variables están correlacionadas



Visualización

Representación de relaciones y conclusiones sacadas de los datos.

EDA, desde otra perspectiva



¿Cuándo se hace un EDA?

- Análisis de datos de la empresa/administración/institución para entender el negocio/situación/problemática y los clientes/ciudadanos/contextos.
- Análisis de la calidad de los datos
- Análisis de ventas y de campañas de marketing
- Preparamos los datos para realizar un modelo de machine learning

¿Por dónde empieza(empiezo) un EDA?

- Por las hipótesis o por los datos
- Ejemplos de preguntas/analisis:

- Analizar un producto: ¿Ha tenido éxito? ¿Por qué? ¿Qué falla? ¿Qué funciona? ¿En qué público funciona?
- Analizar una situación: ¿Hay cambio climático? ¿Es lo mismo que el calentamiento global? ¿Dónde se refleja?
- Analizar un problema: ¿Hay discriminación en mi empresa?
- Analizar una estrategia/moneyball: ¿Es posible predecir quién ganará la liga? ¿un partido? ¿varios? ¿la quiénela?....

¿Por dónde empieza(empiezo) un EDA?

- Por las hipótesis o por los datos
- Ejemplos de datasets de partida

- Dataset de películas de IMDB, explorar y ver qué puede ser interesante
- Dataset de cotización de bitcoin: ¿se puede saber cuándo sube o baja?
- Datasets del INE: ¿quién ganó las elecciones? ¿cómo ha cambiado el IPC? ¿cómo se distribuyen los nombres en España?
- Etc,etc

¿Por dónde empieza(empiezo) un EDA?

- Por las hipótesis o por los datos
- Si ya tienes una idea, consulta con los TA o con el Lead Instructor para saber cómo conseguir datos o cómo desarrollar las hipótesis
- Presta atención a las guías que os iremos dando a lo largo de las sesiones en vivo

