

Taller limpieza y exploración

Programa en Analítica

Curso *Capstone* - Inteligencia Artificial & *Deep Learning*

Analítica Prescriptiva

Educación continua | Universidad de los Andes
octubre 6 – noviembre 24
2021

Hoy

¿Qué haremos hoy?

1. Comentario metodología

2. Manos al código



Photo by [nappy](#) from [Pexels](#)

El comentario

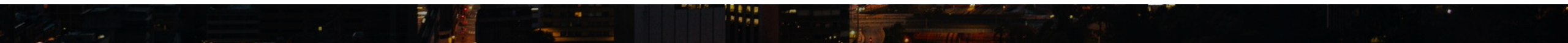
5 minutos



Photo by [nappy](#) from [Pexels](#)

El problema: puede haber preguntas

1. ¿Qué decisiones vamos a apoyar?
2. ¿Cuáles son las opciones que tiene ese/a tomador de decisión?
3. ¿Qué implicaciones trae tomar “mejores decisiones”?
4. ¿Qué criterio de decisión puede ser bueno para escoger entre las opciones? (“escogemos la decisión que maximiza/minimiza tal cosa”)
5. ¿Qué restricciones tenemos para escoger la mejor decisión posible?



El problema: puede haber preguntas

1. ¿Qué decisiones vamos a apoyar?
2. ¿Cuáles son las opciones que tiene ese/a tomador de decisión?
3. ¿Qué implicaciones trae tomar “mejores decisiones”?
4. ¿Qué criterio de decisión puede ser bueno para escoger entre las opciones? (“escogemos la decisión que maximiza/minimiza tal cosa”)
5. ¿Qué restricciones tenemos para escoger la mejor decisión posible?

Noten: algunas son respuestas muy evidentes para algunos grupos, o algunas no están definidas en esta etapa del proyecto. Está bien. Lo importante es ser consciente de estas preguntas.

Limpieza & exploración

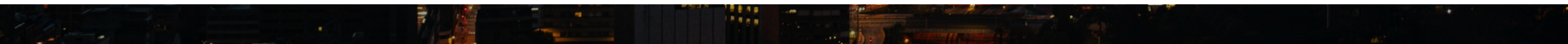
1. Carga de datos a R: ¿qué estamos viendo?
¿Qué significa mi unidad de observación?
2. Limpieza de forma:
 1. Nombres de las columnas
 2. Tipos de datos está bien
3. Limpieza de calidad:
 1. ¿Hay observaciones atípicas? ¿Por qué pueden ser?
 2. ¿Las descriptivas de las variables tienen sentido? ¿Hay errores de diligenciamiento en la base de datos?
4. Exploraciones descriptivas
 1. Caracterizar las columnas: composición de las observaciones en grupos, medias, frecuencias, mínimos, máximos.
 2. ¿Qué columnas son más útiles?: correlaciones entre variables, diagramas de dispersión, correlogramas.
 3. Piensen qué descriptivas pueden ser útiles para su problema particular: que informen la técnica, o que informen el problema.

NOTA: Ojo! Vamos a tratar de ser concretos/as, si una variable requiere tratamiento manual por 3 horas para estar lista, ¿Es realmente importante?

¿Cómo funciona? Imagínense que es MasterChef

1. Conversación en el grupo: qué tenemos que hacer, ¿cómo nos organizamos? ¿Tratamos de cargar cada uno/a los datos y luego...?
2. *Brainstorm*: Lista **concreta** de cosas que vamos a intentar (usen papel y lápiz si les es más cómodo)
3. Manos al código: ejecutamos tareas individuales, nos apoyamos grupalmente, **surgen preguntas** vamos consultando en internet.
4. Reporte: cómo nos fue, ¿encontramos hallazgos / resultados notables?

Alfredo va pasando, dando tips, facilitando caminos enredados...



¡Manos al código!

A los grupos... Recomendable: descarguen esta presentación del repositorio.



Photo by [Andrea Piacquadio](#) from [Pexels](#)