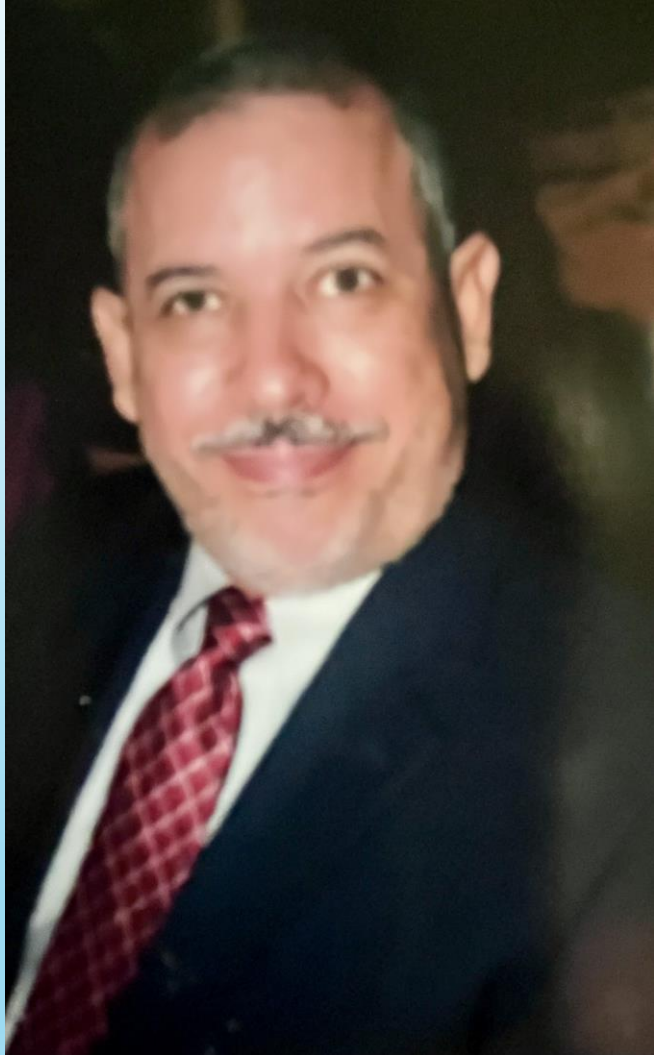




TEORÍAS Y TÉCNICAS PARA IMPUTACIÓN DE DATA FALTANTES

Facilitador: Ricardo J. Serrano



CV

- Ingeniero industrial/evaluador de propiedad inmueble
- Adventurous Analytics – analista de negocios y científico de datos
- Co-organizador de OMLDS (Orlando Machine Learning and Data Science) meetup
- Hobbies: ciencia ficción, música, historia

AGENDA

- Definición de datos faltantes (“missing data”)
- Mecanismos que explican la procedencia e interacción de los datos faltantes con el resto de los datos (MCAR, MAR, MNAR)
- Técnicas para lidiar con datos faltantes (remoción, imputación)
- Utilización de R/Python/Power BI para remover/imputar datos faltantes
- Conclusiones



DEFINICIÓN



- Datos faltantes

“valores no disponibles que serían útiles o significativos para el análisis de los resultados”

- Causas
 - errores en la entrada de datos
 - desperfecto de sensores
 - omisiones intencionales (ej. encuestas)

MECANISMOS

- MCAR (completamente al azar)
 - la falta de una observación **no** está relacionada con los valores faltantes o existentes
 - Ejemplos:
 - fallas ocasionales de equipos que impiden hacer una medición
 - olvido ocasional en registrar un dato
 - encargado de hacer la medición se enfermó
 - 'mcar_test' (R naniar)

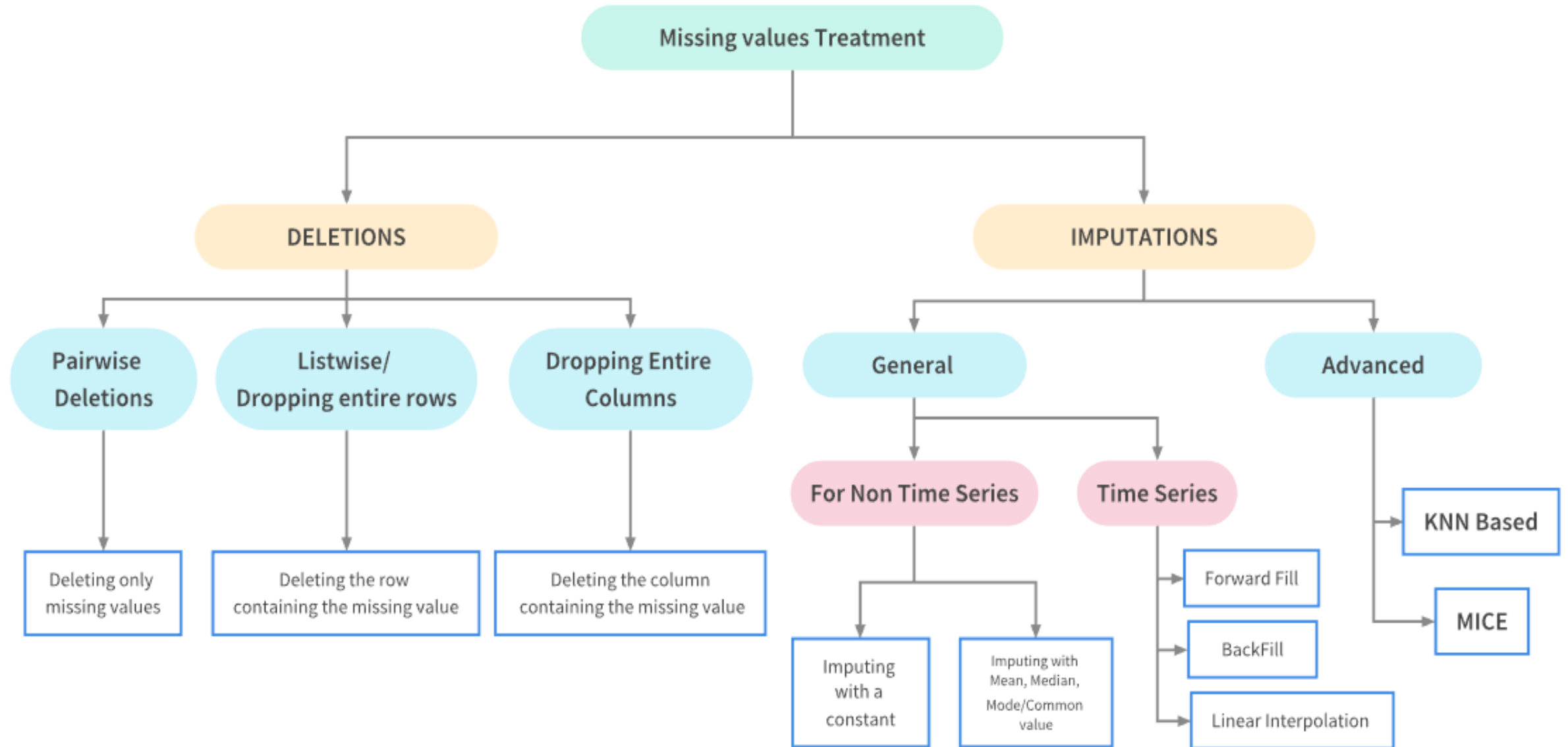
MECANISMOS

- MAR (al azar)
 - la falta de una observación **está** relacionada con los valores existentes
 - Ejemplos:
 - el nivel de respuestas faltantes en una encuesta está relacionado con el nivel socio-económico
 - un sensor de temperatura requiere ser apagado por mantenimiento preventivo. Debido a que el itinerario de mantenimiento ocurre los días laborales, existe una relación entre los datos faltantes y los días de la semana.
 - Visualización entre los datos faltantes y los existentes

MECANISMOS

- MNAR (no al azar)
 - los datos faltantes probablemente dependen o están relacionados con datos **no observados**
 - Ejemplos:
 - paciente decide no continuar el ensayo clínico debido a que su salud se deterioró durante el término del ensayo
 - un sensor de temperatura no registra cuando la temperatura desciende y congela el sensor, lo que resulta en datos faltantes. En este caso, los datos faltantes están relacionados con la capacidad física del sensor.
 - Visualización entre los datos faltantes y los existentes

REMEDIOS/ALTERNATIVAS





DEMO

RECURSOS/CONTACTO

- GitHub repo
https://github.com/rserran/power_bi_panama_ug
- OMLDS
<https://www.meetup.com/orlando-mlds/members/>
- R4DS Community Slack
<http://r4ds.io/join>
- LinkedIn
<https://www.linkedin.com/in/ricardojserrano/>
- Mastodon
<https://techhub.social/@ricsam>