**Mentoría Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus aplicaciones**

**“Modelo de recomendación de profesionales médicos para una mejor experiencia de usuarios”**

**Mentor:**

Juan Manuel Majul.

**Equipo:**

Arrua, Esteban.

Galarza, Alexis.

Valeria, Silvina.

Práctico n° 2



**Exploración y curación de datos**

**Consignas**

1. Generar tabla\_consumos\_consolidada.

El objetivo es generar un dataset de consumos con todos los datos disponibles de pacientes y centros médicos. Para ello, resultará necesario un join de las tablas consumos, pacientes y centros médicos.

1. Breve análisis exploratorio de aquellas variables no analizadas en el práctico 1.
2. Feature engineering.

1. Aplicar técnicas de imputación en datos faltantes, erróneos y/o atípicos.

2. Variables numéricas:

Aplicar transformación (min-max scaler, max-abs scaler, standarization).

3. Variables categóricas:

Crear nuevas categorías agrupando categorías que se encuentren aperturadas (ejemplo segmento empresas - segmento individuos).

Crear nuevas categorías agrupando en “Otros” a aquellos valores que presenten escasa cantidad de observaciones.

Aplicar one-hot encoding.

4. Aplicar transformaciones que resulten convenientes a partir de las conclusiones obtenidas en el análisis exploratorio.