

### **Taller Preparación de Datos**

**Objetivo:** Aplicar los macro-procesos típicos durante la preparación de datos antes de aplicar técnicas de analítica.

#### **Etapa 1 Carga y Exploración**

1. Seleccionar un conjunto de datos del portal <https://www.datos.gov.co/> . Para esta selección considere que el conjunto de datos contenga atributos continuos y discretos y que contengan al menos 10 atributos. Ejemplo: Procesos de compra pública registrados en la plataforma SECOP I
2. Realizar la carga del archivo en la herramienta seleccionada
3. [35%] Realizar el análisis exploratorio de los datos

**Mínimo** incluir:

- Tabla resumen
- Métricas de tendencia central
- Métricas de dispersión
- Visualización de métricas de dispersión
- Análisis de las visualizaciones
- Análisis de Correlación
- Recuerde incluir análisis de todo lo visualizado y no solo el gráfico

#### **Etapa 2 Limpieza de datos**

4. [35%] Realizar el análisis y limpieza de valores faltantes e inconsistentes
- Mínimo incluir:

- Reconocimiento y tratamiento de atributos con valores únicos o distintos
- Reconocimiento y tratamiento de atributos con valores faltantes
- Reconocimiento y tratamiento de atributos con valores atípicos
- Reconocimiento y tratamiento de registros atípicos

- Reconocimiento y tratamiento de atributos redundantes

5. Análisis

A partir de los datos genere conclusiones válidas y útiles que puedan contribuir a nuestro país. Genere una infografía.

**Etapas 3 Creación de la vista minable**

6. [15%] Construir la vista minable.

Mínimo incluir:

- a. Generación de variables derivadas tipo 1 y 2
- b. Normalización de al menos un atributo
- c. Discretización de al menos un atributo
- d. Numerización 1 a n de al menos un atributo

7. Complemente la infografía, si obtuvo nuevo conocimiento a partir de los datos.

**Etapas 4 Presentación de infografía en clase**

8. [15%] Realizar una presentación en grupo de máximo 5 minutos, presentando los principales hallazgos de cada etapa anterior usando la infografía.

**Qué se debe entregar:**

Reporte con análisis pedidos, transformaciones, archivos fuente usados para ejecutar cada fase y una página con la Infografía. Incluya en su reporte elementos básicos para comprensión de un documento como portada, introducción, conclusiones del trabajo realizado. En cada etapa incluya una explicación de lo realizado.

**Grupos:** 5 integrantes.