**BigData**

**EIS**

**Yurany Cortes 865100603**

**Fernanda Otalora 865100607**

**Propuesta de proyecto Big Data Analytics**

1. **Identificar fuente de datos, tipos de datos, formato de los datos, características y ejemplos.**
   1. **Fuente datos:** Base de datos de los usuarios de los usuarios de los servicios públicos de agua y luz del municipio de Villavicencio.
   2. **Tipos de datos:**

* **Numericos:** cuenta,lectura,consumo, longitud, latitud, anomalía
* **Texto:** nombre,direccion,observación
* **Date:** fecha\_lectura\_actual, fecha\_lectura\_anterior

**Formato de los datos:** Base de Datos estructurada (Postgres)

**Características:** Datos básicos de usuarios, ubicación GPS, consumos mensuales.

**Ejemplos:** servicio de agua: numero de cuenta asignado al usuario, nombre, dirección, datos del equipo de medida, consumos.



**2. Descripción de la motivación, problema y justificación.**

En el trabajo realizado se cuenta con los datos reales de los consumos de los servicios de agua y luz en la ciudad de Villavicencio, es notable que esta información es valiosa y puede explotarse para generar campañas de concienciación para incentivar el consumo responsable de estos servicios, a través de la identificación de los perfiles de los usuarios a partir de la información de histórico de los consumos.

**3. Objetivo y alcance.**

Identificar las características de los usuarios en cuanto a sus consumos de agua y luz para generar campañas dirigidas al cuidado del medio ambiente y el ahorro, a través de notificaciones o mensajes individualizados según el perfil definido.

**4. Esquema del sistema propuesto**

BD Consumos PostgreSQL

Po

Generar archivos planos

Definir los centroides

Procesar información KMeans

Determinar cantidad de perfiles

Calcular la distancia de los objetos con respecto a cada centroide

Determinar a qué grupo pertenece cada objeto



Perfiles de usuarios

Archivos planos consumos

Descripción de perfiles de usuario:

* Usuarios derrochadores: 60% o más de consumos anuales el valor de la crítica calculada es excesivo o alto, siempre y cuando no presente visita por desviación.
* Usuarios ahorradores: 60% o más de consumos anuales el valor de la crítica calculada es normal o bajo, siempre y cuando no presente anomalías fraudulentas o visitas por desviación.
* Usuarios variables: usuarios cuyos consumos no se pueden ubicar dentro de los porcentajes definidos anteriormente, ya que varía mucho la critica.