**Avance propuesta proyecto final (semana 3)**

**Grupo 4**

Iván Camilo Barriga Gómez Código: 202121106

Andrés Felipe Arteta Isaacs. Código: 201124652

David Santiago Muñoz Carrillo. Código: 202121443

Edison Arcángel Giraldo Martínez. Código: 202124624

1. **Elegir una de las preguntas utilizando la retroalimentación recibida en el avance anterior y expandir sobre su motivación a responderla.**

**Monitoreo de Red de telefonía celular**

La red de telefonía móvil celular se encuentra en constante cambio principalmente por los trabajos de expansión de infraestructura, optimización de la red y cambios en la demanda de servicios de voz y datos, por lo que realizar labores de monitoreo y control son fundamentales para garantizar el correcto funcionamiento y aseguramiento de la calidad del servicio por parte del operador.

La aplicación que proponemos para estos datos es el descubrimiento o segmentación de grupos de celdas basado en tres agrupadores de información disponible: Identificación de la celda, Parametrización de la celda e Indicadores de desempeño.

**Motivación**: La red LTE tiene más de 70mil celdas para dar cobertura a nivel nacional y por lo tanto se requieren herramientas que puedan procesar y generar información valiosa para los ingenieros de radiofrecuencia, que les permita trabajar de manera más eficiente. Las tareas de auditoría de la red son un requerimiento constante para mantener identificado los cambios en valores atípicos, tanto de parametrización como de indicadores de desempeño.

1. **Describir si el problema pertenece a una tarea de reducción de dimensión, clustering, o una combinación de los dos, explicando el por qué.**

Análisis de Clustering porque con más de 70mil registros y 12 variables de diferente naturaleza, se quieren encontrar patrones que puedan responder preguntas como:

¿Existe algún patrón en las celdas con mayor velocidad de navegación? ¿y con menor velocidad?

¿Qué tanto influye la disponibilidad de las celdas en la velocidad? ¿y en las llamadas exitosas?

¿Existe algún patrón en las celdas con mayor porcentaje de llamadas exitosas? ¿y con menor porcentaje?

¿Existen celdas atípicas en cuanto a parámetros de identificación?

¿Existen celdas atípicas en cuanto a parámetros de configuración?

¿Cómo es el desempeño (velocidad, disponibilidad y llamadas exitosas) en estas celdas atípicas?

1. **Proponer al menos un algoritmo/técnica de aprendizaje no supervisado que ustedes crean que es la adecuada para responder la pregunta.**

k-medoides, para establecer grupos de K clústeres que representen la población con diferentes características según la información disponible. Esto nos podría proporcionar insights sobre la relación entre las características propias de los registros y las variables que describen resultados o características asociadas al desempeño del registro.

1. **Presentar estadísticas descriptivas utilizando tablas y/o visualizaciones de los datos crudos que tengan a su disposición. Describir el plan que van a seguir para tener los datos listos. Por ejemplo, cómo van a limpiarlos, que van a hacer con los datos faltantes, etc.**

Origen de la data cruda:

* Base de datos privada en donde se puede extraer la información de indicadores de desempeño de la red móvil a través de consultas SQL. (Información solicitada en proceso de extracción: 3 variables).
* Repositorio privado con información de identificación y de parametrización. (Información obtenida:10 Variables).

Estas dos fuentes de información no tienen barreras, sin embargo, al ser información privada debe ser utilizada únicamente para el ejercicio académico y sin compartir información sensible.

Información de indicadores de desempeño:

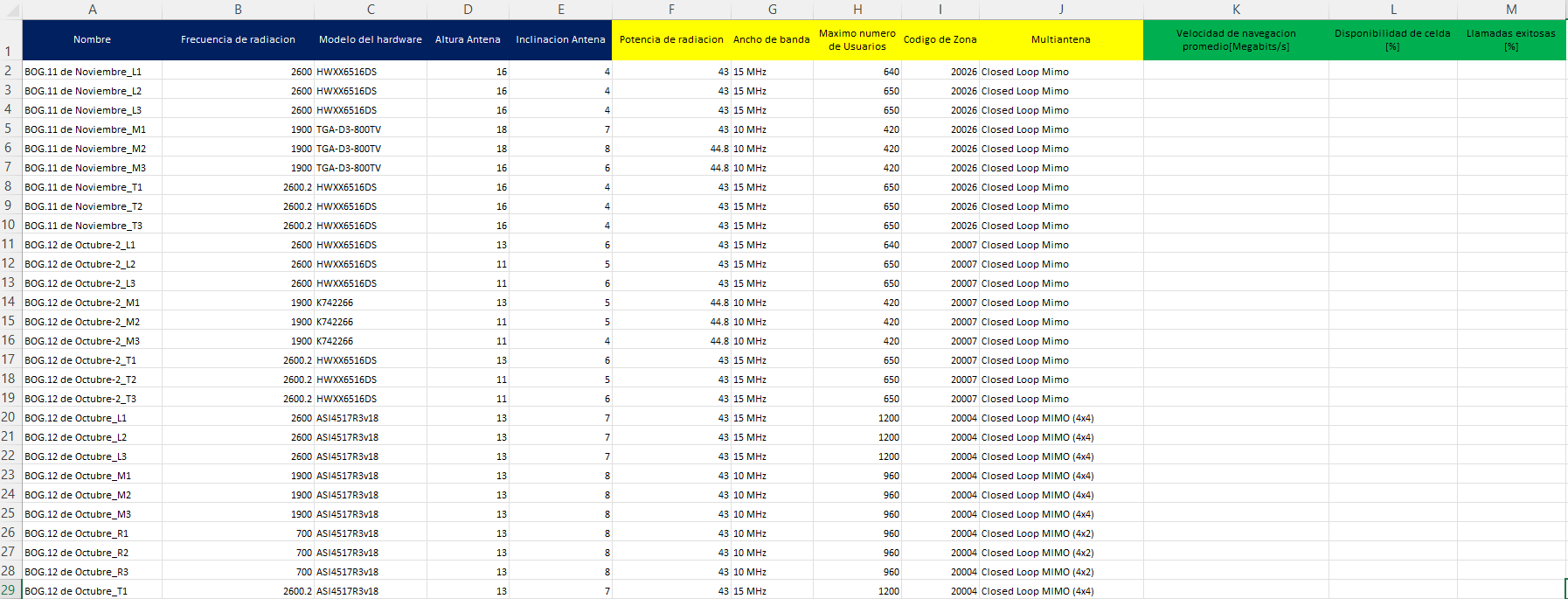
* Velocidad de navegación promedio [Megabits/s]
* Disponibilidad de celda [%]
* Llamadas exitosas [%]

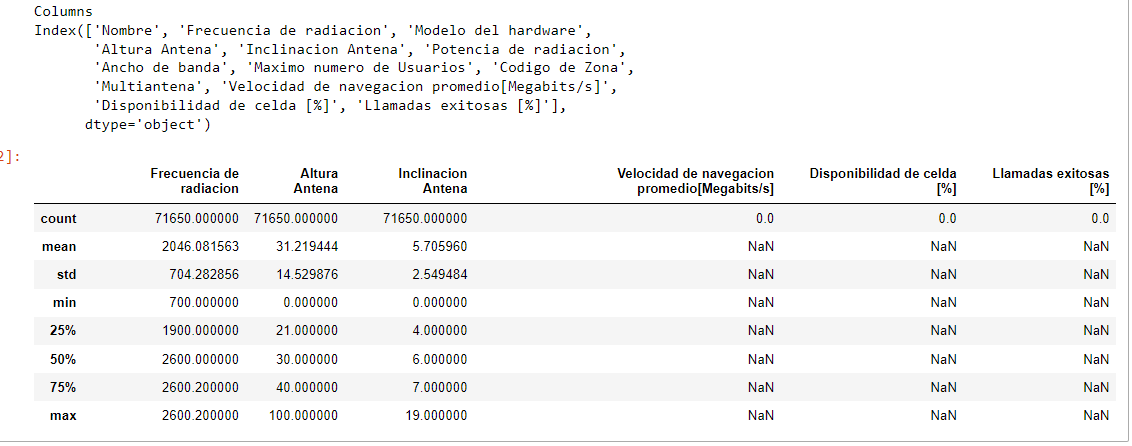
Información de parametrización de la celda:

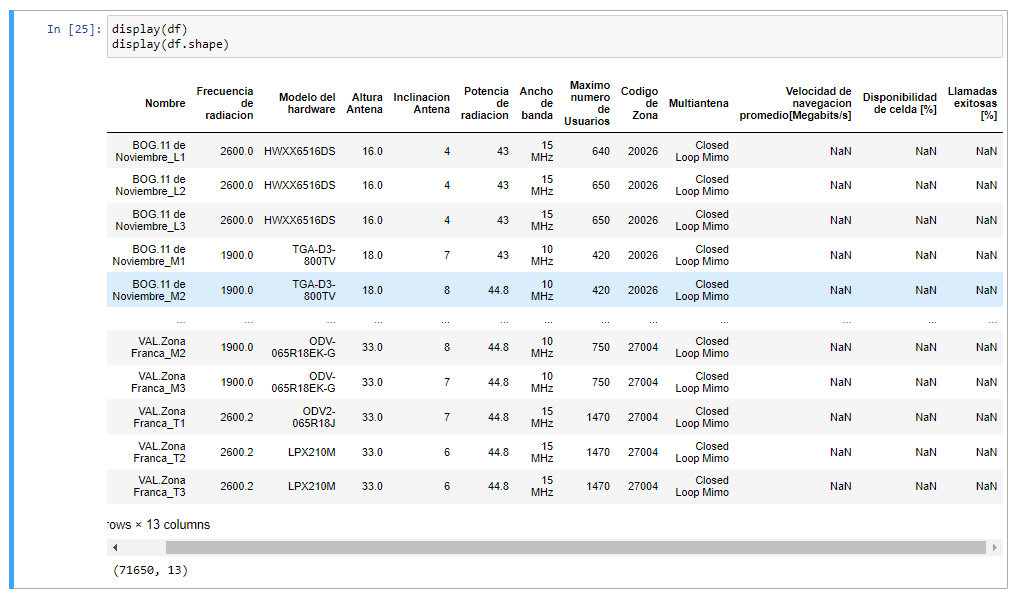
* Potencia de radiación
* Ancho de banda
* Máximo número de Usuarios
* Código de Zona
* Multiantena

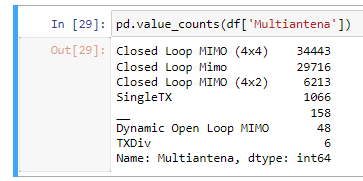
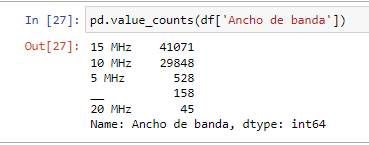
Información de Identificación de la celda:

* Nombre
* Frecuencia de radiación
* Modelo del hardware
* Altura Antena
* Inclinación Antena









Se debe realizar proceso de limpieza de datos faltantes sin imputación, ya que el objetivo es mantener sin alteración los registros existentes.

1. **Con base en los roles definidos en el documento de la semana anterior, delinear las actividades que se llevaron y llevarán a cabo para la primera entrega calificada del proyecto. Ser preciso, y tomar este punto como un contrato entre los miembros del equipo.**

* Exploración de datos, preprocesamiento: Iván Barriga y Edison Giraldo
* Conceptualización del modelo a implementar: Todos
* Modelación e implementación en Python: Andrés Arteta y Santiago Muñoz
* Análisis de resultados: Todos

Referencias

[Colombia, Julio 2022, Experiencia de la Red Móvil Report | Opensignal](https://www.opensignal.com/es/reports/2022/07/colombia/mobile-network-experience)

[¿Qué es LTE? Diferencias entre LTE y 4G // SoftwareLab](https://softwarelab.org/es/lte-4g/#:~:text=En%20concreto%2C%20el%20LTE%20permite,y%20cuatro%20en%20el%20equipo.)

[What is a Radio Access Network (RAN)? (techtarget.com)](https://www.techtarget.com/searchnetworking/definition/radio-access-network-RAN)

<https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Territorial/Guia-Corta-Metodlogia-Medicion-Nuevo-IDF.pdf>