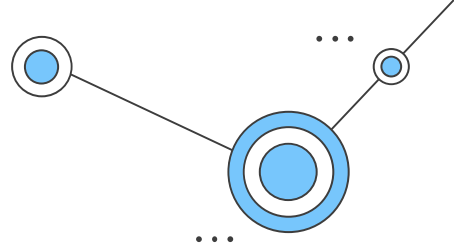


Informe EDA

Análisis
Exploratorio
de
Datos

Introducción

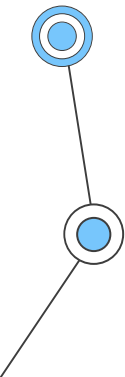


Para comenzar con esta [Análisis Exploratorio de Datos \(EDA\)](#), obtendremos información a partir de datasets publicados en la página oficial del [Ente Nacional de Comunicaciones de la Republica Argentina \(ENACOM\)](#).

Los datasets analizados son los siguientes:

- Internet_Penetracion
- Internet_Penetracion(1)
- Internet_BAF_prov
- Internet_BAF_nacional
- Velocidad_nacional
- Velocidad_prov
- Accesos_tecno_prov
- ConectividadalserviciodelInternet

Se realiza la descarga de los mismos en formato “.csv” para luego importarlos en la herramienta [PowerBI](#).



Generalidades

01

...

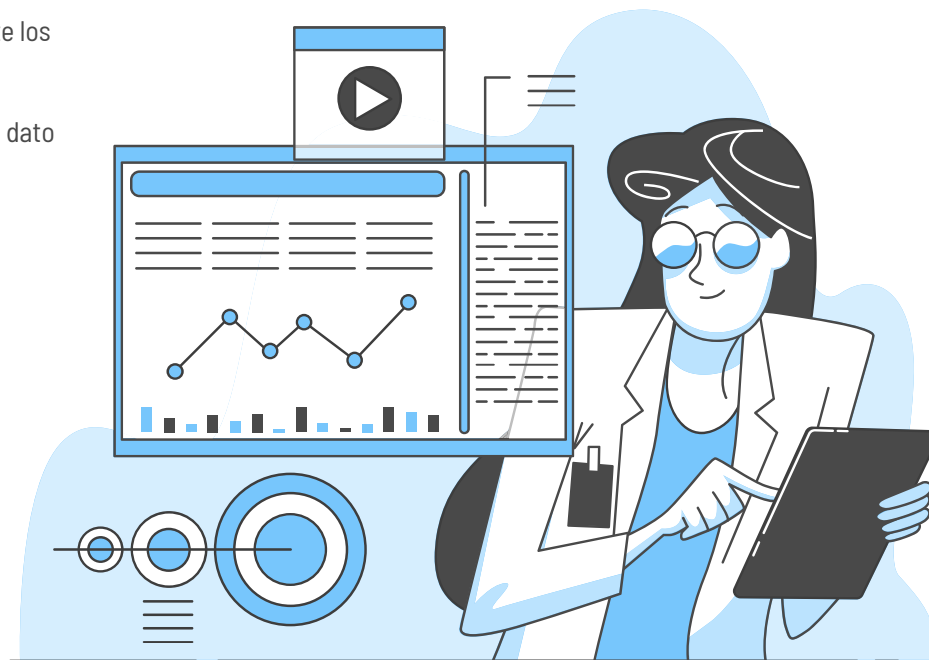
PowerBI realiza por defecto 3 operaciones para cada tabla:

- Origen: facilita la obtención de los datos y los convierte a tipo tabla.
- Encabezados promovidos: determina automáticamente los encabezados de las tablas a partir de la primera fila del dataset.
- Tipo cambiado: determina automáticamente el tipo de dato de cada columna

02

...

En *PowerQuery* activo en el apartado "Vista" la opción "Calidad de columnas" para poder analizar de manera rápida y optima los valores de cada una de las tablas. Dicha característica refleja si existen datos Validos, con Error o Vacíos y a partir de estos comenzar a corregir los que sean necesarios. También activo "Distribución de columnas" para analizar a grandes rasgos la tendencia de los datos de las tablas.



Particularidades

01

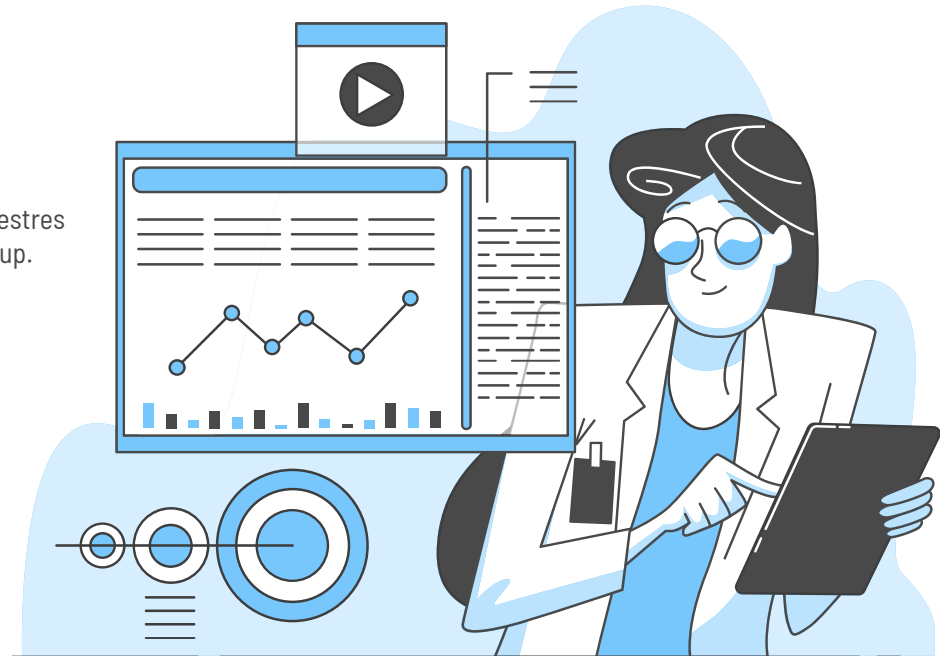
De los datasets Internet_Penetracion e Internet_Penetracion(1) elimino las columnas Periodo, ya que considero que las mismas poseen datos redundantes.

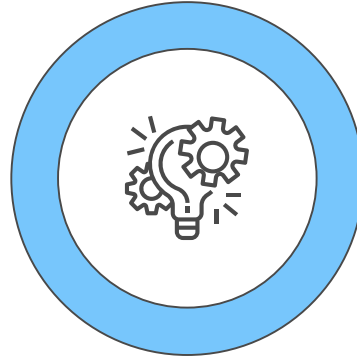
02

En la tabla Internet_BAF_prov en la columna "dialup" reemplazo los vacíos por 0 luego de determinar que los valores vacíos pertenecían a provincias que en los trimestres anteriores y posteriores no contaban con conexión dialup.

03

En la tabla Accesos_tecno_prov en las columnas Año y Trimestre encontré valores con errores (tenían el símbolo * a la derecha del número). Esto impedía convertir correctamente toda la columna a tipo numérico. Para resolverlo, antes de convertir el tipo de dato de la columna, debo reemplazar dichos valores (que pertenecían a los trimestres 1,2 y 3 del año 2019) eliminando los "*". Para ello utilice la función RECORTAR de PowerQuery.





CONCLUSIÓN

Las tablas seleccionadas y normalizadas para ser utilizadas en el análisis son las siguientes:

- Internet_Penetracion (renombrada como Accesos_prov)
- Internet_Penetracion(1) (renombrada como Accesos_nacional)
- Internet_BAF_prov
- Internet_BAF_nacional
- Accesos_tecno_prov
- Velocidad_nacional