Análisis Exploratorio del Dataset: Internet Fijo Accesos por tecnología y segmento - Elaborado por Mónica L. Dorado

https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Internet-Fijo-Accesos-por-tecnolog-a-y-segmento/n48w-gutb

Primer paso: En primer lugar debemos instalar las siguientes librerías: Pandas: Librería para la manipulación y visualización de grandes volúmenes de datos. Numpy: Sirve para trabajar funciones matemáticas algebráicas. Matplotlib: permite crear y personalizar los tipos de gráfico. Seaborn: proporciona varias funciones para personalizar los gráficos. Plotly: Para crear gráficos dinámicos

In [62]:

```
# Importamos las librerías

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Segundo paso: Ahora cargamos el dataset con el que vamos a trabajar y con el parámetro index_col=0 le indicamos que la primer columna tiene los nombres de c/u de las filas

In [58]:

```
df=pd.read_csv('Internet_Fijo_Accesos_por_tecnolog_a_y_segmento.csv', index_col=0)
df
```

Out[58]:

	TRIMESTRE	PROVEEDOR	COD_DEPARTAMENTO	DEPARTAMENTO	COD_MUNICIPIO	MUNICIPIO	SEGMENTO	TECNOLOGIA	VE
AÑO									
2021	3	DIRECTV COLOMBIA LTDA	52	NARI�O	52835	SAN ANDRES DE TUMACO	RESIDENCIAL - ESTRATO 1	OTRAS TECNOLOG�AS INAL�MBRICAS	
2021	3	CABLEMAS S.A.S	25	CUNDINAMARCA	25785	TABIO	RESIDENCIAL - ESTRATO 3	FIBER TO THE HOME (FTTH)	
2022	1	COLOMBIA TELECOMUNICACIONES S.A. E.S.P.	81	ARAUCA	81001	ARAUCA	RESIDENCIAL - ESTRATO 2	XDSL	
2021	3	COMUNICACION CELULAR S A COMCEL S A	23	CORDOBA	23001	MONTERIA	RESIDENCIAL - ESTRATO 4	CABLE	
2021	3	AZTECA COMUNICACIONES COLOMBIA S.A.S	50	META	50400	LEJANIAS	CORPORATIVO	FIBER TO THE HOME (FTTH)	
...	...	...	...	...	...	...	...	...	
2022	2	LEGON TELECOMUNICACIONES S.A.S.	66	RISARALDA	66400	LA VIRGINIA	RESIDENCIAL - ESTRATO 1	FIBER TO THE HOME (FTTH)	
2022	2	ESG COMUNICACIONES S.A.S	76	VALLE DEL CAUCA	76520	PALMIRA	RESIDENCIAL - ESTRATO 2	FIBER TO THE HOME (FTTH)	
2022	2	COMUNICACION CELULAR S A COMCEL S A	25	CUNDINAMARCA	25377	LA CALERA	CORPORATIVO	CABLE	
2022	2	UNE EPM TELECOMUNICACIONES S.A.	66	RISARALDA	66682	SANTA ROSA DE CABAL	RESIDENCIAL - ESTRATO 4	HYBRID FIBER COAXIAL (HFC)	
2022	2	COMUNICACION CELULAR S A COMCEL S A	76	VALLE DEL CAUCA	76001	CALI	CORPORATIVO	CABLE	

1181673 rows × 11 columns

In [20]:

```
df.shape
```

Out[20]:

```
(1181673, 12)
```

Aquí evidenciamos que tenemos 1181673 registros en 12 columnas

Ahora vamos a verificar los tipos de datos que tengo

In [22]:

```
df.dtypes
print('\nLos datos son de tipo:\n', df.dtypes)
```

```
Los datos son de tipo:
 AÑO                  int64
TRIMESTRE            int64
PROVEEDOR           object
COD_DEPARTAMENTO     int64
DEPARTAMENTO        object
COD_MUNICIPIO        int64
MUNICIPIO           object
SEGMENTO            object
TECNOLOGIA          object
VELOCIDAD_BAJADA     int64
VELOCIDAD_SUBIDA     int64
No DE ACCESOS        int64
dtype: object
```

In [23]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1181673 entries, 0 to 1181672
Data columns (total 12 columns):
 #   Column            Non-Null Count    Dtype 
---  ------            --------------    ----- 
 0   AÑO               1181673 non-null  int64 
 1   TRIMESTRE         1181673 non-null  int64 
 2   PROVEEDOR         1181673 non-null  object
 3   COD_DEPARTAMENTO  1181673 non-null  int64 
 4   DEPARTAMENTO      1181673 non-null  object
 5   COD_MUNICIPIO     1181673 non-null  int64 
 6   MUNICIPIO         1181673 non-null  object
 7   SEGMENTO          1181673 non-null  object
 8   TECNOLOGIA        1181673 non-null  object
 9   VELOCIDAD_BAJADA  1181673 non-null  int64 
 10  VELOCIDAD_SUBIDA  1181673 non-null  int64 
 11  No DE ACCESOS     1181673 non-null  int64 
dtypes: int64(7), object(5)
memory usage: 108.2+ MB
```

Con la función describe le pedimos que nos muestre la principales estadísticas tales como la media, la desviación estandar, el mínimo, el máximo y los cuartiles

In [59]:

```
df.describe()
```

Out[59]:

	TRIMESTRE	COD_DEPARTAMENTO	COD_MUNICIPIO	VELOCIDAD_BAJADA	VELOCIDAD_SUBIDA	No DE ACCESOS
count	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06
mean	2.440076e+00	3.731018e+01	3.764078e+04	1.037707e+02	7.406414e+01	7.984898e+01
std	1.085487e+00	2.653242e+01	2.651420e+04	5.928244e+03	5.869491e+03	8.947930e+02
min	1.000000e+00	5.000000e+00	5.001000e+03	0.000000e+00	0.000000e+00	0.000000e+00
25%	2.000000e+00	1.500000e+01	1.523800e+04	5.000000e+00	1.000000e+00	1.000000e+00
50%	2.000000e+00	2.500000e+01	2.573600e+04	1.000000e+01	3.000000e+00	3.000000e+00
75%	3.000000e+00	6.600000e+01	6.640000e+04	4.000000e+01	1.000000e+01	1.700000e+01
max	4.000000e+00	9.900000e+01	9.977300e+04	3.450300e+06	3.450300e+06	1.748250e+05

Revisemos cuantas filas tiene este dataset

In [51]:

```
len(df)
```

Out[51]:

```
1167924
```

Tiene 1167924 filas

Revisemos las primeras 5 filas

In [24]:

```
df.head(5)
```

Out[24]:

	AÑO	TRIMESTRE	PROVEEDOR	COD_DEPARTAMENTO	DEPARTAMENTO	COD_MUNICIPIO	MUNICIPIO	SEGMENTO	TECNOLOGIA
0	2021	3	DIRECTV COLOMBIA LTDA	52	NARI�O	52835	SAN ANDRES DE TUMACO	RESIDENCIAL - ESTRATO 1	OTRAS TECNOLOG�AS INAL�MBRICAS
1	2021	3	CABLEMAS S.A.S	25	CUNDINAMARCA	25785	TABIO	RESIDENCIAL - ESTRATO 3	FIBER TO THE HOME (FTTH)
2	2022	1	COLOMBIA TELECOMUNICACIONES S.A. E.S.P.	81	ARAUCA	81001	ARAUCA	RESIDENCIAL - ESTRATO 2	XDSL
3	2021	3	COMUNICACION CELULAR S A COMCEL S A	23	CORDOBA	23001	MONTERIA	RESIDENCIAL - ESTRATO 4	CABLE
4	2021	3	AZTECA COMUNICACIONES COLOMBIA S.A.S	50	META	50400	LEJANIAS	CORPORATIVO	FIBER TO THE HOME (FTTH)

In [25]:
```python
df.describe()
```

Out[25]:

	AÑO	TRIMESTRE	COD_DEPARTAMENTO	COD_MUNICIPIO	VELOCIDAD_BAJADA	VELOCIDAD_SUBIDA	No DE ACCESOS
count	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06	1.181673e+06
mean	2.020693e+03	2.440076e+00	3.731018e+01	3.764078e+04	1.037707e+02	7.406414e+01	7.984898e+01
std	9.809593e-01	1.085487e+00	2.653242e+01	2.651420e+04	5.928244e+03	5.869491e+03	8.947930e+02
min	2.019000e+03	1.000000e+00	5.000000e+00	5.001000e+03	0.000000e+00	0.000000e+00	0.000000e+00
25%	2.020000e+03	2.000000e+00	1.500000e+01	1.523800e+04	5.000000e+00	1.000000e+00	1.000000e+00
50%	2.021000e+03	2.000000e+00	2.500000e+01	2.573600e+04	1.000000e+01	3.000000e+00	3.000000e+00
75%	2.021000e+03	3.000000e+00	6.600000e+01	6.640000e+04	4.000000e+01	1.000000e+01	1.700000e+01
max	2.022000e+03	4.000000e+00	9.900000e+01	9.977300e+04	3.450300e+06	3.450300e+06	1.748250e+05

Ahora revisemos las últimas filas

In [26]:
```python
df.tail()
```

Out[26]:

	AÑO	TRIMESTRE	PROVEEDOR	COD_DEPARTAMENTO	DEPARTAMENTO	COD_MUNICIPIO	MUNICIPIO	SEGMENTO	TECNOLO
1181668	2022	2	LEGON TELECOMUNICACIONES S.A.S.	66	RISARALDA	66400	LA VIRGINIA	RESIDENCIAL - ESTRATO 1	FIBER TO HOME (FT
1181669	2022	2	ESG COMUNICACIONES S.A.S	76	VALLE DEL CAUCA	76520	PALMIRA	RESIDENCIAL - ESTRATO 2	FIBER TO HOME (FT
1181670	2022	2	COMUNICACION CELULAR S A COMCEL S A	25	CUNDINAMARCA	25377	LA CALERA	CORPORATIVO	CA
1181671	2022	2	UNE EPM TELECOMUNICACIONES S.A.	66	RISARALDA	66682	SANTA ROSA DE CABAL	RESIDENCIAL - ESTRATO 4	HYBRID FI COAX (H
1181672	2022	2	COMUNICACION CELULAR S A COMCEL S A	76	VALLE DEL CAUCA	76001	CALI	CORPORATIVO	CA

Tercer paso: Procederemos a borrar las filas con información nula

In [31]:
```python
df=df.dropna()
```

In [33]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1181673 entries, 0 to 1181672
Data columns (total 12 columns):
 #   Column            Non-Null Count    Dtype 
---  ------            --------------    ----- 
 0   AÑO               1181673 non-null  int64 
 1   TRIMESTRE         1181673 non-null  int64 
 2   PROVEEDOR         1181673 non-null  object
 3   COD_DEPARTAMENTO  1181673 non-null  int64 
 4   DEPARTAMENTO      1181673 non-null  object
 5   COD_MUNICIPIO     1181673 non-null  int64 
 6   MUNICIPIO         1181673 non-null  object
 7   SEGMENTO          1181673 non-null  object
 8   TECNOLOGIA        1181673 non-null  object
 9   VELOCIDAD_BAJADA  1181673 non-null  int64 
 10  VELOCIDAD_SUBIDA  1181673 non-null  int64 
 11  No DE ACCESOS     1181673 non-null  int64 
dtypes: int64(7), object(5)
memory usage: 108.2+ MB
```

Cuarto paso: Procederemos a borrar las filas duplicadas

In [35]:
```
df=df.drop_duplicates()
```

In [36]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1167924 entries, 0 to 1181672
Data columns (total 12 columns):
 #   Column            Non-Null Count    Dtype 
---  ------            --------------    ----- 
 0   AÑO               1167924 non-null  int64 
 1   TRIMESTRE         1167924 non-null  int64 
 2   PROVEEDOR         1167924 non-null  object
 3   COD_DEPARTAMENTO  1167924 non-null  int64 
 4   DEPARTAMENTO      1167924 non-null  object
 5   COD_MUNICIPIO     1167924 non-null  int64 
 6   MUNICIPIO         1167924 non-null  object
 7   SEGMENTO          1167924 non-null  object
 8   TECNOLOGIA        1167924 non-null  object
 9   VELOCIDAD_BAJADA  1167924 non-null  int64 
 10  VELOCIDAD_SUBIDA  1167924 non-null  int64 
 11  No DE ACCESOS     1167924 non-null  int64 
dtypes: int64(7), object(5)
memory usage: 115.8+ MB
```

Vamos a realizar un resumen que cuente cuantas filas hay por Proveedor

In [41]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1167924 entries, 0 to 1181672
Data columns (total 12 columns):
 #   Column            Non-Null Count    Dtype 
---  ------            --------------    ----- 
 0   AÑO               1167924 non-null  int64 
 1   TRIMESTRE         1167924 non-null  int64 
 2   PROVEEDOR         1167924 non-null  object
 3   COD_DEPARTAMENTO  1167924 non-null  int64 
 4   DEPARTAMENTO      1167924 non-null  object
 5   COD_MUNICIPIO     1167924 non-null  int64 
 6   MUNICIPIO         1167924 non-null  object
 7   SEGMENTO          1167924 non-null  object
 8   TECNOLOGIA        1167924 non-null  object
 9   VELOCIDAD_BAJADA  1167924 non-null  int64 
 10  VELOCIDAD_SUBIDA  1167924 non-null  int64 
 11  No DE ACCESOS     1167924 non-null  int64 
dtypes: int64(7), object(5)
memory usage: 115.8+ MB
```

Quinto paso: Usaremos el paquete pandas profile que nos ayuda a perfilar los datos en pandas en una sola línea, para eso debemos usar pip install pandas profiling

In [69]:
```
pip install pandas-profiling
```

```
Collecting pandas-profiling
  Downloading pandas_profiling-3.6.6-py2.py3-none-any.whl (324 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 324.4/324.4 kB 5.9 MB/s eta 0:00:0000:01
Collecting ydata-profiling
  Downloading ydata_profiling-4.0.0-py2.py3-none-any.whl (344 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 344.5/344.5 kB 17.2 MB/s eta 0:00:00
Requirement already satisfied: requests<2.29,>=2.24.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from
ydata-profiling->pandas-profiling) (2.28.1)
Requirement already satisfied: pandas!=1.4.0,<1.6,>1.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (fro
m ydata-profiling->pandas-profiling) (1.4.4)
Requirement already satisfied: scipy<1.10,>=1.4.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from yda
ta-profiling->pandas-profiling) (1.7.3)
Requirement already satisfied: pydantic<1.11,>=1.8.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from
ydata-profiling->pandas-profiling) (1.10.2)
Collecting phik<0.13,>=0.11.1
  Downloading phik-0.12.3-cp39-cp39-macosx_10_13_x86_64.whl (652 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 653.0/653.0 kB 19.2 MB/s eta 0:00:0000:01
Collecting visions[type_image_path]==0.7.5
  Downloading visions-0.7.5-py3-none-any.whl (102 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 102.7/102.7 kB 6.0 MB/s eta 0:00:00
Requirement already satisfied: numpy<1.24,>=1.16.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from yd
ata-profiling->pandas-profiling) (1.21.6)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from yda
ta-profiling->pandas-profiling) (6.0)
Requirement already satisfied: tqdm<4.65,>=4.48.2 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from yda
ta-profiling->pandas-profiling) (4.64.0)
Collecting multimethod<1.10,>=1.4
  Downloading multimethod-1.9.1-py3-none-any.whl (10 kB)
Requirement already satisfied: statsmodels<0.14,>=0.13.2 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (f
rom ydata-profiling->pandas-profiling) (0.13.2)
Collecting typeguard<2.14,>=2.13.2
  Downloading typeguard-2.13.3-py3-none-any.whl (17 kB)
Collecting htmlmin==0.1.12
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: jinja2<3.2,>=2.11.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from yd
ata-profiling->pandas-profiling) (2.11.3)
Requirement already satisfied: matplotlib<3.7,>=3.2 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from y
data-profiling->pandas-profiling) (3.5.3)
Requirement already satisfied: seaborn<0.13,>=0.10.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from
ydata-profiling->pandas-profiling) (0.11.2)
Requirement already satisfied: attrs>=19.3.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from visions
[type_image_path]==0.7.5->ydata-profiling->pandas-profiling) (21.4.0)
Collecting tangled-up-in-unicode>=0.0.4
  Downloading tangled_up_in_unicode-0.2.0-py3-none-any.whl (4.7 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 4.7/4.7 MB 18.4 MB/s eta 0:00:0000:0100:01
Requirement already satisfied: networkx>=2.4 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from visions
[type_image_path]==0.7.5->ydata-profiling->pandas-profiling) (2.8.4)
Requirement already satisfied: imagehash in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from visions[type
_image_path]==0.7.5->ydata-profiling->pandas-profiling) (4.2.1)
Requirement already satisfied: Pillow in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from visions[type_im
age_path]==0.7.5->ydata-profiling->pandas-profiling) (9.2.0)
Requirement already satisfied: MarkupSafe>=0.23 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from jinja
2<3.2,>=2.11.1->ydata-profiling->pandas-profiling) (2.0.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from matp
lotlib<3.7,>=3.2->ydata-profiling->pandas-profiling) (1.4.2)
Requirement already satisfied: pyparsing>=2.2.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from matpl
otlib<3.7,>=3.2->ydata-profiling->pandas-profiling) (3.0.9)
Requirement already satisfied: fonttools>=4.22.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from matp
lotlib<3.7,>=3.2->ydata-profiling->pandas-profiling) (4.25.0)
Requirement already satisfied: cycler>=0.10 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from matplotli
b<3.7,>=3.2->ydata-profiling->pandas-profiling) (0.11.0)
Requirement already satisfied: packaging>=20.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from matplo
tlib<3.7,>=3.2->ydata-profiling->pandas-profiling) (21.3)
Requirement already satisfied: python-dateutil>=2.7 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from m
atplotlib<3.7,>=3.2->ydata-profiling->pandas-profiling) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from pandas!=
1.4.0,<1.6,>1.1->ydata-profiling->pandas-profiling) (2022.1)
Requirement already satisfied: joblib>=0.14.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from phik<0.
13,>=0.11.1->ydata-profiling->pandas-profiling) (1.1.0)
Requirement already satisfied: typing-extensions>=4.1.0 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (fr
om pydantic<1.11,>=1.8.1->ydata-profiling->pandas-profiling) (4.3.0)
Requirement already satisfied: charset-normalizer<3,>=2 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (fr
om requests<2.29,>=2.24.0->ydata-profiling->pandas-profiling) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from
requests<2.29,>=2.24.0->ydata-profiling->pandas-profiling) (1.26.11)
Requirement already satisfied: idna<4,>=2.5 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from requests<
2.29,>=2.24.0->ydata-profiling->pandas-profiling) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from req
uests<2.29,>=2.24.0->ydata-profiling->pandas-profiling) (2022.9.24)
Requirement already satisfied: patsy>=0.5.2 in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from statsmode
ls<0.14,>=0.13.2->ydata-profiling->pandas-profiling) (0.5.2)
Requirement already satisfied: six in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from patsy>=0.5.2->stat
smodels<0.14,>=0.13.2->ydata-profiling->pandas-profiling) (1.16.0)
Requirement already satisfied: PyWavelets in /Users/monicador/opt/anaconda3/lib/python3.9/site-packages (from imagehash->
visions[type_image_path]==0.7.5->ydata-profiling->pandas-profiling) (1.3.0)
Building wheels for collected packages: htmlmin
```

```
  Building wheel for htmlmin (setup.py) ... done
  Created wheel for htmlmin: filename=htmlmin-0.1.12-py3-none-any.whl size=27082 sha256=e44021629818a7b39a374de59bb756e02
28ea6fbc6ea7a55337d3c45af188408
  Stored in directory: /Users/monicador/Library/Caches/pip/wheels/ab/a0/78/885e94cd7af32ff120febdad1870c5381c884d7f4b332d
58dd
Successfully built htmlmin
Installing collected packages: htmlmin, typeguard, tangled-up-in-unicode, multimethod, visions, phik, ydata-profiling, pa
ndas-profiling
Successfully installed htmlmin-0.1.12 multimethod-1.9.1 pandas-profiling-3.6.6 phik-0.12.3 tangled-up-in-unicode-0.2.0 ty
peguard-2.13.3 visions-0.7.5 ydata-profiling-4.0.0

[notice] A new release of pip available: 22.3.1 -> 23.0.1
[notice] To update, run: pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

Sexto paso: Generamos el reporte del análisis exploratorio, para esto vamos a ponerle un nombre al reporte, en este caso lo llamé profile y luego vamos a usar la función Profile_Report y entre () le vamos a indicar el nombre del dataset que queremos que utilice, en este caso el df y se pueden agragar otros datos como el estilo de reporte, pero no son obligatorios. Y por ultimo como profile llamamos a ese reporte que acabamos de acabamos de crear.

In [70]:

```
from pandas_profiling import ProfileReport

profile =  ProfileReport(df, title="Análisis Exploratorio del Dataset: Internet Fijo Accesos por tecnología y segmento", l
profile
```

```
/var/folders/1c/n9t4cdsn7j18cgz1mdd23xrc0000gn/T/ipykernel_20734/4007692417.py:1: DeprecationWarning: `import pandas_prof
iling` is going to be deprecated by April 1st. Please use `import ydata_profiling` instead.
  from pandas_profiling import ProfileReport
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]
Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]
```

Overview

Dataset statistics

Number of variables	11
Number of observations	1181673
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	125456
Duplicate rows (%)	10.6%
Total size in memory	140.4 MiB
Average record size in memory	124.6 B

Variable types

Categorical	6
Numeric	5

Alerts

Dataset has 125456 (10.6%) duplicate rows	**Duplicates**
`PROVEEDOR` has a high cardinality: 1014 distinct values	**High cardinality**
`MUNICIPIO` has a high cardinality: 1032 distinct values	**High cardinality**
`PROVEEDOR` is highly imbalanced (52.6%)	**Imbalance**
`VELOCIDAD_BAJADA` is highly skewed (γ1 = 368.2238432)	**Skewed**
`VELOCIDAD_SUBIDA` is highly skewed (γ1 = 378.2588039)	**Skewed**
`No DE ACCESOS` is highly skewed (γ1 = 59.45343284)	**Skewed**

Out[70]:

Séptimo paso: Mi análisis hasta ahora: Podémos ver que hay 11 variables y el número de total de observación es 1181673 registros, ese es el número de filas y hay 0 celdas faltantes o valores faltantes, 125456 registros duplicados y el tamaño total de los datos es 140.4 MiB y el tamaño medio de los registros en la memoria es de 124.6 bytes

In [ ]: