

## Oswaldo González tesis

### Caso de estudio de la red nacional "Internet para todos"

Los dos archivos seleccionados han sido descargado de la plataforma de Datos Abiertos del gobierno. La finalidad de este proyecto es facilitar el acceso a la información de carácter público que genera el Estado, fomentando la participación ciudadana mediante el uso y reutilización de esta información en pro de la transparencia e innovación de los servicios públicos, la interoperabilidad del Gobierno y la rendición de cuentas con sujeción a las restricciones que establece la Ley.

Aclarado esto los datos seleccionados son los siguientes: Puntos de acceso vigentes y data transferida.

<https://www.datosabiertos.gob.pa/>

Responderemos algunas interrogantes previamente elaboradas:

- 1.) ¿Cuál es la distribución por tipo de conexión?
- 2.) ¿Cuál es el tipo de conexión por provincia?
- 3.) ¿Hay relación entre el tipo de conexión y la latitud?
- 4.) ¿Cuáles son las cantidades de puntos de acceso por provincia?
- 5.) ¿Cuál es el historial de activación por región?
- 6.) ¿Cuál es el tipo de conexión y cantidad de data transferida por región?
- 7.) Localizar los puntos de acceso a través de su latitud y longitud
- 8.) ¿Cuál es el conteo de punto de acceso por proveedor?
- 9.) ¿Cuales son los distritos con más data transferida?
- 10.) ¿Cantidad de data transferida por proveedor ?
- 11.) Relación entre Provincia y Data Transferida

Importamos las librerías con sus correspondientes llamados para cada vez que la usemos

```
In [41]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
import seaborn as sns
```

Llamamos al archivo csv y lo cargamos con el nombre de "df"

```
In [8]: df = pd.read_csv('aig-puntos-de-acceso-de-la-rni2.0-vigentes-en-agosto2022.csv'),
```

Se despliega el llamado del encabezado del archivo y comenzamos con la limpieza del mismo.

```
In [9]: df.head(3)
```

```
Out[9]:
```

	Región	PA	Nombre	Provincia	Distrito	Corregimiento	Tipo UM	Latitud	Longitud	Fecha de Activación
0	1	1	Colegio Rogelio Josué Ibarra	Bocas del Toro	Bocas del Toro	Bocas del Toro	FO	9.340654	-82.242499	12/07
1	1	3	Escuela República de Nicaragua	Bocas del Toro	Bocas del Toro	Bocas del Toro	FO	9.338938	-82.242668	12/07
2	1	5	Parque Simón Bolívar	Bocas del Toro	Bocas del Toro	Bocas del Toro	FO	9.340183	-82.240631	12/07

```
In [10]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1320 entries, 0 to 1319
Data columns (total 10 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Región                1320 non-null   int64
 1   PA                    1320 non-null   int64
 2   Nombre                1320 non-null   object
 3   Provincia              1320 non-null   object
 4   Distrito              1320 non-null   object
 5   Corregimiento          1317 non-null   object
 6   Tipo UM                1320 non-null   object
 7   Latitud                1316 non-null   float64
 8   Longitud               1316 non-null   float64
 9   Fecha de Activación    1282 non-null   object
dtypes: float64(2), int64(2), object(6)
memory usage: 103.2+ KB
```

```
In [11]: df.describe()
```

```
Out[11]:
```

	Región	PA	Latitud	Longitud
<b>count</b>	1320.000000	1320.000000	1316.000000	1316.000000
<b>mean</b>	1.986364	825.450758	8.653626	-80.509939
<b>std</b>	0.795054	492.869627	0.466552	1.268377
<b>min</b>	1.000000	1.000000	7.348218	-82.950249
<b>25%</b>	1.000000	380.750000	8.338338	-81.313797
<b>50%</b>	2.000000	821.500000	8.662533	-80.271723
<b>75%</b>	3.000000	1254.250000	9.024478	-79.539383
<b>max</b>	3.000000	1665.000000	9.580209	-77.354799

```
In [12]: df.isna().sum()
```

```
Out[12]:
```

Región	0
PA	0
Nombre	0
Provincia	0
Distrito	0
Corregimiento	3
Tipo UM	0
Latitud	4
Longitud	4
Fecha de Activación	38
dtype:	int64

Tenemos celdas con espacio en blanco, podemos eliminar las filas complementamente o realizar otro tipo de limpieza para que no se vea afectado el análisis.

## 1.) ¿Cuál es la distribución por tipo de conexión?

```
In [42]: # Cuenta la cantidad de cada tipo de conexión
tipo_um_counts = df['Tipo UM'].value_counts()

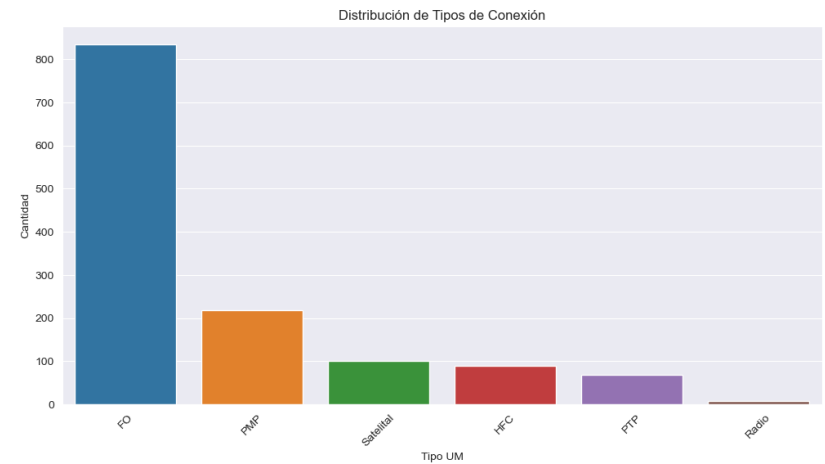
# Crea un gráfico de barras
plt.figure(figsize=(12, 6))
sns.barplot(x=tipo_um_counts.index, y=tipo_um_counts.values)

# Título del gráfico
plt.title('Distribución de Tipos de Conexión')

# Etiquetas de Los ejes
plt.xlabel('Tipo UM')
plt.ylabel('Cantidad')

# Rota las etiquetas del eje x para facilitar la legibilidad
plt.xticks(rotation=45)

# Muestra el gráfico
plt.show()
```



El tipo de conexión más dominante es el FO comunmente llamado fibra óptica, seguido por el PMP y el satelital.

## 2.) ¿Cuál es el tipo de conexión por provincia?

```
In [14]: # Crea un DataFrame que agrupa por provincia y tipo de conexión y cuenta la cant
provincia_tipo_counts = df.groupby(['Provincia', 'Tipo UM']).size().unstack(fill

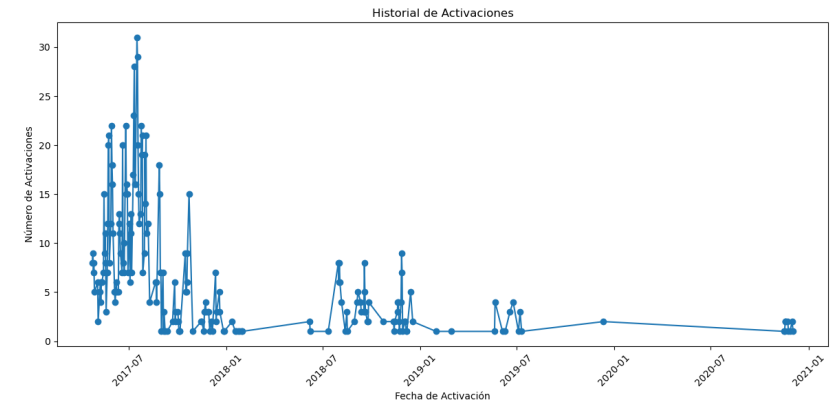
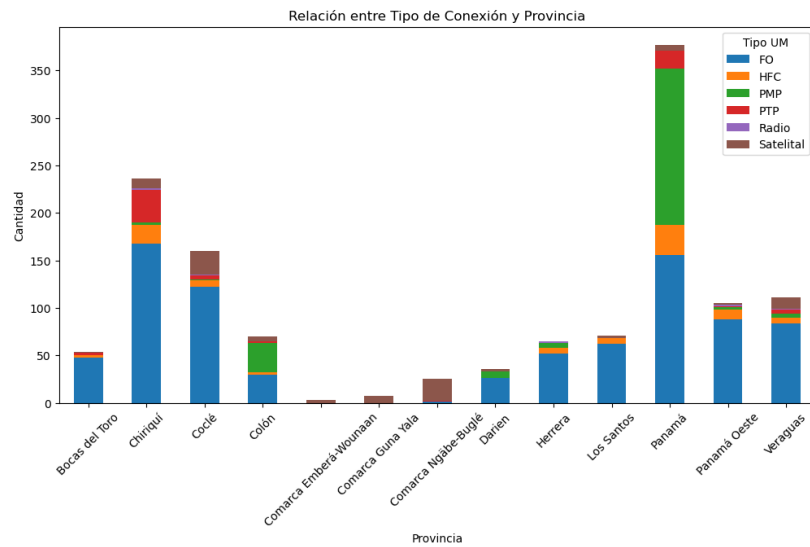
# Crea un gráfico de barras apiladas
provincia_tipo_counts.plot(kind='bar', stacked=True, figsize=(12, 6))

# Título del gráfico
plt.title('Relación entre Tipo de Conexión y Provincia')

# Etiquetas de Los ejes
plt.xlabel('Provincia')
plt.ylabel('Cantidad')

# Rota las etiquetas del eje x para facilitar la legibilidad
plt.xticks(rotation=45)

# Muestra el gráfico
plt.show()
```



Esta gráfica es útil para comprender la distribución de las activaciones de Internet a lo largo del tiempo. Puedes identificar tendencias, picos, o cualquier patrón en la cantidad de activaciones en diferentes momentos.

### 3.) ¿Cuál es el historial de activación por región?

```
In [15]: # Convierte la columna 'Fecha de Activación' en formato de fecha
df['Fecha de Activación'] = pd.to_datetime(df['Fecha de Activación'], format='%d

# Agrupa las activaciones por fecha y cuenta cuántas ocurrieron en cada fecha
activaciones_por_fecha = df.groupby('Fecha de Activación').size()

# Crea un gráfico de líneas
plt.figure(figsize=(12, 6))
plt.plot(activaciones_por_fecha.index, activaciones_por_fecha.values, marker='o')

# Etiquetas y título
plt.xlabel('Fecha de Activación')
plt.ylabel('Número de Activaciones')
plt.title('Historial de Activaciones')

# Rota las etiquetas del eje x para una mejor legibilidad
plt.xticks(rotation=45)

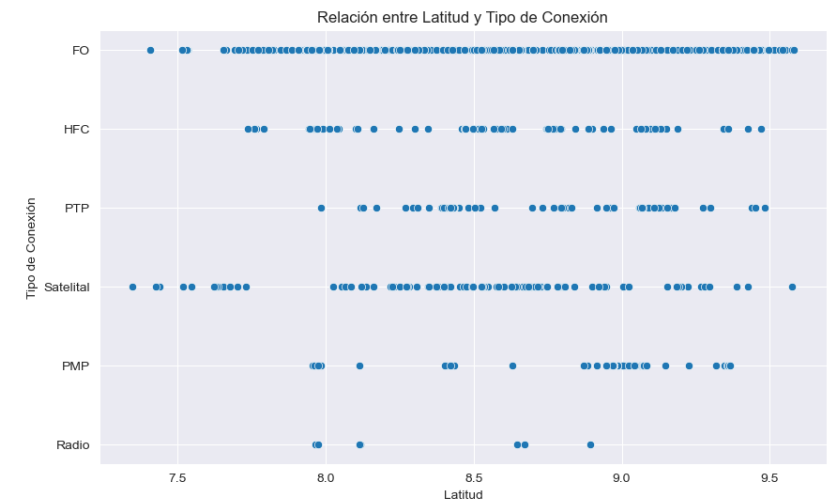
# Muestra el gráfico
plt.tight_layout()
plt.show()
```

```
In [16]: sns.set_style("darkgrid")
plt.figure(figsize=(10, 6))

sns.scatterplot(x='Latitud', y='Tipo UM', data=df)

plt.title('Relación entre Latitud y Tipo de Conexión')
plt.xlabel('Latitud')
plt.ylabel('Tipo de Conexión')

plt.show()
```



Esta gráfica te permite visualizar si existe alguna relación aparente entre la ubicación geográfica (latitud) y el tipo de conexión a Internet en tu conjunto de datos. Sin embargo, en esta gráfica, los puntos parecen estar dispersos y no sigue un patrón claro. Esto sugiere que la latitud por sí sola no determina el tipo de conexión, ya que los puntos de acceso con diferentes tipos de conexión están distribuidos por toda la gama de latitudes.

## 5.) ¿Cuáles son las cantidades de puntos de acceso por provincia?

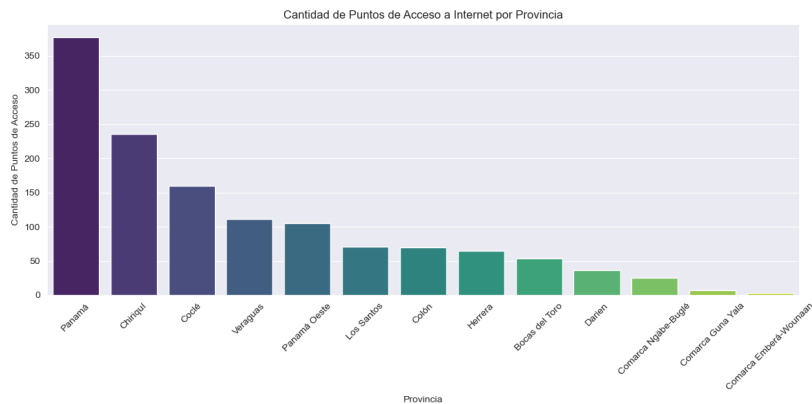
```
In [17]: # Calcula la cantidad de puntos de acceso por provincia
puntos_por_provincia = df['Provincia'].value_counts()

# Configuración del gráfico
plt.figure(figsize=(12, 6))
sns.barplot(x=puntos_por_provincia.index, y=puntos_por_provincia.values, palette='magma')

plt.title('Cantidad de Puntos de Acceso a Internet por Provincia')
plt.xlabel('Provincia')
plt.ylabel('Cantidad de Puntos de Acceso')

# Rota las etiquetas del eje x para facilitar la legibilidad
plt.xticks(rotation=45)

# Muestra el gráfico
plt.tight_layout()
plt.show()
```



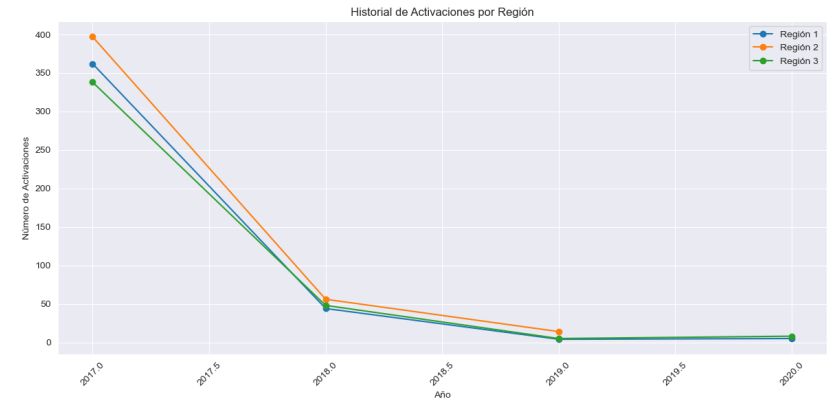
La región de las comarcas es el que menos puntos de acceso a internet tiene, posee menos de 50, incluso Darien no llega a esa cifra.

## 6.) ¿Cuál es el historial de activación por región?

```
In [43]: # Convierte la columna 'Fecha de Activación' en tipo datetime
df['Fecha de Activación'] = pd.to_datetime(df['Fecha de Activación'], format='%d/%m/%Y')

# Agrupa las activaciones por región
regiones = df['Región'].unique()
```

b



La gráfica muestra que el número de activaciones en las tres regiones ha disminuido significativamente en los últimos tres años. Esto podría ser un indicativo de una disminución en la demanda de nuevas activaciones de Internet en estas regiones o de que se ha alcanzado una cierta saturación en el mercado. Es importante para las empresas y proveedores de servicios de Internet comprender estas tendencias para tomar decisiones informadas sobre la inversión y la expansión de servicios en estas regiones. También tomar en cuenta la región 2 que dejó de activarse nuevos puntos desde el 2019.

Cargaremos el siguiente archivo de data transferida llamándolo "df1" exploraremos y limpiaremos los datos si es debido.

```
In [44]: df1 = pd.read_csv('aig-data-transferida-en-abril-2021.csv', delimiter=';', encoding='utf-8')
```

```
In [45]: df1.head()
```

Out[45]:

	Región	Proveedor	PA	Nombre	Provincia	Distrito	Corregimiento	Data (GB)
0	1	C&W Panam	1	Colegio Rogelio Josu Ibarra	Bocas del Toro	Bocas del Toro	Bocas del Toro	5
1	1	C&W Panam	3	Escuela República de Nicaragua	Bocas del Toro	Bocas del Toro	Bocas del Toro	16
2	1	C&W Panam	5	Parque Simón Bolívar	Bocas del Toro	Bocas del Toro	Bocas del Toro	75
3	1	C&W Panam	6	Col. Sec. de Almirante	Bocas del Toro	Almirante	Puerto Almirante	37
4	1	C&W Panam	7	Escuela Almirante	Bocas del Toro	Almirante	Puerto Almirante	30

Como podemos observar el archivo se encuentra con errores de escrituras, esto puede ser que el archivo fue puesto de esta manera o es el tipo de formato de lectura de Python que lo muestra erroneamente.

```
In [61]: provincias = df1['Provincia']
print(provincias.unique())

['Bocas del Toro' 'Chiriquí' 'Comarca Ngábe-Buglé' 'Veraguas' 'Colón'
'Comarca Ember -Wounaan' 'Comarca Guna Yala' 'Darién' 'Panamá' 'Coclé'
'Herrera' 'Los Santos' 'Panamá Oeste']

In [60]: # Función para corregir caracteres no válidos
def corregir_provincia(nombre):
    nombre = nombre.replace('í', 'i')
    nombre = nombre.replace('ó', 'o')
    nombre = nombre.replace('\xa0', ' ')
    nombre = nombre.replace('\x84', 'á')
    nombre = nombre.replace('\x82', 'é')
    nombre = nombre.replace('\xa0', ' ')
    return nombre

# Aplica la función a la columna 'Provincia'
df1['Provincia'] = df1['Provincia'].apply(corregir_provincia)

# Muestra los valores únicos en la columna 'Provincia'
print(df1['Provincia'].unique())

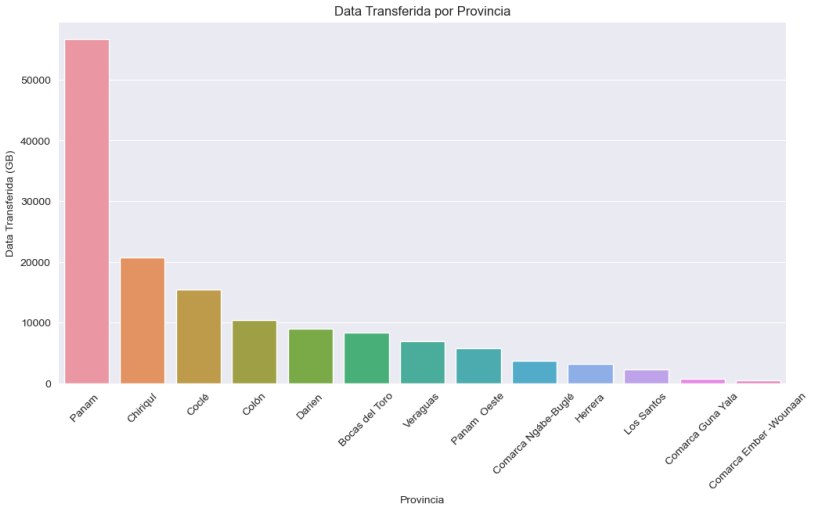
['Bocas del Toro' 'Chiriquí' 'Comarca Ngábe-Buglé' 'Veraguas' 'Colón'
'Comarca Ember -Wounaan' 'Comarca Guna Yala' 'Darién' 'Panamá' 'Coclé'
'Herrera' 'Los Santos' 'Panamá Oeste']
```

Corregimos el nombre de las provincias para que sea comprendido.

Tipo de conexión por provincia, región y cantidad de data transferida.

```
In [65]: # Ordena Las provincias por La suma de Data (GB) de forma descendente
provincias_y_data = df1.groupby('Provincia')['Data (GB)'].sum().reset_index()
provincias_y_data = provincias_y_data.sort_values(by='Data (GB)', ascending=False)

# Gráfico de barras
plt.figure(figsize=(12, 6))
sns.barplot(data=provincias_y_data, x='Provincia', y='Data (GB)', errorbar=None)
plt.title('Data Transferida por Provincia')
plt.xlabel('Provincia')
plt.ylabel('Data Transferida (GB)')
plt.xticks(rotation=45)
plt.show()
```



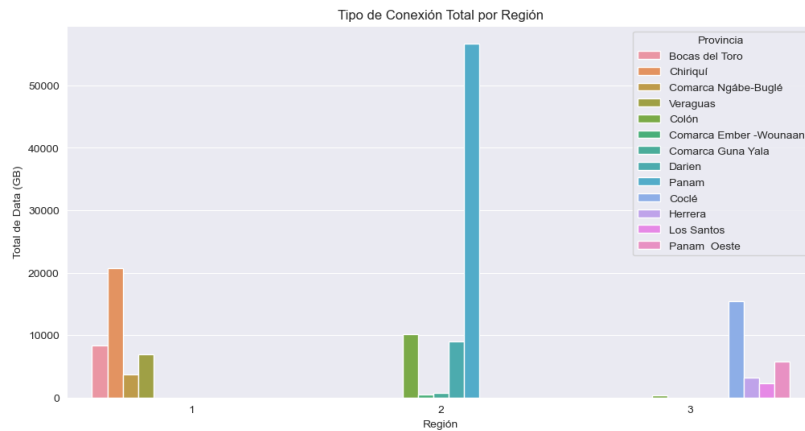
```
In [51]: # Agrupa Los datos por Región y Tipo UM y suma Los valores de Data (GB)
total_por_region = df1.groupby(['Región', 'Provincia'])['Data (GB)'].sum().reset_index()

# Crea un gráfico de barras
plt.figure(figsize=(12, 6))
sns.barplot(x='Región', y='Data (GB)', hue='Provincia', data=total_por_region)

# Título del gráfico
plt.title('Tipo de Conexión Total por Región')

# Etiquetas de Los ejes
plt.xlabel('Región')
plt.ylabel('Total de Data (GB)')

# Muestra el gráfico
plt.show()
```



Como se observa la región 2 tiene un alto consumo de datos, esto se debe a la concentración de puntos en la capital y se muestra una diferencia entre el ambiente rural como son las comarcas y las urbanas como la provincia de Panamá.

## Localizar los puntos de acceso a través de su latitud y longitud

```
In [23]: import folium

# Elimina filas con valores NaN en Latitud o Longitud
df = df.dropna(subset=['Latitud', 'Longitud'])

# Crea un mapa centrado en las coordenadas promedio de Latitud y Longitud
mapa = folium.Map(location=[df['Latitud'].mean(), df['Longitud'].mean()], zoom_start=15)

# Itera a través de los puntos de acceso y añade marcadores al mapa
for index, row in df.iterrows():
    folium.Marker([row['Latitud'], row['Longitud']], popup=row['Nombre']).add_to(mapa)

# Muestra el mapa
mapa
```

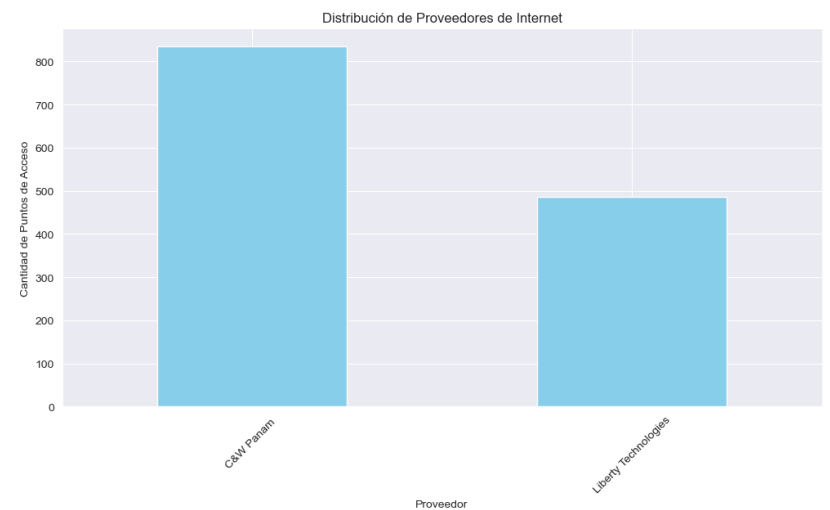


Puntos de acceso a lo largo del país

## 8.) ¿Cuál es el conteo de punto de acceso por proveedor?

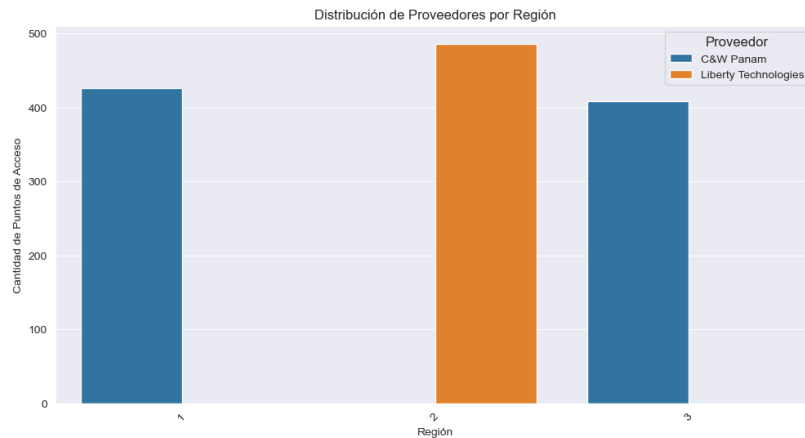
```
In [25]: # Conteo de puntos de acceso por proveedor
proveedor_counts = df1['Proveedor'].value_counts()

# Gráfico de barras
plt.figure(figsize=(12, 6))
proveedor_counts.plot(kind='bar', color='skyblue')
plt.title('Distribución de Proveedores de Internet')
plt.xlabel('Proveedor')
plt.ylabel('Cantidad de Puntos de Acceso')
plt.xticks(rotation=45)
plt.show()
```



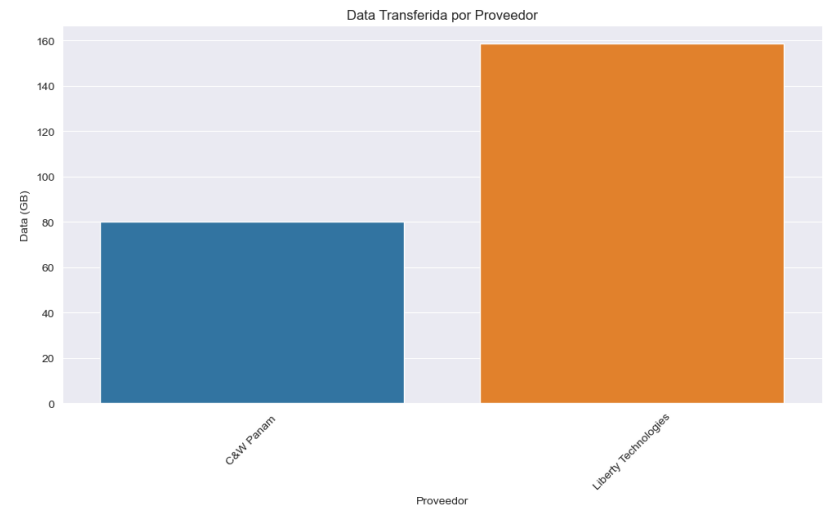
## Distribución de proveedores por región

```
In [26]: # Gráfica de Región por Proveedor
plt.figure(figsize=(12, 6))
sns.countplot(data=df1, x='Región', hue='Proveedor')
plt.title('Distribución de Proveedores por Región')
plt.xlabel('Región')
plt.ylabel('Cantidad de Puntos de Acceso')
plt.legend(title='Proveedor', title_fontsize='12', loc='upper right')
plt.xticks(rotation=45)
plt.show()
```



## Data transferida por proveedor.

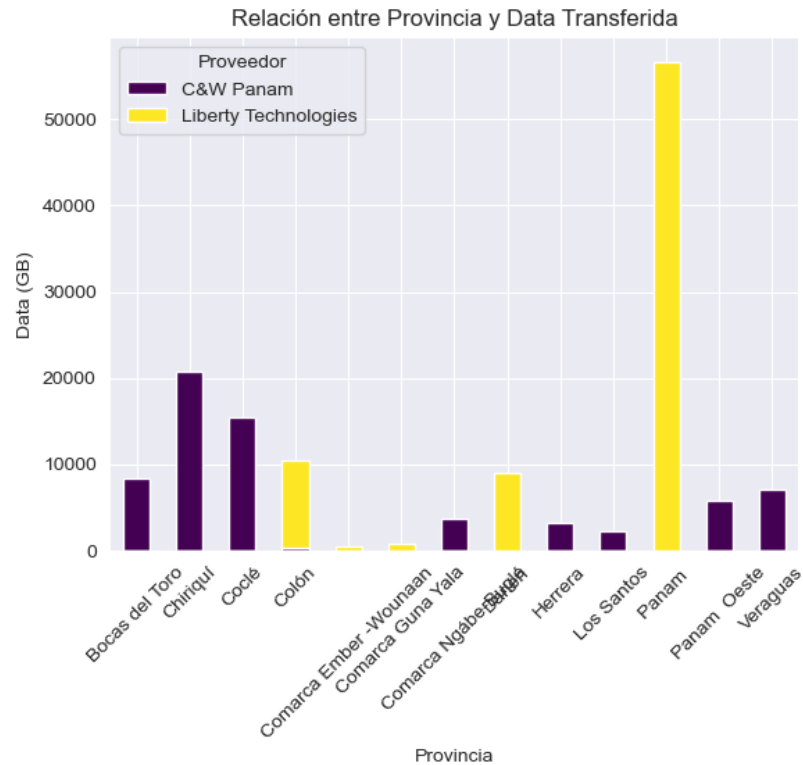
```
In [62]: # Gráfica de Proveedor por Data (GB)
plt.figure(figsize=(12, 6))
sns.barplot(data=df1, x='Proveedor', y='Data (GB)', errorbar=None)
plt.title('Data Transferida por Proveedor')
plt.xlabel('Proveedor')
plt.ylabel('Data (GB)')
plt.xticks(rotation=45)
plt.show()
```



## Gráfica de Provincia, Data transferida y proveedor.

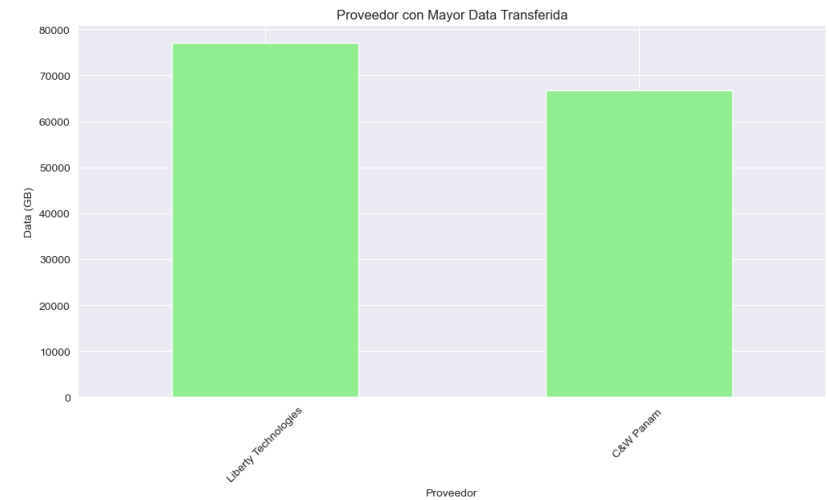
```
In [54]: # Gráfico de barras apilado por provincia y proveedor
plt.figure(figsize=(12, 6))
provincia_proveedor = df1.groupby(['Provincia', 'Proveedor'])['Data (GB)'].sum()
provincia_proveedor.plot(kind='bar', stacked=True, colormap='viridis')
plt.title('Relación entre Provincia y Data Transferida')
plt.xlabel('Provincia')
plt.ylabel('Data (GB)')
plt.xticks(rotation=45)
plt.show()
```

<Figure size 1200x600 with 0 Axes>



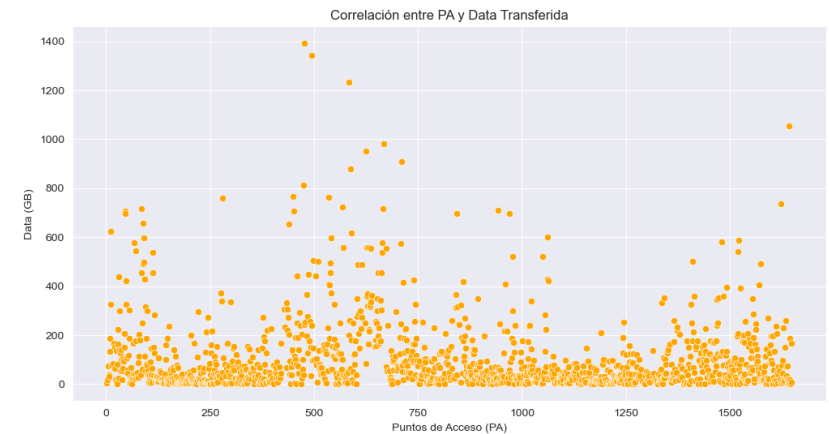
Data transferida según proveedores.

```
In [30]: # Gráfico de barras
plt.figure(figsize=(12, 6))
proveedor_data = df1.groupby('Proveedor')['Data (GB)'].sum().sort_values(ascending=False)
proveedor_data.plot(kind='bar', color='lightgreen')
plt.title('Proveedor con Mayor Data Transferida')
plt.xlabel('Proveedor')
plt.ylabel('Data (GB)')
plt.xticks(rotation=45)
plt.show()
```



Correlación entre la cantidad de Puntos de Acceso y la cantidad de Data transferida.

```
In [63]: # Gráfico de dispersión
plt.figure(figsize=(12, 6))
sns.scatterplot(data=df1, x='PA', y='Data (GB)', color='orange')
plt.title('Correlación entre PA y Data Transferida')
plt.xlabel('Puntos de Acceso (PA)')
plt.ylabel('Data (GB)')
plt.show()
```



Los puntos están dispersos de manera aleatoria y no siguen un patrón específico, no hay una correlación clara entre estas dos variables.

Conclusiones del analisis:



Concentración en la Provincia de Panamá: La mayoría de los puntos de acceso y la Data Transferida se concentran en la Provincia de Panamá. Esto sugiere un desequilibrio en la distribución de recursos y servicios de Internet en el país.

Escasa Infraestructura en Zonas Rurales: La observación de que hay un escaso uso de implementos de red a larga distancia en zonas rurales apunta a una brecha digital. Las áreas rurales suelen tener menos infraestructura de Internet debido a desafíos logísticos y económicos.

Baja Tasa de Instalación de Nuevos Puntos: El hecho de que haya pocos puntos nuevos instalados puede deberse a limitaciones presupuestarias o a la saturación del mercado. Es importante explorar por qué no se están implementando más puntos nuevos.

Posible Satisfacción de la Demanda: La baja instalación de nuevos puntos podría deberse a que se ha satisfecho la demanda actual de servicios de Internet en Panamá. Esto puede ser una señal de madurez en el mercado.