

## **DATA ANALYSIS PROJECT: CONEXIÓN A INTERNET EN PERÚ**

El objetivo de este análisis es demostrar cuál es el porcentaje de las viviendas en todas las regiones del Perú que tienen acceso a internet. Para hacer esta medición, además de la demografía oficial del país, se tomará en cuenta, como tipo de acceso a internet, la conexión fija residencial durante el primer trimestre del año 2022.

### **Alcance Técnico:**

- SQL BigQuery (Data Collecting, Data cleaning, Data Analysis)
- Power BI (Data Visualization)

### **Alcance de Trabajo:** Reporte y Dashboard.

- ¿Cuál es el porcentaje de las viviendas que tienen acceso a internet en las distintas regiones del Perú?
- ¿Cuáles son las 3 principales empresas que brindan servicio de Internet al Perú?
- ¿Qué tecnología de acceso a internet es la más usada en el Perú?

**Fuente de Información:** Los datos recopilados para el análisis de este proyecto son de alta credibilidad gracias a haber sido elaborados por instituciones gubernamentales del Perú. Así mismo, están actualizados a la fecha actual.

BASE DE DATOS	FUENTE	ÚLTIMA FECHA DE ACTUALIZACIÓN
Conexiones de Internet Fijo	Organismo Supervisor de Inversión Privada en Telecomunicaciones – OSIPTEL	2022-09-19
Perú. Estimaciones y Proyecciones de la Población por Departamento, 1995-2030	Instituto Nacional de Estadística e Informática – INEI	2020
Perú. Viviendas Particulares Censadas, Según Departamento, 2007 y 2017	Instituto Nacional de Estadística e Informática – INEI	2017

### **Project Activities:**

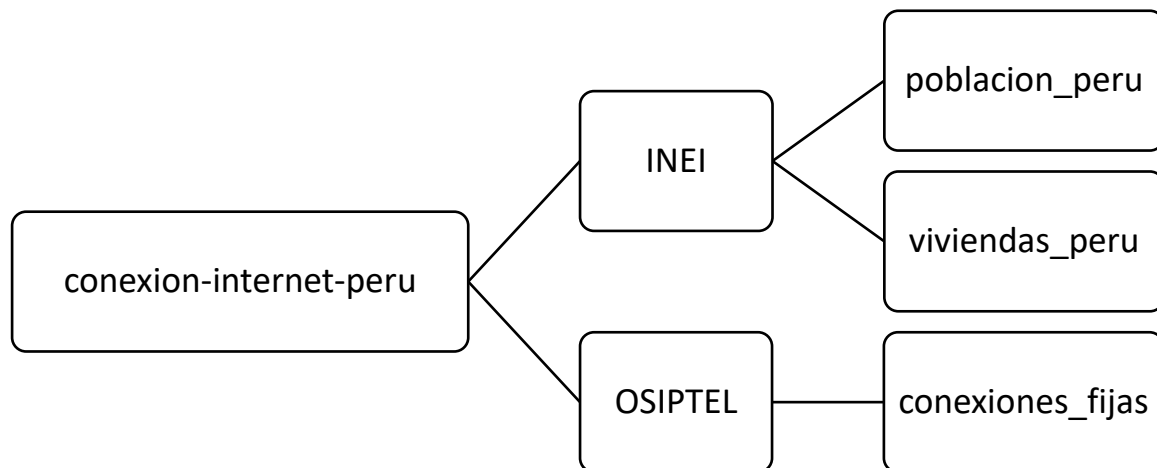
ACTIVITY	DESCRIPTION
Data Collection	Crear el Data Warehouse con todas las bases de datos seleccionadas.
Data Cleaning	Limpiar todas las columnas de cada Base de datos.
Data Analysis	Crear reporte para responder las preguntas del alcance del proyecto identificando y filtrando la información recolectada.
Data Visualization	Crear un dashboard interactivo exponiendo los resultados del análisis.

## Proceso:

### Data Collecting:

Para comenzar con el desarrollo de este proyecto de análisis primero hemos creado el Data Warehouse. Un compilado de las bases de datos que utilizaremos para los análisis correspondientes. Hemos ubicado data sobre las conexiones a internet fijo y la demografía del Perú en repositorios públicos y gubernamentales del Perú. Al ser estos documentos oficiales y actualizados, garantizamos la credibilidad de la información que obtendremos por medio de la lectura de estos datos.

Los repositorios ubicados nos dan la posibilidad de descargar las bases de datos en formato .CSV, lo cual facilita la importación de estos archivos en la plataforma BigQuery, en donde crearemos nuestro Data Warehouse. El nombre de nuestro proyecto es “conexión-internet-peru”, y dentro hemos creados dos Data Sets, una por cada fuente de donde hemos recolectado las bases de datos: INEI y OSIPTEL.



## Data Cleaning:

Al hacer lectura de las bases de datos nos hemos percatado que existe una incongruencia mínima en el nombre de los departamentos o regiones. Por ejemplo: Los departamentos Junín, Huánuco y San Martín están almacenados con tilde en la base de datos de las conexiones fijas a internet, mientras que, en la demografía del Perú, estas regiones aparecen sin tilde.

```
SELECT Departamento
FROM `conexion-internet-
peru.OSIPTEL.conexiones_fijas`
GROUP BY Departamento;
```

Departamento
Amazonas
Ancash
Apurímac
Arequipa
Ayacucho
Cajamarca
Callao
Cusco
Huancavelica
Huánuco
Ica
Junín
La Libertad
Lambayeque
Lima
Loreto
Madre de Dios
Moquegua
Pasco
Piura
Puno
San Martín
Tacna
Tumbes
Ucayali

```
SELECT Departamento
FROM `conexion-internet-
peru.INEI.viviendas_peru`;
```

Departamento
AMAZONAS
ANCASH
APURIMAC
AREQUIPA
AYACUCHO
CAJAMARCA
CALLAO
CUSCO
HUANCAVELICA
HUANUCO
ICA
JUNIN
LA LIBERTAD
LAMBAYEQUE
LIMA
LORETO
MADRE DE DIOS
MOQUEGUA
PASCO
PIURA
PUNO
SAN MARTIN
TACNA
TUMBES
UCAYALI

Esta leve incongruencia ocasionará una mala agrupación de datos en el momento que hagamos el modelado y análisis; por tal motivo con una breve sentencia CASE en SQL limpiamos cada uno de los nombres de las regiones para evitar las tildes.

```
CASE
    WHEN Departamento LIKE '%á%' THEN REPLACE(Departamento, 'á', 'a')
    WHEN Departamento LIKE '%í%' THEN REPLACE(Departamento, 'í', 'i')
    ELSE Departamento
END AS Departamento
```

## Data Analysis

Para empezar con el análisis es crucial filtrar los datos que necesitamos para responder a las preguntas del proyecto. Como nuestro objetivo es determinar el porcentaje de viviendas que tienen acceso a internet fijo, no nos interesa saber cuantas conexiones a nivel “Comercial” o de empresas existen, sino más bien las del segmento “Residencial”. Así mismo, solo queremos información del primer trimestre del 2022. Por tal motivo dentro de nuestra sentencia WHERE en SQL aplicamos los siguientes filtros:

```
WHERE
  Mes BETWEEN '2022-01-01' AND '2022-03-31'
  AND Segmento = 'Residencial'
```

**Análisis 1: Porcentaje de Viviendas que tienen acceso a internet en las distintas regiones del Perú al primer trimestre del 2022 en el segmento Residencial.**

Sentencia SQL:

```
WITH
  ConexionDepartamento AS(
    SELECT
      CASE
        WHEN Departamento LIKE '%á%' THEN REPLACE(Departamento, 'á', 'a')
        WHEN Departamento LIKE '%í%' THEN REPLACE(Departamento, 'í', 'i')
        ELSE Departamento
      END AS Departamento,
      SUM(Conexiones) AS Conexiones
    FROM `conexion-internet-peru.OSIPTEL.conexiones_fijas`
    WHERE
      Mes BETWEEN '2022-01-01' AND '2022-03-31'
      AND Segmento = 'Residencial'
    GROUP BY Departamento
    ORDER BY Departamento, Conexiones DESC
  ),

  Viviendas_2022 AS(
    SELECT
      DEPARTAMENTO,
      VIVIENDAS_2017 + (INCREMENTO_ANUAL * 4) AS VIVIENDAS_2022
    FROM `conexion-internet-peru.INEI.viviendas_peru`
  )
SELECT
  cD.Departamento,
  cD.Conexiones,
  viviendas.VIVIENDAS_2022,
  ROUND((cD.Conexiones / viviendas.VIVIENDAS_2022) * 100, 2) AS Porcentaje
FROM ConexionDepartamento AS cD
FULL JOIN Viviendas_2022 AS viviendas
  ON viviendas.DEPARTAMENTO = UPPER(cD.Departamento)
ORDER BY Porcentaje DESC;
```

En la **primera parte** de la sentencia nos enfocamos en crear un primer bloque sub-query sobre todas las conexiones a internet fijas por departamento. Hemos limpiado los nombres de los departamentos quitando las tildes y además en la clausula WHERE aplicamos el filtro de tiempo y segmento correspondiente. Con una agrupación y orden descendente por departamento obtenemos el siguiente bloque:

Departamento	Conexiones
Amazonas	6594
Ancash	72203
Apurimac	11397
Arequipa	161794
Ayacucho	27427
Cajamarca	40444
Callao	156672
Cusco	77705
Huancavelica	6220
Huanuco	30852
Ica	82174
Junin	71757
La Libertad	156586
Lambayeque	98313
Lima	1594118
Loreto	8033
Madre de Dios	10116
Moquegua	17384
Pasco	6210
Piura	104464
Puno	37784
San Martin	28902
Tacna	36816
Tumbes	15394
Ucayali	22158

En la **segunda parte** de la sentencia creamos el segundo bloque sub-query sobre el número de viviendas por departamento. Aquí hacemos una estimación del total de viviendas que habría en el año 2022 según el incremento anual establecido por el INEI en el año 2017.

DEPARTAMENTO	VIVIENDAS_2022
AMAZONAS	149258
ANCASH	451587
APURIMAC	194698
AREQUIPA	654667
AYACUCHO	299408
CAJAMARCA	539846
CALLAO	309590
CUSCO	479582
HUANCAVELICA	182540
HUANUCO	278445
ICA	337987
JUNIN	475550
LA LIBERTAD	602685
LAMBAYEQUE	388269
LIMA	3308317
LORETO	235956
MADRE DE DIOS	60018
MOQUEGUA	92212
PASCO	100506
PIURA	617974
PUNO	648243
SAN MARTIN	284793
TACNA	171415
TUMBES	90475
UCAYALI	163516

**Finalmente**, unimos ambos bloques en una sentencia SQL principal. Considerando que en una tabla los valores están en mayúscula y en la otra no, al momento de realizar el JOIN correspondiente aplicamos la función UPPER() para que los valores puedan igualarse. Por otro lado, para calcular el porcentaje de viviendas con conexión a internet aplicamos una división: Número de conexiones entre el número de viviendas. Así mismo, redondeamos el resultado a un máximo de dos decimales, y ordenamos de mayor a menor según porcentaje.

Departamento	Conexiones	VIVIENDAS_2022	Porcentaje
Callao	156672	309590	50.61
Lima	1594118	3308317	48.19
La Libertad	156586	602685	25.98
Lambayeque	98313	388269	25.32
Arequipa	161794	654667	24.71
Ica	82174	337987	24.31
Tacna	36816	171415	21.48
Moquegua	17384	92212	18.85
Tumbes	15394	90475	17.01
Piura	104464	617974	16.9
Madre de Dios	10116	60018	16.85
Cusco	77705	479582	16.2
Ancash	72203	451587	15.99
Junin	71757	475550	15.09
Ucayali	22158	163516	13.55
Huanuco	30852	278445	11.08
San Martin	28902	284793	10.15
Ayacucho	27427	299408	9.16
Cajamarca	40444	539846	7.49
Pasco	6210	100506	6.18
Apurimac	11397	194698	5.85
Puno	37784	648243	5.83
Amazonas	6594	149258	4.42
Huancavelica	6220	182540	3.41
Loreto	8033	235956	3.4

La mayoría de los departamentos del Perú no superan el 20% en cuanto a conexiones a internet fijo por viviendas. Las principales regiones con conexión fija son Callao y Lima con 50.6% y 48% respectivamente; mientras que las regiones con menor acceso a este servicio son Huancavelica y Loreto, ambas con 3.4%.



## Análisis 2: Las tres principales empresas que brindan servicio de internet en el Perú al primer trimestre del 2022 en el segmento Residencial:

Sentencia SQL:

```
WITH empresasConexiones AS (  
  SELECT Empresa, SUM(Conexiones) AS Conexiones  
  FROM `conexion-internet-peru.OSIPTEL.conexiones_fijas`  
  WHERE  
    Mes BETWEEN '2022-01-01' AND '2022-03-31'  
    AND Segmento = 'Residencial'  
  GROUP BY Empresa  
)  
SELECT  
  Empresa,  
  Conexiones,  
  ROUND(Conexiones / SUM(Conexiones) OVER() * 100,2) AS Porcentaje  
FROM empresasConexiones  
ORDER BY Conexiones DESC  
LIMIT 3;
```

**Primero** creamos un bloque sub-query sobre el número de conexiones por empresas que dan servicio de internet fijo. Filtramos la información con el segmento residencial y el periodo de tiempo del primer trimestre del año 2022.

Empresa	Conexiones
INTERMAX S.A.C.	7
IWAY TELECOM S.A.C.	6
Winner Systems S.A.C.	1008
Gilat to Home Perú S.A.	118
Telefónica del Perú S.A.A.	1623467
Maral Telecomunicaciones S.A.C.	1308
BFT S.A.C.	77
BANTEL S.A.C.	10673
WOW TEL S.A.C.	37047
Entel Perú S.A.	97277
Cablenortv S.A.C.	1178
Econocable S.A.C.	2093
Chasqui@net S.R.L.	683
COMPUNETWORK S.A.C.	1153
P Y D TELECOM S.R.L	16173
VIETTEL PERU S.A.C.	378
CABLE NETWORK S.R.L.	908
North Telecom S.A.C.	142
FIBERLINE PERU S.A.C.	11918
J.R. TELECOM S.R.LTDA	2820

WI-NET TELECOM S.A.C.	169841
BEST CABLE PERU S.A.C.	993
CABLE VISION ICA S.A.C	1390
CABLE VISION MAGES SAC	1654
HUANUCO TELECOM S.A.C.	264
Hughes de Perú S.R.L.	38982
Econocable Media S.A.C.	6492
Econocable Perú S.A.C.	1042
MEGACABLE NETWORK S.A.C	410
CABLE VIDEO PERÚ S.A.C.	5100
AMERICA MOVIL PERU S.A.C.	820914
TELEVISORA DEL SUR S.A.C.	2090
Giga & Mega Plus TV S.A.C.	148
INVERSIONES TELCOTEL S.A.C	2817
OPTICAL TECHNOLOGIES S.A.C	862
RED INTERCABLE PERU S.A.C.	6864
TV Nor Comunicaciones S.A.C.	2120
QITV TELECOMUNICACIONES S.A.C.	6
CALA SERVICIOS INTEGRALES E.I.R.L	6837
CABLE MASTER COMUNICACIONES S.A.C.	2749
Comunicaciones J&F Cable TV S.A.C.	145
INVENTA TELECOMUNICACIONES E.I.R.L	620
TELECOMUNICACIONES EL TALLAN E.I.R.L.	26
CABLE PUCALLPA SRL	54
TVS Wireless S.A.C.	11
CABLE SISTEMAS S.R.L.	468
Telecomunicaciones Argonza S.R.L.	51
Fibertim S.A.C.	74
OPTICOM S.A.C.	59

**Finalmente**, en la sentencia SQL principal aplicamos una división para hallar el porcentaje por total de conexiones según empresa. Dividimos el número de conexiones por cada empresa sobre el total de todas las conexiones. Así mismo, redondeamos la división a un máximo de dos decimales y limitamos el número de resultados a tres.

Empresa	Conexiones	Porcentaje
Telefónica del Perú S.A.A.	1623467	56.34
AMERICA MOVIL PERU S.A.C.	820914	28.49
WI-NET TELECOM S.A.C.	169841	5.89

La principal empresa que brinda servicio de internet fijo en el Perú es Telefónica del Perú S.A.A., con un 56% del total de conexiones. Seguida por AMERICA MOVIL PERU S.A.C. y WI-NET TELECOM S.A.C. con 28% y 5.8% respectivamente.

### Análisis 3: Tecnología de acceso a internet fijo más usada en el Perú durante el primer trimestre del 2022 en el segmento Residencial

Sentencia SQL:

```
WITH conTec AS (  
    SELECT Tecnologia, SUM(Conexiones) AS Conexiones  
    FROM `conexion-internet-peru.OSIPTEL.conexiones_fijas`  
    WHERE  
        Mes BETWEEN '2022-01-01' AND '2022-03-31'  
        AND Segmento = 'Residencial'  
    GROUP BY Tecnologia  
)  
SELECT  
    Tecnologia,  
    Conexiones,  
    ROUND(Conexiones/SUM(Conexiones) OVER() * 100,2) AS Porcentaje  
FROM conTec  
ORDER BY Conexiones DESC;
```

**Primero** creamos un bloque sub-query con el total de conexiones a internet fijo según tecnología. Así mismo, aplicamos los filtros de tiempo y segmento correspondientes.

Tecnología	Conexiones
Otras Tecnologías	965516
xDSL	73708
Cablemódem	1842293

**Después**, en la sentencia SQL principal, hacemos una división del número de conexiones por tecnología sobre el total de conexiones para hallar el porcentaje correspondiente, y redondeamos el resultado un máximo de dos decimales.

Tecnología	Conexiones	Porcentaje
Cablemódem	1842293	63.93
Otras Tecnologías	965516	33.51
xDSL	73708	2.56

La principal tecnología usada para la conexión de internet fijo en el Perú es el “Cablemódem”, con un 63.9% del total de conexiones.

## Data Visualization:

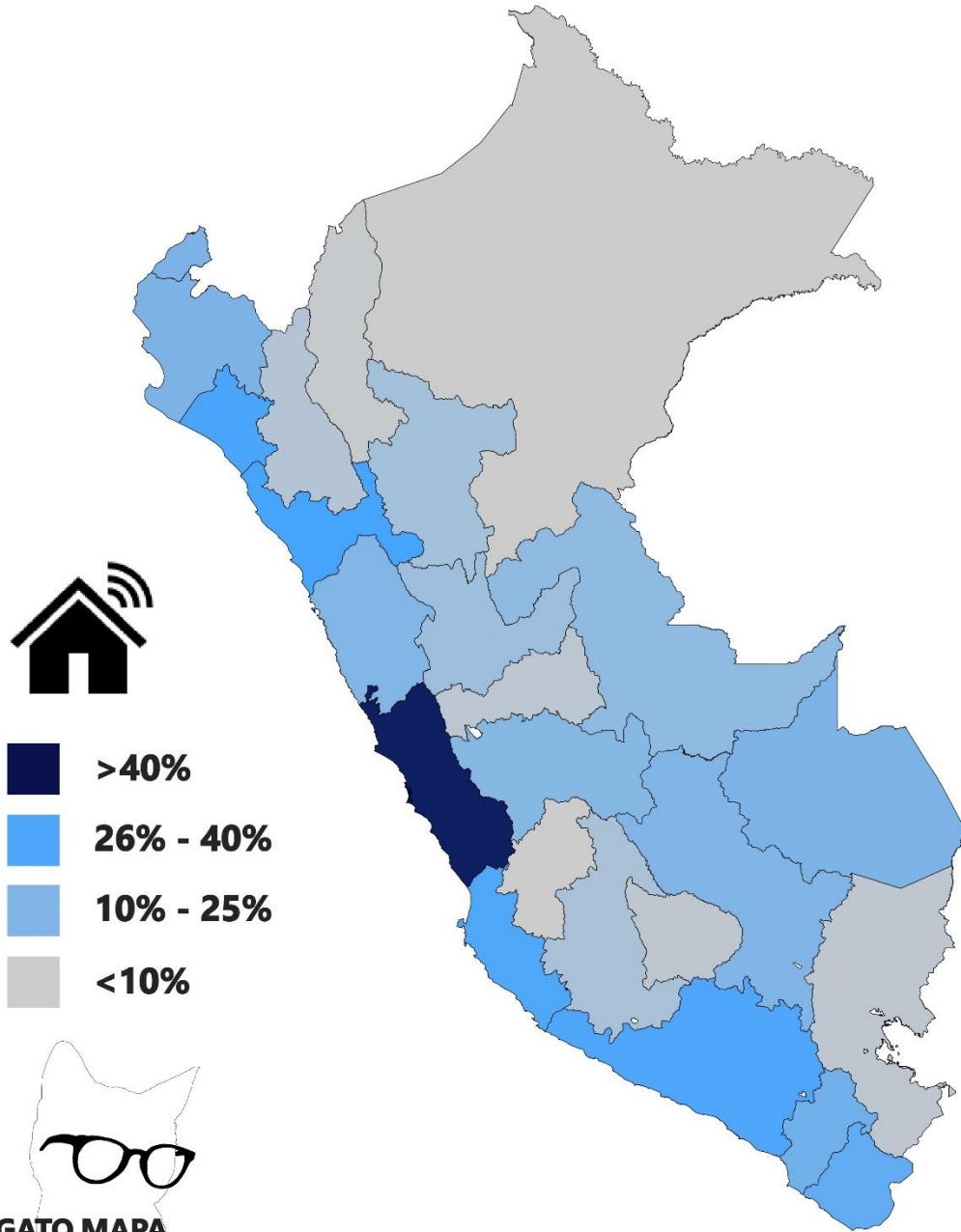
Para importar los datos en nuestra plataforma de visualización, nos hemos conectado directamente desde Power BI hacia Google BigQuery a través de una sentencia SQL. El query que utilizamos agrupa todos los datos en total de conexiones de internet fijo por departamento y empresa operadora. Así mismo, hace limpieza de las tildes correspondientes y realiza un filtro según tiempo y segmento.

```
SELECT
CASE
  WHEN Departamento LIKE '%á%' THEN REPLACE(Departamento, 'á', 'a')
  WHEN Departamento LIKE '%í%' THEN REPLACE(Departamento, 'í', 'i')
  ELSE Departamento
END AS Departamento,
CASE
  WHEN Empresa IN ('Telefónica del Perú S.A.A.', 'AMERICA MOVIL PERU S.A.C.', '
WI-NET TELECOM S.A.C.') THEN Empresa
  ELSE 'Otras Empresas'
END AS Empresa,
Tecnologia,
SUM(Conexiones) AS Total_Conexiones
FROM `conexion-internet-peru.OSIPTEL.conexiones_fijas`
WHERE
  Mes BETWEEN '2022-01-01' AND '2022-03-31'
  AND Segmento = 'Residencial'
GROUP BY Departamento, Empresa, Tecnologia
ORDER BY Departamento, Empresa, Tecnologia;
```

Con el uso de un mapa personalizado del Perú, y la saturación del color según el porcentaje de viviendas con acceso a internet fijo por departamento, creamos un mapa regional en donde se evidencia cada segmento del país con su porcentaje particular.

## PORCENTAJE DE VIVIENDAS CON ACCESO A INTERNET FIJO

## DATA POR DEPARTAMENTO



**GATO.MAPA**

FUENTE: INEI, OSIPTEL 2022 (SEGMENTO RESIDENCIAL)