

# *Adopcion del internet atravez de los años*

*Data scientist : Nicolas Felipe Mogollon Granda*

*Tipo de Proyecto:*

*Investigativo*

## *Introduccion—Contexto del proyecto*

*En la última década, el acceso a Internet se ha convertido en un pilar fundamental para el desarrollo económico, social y educativo de las naciones. Sin embargo, la velocidad y la profundidad con las que los países adoptan esta tecnología varían considerablemente. Esta disparidad plantea interrogantes clave sobre los factores que impulsan —o limitan— la adopción digital, especialmente al comparar naciones desarrolladas con aquellas en vías de desarrollo.*

*El presente proyecto aborda la pregunta:*

*¿Qué factores determinan el ritmo de adopción de Internet en países desarrollados y en desarrollo, y cuáles son las probabilidades condicionales de que se supere el 95 % de penetración antes de 2025, dadas sus condiciones socioeconómicas y de inversión digital en ambos grupos ?*

*Para responder esta pregunta, se realizó un análisis integral utilizando un dataset sintético de alta fidelidad que simula registros diarios de adopción de Internet desde 2010 hasta 2025 para diez países representativos. El estudio combina técnicas avanzadas de estadística descriptiva, inferencial, series temporales con el fin de:*

- *Comparar dinámicamente los niveles de penetración digital en países desarrollados y en desarrollo.*
- *Identificar los factores socioeconómicos y tecnológicos que inciden significativamente en el ritmo de adopción.*
- *Estimar elasticidades y relaciones causales entre inversión digital, alfabetización tecnológica y conectividad.*
- *Evaluar modelos de predicción y estimar probabilidades condicionales de alcanzar metas de cobertura ( $\geq 95\%$ ) antes del año 2025.*

*Este proyecto no solo busca generar hallazgos analíticos, sino también proporcionar una base para la toma de decisiones en políticas públicas, inversión digital y reducción de la brecha tecnológica global.*

## *Pregunta de investigación*

*¿Qué factores determinan el ritmo de adopción de Internet en países desarrollados y en desarrollo, y cuáles son las probabilidades condicionales de que se supere el 95 % de penetración antes de 2025, dadas sus condiciones socioeconómicas y de inversión digital en ambos grupos ?*

## *Preguntas de analisis*

*“este fue el approach que se tuvo para llegar a la conclusion del proyecto”*

- 1. ¿Cómo es el comportamiento general de la adopción en el tiempo y por grupo (desarrollado vs en-desarrollo)*
- 2. ¿Cómo ha evolucionado la adopción y cómo predecimos su futuro?*
- 3. "¿que tan diferentes son ambos grupos?*
- 4. ¿Hay evidencia estadística de que ciertos factores afectan la adopción?*
- 5. ¿Cuál es la probabilidad condicional de alcanzar el 95 % de penetración dado cierto nivel de inversión digital y alfabetización digital en ambos grupos?*

## *Contenido del proyecto*

<i>¿Cómo es el comportamiento general de la adopción en el tiempo y por grupo? .....</i>	<i>6</i>
<i>los países en el que va enfocado este análisis.....</i>	<i>6</i>
<i>Países desarrollados --- en desarrollo.....</i>	<i>7</i>
<i>Adopción del internet a través de los años por país.....</i>	<i>8</i>
<i>¿qué tan rápido creció la penetración de internet en cada país entre 2010 y 2025? .....</i>	<i>10</i>
<i>Cuales son las variables relevantes en cuanto la influencia en la adopción a internet .....</i>	<i>11</i>
<i>¿Cómo ha evolucionado la adopción y cómo predecimos su futuro?.....</i>	<i>12</i>
<i>Tendencia General.....</i>	<i>13</i>
<i>Estacionalidad y Estabilización.....</i>	<i>13</i>
<i>Cuántos registros históricos del pasado influirían en lo que pasaría en el futuro....</i>	<i>13</i>
<i>Correlación entre variables individuales vs Internet adoption % (y) .....</i>	<i>14</i>
<i>Conclusión general CCF.....</i>	<i>15</i>
<i>Conclusión general del test de Granger.....</i>	<i>16</i>
<i>Resumen final.....</i>	<i>17</i>
<i>Prediciendo el futuro de la adopción del internet.....</i>	<i>18</i>
<i>(SARIMAX).....</i>	<i>18</i>
<i>Visualización del modelo.....</i>	<i>19</i>
<i>"¿qué tan diferentes son ambos grupos?".....</i>	<i>21</i>
<i>"validación de supuestos (normalidad)".....</i>	<i>21</i>

<i>“Validacion de supuestos (homocedasticidad)”</i> .....	22
<i>Prueba t-student Welch</i> .....	22
<i>¿Hay evidencia estadística de que ciertos factores afectan la adopción?</i> .....	23
<i>Mobile_Data_Usage (GB)</i> .....	23
<i>Urban_Rural_Urban</i> .....	24
<i>5G_Rollout</i> .....	24
<i>Digital_Investment (M USD)</i> .....	24
<i>GDP_Per_Capita (USD)</i> .....	25
<i>Education_Level (%)</i> .....	25
<i>Digital_Literacy (%)</i> .....	25
<i>Broadband_Speed (Mbps)</i> .....	26
<i>X_Sentiment_Score</i> .....	26
<i>Conclusion final ---coef de regresion</i> .....	26
<i>Elasticidad</i> .....	27
<i>Conclusion</i> .....	27
<i>¿Cuál es la probabilidad condicional de alcanzar el 95 % de penetración dado cierto nivel de inversión digital y alfabetización digital?</i> .....	28
<i>El proceso consistió en:</i> .....	29
<i>Pregunta bonus</i> .....	30
<i>Conclusion final del proyecto</i> .....	31
<i>Recomendaciones</i> .....	33

## *Solucion*

*¿Cómo es el comportamiento general de la adopción en el tiempo y por grupo?*

*(desarrollado vs en desarrollo)*

*los países en el que va enfocado este analisis*

```
array(['USA', 'India', 'Brazil', 'Nigeria', 'Indonesia', 'Ethiopia',  
      'China', 'Kenya', 'Germany', 'South Africa'], dtype=object)
```

*Criterios para determinar si un país es desarrollado o no*

*Los criterios para determinar si un país es desarrollado o en desarrollo incluyen principalmente el PIB per cápita alto (generalmente más de USD 20,000 anuales), un Índice de Desarrollo Humano (IDH) superior a 0.800 (que mide salud, educación y nivel de vida), una economía basada en servicios y tecnología en lugar de agricultura o industria básica, infraestructura avanzada, baja desigualdad económica, alta esperanza de vida y acceso generalizado a servicios básicos como sanidad y educación. También se considera la estabilidad política, la innovación tecnológica y la sostenibilidad ambiental. China, aunque es una potencia económica, aún no se considera desarrollada porque su PIB per cápita es bajo (USD 12,720), su IDH es de 0.768 (por debajo del umbral) y tiene grandes desigualdades regionales, con zonas rurales en pobreza. Otros países como India, Brasil o Nigeria están aún más atrás en estos indicadores.*

### Países desarrollados --- en desarrollo

```
desarrollados = ['USA', 'Germany']  
en_desarrollo = ['India', 'Brazil', 'Nigeria', 'Indonesia', 'Ethiopia', 'China', 'Kenya', 'South Africa']
```

*Países desarrollados % de penetración de internet      Países en desarrollo penetración de internet*

count	11272.000000
mean	70.701282
std	23.430270
min	24.120000
25%	52.397500
50%	69.920000
75%	94.352500
max	100.000000

count	45088.00000
mean	53.71672
std	28.22167
min	7.28000
25%	29.65000
50%	49.98000
75%	84.79000
max	100.00000

**Coefficiente de variación países\_desarrollados:33.13980926116729**

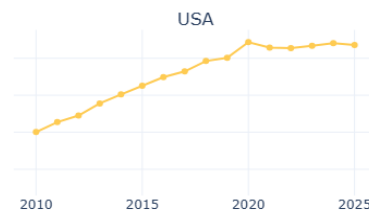
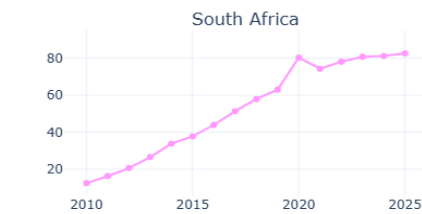
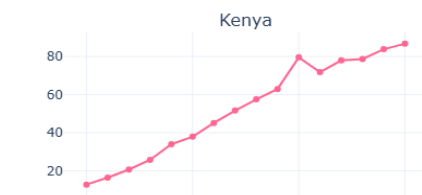
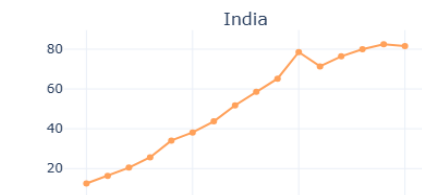
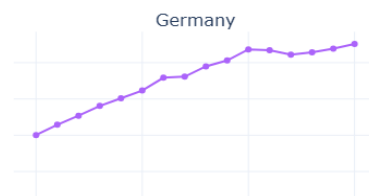
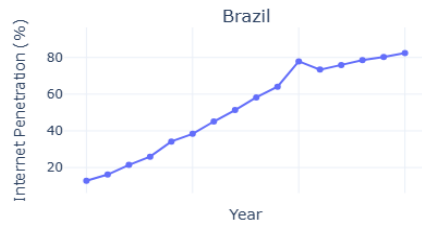
**Coefficiente de variación países\_en\_desarrollo:52.53796210937674**

Los datos muestran una clara brecha digital entre países desarrollados y en desarrollo. En promedio, los países desarrollados tienen una penetración de internet del 70.7%, superando en 17 puntos porcentuales a los países en desarrollo (53.7%). La desigualdad relativa en la adopción de internet es aún más marcada al analizar el coeficiente de variación (CV): los países en desarrollo presentan un CV del 52%, significativamente mayor que el 33% de los desarrollados. Esto confirma que, proporcionalmente, las diferencias internas en acceso a internet son 1.6 veces más pronunciadas en los países en desarrollo.

Además, la dispersión absoluta (desviación estándar de 28.2 en países en desarrollo frente a 23.4 en desarrollados) refleja contrastes extremos: en el 75% de los registros históricos (desde 2010), los países desarrollados mantuvieron tasas entre 24% y 94%, mientras que en los países en desarrollo el rango fue más amplio y bajo (7% a 84%). Aunque algunos países en desarrollo logran cobertura total (100%), el CV elevado revela que estos casos son excepciones en un panorama donde grandes poblaciones quedan rezagadas. La muestra, con 45,088 registros de países en desarrollo frente a 11,272 de desarrollados, subraya la diversidad de realidades en el primer grupo: desde economías emergentes con alta conectividad hasta regiones con acceso marginal.

## Adopcion del internet atravez de los años por pais

Internet Adoption Over Time by Country





### 1. Tendencias globales (2010-2025):

- a. Todos los países muestran un crecimiento sostenido en la adopción de internet a lo largo de los años.
- b. Entre 2019 y 2020 ocurre un salto notable en todos los países. Esto relacionado con la pandemia de COVID-19, que impulsó la digitalización en muchas regiones.

### 2. Países Desarrollados

---Alemania , EE. UU.:

- a. inician con alta penetración (más del 40% en 2010).
- b. Superan rápidamente el 80% antes del 2020.
- c. Para este ultimo año "2025" alemania termino con una adopcion de internet del 90 % y estados unidos del 87 % si miramos bien el grafico alemania parece todavia seguir aumentando su adopcion a internet como estados unidos parece estabilizarse en mira de los proximos años

### 3. Países en Desarrollo – Análisis General

- a. La mayoría comienza con tasas de penetración inferiores al 30%, países como Etiopía, Nigeria ,Kenia, south africa, indonesia, india, brazil apenas alcanzaban el 10%-15%.
- b. Entre los países en desarrollo el que mas alta adopcion a internet tuvo desde el 1 registro en el 2010 fue china con una adopcion de un poco mas del 40 % para este 2025 % acabo con una adopciona internet del 89 %
- c. El impacto que tuvo la pandemia en países como Etiopía, Kenia y Sudáfrica fue un salto grande en adopcion al internet pasaron del 60% al 80% en un solo año.

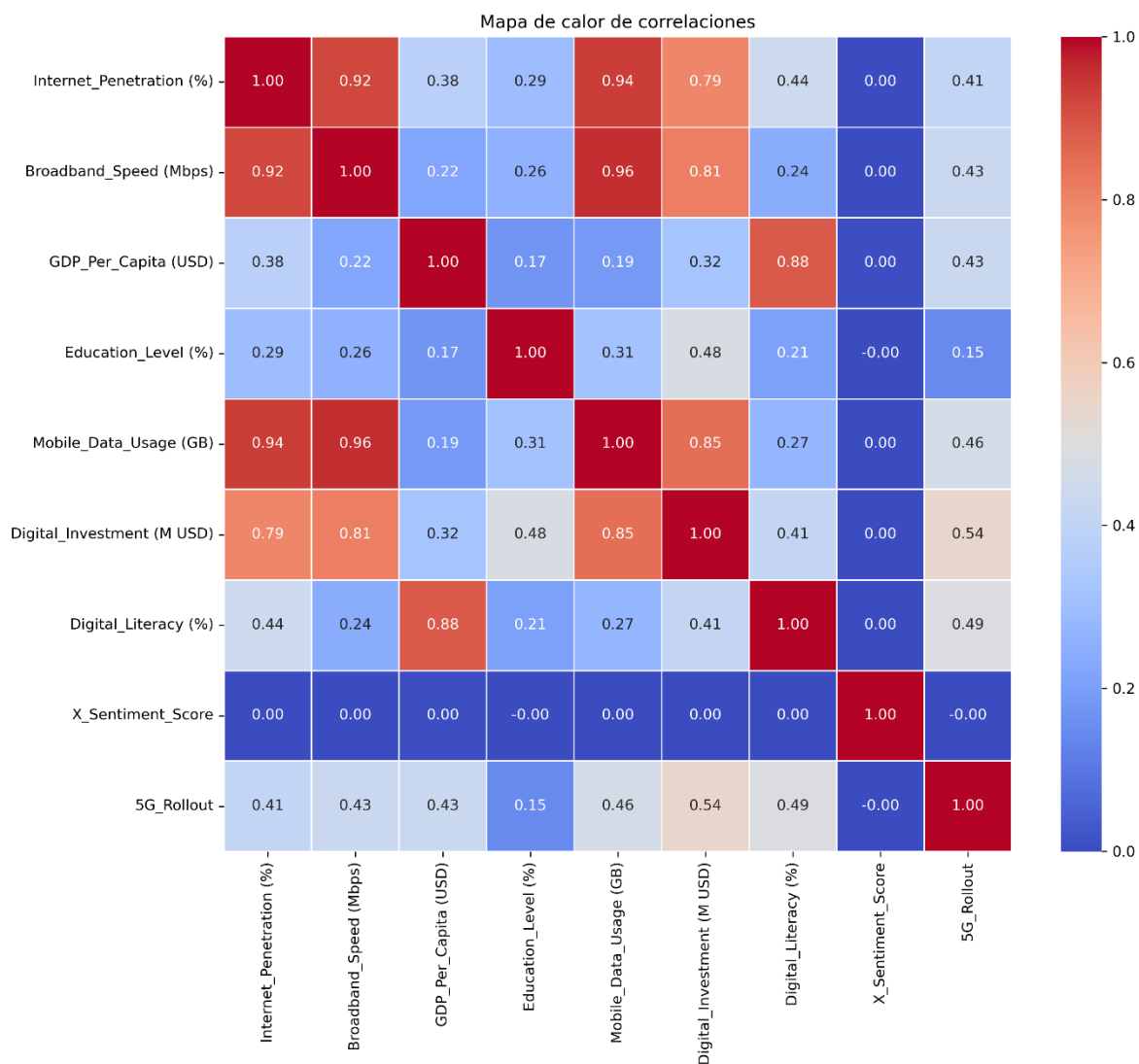
- d. En esta muestra de países, el país en desarrollo con menor adopción de internet ya en el año 2025 fue India con un 81 por ciento

¿qué tan rápido creció la penetración de internet en cada país entre 2010 y 2025?

	Country	Pendiente_promedio
2	Ethiopia	5.978000
6	Kenya	5.964000
0	Brazil	5.529333
5	Indonesia	5.504667
4	India	5.486000
8	South Africa	5.449333
1	China	4.974000
9	USA	4.970000
7	Nigeria	2.538000
3	Germany	0.808667

Al calcular la pendiente promedio anual de adopción de internet entre 2010 y 2025, se evidencia que países en desarrollo como Etiopía (5.97), Kenia (5.96), Brasil (5.53), Indonesia (5.50), India (5.49) y Sudáfrica (5.45) presentan las tasas de crecimiento más altas en cuanto a penetración de internet, todas por encima del 5% anual. Este ritmo acelerado contrasta fuertemente con el comportamiento observado en países desarrollados como Alemania (0.88) y Estados Unidos (4.97), cuyos valores de pendiente son considerablemente más bajos. Estos resultados indican que los países en desarrollo, aunque partieron desde niveles más bajos de conectividad en 2010, han tenido una evolución mucho más dinámica, reflejando una expansión sostenida en su infraestructura digital y en el acceso de la población a internet. La pendiente promedio funciona como un indicador de velocidad de adopción, y en este caso muestra claramente que los países desarrollados ya se encontraban cerca de su techo de adopción, lo cual limita su tasa de crecimiento, mientras que los países en desarrollo aún tenían amplio margen para crecer. El caso de Etiopía es particularmente notable, ya que no solo alcanza una alta pendiente, sino que lidera el ranking de crecimiento, en cuanto a Nigeria fue un país que su crecimiento % anual fue el más bajo entre los países en desarrollo pero para este último año llegó con una adopción del 83 %.

Cuales son las variables relevantes en cuanto la influencia en la adopcion a internet



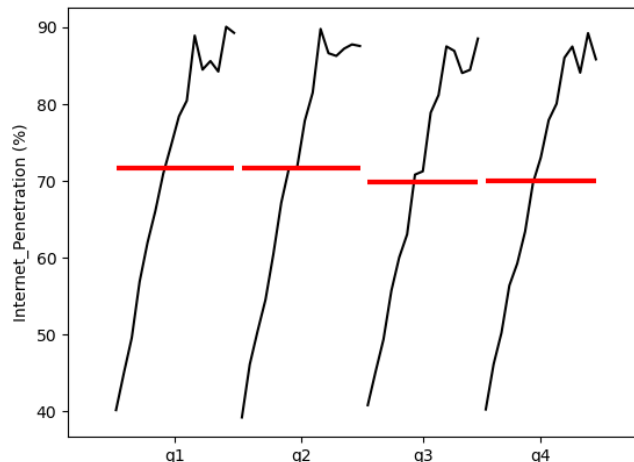
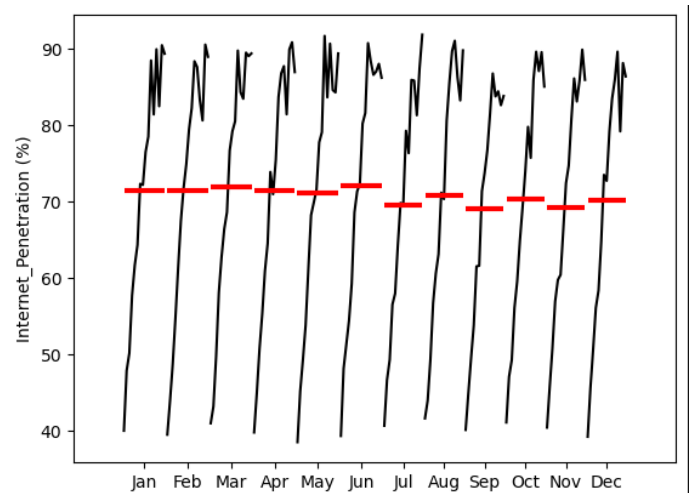
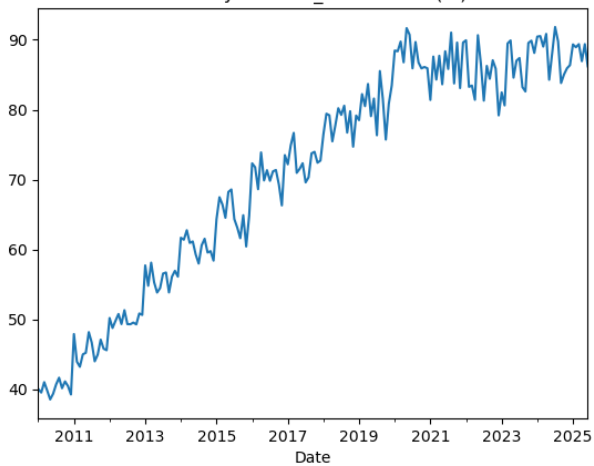
las variables que tienen una relación más fuerte con la adopción de internet son *Mobile\_Data\_Usage* (0.94) y *Broadband\_Speed* (0.92), seguidas por *Digital\_Investment* (0.79). Estas tres variables muestran los valores de correlación más altos con la penetración de internet, lo que indica que a mayor uso de datos móviles, mayor velocidad de conexión, y mayor inversión

digital, también es mayor el porcentaje de adopción de internet en un país. Esto sugiere que la infraestructura técnica y la disponibilidad de servicios son factores clave: cuando las personas utilizan más datos móviles y cuentan con conexiones más rápidas, la penetración crece. Además, la inversión digital mantiene una correlación alta, lo que sugiere que donde hay más recursos destinados a tecnologías digitales, el acceso también es más alto. En cuanto al despliegue de 5G, su correlación negativa y baja sugiere que esta tecnología aún no está asociada a una mayor adopción, lo cual podría deberse a que en muchos países donde hay 5G, la penetración ya es alta y por tanto no hay un crecimiento adicional asociado a esa variable, o bien que su implementación todavía es reciente y no ha impactado significativamente en la adopción global.

En resumen, lo que se observa en la gráfica es que los factores técnicos y de uso están fuertemente ligados a la adopción, mientras que los factores educativos, sentimentales o incluso ciertos avances tecnológicos más recientes como el 5G no muestran una relación clara o positiva.

*¿Cómo ha evolucionado la adopción y cómo predecimos su futuro?*

monthly Internet\_Penetration (%)



### Tendencia General

La evolución de la penetración de Internet entre 2010 y 2025 presenta una tendencia creciente sostenida hasta el año 2020. A partir de este punto, la curva se aplana, lo cual sugiere una posible saturación del mercado, en la que la mayoría de la población objetivo ya ha adoptado el servicio, y el crecimiento marginal se vuelve mínimo.

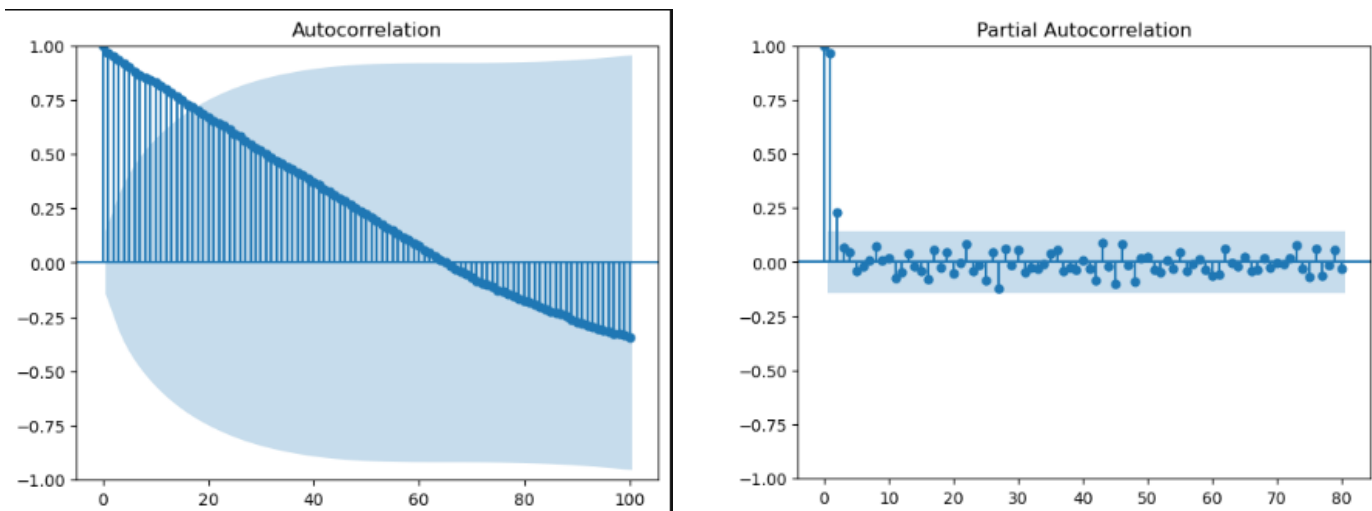
### Estacionalidad y Estabilización

En cuanto a la estacionalidad, los patrones mensuales no evidencian una estacionalidad marcada y sistemática. No obstante, se puede notar una ligera tendencia a que los primeros meses del año presentan niveles de penetración ligeramente superiores en comparación con el segundo semestre. Esto sugiere una estacionalidad débil.

Este patrón se hace más evidente en el gráfico por trimestres, donde se aprecia que el primer y segundo trimestre suelen registrar una mayor adopción en comparación con el tercero y cuarto

### Cuántos registros históricos del pasado influirían en lo que pasaría en el futuro

(Análisis de Autocorrelación y Autocorrelación Parcial)



*El gráfico de autocorrelación (ACF) indica que los valores futuros de la serie de adopción de Internet están fuertemente correlacionados con sus valores pasados, incluso con rezagos de hasta 15-16 meses atrás. Esto significa que el comportamiento pasado de la serie tiene una influencia acumulada importante sobre su comportamiento futuro. Es decir, la serie muestra una alta persistencia y memoria, lo cual es común en fenómenos con crecimiento progresivo como la penetración tecnológica.*

*Por otro lado, el gráfico de autocorrelación parcial (PACF) nos muestra hasta qué punto un valor actual está relacionado directamente con un valor específico en el pasado, sin considerar la influencia de los meses intermedios. En este caso, se observa que la correlación parcial es significativa solo hasta aproximadamente el tercer mes hacia atrás, lo que indica que los valores recientes (últimos 3 meses) son los que realmente tienen un impacto directo sobre el valor actual, mientras que las correlaciones a más largo plazo son indirectas (a través de la influencia acumulada de los meses intermedios).*

### *Correlacion entre variables individuales vs Internet adoption % (y)*

*Se tomaron 2 approaches uno fue el de usar la correlacion de series temporales "lags" CCF con el que se miraria que tan fuerte es la relacion en lags (fechas pasadas) con la serie temporal por variable y el otro approach fue el de usar el test de grangercausality con el se sabia si esa correlacion es significativa o no*

### Conclusion general CCF

*A Education\_Level (%)\_dff y Digital\_Literacy (%)\_dff*

- Correlaciones muy bajas (0.009 – 0.014 en todos los lags)
- Relación lineal prácticamente inexistente con la adopción de Internet.

*B GDP\_Per\_Capita (USD)*

- Correlación moderada baja y estable ( $\approx 0.38$  en todos los lags)
- Asociación constante pero de intensidad media-baja.

*C Mobile\_Data\_Usage (GB)*

- Correlación altísima en lag 0 ( $\approx 0.94$ ) y aún elevada en los siguientes ( $\approx 0.69$ )
- Muy fuerte vinculación inmediata y sostenida con la adopción.

*D Digital\_Investment (M USD)*

- Correlación alta en lag 0 ( $\approx 0.79$ ) y mantenida en los lags posteriores ( $\approx 0.79$ )
- Fuerte asociación consistente a lo largo del tiempo.

*E Broadband\_Speed (Mbps)*

- Correlación muy alta en lag 0 ( $\approx 0.92$ ), luego estable alrededor de 0.64
- Gran relación inmediata, que declina algo pero sigue siendo alta.

### Conclusión general del test de Granger

#### A *Education\_Level (%)\_dff*

-No significativo en lag 1 ( $p = 0.5275$ ).

-Muy significativo a partir de lag 2 hasta lag 12 ( $p \ll 0.01$ ).

*Sugiere un efecto retardado: los cambios en nivel educativo influyen, pero con cierto desfase.*

#### B *Digital\_Literacy (%)\_dff*

-Igual patrón: no significativo en lag 1 ( $p = 0.4158$ ), muy significativo desde lag 2 en adelante.

-Alfabetización digital con impacto diferido.

#### C *GDP\_Per\_Capita (USD)*

-Significativo en todos los lags (1–12).

-El PIB per cápita ejerce un efecto constante y sólido.

#### D *Mobile\_Data\_Usage (GB)*

-Significativo en todos los lags (1–12).

-Uso de datos móviles muy predictivo de la adopción de Internet.



E      *Digital\_Investment (M USD)*

-Significativo en todos los lags (1–12).

-*Inversión digital con fuerte poder predictivo.*

F      *Broadband\_Speed (Mbps)*

-Significativo en lags 1–6.

-No significativo a partir de lag 7.

-*Velocidad de banda ancha tiene un impacto inmediato, pero que decae tras unos meses.*

### *Resumen final*

*1. Variables con poder predictivo y consistencia:*

-*Mobile\_Data\_Usage, Digital\_Investment y GDP\_Per\_Capita.*

*2. Variables con efecto retardado RELACION con “y” baja:*

-*Education\_Level, Digital\_Literacy* (significativas desde lag 2).

*3. Variable de impacto corto plazo ALTA RELACION con “y”:*

-*Broadband\_Speed* (solo hasta lag 6).

### *Pasos a seguir*

Decidi utilizar como variables predictoras para el modelo de serie de tiempo las variables *Mobile\_Data\_Usage*, *Broadband\_Speed*, *Digital\_Investment* ya que estas muestran la relacion mas fuerte tanto en la correlacion normal como en la correlacion por lags aunque , *Broadband\_Speed* empieze a perder poder predictivo apartir del 7 mes en si la naturaleza de esta variable va muy enfocada a la adopcion de internet por ser el ancho de banda osea la velocidad a internet,por eso como primer modelo de prueba o final decidi dejarla.

“ la conclusion final del modelo a elegir se dara despues de entrenarlo y evaluarlo “

### Prediciendo el futuro de la adopcion del internet

#### (SARIMAX)

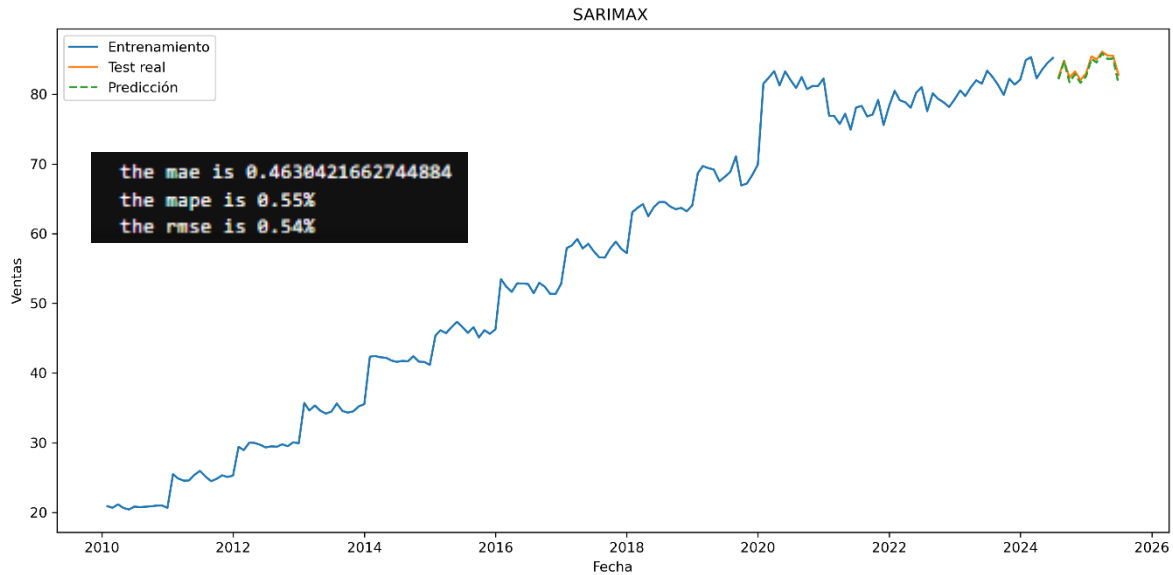
SARIMAX Results						
=====						
Dep. Variable:	Internet_Penetration (%)		No. Observations:		174	
Model:	SARIMAX(0, 1, 1)x(1, 0, [1, 2], 12)		Log Likelihood		-269.759	
Date:	Thu, 17 Jul 2025		AIC		555.518	
Time:	22:17:27		BIC		580.744	
Sample:	01-31-2010		HQIC		565.752	
	- 06-30-2024					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Mobile_Data_Usage (GB)	13.1334	8.698	1.510	0.131	-3.914	30.181
Digital_Investment (M USD)	-0.0555	0.078	-0.711	0.477	-0.208	0.097
Broadband_Speed (Mbps)	-0.1485	0.714	-0.208	0.835	-1.549	1.252
ma.L1	-0.0087	0.471	-0.018	0.985	-0.932	0.915
ar.S.L12	0.9340	0.038	24.828	0.000	0.860	1.008
ma.S.L12	-0.9861	0.074	-13.278	0.000	-1.132	-0.841
ma.S.L24	0.3900	0.063	6.203	0.000	0.267	0.513
sigma2	1.2057	0.059	20.317	0.000	1.089	1.322
=====						
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	12252.55			
Prob(Q):	1.00	Prob(JB):	0.00			
Heteroskedasticity (H):	5.25	Skew:	-0.82			
Prob(H) (two-sided):	0.00	Kurtosis:	44.20			
=====						

Como se puede observar, el modelo no necesit  de las variables ex genas para capturar la din mica temporal de la serie. Aunque el coeficiente m s alto fue el de *mobil\_data\_usage*, con un valor de 13 —lo que indicaría que un aumento de 1 GB en el uso de datos m viles estar  asociado con un incremento de 13 puntos porcentuales en la penetraci n de Internet— este resultado no fue estad sticamente significativo, ya que su p-value fue mayor a 0.05. En general,

*las tres variables exógenas incluidas presentaron p-valores por encima de 0.05, lo que sugiere una baja potencia estadística para explicar el patrón temporal de adopción de Internet. Esto puede deberse a que el propio modelo SARIMAX ya logra capturar suficientemente la dinámica temporal de la serie sin necesidad de depender de las variables exógenas.*

*Los componentes internos del modelo SARIMA —que incluyen los términos autorregresivos (AR), de media móvil (MA) y estacionales— fueron los que sí lograron capturar esa dinámica de manera significativa. El término autorregresivo con 12 rezagos (AR(12)) presentó un coeficiente de 0.93, indicando que el valor actual de la serie depende fuertemente del valor observado hace 12 meses, es decir, del mismo mes del año anterior. Esta relación ya se había anticipado en el análisis previo de correlación (ACF y PACF). Por otro lado, los términos de media móvil  $ma.S.L12$  y  $ma.S.L24$  también fueron estadísticamente significativos, con coeficientes de -0.98 y 0.39 respectivamente, y p-valores menores a 0.05. Esto evidencia que el modelo está corrigiendo de forma activa los errores de predicción ocurridos hace 12 y 24 meses, siendo el rezago de 12 meses (lag 12) el que tiene mayor peso como referencia para ajustar las estimaciones actuales. Esto refuerza la capacidad del modelo para capturar patrones estacionales tanto anuales como bienales. En contraste, el término  $ma.L1$ , que representa los errores del mes inmediatamente anterior, presentó un coeficiente muy bajo (-0.0087) y un p-valor superior a 0.05, lo que indica que no tiene un aporte significativo en la predicción y que el modelo no considera útil corregir errores tan recientes para mejorar sus estimaciones.*

### *Visualización del modelo*



*El modelo logra capturar con notable precisión la tendencia general, la estacionalidad y la variabilidad de la serie. Se aprecia que las predicciones (verde) siguen de manera muy cercana a los datos reales (naranja), lo cual es evidencia visual de un excelente ajuste. No se observan desfases importantes ni errores sistemáticos, y tanto los picos como los valles son replicados de manera adecuada.*

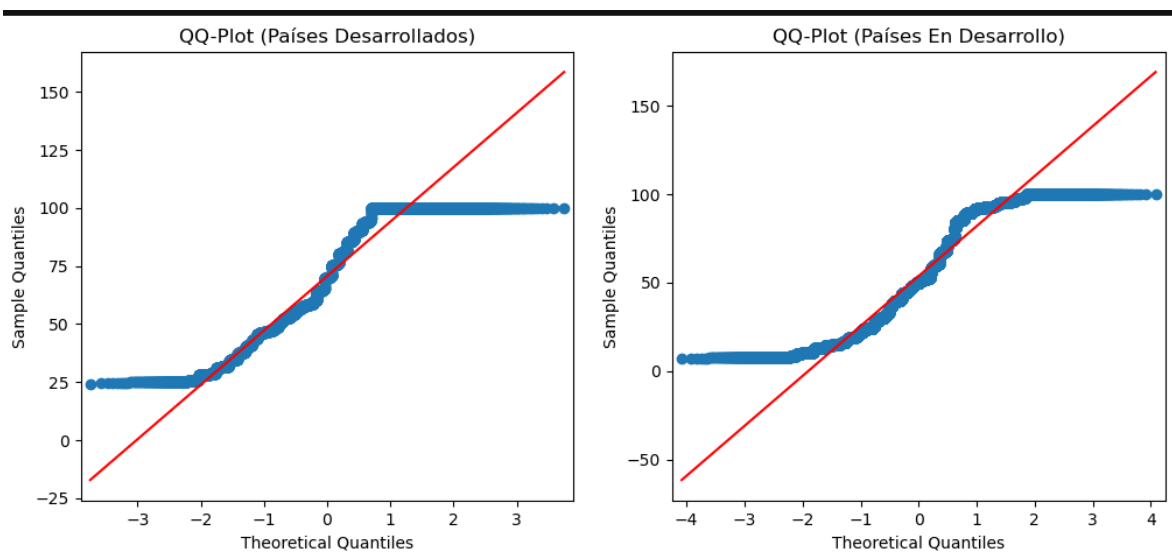
*Como podemos ver en las metricas del modelo esta muy bien ajustado en promedio el modelo se equivoca solo 0.46 puntos en sus predicciones promedio "MAE" o si miramos por medio del error porcentual promedio este fue de 0.55 % osea menos del 1 % lo cual en conjunto estas metricas nos indican el gran performance a la hora de predecir*

*"el modelo esta listo para entrar en produccion "*

*"¿ que tan diferentes son ambos grupos?"*

*"validacion de supuestos (normalidad)"*

*Esta pregunta se respondera Comparando la media de los paises desarrollados vs en desarrollo por medio de la prueba t student con un nivel de significancia de 0.05 % pero primero se tuvo que mirar si ambos grupos son normal o no.*



*Como se observa en el gráfico QQ-Plot, la adopción de internet en ambos grupos no sigue una distribución normal, ya que la línea azul (los datos) se desvía de la línea roja (la normal teórica), especialmente en los extremos. Sin embargo, a pesar de esta falta de normalidad, el Teorema del Límite Central (TLC) nos permite utilizar la prueba  $t^*$ -Student, dado que ambas muestras son lo suficientemente grandes ( $n > 30$ ): 11.272 países desarrollados y 45.088 países en desarrollo. Según el TLC, al trabajar con muestras grandes, las medias muestrales de ambos grupos tenderán a distribuirse de forma aproximadamente normal, lo que valida el uso de la prueba  $t$  para compararlas, incluso si los datos originales no son normales.*

### "Validación de supuestos (homocedasticidad)"

Mediante el test de Levene evaluamos la igualdad de varianzas entre grupos para determinar qué versión de la prueba t-student aplicar:

Si  $p\text{-valor} > 0.05$ : No se rechaza  $H_0 \rightarrow$  varianzas iguales  $\rightarrow$  usar t-student clásica

Si  $p\text{-valor} \leq 0.05$ : Se rechaza  $H_0 \rightarrow$  varianzas desiguales  $\rightarrow$  usar t-student de Welch

resultado:

```
Estadístico t = 65.929, p-valor = 0.0000000000  
Resultado: Existe evidencia estadísticamente significativa de diferencia entre grupos (p value = 0.0000)
```

*"las varianzas -entre grupos difieren significativamente"*

### Prueba t-student Welch

```
# Comparar las medias  
cm = sm.stats.CompareMeans(grupo2, grupo1)  
  
# Calcular el intervalo de confianza (varianzas desiguales) usevar='unequal'  
conf_int = cm.tconfint_diff(usevar='unequal')
```

Utilizando los intervalos de confianza con un 95 % de confianza la diferencia de medias (desarrollados-en desarrollo): (16.4,17.4) lo que implica que los países en desarrollo es entre

16.4 y 17.4 puntos porcentuales menor que la de los países desarrollados. (si hay diferencia de media entre ambos grupos)"

*¿Hay evidencia estadística de que ciertos factores afectan la adopción?*

*Dado que la variable objetivo (y) es continua, y las variables independientes del modelo son en su mayoría numéricas continuas, se optó por emplear un modelo de regresión lineal. Esto permite interpretar los coeficientes como el cambio esperado en la variable de adopción de Internet por cada unidad de cambio en las variables explicativas, manteniendo constantes las demás.*

	features	coeficientes
2	numerical__Mobile_Data_Usage (GB)	20.916558
5	categorical__Urban_Rural_Urban	14.245591
8	remainder__5G_Rollout	-11.627193
3	numerical__Digital_Investment (M USD)	7.190693
0	numerical__GDP_Per_Capita (USD)	5.019504
1	numerical__Education_Level (%)	-1.825511
7	remainder__Digital_Literacy (%)	0.129594
6	remainder__Broadband_Speed (Mbps)	-0.081799
4	numerical__X_Sentiment_Score	-0.006355

*Es el factor con mayor impacto positivo. Cada incremento en el uso de datos móviles por persona se asocia con un aumento de 20.92 puntos porcentuales en la adopción de Internet. Esto resalta que la conectividad móvil es un motor clave, especialmente en regiones con acceso limitado a redes fijas.*

### Urban\_Rural\_Urban

(+14.25)

*Vivir en una zona urbana incrementa en promedio 14.25 puntos porcentuales la adopción de Internet en comparación con zonas rurales. Esto refleja ventajas urbanas en infraestructura, cobertura de servicios y alfabetización digital, lo que facilita la penetración tecnológica.*

### 5G\_Rollout

(-11.63)

*Si hay despliegue del 5G (por ejemplo, más torres instaladas o cobertura técnica), la adopción de Internet se reduce en 11.63 puntos porcentuales. Esto puede parecer contradictorio, pero puede explicarse porque el despliegue de 5G aún está en fases iniciales, y puede estar ocurriendo en países con menor adopción global o grandes desigualdades digitales.*

### Digital\_Investment (M USD)

(+7.19)

*Un aumento de una desviación estandar en digitalización se asocia con un aumento de 7.19 puntos porcentuales en la adopción de Internet. Este resultado destaca la importancia de políticas públicas y privadas que incentiven la infraestructura digital y el acceso a la tecnología.*



GDP\_Per\_Capita (USD)

(+5.02)

*Un aumento de una desviación estándar en PIB per cápita está vinculado a un crecimiento de 5.02 puntos porcentuales en la adopción. El poder adquisitivo mejora la posibilidad de acceder a dispositivos, planes de datos y servicios digitales.*

Education\_Level (%)

(-1.83)

*Un aumento de 1 desviación estándar en el nivel educativo promedio se asocia con una disminución de 1.83 puntos porcentuales en la adopción de Internet. De forma contraintuitiva, la educación muestra un coeficiente negativo no toda educación formal se traduce directamente en competencias digitales.*

Digital\_Literacy (%)

(+0.13)

*Cuando la alfabetización digital aumenta en 1 desviación estándar, la adopción de Internet crece en promedio 0.13 puntos porcentuales. Aunque el impacto es pequeño, es positivo. Esto sugiere que la alfabetización digital es necesaria, pero no suficiente por sí sola, y que debe complementarse con infraestructura y acceso real.*

### Broadband\_Speed (Mbps)

(-0.08)

*Una mejora de 1 desviación estándar en la velocidad promedio de banda ancha produce una leve disminución de 0.08 puntos porcentuales en la adopción. Este resultado sugiere que la velocidad por sí sola no impulsa el acceso, y que podría estar correlacionada con países donde hay buena infraestructura técnica pero desigualdades en el acceso económico o social.*

### X\_Sentiment\_Score

(-0.006)

*El sentimiento (probablemente asociado a percepción pública o narrativa digital) tiene una influencia casi nula, indicando que la adopción se guía más por condiciones objetivas que por percepciones.*

### Conclusion final ---coef de regresion

R-squared:	0.956
Adj. R-squared:	0.956

*El modelo muestra un excelente ajuste ( $R^2 = 0.956$ ) y la mayoría de los predictores son estadísticamente significativos. El uso de datos móviles, el contexto urbano, y la inversión digital*

son los factores con mayor impacto en la adopción de Internet. Mientras tanto, variables como educación formal y sentimiento social y el ancho de banda del internet (velocidad ) tienen un impacto menor o nulo, debido a la falta de conexión directa con el comportamiento de adopción.

### Elasticidad

Tomando en cuenta dos de las variables mas importantes a la hora de explicar la adopcion a internet *mobil\_data\_usage*, *digital investment*, con la elasticidad se mirara que tanto aumentaria la adopcion a internet en general si por ejemplo en este caso la variable *mobil\_data\_usage* y la variable *digital investment* aumenta en un 20 % “

Resultado *digital investment* :

```
Elasticidad constante estimada: 0.9097370156371464
por un aumento del 20 % en 'Digital_Investment (M USD) el internet adoption sube un: 18.194740312742926
```

Resultado *mobil\_data\_usage*

```
Elasticidad estimada: 1.7261
por un aumento del 20 % en Mobile_Data_Usage (GB) el internet adoption sube un: 34.52222021271235
```

### Conclusion

A partir del análisis estadístico realizado, se puede concluir que sí existe evidencia sólida de que ciertos factores afectan significativamente la adopción de Internet en los países evaluados. Mediante el uso de regresion y análisis de elasticidad, se identificaron variables clave como:

- La inversión digital (*Digital\_Investment*)
- El uso de datos móviles (*Mobile\_Data\_Usage*)

- El nivel de urbanización “rural o rubano” (Urban\_Rural\_Urban)

Estas variables mostraron una relación directa y estadísticamente significativa con los niveles de adopción de Internet. Por ejemplo:

- Un incremento del 20 % en la inversión digital se asocia con un aumento del 18.19 % en la adopción de Internet.
- Un incremento del 20 % en el uso de datos móviles resulta en un aumento estimado del 34.52 % en la adopción, lo que refleja un efecto especialmente fuerte.

En cuanto al nivel de urbanización, el resultado `urban_rural_urban = 14.25` indica que, en promedio, los países con mayor proporción de población urbana tienden a presentar niveles de adopción de Internet considerablemente más altos. Esto refuerza la idea de que el acceso a infraestructura en zonas urbanas desempeña un papel fundamental en el cierre de la brecha digital.

*¿Cuál es la probabilidad condicional de alcanzar el 95 % de penetración dado cierto nivel de inversión digital y alfabetización digital?*

Para resolver esta pregunta, se implementó un modelo de regresión logística. Las variables independientes seleccionadas fueron `Digital_Investment` (en millones de USD) y `Digital_Literacy` (%), mientras que la variable dependiente se definió como un indicador binario que toma el valor 1 si la penetración de internet es  $\geq 95\%$  y 0 en caso contrario.

```
y = (df['Internet_Penetration (%)'] >= 95).astype(int)
x = df[['Digital_Investment (M USD)', 'Digital_Literacy (%)']]
```

El proceso consistió en:

1. *Estandarización de características: Las variables predictoras se normalizaron para asegurar que tuvieran media cero y desviación estándar unitaria, mejorando así el rendimiento del modelo.*

```
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X)
```

2. *Entrenamiento del modelo: Se ajustó una regresión logística a los datos escalados para aprender la relación entre la inversión digital, la alfabetización y la probabilidad de alcanzar una alta penetración de internet.*

```
model = LogisticRegression()  
model.fit(X_scaled, y)
```

3. *Predicción para nuevos casos: en el ejemplo se utilizó los valores de (1,944.72 millones USD de inversión y 60.5% de alfabetización digital) para predecir cuanto en probabilidad sería el nivel de absorción teniendo estos dos valores en cada variable*

```
nueva_condicion = [[1944.72 , 60.5]]  
nueva_condicion_scaled = scaler.transform(nueva_condicion)  
  
|  
proba = model.predict_proba(nueva_condicion_scaled)[0][1]  
  
print(f"Probabilidad estimada de alcanzar ≥95%: {proba:.2f}")  
  
Probabilidad estimada de alcanzar ≥95%: 0.77
```

*Como resultado dio que habria un 77 % de probabilidad de que un pais en cierto tiempo "fecha" tuviera un 95 % si ese pais tuviera 1,944.72 millones USD de inversión y 60.5% de alfabetización digital*

*"Apartir de este modelo entrenado se puede seguir prediciendo la adopcion a internet según los valores en inversionUSD y alfabetizacion digital que se requiera"*

*"la exactitud de prediccion del modelo es de el 92 % "*

```
y_pred = model.predict(X_scaled)
accuracy = accuracy_score(y, y_pred)
print(f"Exactitud de la prediccion: {accuracy:.2f}")
Exactitud de la prediccion: 0.92
```

### Pregunta bonus

¿Si se sabe que un país ha superado el 90 % en 2024, ¿cuál es la probabilidad de superar el 97 % en 2025?

*Para responder esta pregunta, se estimó una probabilidad condicional. Primero, se identificaron los países que en el año 2024 alcanzaron una tasa de penetración de Internet igual o superior al 90 %. Luego, se filtraron aquellos países que, en 2025, superaron el 97 % de penetración, pero considerando únicamente los países que ya habían estado en el grupo anterior (los que superaron el 90 % en 2024).*

*A partir de esto, se calculó la proporción entre el número de países que cumplieron ambas condiciones (más del 90 % en 2024 y más del 97 % en 2025) y el total de países que superaron el 90 % en 2024. Este valor representa la probabilidad condicional estimada de que un país que ha superado el 90 % en 2024 también supere el 97 % en 2025.*

*Resultado:*

*La probabilidad condicional estimada fue de 0.40, lo que indica que aproximadamente el 40 % de los países con alta penetración en 2024 lograron superar el 97 % al año siguiente.*

```
grupo_90_2024 = df[(df['Year'] == 2024) & (df['Internet_Penetration (%)'] >= 90)]['Country']  
grupo_95_2025 = df[(df['Year'] == 2025) & (df['Country'].isin(grupo_90_2024)) & (df['Internet_Penetration (%)'] > 97)]  
prob = len(grupo_95_2025) / len(grupo_90_2024)  
print(f"Probabilidad condicional estimada: {prob:.2f}")  
Probabilidad condicional estimada: 0.40
```

### *Conclusion final del proyecto*

*A lo largo de este proyecto de investigación, se analizaron los factores determinantes en la adopción de Internet, comparando países desarrollados con países en desarrollo. Una de las primeras evidencias empíricas encontradas fue la diferencia promedio de 17 puntos porcentuales en el nivel de adopción digital entre ambos grupos. Este hallazgo fue respaldado mediante una prueba t de Student, la cual confirmó, con un intervalo de confianza del 95 %, que la diferencia de medias se encuentra entre 16.4 y 17.4 puntos porcentuales, lo que refuerza la existencia de una brecha digital estructural significativa.*

*Además, se observó una alta variabilidad en la penetración de Internet dentro del grupo de países en desarrollo (variación del 53 %), en comparación con una menor dispersión en países desarrollados (33 %). Este hallazgo refleja la heterogeneidad interna entre los países en desarrollo en cuanto al acceso digital.*

*Un fenómeno notorio fue el impacto estructural de la pandemia de COVID-19 (2019–2020), el cual aceleró la adopción digital en todos los países analizados. Sin embargo, aunque los niveles de adopción se han mantenido relativamente altos en 2025 (por encima del 81 %), no se han*

*vuelto a alcanzar los picos del 90 % registrados durante la pandemia. Este desajuste evidencia que el impulso forzado hacia la digitalización no se tradujo en una estabilización estructural a largo plazo.*

*No obstante, se identificó una tendencia de crecimiento sostenido en los países en desarrollo, con una tasa promedio anual de incremento del 5–6 %, mientras que en países desarrollados como Alemania el crecimiento es marginal (menor al 1 % anual). Esto sugiere que, si se mantiene esta dinámica, los países en desarrollo podrían eventualmente cerrar la brecha digital en los próximos años.*

*Desde una perspectiva temporal, no se identificaron patrones estacionales claros en la adopción mensual de Internet, aunque se detectó una leve mayor adopción en el primer semestre del año.*

*En términos predictivos, se logró ajustar exitosamente un modelo SARIMAX que permitió proyectar la adopción futura con una precisión alta, manteniendo un error medio absoluto (MAE) menor al 1 % en horizontes de hasta un año, lo cual indica que la trayectoria futura de la adopción digital está bien definida a corto plazo.*

*El análisis multivariable identificó los factores más determinantes en la adopción digital. Entre ellos destacan:*

- Uso de datos móviles (GB): Principal indicador, reflejando la importancia del acceso a Internet desde dispositivos móviles como canal dominante de conexión.*
- Nivel de alfabetización digital: Especialmente diferenciado entre zonas urbanas y rurales, donde las ciudades muestran un avance mayor.*
- Inversión digital anual: Se confirmó como un predictor robusto de la infraestructura y penetración tecnológica.*

*Finalmente, al evaluar las probabilidades condicionales de alcanzar un 97 % de adopción, se encontró que, entre los países que ya habían superado el 90 % en 2024, la probabilidad de llegar al 97 % en 2025 es del 40 %. Este resultado sugiere que es altamente probable que los países desarrollados logren una adopción casi universal en los próximos años, consolidando así una transición completa hacia lo digital.*



## *Recomendaciones*

- 1. Priorizar la inversión en infraestructura digital en zonas rurales: Los datos revelan que una de las principales barreras en los países en desarrollo es el bajo acceso en regiones rurales. Se recomienda direccionar esfuerzos a garantizar cobertura y conectividad de calidad en estas zonas.*
- 2. Fortalecer programas de alfabetización digital: El nivel de adopción de Internet está altamente correlacionado con el grado de habilidades digitales de la población. Implementar programas educativos y capacitaciones masivas contribuiría a acelerar el cierre de la brecha.*
- 3. Promover el acceso económico a dispositivos móviles y datos: Dado que el uso de datos móviles fue el principal predictor de adopción digital, facilitar el acceso asequible a planes de datos y dispositivos conectados se vuelve crucial.*