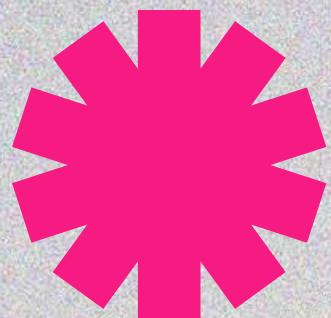


ANÁLISIS DE INFRAESTRUCTURA DE INTERNET EN SUDAMÉRICA

para Expansión Estratégica de Servicios



Angela Daniella Barrutia Rios

SOBRE EL PROYECTO

Abstract

El presente proyecto tiene como objetivo analizar la infraestructura de Internet en países de Sudamérica con el fin de identificar los mercados prioritarios y más estratégicos para la expansión de una empresa peruana de servicios de Internet.

Motivación

La motivación surge de la necesidad de tomar decisiones estratégicas basadas en datos, optimizando recursos y minimizando riesgos al momento de desplegar servicios en nuevos países.

Audiencia

Enfocado en equipos de Estrategia de Negocio, Infraestructura y Marketing/Ventas. Su objetivo es validar crecimiento, priorizar inversión y optimizar campañas. La lectura es apta para perfiles técnicos y no técnicos.

Se utilizaron datos obtenidos mediante la **API de Cloudflare Radar**, en el periodo anual de 2023 y 2024 para considerar solo años ya finalizados, incluyendo métricas de velocidad de descarga y subida, latencia, jitter y la concentración de ASNs por país. Se realizó un análisis exploratorio de datos (EDA) para comprender la distribución de estas métricas y detectar patrones relevantes, complementado con técnicas de Machine Learning no supervisado, específicamente clustering mediante KMeans, para agrupar países según la calidad de la red y la estructura del mercado.

DATOS Y FUENTES API

HIPÓTESES Y PREGUNTAS

Países con mejor velocidad y menor latencia

Hipótesis: Los países con mayor velocidad promedio de descarga y menor latencia son los más adecuados para la expansión inicial de la empresa.

Preguntas:

- ¿Qué países presentan las mejores métricas de velocidad, latencia y jitter?
- ¿Cómo se comparan las velocidades promedio de descarga y subida entre los países analizados?

Datos utilizados: df_speed → barplots de download, upload, latency y jitter; scatter plots multivariados

Concentración de ASNs por país

Hipótesis: Los países donde uno o pocos ASNs concentran la mayor parte de las IPs facilitan estrategias de entrada más focalizadas.

Preguntas:

- ¿Qué países tienen mercados dominados por pocos ASNs?
- ¿Cuál es la participación de los principales ASNs por país para diseñar estrategias de entrada?

Datos utilizados:

df_asn → Top 5 ASNs por país, participación relativa (%) de los ASNs

Relación entre latencia, jitter y velocidad

Hipótesis: A mayor latencia y jitter, menor velocidad de descarga y subida, afectando la consistencia de la red.

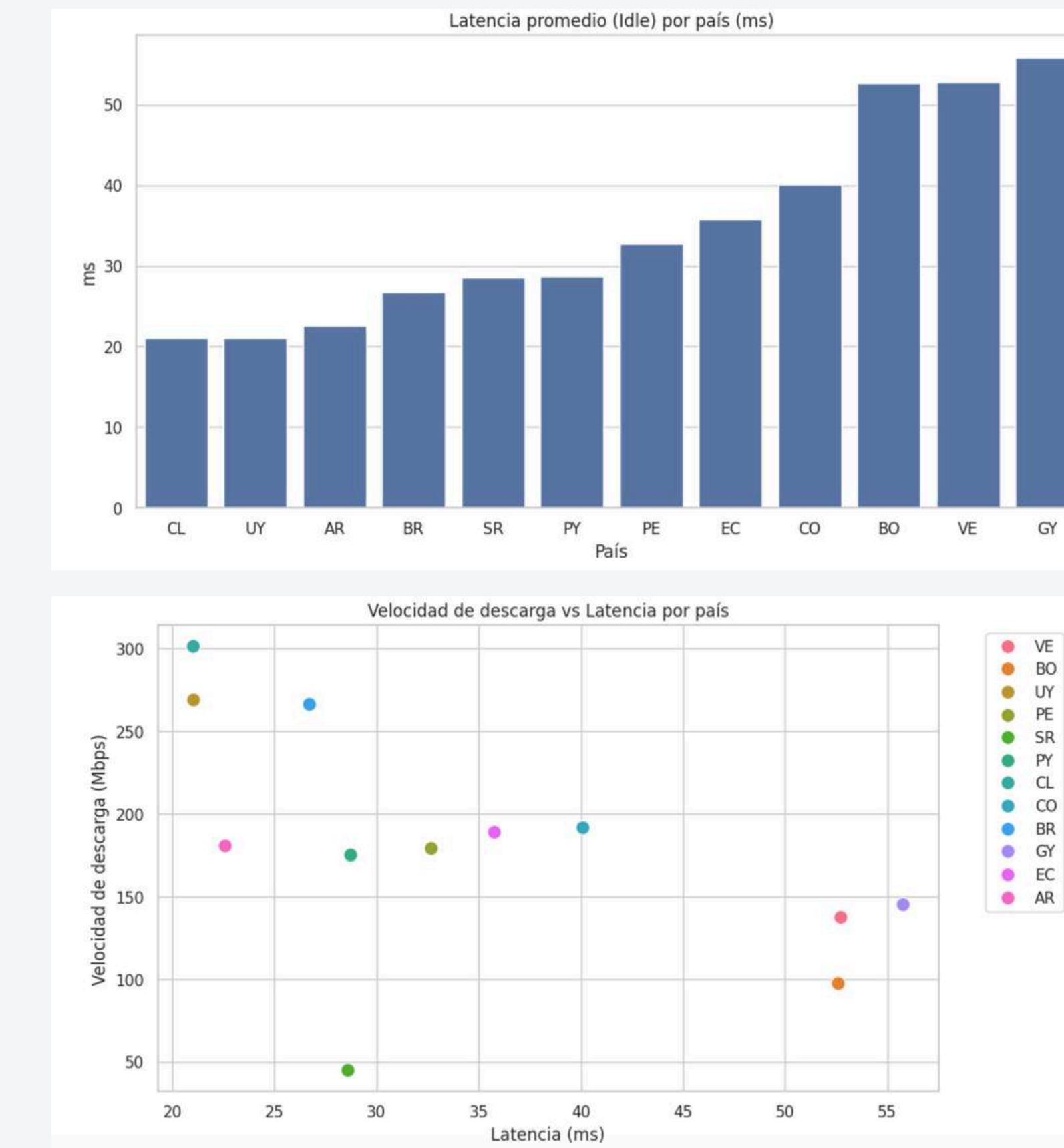
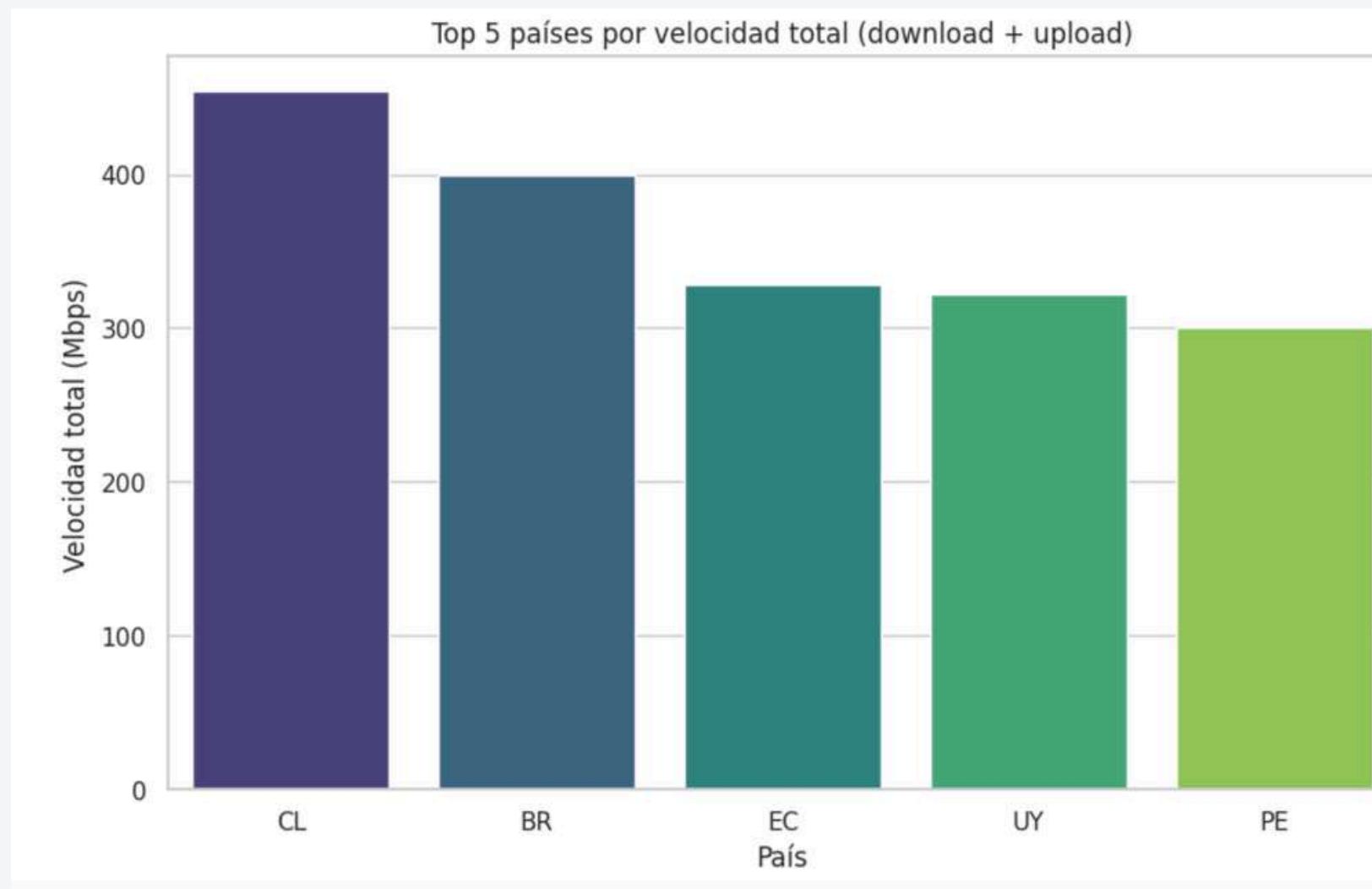
Preguntas:

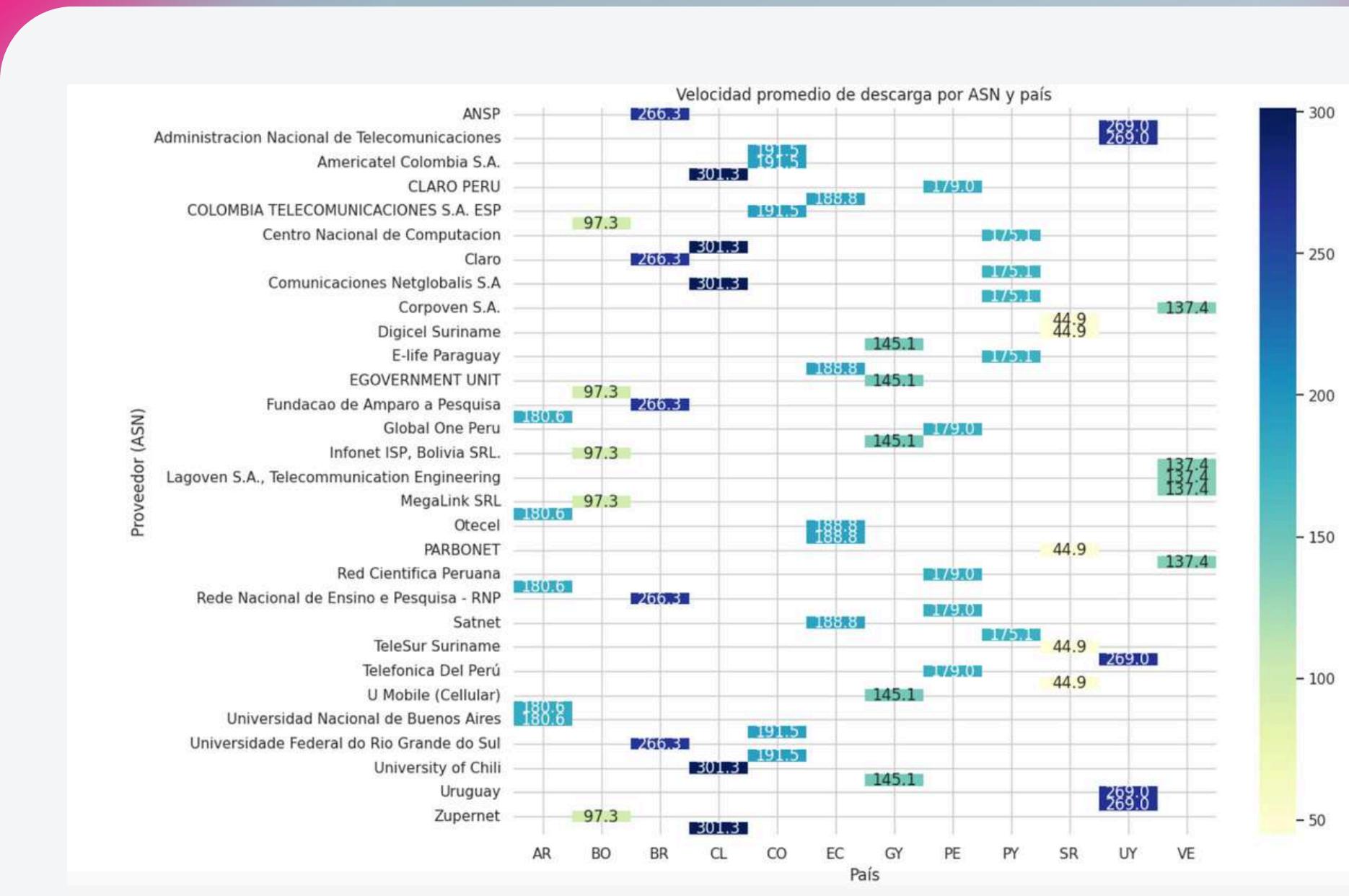
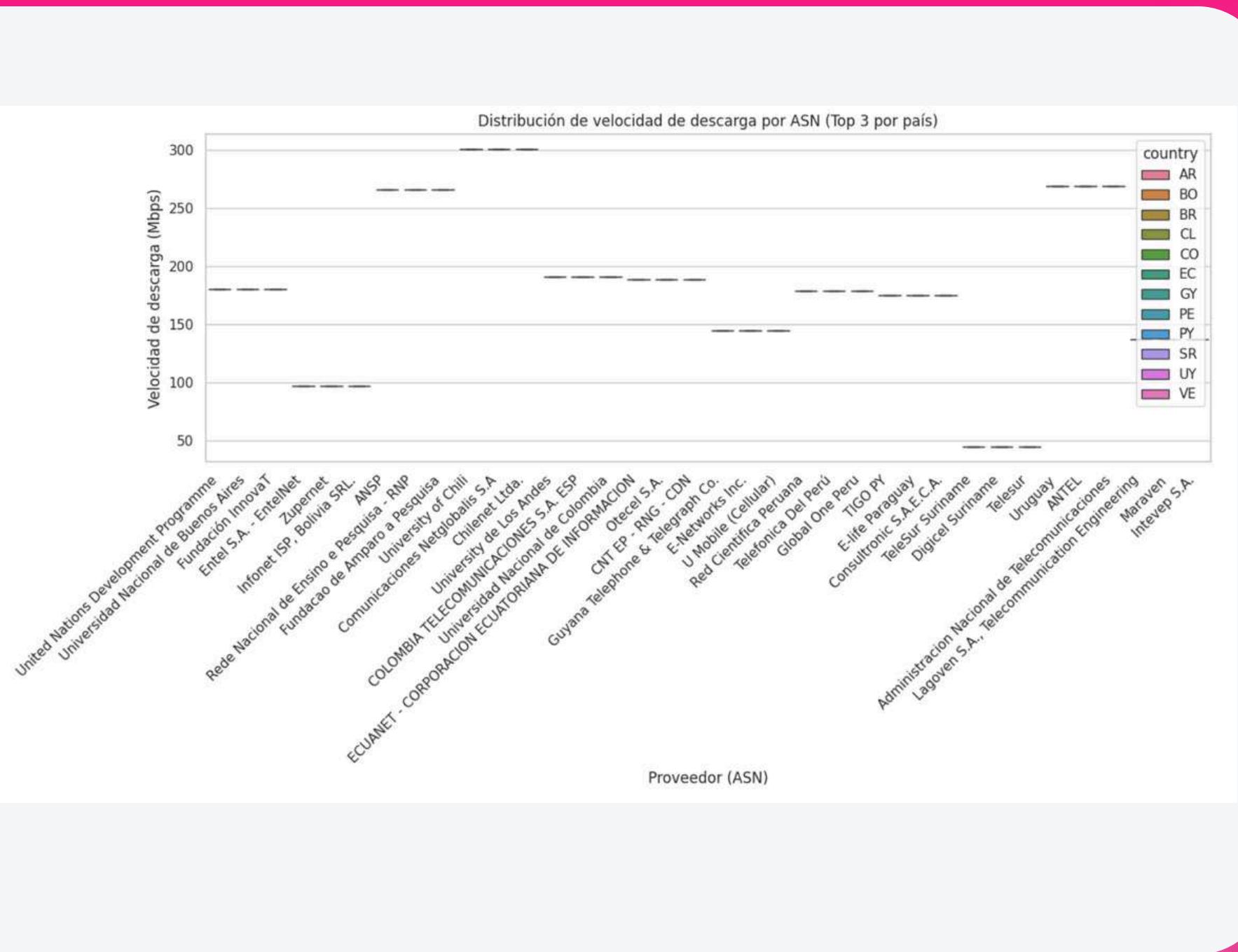
- ¿Existen correlaciones significativas entre latencia, jitter y velocidad por país?
- ¿Qué países presentan mejores métricas conjuntas de latencia, jitter y velocidad?

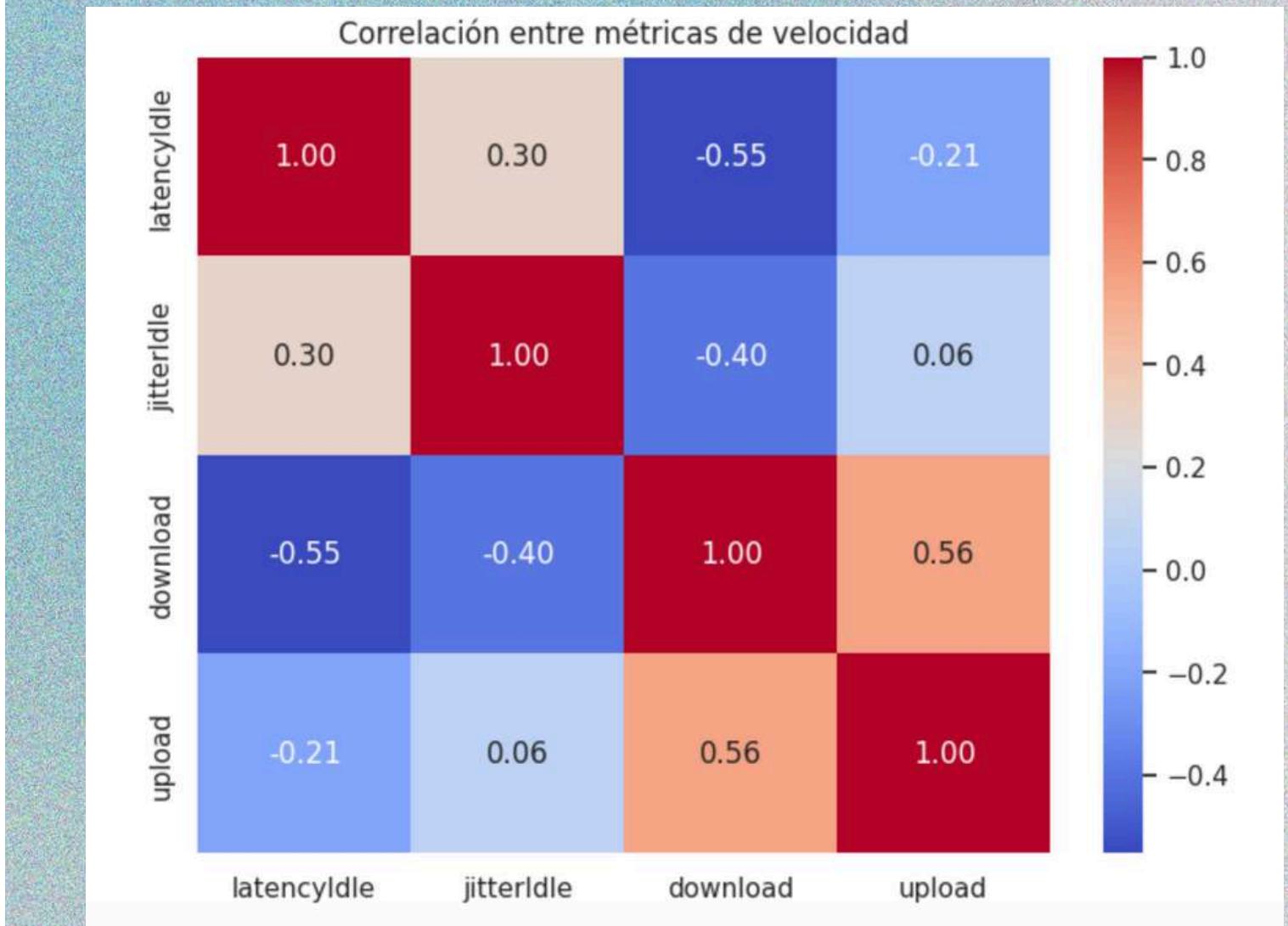
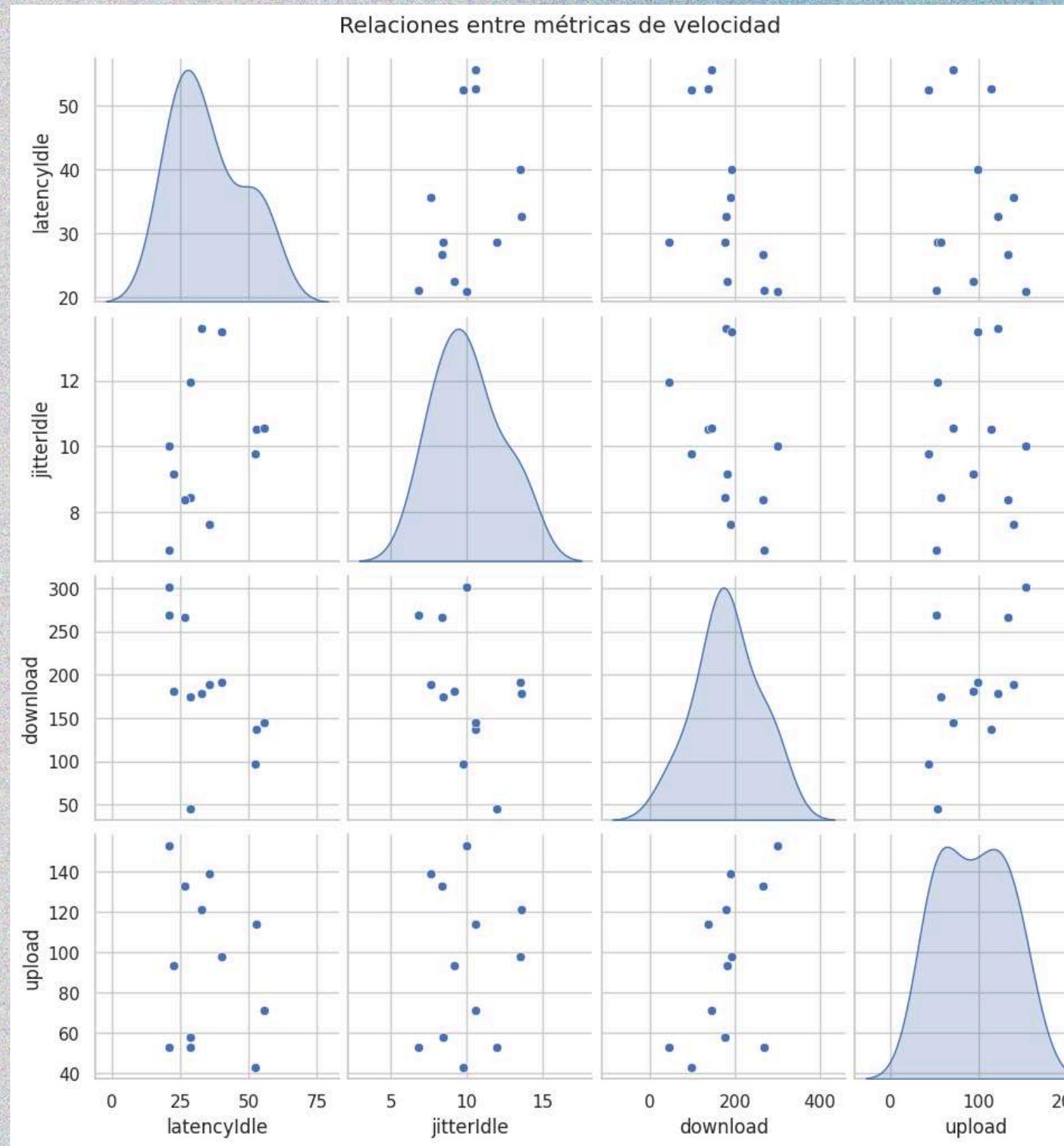
Datos utilizados:

df_speed → scatter plot multivariado y heatmap de correlaciones

ANÁLISIS DE VELOCIDAD Y LATENCIA

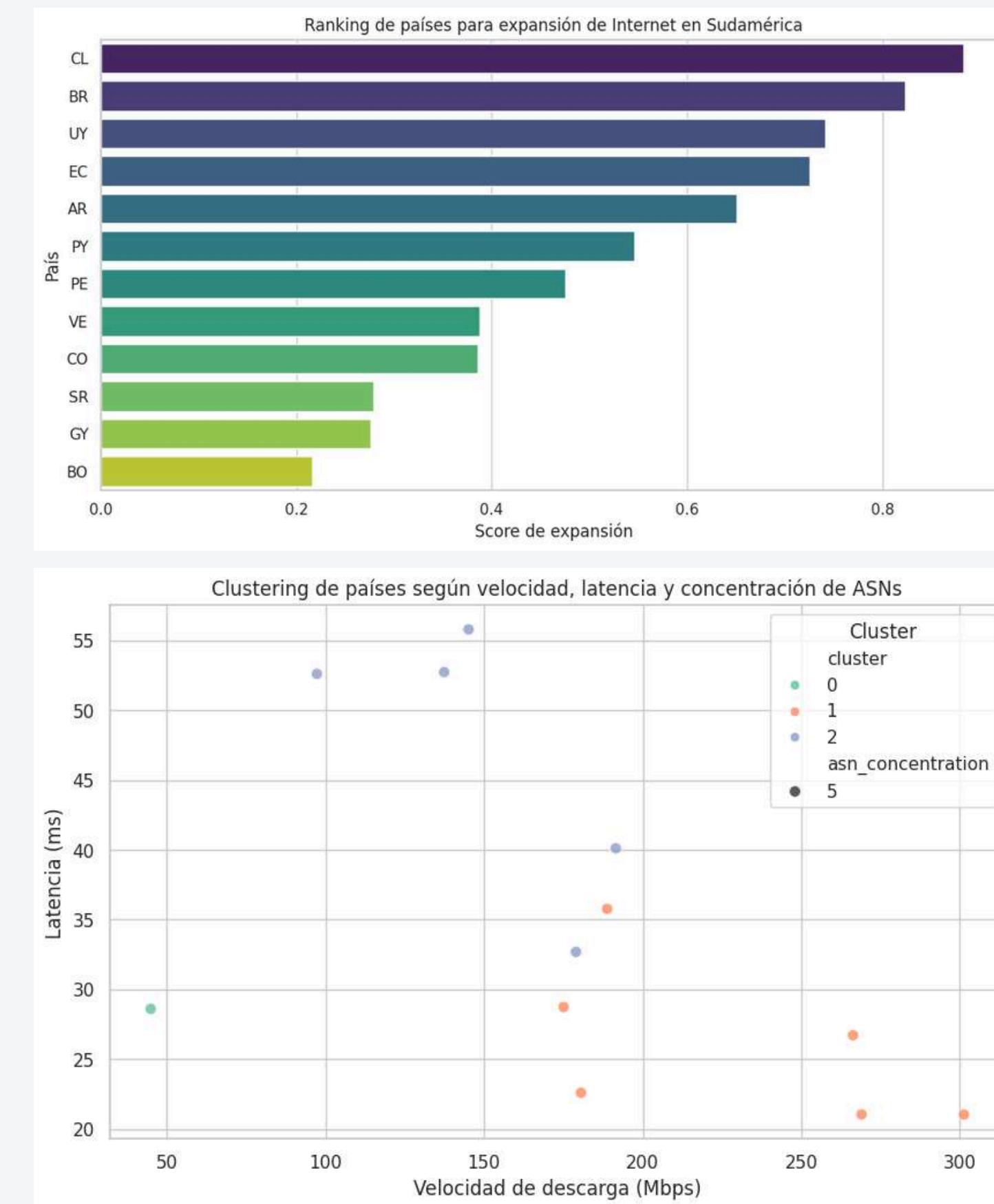
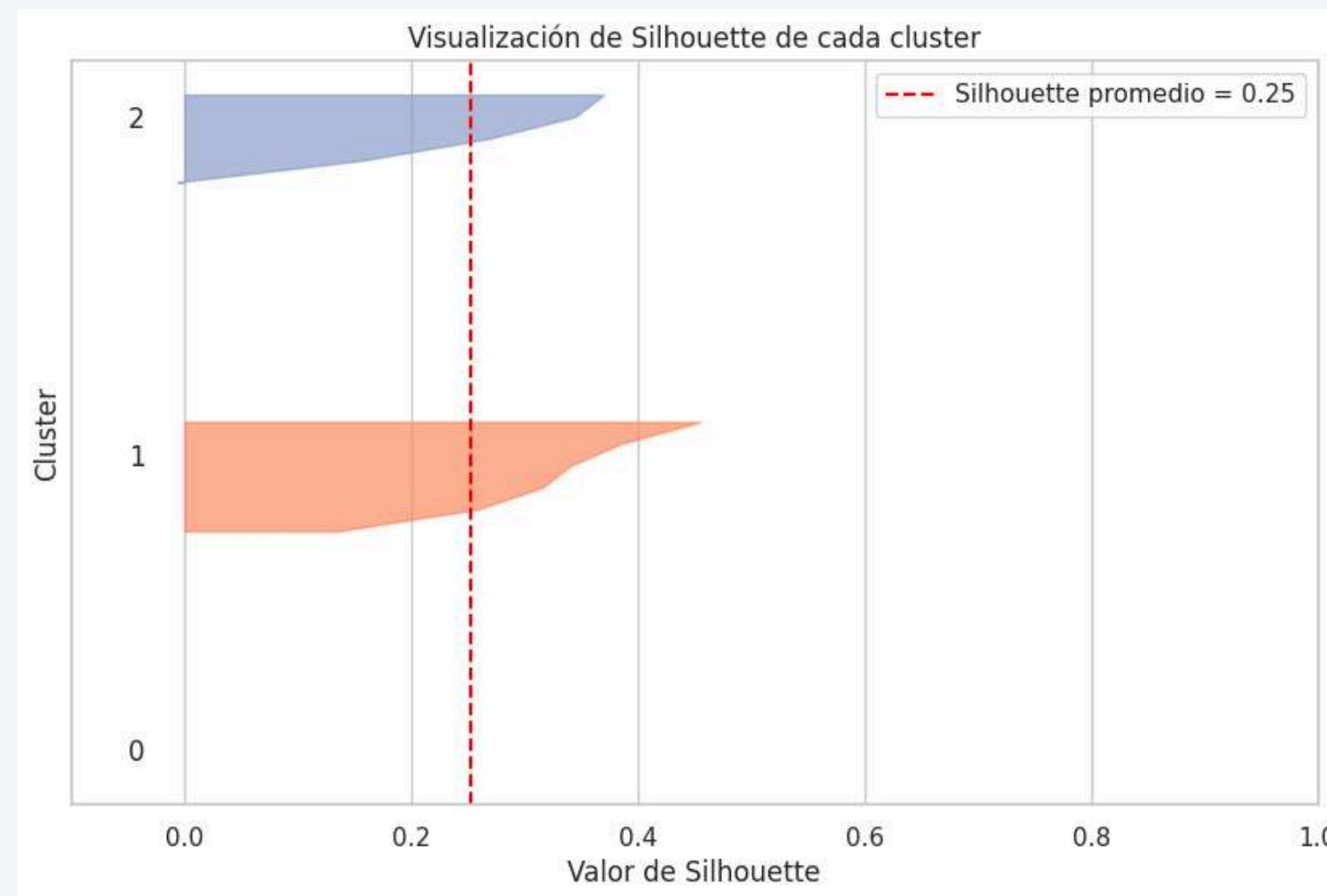






ANÁLISIS DE LATENCIA, JITTER Y VELOCIDAD

MODELADO ML: CLUSTERING



INSIGHTS PREELIMINARES

Países con mejor rendimiento de red

- Perú y Chile muestran las mayores velocidades promedio de descarga y subida y baja latencia, destacando como mercados sólidos para la expansión inicial.
- Colombia y Brasil presentan buenas velocidades, pero con mayor variabilidad en latencia y jitter.

Concentración de proveedores (ASNs)

- Algunos países muestran que uno o pocos ASNs concentran la mayoría de las IPs, facilitando estrategias de entrada mediante alianzas estratégicas.
- Países con ASNs más fragmentados (Brasil, Argentina) requieren negociaciones con varios proveedores.

Relación entre latencia, jitter y velocidad

Se observan correlaciones entre latencia, jitter y velocidad de descarga/subida. Países con menor latencia y jitter relativo a la velocidad ofrecen una red más consistente y confiable.

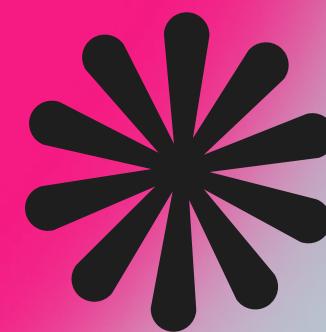
Priorización de países basada en ML

El clustering y ranking de expansión confirman que Perú y Chile son los mercados más adecuados para la expansión inicial. Uruguay y Argentina son opciones secundarias por buena infraestructura y estabilidad de red. Venezuela y Bolivia presentan bajas velocidades y alta latencia, indicando que la expansión allí necesitará inversión significativa.

Validación del modelo ML

La evaluación con Silhouette Score y visualización de siluetas confirma que los clusters están bien definidos y permiten distinguir claramente países con alta, media y baja calidad de red.

El modelo de ML proporciona un marco objetivo para priorizar mercados y tomar decisiones estratégicas basadas en datos.



Proyecto Completo:

<https://github.com/Angebarrutia/Coderhouse-Curso-Data-Science-II/tree/main/Entregas/Primera%20Entrega>