

TALLER 2

El objetivo de este taller es evaluar las capacidades del estudiante para aplicar técnicas de regresión, las cuales permita descubrir *insights* y sugerir accionables al negocio. No olvide realizar los ejercicios previos pertinentes respecto a la preparación y entendimiento de los datos.

Contexto de negocio: Apoyo del Banco Mundial a países en desarrollo.

El Banco Mundial es una organización internacional que se dedica a brindar apoyo financiero y técnico a países en desarrollo mediante préstamos para proyectos de infraestructura, sectoriales, institucionales, entre otros, buscando mejorar el bienestar económico y social de su población. El diseño de políticas de préstamos más informadas y adaptadas a las necesidades y realidades de cada país, no solo maximiza la probabilidad de que los préstamos sean reembolsados en su totalidad, sino que también asegura que los fondos se utilicen de manera efectiva en las necesidades reales del territorio.

Históricamente, el Producto Interno Bruto (PIB) ha sido uno de los indicadores clave para evaluar la capacidad de pago de un país. Sin embargo, si bien existen reportes anuales del PIB de los países, aún no se entiende a la perfección con qué otros indicadores como tasa de empleo, acceso a internet, expectativa de vida, entre otros, se encuentra relacionado.

Como consultor científico de datos del Banco Mundial, se le ha dado la responsabilidad de desarrollar análisis que permitan dar respuesta a esta necesidad. Se le ha proporcionado un dataset construido por la organización [Gapminder](#), quienes han hecho la labor de, a partir de múltiples fuentes, integrar indicadores por país para un periodo aproximado de tiempo. A continuación encuentra el dataset y el diccionario de datos correspondiente:

- [Dataset](#)
- [Diccionario](#)

Mediante el uso de modelos de regresión lineal, en conjunto con técnicas de preparación y exploración de datos, se espera que usted esté en capacidad de extraer y entender las relaciones que puedan existir entre el PIB de un país y otros indicadores disponibles.

ACTIVIDADES Y ENTREGABLES

1. **[20%] Entendimiento de los datos:** Genere un breve reporte de entendimiento de datos en donde destaque las dimensiones del dataset, los tipos de datos que contiene y el top 5 de los indicadores que considera más importantes para el ejercicio de regresión propuesto. Por cada indicador, incluya algunos elementos básicos de su comportamiento o distribución (análisis univariado). También incluya algunos elementos complementarios que permitan entender la relación entre pares de indicadores, el país o la región a la que pertenecen (análisis bivariado).
2. **[25%] Preparación de datos:** Realice un análisis de calidad de datos teniendo en cuenta las diferentes dimensiones de calidad y determine las estrategias más apropiadas para su limpieza. Como estrategia de imputación de datos, se recomienda revisar el portal de Gapminder en donde se encuentra mayor detalle de la información contenida en el dataset. Complementariamente, evalúe la posibilidad de complementar el dataset con información categórica relativa al continente o a la región a la que pertenece, si es un país desarrollado o subdesarrollado, entre otros.
3. **[30%] Entrenamiento del modelo de regresión:** Utilizando los indicadores que considere más relevantes, además de las variables categóricas construidas, entrene un modelo de regresión lineal que prediga el PIB per cápita de un país. No olvide realizar las transformaciones de datos necesarias y la validación de supuestos para asegurar no solo un buen ajuste del modelo sino también una correcta interpretación de los coeficientes.
4. **[25%] Generación de resultados:** Redacte un informe ejecutivo o una presentación corta en donde muestre los hallazgos o *insights* más importantes en términos de entendimiento de los datos, métricas de error e interpretación del modelo. Responda a la pregunta: dado que un alto PIB puede representar un buen comportamiento de pago de un país, a partir de las relaciones más importantes que, según el modelo de regresión, existe entre el PIB y otros indicadores, ¿qué conjunto de políticas públicas recomendaría implementar, a partir de la premisa de que la mejora en estas áreas indicaría al Banco Mundial que el país es estable, está en una trayectoria de desarrollo sostenible y tiene la capacidad de administrar y reembolsar préstamos de manera efectiva?

FORMATO DE ENTREGA

- El taller debe desarrollarse de forma individual.
- Todos los recursos generados deben entregarse mediante un repositorio de GitHub con su estructura apropiadamente documentada mediante el archivo README.

- **El repositorio de GitHub debe ser público.**
- El o los notebooks generados deben poderse ejecutar secuencialmente sin ningún error. En el caso de haber creado varios notebooks, se debe clarificar el orden en que deben ser ejecutados. No olvide también incluir el dataset utilizado.
- Se penalizará a los estudiantes que no cumplan con este formato de entrega.