

TEST DE PROGRAMACIÓN – JR. DATA ENGINEER

INSTRUCCIONES

Crear un pipeline con Python, el cual debe:

1. Extraer los datos correspondientes al total de población de cada país a través de los años usando la API JSON de Indicadores del Banco Mundial.

Documentación:

<https://datahelpdesk.worldbank.org/knowledgebase/articles/889392-about-the-indicators-api-documentation>

URL para extraer información sobre totales de población:

<http://api.worldbank.org/v2/country/all/indicator/SP.POP.TOTL?format=json>

Nota: Puedes usar el parámetro page=n para obtener la página número n.

2. Almacenar los datos extraídos en un objeto DataFrame de la librería Pandas.
3. Pronosticar la población de cada país usando los modelos que creas más convenientes. Debes crear pronósticos para los próximos 4 años, utilizando un tamaño de 4 para el periodo de validación. Para cada país y modelo utilizado, es necesario reportar los errores de tipo MSE, MAE y MAPE.
4. Guardar los resultados en dos archivos CSV:
 - forecasts.csv – Pronósticos generados para cada país y modelo utilizado.
 - errors.csv – Métricas de error obtenidas por cada modelo y país.
5. Analizar los resultados y proporcionar conclusiones generales sobre el desempeño de los modelos que utilizaste. Siéntete libre de usar herramientas de visualización de datos para presentar tus resultados.

Haz tu mejor esfuerzo para crear código simple, leíble, eficiente y modular. Debes enviar tu código y resultados al correo sebastian.gonzalez@ventagium.com antes de tu entrevista.