

Prueba técnica Data Scientist

Objetivo: Evaluar las capacidades del candidato en aspectos técnicos relacionados al proyecto al que aspira.

Secciones:

1. ETL(20%)

Para esta tarea, utilice el siguiente set de datos para el desarrollo de este numeral:

https://drive.google.com/file/d/1AR6tx6g_vmkB-3_IUhCOsCGVDD0N2sf4/view?usp=sharing

- Desarrolle un pipeline que permita procesar el conjunto de datos. El pipeline debe incluir pasos de limpieza, transformación. Realice un breve análisis estadístico y de correlación y proponga estrategias para mejorar el CSAT (Customer satisfaction), SLA (Service Level Agreement) y AHT (Average Handle Time)

2. NLP (20%)

Para esta tarea, utilice el siguiente set de datos, utilice el 30% de los datos para probar el desarrollo de este numeral y el del punto 5:

<https://huggingface.co/datasets/alexcom/analisis-sentimientos-textos-turisitcos-mx-polaridad>

- Realice un análisis de sentimiento en un conjunto de reseñas de productos. Debe crear un modelo capaz de realizar una clasificación de sentimiento

3. Series forecasting (20%)

Para esta tarea, utilice el siguiente set de datos, utilice el 20% de los datos para probar el desarrollo de este numeral:

<https://huggingface.co/datasets/pkr7098/time-series-forecasting-datasets>

- Desarrolle y entrene un modelo de Forecasting, que reciba como datos de entrada el histórico de un día completo y devuelva un vector con un horizonte de pronóstico de una semana completa, en intervalos de tiempo de **2 horas**. Calcule las métricas de error del modelo y muestre sus resultados.

4. Speech to Text (20%)

Para esta tarea, utilice el siguiente set de audios:

https://huggingface.co/datasets/charris/hubert_process_filter_spotify/tree/main

- Desarrolle un algoritmo que permita descargar y transcribir los siguientes archivos de audio.

5. Entregable final (20%)

Asuma que todos los sets de datos anteriores hacen parte de un mismo contexto:

- Cree un informe dinámico e interactivo que presente los resultados de las tareas anteriores. El informe debe ser claro, conciso y permitir la toma de decisiones efectivas con facilidad.

Evaluación:

Entregue una carpeta con todos los archivos (Notebook (.ipynb) con el desarrollo de los puntos 1,2,3,4 y entregue un archivo de PBI, Looker o alguna herramienta de su elección para el punto 5) adicionalmente prepare una presentación gerencial de 15 minutos para presentar sus resultados.

Cada tarea se evaluará en función de la precisión, la eficiencia y la claridad de las soluciones del candidato. La prueba también evaluará la capacidad del candidato para comunicar sus hallazgos de manera efectiva **(-50%)**. Recuerde presentar entregables ordenados y bien documentados, también será considerado dentro de la evaluación **(-20%)**.