Trully

Data Scientist Challenge

Welcome to the **Trully** challenge for **Data Scientists.** From the interview we had, we can already tell that you're a very knowledgeable professional that we would love to call a teammate.

For this challenge you will need to design an end-to-end solution to do the following:

- See the attached data **"Challenge CDMX Mapa (precio m²).csv"**, this file contains the neighborhood, municipality and their average price per square meter (real estate market value). Some of the prices have been removed on purpose.

- Enrich the data with information that might help create a prediction model. *Try government and open data!*

- Clean the data and perform EDA on the data. Remember to remove outliers.

- Create features from the information you collected that might explain the value of the price.

- Create a model to predict the average price per square meter for the missing neighborhoods.

- Present the results, and show the **most significant variables** in the prediction.

- Propose how you will productivize the model using an open-source middleware. Show that you can run inferences to the model **using latitude and longitude as an input**. *How would you infer in new zones that hold no value?*

- Propose how you would maintain and monitor the performance of the model.

Use preferably open-source solutions.

PRO TIP: Explore geospatial information, government databases and use tools such as geopandas to make joins, enrich and predict based on geospatial information..

We wish you the best of luck.

Trully.