

ZestyAI Data Scientist - Analytics Role Homework Assignment

At ZestyAI, we've developed an AI-powered wildfire risk model capable of predicting the likelihood of a property being affected by wildfires. The model assigns a risk score ranging from 1 to 10, where 1 indicates the lowest risk and 10 indicates the highest.

For this assignment, you'll work with the provided dataset ([wildfire_risk_data_set](#)), which contains details of 6,000 properties in the US, including their latitude, longitude, and the wildfire risk scores predicted by our model for 2018 and 2022.

Your Tasks:

1. **Evaluate the 2018 Predictions:**

Using data from 2019 US wildfires, evaluate how well the 2018 predictions align with actual wildfire occurrences. The wildfire data is available from the USGS (United States Geological Survey) website.

2. **Analyze Model Drift:**

Compare the 2018 risk scores with the 2022 scores to assess how the model's predictions have changed over time.

This exercise is designed to evaluate your skills in:

- finding and integrating external data sources
- exploratory data analysis (EDA)
- geospatial data handling
- statistics and machine learning model valuation
- programming and communication

Submission guidelines:

- Present your work (preferably written in Python) in a Jupyter Notebook that includes
 - Data sources (e.g., links to external data)
 - Clear EDA steps with visuals
 - Insights and explanations demonstrating your thought process
- Submit your completed notebook to the link provided in the original email
- You have **7 days** from the date of receiving this assignment to complete and submit your work.