[redacted] Data Science Challenge

The purpose of this exercise is to provide you with a problem similar to the types of challenges we face at Aurora Insight, so you will have a better idea about the types of data and problems we work on. The purpose for us is to see how you approach a problem, how you structure a solution in code, and how you communicate your thinking. The performance or sophistication of your solution does not matter.

Consider the problem described below and spend some time devising and implementing a solution in code. We prefer Python but use whatever language or tools you're comfortable with. Please don't spend more than a couple hours developing a solution. We suggest that you start with a simple solution and then discuss alternatives and areas for improvement. If you have any questions about the problem or would like some help getting started, please let us know. One of our goals is to gauge how we would collaborate, so we value discussion.

Lastly, if you already have an example project solving a data problem with code, you are welcome to submit it instead of this. We've found discussing example work to be very valuable, but we value everyone's time, so we don't want you to work on this project if it's duplicative. You're also welcome to publish your solution to this problem and use it in your portfolio as long as you remove references to [redacted].

Data:
spectrogram.tif
This file contains a 2D array of float values which are the frequency domain representation of one part of the radio spectrum. The shape of the array is 2097 rows by 8000 columns. The first dimension of the array (rows) represents time, which spans 0 to 20ms. The most important dimension is the second (columns), which represents frequency and spans 1000 MHz to 1800 MHz. The values in the array are related to radio signal strength.

channel-metadata.csv
This file contains information about different frequency "channels" in the *spectrogram.tif* file. A "channel" refers to a range of frequencies in the radio spectrum that are used for some purpose. For example, a digital television station might broadcast using frequencies 470 – 475 MHz. Each row of this file is a channel with corresponding ID, start frequency, end frequency, and occupation. Whether a channel is "occupied" or not is based on signal strength. Note that the channels cover most, but not all of the frequencies represented in *spectrogram.tif*.

Problem:
Given the data in *spectrogram.tif*, develop a method/algorithm to predict the "occupied" attribute for each channel in *channel-metadata.csv*.

Submit:
1) A brief description of your solution
2) Any code and instructions for running it if necessary
3) A brief discussion (~1 paragraph) of the strengths and weakness of your solution and any ideas for improvements.