**Write a SQL query that identifies cities with higher than average home prices when compared to the national average. Output the city names.**

SELECT city

FROM zillow\_transactions a

GROUP BY city

HAVING avg(a.mkt\_price) >

(SELECT avg(mkt\_price)

FROM zillow\_transactions)

ORDER BY city ASC

We use the HAVING clause in the above solution whose inner query calculates the average market price using the avg(); the column *mkt\_price* is selected from the table and averaged for the purpose of comparison.

### You are given a dataset on cancer detection. You have built a classification model and achieved an accuracy of 96 percent. Why shouldn't you be happy with your model performance? What can you do about it?

In the dataset on cancer detection, we come across the problem of imbalanced dataset as there is an uneven distribution of classes i.e the number of people who actually have cancer is way less than the number of people who do not have cancer. The 96% accuracy rate might be only predicting the majority class correctly. But our focus here should be the correct prediction of the minority class i.e people who actually have cancer. Therefore accuracy of the model cannot be considered as a measure of its performance.

Here, the sensitivity and specificity (True positive and True negative rates) must be considered to measure the performance of the model. The F1 score, which is the harmonic mean of precision and recall can also be used to determine the class-wise performance of the model.

### What is CNN and what are the different layers in it?

CNN stands for Convolutional neural network. Similar to neural networks, it has neurons that can receive many inputs, takes the weighted sum of each neuron's input and passes it through an activation function. There is also a loss function associated to it at the end. CNN has 4 main layer. They are:

* Convolutional layer - This layer performs the convolution operation i.e different feature maps are convoluted over the dataset.
* ReLU layer - Rectified Linear Unit layer is an activation function layer that smoothens the inputs from convolutional layer by converting all negative pixels to 0.
* Pooling layer - This layer reduces the dimensionality of the stack of outputs from the activation layer.
* Fully connected layer - This layer predicts the image and classifies objects in it.

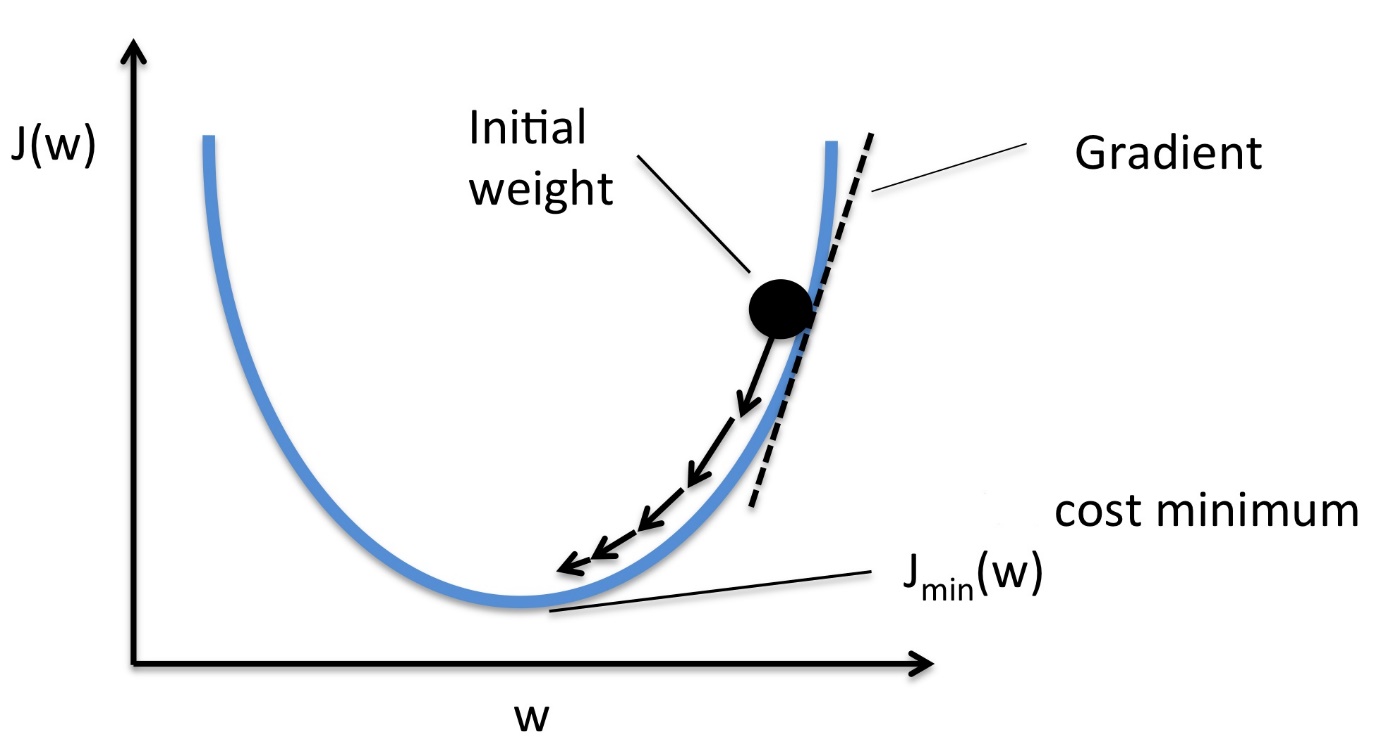
**Explain Gradient descent.**

A gradient can be thought of as the slope of a function. Basically, it measures the amount of change in the output function when the inputs are changed a little bit.

Gradient descent is an optimization algorithm used to find values of parameters of an activation function that minimizes the function. In layman's language, it can be understood by the following 4 steps:

1. Pick a random point.
2. Find its slope.
3. Move towards the downhill direction.
4. Repeat steps 2-3 until the minimum is reached.

Calculating gradient requires looking at every single data point.



**4. How are the time series problems different from other regression problems?**

Every time a problem contains a time variable, it is not necessarily a time series problem. For example, we have a housing dataset that contains the year in which the houses were built, area, locality and we are required to predict their prices. Even though the year here is a time variable, it is a regression problem.

For a problem to be a time series problem, we are looking into a target variable that changes with time. One more characteristic of a time series problem is that the the observations that are closer in time are similar than the observations that are distant in time span. For example, the the market trends of today and tomorrow might be similar. But the market trend 4 months ago won't be similar.

To summarize, time series problem is extrapolation whereas regression problem is interpolation.

**5. Explain content-based and collaborative filtering in recommender systems.**

Both collaborative and content based filtering are different methods to build a recommender system.

In content-based filtering, recommendations are given to users based on the properties of the content the user is interested in. For example, if a user purchases a certain art piece and gives it good ratings, it tells us that the user is interested in such art pieces. So if more such art pieces are recommended to the user, there is a higher chance of the user liking it and purchasing it.