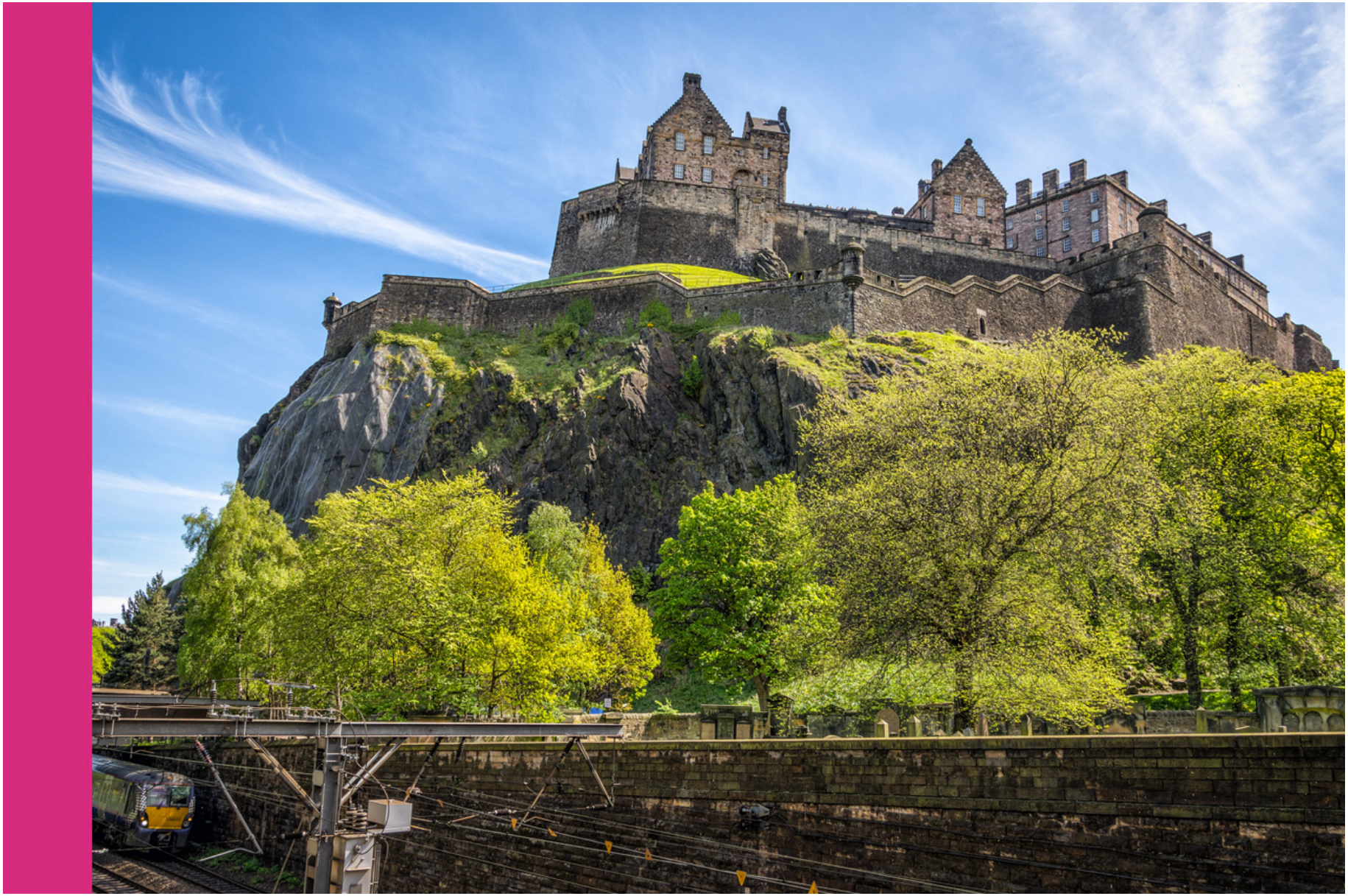
Predicting visitor count

Predicting the visitor count of Edinburgh Castle

Now that we are all setup, let's look into what predictive modelling is all about. In a tailored example, we will discover the different aspects that are important to analyse a certain problem description, and how modelling can help us.

The scenario

You are the marketing director of Edinburgh Castle, standing proudly on Castle Rock for centuries, and you are tasked with analysing the visitor behaviour for marketing purposes. You would like to have an idea of how many people will visit in the following years to determine visitor sustainability alongside the preservation and space constraints that are imposed on you. Furthermore, you want to continue to tailor your marketing campaigns. Getting a grasp of your customers will allow you to see what types of people are more likely to visit the castle, respond to marketing campaigns, and so on. You are aware that visitors come to the castle for different purposes: tourism, historic research, patriotism, and so on.



Edinburgh Castle (2018) by Gary Campbell-Hall on Flickr, licensed under CC BY 2.0 (modified image)

Many businesses are familiar with situations like ours. It is not hard to bend this scenario into something closer to home. It is also not dissimilar to what Amazon does, but on a much bigger scale.

Now, let's take some time to think about all the different components of this exercise. First of all there is you, a director and decision maker. Secondly, there are the different tasks:

- Predicting how many visitors will come in the next few years;
- Profiling your customers; and
- Segmenting those customers.

Starting your analysis

Whenever you start your analysis, it is important to keep two things in mind:

1. We should not analyse or build models for building models' sake.
2. With so many applications being marketed and talked about nowadays, it is tempting to immediately involve automated learning.

In general, it is helpful to take a few steps back and use common sense. Most importantly, this will help you set out the scope which you need to structure your predictive activities.

Breaking it all down

So, you now have your goals, tasks and your research questions. Now you need to break them down. Predicting how many visitors will come in the next few years requires a few things. First of all, there is the assumption that a prediction is possible. Most importantly in the context of this course, it is assumed that we can predict the future based on observations of the past, possibly extended with our own insights. Secondly, it requires that past observations are available. This data also needs to be ample enough to construct a strong model. Simply put, you cannot build a model without data. Although many techniques exist for smaller datasets, our data-driven business environments nowadays are driving researchers to make better predictions by making use of huge amounts of data.

In our case, it would be nice to have data of the last 20 years, with a visitor count perhaps measured yearly, monthly, or perhaps even daily and hourly. If we see that our visitor count is growing annually, we can assume that it will continue to do so in the future. If we can extract a trend, it might give us a good clue about where we will end up next year, in two years, and perhaps even in five years.

We might also have to consider other things: perhaps tourism is growing rapidly in Scotland, but not necessarily in Edinburgh. We might see that the highlands are attracting hikers and nature enthusiasts, which might explain the visitor count to some extent for people making a stop on their trip in Edinburgh. If this visitor count is declining as of this year, we might have to take this into consideration as well.

Highlands, Scotland (2017) by Gregorio Puga Bailón on Flickr, licensed under CC BY 2.0 (modified image)

Furthermore, trends are not a golden ticket: if Brexit suddenly prevents people from entering the UK for tourist purposes (however unlikely), or if Glasgow opens another historic site with a nicer castle, the predictions might have no connection with reality due to disruptive events.

Interpreting the results

Gathering data and building models is one thing. The most important part of your analysis comes at the end: interpreting the results. You can build as many intricate models that have very good performance, but if we cannot explain its implications or discuss the nuances in the results, it becomes useless in many cases. Especially in a business environment, one needs to be able to motivate the directions that the outcome of an analysis is pointing towards to convince management that this makes sense. Again, common sense helps out a lot in this situation. The rise of deep learning and machine learning has brought along many modelling approaches that are very competitive in terms of performance, but not in terms of interpretability. Widely used statistical techniques such as linear and logistic regression are often a good enough alternative which still offers insights into the factors driving the results.

If we can accurately predict that the visitor count is likely to increase by 15.67%, but it is impossible to tell what factors are causing this. It might be the popularity of Edinburgh as a tourist destination, interest in the castle, the continuous efforts to renovate the building, or the number of single malt whisky bottles sold annually or so on. Our analysis is somewhat limited in terms of usability.

In summary

Regardless, both sides of this story will be covered in this course, and the picture is not black and white. There is ample research into uncovering the driving factors in deep learning models as well, although they are not as straightforward. Then again, many statistical models are struggling when dealing with many applications due to their underlying assumptions.

This sums up what you can expect when taking the course: investigating what is appropriate in various situations, making trade-offs, and discovering the implications of the models that are constructed.

Learn About Verified Certificates