Hopefully, we can all agree that data science isn't boring, but if not I hope we will have convinced you by the end

Sometimes it feels like a bit of a wild west when it comes to data. Lots of data in lots of places. In amongst it all, there is definitely some value and insight to help us with our daily work of reaching audiences. But getting at it, organising it, connecting it up and making sense of it can be more work than many of us have the time and resources to be able to do.

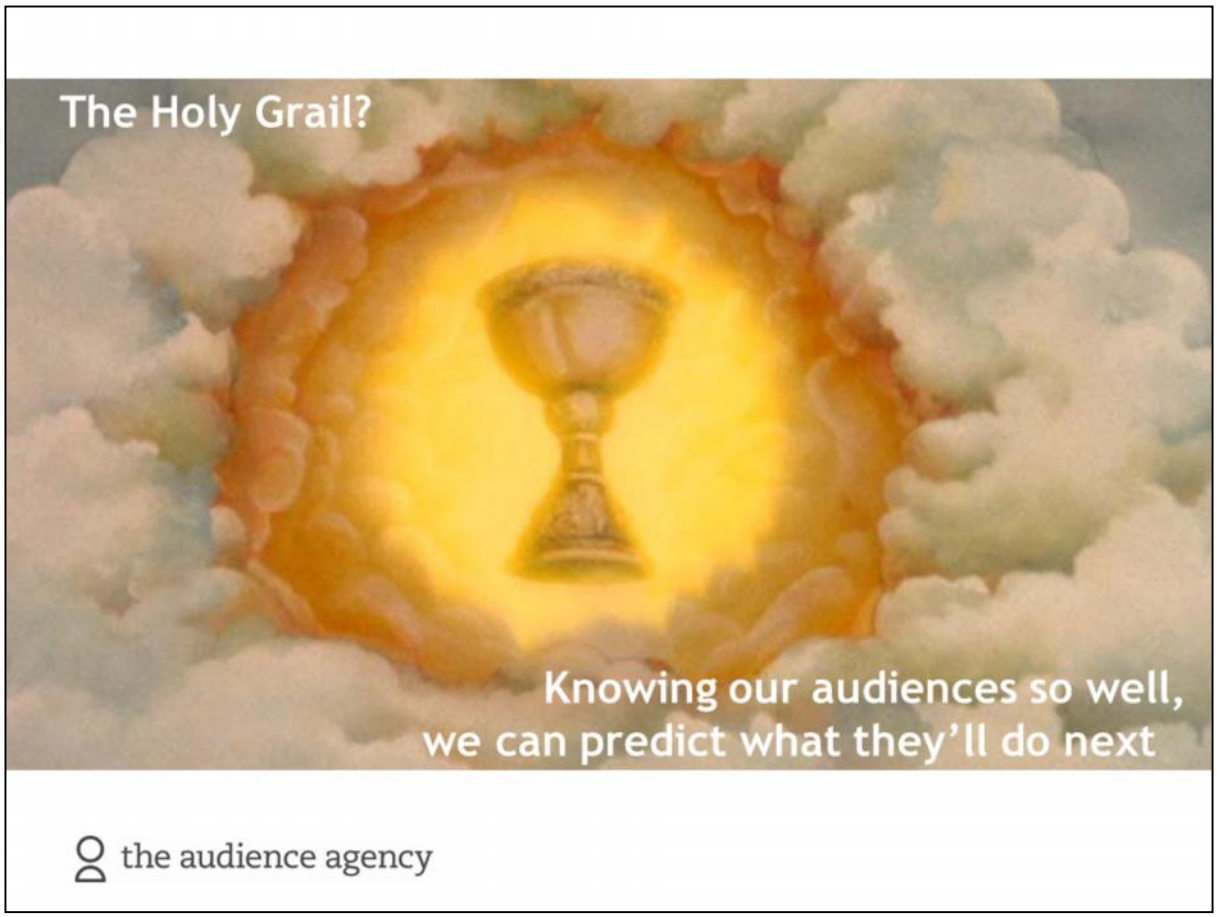
Not only that, but once we do have it, do we have the decision-making processes to be able to act on that data

There's no use us knowing our audiences down to the last minute detail, but that having no impact on how our organisations behave

Many of you will be familiar with the Digital R&D Fund for the Arts, but the Arts Data Impact or ADI project was not really like the other projects. The Arts Data Impact project spent a year working with ENO, NT and Barbican to bring data science techniques to bear on decision-making and in particular to look at the possibilities of Data Driven Decision-making (DDD) that Big Data could unleash in the arts..

We aren't like supermarkets or telecoms, our offer is different and our customers behave differently. But we can still use data science to have a direct impact on day-to-day decision-making, be that digital marketing, fundraising or programming.

In order to bring data science to life, we embedded the first ever data scientist in-residence programme witht the partners and developed some prototype tools that provide solutions to some common data-driven questions. A couple of which we're going to talk to you about today.



At the beginning of the project, we asked all the participants what were the most important questions they wanted to be able to answer

One of the most common question was to better understand their audiences in a more rounded and holistic way. Whether that was better knowledge of the audiences for non-ticketed activity such as art exhibitions.

Or whether that was being able to link bookers and social media followers. The holy grail for most of us seems to be a 360 degree view of our audiences.

The beauty of a data science approach is to look at the patterns and relationships between those patterns to give us 360 degree insight about our audiences, and even predict their behaviour.

The “problem” with new audiences

the audience agency

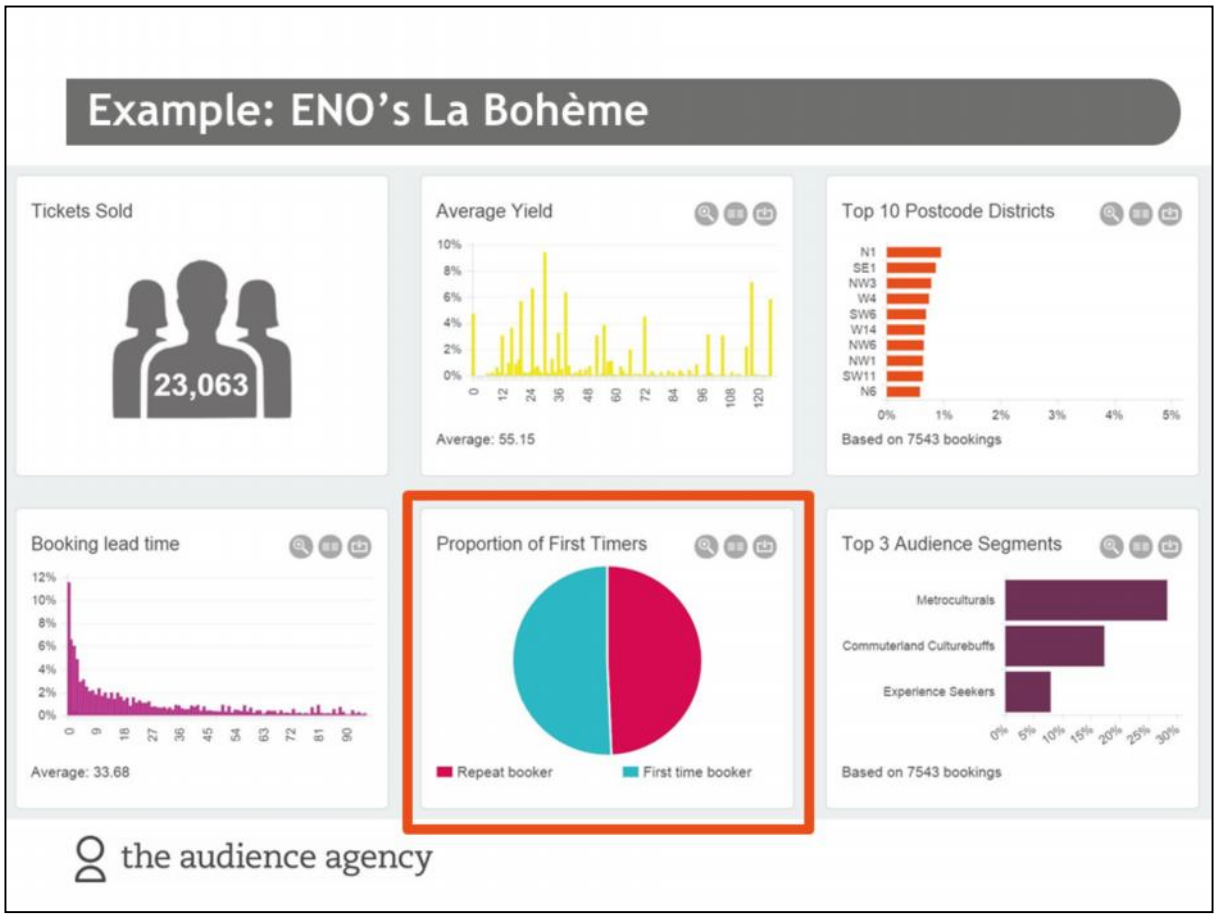
Not only is it difficult to even try to know every single detail of your audiences, but this is doubly difficult with those audiences who are new to your organisation.

ENO’s primary concern was how to address the volume of new one-off audiences, or as I call it, the problem of first timers.

If 50% of your audience are new to you, how can you understand what they’re like at all, let alone get a 360 degree understanding

- Shallow depth of information of individuals
- Limited ability to analyse buying behaviours
- Limited ability to segment
- See audience only as ‘transactors’, not fully rounded people
- Limited data points to capture more information

Therefore even the most data rich organisation (such as ENO) has limited data on much of its audience.



What can we know simply from the transaction data?

Over 23k tickets

Over 50% of bookers from the top three most engaged Audience Spectrum groups – NB Audience Spectrum is a segmentation of the UK population according to their propensity to engage with arts and culture

Many bookers from London

Over 50% (53.59) first time bookers

That's all well and good, but how do we flesh that out and make it more detailed? This is where the power of the aggregate comes in – something we like to call "elsewhere analysis"

How do you solve a problem like Maria?

Maria's night at ENO - what the data tells us

- Her first booking at the London Coliseum
- Out for an evening with her guest at *La Bohème*
- It's likely to be a special night:
  - She planned it 6 months in advance
  - She chose the best seats in the house
  - She treated them to champagne in the interval
- Lives in Islington, North London

the audience agency

Remembering that we're talking about people here, let's imagine one of ENO's audience members for La Boheme

We'll call her Metrocultural Maria.

Maria Loves arts and culture BUT her first time at the London Coliseum

Out for an evening with her partner at *La bohème*

It's likely to be a special night:

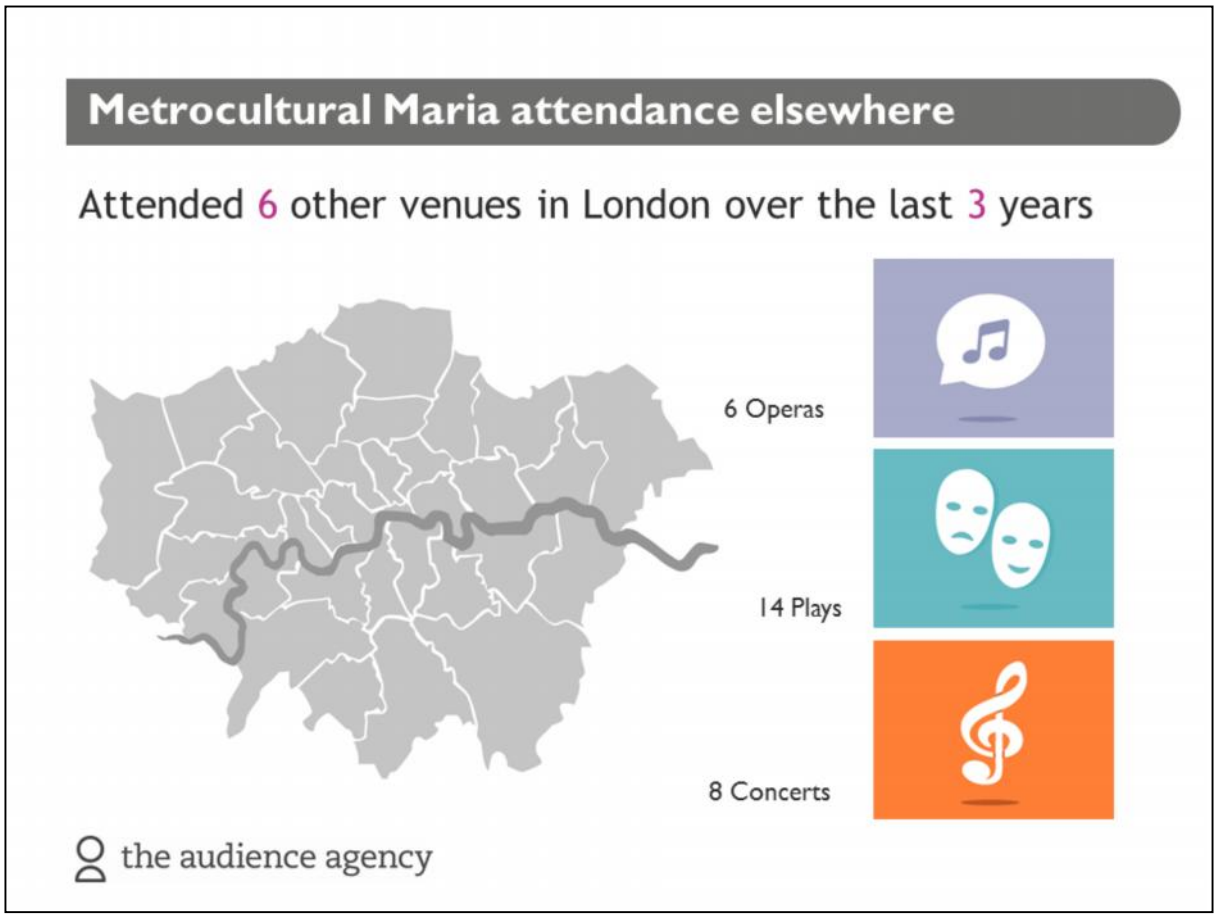
She planned it 6 months in advance

She chose the best seats in the house

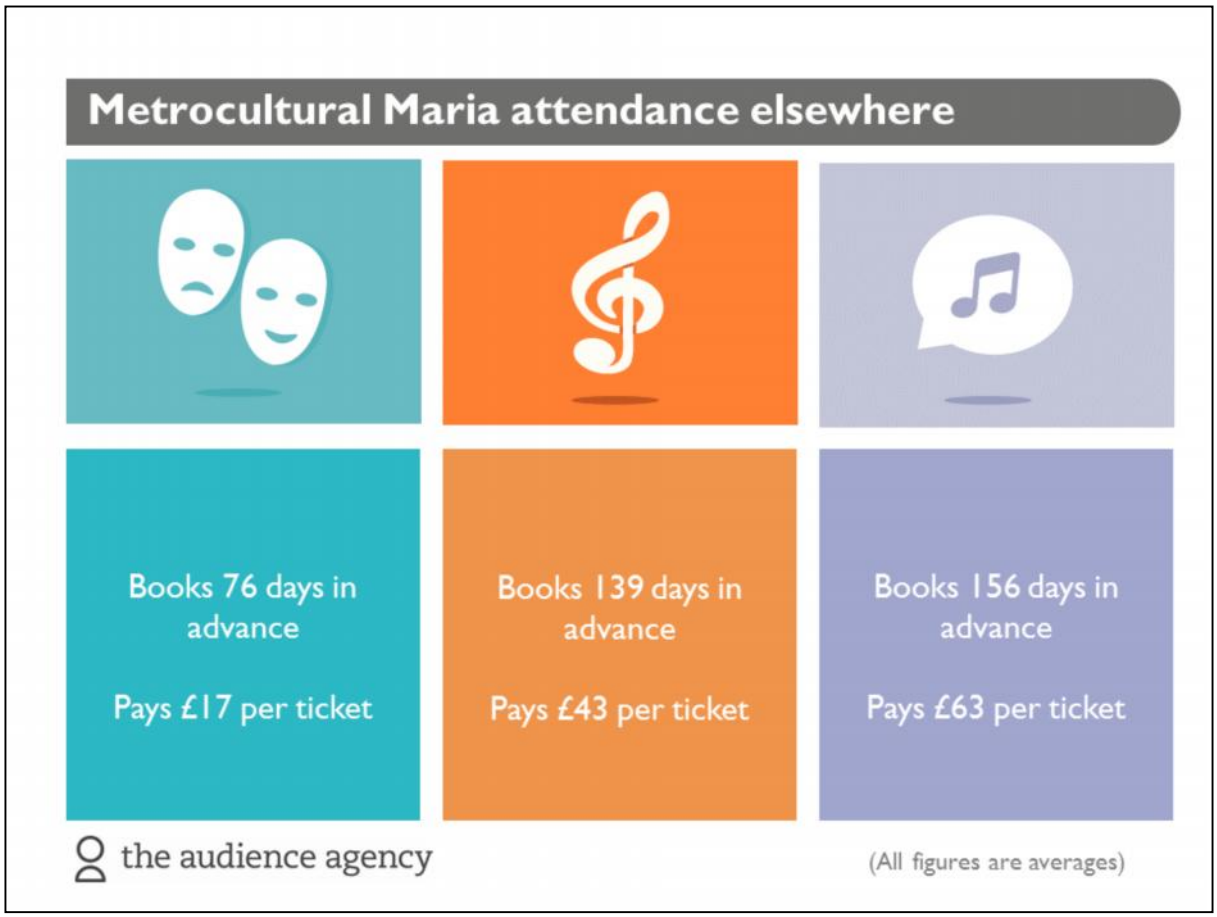
She treated them to champagne in the interval

Lives in Islington, North London

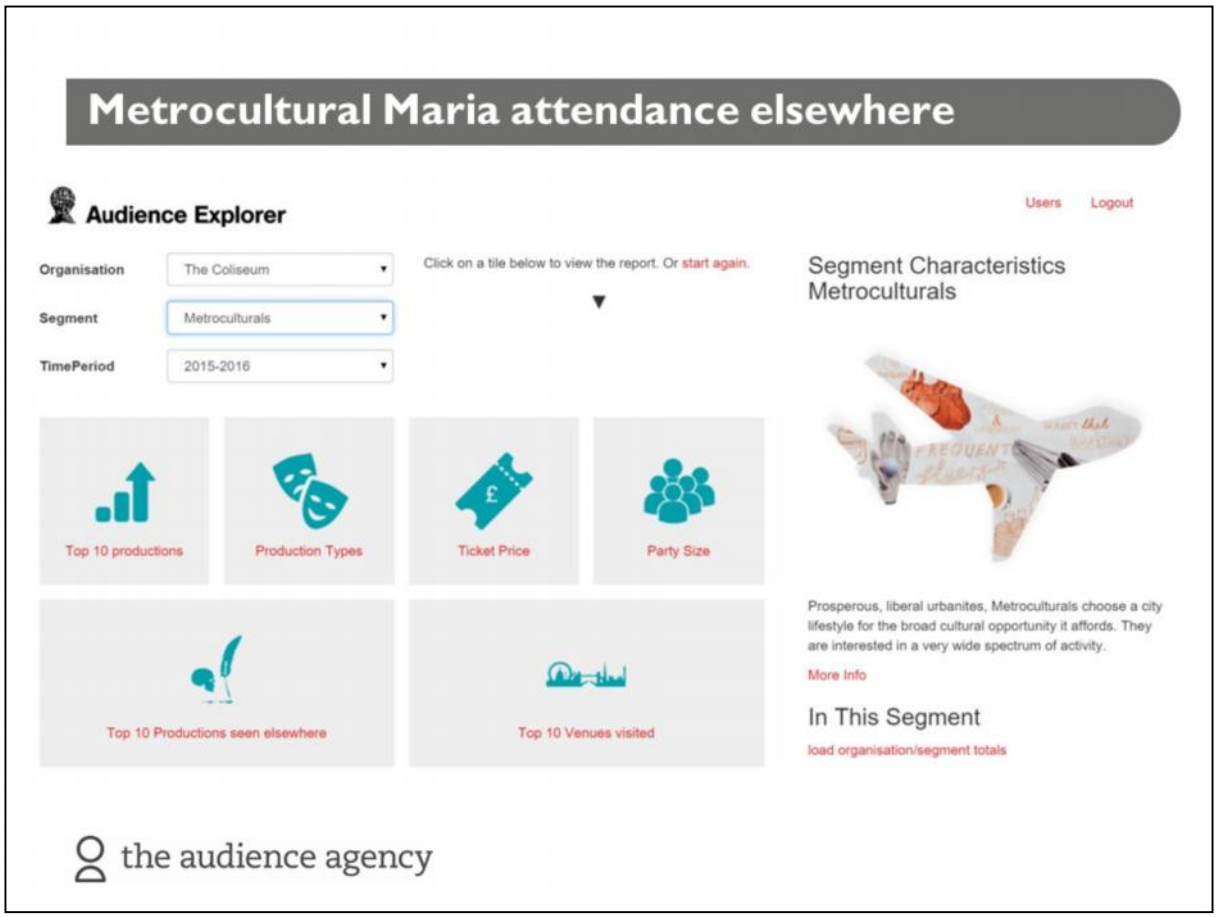
You could say, we now have a 90 degree view of Maria – but how can we go further?



Taking another view we can see a typical pattern for people similar to Maria – this means that she is attending a performing arts event about once every five weeks. And because this data is from ticketing transactions, it is likely that we're not capturing her visits to museums and galleries for example.



Looking in more detail at this data, we can understand Maria's patterns of behaviour.
On average, Maria books her tickets really far in advance to the Opera – about 156 days, that's nearly 6 months before the performance.
However, Maria only books theatre tickets half as far in advance – 76 days – that's a little over 2 months and spends about a quarter on each ticket.
Of course, we need to think about the other factors here – are theatre tickets just released later or generally cheaper? But at least we're asking the right questions and we've started to build a rich picture of Metrocultural Maria even though she's only attended ENO once.



What if we could follow Maria and people like her to see what else they get up to?

Remembering that the arts Data Impact project wasn't just about data science techniques, but also about prototyping tools we used the Audience Finder aggregate data set to provide the elsewhere analysis of all the Metrocultural Marias, Commuterland Clives and the rest all in one place.

Data scientist is about helping the data to speak, and this is a very simple storytelling tool. It makes that sometimes difficult link between a segmentation that describes propensity and the actual behaviour.

Only through aggregating the data do we have the ability to look beyond our own view of Maria and our audiences to see Maria in a more rounded way and very practically to find more like her, ourselves or theough potential partners and collaborators.

What else do we now know about Maria?

Maria - what the combined data tells us

- she's a regular opera-goer in London
- She also enjoys plays and classical concerts
- She's a frequent attender
- Her booking lead time varies between about 3 and 6 months
- She likes good seats and will pay for them
- She's quite like a substantial number of other bookers in the ENO database - which means there are **more like her to engage**

It's her first booking at the London Coliseum, but she's a regular opera-goer in London, so she's likely to be well informed about the artform and interested in the specifics of the programme.

She also enjoys plays and classical concerts, so joint marketing with venues presenting those artforms might help keep ENO in mind between visits.

Her booking lead time varies between about 3 and 6 months, so getting the offer in her mind early is most likely to have best results in terms of frequency of attendance

Her ticket yield is at the top end of normal or in other words, she isn't averse to paying for good seats, and seems to be responsive to add-ons and upgrades, so emphasising quality and hospitality is probably a good approach to take with the marketing.

Maria is starting to sound like a really rounded person now, and part of quite a substantial group of similar audiences in London and in the ENO database – which means there are definitely more audiences like Metrocultural Maria to re-engage.

So how have ENO worked with this and other tools, and in particular a data driven approach to reaching their audiences?

Slide 10

CE1 How do we know it's her husband?
Cimeon Ellerton, 24/11/2015

CE2 Cimeon Ellerton, 24/11/2015

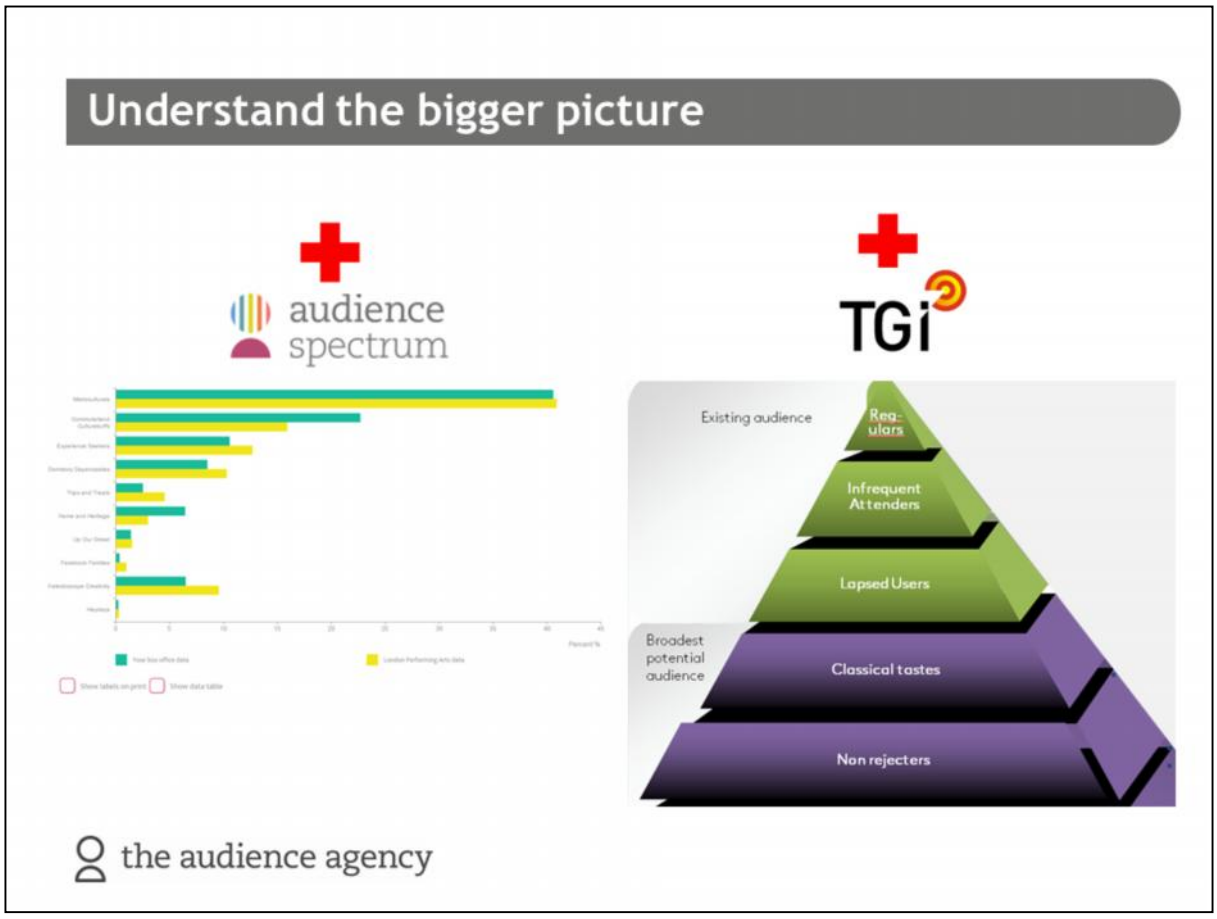
Claire Round from the ENO.

Thanks Cimeon. As we saw earlier, one of the challenges promoting what we call the 'blockbuster' operas at ENO, is that past sales history shows that they tend to attract a high proportion of first time bookers. That's great for us in terms of bringing in box office income and converting new audiences to opera. But it's also challenging when you consider the combination of relatively small marketing budgets, ticket purchases being made ever later and the limited opportunities for data capture we identified earlier. I don't need an algorithm for that to prove to you all that the possibility of sleepless nights is pretty high.

While Cimeon and Eva can tell you a lot about the theory behind data science, I'm here to talk you through a case study of how we've applied it at ENO. This campaign took place during the period that the company was involved in the ADI project, so data-informed thinking and tools were very well embedded in the marketing department and our agencies.

Earlier, we met ENO customer Maria, and were able to learn a fair amount about her booking and cultural preferences. That's all really interesting and useful - but how to turn finding out about people like her into a campaign to reach them and hit our business targets?

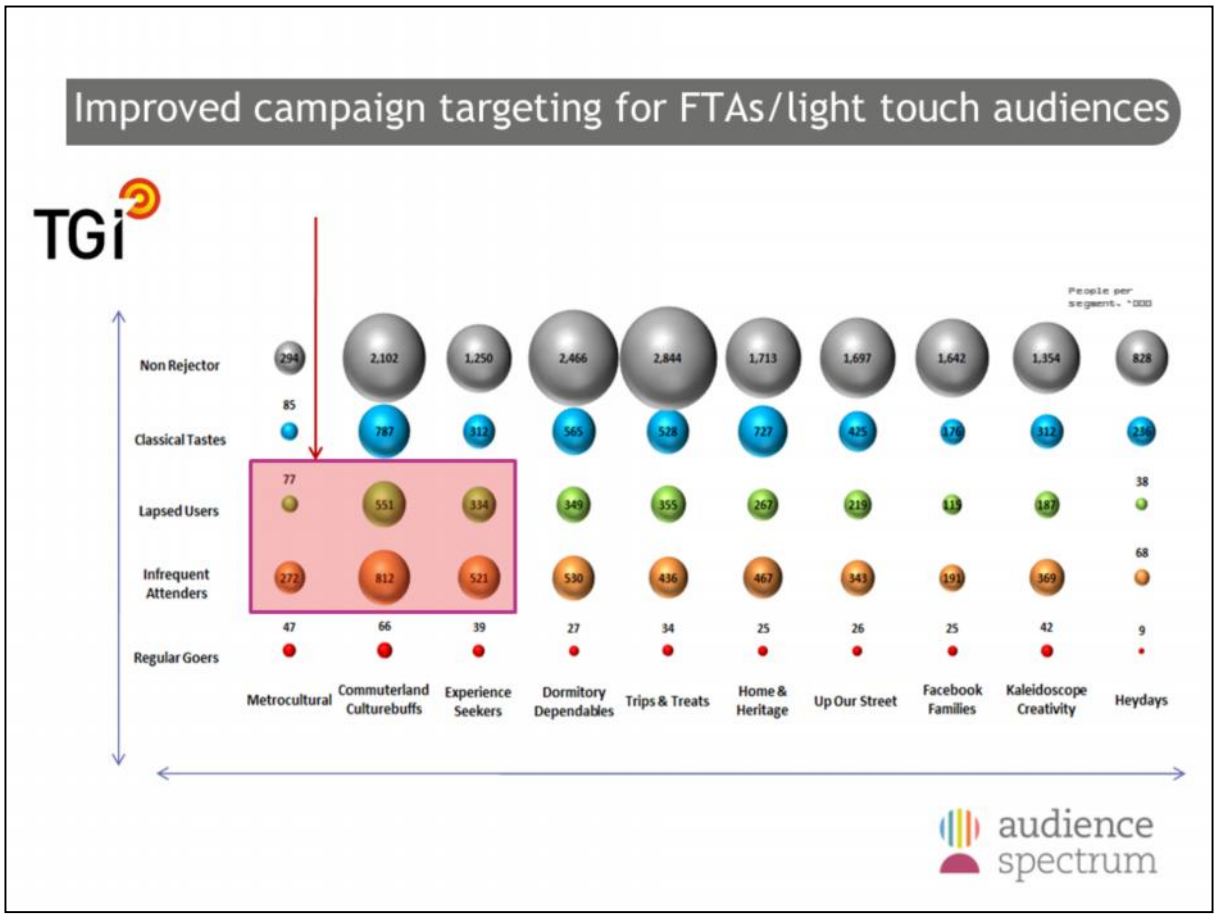
Essentially, that's the question the ENO marketing team was faced with in the spring. Like *La boheme, Carmen* tends to attract a lot of newcomers to opera. But early sales were below the level we'd projected, and we wanted to get creative with our use of data to try and reverse that trend.



By this stage, we'd carried out our data audit, and were into the 'aggregation' phase of our project. Working with the data scientists - as well as pricing and data specialists, Baker Richards and our media agency, Total Media, we wanted to find a way of making our data more actionable so we could use our knowledge to create campaigns.

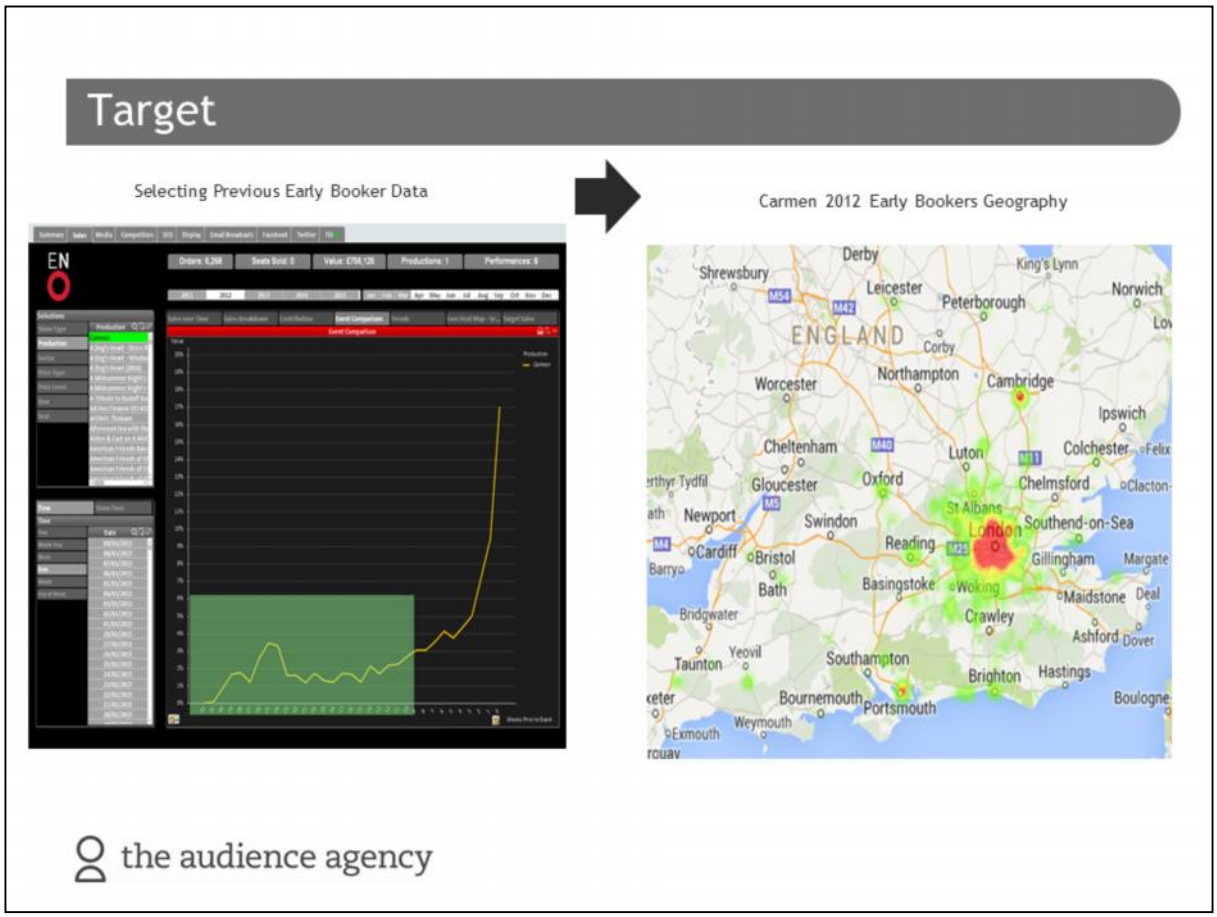
As Eva showed earlier, the key is to find a linking point between the different sets of data. As part of our ‘audit’ phase, we had tagged up our purchasing database with Audience Spectrum profiles. Total Media had also used the TGI omnibus survey to build a broad view of cultural audiences from the actively engaged regulars right out to people who aren’t currently attending but ‘never say never’. Behind those profiles, we could match up the right marketing and communications plans and customise them to our individual campaign objectives.

The aggregate task was to bring those two data sets together so we could target both by demographic and by attitude for maximum effectiveness. Luckily, Mosaic – one of the bases for Audience Spectrum – can be matched up to TGI.



And as a result we were able to create a visualisation of the Fused Audiences which shows the relative size of each group. Behind each bubble, we can drill in to see what the media and lifestyle preferences of each group is, and develop plans accordingly (although we do have to be careful with some of the smaller 'bubbles', as the data sets are quite small and they are therefore not quite reliable and robust).

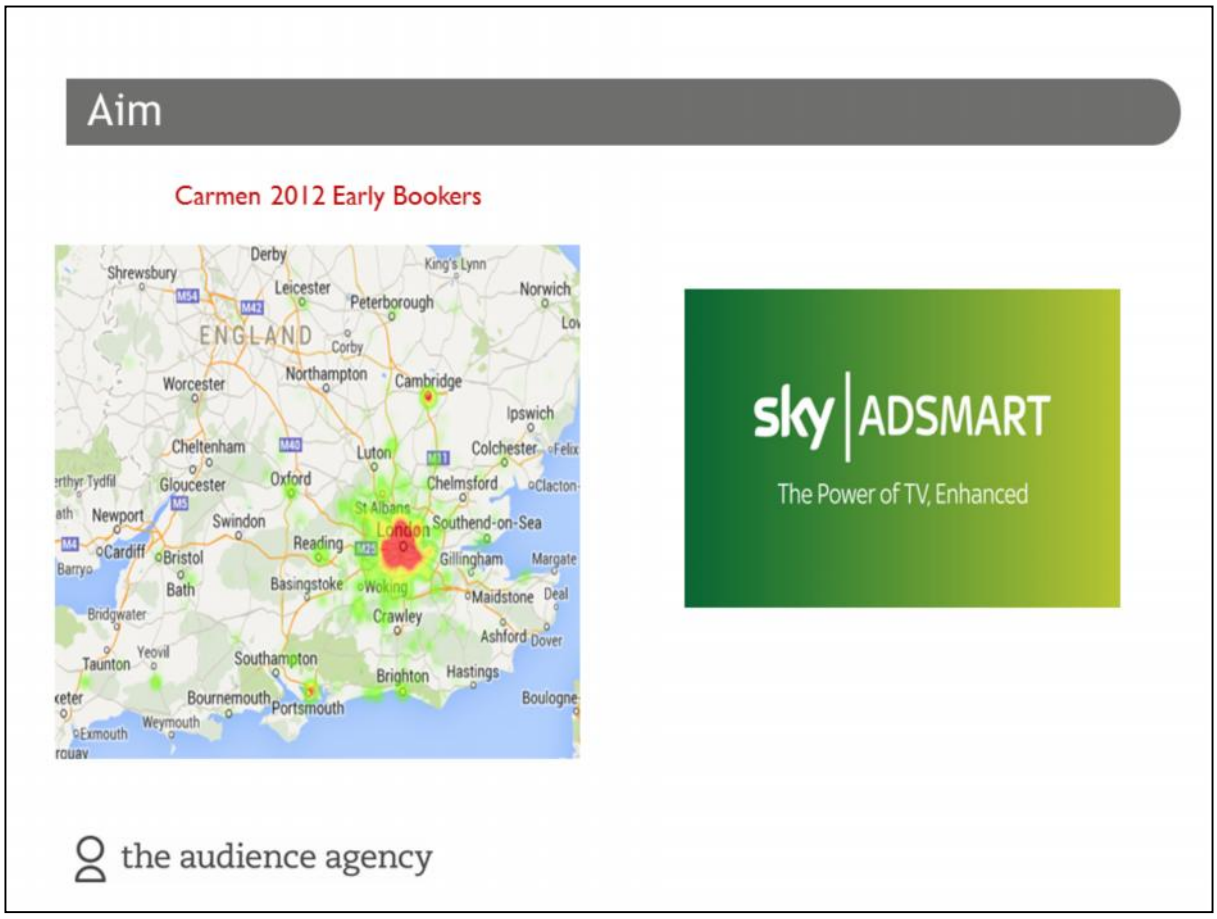
This visualisation is used by the marketing department in our planning and agency briefings – and increasingly when we're talking about audience development internally.



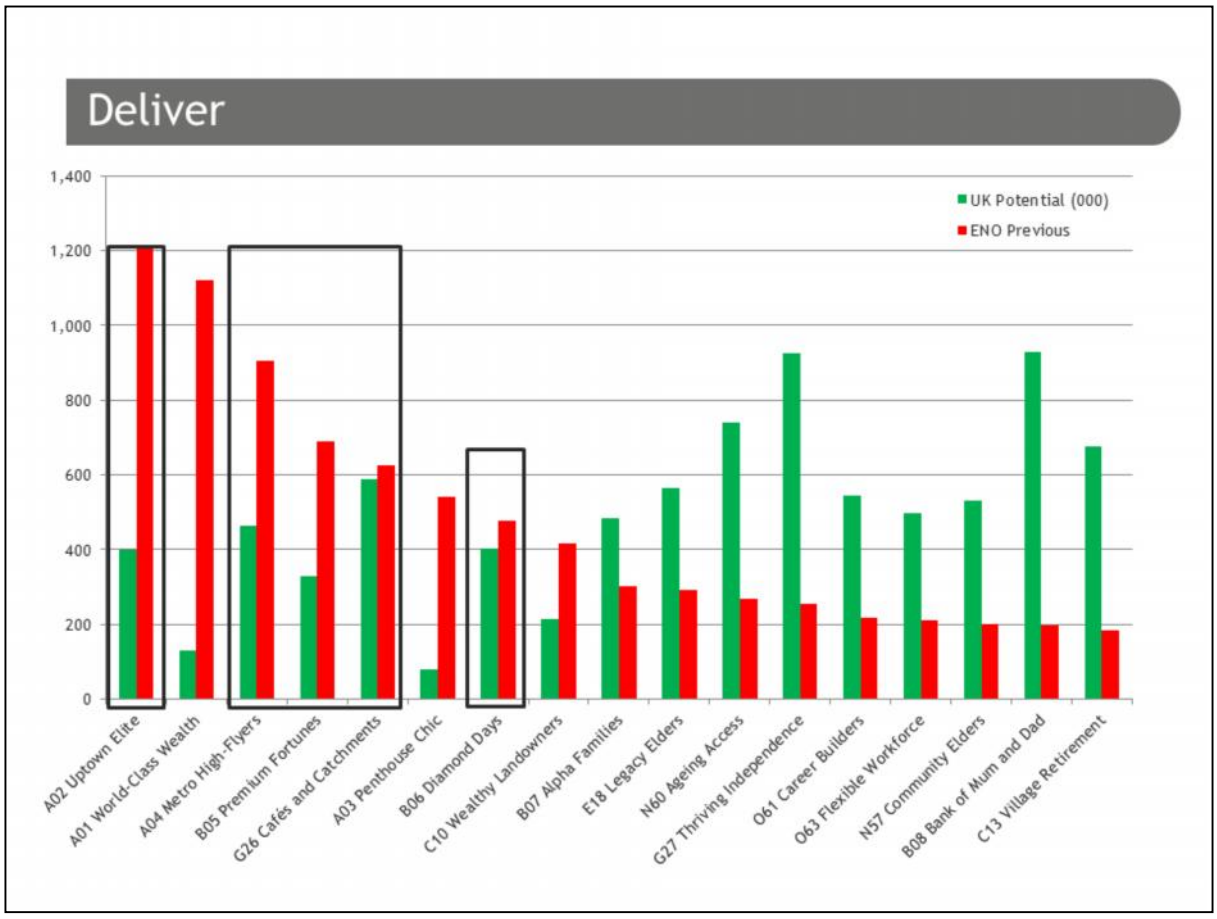
In the case of *Carmen,* we started off by looking at the data of people who had already booked for the production, as well as other similar works. We were looking for shared characteristics that could then be applied in another environment, so we could then go out and find more people like that. So Maria, our first time *La bohème* booker, who'd booked six months ahead, would certainly have fallen into the set of data we pulled.

Using a tool we license from Total Media called Qlikview, we could see how those people fit into our audience profiles and match that up to a media plan. And we used the initial booker data to target that on a postcode level.

We also ran a small survey to find out what creative approach would most appeal to our target audience groups, and learned that video was likely to be very effective – people responded well to the production style, and *Carmen* is full of hummable songs that are already familiar.



Surprisingly, we learned that because we could be very focused on the audiences we wanted to target, TV advertising was a viable option for the campaign using Sky's new Adsmart product which is bought more like digital advertising and can be targeted on a household level. As a guide, we had a budget of £10,000 for the media – and because the production was a revival, we already had content we could repurpose cost efficiently.



Using the Mosaic data again, we could see how our first party data stacked up against the population data, to inform the TV buy.

Outcome

- 25% overall sales uplift on previous production; 32% in areas where TV campaign was active
- 50% new bookers (11% above season average)
- Box office target achieved, reversing earlier sales trend

Postcode	Tickets 2012	Tickets 2015	Diff
BA	27	57	111.11%
LE	35	62	77.14%
OX	91	136	49.45%
HP	171	253	47.95%
GU	220	315	43.18%
AL	144	194	34.72%
CB	143	174	21.68%
PE	95	109	14.74%
CO	62	69	11.29%
PO	66	73	10.61%
RG	190	199	4.74%
Total	**1,244**	**1641**	**31.91%**

the audience agency

And the postcode data meant the ads were shown in the correct areas. Applying the 'analyse and act' principles, the campaign was modified once live to deliver the best results.

But did it work?

Well, sales overall for the production went up by 25% – this wasn't the only piece of activity we ran. However sales went up by 32% in the areas where the campaign was live.

And we managed to hit our target and overcome the initial lag, while also showcasing ENO's work in a creative and accessible way – bringing in the 50% new audiences we'd set out to achieve.

We can clearly see the opportunity of using data to both understand, target and convert audiences.

We now need to think about what to do next. Naturally we'll keep experimenting with different applications of data in our media campaign.

But I'm also interested in how we can bring data and understanding into the organization at an earlier stage – and improve our audience retention rates as well as acquisition.

From projecting who the audiences of the future might be, working with our artistic and executive teams to build the programmes and wider relationships for those audiences, as well as analysing customer journeys to make better predictive recommendations in CRM and to join up our systems and knowledge across the organization and improve our audience retention rates as well as acquisition. The more we interact, the more we can learn.

From description to prediction

the audience agency

As we could definitely see from what Claire and Cimeon said, when we gather all our data sources together, we know quite a lot about our audiences. Let's take the example of Maria once more. We are able to describe her in a variety of ways - we can focus on her demographic profile and lifestyle, find people similar to her through various audience segmentation systems and look at the ways in which they participate in culture. We can analyse Maria's repertoire journeys and her journeys between organisations as well as see how loyal a customer she is. We can tell Claire which other venues Maria visits and which artforms she likes. These are the standard analysis that we do and I'm sure may of you do on a daily basis.

What we are working towards

- Predicting audience types for performances
- Predicting success and failure
- Predicting effects of different pricing strategies
- Predicting audience flows, audience retention and acquisition
- Knowing the predictive power of audience segmentation systems
- Knowing the precision of prediction
- Using the predictions to inform planning and campaigns

the audience agency

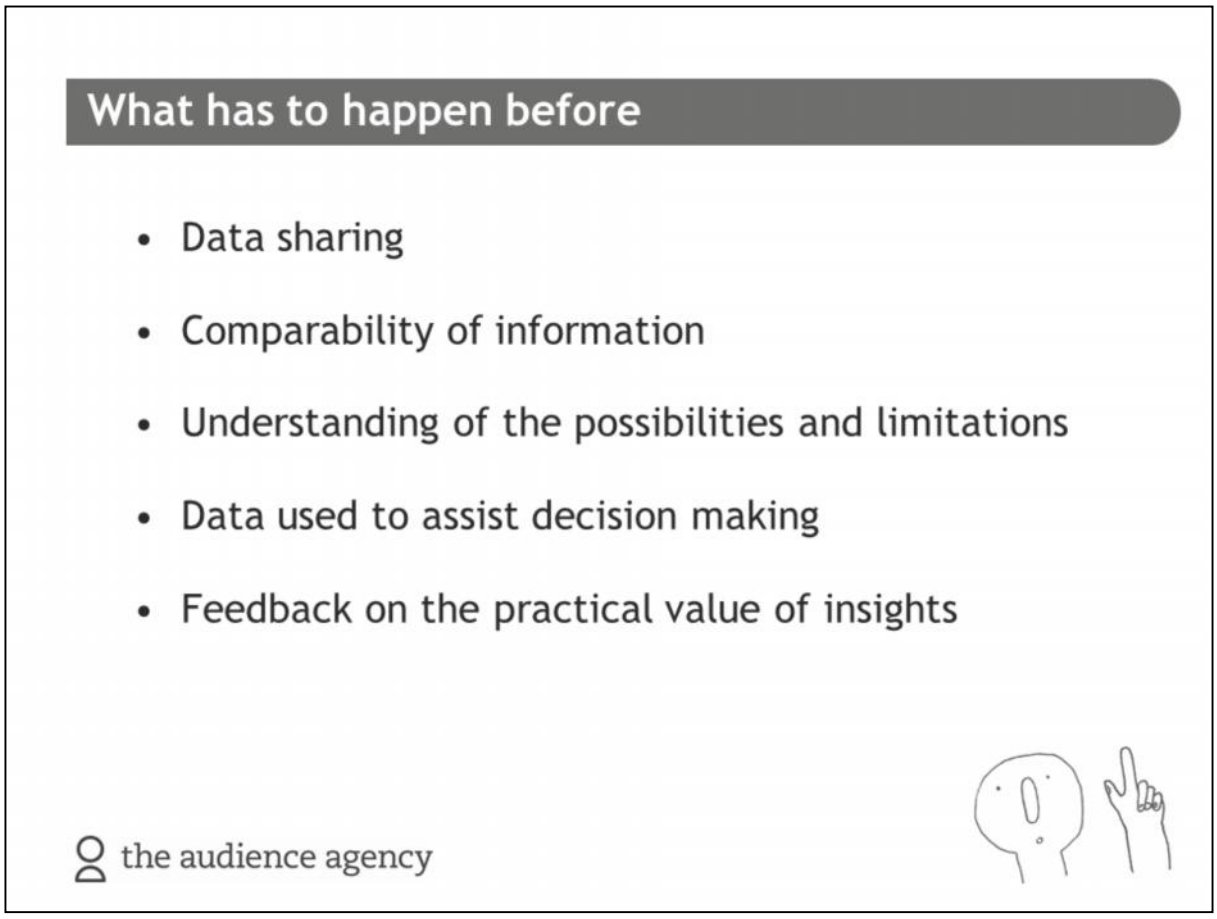
But what is the reason why we are doing all these analysis? Surely, we do this to inform our FUTURE actions. This means that what we would really like to do is predicting the future. It's great to know what Maria attended in the past, but it would be much more useful to know what she is likely to attend in the future - both immediate future and in the next seasons and to know what our current audience will be up to in a few years.

Similarly, it's always good to know whether our past performances were successful, however we define success, but it would be much better to be able to predict how well our future events will do in achieving whatever we would like them to achieve - e.g. attracting a specific socio-demographic group of visitors, bringing in new audiences or simply bringing more income.

Whenever we change anything we do, we would like to know what effect it will have - in many dimensions. For example, when we change price structure - what will the effect be on the type of audience we attract? Can we go beyond common knowledge here?

We would also like to predict how good we will be in the following season in keeping our audiences with us or in attracting new ones.

Finally, we would like to know which bits of information that we have are useful for predicting audience behaviour how precise we can be in our predictions, so that we can use them as an aid in planning and campaigns.



I do think that we are slowly approaching the world where we will be able to predict various aspects of audience behaviour - and we will have an idea about how good this prediction is. Before this happens, though, a few things have to be in place.

We certainly need datasets large and comprehensive enough to have the predictive power. This can probably only be achieved by sharing data. Sharing data doesn't only allow every single organisation to use the information gathered by others – to an agreed extend, but it also makes it possible to share the effort necessary to analyse the data.

In order to share data we need some sort of comparability, but this doesn't necessary mean that we all need to keep data in the same structure. It's enough if we agree on what we mean by certain things and what information we collect.

Another thing that seems important to me is developing a common understanding of what's possible and what's not possible, or not yet possible, or not possible with the resources we have. This works both ways – we wouldn't like to waste time trying to solve problems google isn't able to solve , but we also wouldn't like to miss on doing things that are certainly possible and may prove surprisingly easy to do.

Once this all happens, we can start using the predictions we make to help with decision making and from this we will be able to get a better idea of which pieces of information have a practical value, so that we can refine the list of things that would be interesting to predict to prioritise those that would be useful to predict.

In the ADI project, we've made our first step towards predicting audience behaviour. We used the aggregated AF data source to create a tool which will let organisations see what type of audience is likely to attend performances that they plan.

There were two main ideas to this project. First, we wanted to create something that allows to fail in the "imaginary world of the model". Second, we tried to create something that goes beyond what every single organisation can find out about their visitors and let different organisations in the sector "borrow strength" from each other to inform decision making.

So suppose that we wanted to know how likely we are to get Maria at our next performance. The tool looks at past performances that were similar to the one that we plan and finds out what their audiences were.

Audience Predictor - next steps

- The richer the aggregated dataset, the more accurate the predictions
- Testing and evaluation of what has the predictive power
- Tweaking the prediction model
- Extending the model

the audience agency

I said that the main idea of the tool was to let users see a likely audience of their planned performances, but that's not the only aim. The nice feature of an R&D project is that it allows you to ask questions and test ideas.

The thing that we wanted to test was how well we are able to predict audiences right now. It proved that with the current data, we can do well for some artforms, but there are still some niche cases when we really could do with more data.

So the first thing to consider in the future is how to make the aggregated dataset bigger and richer. As the dataset grows, we would like to test and tweak the prediction model and to find out which characteristics of performances really have the prediction power.

There are also many ways in which I would like to extend this prediction. For example, we could ask the opposite question: instead of trying to predict how likely Maria is to attend our performance, we may want to know what we need to do in order to attract her.

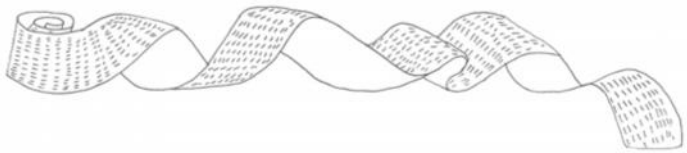
That's all very well, but how does this help me?

the audience agency

Our commitment to you

Work with the sector to:

- Share learning and access to any tools as widely and freely as possible
- Open up the aggregated data for all to benefit and experiment
- Widen the datasets to include:
  - Exhibitions and survey data
  - Social media, websites and digital content



the audience agency

Final thought

'We need to debunk the myth that data is uncreative, impersonal, or a threat to artistic decision making. Data is only useful where it enhances existing ideas or current understanding.'

Professor Paul Moore

Head of the School of Creative Arts & Technologies, Ulster University

the audience agency

Thank you!

cimeon.ellerton@theaudienceagency.org
eva.kabzinska@theaudienceagency.org
cround@eno.org

the audience agency