**People Analytics Special Interest Group**

**Minutes**

Spring Meeting 30 April 2019

Ministry of Justice 102 Petty France

**WELCOME**

Jill Millichamp introduced the third event of the People Analytics Special Interest Group. This event focuses on areas where government are applying People Analytics and journey taken in developing these approaches.

**PRISON OFFICER RECRUITMENT: GRAHAM KILVINGTON, MINISTRY OF JUSTICE (MOJ)**

The first talk by Graham Kilvington was on how analysis was developed in the MoJ to assist in the delivery of a high profile target for Prison Officer recruitment. There are just over 100 public sector prisons in England and Wales, with prison officers the key operational grade in prisons. In 2010 there were approximately 25,000 full time equivalent (FTE) officers but this was reduced following SDSR 2010 to around 18,000, driven by a reduction in the number of prisons, cost savings measures and a new benchmarked staffing level. However, recently levels of violent incidents in prisons have been rising.

In late 2016 the MoJ secured £100 million to hire an additional 2,500 officers by the end of 2018. With also replacing forecast future leavers and filling existing vacancies, around 7,500 prison officers would need to be recruited over two years. A multidisciplinary project team was brought together to deliver the target consisting of around 30 members of staff, including 4 analysts embedded in the programme team.

In **setting up the analysis** it was important to be clear about target and what was and wasn’t included. In defining the target it was made clear that it represented a net increase of 2,500.

To begin with there was very active ministerial involvement with requirements for weekly reporting which needed to be met. To inform early reporting the stages of the recruiting process were considered, and assessment of how many applications were being received compared to what was required to deliver the number of prison officers required at each prison was used to quickly provide reassurance and identify areas of concern. Leaving rates were also considered. Over time a suite of basic data required to cover all parts of the process was identified, with data coverage gradually built up until all parts of the recruitment process were covered.

**Development of analysis**: The team forecasted recruitment need, monitored application levels and which advertising channels worked best. Recruitment pipeline data, combining with data on vacancies to see which prisons were struggling, resulted in speeding up pre-employment checks for priority prisons. A red/amber/green ‘Cinema seat’ view of staffing levels across each of the prisons was developed. Analysis also included a deep dive in to why prison officers were leaving. Cluster analysis of applications withdrawn free text data was carried out to identify key messages. A pipeline model to forecast different scenarios was developed. A Power BI dashboard was also developed.

A prison officer recruitment Annex was published as part of Official Statistics, to help answer PQs and FoIs. Analysts briefed key stakeholders on what would be publically published.

**What worked well**: being embedded in the programme team worked well. Maintaining independence of the data team within the wider programme, to enable the independence of the data service. Providing key data to stakeholders, to provide reassurance that things were moving in the right direction. Once regular data was established, key decision points were identified, with work carried out to help identify priorities for intervention e.g. which prisons need additional advertising spend.

**Final stage of learning**: Programme ending, team members reassigned, some products passed to other teams, some stopped.

Key Learnings:

* Clearly define target at start of the process
* Be part of embedded project team, including having a seat at team leads meetings and presenting latest data there
* Identify key decision points and highlight the prisons that the data shows need support e.g. with extra advertising spend, or with retention of prison officers

The target was achieved ahead of schedule in early 2018. The data driven approach to targeting advertising spend won an AMEC award. Impact on prison violence is not yet clear, there has been a recent small decline but it is too early to establish whether this is an emerging trend.

**THE ROAD TO PEOPLE ANALYTICS: KATIE GRONOW, DEPARTMENT FOR TRANSPORT (DFT)**

In the second presentation Katie Gronow talked to group about the work done on people analytics at DfT over the last few years, challenges that have been faced and where they are going in the future. Work was carried out by the Analytics Unit which is a central analysis team for the whole Department.

Looking back over the last 10 years there has been an evolution in analysis from univariate statistics in Excel spreadsheets to multivariate statistics, R, RAP, interactive dashboards and narrative storyboards. The talk considered:

* What changes have been made and why
* Challenges and benefits

Focusing on the example of the Public Sector Equality Duty. This is a snap shot of data taken every March to see if DfT staff is representative of the local population what has changed and is there equality of outcomes. This feeds into the Departments diversity and inclusion strategy. The Equality Monitoring storyboard for 2017/18 was published in April 2019 on www.gov.uk.

Equality Monitoring storyboard used to be a lengthy word report. The report has moved away from traditional reports, becoming more accessible. Now key narrative from results added to a dashboard to create a “storyboard” which is a cross between a report and a dashboard. Structured by protected characteristics (age, race gender etc.…) but covering all the HR processes. It is a legal requirement to publish the basic statistics but have increased the context and interpretation. There is a wide audience, HR, trade unions, HR Functions, Staff and staff networks and public. This is a journey and will continue to evolve over time.

**Changes to Analysis:** 15 months ago started to move to R, this year all multivariate analysis in R. Now also carry out logistic regression, to see if there are differences across group’s. Still use VBA and excel to generate tables. Making changes to data collection: still using four different sources together, requires a lot of data cleansing. Reconsidered data cleansing e.g. people who appeared in one data set and not another, small numbers and disproportionate time. This year data on socio-economic background will be available. Data still in MS Access databases, need consider how data will be sorted in the future e.g. Sql, google cloud platform.

**Benefits:** Reports are being used, which means more evidence based decision making. Improved understanding of HR needs, analytical tools are efficient and easy to quality assure, streamlined data preparation process.

**Challenges:** Identifying an alternative to using PowerPoint and pdfs, which are not accessible to assistive technologies, time consuming, and static. No ideal way to structure the reporting characteristics / processes. Finding the right narrative to describe analysis has been hard.

DfT are moving forward with analysis including now looking at workforce planning. Katie asked audience about methods used to publish interactive visualisation.

**DIGITAL FOOTPRINTS INSIGHTS TOOL FOR WORKFORCE ANALYSIS: DR HANNAH STATE-DAVEY, QINETIQ**

Dr Hannah State-Davey spoke to the group about research carried out in the Ministry of Defence (MOD) that could represent an analytical disruptive capability applied to the future workforce. The work form parts of Dstl's Achieving the Strategic Edge Through People (SETP) 2040 disruptive capability project and Future Workforce and Human Performance programme. The project developed a proof of concept behavioural profiling tool that can provide information on an individual's behaviours and characteristics from their ‘digital footprints’.

Measuring behaviours form digital data is a novel and emerging research area. Behaviours that were being looked at were personality, resilience, EI, organisational commitment, job satisfaction and satisfaction with life. There are a number of ways the data can be used in workforce management for example selection, career management, training, employee engagement and wellbeing. This application was looking at employee engagement, and whether the approach could be utilised in the nearer term.

Employee engagement and organisational climate: There is considerable disagreement on employee engagement and organisational climate in the literature, although there is more agreement on the drivers of engagement. This approach could compliment current measures, combating some of the disadvantages of the current measures e.g. survey fatigue, time consuming to fill in and time consuming to analyse. This approach has the potential to collect data passively from individuals, allowing MoD to look at things at a deeper level on a more frequent basis and also carry out more segmentation.

To conduct the behavioural profiling, social media and self-report survey data was extracted from 774 participants who consented to the extraction of their Facebook and Twitter posts. This provided the data to develop the measures. Also, due to the proportion of males and female respondents, the team controlled for over representation of females in analysis.

Stage 1 feature extraction - both a closed and open-vocabulary approach was used to extract language features from the social media data. Correlational analysis was conducted between the language features and self-report survey scores to develop the machine learning models. Validation analysis was conducted as part of the process. A language readability index. However, it was noted that the index was developed for more formal writing. This provided an indication of reading levels for the text and its applicability to a Defence context..

In development of the machine learning models there is no single machine learning algorithm. As such, the research compared multiple algorithms including; penalised regression, Support Vector Machines (SVM) and Random Forests. Random Forests and SVM performed best on unseen data, but they are not transparent. Therefore, the research looked to identify algorithms that were more transparent whilst maintaining accuracy. BoostSRL offers complex, transparent algorithm that shows the potential to outperform other algorithms.

Validation, preliminary validity tests were conducted to ensure automated models measure what they intended to measure. Validity is a matter of degree and it can take a long time to assess validity. Validation was difficult to establish due to a lack of agreement on definitions of the behaviours and different ‘ground truth’ measures. The research identified striking language variation across behaviours, which were found to make conceptual sense when linked backed to psychological theory. For example, those with high job satisfaction had a strong relationship with co-workers.

The Digital Footprints Insights Tool, uses Power BI visualisation. There is a management interface, with data streaming in with models running on the data. There is also potential to input other data sources. There are a number of dashboards where you can segment, carry out population comparisons, geographical analysis etc... Preliminary user evaluations have been carried out.

The future

* Investigate alternative sources of content data, to build behavioural models for job satisfaction and commitment in particular.
* Further investigation the use of BoostSRL
* Pilot test the approach on real world mod data.
* Validate models.

**PEOPLE SPRINT – JAMIE LENDRUM (DEFENCE SCIENCE AND TECHNOLOGY LABORATORY (DSTL) AND CATRIONA SMITH (MOD)**

The final presentation by Jamie Lendrum and Catriona Smith outlined the ‘People Sprint’ an application of the Scrum approach in the MOD. Defence is developing a People Strategy narrative which will be underpinned by making better use of the people data. This will require building Defence's capacity to bring data together to provide analysis and insight to give a Defence wide view of the workforce.

In 2015 the delivery of science and technology (S&T) in Defence was reviewed. As part of the resulting research portfolio there was a requirement S&T programme to produce a step change in people data analytics. Dstl and broader Defence has a strong human science capability and growing data science capability. Until now the data science capability was focused on intelligence community. There has been a considerable effort to raise the level of skills. People analytics sits between data science and human science.

The **People Sprint** was a capacity building exercise, identifying what approaches can be used to visualise and understand the data - in this case 10 years of Royal Air Force personnel data. **Scrum** is a form of agile development, a framework for getting the most out of the limited time available. The Scrum allowed a multi-disciplinary team to quickly come together to produce customer-focused outputs, to fail fast, and cut out bureaucratic layers in order to deliver within a short timeframe (1 week).

The intention was around **understanding the data** with descriptive analytics as a precursor to more complex analysis, such as building predictive models. Lessons included; there is a danger it can be seen as a ‘silver bullet’, and therefore requires some expectation management. It also requires significant preparation before the event to articulate customer needs, gain access to the data, and set up the infrastructure.

The benefit of having people from a range of disciplines (and organisations) in the same room working on the same problems, this represented a step change in capability.Positive **feedback** was received from analysts who attended the sprint, stating that it provided the opportunity to focus on the data they deal with everyday. In their day job a lot of pressure placed on delivering outputs the sprint provided an opportunity to really explore and understand the data and providing an insight into how data scientists approach the analysis.

The people sprint was a week, to show what could be produced, as progress understand domains more might want to have more slow burn sprints with a smaller number of analysts to develop the skills that they need. The sprint environment can also be high pressure, therefore there is a need to adapt scrum approach to suite Defence's needs. Going forward we will experiment on how to apply these approaches.

**WRAP UP AND DISCUSSION**

The SIG’s AGM will be held later in the year, Jill encouraged anyone interested in joining the committee to contact her or any other the other members of the committee, volunteers are always need to help with organising events. There was also a suggestion for the AGM to consider People Analytics and GDPR.