# Notes for interview

**Cluster modelling work for the department of health;**

**Problem definition:** DH wanted a data driven approach to clustering hospital ED and inpatient wards based on their capability levels.

The stakeholder needed the clusters to equal 4 & 5 levels to match the nursing ratio required to handle certain patient volumes and complexity.

**Context:** previously this was determined by legislation and SME’s to determine the nursing ratio and the department was a data driven approach sued for new ED and inpatient wards not included in the legislation.

**Output:** After testing numerous cluster methods it was clear that so build in enough complexity into the models and create pairing which made sense I would need to use an advanced cluster technique called mixture models**.** These models allow you to select probability distn i.e. Poisson distrn for cnt vars and pair obs based on max likelihood functions.

This allowed me to present cluster pairing using visualization like a PCA chart (illustrate), and the feedback from SME played a crucial part in ID inappropriate pairings.

**End result:** the produced a formal report for the policy team and they presented my findings to health unions and the minister’s office are part of a negotiation on nursing ratios.

**Predictive analytics products:**

Drivers of ED & Ambulance waiting times

**Problem:** ID the key factors driving of performance at large metro hospitals

**Context:** post the covid restriction the demand for service in ED’s had increased dramatically and ward where struggling to meet target waiting time for cate 1 & 2 patients (urgent care patients)

**Output:** The project team needed a quantifiable statistic which could be included in an intuitive graphic to included in a slide pack for the ministers office.

To build feature which will actually solve a problem you need to shape the data is a way in which the models will actually give you useful lnfo, i.e. in this case I needed to capture patient level data (target variable is patient waiting times) whilst also adding campus level factors to the model i.e, number of available bed in the ED/inpatient ward, number of staff, cnt of different patient complexities using funding codes as a proxy etc.

To handle the combination of patient and campus level factors I decided to use a mixed model which allows you to split variables up into fixed on random so the fix variables being things are constant throughout each campus and the random variables is where you allow some variability at a campus level such as the number of available beds.

Once I had fitted as many meaningful variables I could I used an AOV to qualify the amount of variation explained by each factor (fixed variables), the stats very intuitive labels i.e no. of available beds both in the ED and lagged inpatient from the night before explained 20% of an ED performance on the time window.

**Live examples:**

**Med cert:** was tasked with reproducing a complex python workflow designed by the former data science May which used sophisticated business rules and stats id claimants who had a history of med cert being issued with invoice over a certain time window.

Working with insurance assessor Rebecca Summers who was validating the outputs of the workflow it become evitable that we would need to re-engineer the business rules and stats used to id claimants of interest whilst also remove false positive (docs who bulk bill).

**Avoidance** uses what can be desc as unsupervised learning, are getting a huge amount of info from Bec and other SME’s I id a rich list of avoidance features which could be used to cluster claims into homogenous group of interest i.e. combination of high avoidance stats, due to complexities which I have had in the past I shift to using PCA, stat modelling and applying weights to variables to each claims has a avoidance ranking which was more intuitive and useful for insurance agents i.e. 730 claims which have some sort of construction license sort them by avoidance rankings to help priorities which claims to reviews.

Features like mental health, comp, avoidance flags and cnts, so you have see where cluster model will struggle to due all the different variable type, binary flags, cnts, factors, etc.

Creating a scoring system using survey result from SME and Bec. Input I was able to produce an output which will be used as part of quarterly reporting of constructions license reports.