

Using AI to mine verbatim comments from the National Rail Passenger Survey

Wherever one looks, AI is the topic of the moment and supposedly the answer to many of the challenges facing businesses as we risk drowning in data. The appeal of this resonated with us as we have always found it a challenge to milk the wealth of verbatim comments passengers leave on our questionnaires such as for the National Rail Passenger Survey (NRPS).

To begin to understand what value AI might add to our evidence gathering, Transport Focus partnered with *Signoi*, AI specialists with its roots in survey research. We saw their combination of AI expertise and a deep understanding of the research process as ideal for seeing what value might be released from the NRPS verbatim comments.

NRPS generates some 30,000 completed questionnaires twice each year. The majority of questions are numerical rating scales, but there is a single open-ended question that asks: *“Do you have any further comments about your trip today or the rail service generally?”* This allows passengers to comment on something they don't feel the structured questions have addressed as well as commenting on issues of a more general nature and not related to the specific journey which we get them to focus on in the rest of the questionnaire. (Indeed, we were already conscious that many of the verbatim comments relate to ways in which that one journey was atypical and not what they usually experience.)

The Signoi algorithm looks not only at the free text answers to this one question; it also takes into account all the other NRPS questions. So, put simply, amongst other things it knows how satisfied (or otherwise) passengers were with their journey and the various station and journey factors covered by the questionnaire. It also utilises the demographic and journey data (for example: sex, age, journey purpose, ticket type). The combination of structured (numerical) and unstructured (free text) data adds to the value of the analysis in a way that a simple 'word search' of the verbatim comments cannot.

Signoi has produced the report detailing its findings from an analysis of the four waves of NRPS conducted in 2017/18. It is probably reassuring that, at the macro level, their summary replicates much of what we already know from the structured data and has not revealed any particular surprises. We regard this as proof, if any were needed, that NRPS asks the right questions and addresses the key factors impacting satisfaction with any rail journey. The analysis confirms that commuters are the most critical passengers. Comments on the basics: punctuality/reliability, capacity, dealing with delays, reinforce just how critical these are. But passengers also have a lot to say – both good and bad – about staff, information, the state of the rolling stock, and accessibility (whether they are disabled or not).

Signoi has also provided us with a number of dashboards that enable us to drill deeper down into the data set – so by train operating company and key stations. We are currently exploring the value of these dashboards with interested parties.