From the automotive aftermarket data, it is possible to derive patterns of buying habits, preferences, return rates and product attributes

How data can change the automotive aftermarket

Rain Data Ltd and Newcastle University share analyses from datasets collected on every aspect of buying car parts online.

Data science and big data analytics are currently being applied to virtually every aspect of modern life, including online security, finance, insurance, advertising, medicine, streaming services, town/traffic planning and many other domains. Whenever and wherever data is collected, statistical analysis and pattern discovery can be applied to provide unique and novel insight, usually with the aim of gaining a competitive advantage.

The automotive aftermarket sector is no exception, collecting large datasets on every aspect of buying car parts online via web-hosted catalogues. These datasets contain a plethora of information about the transactions of searching for and buying automotive parts – when, where, what – and details of the parts themselves. From this, it is possible to derive patterns of buying habits, preferences, return rates and product attributes from the data.

The results described below arise from a knowledge transfer partnership between Rain Data Ltd and Newcastle University's School of Mathematics, Statistics and Physics and the School of Computing.

Rain Data works in partnership with MAM Software, which has provided the data for the analyses described below.

The types of analyses we have applied to transactional automotive datasets include:

- Determining the return rates of auto parts
- Determining measurements of factory-fitted parts (standards of the original equipment manufacturer (OEM)) for the automotive aftermarket
- Measuring and predicting when auto parts need to be replaced
- Gaining insight from big data on electronic transactions about which combinations of cars, models, auto parts and suppliers account for the largest proportion of invoice entries
- Tracking profit margins
- Optimising pricing
- Detecting seasonal trends in sales and returns
- Segmentation analysis to rank auto parts, prioritising which auto parts are most important to a business

AUTHORS: **WAYNE SMITH, SHIRLEY COLEMAN JAUME BACARDIT, SYD COXON**

Three examples from the above list are described in further detail in the following case studies. The first two were conducted between October 2017 and July 2018 and the third study was conducted between May and July 2018.

Case study 1: determining the return rates of auto parts

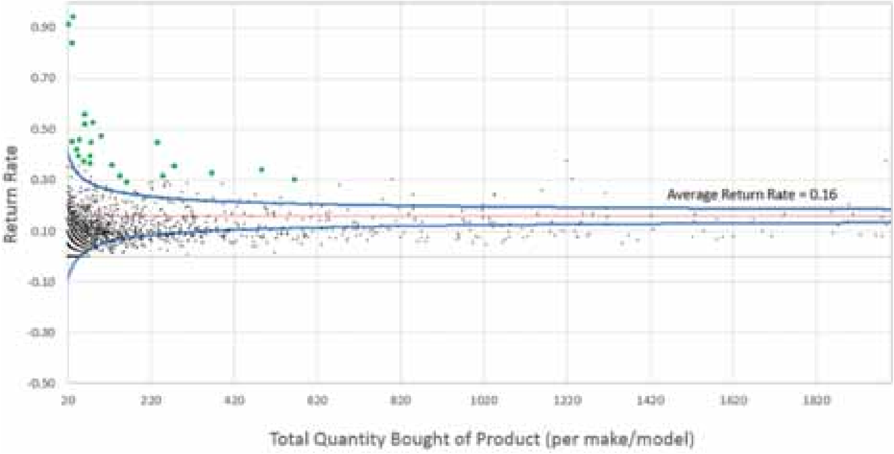
Understanding the nature of product returns is a key factor in running any successful business that supplies goods. They can inconvenience a company for several reasons, including the cost of reposting, identifying if a product is fit for resale and the cost of future sales. When dealing with very large datasets that include hundreds of thousands of products and millions of units sold, it can be difficult to prioritise product returns – that is, which auto part return rates should be attended to first. One way is to consider return rates using control-limit charts.

This type of analysis – see Figure 1 – considers the return rate for each category of auto part according to the number of units sold. In doing so, it is possible to plot statistical limits on these charts, which then reveal which auto parts' return rates lie beyond these limits. In Figure 1, the return rates of frequently bought air filters are plotted as a funnel chart with statistical limits to identify which return rates require immediate investigation. This allows a method to prioritise attention to parts according to whether and by how much any return rates are positioned above the action limits of the control chart.

Case study 2: determining measurements of factory-fitted parts

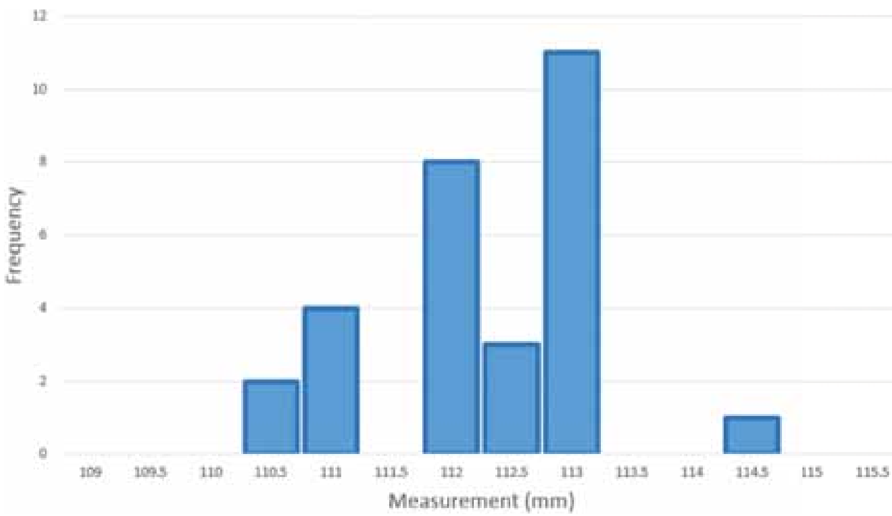
When referring to auto parts, the OEM standards refer to the time a car is made – that is, the parts assembled and installed during the construction of a new vehicle. Aftermarket parts are those made by companies trying to match the OEM factory standards. There is considerable uncertainty about which parts match to which OEM numbers – for example, many suppliers will often manufacture parts for the same specific vehicle, yet the dimensions and properties listed against those parts made by suppliers can vary greatly – see Figure 2.

Determining the correct measurements for an OEM confirms which parts should be affiliated to a car and identifies those parts which may be wrongly matched to a car – see Figure 3. Estimations of the factory standards can be derived via →



The plot above shows quantities of air filters bought (horizontal axis) along with the return rates for each part (vertical axis). Plotting the expected limits of the return rates (blue lines) highlights those parts that have a statistically high rate of return and therefore require further investigation.

Identifying the return rates of air filters
Figure 1



The above plot demonstrates a frequent problem in the automotive aftermarket: trying to determine the correct measurements of a part for a given car. Here, 29 height measurements of fuel filters (all listed against the same Ford car) are plotted as a frequency histogram. In this instance, 113mm appears to be the correct measurement, but there is also range of other values (from 110.5 to 114.5mm). The extent to which these different measurements are important will depend on how critical the height measurement is for this filter on this particular Ford car.

Frequency of height measurements for Ford fuel filters
Figure 2

Aftermarket parts are those made by companies trying to match the OEM factory standards inspection of all the supplier's parts associated with each OEM standard. Using fuzzy matching to compare supplier parts to these estimated values of the OEM factory standards reveals those supplier parts that can be matched against other cars to which they were not originally intended to be matched, and in doing so highlights potential revenue for a part supplier, as they can sell these parts against more cars with very little modification. This is only possible due to the analysis of vast quantities of supplier data.

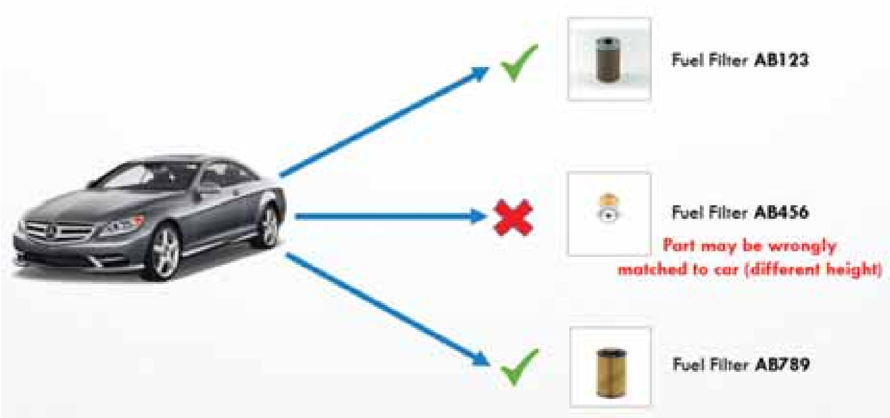
Case study 3: predicting auto part replacement

Comparing invoices for the replacement of auto parts and noting the mileage of a car at the time of replacement highlights the potential failure rate of non-serviceable car parts. Inspecting those parts that were replaced, along with the mileage at the point of replacement, and comparing these across different car models also demonstrates the reliability of different auto part manufacturers. Fitting an empirical cumulative density function (ECDF) to the replacement of an auto part and the mileage at replacement aids in the process of identifying key points in a car's history when replacement is needed and also potentially when failure may occur. This statistical technique determines the probability of an event using the observed data rather than a theoretical estimate of the probability.

Figure 4 shows the ECDF based on mileage data for brake disc replacement. It indicates the probability of a replacement/repair (vertical axis) when arriving at a garage with a given mileage (horizontal axis). The three plots demonstrate the probability of brake disc replacement for a range of small cars at 50,000 miles (dotted vertical black line). There is a 23% difference in the range of probabilities for these cars at 50,000 miles. Given that these cars are similar in size and may conduct similar journeys over their life, the difference in probability of replacement for these parts could be either from manufacturer/garage recommendation or part quality.

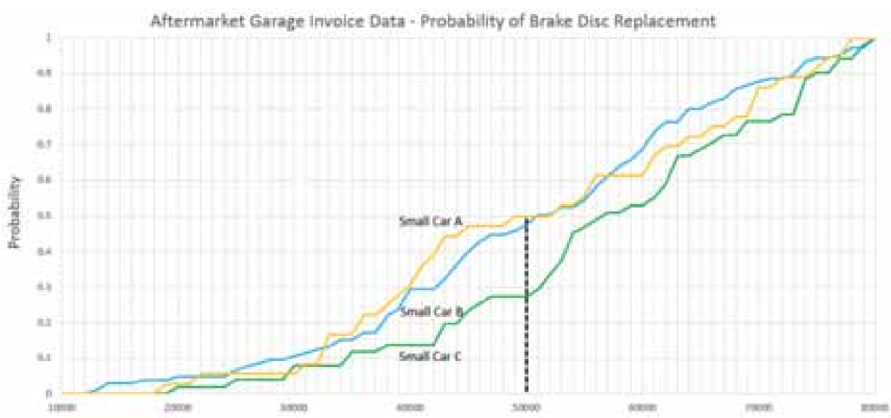
In conclusion

These examples show just some of the insight that can be obtained from the massive datasets generated within the automotive aftersales sector and show the potential for increased understanding of the market and increased efficiency and profitability for auto traders.



Comparing part measurements that are applicated against a particular car helps to derive the typical/most likely measurements for that part. This analysis highlights the parts that should not be listed against a car (as above) and helps to match parts to other cars that were not originally considered.

Identifying correct and incorrect car parts
Figure 3



The above plots show the probability (vertical axis) of brake disc replacement for a range of small cars at 50,000 miles (dotted vertical black line positioned on the horizontal axis). There is a 23% difference in the range of probabilities for these cars.

Aftermarket garage invoice date: probability of brake disc replacement
Figure 4

Original equipment manufacturer standards refer to the time a car is made

Wayne Smith
School of Computing Science, Rain Data/Newcastle University.
wayne@raindata.co.uk

Shirley Coleman
Technical Director, Newcastle University.
shirley.coleman@newcastle.ac.uk

Jaume Bacardit
Reader in Machine Learning, Newcastle University.
jaume.bacardit@ncl.ac.uk

Syd Coxon
Director, Rain Data.
syd@raindata.co.uk