

Forecasting the survival rate of tenants using historical occupier data and key performance drivers

Islam Ira Ashia

University College London, Local Data Company

Background and Motivation

Retail has been at the forefront of countless debates and studies over the past 24 months, with an increase headlight amidst the COVID-19 pandemic where many small and big well-known brands like Debenhams have incurred losses and closed down, creating more empty stores. The combination of the explosion in e-commerce together with the pandemic and its impact on offices and retail is prompting a wide-ranging and difficult rethink about what the purpose of town and city centres will be in the future. Nationally, the vacancy rate in retail is now 13.2%, according to LDC. But in some towns, especially in the north of England and parts of Scotland, it is almost the double. Vacancy rates in shopping centres are almost twice those of retail parks. In addition, property company JLL estimates that there will be 80,000 redundant shops littering UK high streets in the next few years (Financial Times, 2020). Landlords are experiencing challenges finding tenants to fill vacant space and are also struggling to push up rents for existing occupiers given the low levels of occupational demand in the wider market. In fact, according to Financial Times, a growing number of retailers are demanding leases linked to store sales rather than fixed amounts that can only be changed when the lease ends. It is, therefore, that in such climate that becomes essential to understand and gain insights on the long-term fortunes of occupiers. The client's uniquely detailed and historical data allows to analyse the average length of operation for a variety of occupiers to understand the risks and opportunities landlords face during periods of tenant churn. The aim of this project is to build a model that can predict the survival rate, that can be used as a forecasting tool retrospectively testing the model against 10 years of data to help predict future vacancy rates.

Data and Methods

The original dataset supplied by the client consisted of 31 variables and 1.3m rows with a total of 42 million observation. This dataset contains information of different retail and leisure units from the period of 2011-2021 across England. They follow the footfall

technology where they employ a team of field researchers to physical audit the occupancy and vacancy of every unit across the UK on a regular basis. This allows to have a very close real-time view of the market. The initial analysis consisted into broadly exploring England and understand the market. Afterwards, the main analysis consisted into exploring Greater London region in detail, which means analysis of different towns within the region such as Essex, London, Hertfordshire, Kent, and Surrey. This is because the nature of survival differs geographically, thus a model should be created for each region separately to avoid bias.

Kaplan Meier: In real scenarios, the actual survival function of a population is not available. Hence survival function is usually estimated from the observed data using Kaplan-Meier curve with time on its x-axis and the probability of survival on the y-axis. It is a non-parametric statistic that estimates the survival function of time-to-event data. It constructs a survival curve from collected data and does not assume underlying distribution. It is also known as the product-limit estimator. As per client request in creating a model, a For Loop was generated which allows to plot the survival curve for each category that a retail unit belongs to by fitting the model to Kaplan Meier and identify their survival trend. For each category, it also gives details on the total number of observations fitted, and how many of these are right-censored observations. In addition, another For Loop was produced which would print a table with survival probabilities instead of plotting them. This is another efficient way to analyse their trends as it shows accurate details of survival probabilities for each year. However, it would also be interesting to understand what specific subcategories, amongst other factors, that drive the survival probability curve. This is because the categories are very broad and may not give the full picture. For instance, the category 'Clothes and Fashion' have more than 20 sub-categories such as maternity wear, leather goods and many more. Therefore, it becomes essential to analyse the sub-categories so that landlords can make informative decisions on which sections are worthy of being invested in the future. To create a For Loop for this analysis, it was firstly necessary to create subset data for each category and subsequently call the code. This step was necessary so that when code is called,

it will be able to identify which sub-category within the category to go to.

Cox Regression: The purpose of the Cox Regression is to evaluate simultaneously the effect of several factors on survival. It allows to examine how specified factors influence the rate of a particular event happening (i.e., when a retail company is most likely to close) at a particular point in time. It is a semiparametric model that is also referred to as the Cox proportional hazard model $h(t)$, as one of the assumptions of the model is that over the duration period, the hazards in the groups compared should be proportional to each other and that consequentially the hazard rate should be the same during follow-up. Another assumption is that the effects of the predictor variables upon survival are constant over time and are additive is one scale.

Weibull Model: The Weibull distribution is a parametric model, which is also known as the life data analysis. It is a continuous probability distribution that models time-to-event data precisely. The parameterised distribution for the dataset can be used to estimate not only the probability of failure at specific time (hazards), but also important life characteristics of a product, the mean life, the reliability, and the failure rate. The term 'life data' refers to measurements of product life which can be calculated in hours, miles, cycles, or any other metric that applies to the period of successful operation of a particular product. Since time is a common measure of product life, life data points are often referred as 'time-to-failure' or 'time-to-event'.

Key Findings

Kaplan Meier analysis helped to identify the survival probabilities for different categories and subcategories. This is a useful model for the client as it can now give informative recommendations to future investors on where to invest. It was identified that Hotel has the highest probability of survival, while Clothes and Fashion is one of the lowest. It was also identified that unit stores that have expanded internationally are more likely to survive more and their business rates are higher compared to store that have only expanded nationwide. Further analysis also showed business rates are highest in Greater London Region, in fact, the average survival duration in this region is relatively higher. It also showed that Hertfordshire is the county that with highest survival probabilities for the long term across all categories as it is one of the wealthiest

counties. Additional statistical analysis such as Weibull and Cox regression were able to identify the magnitude of the covariates on the survival rate. Both models resulted similar results where Hotel is the category that can survive the longest compared to any other category. Through AIC, it was established that Weibull is the model that can give better predictability since it scored the lowest

Value of the research

It is, therefore, that in such climate that becomes essential to understand and gain insights on the long-term fortunes of occupiers. The client's uniquely detailed and historical data allows to analyse the average length of operation for a variety of occupiers to understand the risks and opportunities landlords face during periods of tenant churn. The model that has been created will bring fruitful results for the company as it will allow them to make informed suggestions for their respective tenants and landlords.