CDRC Retail Centre Typology (Version 2)

Patrick Ballantyne1, Les Dolega1 and Alex Singleton1

1Geographic Data Science Lab, University of Liverpool

May 2022

Consumer Data Research Centre (CDRC)

[GitHub Repository](https://github.com/patrickballantyne/CDRC-Retail-Indicators/blob/main/README.md)

**1. Introduction**

The 2022 Retail Centre Typology presents an updated multidimensional typology of retail and consumption spaces across Great Britain. Retail presence within towns and city centres has become increasingly combined with other uses such as services and leisure activities, and the expansion of online retailing is causing a net loss of demand for some forms of retail floor space and change in function for others. This data product can be used to understand and comprehend the spatial provision of retailing and service activity, and better unpack the transformed roles and functions of consumption spaces.

As with the previous version of the CDRC Retail Centre Typology, available [here](https://data.cdrc.ac.uk/dataset/historic-retail-centre-boundaries) - construction of the typology focuses on four key domains; the *composition, diversity, function* and *economic health* of centres, and comprises two tiers of Supergroups and nested Groups. With this second iteration of the typology, we instead have four Supergroups, and eight nested Groups, for which descriptions in the form of pen portraits are provided.

**2. Methodology**

**2.1 Input Variables**

The Retail Centre Typology is available for 1,245 retail centres across Great Britain. This number is in line with those centres for which the CDRC [Retail Centre Indicators](https://data.cdrc.ac.uk/dataset/retail-centre-indicators) data product is available. These 1,245 centres comprise those not hierarchically classified as ‘Small Local Centres’, and those with sufficient coverage of the Local Data Company (LDC) dataset; a key contributing data source for this data product, as with the Indicators.

A total of 46 variables were gathered about the retail centres (Table 1), covering the four classification domains. The majority of these variables matched those used to construct the original CDRC Retail Centre Typology, with some of these calculated using open access data such as the Index of Multiple Deprivation and Open Retail Centre Indicators, code for which can be found on the associated [GitHub repository](https://github.com/patrickballantyne/CDRC-Retail-Indicators/blob/main/Analysis%20Code/Assembling%20OPEN%20typology%20variables.R). Many of the variables were calculated inside the CDRC Safe Researcher environment (DSH), using the Local Data Company (LDC) Retail Type, Location and Vacancy datasets. A number of variables were not calculated for this version, and a description and justification of this can be section 4.5. For detailed information about the construction of each of the 46 variables, see the Appendix.

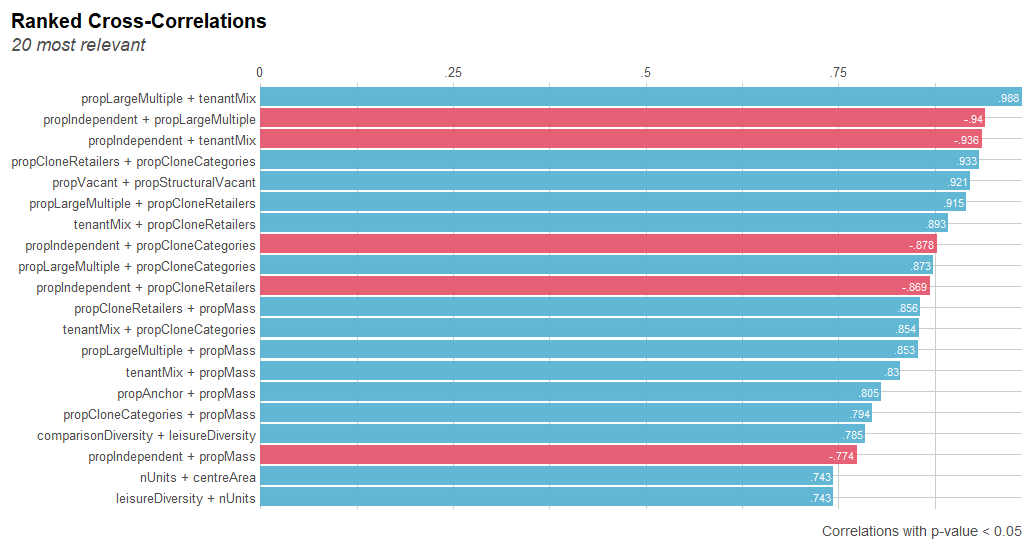
**Table 1.** Variables used to assemble typology.

|  |  |
| --- | --- |
| **Domain** | **Variables** |
| **Composition** | propClothingandFootwear, propDIYandHousehold, propElectricalandHomeEntertainment, propRecreational, propFood, propCTNOffLicence, propChemist, propRestaurantsandFastFood, propPubsBarsandClubs, propCafe, propEntertainmentandFitness, propHealthandBeauty, propConsumerServices, propHouseholdServices, propBusinessServices. |
| **Diversity** | propIndependents, propSmallMultiple, propLargeMultiple, tenantMix, comparisonDiversity, convenienceDiversity, leisureDiversity, serviceDiversity, propCloneRetailers, propCloneCategories. |
| **Size & Function** | nUnits, centreArea, geographicReach, roeckScore, attractivenessScore, propAnchor, propPremium, propMass, propValue, propCharity, propPawnbrokersandBetting, propConcession. |
| **Economic Health** | nCompeting, propVacant, propStructuralVacant, vacancyChange, onlineExposure, retailService, retailChange, averageIMD, nightPopulation. |

**2.2 Feature Selection**

Once the variables had been assembled, we performed a series of checks on the variables to ensure none were highly correlated with each other, which would otherwise skew the results of this classification. We plotted the highest/lowest correlations, as below in Figure 1, and used this to inform removal of four of the 46 variables; *propLargeMultiple, tenantMix, propCloneCategories* and *propStructuralVacant*.

Once these variables had been removed, we were left with 42 variables, which were supplied as inputs to the typology. The variables were range-standardised, but not weighted as this was found to have a limited effect on the construction of the typology in the [previous iteration](https://journals.sagepub.com/doi/full/10.1177/2399808319840666). Code for the feature selection can be found in the associated [GitHub repository](https://github.com/patrickballantyne/CDRC-Retail-Indicators/blob/main/Analysis%20Code/Multidimensional%20Typology.R).



**Figure 1.** Variable correlations

**2.3 Typology Construction**

The Retail Centre Typology was constructed using an unsupervised machine-learning algorithm called PAM; partition around (the) medoids, in line with the methods used to construct the first iteration of the CDRC Retail Centre Typology. It works similarly to k-means, but performs better when input variables are heavily skewed, as in our case. The algorithm works by taking the 42 input variables about the retail centres, and splitting them into ‘clusters’ or Supergroups/Groups based on identified similarities and differences between them. Code for typology construction can be found in the associated [GitHub repository](https://github.com/patrickballantyne/CDRC-Retail-Indicators/blob/main/Analysis%20Code/Multidimensional%20Typology.R).

However, in order to split the retail centres into typology Supergroups and Groups, we have to identify the number of ‘clusters’ or Supergroups/Groups, supplied as *k* to PAM. To do this, we utilised clustergrams and average silhouette scores, to identify the optimal value of *k* when identifying the number of Supergroups, and then when identifying the number of nested Groups within each Supergroup.

The average silhouette scores and clustergram for the Supergroups can be seen below in Table 2 and Figure 2. The maximum average silhouette score was used to identify the optimal *k* value when splitting the Supergroups into nested Groups.

**Table 2.** Determination of optimal *k* value for Supergroups

|  |  |
| --- | --- |
| ***k*** | **Average silhouette score** |
| 2 | 0.081 |
| 3 | 0.094 |
| **4** | **0.096** |
| 5 | 0.073 |
| 6 | 0.065 |
| 7 | 0.070 |
| 8 | 0.065 |
| 9 | 0.054 |
| 10 | 0.055 |

A picture containing chart

Description automatically generated

**Figure 2.** Determination of optimal *k* value for Supergroups

**3. Retail Centre Typology**

As above, the Retail Centre Typology consisted of four Supergroups and eight nested Groups. Examination of variable values within each of the ‘clusters’ or Supergroups/Groups resulted in the creation of Supergroup/Group labels and associated pen portraits summarising their key characteristics. A breakdown of the Retail Centre Typology can be seen below in Table 3.

**Table 3.** Retail Centre Typology

|  |  |  |  |
| --- | --- | --- | --- |
| **Supergroup Code** | **Supergroup Name** | **Group Code** | **Group Name** |
| 1. | Local 'everyday' goods and service centres | 1.1. | Local urban convenience centres |
| 1.2. | District and urban service centres |
| 2. | Retail and shopping parks | 2.1. | Primary shopping centres and premium destinations |
| 2.2. | Secondary retail parks and shopping centres |
| 3. | Leading comparison and leisure destinations | 3.1. | Large regional retail and leisure destinations |
| 3.2. | Sub-regional retail and leisure destinations |
| 4. | Traditional high streets and market towns | 4.1. | Mass and value traditional high streets |
| 4.2. | 'Indie' traditionalhigh streets |

**3.1 Supergroup 1 –** Local *‘*everyday*’* goods and service centres(n = 314)

Predominantly local and retail service centres, just under half of which are located in Greater London. They are often linear and provide a generally high independent offer, focusing on ‘everyday’ goods (e.g. groceries, confectionary), and offering a diverse range of services, with limited provision of comparison retail. They are typically characterised by higher proportions of convenience retail, business services, have seen recent increases and relatively high vacancy rates, and are typically comprised of higher average numbers of pawnbrokers, betting shops and fast food/takeaway outlets.

**Examples:** *Smithdown Road (Liverpool), Dunstable Road, Bury Park (Luton), Caledonian Road, Islington (London), Whitby Road (Ellesmere Port).*

**FChart, histogram

Description automatically generatedigure 3.** Characteristics of retail centres in the first supergroup

**Table 3.** Characteristics of groups nested within the first supergroup



**3.2 Supergroup 2 –** Retail and shopping parks (n = 105)

A very distinctive cluster of typically out-of-town retail developments, occupied by high proportions of ‘clone’ retailers and anchor stores. These centres specialise in mass comparison retail goods, offer very limited business and consumer services, convenience retail and generally have a low vacancy rate. They are more attractive than most other retail centres, providing free car parking, and attracting typically car borne consumers from a much wider area.

Chart, histogram

Description automatically generated**Examples:** Cheshire Oaks Designer Outlet (Ellesmere Port),Bluewater Shopping Centre (Dartford), Lakeside Shopping Centre (Grays), Trafford Centre (Manchester), York Designer Outlet (York).

**Figure 4.** Characteristic of retail centres in the second supergroup

**Table 5.** Characteristics of groups nested within the second supergroup



**3.3 Supergroup 3 –** Leading comparison and leisure destinations (n = 215)

These are the main retail destinations of regional and sub-regional importance, and also generally including the larger market towns. These large, expansive and highly attractive centres have large catchment areas and have diverse and comprehensive retail, leisure and hospitality offers, comprising large numbers of national chains. They have smaller proportions of independent retailers, less spatial competition than other centres, but have experienced relatively high increases in vacancy rates.

**Examples:** Liverpool City, City of London, Chester City, Manchester City, Sheffield City, Cardiff City, Blackpool City.

**Chart, histogram

Description automatically generated**

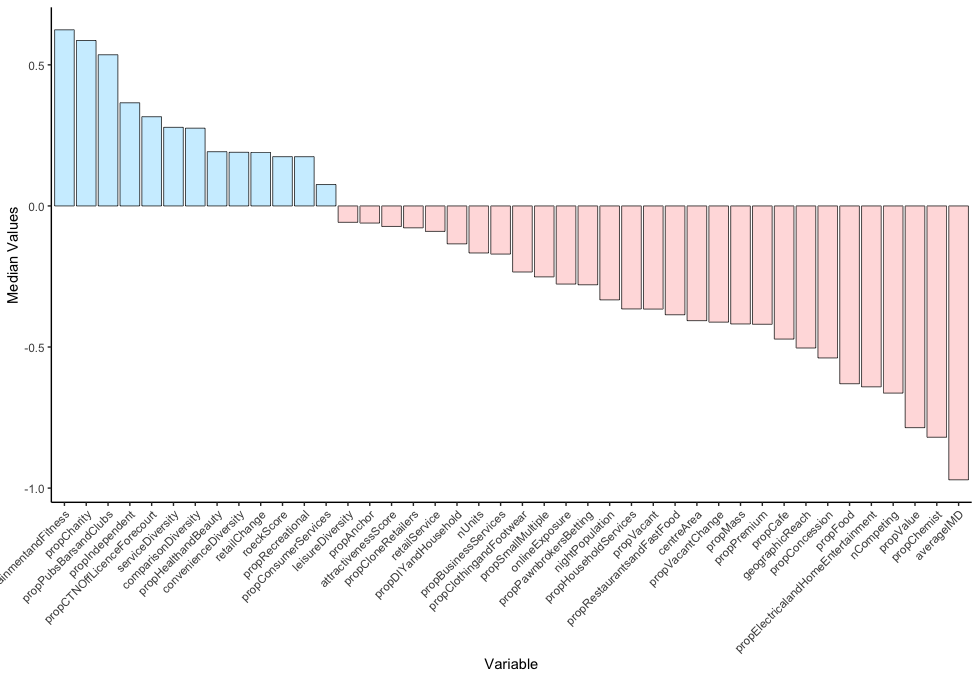
**Figure 5.** Characteristics of centres in the third supergroup

**Table 6.** Characteristics of groups nested within the third supergroup

**3.4 Supergroup 4 –** Traditional high streets and market towns (n = 611)

These are highly diverse but traditional high streets and market towns, which focus on hospitality, services, convenience goods and non-fashion comparison goods. These centres are characterised by high proportion of independent retailers, low vacancy rates, with typically low levels of premium, mass and value retailers. The spatial competition is relatively low and their catchments are characterised by low levels of deprivation.

**Examples:** Liverpool Road, Great Crosby (Liverpool), Wilmslow (Cheshire), Sandbach (Cheshire), High Street (Harpenden) Kings Road (Chelsea).



**Figure 6.** Characteristics of centres in the fourth supergroup

**Table 7.** Characteristics of groups nested within the fourth supergroup



**4. Appendix – Assembly of Variables**

**4.1 Domain 1 - Composition**

A number of variables were used to capture differences in the overall composition of retail centres, using the 2020 release of the Local Data Company dataset. Each variable was calculated by extracting the total number of each type of retail (e.g. ClothingandFootwear, ConsumerServices) in each retail centre, before calculating the proportion of these relative to the total number of units in each retail centre. Only the active LDC units   
(e.g. PremiseStatus == “Live”) were used to construct these different variables. The full list and provenance of these is listed below.

*propClothingandFootwear* – all LDC units pertaining to clothing and footwear retail goods, extracted by identifying all LDC units in the c(“Fashion & General Clothing”, “Clothes & Fashion”, “Footwear”) categories.

*propDIYandHousehold* – all LDC units in the following categories; c(“DIY, Hardware, Builder’s Merchants & Household Good”, “Furniture, Carpets, Textiles, Bathrooms & Kitchens”, “Home & Garden”, “Household & Home).

*propElectricalandHomeEntertainment* – all LDC units in the following categories; c(“Electrical Goods & Home Entertainment”), and some additional subcategories; c(“Audio Visual Rental”, “Computers”, “Electrical Goods”, “Home Entertainment”, “Mobile Phones”, “Printer Cartridges”, “Printers”, “Records, Tapes & CD’s”, “Sewing Machines”, “Video Library”).

*propRecreational* **–** all LDC units in the following categories; c(“Books, Arts & Crafts, Stationery, Printers”, “Sports, Toys, Cycle Shops & Hobbies”), and a number of additional subcategories; c(“Camping Goods & Outdoor Wear”, “Model Shops”, “Pianos & Accessories”, “Potteries & Ceramics”, “Skiing Equipment”, “Sports Goods Stores”).

*propFood* – All LDC units pertaining to non-service Food (e.g. excluding takeaways, restaurants etc.) in the following categories; c(“Groceries, Supermarkets & Food Shops”, “Food & Drink”, “Butchers & Fishmongers, “Bakers”), and a number of additional subcategories; c(“Convenience Stores”, “Grocers”).

*propCTNOffLicenceForecourt* **–** all LDC units in the following categories; c(“Confectionary, Tobacco, Newsagents”, “Off Licences”, “Petrol Filling Stations”), with all Petrol Filling Stations attributed to a Supermarket retailer extracted by removing LDC units with the following MultipleNames: c(“Tesco Petrol Station”, “Sainsbury’s Filling Stations”, “Asda Petrol Station”, Morrisons Petrol Filling Stations”) removed, to keep only Petrol forecourts.

*propChemist* – all LDC units in the “Chemists / Toiletries” Subcategory.

*propRestaurantsandFastFood* – all LDC units in the “Restaurants” category and a number of cubcategories containing Fast Food and Takeaway units; e.g. (“Fish & Chips Shops”, “Pizza Takeaway”, “Indian Takeaway”). A filter(grepl(“Fast Food”, Subcategory)) and filter(grepl(“Takeaway”, Subcategory)) will pull most of these up.

*propPubsBarsandClubs* **–** all LDC units in the “Pubs, Bars & Clubs” and “Bars, Pubs & Clubs” categories.

*propCafe* **–** all LDC units in the following subcategories; c(“Café & Tearoom”, “Coffee Shops”, “Internet Cafes”).

*propEntertainmentandFitness*– all LDC units in the “Entertainment” and “Events & Attractions” categories (e.g. Cinemas, Tourist Attractions), and all LDC units in the “Health Clubs” subcategory (e.g. gyms).

*propHealthandBeauty*– all LDC units in the following Subcategories: c(“Alternative & Complementary Medicines”, “Barbers”, “Beauty Salons”, “Hair & Beauty Salons”, “Hairdressers”, “Herbalists”, “Nail Salons”, “Tanning Shops”, “Tattooing & Piercing”, “Beauty Products”, “Health & Beauty Shops”, “Health Foods & Products”).

*propConsumerServices*– all LDC units in the “Travel Agents”, “Banks & Other Financial Institutions” and “Estate Agents” subcategories.

*propHouseholdServices*– all LDC units in the “Laundries & Laundrettes”, “Locksmiths” and “Car Dealers” subcategories.

*propBusinessServices*– all LDC units in the “Recruitment Agencies”, “Wholesalers” and “Solicitors” subcategories.

**4.2 Domain 2 - Diversity**

A number of variables were used to capture the overall diversityof retail centres. Many are based on the original variables used in the previous typology, however some are new additions from the recent ‘retail centre indicators’ data product. The full list and provenance of these can be seen below.

*propIndependent* – all independent retailers/service providers as defined by LDC (units containing an NA value in the MultipleID and MultipleName columns).

*propSmallMultiple* – all small multiples comprising LDC defined multiple units (not containing an NA value in the MultipleID and MultipleName columns) and belonging to a ‘chain’ with 10 or less stores across the UK.

*propLargeMultiple* - all LDC units not containing an NA value in the MultipleID and MultipleName columns and belonging to a ‘chain’ with more than 10 stores across the UK.

*tenantMix* **–** the proportion of unique brands (i.e. those not containing an NA value in MultipleID/MultipleName) in each retail centre, in relation to the total number of units. Higher values indicate a greater dominance and diversity of major brands in each retail centre.

*comparisonDiversity*– proportion of ‘comparison’ retail subcategories present in each retail centre, relative to the total number of ‘comparison’ retail subcategories nationally.

*convenienceDiversity*– proportion of ‘convenience’ retail subcategories present in each retail centre, relative to the total number of ‘convenience’ retail subcategories nationally.

*leisureDiversity* – proportion of ‘leisure’ retail subcategories present in each retail centre, relative to the total number of ‘leisure’ retail subcategories nationally.

*serviceDiversity*– proportion of ‘service’ retail subcategories present in each retail centre, relative to the total number of ‘service’ retail subcategories nationally.

*propCloneRetailers* – proportion of total LDC units in each retail centre made up by “clone” retailers, identified from the list obtained when calculating the ‘Clone Town’ measure for retail centres.

*propCloneCategories* – proportion of total distinct categories in each retail centre made up of “clone” retail categories, identified from the list obtained when calculating the ‘Clone Town’ measure for retail centres.

**4.3 Domain 3 – Size & Function**

A number of variables were used to capture the overall size and functionof retail centres. Many are based on the original variables used in the previous typology, however some are completely new (geographicalReach), and some have been modified to account for changes in the retail sector and LDC data (e.g. propAnchor, propPremium, propConcessions). The full list and provenance of these can be seen below.

*nUnits* **–** these values are taken from the ‘retail centre indicators’ data product, summarising the total number of units in each retail centre (active & vacant).

*centreArea* – the total area (km2) of each retail centre boundary.

*geographicalReach* – the total area (km2) of the drive-time catchment for each retail centre.

*roeckScore* **–** a summary of the morphology of the retail centre boundary, which is calculated by calculating the area of the retail centre boundary, relative to the area of its minimum bounding circle.

*attractivenessScore* **–** a composite index calculated by bringing together five individual variables that are deemed to determine the overall ‘attractiveness’ of retail centres; nUnits, propAnchor, propLeisure, tenantMix and propVacant (with a negative weight).

*propAnchor* – all LDC units in a list of anchor fascia were extracted. The list of anchor fascia comprised comparison, convenience and leisure fascia, and was compiled from the original list from 2019, adding in some additional anchors that were originally missed out and removing those that have since closed down (e.g. Debenhams). It is important to note that all c-stores were excluded from the list of anchors (e.g. Tesco Express), keeping only the larger supermarket stores as convenience anchors (e.g. Tesco Metro).

The list was as follows; c(“John Lewis & Partners”, “John Lewis & Partners at Home”, “Marks & Spencer”, “Harvey Nichols”, “House of Fraser”, “Selfridges”, “Harrods”, “IKEA”, “H&M”, “H&M Kids”, “Primark”, “Zara”, “Zara Home”, “Boots the Chemist”, “Next”, “Next Home”, “B&Q”, “Homebase”, “Tesco”, “Tesco Metro”, “Tesco Extra”, “Sainsbury’s”, “ASDA”, “ASDA Supercentre”, “Morrisons”, “Odeon Cinema”, “Vue”, “Cineworld”).

*propPremium*– all LDC units in a list of premium fascia were extracted. The list of premium fascia was compiled from the original list from 2019, adding in some additional premium retailers that were originally missed out using this report[[1]](#footnote-1), and removing those that have closed down (e.g. Cath Kidston).

The list was as follows; c(“Waitrose & Partners”, “Little Waitrose & Partners”, “John Lewis & Partners”, “John Lewis & Partners at Home”, “Harvey Nichols”, “Laura Ashley”, “Laura Ashley Home”, “Ted Baker”, “Tommy Hilfiger”, “Fat Face”, “Superdry”, “Seasalt Cornwall”, “Jack Wills”, “White Stuff”, “Crew Clothing Co.”, “Hugo Boss”, “Joules”, “Swarovski”, “Lacoste”, “Diesel”, “Bose”, “Ralph Lauren”, “Calvin Klein”, “Dior”, “Gucci”, “Pandora”, “Tiffany & Co”, “Barbour”, “Dolce & Gabbana”, “Emporio Armani”, “Armani Exchange”, “Cartier”, “Louis Vuitton”, “Omega”, “Versace”, “Burberry”, “Hermes”, “Prada”, “Michael Kors”, “TAG Heuer”, “L’Occitane”, “Jo Malone London”, “Jimmy Choo”, “Paul Smith”, “Mulberry”, “Montblanc”.

*propMass*– all LDC units in a list of mass fascia were extracted.

The list of mass fascia was; c(“Tesco”, “Tesco Express”, “Tesco Metro”, “Tesco Extra”, “Sainsbury’s”, “Sainsbury’s Local”, “ASDA”, “ASDA Supercentre”, “Greggs”, “Boots the Chemist”, “Superdrug”, “Holland & Barrett”, “Argos”, “WHSmith”, “Next”, “Next Home”, “H.Samuel”, “JD Sports”, “Claire’s”, “River Island”, “TK Maxx”, “H&M”, “H&M Kids”, “Zara”, “Zara Home”, “Topshop”, “Game”, “The Perfume Shop”, “Starbucks Coffee”, “Costa”, “Caffe Nero”, “McDonalds”, “KFC”, “Subway”, “Domino’s Pizza”, “Pizza Express”, “Currys PC World”, “The Fragrance Shop”).

*propValue*- all LDC units in a list of value fascia were extracted.

Value fascia were; c(“Aldi”, “Lidl”, “Iceland”, “Primark”, “Poundland”, “Farmfoods”, “Sam 99p”, “Poundworld Plus”, “Poundstretcher”, “Home Bargains”, “Savers”). Once the list of value stores had been assembled, the proportion of these relative to total units in a retail centre was calculated.

*propCharity*- all LDC units in the subcategory ‘Charity Shops’ were extracted, and the proportion of these relative to all units in a retail centre was calculated.

*propPawnbrokersBetting*– all LDC units in the subcategories “Pawnbrokers” and “Bookmakers” were extracted, and the proportion of these relative to all units in a retail centre was calculated.

*propConcession* **–** all LDC units classified as “1” in the ‘concession’ field were extracted   
(c. 21,000), and the proportion of these to all units in a retail centre was calculated.

**4.4 Domain 4 – Economic Health**

A number of variables were used to capture the overall economic healthof retail centres. Many are based on the original variables used in the previous typology, however some are new additions from the recent ‘retail centre indicators’ data product. The full list and provenance of each can be seen below.

*nCompeting*– here we count the total number of directly competing functions within the drive-time catchment of each retail centre. We treat regional centres, major town centres, town centres, large retail parks and large shopping centres as directly competing entities; i.e. they can only compete with eachother. In addition, we calculate it so that district centres, market towns, small shopping centres and small retail parks can compete with all types of retail centres.

*propVacant*– proportion of vacant units to all LDC units in each retail centre, obtained from the safeguarded ‘retail centre indicators’ data product.

*propStructuralVacant*–proportion of vacant units that have been vacant since 2018 to all LDC units in each retail centre. This is obtained by tracking whether the OccupierID remains vacant between 2018 and 2020.

*vacancyChange* – change in the proportion of vacant units between 2017 and 2020, obtained from the safeguarded ‘retail centre indicators’ data product.

*onlineExposure*– online exposure score for each retail centre, obtained from the ‘retail centre indicators’ data product.

*retailService*– ratio of retail (comparison + convenience): services in 2020.

*retailChange*– the ratio of retail (comparison + convenience): services in 2017 subtracted from the retail:service ratio in 2020.

*averageIMD* – average IMD score of the retail centre walking catchment, obtained from the latest version of the deprivation profiles. This is used as a proxy for relative affluence of the retail centre, as use of census variables is problematic when developing a cross-national typology.

*nightPopulation*– here we calculate the total residential population within the walking catchment of each retail centre, using population estimates from 2019 for England and Wales at LSOA level and Scotland at Data Zone level.

* 1. **Notes**

It was not possible to compute *rateable value* for the retail centres, as the VOA dataset is not available for Scotland.

It was also not possible to calculate *crime and unemployment rates* for the retail centres, as the measurement of each varies between England/Wales and Scotland.

*Economic activity* would be a useful measure, when we receive data from Geolytix we might look into this.

1. The Most Popular Luxury Brands UK - <https://yougov.co.uk/ratings/consumer/popularity/luxury-brands/all> [↑](#footnote-ref-1)