

Top 5 External Data Types for the Retail Industry

Learn more about the top five external datasets that retailers can collect to better support store-level strategic and tactical decisions for their entire network.

Covid-19 is having an extraordinary effect on the retail industry. Multinational companies with significant retail operations have had to close their physical stores and are waiting to see when and how consumption will resume and revenue will recover. To support store level sales & marketing decisions, teams of data scientists are building models, predictions, forecasts, and simulations. The quality and depth of these models mainly depend on the available data.

In this article, we briefly introduce the top five external datasets that retailers can collect to better support store-level strategic and tactical decisions for their entire network. It is important to note that these datasets are very useful in more “normal” times, and there are multiple business decisions where these external datasets can play an essential role.

Catchment area

Every store has its own catchment area (the region where a store is collecting the majority of its shoppers). Unfortunately, catchment areas do not follow urban planning boundaries, district borders, electoral or census regions, etc. In most cases, there are three opportunities to attract customers to a physical store: proximity to their home (residential area), proximity to their workplace (e.g. CBD) and proximity to a commuting hub (airport, underground, etc.). The first external dataset one should try to collect is a GPS- level location dataset, identifying residential areas, work areas and commuting hubs. Some of these datasets can be purchased (e.g. residential address book, company registry, etc.), but there are many open sourced and crowd-sourced data available (4square, open street map, etc.)

Competitors and Externalities

While the catchment area can help to gain some understanding for local demand and customer profiles, it is essential to know how many

competitors (and which ones) are currently supplying these customers. The more is known about direct competitors with whom you are sharing the catchment area, the more precise can consumption at your store can be modelled. For example, you can easily collect basic information from Google maps (geo-location, opening hours, popular hours, etc.), but in many cases, the competitor's website can also provide valuable information. A few data vendors specialize in the collection, organization, and historical analysis for specific Point of Interests (POIs). One example is Nokia, which sells data about thousands of categories (shops, petrol stations, banks, post offices, etc.) Keep in mind that some POI categories may not behave as competitors but rather boosters for consumption. For example, the nearest store to a hospital might benefit from being the first one on the way, if selling the right items for a relevant segment.

Prices

We all know that price is the main driver for purchasing decisions. Prices and discounts for competing and complementary products are key. And this applies for both current and historical prices. Web stores and price aggregator websites are maintaining item level prices and discounts accurately and can also help to identify stock-outs. Naturally, at the store level it is possible to deviate from "list prices" if needed, based on inventory levels, specific promotions run by local competing stores, etc. Store-level discounts and promotions for competitors can be acquired using your own store employees, with a very simple and old-fashioned survey method.

Voice of the Customers

Customer complaints, reactions and questions related to your brand (and your competitors') are other important pieces of external information. Blogs, forums, e-commerce sites, social media and digital news are easily accessible sources of information for this type of data. However, transforming unstructured textual data to structured and organized tables is a massive and sometimes painful procedure, including the building and maintenance of different web-crawling

engines with Natural Language Processing (NLP) capabilities for multiple languages. There are a few vendors who are ready to provide already structured reports and databases from selected and researched digital media, and others who are happy to help in developing and maintain a solution for your own focus areas.

Digital Footprint

It is possible to obtain very large and detailed datasets about customer digital footprints. Different vendors are selling different information resolutions, depending on the data they collect and what they are allowed to share by applicable laws and regulations. These are social media and internet companies, mobile telecom operators, digital survey and crowd-sourced data vendors, digital ad networks, etc. Depending on the regulations in your country (and your appetite), you can purchase segment level information, or even customer device level information, on potential customers who passed within a given radius of your stores.

These are just a few examples of supplemental datasets that can be used for retail network data science models. There are hundreds of many other types. In our extremely digitized world, we can probably say that there is no data that cannot be accessed, collected, or approximated with relatively high accuracy. Of course, the costs of these external datasets can add up quickly and the benefits are uncertain until you tried them in your models and hopefully improved your business outcomes.