REGION
The Journal of ERSA
Powered by WU

Volume 12, Number 2, 2025, 1–22
DOI: 10.18335/region.v12i2.573

journal homepage: region.ersa.org
ISSN: 2409-5370

Know your competitor! Analyzing and predicting the location of competing stores: The case study of Valora at Swiss railway stations

Thomas Wieland[1]

[1] Independent researcher, Freiburg, Germany

Received: 30 October 2024/Accepted: 10 July 2025

Abstract. Location choice in retailing is a key subject of retail location theory, but is also of great practical relevance. Retail companies must assess the demand and competition situation and try to anticipate the behavior of their competitors. This study examines location choice patterns of two convenience food formats from Valora, Avec and k kiosk, at Swiss train stations. The study combines an analytical and a predictive modeling approach using econometric and AI/machine learning techniques. Possible location factors for the two formats are derived from the literature. Publicly available data from the SBB (Swiss Federal Railways) serve as the basis of the analysis. Binary logit models are built for the formats examined in order to identify the determinants of location choice. Machine learning algorithms are used to check and optimize the predictive ability of the models. It turns out that people boarding, alighting, and changing trains at train stations (which represent the main demand for convenience stores at railway stations) are an important determinant of location choice. The more frequent a train station is, the more likely it is that Avec or k kiosk will be present there. Furthermore, format-specific clustering and avoidance patterns emerge. Both Valora formats show an avoidance of each other. While Avec tends to avoid competing convenience supermarkets, this is not the case with k kiosk. With the help of machine learning, the predictive ability of the models can be greatly improved. A prediction model with high specificity and sensitivity is built for k kiosk and applied to a real case.

1 Introduction

A retailer's location choice, along with consumers' store choice, is a central topic of retail location theory, with both aspects being closely related (Reigadinha et al. 2017, Wieland 2023). Location planning is a complex process. Retail companies have been using quantitative methods in location planning for decades in order to estimate the performance of new stores (Aversa et al. 2018, Reynolds, Wood 2010). Retailers must consider demand conditions when choosing locations, and the behavior of competitors must be anticipated as well. On the one hand, a retailer competes with another retailer for the same retail space and must, for example, base its rental offer on what a competitor would pay for it. On the other hand, it must be taken into account that competitors or other non-competing stores are already located at the respective location or may expand there later. The effects of co-location with these other stores must be predicted as accurately as possible. For example, the presence of certain competitors can reduce expected sales, while others may have a positive effect. All retailers are operating under uncertainty because they do not have complete information (Orhun 2013, Zhu, Singh 2009).

Railway stations were originally purely transport hubs for people and goods. In recent decades, however, they have become important locations for retail, catering, and other services. This particularly affects large train stations that are connected to the high-speed long-distance network, regardless of whether they are centrally or peripherally located within a large city (Bills 1998, EHI Retail Institute 2023, Office of Rail and Road 2024). This is also due to the fact that mobility in European societies is increasing, particularly due to commuting, and in particular – with a short decline in the context of the Corona pandemic – train traffic is gaining (Eurostat 2024). Although transport hubs, particularly train stations, are important retail locations, they have been neglected in location research, which has mainly focused on city centers, shopping malls, and locations of supermarkets. Train stations are often seen as traffic generators for retail but are not treated as retail locations in their own right (Nilsson, Smirnov 2016, Rao, Pafka 2021). This is also problematic because several retail chains are focusing on station locations, and European national railway companies, through their own property companies, are promoting their stations as built retail destinations, similar to shopping malls (DB Station&Service AG 2017, OEBB 2024, SBB 2024b, SNCF Gares & Connexions 2024).

The *Swiss Federal Railways* (*SBB*) is the main railway company in Switzerland and has around 1,200 train stations, which are frequented by around 1.3 million travelers every day (SBB 2024c). It is also one of the largest real estate operators in Switzerland and is developing its railway stations into business centers, which makes the transport company a key player in the Swiss retail market. Commercial space in train stations is regularly advertised for retail or catering use (Neue Zürcher Zeitung 2024, SBB 2024b). The retail spaces in *SBB* train stations are highly sought after for the expansion of several food retail companies such as *Migros*, *Coop*, *Aldi*, and *Lidl*. *SBB* generally grants only fixed-term leases, which are re-tendered or renegotiated with the existing tenant after, for example, five or ten years. It often happens that the space is then taken over by a competitor (Blick 2025, Radio Frequence Jura 2024). Accordingly, there is strong competition both between retail stores in the stations and during expansion for retail space.

This study examines the location choice of the *Valora Holding AG* with regard to two convenience formats, *Avec* and *k kiosk*, at *SBB* train stations in Switzerland. *Valora* is an internationally active Swiss retail company and specializes in food convenience stores, particularly in high-frequency locations. *Valora* currently has 13 formats and around 2,800 stores in Switzerland and other European countries. In addition to *Avec* and *k kiosk*, these include catering formats such as *Caffé Spettacolo*, or bakery chains such as *Back Factory* and the Swiss branch of *Backwerk* (Valora Holding AG 2024d). *Avec* is a convenience store format with fresh to-go food and a narrow supermarket assortment that advertises with the slogan “Handmade with Love”. The largest *Avec* stores in Switzerland are located in the St. Gallen (360 m2) and Andermatt (227 m2) train stations (Valora Holding AG 2019, 2024a). The kiosk format *k kiosk* advertises with the slogan “Gönn Dir was!” (“Treat yourself!”) and focuses on tobacco, lottery products, snacks, and press. Originally, *k kiosk* stores were “kiosks” in the literal sense, i.e., small, often free-standing outlets in the form of a tiny house or booth that customers cannot enter. However, in recent years, more and more stores have been opened that are accessible to customers and have been expanded in terms of their (food) product range (Valora Holding AG 2024c).

This study follows a two-pronged strategy, namely an analytical and a predictive modeling approach. The first research question is of an analytical nature: *What are the determinants of the location choice of Valora convenience formats at Swiss train stations?* For this purpose, potential location factors are derived from the literature and empirically tested for their significant effect using a micro-econometric model. The second question relates to the applicability of the results for predictions: *How well can the location choice of Valora convenience formats be predicted for new situations?* For this purpose, a set of machine learning models are created, compared, and tested for their predictive ability. The study uses publicly available data sets published by *SBB*.

The paper is structured as follows. Section 2 provides an overview of the existing literature. The broad outlines of retail location theory as well as empirical studies on location choice and store performance in retailing are presented. In Section 3, independent variables of location choice with respect to the store formats examined are derived. This is followed by a description of the data sets used and the analysis and prediction models. Section 4 presents the results of the analytical model, a review of several modeling approaches in terms of their predictive ability, and an application example in which the best model is used to predict site selection. In Section 5, the conclusions are summarized and the limitations of the study are addressed.

2 Literature review

2.1 Retail location theory

Four approaches are usually attributed to retail location theory, namely (1) *central place theory*, (2) *market area models*, (3) *bid rent theory*, and (4) models of *retail agglomeration* (Reigadinha et al. 2017, Wieland 2023). The central place theory by Christaller (1933) is primarily concerned with spatial consumer behavior. Utility maximization is assumed for consumers, which means minimizing the transport costs they have to bear for a shopping trip. The demand for a specific good decreases as transport costs increase (*distance decay*). The willingness to accept traveling varies between goods depending on the frequency with which the goods are bought. The *lower range* is the minimum demand of a good that is necessary to maintain it in an economically viable manner (*demand threshold*). The *upper range* is the furthest distance up to which consumers will purchase a good offered. Profit maximization is assumed for the providers of central goods. This includes avoiding direct competitors, while clustering is assumed for suppliers of complementary goods. The result is a hierarchical system of central places of different sizes, with several goods being offered in each of these locations. Each central place has a supplementary area whose spatial extent corresponds to the upper range of the highest-ranking good offered. The theory has been formalized and extended over decades, for example, with special consideration of multi-purpose shopping (Eaton, Lipsey 1982, Ghosh 1986).

Market area models are mathematical models for calculating customer and sales flows for locations, with distance decay playing a central role. The first approaches by Reilly (1931) and Converse (1949) are deterministic and divide a market area between two locations. The probabilistic model by Huff (1962) determines the probabilities that customers from a set of origins will shop there in a system of supply locations. Consumer utility is explained by two variables that are based on microeconomic assumptions. The size of a location is regarded as an attractiveness indicator because shopping decisions are made under uncertainty, and the probability of being able to purchase the desired goods increases with the size of the location. However, *diminishing marginal utility* is assumed for the size. Consumer travel time has a non-linear negative effect on store choice because the trip to the shopping location is interpreted as *opportunity cost*. There are countless extensions to this model, e.g., to take into account the image of store chains (Stanley, Sewall 1976) or agglomeration effects (Fotheringham 1985).

The *bid rent theory* by Alonso (1964) explains urban land use and rent dynamics based on accessibility and distance from the city center. It posits that land value decreases as distance from the center increases, due to transportation costs and demand for proximity to amenities. Rent is determined by the willingness of different users to pay for location. Residential and commercial users compete for space, leading to higher rents in more desirable areas. Land near the center tends to be used for activities with higher output per area unit (e.g., retail), while outer areas are utilized for lower-value uses. The model illustrates the trade-off between land use and transportation costs, emphasizing that urban growth patterns depend on economic activities and population density.

The fourth strand of retail location theory goes back primarily to Hotelling (1929), who describes a duopoly in a linear market, where suppliers can change their location to maximize their demand. In this specific case, the best location structure for both providers is that they are located right next to each other and serve the left or right half of the market (*principle of minimum differentiation*). The influential work by Nelson (1958) stems from empirical-inductive location research. Based on customer surveys at retail locations, Nelson derives three elements of retail location success. Apart from their own attractiveness (*generative business*), the sales of retailers also depend on the attraction of compatible stores at the same location (*shared business*) and external customer frequency generators such as workplaces or public transport stops (*suscipient business*). Shared business consists, on the one hand, of the *cumulative attraction* of competitive suppliers, which arises from the fact that customers have the opportunity for *comparison shopping*, and, on the other hand, of the advantages resulting from the compatibility with other stores, which enable *multi-purpose shopping*. Nelson also derives a mathematical formula for calculating the customer exchange between two compatible retailers (*rule of retail compatibility*) and creates compatibility tables for a number of retail industries.

These topics were also dealt with early in microeconomics. Chamberlin (1933) discusses the clustering of complementary and competing stores in his seminal work towards the *theory of monopolistic competition*. Following this, it always makes sense for suppliers whose products are perfect substitutes to avoid competitors because this means they have a *monopoly on location*. In contrast, spatial clustering is favorable for suppliers of imperfect substitutes or complementary goods because the former enables comparison shopping and the latter allows for multi-purpose shopping. These considerations were later expanded and formalized with regard to incomplete consumer information (Nelson 1970, Wolinsky 1983).

Later, many of these older theories were incorporated into the microeconomic models of the "*New Economic Geography*". Here, central elements, some of which were only formulated verbally, were converted into fully mathematical equilibrium models, for example by Fujita et al. (2002) and Tabuchi, Thisse (2011).

2.2 Empirical studies regarding location choice and store network expansion

There is a heterogeneous collection of literature from economic geography and regional economics on the topic of retail location choice and store network expansion, respectively. Typically, these studies are concerned with drawing conclusions about the determinants of location choice from the empirical distribution of specific retail chains or store types. In many cases, hypotheses derived from retail location theory are empirically tested.

Larsson, Oener (2014) examine the location patterns of three retail business types in Swedish cities, especially with respect to clustering of stores. They use a geocoded database of all workplaces, and the urban areas are divided into small-scale grids. The degree of clustering is analyzed using Poisson count data models, with the number of stores in a particular industry in the grids being the dependent variable and the number of other stores and other locational variables acting as the independent variables. The presence of clothing stores is positively influenced by the presence of specialty shops and second-hand shops, while there is a negative relationship with household stores. The authors conclude that the complementarity of providers is based on the same or similar shopping frequency. It is also found that store presence is positively explained by small-scale demand (surrounding residents). Wieland (2017) applies a similar research concept to healthcare services in a rural German region. Special count data models (hurdle models) are used to investigate which location characteristics explain the number of general practitioners, psychotherapists, and pharmacies. The spatial aggregation level of the study is villages or districts. This shows that it is primarily the local demand potential that explains the number of providers examined. At the same time, clustering patterns can also be seen here; in particular, pharmacies tend to choose their location depending on the presence of medical practices.

In Canadian cities, Krider, Putler (2013) examine the location distribution of 54 retail industries and other consumer-oriented services in order to identify industry-specific clustering and avoidance patterns. Geocoded store addresses serve as the data basis to locate the individual stores. The authors use geostatistical measures to identify excessively random clustering of providers (Ripley's K, Kulldorff's D). Apart from differences between the cities, clear tendencies emerge: In particular, stores selling medium- and long-term goods (e.g., clothing, shoes, electrical goods, furniture) and specialist shops (e.g., delicatessens) tend to have a more or less strong small-area concentration. In contrast, non-specialized food retailers (e.g., supermarkets) as well as gas stations, liquor stores, pharmacies, and catering providers (e.g., ice cream shops) tend to avoid competitors.

Reigadinha et al. (2017) investigate the location structures of food retailers in a Portuguese region in order to test statements from classic retail location theory. They use point data from 273 stores that belong to eight large food retail chains. The evaluation is carried out using GIS analyses and regression models, where, among other things, store density and the distance to the nearest competitor are the dependent variables. They note a correlation between store density and population density as well as a tendency for competitors to cluster and interpret this as confirmation of the theories tested. Seong et al. (2022) examine location patterns of convenience shops in urban districts in South Korea. In their regression models, they examine the determinants of average convenience store sales and test for, among other things, the influence of convenience store density and supermarket density, while using footfall and local demand (surrounding residents and employees) as control variables. Convenience store density tends to have a positive influence on average sales, which is interpreted as a positive agglomeration effect. Supermarket density reduces sales, which can be understood as a competition effect. The footfall also increases average sales.

In a series of papers, Joseph, Kuby (2013, 2015, 2016) examine the location patterns and expansion strategies of US retail chains. Among other things, they identify that the chains have different expansion strategies and that the expansion is sometimes based on the location of the headquarters. Additionally, some chains began their expansion in large markets and continue to expand in large markets as well. The same applies vice versa to chains that initially focus on small markets. Rice et al. (2016) compare the expansion of *Walmart* and *Carrefour*. They note that *Walmart* tends to avoid competition and also serves more remote markets, while *Carrefour* primarily expands in urban areas and their metropolitan surroundings. Zhou et al. (2024) examine the expansion of Chinese electronics retailer *Suning*, counting the number of its stores at the prefecture level. They use a geographically weighted Poisson count data model. They note regionally different patterns of shrinkage and expansion. They also find that internet penetration is a predictor of the regional number of stores, in the sense that a high penetration rate can also lead to a reduction in store density.

2.3 Store performance models and predicting store sales

Another strand of literature deals with the determinants of store performance and the prediction of new store sales. The practical purpose here is to provide decision-making aids in operational location planning. Many large retail chains have been using model-based forecasts for decades (Aversa et al. 2018, Reynolds, Wood 2010). Here, regression models and/or machine learning approaches are used as well, incorporating store sales or store customer numbers as the dependent variable. Mostly, independent variables derived from retail location theory are tested empirically as well.

Possibly the first comprehensive scientific work on this topic comes from Taylor (1978), who examined the influence of location characteristics on the sales of two chains in the U.S.A., *Pizza Hut* and *Zale*. Among other things, direct competitors, population, employees, and the median income of residents in the area, as well as various aspects of micro-location, are examined as independent variables. As expected, there are positive effects from local demand and a sales-reducing effect from competition. A very early study on this is also that of Weber (1979), who examined the customer frequencies of pharmacies in a German city. Linear and intrinsically linear regression models are used to check which location factors have a significant influence. A distinction is made between locations in the city center and in the outskirts. In the first case, there is a positive influence of footfall and doctors practicing around the pharmacies, as well as a negative effect of other surrounding pharmacies. In the second case, the footfall (positive) and the distance to the nearest competitor (negative) also influence the number of customers.

Müller-Hagedorn (1991) deals with stove businesses and, unlike the previously mentioned authors, derives the location factors from the specific retail industry instead of from classic location theories. A theoretical distinction is made between, on the one hand, the situation of the consumer (e.g., level of knowledge about the products and the providers) and, on the other hand, the function of the location (time-saving or information function). Locations should therefore be evaluated differently depending on the product and type of buyer. In this case, it is argued that the information function of a location in particular makes a decisive contribution to sales, i.e., that the location should make the existence of the respective specialist store known. The result shows that both pedestrian and vehicle frequency as well as the size of the shop window correlate positively with sales.

Statistical forecasts of store performance became famous thanks to the *SLAM* (*Store Location Assessment Model*) by Simkin (1989), which – according to the author's own statement – was used for location planning after its development in several British retail chains. This is a linear regression model with, among other things, market size, accessibility, and the competitive situation of the stores as independent variables. Such models were frequently built and used in the following decades (Chang, Hsieh 2018, Themido et al. 1998). Wieland (2018) first used a panel data model to be able to take temporal effects into account; in this specific case, the yearly turnover of consumer electronics stores is investigated, with competition, regional demand, and time (as a proxy for the gaining relevance of online retailing) being the most important impacts on store performance. The topic received great attention again, especially with the emerging relevance of machine learning, which from then on was regularly used to optimize the predictive ability of such models (Ge et al. 2019, Lu et al. 2024, Ting, Jie 2022, Wang et al. 2018, Zhou et al. 2015).

Broadly speaking, almost all of these studies show positive effects of market size (e.g., residents within a travel time of X minutes) and negative effects of the presence of competitors (e.g., number of competitors in the same municipality or within a travel time of X minutes). Store characteristics are often also taken into account, e.g., store size, which typically has a positive effect on sales in terms of the store's own attractiveness. Therefore, fundamental assumptions of location theory were regularly confirmed (Turhan et al. 2013, Wieland 2018).

3 Research approach and methodology

3.1 Identification of relevant explanatory variables

In order to build a meaningful model that explains *Valora*'s choice of location in the train stations, it is necessary to derive variables from the previous location literature that can be assumed to influence the decision for or against opening. The work from location theory and empirical retail research briefly summarized in Section 2 is very heterogeneous but essentially identifies three aspects of retail locations that influence the location choice of retail companies, namely 1) the market size, i.e., local or regional demand, 2) the competitive situation, and 3) possible positive agglomeration effects due to clustering of competitive or complementary stores. However, train stations are a special type of retail location where not all of the commonly identified location factors can be directly adopted. This is mainly because train stations are transport hubs in their main function, and the retail offering there is only an "additional" service. Therefore, the essentially known location factors have to be adapted to the train station situation.

A fundamental axiom of central place theory (Christaller 1933), as well as many subsequent location theories, is that a retail store requires minimal demand to be economically viable. This minimum demand is unknown; however, one can in any case assume that the impact of demand is positive. Both studies on location selection and store performance have regularly demonstrated empirically that local demand has a positive influence on both the opening and sales of a retail store (Larsson, Oener 2014, Seong et al. 2022, Wieland 2018). Therefore, it can be expected that the probability of *Avec* and *k kiosk* being present increases the greater the demand at the station. Since the main function of a train station is that of a transport hub for loading and unloading passengers, the demand at the station is primarily determined by the number of these passengers. This also fits with the statements of Nelson (1958), for whom train stations are external frequency generators that have a positive influence on the sales of stores (*suscipient business*). Store performance studies that investigate retailers in high-frequency locations have found that footfall is a positive driver of sales in different retail industries, including convenience stores (Müller-Hagedorn 1991, Seong et al. 2022, Weber 1979). *Valora* itself explicitly states that it is looking for new retail spaces "in a highly frequented location" (Valora Holding AG 2024b). Train passengers may also be considered as a proxy variable for the (unknown) footfall within the train stations. Therefore, the average daily number of passengers is used as the independent variable of demand volume.

The aspects of the competition and positive agglomeration effects cannot be clearly separated from one another, since a competitive store may certainly increase the competitive pressure or, on the contrary, can even increase frequency. Retail location theory describes both clustering and avoidance strategies of competing retailers when choosing a location as well as positive agglomeration effects due to clustering with competitive and/or complementary stores (see Section 2.1). Nelson (1958) in particular regards competitive stores partly as a source of frequency (positive agglomeration effects) and partly as damaging to business because they increase competitive pressure. Which of the two effects predominates depends on the retail industry under consideration and cannot be determined a priori. Empirical location studies also find effects of clustering with competitors or other providers, although the effect is very industry-specific (see Sections 2.2 and 2.3). For example, Seong et al. (2022) find with regard to convenience stores (which most closely corresponds to the case examined here) that their small-scale density has a positive effect on average sales, while their proximity to supermarkets has a negative effect. However, Krider, Putler (2013) find that food retailers tend to avoid competition. In train stations, the focus of the retail offering is usually on to-go food and groceries. Other retail chains that are often present at train stations include *Coop* and *Migros* with different convenience formats. There are also other kiosks, bakeries (including those of *Valora*), and fast food and takeaway restaurants. All of these types of offerings mentioned can in principle be considered as competitors for the two *Valora* formats examined, especially for *Avec*, which is a convenience supermarket, similar to *Pronto* (*Coop*) or *Migrolino* (*Migros*), for example. However, that doesn't mean that their presence will necessarily stop *Valora* from opening a store there. Instead, *Valora* may have a specific clustering and avoidance strategy, which is, however, unknown to the public. Thus, it is in no way clear a priori which of these competitors will have a positive or negative effect on *Valora* setting up at a train station. Therefore, the numbers of all mentioned competitors are taken into account as independent variables. It is expected that at least the presence of direct competitors – especially other convenience stores like *Pronto* or *Migrolino* – will reduce the likelihood of locating in a given micro-location.

Furthermore, *Avec* and *k kiosk* themselves also compete with each other to a certain extent. Since *k kiosk* stores tend to expand further and offer more food products (Valora Holding AG 2024c), it is to be expected that *Valora* will coordinate the location planning of these two formats in order to avoid self-cannibalization. It is therefore expected that the presence of *k kiosk* at a micro-location decreases the probability of opening *Avec*, and vice versa.

In addition, other characteristics of the micro-locations in the *SBB* train stations must be taken into account. For example, the passenger frequencies are only available at the level of the entire station (see Section 3.2). However, especially in large train stations with many micro-locations (e.g., waiting hall, platform area, secondary entrances), it cannot be assumed that the frequency is the same everywhere. Therefore, other available attributes of the respective micro-location (floor, type of micro-location) are included as explanatory variables. Furthermore, the number of ticket machines at the micro-location is taken into account as an independent variable, as it can be assumed that this represents an indication of the frequency at the micro-location.

3.2 Data collection and preprocessing

Two freely available data sets published by the *SBB* were used. The first data set contains all commercial space uses in Swiss train stations, including the name, the associated train station (marked with name and unique identification code `BPUIC`) and other information (SBB 2024d). For the stores, the dataset contains, among other things, their name or chain (`Name`), a categorization (`category`, e.g., `shopping`, `sbb_services`) and subcategorization (`subcategory`; the `shopping` category includes, for example, `food`, `bakery`, or `kiosk`), the micro-location within the train station (`location_details_en`, e.g., city level, hall XY, underpass XY, floor XY), the level of the train station in which the store is located (`Ebene`), and the opening hours (`openinghours`). Ticket machines also have their own entry in the classification. The data set has 5,254 entries (download from April 3, 2024).

The second data set contains the passenger frequencies of the SBB stations (SBB 2024a). There is data on those boarding and alighting at the train stations (train passengers only), namely the annual average daily traffic (`DTV_TJM_TGM`), the annual average weekday traffic (`DWV_TMJO_TFM`) and the average non-weekday traffic (`DNWV_TMJNO_TMGNL`). This data set contains the information mentioned for the years 2018 (1,160 stations), 2022 (1,161 stations), and 2023 (1,159 stations). The data also includes the station's unique identification code (`UIC`) and information about which data from which transport companies was included in the numbers. This data set is updated annually (download from June 14, 2024). In some cases, the counts do not include passenger numbers from specific railway companies, especially those from *RBS* (*Regionalverkehr Bern-Solothurn*).

The 1,159 train stations were divided based on their micro-locations recorded in the variable `location_details_en` in the *SBB* data set, which is the basis of the analysis (n=1,443). Most railway stations only consist of one micro-location, although the large stations (e.g., Zurich, Bern, and Basel) are divided into up to 30 micro-locations on several floors. The stores were summed up by their name and their subcategory at the level of train station micro-locations. At the time of data collection, 264 *Avec* or *k kiosk* shops were located at *SBB* train stations, of which 127 were *Avec* and 137 were *k kiosk*. At least one of both formats can be found at 249 micro-locations in 203 *SBB* railway stations. The highest density of *Valora* convenience stores is with 14 stores (1 x *Avec*, 13 x *k kiosk*) at Zürich main station (average daily passenger frequency 2023: 398,300) at ten different micro-locations.

3.3 Analytical model

In the first step, a binary logit model is used for the microeconometric analysis of the determinants of location choice. The dependent variable in each model is coded in binary for the respective chain examined. It is equal to one if at least one store of the respective chain (*Avec*, *k kiosk*, and both) is present at a micro-location, and zero otherwise. The following representation of the model is based on that in Cameron, Trivedi (2005) and Greene (2012). The target variable of a binary logit model is the probability that the examined condition is true or not, which is derived from the empirical distribution of positive (1) and negative (0) events, taking into account the explanatory variables. Here the target variable is the probability that the respective chain is present at the respective micro-location, which means that the number of stores of the chain c at the micro-location m in the train station s is greater than zero:

$$\Pr(Y_{cms} > 0|\mathbf{X}_s) = p_{cms} = \frac{\exp(\boldsymbol{\beta}\mathbf{X}_{ms})}{1 + \exp(\boldsymbol{\beta}\mathbf{X}_{ms})}$$

where Y_{cms} is the number of stores belonging to chain c at micro-location m in railway station s, p_{cms} is the probability that chain c is located at micro-location m in railway station s, $\mathbf{X}_{ms}$ is a set of explanatory variables (attributes of micro-location m and/or railway station s), and $\boldsymbol{\beta}$ is a set of corresponding regression coefficients.

Exponentiating both sides leads to the odds (ratio of the probability that the event occurs to the probability that the respective event does not occur): $p_{cms}/(1 - p_{cms}) = \exp(\boldsymbol{\beta}\mathbf{X}_{ms})$. The logit (log-odds) equals the linear combination of parameters and independent variables: $\ln p_{cms}(1 - p_{cms}) = \boldsymbol{\beta}\mathbf{X}_{ms}$. The coefficient of independent variable x_n, β_n, may also be interpreted as (semi-)elasticity with respect to the odds. The marginal effect (probability change due to a one-unit change in independent variable x_n) is the partial derivative with respect to x_n: $\partial p_{cms}/\partial x_n = p_{cms}(1 - p_{cms})\beta_n$.

Table 1 shows the independent variables of the model analysis. The passenger frequency was transformed with the natural logarithm in order to achieve an approximately normal distribution and to be able to interpret the associated model coefficients as elasticity. The levels of the station and the types of micro-locations have been converted into a simplified classification with three categories each. Competitors from the same company

Table 1: Independent variables in the model analysis

Variable name	Description
	Micro-location characteristics
DTV_TJM_TGM	Station frequency (average daily boarding and alighting 2023)
level_cat	Floor in the train station (categorized)
	Cat 1 First floor or above
	2 First basement floor or below
	0 Ground floor (city level)*
microlocation_cat	Type of micro-location (categorized)
	Cat 1 Underpass or pedestrian bridge
	2 Shopping street, gallery, passage, shopping center
	0 All others*
	Competitors and other suppliers
K_Kiosk_count	No. of *k kiosk* stores at the micro-location (*Avec* model)
Avec_count	No. of *Avec* and *Avec express* stores at the micro-location (*k kiosk* model)
Migros_all_count	No. of *Migros* and *Migrolino* stores at the micro-location (Eating and drinking formats from *Migros* such as *Migros Daily*, *Migros Eatery*, *Migros Restaurant*, and *Migros Take Away* are not included; these are included in the variable *catering_count*)
Coop_all_count	No. of *Coop*, *Coop to go*, and *Coop Pronto* stores at the micro-location (Other *Coop* food formats such as *Karma* and *Sapori d'Italia* are not included, but belong to *Food_other_count* or *catering_count*, depending on the classification)
Discounter_count	No. of *Lidl*, *Aldi*, and *Spar* stores at the micro-location
Kiosk_other_count	No. of other stores of type "kiosk" at the micro-location except *k kiosk*
Food_other_count	No. of other stores of subcategory "bakery", "beverages", "butcher", "food", and "supermarket" at the micro-location except those mentioned above
catering_count	No. of other stores of subcategory "bar", "cafe", "fast food", "restaurant", and "take away" at the micro-location
vend_machine_count	No. of vending machines at the micro-location

Note: * Reference category

(e.g., different *Migros* convenience formats) were grouped together so that the distribution of these independent variables is less skewed. The linear combination of the independent variables and their empirically determined coefficients is thus:

$$\begin{aligned} \mathbf{X}_{ms} &= \alpha + \beta \ln \mathrm{DTV\_TJM\_TGM} + \gamma_m \sum_{m=1}^{M} \mathrm{level\_cat}_{ms} + \\ &\quad \delta_n \sum_{n=1}^{N} \mathrm{microlocation\_cat}_{ms} + \lambda_c \sum_{c=1}^{C} \mathrm{Comp}_{ms} \end{aligned}$$

Binary logit models are estimated using the maximum likelihood method. The log-likelihood in this case is:

$$LL = \sum_{i=1}^{n} y_i \ln f(\boldsymbol{\beta}\mathbf{X}_{ms}) + (1 - y_i) \ln(1 - f(\boldsymbol{\beta}\mathbf{X}_{ms})) \tag{1}$$

where y_i is the i-th observation and n is the number of observations.

The iteratively reweighted least squares (IWLS) algorithm is used for the estimation. The significance level was set to 90%. The analysis was conducted in *R* version 4.4.0 (R Core Team 2024), including the help of the *stargazer* package (Hlavac 2022).

3.4 Optimization of predictive ability

The second step, after the microeconometric analysis, is about optimizing the predictive ability of the model using machine learning techniques. Machine learning (ML) is a subset of artificial intelligence (AI) and enables systems to learn from data and improve over time without being explicitly programmed for each specific task. From an ML perspective, this is a (binary) classification problem (Boehmke, Greenwell 2020, Kuhn 2008). Such questions arise in very different disciplines, for example, in banking with regard to the probability of loan default or in the medical context when predicting possible complications after treatments or assessing the risk of death. In these cases, different ML modeling approaches are used and compared with each other in terms of the accuracy of their predictions (Celio Di Cellio Dias et al. 2018, Omar et al. 2024, Shahidi et al. 2023). Here, five ML algorithms are implemented. Four of them are ensemble methods, which means that they combine a given number of (weak) learners into one aggregated learner with high accuracy, sometimes referred to as the "wisdom of the crowd" effect (Boehmke, Greenwell 2020):

1. *Decision tree* (DT): The tree consists of nodes (decisions, more precisely divisions of the independent variables) and leaves (predictions). Each node divides the data based on the input features, creating a hierarchical structure. During the training process, the tree is built by dividing the data into different groups (for categorical independent variables) or intervals (for continuous independent variables) based on the input features. Unlike the binary logit model, modeling the relationship between a dependent variable and the independent variables using one (or more) decision tree(s) is a non-parametric algorithm. The division of data in the classification tree is done using Gini impurity, which is an indicator of how mixed a node is in terms of categories:
$$Gini = 1 - \sum_{c=1}^{C} (p_c)^2$$
where C is the number of classes and p_c is the proportion of class c.

   The algorithm looks for the split that leads to nodes that are mixed as little as possible; that is, the Gini impurity is minimized. A decision tree may, but not necessarily, contain all explanatory variables.

2. *Decision trees with bagging* (DTBG): The second model is an ensemble method that combines decision trees and bagging (bootstrap aggregating). In this algorithm, t decision trees are formed, always with a different bootstrap sample from the data. These individual models are combined into one prediction by averaging the estimated class probabilities together. The minimum number of observations that must be present in a node for this node to be further split was set to $Split_{min} = 2$.

3. *Random forest* (RF): A random forest algorithm is an extension of bagged decision trees. A further random component is implemented here, namely only a randomly selected subset of the explanatory variables, m_{try}, is implemented for each split (here $m_{try} = \sqrt{p}$, with p being the number of explanatory variables). Bagging and random forest algorithms were tested with $t = 20$, 50, and 100 trees.

4. *Gradient boosted logit* and *gradient boosted trees*: Gradient boosting (GB) can be used for various basic models and is also an ensemble method. Here, new learners are added sequentially to the ensemble; namely, in each step i a new learner is added who is specifically fit to address the errors (residuals) of the previous one as well as possible. More precisely, this means that the algorithm iteratively adjusts the predictions to minimize a specific loss function. Here, the logistic loss function (Log loss) is used, which is a normalization of the log likelihood (Equation 1) and the default metric for binary classification problems in the used estimation package:
$$Logloss = -\frac{1}{n} LL \tag{2}$$
where n is the number of observations.

   The result is an aggregate of the learners created over I iterations. The fit to the overall data set usually improves with each iteration, but this does not necessarily apply to out-of-sample fitting (overfitting). The algorithm is used with both binary logit models (BLGB) and decision trees (DTGB). Both gradient boosting algorithms were tested with $I = 100$ and 200 iterations.

5. *Artificial neural network* (ANN): Artificial neural networks imitate the structure of biological neurons in brains. They consist of input, output, and at least one hidden layer, where each node (neuron) computes a weighted sum of its inputs (original input features), applies an activation function, and passes the result forward. During training, backpropagation and gradient descent minimize the loss function by updating the weights. In the present binary case, the loss function is also log loss (see Equation 2). The ANN algorithm is applied using default values (5 neurons, decay parameter equal to 0.1 for regularization) and is tested with 100 and 200 iterations.

In line with other ML classification problems, the models are assessed by a confusion matrix, which includes two performance indicators that have been used for decades to check the quality of diagnostic tests in medicine, *specificity* and *sensitivity*. Sensitivity is the share of true positives that are correctly predicted by the models. Specificity is the share of true negatives that are correctly predicted by the models (Altman, Bland 1994, Boehmke, Greenwell 2020, Trevethan 2017):

$$Sensitivity = \frac{TP}{TP + FN} = \frac{TP}{P}$$

$$Specificity = \frac{TN}{TN + FP} = \frac{TN}{N}$$

where TP is the number of true positives (number of micro-locations where $Y_{cms} > 0$ and $p_{cms} > 0.5$), TN is the number of true negatives (number of micro-locations where $Y_{cms} = 0$ and $p_{cms} < 0.5$), FN is the number of false negatives (number of micro-locations where $Y_{cms} > 0$ but $p_{cms} < 0.5$), FP is the number of false positives (number of micro-locations where $Y_{cms} = 0$ but $p_{cms} > 0.5$), and P and N are the numbers of total positives and negatives, respectively.

Sensitivity therefore makes a statement about how well the individual model predicts the cases in which a store is actually located in a micro-location. Specificity, on the other hand, documents how well the individual model predicts the cases in which there is no store. The metrics are calculated for the models based on the training data for the test data set (out-of-sample). Additionally, the $ROC - AUC$ (*Receiver Operating Characteristic - Area Under the Curve*) metric is calculated, which represents a trade-off between sensitivity and specificity. The ROC curve plots sensitivity against $1 - specificity$ at various thresholds, and the AUC measures the overall ability of the model to distinguish between classes. A higher AUC value indicates better model performance, with values closer to 1 signifying high sensitivity and specificity, while values closer to 0.5 suggest a random classification. Since in the dataset the case of an outcome of 1 is less likely than an outcome of 0 (the dependent variable is skewed), sensitivity is used as a metric for model selection. The division into training and test data sets is 90 to 10%. Model validation is undergone with 10-fold repeated cross validation with five repeats (Boehmke, Greenwell 2020). The analysis was conducted in *R* version 4.4.0 (R Core Team 2024) using the package *caret* (Kuhn 2008) and related packages such as *randomForest* (Liaw, Wiener 2002) and *nnet* (Venables, Ripley 2002), as well as own functions.

4 Results

4.1 Analytical model: Determinants of location choice

Table 2 shows the results of three binary logit models, with the presence of *Avec*, *k kiosk* or at least one of the two acting as the dependent variable (1=yes, 0=no).

The daily passenger frequency has a significant and positive influence on the probability of choosing a location at the respective micro-location, which is similar in all three models. A 1% increase in passenger frequency increases the odds of opening a *Valora* convenience store by approximately 0.4% (*Avec*: 0.364, *k kiosk*: 0.406, both: 0.434). This is not

Table 2: Binary logit model results

	Dependent variable:		
	Avec	K_Kiosk	Valora
	(1)	(2)	(3)
log_DTV_TJM_TGM	0.364***	0.406***	0.434***
	(0.051)	(0.101)	(0.054)
level_cat1	−2.667***	−0.215	−1.648***
	(0.926)	(0.689)	(0.520)
level_cat2	−16.238	−1.436	−2.289***
	(527.091)	(0.944)	(0.686)
microlocation_cat1	0.855	0.036	−0.304
	(1.114)	(0.773)	(0.668)
microlocation_cat2	0.444	2.162**	1.251
	(1.397)	(1.094)	(0.895)
K_Kiosk_count	−1.576***		
	(0.527)		
Avec_count		−1.469**	
		(0.612)	
Migros_all_count	−1.683	1.214**	0.099
	(1.047)	(0.576)	(0.453)
Coop_all_count	−0.924	3.991***	2.264***
	(0.652)	(0.627)	(0.547)
Discounter_count	−16.097	1.061	0.774
	(1,909.183)	(5.341)	(3.634)
Kiosk_other_count	−15.403	1.194	−1.053
	(919.552)	(1.153)	(1.073)
Food_other_count	0.054	−4.434***	−2.260***
	(0.320)	(0.400)	(0.273)
catering_count	0.326**	2.321***	1.373***
	(0.147)	(0.250)	(0.171)
vend_machine_count	0.842***	0.954***	1.308***
	(0.166)	(0.243)	(0.202)
Constant	−5.583***	−8.443***	−6.462***
	(0.452)	(0.958)	(0.512)
Observations	1,443	1,443	1,443
Log Likelihood	−368.285	−130.503	−404.544
Akaike Inf. Crit.	764.571	289.007	835.088

Note: *p<0.1; **p<0.05; ***p<0.01

surprising, as any store requires a minimum level of demand, and in this case this is represented by the passenger frequencies. Empirical location research with regard to stores at frequent locations has shown that footfall is a determinant of store sales (Müller-Hagedorn 1991, Weber 1979, Seong et al. 2022). However, it must be taken into account that studies that include pedestrian frequencies in their models may have an endogeneity problem. It cannot be clearly clarified which part of the frequency explains the sales or the number of customers of the stores examined and which part of it is caused *by* these stores. The direction of the causal relationship cannot therefore be fully explained (*chicken and egg problem*). However, this problem does not exist in the current case of train stations because we consider the train passengers and not the footfall (which is not available). It is plausible to assume that those boarding, leaving, or changing trains are moving from an origin to a destination (e.g., work) while using a train and typically do not take the train to a train station just to buy groceries there. That's why train stations act as an external frequency generator, which is, so to speak, a "model example" of *suscipient business* in the sense of Nelson (1958).

A minimum demand in the sense of a minimum necessary passenger frequency can also be implicitly derived from the empirical data. The train station with the lowest frequency, where *Avec* is located, has a daily passenger volume of 540 people (Murgenthal), closely followed by Flums (560 people) and Aarberg (580 people). In all three cases mentioned there is no competition (*Coop*, *Migros*, or other food or kiosk) located at the station. The smallest train station with a *k kiosk* store has 640 passengers daily and also no competition (Saanen).

The type of micro-location within the train station also has an influence on the probability of location choice, although not every characteristic is statistically significant. Both *Avec* and *k kiosk* tend to be located less often on the upper or lower floors (level category 1 or 2) of train stations. The *k kiosk* model also shows that this format is found significantly more frequently in micro-location category 2 (shopping streets, etc.). If the micro-location is a shopping street or something similar, this increases the odds of a *k kiosk* being present by a factor of exp(2.162), which equals approximately 8.688. These results can be explained by the fact that footfall is, of course, not evenly distributed within a railway station. The passenger frequencies are only collected for the entire station. It is very likely that there will be a higher frequency on the ground floor of a station than on the upper or lower floors. The same applies to shopping streets, etc., within the train stations, where many other shops are located, which in turn brings frequency.

For every *k kiosk* store that is present at the micro-location, the chance that an *Avec* store will also be located there reduces by the factor exp(-1.576), which equals approximately 0.207, all other things being equal. The reverse effect is very similar. Each *Avec* store reduces the chance that a *k kiosk* will be opened at a micro-location by a factor of exp(-1.469) $\approx$ 0.230. Since these two formats do not occupy the same sales area sizes, it is very likely that an avoidance strategy is being deliberately pursued in order to prevent internal competition (self-cannibalization). This is also plausible because *k kiosk* stores have expanded their assortment in recent years (Valora Holding AG 2024c), and it is to be expected that the markets that serve these two formats will overlap to a considerable extent. However, there are no explicit statements from the *Valora Group* that the two formats are deliberately localized according to the principle of avoidance. That this is the case remains a reasonable assumption but cannot be directly proven.

When it comes to clustering and avoidance strategies with regard to other competitors at train stations, there are apparently differences between the two formats examined. The coefficients of the variables for the presence of *Coop* and *Migros* stores as well as discounters and kiosks are all negative in the case of *Avec*, but not statistically significant. This result may also be due to the fact that *Avec* tends to compete with the above-mentioned competitors for sales space of similar size and that the *SBB* only awards the contract to one of the two if only one store is available. The format *k kiosk*, on the other hand, occupies much smaller selling spaces (see Section 1). In the latter case, the presence of *Coop* or *Migros* seems to increase the probability that *k kiosk* is located at a micro-location. Avoidance patterns of food competitors, which have been shown by, e.g., Krider, Putler (2013), cannot be confirmed in this specific case. However, the result of Seong et al. (2022), who found positive sales effects of clustering food convenience stores, may be confirmed here. Convenience supermarkets from competing companies may also act as external frequency generators. Gastronomic providers and *SBB* vending machines increase this probability in both cases. In the latter case, this is probably because the presence of ticket machines is a proxy variable for the frequency in the respective part of the railway station (see Section 3.1).

4.2 Evaluation of external model validity

Table 3 presents the three metrics sensitivity, specificity, and ROC-AUC for the three models (out-of-sample with a test dataset of 10% of all cases). The analytical model (binary logit) acts as a baseline against which the machine learning models are assessed. First of all, it can be seen in general that, as expected, the ML algorithms perform significantly better than the binary logit model. These results are consistent with those of other ML applications for (binary) classification problems (Celio Di Cellio Dias et al. 2018, Omar et al. 2024, Shahidi et al. 2023). This is not surprising since a binary logit model is not estimated with the aim of optimal predictive ability, while ML algorithms are designed for exactly that. This applies in particular to sensitivity, i.e., the correct prediction of the true positives, in this case the micro-locations where a *Valora* convenience format is actually located. In the *Avec* case in particular, a significant improvement in this metric can be seen, as here the binary logit model correctly predicts only 1.6% of the positive cases, which makes this analytical model completely unsuitable for a forecast.

The specificity is close to 100% for all models, meaning that all models almost perfectly predict the micro-locations where no *Valora* convenience format is present. However, this can be explained by the distribution of positive and negative events in the data. An *Avec* is located in only 8.8% of the micro-locations and a *k kiosk* in 9.5%, while one of the two occurs in 17.3% of all cases. Thus, at most of the 1,443 micro-locations there is neither an *Avec* nor a *k kiosk* store, i.e., the expression "no" or 0 is the most common case. A model that always predicts "no" would therefore in 91.2% or 90.5% or 82.7% of the cases make a correct prediction. This is the so-called *no information rate* that should be taken into account when assessing the accuracy of binary outcome models (Kuhn 2008). In this case, specificity cannot be used meaningfully for a comparison of predictive ability. Thus, the best model for each case is chosen with respect to sensitivity (marked bold in Table 3).

In the *Avec* case, a bagged model with 20 decision trees leads to the highest sensitivity value, with 31.6% of the true positives being predicted correctly. However, even the hit rate of less than a third can still be described as rather weak, so for practical purposes it probably wouldn't make sense to use this model for forecasting. The random forest models with 50 or 100 trees achieve almost the same sensitivity (30.5 and 31.4%). In contrast, gradient boosted trees have slightly higher specificity but lower sensitivity. Both artificial neural networks produce lower sensitivity values compared to the tree-based models. However, the trade-off between sensitivity and specificity in terms of ROC-AUC is the highest for the ANN results. When predicting the *k kiosk* locations, the highest sensitivity of 84.7% is achieved with the decision tree bagging models with 50 or 100 trees. In this case, the model with the smaller number of trees given the same performance is considered the best model. The sensitivity of gradient boosting and ANN models is slightly lower. In the case of predicting one of the two *Valora* chains, the best performance in terms of sensitivity is achieved by the gradient boosted logit model approach with 100 iterations (67.7%). However, the specificity is the lowest of all model variants. In the last two cases, the ANN models also provide a (sometimes much) better balance between sensitivity and specificity, but not the highest sensitivity, which serves as a selection criterion here.

It turns out that, depending on the respective case, different ML model approaches lead to the best result and that a higher number of learners does not necessarily lead to a higher predictive ability. This is usually because the trained model reflects a lot of the variance in the training data, which at the same time reduces the external validity (*overfitting*) (Boehmke, Greenwell 2020). When searching for the best model, it is therefore necessary to test different algorithms with different configurations, using a performance metric that is suitable for the case at hand (in this application this is sensitivity; see above). Regarding sensitivity and specificity, it should also be said that, normally, a trade-off between the two must be made because, in any test (or model), an optimization of one indicator leads to a deterioration of the other indicator (Trevethan 2017). In the present case, this effect hardly occurs because the specificity is automatically very high, as a negative result is much more common in the empirical data. If sensitivity would not have been the decisive factor in this case, an ANN approach would have been the best model in all three cases.

4.3 Model simulation: Which station commercial spaces are suitable for k kiosks*?*

The model analysis is now used for a practical case: The *SBB* permanently advertises commercial space in train stations (and on other properties that belong to the *SBB*) to the public for rent. In some cases, space uses are already determined in advance (e.g., retail, catering). Based on currently advertised commercial spaces, it is now being examined how high the probability is that a given store will be located there. The *k kiosk* location choice prediction model (DTBG with 20 trees) was very good in terms of both specificity and sensitivity (see Section 4.2), which is why this model is used as an example. All retail spaces advertised for rent were researched from the SBB website (SBB 2024b) for which a different shop concept (e.g., catering) was not already expressly specified and which are located within train stations. Of 32 offers on the day of access (accessed on September 17,

Table 3: Performance metrics for the machine learning models

Chain	Model	Out-of-sample performance		
		Sensitivity	Specificity	ROC-AUC
Avec	BL	0.016	0.986	0.807
	DT	0.308	0.982	0.815
	DTBG*	**0.316**	**0.968**	**0.789**
	DTBG**	0.302	0.970	0.801
	DTBG***	0.294	0.971	0.805
	RF*	0.305	0.979	0.781
	RF**	0.314	0.979	0.803
	RF***	0.314	0.978	0.811
	BLGB#	0.144	0.991	0.828
	BLGB##	0.164	0.990	0.828
	DTGB#	0.090	0.998	0.876
	DTGB##	0.216	0.986	0.878
	ANN#	0.170	0.990	0.891
	ANN##	0.172	0.989	0.892
k kiosk	BL	0.775	0.987	0.962
	DT	0.772	1.000	0.890
	DTBG*	0.844	0.987	0.961
	DTBG**	**0.847**	**0.988**	**0.966**
	DTBG***	0.847	0.988	0.968
	RF*	0.781	0.995	0.970
	RF**	0.807	0.996	0.977
	RF***	0.804	0.996	0.978
	BLGB#	0.836	0.993	0.979
	BLGB##	0.824	0.992	0.977
	DTGB#	0.820	0.995	0.979
	DTGB##	0.824	0.994	0.982
	ANN#	0.821	0.990	0.983
	ANN##	0.815	0.991	0.986
Valora	BL	0.476	0.973	0.885
	DT	0.389	1.000	0.746
	DTBG*	0.627	0.951	0.878
	DTBG**	0.627	0.952	0.891
	DTBG***	0.627	0.953	0.894
	RF*	0.610	0.966	0.887
	RF**	0.622	0.966	0.897
	RF***	0.619	0.967	0.900
	BLGB#	**0.677**	**0.936**	**0.907**
	BLGB##	0.590	0.987	0.904
	DTGB#	0.525	0.985	0.932
	DTGB##	0.571	0.978	0.934
	ANN#	0.550	0.976	0.936
	ANN##	0.552	0.977	0.938

Notes: Models: BL = Binary Logit, DT = Decision Tree, BG = Decision Trees with Bagging, RF = Random Forest, BLGB = Binary Logit with Gradient Boosting, DTGB = Decision Trees with Gradient Boosting, ANN = Artifical Neural Network.
Flags: *, **, *** = 20, 50, or 100 trees; / #, ## = 100 or 200 iterations. / The best model is marked in **bold.**

2024), this applied to nine space offers. These are space offers at the following Swiss train stations (in alphabetical order): Altdorf UR, Bex, Chiasso, Genève-Eaux-Vives, Glovelier, Hedingen, Hunzenschwil, Münsingen, and St. Gallen Winkeln.

In most cases there is only one micro-location in the train station (city level), with Chiasso and Genève-Eaux-Vives being exceptions. The frequency numbers (*DTV_TJM_TGM*) of the train stations in 2023 were between 680 (Hunzenschwil) and 8,600 (Chiasso). In the case of Genève-Eaux-Vives, a commercial space was put out to tender to replace an existing food provider (*Tekoe*), which was communicated in the tender documents. This results in a change in the independent variables, namely that the value of the variable *Food_other_count* drops from 3 to 2. In the remaining cases, there is no obvious change in the business structure. In three cases, the other *Valora* format examined (*Avec*) is already located in the respective micro-location. This is important because the econometric model analysis has shown that an avoidance strategy appears to apply to these two *Valora* formats (see Section 4.1). Table 4 shows a summary of the independent variables for the train station or micro-location as well as the result of the prediction model (DTBG).

Table 4: Results of the model prediction for the opening of a *k kiosk* store

Train station		*Micro-location*			*k kiosk*
Station	DTV_TJM_TGM	level_cat	microlocation_cat	Competitors*	Prediction
Altdorf UR	2,200	0	0	2	**NO**
Bex	2,400	0	0	3**	**NO**
Chiasso	8,600	0	0	1**	**NO**
Genève-Eaux-Vives	8,400	0	2	8	**YES**
Glovelier	1,000	0	0	1	**NO**
Hedingen	2,300	0	0	2	**NO**
Hunzenschwil	680	0	0	1	**NO**
Münsingen	6,400	0	0	1**	**NO**
St Gallen Winkeln	1,500	0	0	1	**NO**

Notes: *Sum of *Avec_count*, *Migros_all_count*, *Coop_all_count*, *Disocunter_count*, *Kiosk_other_count*, *Food_other_count*, *catering_count*, and *vend_machine_count*
**Including one *Avec* or *Avec express* store ($Avec\_count > 0$)

It turns out that in the nine micro-locations, a positive location decision is only predicted in one case, namely Genève-Eaux-Vives. It is unlikely that *k kiosk* stores will be opened in the remaining micro-locations, which is certainly not primarily due to insufficient demand, as the passenger frequencies in all train stations reach an acceptable level (see Section 4.1). Instead, the characteristics of the micro-locations reduce the probability of a positive result: It was already determined in the econometric analysis that *k kiosk* stores are preferred to be located in category 2 micro-locations (shopping streets, etc.), which is only the case with the commercial space on offer in Genève-Eaux-Vives. Three micro-locations are already occupied by *Avec*. The results of the model analysis suggest that there is an avoidance strategy between the two *Valora* formats *Avec* and *k kiosk*, which is why it is not surprising that *k kiosk* is unlikely to open in these locations. The fact that a food space is being abandoned in Genève-Eaux-Vives also increases the likelihood of a *k kiosk* opening. However, there is an important limitation in the results of the model forecast: In contrast to the other eight train stations, this micro-location already has a *k kiosk* store, which is not explicitly covered by the model. Although it is in principle conceivable that more than one *k kiosk* will be located at a micro-location (such as in Basel SBB, Chur, or Olten), it cannot be predicted on this basis.

5 Conclusions and limitations

The study on location choice of *Valora* convenience formats in Swiss train stations has an analytical and a predictive part. To answer the first research question, binary logit models were built for analytical purposes. The presence of *Avec* and *k kiosk* stores at the level of micro-locations within the railway stations was examined against the background of location-specific independent variables, which were derived from location theory and previous empirical work with respect to other location types. Local demand was measured by passenger frequency. The more people boarding, alighting, and transferring at a station, the more likely it is that an *Avec* or *k kiosk* store is located there. There is also a mutual avoidance strategy for both *Valora* formats. Competitors' location decisions influence the likelihood of an opening, but not to the same extent in both formats. With respect to *k kiosk*, there is no evidence of any avoidance of competition with respect to convenience supermarkets. Rather, the presence of *Coop* or *Migros*, all other things being equal, increases the probability of the presence of *k kiosk*. In the case of *Avec*, this effect is diffuse because many model parameters are not statistically significant. The most important determinants of location choice are, thus, demand and chain-specific clustering and avoidance patterns.

The positive impact of station frequency on the probability of an opening, clearly identified in all cases examined, is congruent with the statements of location theories and empirical work on both location choice and store performance, according to which market size or demand is always identified as a positive location factor. With regard to the interplay between competition and agglomeration effects, which is discussed in many theoretical and empirical contributions to retail locations, clear statements cannot be made in any case. Many approaches predict that very similar providers engage in competitor avoidance, e.g., Christaller (1933). This is obviously not the case with *k kiosk*. On the contrary, this format tends to be located where (larger) food competitors such as *Coop* or *Migros* are already located. Here, *k kiosk* may benefit from *shared business* in the sense of the *theory of cumulative attraction* (Nelson 1958) or with respect to the clustering of competitors that are imperfect substitutes (Chamberlin 1933). However, interformal competition avoidance with each other is likely for both formats, although this cannot be directly explained by location theories, as the two formats, while offering overlaps, cannot be considered completely substitutable. Rather, it is likely that a company-specific avoidance strategy is the cause.

The second research question concerned the degree to which these location decisions can be predicted in new cases. For this purpose, based on the model mentioned above, various machine learning algorithms were used to optimize the prediction ability, and the models were checked with regard to their out-of-sample accuracy. This showed that AI/ML models make a huge contribution to significantly improving the predictive ability of these models. In one case (*k kiosk*) a model was built that showed very good results in terms of both sensitivity and specificity. In the other case (*Avec*), however, it must be admitted that even the best model solution is not so good that it would be suitable for practical purposes. The suitable model was used for a forecast with real data. However, it should be noted that the selection of the best model must also follow logical considerations related to the study case. In the present case, for example, it was argued that sensitivity, i.e., predicting positive location decisions, is more important than predicting negative values. In other cases, specificity or a trade-off between the two metrics may be the decisive factor. Furthermore, it has been shown that model quality does not necessarily increase with model complexity (e.g., number of estimators) and that it is always useful to test a number of tuning parameters.

The study also faces some theoretical and methodological limitations. *Firstly*, the entire analytical and predictive model approach is based on the premise that *Valora*'s location decision is based on the evaluation of location characteristics (especially demand) and the behavior of competitors. However, it could not be taken into account whether free retail space is available at all, as there is no data on the total amount of retail space in the *SBB* train stations (i.e., including possible vacancies). It cannot therefore be clarified whether the non-presence of the examined formats at certain train stations is possibly due to the fact that opening there is not possible because of a lack of retail space. Since *Avec* is implemented on much larger sales areas than *k kiosk*, it is plausible to assume that this problem is much greater in the *Avec* model. This could in turn explain why this model performs significantly worse in terms of predictive performance than the *k kiosk* model.

Secondly, for the quantification of the demand volume, only train passenger frequencies were available, but not the actual footfall at the station (although this could potentially lead to an endogeneity problem; see Section 4.1). This is likely to underestimate the actual demand, especially in large train stations with integrated shopping streets. Furthermore, the passenger frequencies are naturally only available for the entire station, although the same frequency does not prevail in every part of the station. Passenger frequencies, therefore, do not provide a complete picture of local demand. Other variables in the model, such as the number of ticket machines, most likely partially compensate for this. Both of these limitations could be addressed in future studies. However, this would require data that is not currently (publicly) available, namely all retail spaces within stations (not just occupied ones) as well as small-scale pedestrian traffic.

Thirdly, as already mentioned, the predictive ability of the *Avec* predictive model is rather weak. At the same time, certain difficulties also appear in the analytical model in the form of some high coefficients with very high standard errors. Both problems can arise from an unfavorable combination of the explanatory variables, e.g., in terms of underspecification and/or multicollinearity. It is likely that other variables are missing here, apart from the deficit mentioned in the first point, which affects all models. For example, it could make sense to differentiate between different *Avec* subformats, e.g., *Avec* and *Avec express*, or to distinguish which *Avec* stores are open 24/7 (without service after regular opening hours) and which are not. This distinction was not made in the current study, as all *Avec* stores were treated equally. In this study, the same models were built for both convenience formats in order to be able to compare the results. However, it becomes apparent that the explanatory variables for *k kiosk* are very good but are obviously not sufficient for *Avec* and/or would have to be arranged differently in order to obtain a better model result. In principle, it's also conceivable that *Avec*'s expansion is simply less structured than *k kiosk*'s. To achieve better predictive ability for *Avec*, additional variables should be considered and/or the influence of competitors should be further differentiated, unless, as in this case, the comparison between multiple chains is the primary focus.

Fourthly, in the models, the dependent variable was coded as binary (*Valora* chain is present or not), which was calculated from the sums of the respective chains at the micro-locations. No distinction is made here as to whether one or more stores are located at the same micro-location. It is very unlikely, however, that this induces a substantial bias. In the case of *Avec*, there is no instance where more than one store is located at a micro-location. For *k kiosk*, there are only 10 micro-locations where two *k kiosk* stores are located. However, in future studies, count data models (Larsson, Oener 2014, Wieland 2017) could be used instead of binary outcome models.

References

Alonso W (1964) *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press, Cambridge, MA. CrossRef

Altman DG, Bland JM (1994) Diagnostic tests. 1: Sensitivity and specificity. *British Medical Journal* 308: 1552. CrossRef

Aversa J, Doherty S, Hernandez T (2018) Big data analytics: The new boundaries of retail location decision making. *Papers in Applied Geography* 4: 390–408. CrossRef

Bills SJ (1998) New geographies of retailing: An investigation of developments at airports, railway stations, hospitals and service stations. PhD thesis, https://www.valora.com-/en/brands/kkiosk/

Blick (2025) Deutscher Discounter Aldi verdrängt Coop im Bahnhof Basel – Kampf um Flächen geht los. Press article, https://www.blick.ch/wirtschaft/sbb-mischen-shopping-karten-neu-deutscher-discounter-aldi-verdraengt-coop-im-bahnhof-basel-kampf-um-flaechen-geht-los-id20496591.html

Boehmke B, Greenwell B (2020) *Hands-On Machine Learning with R* (1 ed.). Taylor & Francis, New York, NY. CrossRef

Cameron AC, Trivedi PK (2005) *Microeconometrics. Methods and Applications*. Cambridge University Press, Cambridge. CrossRef

Celio Di Cellio Dias P, Forti M, Witarsa M (2018) A comparison of Gradient Boosting with Logistic Regression in Practical Cases. Working paper, https://support.sas.com/resources/papers/proceedings18/1857-2018.pdf

Chamberlin EH (1933) *The Theory of Monopolistic Competition*. Harvard University Press, Cambridge, MA

Chang HJ, Hsieh CM (2018) A new model for selecting sites for chain stores in China. *International Journal of Industrial and Systems Engineering* 28: 346–359. CrossRef

Christaller W (1933) *Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Gustav Fischer, Jena

Converse PD (1949) New laws of retail gravitation. *Journal of Marketing* 14: 379–384. CrossRef

DB Station&Service AG (2017) Germany's stations: Top locations for gastronomy and retail. Brochure, https://www.deutschebahn.com/resource/blob/284664/fa59e6114-fa1f1147eb699b9fe494c1f/vermietungsbroschuere˙bahnhoefe-data.pdf

Eaton BC, Lipsey RG (1982) An economic theory of central places. *The Economic Journal* 92: 56–72. CrossRef

EHI Retail Institute (2023) *Travel Retail 2023*. EHI Retail Institute

Eurostat (2024) Rail passenger transport reaches new peak in 2023. Press release, https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20241030-1

Fotheringham AS (1985) Spatial competition and agglomeration in urban modelling. *Environment and Planning A: Economy and Space* 17: 213–230. CrossRef

Fujita M, , Thisse JF (2002) *Economics of Agglomeration. Cities, Industrial Location, and Regional Growth*. Cambridge University Press, Cambridge. CrossRef

Ge D, Hu L, Jiang B, Su G, Wu X (2019) Intelligent site selection for bricks-and-mortar stores. *Modern Supply Chain Research and Applications* 1: 88–102. CrossRef

Ghosh A (1986) The value of a mall and other insights from a revised central place model. *Journal of Retailing* 62: 79–97

Greene WJ (2012) *Econometric Analysis* (7 ed.). Pearson

Hlavac M (2022) stargazer: Well-formatted regression and summary statistics tables. R package version 5.2.3. Software, https://CRAN.R-project.org/package=stargazer, Bratislava, Slovakia

Hotelling H (1929) Stability in competition. *The Economic Journal* 39: 41–57. CrossRef

Huff DL (1962) *Determination of Intra-Urban Retail Trade Areas*. University of California

Joseph L, Kuby M (2013) Regionalism in US retailing. *Applied Geography* 37: 150–159. CrossRef

Joseph L, Kuby M (2015) Modeling retail chain expansion and maturity through wave analysis: Theory and application to Walmart and Target. *International Journal of Applied Geospatial Research* 6: 1–26. CrossRef

Joseph L, Kuby M (2016) The location types of US retailers. *International Journal of Applied Geospatial Research* 7: 1–22. CrossRef

Krider R, Putler D (2013, 04) Which birds of a feather flock together? Clustering and avoidance patterns of similar retail outlets. *Geographical Analysis* 45: 123–149. CrossRef

Kuhn M (2008) Building predictive models in R using the caret package. *Journal of Statistical Software* 28: 1–26. CrossRef

Larsson JP, Oener O (2014) Location and co-location in retail: A probabilistic approach using geo-coded data for metropolitan retail markets. *The Annals of Regional Science* 52: 385–408. CrossRef

Liaw A, Wiener M (2002) Classification and regression by randomforest. *R News* 2: 18–22

Lu J, Zheng X, Nervino E, Li Y, Xu Z, Xu Y (2024) Retail store location screening: A machine learning-based approach. *Journal of Retailing and Consumer Services* 77: 103620. CrossRef

Müller-Hagedorn L (1991) Moderne Verfahren zur Ermittlung der Bedeutung einzelner Standortfaktoren. In: DHI (ed), *Standortpolitik des Einzelhandels*. Köln, 100–105

Nelson P (1970) Information and consumer behavior. *Journal of Political Economy* 78: 311–329. CrossRef

Nelson RL (1958) *The Selection of Retail Locations*. F.W. Dodge, West Palm Beach

Neue Zürcher Zeitung (2024) Discounter unerwünscht? Die SBB lassen Aldi und Lidl abblitzen. Press article, https://www.nzz.ch/wirtschaft/discounter-unerwuenscht-lidl-und-aldi-bewerben-sich-bei-den-sbb-um-ladenflaechen-aber-blitzen-ab-ld.1803620

Nilsson IM, Smirnov OA (2016) Measuring the effect of transportation infrastructure on retail firm co-location patterns. *Journal of Transport Geography* 51: 110–118. CrossRef

OEBB (2024) Immobilien-Angebote. Website, https://immobilien.oebb.at/de/angebote

Office of Rail and Road (2024) Railway station catering market. Final Report. https://www.orr.gov.uk/sites/default/files/2024-06/railway-station-catering-market-study-final-report-june-2024_0.pdf

Omar ED, Mat H, Zafirah Abd Karim A, Sanaudi R, Ibrahim FH, Azahadi Omar M, Zulfadli Hafiz Ismail M, Jayaraj VJ, Leong Goh B (2024) Comparative analysis of logistic regression, gradient boosted trees, svm, and random forest algorithms for prediction of acute kidney injury requiring dialysis after cardiac surgery. *International Journal of Nephrology and Renovascular Disease* 17: 197–204. CrossRef

Orhun AY (2013) Spatial differentiation in the supermarket industry: The role of common information. *Quantitative Marketing and Economics* 11: 3–37. CrossRef

R Core Team (2024) R: A Language and Environment for Statistical Computing. Software, https://www.R-project.org/, Vienna, Austria

Radio Frequence Jura (2024) La gare de Delémont se sépare de son Coop Pronto. Press article, https://www.rfj.ch/rfj/Actualite/Region/20240127-La-gare-de-Delemont-se-separe-de-son-Coop-Pronto.html

Rao F, Pafka E (2021) Shopping morphologies of urban transit station areas: A comparative study of central city station catchments in Toronto, San Francisco, and Melbourne. *Journal of Transport Geography* 96: 103156. CrossRef

Reigadinha T, Godinho P, Dias J (2017) Portuguese food retailers – Exploring three classic theories of retail location. *Journal of Retailing and Consumer Services* 34: 102–116. CrossRef

Reilly WJ (1931) *The Law of Retail Gravitation*. Knickerbocker Press

Reynolds J, Wood S (2010) Location decision making in retail firms: Evolution and challenge. *International Journal of Retail & Distribution Management* 38: 828–845. CrossRef

Rice MD, Ostrander A, Tiwari C (2016) Decoding the development strategy of a major retailer: Wal-Mart's expansion in the United States. *The Professional Geographer* 68: 640–649. CrossRef

SBB (2024a) Ein- und Aussteigende an Bahnhöfen. Dataset, https://data.sbb.ch/explore-/dataset/passagierfrequenz

SBB (2024b) Freie Retailflächen. Website, https://sbb-immobilien.ch/mieten/retail/

SBB (2024c) SBB Company. Website, https://company.sbb.ch/en/home.html

SBB (2024d) Öffnungszeiten Shops. Dataset, https://data.sbb.ch/explore/dataset/offnungszeiten-shops/

Seong EY, Lim Y, Choi CG (2022) Why are convenience stores clustered? The reasons behind the clustering of similar shops and the effect of increased competition. *Environment and Planning B: Urban Analytics and City Science* 49: 834–846. CrossRef

Shahidi F, Rennert-May E, D'Souza AG, Crocker A, Faris P, Leal J (2023) Machine learning risk estimation and prediction of death in continuing care facilities using administrative data. *Scientific Reports* 13: 17708. CrossRef

Simkin L (1989) SLAM: Store location assessment model - Theory and practice. *Omega-international Journal of Management Science* 17: 53–58. CrossRef

SNCF Gares & Connexions (2024) The station, new town centre venue. Website, https:-//www.garesetconnexions.sncf/en/retail/retail-commercial-activity

Stanley TJ, Sewall MA (1976) Image inputs to a probabilistic model: Predicting retail potential. *Journal of Marketing* 40: 48–53. CrossRef

Tabuchi T, Thisse JF (2011) A new economic geography model of central places. *Journal of Urban Economics* 69: 240–252. CrossRef

Taylor RD (1978) Retail site selection using multiple regression analysis. Phd thesis, https://sbb-immobilien.ch/mieten/retail/

Themido IH, Quintino A, Leitão J (1998) Modelling the retail sales of gasoline in a Portuguese metropolitan area. *International Transactions in Operational Research* 5: 89–102. CrossRef

Ting CY, Jie MY (2022) Location profiling for retail-site recommendation using machine learning approach. In: Haw SC, Muthu KS (eds), *Proceedings of the International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022)*, Volume 10 of *Atlantis Highlights in Computer Sciences*. Atlantis Press, 85–116. CrossRef

Trevethan R (2017) Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health* 5. CrossRef

Turhan G, Akalın M, Zehir C (2013) Literature review on selection criteria of store location based on performance measures. *Procedia - Social and Behavioral Sciences* 99: 391–402. CrossRef

Valora Holding AG (2019) Andermatt: Zweigrösster neuer avec Store der Schweiz feierlich eingeweiht. Press release, https://www.avec.ch/media/standorte/eroeffnungen/-20191220_val_mm_avec_andermatt.pdf

Valora Holding AG (2024a) avec - "Handmade with Love". Website, https://www.valora.com/en/brands/avec/

Valora Holding AG (2024b) Immobilien. Website, https://www.valora.com/de/contact/-immobilien/

Valora Holding AG (2024c) k kiosk - “Treat yourself”. Website, https://www.valora.com/en/brands/kkiosk/

Valora Holding AG (2024d) Our brands. Website, https://www.valora.com/en/brands/

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Fourth ed.). Springer, New York. CrossRef

Wang L, Fan H, Wang Y (2018) Site selection of retail shops based on spatial accessibility and hybrid BP neural network. *ISPRS International Journal of Geo-Information* 7. CrossRef

Weber B (1979) *Eine statistische Analyse der Abhängigkeiten des Kundenaufkommens von Standorteinflüssen bei Einzelhandelsgeschäften. Dargestellt an ausgewählten Apotheken der Stadt Münster*, Volume 45 of *Schriftenreihe wirtschaftswissenschaftliche Forschung und Entwicklung*. Florentz

Wieland T (2017) Versorgungsstrukturen und Tragfähigkeit von Gesundheitseinrichtungen aus einer standortökonomischen Perspektive. In: Harteisen U, Dittrich C, Reeh T, Eigner-Thiel S (eds), *Land und Stadt - Lebenswelten und planerische Praxis*, Volume 121 of *Göttinger geographische Abhandlungen*. Goltze, 85–116

Wieland T (2018) Standorterfolg in Zeiten des Onlinehandels - Aufbau, Ergebnisse und planungsbezogene Implikationen einer modellgestützten Standortanalyse für die Elektrofachmärkte in der Region Mittlerer Oberrhein. *Berichte. Geographie und Landeskunde* 92: 5–26. https://www.geographische-regionalforschung.de/download/1254/bd-92-heft-1/1481/bgl_bd_92_heft_1_2018_01_wieland.pdf

Wieland T (2023) Spatial shopping behavior during the Corona pandemic: Insights from a micro-econometric store choice model for consumer electronics and furniture retailing in Germany. *Journal of Geographical Systems* 25: 291–326. CrossRef

Wolinsky A (1983) Retail trade concentration due to consumers’ imperfect information. *The Bell Journal of Economics* 14: 275–282. CrossRef

Zhou L, Wang S, Li H (2024) Store network expansion in the era of online consumption: Evidence from the suning appliance retail chain in China. *Applied Geography* 165: 103225. CrossRef

Zhou Q, Huang K, Huang D (2015) Forecasting sales using store, promotion, and competitor data. Report, https://jmcauley.ucsd.edu/cse255/projects/fa15/022.pdf

Zhu T, Singh V (2009) Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics* 7: 1–35. CrossRef