ROUTE AND ACTIVITY LOCATION CHOICE BEHAVIOUR OF DEPARTING PASSENGERS IN TRAIN STATIONS

Danique Ton [1)*], Jeroen van den Heuvel [1) 2)], Winnie Daamen [2)] and Serge Hoogendoorn [2)]
[1)] NS Stations, Netherlands Railways
[2)] Delft University of Technology, the Netherlands
[*] e-mail: danique.ton@nsstations.nl

ABSTRACT

In the Netherlands, rail passenger transport has grown considerably during the last decades. Although growth has been seen nationwide, it has concentrated at a few of the largest stations in the country (e.g. Amsterdam Central and Utrecht Central station). As a result, these stations are being redeveloped. Furthermore, this growth trend is expected to continue into the future. This also means that (temporary) overcrowding can arise at these stations. These trends have triggered a need for more insights into the behaviour of passengers inside train stations. With these insights, existing stations can be better managed, and redevelopments can be better designed and planned.

This paper presents quantitative research results regarding the factors that influence route and activity location choice behaviour of departing passengers at train stations. For this study, SMART Station data has been used. This automated data collection system is based on Bluetooth, Wi-Fi and infrared sensors. It generates revealed preference data on passenger routes and activities in train stations. Estimated discrete choice models based on these data show which factors influence the choice of departing passengers for vertical infrastructure facilities (i.e. escalators and stairways; route choice) and the choice of retail outlets in the station building (activity location choice).

The factors significantly influencing route choice behaviour of departing passengers are: travel time, walking distance, train stop location alongside the platform and right side orientation of the vertical infrastructure facility which provides access to the platforms. Travel time and walking distance appeared to be highly correlated. To get a better model fit, only the first has been included in the final model, together with train stop location and orientation. Travel time is the most important factor.

In the activity location choice model, also four factors turned out to be of significant influence: travel time from station entrance to the retail outlet, total distance (from entrance via outlet to the platform), the requirement to make a detour for a shop visit, and right side orientation. Distance is the most important factor in this model.

To the best of our knowledge, before this study there has been no empirical, quantitative evidence of the factors which contribute to route and activity location choices inside a train station. Due to the large amount of revealed preference data which has been collected by Bluetooth/Wi-Fi-sensors, this study can be considered as first of its kind in the field of train passenger route and activity choice behaviour in train stations.

1. Introduction

Since 1970, train ridership at Netherlands Railways (NS) has more than doubled, from 8 billion passenger kilometres in 1970 to over 17 billion in 2014 (Figure 1). Growth has been driven by social-economic factors which also have pushed growth of other transport modes, particularly car transport. Specifically for rail, the introduction of the free travel card for students caused a leap in ridership in the early '90s. Moreover, increasing congestion in the urban areas of the Netherlands and improvements in the railway network have stimulated train use (Veenendaal, 2004).

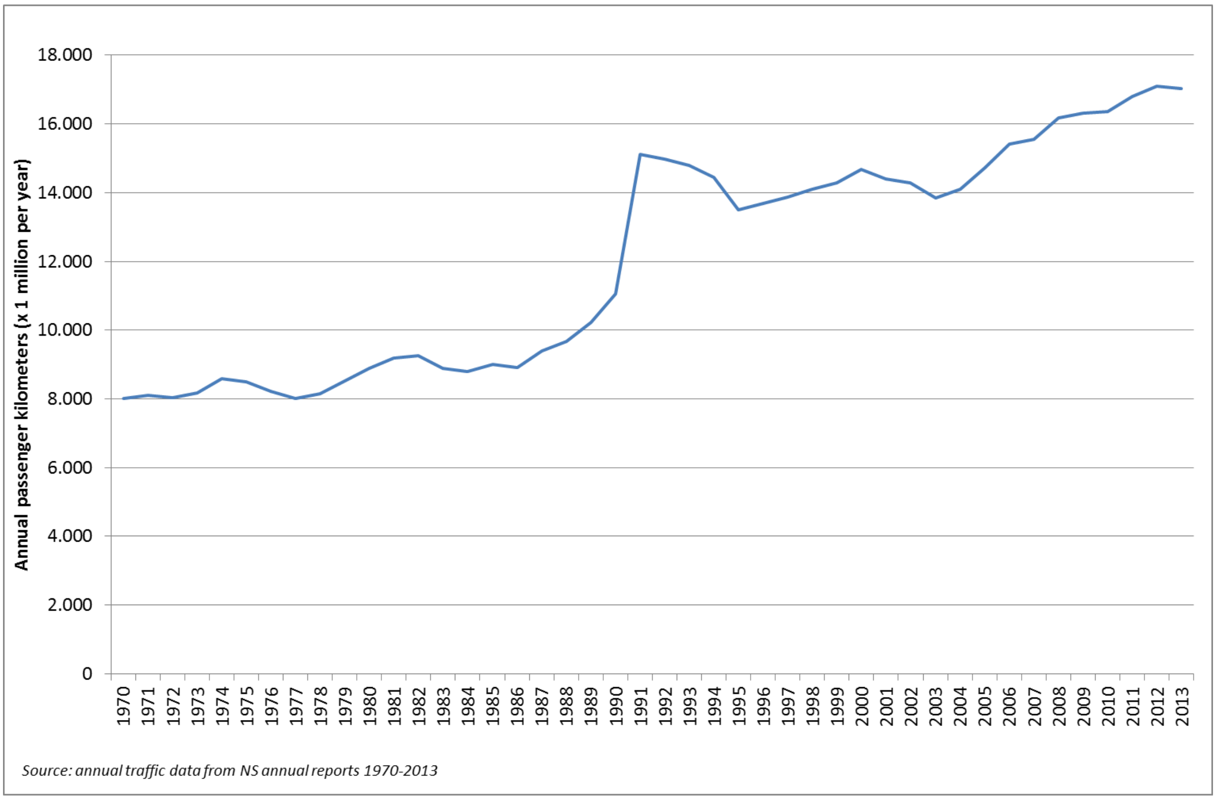

Figure 1 - Trend in rail passenger transport between 1970 and 2013

The number of passengers has grown by approximately 20% from 1992 until 2011, nationwide from approximately 2 million to 2.4 million arriving and departing passengers per average workday. This growth has been concentrated at the large stations (see Figure 2). For example, Utrecht Central station – the nation's largest station in terms of arrivals and departures – has contributed to national growth by 55,000 arrivals and departures per average workday (14% of the total growth). Whereas for example, Amsterdam Central station – the second largest station – has contributed 24,000 (6% of the total growth).

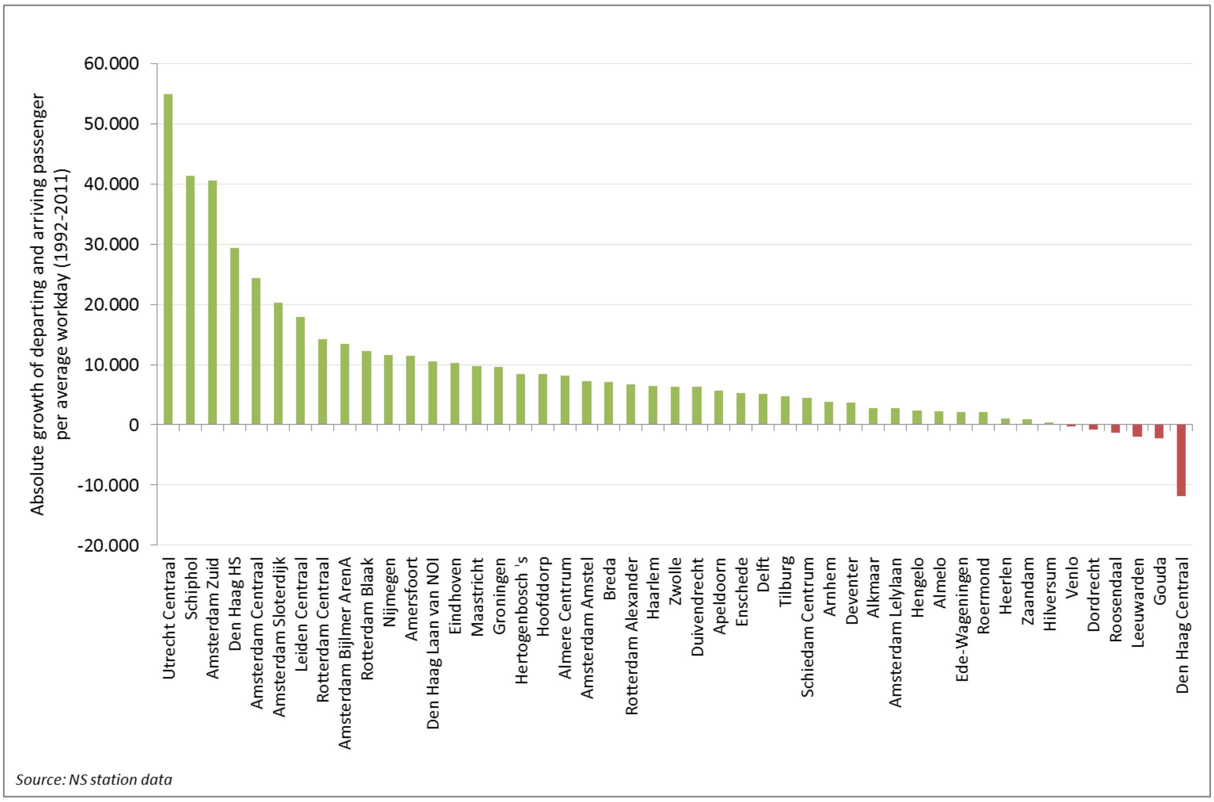

Figure 2 - Growth distribution over the 47 largest train stations in the Netherlands

Parallel to the growth in the number of passengers and passenger kilometres, the largest train stations were also transforming from transport-only hubs to places where people travel *and* perform activities (i.e. working, meeting). The combination of these trends implies that the space inside the large train stations is shared by more people, who perform more and more diverse activities and stay for longer time periods. As a result, the available space per passenger becomes scarcer, whilst capacity expansion is extremely expensive due to the dense urban environment in which these stations are situated.

To get more added (functional) value of existing facilities and/or new investments, more insights are required into passenger behaviour. An important aspect of passenger behaviour is choice behaviour, more specifically route and activity location choice behaviour. We define a route as the way or road taken from a starting point to a destination. The definition of activity we accepted is a thing a person does or has done. The NS - as the largest train operator and the station operator in the Netherlands – is continuously monitoring the performance of its train stations and looking for new knowledge and best practices for improvements in design and operations.

Most existing (practical) knowledge on route choice and activity location choice is based on qualitative experience which took years to develop (tacit knowledge). Also, research about this subject is only available to a limited extent, particularly related to passengers inside multi-functional public transportation hubs. NS and Delft University of Technology are currently investigating the behaviour of passengers at train stations. This paper aims to determine the relevant factors for the route and activity location choice behaviour of departing passengers in train stations, and to quantify their influence.

The paper is structured as follows: section 2 describes choice behaviours of passengers in the train station. Section 3 describes route and activity location choice factors which are relevant in the context of a train station. In order to quantify the influence of the factors, data needs to be collected. The data collection method is described in Section 4. Section 5 introduces the case study of Utrecht Central station. The case is described and the dataset is presented. The data is analysed using discrete choice analysis. Therefore a brief introduction to the models used is provided. Section 6 captures the case study on route choice behaviour. In this section the case study, analysis and results are provided. Next, Section 7 provides the case study on activity location behaviour including the analysis and results. This paper concludes with several conclusions and recommendations for further research.

2. CHOICE BEHAVIOUR AND PASSENGER TYPES

Knowledge about passenger behaviour can be derived from researching choices which are made at the station. Based on a literature survey, five choice behaviours have been distinguished. Please note that these choices are related to one another (Ton, 2014):

1. *Activity choice*: the selection of activities to be performed at the train station. When combining the selected activities, an activity set is created;
2. *Activity hierarchy*: the assignment of a relative priority to the activities in the activity set;
3. *Activity sequence*: the order in which the activities of the activity set will be performed;
4. *Activity location choice*: the selection of the location at which each activity in the activity set will be performed;
5. *Route choice*: the selection of the route connecting all activities in the activity set.

Activity choice and activity hierarchy are mainly determined on a strategic level (Hoogendoorn et al., 2001). This strategic level refers to choices made before arriving at the train station. An example of an activity set which results from activity choice is buying a train ticket and buying a cup of coffee in a retail outlet. The ticket is more important to the passenger than the cup of coffee. Given the activity hierarchy, the activity sequence can be determined on a strategic level, and updated on a tactical level (Root & Recker, 1981). Extending the example, the passenger plans to buy the cup of coffee before buying the train ticket (strategic), but changes the order of activities when he is confronted with a queue at the retail outlet (tactical). Activity location choice and route choice take place on a tactical level (Hoogendoorn et al., 2001). When a passenger is familiar with the station, he is able to plan the activity locations and routes before arriving at the station. Potentially faced with unexpected (passenger) traffic conditions or changes in the train schedule (i.e. delays), a familiar passenger can adapt the plans during the trip. When a passenger is unfamiliar with the station, he is not able to plan ahead (van Hagen, 2011) and always decides on the activity location and route during the trip. Following the previous example, a familiar passenger finds his preferred ticket vending machine out of order, is confronted with large queues at the adjacent machines, and decides to walk to another station entrance where he knows more ticketing vending machines are located. The result is a change of the planned route to the retail outlet. The unfamiliar passenger decides to queue at the ticket vending machines which he finds at the entrance of arrival, unless another ticket vending machine is visible to him. After buying the ticket, he starts to look for the retail outlet.

All pedestrians in a train station – train passengers and non-passengers - make activity and route choices when present in the train station. These choices partially depend on the type of pedestrian, four of which can be distinguished in the context of a train station: departing passengers, arriving passengers, transferring passengers and non-train passengers (see Figure 3). Passengers spend more time in the station before departure than after arrival, and most activities are performed by departing passengers (NS Stations, 2013). Regarding route and activity location choice, departing passengers are most interesting for our research. Route and activity location choices are made on a tactical level for unfamiliar passengers, while they can be planned (strategic) and updated (tactical) by familiar passengers. Because all passengers are able to make decisions regarding route and activity location on the tactical level, this is chosen as the focus of our research.

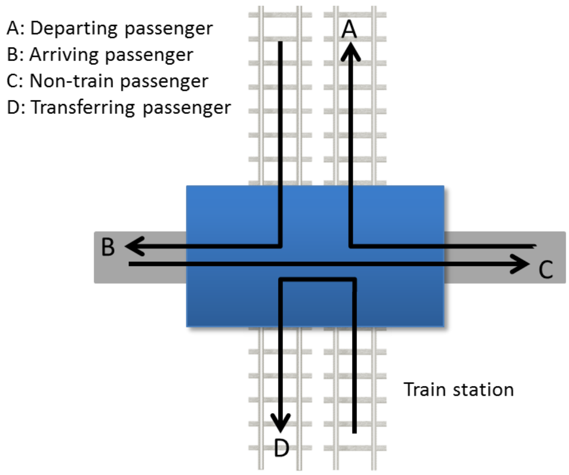

Figure 3: Passenger types in the train station (adapted from Ton (2014))

3. Route and activity location choice factors

Our research aims at determining the factors which influence route and activity location choice of departing passengers in train stations. A literature survey was performed in order to determine which factors are known to influence choice behaviour in general (Ton, 2014). A total of 32 factors has been reviewed for our research. Based on the limited availability of knowledge in the context of train stations and applicability of data collected by the SMART station system (see section 3), eight factors were selected for quantitative research. These factors are:

- ***Orientation***. The orientation of a passenger relates to the preference to walk on the left or right side or to perform activities on the left or right side. Literature shows that this preference strongly depends on the environment or country where people live. For example, in the Netherlands cars drive on the right-hand side of the road. Departing train passengers are therefore expected to have a preference for right-hand side orientation concerning route and activity location choice.

- ***Time spent in the station***. The time spent in the station depends on the arrival time of the departing passenger at the train station. It is a function of the walking time inside the train station and the waiting time which serves as an insurance against missing a train. This buffer time depends on the train frequency and the implications of not being on-time for the desired train service (Taylor, 1994). Missing a local train which runs every 10 minutes has a different impact than missing a 2-hourly Thalys service to Paris which requires a seat-reservation. Research in the Netherlands has confirmed the size of the buffer is related to the perceived risk of missing a train (van Hagen, 2011). Risk prone people do not mind much to miss their train, while risk-averse people make sure they do not miss the train.

- ***Time of day or week***. Golledge (1999) claims that choices and reasons for choosing a route may change during the day or week. This is related to the purpose of the trip (Seneviratne & Morrall, 1985) and the crowdedness in the station. During a weekday morning peak, most passengers travel to work or school, while during the weekend leisure-related trip motives are dominant. The time of day also relates to the crowdedness in the station. During peak hours large train stations (like Utrecht Central station) are crowded, possibly resulting in a different route and/or activity location choice than during off-peak hours.

- ***Visibility***. Visibility relates to whether a route or location is visible for the passenger inside the train station. When a passenger is unfamiliar with the station lay-out, routes and locations that are only partially visible, are less frequently chosen. The distance to the less visible routes and location of less visible outlets is often misjudged by passengers (Montello, 1991). They overestimate it, therefore those routes and locations are less frequently chosen. The hypothesis is that because a departing passenger observes the visible locations and routes first, he or she is more tempted to choose those over the (partially) invisible ones.

- ***Travel time***. This factor has been addressed in many studies, e.g. Cheung & Lam (1998) and Daamen & Hoogendoorn (2003), although all were related to arriving passengers. They claim that travel time (walking time + waiting time) is an important factor when queueing occurs. Various case studies have shown that arriving passengers tend to adjust their route to avoid queues as these can cause a considerable (perceived) increase in travel time (Voskamp, 2012). As travel time is directly related to the transport function of the facility, it can be hypothesised that travel time in a train station is also relevant for departing passengers. Next to that, it is expected that departing passengers try to optimise their travel time when choosing the activity location.

- ***Walking distance***. The walking distance is the distance a departing passenger walks in the train station, from entering the station building to boarding the train. According to Seneviratne & Morrall (1985) the length of a route in terms of distance (meters) and time (minutes) is ambiguous for pedestrians, as they do not know whether they optimise for time or distance. In most cases pedestrians refer to choosing the shortest route. However, in some studies, e.g. Borgers & Timmermans (1986), distance has been used as a factor for choosing a route. In the station environment departing passengers need to catch a train. Therefore, it can be expected that they will optimise on the distance that needs to be covered. The same accounts for activity location choice. If there is a choice between two comparable services at different distances, the closest is often selected unless there is a specific preference for the other service.

- ***Timetable***. The timetable describes the planned platform (spatial) and planned departure times (temporal) of trains. Moreover, the timetable includes the type of service (e.g. local/intercity train) and the service frequency (e.g. a 15-minute interval). Because departing passengers are on their way to catch a train, we expect that the timetable influences both route and activity location chosen. In existing research, the timetable has not been addressed and therefore provides interesting opportunities for this research.

- ***Train operations***. The train operations are defined as the degree to which the timetable has been executed (plan versus realization). This factor can be described as a train departing on time or with a delay, a train departing from a different platform than planned, and the cancellation of a train. It can be expected that changes in the time table affect the route and activity location chosen (for example due to platform changes). Van Hagen (2011) has addressed this issue in the context of waiting behaviour of passengers.

For quantitative analysis, the previously mentioned factors have been operationalised as presented in Table 1.

Table 1: Operationalisation of the influential factors

	Operationalisation			
Factor	**Route choice**	**Units**	**Activity location choice**	**Units**
Orientation	Orientation to the right side	left/right	Orientation to the right side	left/right
Time spent in the station	-	-	Time spent at the location	minutes
Time of day or week	Peak hours versus off-peak hours	peak/off-peak	Peak hours versus off-peak hours	peak/off-peak
Visibility	Visibility of vertical infrastructure	visible/invisible	-*	-
Travel time	Travel time to the platform	minutes	Travel time before the activity	minutes
			Travel time after the activity	minutes
Walking distance	Distance	meters	Total distance	meters
			Detour in the route	detour/no-detour
Timetable	Stop location of the train	section A/section B/section A+B	Service type of the train	IC/Local/INT
			Platform of departure	1-4/ 5-7/ 8-9/ 11-12/ 14-15/ 18-19
Train operations	Delay	delay/no delay	Delay	delay/no delay

*The visibility related to activity location choice is not operationalised because in train stations retail outlets are often located in plain sight.

4. Data collection method

In order to quantify the factors influencing route and activity location choice of departing passengers in train stations, data about these choices is required. For data collection, both stated preference and revealed preference data can be considered (Bovy & Stern, 1990). In general, revealed preference data is preferred over stated preference data, because it generates data on what passengers actually did. Stated preference data cover what passengers would do. However in pedestrian behaviour research, one of the most important and pressing issues is the measurability of choice behaviour in real life situations. Traditional methods, such as manual counts and surveys, are expensive, time-consuming, and often result in biases due to selective sampling and/or small sample sizes (e.g. Helbing et al., 2002).

Due to technological developments, it has become possible and affordable to measure revealed pedestrian behaviour of large samples in an automated way, for example by Bluetooth tracking or analysing mobile phone data (e.g. Versichele et al. (2012)). Privacy is an important concern for automatic measurements of revealed behaviour of a large number of pedestrians. This issue can be overcome when the measurement system and the data processing procedures have been designed properly, and all researchers involved are working privacy-aware (Van den Heuvel et al, 2013).

In our research, we have used data from the automated measurement system SMART Station, which has been developed by NS Stations and its technology partners. Data is collected by Bluetooth, Wi-Fi and infrared sensors which track and count pedestrians in train stations (Daamen et al., 2015). The Bluetooth and Wi-Fi sensors measure movements by pedestrians carrying a mobile device with Bluetooth and/or Wi-Fi enabled. By placing these sensors at multiple points in the train station, Bluetooth/Wi-Fi enabled mobile devices - which act as a proxy for a pedestrian – are tracked during their route through the station. The amount of sensors and their distribution over the train station determine scope, accuracy and resolution of the collected route data. Not every pedestrian has a Bluetooth/Wi-Fi enabled device. Therefore, infrared counters are used to calibrate for the share of pedestrians with a Bluetooth/Wi-Fi enabled device, by counting all pedestrians at one or more strategic points in the train station. Total pedestrian flows, walking times, waiting times and routes are derived by combining the tracking and counting data.

The use of SMART Station data has two limitations. Firstly, a pedestrian (or device) cannot be tracked in a continuous way. Therefore, exact movements (left or right from a sensor) or local trajectories are not captured. With SMART Station, a pedestrian movement is detected as a sequence of confirmations of the presence of a pedestrian in time at several points in the train station. Secondly, due to privacy limitations, demographic factors are unknown and identification of the same person on different days is not possible (van den Heuvel et al., 2013). This means that only the influence of alternative specific factors can be determined using SMART station data and not the influence of demographic factors.

The first SMART Station system has been installed at Utrecht Central station in the fall of 2013. For this study, SMART Station data from that station has been used. At this station, sensors are located at every entry, exit, escalator and stairs in the station building and many retail outlets. Due to the high penetration of sensors at this station, the measurement system is able to generate detailed data regarding route and activity location of the station users.

5. CASE STUDY INTRODUCTION

In this section, background information on the case study location (Utrecht Central station) is provided. Next to that, the data selection procedure for our dataset is described. Finally, the data analysis method is introduced: discrete choice models.

Station Utrecht Central

On an average weekday approximately 85,000 train passengers depart by train from Utrecht Central station. Approximately 60% travels during off-peak hours, and 40% during peak hours (7.00-9.00h and 16.00-18.00h). For departures, the evening peak-hour traffic is dominant over the morning peak (25% vs. 16%). The majority of the departing passengers (around 75%) travels with work or school related trip motives. The other departing passengers travel with social/recreational/leisure motives (NS, 2014).

Inside the station building many retail outlets are situated. Most of these shops are catered for the ‘to go’ segment. These shops are targeted at departing passengers, as this segment prefers an as low as possible in-store time, to allow them to catch their train. Examples are the AH to Go (convenience), Brooodzaak (sandwich) and Smullers (fast food). Other shops are targeted more at the ‘to stay’ segment. Examples are Starbucks (coffee), Burger King (burgers) and Julia’s (Italian food).

Dataset

The data collected by the SMART station system covered a period of three weeks for our research, from August 30, 2013 to September 19, 2013. As mentioned before, every sensor in the station building captures mobile devices with Bluetooth and/or Wi-Fi enabled. Figure 4 shows the selection procedure of the dataset. In those three weeks, a total of 92 million scans were registered. The sensors also register devices that are relatively far away. Therefore, a limit was established which filtered the devices that were located too far from the sensor. This limit is estimated at -70 dBm inside the station and -90 dBm at the entrances and exits. Inside the station we require more detailed information than on the borders. Therefore that limit is higher. All scans are then combined into routes of individual devices. A total of approximately 1.3 million individuals was seen in those three weeks. Next the dataset needs to be filtered on complete routes, departing passengers and travel time. This selection procedure results in 240,949 individual devices in our general dataset.

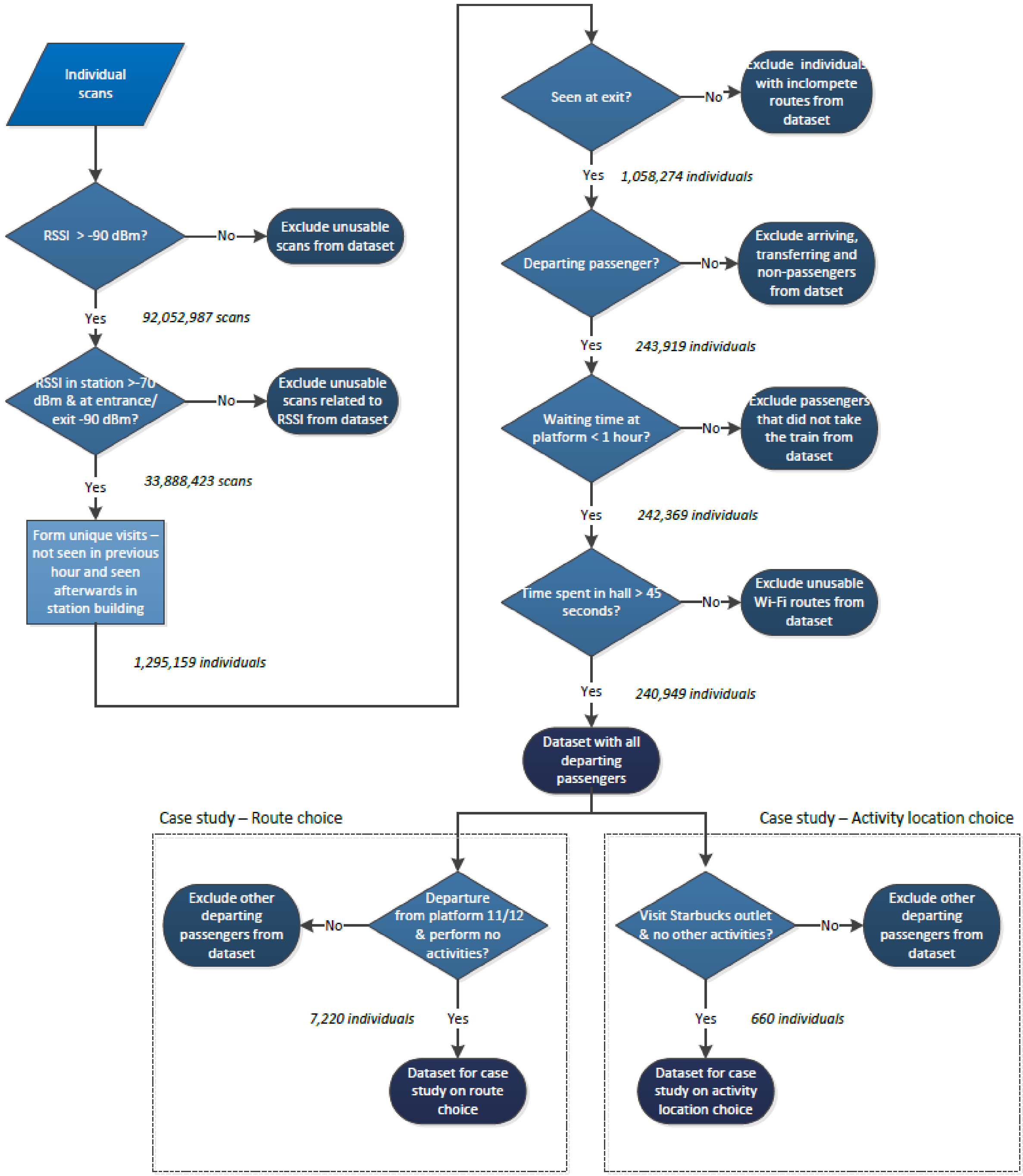

Figure 4: Selection of the datasets

Discrete choice modelling

A decision such as choosing a route or location is mutually exclusive, meaning that only one of the choice options can be chosen (Ben-Akiva & Bierlaire, 1999). Therefore, discrete choice models provide a suitable modelling class for estimating the relevance of factors for route choice and activity location choice. The discrete choice models in our case study are based on the concept of utility, which represents the value or satisfaction of a good or service to a user. A frequently used assumption in discrete choice analysis is the assumption of utility maximisation, which implies that the decision maker chooses the option with the highest utility. In the context of a train station, where efficiency is essential, this concept is applicable as well. For this study, the basic Multinomial Logit Model (MNL) has been applied.

6. CASE STUDY - ROUTE CHOICE BEHAVIOUR

It has been argued that route choice and activity location choice take place simultaneously (Hoogendoorn & Bovy, 2005), particularly related to pedestrians who are familiar with the spatial environment. However, for departing passengers who do not visit any retail outlets in the train station, only route choice is relevant. As these passengers go straight to the platform where their train will depart. These passengers are therefore incorporated in the case study on route choice behaviour.

Many route alternatives have already been excluded before a passenger arrives at the station. Firstly, the location of the station entrance is (for most passengers) set by the location where access modes (transport mode to travel to the station) are linked to the station. At Utrecht Central station, the most important access mode locations are the bicycle storage points, bus/tram stops and the city pedestrian infrastructure. These locations define whether departing passengers enter the station from the eastern or western entrance (see Figure 5). Secondly, a train passenger has already decided on his travel destination. Because the platform of departure is included in the timetable, the platform is known in advance. Together, these fixed origin and destination limit the degree of freedom in movement through the station. As mentioned before, SMART station sensors are present at each entrance and exit point and vertical infrastructure (stairs/escalators). Therefore, the choice of vertical infrastructure facility to get to the departure platform is used as the research aspect for our study (see Figure 5).

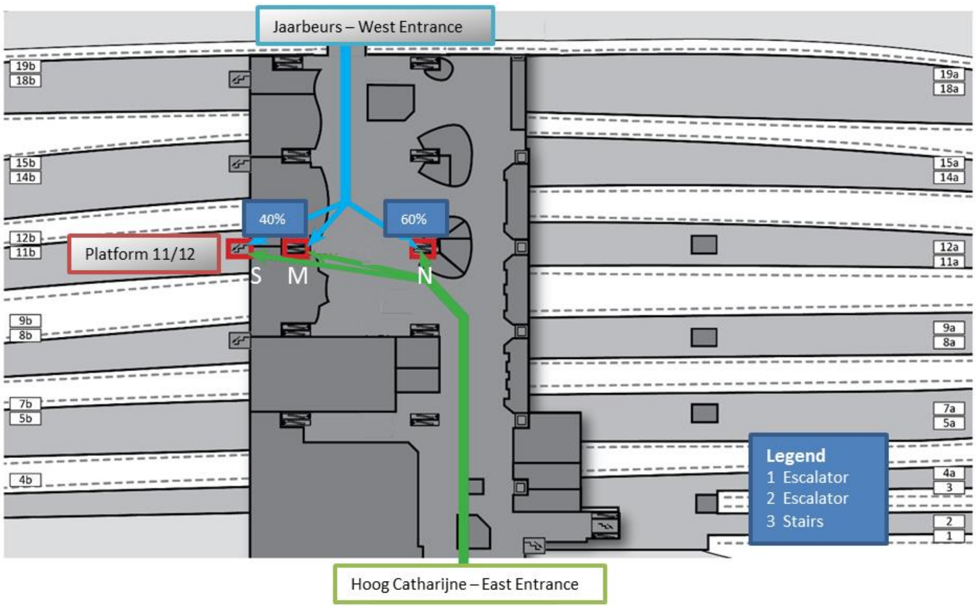

Figure 5: Case study on route choice within Utrecht Central station

Not all platforms can be included in the model. If the train of a departing passenger departs from track 11, he does not include a route towards track 5 in his alternatives. Therefore, we have selected one platform for our research. To get valid research results, we used a multi-criteria analysis (MCA) to select the most representative platform. Table 2 shows the MCA for the platform selection.

Table 2: Multi-criteria analysis on platform selection

No.	Criteria	Best score?	Weight
1	Distance from entrance to platform	Most comparable for both entrances	1
2	Distribution of trains over time	Most uniform distribution	4
3	Distribution of service type of the trains	Most uniform	2
4	Amount of disrupted trains	Least number of disrupted trains	3
5	Distribution of travel time	Most representative for total population	5
6	Distribution of peak/off-peak	Most representative for total population	5

Platform 11/12 shows the best score on these criteria (78 points vs. platform 8/9 (no. 2; 45 points). Therefore this platform is used in our further research (see Figure 5). All departing passengers who have departed to platform 11/12 and did not perform any activity during their stay in the station have been included in the dataset. This leaves 7,220 departing passengers in the dataset out of the 240,949 departing passengers that were present in the general dataset (see Figure 4).

For this platform, SMART Station data has revealed that 60% of the departing passengers chooses the north escalator, while the other 40% is distributed over the middle escalator and south stairways. This shows a preference towards the north side. The research question in this case study is: *which factors have a significant influence on the route choice behaviour of the departing passengers who do not perform activities in the station?*

A first, explorative analysis on the route choice dataset has provided an impression of the potential influence of the various factors on route choice, as well as insight in the data itself (see Table 3).

Table 3: Exploring the route choice factors

Factor	Via north escalator	Via middle escalator	Via south stairways	Total
Travel time: mean (std. deviation)	1:40 min (0:24 min)	1:59 min (0:23 min)	1:49 min (0:34 min)	1:44 min (0:28 min)
Peak – Off-peak distribution	44.8% - 55.2%	36.5% - 63.5%	47.3% - 52.7%	55.1% - 44.9%
Walking distance from entrance: East - West	73.0% - 27.0%	40.4% - 59.6%	54.8% - 45.2%	64.6% - 35.4%
Stop location of train: section A –section B – section A&B	22.2% - 4.8% - 73.0%	13.0% - 7.4% - 79.5%	17.9% - 7.2% - 74.8%	20.1% - 5.8% - 74.1%
Delay – No delay	9.5% - 90.5%	10.7% - 89.3%	13.0% - 87.0%	10.7% - 89.3%
Orientation: left – right	27.0% - 73.0%	40.4% - 59.6%	54.8% - 45.2%	36.9% - 63.1%

Table 3 shows the results of the explorative analysis of the potentially relevant factors for route choice. Please note that the percentages in each cell sum up to 100%. Several interesting observations can be made. Firstly, the travel time is different for all routes. The travel time is measured from the entrance up to entering the platform. This means that the waiting and walking time on the platform are excluded from the travel time. The average travel time towards to southern stairways is lower than to the middle escalator while the physical distance is slightly larger. This could indicate that departing passengers in a hurry have a preference for stairs over an escalator. We observed that passengers using the escalator to go to the platform often stand still on the escalator. Therefore, the choice for the stairs could be to avoid the potential blocking of other passengers. This hypothesis is supported by a relatively high standard deviation of the south stairs. Secondly, distance seems to be an important choice factor. From the eastern entrance, the north escalator clearly is the closest vertical infrastructure facility. In contrast, from the western entrance the distances to all three facilities are not significantly different, this shows for the middle and south vertical infrastructure facilities. Thirdly, peak/off-peak distribution does not seem to differ much, indicating a limited influence of the time of the day and passenger familiarity on route choice. Finally, orientation shows an interesting difference. Departing passengers following the routes via the north and middle vertical infrastructure facilities show a preference for the right-side, whereas for the south stairways a slight preference towards the left-side is shown. These differences are an indication that orientation is of influence to route choice.

The first step in the discrete choice analysis is the estimation of MNL models that create an utility function for each individual factor. This step is taken to do a first selection of the factors that show a significant influence on the route choice. An overview of the results of each single factor choice model is shown in Table 4.

Table 4: Results of single-factor route choice models

Parameter	Value	Robust t-test	p-value	Log-likelihood (LLH)	Adjusted rho-square
Travel time	-2.18	-33.10	0.00*	-6,070	0.234
Peak	$3.56e^{-15}$	0.00	1.00	-6,315	0.203
Off-peak	$2.52e^{-15}$	0.00	1.00		
Walking distance	-0.0269	-27.51	0.00*	-6,234	0.225
Stop location of the train	0.166	6.60	0.00*	-6,293	0.206
Delay of the train	$-2.42e^{-15}$	-0.00	1.00	-6,315	0.203
Visibility	$1.06e^{-15}$	0.00	1.00	-6,315	0.203
Orientation	0.230	18.19	0.00*	-6,147	0.225

*Significant on a 95% confidence interval

Three factors turned out to be not significantly different from zero: peak/off-peak, train delays and visibility. Note that the peak/off-peak factor has been included in the model as a dummy variable, where the other nominal variables are included with effect coding (Ben-Akiva & Lerman, 1985). This results in two parameters for the peak/off-peak utility function and one parameter for the effect coded variables. These results indicate that the time of day does not influence route choice. Indirectly, this outcome is an indication that the familiarity is no significant factor in route choice in the central station of Utrecht. This result is reasoned by the fact that during peak hours familiar passengers are the dominant users of the train station, while during off-peak hours unfamiliar passengers are the dominant users. Regarding visibility, this indicates that departing passengers seem to be aware of the locations of the vertical infrastructure facilities, despite the fact that not all facilities are visible from every direction. As every vertical infrastructure facility is referred to in the station's signing system, this result could be an indication of the added value of signing.

Four factors have been found to be significant on a 95% confidence interval: travel time, walking distance, stop location of the train and orientation. If the vertical infrastructure facility provides direct access to the train – so there is no need to walk alongside to platform to get to the train – this has a positive effect on the utility of the alternative. Also, the orientation provides a positive value to the utility function if the vertical infrastructure is located on the right-hand side of the pedestrian flow. This outcome is consistent with expectations, as people in the Netherlands are mainly right-side oriented. The travel time shows the highest model fit and therefore seems to be the most important factor. The higher the travel time, the lower the utility for that alternative. The same is applicable for distance, although to a smaller extent. This is an indication that departing passengers optimise travel time in the station instead of walking distance. They adapt speed to get to the platform faster in order to catch their train. This seems to be consistent with the logic that time is more important when catching a train. Walking distance and travel time are highly correlated (significant on 99% confidence interval).

The second step is the combination of the significant factors into one model in order to determine the relevance of each factor. Due to their high correlation, walking distance and travel time cannot be combined into one model. Since travel time has a higher model fit than walking distance, travel time is used for further analysis and distance is excluded. Combining the three remaining factors into one model indeed results in a better model fit. The utility function for alternative *i* is expressed by the following equation:

$$U_i = ASC_i + \beta_{tt} \times Travel\ time_i + \beta_{tr} \times Train\ stop_i + \beta_o \times Orientation_i$$

As shown in Table 3 the southern stairway has a considerably higher travel time than the northern escalator. Still, the south stairway is relatively often used by departing passengers (31%). Because the travel time is found to be a dominant factor in the model, it has a large, negative impact on the utility of the south stairs in the model. This impact is compensated in the alternative specific constant of the model

(ASC_{south}). An effect of this compensation is a high correlation of the ASC_{north} (northern escalator) with the travel time factor. To avoid this correlation issue, we have excluded the ASC_{north} from the final model.

Table 5: General information on the multi-factor route choice model

Model	Multinomial Logit
Number of estimated parameters	4
Number of observations	7,220
Null log-likelihood	-7,931
Final log-likelihood	-5,991
Likelihood ratio test (null situation)	3,881
Likelihood ratio test (individual model)	157
Adjusted rho-square	0.244

Table 5 shows information about the final route choice model. The performance of the model has been tested with the likelihood ratio test. This value is provided in the table for the comparison of this model with the null situation. The critical value at 95% confidence interval is 9.49 (at four degrees of freedom). Therefore the model is significantly better than the null situation. A comparison of this model with the best single-factor model – the travel time model – is also made. The resulting likelihood ratio is 157. This is higher than the critical value of 5.99 on a 95% confidence interval. Therefore, it can be concluded that the multi-factor model is significantly better than the best single-factor model.

Table 6: Utility parameters for the multi-factor MNL model

Name	Value	Robust Std err	Robust t-test	p-value	Relative importance
ASC_M	-1.64	0.0436	-37.53	0.00*	-11.31
ASC_S	0.00	fixed			0.00
Beta_o	0.156	0.0135	11.61	0.00*	1.08
Beta_tr	0.145	0.0261	5.56	0.00*	1.00 (reference)
Beta_tt	-1.76	0.0736	-23.89	0.00*	-12.14

*Significant on a 95% confidence interval

The estimated parameters of the multi-factor MNL model are presented in Table 6. The ASC_{south} is fixed to zero. Therefore, the southern and northern vertical infrastructure facility have the same base preference in this model. All attributes are significant on the 95% confidence interval. The last column of Table 5 shows the relative importance of the parameters. The parameter for train stop location is used as a reference and forced to 1.00. It shows that travel time is the most important parameter in the model (12 times higher than the reference parameter). Orientation and train stop location are of equal importance in the model and contribute to a limited extent to the utility of each alternative.

Concluding, the results from the discrete choice analysis for route choice have been visualised in Figure 6. A red colour indicates a negative relationship, whilst a green colour indicates a positive relationship. Next to that, the bigger the arrow, the higher the relevance of that factor regarding route choice. The factor travel time is found to be the most important factor for route choice. This finding is consistent with various theories and models (e.g. Daamen & Hoogendoorn (2003)). A second important route choice factor is distance. Due to a high correlation between travel time and distance, only the travel time has been incorporated into the final multi-factor model. The right-side orientation of the vertical infrastructure facility related to the pedestrian flow and train stop location relative to the vertical infrastructure facility also contribute to the utility, however to a much less extent than the travel time factor. The other factors - peak/off-peak, visibility and delay – are found not to have a significant influence on route choice behaviour in our case study.

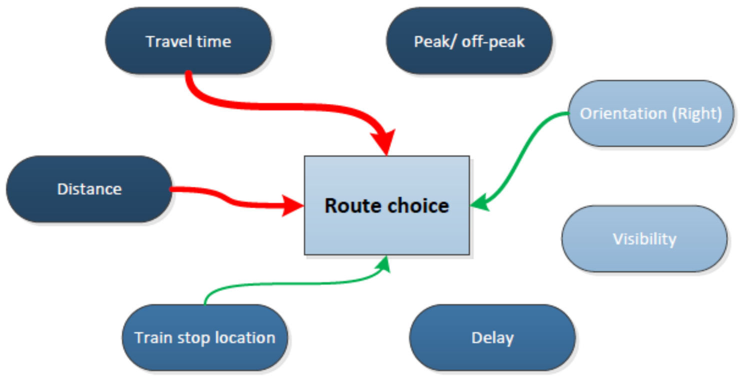

Figure 6: Factors influencing route choice

7. CASE STUDY - ACTIVITY LOCATION CHOICE BEHAVIOUR

The case study for activity location choice behaviour covers departing passengers that visit a retail outlet and then head to the platform. Contrary to the case study on route choice, now passengers heading to all platforms will be included. The main reason for this approach is the sample size. The number of observations of departing passengers from platform 11/12 who also have performed an activity is too small to get valid research results.

Activity location choice is related to the activity departing passengers perform in the station. This choice can have different influential factors for different activities. Therefore, only one activity is selected for our research. The activities related to station retail are grouped into six different categories: coffee, burgers/fries, pasta, train information/train tickets, sandwich and non-food. At Utrecht Central station, eleven shops have been equipped with the SMART station sensors. By means of an MCA, the most representative activity category is selected (see Table 7).

Table 7: Multi-criteria analysis on activity category selection

No.	Criteria	Best score?	Weight
1	Distance from entrance to location to platform	Most variation	5
2	Distribution of service types of the train	Most variation	3
3	Distribution of time spent	Most variation	3
4	Distribution of peak/off-peak	Most representative	4

Based on these criteria the shopping category coffee is selected (27 points vs. burger/fries (no. 2; 18 points). In Utrecht Central station several retail outlets are present which sell coffee. However, most also sell other products. In order to make a sound comparison, only Starbucks is included in the research. This is the only retail shop which focuses on coffee. Two different Starbucks outlets are present in Utrecht Central station (see Figure 7). All departing passengers visiting either of the two Starbucks locations (and do not perform any other activities) before heading to the platform have been included in the dataset. This resulted in a selection of 660 departing passengers (see Figure 4).

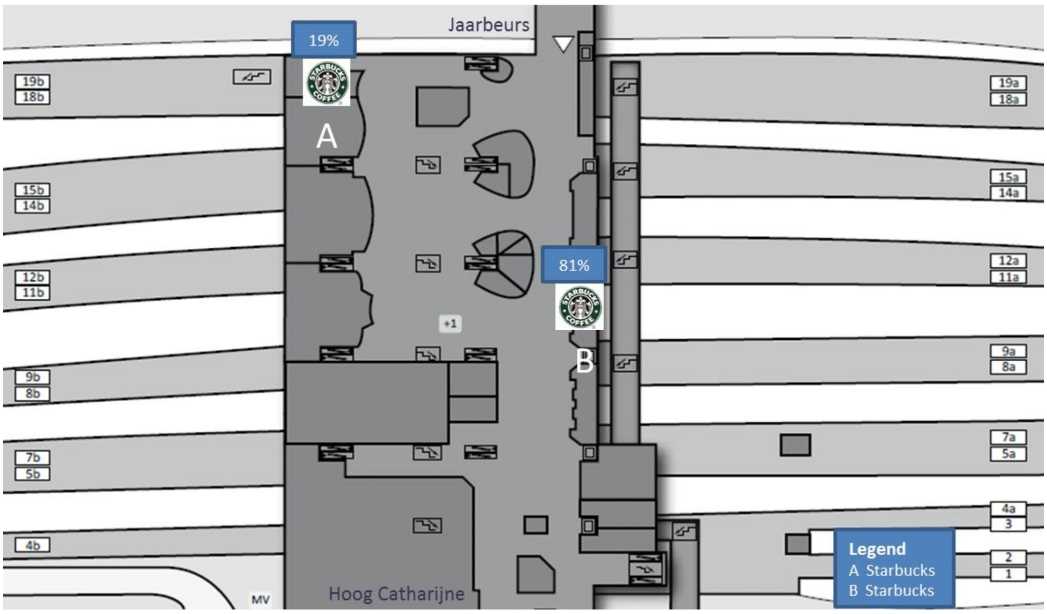

Figure 7: Case study on activity location choice within Utrecht Central station

SMART Station data has revealed that Starbucks outlet B is used by far more clients than outlet A (80%-20%; Figure 7). Although outlet A is smaller in terms of floor size than outlet B, both stores sell exactly the same products and offer a similar experience. This shows a preference towards outlet B. The research question in this case study is: *which factors have a significant influence on the Starbucks location choice behaviour of the departing passengers?*

We start by exploring the data to get a first impression of the relevance of each activity location choice factor that has been identified in the literature survey (Table 8).

Table 8: Exploring the activity location choice factors

Factor	Starbucks A	Starbucks B	Total
Mean travel time from entrance to Starbucks (std. dev.)	1:43 min (0:32 min)	1:22 min (0:40 min)	1:33 min (0:38 min)
Mean time spent in Starbucks (std. dev.)	4:26 min (3:29 min)	7:55 min (7:36 min)	5:06 min (7:01 min)
Mean travel time from Starbucks to platform (std. dev.)	5:33 min (6:03 min)	6:21 min (6:14 min)	5:42 min (6:05 min)
Peak – off-peak	44.1% - 55.9%	29.5% - 70.5%	32.3% - 67.7%
Distance from entrance to outlet to platform: average (std. deviation)	185 m (45 m)	188 m (49 m)	186 m (47 m)
Detour – no detour	60.6% - 39.4%	71.7% - 28.3%	69.5% - 30.5%
Train service: International – local – intercity	0.8% - 36.2% - 63.0%	0.9% - 28.3% - 70.7%	0.9% - 29.8% - 69.2%
Platform of departure: 1-4 – 5/7 – 8/9 – 11/12 – 14/15 – 18/19	4.7% - 13.4% - 6.3% - 17.3% - 22.0% - 36.2%	8.1% - 21.6% - 16.5% - 20.8% - 19.3% - 13.7%	7.4% - 20.0% - 14.5% - 20.2% - 19.8% - 18.0%
Delay – no delay	11.0% - 89.0%	16.5% - 83.5%	15.5% - 84.5%
Orientation: left – right	56.7% - 43.3%	22.3% - 77.7%	28.9% - 71.1%

The explorative analysis shown in Table 8 reveals various relevant observations. Please note that the percentages sum up to 100% in each cell. Firstly, the mean travel time from station entrance to either one of the Starbucks outlets is small (1:22/1:43 min), indicating that these departing passengers head straight to the Starbucks outlet without including waiting or decision time. Secondly, looking at the differences in the time spent between Starbucks A and B, two things can be noted. Departing passengers stay longer in

Starbucks B (7:55 min) than in Starbucks A (4:26 min). This difference could be explained by the size of the outlet: outlet B is larger and has more seating capacity, allowing more people to finish their coffee before going to the platform. Next to that, the time spent of both outlets is rather large. This indicates that the Starbucks outlet is indeed used as a 'to stay' shop, passengers are enjoying their coffee in the outlet. Thirdly, the travel time between the Starbucks outlets and the platform is relatively high (5:33/6:21 min). This is an indication that outlet visitors wait in the station building, before getting to their platform. The average walking time heading direct to the platform – given a speed of 5 km/h – would be 0:56 minutes. Fourthly, in a lot of cases a detour is necessary for visiting a Starbucks outlet, this depends on the entrance and platform of departure. In case of a detour, Outlet B is more often chosen than outlet A. This again might be due to the available seating capacity. Fifthly and finally, the orientation factor shows a large difference. When left-side orientation is involved outlet A is more attractive than with right-side orientation. This indicates an influential relation with activity location choice.

For estimating the activity location choice models, the same stepwise approach is used as for the route choice model. First, MNL models for every variable are estimated, providing an utility function for every variable. Then the significant factors are combined in one, multi-factor model. Table 9 gives an overview of all single-factor models.

Table 9: results of single-factor activity location choice models

Parameter	Value	Robust t-test	p-value	Log-likelihood	Adjusted rho-square
Travel time before visit	-0.408	-3.62	0.00*	-316	0.304
Travel time after visit	$3.33e^{-15}$	0.00	1.00	-323	0.289
Time spent on location	$3.88e^{-15}$	0.00	1.00	-323	0.289
Peak	$1.73e^{-16}$	0.00	1.00	-323	0.287
Off-peak	$3.63e^{-16}$	0.00	1.00		
Total distance	-1.21	-7.70	0.00*	-288	0.365
Detour	-0.424	-5.62	0.00*	-306	0.326
Service type: Local	$3.66e^{-16}$	0.00	1.00	-323	0.287
Service type: Intercity	$1.55e^{-16}$	0.00	1.00		
Platform 1-4	$-5.68e^{-17}$	0.00	1.00	-323	0.280
Platform 5/7	$1.06e^{-17}$	0.00	1.00		
Platform 8/9	$-1.87e^{-17}$	0.00	1.00		
Platform 11/12	$1.14e^{-17}$	0.00	1.00		
Platform 14/15	$9.74e^{-18}$	0.00	1.00		
Delay	$-3.70e^{-16}$	0.00	1.00	-323	0.289
Orientation	0.244	4.72	0.00*	-312	0.313

*Significant on a 95% confidence interval

Six factors proved not to have a significant influence on the activity location choice: travel time after visit, time spent on location, peak/off-peak, service type, platform and delay. Please note that peak/off-peak and platform are included in the models as dummy variables, whereas other nominal variables are included using effect coding (Ben-Akiva & Lerman, 1985). Because the travel time after visit consists of both walking time and waiting time, the variation in this variable is too large to explain the choice for either of the outlets (see Table 8). The time of the day does not influence the choice of activity location, this also indicates that familiarity is not an issue in the choice behaviour. Next to that, the platform of departure, possible delays and the service type of the train do not influence the location choice of the activity in the station building.

Four factors have a significant influence on the activity location choice (95% confidence interval): travel time before visit, total distance, detour and orientation. The travel time before visit has a negative value, which means that the alternative (outlet) becomes less attractive when the travel time increases. As expected, the total distance also has a negative influence on the attractiveness of the outlets. The effect of this factor on the activity location choice is strongest. If a departing passenger is required to make a detour to visit an activity location (Starbucks outlet), the outlet becomes less attractive. As expected, the locations on the right side of the walking direction generate a higher utility. This is due to the Dutch population being generally right-side oriented.

All factors referring to attributes of the trip before the outlet visit or the total trip are of significant influence. In contrast, factors describing attributes after the outlet visit are not significant. This seems to indicate that departing passengers do not base their decisions on the events that happen after the outlet is visited. Since detour and total distance are of significant influence, it can be expected that the departing passengers do take the entire route into account, with emphasis on the events before the outlet is visited.

The second step is to combine all significant factors into one model, to see if a better model can be found. Similar to the route choice model, travel time before and total distance are highly correlated. However, this correlation is lower than in the route choice model. In this case, only a segment of the travel time (from entrance to outlet) is incorporated while the distance factor describes the full distance between entrance-outlet-platform. In this case, study distance shows a better model fit. This could be due to the large variation in travel time for all departing passengers (especially after the outlet visit). Therefore, the total distance is included in the model.

Distance is a dominant factor, meaning that orientation and detour are not significant when all three factors are included in the model (see Table 10).

Table 10: Parameters of the combined activity location model

Parameter	Value	Robust std error	Robust t-test	p-value
ASC_A	-1.26	0.112	-11.30	0.00*
ASC_B	0.00			
Beta_distance	-0.0137	0.002	-5.54	0.00*
Beta_detour	-0.110	0.099	-1.10	0.27
Beta_orientation	-0.153	0.078	-1.96	0.05

*Significant on a 95% confidence interval

Because orientation and detour can be neglected in the model, the single-factor model for distance is the best possible model. This model uses the following utility function:

$$U_i = ASC_i + \beta_d \times Distance_i$$

Table 11 shows information on the final activity location model with only total distance included. The likelihood ratio test of this model is 338, which is higher than the threshold of 5.99 at 95% confidence interval. This means that the model is better than the null situation. In addition, the model has an adjusted rho-square of 0.365.

Table 11: General information on the activity location choice model

Model	Multinomial Logit
Number of estimated parameters	2
Number of observations	660
Null log-likelihood	-457
Final log-likelihood	-288
Likelihood ratio test	338
Adjusted rho-square	0.365

The utility parameters of this model are provided in Table 12. The ASC of outlet B is fixed to zero. The ASC of outlet A indicates that this location is preferred less by departing passengers when all other values in the utility function are the same. This corresponds with the distribution of the travellers over the locations. The parameter for distance is -1.21 where distance can range from 1.05 to 2.95 (hectometre). This means that the impact of the attribute ranges from -1.27 to -3.56.

Table 12: Utility parameters of single-factor MNL model

Parameter	Value	Robust std error	Robust t-test	p-value
ASC_A	-1.20	0.105	-11.40	0.00*
ASC_B	0.00			
Beta_d	-1.21	0.157	-7.70	0.00*

*Significant on a 95% confidence interval

Concluding, the results of the discrete choice analysis for activity location choice have been visualised in Figure 8. Again, the red arrow indicates a negative relationship, whereas the green arrow indicates a positive relationship. Also, the bigger the arrow, the higher the impact of that factor on activity location choice. Total distance is the most important factor regarding activity location choice. Travel time from entrance to activity location has a significant and negative effect on the choice. Minor, but still significant factors for activity location choice are the right-side orientation of the retail outlet (similar to route choice) and the requirement to make a detour to visit the shop on the route from entrance to platform. Travel time and total distance cannot both be incorporated in the multi-factor model, due to high correlation. Since total distance is the most important factor, only that factor is included. Due to the dominance of the distance factor, the best model is based on this factor only. Detour and orientation are neglectable when all three factors are combined into a model. The other factors - peak/off-peak, service type, delay, platform, time spent on location and travel time from location to platform – are found not to have a significant influence on activity location choice.

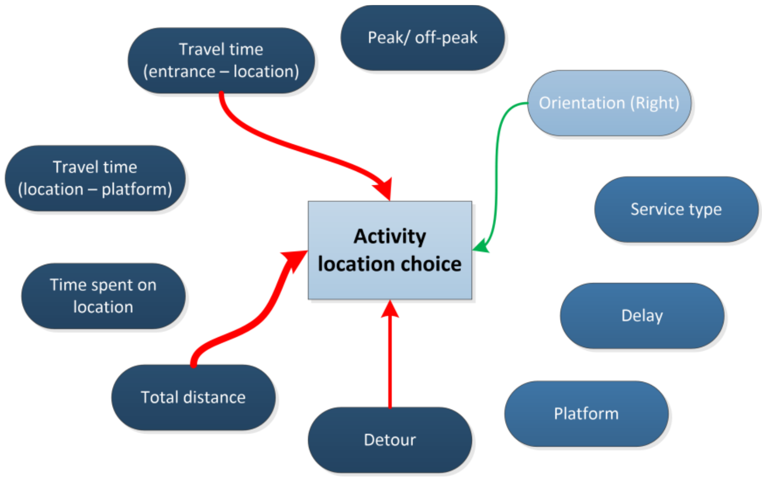

Figure 8: Factors influencing activity location choice

8. Conclusions and Recommendations

The results of this study have several interesting implications for station design and operations. The location and capacity of the vertical infrastructure facilities can be directly linked to the route choice factors of travel time and right-side orientation of the facilities. When large differences in needed travel time towards vertical infrastructure facilities exist, suboptimal usage of some facilities could be the result. This in turn could cause unnecessary divestments or additional investments when during the functional lifetime, relocation or additional capacity is required due to overcrowding. Moreover, the research findings point to the added value for passengers to have trains stopping next to the vertical infrastructure as much as possible. The same argument holds for the position of retail outlets, in our study operationalized by two Starbucks outlets in the station building. Our study has given empirical indications that departing train passengers in general prefer to visit an activity location that is on their route to the platform (no detour necessary).

The scope of this study has been limited to a single station (Utrecht Central station) during a time period of three weeks. Since e.g. station lay-out, number of arriving and departing passengers per day and number and variety of retail outlets differ for other stations in the Netherlands, further research is required to determine the extent to which the findings can be generalized. It is expected that the factors found in this study will also account for other stations. However, the relative importance might differ. Moreover, qualitative research focussed on the most important factors – time, distance and orientation – could provide more insights into the motivation of passengers for their route and activity location choices. These insights will be relevant for managing and redeveloping the station areas.

References

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In R. W. Hall, *Handbook of transportation science* (pp. 5-33). New York: Springer US.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and Application to Travel Demand.* Camebridge : The MIT Press.

Borgers, A., & Timmermans, H. (1986). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical Analysis*, 115-128.

Bovy, P., & Stern, E. (1990). *Route choice: Wayfinding in transport networks.* Kluwer Academic Publishers: Dordrecht.

Cheung, C. Y., & Lam, W. H. (1998). Pedestrian route choices between escalator and stairway in MTR stations. *Journal of Transportation Engineering*, 277-285.

Daamen, W., & Hoogendoorn, S. P. (2003). Research on pedestrian traffic flow in the Netherlands. *Proceedings Walk 21 IV* (pp. 101-117). Portland: Walk 21 conference.

Daamen, W., van den Heuvel, J., Ton, D., & Hoogendoorn, S. (2015). Using Bluetooth and WiFi to unravel real-world slow mode activity travel behaviour. Published at International conference on travel behaviour research, Windsor, United Kingdom.

Golledge, R. G. (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes.* Baltimore: Johns Hopkins University Press.

Helbing, D., Farkas, I., Molnar, P., & Vicsek, T. (2002). Simulation of pedestrian crowds in normal and evacuation situations. In M. Schreckenberg, & S. D. Scharma, *Pedestrian and Evacuation Dynamics* (pp. 21-58). Berlin: Springer.

Hoogendoorn, S. P., & Bovy, P. H. (2005). Pedestrian travel behavior modeling. *Network and Spatial Economics*, 193-216.

Hoogendoorn, S. P., Daamen, W., & Bovy, P. H. (2001). Microscopic pedestrian wayfinding and dynamics modelling. In M. Schreckenberg, & S. Sharma, *Pedestrian and Evacuation Dynamics* (pp. 123-154). Berlin: Springer.

Montello, D. R. (1991). Spatial orientation and the angularity of urban routes: a field study. *Environment and Behavior*, 23-47.

NS. (2014). *Keten Informatie Systeem.* Utrecht: NS [in Dutch].

NS Stations. (2013). *Koopgedragonderzoek 2013: High level rapportage.* Utrecht: NS stations [in Dutch].

Root, G. S., & Recker, W. (1981). Toward a dynamic model of individual activity pattern formulation. *International Conference on Travel Demand Analysis*, (pp. 2-22). Oxford, England.

Seneviratne, P. N., & Morrall, J. F. (1985). Analysis of factors affecting the choice of route of pedestrians. *Transportation Planning and Technology*, 147-159.

Taylor, S. (1994). Waiting for services: the relationship between delays and evaluations of service. *Journal of Marketing*, 56-69.

Ton, D. (2014). *Navistation: A study into the route and activity location choice behaviour of departing pedestrians in train stations.* Delft: Delft University of Technology.

van den Heuvel, J., Thiellier, E., & van Gerwen, N. (2013, March). Privacy by design; Bij reizigersmetingen op stations. *Privacy & Compliance*, pp. 17-21 [in Dutch].

van Hagen, M. (2011). *Waiting experience at train stations.* Delft: Eburon Academic Publishers.

Veenendaal, A. J. (2004). *Spoorwegen in Nederland; Van 1834 tot nu.* Boom [in Dutch].

Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 208-220.

Voskamp, A. (2012). *Measuring the influence of congested bottlenecks on route choice behavior of pedestrians at Utrecht Centraal.* Delft: Delft University of Technology.