University of Groningen

Developing a Service Improvement System for the National Dutch Railways

Verhoef, Peter C.; Heijnsbroek, Martin; Bosma, Joost

Published in:
Interfaces

DOI:
[10.1287/inte.2017.0915](https://doi.org/10.1287/inte.2017.0915)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Verhoef, P. C., Heijnsbroek, M., & Bosma, J. (2017). Developing a Service Improvement System for the National Dutch Railways. *Interfaces*, 47(6), 489-504. <https://doi.org/10.1287/inte.2017.0915>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Developing An Analytical Based Service Improvement System for the National

Dutch Railways¹

Peter C. Verhoef²³

University of Groningen

Customer Insights Center

Martin Heijnsbroek

MICompany

Joost Bosma⁴

NS

Submission for Interfaces

Section Marketing

¹ We acknowledge a larger team that have been involved in multiple phases of this project: Hanneke van de Boog (NS), Menno de Bruyn (NS), Thijs Urlings (NS), Mark van Hagen (NS), Daan de Bruin (MICompany), Sven van Veen (MICompany), Jeroen Boesmans (MICompany), Helen Rijkens (MICompany), Natasha Walk, Hans Risselada (University of Groningen) and Maarten Gijsenberg (University of Groningen). We also acknowledge the helpful comments of the Area Editor and the review team.

² This project was based on a consulting project from MICompany and the Customer Insights Center in joint cooperation with the MI department of the NS. The Customer Insights Center received a fee for this project from the Dutch National Railways. This project was a finalist in the Gary L. Lilien ISMS-MSI Practice Prize. A summary of this submission will also be published in Roberts (2017). This paper concerns a full version of our project for the Dutch National Railways.

³ Corresponding author: Peter C. Verhoef, University of Groningen, Faculty of Economics and Business, Duisenberg building 329; P.O. Box 800, NL-9700 AV Groningen, The Netherlands; E-mail: p.c.verhoef@rug.nl; Phone: +31 50 363 7320

⁴ The authors are displayed in alphabetical order and each of the authors contributed equally to the development of this project.

Developing an Analytical Improvement System for the Dutch National Railways

Abstract

Customer satisfaction is essential for public and railway services, as firms in these industries have contracts with governments requiring them to achieve specific customer satisfaction targets. In this paper, we describe a project for the Dutch National Railways in which we identify the major determinants of customer satisfaction. By combining multiple data sources, we are able to link operational data and marketing data to customer satisfaction. The models show that punctuality and sufficient seating are important to satisfaction, as are other service elements such as the presence of Wi-Fi in a train and the condition of facilities at the station. Drawing on the model results, the Dutch National Railways has pursued initiatives to increase satisfaction. Among these initiatives is development of an app that allows passengers to check on seating availability, design of a marketing dashboard reflecting developments in customer satisfaction, and creation of a tool to identify the most critical determinants of customer satisfaction.

Keywords: Customer Satisfaction, Public Transport, Big Data,

Developing an Analytical Improvement System for the Dutch National Railways

The Dutch National Railways (NS) is the railway provider in the Netherlands. For years, the NS has been a monopolist running all public and business train connections (cargo transport). In the 1990s, reforms were introduced, resulting in a system of concessions, or large contracts with governments. As in multiple European Union countries, the Dutch public transport market is liberalized. In these liberalized markets public transport firms usually sign concessions that allow them to organize the transport for a specific time period (e.g., 10 years). The NS has been able to achieve concessions for the major connections between the large cities in The Netherlands (Major Railway Net), while small regional competitors, such as Arriva, Syntus and Veolia, run trains on smaller regional connections. These companies are only responsible for train operations, while an independent Government-owned organization (Prorail) is responsible for the train infrastructure. The NS now serves around 1.2 million customers daily, leading to almost 9 million unique passengers serviced during the year traveling in total around 17 billion kilometers per year. Despite the reforms, the Government remains the only shareholder of the NS, although the NS has its own governance structure that includes a top management team as well as an independent governance board. Nevertheless, within both the government and the Dutch parliament strong attention is directed at performance issues of the NS.

Importantly, as part of the contract public transport firms have with the governments, both the sales objectives and the delivered service are evaluated. This evaluation may include specific contracts on service levels (e.g., punctuality as measured by the percentage of delayed trains) and on the delivered customer satisfaction level. These contracts state specific objectives, such as 74% of the customers should give a 7 or higher (on a 10-point scale) on a specific customer satisfaction metric. If these objectives are not met a firm may be penalized, and in addition the firm's reputation may be damaged and negotiations for the next long-term

contract may be impaired. Firms may even lose the concession if customer satisfaction drops below an agreed minimum level. Hence, as strong drops in customer satisfaction can have severe negative consequences for the Dutch National Railways, achieving acceptable levels of customer satisfaction is very important.

At the outset of our project, NS had been confronted with multiple service crises over recent years. Major problems occurred during winter, in which many trains were not able to function, resulting in severe delays for travelers. This situation led to strongly negative media exposure and finally to serious decreases in customer performance (Bügel et al. 2011), which typically gains strong press coverage⁵ (see Figure 1). The highly negative effects of these service crises have been reported in the service marketing and customer satisfaction literature (e.g., Gijsenberg, van Heerde and Verhoef 2015; Smith and Bolton 1998). Not surprisingly, customer dissatisfaction also has negative financial consequences. If the firm cannot meet the performance level agreed upon in a concession, the company must pay fines of millions of Euros to the Dutch government.

Probably the most important driving factor of customer satisfaction with railway firms is punctuality. A recent study finds that punctuality and past satisfaction levels explain 56.8% of the variance of current satisfaction (Gijsenberg, van Heerde and Verhoef 2015). However, punctuality is sometimes hard to control owing to external unforeseen circumstances such as bad weather or accidents. Also, delays often result from issues around the railway infrastructure, for which in the case of the NS an external party (ProRail) is responsible.

In the past, the NS has always focused on punctuality. The NS has, for example, successfully implemented new train scheduling (thereby winning an Edelman Award for innovation), which resulted in improved punctuality as well as an annual improvement in profit (Kroon et al. 2009). The importance of customer satisfaction with punctuality and other

⁵ See for news item: <http://nos.nl/artikel/215838-onderzoek-trein-en-bus-presteren-slecht.html> (in Dutch)

elements of service is reflected in the statements of the board of directors: “Without customers we would be nowhere. Their satisfaction determines our success.” Therefore, given the NS’s close attention to customer satisfaction as well as the close government attention to the NS, a strong customer performance is extremely important for the top management of the NS.

< Insert Figure 1 about here >

Aim of the Project

In a boardroom session in Spring 2011, we were asked to analyze the major forces underlying the strong drop in customer satisfaction and to suggest potential remedies. The main objective of the NS was to get a better understanding of how customers evaluate their service. Within this overarching objective, the NS has three sub-objectives:

- (1) Establish the right metric for customer evaluations.
- (2) Assess the effects of determinants of customer evaluations.
- (3) Provide the top management of the NS with effective information on customer service as well as directions for improvements through a marketing dashboard.

Simultaneously this investigation should result in a customer investment model (internally referred to as KIM).

The Project

The project consists of four phases (Figure 2). In phase 1, which can be considered as a kind of pre-phase, we select the key customer metric. In phase 2, we aim to develop an individual-based big-data model using existing data within the firm that would allow the firm to assess the impact of multiple service attributes, thereby adopting the multi-attribute model. Thereby, we aim to benefit from new data sources within the firm that allow actual internal service

operations data as well as external data on social media to be linked to individual customers. This approach builds on studies within service marketing and customer loyalty, in which for example service operations data on handling of service requests can be linked to individual customer attitudes and behavior (e.g., Bolton, Lemon and Verhoef 2008). Hence, using a big-data approach we integrate several data sources to understand the impact of service operations on customer satisfaction. In phase 3, we apply a survey-based approach, where we collected data on customer satisfaction as well as use of services and service experiences. Finally, in phase 4 we ask input from managers as to their understanding of the impact of different service attributes. We combine the results of the phases to arrive at a presumed impact of service attributes on customer satisfaction and subsequently use our validated model to derive implications for the management of the NS to improve customer satisfaction.

< Insert Figure 2 about here >

Phase 1: Metric Selection

An initial required step in our project is the selection of the key customer metric. As noted, the NS collects multiple customer metrics over time, which induces ambiguity about the goals to achieve and ongoing debates as to which metrics to use. Although probably no silver customer metric exists (e.g., Ambler and Roberts 2008), from a firm perspective it is preferable to focus on a single accepted customer metric. The main metrics used here are customer satisfaction, net promoter score, and corporate reputation.

Customer Satisfaction

The NS measures customer satisfaction in trains on a daily basis with the item, “What is your general opinion/judgment about traveling by train?” Respondents answer this question on a ten-point scale (1 = could not be worse, 2 = very bad, 3 = bad, 4 = very inadequate, 5 = inadequate, 6 = sufficient/satisfactory, 7 = more than sufficient/satisfactory, 8 = good, 9 = very good, and 10 = excellent). This item is the official survey question for measuring service

quality judgments, and the NS has used it for years. The company's agreement with the Government stipulates that its contractually required performance criteria (including customer satisfaction) be evaluated with this question, with the firm having on average at least 74% of respondents providing a 7 or higher on this question. The NS therefore mainly focuses on the percentage of respondents scoring 7, 8, 9, and 10. We note that within the firm this metric is referred to as customer satisfaction, although it has components of both customer satisfaction and perceived service quality and therefore can also be labeled as measuring perceived service quality (Gijzenberg, van Heerde and Verhoef 2015).

Net Promoter Score

The net promoter score (NPS) is based on a single item assessing the likelihood (from 0 to 10) the respondent will recommend the firm to friends, colleagues, and others (Reichheld 2003). Using a transformation this metric is calculated as the percentage of promoters (score >8) minus the percentage of detractors (score < 7). This metric is measured on a monthly basis using online surveys. While the net performance score is not theory-based and is strongly debated in the academic literature (e.g., Keiningham et al. 2007), some recent evidence is more positive (e.g., De Haan, Verhoef and Wiesel 2015).

Corporate Reputation

This metric, referred to as the RepTrack metric, is collected on a quarterly basis by the reputation institute using an online survey of customers and non-customers. It is based on work of Fombrun and van Riel (1997), and is regarded as an important metric by the board of the NS (for more information, see <https://www.reputationinstitute.com/>). RepTrack is considered to be a higher level corporate metric that is less directly linked to service performance and value indicators such as customer loyalty.

Evaluation of the Metric

In selecting the metric, we apply criteria from prior work (Ailawadi, Lehmann and Neslin 2003) as adapted to the context of the Dutch National Railways. We summarize this evaluation in Table 1. This selection is made in 2010 and we use the literature available at that time. A clear advantage of the customer satisfaction metric is that it is theory-based, which clearly links the measure to business performance in regular markets and monopolies such as public services (e.g., Anderson, Fornell and Mazvancheryl, 2004; Bhattacharya, Morgan and Rego 2016).

Importantly, analysis of the application of all three metrics over time shows more robust data on customer satisfaction. Specifically, for NPS we observed relatively stronger and not clearly understood variations over time than we observed for customer satisfaction, making the NPS metric less diagnostic and reliable. One potential explanation for this result could be the transformation used, which is being debated as arbitrary (e.g., Keiningham et al. 2007). Specifically, for the NS many of the scores on the underlying question were between 6 and 8, inducing stronger variations (i.e., increased frequency of 6 making more customers detractors) in the official NPS score. Our analyses also strongly link customer satisfaction to service crises (Gijzenberg, van Heerde and Verhoef 2015). Importantly, as other stakeholders will find NPS and RepTrack less acceptable, the use of these two metrics will not satisfy the Government. Finally, the data presence of customer satisfaction is very strong, given that data have been collected for many years on a daily basis (with a sample size of around 60,000 customers per year). For NPS and RepTrack, the data frequency is lower and the sample sizes are much smaller.

On the basis of this systematic scoring of the metrics on these criteria, the board unanimously accepted the customer satisfaction metric as their key-customer metric. *This first achievement of the project is very important, as it brought to a halt the time-consuming internal debates concerning which customer metric to use.*

<Insert Table 1 about here>

Phase 2: Big-Data Model for Customer Satisfaction

A principal aim of the NS management is to understand how to influence customer satisfaction, as despite numerous studies top level executives still have no clear view as to how to approach this issue systematically. Given corporations' emerging attention to big data, we built on the movement to use more internal data by integrating different data sources and began to develop a big-data model.

Big-Data Model

As mentioned, the NS collects satisfaction data in trains at the individual customer level on a daily basis. Individual-specific questions are asked in the survey, and in addition the interviewer provides information on the train trip. The individual-level data can also be linked to specific operational data relevant for the train experience of each individual respondent (e.g., presence of a specific facility at a station visited). These data can be linked using the train number, date and time, as the NS has an operations database that includes train numbers, time of day, information on delays, stations visited, and so on. In our study, we also aim to account for social and media exposure effects.

In presenting our big-data model, we discuss up front specific categories of determinants that can influence customer satisfaction. This discussion relies on four presumed categories of determinants: train, station, service, and exposure. For each category we collect indicators from internal operational data sources as well as external data sources. As indicators we include the actual presence of events and not perceptions to overcome issues surrounding common-method variance. This approach is in line with recent research (Hamilton et al. 2016).

Our sample consists of approximately 144,000 respondents answering surveys between 2010 and 2012. Our main dependent variable is whether the customer responds with a 7 or

higher on the satisfaction question. We use this binary variable ($0 = \text{satisfaction} < 7$; $1 = \text{satisfaction} \geq 7$) because the firm's key performance indicator in its contract with the government is actually the percentage of customers giving a 7 or higher. A list of the independent variables and sources is given in Table 2. We integrate the data using two linking variables: (1) the traveled trajectory (from train station A to train station B) and the timing of the travel. Data sources include the survey, observations of the interviewer on some operational issues (e.g., cleanliness of the train); internal operational data on punctuality, stations, and so on; internal marketing data; and external data on social media presence. As we have large volumes of data, use unstructured data, and integrate multiple data sources, we consider this approach to be an analysis of big data (Verhoef, Kooge and Walk 2016).

<Insert Table 2 about here>

Model Approach

Given that our dependent variable is binary, we use a logit model to estimate the effects of the independent variables on customer satisfaction (Franses and Paap 2001). In principle, a normal regression model (or ordered logit model) is preferable, as satisfaction could be considered to be a continuous (or ordinal) variable. However, as the NS management clearly considers a 7 or higher to reflect successful customer performance, we treat customer satisfaction as a binary variable. Since we have a large number of independent variables that could also be correlated strongly, we first execute a principal components analysis using varimax rotation. This analysis results in a number of components regarding the cleanliness of the train, cleanliness of the station, exposure in media, marketing efforts, presence of shops and service points at stations, and the presence of parking and taxis at stations. Details of the principal components analysis are available from the authors on request.

In our data we also have missing values. Given the large number of data points, we delete cases that have random missing values on variables. However, for some variables the

missing cases are systematic, particularly for some operational data (e.g., stations). Here we include a dummy variable to indicate a missing observation. The effect of that specific variable is estimated only for customers having valid observations on that specific variable, while we also include the dummy in our estimation to account for a potential effect of the missing observation (Verhoef and Donkers 2001). In our model we also control for some specific effects regarding the reason for travel and whether a service crisis occurred in that specific period (Gijzenberg, van Heerde and Verhoef 2015). With regard to travel motive, the company distinguishes four main segments: leisure, work, business, and school. Importantly, the satisfaction scores differ per segment, with the leisure segment typically more satisfied than the other segments. We also control for two complaint variables: the presence of a complaint and the provision of money back to the customer.

Our dataset can be considered to be repeated cross-sections. An advantage of repeated cross-sections over a single cross-section is that customer satisfaction can be studied over time. However, individual customers cannot be followed over time. Satisfaction research has shown that past satisfaction influences current satisfaction – an effect conventionally accounted for by using a lagged satisfaction variable (e.g., Gijzenberg, van Heerde and Verhoef 2015). However, as these respondents are not observed over time, an individual lagged satisfaction term cannot be included. We therefore include average satisfaction at $t-1$ to account for lagged effects (Verbeek 2007; Moffit 1993). We specifically include the lagged effect per travel motive segment. The model description can be requested from the authors.

Model Results

The estimated model is significant, although the pseudo R^2 (Nagelkerke) is rather low with a value of 0.037 (McFadden pseudo $R^2 = 0.022$). The estimation results of the logit model appear in Table 3, which shows many significant variables – a common issue with the large

number of observations being analyzed. However, the Wald statistic suggests that some variables have a much stronger effect.

With regard to the control variables, we find expected negative effects of the occurrence of a crisis and two complaint variables. We also find expected differences between the travel motives, where the school and the work motives score lower than the leisure motive. In addition, we observe a carry-over effect, as the lagged average satisfaction variable is significantly positive.

Almost all included train variables are significant. Punctuality has the expected positive effect, whereas delays, the fullness of the train, and the number of complaints have the expected negative effects. The presence of good station facilities has a relatively strong positive effect. Interestingly, the service variables are not very significant. Finally, we find positive effects of social media. The majority of the dummies included to account for missing data are not significant.

To assess the importance of each of the four groups of determinants of satisfaction, we sum the Wald statistics per group of variables and divide by the sum of the Wald statistics. The results appear in Figure 3, which shows that the train factor is the most important determinant, followed by the station and service.

<Insert Table 3 and Figure 3 about here>

Evaluation and Problems Faced

With a very large database we build a big-data model that includes many potential determinants of customer satisfaction. We also assess which parts of the offered train service hold relatively strong importance for customers. Results show that some presumed determinants, such as cleanliness of trains and stations, are not particularly important, whereas others such as station facilities have a relatively strong relationship with satisfaction.

The outcome of this second phase also reveals three problems. First, given the occurrence of missing values, the integration of data is not perfect. Second, the R^2 of the model is relatively low, which can negatively affect management's acceptance of our model results. Third, existing data are limited in terms of scope. As a result, we believe that our model is not complete and we might have missed some important antecedents. In Phase 3, we therefore collect additional survey data to build a survey-based model.

Phase 3: Survey Based Model

On the basis of the big-data driver model, we now explicitly include variables that account for the profile of customers. We also use qualitative insights from studies on the customer journey to further develop the model (Lemon and Verhoef 2016). Thereby we also focus more on the inclusion of less concrete elements of the service that go beyond traditional dissatisfiers such as train delays and insufficient seating availability. Again we distinguish between variables related to the train and stations, and we include data on the atmosphere of stations. These data are separately collected in marketing research and we can include the average score per station. Service is now included under "train" since these variables also occur during the train trip. We include exposure again, but extend it to include the experiences of customers before and after their train trip. Importantly, we add a new set of variables related to pre- and post-transport to and from the train, which mainly focus on the accessibility of the train station. We control for external effects, such as weather and the occurrence of incidents.

A second change with regard to the prior model approach is that we execute separate analyses for two customer segments. We label the work, business, and school passengers as the commuter segment and the leisure travelers as the non-commuters segment in our analysis.

Data and Model

The main data source in this part of the project is a survey. Data are collected from train passengers over 8 weeks in the spring and summer of 2014. We ultimately collect data from 3,832 customers, deleting 106 observations owing to missing values on the satisfaction variable for a final sample size of 3,726 customers. In the survey, we ask the satisfaction question as before. Next, we ask factual information on their experiences. Rather than measuring opinions and attitudes toward specific service components, we ask customers about, for example, the delay of the train or the availability of sufficient seats. Although the survey is the major source of our data, we also consider other sources of data, such as operational data. Drawing on insights from the internal journey studies, we also include more variables on, for example, the time spent at the station.

To explain customer satisfaction scores higher than 7, we again use a logit model similar to the model that we used in phase 2. The general difference from the model in phase 2 is that we do not include a lagged term of customer satisfaction, as we analyze cross-sectional data. Moreover, we now estimate two separate models for the commuter segment (work, school, business) and the non-commuter segment (leisure).

One of the issues we faced is the large number of variables that potentially could enter our model. To solve this issue we first factor-analyze our data, which results in some meaningful factors. However, when these factors are included in our model none is significant. We then included all variables in a first model, and as expected many variables are insignificant. Next, we ran stepwise selection techniques multiple times, to arrive at a final set of variables that turned out to be significant predictors of customer satisfaction (e.g., Feld et al. 2013).

Model Results

Tables 4 and 5 show the results of our models for the two segments. As the tables show, the significant variables differ substantially per segment. The R^2 values of the two models also

differ, with a pseudo R^2 (Nagelkerke) of 0.183 for commuters and 0.141 for non-commuters (McFadden $R^2 = 0.114$ and 0.091). This result is a substantial increase over the R^2 of the big-data model, and could improve the acceptance of the developed model.

For the commuter segment, we find many significant train variables, with especially strong effects for variables related to punctuality and the absence of delay (arrival as planned). Interestingly, the fact that Wi-Fi service in the train is not working is a relatively strong dissatisfier. However, providing information on connections improves satisfaction, and our results confirm the importance of sufficient seating space. The presence of service personnel (i.e., the conductor) in the train does not create satisfaction, but decreases satisfaction when tickets are not checked, possibly because customers expect personnel to check tickets and to take a service role. Not surprisingly, waiting too long creates dissatisfaction. Hence, for commuters the station is an important element in creating satisfaction. Customers having relatively short trips to the departure station tend to be more satisfied, but if they have to seek too long for a parking spot they become less satisfied. With regard to the included exposure variables, our results show that customers searching for no information before the trip are more satisfied. Finally, we found some effects of the time of traveling (night) and the presence of windy weather. Also, customers who have no alternative to using the train are less satisfied, while within the commuter segment the customers with a school motive are less satisfied, similar to the model results in phase 2.

For the non-commuter segment, the number of seats occupied by unfamiliar people reduces satisfaction, and non-commuters also value empty seats. As with commuters, a non-working Wi-Fi creates dissatisfaction. With regard to stations, browsing and using a shop and a short waiting time again improve satisfaction, as does the atmosphere and presence of a building. For non-commuters the travel time after the trip to reach the final destination is important, as customers are more satisfied when they reach that destination faster. Direct

mailings positively influence satisfaction for this segment. Perhaps because non-commuters travel less frequently, communication from the firm tends to improve their relationship. Finally, we find again that windy weather reduces satisfaction.

We also calculate the average importance of each group of determinants per segment and their underlying sub-drivers. These results are shown in Figures 4A and 4B. For both segments the train determinants are most important, followed by the station, while pre- and post-transport as well as exposure are less important. However, the segments have some differences. For example, for the non-commuter segment, pre- and post-transport are not very important, while exposure is more important.

<Insert Tables 4 and 5 about here>

<Insert Figures 4A and 4B about here>

Evaluation and Problems Faced

The survey-based model is a richer model than the big-data model, as we are able to collect additional data. An important advantage is that we also include less concrete elements, such as the atmosphere at the station. This inclusion strongly improves the acceptance of this model within the organization. However, one disadvantage is that our data are purely cross-sectional and only cover a specific time period in a specific season that has more stable punctuality (i.e., it is not winter). Overall, the results confirm the importance of the train and the station and its underlying variables as important determinants. Given the richer nature of the cross-sectional model and the expected stronger acceptance of the model owing to the higher R^2 , we have chosen to use this model.

Phase 4: Input from the NS on Models

To further validate our model and gain a stronger acceptance, we ask for input from employees within the NS. We consider two target groups to further fine-tune our model: (1) experts within the marketing and market research and intelligence departments and (2) service employees on the train. The latter group is important because it is a key internal stakeholder within the firm and has day-to-day interactions with customers.

We ask for feedback on the model results from the experts because they understand the model results. The experts consider the majority of the results in the survey-based model to be as expected. However, they believe that the importance of seating was too low, while the importance of the atmosphere at the departure station was too high.

Using a different method, we ask for input from nine service employees. Using a color system, we show the importance of each of the factors (based on Figure 3) (not important and low, average, and high importance). We then ask them to rate whether the importance given to a specific factor is too low, appropriate, or too high. Their feedback suggests that the cleanliness of trains should be more important, while the importance of shops and facilities at the station is overrated.

The input from the management and employees from the NS confirms the results of the model, which will increase the acceptance of the model results and the implications derived from it.

Deriving Implications

The model results provide guidance for improving customer satisfaction and specifically the effect sizes of various determinants. However, the NS should not only consider the effect size, but also the improvement potential (e.g., L’Hoest-Snoeck, van Nierop and Verhoef 2015).

The improvement potential can be defined as the extent to which a current performance differs from the maximum performance. We calculate a potential improvement score by

dividing the maximum score by the current realized performance. Combining the potential improvement score with the effect size, we create a management framework that provides a clear direction as to which determinants would benefit from potential initiatives to improve satisfaction (see Figure 5). This matrix can also be used to assess existing ideas within the firm.

Within the NS twenty new ideas emerge, including an information solution for customers to check seating availability and a new service formula for shops. Using our approach we can clearly show that an information system on seating availability could be very useful in helping customers to find seating in a relatively full train. This initiative clearly links to the importance of seating in our model, and could provide important benefits to customers. The result is an app of the NS, through which customers can now check seating availability within trains.

<Insert Figure 5 about here>

Impact of the Project

The described project had a large impact on the NS and serves as the initial step in a much stronger focus on customer satisfaction within the firm, which is resulting in strategic programs to better serve the customer. In addition, the program has had multiple other direct impacts on specific marketing initiatives and selected metrics, as well as the role of the marketing intelligence function. We describe these effects in more detail below.

Marketing and Operations Impacts

The models result in specific marketing and service operation initiatives to improve customer satisfaction.

First, as an example of how our model results transfer to improvement initiatives, we discuss the introduction of the seating app. Customer satisfaction decreased substantially

during fall 2013. Complaints from customers and social media hinted that insufficient seating could have been the problem. Indeed, while the model results show that full trains with insufficient seating decrease satisfaction (see Tables 4 and 5), there also seems to be sufficient improvement potential. Initiatives related to seating availability end up in the upper right of Figure 5. As a consequence, a taskforce is now responsible for tackling the operational issues surrounding seating availability. As seating availability cannot immediately be solved because investing in new trains takes time, the NS has introduced a seating app that communicates to customers the seating availability of their chosen train (see Figure 6). Access to this information gives consumers a stronger feeling of control of their situation, which can also be helpful in feeling more satisfied. This example shows that the NS is attempting to understand the causes of decreasing satisfaction. By drawing on the insights of the model, the NS can assess whether assumed problems are really causing dissatisfaction. Initiatives can be developed to solve the problem and a potential impact of customer satisfaction can be assessed. In that way a kind of return on satisfaction can be calculated.

Second, the big-data model shows an impact of exposure on satisfaction and specifically an important role of social media. Negative mentions on social media account for 2.6% of the explained variation of the included driver variables in that model. A 10% improvement on the social media factor could potentially result in an increase of .36% in the satisfaction KPI. Based on this insight the NS invests in social media to improve the balance between positive and negative messages on Twitter. Third, the models clearly show the impact of crises or disturbances in the operations owing to external events (e.g., winter, problems owing to an external provider). The NS now creates plans for periods of crisis. For example, in the event of an expected crisis, the number of operating trains is reduced significantly. There is also extensive communication to customers on expected problems such as punctuality.

Organizational Impact

While the NS has measured customer satisfaction for years, this metric has been mainly used as a report metric with no strong focus on how to improve or predict customer satisfaction. Each month the board of the NS observes customer satisfaction but never pro-actively aims to influence that score. Moreover, the responsibility for customer satisfaction is very fragmented, with every department (service, trains, etc.) having responsibility for its customer satisfaction score. The presented project induces a much stronger focus on customer satisfaction within the firm accompanied by a focused and more centralized way of steering customer satisfaction. More specifically, it results in fact-based target setting, as well as customer satisfaction-based priority setting for specific initiatives to improve customer satisfaction involving multiple aspects of the service delivered in a customer's journey. In this respect the model is very useful in moving beyond traditional dissatisfiers (e.g., delays), and also in assessing the impact of usually less concrete satisfiers (e.g., WIFI in train, atmosphere at the station).

<Insert Figures 6 and 7 about here>

Owing to this project, customer satisfaction is now clearly a core metric for the NS and the NS is becoming more customer-focused (e.g., Shah et al. 2006). Importantly, the NS strategy gives customers top priority.

Business and Marketing Intelligence Impact

The project directly results in the development of a marketing dashboard with a strong focus on developing customer satisfaction (see Figure 7). This dashboard can be used to monitor and forecast customer satisfaction scores, as well as to gain insights in the potential impact of determinants of satisfaction.

An important result of this project from a marketing intelligence point of view is that it ends the ongoing discussion on the metrics to be used. Phase 1 of this project clearly resulted

in the choice of customer satisfaction as the major KPI. This KPI is now the core KPI in the concession agreements with the Dutch Government and in discussions with the government improvement initiatives that are based on the model insights. This KPI is now also foremost in NS International, which is responsible for train connections between the Netherlands and other European countries such as Germany, Belgium, and France.

The project also has an organizational impact on the marketing intelligence function. Before and during this project, two departments were responsible for gaining customer and market insights: marketing research and marketing intelligence. Marketing research is traditionally responsible for survey research, whereas marketing intelligence focuses more on analyzing available customer databases. This project resulted in cooperation between these departments, leading to their ability to present one story to the management of the NS. This successful cooperation motivates the board to merge these two departments into one intelligence function, thereby improving efficiency and effectiveness.

References

- Ailawadi KL, Lehmann DR, Neslin SA. (2003) Revenue premium as an outcome measure of brand equity. *J. of Mark.* 67(4):1-7.
- Anderson, EW., Fornell, C., & Mazvancheryl, SK. (2004). Customer satisfaction and shareholder value. *J. of Mark.* 68(4): 172-185.
- Ambler T, Roberts JH. (2008) Assessing marketing performance: don't settle for a silver metric. *J. of Mark. Man.* 24(7-8):733-50.
- Bhattacharya, A., Rego, L., & Morgan, N. (2016). Customer Satisfaction in Monopolies: Does it Matter? *Working paper*, Indiana University
- Bolton RN, Lemon KN, Verhoef PC. (2008) Expanding business-to-business customer relationships: Modeling the customer's upgrade decision. *J. of Mark.* 72(1):46-64.
- Bügel, MS., Verhoef, PC, Hoving-Wesselius, T, Wiesel, T, Bouma, JT Alleman, T. (2011). Dutch customer performance index: IKEA levert beste klantprestatie. *Tijdschrift voor Marketing*, 2, 44 - 48.
- De Haan E, Verhoef PC, Wiesel T. (2015) The predictive ability of different customer feedback metrics for retention. *Int. J. of Res. in Mark.* 32(2):195-206.
- Feld, S., Frenzen, H., Krafft, M., Peters, K., & Verhoef, PC. (2013). The effects of mailing design characteristics on direct mail campaign performance. *Int. J. of Res. in Mark.* 30(2): 143-159.
- Fombrun C, Van Riel C. (1997), The reputational landscape. *Corp. Reput. Rev.* 1 (1): 1:1-6.
- Franses PH, Paap R. (2001) *Quantitative models in marketing research*. (Cambridge University Press, Cambridge)
- Gijzenberg MJ, Van Heerde HJ, Verhoef PC (2015) Losses loom longer than gains: Modeling the impact of service crises on customer satisfaction over time. *J. of Mark. Res.* 52 (5): 642-656

- Hamilton RW, Rust RT, Wedel M, Dev CS. (2016) Return on Service Amenities. *J. of Mark. Res.* forthcoming
- Keiningham, TL., Cooil, B., Andreassen, TW., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *J. of Marketing*, 71(3), 39-51.
- Kroon, L, Huisman, D, Abbink, E, Fioole, PJ, Fischetti, M, Maróti, G, Schrijver A, Steenbeek A, Ybema, R (2009). The new Dutch timetable: The OR revolution. *Interfaces*, 39(1), 6-17.
- Lemon KN Verhoef P.C. (2016) Understanding Customer Experience and the Customer Journey. *J. of Mark.* 80(6): 69-96.
- Lhoest-Snoeck S, Nierop van E, Verhoef PC (2015) Customer value modelling in the energy market and a practical application for marketing decision making. *Int. J. of Elect. Cust. Rel. Man.* 9(1):1-32.
- Moffit R (1993) Identification and estimation of dynamic models with a time series of repeated cross-sections. *J. of Econometrics* 59 (1-2): 99-123
- Roberts, J.R. (2017), Marketing science in practice: Examples of Impact from Insight from the ISMS/MSI Gary Lilien Marketing Science Practice Prize, *Working Paper*, University of New South Wales.
- Reichheld, FF (2003). The one number you need to grow. *Harvard Bus. Rev.*, 81(12), 46-55.
- Shah D, Rust RT, Parasuraman A, Staelin R, Day GS. (2006) The path to customer centricity. *J. of Serv. Res.* 9(2): 113-24.
- Smith AK., Bolton R.N. (1998), An Experimental Investigation of Customer Reactions to Service Failure and Recovery Encounters: Paradox or Peril? *J. of Serv. Res.* 1 (1): 65–81.
- Verbeek M. (2008) Pseudo panels and repeated cross-sections. *The Econometrics of Panel Data* 46 (2): 369-383

Verhoef PC, Donkers B. (2001) Predicting customer potential value an application in the insurance industry. *Dec. Supp. Syst.* 32(2):189-99.

Verhoef PC, Kooge E., Walk N. (2016) *Creating Value with Big Data Analytics* (Routledge, London)

Table 1: The criteria for the election of the key-customer metric are theory-based. Each of three customer metrics is evaluated on these criteria and receives a score.

Criterion	Customer Satisfaction	NPS	RepTrack
Theory-based	+	-	-
Complete: multi-dimensional	-	-	-
Diagnostic for crises	+	+	-
Robust and reliable	++	+/-	+/-
Single number	+	+	+
Intuitive and trustworthy for top management	++	++	++
Intuitive and trustworthy for stakeholders (i.e. government)	++	-	-
Validated with outcome measure: customer value development	++	+	??
Based on existing data	++	+	+

Notes:

++ = very good score on criterion; + = scores well on criterion; +/- no clear score on criterion (not negative nor positive); - = negative score on criterion; -- = very negative score on criterion.

Table 2: The big-data model include multiple determinants. Per determinant several variables are considered. We describe these variables per category and show the source.

Determinant Group	Variable	Source
Train	Average punctuality on last 4 weeks on trajectory and in region	Internal operational
	Minutes delay on traveled trajectory	Observation interviewer
	Fullness of train (i.e., sufficient seating places) on traveled trajectory	Observation interviewer
	Number of complaints on full trains on this trajectory	Internal operational
	Cleanliness outside train good	Observations interviewer
	Outside train free of graffiti	Observations interviewer
	Facilities	Internal operational
	<ul style="list-style-type: none"> • Number of food shops • Number of non-food shops • Number of bicycle arrangements • Number of service and sales points • Number of other facilities • Taxi presence • Parking presence 	
	Cleanliness of station	Observations of interviewer
	<ul style="list-style-type: none"> • Hall • Platform • Waiting 	
Service	Timeliness of station	Observation of interviewer
	Information on delays	
	Information on delays within train	Observation of interviewer
Media and Marketing exposure	Presence and visibility of train employee within train	Observation of interviewer
	Number of negative social media mentions (excl. Twitter)	OXYME Market Research
	Number of negative media mentions	OXYME Market Research
	Number of negative twitter mentions	OXYME Market Research
	Presence in door-to-door print advertising	Internal marketing data
	Retail action	Internal marketing data
	Radio/TV advertising	Internal marketing data

Table 3: In our big-data model we use a logit model to assess the impact of different variables on customer satisfaction being 7 or larger. We show the estimated parameters, the accompanying wald statistic and the significance level (p-value)

Variable	Parameter	Wald	p-value
Constant	.488	2.385	.123
<i>Controls</i>			
Dummy crisis (yes =1; no =0)	-.167	51.406	.000
Dummy complaint (yes=1 , no =0)	-.515	800.115	.000
Dummy money back (yes=1, no=0)	-.084	15.157	.000
Dummy motive school (yes=1, no =0)	-.355	101.333	.000
Dummy motive work (yes=1, no=0)	-.167	26.482	.000
Dummy motive leisure (yes=1, no=0)	.050	2.390	.122
Lagged satisfaction per motive	.928	14.168	.000
<i>Train</i>			
Average punctuality on traveled trajectory	.133	33.104	.000
Minutes of delay on traveled trajectory	-.009	15.280	.000
Dummy on fullness of train (no seating availability =1, seating available=0)	-.204	43.155	.000
Number of complaints for traveled trajectory	-.003	57.571	.000
Factor train cleanliness	.0020	6.627	.010
<i>Station</i>			
Factor station cleanliness	.027	6.575	.010
Factor station facilities	.042	36.781	.000
Factor station taxi-car parking	.025	13.549	.000
Timeliness of information on delays on station	.062	1.492	.222
<i>Service</i>			
Information on delays within train	.054	2.025	.155
Presence/visibility of train employee	.016	4.270	.039
<i>Marketing and Media</i>			
Factor marketing	.017	5.728	.017
Factor (social) media	-.032	21.820	.000
<i>Dummies for missing's on specific variables:</i>			
Station cleanliness	-.027	3.173	.075
Train cleanliness	-.002	.005	.946
Timeliness of information on delays on station	.039	.950	.330
Information on delays within train	.031	1.167	.280
Presence/visibility train employee	-.040	2.425	.119

N=144.228, McFadden $R^2=0,022$

Table 4: In the survey-based model we use a logit model to assess the impact of different variables on customer satisfaction being 7 or larger for commuters. We show the estimated parameters, the accompanying wald statistic and the significance level (p-value).

Variable	Parameter	Wald	p-value
Constant	-.526	38.044	.000
<i>Train</i>			
Used time in train as planned	.302	38.601	.000
Punctuality of train on trajectory	.121	6.103	.013
Observed train personnel but did not check tickets	-.131	8.504	.004
Free space in train	.166	10.703	.001
Comfortable seating	.102	3.789	.052
Cleanliness of window	.097	3.627	.057
Wi-Fi tried but did not work	-.151	10.295	.001
Information in train on connection	.173	8.420	.004
Travel with intercity	.118	5.022	.025
<i>Station</i>			
Cleanliness of waiting rooms	.096	3.895	.048
Ambiance of station	.112	4.697	.030
Used shop	.119	4.476	.034
Looked around in shop	.133	5.732	.017
Waited for 2-8 minutes	.118	5.111	.024
Waited for more than 13 minutes	-.129	10.810	.001
On arrival station for 1-2 minutes	.189	16.351	.000
Use of service desk (1=yes, 0=no)	.122	5.262	.022
<i>Pre- and post-transport</i>			
Transport time from departure station to destination is 0-5 minutes	.147	9.667	.002
Transport time from departure station to destination is 5-15minutes	.159	9.746	.002
Parking time for car or bike was long	-.115	5.665	.017
<i>Before/after journey</i>			
Customer pays	-.110	3.626	.057
No information searched before trip	.162	9.736	.002
Two times contact about same complaint	-.137	9.949	.002
<i>External</i>			
Night	-.142	6.410	.011
Windy weather	-.096	3.740	.053
<i>Controls</i>			
Alternative for train or not (1=no, 0=yes)	-.168	10.314	.001
Travel motive (work =1, school =0)	.415	15.739	.000

N = 2683; McFadden R²=0,114

Table 5: In the survey-based model we use a logit model to assess the impact of different variables on customer satisfaction being 7 or larger for non-commuters. We show the estimated parameters, the accompanying Wald statistic and the significance level (p-value)

Variable	Parameter	Wald	p-value
Constant	-1.233	299.799	.000
<i>Train</i>			
Number of unoccupied seats with unfamiliar people	-.289	18.031	.000
Incidents on trajectory history	-.107	3.269	.071
Information in train on connection	.151	5.102	.024
Seating where customer liked	.127	4.058	.044
Wi-Fi tried but did not work	-.147	5.030	.025
Used time in train as planned	.116	3.395	.065
Witness or victim of incident in train or station	-.210	11.959	.001
<i>Station</i>			
Looking around in shops	.125	3.595	.058
Non-food shop available on destination station	.190	8.346	.004
Ambiance of departure station	.145	4.527	.033
Presence of building on departure station	.158	7.788	.005
Waited for 2-3 minutes on station	.082	3.055	.080
<i>Pre-post transport</i>			
Transport time from departure station to destination is 0-8 minutes	.145	8.761	.003
<i>Exposure</i>			
Used customer service	.136	3.671	.055
Received newsletter	.295	18.164	.000
<i>External</i>			
Travel outside rush hours	.149	4.663	.031
Weather windy	-.215	12.504	.000
N = 1043; McFadden R ² =0,091			

Figure 1: The satisfaction score and the operational performance over time for the NS. Satisfaction scores vary over time, but strongly decline when operational performance is very low, such as during specific winter periods.

Note: Operational performance is a combination of punctuality and the percentage of scheduled trains that were operating. A similar picture can be found in Gijsenberg, van Heerde and Verhoef (2015), that only considers punctuality and reports the average satisfaction level.

Figure 2: Process chart of the four phases of the project. We describe each of these phases with its' objectives, used data, used analyses and the resulting outcome.

Figure 3: On the basis of the estimation results of the big-data model, we calculate the relative importance per determinant group. Train variables have strongest impact.

Figures 4A and 4B: On the basis of the estimation results of the survey-based model we calculate the relative importance per determinant group and per segments (commuters vs. non-commuters). Train variables still have the strongest impact.

4A: Commuters

4B: Non-Commuters

Figure 5: On the basis of our model results, we can calculate the relative effect of specific initiatives related to specific determinants on customer satisfaction. We also can calculate an improvement potential. If we combine these two insights, this figure suggests selection of these initiatives in the top-right of the matrix. In total the management came up with 20 initiatives.

Figure 6: The NS introduced an app on seating availability, based on the insight that seating is an important determinants of customer satisfaction.

The screenshot shows a mobile app interface for reporting train occupancy. At the top, a yellow header bar contains the text "Cancel" and "Report train occupancy". Below this, three circular icons represent different levels of satisfaction: a green smiley face, an orange neutral face, and a red angry face. A blue tooltip points to the orange face with the text "Why?". Below the tooltip, there are three text input fields for reasons: "I had to stand", "I could find a seat", and "I could not board". The "I could not board" field is currently selected. Below these fields, the text "Your route" is displayed, followed by a vertical line connecting two station names: "Amsterdam CS" and "Utrecht C.". At the bottom of the screen, a large blue button labeled "Submit" is visible.

●●○○ KPN NL 00:33 40%

Cancel Report train occupancy

Why?

I had to stand I could find a seat I could not board

Your route

Amsterdam CS
Utrecht C.

Submit

Figure 7: This project resulted in a market dashboard that helps the managers of the NS to focus more on customer satisfaction and to understand how they can influence it.