PROFIT
BIG DATA
EXPERIMENTS

APPROACH AND
LEARNINGS

UNIVERSITY
OF APPLIED SCIENCES

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

The aim of PROFIT is to stimulate and facilitate SMEs in their innovation process. A very important starting point is the understanding of consumer behaviour and consumer needs. Unlike large companies, SMEs do not have access to large datasets about their customers, nor the skills to analyse and interpret consumer data. Moreover, SMEs in the tourism economy are strongly interconnected, as tourists combine products from different SMEs during their stay, like accommodation, restaurants, retail, leisure activities, travel et cetera. This means that data about consumer behaviour in the tourism economy is fragmented too, owned by many different parties. As part of PROFIT, we are therefore experimenting with (big) data from SMEs and other relevant stakeholders, to see if these data lead to more insight in consumer behaviour.

This work package has been led by HZ University of Applied Sciences, more specifically by the research group Data Sciences. Activities have been carried out according the CRISP-DM approach. CRISP-DM (Cross Industry Standard Process for Data Mining) is a data mining process model that describes the most commonly used approach for data mining. CRISP-DM breaks the process of data mining into six major phases. The sequence of the phases is not strict and moving back and forth between different phases is always required. The arrows in the process diagram indicate the most important and frequent dependencies between phases.

Business Understanding

This initial phase focuses on creating a better understanding of the need for information of SMEs and tourism destinations. We already have a basic understanding of those needs: how many people visit the destination, how do they move around the destination, can we recognize behavioural patterns, based on interests or other characteristics? But we do need to define these questions more specifically. This has been carried out by organising meetings with SMEs and stakeholders in each partner region, in March 2017. Looking at the outcomes of these meetings, business needs can be clustered into four topics: Motivation and experience, Visitor profiles, Visitor behaviour and Visitor expenditure. All these topics can be divided in many sub questions, that are strongly interconnected.

Data Understanding

The data understanding phase starts with an initial data collection: what data is available? We have tried to focus on data that could be available in all partner regions, to optimise comparability. In this phase, it turned out to be challenging to convince SMEs in all partner regions to contribute their data.

Data Preparation

The data preparation phase covers all activities to make the data ready for analysis and modelling. This involves the cleaning of data and making sure that data is in a format that can be compared to other available data. As always, this was the biggest job in the whole data mining process. Data of SMEs turned out to be very different, even when the source of data or type of data was defined beforehand.

Modelling

In this phase, various modelling techniques were selected and applied, to analyse the data and find patterns in data. This has been applied mostly to text data from reviews and booking data.

Evaluation

At this stage in the process, we went back to the stakeholders at each destination: do the findings answer the information need? This has been done in meetings with SMEs in several partner regions, during autumn and winter 2017, as well as in autumn 2018.

Deployment

The data and insights gained from review data have been organized and presented in a way that is useful to the SMEs and other stakeholders at the destinations via the PROFIT digital platform.
Data and insights gained from booking data have been analysed and presented through a personalised factsheet, which compares the data of one specific SME with data of a benchmark group of SMEs at the same destination.

This document describes the different data experiments carried out in Work Package I.
Information is given about the data in general, the available data used in the experiment and the following results. For each experiment, a short conclusion and evaluation is included, describing some lessons learned as well as specifying some terms and conditions in case of a continuation of an experiment at some point in the future.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 4 OF 30

WHAT IS FACEBOOK INSIGHTS?

A common way for a brand or business owner to communicate with their audience via Facebook, is through a Facebook page. Facebook Insights is a free module that, as the name suggests, gives insight into the interactions Facebook users have with a certain page.

WHAT KIND OF DATA DOES FACEBOOK INSIGHTS CONTAIN?

Facebook Insights provides a page's owner with specific metrics. These include:

Visits	Tells you how many people come to your page, view one of your pages or a page's tab and the amount of visitors coming in through an external referrer.
Likes	Gives insight into the total amount of page likes and the amount of new page likes for a specific time span. Additionally, it gives information about the percentage of page likes lost or gained in comparison to an earlier point in history.
Reach	Contains information about the number of Facebook users who got to see your page's posts while using Facebook.
Engagement	Facebook Insights provides information about users' engagement with your page and posts and this contains all the actions users performed on your page or post (actively interaction with instead of just passively viewing). Facebook Insights shows both positive engagement as well as negative engagement, where positive engagement is made up of interactions such as comments, likes and share and negative engagement is made up of interactions such as hiding, reporting or unliking.
Audience information	Facebook users that 'like' a Facebook page are called 'Fans'. A page's manager can get more information about the gender, age, language and location of their page's Fans. Facebook Insights also tells you when the page's Fans are usually online

Besides the metrics mentioned above, Facebook in general also contains interesting qualitative data: Facebook users can leave comments on posts or post about a topic, which can give valuable insight into motivations or experience.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 5 OF 30

WHAT KIND OF DATA DID WE HAVE?

Using Facebook Insights we were able to fetch data containing the amount of posts per business owner as well as the average amount of comments per post. Adding Facepager to the toolset allowed us to investigate the content of a page's posts and the replies on posts. We set up a manual for SMEs with instructions for collecting the data from Facepager. Based on the instructions, we gathered data for the last 500 posts and the comments made on those posts.

WHAT DID WE DO WITH THE DATA?

SMEs were sorted per region and we were able to gain some insights into the total number of SMEs per region, their combined amount of posts and the average number of comments per post. This gives some basic insight into the Facebook activity for SMEs in a specific region and the engagement their posts elicit based on the amount of comments per post.

Region	SMEs	posts	comments per post
Renesse	18	5263	3.4
Goes	14	4870	3.5
Medway	8	3350	4.3
Oostende	13	2846	4.2
Southend On Sea	5	2525	2.9
Sourc. de donnees	5	1889	4.1
Oostkapelle	9	1803	3.5
Nieuwpoort	7	2337	2.3
Source a lechelle	1	34	2.1

Besides that we were able to get the data from the actual posts using FacePager, allowing us to find the most used words in a post per region as well as give insights into which words are most commonly found together in a post.

WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

Facebook Insights offers many statistics and dashboards based on quantitative information, providing a SME with useful and informative visual summaries of their Facebook Page's performance in terms of people visiting the page and interacting with it.

Using Facepager we were able to analyse the content of the messages posted on a page, as illustrated by the examples above. Based on our analysis, however, the results did not prove useful and relevant information considering the questions of SMEs.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

Facebook Insights is a fully developed tool available to SMEs using Facebook as a means to promote their business. For continued use, simply providing a manual would suffice (if needed). Joint analysis for multiple pages proves to be difficult, since the content of every page is different and is set by its administrator. Individual use is recommended for SMEs who actively use Facebook as a means to promote their business and communicate with (potential) customers. A manual could be provided for those SMEs as to optimally use this tool.

Considering the somewhat disappointing FacePager results, it is not recommended to use in the future at this point. Facebook users responding to posts simply do not seem to communicate valuable information in regards to SME business questions. Sometimes replies to a post were made by the SME themselves, giving a distorted image of the content of replies to a post. If continued in the future, it would be wise to also include a linguistic expert, because sometimes words that seem to have the same meaning, make up most of the commonly encountered words in post replies.

BIG DATA EXPERIMENTS

UNIVERSITY OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 7 OF 30

WHAT ARE GOOGLE ANALYTICS DATA?

Google Analytics is a tool an SME can implement in their website, allowing for the collection of web statistics. It is a free service that helps users gain insight into the behaviour of website visitors. A piece of code (tracker) is implemented on the different pages of the website to gather visitor data. Website performance can be measured with Google Analytics and the tool offers visualisations for different kinds of unique as well as combined metrics. A website owner can choose to use more Google tools to integrate with Google Analytics, such as Google Ads, Data Studio and Optimize.

WHAT KIND OF DATA DO GOOGLE ANALYTICS DATA CONTAIN?

Google Analytics gathers information of visitors' paths through the website, such as the amount of time spent on the website, information on how a visitor reached the website and the visitor flow through the website. Besides that, it shows demographic details of the visitors to a certain degree, such as age, gender and interests.

WHAT KIND OF DATA DID WE HAVE?

To avoid having to request access to all the different Google Analytics account, a specific set of metrics was requested from the SMEs. A manual was distributed containing information on how to export Google Analytics data containing the right variables. SMEs were asked to upload data for the top 5000 most viewed pages of their website per month for the years 2013-2017. When starting data analysis, not all data was complete (missing values and years).

WHAT DID WE DO WITH THE DATA?
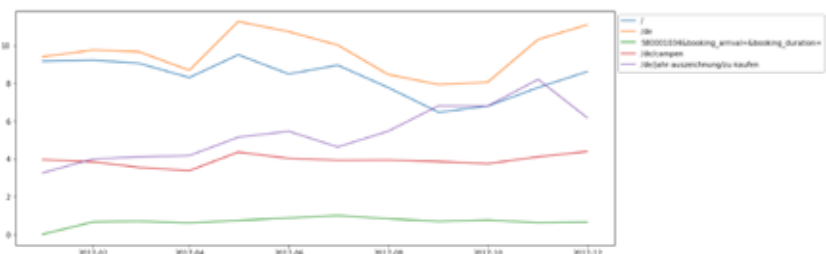
Using the requested data, we were able to gain insight into website activity throughout the year(s) as well as the distribution of visitors from different countries (see pie chart). Combining these variables, we were also able to plot

the number of visits from specific countries (for example the top 5 countries) throughout the year(s), allowing us to visualize the online behaviour for different countries.
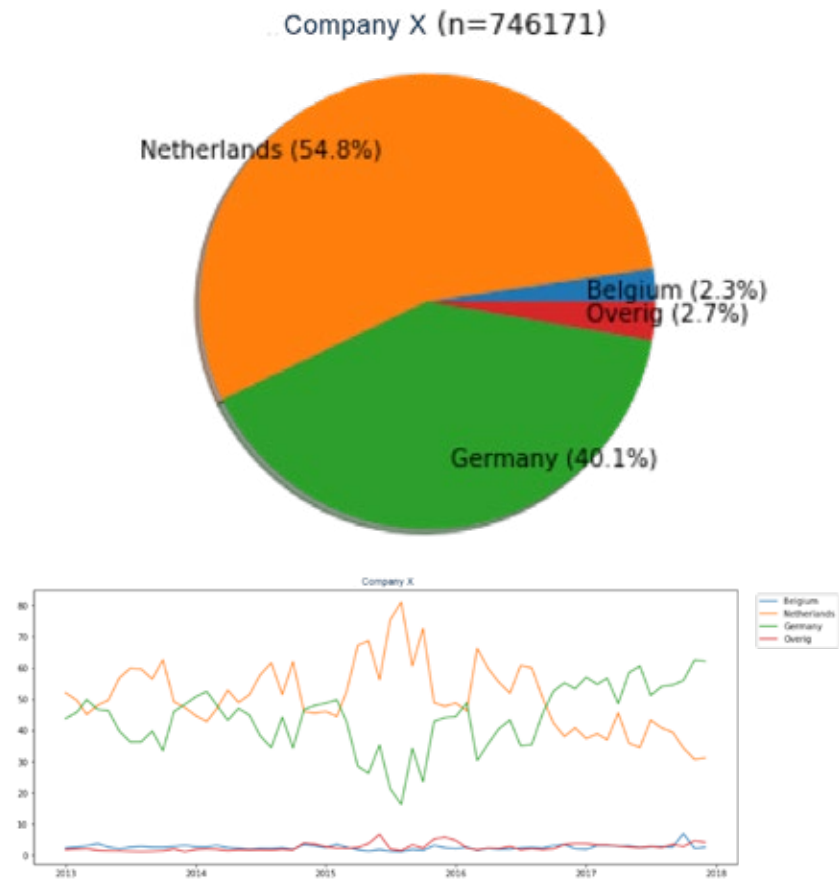
WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

One of the insights was that many of the SMEs do not use Google Analytics actively and to varied extents. Most SMEs did not know about all the different opportunities Google Analytics itself offers.

This experiment did not give information not already accessible. As mentioned before in the description of Google Analytics, the tool is aimed at giving users visual insight in website metrics. The generated graphs are already accessible via this tool.

There is no way to relate the website visitors to the actual customers who complete a booking. Some interesting patterns may be discovered (for example when website visits peak just before bookings peak, in general or for a specific country), but there is no way to claim with certainty that those are the same people.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

When a booking system is more incorporated into the website, allowing for detection of interactions via Google Analytics, a more certain statement can be made about the relation between lead time on a website and actually booking an accommodation. In the booking data there should be a variable containing information about where a customer made the booking.

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

WHAT ARE BOOKINGS DATA?

Data about bookings gives insights in different important metrics generated in the process where a guest books a room, camping site or cabin. The data is usually generated when a customer completes a booking online by entering and confirming the details of their stay as well as some personal information needed to confirm the booking. A different way the data can be gathered is when the guest registers and completes their booking right at the accommodation they are staying in. These data are saved via the reservation system the business owner uses either stored locally or in the cloud.

The bookings data used by the Data Science research group were requested with specific requirements.

WHAT KIND OF DATA DO BOOKINGS DATA CONTAIN?

Bookings data varies based on the booking software system used by SMEs, allowing for different kinds to data to be collected when registering a booking. Even for similar variables, the data format may be different.

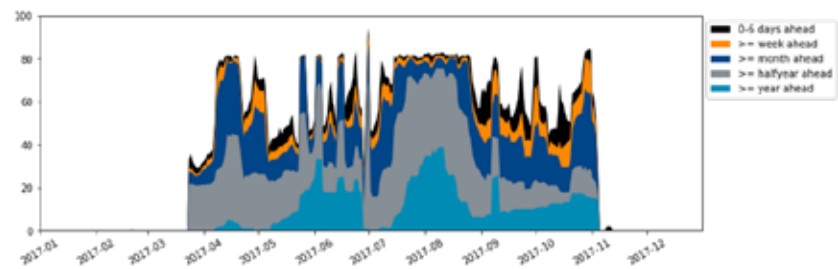|	
Booking date	The date the booking was made.
Arrival date	The date of arrival; first day of stay.
Departure date	The day of departure; last day of stay.
Number of guests	The amount of guests in one booking
Object type	Depending on accommodation, this could give information about the specific type of room, camping site or cabin that was booked.
Origin	Origin of the customer that made the booking (by country)
Postal code	Origin of the customer that made the booking (by postal code)

WHAT KIND OF DATA DID WE HAVE?

As mentioned before, the Data Science research group requested this data with the metrics as mentioned above. Data was to be uploaded in a web application and a manual was provided, giving step-by-step instructions on how to upload the right data in the requested format. We were not able to collect data from all SMEs and the SMEs that did upload their data were mainly located in The Netherlands.

We explicitly asked SMEs to not include personal information that can be traced back to the customers, to comply with privacy regulations.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 11 OF 30

WHAT DID WE DO WITH THE DATA?

These data allowed us to try and see if we could find patterns in booking behaviour. Using booking date and arrival date, we were able to see how many days visitors booked ahead and how this changes throughout the year.

With the inclusion of data for multiple years, we were able to give an SME insight in their performance when looking at the booking rate and how well they performed compared to the year before. Besides comparing to a previous year, we were also able to give a comparison of the booking rate as compared to the SMEs competitors.

Because the dataset also included the visitors' origin, we were able to distinguish different groups, allowing us to split a chart and providing more information on booking behaviour for customers of different origins.
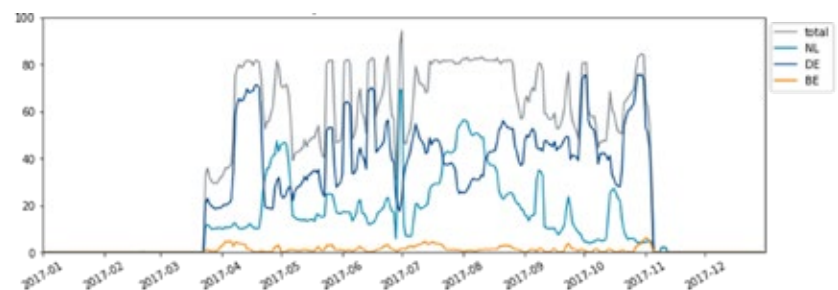
WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

This experiment used quantitative data to provide insight in booking behaviour throughout multiple years (ideally 2013-2017), also allowing us to give more detailed insights in booking behaviour based on customer origin.
The instructions provided for uploading the data ensured that we were able to do joint analysis and make comparisons between SMEs. This gives SMEs insight into certain patterns and specific points in time where they might seem to miss out (either based on their performance of previous years or based on the performance of competitors).

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

In this experiment, data were contributed by SMEs who understand the need for data analysis, who were willing to share their data and were able to export the data needed from their own booking software. This means we were working with SMEs that are not representative to all SMEs in the tourism industry. The participating SMEs are used to looking at their customer data, which means that some of the results of this experiment were not that surprising to them. However, some results were and provided completely new insights into their business. This type of analysis seems to be very valuable in providing new insights in customer behaviour for SMEs who are not that (digitally) skilled or simply do not have the means to study their data.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

Based on quantitative data like this, a number of different type of graphs can be generated, giving insight in different variables and how they may seem to relate to each other. When continuing this data experiment, more variables can be added to the data sets. Adding data needs to be done in accordance with the SMEs as to provide the most useful data insights. For this, SME engagement and commitment is essential.

It proved to be very difficult to get data from SMEs, making this a time consuming process. Sometimes SMEs did not know how to export their data, so further instructions were necessary. In terms of optimizing this part, manuals can be made for the booking software systems most commonly used. Another key point to be taken from this experiment is the timing. During 2018, SMEs were approached during peak season when they did not have time to contribute to this experiment. For a continuation, agreements have to be made beforehand to take into account the schedules and workload of SMEs.

When a file was uploaded containing the requested data, the format was usually different for every SME. We conclude that this is because of the different booking software systems used by the SMEs. Different formats required individual attention and effort in processing the files for joint analysis, which was also time-consuming for the data analysts. For optimal processing, it is important that the data upload process and corresponding manual is revised to ensure that uploaded files are as similar as possible in format. As mentioned before, not all SMEs use the same booking software system, therefore every data export each SME generates will probably differ in some way. If providing one and the same format is impossible, a realistic estimate for the duration of data processing needs to be taken into account. We conclude that it is not possible to automate this process, it will always include manual work.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

WHAT ARE GOOGLE REVIEWS DATA?

For companies listed on Google, people can leave reviews that show up when users of the Google search engine look for a specific company.

WHAT KIND OF DATA DO GOOGLE REVIEWS DATA CONTAIN?

Google Reviews contains a score on a 5-point Likert scale. In addition to giving a score, a reviewer can leave an actual review where Google asks the reviewer to provide information based on a their experience with the company. Google Review always contains quantitative data and sometimes it additionally contains qualitative data.

WHAT KIND OF DATA DID WE HAVE?

Google allows only 5 reviews to be fetched at a time using an API. Executing multiple requests allowed for more reviews to be fetched. For this experiment we focused on the qualitative part of the Google Reviews: the actual written out reviews. Our dataset contained the various written reviews for different SMEs. Reviewers are free to write whatever they want, but reviews were considered useful to answer questions related to visitor motivation and visitor experience.

WHAT DID WE DO WITH THE DATA?


PROFIT - Google Places Review Analysis

language: EN FR DE NL

Contrasts
Create a contrast between two places, by finding the most discriminative words. Click through to summarize what is said about a discriminative word.

Summaries
Create a summary of the reviews of a place. Provide a context word to select a subset of reviews to summarize.

Concepts
Visualize concepts and see the difference between
- SME's in a place
- places in a region
and select a SME or place to compare it to the others.

The data allowed us to gain insight into the most frequently occurring words as well as most frequently occurring combinations of certain words, allowing for the creation of summaries. We were able to determine topics or concepts for which we were able to compare different SMEs and different regions to each other. We were also able to detect words with either a positive or negative connotation, allowing us to get an idea of the overall sentiment of a review or a collection of reviews.

Automatic summaries could be generated based on all reviews in a destination, based on all reviews of a certain business type in a destination, based on all reviews of a specific company, based on positive or negative reviews or based on reviews that contain a certain word or topic.

Summaries could be formed in sentences, such as "A very nice place to stay - lovely building and quiet and clean - breakfast top-notch as well.". In addition to this, we were able to determine the frequency of the different concepts and visualize this for different destinations.
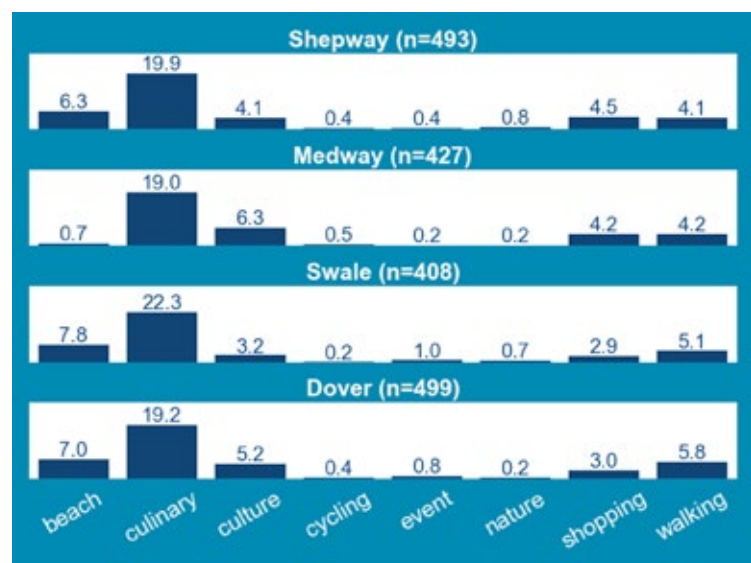
WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

For other experiments the different variables and metrics could be anticipated to some degree beforehand, for Google Reviews this was different because of the qualitative nature of the data. During data analysis and interpretation, patterns were discovered as the data understanding phase progressed.

It is a great opportunity for an SME to gain basic insight into the general opinion of their company (either positive or negative) and what seems to contribute to this. When looking at the different topics, SMEs can see how frequently visitors in their region talk online about a specific topic. This could give some indication about what motivates visitors to visit a specific region.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

Because the format of Google Reviews dos not require a reviewer to fill in specific information, the content and quality of the review may vary. These reviews seem great to gather more information about visitor motivation and visitor experience, but reviewers do not have to explicitly make a statement about either of these things. This makes it difficult to determine the value of the data beforehand, but the experiment was successful in generating new insights. When the template for the review is more structured based on SME questions, more valuable information could be gathered. Unfortunately, Google Reviews does not facilitate this.

Besides Google Reviews there are also other platforms that gather and present reviews, such as TripAdvisor, Booking.com and Zoover. These companies were approached, but since no collaboration could be established, we were not able to gather any more, additional, or more structured review data.

BIG DATA EXPERIMENTS

UNIVERSITY OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 16 OF 30

Cropland – research through femtocells & wifi-scanners

westtoer

WHAT IS THE SCOPE OF THE WESTTOER-CROPLAND RESEARCH?

The scope of the research was to detect and measure visitor flows and obtain more information on cross-visits in tourism POIs in 2 pilot areas at the Belgian Coast. The methodology used to detect these visitor patterns was analysing mobile phone data, Wi-Fi data and the combination of data from Wi-Fi-scanners and Femtocells. Mobile phone data was used to determine the profile of the visitor (origin and type). In addition, a study was included to determine whether anonymous Wi-Fi and mobile data can be linked without breaking the privacy of the users.

The aim was to provide at least 200 SMEs at the Belgian Coast with relevant information about their visitors and the different locations they visit during their stay.

Westtoer cooperated with an external supplier 'Cropland' in this research. Cropland is a company that specializes in data driven solutions. It applies advanced pattern analyses and correlates various data sources with each other in order to gain new insights.

WHAT ARE FEMTOCELLS AND WI-FI-SCANNERS AND WHAT KIND OF DATA DO THEY CONTAIN?

Femtocells and Wi-Fi-scanners are technologies mainly used for creating complementary networks, providing better connectivity and are often used in situations where local cell reception may be limited or unavailable.

Both Femtocells and Wi-Fi-scanners are devices that register the presence of mobile phones.

WHAT KIND OF DATA DID WE HAVE?

We installed 10 Femtocells and 20 Wi-Fi scanners in different types of tourism SMEs/ POIs: hotels, restaurants, beach clubs, attractions, train stations, holiday centres, shops … We chose two pilot areas, two cities at the Belgian coast: Oostende and Nieuwpoort. In this specific experiment, we could also use the data from Wi-Fi scanners that were already present in the pilot areas.

Devices, like mobile phones carried by visitors, always connect with the strongest network, in this case the Femtocells and/ or Wi-Fi-scanners (if the Wi-Fi is on). The Femtocells and Wi-Fi-scanners provided us with data in the different locations.

BIG DATA EXPERIMENTS

UNIVERSITY OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 17 OF 30

▶ The reason we used Femtocells, is because they can give us information on a specific location (on cell level), so we are able to detect if a visitor is actually present in a specific restaurant or hotel.
▶ Wi-Fi scanners give us information on anonymised, aggregated personal data (within the framework of the Belgian privacy law).

WHAT DID WE DO WITH THE DATA?

We wanted to see if we could detect and analyse specific visitor flows at the Belgian Coast by using data from Femtocells and data from Wi-Fi scanners. The Wi-Fi data was used to detect the visitor flows, the data from the Femtocells to establish the link with the mobile phone data.

Westtoer and Cropland have made agreements to determine when a mobile phone was considered "present" at a certain location and agreed on definitions for visitors by type (locals, day visitor, long term visitor).

During the study, the question whether anonymous Wi-Fi and mobile data can be linked without breaking the privacy of the users, was answered negatively: it appears that it is not possible to link the IDs between Wi-Fi and mobile data. This resulted in creating new definitions for visitors by type based on the Wi-Fi data. This proved to be quite challenging: defining different types of visitors depends from location to location, additionally both the positioning of the scanners and the type of POI, are determining factors for those definitions.

Cropland combined all the data and created reports on the visitors based on the Wi-Fi data obtained.

The analyses gave us information on

▶ Number of visitors per location
▶ Origin of the visitors
▶ Visitors by type (locals, day visitor, long term visitor)
▶ The starting point of the visit and in 2-3% of the cases information on a second location the visitors visited during their stay

All data was anonymised and aggregated in line with the Belgian privacy law.

WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

The combination of Femtocells and Wi-Fi-scanners did not provide us with enough relevant data sets to draw conclusions. We can detect some movements in a few cases, but the information is limited.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 18 OF 30

What we learned from the results based on data from Wi-Fi-scanners:

▶ Problems with adjusting the device + tuning the number of decibels (-70dB / -65dB)
▶ Problems with determining correct definitions (converting figures to people)
▶ Difficult to determine and control the range of the devices
▶ Devices cannot always be installed in the right place (approval, technical provisions (socket and Internet connection needed) ...)

We notice that there is still too much noise on the data sets and results; e.g. the devices capture more people than actually present.

The (small) scale of this experiment/ research is a disadvantage. In order to have good results and to detect visitor flows in tourism businesses, you need to be able to include enough POIs in the pilot. In this pilot, the data sets received were limited due to the number of locations.

This research is too expensive in comparison with the results delivered. The installation costs are too high to include more businesses in the research and upscale the experiment.

Per location the (time) investment is too big. You have to work on an individual level, check every device separately and adjust it to the location and situation to obtain correct results. Defining different types of visitors should be done by POI, which would be a very time-consuming exercise (apart from the investments for placing the devices).

We learned a lot from this experiment and it has had its value, but the investment is too big to continue or expand it in the same way.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

The opportunities we see for the future:

▶ We have noticed a great interest among our SMEs to obtain more information on visitor flows and cross-visits. This information can be interesting for SMEs for
  ▷ Product development
  ▷ Detecting what arrangements/ collaborations with colleagues work in a specific period and if they need adjustment
▶ Wi-Fi research can be very interesting and can have added value for certain SMEs if the device is installed correctly and appropriate definitions are applied to the data.
▶ The Wi-Fi data did not provide us with enough added value to detect cross-visits on a large scale, so we are looking for other methodologies to do more research on cross-visits in tourism POIs.

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

WHAT IS THE SCOPE
OF THE WESTTOER-TELENET PILOT?

The scope of the pilot was to find a methodology and concept with technology that can provide tourism SMEs with relevant business info on their visitors and visitor flows. The methodology

▶ Needs to be financially feasible
▶ Should be easy to upscale to a big group of POIs/ SMEs
▶ Should not ask for a major installation cost or time investment

The focus of this pilot was building a self-service platform for business owners. A platform that shows SMEs analytics dashboards containing various insights on his business, which he can use to improve profitability. The platform will only provide insights and no suggested actions, how the owner chooses to act on the provided insights is completely up to him. Telenet and Westtoer worked together on this to make the platform user-friendly and fit the needs and expectations of business owners.

WHAT KIND OF DATA DOES
THE TELENET PILOT CONTAIN?

Telenet Group is the largest provider of cable broadband services in Belgium. Its business comprises the provision of analog and digital cable television, fixed and mobile telephone services, primarily to residential customers in Flanders and Brussels. In addition, Telenet offers services to business customers all across Belgium and in Luxembourg under its brand Telenet Solutions.

In this pilot, Telenet used the mobile phone data captured via the Telenet WiFi network.

WHAT KIND OF DATA DID WE HAVE?

The deliverable of the pilot is a mock-up dashboard containing real data. The goal of this dashboard is illustrating the possible capabilities.

The data used in the dashboard only use WiFi logins, because a granular level of localization is required. For example, data from the mobile network is accurate up to 50 m, which is often not accurate enough for restaurants or cafés. For WiFi, a user must have a maximum distance of 10 m from the modem to be able to log in.

Telenet used the real data from existing technologies (modems) in their network in more than 50 POIs/ tourism SMEs along the Belgian Coast. In this experiment as well, all data was anonymised and aggregated in line with the Belgian privacy law.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 20 OF 30

WHAT DID WE DO WITH THE DATA?

Telenet combined all data captured and presented it in a user-friendly dashboard with nice and easy-to-read graphs. The dashboard consists of different panels providing information on visitor trends, visitor origin and cross-visits.
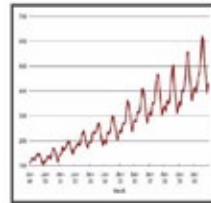
Visitor trend   Visitor origin   Crossvisits

Dashboard

▶ Visitor overview
  ▷ Number of visitors by day (combined with weather data)
  ▷ Recurrence of visitors
  ▷ Visitors by type (locals, day visitor, long term visitor)

▶ What other places did my visitors go to?
  ▷ Accommodation, restaurant and/or attractions (absolute and relative numbers) including a map to situate the locations

▶ What other places did my visitors go to in other cities?
  ▷ See above

BIG DATA EXPERIMENTS

UNIVERSITY OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 21 OF 30

▶ What are visit patterns depending on different factors?
   ▷ Popularity of visited location during weekends or midweeks
   ▷ Is a location more popular in a certain season?

▶ Where do my visitors come from?
   ▷ Number of visitors (absolute and relative numbers) including a map to situate the origin of the visitors

In each tab/ panel the user has the possibility to select a specific date range and can select weekends versus midweeks.

WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

This pilot proved to be a very valuable and successful experiment. The dashboard created by Telenet presents a lot of information on visitor typology, trends and cross-visits and is very user-friendly.

Telenet was able to perform an experiment:

▶ By using the existing technologies within the tourism POIs/ SMEs (no installation cost)
▶ Without burdening SMEs with any form of investment
▶ With the possibility to upscale to a bigger group of POIs/ SMEs

We tested the dashboard with a number of SMEs to check the added value for the sector and received a lot of enthusiastic reactions.

We noticed a significant difference between visits during weekends and midweeks and would like to further investigate it.

We are now looking into several possibilities to dive deeper into this experiment and further develop the dashboard.

BIG DATA EXPERIMENTS

UNIVERSITY OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 22 OF 30

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

We asked Telenet to further investigate the possibilities of the dashboard.

- ▶ We would like to do additional research and look into the possibilities of upscaling this model.
- ▶ We can conclude that the difference between visits during weekends and midweeks is very relevant to further investigate and we want to look into a more detailed model to specify the difference between weekends and midweeks, taking into account holidays and long weekends.
- ▶ We will inform our SMEs on the possibilities of the dashboard and the whole research and will define & suggest concrete actions for our SMEs based on the results and insights gained.

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 23 OF 30

WHAT IS WEATHER AND LAST MINUTES BOOKINGS DATA

In this experiment we tried to see if we could find information about the effect of weather on last minute bookings. A last minute booking was defined as a booking made the day of arrival or the day before arrival. For weather data we used data from the Royal Netherlands Meteorological Institute (KNMI) for the arrival day. We argued we could use the actual weather instead of weather predictions, because of the likely high accuracy of the prediction.

What kind of data does weather and last minute bookings data contain? Weather data contains different kinds of information regarding the weather, such as temperature and amount of rainfall (in millimetres). Last minute booking data is equivalent to booking data. As such, last minute booking data contains information about the dates concerning the booking (date of booking, arrival date, departure date), the number of guests, detailed guest information and detailed accommodation information.

WHAT KIND OF DATA DID WE HAVE?

Weather data

For weather data we relied on data from the Royal Netherlands Meteorological (KNMI) data. From this data we used temperature and rainfall as indicators for weather. The KNMI does not provide information about weather predictions. However, we argued that since we were only focusing on lastminute bookings, the weather report for the arrival date on the booking date would be highly accurate. As such, we used the weather information for the arrival dates, trusting that this information would not deviate from the weather predicted on the date the booking was made.

Last minute bookings data

For last minute bookings data we used the same data as bookings data, but only for bookings that we consider last minute bookings. For last minute bookings we narrowed down the dataset to bookings made a maximum of two days before the arrival time (maximum lead time of two days)

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 24 OF 30

WHAT DID WE DO WITH THE DATA?

For this experiment we tried to see what different variables had effect on the likelihood of making a last minute reservation.
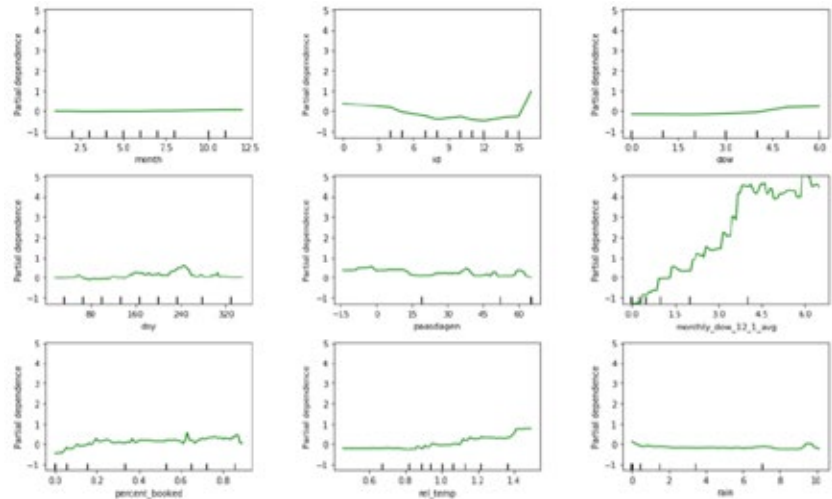
	Dependency	Variable
1	0,29	Bookings made in same month of previous year
2	0,20	Occupancy rate
3	0,16	Bookings made for same day of previous year
4	0,12	Relative temperature (day temperature / average temperature month)
5	0,07	ID
6	0,05	Amount of rainfall in mm
7	0,05	Easter holidays
8	0,01	Month of booking

Different models were created and comparisons were made based on how accurate they were in predicting last minute bookings. A model either did not use weather variables, one of the weather variables, or a combination of more than one weather variable.

WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

This experiment concluded that the model using only rainfall as a weather variable was best in predicting last minute bookings. In second place came the model using only relative temperature as a weather variable and lastly the model that used the combination of both rainfall in mm and relative temperature.

The models either predict accurately (error of 0) or they predict wrongly, which would result in a prediction that deviates from the real number of lastminute booking by 1 or 2 guests. The differences in performance between

the different models turned out to be very small. The "best" model (the model using only rainfall) predicted last minute bookings inaccurately by an average of 0,88 guests (73%). The worst model when it comes to predicting last minute bookings with weather data (the model using both rainfall and temperature) made an average inaccurate prediction of 0,92 guests (75%).

Other interesting observations include the fact that a relatively colder temperature leads to less last minute bookings. The average number of guests booking last minute is usually 1 or two (with the occasional outlier up to 6 guests). Last minute bookings seem to occur most frequently starting the month of September.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

Based on the data explored we did not find anything particularly surprising. Weather does seem to be a attributing factor in last minute bookings, but the amount that can be attributed to this is not the most determining factor.

The models that have been trained to determine if we can predict last minute bookings based on weather variables (rainfall and relative temperature). The experiment does not seem to give any new or surprising insights that can be used by SME's to adjust their business strategy for last minute bookings. Therefor we would not recommend a continuation of this experiment.

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

WHAT IS SEARCH AND BOOKING BEHAVIOUR?

In this experiment we tried to see if it was possible to gain insight into the online search behaviour of website visitors and their actual bookings. We defined online search behaviour as the keywords and search strings people used in search machines.

WHAT KIND OF DATA DOES SEARCH AND BOOKING DATA CONTAIN?

Ideally, we would have access to search information: keywords (or strings), information about the search machine used, and other informative information about the user's search, such as device used to make the search. For Booking behaviour, ideally we would have access to information about which online users make a booking. To complete this experiment, we need to be able to combine information about search behaviour with information about booking behaviour for specific users.

WHAT KIND OF DATA DID WE HAVE?

Acquisition

Google Analytics does provide information about so called 'acquisition': the way websites visitors 'find' the website. Different types of acquisition can be distinguished: Organic Search, Referral, Social, Direct, Paid Search and E-mail.
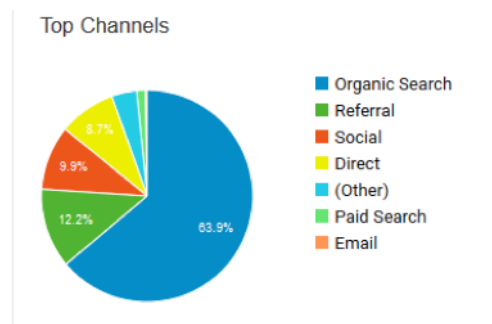
Image 1   Distinction between the different ways (channels) website visitors find the website.

▶ **Organic Search**: Visitors who find the website through the use of a search engine.
▶ **Referral:** Visitors who find the website through another website.
▶ **Social:** Visitors who find the website via a social media network such as Facebook.
▶ **Direct**: Visitors who find the website by directly typing in the URL in their browser navigation bar.

▶ **Paid Search**: Visitors who find the website through paid advertisements (like Google Adwords).
▶ **E-mail**: Visitors who find the website using the clickthrough link in an email(campaign) they received.

We can zoom in on the different channels. The channel 'Organic Search' is especially interesting in the case of searching and booking behaviour: visitors use search engines by typing in certain words and find an SME's website. Google Analytics provides some insight in the different keywords used in these searches. The overview below gives an example of what this might look like:

Image 2 Keywords used in searches, sorted by frequency

Keywords are sorted by the amount of times they have been used (see Image 2). It is notable that the number one keyword is defined as "(not provided)" and it has a very high frequency (94% in the example above). Google secures the searches made by their users because of their privacy (Google, 2011). Consequently, these searches are no longer visible in Google Analytics, making the practical application of this experiment very difficult.

Conversions

Google Analytics allows for the tracking of specific occurrences, called conversions (Image 3). Usually these also included things like creating an account or subscribing to a newsletter, but these may also include making a booking or buying a product or service. The customer journey towards these conversions can be visualized, but Google Analytics does not show the different search words these visitors use if they used organic search to arrive at the website. Within the website it allows an SME to see how their website visitors navigate through the website, but it does not show any information about those visitors when their location is outside of the SME's website.
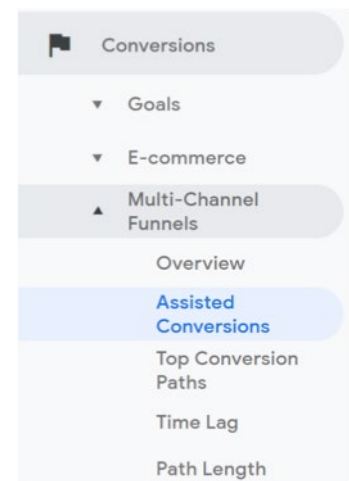

Image 3 Conversions in Google Analytics

Google Trends

Google Trends (https://trends.google.nl/) gives insight into key words used in search. An overview is generated visualizing the popularity of the key word or combination of key words (based on the frequency). When a key word or a combination of key words passes a certain frequency threshold, Google Trends also shows information about the frequency of that search for different regions. It also allows you to compare different keywords with each other.
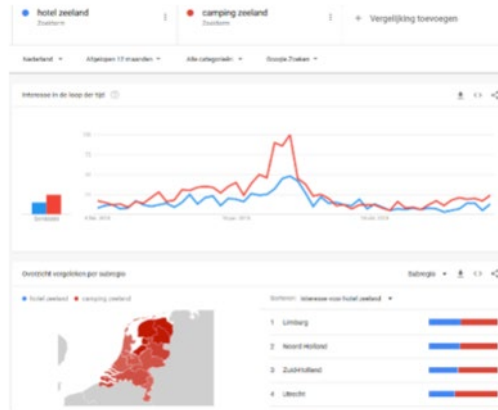

Image 4   Comparing two searches in Google Trends

Google Search Console

The Google Search Console is mainly interesting from the viewpoint of Search Engine Optimization (SEO). It offers insights in the performance of a website when searching for specific keywords (based on ranking) (Google, n.d.). Google Search Console mainly gives information about the clickthrough rate (CTR) en not necessarily which ones of these actually lead to conversions (like making a booking). It does allow an SME to gain more insight into the performance of their website or a specific page for certain key words used in search. These key words are not linked in any way to individual users (for example through the use of a User ID) and for this reason, this data cannot be linked to information about bookings (conversions).

WHAT DID WE DO WITH THE DATA?

For this specific experiment different types of data exist. However, due to the limitations Google puts on the data, is it not possible to combine these different data sources in a meaningful way in regards to the definition of this experiment.

WHAT IS THE CONCLUSION OF THIS EXPERIMENT?

An SME can gain insight into booking behaviour using the conversions module in Google Analytics to see the customer journeys of their digital visitors. For search behaviour an SME can use Google Trends, Google Search Console and,

BIG DATA
EXPERIMENTS

UNIVERSITY
OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT

to a very limited degree, Google Analytics. Is it not possible to combine these different sources to make a meaningful conclusion about the relation between search behaviour and booking behaviour.

WHAT ARE SOME OPPORTUNITIES FOR CONTINUED USE OF THIS METHOD?

If an SME wants to use these different types of insights to optimize their website, it is absolutely possible. However, Search Engine Optimization (SEO) and website optimization are professions in their own right. Because of this, we would recommend consulting experts in these fields to support in these optimization processes.

SOURCES

Google. (n.d.). Google Search Console. Retrieved February 1, 2019, from https://search.google.com/search-console/about

Google. (2011, October 18). Making search more secure. Retrieved February 15, 2019, from https://googleblog.blogspot.com/2011/10/making-search-more-secure.html

BIG DATA EXPERIMENTS

UNIVERSITY OF APPLIED SCIENCES

Interreg
2 Seas Mers Zeeën
PROFIT
European Regional Development Fund

P. 30 OF 30