UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Customer Journey Dutch Employee Insurance Agency

Master in Data Science, FIB
Process-Oriented Data Science
BPI Challenge 2016

December 20th, 2022

Group 9
Ander Barrio Campos - Nora Hérault - Odysseas Kyparissis - Perrine Lafaye - Emma Salvan

Table of Contents

Introduction

Every organization is structured on a set of processes that define its operation. In order to be able to manage their processes, organizations use log reporting mechanisms which keep track of all processes' steps and create logs describing their status. In that way, organizations can create models allowing them to analyze and continuously improve the organization's operations. The Process-Oriented Data Science course at the FIB-UPC focuses on how data science can significantly improve the way organizations manage and improve their business processes.

Since process mining is a data driven research field, real-life datasets have always been a cornerstone of the work in this field. The Business Process Intelligence Challenge (BPIC) convinces companies to share their datasets publicly and researchers to actively use these datasets. This challenge provides participants with a real-life event log and asks them to analyze these data using whatever techniques available, focusing on one or more of the process owner's questions or providing unique insights into the process captured in the event log.
As part of the annual conference of Business Process Intelligence, the sixth International Business Process Intelligence Challenge (BPIC'16) was organized using public data from UWV. The UWV is the Dutch government agency that is responsible for the distribution of unemployment benefits and for the assistance of unemployed in finding a new job. UWV's goal for BPIC'16 is to improve the customer experience provided to Netherlands citizens.

Every business must be customer-centric to sustain in the competitive market. Investing in customer support will help businesses to understand their complaints and improve their services and experience. Consequently, this paper will focus on providing general insights about UWV customer's demographics and their general behaviors while using the organization's services, analyzing the changes in usage patterns of the website service over time. Finally, special attention will be paid to the customers' complaints. The different age categories, as well as the gender of the customers, are considered while tackling the above domain topics. Thus, the research questions addressed will be the following:

1. Are there clear distinct usage patterns of the services to be recognized? Insights into the way various customer demographics use the services are of interest.
2. Do the usage patterns of the website by customers change over time? How does the usage change over time?
3. What are the general insights about complaints over time by age category and gender?
4. Does the behavior of the customers change after they have filed a first complaint? Are customers more likely to use these channels again after they have complained for the first time? Do complainers enter the website afterwards?
5. Are there any insights regarding the frequency of a user complaining?

This paper is organized as follows: Section 2 gives a complete description of the context and the datasets. Section 3 elaborates the methodology taken, detailing data preprocessing and the tools used. Sections 4 discuss the analysis results for general insights, website usage patterns and complaints. Finally, the paper ends in Section 5 with conclusions and recommendations.

Context

Case study

The Ministry of Social Affairs and Employment (SZW) has given the UWV (Employee Insurance Agency), a mandate to execute employee insurances and offer labor market and data services in the Netherlands. UWV is searching for insights into the customer journeys, and the data in this collection relates to customer contacts during an 8-month period from July 2015 to February 2016. Customers receiving WW (unemployment benefits) are the object of this study. UWV collected data from several sources. They refer to clickdata from the site www.werk.nl collected from visitors that were logged in and not logged in, Werkmap message data, showing when customers contacted the UWV through a digital channel, call data from the call center, showing when customers contacted the call center by phone, and complaint data showing when customers complained.

All data are complemented by fields including anonymized client information, information about the website visited, the nature of the call, and/or information about the complaint. The data was collected based on the site's structure at the time and visited URLs are included. UWV is interested in learning more about how their channels are being utilized, when customers switch from one contact channel to another and why, as well as whether it is possible to identify distinct customer profiles in the behavioral data.

Datasets Description

For this project, 5 datasets were available and are presented in Table 1. Due to the limitation of the assignment's size, it was decided to focus the analyses on four of them. The ones used are *Logged_In*, *Complaints, Werkmap_Messages* and *Questions*. The metadata file describing the content of each dataset's variables can be found by following this link[1].

Table 1 Datasets Description

Name of the dataset	Description	Volumetry
BPI2016_Clicks_Logged_In	Interaction data from registered customers who have logged into the website	7,174,934 rows
BPI2016_Complaints	Dataset that identifies all the complaints filled by the customers	289 rows
BPI2016_Werkmap_Messages	Data that are from the Werkmap system that is used as an internal instrument to keep track of customers' obligations	66,058 rows
BPI2016_Questions	Informations about questions asked by the customers	123,403 rows

[1] Metadata file

Methodology

This section includes the methodology steps which were followed, as well as the tools used during the implementation of the project. Data preprocessing and creation of new datasets were necessary in order to deal with the aforementioned questions of UWV. In addition, process mining techniques were used in order to provide answers to those questions. All steps are described in detail in the next subsections.

Tools Used

In order to preprocess the chosen datasets (*Complaints*, *Werkmap_Messages*, *Questions* and *Logged_In*) *Python*[2] and *Jupyter Notebooks*[3] were used. *Jupyter Notebook* is the original web application for creating and sharing computational documents. It provides a straightforward, efficient, document-focused experience. In that way, users can configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning using the interface's flexibility. *Jupyter Notebooks* are based on *Python* which is a computer programming language essentially used to conduct data analysis (cleaning, analyzing and visualizing data). It is a state-of-the-art language used for data science solutions, due to the existence of several libraries and frameworks focusing on such topics. Finally, *Python* is capable of managing a very huge amount of data.

After the preprocessing step, the new derived datasets are imported to *Disco*[4] software, for further analysis on process mining. *Disco* was chosen instead of other similar state of the art softwares, since the provided datasets for this assignment contained a large number of observations. At the same time, *Disco* is a great tool that provides several process mining functionalities such as: automated process discovery, process map animation, detailed statistics and several filters. Moreover, it can deal with large event logs and complex models. Ultimately, *Disco* can create insightful process maps directly from raw data.

Data Preprocessing

In order to generate insights from the data and be able to provide answers to the UWV questions, the following preprocessing steps were completed.

1) Imported the 4 datasets into *Jupyter Notebooks*.
2) Dropped all the columns with more than 70% of missing values in each file.
    a) In the *Logged_In* file, the following 8 variables were dropped because they had more than 70% missing values: *REF_URL_category*, *page_action_detail*, *tip*, *service_detail*, *xps_info*, *page_action_detail_EN*, *service_detail_EN*, *tip_EN*.
    b) In the other 3 files (*Complaints*, *Werkmap_Messages*, *Questions)*, there is no missing data.

[2] https://www.python.org/
[3] https://jupyter.org/
[4] https://fluxicon.com/disco/

3) Added descriptive text for each dataset's activity in order to achieve higher interpretability of the event log analysis in Disco, once we merge the 4 datasets in the next steps.
4) Checked the date format of each observation in all files. In some cases, it was necessary to create a timestamp with datetime format, in order to follow the same structure through all datasets.
    a) In the *Werkmap_Messages* and *Logged_In* files, there are already fields with a datetime format, *EventDateTime* and *TIMESTAMP,* respectively. However, in the *Complaints* file there is only a column with the date (ContactDate). Thus, an artificial time was added to this date. For the *Questions* dataset, there is one column for the date (ContactDate) and two other columns with time format (*ContactTimeStart* and *ContactTimeEnd*). *ContactTimeStart* was chosen because it indicates when an observation's event has begun. Finally, we concatenated both date and time in the same column.
5) Selected only the necessary variables needed for the analysis.
    a) In all datasets, *CustomerID* was saved in order to use as a *Case ID* for *Disco*. Also, the datetime of each observation was selected in order to be used as the *Timestamp* variable. As for the activity field, the following variables of each dataset were used: *PAGE_NAME* (*Logged_In* file), *QuestionTopic_EN* (*Questions* file), *ComplaintTopic_EN* (*Complaints* file) and *HandlingChannelID* (*Werkmap_Messages* file).
6) Renamed the columns in order to correctly concatenate the data.
7) Concatenated the data and merged it with customer information based on *CustomerID* values from all the datasets.
8) Filtered the final merged data according to the *AgeCategory* variable.
    a) The dataset is too large to be used in *Disco* or *Appromore*, so we segmented the data into the 4 age categories: 18-29, 30-39, 40-49 and 50-65.
9) Exported all new generated datasets (the merged one and the ones splitted by age).
    a) In the end, we have 5 csv files named as follows: final, *age_category_18-29_final*, *age_category_30-39_final*, *age_category_40-49_final* and *age_category_50-65_fina*l.

Most of the project's analysis has been focused on the new derived files, but for dealing with questions 2 and 4 as presented in the Introduction, some different preprocessing steps were necessary. More details can be found in the appropriate subsections of the report, which describe in detail the process of reaching the final answers. Eventually, the scripts produced include the generation of several graphs used in this paper.

Results

General Insights

In the graph below (Figure 1), we can see that the total numbers of males and females are balanced for each age category except for 50-65. Indeed, in this category, there are 4,150 males compared to 3,339 females approximately. In total, there are 13,305 females and 14,107 males. There is a slight difference between age categories in terms of number of customers. The groups 18-29 and 50-65 are more represented (roughly 27% each) than the 2 others (roughly 23% each).

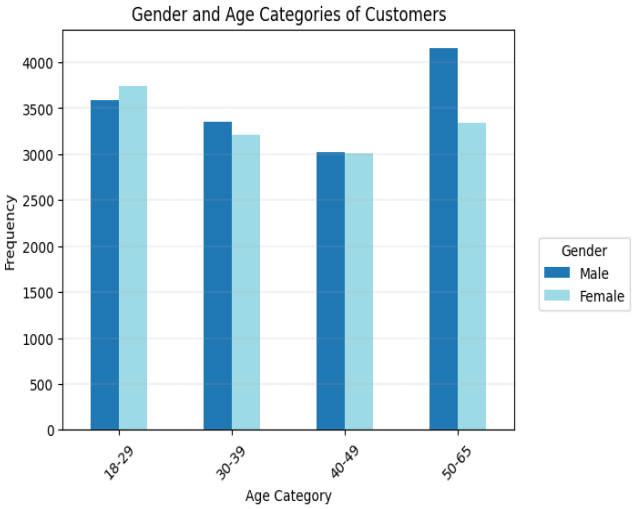

Figure 1. Number of cases by Age Category and Gender

Number of Events by Group of Age and Gender

In Table 2, the group 18-29 has the lowest number of total events, and the group 50-65 has the biggest one followed by the group 40-49. Thus, we can conclude that the number of customers per group is not related to the number of total events. In addition, women are using the UWV services more than men.

Table 2. Number of Total Events by Age Category and Gender

Age category	Nbr of events (frequency)			% (relative frequency)	
	Male (M)	Female (V)	Total (% of the total)	Male (M)	Female (V)
18-29	489,144	657,596	1,146,740 (15.57%)	42.66	57.34
30-39	681,844	795,641	1,477,485 (20.06%)	46.15	53.85
40-49	840,874	984,253	1,825,127 (24.78%)	46.07	53.93
50-65	1,552,642	1,362,690	2,915,332 (39.59%)	53.26	46.74
Total	**3,564,504**	**3,800,180**	**7,364,684 (100%)**	**48.40**	**51.60**

Mean Duration of Usage of UWV's Services by Age Category and Gender

In this subsection, an analysis on the duration of usage of UWV's services is completed. Figure 2 presents the mean duration of usage of UWV's services grouped by both age category and gender. The duration has been calculated, in days, by subtracting the minimum date of a customer's events from the maximum date. The final result is produced by grouping the data on variables: *AgeCategory* and *Gender* and by aggregating the duration values to their mean value.

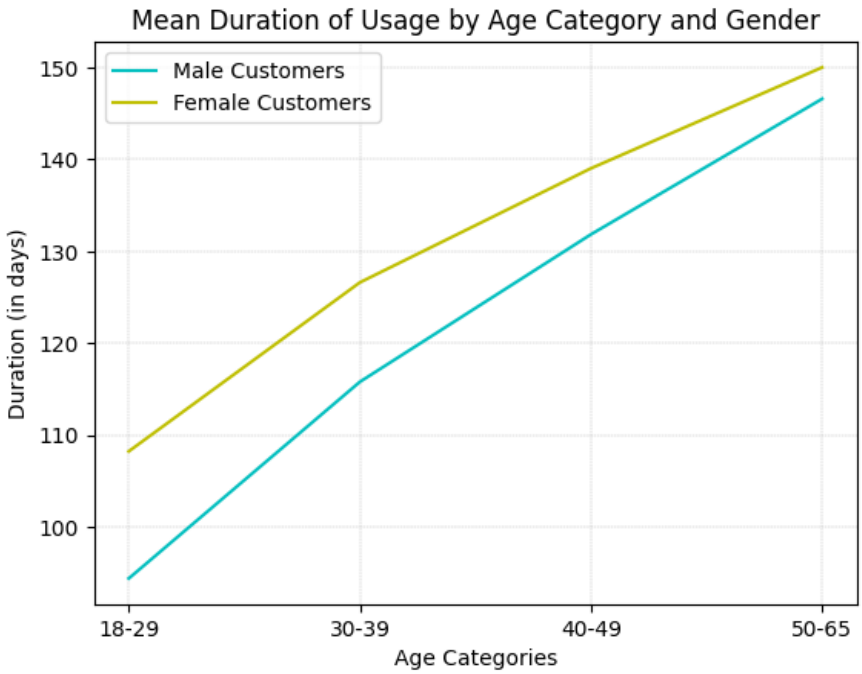

Figure 2. Mean duration of the UWV's services by Age Category and Gender

By taking a closer look at the figure, an increasing pattern of the mean duration can be observed while the age category changes from younger ages to older ones. Moreover, one interesting insight produced by the specific graph is that, for all age categories, male customers finish their customer journey sooner than the female ones. The difference between the two categories of ages (18-29 and 50-65) can be understandable because companies might prioritize hiring younger people than older people.

Customer Behavior on UWV's Services by Age Category

In this part of the report, an overview of the usage of UWV's services is brought about. Figure 3 depicts the general behavior of customers while using the services offered by UWV, as well as the interactions between those services. In order to achieve the result presented below, the datasets were loaded into *Disco*, with the column *Source* chosen as the activity of the analysis.

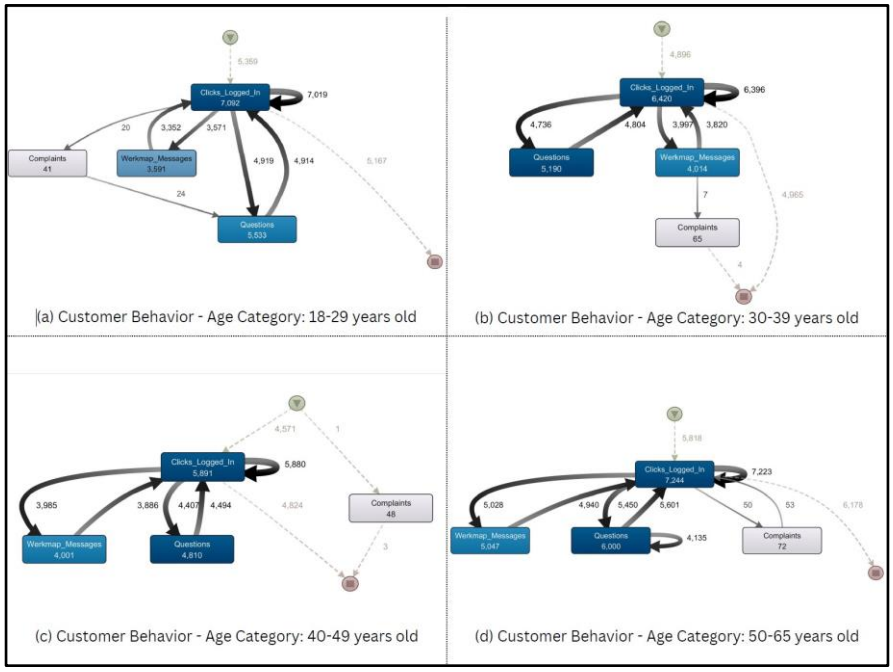

Figure 3. General Behavior of Customers by Age Category

As it is clear, in most of the cases, all age categories follow quite the same behavioral patterns. In addition, from the illustration one can understand that the customers jump from one service to another several times, before ending their customer journey. In more detail, the service, which is being used the most, in all 4 categories, is the website, followed by the questions service, werkmap messages and complaints respectively. The total number of cases fluctuates throughout the services for the different age groups, but the differences are not significant enough. Finally, it is bizarre that in *AgeCategory*: 40-49 years old, there is one case that starts directly by using the complaints service. This case will be analyzed thoroughly in the subsection Most Complaining Customers Analysis.

Website Usage Pattern Changes Over Time

In this subsection, a deeper dive into the changes, over time, of the website usage patterns is taking place. To begin with, in order to obtain information about the total number of sessions per customer, the count of unique sessions (variable *SessionID* of *Logged_in*) was conducted. The same approach has been followed by Jalali, A. [1], but in this specific case, a deeper analysis of the results is presented. Figure 4 presents the distribution of the unique number of sessions for all customers. It is obvious that the distribution is described by right skewness, meaning that the unique number of sessions is low for most of the customers.

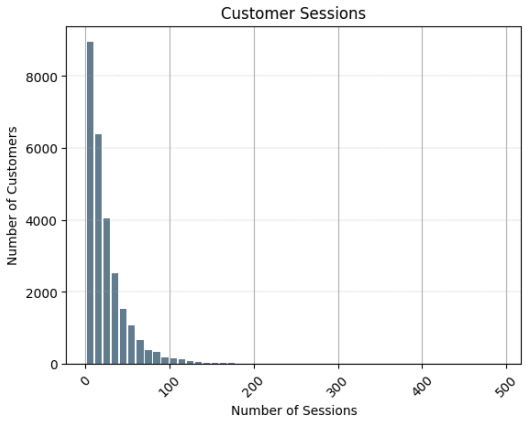

Figure 4. Distribution of the Unique Number of Sessions for all Customers

Table 3 provides a clearer view of the statistics for the unique number of sessions per customer.

Table 3. Statistics of Unique Number of Sessions per Customer

Statistics	**Mean**	**Standard Deviation**	**Min**	**25%**	**50%**	**75%**	**Max**
# Sessions	24.79	26.70	1.00	8.00	17.00	33.00	494.00

To continue with, by taking into account Figure 4 and Table 3 the thresholds for grouping the data based on the session number per customer are selected. Moreover, the dataset *Logged_in* was divided into three groups: pages visited during the first 10 website sessions of a customer, between session number 11 and 30, and finally, during session numbers of 30 and higher. To achieve the above-mentioned division, firstly, it was necessary to create a new variable indicating the order of the session for each specific event per customer. Thus, if data is sorted by website session, per customer, the division described above can be completed. Then by applying this specific filtering the dataset is splitted into the three mentioned groups. In addition, the datasets of the generated groups were loaded into *Disco*, with the variable *Page_Name* being the activity variable. Consequently, the discovery of insights for the top pages used into the 3 different groups is accomplished and described in the following subsections. Only the most important insights of this specific analysis are presented, due to space limitation. Finally, due to the enormous amount of available data, even after the filtering and creation of the groups, the usage of model comparison technique (provided by *Apromore*[5] or *ProM*[6]) was not available in *Disco*. For that reason, conclusions were derived only by following process discovery techniques in *Disco* and manual comparison.

Pages Visited During the First 10 Website Sessions

The first group of events corresponds to the top pages that clients visit throughout their first ten sessions. The mapping of the event log for this specific group is presented in Figure 5 below.

[5] https://apromore.com/
[6] https://www.promtools.org/doku.php?id=prom612

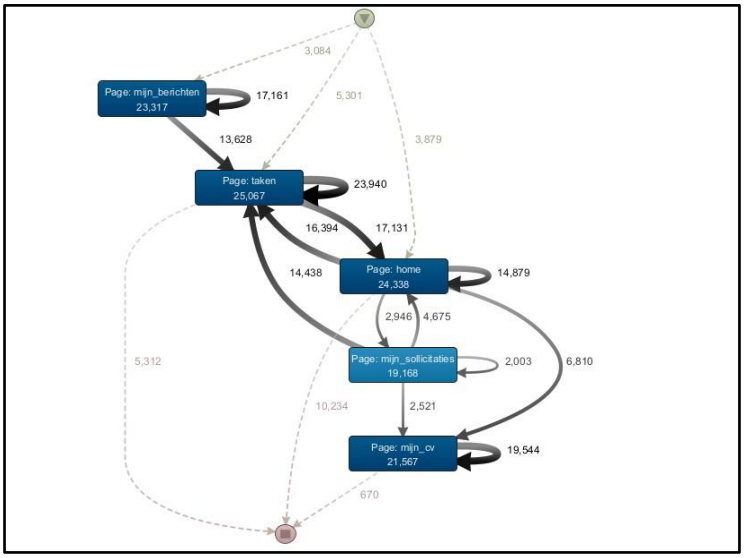

Figure 5. Top Pages Visited in the First 5 Sessions

Based on the illustration, one can understand that customers of this group use the *taken* (tasks) and *home* pages, followed by *mijn_berichten* (my_messages), *mijn_cv* (my_cv) and *mijn_sollicitaties* (my application), respectively. This group may reflect how a new consumer acts when asked to enter their information on a website. For example, this group could demonstrate how clients build their resumes when they are just starting out before applying for any job.

Pages Visited During Website Sessions: 11-30

The top pages that clients visit during their session numbers 11 through 30 are represented by the second group of events. The map of this group is produced again by using Disco mining and it is illustrated in Figure 6.

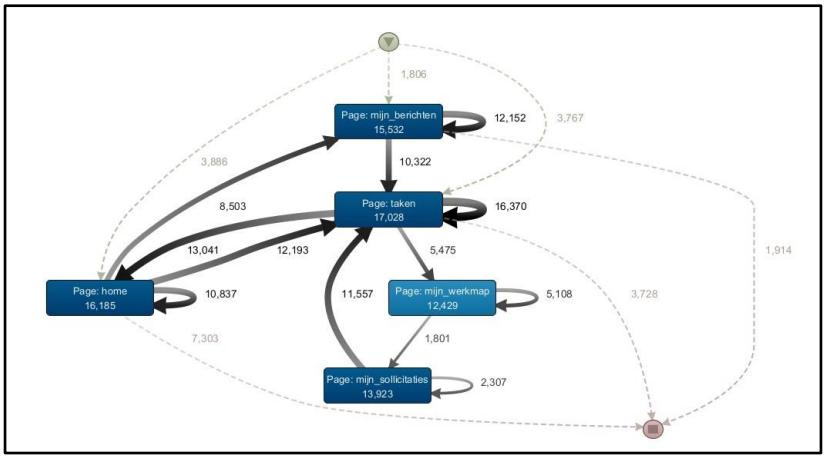

Figure 6. Top Pages Visited in Sessions 11-30

The mined process model demonstrates that the fundamental steps of the process remain the same. The most used pages presented by descending ordering are *taken* (tasks), *home*, *mijn_berichten* (my_messages), *mijn_sollicitaties* (my application), but compared with the previous model, instead of *mijn_cv* (my_cv), *mijn_werkmap* (my work folder) is included. Consequently, the *my_cv* page is less frequently used, which may be a sign that clients have already sufficiently developed their CVs. It

should be noted that while this page is still included in the process model, its frequency isn't high enough for it to be among the top most frequently visited pages. However, we can observe that the page *mijn sollicitaties* (my application) is included among the top five most visited pages. This may be because customers are motivated to pursue their submitted applications. Finally, it can be observed that customers of this group start to use more frequently the *mijn_werkmap* (my work folder) page.

Pages Visited During Website Sessions: 30 - Max

The final batch of events corresponds to the pages that clients view after session 30. Based on this sequence of events, the process map is mined using *Disco*, as shown in Figure 7 below.

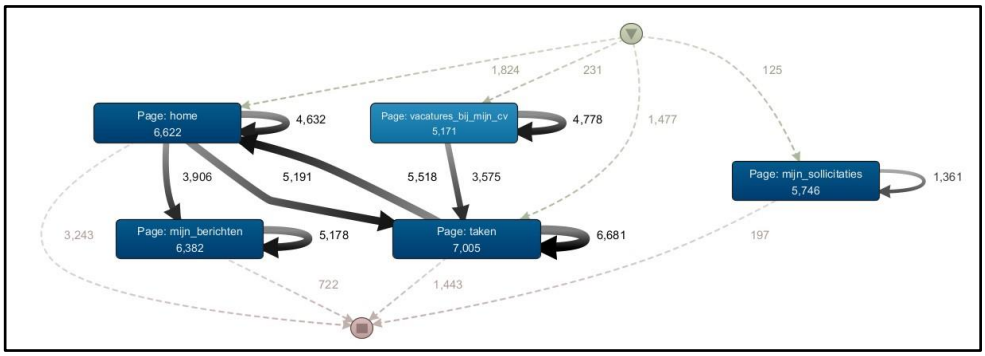

Figure 7. Top Pages Visited in Sessions 30-Max

As it is shown in the graph, the top pages used for this group are the following: *taken* (tasks), *home*, *mijn_berichten* (my messages), *mijn_sollicitaties* (my application), and finally, *vacatures_bij_mijn_cv* (jobs at my CV). The results of this group may indicate a representation of a customer who has been involved in the process for a while. The use of the *mijn_sollicitaties* (my application) page has decreased, though, which may mean that clients have applied for enough positions and are no longer applying as frequently. Contrarily, we can see that the page *vacatures_bij_mijn_cv* (jobs at my CV) is among the top five most often visited pages. This may be because of the follow-up actions taken by these clients.

Complaints Analysis

Generic complaint profiles

Number of Complaints by Age Category and Gender

Overall, as shown in Table 4 women seem to complain more than men. Indeed, 55.7% of the complaints are made by women compared to 44.3% for men. For each group of age, the number of females complaining is higher than men except for the men between 50 and 65 years old.

Table 4. Number of Events (Complaints) by Age Category and Gender

Age category	**Nbr of events**		
	Male (M)	**Female (V)**	**Total cases** (events)
18-29	13	33	46
30-39	29	56	85
40-49	27	33	60
50-65	59	39	98
Total	**128 (44.3%)**	**161 (55.7%)**	**289**

Most Common Complaints by Age Category and Gender

As described in Table 5 the complaint that comes up the most in every group of age, especially for males, is *Information: incorrect/inconsistent.*

Table 5 Most Frequent Complaints by Age Group and Gender

Age category	**Most common complaints**	
	Male (M)	**Female (V)**
18-29	- Information: no/insufficient - Payment over a certain period is missing - not/hard to reach	- Information: no/insufficient
30-39	- **Information: incorrect/inconsistent** - No respect/not taken seriously - Uninterested/received too little attention	- **Information: incorrect/inconsistent**
40-49	- **Information: incorrect/inconsistent**	- No respect/not taken seriously
50-65	- **Information: incorrect/inconsistent**	- Website not available - Payment over a certain period is missing - Income form ww unreachable

The 8 complaint topics (out of 70) that cover 51.2% of all the complaints are the ones presented in the graph below (Figure 8):

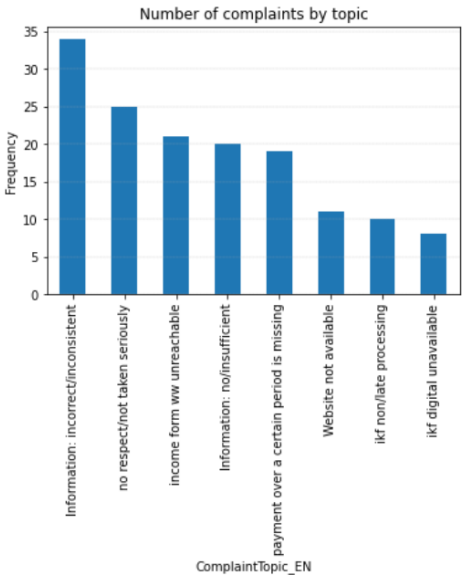

Figure 8. Most Frequent Complaints

Number of Complaints by Age Category Each Month

In Figure 9, we noticed that the overall trend is the same for each age group. Indeed, during summer the number of complaints is lower than in the other months present in the plot. Moreover, the age groups 30-39 and 50-65 are the ones that complain the most as we figured out in the previous table. In January, the 50-65 age group complained a lot compared to the other groups and the general trend.

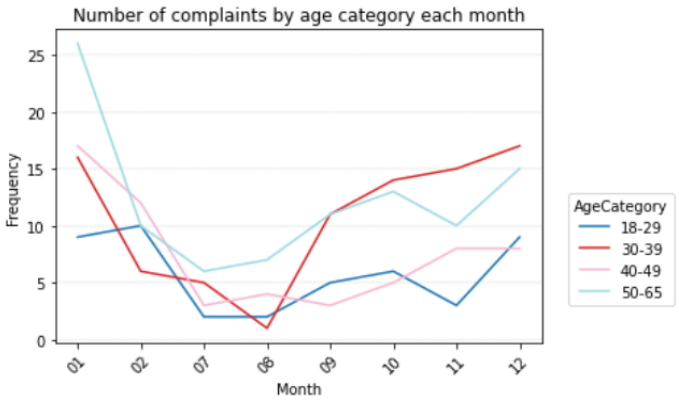

Figure 9. Number of Complaints by Age Category

By having a closer look at this strange peak, we find that a customer (customerID = 1881904) complained twice in the same day for the same issue (Table 6). This kind of behavior must be avoided, for example by adding error messages when customers want to complain on the same day for the same topic on purpose or by mistake.

Table 6. CustomerID 1881904 - Same Complaint on the Same Day

Activity	Date	Time	Source	Gender
Question: General: When should I send the form Revenue Problem?	07.01.2016	11:51:14	Questions	M
Complain: no respect/not taken seriously	12.01.2016	00:00:00	Complaints	M
Complain: no respect/not taken seriously	12.01.2016	00:00:00	Complaints	M

Most Common Complaints per Month

Over the period covered by the dataset, the most redundant complaint is *Information: incorrect/inconsistent.* However, in December the number of complaints for *income form ww unreachable* is very high and different compared to the other months. Out of the 21 complaints, 15 of them happened in early December (the first, the second or the eleventh). These complaints are done by every category of age and both males and females. We can conclude that the availability of the website is an important cause of complaints and must be fixed by UWV's services.

Table 7. Monthly Most Frequent Complaints

Month	Complaint	Number of events (max)
1	Information: incorrect/inconsistent	8
2	no respect/not taken seriously	7
7	Information: incorrect/inconsistent	3
8	no respect/not taken seriously	2
9	Information: incorrect/inconsistent	6
10	Information: no/insufficient	4
11	payment over a certain period is missing	6
12	income form ww unreachable	15

Behavior of the customers after filing the first complaint

Knowing the path of actions taken before filing a complaint is one of the most powerful measures to prevent bad user experiences, but also, knowing which actions are the ones taken after filing the first complaint is also a really interesting analysis to check if those complaints have affected the user experience.

In one of the papers related to this project [2], this analysis is performed taking into account all users. However, in order to acquire better conclusions, it has been decided to improve this strategy and perform the analysis segmented by age and gender. That is why, in this section, a study will be carried out on the path taken by users (segmented by age and gender) before and after filing a complaint. Although the segmented analysis has been made for the cases before filing the first complaint, only the graphs related to the cases after filing the first complaint will be presented; due to the main analysis made in Figure 3 and due to lack of space in this report.

The process carried out to create this section of the document was as follows: once the four different datasets associated with the different age categories had been obtained, it was necessary to know the exact time at which a user made the first complaint. For this reason, an extra column has been added whose entries are composed of the values *Before* or *After*. After grouping the *IDs* of each user together, the rows were sorted according to the *Timestamp*. A Boolean variable (*firstcomplaint*) was used as an indicator to know when the first complaint occurred. Therefore, before our boolean indicator has changed, all values of the new column are filled with the value *Before*. That is why as soon as we find the first complaint from a user, our Boolean flag changes and we proceed to enter the value *After* in the rest of the values of the new column until the end of the iteration of the corresponding user. Note that the first complaint is taken into account as *Before*. However, if a user's next action is to make a complaint again, it is processed as *After*. Once this process has been completed for all users, four new datasets have been created with the new column mentioned before. These datasets are called *beforeafter1829.csv*, *beforeafter3039.csv*, *beforeafter4049.csv* and *beforeafter5065.csv*. The datasets have been uploaded to *Disco* and the *Source* column has been taken as the activity. Once inside *Disco*, two *Attribute* filters have been used together to analyze the behavior of men or women before or after making the first complaint.

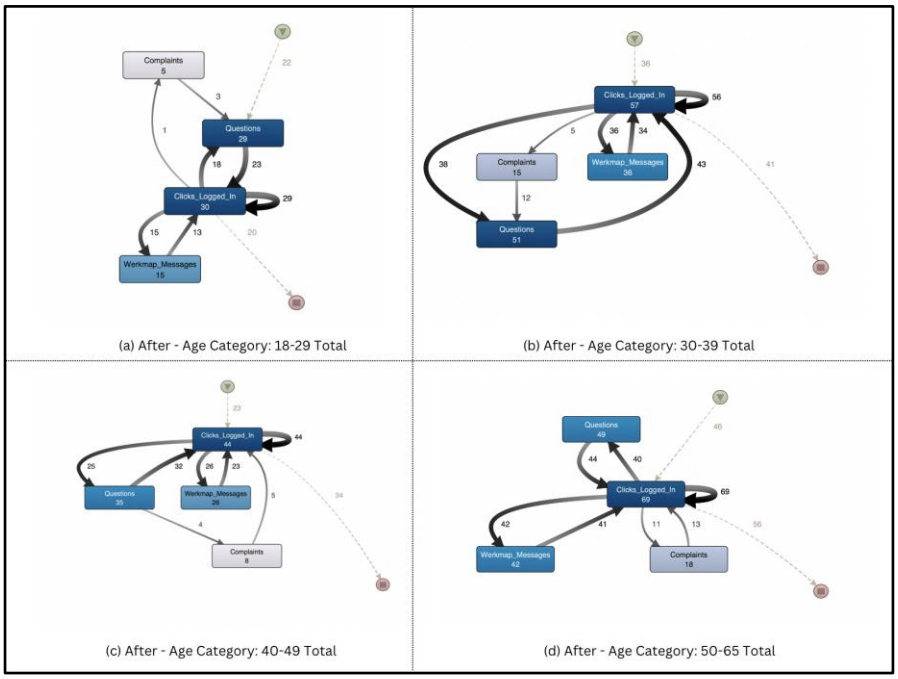

Figure 10. Path of Sources After the First Complaint

The Figure 10 shows that the behavior of the different age categories is similar among them after the first complaint. It is clear that there is a considerable decrease in the use of the web page. This shows us that the behavior of the users becomes negative when it comes to using the services of the website again. Even less frequently, users keep moving from one service to another before ending their session, following the same patterns as before complaining for the first time. Again, in all four categories, the most used service is the website followed by questions, messages and complaints. Users between 50 and 65 years old are the ones that use the complaints services the most while users between 18 and 29 years old are the ones using this service less. Immediately after the first complaint,

all categories return to using the website as the first service except users between 18 and 29 years old. These opt to use the inquiry service as their first choice. This is why this behavior will be analyzed below.

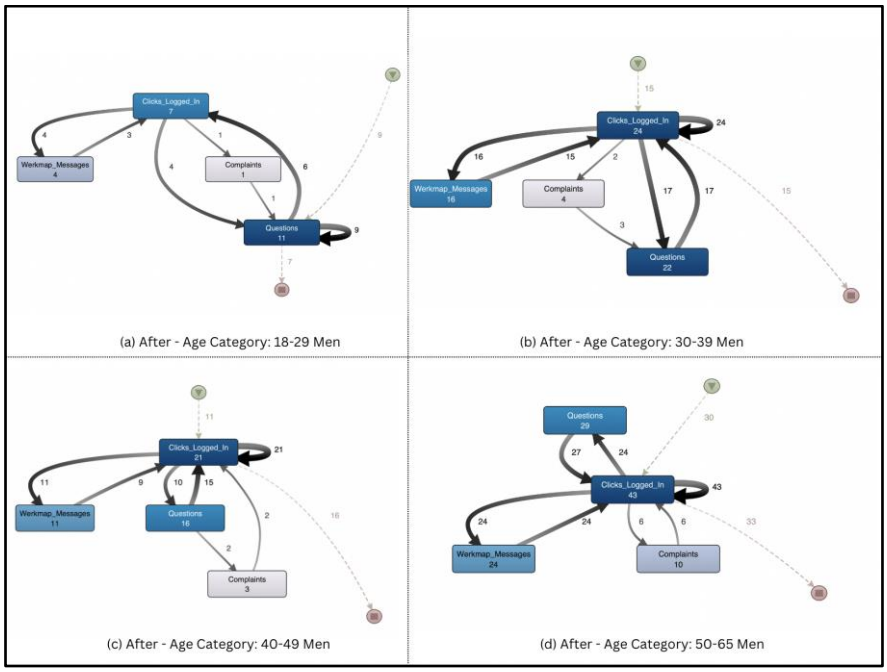

Figure 11. Path of Sources for Men After the First Complaint

The Figure 11 shows the behavior of men in different age categories after making their first complaint. Continuing with the analysis made in the previous graph, younger men complain the least while older men top the list of complaints. The order of service use is the same as above except in the 18-29 years category. Not only do they use the inquiry service as the first option after the first complaint, but it is the most used service ahead of the website. In this case, one of the conclusions that can be drawn is that younger men go directly to the questions service due to their lack of experience in the working world. However, older men tend to ask more questions and to make more complaints after filing the first complaint. So, two main recommendations will be to explore what is the real case of why young people go directly to ask questions after filing the first complaint and to make the services more accessible for older people as they have the biggest numbers in the most expensive channels.

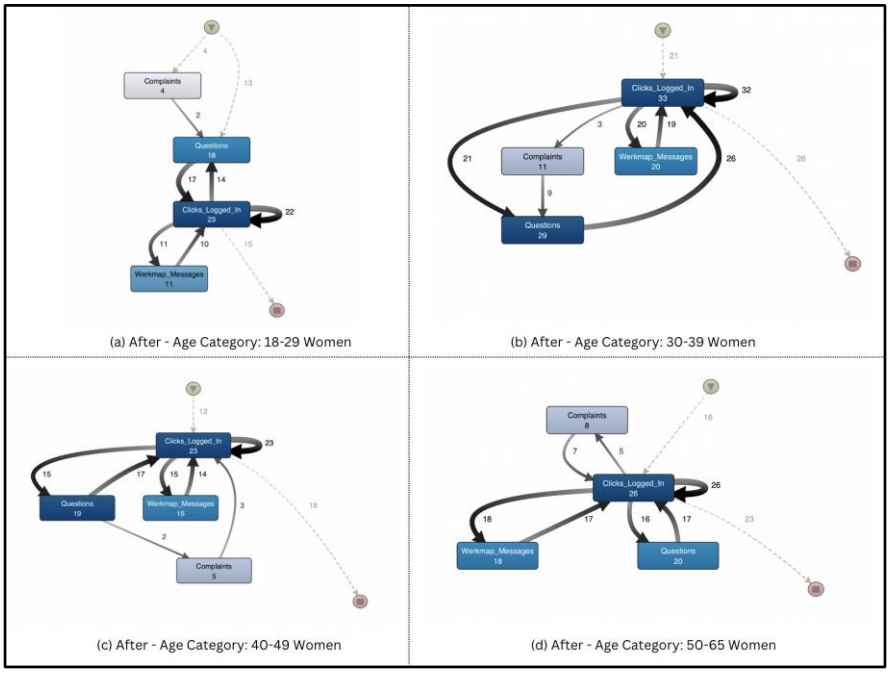

Figure 12. Path of Sources for Women After the First Complaint

Analyzing the behavior of women in the different age categories (Figure 12) the results are quite similar from those obtained for men. In all age categories the most used service was the web page, followed by questions, messages and complaints. Once again, women between 50 and 65 years of age made the most complaints after the first one. Only in the group of younger women can it be observed that after the first complaint they use the question and complaint services as their first choice.

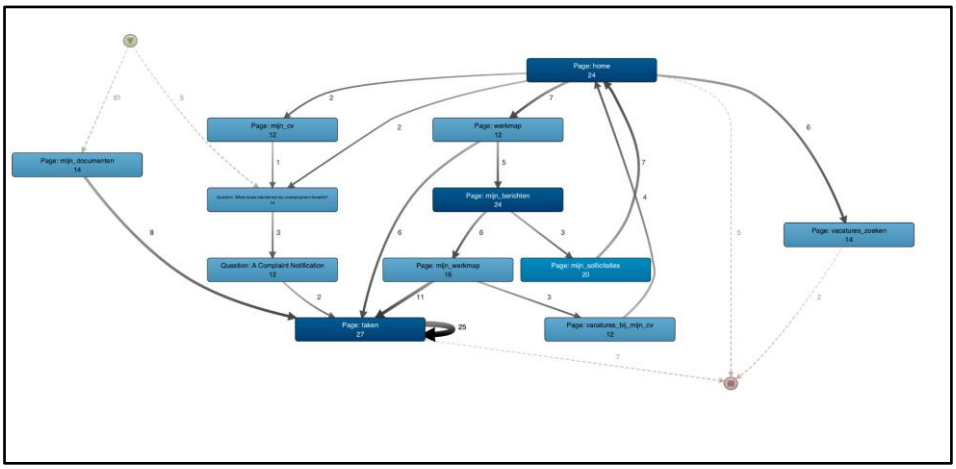

Figure 13. Path of Activities After the First Complaint

Since the behavior of younger users is the most unusual after making the first complaint, it has been decided to change the *Activity* column to try to find out the curiosities of young people. As can be seen in Figure 12 the most frequently asked question is *When is/are my unemployment benefits transferred?* followed by the question *A Complaint Notification*. This further supports our conclusion that young people do not yet have enough work experience to know, in this case, information about the benefits of a layoff, and also seek information about the complaints themselves. On the other hand, the pages most used by this user group are *taken* (tasks), *home*, *mijn_berichten* (my messages) and *mijn_solicitaties* (my application).

Frequency of a Customer Complaining

General Frequency Analysis

Complaints should be avoided as handling them for the website's support team is quite expensive. The investigation of complaints' frequency can bring more insights, help enhance UVW services and prevent further multiple complaints in the future.

Table 8. Number of Customers with One or More Complaints

Number of complaints	Number of distinct cases	Total number of complaints
1	180	180
2	35	70
3	6	18
4	4	16
5	1	5
Total	**226**	**289**

The Table 8 summarizes the number of complaints per case (customer). The log gives a total of 289 complaints from 226 different people. Overall, most users (180) complained only once nevertheless 46 complained multiple times.

Table 9. Statistics about Multiple Complainers

Age category	Total cases	Total cases with multiple complaints	% of multiple complainers	Average number of complaints	Duration interval between 1st and last complaint (days)	Average duration between 1st and last complaint (days)
18-29	41	5	12.2%	2	[0;21]	5.25
30-39	65	15	23.1%	2.33	[0;122]	18.5
40-49	48	8	16.7%	2.5	[0;113]	30.25
50-65	72	18	25%	2.44	[0;120]	33
Total	**226**	**46**	**20.35%**	**2.32**	**[0;122]**	**21.75**

Because of incompatibility in the time granularity of the complaints events and the insertion of artificial times, the statistics related to complaints in terms of hours may not be reliable. We will do the analysis with regards to days.

The Table 9 shows statistics about customers complaining more than once, found thanks to a *Disco Follower* filter. Like we just saw, among the 226 complaining customers, 46 complained multiple times. In total, this result corresponds to 20.35% of distinct complainers with an average of 2.32 complaints. With regards to age categories, the most multiple complainers are included in the 30-39 and 50-65 age groups. However, the number of multiple complaints is higher for the 40-49 years old category with 2.5 complaints on average. Young people tend to make fewer complaints with respect to the total number of complaints and the total number of complainers.
The timespan between the first and the last complaint also differs according to age. Indeed, the younger the customers are, the shorter the duration between the complaints. These results are correlated with the mean duration of usage of UWV's services of each category, where an increasing pattern of the mean duration can be discerned while the age category changes from younger ages to older ones.

Most Complaining Customers Analysis

The volume of complaints might not be a significant problem, but from a quality point of view, it is obvious that all process improvements should be sought.

Case ID	Events	Variant	Started	Finished	Duration
1202227	5	Variant 1	07.07.2015 00:00:00	07.07.2015 00:00:00	0 millis
2016564	4	Variant 2	09.07.2015 00:00:00	30.10.2015 00:00:00	113 days, 1 hour
2060596	4	Variant 3	18.08.2015 00:00:00	17.09.2015 00:00:00	30 days
1604792	4	Variant 4	16.10.2015 00:00:00	23.10.2015 00:00:00	7 days
1373382	4	Variant 5	05.10.2015 00:00:00	01.12.2015 00:00:00	57 days, 1 hour

Figure 14. Top Five Most Complaining Customers

In order to find insights about the profile of complainers and usage patterns, the top five most complaining customers are investigated using an *Attribute* filter in *Disco*. This further analysis, never done in previous studies, can determine whether the complaints are justified or correspond to misuses from complainers.
You can find, in Figure 14 the most complaining users with the corresponding number of complaints, the first complaint's timestamp, the last complaint's timestamp and the related timespan.

Case ID 1202227

12	Channel 1	05.07.2015	11:54:48	Werkmap_Messag
13	Complain: (Multiple) requests, to little avail	07.07.2015	00:00:00	Complaints
14	Complain: no respect/not taken seriously	07.07.2015	00:00:00	Complaints
15	Complain: Information: incorrect/inconsistent	07.07.2015	00:00:00	Complaints
16	Complain: Information: incorrect/inconsistent	07.07.2015	00:00:00	Complaints
17	Complain: uninterested/received too little attention	07.07.2015	00:00:00	Complaints
18	Question: What time/when/with whom I have my appointment?	07.07.2015	11:22:08	Questions
19	Question: I receive a benefit from UWV or have one made and want to cancel my appointment (with the consultant work) (valid reason)	07.07.2015	11:25:29	Questions

Figure 15. Case ID 1202227 Complaints

The maximum number of complaints was done by the customer with ID 1202227, a 30-39 year old woman, with a total of 5 complaints on July 7th 2015 (Figure 15). These complaints were all made the same day, 3 days after her registration on the website. Also, before making complaints, she only did 12 events including one werkmap message and she decided to ask questions after complaining. Thus, we can suppose that this customer may not have used the website a lot before, and in a hurry sent a lot of complaints.

Case ID 1373382

120	Channel 2	15.09.2015	14:33:34	Werkmap_Messag
121	Complain: no respect/not taken seriously	05.10.2015	00:00:00	Complaints
122	Complain: Information: no/insufficient	05.10.2015	00:00:00	Complaints
123	Page: werkmap	05.10.2015	11:37:16	Clicks_Logged_In
216	Page: home	04.11.2015	15:17:48	Clicks_Logged_In
217	Complain: change form unreachable	05.11.2015	00:00:00	Complaints
218	Question: To which address should I send the documents?	05.11.2015	10:44:15	Questions
265	Page: home	30.11.2015	16:27:51	Clicks_Logged_In
266	Page: online-training	30.11.2015	16:28:15	Clicks_Logged_In
267	Complain: income form ww unreachable	01.12.2015	00:00:00	Complaints
268	Page: home	01.12.2015	10:25:55	Clicks_Logged_In
269	Page: taken	01.12.2015	10:27:13	Clicks_Logged_In

Figure 16. Case ID 1373382 Complaints

In the same age category, another woman complained 4 times in total (Figure 16). On October 5th 2015, she complained about a lack of respect and a lack of information after sending a Werkmap message 20 days before, maybe without any answer.
Another complaint was made on November 5th for a *change form unreachable* and a last one on December 1st for an *income form ww unreachable*. This total timespan is around 113 days. We can assume that the customer complaints are pertinent and should be considered by the UWV.

Case ID 2016564

	Activity	Date	Time	Source
1	Complain: (Multiple) requests, to little avail	09.07.2015	00:00:00	Complaints
2	Complain: Information: no/insufficient	09.07.2015	00:00:00	Complaints
3	Question: Why was my application rejected for unemployment benefits?	09.07.2015	15:53:11	Questions
4	Question: I received a request to submit documents, but I have a question about that.	30.09.2015	12:42:36	Questions
39	Question: A Complaint Notification	28.10.2015	13:55:34	Questions
40	Complain: Information: no/insufficient	30.10.2015	00:00:00	Complaints
41	Complain: missed appointments	30.10.2015	00:00:00	Complaints
42	Question: I got a call from one of the UWV in connection with applying for unemployment benefits.	30.10.2015	09:12:28	Questions

Figure 17. Case ID 2016564 complaints

Regarding the 40-49 age group, a new case arises with a woman starting directly by complaining twice just after her registration on July 9th, 2015 (Figure 17). Then, she stopped using the website for almost three months. These events can demonstrate a misuse of the website. She may have made a mistake or even created a new account and complained about events happening with her previous account. The last proposition may be more plausible since she continued by asking questions about a rejected application.
The two other complaints were made on October 10th, after asking several questions two days before, mainly about *complaint notification*. We can notice that, as it was shown in other studies [3, 4], the filing of complaints was found to be correlated with the presence of messages and questions. In contrast to the two first odd complaints, the two last ones seem relevant.

Case ID 2060596

272	Page: Portal (ignore_request_page)	11.08.2015	13:46:53	Clicks_Logged_In
273	Complain: (Multiple) requests, to little avail	18.08.2015	00:00:00	Complaints
274	Complain: no respect/not taken seriously	18.08.2015	00:00:00	Complaints
275	Page: werkmap	18.08.2015	05:16:01	Clicks_Logged_In
343	Page: mijn_werkmap	01.09.2015	11:12:41	Clicks_Logged_In
344	Complain: no respect/not taken seriously	17.09.2015	00:00:00	Complaints
345	Complain: employee not competent	17.09.2015	00:00:00	Complaints
346	Page: mijn_berichten	17.09.2015	11:55:09	Clicks_Logged_In

Figure 18. Case ID 2060596 Complaints

A 50-65 year old man made 4 complaints about the customer service. Indeed, after sending 4 messages on August 8th and 4 messages on August 10th, he complained about *multiple requests, too little availability* and *no respect/not taken seriously* on August 18th (Figure 18). At the beginning of September, he logged in the website to complain again about the lack of respect and the employee incompetence. We can once again think that he did not have an answer to his demands.

Case ID 1604792

129	Page: taken	02.10.2015	10:12:34	Clicks_Logged_In
130	Question: Why did you change the amount of my payment?	15.10.2015	16:49:48	Questions
131	Complain: change form is not processed, incorrect payment	16.10.2015	00:00:00	Complaints
132	Complain: change form non/late processing	16.10.2015	00:00:00	Complaints
133	Page: taken	16.10.2015	10:52:49	Clicks_Logged_In
145	Page: taken	16.10.2015	17:21:04	Clicks_Logged_In
146	Complain: missed appointments	20.10.2015	00:00:00	Complaints
147	Question: specific question	22.10.2015	08:50:43	Questions
148	Complain: no internal communication/data transfer	23.10.2015	00:00:00	Complaints
149	Question: specific question	23.10.2015	10:21:24	Questions

Figure 19. Case ID 1604792 Complaints

Finally, the last most complaining customer is a 50-65 year old woman with 4 complaints. The day before the two first complaints about *incorrect payment* and *change form non/late processing*, she asked a question about the amount of her payment (Figure 19). As seen in paper [5] the majority of the questions fall in the category of "Payment" and considering the number of complaints, improvement in payment procedures and its reliability can be taken into account.
Thus, we might conclude that the two other complaints concerning *missed appointments* and *no internal communication* are related to the previous payment events.

Overall Findings

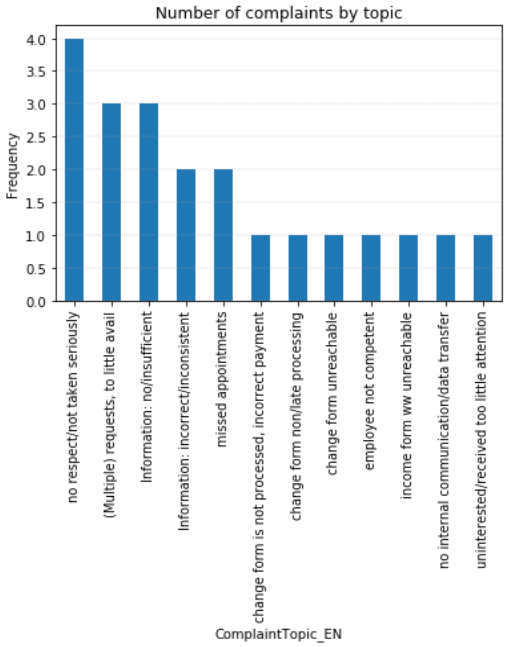

Figure 20. Number of Complaints by Topic for the top 5 Most Complaining Customers

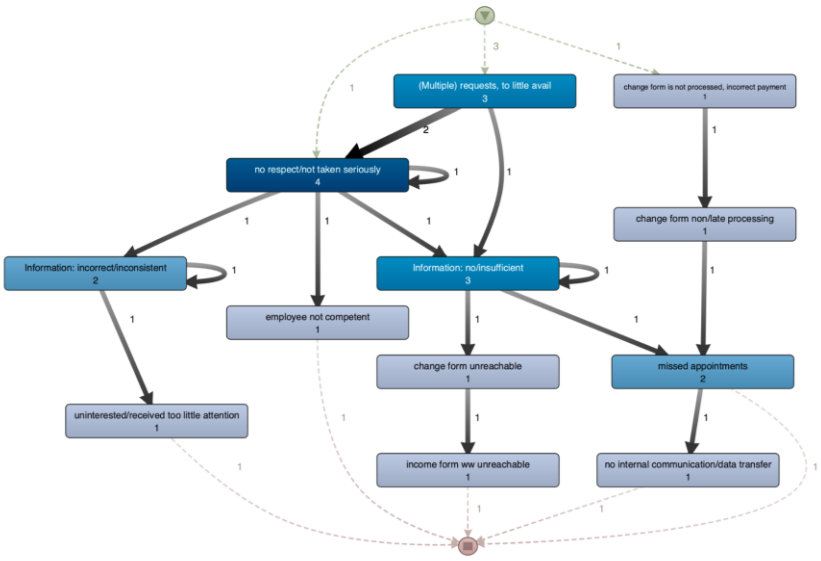

Figure 21. Process Map of Complaints for the Top 5 most Complaining Customers

To summarize, the most complaining customers tend to complain about various topics on the same day. The most frequent types of complaint are about the lack of respect/seriousness, the too little availability regarding multiple requests and the lack of information (Figure 20). Regarding the usage patterns, Figure 21 shows the complaints *information: no/insufficient* and *no respect/not taken seriously* stem from another complaint: *(multiple) requests, too little availability*.

Some of the multiple complaints cases, related to missing information or impolite treatment, seem justified, meaning that the UWV should enhance its customer support and the access to information. By improving these areas, the UWV can decrease the amount of complaints and avoid multiple correlated complaints by the same customers. The top five customers have often used the *Werkmap Messages* and *Questions* a lot, so they seem seriously committed.

We also noticed that the majority (4 out of 5) are women, which again supports that women complain the most.

Conclusion

In this report, we explored the general insights of the users of the UWV, the website usage patterns and the analysis about the complaints.

First of all, we have done preprocessing work of the four files that we used before loading them on Disco. To analyze the data and find some new insights we used two tools: Disco and Python.

For the general insights, we analyzed gender and four groups of ages. Regarding the sex, the percentage of its repartition gender is quite similar but women are the ones that use more services than men.

Concerning the changes in the usage patterns over time, it can be concluded that customers during their first 10 sessions spend most of their time creating their CV. The group of customers with session numbers 11-30 focus mainly on applying for jobs, while the last group (30-max sessions), mainly focuses on page *Jobs On My CV.*

About the number of users, the categories that are the most represented are the 18-29 and 50-65. Even if they are the most represented on the UWV platform, their mean duration are very different. The 18-29 users are the ones that stay less time on this platform whereas the 50-65 age group, indeed, this last category might have more difficulties finding new jobs.

Complaints are a really interesting topic to analyze because it shows if the customers are satisfied or not about the services. Answering complaints takes a lot of time for the UWV employees so it can be important to enhance UWV services, and thus, reduce the amount of complaints. We studied BPI'16 previous papers that did not go deeper into the analysis of this issue .

Women are the ones that complain the most, except for the 50-65 category. For all ages, the complaint that is more redundant is Information: incorrect/inconsistent. This complaint is more likely used by men than women. In December, we noticed that the most common complaint was income form ww unreachable. About this complaint, we can conclude that the availability of the website is an important cause of complaints and must be fixed by the UWV services.

After their first complaint, young men use the inquiry service as the first option whereas young women use the question and complaint services. Thus, younger people might go directly to the questions service due to their lack of experience in the world of work.

In this 8-month analysis, 226 customers made one or more complaints. A small proportion of the complaining users complained more than once. The most multiple complainers belong to the 30-39 and 50-65 age groups, with, again, a majority of women. Moreover, some outlier cases were found, showing that complaining should be regulated. Indeed, two cases demonstrated a strange behavior with two complaints just after the registration for one. Looking at the other one, two complaints were made the identical day about the same topic.

From the analysis of this paper, it is recommended to the management of UWV to analyze why young people go directly on questions after their first complaint. They can also focus on getting the services more friendly to older people because they are the ones that have the biggest numbers in the most expensive channels. To avoid some complaints the management of UWV could take a look at the customer service because multiple complaints cases are related to missing information or impolite treatment. Finally, to avoid misuse and abuse of complaints, an error message could be included in the website.

To go further on the research, we could analyze the duration between the time they register and the time they make their first complaint. Moreover, it would have been interesting to analyze the duration between the time they make their first complaint and the time the company answers. Finally, we thought about including more data to analyze the answer of the company about the complaints and see if the future complaints are fair or not.

References

[1] A. Jalali, Exploring different aspects of users behaviours in the Dutch autonomous administrative authority through process cubes. Business Process Intelligence (BPI) Challenge., 2016.

[2] Janssenswillen G. et al., Does werk.nl work?. Business Process Intelligence (BPI) Challenge., 2016.

[3] Dadashnia S. et al., Identification of Distinct Usage Patterns and Prediction of Customer Behavior. Business Process Intelligence (BPI) Challenge., 2016.

[4] V. D. H. U., Marking up the right tree: understanding the customer process at UWV. Business Process Intelligence (BPI) Challenge., 2016.

[5] Heidari F. et al., Usage analytics Using Process Mining. Key Findings for the Dutch Employee Insurance Agency. Business Process Intelligence (BPI) Challenge., 2016.

Contribution of Group Members

In this section the contribution of each group member is described, indicating who participated in writing the sections and subsections of the report, as well as
the creation of the preprocessing and visualization *Python* scripts.

Ander Barrio Campos

- Behavior of Customers After Filing the First Complaint (segmented by age category and sex)
- Behavior of Customers Before Filing the First Complaint (segmented by age category and sex) → not added due to lack of space
- Creation of *before_after_creation.ipynb*

Nora Hérault

- Methodology (Tools Used, Data Preprocessing)
- General Insights (Number of Events by Group of Age and Gender)
- Generic Complaints profiles (Number of Complaints by Age and Gender, Most Common Complaints by Age Category and Gender, Number of Complaints by Age Each Month, Most Common Complaints per Month)
- Creation of *data_preprocessing.ipynb*
- Creation of *graphs_general_insights_complaints.ipynb*

Odysseas Kyparissis

- Methodology (Tools Used, Data Preprocessing)
- General Insights (Mean Duration of Usage of UWV's Services by Age Category and Gender, Customer Behavior on UWV's Services by Age Category)
- Website Usage Pattern Changes Over Time (Pages Visited During the First 10 Website Sessions, Pages Visited During Website Sessions in Interval 11-30, Pages Visited During Website Sessions with Min 30 Sessions)
- Creation of *data_preprocessing.ipynb*

Perrine Lafaye

- Dataset description
- Tools used
- Conclusion
- Prospect

Emma Salvan

- Introduction
- Context
- Frequency of a Customer Complaining (General Frequency Analysis, Most Complaining Customers Analysis)
- Creation of *graphs_frequency_complaints.ipynb*