TECHNICAL UNIVERSITY OF MUNICH

TUM Data Innovation Lab

End-to-End Process Enhancement for Crew Management

Authors	B.Sc. Julia Winona Fiebich (03670068)
	B.Eng. Lennart Lehmann (03707777)
	B.Sc. Mitali Agarwal (03722893)
	B.Sc. Sebastian Stohrer (03671954)
	B.Sc. Siwar Ben Chaabene (03687915)
Mentors Lufthansa CityLine	Dipl.-Ing. Philipp Grindemann
	M.Sc. Maximilian Hoffmann
Mentors Celonis	Dr.-Ing. Josua Vieten
	M.Sc. Janina Nakladal
Co-Mentor	M.Sc. Leonardo Cancissu Araujo (Department of Mathematics)
Project Lead	Dr. Ricardo Acevedo Cabra (Department of Mathematics)
Supervisor	Prof. Dr. Massimo Fornasier (Department of Mathematics)

Feb 2021

Contents

1 Introduction

1.1 Content and Background

In this paper we have documented the work we have done in between November 2020 and February 2021 on the project to create end-to-end transparency within the crew management process for Lufthansa CityLine (CLH) leveraging real company data from CLH and the Celonis software by creating a process mining model. In the following, we give a short introduction about our project partners as well as about the topic of process mining. Next, we explain the starting point and goal of our project and give details into our methodology. Following that, we describe the crew management process and derive use cases. The next chapter offers details about the creation and implementation of our process mining model and dashboards for the crew management process. Based on the Celonis dashboards, the next chapter deals with the analysis of the crew management process and illustrates optimization potentials. In the final chapter, we give a conclusion and offer an outlook.

Lufthansa CityLine GmbH (CLH), a subsidiary of the Deutsche Lufthansa AG, is a regional airline completing feeder flights under the Lufthansa name. This includes flights primarily from European destinations to the Lufthansa hubs in Munich and Frankfurt. Lufthansa CityLine enhances the Lufthansa offering with flights to European metropoles and transports over eight million passengers annually. With 2,250 employees, they provide over 300 flights daily between 85 destinations in 22 countries. The employees are composed of cockpit members, cabin crew, technical crew and administrative staff. [1]

Our second project partner is Celonis SE, a software company focusing on process mining and one of the market leaders in this area. The Celonis software enhances business performance by analyzing digital footprints (e.g. from ERP systems) to visualize process chains and derive process knowledge. Due to the increase of collectible data, it is becoming increasingly important to effectively harness this potential. Process mining is a technology leveraging system-generated data to improve the overall efficiency of various processes. With this technology, work does not have to be re-imagined, as action potentials can be derived to improve processes. [2]

1.2 Process Mining Fundamentals

Digital transformation is a key driver that businesses focus on today. In order to start the digital transformation, transparency is a prerequisite, which can be achieved with process mining. [3] Understanding actual and real processes across various industries and organizational departments allows for sustainable efficiency improvement, driving transformation and creating value. [4] Process mining provides a means of generating data-driven transparency on how business processes are executed within companies. This helps companies to identify bottlenecks in their process chains, identify non-conforming variations and thus, they can counteract and improve their processes. [4, 3]

In order to make use of process mining, there needs to be at least three crucial data properties in place. The data must comprise a unique identifier (ID), an activity name as well as a timestamp tied to the respective activity name. This is called the event log. [3] After data-specific transformations to the event log, the activity table is generated. The activity table contains all relevant process tasks along with its ID, timestamp and optionally some more specific information such as whether the activity was system generated or human-made, or the location of the activity. Each new process mining process requires a mapping of the activity names (the ellipses seen in a process mining analysis) to the underlying timestamps within the data set. Mapping is a complex process, sometimes lasting for months, and involves the expert knowledge of the process owners as well of the IT architects. Furthermore, it is necessary to come up with a distinct ID, which can be traced for the entire process.

Process mining is a widely used technique for different problems, varying from purchase process analytics [3] to supply-chain analytics [5] to software process mining [6, 7]. Process mining will give a holistic and transparent overview of the entire crew management process with the recorded data comprising an ID, activity name, and timestamp. The vast amount of data is statistically analyzed as well as visualized and thus, process mining provides an extensive view on how the crew management process is enforced. Process mining enables organizations to analyze and evaluate millions of data points and helps to better understand the process end-to-end. Due to this process transparency it is possible to analyze bottlenecks and counteract against the identified problems.

Celonis software allows to intuitively explore the business process, checks for conforming cases, and has additional functionalities, like a machine learning workbench or an action engine that proactively suggests process improvements. This work will make use of Celonis process mining to analyze crew management data and to derive action potentials for process improvements. A Celonis on-premise 4.6 version was used to build every process mining analysis and model in this project.

2 Motivation

In the past, Lufthansa CityLine has successfully implemented process mining to enhance their ground operations process. Flight punctuality has been increased by 300,000 minutes by identifying bottlenecks and sup-optimal performance and by offering enhancement recommendations. The AI-based software simultaneously monitors process conformity to ensure continuous process improvement [8].

As the next step going forward, Lufthansa CityLine wants to extend their work with Celonis into other processes. With over 2,250 employees, the crew management process is rather complex, especially during an exceptional situation like the one we find ourselves in during the current pandemic. In this project, we will create a process mining model in the Celonis environment for the crew management process. The aim is to visualize process chains, derive bottlenecks and sup-optimal process flows and leverage optimization potentials for several use cases. Overall, we will add value by further optimizing operation for Lufthansa CityLine by holistically analyzing the crew management process and deriving recommendations for action from optimization potentials.

2.1 Goal of this Project

All stakeholders in this project had different interests, therefore it was very important to define common desired outputs at the beginning of this project. In a first meeting with all stakeholders, the following desired outputs were put forth by the stakeholders. The TUM is interested in having their students gain experience whilst implementing and extending their knowledge in research projects. Lufthansa CityLine is interested in gaining more knowledge about their crew management process. They want a transparent overview of all processes involved in order to be able to determine inefficiencies and potentials for improvement. To do this, they wanted us to analyze the process in Celonis and derive knowledge and improvement potentials from the analysis of KPIs, process flows and problems in the process chain. This overlaps partly with the goals of Celonis, as they are interested in further implementing their software with Lufthansa CityLine to help them increase their process efficiency via process mining. Lastly, our student team was interested in expanding our skillset, gaining experience in a new field and further developing our team working abilities.
These differing desired outputs were combined in our overarching SMART goal for the project:

We will implement a process mining model in Celonis that encompasses transparency in the end-2-end crew management process of Lufthansa Cityline by analyzing and optimizing the entire process chain based on historical flight data until Feb 2021.

This was the initial project goal determined in concurrence with all stakeholders. All future changes to the scope of the project, and therefore the project goal, will be discussed in chronological order of appearance throughout the report.

2.2 Methodology and Approach

The SMART goal outlines the scope of this project. The defined tasks were complex and overarching and thus, were to be broken down according to necessity. We came up with three areas in which we needed to work: Analyzing the current crew management process at Lufthansa CityLine, designing and implementing a process mining model in the Celonis environment, and deriving improvement potentials from our analyses and dashboards, which are based on the process mining model.

Using the main tasks as a directory, we mapped out the tasks within each main task. From this, we developed a timeline for the upcoming months including large milestones. You can see this timeline in figure 1 below.

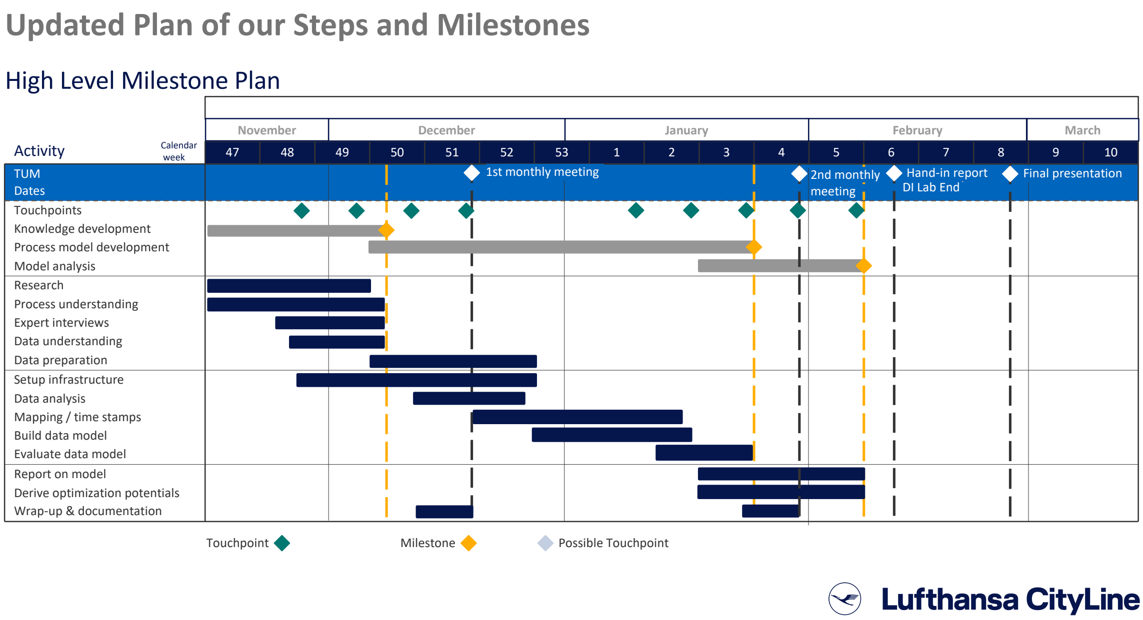

Figure 1: Timeline with Milestones

The first part of our project was knowledge development. To foster this, we conducted interviews with experts in the crew management process at Lufthansa CityLine. We also completed virtual trainings with Celonis to learn how to use their software to build process mining analyses. In the second part of our project we created a data model in the Lufthansa CityLine SQL server. Following this, we created several analyses of the process mining model in Celonis. In evaluating our data model with the Lufthansa CityLine experts, we improved our data model in several iterations. The last part of our project was spent analyzing the process mining model and defining optimization potentials for Lufthansa CityLine.

In general, we worked in an agile manner, in which we had a biweekly sprint, a so-called *touchpoint*, with all the experts that we interviewed in the beginning of this project. During these *touchpoints* experts from various backgrounds validated our results from our Celonis analysis and gave us constructive feedback to further improve the process mining model. This helped a lot, since often times it is the details within columns i.e., a different value in a column that makes the difference to the process mining model.Through receiving the feedback and answers to our questions, we were able to build our data model and could continuously improve the data model itself as well as the results. Additionally, once we had created the data model, we were able to continuously receive feedback from the experts, who will be the future user group of the dashboards, which helped us to adjust the dashboards specifically to their user needs. Furthermore, we had a weekly meeting with the project owner from CLH, who answered all the open points that we faced during the implementation phase as well from our Celonis mentor, who helped us with any subjects related to the Celonis software. Hence, we were able to continuously enhance our data model.

3 Crew Management Process

3.1 Understanding the Crew Management Process

The first step in our project was the analysis of the current processes. Thus, we started by analyzing the process schemes of the crew management and conducted expert interviews to develop a general understanding of the different process steps. The crew management process is made up of four sub-processes: the crew capacity planning, crew scheduling, updating the crew schedules and finally crew control. In the following, we will go into detail about the content, scope and timeline of all four sub-processes to provide the reader with a better understanding of the subject. A visualization of the four sub-processes can be found in the diagram below.

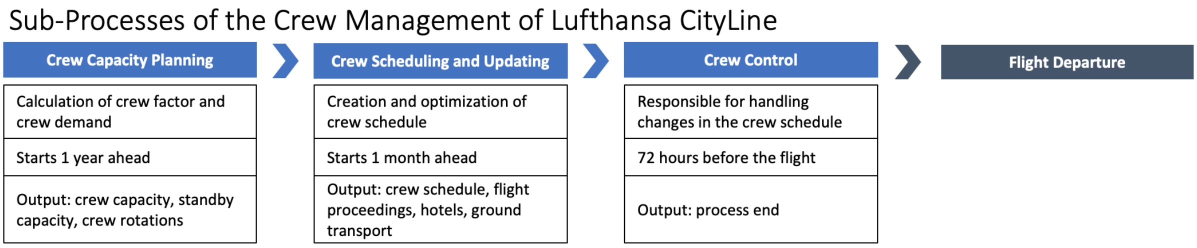

Figure 2: Overview of the Sub-Processes

3.1.1 Crew Capacity Planning

Crew capacity planning is the first sub-process and starts one year ahead with the calculation of the crew factor and the crew demand. These calculations are mainly manually done in Excel. The crew factor determines how many crew members per aircraft will be needed to satisfy the crew demand. In the next steps, requests for part-time and vacation as well as calculation of demand for training and recruitment is done in order to match the crew demand. At the end of this process the monthly crew capacity is provisioned. Based on the monthly crew capacity, the monthly standby capacity is calculated. Standbys are additional crew members that act as operational reserve, ready to replace operating crew members. For example in case of sick crew members or in case of flight plan changes in the short term.

Simultaneously, the crew rotations are created. Crew rotations indicate how much and when a crew member will be working - which is crucial in the highly regulated airline industry. At end of the crew capacity planning, three outputs have been generated: the necessary monthly crew capacity, the required standby capacity and the monthly crew rotations.

3.1.2 Crew Scheduling and Updating

The creation of the crew schedules is done one month in advance. This process is done automatically in the software Netline and optimally assigns crew members to their duty days for the following month according to capacity demand and flight schedules. Given the inputs of crew capacity planning further factors are taken into account such as absences or trainings of crew members. The crew schedules go through an optimization process taking into account regulatory and internal requirements such as equal distribution of working hours and days for crew members.

After the approval of the employee representatives, the crew schedule is released for publishing and flight proceedings are booked. Due to Lufthansa CityLines former stations in other German cities, not all crew members live in the two main hubs Frankfurt and Munich. Therefore, flight proceedings are needed to bring these crew members to the two hubs from where they can start their flight duties. The sub-process crew scheduling ends with the published crew schedule for the upcoming month as well as booked flight proceedings for the upcoming month. In the sub-process crew schedule updating ground transports, hotels as well as rental cars are booked for crew members in advance.

3.1.3 Crew Control

The last sub-process, crew control, starts after the crew schedules are published. Crew control is responsible for any changes that happen between the publication of the crew schedules and the departure of an aircraft. It is divided up into schedule management and crew control. In schedule management, any changes that happen more than 72 hours before the flight departure are included such as flight plan changes and re-planning of trainings. Crew control is responsible for short-term changes 72 hours before the flight departure. Examples for short-term changes are crew incidents, technical incidents or operational incidents. In both schedule management and crew control, if sufficient capacity is available, the crew schedule changes accordingly. Additionally, crew members and hotels are informed, and flight proceedings are updated.

3.2 Reality of existing Data

In order to build a process mining model of the overarching process, it is essential to have data inside each sub-process. However, some departments responsible for a sub-process, do not have their data recorded in a database. This unfortunately means we were unable to build and visualize the entire crew management process in our data model. Hence, we had to focus on a smaller part of the process, namely the remaining three sub-processes crew plan scheduling, crew plan updating and crew control. Unfortunately, the crew capacity planning does not leave any valuable trace of data, which means this sub-process could not be incorporated in this project. This is due to the fact that the scheduling is still partly completed manually via Excel. Thus, it does not generate the data points necessary to build a process mining data model. Therefore, we were forced to adapt our project goal: We were not able to generate end-to-end transparency for the entire crew management process. Instead we created transparency for the three sub-processes *Crew Scheduling*, *Updating Crew Schedules*, and *Crew Control*.

4 Use Cases

4.1 Derivation of Use-Cases

After understanding the crew management process and conducting the expert interviews, we tried to identify the most important pain points and KPIs to derive optimization potentials for every process. These potentials were then sorted into a general category or one of the crew management sub-processes and divided into the two dimensions *Increase Efficiency* and *Increase Effectivity*. In total, our team was able to collect 12 optimization potentials, which represented possible use cases for different sub-processes in the crew management process. Considering the feasibility of addressing all the 12 use cases, we conducted an expert workshop to identify the most impactful and relevant problems which would be the use cases we work on for our project.

At the end of the discussion we were able to identify one use case for each sub-process. The use cases we determined together with the experts are *Standby-Crew utilization, Distribution of Duty Days, Crew Plan Rejection Rate and Automation Rate* and can be seen in the figure 3 below in the colors green, pink, and yellow. The blue-colored use case represents our initial project goal of achieving end-to-end transparency in the crew management process. After creating our data model, we realized we did not have data points connected to rostering plan rejection. Therefore, we were forced to remove it as a use case. Going forward, we will describe the use cases we were able to tackle.

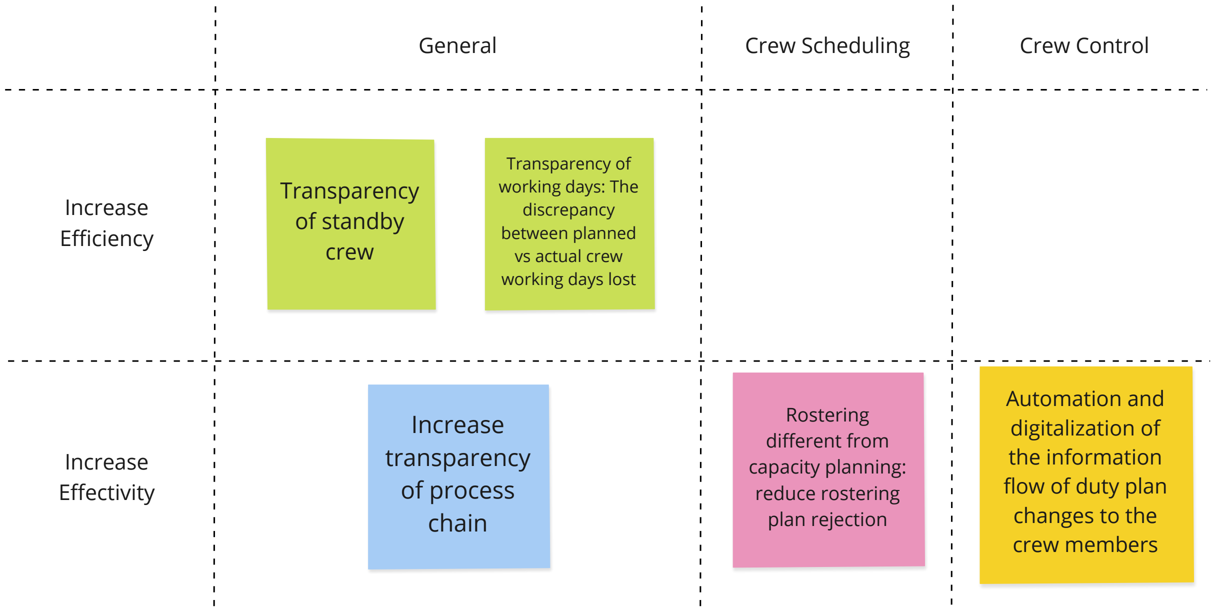

Figure 3: Use Cases

4.2 Description of Use-Cases

A general problem that companies across all industries face is the lack of transparency between the planned and actual processes. While the crew management departments have an in-depth process description, they are currently unable to perform process analysis and identify optimization potentials because they do not have transparency across the whole crew management process.

Thus, our main project goal at Lufthansa CityLine is to use process mining to create a Celonis data model that aids in achieving end-to-end transparency in the crew management process. In addition to this, there are two more use cases in the *General* category: Standby crew utilization and distribution of duty days. Therefore, one general use case of our team is to create a dashboard in Celonis showing the distribution of tasks of employees during their duty days. The different tasks include duty days working on an airplane as well as ground days which are used for trainings or meetings. Particularly the number of days spent for standby and how often standbys were actually transferred to flight duty was of great interest for Lufthansa CityLine. Detailed information about the cities in which the standbys were positioned as well activation rates of standbys was also highly welcomed by the product owners.

Our last use case belongs to crew control and aims at finding automation and digitalization potentials. After the publication of the crew plan, crew control takes over and is responsible for handling changes in the crew schedule in the short-term up to three days before the flight departure. In case of an incident (such as weather, sickness or technical problems) it is not only the responsibility of crew control to renew the crew schedule but also to notify crew members that are affected by the crew schedule changes. The standard process for the publication of the crew schedule is already automated: crew members receive their monthly schedule per mail and hotels and crew buses are automatically booked one month in advance. However, if the crew schedule requires changes due to the reasons mentioned above, the majority of changes are currently processed manually.

As a consequence, additional hotels have to be manually booked in the short-term by members of the crew control. Another manual process that is even more time consuming is notifying crew members via phone about changes in the crew schedule. Up to hundreds of crew members can be affected by crew schedule changes and most of them are currently contacted manually, which means that crew control has to make several phone calls to inform the crew of the duty changes. This is not only additional manual effort but also a very time intensive task, which makes the whole process inefficient. Hence, our use case for crew control is to visualize and specify automation and digitalization potentials for the information flow of duty plan changes to crew members.

5 Process Mining Implementation

This chapter elaborates on how the process mining project for the crew management process has been implemented within Lufthansa CityLine (CLH).

All following tasks have been performed on the CLH SQL Server with SQL statements to create all tables necessary for our analysis. Due to the setup of the CLH SQL Server, the Celonis on-premise system can connect to the SQL server. Hence, whenever we create new tables on the CLH SQL server with a specific schema name, they become available in Celonis as well. Fortunately, the CLH team already had a database with all relevant raw tables and data. Most of the tables had the crucial data properties of ID, activity code and timestamp in place, which are needed for process mining. Furthermore, we also received a rough mapping of the activity codes with their corresponding activity names from the CLH team. In conclusion, the raw tables include a TLC, a timestamp including begin date and begin time, and an activity code. The TLC is a pseudonymized employee number. This information allows us to track a crew member across all their assigned tasks for each duty day without including any personal information about the crew member.

A similar process mining model has already been validated with a subsidiary of Lufthansa. Thus, it was possible to benefit from their project output, for example through templates for analyses. However, there were some differences not only in the database itself but also in the data structure. Specifically, our raw tables contained different columns and some tables employed in their process mining model were not available to us.

5.1 Data Analysis and Preprocessing

In order to setup a process mining model it is important to understand the underlying data. Therefore, we worked with a MySQL database containing crew management data to get detailed insights. The database of CLH for crew management is called *Netline*. Within the *Netline* database there are 8 different tables with information ranging from hotel bookings to vacation or sick days of employees, up to the respective flight events. The data we received spans from mid 2018 to now. We performed data cleaning by categorizing columns into larger categories, which also accounted for human errors made in recording data (e.g. spelling errors). This was done in accordance with CLH experts so assure categories correctly represented the data. We also added tables not included in our *Netline* database to make our data more comprehensible (e.g. adding airport coordinates to map airports). In order to understand the data, SQL queries have been performed to understand the data structure and data type of the underlying data. As an example, a part of the content of the raw event table from *Netline* is illustrated below. This table was one of our main sources of information. As visible in figure 4, the *Netline* event table contains many columns. Of biggest importance to us was TLC, begin date and begin time as well as the column type. While the TLC allowed us to identify a crew member, the begin date and begin time helped us to chronologize events. The column types were used to specify the type of event. The types are very general and contain 8 categories, including *Leg*, *Absence* or *Simulator*. Within CLH a *Leg* refers to a completed flight. Also visible from figure 4 are the differences in data types. In the *Netline* table data points are stored as integers, varchars and datetime formats.

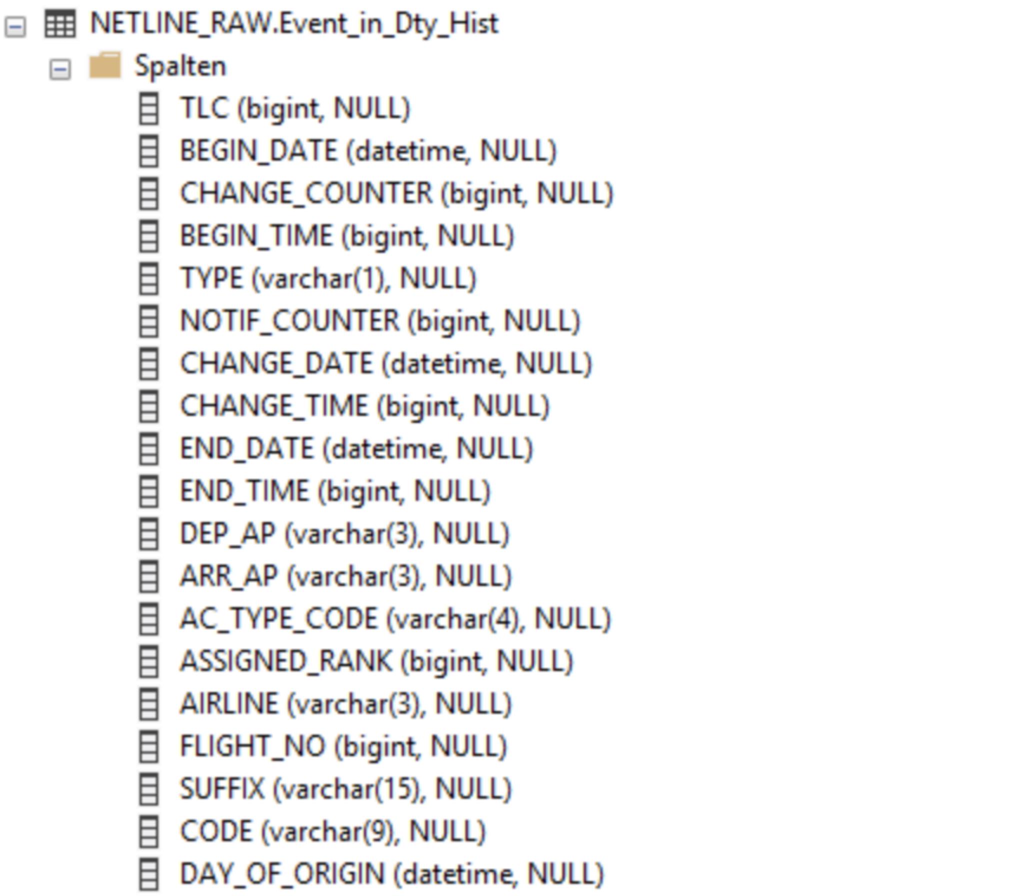

Figure 4: Contents of the Raw Event Table from Netline

In order to get the data into the right shape for process mining it was essential to perform extract, transform and load (ETL) tasks on the data. Not all tables could be used in their raw format, therefore we had to create a set of new tables based on the raw tables. For this, we extracted the necessary columns, thereby also reducing the volume of our set of tables. Next, we performed transformations to standardize the tables to achieve a consistent data set.

First, inconsistencies in the format of dates in the tables would have lead to inconsistent *Case Keys*. *Case Keys* are the unique IDs to trace employees across their duty days. We create one case key for every duty day of every employee. To prevent inconsistencies in our *Case Keys*, we standardized the date format with one of the SQL-supported formats: *yyyy-mm-dd*.

After standardizing the date format, we concatenated it with the time. This allows for a more detailed process flow in the data model, allowing us to track each employee across a given day, with a chronological order of all events due to the precise timestamp. The time was stored as a normal integer value such as "245" representing 02:45 am or "2309" symbolizing 11:09 pm. The dates and the times were concatenated in the final datetime format *yyyy-mm-dd HH:mm:ss*.
Furthermore, it was essential for all tables to have one column in common to be able to join the tables together and infer their relationship to one another. This was accomplished by adding the *Case Key* column to most of the tables. The *Case Key* was a concatenation of the TLC, which is an integer value, with the begin date. Thus, the *Case Key* has the format *TLC_yyyy-mm-dd*. Tables such as the Cancellations table, which contains cancelled flights, did not need to be joined via the *Case Key*, but rather through the flight number.

This resulted in a consistent data set across eight tables containing all important information for every event in the crew management process.

5.2 Creation of the Celonis Data Model

5.2.1 Creation of the Activity and Case Table

Based on the tables we extracted from the *Netline* database, we were able to create the cases table and the activity table; the two most important tables in a process mining analysis.

The activity table contains every activity executed. Our activity table contains over 3.4 million distinct activities. Based on the mapping provided to us by CLH, the initial activity table contained the following columns: ID, activity name, timestamp, flight, departure and arrival. A template for the activity table, which was used in the past for another process mining project in one of CLHs subsidaries, contained numerous restrictions: i.e. only specific types of flights, duties or captains with a specific rank. This served as the basis for our first analysis of the crew management process. However, after receiving feedback from experts from CLH, especially regarding the internal codes used at CLH, we wanted to add more details to enhance our data model. It was an iterative process to add more activities as information was made available to us over time. We wanted to present the crew management process in as much detail as possible in our process mining analysis. In the final analysis, many activities from the template were removed and replaced by more precise ones which better represent the process flow. Therefore, the final activity table contained additional columns such as longitude and latitude coordinates of all European airports targeted by CLH and a duty code category, which describes the duty activity in more detail such as proceeding flight, vacation, sick days etc. This greatly increased the amount of detailed information we were able to provide in our analyses, as we were able to target specific events through the duty code categories.

The cases table contains every case we examine in the process mining model. Our cases table consists of over 1.3 million distinct cases. Every case contains all activities one crew member completes during one duty day. Thus, the cases table has one entry for each crew member on every duty day. Therefore, each case can be identified via its distinct *Case Key*, meaning the cases table creates a list of all distinct case keys. An inner join of the cases table with the activity table has been performed for each of the activities to achieve a consistent data set. This means we keep only the activities related to cases in our cases table. Thereby we created a 1:N relationship between the cases table and the activities table. The cases table contains further information, such as the home base of a TLC, the TLC itself as well as cancelled flights. This enabled our team to build relationships between the cases table and different tables. These relationships were necessary for a detailed analysis of the crew management process.

5.2.2 Creation of the Data Model

Since all relevant tables were created in conformance to Celonis guideline, it was possible to upload the tables directly into Celonis. There, the tables were connected by common data points to obtain the final data model. The data model with all relationships between tables is depicted below in figure 5.

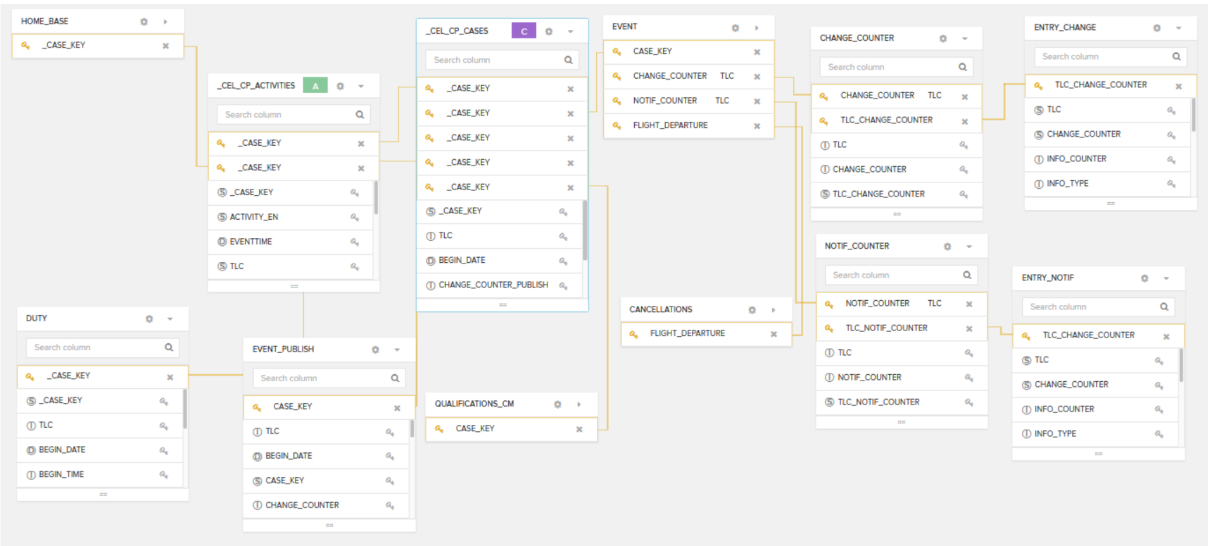

Figure 5: Tables and Relationships of our Data Model

Celonis only supports 1:1 or 1:N relationships among the tables. The relationships among the tables we inferred were verified by the CLH IT experts in a separate meeting. This ensured a proper setup of the data model, which was imperative to achieve a meaningful analysis. Due to the connections we realized among the tables it was possible to retrieve more detailed information about the activities, such as the name of a hotel where a crew member stayed or from which city a crew member took a proceeding flight to start their duty.

This data model, with all its internal relationships, is used within Celonis to visualize the process flow and to create the process analyses. Hence, the data model is a crucial part of process mining. During data model creation, we had to manually reload the model in Celonis whenever we implemented changes in the SQL server. However, once the data model was stable and any inconsistencies had been removed, we were able to automate data loads and ensure our analyses were always up-to-date.

5.2.3 PQL introduction and JSON import

After the data model has been successfully loaded into Celonis, the data becomes accessible in the Celonis front end, which are the analyses. Every analysis has to be set up by data engineers. The analyses are based on an own programming language called Process Query Language (PQL). PQL focuses on the process specific syntax but shows similarities to SQL. An example for a statement that would calculate the amount of a specific process flow from activity 1 to activity 2 would look like the following:

```
CASE WHEN PROCESS EQUALS "Activity 1"
TO ANY TO "Activity 2"
THEN 1.0 ELSE 0.0
END
```

With PQL, KPIs can be defined and retrieved from the data model. Once a KPI has been specified, one can make use of the KPI and visualize it on the Celonis front end. Therefore, the analyses serve as dashboards to see the historic changes of that respective KPI over the recorded time inside the data model or as a single digit on a daily basis - it depends on the setup by the data engineer who designs the analysis. Celonis allows for arbitrarily many KPIs and analysis sheets. Due to the similar work of the subsidiary of Lufthansa, it was possible to transfer the analysis sheets from the subsidiary as a template and adapt it to our use cases. However, the analysis was hosted in the cloud version of Celonis, and CLH has an on-premise Celonis version in their entire IT landscape. Nevertheless, it is possible to transfer analyses across different versions of Celonis. Hence, we exported a JSON file containing all the analysis descriptions from the cloud system and imported it into our on-premise version. This also allowed us to replace column names,column values, and color schema to fit CLH design inside the JSON file to reduce the amount of necessary changes. From that point on, additional analysis sheets were created with PQL statements in order to adapt the analysis sheets to the needs of our project.

5.3 Creation of the Celonis Dashboards

Based on the template we received from the Lufthansa subsidiary, we were able to create 12 dashboards. Because the underlying data of the template we received was different, we had to adapt the sheets to fit our data. Specifically, the dashboards concerning standby and proceedings were extended by us to include additional data available in our data model. In addition to the analysis sheets from the template, we created sheets tailored to our use cases. We also created a sheet to monitor data quality, which we reviewed with the Lufthansa crew management experts to validate our data model. In all dashboards we created, data can be filtered on attributes such as dates, home base, crew member rank or specific duty codes to name a few. Additionally, Celonis also offers the possibility to filter processes by a variety of criteria, such as flowing through a certain activity in a specific order. Due to the complexity and volume of the analyses, we want to offer a glimpse into a few dashboards. Before explaining our four most relevant dashboards in chapter 6.1, we want to show how our team created KPIs on which our dashboards are based. For the sake of simplicity, we will show the calculation of two KPIs related to the use cases identified in chapter 4.2 in the following chapter.

5.3.1 Standby Crew Utilization KPI

In order to check the standby crew utilization within the given data, we calculated how many people were activated given that they were assigned as standby. Hence, we calculated the ratio of cases where a standby was activated in comparison to the total amount of standby crew as the following PQL formula clarifies:

```
SUM(CASE WHEN
PROCESS EQUALS "Publish Plan: Standby" TO ANY TO "Legs" OR
PROCESS EQUALS "Publish Plan: Standby" TO ANY TO "Simulator" OR
PROCESS EQUALS "Standby" AND PROCESS EQUALS "Legs"
THEN 1.0 ELSE 0.0 END) /
SUM(MATCH_ACTIVITIES(NODE["Publish Plan: Standby", "Standby]))
```

This formula counts all activated standbys, which had to fly (= "Legs") and divides this sum by the number of published standby plans. For example, a captain is assigned as a standby for a given date. This means that the captain needs to be available at the respective airport within 3 hours after being notified that he needs to fly because a captain assigned to fly is unavailable. This would be reflected as one case in our data model where the standby member was activated.

Thus, we aggregate all such activities, where any standby crew member got activated, and divide it by the total number of standbys. This ratio is then being used to monitor the standby crew utilization across the recorded time frame. Furthermore, it is also possible to conduct a root cause analysis, to better understand why the activations took place, i.e. sickness of a colleague or long delay of another flight. This creates transparency and can aid in increasing efficiency and effectivity in the long-term.

5.3.2 Notification Automation Rate KPI

This KPI monitors the automation rate for notifications in the crew control sub-process. To calculate this KPI it is essential to understand how an activity is created in Celonis. Whenever an activity is system generated, the activity name will have *CIT* in its respective name. *CIT* means crew information terminal and is the CLH system that sends push notifications to employees affected by schedule changes via their smartphone. Thus, inside the activity table there is a data point to monitor whether a specific activity was performed manually by an employee (='User') or automatically by the system (='CIT'). This specific value can be used as an indicator for the automation rate. This is accomplished by aggregating all automated activities and dividing it by the total number of activities. Hence, the KPI notification automation rate is calculated by the following PQL formula.

```
SUM( CASE WHEN
PU_COUNT("_CEL_CP_CASES", "_CEL_CP_ACTIVITIES"."ACTIVITY_EN",
"_CEL_CP_ACTIVITIES"."ACTIVITY_EN" LIKE ("%notified by CIT%")) > 0
THEN 1.0
ELSE 0.0 END) /
SUM(CASE WHEN
PU_COUNT("_CEL_CP_CASES", "_CEL_CP_ACTIVITIES"."ACTIVITY_EN",
"_CEL_CP_ACTIVITIES"."ACTIVITY_EN" NOT LIKE ("%notified by CIT%")) > 0
THEN 1.0
ELSE 0.0 END)
```

This KPI enables CLH to get an overview of how automated the crew member notification process is after changes in the schedule occur. A higher automation rate in the notification process means less manual labor and less time invested. This can increase productivity in the crew control department.

6 Analysis and Optimization Potential of the Process

6.1 Analysis of the current Crew Management process

Due to the large amount of data we analyzed, our analysis is multi-facetted and complex. Our data model includes over 3.4 million distinct activities that Lufthansa CityLine crew members have engaged in since mid 2018. The activities included are all crew management processes that have been recorded via the *Netline* system. All following analyses and conclusions are based on this data.

The complexity of the relationships between activities in the crew management process can be seen in figure 6, which depicts the process explorer in Celonis. In the upper part of the figure, while 93 percent of activities are visible, only 60 percent of connections are shown. Here, the main process flows can be observed. In the lower part of the figure, 100 percent of the activities and connections are shown. Their complexity makes it difficult to grasp the process as a whole. Therefore, a short description of the process flow with reduced complexity will follow.

In the upper part of the figure, the most frequent activities and connections are visible. On the outer edges, activities such as off days, vacation, sickness, and ground activities can be seen. These are grouped under the category *Absence*. In the middle, the core processes are visible. In the section visible they are divided into two activities after the process start: *Publish Plan: Standby* and *Publish Plan: Legs*. This means that in the published schedule the crew member was planned to a standby duty or a flight duty, respectively. If a crew member was assigned to a flight duty, there are two possibilities as a following activity: either they fly their planned flights or they were notified that their duty had changed to an absence. If they did fly their flights they were most likely to have either a hotel stop or a flight to their home base as their next activity before the process ends.

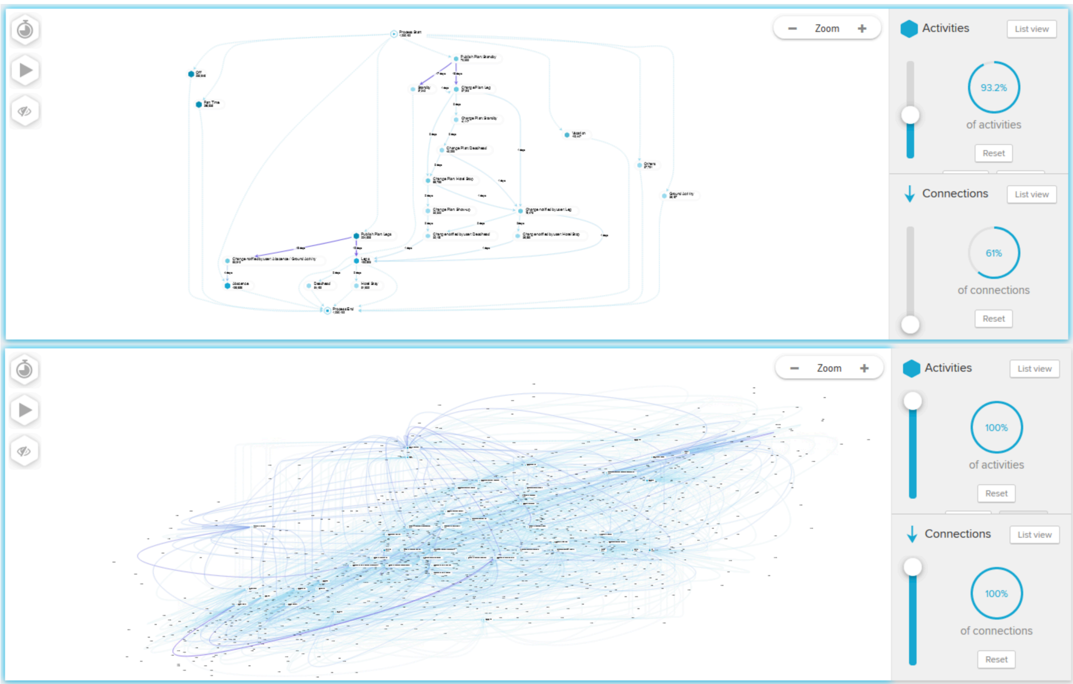

Figure 6: Process Explorer at Varying Complexities

In addition to the process explorer we created 12 dashboards. In this chapter, we will take a closer look at five dashboards. Two dashboards related to the KPIs explained in chapter 5.3, the Standby Utilization Dashboard and the Notification Dashboard, and two additional insightful dashboards, the Data Quality Dashboard, the Flight Proceedings Dashboard and the Duty Changes Dashboard. To ensure data security, all dashboard examples will be censored.

Our first dashboard Data Quality can be seen in figure 7 and can be used by the experts to get an overview of overarching KPIs as well as to verify the data quality of the data model and its underlying data set. The total number of *Leg Duties*, *Standby Activations*, *Cancellations* and *distinct TLCs* can be seen in the top left of the dashboard. The experts work with these KPIs on a daily basis, which enables them to assess the quality of our data model reliably. For example, the number of employees with flight duties, and thus registered in the Netline system, is equal to the number of distinct TLCs. Furthermore, the diagram *Leg Duties and Standby Activations per Month* can be used to get more specific insights about the number of leg duties and standby activations in each month and also gives an overview of their development over time. The map *Departure Airports* shows the different locations of departure airports in Europe and different colors indicate the number of departure flights per airport. As the number of flights from each airport are easy to track, this is a great indicator for data accuracy. Moreover, three other charts on the bottom of the dashboard give insights about the distribution of *Cockpit / Cabin Crew*, *Duties on Aircraft Types* and *Home Bases*. In addition, filters on rank, home base, leg chain and timeframe can be used to get insights into more specific cases.

Figure 7: Data Quality Dashboard

The second dashboard highlights standby utilization KPIs. As can be seen in figure 8, the dashboard depicts key performance indicators like *Standby Utilization*. It represents the percentage of standby crew members that were activated into duty. This is a useful indicator in capacity planning. If standby crew utilization as well as flight cancellations caused by missing crew members are low, the standby crew capacity can be decreased. Another important KPI is the *Standby Activations by Leg Duties*. This number shows the percentage of crew members that were activated from standby to fly in comparison to total flights flown. A lower percentage of standby activations means relatively fewer crew members had to be replaced with standby crew. In addition to these KPIs, it also shows the *Share of Standby Crews* (long-haul and short-haul flights), the *Standby Activations by Reason*, the number of *Flight Cancellations*, and the *Standby Activation filtered* which shows the utilization rate of standby crews for different filters, such as home base or weekday. Also visible are a process explorer, a column graph describing the process flow around standby activation and a comparison of planned to activated standby crew. All this information will be extremely valuable in planning and adjusting standby capacity in the future. Additionally, the analysis includes a number of filters a user can apply to adjust time frame, crew member rank, or home base in order to customize the dashboards to individual needs.

Figure 8: Standby Utilization Dashboard

The third dashboard shows the KPI notification automation rate which was calculated with PQL as shown in the chapter above. The most relevant chart in figure 9 is the donut chart *Notification % User vs CIT*. It shows the distribution of the manual and automated notifications sent to crew members. A lower percentage of automated notifications via CIT means that the majority of the crew members are still notified via a manual call by crew control. In addition, the chart *User Notification Rate per Month* can be used to track this distribution over time. It also shows the number of leg duties in the respective month. Two other important numbers of the notification dashboard are *CIT Notice Before Leg* and *User Notice Before Leg*. Both numbers show how many hours prior to their duty start crew members are notified. By comparing both numbers, it can be seen that on average crew members are notified several hours earlier via CIT. The higher these two numbers are in general, the earlier crew members are notified about changes in their crew schedule plan. Furthermore, user notifications can be shown in more detail with different filters, such as home base or weekday, in the table *User Notifications by filter*. In addition, a process explorer as well as filters for time frame, rank, home base and leg chain are available to have a closer look at more specific cases.

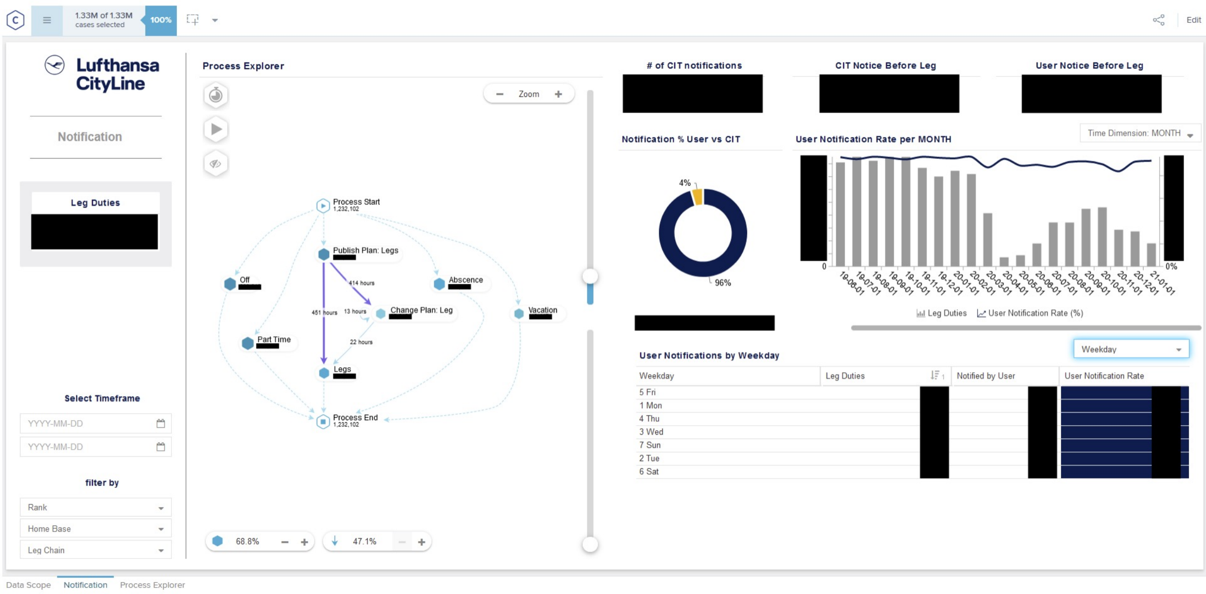

Figure 9: Notification Automation Dashboard

Our fourth dashboard can be seen in figure 10 and gives an overview of the flight proceedings. The top right of the dashboard highlights the most relevant information. It shows the total number of *Flights*, *Flight Proceedings* and *Flight Proceedings to Duty*. Flight proceedings are bidirectional and thus can be separated into *To Duty Flights* and *From Duty Flights* as crew members can take a flight from their home base to the starting point of their duty as well as in the opposite direction. Furthermore, information about the number of *Hotel Stays*, *Ground Transports* and *Ground Transport Hours* is displayed. Ground transports is a category we created which summarizes among others taxis, trains and buses. Additionally, we were able to calculate the number of *Deadheads* as well as *Deadhead Hours*. Deadheads are proceeding flights for crew members within their duty to the location where their succeeding flight duty will start. The number of total duty days is displayed on the left. Furthermore, the diagram *Proceedings per Month* shows the number of proceeding flights per month as well as the number of hotel stays in the respective month. Lastly, the table *Proceedings by filter* offers information on the number of legs, ground transports and hotel stays for different filters, such as home base or weekday. The process explorer as well as filters such as time frame filters can be used to limit the data shown to more specific cases.

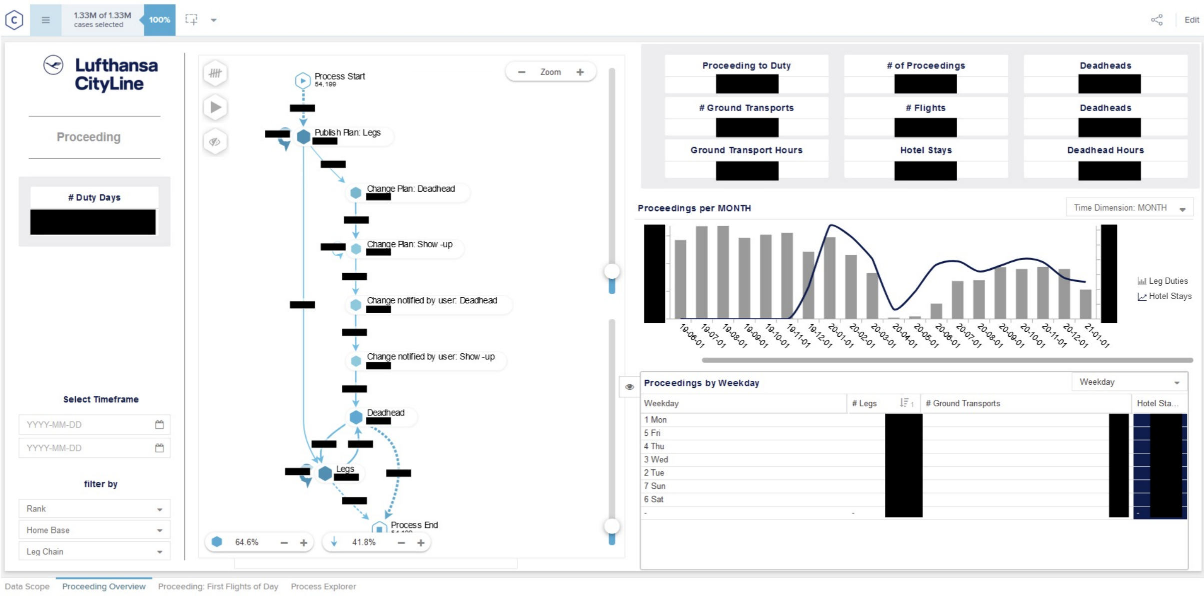

Figure 10: Flight Proceedings Dashboard

Another dashboard our team created is shown in figure 11. The dashboard Duty Changes can be used to get a detailed overview of the duty changes in comparison to the total number of duties, represented in the dashboard by the KPI *Changed Duties*. Another important figure is *Leg Duty Extensions* showing the percentage of leg duty changes resulting in a higher amount of legs than planned legs. A detailed explanation of Leg Duty Extensions can be found in chapter 6.2. More detailed information about duty changes is displayed in the diagrams *Changes from Plan to Duty*, *Leg Duty Changes* and *Leg Duty Extensions over Time*. The table *Drilldown by Filter* shows Duty Days, Duty Changes, Leg Duty Extensions, Change Rates and Extension Rates corresponding to the chosen filter. The Change Rate is the ratio of changes to duty days whereas Extension Rate is calculated by dividing leg duty extensions by duty days. Once our team created all 12 dashboards, we derived optimization potentials based on the dashboards in a next step.

Figure 11: Duty Changes Dashboard

6.2 Optimization Potentials of the Crew Management Process

Our team identified several opportunities for employees of Lufthansa CityLine in the crew management department to effortlessly exploit optimization potentials via our Celonis dashboards. One opportunity is provided by the standby dashboards which can be used to monitor, and later on optimize, the standby utilization in general as well as for specific periods of time or locations. The standby dashboard has been discussed in 6.1 and can be seen in figure 8. A low standby utilization means that many of the planned standby crew members are never activated for a flight duty. Thus, Lufthansa CityLine can try to reduce their standby capacity based on historical data to increase the standby utilization rate and to save labor costs. However, Lufthansa CityLine has to also take into account the trade-off between standby utilization and flight cancellations. Thus, while increasing the standby utilization, Lufthansa CityLine has to track the utilization on the Celonis dashboard to assure sufficient standby capacity to prevent flight cancellations due to a lack of standby crew members. Additionally, crew capacity planning employees are enabled to conduct root cause analysis on reasons causing the standby activations. Insights about whether the standby activations are caused for example by the absence of other crew members or by flight plan changes are of high value. These insights into the roots of standby activations can be used to implement measures to reduce the amount of standby activations and therefore, to be able to further reduce standby capacity.

The dashboards about flight proceedings provide another opportunity for optimization potentials. Our team created three dashboards dealing with flight proceedings of which we explained one dashboard in chapter 6.1. The total number of proceedings can be tracked and should be minimized in the long term. By reducing the total number of proceedings, Lufthansa CityLine can reduce their duty hours and therefore realize cost savings. Additionally, it can also reduce the number of flights and groundtransports taken by crew members as well as the number of hotel stays.

Furthermore, it is especially difficult for the crew scheduling department to receive feedback. In general, the better optimized the plan, the fewer changes to the plan are necessary. But the schedules are heavily influenced by unforeseen events such as bad weather and crew sickness. Therefore, a certain percentage of changes in the schedule is inevitable. Indeed, with our *Duty Changes* dashboard employees can transparently see in how many cases schedule changes occurred and also compare it to the total number of duty days.

For Lufthansa CityLine, changes in the leg duty are of particular importance as it requires high flexibility of the crew members in their work schedule. Additionally, the salary of a crew member heavily depends on duty time and thus also on the amount of legs flown. Therefore, it is in the best interest of Lufthansa CityLine that in cases of a leg duty change, the amount of legs for crew members stays the same. For example, if a crew member had a flight schedule with 3 legs, after the flight schedule change he/she should still have 3 legs in their schedule. However, if the crew member receives a new flight schedule with 4 legs, he/she will be paid overtime due to the additional leg duty. Hence, reducing the number of flight schedule changes with a different number of legs at a minimum could be of great interest for Lufthansa CityLine to optimize their crew management process and also represents a high saving potential.

Our last suggestion for an optimization potential can be seen in the notification dashboard in figure 9 which we presented in chapter 6.1. With the Celonis software, it is possible to differentiate between manual and automated processes. Our use case for crew control was to show automation and digitalization potentials specifically for this sub-process. Thus, our team was able to calculate the current automation rate of notifications to crew members regarding crew schedule changes. A higher usage of the CIT will save time and personnel resources spent for manually notifying crew members and this will make the crew control process more efficient. Additionally, crew members can get notified even earlier about duty changes by CIT, and thus require less flexibility in their work schedule. Celonis recently acquired a Czech startup focusing on automation, which has been incorporated in Celonis latest software version, called the "Executive Management System". This feature allows to automate specific activities, which would be a great fit to achieve higher automation rates on specific activities such as crew member notifications. An overview of the improvement potentials our team identified for Lufthansa CityLine can be found below in figure 12.

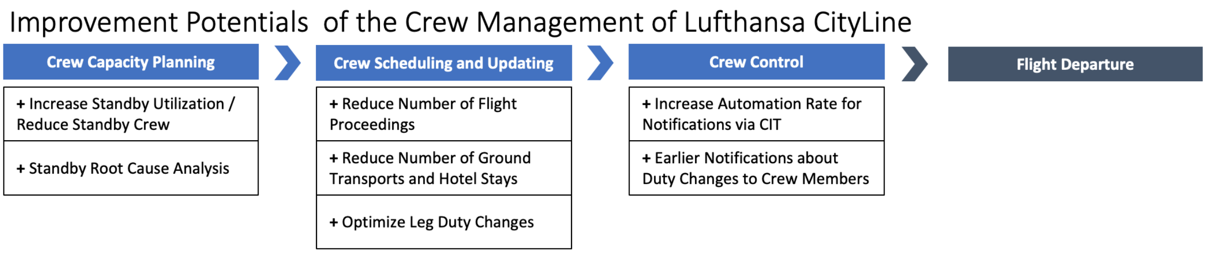

Figure 12: Overview of Improvement Potentials

7 Conclusion and Outlook

After understanding the complexity of the crew management and the requirements to increase transparency in the entire process we set our SMART goal of:

We will implement a process mining model in Celonis that encompasses transparency in the end-2-end crew management process of Lufthansa Cityline by analyzing and optimizing the entire process chain based on historical flight data until Feb 2021.

After having to reduce the scope to only three of the four sub-processes of the crew management process as well as suspending use cases that were not feasible due to data constraints, we achieved our overall project goal.

We successfully created a data model, implemented an analysis related to our use cases and beyond and were able to derive improvement potentials in many areas of the crew management process. Our team created transparency in the crew management process that will assist experts from all departments. The KPIs in our analyses show information that can serve as a basis for long-term process enhancement. The knowledge gain of our project output will serve as the foundation of process efficiency increases and we have created a solid base on which process mining in the crew management process can be further developed at Lufthansa CityLine.

Due to the short duration of our project, we faced several limitations. Missing data reduced the scope of the model and the analysis. In detail, the data did not cover the whole process that was intended to be analyzed. As a consequence, some use cases set at the beginning of the project were not feasible anymore, as the sub-processes they represented were not part of the data available to us. Additional data points that could have enhanced the analysis were not available. This includes the scope of data. While our dashboards give an accurate representation of ratios of, for example, crew members on duty and off duty, which is an indicator that our data model is correct, we noticed that it includes less flights than reality. This means we are missing data. This has been discussed with Lufthansa CityLine as is noted as a point of further investigation.

However, the data we received was sufficient to create an analysis covering most of the crew management process. Our analyses can serve as a foundation for both, first process enhancements and further additions to the crew management process analysis. Adding additional data would broaden the scope of the process analysis. This can be done both vertically and horizontally to enrich the content of the analysis. Working with the analyses we created for a longer time frame will also offer many improvement potentials. Monitoring KPIs and process steps over an extended time will allow for the implementation of initial optimization measures. This will be ideally supplemented by Lufthansa CityLine's intention to automate data loads to keep our analyses up-to-date. The automated data load will ensure ideal monitoring for Lufthansa CityLine experts in the future with the latest data. Initiating process mining in the crew management process at Lufthansa CityLine was the first step in the ongoing development to increase efficiency across the entire organization and we are confident our project output can foster augmentation in the long-term.

References

[1] Lufthansa. *Lufthansa CityLine GmbH Website*. https://www.lufthansacityline.com/de/home.html. 2021. (Visited on 12/19/2020).

[2] Celonis SE. *Celonis*. https://www.celonis.com/blog/celonis-a-leader-in-process-mining-assessment/. 2021. (Visited on 03/19/2020).

[3] Wil Van Der Aalst. "Data science in action". In: *Process mining*. Springer, 2016, pp. 3–23.

[4] Lars Reinkemeyer. *Process Mining in Action: Principles, Use Cases and Outlook*. Springer Nature, 2020.

[5] Henry CW Lau et al. "Development of a process mining system for supporting knowledge discovery in a supply chain network". In: *International Journal of Production Economics* 122.1 (2009), pp. 176–187.

[6] Vladimir Rubin, Irina Lomazova, and Wil MP van der Aalst. "Agile development with software process mining". In: *Proceedings of the 2014 international conference on software and system process*. 2014, pp. 70–74.

[7] Vladimir A Rubin et al. "Process mining can be applied to software too!" In: *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. 2014, pp. 1–8.

[8] Celonis. *Celonis customer Lufthansa CityLine crowned the winner for 'Data Powered Business' in IDC European Data Strategy & Innovation Awards 2020*. https://www.celonis.com/press/celonis-customer-lufthansa-cityline-crowned-the-winner-for-data-powered-business-in-idc-european-data-strategy-innovation-awards-2020. 2020.