

# CHAPTER 1

## Introduction

### 1.1 Introduction to Process Mining:

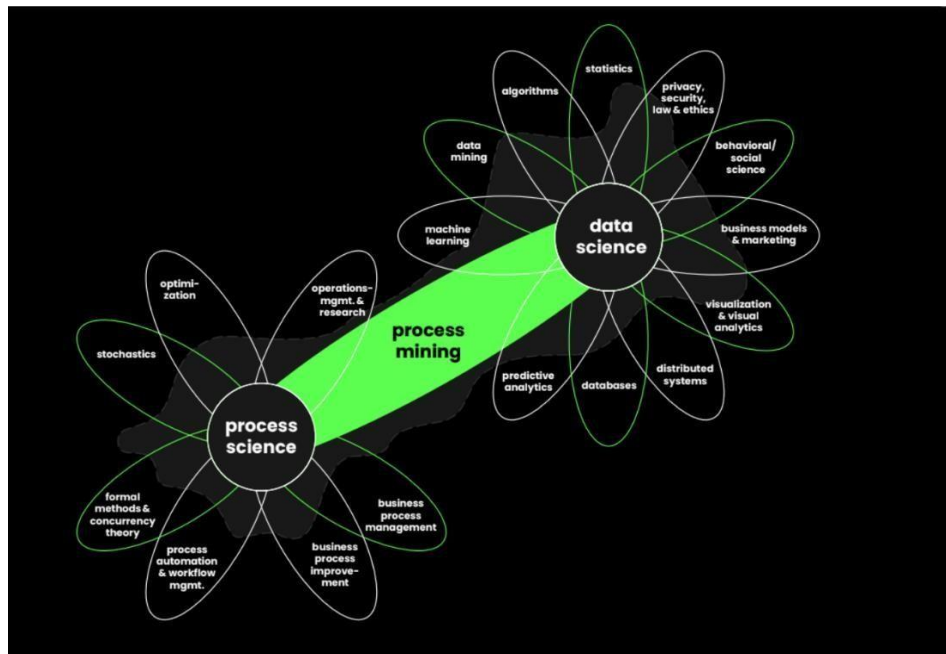
Process mining is the practice of examining processes, cases, and events logged by enterprise applications by using specialized data mining algorithms. It provides you data-driven insights to manage, monitor, and control your business processes. The insights from process mining include process models and statistics, model comparisons and conformance check with a reference model, process deviations, cost and time estimation, process delays and inefficiencies, resource mapping, and task analysis. In addition, the process mining helps you to make intelligent decisions for process improvements by creating process simulations based on the data-driven insights and analytics.

### What is Process Mining?

Process mining is a technique that examines a business's activities and enhances workflows for better productivity. It is an analytical system that observes, monitors and improves business functionalities. Process mining applies data science to discover, validate and improve workflows. By combining data mining and process analytics, organizations can mine log data from their information systems to understand the performance of their processes, revealing bottlenecks and other areas of improvement.

Process mining, also known as performance or organisational mining, is a multidisciplinary system that detects, tracks and optimises the activities in a company for a better yield.

Process Mining combines data science and business process management using advanced algorithms and digital tools to resolve multiple system-related issues. Algorithms can identify and analyse a process's patterns, details and trends and form insights to reduce inefficiency.



**Fig 1.1:** Process Mining

### Importance Of Process Mining:

Process mining can allow companies to streamline internal evaluations of daily operations using automated technology, meaning they can get faster, more accurate data. As a result, executives can discover key insights about each process, such as how long they take, how much they cost and what issues they help a business solve. Here are some more reasons process mining is important:

**Reducing expenses:** Process mining software can show executives whether a company spends money inefficiently. By reviewing this information, they can better manage resources and streamline production steps.

**Increasing ROI:** Reducing costs associated with business processes and removing ineffective steps can allow companies to improve their return on investment (ROI). This allows companies to improve their overall revenue and profits.

**Enhancing quality:** This method can help companies improve the overall efficiency of their internal processes, which may encourage staff members to think more creatively when developing new project ideas. Companies may also receive more high-quality data and better customer management

## CHAPTER – 2

### Foundations of Process Mining

**Process mining** is an analytical discipline for discovering, monitoring, and improving processes as they actually are and not as you think they might be. Process Mining works by extracting knowledge from event logs (also called digital footprints) readily available in today's information systems, in order to visualize business processes—and their every variation—as they run.

#### 2.1 Review and interpret Analysis

##### 2.1.1 Get to know Celonis Analysis:

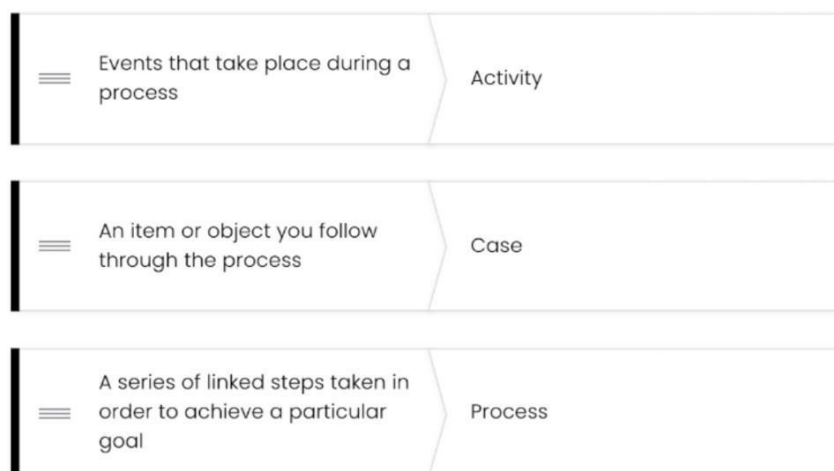
The Celonis Execution Management System (EMS) extends process mining by executing on insights automatically and orchestrating your existing technologies.

##### Process, Activity, Case:

**Process:** A series of linked steps taken in order to achieve a particular goal

**Case:** An item or object you follow through the process

**Activity:** Events that take place during a process



**Fig 2.1:** Process, Activity, case

Before interacting with your organization's process analysis, you'll want the right perspective on what makes business sense. With that, think about the following:

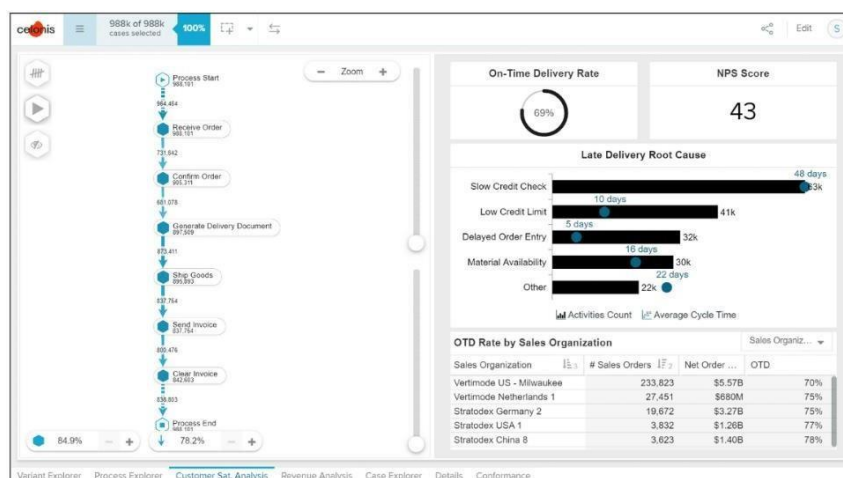
- Think about your process' ideal set of activities, then list them in the desired order they should take place.
- Write a list of undesired activities that negatively impact your organization's goals so that you can spot them in the analysis and then, using Celonis Analysis tools, identify potential root causes.
- Another angle to consider: Is it possible that a case goes through desired activities but not in the right order?
- Do some cases go through one activity more than once? Which activity or activities are usually associated with "rework?"
- Is it possible a crucial activity is missing from some cases' journey?

## 2.1.2 Navigate to an analysis

This is one of several courses that together form the "Review and Interpret Analyses" and the "Monitor KPIs in Analysis Dashboards" training tracks.

### 2.1.2.1 What is in an Analysis?

Once you have accessed the analysis, you may see anywhere from one to several sheets in it. The person building the analyses creates each analysis with the specific user(s) needs in mind.



**Fig 2.1.2:** example of an process expert analysis with seven sheet

## 2.1.3 Use the variant explorer

### 2.1.3.1 what is variant explorer?

As the name implies, using the Variant Explorer, you can discover all the process variants—that is all the different ways the process flows in your organization.

The Variant Explorer is one of the Analysis tools to help you take an "exploratory" approach to find out how your process is performing.

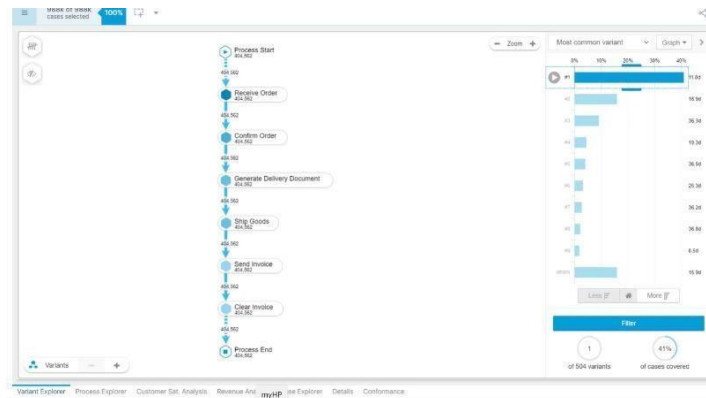


Fig 2.1.3: Variant explorer

## 2.1.4 Use the process Explorer

### 2.1.4.1 what is process Explorer?

- The Process Explorer is another analysis tool to use when taking an exploratory approach. It's especially useful for quickly revealing activities beyond the most common ones.
- It also allows you to narrow your focus on a single activity especially useful for quickly revealing activities beyond the most common ones. It also allows you to narrow your focus on a single activity.

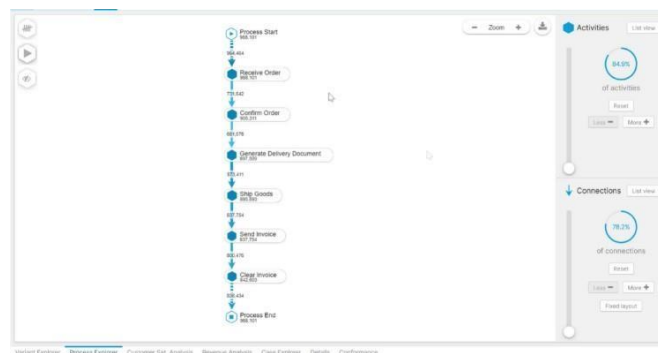
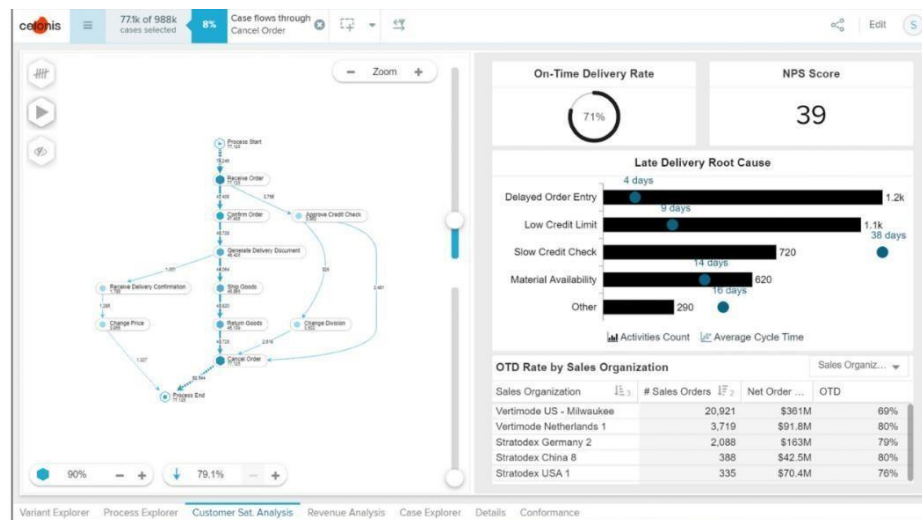


Fig 2.1.4: Process Explorer

Process Explorer is not showing the process path of any particular group of cases. It is showing the most common activities and connections, by default based on Case Frequency.

Once you learn about working with charts and tables, you'll find even more use cases for the Process Explorer. For example, after using the Process Explorer to filter on cases associated with an undesired activity (such as "Cancel Order") you can then drill into charts and tables to identify associated attributes (such as Sales Organization or Customer).



**Fig 2.1.4.1:** Process Explorer as a Component Along with Charts and Tables

### Through Process Explorer we learnt about:

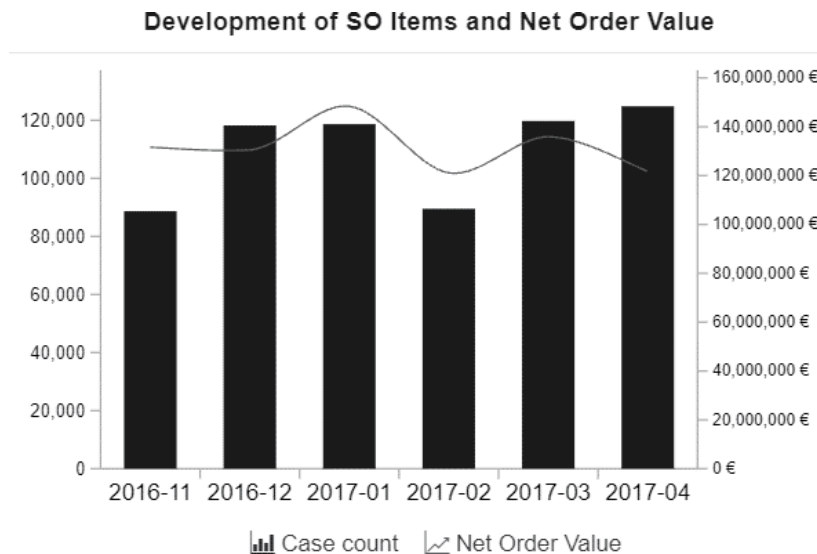
- Interpret the variation in case frequency numbers noted on activities and connection lines.
- get a transparent perspective on the flow of activities into and out of a single activity by reviewing where cases are coming from and where they're going to.
- control the activity and connection coverage beyond the initial ones to quickly surface potentially undesired activities and connections.
- filter on activities and connections and switch the KPI displayed in a similar fashion as in the Variant Explorer.

### 2.1.5 Use Charts and Tables review KPIs

### 2.1.5.1 What are Analysis Charts and Tables?

- A dimension is a category of attributes; for example, the dimension "customer name" is a category for individual customer names. Other examples of dimensions, depending on the nature of the process, can include vendor name, sales organization, region, and material group.

- **Key Performance Indicators (KPIs)** are used to calculate and add aggregated values; for example, case count, order value, invoice value, throughput time, and automation rate.



**Fig 2.1.5:** Development of SO items and Net Order Value

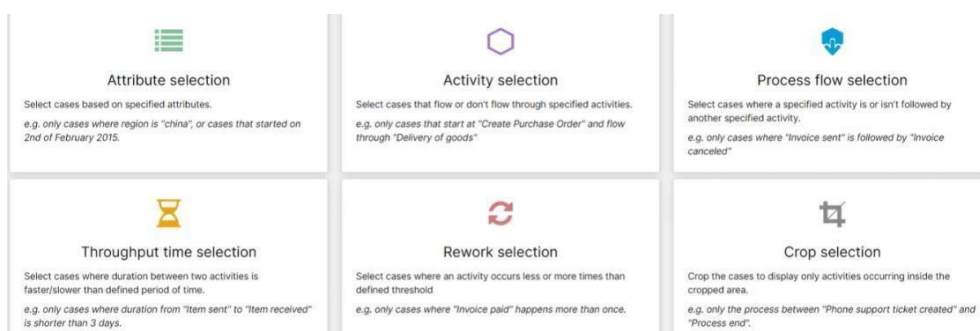
This chart shows the development of sales order items (KPI) and the corresponding net order value (KPI) over a period of time (dimension).

The x-axis displays the dimension, the creation date of sales order, grouped by months.

The two y-axes display the KPIs: The columns display the number of sales order items (case count) and the line displays the net order value.

### 2.1.6 Use Selection views

Selection Views offer a more comprehensive set of options to filter on cases as compared to what you can do in the components in analysis sheets.



**Fig 2.1.6:** Selection views

**There are six selection views**

- Attribute selection
- Activity Selection
- Process flow selection
- Throughput time selection
- Rework selection
- Crop selection

**2.1.7 Use the Case Explorer**

The Case Explorer is useful once you've narrowed down the analysis to a few cases that you want to investigate further. You can view specific case details such as timestamp of activities, user type (manual or automatic), possibly even user name (depending on your setup), and other useful pieces of info.

Case Explorer can also be used for single-case analysis once you have narrowed down your analysis to a few extreme cases. Additionally, if you're involved with process and data validation when a new process is brought on, you'll most likely end up using the Case Explorer

Case Explorer is not an OLAP table. It's a representation of raw data and it doesn't display KPIs.

**2.1.8 Use the Conformance Checker**

Every organization has an optimal process in mind that they want to achieve. With the conformance checker, you can see how far away the organization is from reaching that goal and investigate common patterns for inefficiency.

We can use conformance checker to

- get perspective on the percentage of cases that conform to the idea flow of activities specified in the Analysis (target process model).
- get automated insights into potential root causes of inefficiency by reviewing the most common process violations and the attributes they're associated with.
- mark certain process violations as acceptable to include them in the conforming cases statistics.



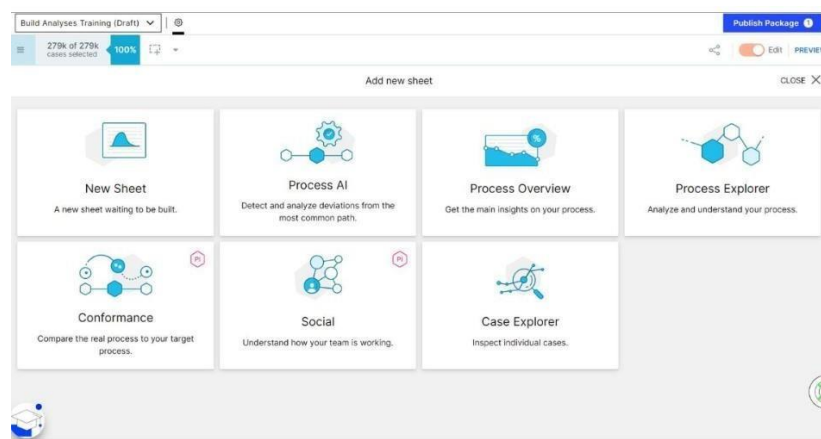
## 2.2 Build Analyses

### 2.2.1 Get ready to build Analyses

As Celonis Analysis builder, the analyses you create are the interface between the data and the end-users. You help make digital business processes transparent so that users in operational and leadership roles can make data-driven decisions.

Everyone become familiar with best practices in gathering user requirements as well as in data visualization—to make data understandable, easy to work with, and visually appealing.

Beyond uncovering process execution gaps and causes of inefficiency, some of our customers choose to use Celonis services such as process automation (using Action Flows) and Apps (using Knowledge Models and Views) to address inefficiencies and maximize their organization's execution capacity. You can keep them in mind as you think about business users' needs and organizational efficiency, and what can be addressed with analyses, and what should be addressed with other solutions.



**Fig 2.2.1:** Analysis Sheet

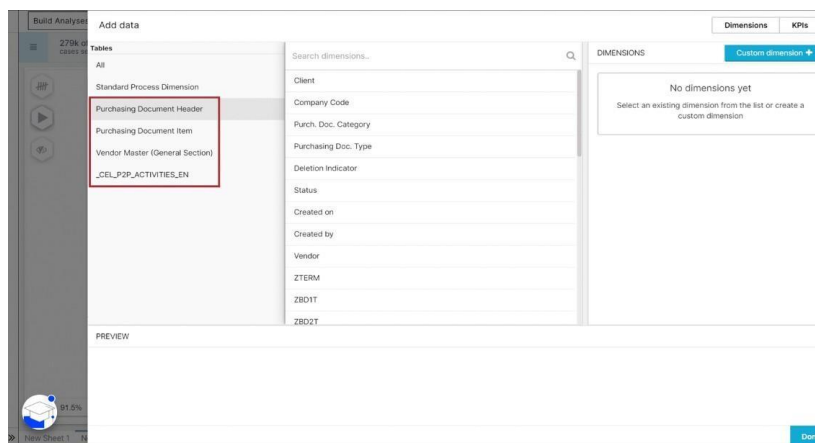
### 2.2.1 Add the first analysis sheet and publish

Go to the Analysis you previously created, and switch to "Edit" mode if you need to, in the upper right.

- Add a new sheet.
- Add the "Variant Explorer" component, then drag the edges to fill the entire sheet.
- In the component settings on the right, enter "Variant Explorer" in the "Title" field so it appears above the Explorer.
- Click the gear icon on the sheet name, then rename the sheet to "Variant Explorer."
- Click "Publish Package."

## 2.2.2 Configure tables and charts in analysis

To add dimensions and KPIs to an analysis component, you'll need to work with the data tables in the analysis. In the SAP Purchase-to-Pay (P2P) data model we work with for this training, we have four tables regardless of whether we're selecting dimensions or KPIs.



**Fig 2.2.3: Data tables**

- Purchasing Document Header (EKKO)
- Purchasing Document Item (EKPO)
- Vendor Master (LFA1)
- Activity Table—also known as the Event Log

The principle of creating table and chart components in Celonis is always the same. Although, you'll need to select **only one KPI** to display in a Pie Chart. Celonis Analysis includes four types of single KPI components. The most common use cases for the single KPI component include the case count and net value. For both, you would use the Number KPI.

### 2.2.3 Configure single KPI, Selection and Design components

Celonis Analysis includes four types of single KPI components. The most common use cases for the single KPI component include the case count and net value. For both, you would use the Number KPI.



**Fig 2.2.4: Single KPI componenets**

### 2.2.4 Configure standard process KPIs in the Visual Editor

Standard Process KPIs are generic KPIs that apply across any process you analyze. They are accessible in all Celonis Analyses, regardless of process, in the KPIs selection window. Standard Process KPI, Total throughput time in days comes in. By default, this KPI calculates the average total throughput time between the first and last activities of the process. But, using the Visual Editor, you can customize it to calculate the throughput time between specific activities in the process.

### 2.2.5 Configure custom KPIs in the Visual Editor

“Configure Tables and Charts,” you became familiar with the three simplified data tables we use in this training when configuring components that require dimensions and *KPIs*. These tables are:

- The Purchasing Document Header Table (EKKO).
- The Purchasing Document Item Table (EKPO).
- The General Vendor Master Table (LFA1).

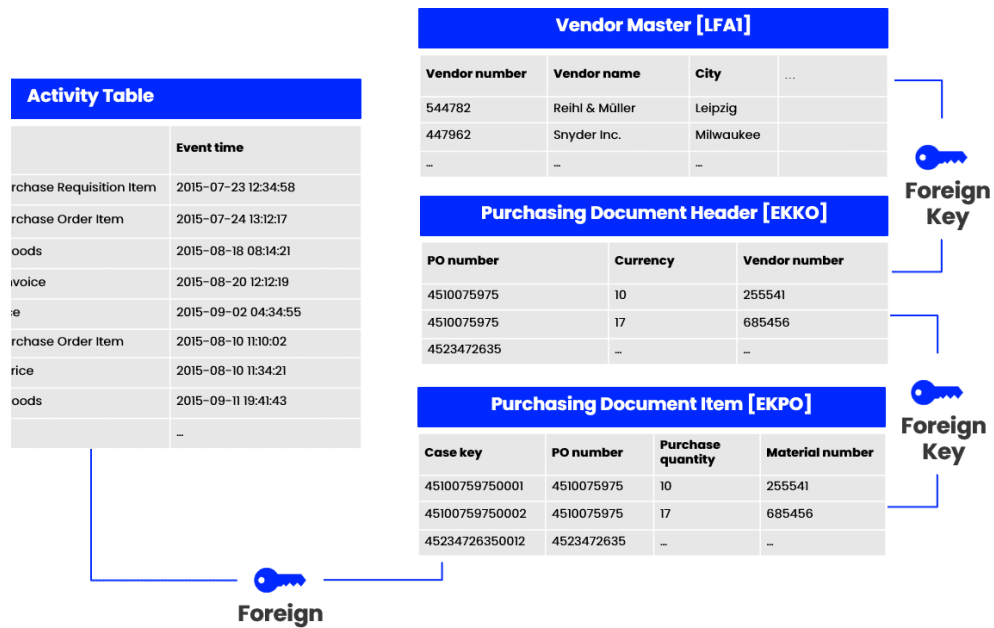


Fig 2.2.6: Visual editor

### What does Activity table contain?

At a minimum, the Activity Table contains these three columns.

- **Case**
- **Activity**
- **Event time**

**Case:** The “Case key” column specifies which object you follow through the process. As we are analyzing a Purchase-to-Pay process in this training, the objects we are following through the process are single purchase order items.

**Activity:** The “Activity” column contains all the activities that have been carried out for the purchase order items.

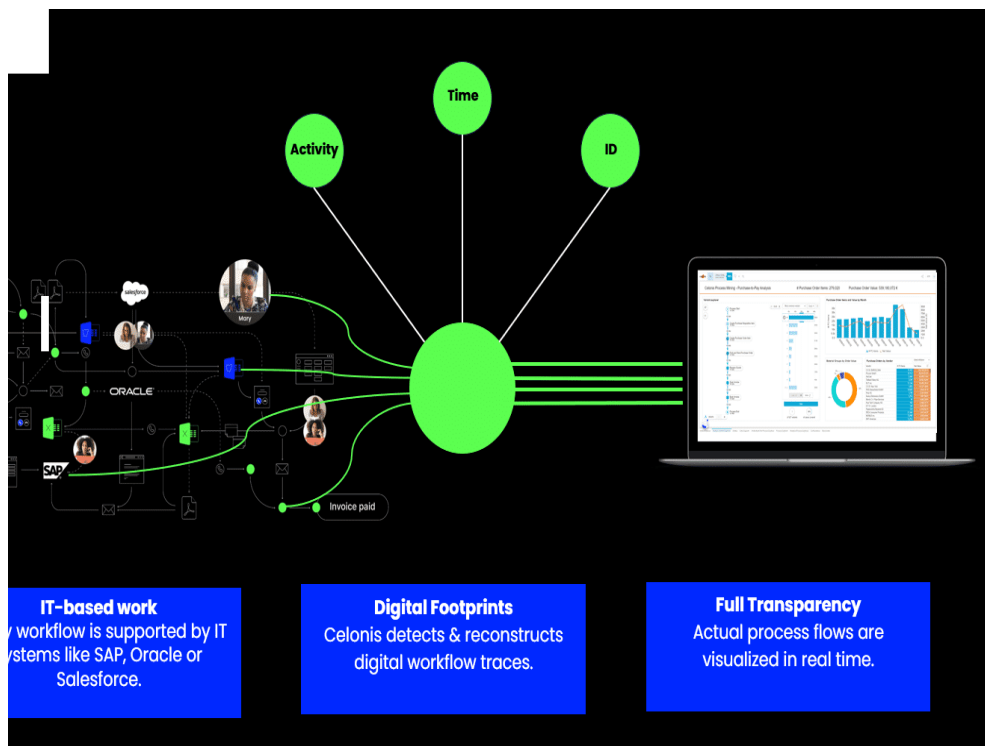
**Event time:** The “Event Time” column contains the point in time at which the activities have been conducted.

The Data Engineer can add more columns to the Activity Table with activity-specific information, for example, add a “User Type” column which indicates whether an activity has been conducted by an automated or a manual user. In an SAP system, automated users are

marked with a “B” while manual users are marked with an “A”. You can use the “User Type” to calculate automation rates.

### How is the Activity Table Created?

Unlike with the other tables in the data model, Data Engineers cannot directly extract the Activity Table from the source system. Instead, they identify the digital footprints in the source system and then consolidate them in the Activity Table. A digital footprint, at minimum, includes what has been done (activity), when it has been done (time), and which unique identifier it is associated with (e.g. sales order number).



**Fig 2.2.6:** Activity table

PQL allows you to further customize Analyses beyond what is possible with the Visual Editor. But remember that you can use PQL across several Celonis services, not just Analysis!

### 2.3 Case Study: Pizzeria mamma mia

We are looking at the digitization journey of the Pizzeria Mamma Mia from the perspective of Giovanni, the owner of the business, and Martin, his Junior Manager.

The focus will be on the Order-to-Cash process which is the core process of the Munich-based business. The journey starts with the digitization of all process steps, continues with the

discovery of inefficiencies and bottlenecks, and closes with recommendations for short-term enhancements as well as for sustaining the businesses success.

The case study is based on the Celonis Execution Management System (Academic version) and a data model provided by the Celonis Academic Alliance . The course will enable you to use Process Mining in a realistic environment in order to understand the functionalities and potentials of the Celonis EMS. You will learn how to build and interpret analyses in Celonis!

The Pizzeria Mamma Mia case study is an example of review analysis is built up in a modular structure consisting of different lessons and quizzes as displayed below. You have the freedom to jump directly to a specific lesson or do the full training in the recommended order. However, to pass the training successfully, you must complete all of the lessons and exercises.

## CHAPTER -3

### Enhance the analysis building skills and learn to get data into Celonis

#### 3.1 Write PQL Queries

##### 3.1.1 Basic queries in PQL

In the course of digitization, an increasing number of log data is recorded in IT systems of companies worldwide. This data is precious, as it represents how business processes are running inside a company. Process Mining comprises data-driven methods to discover, enhance and monitor processes based on such data. The heart of Process Mining are the Event Logs.

Those Event Logs are a collection of process events that can be described by the following attributes:

- Case
- Activity
- TimeStamp

##### **Case:**

The case attribute indicates which process instance the event belongs to. A process instance is called a **case**, usually consisting of multiple events.

##### **Activity:**

The activity attribute describes the action that is captured by the event.

##### **Time Stamp:**

Each activity leaves a digital footprint with a timestamp, indicating precisely when each event took place. With the help of timestamps, we know precisely in which chronological order the different activities have run off.

#### **Executable Queries in Process Mining**

To gain valuable process insights, it is essential for Process Mining users to formalize their process questions as executable queries. For this purpose, we present the Celonis Process Query Language (Celonis PQL), which is:

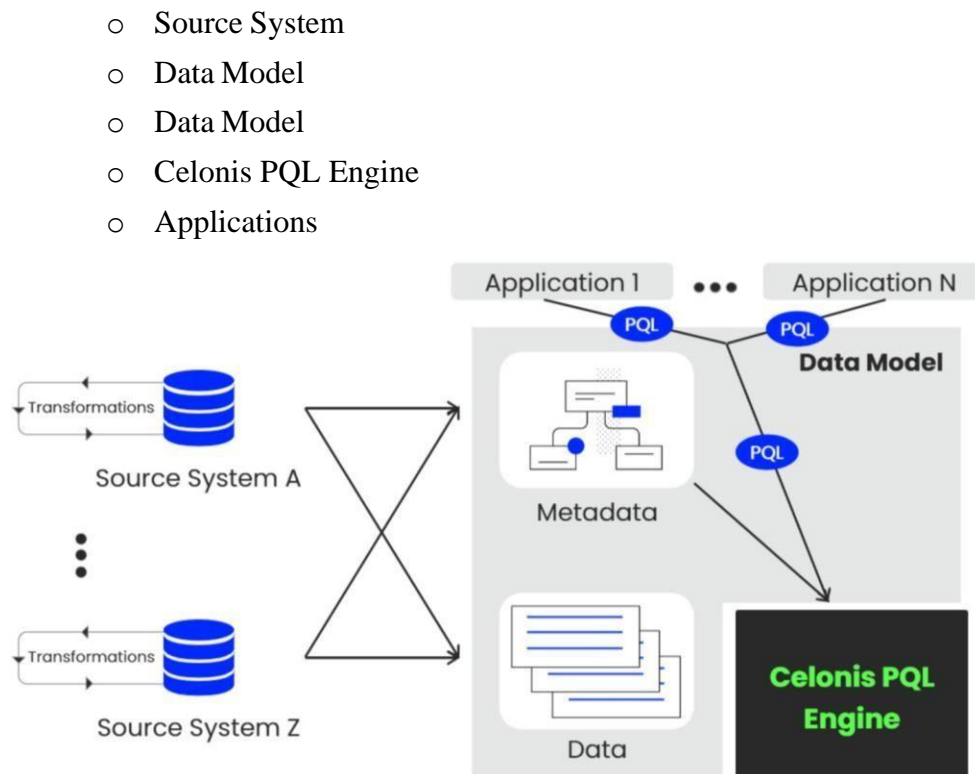
- a domain-specific language
- tailored towards a particular process data model and
- designed for business users.

It translates process-related business questions into queries and executes them on a custom-built query engine, the Celonis PQL Engine.

## Celonis Software Architecture

### The Celonis PQL Engine

As you can observe in the graphic below, Celonis PQL is an integral component of the Celonis Software Architecture. All Celonis applications use this language to query data from a data model. Hover over the different components in the architecture overview to find out more. Click on "example" to explore a data model of a procurement process.



**ig 3.1:** PQL Engine



## SQL vs. PQL

Even though Celonis PQL is inspired by SQL, there are major differences between the two query languages.

On a high level, Celonis PQL varies along four key dimensions:

- ☐ Celonis PQL does not support all operators that are available in SQL.

This is because customer requirements drive the development of the language, and only operators needed for the target use cases are implemented

- ☐ Celonis PQL is not supported by a data manipulation language (DML).

As all updates in the Process Mining scenario should come from the source systems, there is no need to manipulate and update the data through the query language directly.

- ☐ Celonis PQL does not provide any data definition language (DDL).

As the data model is created by a visual data model editor and stored internally, there has not been any need for creating and modifying database objects.

- ☐ In contrast to SQL, Celonis PQL is domain-specific and offers a wide range of Process Mining operators not available in SQL.

Consequently, Celonis PQL seamlessly integrates the data with the process perspective.

### **Across all classes of operators, Celonis PQL follows four language features:**

First, operators usually create and return a single column that is either added to an existing table (e.g., the case or activity table) or to a new, temporary result table.

But note, there are also a few operators that create and return one or more tables with multiple columns (e.g., for computing a process graph)

Second, the supported data types comprise STRING, INT, FLOAT, and DATE.

Third, Boolean values are not directly supported, but can be represented as integers.

Fourth, each data type can hold NULL values. Celonis PQL operates as follows:

**In aggregations:** treats NULL values as non-existing and ignores them.

**In row-wise operations:** returns NULL if one of its inputs is NULL.

### **Applications in the EMS**

PQL is the query language to formulate your process questions and calculate KPIs. This is why you can apply it in a multitude of applications in the EMS.

Then, Data Explorer is your place-to-be.

It offers not only the possibility to use a visual editor and switch between visual editor and Code editor, but also visualizations to validate your queries and to make sure that this is exactly the query you need.

If you're happy with the query, you can directly save it to your Knowledge Model.

The central place for all your records, KPIs, filters, etc. is the Knowledge Model. This is where you can add new PQL queries and save them for future usage across the different assets.

Whenever you configure Views, you use the PQL queries you defined in your Knowledge Model.

Although your analysis should ideally be connected to your Knowledge Model, you can also write individual PQL queries inside analyses.

When building Action Flows, sometimes you have to filter on specific subsets in your data or build a logic based on values in your data model.

Guess what? PQL is your tool of choice to phrase those filters or dimensions/attributes.

When working with ML Workbench and the Pycelonis package, you can interact with Celonis objects as native objects, e.g. copy an analysis, pull&push data, reload data models, etc.

Whenever you want to query specific data, PQL will be your friend.

### **3.2.2Get Data into EMS**

In this topic we will study about two types they are Set up a data pipeline Refine your Data Pipeline In the set up a data pipeline again divide into sub parts they are

1. Data Integration basics
2. Connect to Systems
3. Extract Data
4. Transform Data
5. Load a Data Model

In the Refine your Data Pipeline divide into parts they are

1. Schedule Data Jobs
2. Monitor and validate your Data pipeline
3. Multiple Process and Systems
4. Boost your EMS SQL Transformations
5. Connect Custom processes
6. Quality Assuring your Data Pipeline

#### 1. Data Integration basics

What is Data Integration for?

As a data engineer or analyst working in Data Integration (formerly known as Event Collection), you're responsible for bringing in clean, real-time process data into the EMS. In other words, you build the data pipeline. Computer Science and Engineering Page 10 Process Mining Virtual Internship

2. Connect to Systems Connecting to source systems is your very first step to pull process data into the Celonis EMS. The EMS utilizes a broad set of technologies like message queues, Restful APIs, Soap APIs, direct database access, or system-specific solutions to connect. Data Integration in the EMS Data Integration is where you set up connections and your datapipeline. The main ways you can bring data into the EMS are: Process Connectors Extractors (Data Connections) Extractor Builder File Uploads Data Push API Celox tractor

#### 3. Extract Data:

No matter which system you're working with when extracting data, it's a good idea to first understand the business process to know exactly which tables you need. Why don't we extract entire databases and make our lives simple? For simple reasons—entire database extractions would: take too long, be taxing on source systems, take up unnecessary cloud storage, and be expensive!

#### 4. Transform Data:

The Activity table represents your process and always contains at least these three columns that map your process:

The object ID or case key, the process steps or activities that took place for the different case keys and the timestamps or event time of each activity In the Purchase-to-Pay process, the Purchase Order Item Number is the central case key we follow. Every Purchase Order Item goes through different activities such as creating the request, creating the item, receiving goods, and paying the invoice. And every activity has a corresponding event time. In short,

every Purchase Order Item has a unique case key that goes through various activities at different points in time. Together these three columns build the core of your process flow.

5. Load a Data Model Just the Activity table on its own in a Data Model is not enough. To be able to drill down into case information, we need the Case table and other master data tables. As you know, in Celonis, the Case table is a table containing one row for each case. In other words, this table contains a row for each "process path" (a path following a case) being analyzed in the application.

## **CHAPTER-4**

### **Real time examples of Process Mining**

Financial services, telecommunications, healthcare, and retail are just a few examples of industries where process mining can be used for business process management and process improvement. These sectors have a wealth of data that can be used as a starting point, and process deviations from their intended behavior can have expensive repercussions.

#### **Financial Services:**

Because of the rise in transaction volume and the digitization of more industries, aberrant activity is harder to detect using manual methods. Companies in the financial services sector have the chance to continually and thoroughly identify issues within high-volume processes thanks to process mining, which is a solution to the increased regulatory and audit requirements.

#### **Telecommunications:**

As subscriber quantities increase and activations become more and more automated, there is a greater danger of unsuccessful activations. When telecom companies get more orders, process mining gives them the chance to identify pricey issues and client blowback in their Order-to- Activation processes.

#### **Healthcare:**

The risks associated with preserving population health and achieving individual patient journey objectives rise as data about patient experiences and results keep growing. Process mining supports the delivery of effective and high-quality end-to-end patient journeys for healthcare organizations dealing with the exponential growth of data, from before a first doctor appointment through treatment regimens to closed treatment cases.

#### **Retail:**

Due to technology or process problems, retail businesses have seen expensive consumer fallout from complicated e-commerce operations. Process mining assists merchants in ensuring that consumers can complete transactions efficiently and without issues despite rising transaction volumes.

#### **Digital Transformation:**

Process mining is frequently used in larger-scale digital transformation initiatives because it can give you the precise insights needed for process improvement, allowing systems to run

more quickly, smoothly, and efficiently, as well as objective data-driven insights into the causes of delays and inefficiencies within business processes.

As a result, process mining may assist in identifying the digital transformation opportunities with the greatest potential for value addition and determining whether or not transformation activities have really produced the desired results. To optimize returns on investments in projects for digital transformation, process mining becomes a crucial instrument

## **CHAPTER-5**

### **Learning outcomes of the internship**

- Get familiar with the technical and academic theory that forms the critical infrastructure of ProcessMining.
- Learn further the practical application of Process Mining with real business cases, data and insidertips from Celonis.
- Understand the most important concepts behind Process Mining and the respective softwarefunctionalities building on them and what they can be applied for in a business context.
- Learn the key process discovery and conformance checking algorithms.
- Study comparative and predictive process mining techniques allowing organizations to perform rootcause analysis of performance and compliance issues.
- Apply basic Celonis PQL queries containing conditions and basic functions
- Modify string input and timestamps for further usage
- Leverage basic process-related functions like PROCESS EQUALS, CALC\_THROUGHPUTand many more
- Capitalize on the different best practice and library sources for your future Celonis PQL usage

## **CHAPTER-6**

### **Conclusion**

In conclusion, process mining has the potential to bring significant benefits to organizations by improving efficiency, increasing visibility, and enabling data-driven decision making.

As technology continues to advance, process mining techniques are likely to become even more sophisticated and integrated with other data-driven approaches, further enhancing their ability to drive process excellence. However, successful implementation of process mining requires a comprehensive understanding of both the technology and the underlying business processes. Organizations that embrace process mining stand to gain a competitive edge by harnessing the power of data-driven insights to continuously refine their operations and achieve higher levels of efficiency and effectiveness.

By analyzing data from information systems, process mining can identify areas of inefficiency and help streamline processes. Process mining provides a detailed view of business processes, making it easier to identify bottlenecks and areas for improvement. Process mining can help organizations ensure they are following established procedures and regulations.



## CHAPTER-7

### Internship certificate



## **REFERENCES**

- **Login page:** <https://academy-login.celonis.com/s/login/>
- <https://academy.celonis.com/learn/course/introduction-to-process-mining/introduction-to-process-mining/course-outline?client=academic-alliance-celonis&page=2>  
<https://www.ibm.com/topics/process-mining>