การวิเคราะห์เส้นทางการเข้าชมในสวนสนุกด้วยเทคนิคเหมืองกระบวนการ (รูปแบบ ความผิดปกติ และข้อมูลเชิงลึกจากการวิเคราะห์บันทึกเหตุการณ์)

Unveiling Visitor Journeys in Theme Parks through Process Mining (Patterns, Anomalies, and Insights from Event Log Analytics)

เอนก นามขันธ์[1*] และ พาฮอม โพโรฮาน[2]

สาขาวิชาคอมพิวเตอร์ธุรกิจดิจิทัล คณะบริหารธุรกิจ มหาวิทยาลัยธนบุรี[1]

สาขาวิชาเทคโนโลยีสารสนเทศ คณะบัณฑิตวิทยาลัย มหาวิทยาลัยสยาม[2]

E-mail: anake_cc@thonburi-u.ac.th[1], parham@siam.edu[2]

บทคัดย่อ

การเปลี่ยนผ่านสู่ดิจิทัลที่เพิ่มขึ้นและความต้องการในการตัดสินใจโดยอิงจากข้อมูลที่มากขึ้น ได้นำไปสู่การประยุกต์ใช้เทคนิคเหมืองกระบวนการในหลากหลายอุตสาหกรรม งานวิจัยนี้นำเทคนิคเหมืองกระบวนการมาใช้เพื่อวิเคราะห์พฤติกรรมของผู้เข้าชมในบริบทของสวนสนุก โดยใช้ข้อมูลบันทึกเหตุการณ์จริงเพื่อค้นหาข้อมูลเชิงลึกเกี่ยวกับรูปแบบกิจกรรม ประสิทธิภาพของกระบวนการ และความผิดปกติในการดำเนินงาน

ปัญหาหลักที่งานวิจัยนี้มุ่งแก้ไขคือ การขาดความสามารถในการมองเห็นปฏิสัมพันธ์ของผู้เข้าชมแบบครบวงจร ซึ่งเป็นอุปสรรคต่อการปรับปรุงบริการและการส่งมอบประสบการณ์เฉพาะบุคคล วิธีการวิเคราะห์แบบดั้งเดิมมักไม่สามารถเปิดเผยจุดคอขวดที่ซ่อนอยู่ ความแตกต่างของรูปแบบ หรือเส้นทางที่ผิดปกติในกระบวนการไหลของผู้เข้าชมได้ ดังนั้นวัตถุประสงค์ของงานวิจัยนี้คือการใช้เครื่องมือเหมืองกระบวนการเพื่อดึงข้อมูลเชิงลึกจากบันทึกเหตุการณ์ ประเมินลำดับของกิจกรรม จัดกลุ่มรูปแบบ ระบุจุดที่ไม่มีประสิทธิภาพ และเสนอข้อเสนอแนะเชิงปฏิบัติที่สามารถนำไปใช้ได้จริงเพื่อการปรับปรุง

การวิเคราะห์ดำเนินการกับชุดข้อมูลที่ประกอบด้วย 100 กรณี และ 684 เหตุการณ์ ครอบคลุมกิจกรรมที่แตกต่างกัน 15 รายการ โดยใช้เครื่องมือ เช่น แผนผังกระบวนการ การวิเคราะห์ความแตกต่าง การจัดกลุ่ม และการกรองตามเวลา ผลการวิเคราะห์ที่ได้มีความน่าสนใจ ได้แก่ การค้นพบรูปแบบกระบวนการหลัก 2 กลุ่ม (คลัสเตอร์) ที่ครอบงำพฤติกรรมของผู้เข้าชม โดยแต่ละกลุ่มมี 13 กรณี นอกจากนี้ยังตรวจพบรูปแบบที่ผิดปกติ พร้อมกับข้อมูลเชิงลึกด้านความถี่ที่มีคุณค่า เช่น ความนิยมของกิจกรรม “มาถึงสวนสนุก” และ “ออกจากสวนสนุก” และเส้นทางที่พบได้ยาก เช่น “ไถลลงจากสไลเดอร์น้ำ” ตามด้วยการออกจากสวนทันที การวิเคราะห์ตามเวลาเผยให้เห็นว่า 93% ของกรณีใช้เวลามากกว่า 5 ชั่วโมง โดยมีเพียงกรณีเดียวที่ใช้เวลาเกิน 9 ชั่วโมง นอกจากนี้ การจับคู่กิจกรรมกับทรัพยากรที่รับผิดชอบทำให้สามารถเข้าใจการกระจายภาระงานและความไม่มีประสิทธิภาพของบริการที่อาจเกิดขึ้นได้

ประโยชน์ของงานวิจัยนี้อยู่ที่ความสามารถในการปรับปรุงแผนที่เส้นทางลูกค้า ปรับการจัดสรรทรัพยากรให้เหมาะสม และสนับสนุนการตัดสินใจเชิงกลยุทธ์เพื่อการยกระดับบริการ ด้วยการเปิดเผยรูปแบบและความเบี่ยงเบน ผู้จัดการสวนสนุกสามารถจัดสรรพนักงานได้ดีขึ้น ออกแบบโปรโมชั่นเฉพาะกลุ่ม และลดเวลารอของผู้เข้าชม

ขอบเขตของงานวิจัยนี้สามารถขยายไปยังสภาพแวดล้อมอื่นที่มุ่งเน้นลูกค้า เช่น ธุรกิจค้าปลีก การดูแลสุขภาพ และการศึกษา ซึ่งการทำเหมืองกระบวนการสามารถเปลี่ยนแปลงประสิทธิภาพในการดำเนินงานได้ งานวิจัยในอนาคตจะเน้นการเฝ้าติดตามแบบเรียลไทม์ การวิเคราะห์เชิงพยากรณ์ และการบูรณาการตัวชี้วัดความพึงพอใจของลูกค้าเข้ากับแบบจำลองเหมืองกระบวนการ เพื่อส่งเสริมกรอบการดำเนินงานที่ชาญฉลาดและเชิงรุกมากยิ่งขึ้น

คำสำคัญ: เหมืองกระบวนการ, พฤติกรรมผู้เข้าชม, การวิเคราะห์บันทึกเหตุการณ์, สวนสนุก, รูปแบบกิจกรรม, การจัดกลุ่ม, ความแตกต่างของรูปแบบ, การจับคู่ทรัพยากร, ระยะเวลากรณี, การปรับปรุงบริการ

Abstract

The rise of digital transformation and the growing demand for data-driven decision-making have led to the increasing adoption of process mining techniques across various industries. This research applies process mining to analyze visitor behavior in a theme park setting, using real event log data to uncover insights into activity patterns, process efficiency, and operational anomalies.

The primary problem addressed in this study is the lack of visibility into end-to-end visitor interactions, which hinders service optimization and personalized experience delivery. Traditional analysis methods often fail to reveal hidden bottlenecks, variants, or abnormal paths within visitor flows. Therefore, the objective of this study is to utilize process mining tools to extract meaningful insights from event logs, evaluate activity sequences, cluster patterns, identify inefficiencies, and provide actionable recommendations for improvement.

The analysis was conducted on a dataset comprising 100 cases and 684 events, covering a range of 15 distinct activities. Using tools such as process maps, variant analysis, clustering, and time-based filtering, several findings emerged. Notably, two core process patterns (clusters) were found to dominate visitor behavior, each involving 13 cases. Additionally, abnormal variants were detected, along with valuable frequency insights, such as the popularity of “Arrive to Theme Park” and “Leave Theme Park” activities, and rare paths like “Slide Down the Water Slide” directly followed by exit. Time-based analysis revealed that 93% of cases lasted over 5 hours, with only one exceeding 9 hours. Furthermore, mapping activities to

responsible resources enabled an understanding of workload distribution and possible service inefficiencies.

The benefits of this research lie in its ability to improve customer journey mapping, optimize resource allocation, and support strategic decision-making for service enhancements. By uncovering patterns and deviations, theme park managers can better allocate staff, design targeted promotions, and reduce visitor waiting times.

The scope of this study can extend to other customer-centric environments such as retail, healthcare, and education, where process mining can transform operational efficiency. Future work will focus on real-time monitoring, predictive analytics, and integrating customer satisfaction metrics into the mining models to foster a more proactive and intelligent operational framework.

Keywords: Process Mining, Visitor Behavior, Event Log Analysis, Theme Park, Activity Patterns, Clustering, Variants, Resource Mapping, Case Duration, Service Optimization

1. Introduction

1.1 Background and Industry Context

The digital transformation wave has significantly reshaped how industries optimize operational efficiency and improve customer experience. In the entertainment and leisure sectors, especially in theme parks, the ability to understand visitor behavior, service bottlenecks, and resource utilization is critical to enhancing operational excellence and delivering a memorable experience. With the increasing complexity of customer journeys and service processes, traditional analysis methods are no longer sufficient to uncover detailed process insights.

One of the most powerful approaches to understanding such complex service processes is Process Mining, which bridges the gap between traditional model-based process analysis and data-centric analysis techniques. By extracting knowledge from event logs readily available in IT systems, process mining enables organizations to visualize and improve actual processes based on real-world data. In this study, we focus on applying process mining to a simulated theme park environment using Disco, a professional tool designed for fast and intuitive process analysis.

1.2 Problem Statement

Despite the abundance of data generated in customer-facing environments such as theme parks, organizations often lack visibility into how their processes are executed. Visitor flows are complex, with various optional, sequential, and parallel activities. Managers struggle to pinpoint inefficient paths, resource overuse, and anomalies in service delivery. Without clear insight into these issues, improvements are often based on assumptions rather than data.

This study addresses the following core problems:

- Limited transparency of how visitors interact with various services and staff across the park.
- Unclear performance benchmarks, such as case durations and activity frequencies.
- Inability to quickly identify deviations from standard process paths or abnormal visitor behavior.

1.3 Objectives of the Study

The main objective of this research is to apply process mining techniques using Disco to analyze and interpret real-world process behavior captured in an event log from a theme park scenario. Specific objectives include:

- To identify and visualize the most common visitor journeys and process variants.
- To evaluate resource utilization by identifying which staff members (resources) performed which activities.
- To detect unusual or abnormal process executions and understand their causes.
- To quantify the performance of each process case in terms of duration, frequency, and bottlenecks.
- To enable evidence-based decision-making to improve service delivery and operational efficiency.

Fig. 1. Theme Park Overview.

Fig. 1. shows the theme park's layout used in this study, highlighting key attractions such as: thrilling roller coasters, spinning rides, a spooky haunted house, an immersive 8D theater, a classic carousel, fun bumper cars, refreshing water slides and wave pools, adrenaline-pumping drop towers, and a challenging escape room.

1.4 Motivation of the Study

This research is motivated by the need to empower service-oriented organizations, particularly in the entertainment and theme park industries, with data-driven tools to gain operational transparency. As digitalization expands, there is an increasing need for automated methods to audit, analyze, and optimize processes with minimal manual intervention. The ability to analyze complex behavior patterns in real-time, as enabled by process mining, is a key differentiator in customer-centric industries.

In educational settings, this study serves as a practical example of how academic learning can be applied to real-world scenarios through advanced tools and analytics methodologies.

1.5 Scope of the Study

The scope of this study includes:

- Dataset: A structured event log from a theme park, comprising 100 visitor cases and 684 events.
- Tool: Disco, a process mining software used for discovery, performance analysis, and variant identification.
- Analysis Dimensions: Process flows, time-based performance, resource assignment, activity frequency, and case duration.

The findings and outcomes of this study are beneficial to a variety of stakeholders:

- Theme Park Managers and Planners: To enhance visitor experience and staff allocation strategies.
- Business Analysts and Process Improvement Teams: To develop targeted interventions for optimizing operations.
- Academics and Students: To understand practical applications of process mining tools.
- Digital Transformation Consultants: To demonstrate the ROI of applying advanced analytics in customer-facing industries.

2. Literature Review

Process mining has emerged as a pivotal technique for analyzing and improving business processes across various domains. By extracting knowledge from event logs, organizations can visualize actual processes, identify bottlenecks, and enhance operational efficiency. The following studies exemplify the application of process mining techniques in diverse contexts:

Kurniati et al. (2021) conducted a study titled "Process Mining on the Extended Event Log to Analyse the System Usage During Healthcare Processes," focusing on the usage of the General Practitioner (GP) tab during chemotherapy treatments. By applying process mining techniques to extended event logs, the study identified patterns in system usage, providing insights into healthcare professionals' interactions with electronic health records. The findings highlighted areas where system usage could be optimized to enhance patient care.

Wagner et al. (2022) presented "A Combined Approach of Process Mining and Rule-based AI for Study Planning and Monitoring in Higher Education." The study utilized process mining to analyze students' study paths and combined it with rule-based artificial intelligence to provide personalized study recommendations. By integrating event logs from campus management systems with examination regulations, the approach supported students in planning and monitoring their academic progress effectively.

Mohammadi et al. (2022) explored the "Process Mining Approach to Performance Analysis and Bottleneck Finding in Electronic Processes," using the billing process of hospital services as a case study. Employing the Disco tool, the study discovered process models, analyzed performance metrics, and identified bottlenecks, particularly in the insurance verification step. The research provided actionable insights for process improvement in healthcare billing systems.

Thiyagarajan and Prasanna (2023) investigated "Process Mining-Based Behavioral Modeling of Learners in Self-paced Learning Environment." The study applied heuristic and inductive mining techniques to event logs from learning management systems, aiming to model students' learning behaviors. The results demonstrated the effectiveness of processing in understanding and enhancing self-paced learning experiences.

Saad (2023) addressed industrial challenges in "Application of Process Mining and Sequence Clustering in Recognizing an Industrial Issue." Focusing on the weaving process in manufacturing, the study utilized process mining and sequence clustering to identify bottlenecks caused by machine overloading. The analysis led to significant improvements in cycle time, worker performance, and product quality.

Liu et al. (2023) conducted a systematic review titled "Turning Logs into Lumber: Preprocessing Tasks in Process Mining." The study identified and categorized preprocessing tasks essential for effective process mining, emphasizing the importance of structured and transparent event log preparation to enhance the reliability of process mining outcomes.

In the realm of customer experience, a 2023 study titled "Exploring Customer Journey Mining and RPA: Prediction of Customers' Next Touchpoint" applied process mining techniques to predict customers' next interactions. By integrating robotic process automation, the study aimed to enhance customer journey analysis, providing businesses with tools to anticipate and meet customer needs proactively.

Cerezo et al. (2024) explored "Process Mining for Self-Regulated Learning Assessment in E-learning." Utilizing the Inductive Miner algorithm on Moodle platform event logs, the study discovered models of students' self-regulated learning processes. The analysis differentiated between successful and unsuccessful learning strategies, offering insights for instructional design improvements.

Nai et al. (2024) presented "Enhancing E-learning Effectiveness: A Process Mining Approach for Short-term Tutorials." The study applied process mining to user behavior data from short-term web-based tutorials, aiming to extract activity flows and predict learning outcomes. The findings provided descriptive insights into learning processes, informing the design of more effective e-learning experiences.

Halvorsrud et al. (2024) discussed "Customer Journeys and Process Mining – Challenges and Opportunities." The study examined the integration of customer journey analysis with process mining, highlighting the potential for enhanced service quality evaluation and the identification of areas for improvement in customer experience management.

In the study "Process Discovery and Process Optimization of Banking Industry through Visual Mapping and Simulation of Activity Sequences," Parham Porouhan (2023) applied process mining techniques to an anonymous private bank's customer complaint support network. Utilizing Fluxicon Disco, the research involved filtering and clustering techniques to generate process maps and simulate procedures. The analysis revealed bottlenecks, departmental inefficiencies, and rule violations, providing actionable insights for enhancing customer satisfaction and operational efficiency.

In "Applying Process Mining to Minimise Order Waiting Time of FitBox," Porouhan (2022) investigated delays in food delivery services during the COVID-19 pandemic. By cleansing the dataset and employing Fluxicon Disco, the study utilized automated process discovery, filtering, clustering, and simulation techniques. The findings identified key reasons for delivery delays, offering a foundation for future improvements in service efficiency.

The research "Process Modeling and Bottleneck Mining in MXML-based Course Training Event Logs" by Porouhan and Wichian Premchaiswadi (2022) focused on a project management training course at a private university in Thailand. Applying alpha and heuristic algorithms, the study reconstructed process models to analyze student performance. Results indicated that 80% of students achieved certification, while 6% failed after multiple attempts, highlighting areas for curriculum improvement.

In "Using IoT and Mobile Robots to Model and Analyze Work Processes with Process Mining Techniques," Premchaiswadi et al. (2024) explored the integration of IoT and mobile robots in data collection for process analysis. The study developed a system to store event logs generated by IoT devices and analyzed them using process mining techniques. This approach provided insights into individual user activities, demonstrating applications in sectors like healthcare, logistics, and manufacturing.

The study "Creating and Collecting e-Learning Event Logs to Analyze Learning Behavior of Students through Process Mining" by Nammakhunt, Porouhan, and Premchaiswadi (2023) addressed the need for data-driven approaches in online education. Transforming and preprocessing online student datasets, the research applied the Fuzzy Miner algorithm via Fluxicon Disco to understand learning behaviors. Comparative analysis using Dotted Chart Analysis revealed significant differences between high- and low-performing student groups, informing strategies for curriculum enhancement.

In "Emotion Analytics with Process Mining," Palangsantikul et al. (2019) combined process mining with emotion analytics to study students' emotional patterns using data from the Photographic Affection Meter application. Employing Fuzzy Miner and Dotted Chart Analysis, the research categorized emotional trends and provided frequency-based models of student behavior, offering a novel approach to understanding emotional dynamics in educational settings.

The paper "Data-Driven Business Process Improvement" by Arpasat and Premchaiswadi (2024) analyzed outpatient service processes in a hospital setting. Utilizing Fluxicon Disco and the Fuzzy Miner algorithm, the study examined event logs from 12,836 patients, uncovering 4,293 distinct process variants. The analysis facilitated optimization of resource allocation and development of effective staff training plans, demonstrating the efficacy of process mining in healthcare.

In "Process Mining and Learners' Behavior Analytics in a Collaborative and Web-Based Multi-Tabletop Environment," Porouhan and Premchaiswadi (2017) investigated collaborative dynamics in educational settings. Applying social network mining, performance analysis, role hierarchy mining, and dotted chart analysis, the study found that high-performance groups exhibited greater symmetry in actions and roles, suggesting more effective collaboration strategies.

The research "Analysis of Student Learning Behavior using Process Mining and Spectrogram" by Nammakhunt and Premchaiswadi (2024) employed Spectrogram Analysis, Correlation Analysis, and Multiple Linear Regression Analysis to examine factors influencing teaching efficiency. Analyzing event log data with the Fuzzy Miner algorithm and Python's Matplotlib library, the study revealed a positive correlation between attendance frequency and academic achievement, informing strategies for enhancing e-learning systems.

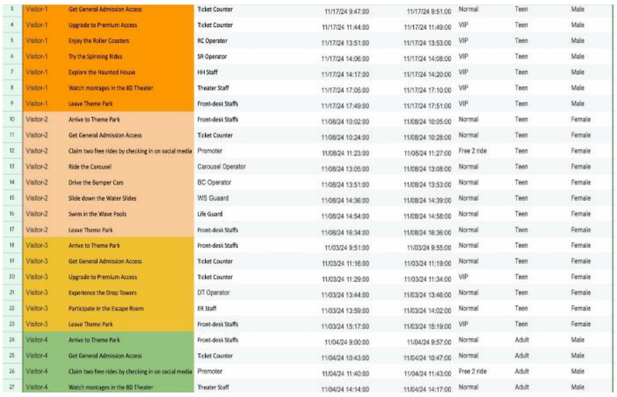
Finally, "Process Modeling and Bottleneck Mining in Online Peer-Review Systems" by Premchaiswadi and Porouhan (2015) focused on improving the efficiency of online peer-review processes. By capturing and formatting event logs from conference proceedings, the study utilized process mining techniques to discover process models and identify bottlenecks. The analysis provided insights into performance and waiting times, contributing to the enhancement of peer-review systems.

3. Methodology

This chapter has detailed the dataset's structure, preprocessing steps, and the mapping strategy used for process mining. By leveraging Disco, we aim to uncover hidden process dynamics, optimize visitor experiences, and generate insights into staff efficiency and operational flow within the simulated environment. The next chapter will present the results derived from the application of these methods.

3.1. Introduction

This chapter outlines the methodology employed in this study, focusing on the application of process mining techniques to analyze event data collected from a simulated theme park environment. The primary objective of the analysis is to uncover behavioral patterns, resource utilization trends, and potential process inefficiencies using the Disco process mining tool. The data used in this study was curated and structured specifically to support event log-based analysis.



3	Visitor-1	Get General Admission Access	Ticket Counter	11/17/24 9:47:00	11/17/24 9:51:00	Normal	Teen	Male
4	Visitor-1	Upgrade to Premium Access	Ticket Counter	11/17/24 11:44:00	11/17/24 11:49:00	VIP	Teen	Male
5	Visitor-1	Enjoy the Roller Coasters	RC Operator	11/17/24 13:51:00	11/17/24 13:53:00	VIP	Teen	Male
6	Visitor-1	Try the Spinning Rides	SR Operator	11/17/24 14:06:00	11/17/24 14:08:00	VIP	Teen	Male
7	Visitor-1	Explore the Haunted House	HH Staff	11/17/24 14:17:00	11/17/24 14:20:00	VIP	Teen	Male
8	Visitor-1	Watch montages in the 8D Theater	Theater Staff	11/17/24 17:05:00	11/17/24 17:10:00	VIP	Teen	Male
9	Visitor-1	Leave Theme Park	Front-desk Staffs	11/17/24 17:49:00	11/17/24 17:51:00	VIP	Teen	Male
10	Visitor-2	Arrive to Theme Park	Front-desk Staffs	11/08/24 10:02:00	11/08/24 10:05:00	Normal	Teen	Female
11	Visitor-2	Get General Admission Access	Ticket Counter	11/08/24 10:24:00	11/08/24 10:28:00	Normal	Teen	Female
12	Visitor-2	Claim two free rides by checking in on social media	Promoter	11/08/24 11:23:00	11/08/24 11:27:00	Free 2 ride	Teen	Female
13	Visitor-2	Ride the Carousel	Carousel Operator	11/08/24 13:05:00	11/08/24 13:08:00	Normal	Teen	Female
14	Visitor-2	Drive the Bumper Cars	BC Operator	11/08/24 13:51:00	11/08/24 13:53:00	Normal	Teen	Female
15	Visitor-2	Slide down the Water Slides	WS Guaard	11/08/24 14:36:00	11/08/24 14:39:00	Normal	Teen	Female
16	Visitor-2	Swim in the Wave Pools	Life Guard	11/08/24 14:54:00	11/08/24 14:58:00	Normal	Teen	Female
17	Visitor-2	Leave Theme Park	Front-desk Staffs	11/08/24 16:34:00	11/08/24 16:36:00	Normal	Teen	Female
18	Visitor-3	Arrive to Theme Park	Front-desk Staffs	11/03/24 9:51:00	11/03/24 9:55:00	Normal	Teen	Female
19	Visitor-3	Get General Admission Access	Ticket Counter	11/03/24 11:16:00	11/03/24 11:19:00	Normal	Teen	Female
20	Visitor-3	Upgrade to Premium Access	Ticket Counter	11/03/24 11:29:00	11/03/24 11:34:00	VIP	Teen	Female
21	Visitor-3	Experience the Drop Towers	DT Operator	11/03/24 13:44:00	11/03/24 13:46:00	Normal	Teen	Female
22	Visitor-3	Participate in the Escape Room	ER Staff	11/03/24 13:59:00	11/03/24 14:02:00	Normal	Teen	Female
23	Visitor-3	Leave Theme Park	Front-desk Staffs	11/03/24 15:17:00	11/03/24 15:19:00	VIP	Teen	Female
24	Visitor-4	Arrive to Theme Park	Front-desk Staffs	11/04/24 9:00:00	11/04/24 9:57:00	Normal	Adult	Male
25	Visitor-4	Get General Admission Access	Ticket Counter	11/04/24 10:43:00	11/04/24 10:47:00	Normal	Adult	Male
26	Visitor-4	Claim two free rides by checking in on social media	Promoter	11/04/24 11:40:00	11/04/24 11:43:00	Free 2 ride	Adult	Male
27	Visitor-4	Watch montages in the 8D Theater	Theater Staff	11/04/24 14:14:00	11/04/24 14:17:00	Normal	Adult	Male

Fig. 2. A print-screenshot of the dataset used in this study.

3.2. Data Collection and Structure

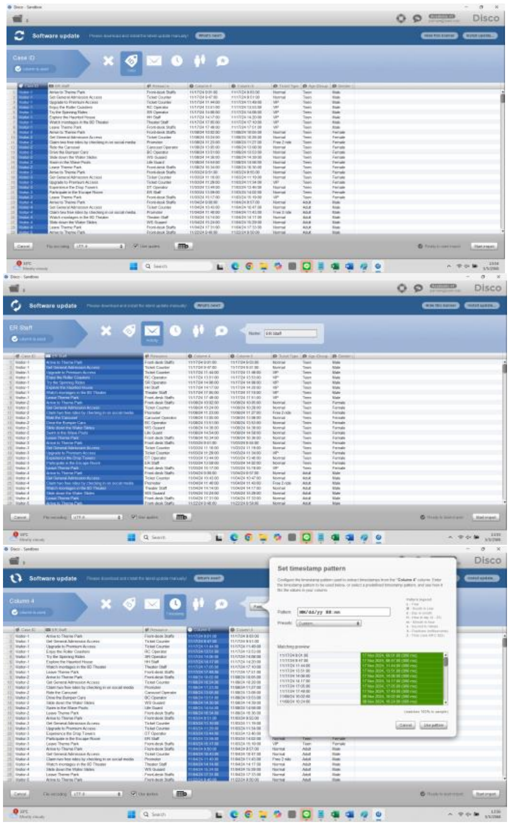


Fig. 3. (up) Assigning "Case" role through Disco. (middle) Assigning "Activity" role via Disco. (down) Assigning "TimeStamp" role using Disco.

The dataset utilized in this research comprises 684 event records representing the chronological sequence of activities performed by visitors within the theme park. Each record corresponds to a single event in a visitor's journey and includes temporal, contextual, and demographic attributes. The dataset was provided in comma-separated values (CSV) format and consists of eight columns, each serving a specific role in enabling process analysis.

A brief description of the dataset's features is provided below:

- Case ID: This attribute serves as the unique identifier for each visitor (e.g., *Visitor-1*, *Visitor-2*). It defines the scope of a single process instance, commonly referred to as a "trace" in process mining. All events associated with the same Case ID are treated as a coherent sequence of actions performed by an individual visitor.
- ER Staff: This column denotes the activity label representing the task performed (e.g., *Arrive to Theme Park, Get General Admission Access, Upgrade to Premium Access, Enjoy the Roller Coasters, Try the Spinning Rides, Explore the Haunted House, Watch montages in the 8D Theater, Leave Theme Park, Claim two*

free rides by checking in on social media, Ride the Carousel, Drive the Bumper Cars, Slide down the Water Slides, Swim in the Wave Pools, Experience the Drop Towers, Participate in the Escape Room). These activity names are essential in identifying the steps constituting the overall process (see Fig.2. and Fig.4.).

- Resource: This attribute captures the role or individual responsible for executing the activity (e.g., *Front-desk Staffs*, *RC Operator*). Mapping this to the "Resource" field in Disco enables the analysis of workload distribution and performance at the staff level.
- Start Timestamp (Column 4) and End Timestamp (Column 5): These two columns record the start and end times of each event, respectively. These timestamps are pivotal for calculating event durations, identifying waiting times, and generating time-based visualizations within the process model. In the Disco tool, Column 4 is designated as the primary timestamp. If supported, Column 5 can be used to enrich the analysis by capturing activity duration.
- Ticket Type: This feature categorizes visitors by the nature of their admission (e.g., *Normal*, *VIP*). It is treated as a case-level attribute in process mining and allows comparative analysis between different visitor types.
- Age /Group: This field provides demographic segmentation (e.g., *Teen*, *Adult*), enabling behavioral comparisons across age groups.
- Gender: This attribute offers additional segmentation by visitor gender (e.g., *Male*, *Female*), which is also used as a case-level attribute to identify any notable differences in process behavior.

Fig. 4. The figure shows 15 different types of activities provided in the dataset.

3.3. Data Preprocessing

Prior to importing the dataset into Disco, minor preprocessing was performed to ensure compatibility and semantic clarity:

- Column headers were renamed to clearly reflect their functional roles in process mining: e.g., "Activity", "Start Time", "End Time".
- Timestamps were verified for formatting consistency to align with Disco's requirements.
- Any redundant whitespaces or irregular text entries were cleaned to maintain data quality and avoid misclassification during analysis.

3.4. Mapping to Disco

Upon preprocessing, the dataset was imported into Disco, a leading tool for exploratory and descriptive process mining. The following configuration (see Fig.3.) was used during the import phase:

- Case ID ⟶ *Mapped to*: Case ID
- Activity (ER Staff) ⟶ *Mapped to*: Activity Name
- Start Time (Column 4) ⟶ *Mapped to*: Timestamp
- Resource ⟶ *Mapped to*: Resource
- Ticket Type, Age /Group, and Gender ⟶ *Mapped as*: Case-level attributes

This mapping allowed for the full reconstruction of each visitor's journey, identification of process variants, and visualization of temporal and resource-based metrics.

3.5. Rationale for Methodological Choice

The rationale behind using process mining and the Disco tool lies in the need to understand the real behavior of process participants beyond static or aggregate data summaries. Unlike traditional statistical techniques, process mining offers a dynamic and time-aware perspective of process execution, capturing actual sequences, loops, delays, and deviations. Disco, in particular, is known for its intuitive interface, automatic process discovery, and rich filtering capabilities, making it highly suitable for both novice and advanced researchers.

4. Findings and Results

This chapter presents the key findings derived from the analysis of the event log file related to visitor activities in a theme park setting. The analysis has been conducted using process mining techniques to uncover the behavior of visitors, their process variants, and event dynamics throughout the duration of the recorded period.

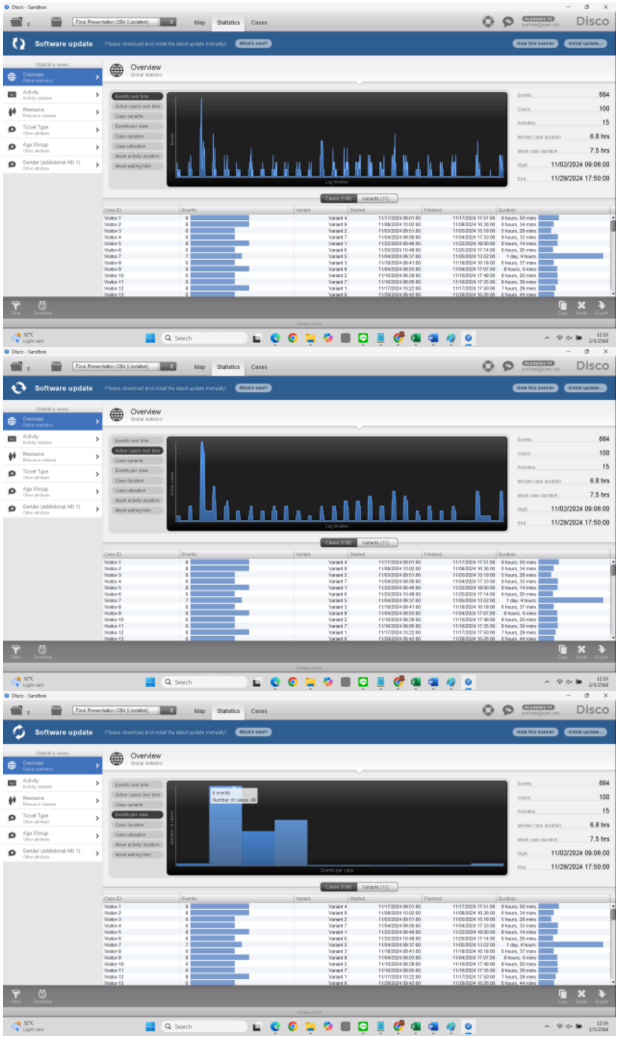


Fig. 5. Global Statistics Overview. (up) Events over Time. (middle) Active Cases over Time. (down) Events per Case.

The event log comprises a total of 100 process instances or cases. Each case represents the activity journey of a unique visitor in the theme park. In total, the event log contains 684 recorded events, which are distributed across 100 cases. This suggests that, on average, each visitor was involved in approximately 6.84 activities during their stay.

A detailed examination of the event log revealed the existence of 15 distinct activity types. These are as follows: Arrive to Theme Park, Leave Theme Park, Get General Admission Access, Upgrade to Premium Access, Slide down the Water Slides, Enjoy the Roller Coasters, Participate in the Escape Room, Claim Two Free Rides by Checking in on Social Media, Watch Montages in the 8D Theater, Drive the Bumper Cars, Experience the Drop Towers, Explore the Haunted House, Try the Spinning Rides, Ride the Carousel, Swim in the Wave Pools.

The first recorded activity in the event log (see Fig.5.) occurred on 2nd November 2024 at 9:00 AM, and the last recorded activity occurred on 29th November 2024 at 5:50 PM. This indicates that the data covers nearly an entire month of operational and visitor activity.

Analysis When activities were analyzed and sorted by their frequency and relative frequency (see Fig.5.), it was observed that:

- "Arrive to Theme Park" and "Leave Theme Park" were the most frequently performed activities. These two are likely included in almost every case, as they mark the beginning and end of each visitor's journey.

- On the other hand, "Explore the Haunted House" was the least performed activity, indicating either lower interest or limited availability of the attraction.

4.1. Visualization at Different Thresholds

when adjusting the visualization thresholds:

- At a 0% threshold for both activities and paths, the process map displayed a simplified version with fewer activities and connections. This helped isolate the main flow of the process and minimize distractions from low-frequency variants.
- At 100% threshold, all activities and paths were shown, including infrequent and optional steps. This version highlighted the overall complexity of the visitor journeys and revealed bottlenecks and alternative paths (see Fig.5.).
- At the 100% threshold setting, it was observed that while all 100 cases were present in the dataset, only Variant 1 aligned completely with the process map, representing the most typical behavior. Other variants diverged due to inclusion of optional or rare activities (see Fig.5.).
- The most frequent execution path, involving Visitor 18, comprised 14 events, making it the most detailed and complete visitor experience in the dataset.

Fig. 5. Fuzzy Miner Graph/Model. (up) Simplified model with thresholds of 0% Activity and 0% Paths. (middle) Generated model with thresholds of 100% Activity and 100% Paths. (down) Customized model with thresholds of 100% Activity and 0% Paths.

4.2. Event Distribution Across Visitors

- The maximum number of events recorded for a single visitor was 14, as seen with Visitor 18.
- The minimum number of events was 5, which occurred for Visitor 88. This

suggests a short or minimally interactive visit.

The analysis of active cases over time revealed:

- A maximum of 16 active cases within a 10-minute interval, indicating peak visitor activity.
- A minimum of 1 active case within a 10-minute interval, possibly during off-peak hours.

4.3. Events Per Case

- The maximum number of events per case was 6, notably in Case 49.
- The minimum number was 5 events, as found in Case 1. This consistency in the lower bound suggests a minimum threshold of activities that every visitor must complete.

4.4. Observations from Process Map

While observing the process model in the map view, a noteworthy loop was identified on 20th November 2024. This repetition could signify a visitor re-engaging with an activity (such as multiple rides) or a re-entry into a particular segment of the park. This kind of behavior highlights a non-linear flow, which is common in entertainment and theme park environments.

4.5. Cluster and Pattern Analysis

To investigate the behavioral patterns of the cases, we analyzed the dataset using the Cases View and examined the resulting Clusters (also referred to as Pattern Groups). These clusters represent commonly occurring sequences of activities across different cases.

- First Cluster/Pattern (see Fig.6.):

  The first cluster comprises 13 cases. The activity sequence for this cluster is consistent across these cases and follows a standardized process flow. The typical sequence (based on visual inspection of the process map and case sequences) is:
    - *Enter Theme Park*
    - *Get Wristband*
    - *Rent Locker*
    - *Change Clothes*
    - *Slide Down the Water Slide*
    - *Leave Theme Park*

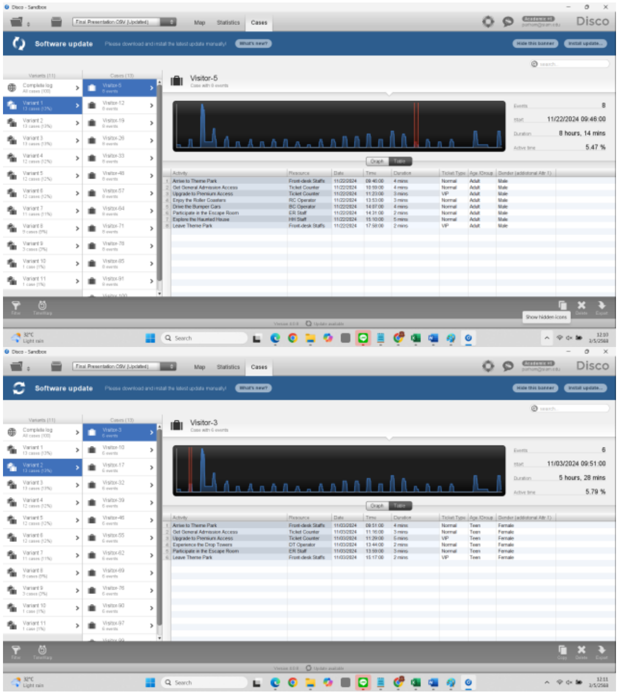


Fig. 6. Clustering Results. (up) Variant/Cluster 1 including 13 cases. (down) Variant/Cluster 2 including 13 cases.

- Second Cluster/Pattern (see Fig.6.): The second cluster also includes 13 cases. While structurally similar to the first cluster, the order or presence of specific activities may slightly differ, indicating variability in behavior. The observed sequence (based on the visualization) is:
  - *Enter Theme Park*
  - *Get Wristband*
  - *Change Clothes*
  - *Rent Locker*
  - *Slide Down the Water Slide*
  - *Leave Theme Park*

These patterns reflect different customer behaviors and slight process variants, possibly due to contextual or operational differences.

4.6. Identification of Abnormal Patterns

Among the detected variants, two abnormal clusters were identified — Variant 10 and Variant 11 (see Fig.7.).

- Variant 10 includes only 1 case.
- Variant 11 also includes only 1 case.

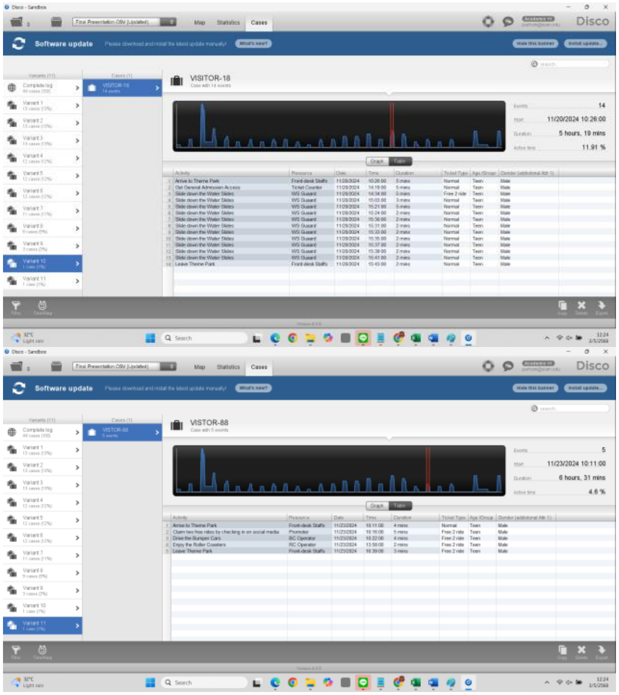


Fig. 7. Clustering Results. (up) Variant/Cluster 10 including 1 case. (down) Variant/Cluster 11 including 1 case.

These variants are deemed *abnormal* due to their low frequency and deviation from common patterns. Their uniqueness could be due to missing activities, unusual activity orders, or resource-related anomalies. Investigating these abnormal cases further may reveal exceptional circumstances, errors, or outlier behaviors within the process.

4.7. Case Duration and Bottleneck Analysis

An important performance indicator assessed in this study was case duration, which measures the total time taken for a process instance (case) to complete from start to end. The data was filtered based on the total duration to uncover time-related insights (see Fig.8.).

- Cases running longer than 5 hours: A total of 93 cases exceeded a runtime of 5 hours, indicating prolonged engagement in the process.
- Cases running between 5 to 9 hours: Among them, 89 cases were identified to have a runtime greater than 5 hours but less than 9 hours, representing most of the extended-duration cases.
- Cases run between 9 to 10 hours: Only 1 case was observed to have a runtime within this range, suggesting that such long durations are rare and may point to specific anomalies or delays.

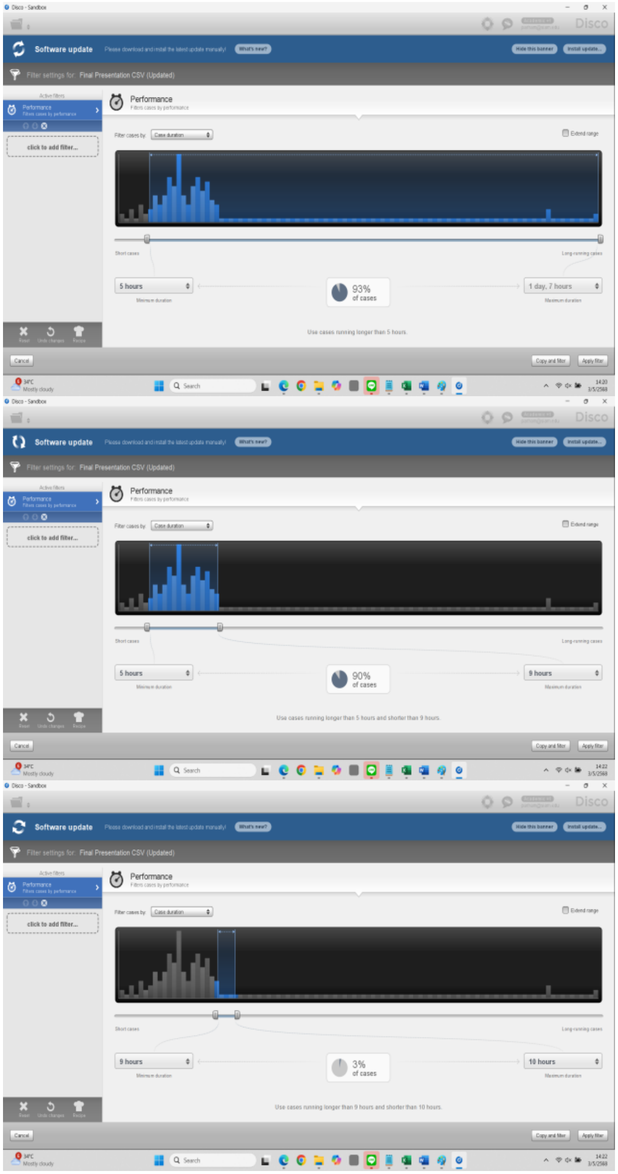


Fig. 8. (up) Cases run longer than 5 hours (middle) Cases run between 5 to 9 hours. (down) Cases running between 9 to 10 hours.

This time-based segmentation highlights where the process is efficient and where it may experience bottlenecks or delays (see Fig.8.).

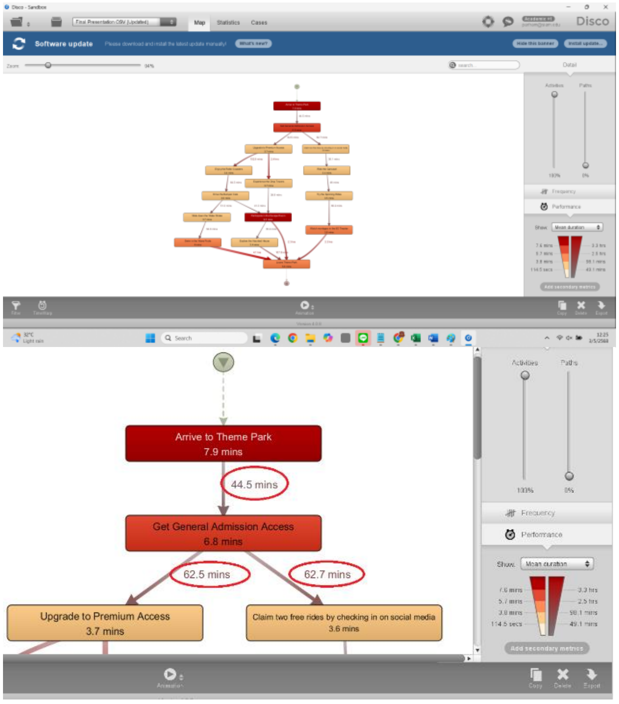


Fig. 9. Time Interval/Gap Analysis Results. (up) Average/Mean duration of time accumulated between Activities. (down) Identification of bottlenecks and slow handling of the processes.

The graphs presented in Fig.9. illustrate a process performance analysis of a typical visitor journey through a theme park, focusing on the mean duration of each step. The data exposes several inefficiencies that significantly affect visitor experience and operational performance. One of the most prominent bottlenecks is the transition from "Arrive to Theme Park" to "Get General Admission Access", which takes an average of 44.5 minutes. This long wait time at the entrance suggests issues such as overcrowding, lack of digital or pre-booked ticketing options, and inefficient manual ticket checks. In modern amusement environments, such delays can drastically reduce visitor satisfaction and time available for actual enjoyment.

The following step, "Get General Admission Access" to "Upgrade to Premium Access", reveals another area of concern, with an average time of 62.5 minutes. This process is likely slowed by unclear upgrade procedures, insufficient staffing, or reliance on manual confirmation steps. The fact that it takes longer than the initial admission is alarming and points toward systemic inefficiencies.

Similarly, claiming promotional rewards—such as "Claim Two Free Rides by Checking in on Social Media"—also consumes around 62.7 minutes on average. Such a delay indicates possible technical hurdles, such as app or system incompatibility, poor signage, or lack of real-time verification tools. This further detracts from the intended benefit of the promotion, which is to create quick engagement and satisfaction.

The color gradient used in the diagram supports this analysis by indicating high mean durations (up to 3.3 hours), showing that the entire admission and benefit-claiming process is significantly delayed. The broader impact of such delays includes reduced throughput of visitors, higher congestion, and a lower capacity for rides and attractions utilization.

To resolve these issues, several improvements are recommended. The implementation of digital ticketing and mobile applications can allow for smoother entry and quicker upgrades. Introducing self-service kiosks for both admissions and upgrades will reduce dependency on staff. Additionally, using automated systems for social media verification can drastically cut down the time to claim promotions. Optimizing the physical layout and providing better signage and queuing systems will also improve the flow.

For future optimization, the integration of AI-based predictive analytics and real-time dashboards for operations teams will enable dynamic management of resources. Process mining tools can continue to monitor and evaluate improvements, ensuring a continuous cycle of performance enhancement.

4.8. Direct Follower Relationship Analysis

To understand specific behavioral patterns, a direct follower analysis was conducted. One of the interesting findings pertains to the sequence:

- “Slide Down The Water Slide” ⟶ “Leave Theme Park”

It was found that 13 cases exhibited this exact sequence within a very long average/mean time of 4.1 hours (see Fig.10.), indicating that for some visitors, sliding down the water slide marked their final activity before exiting the theme park. This could represent a typical "last activity" pattern and may reflect user preferences or fatigue factors.

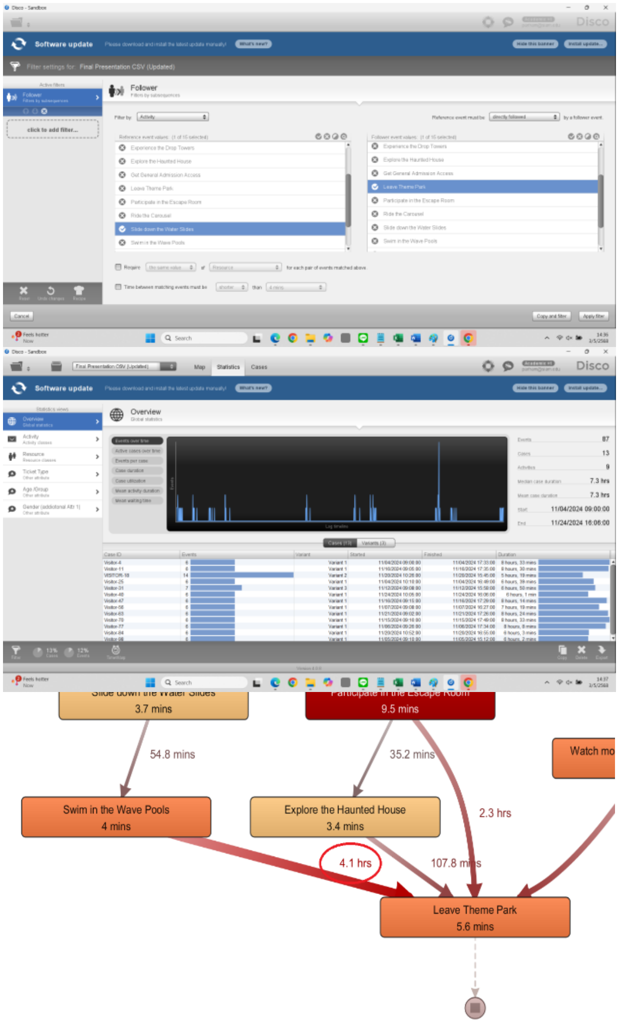


Fig. 10. (up) Application of Follower Filtering feature. (middle) Investigation of direct relationship from “Slide Down The Water Slide” ⟶ “Leave Theme Park”. (down) Long time interval/gap identification between the above two activities.

4.9. Resource and Activity Mapping

In the final analysis, the dataset was re-imported and redefined to incorporate resource information — answering the fundamental question of “Who did what?”

By mapping Resources into Activities, we were able to identify which staff members or system entities were involved in executing specific activities. This mapping helps in:

- Assessing workload distribution across resources.
- Identifying potential overutilization or underutilization.
- Understanding handoffs and collaborative efforts between different roles.

Such insights are crucial for improving human resource allocation, training, and process optimization strategies.

5. Discussions and Conclusions

This chapter synthesizes the findings presented in Chapter 4 to discuss the implications, identify underlying problems, and draw meaningful conclusions based on the process mining analysis of visitor activities within a theme park environment. The discussion also outlines the contributions and limitations of the study and proposes directions for future research.

5.1 Summary of Findings.

The analysis of 100 process instances (cases) and 684 associated events revealed significant patterns and behavioral insights. Fifteen distinct activity types were identified, with "Arrive to Theme Park" and "Leave Theme Park" being the most frequent activities, and "Explore the Haunted House" being the least frequent. The majority of visitors followed a core sequence of experiences, but notable variations and outliers were observed.

Process model visualization at different thresholds helped in understanding the complexity of visitor journeys. The 0% threshold highlighted a basic, linear flow, while the 100% threshold revealed all possible paths and bottlenecks. Case duration analysis indicated that 93% of visitors spent more than 5 hours in the park, with a significant concentration of durations between 5 and 9 hours.

Clusters and patterns revealed that the majority of visitors followed consistent behavioral flows, with the first two clusters each comprising 13 cases. Two abnormal patterns (Variant 10 and 11) suggest outlier behavior, with only one case each, prompting further examination for potential anomalies.

Direct follower analysis showed a common visitor behavior where 13 cases had the "Slide Down The Water Slide" activity immediately followed by "Leave Theme Park," marking a pattern of ending the visit on a high adrenaline note. Resource mapping also provided valuable insights into workload distribution and potential optimization areas for staff deployment.

5.2 Problem Identification.

The analysis uncovered several operational inefficiencies and user behavior patterns that warrant attention:

- Some attractions like the Haunted House had very low participation, suggesting possible issues in appeal, location, or accessibility.
- Certain days (e.g., 20th November 2024) showed loops in the process, hinting at repeated activities that could indicate confusion, repeated entry, or inefficient guidance.
- While prolonged engagement might indicate enjoyment, it could also signify waiting times, lack of signage, or inefficiencies in attraction transitions.
- The presence of unique outlier cases suggests inconsistencies in process enforcement or potential data quality issues.
- The process analysis reveals major delays early in the theme park experience. Visitors spend 44.5 minutes on average from arrival to obtaining general admission. Further delays occur when upgrading to premium access (62.5 minutes) and claiming free rides via social media check-in (62.7 minutes). These prolonged wait times reduce time spent on attractions and may negatively affect visitor satisfaction.

5.3 Proposed Solutions and Implications.

The findings lead to several actionable solutions:

- Conduct surveys or A/B testing to understand visitor preferences.
- Optimize resource allocation on high-traffic days or around specific bottleneck points identified through process maps.
- Provide digital maps or mobile apps to minimize loops and streamline visitor flow.
- Use common patterns (e.g., Cluster 1 and 2) to design personalized packages for different visitor segments.
- Continue using process mining as a real-time monitoring tool for operational optimization and anomaly detection.
- The root causes likely include manual processes, insufficient staffing, and lack of digital systems. To address this, the park should implement digital ticketing, self-service kiosks, and a mobile app to streamline entry, upgrades, and reward

claims. Additionally, smart queue systems and better staff allocation during peak times can balance visitor flow and reduce congestion.

5.4 Limitations and Scope.

While this study provides rich insights, some limitations include:

- Data was limited to one month and one park, which may not generalize to all settings.
- Resource data was minimal and could be enriched for deeper HR analysis.
- Visitor demographics and motivations were not captured, limiting behavioral interpretation.

5.5 Future Work.

Future research can expand on current work by:

- Applying similar analysis across multiple theme parks or seasons.
- Integrating customer satisfaction surveys for correlation with process behavior.
- Leveraging AI techniques to predict visitor paths and optimize layouts dynamically.
- Exploring cross-industry applications, such as healthcare, education, or public administration.

References

[1] Kurniati, A.P., Hall, G., Hogg, D., & Johnson, O. (2021). Process Mining on the Extended Event Log to Analyse the System Usage During Healthcare Processes (Case Study: The GP Tab Usage During Chemotherapy Treatments). In S. Leemans & H. Leopold (Eds.), *Process Mining Workshops* (Vol. 406, pp. 373–385). Springer, Cham.

[2] Wagner, M., Helal, H., Roepke, R., Judel, S., Doveren, J., Goerzen, S., Soudmand, P., Lakemeyer, G., Schroeder, U., & van der Aalst, W. (2022). A Combined Approach of Process Mining and Rule-based AI for Study Planning and Monitoring in Higher Education. preprint :2211.12190.

[3] Mohammadi, F., Kazempourian, S., & Raeesi Vanani, I. (2022). Process Mining Approach to Performance Analysis and Bottleneck Finding in Electronic Processes (Case Study: The Billing Process of Hospital Services). International Journal of Process Management and Benchmarking, 12(3), 212–232.

[4] Thiyagarajan, G., & Prasanna, S. (2023). Process Mining-Based Behavioral Modeling of Learners in Self-paced Learning Environment. In K.P. Ray, A. Dixit, D. Adhikari, & R. Mathew (Eds.), *Proceedings of the 2nd International Conference on Signal*

and Data Processing (pp. 121–132). Springer, Singapore.

[5] Saad, H. (2023). Application of Process Mining and Sequence Clustering in Recognizing an Industrial Issue. preprint :2311.15362.

[6] Liu, Y., Dani, V.S., Beerepoot, I., & Lu, X. (2023). Turning Logs into Lumber: Preprocessing Tasks in Process Mining. preprint :2309.17100.

[7] Anonymous. (2023). Exploring Customer Journey Mining and RPA: Prediction of Customers' Next Touchpoint. In Business Process Management: Blockchain, Robotic Process Automation and Educators Forum *(BPM 2023)* (pp. 181–196). Springer.

[8] Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2024). Process Mining for Self-Regulated Learning Assessment in E-learning. preprint :2403.12068.

[9] Nai, R., Sulis, E., & Genga, L. (2024). Enhancing E-learning Effectiveness: A Process Mining Approach for Short-term Tutorials. Journal of Intelligent Information Systems, 62, 1773–1794.

[10] Halvorsrud, R., Mannhardt, F., Prillard, O., & Boletsis, C. (2024). Customer Journeys and Process Mining – Challenges and Opportunities. ITM Web of Conferences, 62, 05002.

[11] Porouhan, P. (2023), "Process Discovery and Process Optimization of Banking Industry through Visual Mapping and Simulation of Activity Sequences," Progress in Applied Science and Technology, Vol. 13, No. 2, pp. 59–66.

[12] Porouhan, P. (2022), "Applying Process Mining to Minimise Order Waiting Time of FitBox," International Journal of Knowledge Engineering and Data Mining, Vol. 7, No. 3, pp. 190–233.

[13] Porouhan, P. and Premchaiswadi, W. (2022), "Process Modeling and Bottleneck Mining in MXML-based Course Training Event Logs," Progress in Applied Science and Technology, Vol. 13, No. 1, pp. 1–10.

[14] Premchaiswadi, W., Arpasat, P., Palangsantikul, P., Intarasema, S., Kungcharoen, K., and Narksarp, W. (2024), "Using IoT and Mobile Robots to Model and Analyze Work Processes with Process Mining Techniques," Progress in Applied Science and Technology, Vol. 14, No. 1, pp. 1–10.

[15] Nammakhunt, A., Porouhan, P., and Premchaiswadi, W. (2023), "Creating and Collecting e-Learning Event Logs to Analyze Learning Behavior of Students through Process Mining," International Journal of

Information and Education Technology, Vol. 13, No. 2, pp. 211–222.

[16] Palangsantikul, P., Porouhan, P., Premchaiswadi, N., and Premchaiswadi, W. (2019), "Emotion Analytics with Process Mining," Malaysian Journal of Computer Science, Special Issue 2, pp. 109–131.

[17] Arpasat, P. and Premchaiswadi, W. (2024), "Data-Driven Business Process Improvement," Progress in Applied Science and Technology, Vol. 14, No. 3, pp. 11–20.

[18] Porouhan, P. and Premchaiswadi, W. (2017), "Process Mining and Learners' Behavior Analytics in a Collaborative and Web-Based Multi-Tabletop Environment," International Journal of Online Pedagogy and Course Design, Vol. 7, No. 3, pp. 29–53.

[19] Nammakhunt, A. and Premchaiswadi, W. (2024), "Analysis of Student Learning Behavior using Process Mining and Spectrogram," Progress in Applied Science and Technology, Vol. 14, No. 2, pp. 1–11.

[20] Premchaiswadi, W. and Porouhan, P. (2015), "Process Modeling and Bottleneck Mining in Online Peer-Review Systems," SpringerPlus, Vol. 4, Article 441.