DTU Entrepreneurship
DTU Entrepreneurship

Improving crowd safety measurability at Roskilde Festival utilizing AI-enabled video surveillance analysis

Thesis subtitle

Master Thesis

Improving crowd safety measurability at Roskilde Festival utilizing AI-enabled video surveillance analysis
Thesis subtitle

Master Thesis
Date, year

By
Torben Albert-Lindqvist

Cover photo: Vibeke Hempler, 2012
Published by: DTU, DTU Entrepreneurship, Diplomvej, Buildings 371 & 372, 2nd floor, 2800 Kgs. Lyngby Denmark
www.entrepreneurship.dtu.dk/
ISSN: [0000-0000] (electronic version)
ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)
ISBN: [000-00-0000-000-0] (printed version)

Approval

This thesis has been prepared over six months at the Section for Indoor Climate, Department of Civil Engineering, at the Technical University of Denmark, DTU, in partial fulfilment for the degree Master of Science in Engineering, MSc Eng.

It is assumed that the reader has a basic knowledge in the areas of statistics.

Torben Albert-Lindqvist - s233587

....

Signature

....

Date

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgements

Torben Albert-Lindqvist, MSc Civil Engineering, DTU
Creator of this thesis template.

[Name], [Title], [affiliation]
[text]

[Name], [Title], [affiliation]
[text]

Contents

1 Introduction

1.1 Background and motivation

On June 30th 2000, nine young men lost their lies in a crowd crush during a Pearl Jam concert at Roskilde Festival in Denmark [1]. An uncontrolled surge, pushing the crowd towards the scene, caused immense pressure on the front most concert-goers, thrusting them against the barriers. The high-energy mass of people unknowingly trampled the victims, who succumbed under the pressure of the crowd. This incident is unfortunately not the only one of its kind, as crowd crushes continue to occur at mass gatherings around the world.

1.2 Problem definition

No matter the size of the event, crowd safety is a complex and multifaceted problem. No matter the extent of planning and preparation, the unexpected can still happen. Crowd safety professionals have an immense responsibility, as they are tasked with ensuring the safety of thousands of people. Large crowds are often unpredictable, and improper planning and/or response can lead to disastrous consequences. Fortunately, however, crowd safety professionals have a plethora of tools and knowledge at their disposal to help mitigate these risks and keep crowd dynamics within manageable bounds (see section 2.1).

Despite the available tools and established frameworks, current crowd safety management practices face significant limitations. A recurring theme identified through discussions with industry professionals is a heavy reliance on the experience and intuition of the safety team. Key decisions regarding venue layout, capacity planning, resource allocation (e.g., staffing levels and placement), and risk assessment of concerts often depend heavily on estimations derived from past events and anecdotal knowledge rather than objective, quantitative data specific to the event's context. While experience is invaluable, this reliance introduces subjectivity and potential inconsistencies. Calculations conducted for aspects such as stage layout or entrance/exit dimensions are often based on estimations made by the safety team, albeit these usually are adjusted to cover worst-case scenarios. Even these adjustments, however, are still limited by the imagination and, crucially, the past experiences of the safety professionals. This ultimately suggests that younger or less experienced teams may lack the extensive reference points available to seasoned experts, potentially leading to false assumptions and miscalculations. Consequently, the efficacy of safety planning can appear correlated with the cumulative experience within the team.

This complicated problem, has a seemingly simple and obvious solution: more objective data. At present, experiences and observations often go undocumented, as the safety team already has a plethora of responsibilities and tasks to attend to, not to mention the impossibility of having a full overview of the event at one given moment. Moreover, most music festivals are massively scalable operations, going from a relatively small team of full-time employees and contractors throughout the planning stage, to a massive team of volunteers during the event. In the example of Roskilde Festival, this culminates in a team scaled from over 100 employees to over 30 thousand. This drastic staff scaling certainly applies to the crowd safety department as well, potentially contributing to a lack of continuity and knowledge retention. This implies that immediately following the event, or even after a given day, important observations and learnings from staff may be lost if not systematically documented. These factors effectively hinder a comprehensive, data-driven understanding of crowd behavior.

Furthermore, communicating the rationale behind safety decisions and requirements to other internal departments or external stakeholders can be challenging without clear, objective evidence. Provided that safety precautions are based on subjective assessments, conveying a need for specific resources or precautions can prove difficult. For instance, Roskilde Festival's safety team present a recurring challenge of convincing colleagues in the food and beverage department of concerns regarding the placement of food stalls or bars. Without the aid of clear evidence, Roskilde Festival's safety team occasionally find themselves dedicating valuable time and resources to justify their positions, at times even having to deviate from their core competencies to develop visual material to support their arguments.

These scenarios and limitations highlight opportunities for significant improvement, namely through the development and integration of technology capable of providing objective, measurable insights into crowd dynamics at music festivals. This thesis seeks to explore these opportunities, guided by the following hypotheses:

- Challenges in communicating crowd safety requirements and justifying decisions internally are often due to the subjective nature of current assessments, lacking objective, easily understandable evidence.
- The precision and efficacy of safety planning are constrained by a reliance on experience-based estimations rather than quantitative, historical data on actual crowd dynamics.
- The lack of a persistent, easily accessible digital record of crowd dynamics during an event limits post-event analysis, knowledge retention, and continuous improvement within safety teams.

1.3 Brief history of Fluxense

Together with two classmates, I founded a startup, Fluxense, in January 2024. Our initial plan was to solve the crowd safety challenges of music festivals, as presented in section 1.2, by developing an AI-enabled system for monitoring existing CCTV infrastructure to provide automated analyses of crowd behavior. We gained traction quickly, with several large festivals expressing their interest in our proposed product. Development began almost immediately, and we held our first prototype test at DTU's Commemoration Day, where we provided a live count of the number of people in the concert hall. The test gave great results, as well as valuable learnings, and became the first of many. The following summer was very busy, as we attended three of Denmark's largest music festivals – Copenhell, Roskilde Festival and Smukfest – to further test and develop our product.

After the summer, we stood at a crossroads. Our collaborations with the different festivals had revealed that each had their own unique requirements, and the value of our product was not as clear-cut as we had initially thought. We feared that crowd safety was not a large enough market for scaling our business, nor that a generalized product would be attractive in the industry. We decided to pivot, and began exploring other markets where our technology could be of use. We gradually moved away from our initial focus on crowd safety, and found business intelligence to be a much larger and lucrative market. Instead of monitoring crowds, our new product would track individual customers in retail stores, transportation hubs, amusement parks, and museums. We aimed to provide insights into places/products of interest, dwell times, conversion rate, footfall analysis, etc., to help businesses optimize their operations.

As the autumn progressed, we began securing new collaborations in our target industry, and our value proposition became clearer. One important thing had been lost in the process, however: our motivation. We had started Fluxense with the goal of improving crowd safety at music festivals, as it was a mission we shared a passion for. Our new focus on business intelligence made sense fiscally, but didn't evoke the same feeling of purpose. Fluxense ended up dissolving in the winter of 2024/25, as we couldn't see ourselves in the startup's new reality, and struggled to find a common vision.

1.4 Scope and purpose statement

This thesis continues approximately where Fluxense left off before the pivot. However, rather than following the path laid out by the startup and striving to develop a scalable commercial product, the purpose of this work is to create a tool that directly aids crowd safety managers at Roskilde Festival. This is partially due to our already close collaboration throughout the entirety of 2024, as well as their expressed interest in continuing our collaboration through this leg of the project. Additionally, it is the largest music festival in Northern Europe, attracting over 130 thousand guests each year [2]. With considerable prestige in the industry, as well as a passionate dedication to improving crowd safety, Roskilde Festival is an ideal partner for this project.

It is important to note that the majority of crowd safety practices presented in this thesis are gathered through discussions with Roskilde Festival's safety team. While occasional references may be made to interactions or insights gained through collaborations with other festivals, these are included solely to illustrate the broader landscape and are not indicative of the project's applicability beyond Roskilde. These findings are also assumed limited to a Danish context, as all discussions with crowd safety professionals were with Danish festivals. Additionally, it was observed during these consultations that some prominent experts occasionally presented viewpoints that could be interpreted as subjective or opinionated. However, as the explicit goal of this project is to provide a functional tool for Roskilde Festival's specific operational environment, a critical evaluation of the objective validity of these statements falls outside the defined scope and is not deemed essential for achieving the project's objectives.

In summary, the purpose statement of this project is as follows:

> *To enhance Roskilde Festival's crowd safety management by developing an intuitive, data-driven platform that provides actionable insights into crowd dynamics (such as density, flow, and movement patterns), thereby improving planning, internal communication, and documentation for the safety team.*

1.5 Objectives

Formalizing the hypotheses presented in section 1.2, the business objectives of this project are as follows:

1. **Improve internal communication**: offer clear, visual, objective evidence to help the safety team communicate requirements and justify decisions to other departments.

2. **Enhance safety planning**: provide quantitative, historical data on crowd dynamics to enable more accurate planning of layouts, capacities, resource allocation, and facility placement.

3. **Create reliable documentation**: Generate a persistent digital record of crowd dynamics for post-event analysis, debriefing, and knowledge retention.

1.6 Thesis structure

In their book, *Design Science*, Hubka and Eder characterize the design process as intuitive, iterative, innovative, unpredictable and reflective [3]. While these aspects are inherent to design, tackling complex engineering challenges requires more than intuition and creativity alone. To manage the process effectively, ensure thoroughness, and facilitate clear understanding and traceability, a structured approach is beneficial [4]. Therefore, the process needs to be organized, drawing upon established product design methodologies and frameworks to guide this project. Many such frameworks exist, offering different levels of detail and focus. This section will explore the most relevant frameworks, and propose a design and development methodology for this project.

1.6.1 Comparison of frameworks

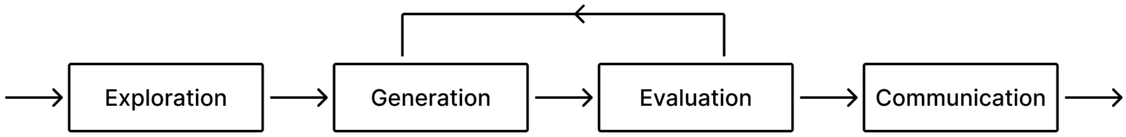


Figure 1.1: Cross' four-stage model of the design process

Cross [5] proposes likely the most simplistic, yet well-known framework: a four-stage model comprised of *exploration*, *generation*, *evaluation* and *communication* (Figure 1.1). Cross describes this type of modal as descriptive, as it merely attempts to model the conventional, heuristic design process. More detailed models of this type exist, such as French's [6] "anatomy of design," (Figure 1.2) detailing four stages, most distinctly underlining the problem analysis and definition, as conducted in section 1.2. According to Cross, these models differ from prescriptive models, which offer a more systematic procedure, as well an emphasis on analyzing and understanding the design problem before generating solution concepts. Perhaps the most well-known of these is offered by Pahl et. al [7], and is based on the following design stages: *clarification of the task*, *conceptual design*, *embodiment design*, and *detail design*. Combining the aforementioned models, Ulrich and Eppinger present a rather comprehensive framework. Their process is based on the following stages: *concept development*, *system-level design*, *detail design*, and *testing and refinement* [8].

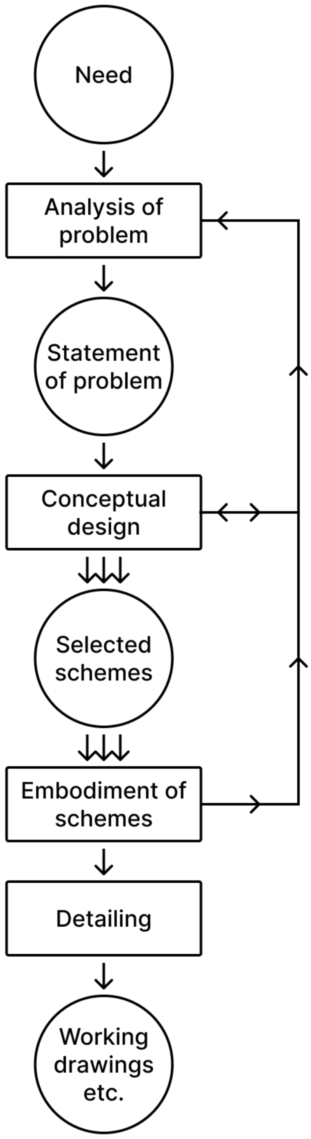


Figure 1.2: A block diagram illustrating the design process according to French. The circles represent stages reached, and the rectangles represent work in progress.

These frameworks provide varying degrees of structure and granularity to the design/development process, but all share the commonality of being highly engineering-focused. In engineering a physical product, a rigid, structured process is often necessary as each iteration must be designed, manufactured and tested. This is costly, both in effort and material costs. Therefore, the design and development process are divided and sequential. Software, on the other hand, is much more flexible, with iterations being a magnitude faster and cheaper to develop. This demonstrates a need for adapting the design/development process to the context of the product being developed. Conveniently, Ulrich and Eppinger present a multitude of adaptations to their framework, including what they refer to as ”Quick-Build Products” and ”Digital Products.” Here the *detail design* and *testing and refinement* stages are omitted, and replaced with a cyclical design-build-test process. Whereas the linear, rigid processes described previously are labelled as ”waterfall methods”, this iterative process is most often referred to as *agile development*.

Agile development has many benefits in contrast to the waterfall approach, especially in the context of software development. As mentioned previously, the waterfall approach is ideal for engineering projects where prototyping is costly. When the cost of prototyping is negligible, however, agile methodology grants the flexibility to iterate quickly, and adapt to evolving requirements. Design and development are sequential in a waterfall model; here they are heavily intertwined. A strong example of this, as well as being the most popular implementation of agile development, is *Scrum*. Scrum defines the following stages: *sprint planning*, *daily stand-up*, *sprint review*, and *sprint retrospective*, with a sprint typically lasting 2-4 weeks [9]. This framework is ideal for large teams, as it ensures that all team members are aligned, and are able to coordinate their efforts efficiently. The daily stand-up is a particularly useful tool for larger teams, preventing overlapping work, or potential blockers from being overlooked. In smaller teams, however, this structure can be cumbersome, and potentially even counterproductive. Especially when considering this project, exploring a singular use-case as a solitary developer, the full-scale implementation of Scrum is evidently not necessary. Instead, a more adaptable and lightweight framework is employed.

In his book, *The Lean Startup*, Eric Ries describes a simple, yet effective agile framework, which he refers to as the *Build-Measure-Learn* loop [10]. This framework is designed for rapid prototyping and iteration, splitting each stage into *build*, *measure* and *learn*. *Build* involves developing a minimal product or feature, which is then tested with the target user(s) in the *measure* stage. The results of this test are then studied in the *learn* stage, where the product/feature is adapted based on these results. This process is then repeated, until the requirements are met. This framework is ideal for this project, as it is quite exploratory in nature, while still aiming to fulfill predetermined requirements.

1.6.2 Design and development methodology

Building upon the exploration of frameworks conducted above, the following design methodology is proposed for this project. The design stage follows an engineering approach, guided by Ulrich and Eppinger’s *concept development* and *system-level design* stages. Subsequently, the development stage is based on a more agile, entrepreneurial approach, as described by Ries. Standing in for the *detail design* and *testing and refinement* stages, Ries’s *Build-Measure-Learn* architecture is employed in order to facilitate rapid prototyping and iteration, albeit with a slight augmentation. The *measure* and *learn* stages are consolidated through periodic feedback sessions with Roskilde Festival, where implemented features are reviewed and emerging requirements are gathered. This eliminates the implied analysis between these two stages, as this project is developed in close collaboration with the target user, and its usage is not intended for a wider audience.

Chapter 2, **Concept Development**, ultimately selects and proposes a conceptual solution to the problem outlined in section 1.2. As defined by Ulrich and Eppinger, a concept is ”a description of the form, function, and features of a product and is usually accompanied by a set of specifications, an analysis of competitive products, and an economic justification of the project.” [8] This selection is initially preceded by a thorough inspection into the intricacies of crowd safety management, as well as a review of potential solutions and existing products. A novel solution is thereafter presented, outlining requirement specifications, as well as its technical, legal, and financial feasibility.

Chapter 3 focuses on the **System-level Design**, including the definition of the product architecture, and a decomposition of the product into subsystems and components. The full workflow from data collection to the resulting user interface is presented, followed by a detailed description of each sub-system, including the data collection, computer vision model, spatial mapping, metric extraction, and the user interface/frontend.

Finally, Chapter 4 presents the **Results** of the development stage, beginning with a showcase of the frontend, including an overview of the iterative feature selection conducted in accordance with the augmented *Build-Measure-Learn* framework. This is followed by a technical performance evaluation, describing the accuracy of the solution. The chapter concludes by revisiting the business objectives outlined in section 1.5, and evaluating the product's business value. This includes a summary of a workshop conducted with Roskilde Festival, where the final product was presented and tested by members of the safety team.

2 Concept Development

2.1 Understanding crowd safety management

In order to develop a solution that supports crowd safety professionals, it is imperative to understand how they operate, and what tools they currently have at their disposal. Together with my co-founders at Fluxense, we conducted interviews with many crowd safety professionals from various different organizations, including Event Safety (Smukfest), smash! bang! pow! (Syd for Solen), Roskilde Festival, and Live Nation (Copenhell, Heartland). Throughout this period, it became clearer that crowd safety management is very complex, and is almost as much a philosophy as it is a science. Music festivals and events vary greatly in size, participant demographics, venues, and budget. Equally varied are the crowd safety professionals themselves, who appeared to have varying levels of experience, as well as distinct approaches to their work.

Most interestingly, the greatest discrepancy was seemingly between a focus on incident-prevention and incident-response, or ”crowd safety vs. security”, as according to Roskilde Festival’s Director of Safety, Morten Therkildsen (Appendix A.1). A security-focused approach often involves less planning, as well as hiring third-party professionals to handle safety during the event. Safety-focused teams, on the other hand, spend most of the year leading up to their events meticulously planning initiatives to ensure the well-being and enjoyment of their guests. The distinction between these two protocols was apparent throughout Fluxense’s collaborations with both Copenhell and Roskilde Festival. At their 2024 events, Live Nation had two full-time employees responsible for crowd safety at Copenhell, whereas Roskilde Festival had a team of 10+ full-time employees. Additionally, our collaboration with Copenhell was focused on real-time analysis, while Roskilde Festival was much more interested in post-event analysis.

This difference in approach is reflective of Denmark’s regulatory landscape regarding crowd safety. While official documentation exists, such as the ”Vejledning om sikkerhed ved udendørs musikarrangementer o.lign.” published by the Ministry of Justice and Ministry of Culture, these serve as guidelines rather than enforceable legislation specifically covering the planning phase of crowd safety management. The document itself emphasizes its role as a tool and catalog of ideas for organizers, who ultimately retain the responsibility for risk-assesment and implementation of appropriate measures based on their specific event [11]. Furthermore, the Danish Police have published supplementary guidance, describing the requirements for receiving a permit. While the police require organizers of large events to submit a safety plan as a condition for obtaining the necessary event permit, this is solely focused on incident-response [12]. Thus, although numerous legal requirements touch upon event safety, there isn’t distinct legislation dictating the specific process and minimum standards for proactive crowd safety planning. The legislation highlighted in these guidelines focuses on structural safety, emergency response, and police involvement, leaving the interpretation and extent of proactive safety planning largely to the individual organizers. This contributes to the observed discrepancies in how crowd safety management is implemented across different Danish festivals, as they interpret the available guidelines and integrate them into their operational workflows individually.

As the scope of this thesis is to develop a solution for Roskilde Festival explicitly, their distinct interpretation of crowd safety will be the exclusive focus in requirement gathering.

The following sections will outline the current frameworks and workflows used by Roskilde Festival, as well as the key metrics they calculate and monitor.

2.1.1 Existing frameworks and workflows

Roskilde Festival's approach to crowd safety management is built upon internationally recognized frameworks, prioritizing proactive planning and analysis. Roskilde Festival's Director of Safety, Morten Therkildsen, considers the United Kingdom to be on the forefront of crowd safety management, and the festival follows several UK-based frameworks to inform their practices. These include *The Purple Guide* [13], as well as the methodologies of Professor G. Keith Still, whose work highlighted in *Applied Crowd Science* has been adopted for mandatory training for public event commanders by the UK College of Policing since 2018 [14]. Roskilde Festival also references the *Event Safety Guide* from the US-based Event Safety Alliance (ESA) [15]. Roskilde's practical implementation of these frameworks is apparent in many of their workflows, several of which are outlined below.

An integral workflow involves conducting artist- and stage-specific risk assessments for every concert. The reasoning is to anticipate potential hazards based on the unique characteristics of each artist and their expected audience. This involves a detailed "band analysis" (Appendix A.1) considering factors like genre, typical crowd behavior, demographics, together with Still's RAMP analysis (evaluating routes, area, and movement specific to the stage, as well as the people/audience specific to the artist) [16]. Subsequently, concerts are categorized using a system reflecting risk based on the RAMP analysis (red/yellow/-green), and expected attendance relative to stage capacity (A: <25%, B: 25%-75%, C: 75%-100%, C+: 100%+). This assessment enables the safety team to tailor resources – such as staffing levels, barrier placement, and egress/flow management – to the predicted risk profile of each concert.

Understanding and managing how attendees move is addressed through flow analysis. Roskilde Festival's safety team analyze ingress/egress patterns and rates, particularly at entrances/exits, aggregating data across various time intervals (1, 15, 60 minutes). The objective is to prevent bottlenecks and ensure that the physical infrastructure can safely accommodate peak movements. This quantitative understanding of flow is crucial for accurately calculating required pathway widths, optimizing crowd control measures like queue management, and deploying staff to guide attendees when necessary. Additionally, crowd density analysis, using metrics like Levels of Service (LoS) inspired by John J. Fruin [**fruin**], is employed to monitor crowd concentration. This is especially conducted in areas with expected high density (e.g., front-of-stage) or potentially congested zones (e.g., queues around vendors). Its purpose is to prevent the high densities that pose a direct risk of crowd crushes, and to evaluate the effectiveness of the site layout in distributing the crowd effectively.

2.1.2 Key metrics

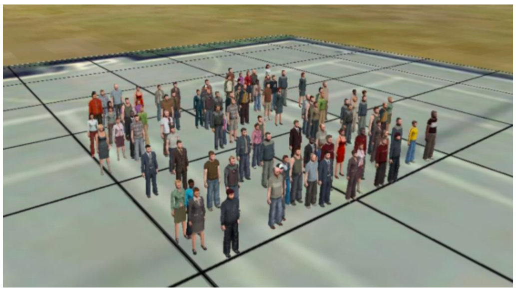
Summarizing the workflows outlined above, the key metrics that Roskilde Festival's safety team monitors include:

2.1.3 Revisiting the problem definition

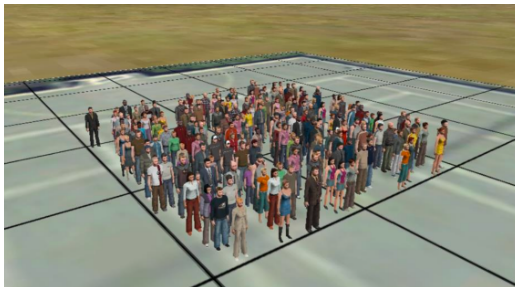
2.2 Comparing technical solutions

2.2.1 Global Positioning System (GPS)

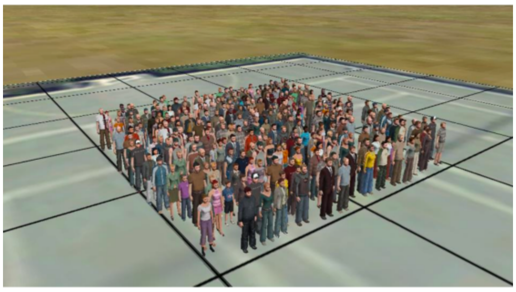
Using GPS to track the location of festival-goers is a common practice, and likely the easiest to implement technology in this comparison. This is typically achieved by providing



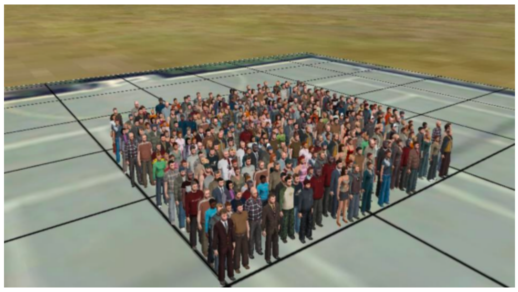
A (0-1 person/m^2). Head, shoulders, chest, and feet are visible.



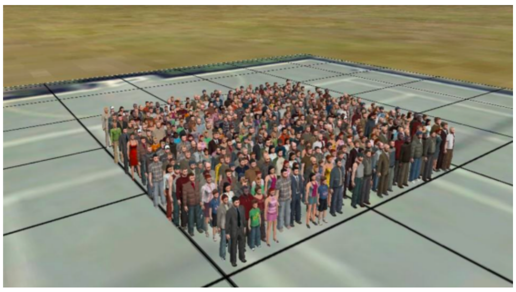
B (1-2 people/m^2). Head, shoulders, chest, and feet are visible.



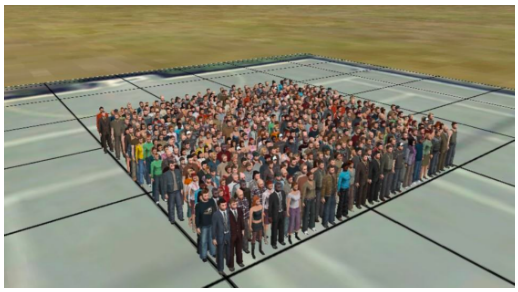
C (2-3 people/m^2). Head, shoulders, and chest are visible.



D (3-4 people/m^2). Head, shoulders, and chest are visible.

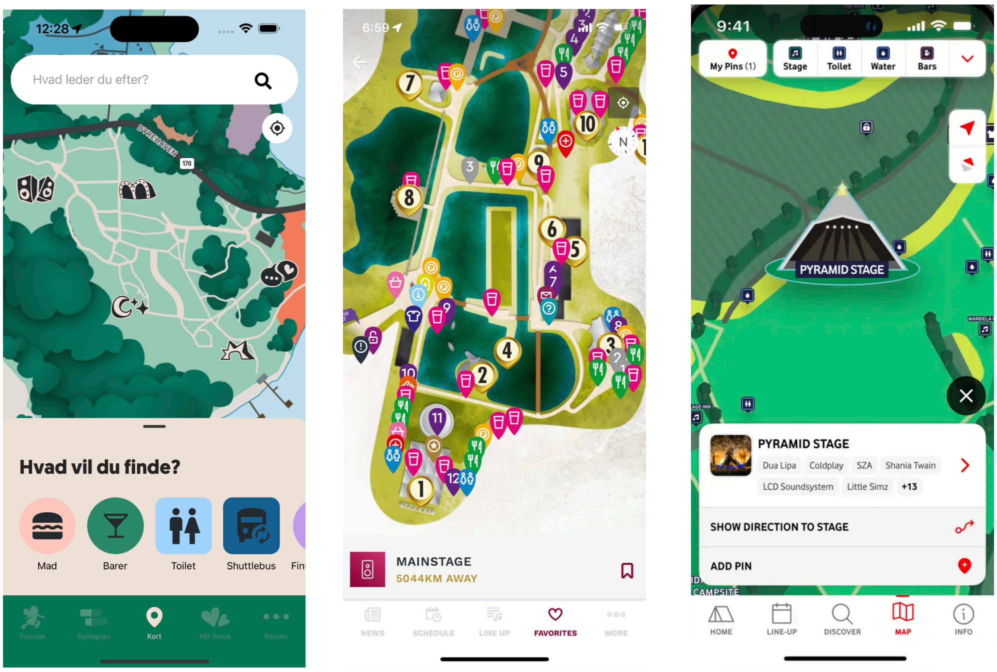


E (4-5 people/m^2). Head and shoulders are visible.



F (5+ people/m^2). Only heads are visible.

Figure 2.1: Roskilde Festival's internal implementation of Levels of Service (LoS), using a scale of A-F. The images serve as visual examples of the different LoS levels, used to estimate crowd density on-site or through video footage. Roskilde relies on visual cues, namely the visibility of heads, shoulders, and feet, to assess crowd density.



(a) Smukfest app [18] (b) Tomorrowland app [19] (c) Glastonbury app [20]

Figure 2.2: Examples of map features in music festival mobile apps (Smukfest, Tomorrowland, Glastonbury), often used to encourage attendees to enable GPS location tracking.

guests with a mobile app that uses their smartphone's GPS to track their location. Of course, this requires the guests to opt in to location tracking, as well as there being a sufficient reason for doing so. In almost all cases, this is attempted by including a map of the festival in the app, as visible in Figure 2.2. This feature, however, still functions without location tracking, and therefore doesn't guarantee users will grant data access. Even before this obstacle is met, there is the question of whether festival-goers will actually use the app. A study on Roskilde Festival 2015 by the Copenhagen Business School found that of the 60 thousand people who installed the festival application, 44 thousand opted-in to allowing anonymous tracking; yielding 38.678 unique users who were present inside the festival area [17]. This equates to slightly under 30% of the total 130 thousand attendees. In a crowd safety context, this is a significant limitation, as the location data gathered is not representative of entire crowds.

Beyond low adoption and potential privacy concerns, the technical limitations of GPS also hinder its utility for detailed crowd analysis. According to GPS.gov, GPS-enabled smartphones are typically accurate only to within a 4.9 m radius under open sky; however, their accuracy worsens near buildings, bridges, and trees [21]. While a 4.9-meter radius might seem acceptable for general location awareness on a festival map, this level of uncertainty significantly hinders the calculation of precise crowd dynamics metrics. Furthermore, the degradation of accuracy near structures is particularly problematic in festival environments, which often feature large stages, tents, and temporary structures – precisely where accurate monitoring is most needed. The effectiveness of GPS tracking is also contingent on factors outside the organizers' control, such as users keeping their phones charged and maintaining a stable mobile data connection.

Compared to infrastructure-based monitoring systems (like cameras or dedicated sensors), GPS relies heavily on user cooperation and device functionality, making it less suitable for generating the consistent, high-resolution data needed for proactive crowd safety management and detailed post-event analysis. Therefore, while mobile app GPS data can offer some high-level insights into general attendee distribution, its inherent limitations in accuracy make it insufficient as a primary tool for gathering crowd dynamics measurements.

2.2.2 Bluetooth beacons

Another potential solution for gathering crowd dynamics data involves utilizing Bluetooth Low Energy (BLE), a wireless communication technology designed for low power consumption. This technology is typically utilized in Bluetooth beacons – devices that periodically transmit signals that can be detected by nearby receivers, such as smartphones or dedicated hardware. These signals can be used to determine the proximity of beacons to the receivers, as well as location estimation through triangulation techniques, requiring three or more receivers. This capability is increasingly utilized for proximity marketing, indoor navigation, and positional tracking, typically in retail environments [22]. It has also been effectively used for guest tracking in environments like museums, as demonstrated in a study at the Galleria Borghese, where visitors were given portable BLE beacons tracked by fixed receivers [23].

Finding examples of BLE beacons utilized at large outdoor events is more challenging, as the technology has its own set of drawbacks. Similar to GPS-based mobile apps, the effectiveness of Bluetooth beacon systems is significantly contingent on user cooperation. Beacons can be implemented either as dedicated physical devices or embedded within smartphone applications. The latter is simpler and cheaper, however motiving users to install an app is challenging, as discussed in section 2.2.1. The former approach has challenges of its own, as requiring event participants to carry a dedicated device demands a value proposition of its own, not to mention the overhead costs involved in developing, deploying, and maintaining the system. Given that beacon signals have an effective range of 30 meters without obstructions, a considerable number of receivers are necessary to ensure adequate coverage [22].

Ultimately, the most critical factor for crowd dynamics analysis is the technology's accuracy. A study conducted at a large exhibition in Japan evaluated the accuracy of trajectory estimation using Bluetooth beacons. They found that in 68% of instances, their positional estimates were accurate within a radius of 18 meters [24]. While still useful for certain applications like indoor tracking or general zone classification, this level of accuracy is significantly less precise than typically achievable with GPS (Section 2.2.1). This limitation, combined with the dependency on user adoption and implementation costs, makes Bluetooth beacons a less favorable option for gathering crowd dynamics data at large outdoor events.

2.2.3 Camera solutions

Camera-based solutions leverage existing CCTV systems in conjunction with computer vision algorithms to analyze crowd dynamics. Utilizing video footage from these camera systems, novel advancements in machine learning and computer vision enable automated extraction of valuable metrics.

A primary advantage of this approach compared to GPS (Section 2.3.1) and Bluetooth beacon (Section 2.3.2) systems is its independence from attendee cooperation. Unlike solutions requiring festival-goers to install an app, enable location services, carry a specific device, or keep their phones charged, camera-based analysis operates passively. It gathers data from anyone within the camera's field of view, potentially offering a more comprehensive and unbiased view of crowd behavior across monitored areas. This circumvents the significant limitations associated with low adoption rates and opt-in requirements inherent in user-device-dependent methods.

(competitor analysis)

2.3 Proposed solution

2.4 Feasibility of solution

2.4.1 Technical feasibility

2.4.2 Legal feasibility

When dealing with CCTV footage, there are a number of legal considerations to take into account.

2.4.3 Financial feasibility

3 System-level Design

3.1 Product architecture

3.2 Sub-systems

3.2.1 Data collection

Data collection constitutes the initial, vital stage of the system, providing the raw data required for all subsequent processing and analysis. This process entailed the on-site deployment of designated cameras, strategically positioned to capture the targeted crowd dynamics.

The selected cameras were Reolink RLC-520A, which are PoE-enabled (Power over Ethernet) and capable of recording 5MP (2560x1920 pixels) video at 30 frames per second. A separate PoE switch (Ubiquiti PoE++ Adapter), connected to a standard power outlet, was required to power the cameras. This setup allowed connection to the cameras via an Ethernet cable linked to a laptop, enabling camera configuration and live feed monitoring. Utilizing the cameras' integrated software, recording windows could be predefined such that footage would automatically be archived on the internal SD card. This setup was designed to ensure that the cameras could operate independently without requiring a constant connection to a computer. Mounting the cameras was achieved with 3D-printed brackets, designed to securely attach the cameras to existing infrastructure, such as fences or poles. Where existing structures were unavailable, aluminum poles were utilized to achieve the necessary height for capturing the entire designated area. Four cameras were deployed, denoted as *CAM1*, *CAM2*, *CAM3*, and *CAM4*.

As agreed upon with Roskilde Festival's safety team, the cameras were installed around the Eos stage during the first three days of the festival, and subsequently moved to the Arena stage for the remainder of the festival.

Determining camera placement for the Eos stage was relatively trivial. During the festival's "First Days," (June 30th to July 2nd) the remainder of the festival site, excluding the Gaia stage, was closed off to guests, leaving only two pathways for entering and exiting the stage area. At each of these two pathways, two cameras were installed, oriented to face one another. *CAM1* and *CAM3* were positioned at Eos' southern entrance, Entrance 10, while *CAM2* and *CAM4* were placed at the eastern entrance, towards the Gaia stage (Figure 3.2). This dual-camera configuration served two purposes: ensuring complete monitoring of the pathway's width, as well as providing a redundant dataset for each location, effectively mitigating the risk of equipment malfunction during the initial deployment.

Camera placement for the Arena stage presented greater complexity due to its significantly larger scale and increased number of entrance and exit points. Arena is located at the far eastern corner of the festival grounds, and it was anticipated that the majority of attendees would approach from the western side, where the remainder of the festival's stages are located. This involved three possible entrances/exits: *the Stables*, *the Graffiti Walk*, and *the Fast-Track*. As the Graffiti Walk is the broadest and most heavily trafficked route, *CAM1* was mounted atop a tall utility pole to ensure comprehensive coverage of this wide pathway. *CAM2* was mounted on an aluminum pole to overlook the fast-track, and *CAM4* was positioned to capture the stables entrance. Finally, *CAM3* was placed at the southeast corner of the stage, at a junction point of the fast-track and Entrances 5 & 6. This camera was positioned to capture individuals entering and exiting the latter two

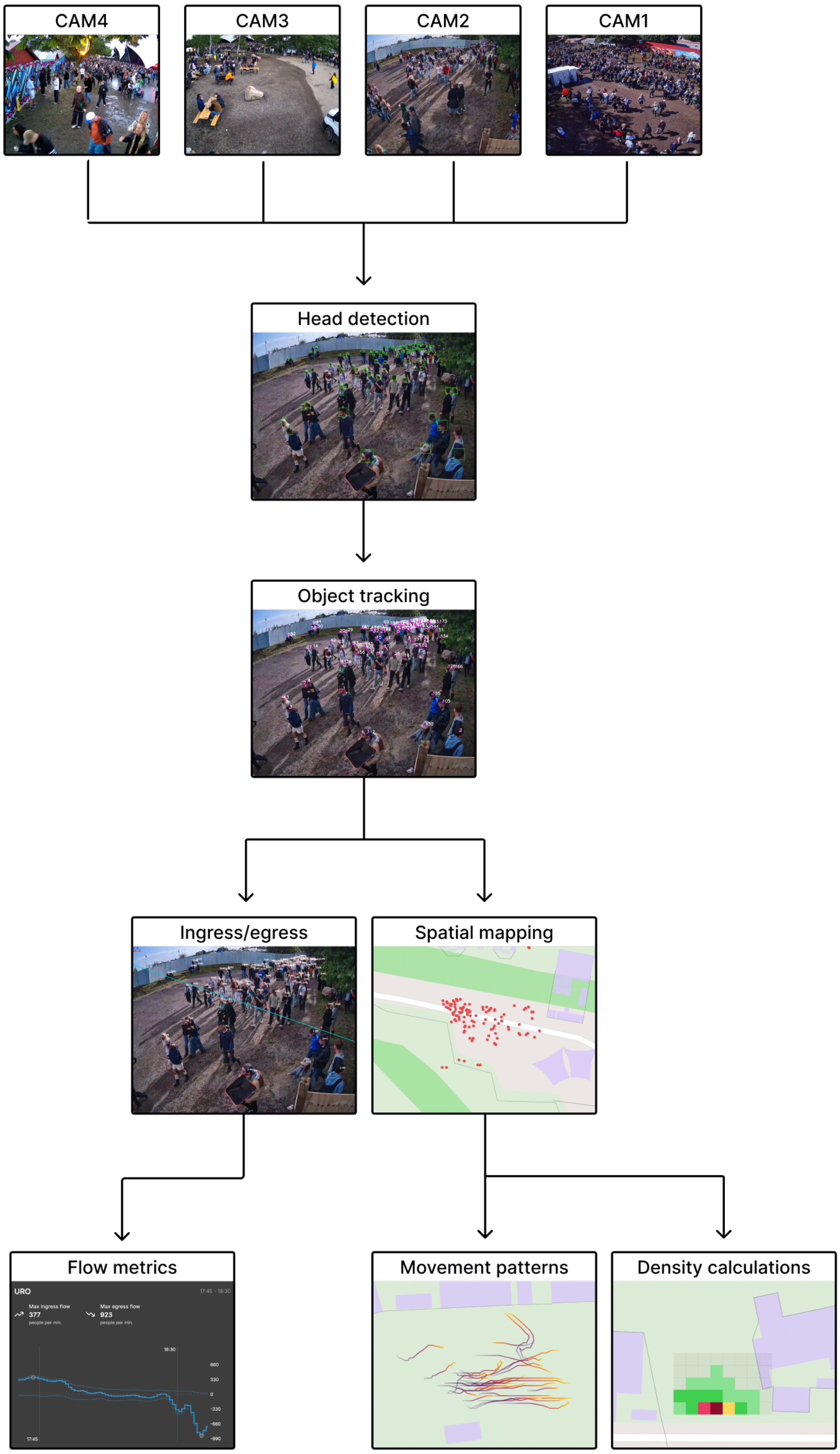
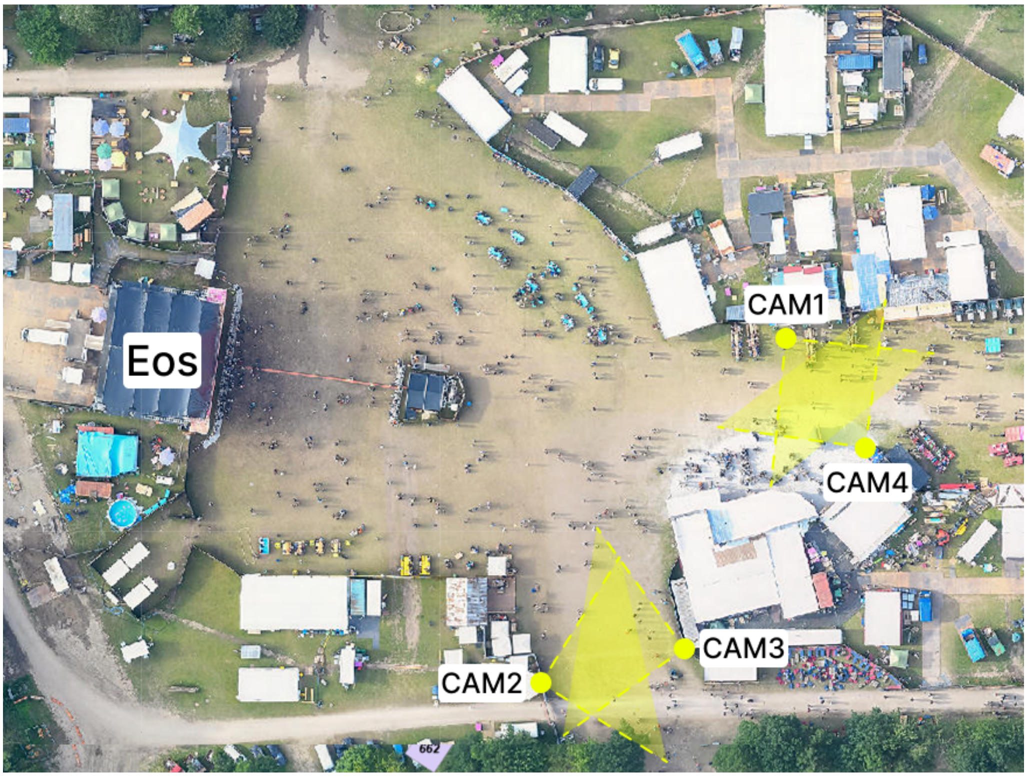
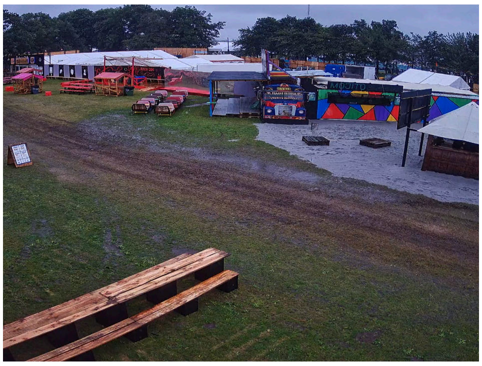


Figure 3.1: Flowchart visualizing the system architecture.



(a) Eos stage area – visualized with camera placements



(b) CAM1 Preview

(c) CAM2 Preview

(d) CAM3 Preview

(e) CAM4 Preview

Figure 3.2: Camera placement at Eos stage, with approximate field of view (FOV) indicated (a). The bottom images show sample frames from the four cameras deployed at the Eos stage, showing the field of view for each camera position (b-e).

pathways. Altogether, these camera positions were designed to theoretically provide full coverage of the Arena stage's entrances and exits, allowing accurate metric extraction (Section 3.2.4). See Figure 3.3 for a visualization of the camera placements.

The Reolink RLC-520A cameras also included infrared (IR) night vision capabilities, theoretically allowing for monitoring in low-light conditions. However, this feature proved ineffective in practice as the integrated software lacked functionality for time-based switching between day and night modes, offering only a subjective slider for ”ambient brightness”. This created unpredictability in terms of when the cameras would switch between modes, sometimes leading to the cameras using IR imaging during daylight hours. As this resulted in reduced image quality and limitations on the video frame rate, the cameras were configured to operate without the IR functionality. Therefore, the cameras were set to record continuously for only 12 hours each day, from 12:00 to 24:00. Furthermore, due to the cameras' maximum SD card storage capacity of 256 GB, archived footage required manual download and purging when relocating the equipment between stages.

3.2.2 Computer vision model

The core of the system's ability to analyze crowd dynamics relies on a robust computer vision model capable of detecting individuals within the video footage. This section details the model selection, training process, as well as its implementation for inference.

Model selection

The primary objective of the computer vision model in this system is the detection and localization of individuals within video frames; both are prerequisites for subsequent tracking and the metric extraction. Therefore, the selected model must provide image coordinates for each detected person, rather than merely an aggregate count. Methodologies in crowd analysis typically follow either density estimation approaches, which generate maps representing crowd concentration, or detection/localization-based approaches, which identify the coordinates of each individual, often via points or bounding boxes. As tracking individual trajectories is fundamental to this project's goals, localization-based methods were deemed most relevant.

The model selection process involved evaluating architectures benchmarked on an established crowd analysis dataset, NWPU-Crowd [25]. Other prominent datasets known for their complexity include ShanghaiTech and JHU-CROWD++ [26] [27]. These datasets encompass diverse scenarios and significant variations in crowd density.

One candidate architecture considered was CrowdHat, a recently proposed model aimed at enhancing the localization performance of standard object detectors within crowded environments [28]. Given its high position on the NWPU-Crowd localization leaderboard as well as its availability as open-source software, CrowdHat was selected for initial evaluation.

Another prominent architecture evaluated was You Only Look Once (YOLO), representing a widely adopted family of object detectors known for high inference speed [29]. This project employed the YOLOv8 implementation by Ultralytics. Standard YOLOv8 models are pre-trained on the large-scale COCO dataset, providing a generalized object detection capability [30] [31]. However, achieving optimal performance for the specific domain of festival crowds necessitates fine-tuning this pre-trained model on representative video data captured during the event. YOLOv8 was chosen for comprehensive testing and ultimate deployment due to its established performance benchmarks, the ease of implementation offered by the Ultralytics library, as well as its significant advantage in processing speed.

(a) Arena stage area – visualized with camera placements

(b) CAM1 Preview

(c) CAM2 Preview

(d) CAM3 Preview

(e) CAM4 Preview

Figure 3.3: Camera placement at Arena stage, with approximate field of view (FOV) indicated (a). The bottom images show sample frames from the four cameras deployed at the Arena stage, showing the field of view for each camera position (b-e).

A comparison of processing efficiency revealed that CrowdHat processes frames at an average of [Insert CrowdHat Inference Time] ms/frame, whereas YOLOv8 achieves an average inference time of [Insert YOLOv8 Inference Time] ms/frame on the identical hardware configuration. Considering the practical requirement for efficient video analysis alongside robust detection accuracy, YOLOv8 offered a superior balance. Consequently, the fine-tuned YOLOv8 model was selected as the definitive detector for integration into the system.

Annotation

While pre-training on large, diverse datasets like COCO provides the YOLOv8 model with a robust general object detection capability, achieving optimal performance necessitates fine-tuning on a custom-annotated dataset. The rationale for this extends beyond simply adapting to the general festival environment; the goal is to develop highly specialized models optimized for the precise conditions and appearance characteristics encountered by each camera at each specific deployment position. It is hypothesized that such hyper-specialization yields superior detection accuracy compared to a more generalized model.

The annotation process utilized Label Studio, an open-source data labeling tool [32]. A self-hosted instance was employed to ensure data privacy and control, preventing the need to upload potentially sensitive video material captured on-site to third-party services.

Random frames were extracted from the video recordings detailed in Section 3.2.1. The annotation target was specifically the heads of individuals, rather than full bodies. This choice was predicated on the assumption that in dense crowd scenarios, heads are more consistently visible than entire bodies, providing a more reliable feature for detection and subsequent tracking. Bounding boxes were drawn around each identifiable head, assigned the single class label ”person”.

Specific annotation guidelines were established to ensure consistency:

- Bounding boxes were drawn tightly around the visible extent of the head, explicitly including hair.
- In cases of occlusion, where one head partially blocks the view of another, overlapping bounding boxes were permitted.
- If individuals were distant (e.g., in the far background) or within extremely dense parts of the crowd, making distinct heads difficult to discern, bounding boxes were only placed where a head could be clearly distinguished. Ambiguous cases were omitted to maintain label quality.

To minimize time spent annotating, an iterative approach was used. After an initial batch of frames was annotated, a YOLOv8 model was trained on this preliminary data. This temporary model was then integrated with Label Studio to pre-annotate subsequent frames by suggesting bounding boxes based on its predictions. This workflow is commonly referred to as human-in-the-loop, and significantly expedited annotation time, as the task increasingly involved refining or validating the model’s suggestions rather than manually creating annotations. Additionally, this methodology offered a visual indicator of model performance, as the diminishing need for manual adjustments indicated that the model was becoming more robust and that enough data had been annotated for final training.

The annotation results are summarized in Table **??**.

Stage	Camera	Images	Bounding Boxes
Eos	CAM1	[Number]	[Number]
	CAM2	[Number]	[Number]
	CAM3	[Number]	[Number]
	CAM4	[Number]	[Number]
Arena	CAM1	[Number]	[Number]
	CAM2	[Number]	[Number]
	CAM3	[Number]	[Number]
	CAM4	[Number]	[Number]
Total		**[Total Number]**	**[Total Number]**

Table 3.1: Annotation statistics per camera deployment

Training

To develop the specialized models required for each camera deployment, the YOLOv8m model, pre-trained on the COCO dataset, was fine-tuned using the custom annotated dataset (detailed in previous section). This process utilized the Ultralytics framework [30].

Training was configured with the following key hyperparameters: an input image size of 1280x1280 pixels, a batch size of 4, and training duration of 600 epochs with early stopping patience of 200 epochs. The process was optimized for the single ”person” class. To accelerate the computationally intensive training process, computations were performed on a personal computer equipped with an Nvidia RTX 4080 SUPER graphics card, leveraging GPU acceleration.

This fine-tuning stage produced specialized model weights adapted to the unique visual characteristics of each camera view at the festival, which were then used for the subsequent inference stage.

Inference and Tracking

Following the training phase, the inference stage employs the specialized YOLOv8 models to detect heads within the recorded video footage, which are then tracked across frames using an object tracking algorithm. This combined inference and tracking pipeline generates the foundational data required for subsequent spatial mapping and metric extraction. The process begins by loading the fine-tuned model weights specific to the camera view being analyzed. Input videos are processed in parallel, analyzing each frame individually to identify and track individuals present. Videos are processed at a rate of 15 frames per second (FPS), which is half the recorded frame rate of 30 FPS.

Each frame is resized to a 1280x1280 pixel resolution, as defined during training, before running inference. This prepared frame is then passed to the fitted YOLOv8 model, which outputs bounding boxes around predicted heads and assigns confidence scores to these detections. These predictions are filtered; detections falling below a predefined confidence threshold are discarded (40%), and Non-Maximum Suppression (NMS) is applied to resolve significant overlaps between bounding boxes, preserving only the most confident prediction.

The resulting bounding boxes and confidence scores are subsequently passed as input to the object tracking module. This project utilizes ByteTrack, a high-performance algorithm chosen for its accuracy, particularly within crowded scenes [33]. ByteTrack associates the current frame's detections with previous frames, assigning a unique ID to each tracked individual.

The final output for each frame is a list containing the bounding box coordinates, the assigned tracking ID, and the detection confidence score for every tracked individual. This data serves as input for the spatial mapping system described in the following section.

3.2.3 Spatial mapping and GIS

While the computer vision model (Section 3.2.2) outputs the locations of individuals in terms of pixel coordinates within the video frame, these coordinates alone are insufficient for thorough analysis of crowd dynamics. In order to derive area-based metrics such as crowd density (people per square meter), movement speeds and distances, it is necessary to translate these pixel positions into real-world geographic coordinates. Spatial mapping has this purpose, providing the planar transformation between the camera's perspective and the geographic context of the festival. This mapping also enables visualization of individual positions, from multiple cameras, onto a single overhead map, significantly enhancing contextual understanding, compared to that which is achievable through video footage alone.

The technique employed for this spatial mapping is *homography*: a transformation that maps points from one plane to another. In this context, it establishes a mathematical translation between the pixel coordinates in the 2D camera image plane and the corresponding real-world coordinates on the ground plane [34]. This allows any detected pixel coordinate within a defined area to be projected onto its actual geographic location.

Performing the homography calculation requires two sets of corresponding points: one set in the camera's pixel coordinates and another in real-world coordinates. The latter is obtained from Roskilde Festival's internal Geographic Information System (GIS) tooling, which provides precise GPS coordinates of all infrastructure on the festival grounds. Selecting corresponding points was achieved by identifying distinct, stationary landmarks visible in the video frame, such as corners of structures or fences, and marking their position in pixel coordinates. These landmarks were then located in the GIS utility, where their GPS coordinates were recorded. Given a minimum of four distinct pairs of corresponding points, the homography transformation can be computed – in this case, utilizing the OpenCV Python library [35]. This resulting transformation is stored in a configuration file associated with that camera deployment. This manual process was performed once for each camera deployment, and the resulting homography matrices were used for all subsequent video footage captured by that camera. See Figure 3.4 for an illustration of the results of this mapping process.

Note a slight limitation of homography in this context; homography assumes that the mapping occurs between two planar surfaces. While the surfaces in the camera and map are treated as planar, the Earth's surface is curved. For the relatively small geographic areas covered by individual camera views, the ground surface can be reasonably approximated as flat. The error introduced by this assumption is considered negligible for the purposes of crowd analysis at this scale.

3.2.4 Metric extraction

Following the spatial mapping process (Section 3.2.3), which translates tracked individuals' pixel coordinates into real-world geographic coordinates, the metric extraction system processes this positional data to derive quantitative insights into crowd dynamics.

(a) Camera view (b) Map view

Figure 3.4: Illustration of homography mapping between camera view and map view.

This stage is the crucial step transforming raw data into the key metrics utilized by crowd safety professionals. The primary metrics extracted, as defined in Section 2.1.2, include ingress/egress counts, flow rates, cumulative counts, crowd density, and movement patterns.

Ingress/Egress and Flow Rate Calculation

To measure the flow of people into and out of specific areas, lines are drawn across each camera view. These lines correspond to the entrances/exits covered by the cameras (as detailed in Section 3.2.1, Figures 3.2 and 3.3). The system follows the trajectory of each tracked individual (identified by a unique ID from the tracking algorithm, Section 3.2.2). When a trajectory crosses the virtual line, the system registers it as either an ingress or egress event based on the direction of crossing.

These individual crossing events are then aggregated over specific time intervals. Intervals of 1 minute, 15 minutes, and 1 hour were identified as most useful by Roskilde Festival (Appendix A.1). This aggregation allows the calculation of the ingress/egress flow rate, representing the number of people entering/exiting the area per time unit. The net flow rate is calculated as the difference between ingress and egress flow rates, indicating the rate of change in the number of people within the area.

Furthermore, a cumulative count provides a running total of the net number of people within the monitored area over time, calculated by aggregating ingress/egress over time. This cumulative count is particularly useful for understanding the overall occupancy of the area. These flow metrics can be calculated for individual entrances or cameras, or they can be aggregated to provide a total flow for a larger area such as an entire stage. The system also identifies and records maximum ingress and egress flow rates observed during specific periods, like during a concert. This data is essential for understanding peak loads, capacity planning, and validating entrance width calculations.

Crowd Density Calculation

Crowd density, measured in people per square meter (people/m²), is another critical metric for assessing safety and comfort. To calculate density, the monitored area is divided into a grid of 3x3 meter cells. At discrete time intervals, the system counts the number of tracked individuals whose mapped geographic coordinates (Section 3.2.3) fall within each grid cell. The density for that cell or zone is then calculated by dividing this count by the known area of the cell or zone.

Movement Patterns

Beyond counts and density calculations, understanding how crowds move is crucial. By analyzing the sequence of time-stamped geographic coordinates, or trajectories, associated with each unique tracking ID, the system can visualize movement patterns. This involves plotting the paths taken by individuals over time, represented as gradient lines on the map.

Analyzing these patterns can reveal dominant flow directions, identifying the primary paths people take when moving between locations. It can also aid in detecting cross-flow, areas where different streams of people intersect. Additionally, this analysis helps understand origin-destination patterns, showing where people come from when approaching an area, such as which pathway they used, and where they head afterwards.

3.2.5 Interface/frontend

The interface/frontend is the final component of the system, and is the only component visible to the end-users. It is therefore designed to condense the complexity of the subsystems described in the previous sections into an accessible and user-friendly interface. The application employs a modern web development stack: namely, React, Next.js, TypeScript, and Tailwind CSS.

Data storage and retrieval is managed through a PostgreSQL database, accessed via the Prisma Object-Relational Mapper (ORM). The database contains predefined information, including camera configurations, processed count data, and timestamped geographic point data. Additionally, the database stores user login information, as well as user-generated labels/notes for specific time intervals and cameras.

The entire frontend application is deployed and hosted using Vercel. See full overview of the frontend application in Section 4.1.

4 Results

4.1 Frontend showcase

4.2 Technical performance evaluation

4.3 Business value

5 Conclusion

5.1 Summary of results

5.2 Market expansion opportunities

5.3 Technical challenges and lessons learned

5.4 Technical improvements

5.5 Closing remarks

References

[1] David Fricke. *Nine Dead At Pearl Jam Concert*. Aug. 2000. URL: https://web.archive.org/web/20180830050458/https://www.rollingstone.com/music/music-news/nine-dead-at-pearl-jam-concert-235167/.

[2] YOUROPE - The European Festival Association. *Roskilde Festival*. URL: https://web.archive.org/web/20241212233325/https://yourope.org/member/roskilde-festival/.

[3] Vladimir Hubka and Wolfgang Ernst Eder. *Design Science: Introduction to the needs, scope and Organization of Engineering Design Knowledge*. Springer London, 1996.

[4] Wolfgang Eder. "Why Systematic Design Engineering?" In: vol. 8. Jan. 2009. DOI: 10.1115/DETC2009-86067.

[5] Nigel Cross. *Engineering Design Methods: Strategies for Product Design*. Third Edition. John Wiley and Sons, Ltd, 2000.

[6] Michael Joseph French. *Conceptual Design for Engineers*. Second Edition. Springer-Verlag Berlin Heidelberg GmbH, 1985.

[7] Gerhard Pahl et al. *Engineering Design: A Systematic Approach*. Third Edition. Springer, 2007.

[8] Karl T. Ulrich, Steven D. Eppinger, and Maria C. Yang. *Product Design and Development*. Seventh Edition. McGraw-Hill Education, 2020.

[9] Ken Schwaber and Jeff Sutherland. *Scrum Guide*. Scrumguides.org. Nov. 2020. URL: https://scrumguides.org/scrum-guide.html.

[10] Eric Ries. *The Lean Startup: How Toda's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business, 2011.

[11] Kulturministeriet. *Vejledning om sikkerhed ved udendørs musikarrangementer o.lign.* URL: https://web.archive.org/web/20211229000406/https://kum.dk/fileadmin/_kum/5_Publikationer/2014/Vejledning_om_sikkerhed_ved_udendoers_musikarrangementer_2014.pdf.

[12] POLITI. *Vejledning om udarbejdelse af sikkerhedsplan for større udendørs musikarrangementer o.lign.* URL: https://web.archive.org/web/20240927194102/https://politi.dk/-/media/mediefiler/landsdaekkende-dokumenter/tilladelser/begivenheder/vejledning-om-udarbejdelse-af-sikkerhedsplan-for-strre-udendrs-musikarrangementer-ol.pdf.

[13] *The Purple Guide to Health, Safety and Welfare at Outdoor Events*. Food, Events and Things (FEAT) Ltd, 2021.

[14] G. Keith Still. *Applied Crowd Science*. CRC Press, 2022.

[15] *The Event Safety Guide*. Event Safety Alliance of USA, Inc., 2016.

[16] G. Keith Still. "Ramp analysis". In: *Applied Crowd Science* (Dec. 2021), pp. 103–117. DOI: 10.1201/9781351053068-5.

[17] Chris Zimmerman et al. "Space vs. Place: Comparing Space-based Movements and Place-based Experiences at the Roskilde Festival 2015". In: vol. 11. 2016, pp. 1–10. DOI: 10.1145/2930971.2930983.

[18] Apple App Store. *Smukfest 2025*. URL: https://web.archive.org/web/20250408172430/https://apps.apple.com/dk/app/smukfest-2025/id6447997155.

[19] Reddit. *Tomorrowland Map*. URL: https://web.archive.org/web/20220709110843/https://preview.redd.it/m5jhl2mvyia91.jpg?width=1170&format=pjpg&auto=webp&s=2c44480b31448b153fc81a2dc11269662ebdd985.

[20] 3 SIDED CUBE. *An App To Truly Elevate The Festival Experience For Every Attendee*. URL: https://web.archive.org/web/20250207080506/https://3sidedcube.com/projects/an-app-to-truly-elevate-the-festival-experience-for-every-attendee.

[21] GPS.gov. *GPS Accuracy*. URL: https://web.archive.org/web/20250421225547/https://www.gps.gov/systems/gps/performance/accuracy/.

[22] Wiliot. *Bluetooth Beacon*. URL: https://web.archive.org/web/20240529070543/https://www.wiliot.com/bluetooth-beacon#applications-of-bluetooth-beacons.

[23] Pietro Centorrino et al. "Managing Crowded Museums: Visitors Flow Measurement, Analysis, Modeling, and Optimization". In: *Journal of Computational Science* 53 (July 2021). URL: https://arxiv.org/abs/2006.16830.

[24] Sakurako Tanida et al. "Quantifying crowding perception at large events using beacons and surveys". In: *Scientific Reports* 15.1 (Jan. 2025). DOI: 10.1038/s41598-024-83065-4.

[25] Qi Wang et al. "NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.6 (2021), pp. 2141–2149. DOI: 10.1109/TPAMI.2020.3013269.

[26] Yingying Zhang et al. "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 589–597. DOI: 10.1109/CVPR.2016.70.

[27] Vishwanath Sindagi, Rajeev Yasarla, and Vishal Patel. "JHU-CROWD++: Large-Scale Crowd Counting Dataset and A Benchmark Method". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP (Nov. 2020), pp. 1–1. DOI: 10.1109/TPAMI.2020.3035969.

[28] Shaokai Wu and Fengyu Yang. "Boosting Detection in Crowd Analysis via Underutilized Output Features". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2023, pp. 15609–15618.

[29] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[30] Ultralytics. *Explore Ultralytics YOLOv8*. URL: https://web.archive.org/web/20250426153402/https://docs.ultralytics.com/models/yolov8/.

[31] Tsung-Yi Lin et al. "Microsoft COCO: Common Objects in Context". In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, 2014, pp. 740–755. ISBN: 978-3-319-10602-1.

[32] Maxim Tkachenko et al. *Label Studio: Data labeling software*. Open source software available from https://github.com/HumanSignal/label-studio. 2020-2025. URL: https://github.com/HumanSignal/label-studio.

[33] Yifu Zhang et al. "ByteTrack: Multi-object Tracking by Associating Every Detection Box". In: *Computer Vision – ECCV 2022*. Springer Nature Switzerland, 2022, pp. 1–21. ISBN: 978-3-031-20047-2.

[34] A. Torralba, P. Isola, and W.T. Freeman. *Foundations of Computer Vision*. Adaptive Computation and Machine Learning series. MIT Press, 2024. Chap. 41. ISBN: 9780262378666. URL: https://mitpress.mit.edu/9780262048972/foundations-of-computer-vision/.

[35] OpenCV. *Camera Calibration and 3D Reconstruction*. URL: https://web.archive.org/web/20250327170811/https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html#ga4abc2ece9fab9398f2e560d53c8c9780.

A Appendix

A.1 Roskilde Festival meeting notes

A.1.1 February 16, 2024

Roskilde currently uses GPS phone signal to get overview of guest counts through their app. Updates every 15 minutes to get a signal. Greentel system, the function is called CrowdView.

Morten says there are many systems available, but all come with a disadvantage - in Denmark, they dont use many cameras, as they are expensive to set up. On top of that they need to move them and this fu**s up the analysis done by cameras.

Morten tested DCM, they use a grid-like approach to calibration of cameras . It can tolerate some camera movement but not much.

- Gratest value of DCM is in his eyes is the after-analysis of events. For example to see how their barrier design influences flow.
- How fast can a scene get emptied, what is the flow of people from A to B. How fast can a scene fill up and what are the connections to the temporal aspects of the show (concert starting etc.)
- NOT interested in real time analysis - not even in front of stages, as there are plenty of people there present as ground crew and they can tell.
- Sometimes, the crowd issues happen outside of view of cameras (example of Ukendt kunstner concert at Arena stage) far away from stage

It is crucial to build this product/app/function in collaboration with educated crowd management professionals, and the accuracy of reporting is key. Following the current industry standards of flow estimation, density calculation in pre-planning of venues. Otherwise he will NOT use it.

DCM is used at Silverstone in areas where they know they have issues with crowd flow as an analysis tool.

Morten says, psychology of people has changed after corona - They go much more to concerts and they are more hesitant to stand close together.

Festivals are scaling operations - during festivals they have a lot of employees, after festival everybody leaves and with that also knowledge of what was done last year. Creating documentation of layouts, crowd management plans and general analysis of how things were will be of great value.

The Police does security - preventing harm where there could be intent of harm

Roskilde does safety - Accomodating people so there is no harm in first place

Safety is about creating risk assessments for possible situations.

Communication: two separate channels 112 for emergencies and 114 for internal communication. They use VHF and Whatsapp groups. Sometimes phone signal is an isssue.

Mads Therkilsen is responsible for flow analysis, and early in researching how AI tools can help with crowd management. (feeding existing information about venues to the …model.)

We will recieve contacts to people from england doing Crowd Science. Keith Still - mathematician with focus in this field.

Takeaway is - analysis is more important than real time insights

RF meeting notes – February 16, 2024

A.1.2 March 10, 2024

Crowd management and crowd safety is something some companies do, and some don't. In any case, to get licensed to make an event, you need to follow standard guides for a risk/event safety assesment that is submitted to the police. Issue is, the standard guide is very high-level and 'doesnt go in depth of what the actual safety plan is. Therefore, things are up for interpretation.

LiveNation 'doesnt plan, and they probably 'dont have the capacity or knowledge to do it either - their safety management is security focused, hiring people to manage things on the go on the ground as they develop. Sometimes, Roskilde offers some small planning if they are participating, however only internally.

Roskilde goes one step further by developing a Venue Manual for each stage/area of the festival. What they are interested in is understanding where people flow in from and mapping risks caused by these flows. They want to have a good understanding of what the flow of people over the whole venue is. People at every festival are different - at Roskilde they are younger, active, move around a lot and see many concerts.

UK is on the forefront of crowd management because they are a fucked up country with many problems, so they have to invest into it.

This summer, Roskilde only provides manpower on ground for Copenhell, no planning. Us making a plan for LiveNation / Copenhell since they don't have people that can do it?

Roskilde makes a **risk assessment for every single concert** - Starting with talking to bookers to understand what kind of a concert it will be (qualitatively speaking). After that, they research the band and do a **Band analysis** (what is the level of popularity/publicum attraction, what sound level they play at, what is the demographic). They try to see if they need to prepare for particular **crowd dynamics** (if the band causes moshpits or storming of the stages to happen etc.) **Crowd types are outlined by Berlongi** [-Crowd types (Berlonghi 1995/EMA 1999) | Download Table (researchgate.net)](https://www.researchgate.net/figure/Crowd-types-Berlonghi-1995-EMA-1999_tbl2_224911893)

They are building a database of bands and concerts to refer to. Word files with filled out features of the band/concert. Can we digitalize/log this data in our product?

Work flow:

Band Analysis > Risk Assesment > Concert Colour (red, green, yellow)

DJ sets are the most unpredictable.

Density mapping is very interesting. Where people leave and at what time, what direction. Setting up more cameras is only a problem when they are CCTV due to cost. Regular cameras is not a problem.

As planners, they have to decide based on their knowledge and intuition what densities are acceptable. There are EU guidelines, but in the end high densities are okay as long as people are having fun at a good concert.

They look at live camera footage to see if there is a crowd collapse, if people are …happy This is included in post evaluation.

Old collaboration with IBM was not good. They were trying to link data from cash registers, spending and concert types/bands to find correlation (Or was this what they wanted from them but didn't get???)

DCM - Cameras cant move to work. They have to be calibrated, and the data outputs you get from it depend on the inputs you give it for raising alarms = too much work for the event organizer, a lot of effort into setting up of the software.

Morten spends time looking at CCTV footage even after festivals.

What would you change if you had the product/analysis - Calculating number of emergency exits, which areas to close off and at what time. Staff allocation - better placement of people on ground in preparation of influx of crowds. Proving their theoretical knowledge with real data.

Roskilde spends approximately 600.000DKK every year on cameras. For places like Royal Arena, this is more of a one-time cost.

They follow the Event Safety Guide and the purple guide [The Purple Guide](https://www.thepurpleguide.co.uk/index.php/the-purple-guide)

[The+Event+Safety+Guide.pdf](Roskilde%20Festival%20 c0150926974744a3a5d39167c77340a2/TheEventSafetyGuide.pdf)

For planning of Roskilde festival, they use GIS map tool (Its also used by the emergency response departments). They can import AutoCAD drawings. Its a map of all services in one place. When they go out and build on site, they pay a land surveyor to map out the whole venue with accurate GPS coordinates. This must be really fucking expensive to do for such a large venue!!!!

Moving onwards - Contact DTU about GDPR and-or non-disclosure agreement for data processing. Create data risk assessment for data processing. Roskilde contacts us back if they have a GDPR data agreement / data transfer agreeement .

What problem is being solved by our product? No clear problem being solved. Feedback on redesign of barriers, Less staffing, and staffing on time where it needs to be.

Business talk from Morten - Do something cheap, place cameras yourselves, **do an analysis and offer a visual, understandable data visualization** - People struggle understanding the theory behind what they preach.

Steen is the name of the head of IT at Roskilde

RF meeting notes – March 10, 2024

A.1.3 May 25, 2024

Meeting with Mads Therkildsen.

Roskilde proposes our involvement at the festival to be composed of two phases . Pre-festival during warmup days they would like us to analyze EOS. After opening of the big festival, Arena stage. We are open and flexible to analyze also other parts of the venue, maybe areas with Food and Beverages.

Mads asks to be sent the project description, and to understand better what our aims with the product are. Ideally in the future, we are coming, setting up cameras, taking them down and just giving a report on the analyzed areas, with minimum work effort needed from the organizers side.

The expectation is to get data/demand curves for the ingress ang egress of people to the individual stages. Secondarily, to get relatively accurate measurements into what the density of crowds is in the analyzed zones that follow Fruins levels of service. Estimate density based in people pr. m2 etc.

It is interesting for Roskilde to get an idea about how a concert starting at Orange stage affects people moving from Arena to other areas.

Our contact people at Roskilde for the beginning will be Niels - working with CCTV systems and Adam from England who is a flow manager. We will setup a way to follow and learn from Adam while on ground working to understand the way they handle things.

We will receive risk assessments and concert schedules to understand key events.

RF meeting notes – May 25, 2024

A.1.4 August 28, 2024

Agreed on data exchange - week between 16th and 20th of September. Follow up meeting regarding progress to be agreed in October.

Soft milestones :

- 14th October - WIP Mockup for ingress/egress
- 25th November - Ingress/egress visualization in user-friendly platform and WIP mockup of crowd flow visualizations
- 20th January - Flow visualization in user-friendly platform integrated

RF is interested in seeing visualizations of direction of flow. Crossflow and density is bad for sales and influences their decisions when planning the placement of bars etc. We will ask for contact to a BI officer at RF to understand more how this data can be used in other wars than only crowd safety planning.

Key questions crowd officers ask: Do we have enough time to move people from one area to another? Low flow means higher densities. Direction of flow is important because they want to know where do people come from.

While they can use time to look at CCTV videos, it is more valuable for them to have a user friendly data based interface that can be understood across the entire organization. They could use the data in a broader sense to place bars , information screens, update their risk assessments, booker assessments of the attraction of the individual concerts.

RF meeting notes – August 28, 2024

A.1.5 September 18, 2024

Agenda:

Showcase the dashboard (cumulative total chart, ingress/egress chart)

Notes:

Nice visualization, only feedback on colour scheme

Morten can use the trends but not the absolute numbers, he wants to know what the error rate is, ho we validate it.

Bookers would LOVE this data

Morten wants a some highlights of the data that he can then use for making changes to the layout and staffing etc.

Overlay concerts from Gaia over this graph too

RF meeting notes – September 18, 2024

A.1.6 February 17, 2025

Levels of service

- Density measurement
- Heads and shoulders -> high density
- Feet, low density -> ~1-2 ppl pr m2

Flow

- Mads knows he needs 2500 guests in before 18
- Mads can count at 16:00, the ingress flow, and capacity of the entrance
- Multiply by an hour. How many come in an hour
- Also 15-min interval
- If Only 1000 people/hr, Need to increase capacity somehow
- 1-minute, 15-minute, 1-hour intervals most interesting

Mads sees a value in having continuous values

He only has initial count, not after changing capacity/flow

For planning purposes at concert:

- Mads needs to check previous flow numbers of venue to know when he needs to up/downscale flow at the entrance
- At Arena, constantly adjusting flow, redirecting crowds
- This data can help them see that 30-min before that need to redirect people
  Or here there is a density that is too high

Resource allocation is a huge part of it. 'Its hard to predict. They use estimates and experience to guess

When to open line up system at scene so that not too many come all at once
Also used to calculate entrance width. Based on experience.

Mads agrees that this tool helps with communicating and knowledge sharing. Internally across departments it helps justify requirements.

Map

- Mads likes the detail-level
- They have had heatmaps before. From app data. Every 15 minutes

```
- Only overview, traditional heatmap. Only colors/blobs. They can see that
there are a lot people based on blob color
- Makes sense around food and beverage. Can see line formation

**Density**

- Mostly discussed in crowds in front of scenes
- Mads prioritises areas they know are high-density, to do their density
calculations.
- 1.5-2 people pr m2 ish for food & beverage areas
- They want to know if there are areas with high density, where not designed
for high density
- Density numbers are still good. Shows real world situation vs. Planned
calculations
- You just 'dont want to see a high density.
- The data is documentation

**Other**

- Mads would like to see movement patterns.
- More specifically: where are people coming from, and where exactly are they
going
- Important to have people hit scene head-on, to prevent imbalance
```

RF meeting notes – February 17, 2025

A.1.7 March 24, 2025

```
**Flow**

- Mads likes the overview, 'its much more intuitive. He feels the numbers are
more useful
- They estimate flow to calculate how many meters are needed for proper
ingress/egress
- This helps document that space was properly allocated, or if it was too
narrow
- Would like to see flow separated for each entrance/camera
- Based on his intuition, the numbers look very realistic

- Would like to know the load of each entrance
- Flow chart should be able to filter based on entrance/camera, but also show
them aggregated

**Density**

- Chosen bins for heatmap are alright, but they use Levels of Service
internally, Mads sends a document

- This tool helps communicated internally in other departments
- Other departments dont understand crowd 'safetys recommendations
```

RF meeting notes – March 24, 2025

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Technical
University of
Denmark

Diplomvej, Buildings 371 & 372, 2nd floor
2800 Kgs. Lyngby
Tlf. 4525 1700

www.entrepreneurship.dtu.dk/