

UNICORN VYSOKÁ ŠKOLA S.R.O.

MASTER'S THESIS

2024

Bc. Filip DITRICH

UNICORN VYSOKÁ ŠKOLA s.r.o.

Softwarový vývoj



MASTER'S THESIS

# Cashless festival data analysis and analytical dashboard development

Author: Bc. Filip Ditrich

Supervisor: Mgr. Václav Alt

Prague 2024

## Statutory Declaration

I hereby declare that I have written my Master's Thesis on the topic of *Cashless festival data analysis and analytical dashboard development* by myself, under the guidance of my thesis supervisor, using only the technical publications and other information sources which are all quoted in the thesis and listed in the bibliography.

I declare that artificial intelligence tools have been used only for support activities and in accordance with the principle of academic ethics.

As the author of this Master's Thesis, I also declare that in association with its writing I have not violated the copyright of any third party or parties and I am fully aware of the consequences of provisions of s. 11 et seq. of Act No. 121/2000 Coll., the Copyright Act.

Furthermore, I hereby declare that the submitted hard copy of this Master's Thesis is identical to the electronic version I have submitted.

In ..... on .....

Bc. Filip Ditrich

## **Acknowledgements**

I would like to express my sincere gratitude to my supervisor, Mgr. Václav Alt, for his guidance, support, and valuable advice throughout the process of writing this thesis. I would also like to thank my family and friends for their encouragement and understanding.



**Cashless festival data analysis  
and analytical dashboard  
development**

Analýza dat bezhotovostního festivalu  
a vývoj analytického dashboardu



## **Abstract**

TODO

*Keywords: TODO*

## **Abstrakt**

TODO

*Klíčová slova: TODO*

# Contents

<b>Introduction</b>	<b>10</b>
Background and Motivation . . . . .	10
Problem Statement . . . . .	12
Objectives of the Work . . . . .	14
Scope of the Study . . . . .	18
<b>1 Data and Methodology</b>	<b>21</b>
1.1 Environment and local setup . . . . .	21
1.1.1 Data Obtaining and Preparation . . . . .	22
1.1.2 Local Database Setup . . . . .	22
1.1.3 Local Database Modifications . . . . .	23
1.2 Data Anonymization . . . . .	24
1.3 Data Sources . . . . .	26
1.4 Conclusion . . . . .	26
<b>2 Data Analysis and Results</b>	<b>27</b>
2.1 Cashflow and Revenue Sources Analysis . . . . .	27
2.2 Performance Indicators Analysis . . . . .	27
2.3 Beverage Consumption Analysis . . . . .	27
2.4 Customer Analysis . . . . .	28
<b>3 Dashboard Implementation</b>	<b>29</b>
<b>4 Conclusion</b>	<b>30</b>
4.1 Summary of Work . . . . .	30
4.2 Reflections . . . . .	30
4.3 Future Directions . . . . .	30
<b>List of Figures</b>	<b>31</b>

<b>List of Tables</b>	<b>32</b>
<b>List of Acronyms</b>	<b>34</b>
<b>List of Appendices</b>	<b>35</b>
Appendix A Source code of the application . . . . .	35



# Introduction

## Background and Motivation

Payments at festivals are a crucial part of the successful event management. The shift from cash to cashless payments has been a significant trend in the last decade that has brought many benefits to both festival organizers and attendees.

However, traditional cashless payment systems utilizing payment terminals are not only expensive to reliably implement at a venue, where the internet connection is often unreliable and requires expensive Base Transceiver Stations (BTS) set-ups or in-place wired/optical internet implemented which can cost up to several million CZK, but also do not provide any insights into the data generated by the transactions. In the best case scenario, the organizers are able to generate a report of processed transactions made by each terminal.

Which, frankly, is not enough to make any actionable decisions based on the data. Moreover, given the event organizers are not data scientists, they often lack the knowledge and tools to analyze the data and extract valuable insights from it.

That is where NFCtron comes in and offers a solution that not only provides a reliable cashless payment system, with credit-based NFC chip bracelets supporting offline mode or card terminal payment solutions, but also provides a comprehensive B2B platform that allows the organizers, the vendors and event the third-party partners to benefit from the data that the system operates with.

The system is a full-scope solution that provides from initial online ticket sales and online credit top-up, through attendee check-in, on-site credit top-up attendee access control, security monitoring, vendor sales and inventory management, most importantly the fast and reliable payment processing with the real-time data analytics and reporting, all the way to the post-event automatic settlement and

reporting.

It simply provides everything an event organizer needs to successfully and efficiently manage their event without the need to worry about any technicalities or staff management. Since NFCtron does not only provide the system as a service, but also provides experienced Event Managers, cashiers, check-in brigadiers and other staff to operate the system and the event itself ensuring that the organizer can focus on the event in terms of communication, marketing, line-ups and other important aspects of the event.

Put, NFCtron offers organizers a peace of mind and a guarantee that their event will be a success.

## **NFCtron Company**

NFCtron is a Czech company that has been operating since around 2019. In its early beginning during a COVID-19 pandemic was on the verge of survival because of the event industry being paralyzed by the government restrictions. However, the company survived and even in these difficult times managed to turn the disadvantage into an advantage by focusing on the core system and product development.

Which years later resulted in a robust and reliable system. It is now used by many event organizers across the Czech Republic and Slovakia and is currently expanding to other countries in Central Europe such as Austria, Poland and Germany. In its primary market – the Czech Republic – NFCtron penetrated the market and is now the leading cashless payment system provider for festivals and other events.

In recent years, the company has also been focusing on expanding to the Payments market focusing both on Card Acquiring and Card Issuing. From the acquiring part, it is now actively developing its own SoftPOS solution that will allow vendors to accept payments via mobile phones. On the other hand, on the Issuing part, it is also working on Card Issuing; that will allow the company to issue its own NFCtron branded payment cards in cooperation with Mastercard.

A big part of the successful market penetration was the company's focus on the business-to-business (B2B) side of the business with event organizers. Giving the organizers amounts of data improving their decision-making and providing them

with insights allowing them to optimize their events. And most importantly, providing them with economic aspects and cashflow optimizations that allows many events and festivals to survive and continue to operate.

## Personal Position and Motivation

I have been with the company from the COVID-19 times. My current position is a Chief Product Officer (CPO), and I am responsible for the product development and the product management of all the products and services that NFCtron offers. This allows me to have a deep insight into the system and most importantly, access to all the data that the system generates. As previously mentioned, the company's success is based on the B2B side of the business and that means services and products provided to the event organizers.

The main product, that organizers have access to is the platform called **NFCtron Hub**. My personal goal and personal motivation to work on this thesis is to discover new ways to improve the platform and provide even more valuable insights to the organizers.

## Problem Statement

Even though the system provides a lot of data, it still has a lot of potential to provide even more valuable insights to the organizers. Currently, the previously mentioned B2B platform, **NFCtron Hub**, provides real time data analytics dashboard presenting the most important KPIs and metrics to the organizers.

These KPIs and metrics summarize:

- **Total sales:** the total amount spent on online ticket sales and on-site payments.
- **Total sales in time:** the total amount above split into time intervals.
- **Total refunds:** the total number of sale reversals or refunds made including refunds from online tickets, refunds of on-site payments and chip credit refunds.

- **Chip balances:** the current balance topped-up on the NFC chip bracelets on-site or pre-topped-up online.
- **Customer orders rating:** customers can rate their orders via a NFC-tron mobile application which provides the organizers with feedback on the vendor's performance and the quality of the products sold.

Moreover, it provides less clear data such as

- **List of vendors:** the list of vendors presents at the event with their sales and rating.
- **List of products:** the list of products sold at the event with their sales and rating.
- **List of points of sale:** the list of selling points at the event with their sales and rating.
- **List of top-up places:** the list of top-up places at the event with the amount of top-ups made.
- **List of customer chips:** the list of unique individual NFC chips issued to the customers with their balance, spending and security status.
- **List of customer ratings:** the list of customer ratings with their feedback from points of sale.

And finally, it also provides unstructured data in the form of data exports in tabular format that can be used for further analysis:

- **Product exports** – list of all products sold with summarized metrics.
- **Points of sale exports** – list of all selling points with summarized metrics.
- **Vendor exports** – list of all vendors with summarized metrics.
- **Deal exports** – list of summarized sales made under a deal<sup>1</sup> Between the organizer and the vendor.
- **Transaction exports** – a heavy export of all transactions made at the event.

---

<sup>1</sup>A deal is an arrangement between the organizer and the vendor that states the terms of the vendor's presence at the event, the products they are allowed to sell, the price of the products, the commission the vendor pays to the organizer and other terms of the deal.

- **Ticket redeems exports** – list of all tickets redeemed at the event.
- **And other exports regarding the online sales** – list of all online ticket sales, top-ups, receipts, customers and other data.

With the above giving some initial picture about the capabilities of the system, the platform (and thus the organizers) still face several problems or challenges that need to be addressed:

1. **Problem 1:** The main KPI metrics may provide some core insights, but the other less or unstructured data is not used to its potential.
2. **Problem 2:** The organizers are not data scientists and need a simple and clear way to understand the data.
3. **Problem 3:** Even with all this data, there is still a lot that can be done to dig deeper and provide more valuable insights.

## Objectives of the Work

With the problems stated above, the main objective of this thesis is to analyze, answer and present results to important questions about the available data and the potential insights that can be extracted from it.

But to achieve this, it was a prerequisite to find a willing event organizer, that would provide the data and would be willing to cooperate on the project. Cooperate in terms of providing valuable insights into what they would like to know more about their event.

## In Cooperation with the Event Organizer

For this purpose, I have chosen a not more undisclosed event organizer, that has been a close and helpful partner of NFCtron for many seasons. Together with the organizer in the first step, we have stated the following requirements to perform the analysis:

- **Requirement 1:** The event and organizer should be kept undisclosed.

- **Requirement 2:** The data should be anonymized to not leak any possible sensitive information about vendors or customers.

The next step was to choose an event from which the data will be used. As it cannot be disclosed any further, we will refer to the event as **The Event** and the organizer as **The Organizer**.

Now the important information about **The Event** for this study is the following:

- **The Event** is a music festival that has been organized for several years now.
- **The Event** takes place in the Czech Republic in the begging of July 2025.
- **The Event** is a 3-day event with multiple stages and multiple vendors.
- **The Event** uses NFCtron system for cashless payments and access control.
- **The Event** had around 7,000 attendees in 2024 and had a roughly 43% increase in 2025 to around 10,000 attendees.

## Data Analysis Objectives

The final step was to define questions or data analysis objectives that should be answered or achieved by the end of the thesis.

Together in the cooperation with **The Organizer** and several internal colleagues in NFCtron, we have defined the following questions for the data analysis:

### Cashflow and Revenue

- **Q1:** *What was the total revenue of the organizer and what does it consist of?*
- **Q2:** *How much and by what means was the balance topped up on the chips?*
- **Q3:** *How much balance remained on all chips after the event and after refunds?*
- **Q4:** *What was the total revenue of the event, how much of it was the revenue of the organizer and how much of external vendors?*

These questions should provide currently unclear insights into the cashflow and revenue sources of the event. Possible answers to these questions could provide the organizer with valuable insights into the economic aspects of the event and could help optimize the cashflow and revenue sources for the next event.

## Performance

- **Q5:** *How many unique customers were processed in total and how many were there at most in one day (and when)?*
- **Q6:** *How many transactions were processed in total and what was the largest “peak” in the volume of transactions processed by the system (and when)?*
- **Q7:** *What were the best-selling points?*
- **Q8:** *What were the best top-up points?*
- **Q9:** *Who were the best vendors?*
- **Q10:** *What were the best products?*

The current platform already provides some performance metrics, but these questions should provide more detailed insights into the performance of the event.

## Beverage Consumption

- **Q11:** *How much was the total consumption of drinks/fluids?*
- **Q12:** *How many returnable cups were issued, and how many were returned or not returned?*
- **Q13:** *What was the most popular drink category?*
- **Q14:** *What was the TOP beer brand, how much was consumed and sold?*
- **Q15:** *What was the TOP brand of other alcoholic beverages, how much was consumed and sold?*
- **Q16:** *What was the TOP brand of non-alcoholic beverages, how much was consumed and sold?*

Currently, a more in-depth product analysis is missing in the platform and the most important part of the product sales analysis at festivals is the beverage consumption. These questions should try to answer and give detailed insights into the beverage consumption, preferences and sales at the event.

## Customers

- **Q17:** *What was the total attendance at the event, how many active customers were at the event each day?*
- **Q18:** *How many customers topped up credit in advance online?*
- **Q19:** *What was the distribution of customers by their type (on-site, online, staff, guest, VIP)?*
- **Q20:** *How many customers used the mobile application?*
- **Q21:** *What is the distribution of bank cards used to refund credit?*
- **Q22:** *What is the distribution of card schemes used to top up credit on-site and online?*
- **Q23:** *What was the course of the event in terms of new visitors and when were the largest “peaks”?*
- **Q24:** *What is the average time of a visitor from arrival in the first transaction?*
- **Q25:** *What was the course of the event in terms of topping up credit on-site and when were the largest “peaks”?*
- **Q26:** *How many customers topped up credit more than once and how much only once?*
- **Q27:** *What was the difference between one-day, two-day and three-day visitors in terms of spending, topping up and refunded credit?*
- **Q28:** *What was the difference between spending and topping up between different types of visitors (primarily online vs. on-site)?*
- **Q29:** *What were the preferences for the type of drink throughout the day?*
- **Q30:** *What were the most common combinations of products?*



The customer analysis is the most important part of the data analysis. It is crucial for the festival organizers to know their customer base and their behavior to optimize the event and make it more attractive for the customers.

Currently, no customer analysis, other than the customer ratings and list of customer chips, is available in the platform.

Answering these questions will possibly lead to the most valuable insights about the event's customer base and their behavior that no other platform or system currently provides.

Making these questions crucial and most valuable for the organizer and for the platform itself.

## Technical Objectives

Answering the above questions will require a technical solution that will be able to process the data and provide the answers.

The scope of this study is not to implement a new system or a new platform and not even to implement any new changes to the existing NFCtron Hub platform. It is to find answers to the above questions and present them in a clear and understandable way in the form of a simple internal dashboard.

The technical goals of this study are:

- Prepare, process and analyze the data from **The Event**.
- Find answers to the above questions.
- Implement a simple internal dashboard that will present the answers to the questions.

## Scope of the Study

To ensure the feasibility and focus of the study, certain boundaries have been defined in terms of what is included and excluded from the scope of the study.

## Included in the Scope

- The study will focus on transactional, customer, and operational data from a specific event, referred to as **The Event**.
- Key areas of analysis include cashflow, revenue sources, performance indicators, beverage consumption, and customer segmentation and behavior.
- The data analyzed includes pre-event (e.g., online top-ups), during-event (e.g., chip transactions, sales), and post-event data (e.g., credit refunds).
- A prototype dashboard will be developed using Python’s Dash and Plotly libraries to present the key insights.
- The dashboard is intended for internal use and post-event analysis by the event organizer.

## Excluded from the Scope

- **Real-Time Monitoring:** While the dashboard may be designed with real-time data potential, this study will focus solely on post-event analysis.
- **Multiple Event Comparisons:** This study is limited to the analysis of a single event (**The Event**) and does not involve comparative studies.
- **Data Collection:** The study does not involve the collection of new data and relies on data provided by **The Organizer** and the NFCtron system.
- **Implementation in NFCtron Hub:** The thesis focuses on analyzing data and developing a standalone prototype dashboard, not on direct integration into the NFCtron Hub platform.

## Limitations

- **Anonymized Data:** To protect privacy, all customer and vendor data has been anonymized, which may limit certain the analysis in some ways.
- **Single Event Focus:** Insights and recommendations are based solely on data from **The Event**, which may limit broader generalizations.

- **Time Constraints:** Given the timeline of the thesis, certain advanced features (e.g., predictive analytics) and technical implementations have been deprioritized but kept in mind for future work.

# 1 Data and Methodology

This chapter address the process and challenges of local environment setup, obtaining, preparing and anonymizing the data. Most importantly, this chapter describes and explains the data that was used for this research. It also briefly describes the tools, technologies and methods employed to answer the research questions.

## 1.1 Environment and local setup

**TODO:** PostgreSQL, Python, DataSpell, ...

To start off, we needed to set up some kind of environment where we would later work with the data. The data we would be working with was stored in a PostgreSQL database.

Having direct access to the production database to perform the analysis was not a secure and ethical way to go. Exporting only the necessary and raw data from the production database was an initial thought, but we initially did not know what data we would need, and by exporting we would lose all the relations between the tables.

Therefore, we decided to set up a local database with the same structure as the production database where we can query and analyze the data safely. The next step was to import the data from the production database to the local database. Importing or simple cloning the full database was also not an option, because only a small fraction of its subset was required.

So a deep internal analysis of the tables that were relevant to our study was performed. This resulted in a list of total 21 tables that held the necessary data

for the study and were necessary to be imported.

### 1.1.1 Data Obtaining and Preparation

Almost every table was easily queried for the event and exported from the production database to a local CSV file. But some tables (for example and not surprisingly, the *transactions* table with over 140k rows) were too large to be exported in one piece, so we had to split them into smaller parts. Later, these parts were joined together to a single CSV file using a simple Python script.

Since no direct access to the production database was used for the export but rather a database management tool, the export was not as fast as it could be and took a significant amount of time. Moreover, the exported data, most importantly the timestamps, were in a different format than we needed. And also all numeric values were exported as a formatted strings with a comma as a decimal separator. So a data preprocessing Python script was written to convert such invalid columns to the correct format.

### 1.1.2 Local Database Setup

Then a step to set up the local PostgreSQL database was needed. Due to the nature of this study, we wanted to keep the setup as simple as possible, so we used the default PostgreSQL installation without using any special environment using Docker or similar. However, during this process I made a mistake and forgot that a PostgreSQL with PostGIS extension was needed. So it required to re-set up the database with the PostGIS extension.

The next step was to import the data from the CSV files to the local database. For further database handling, analysis and visualization, we used DataSpell, a Python IDE with a built-in database explorer and data visualization tools. DataSpell was then used for the local database import, which prior to it required some necessary database relations and constraints modifications, since the data was exported without them and was not relevant for the study.

This whole process resulted in approximately 387k rows of data in the local database that were ready to be queried and analyzed.

### **1.1.3 Local Database Modifications**

Before any analysis was performed, some modifications to the local database were needed due to some known limitations and missing data.

#### **Beverage Volumes**

The first necessary limitation that the Beverage Consumption Analysis section heavily relied on was the missing information about beverage products volume in milliliters. This information was crucial for the analysis, so a new column was added to the relevant product information tables. However, the next step was to back-fill this information which was not easily automated.

The First approach was to write a Python script that would try to find the volume information from the product name. This worked for some products, but not for all since the naming convention was not consistent.

After several attempts to automate this process, it was decided to manually fill in the missing information since only 425 products were present in the database. Only 159 of them were of beverage type and thus eligible for the volume information.

#### **Depositable Products**

Since one of the research questions was to analyze the depositable cups and this information was not easily available in the database, a new column was added to the product information tables.

This was a simple binary column that indicated whether the product was depositable or not. Back-filling this information was also pretty straightforward since only one product was a depositable cup.

#### **Venue Map Visualization**

One of the initial ideas was to visualize the venue map with the locations of the selling places, top-up service points, stages and other important places.

This would be invaluable, the database was partially ready for this, but the data would be significantly time-consuming to back-fill and the later analysis and visualization would require more time.

Since these facts and the fact that this process of preparing the data took place before completing the list of data analysis questions, this idea was later abandoned.

## **Event Program**

To present some time-related data and its correlation with the event program, it would require to have the event program in the database.

Again, the database was ready for this, but no event program was set up, since it was unnecessary for the event. Therefore, this required getting the event program from the festival website and manually insert it into the database.

This was manually a very time-consuming process, but it was necessary for the analysis. For some simplification of the process, an AI tool was used to extract the data from the program schedule screenshots and instructed to prepare an SQL script that would insert the data into the database.

This seemed like a good idea, but the AI tool was initially hallucinating and made up some incorrect data. But after several iterations, it successfully extracted the data and prepared the SQL script which was used and the event program was successfully inserted into the database.

In the end, I doubt that this process was faster than manual data entry, but it was a good exercise and a good example of how AI can be used to automate some processes.

## **1.2 Data Anonymization**

The Data Anonymization process was necessary due to requirements initially set by the data provider and later by the ethical considerations. This step was performed for the already imported data in the local database. It required identifying the sensitive data and replacing them with anonymized values.

In this case, the most sensitive data were:

- **Vendor names:** Since it included the legal names of the vendors, it was necessary to anonymize them.
- **Selling places:** Some selling places were named after the vendors, so it was necessary to anonymize them as well.
- **Customer information:** Some tables included customer information like names, emails, phone numbers, etc.

The process could have been done various ways, but the fact that this study will not be exposing internal database structure, it was decided to perform the anonymization directly in the database.

However, if one-way anonymization were to be performed, it would permanently overwrite the original data, losing the possibility to switch from anonymized to original data. Therefore, a two-way anonymization process was chosen and performed.

This was particularly useful during the analysis phase, where the results would contain the original data for better understanding, fact-checking and for the internal presentation and consultations with the organizer.

It was done on the database level, where two new internal tables were introduced – *public.anonymization\_config* and *public.original\_values*.

Where the *public.anonymization\_config* table held the configuration about which schema, table and column should be anonymized and how. For the usage, a simple SQL function was created to define the anonymization configuration in a simple JSON format, that looked like in the **Anonymization configuration example..**

```
1  SELECT configure_anonymization('[
2    { "table": "schema.seller", "columns": ["legal_name", "name_int",
      ↳ "name_pub"] },
3    ...
4    { "table": "schema.user_account", "columns": ["email", "last_name",
      ↳ "first_name", "phone"] },
5    ]'::JSONB
6  );
```

**Source code 1:** Anonymization configuration example.



The *public.original\_values* table was used to store the original values of the anonymized columns. Again, using a simple SQL function *anonymize\_database()*, it would store the original values and anonymize the configured table columns.

One particular challenge was anonymizing the values smartly. It could have been easily done by replacing the values with random strings, hashes or encrypted values. But working with data, where a vendor is named *fa65165b923e9cc* is not very convenient.

Therefore, a simple SQL function was written to anonymize the value depending on the configuration. This allowed to configure the anonymization to:

- replace vendor names with values like *Vendor 1*,
- customer emails with *03b09592-d0eb-43a3-9941-30d38ade6bce@gmail.com* keeping the original email domain,
- selling places with values like *Place 1* where the original name contained sensitive information, but keep original values for places like *BAR L2*, etc.

In the end, it resulted in a database with anonymized values per stated configuration, that could have been used for the analysis and results presentation safely. However, the original values were still present in the database, and the database could have been anytime easily restored to the original state if needed and vice versa.

## 1.3 Data Sources

**TODO:** Describe the data sources, with what data we are working with.

## 1.4 Conclusion

**TODO:** Conclude.

## 2 Data Analysis and Results

This chapter presents the data analysis and results of the research. The data analysis is divided into four sections: **Cashflow and Revenue Sources Analysis**, **Performance Indicators Analysis**, **Beverage Consumption Analysis** and **Customer Analysis**. Each section focuses on a different aspect of the data analysis trying to answer the research questions as previously defined in the **Introduction** chapter.

### 2.1 Cashflow and Revenue Sources Analysis

**TODO:** Haha.

### 2.2 Performance Indicators Analysis

**TODO:** Haha.

### 2.3 Beverage Consumption Analysis

**TODO:** Haha.

## 2.4 Customer Analysis

**TODO:** Haha.

## 3 Dashboard Implementation

**TODO:** This chapter describes the implementation of the analytical dashboard using Dash and Plotly.

## 4 Conclusion

### 4.1 Summary of Work

**TODO:** Write a summary of the work done in this thesis, recap the main objectives and how they were achieved. Note the key contributions of the research and practical side.

### 4.2 Reflections

**TODO:** Write reflections on the project, what was learned, how the project could change festival operations, etc.

### 4.3 Future Directions

**TODO:** Write about the future directions of the project, what could be done next, what could be improved, what could be added, etc.

# List of Figures

*Note: Unless stated otherwise, all figures are the author's own work.*

# List of Tables

# List of Source Codes

1	Anonymization configuration example. . . . .	25
---	--	----



## List of Acronyms

# List of Appendices

Appendix A   Source code of the application