Narcolands

A data driven approach to event based drugs popularity in the Netherlands

Ana Chaloska, Anisha Sloeserwij, Danny de Vries, Faye Tervoort, Jesse Kommandeur
University of Amsterdam
Amsterdam, The Netherlands

ABSTRACT

KEYWORDS

drugs, google trends, court cases, Netherlands, visualization

1 INTRODUCTION

Every country is affected by crime, and so is the Netherlands. According to the Dutch center for crime prevention and safety statistics (CCV), over 90% of all organized crime is related to illegal drugs (CCV, 2022). The Netherlands owes this percentage to its prominent role in international drug trade (McDermott et al., 2021). Since problems surrounding drugs have increased considerably in recent years, authorities have prioritized and focused on identifying and tackling drugs related behavior (Eski and Buijt, 2016; Ferwerda et al., 2016).

In order to battle drugs related organized crime better, it is important for authorities, such as the police, to understand the modus operandi of criminal organizations through both academic and practical research. One of the organizations that focuses on the discovery of new modus operandi is the Police Academy (PA). In cooperation with the University of Amsterdam, the PA wants to focus on data driven research to identify drug related patterns in the Netherlands.

Previous research within the narcotics space focused on the supply-side of illegal drug trade (Paoli et al., 2013; Magliocca et al., 2019). This domain focuses on understanding the production, distribution, and trafficking of illegal drugs. It includes research on the modus operandi used by criminal organizations to produce and transport drugs, as well as the routes and networks they use to distribute them. This type of research is important for law enforcement agencies as it helps them identify and disrupt the operations of drug traffickers (McDermott et al., 2021).

Another area of research that is also present in prior literature is the demand-side of illegal drugs. This domain focuses on understanding the consumption and impact of illegal drugs on individuals, politics, and society (Flores-Macías & Zarkin, 2019; Riley, 2017). It includes research on the psychological and social factors that contribute to drug addiction, and the health and social consequences of drug use (Gonzalez, 2015; Sesnie et al., 2017).

These research domains are important for public health agencies and policymakers as it helps them understand the extent of the drug problem and develop effective strategies for addressing it. However, most of these studies are less relevant for authorities such as the PA since they focus less on quantitative data driven methods to identify patterns in society. Therefore, this study focuses on a data driven pattern identification of drug usage and events, such as festivals and public holidays. In order to investigate this relationship, this study aims to answer the research question: *"To what extent can we identify event-based drug popularity on online data resources?"*.

To answer this research question, the following sub questions were formulated:

- Which Dutch events are indicative for drug popularity?
- What drugs are used during events in the Netherlands?
- Which online data resources are relevant to identify drug popularity at events?
- How could we identify drug popularity at events?
- What are the differences between the specific events and drugs?
- How could we visualize event based drug popularity on online data resources?

The sub questions are relevant to the main research question as they provide a more detailed and specific understanding of the topic. The subquestion "Which Dutch events are indicative for drug popularity?" helps to narrow the scope of the research since it provides a starting point for identifying which events may be associated with drug use. Second, the subquestion "What drugs are used during events in the Netherlands?" is important for identifying which specific drugs are of interest for the research. "Which online data resources are relevant to identify drug popularity?" and "How could we identify drug popularity at events?" are crucial for understanding the methods and data sources of this research. "What are the differences between the specific events and drugs?" helps to identify any patterns or trends in drug popularity that may vary depending on the specific event or drug being considered. The last subquestion “How could we visualize event based drug popularity on online data resources?” is relevant since it allows us to transform our research data into a user-friendly dashboard to create insights for the PA.

In order to answer the research questions, we first present the theoretical state of affairs, followed by the methodological set-up, after which the results are presented. In the final sections, the most important findings are concluded and limitations are discussed, followed by recommendations for future work.

2 RELATED WORK

The collection and utilization of large amounts of data is a popular resource nowadays and is used in a variety of disciplines. [bron] The use of these large amounts of data, also known as big data, combined with the use of data analysis is having a major impact on the social sciences and humanities in general. [bron] Chan and Moses’ research shows that it has a particularly large impact for the specific field of criminology. [bron]

2.1 Predictive policing

Predictive policing is a term used for predicting certain behaviors or trends based on analyzing various data for law enforcement. [bron] Authorities use predictive policing to predict crime in order

to prevent the criminal activity instead of reacting to the crime that already occurred. For this reason, the domain and effects of predictive policing have been examined for a long time.

However, not all research agrees that predictive policing is a valid method. According to The New York Times' debate, predictive policing is a very effective way when it comes to predicting criminal behavior, but still contains many improvements in terms of ethicality. [bron] In contrast, Hardyns' research says that the analysis of such predictions has indeed shown its worthiness for different predictive systems in different areas, but exactly because it is a new development in the field of criminology, little is yet known whether it is an effective way for law enforcement. [bron]

Several studies have been conducted that have investigated whether Google trends can be a possible resource for predicting particular trends. Previous research has shown that Google Trends can be used as a predictor in different fields. Such as the healthcare industry, where Google Trends can help support the prediction of the outbreak of seasonal influenza and COVID-19 [bron], [bron]. Another study by Kassraie et al. used Google trends data in combination with Twitter data to predict the popularity vote of the 2016 presidential election in the US. They concluded that the combination of these social media platforms could be a mirror for the public opinion on political events [bron].

In the study by Perdue et al. it appears that Google trends can be a possible predictor for drug abuse trends. [bron] This is supported by the research of Gamma et al. which investigated whether there is a comparison between time trends of Google search interests and offenses committed in relation to the drug called Methamphetamine. From this study, it was found that law enforcement could indeed use the Google search feature as a possible predictor of Methamphetamine-related crimes. [bron]

3 METHODS

3.1 Explain why we use these methods and why they are useful for our case, mention what Ana Barros told us

3.2 Google trend research

3.3 Court cases from rechtspraak.nl

3.4 News articles

3.5 Requirements

These user requirements are derived from the case description provided by the client as well as feedback from the stakeholder during the initial ideation and prototyping phase of the project and where further narrowed down during the project design workshops using the MoSCoW priorization method. The term 'user' in the requirements refers to two specific types of similar target audiences that will make use of our prototype, police agents who want to explore the dataset to gain insights and data analysts who want to filter and compare our datasets.

(1) (M) The user must be able to use the prototype on a personal computer and interface with a screen
(2) (M) The user must be able to filter the datasets to compare different years of data
(3) (M) The user must be able to filter the datasets to compare different types of drugs
(4) (M) The user must be able to overlay multiple datasets and trend lines on top of eachother
(5) (M) The system uses open-source software and not locked-in corporate data tools
(6) (M) The system has a user-friendly visual design and interaction design
(7) (S) The user should create an account to store specific and personalized filters
(8) (S) The user should be able to download the raw datasets in specific file formats
(9) (S) The user should be able to navigate between different overviews showing corresponding data
(10) (C) The user could upload their own dataset and sources of specfic drugs and news sources
(11) (C) The system uses real-time up-to-date API data

4 RESULTS

4.0.1 Prototype (dashboard).

Based on these requirements and the datasets a custom web-based visualization was created to allow the analyst to visualize explore, compare and filter the datasets. When filtering the charts update in real-time with keyframed animations. Screenshots and a live version of the prototype included in appendix 2a. For this prototype demo the user is logged-in with a default user account which is shown in the sidebar navigation. From there the user has four different overview pages to navigate to.

Events overview page

The first is an events overview page which shows a line chart with on the x-axis the week numbers per year and a relative score from 0 to 100 on the y-axis for our three datasets, the Google Trends, NOS News data and Twitter tweets. On the left bottom is a legend with a list of drug-related events. The user can click on an event which will highlight the week in which the event occured in the line chart to more clearly visualize peaks in the data which indicate event-based drug popularity. A small description of the event, most popular drug and event dates also is shown. The bottom-right shows the Google Trends data in a relative score over all 5 years of our dataset. This shows more clearly shows the lack of peaks of events in covid pandemic years.

Region overview page

The second page is a chloropleth map page showing the regions of the Netherlands. Each region is color coded using a linear color scale based (blue interpolation) on the relative score of Google searches for that region. On this page the user also has the option to focus on years and drugs but also on specific event dates. This for example shows that in Flevoland when the Lowlands event is happening there are alot of searches in that region for the drugs XTC.

Related queries page

The third page is the related queries page which shows a line chart for our three chosen drugs as filter options. The y-axis is a relative score from 0 to 100 and the x-axis are the week numbers for our specific year. The user can filter between different years. On the

bottom are three polar charts which show the related drugs people are searching for with a relative score and a subset of the drugs in the legend. For example, people who search for cocaine also search for crack very often. It also shows that before 2019 GHB was not searched for by many people. It increased in popularity after that year.

Settings page

A settings page is also included. This page shows the different datasets used in the dashboard and a download button which allows the user to download a specific dataset either as .csv or .json to store on their computer.

4.0.2 Web application. The prototype is a web-based application using web standards and open-source software and libraries. It being web-based allow the dashboard to run operating system independent. Only a webbrowser installed on the users device is required. The application will mainly be used in a desktop environment by the analyst so the dashboard is not fully responsive and not mobile optimized. The web application is created with the open-source front-end framework Svelte [1] and UI framework SvelteKit which allows the application to be build in components, each chart is rendered seperately making it more efficient to add functionality (add datasets, render different chart types) in the future but also makes the dashboard performant when more data and charts are added since Svelte already pre-renders the page offloading work from the browser. Working in components and with a frameworks such as Svelte allows future web developers to get up and running fast and add more functionality in a progressively enhanced manner. Svelte can be downloaded as a module (package) from NPM [2] and uses the JavaScript back-end run-time Node.js [3].

For the charts the JavaScript charting library Chart.js [4] is integrated into the components which allows charts to be rendered in HTML5 Canvas without much configuration. With Chart.js you can add a specific dataset and the scales of the axis will automatically change accordingly to the scale defined by the data. The filter options and updating of the charts is more custom, it uses JavaScript utility functions to allow the data of the to be preprocessed and have only the data changed not the scales of the whole charts. For the map page an additional Chart.js Geo Plugin is used to render the Chloropleth map. It uses a TopoJSON file to render the regions of the Netherlands and filters the properties within that .json file to render a score for each of the region.

The source code for the dashboard is published open-source on GitHub using the MIT license. A live version of the dashboard is contionously deployed on hosting platform Netlify [5]. Corresponding links can be found in appendix 2b.

5 CONCLUSION

5.1 Give an answer to the research question

6 FUTURE WORK

6.0.1 Expand data sources. A subset of drugs and years was used for our research. To allow for more exploration and filterability the dashboard could be expanded with more datasets. Especially more relevant drugs (upcoming drugs, non-legal substances, NSPs) could be added as filter options. We also currently only use NOS news data but other more popular dutch news sources (e.g. RTL nieuws, Nu.nl) could be scraped to give a more accurate representation of the popularity of drugs mentioned in news articles. This will both improve the quality of the dataset by more accurately calculating a relative score as well as quantitify by gathering multiple sources and aggregating.

6.0.2 Real-time API data. Currently the dasboard relies on exported data that is then loaded into the web visualization. As a futher enhancement for the prototype and to make it more dynamic is to have the dataset exposed as an API which the dashboard can then fetch getting up-to-date real-time data. In this beta version datasets need to be added manually to the GitHub repository.

6.0.3 Usability testing. Further usability testing needs to be done to validate the User Experience (UX) and User Interface (UI) of the prototype to more accurately represent the needs of the police analyst. Basic user testing was done on a small group of people and feedback from the client was incorporated. But still, a lot of assumptions about the user have been made. User testing the live version of the prototype to a larger user base will uncover hidden problems. The feedback from the users would further validate the workings of the prototype.

In general further research is required to conclude that online data-gathering tools fare good indicators for predicting the popularity of specfic drugs.

7 DISCUSSION

8 APPENDIX

[1] https://svelte.dev
[2] https://www.npmjs.com
[3] https://nodejs.org/
[4] https://www.chartjs.org
[5] https://www.netlify.com