DRUG TRENDS

DOI: 10.26190/r4dt-aj11

Methods for "Trends in Google searches and social media discussions about cocaine, July 2021-June 2022: A pilot study"

Authors: Qingyuan Linghu, Amy Peacock, Raimondo Bruno, Rachel Sutherland, Monica J. Barratt and Nicola Man

National Drug and Alcohol Research Centre

University of New South Wales Sydney

1. General note

The work in this bulletin is conducted as part of the Drugs and New Technologies (DNeT) project. The DNeT project has been running since 2012 and forms part of Drug Trends, an illicit drug monitoring system in Australia. The DNeT project has been examining cryptomarket drug listings since 1st February 2014 (Mathur et al., 2020; see here for most recent bulletin).

The current bulletin expands the online monitoring of drug-related data to the 'surface web', focussing on the following sources:

1) Google search using the Google Extended Trends API for Health (GETH)
2) Tweets using Twitter academic API version 2
3) Reddit submissions and comments using the Reddit API and Reddit Pushshift database

This bulletin was undertaken to establish the feasibility of using social media data to inform illicit drug surveillance. For this reason, the bulletin is focused only on cocaine. Please see the bulletin for discussion as to future possible work.

In this Methods document, we outline the background to this program of work and the methods underpinning data presented. There are various approaches to collecting, collating, categorising and analysing online data sources in relation to illicit drugs, and inherent challenges in these processes. There are also limitations and constraints on the appropriate interpretation of these data (see below for further detail). For this reason, we have attempted to be as transparent as possible about our procedures.

Our monitoring and analysis of online data is an ongoing process, with ongoing refinements to the monitoring and analysis process. We welcome feedback and

Funded by the Australian Government Department of Health and Aged Care under the Drug and Alcohol Program

DOI: 10.26190/r4dt-aj11

suggestions so that we can continue to improve utility of these data and our reporting on them (drugtrends@unsw.edu.au).

2. Overview of approach

This methods document is for our bulletin on '*Trends in Google searches and social media mentions of cocaine, July 2021-June 2022: A pilot study*' which focuses on the user[1] activity or posts on the Google search engine, Twitter and Reddit in the period from 1st July 2021 to 30th June 2022. Data presented in the bulletin are obtained from retrospective scraping, collation, and analysis of data on user activity in the stated period from the three data sources.

This project has ethical approval from the University of New South Wales Human Research Ethics Committee (HC180004).

3. Method of data collection

3.1. Description of each data source

Google Trends is a website by Google that analyses the popularity of top search queries in Google Search across various regions and languages. The website uses graphs to compare the search volume of different queries over time. It is reported that:

"The numbers returned are the probability of a short search-session (few consecutive searches), to satisfy the corresponding term restriction, given it was done in the restricted geography (if such exist) and during the time represented by that data point."

It is important to note that the value is the probability of a short search session (Google does not define what they consider to be "short"), and not the absolute volume of searches. The Google Trends API (https://trends.google.com) scales the data so that the maximum value in the series is set to 100, and all other values as expressed relative to that. Furthermore, researchers can apply (http://bit.ly/2KyqDYW) to Google for access to the old Google Flu Trends (GFT) API, now named Google Extended Trends API for Health (GETH). Google Flu Trends differ from Google Trends data in that they are not rounded and scaled to 100, which means the data available from GETH represents the raw search probability (multiplied by 10,000,000) for the specified search term. Note that we report on the Google Trends results using search probabilities from the Google Extended Trends API for Health.

[1] A user in the context of online activity refers to someone who uses Google, Reddit or Twitter in this bulletin. When we refer to someone who uses drugs, we use a term such as "people who use drugs".

DOI: 10.26190/r4dt-aj11

Reddit is an American-based social news aggregation, content rating, and discussion website. The discussions on Reddit are organized by subject into user-created boards called "subreddits", e.g., “r/cocaine” is the main community discussing about the drug cocaine. In each subreddit, registered users (commonly referred to as "redditors") post contents to the site such as links, texts, images, and videos etc. In our bulletin, the Reddit posts which users submit to start a posting thread are called “submissions”. Under each submission (or posting thread), other users may follow up with their additional posts, which are named “comments” in our bulletin. We extract and focus on the textual content for both “submissions” and “comments”.

Twitter is a microblogging and social networking service owned by the American company Twitter, Inc., on which users post and interact with messages known as "tweets". Registered users can post, like, and retweet tweets. Tweets were originally restricted to 140 characters, but the limit was doubled to 280 in November 2017. Audio and video tweets remain limited to 140 seconds for most accounts. In our bulletin, we extract and focus on the textual content of each of the tweets.

3.2. Cocaine search term methodology

For monitoring cocaine on our online data sources (i.e., Google Trends, Reddit and Twitter) search terms are the interface to identify the specific content about cocaine. Instead of only using the accurate substance name *‘cocaine’*, many ambiguous variant names of cocaine are also used by people in online activities. Thus, we study the process of collecting, cleaning and effectively using the search terms of cocaine, so as to have more comprehensive and precise content regarding cocaine from the sources.

3.2.1. Sourcing potential search terms

All the sources where we collect different search terms of cocaine are listed as follows:

1. The DNeT cryptomarket drug dictionaries [Mathur et al., 2020]
2. Drug profile from the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA): https://www.emcdda.europa.eu/publications/drug-profiles/cocaine
3. Psychoactive substance index on the Psychonaut Wiki: https://psychonautwiki.org/wiki/Cocaine
4. Drug Facts from the Alcohol and Drug Foundation in Australia: https://adf.org.au/drug-facts/cocaine
5. Wikidata: https://www.wikidata.org/wiki/Q41576

DOI: 10.26190/r4dt-aj11

Note that, all our methodologies for search terms and subsequent analysis (as detailed later in section 4) are case insensitive for all our online media sources, e.g., *'Cocaine'* equals to *'cocaine'* in all circumstances. Thus, all the terms are only in lower case by default.

After collecting all the search terms of cocaine from the above sources, we feed the union set of search terms with deduplication into the following exploratory procedure.

3.2.2. Exploratory analysis on potential search terms

Our search terms investigation procedure has the following four steps:

1) Manual screening by Google search engine in private browsing mode[2]
2) Google Trends search probability comparison
3) Submissions and comments counting in the subreddit "r/cocaine"
4) Tweets counting on the whole Twitter platform

In step 1), our domain experts searched each term by the Google search engine, then manually browsed the results, based on which all the search terms can be categorized into the three categories as follows.

Category①: These search terms are considered predominantly about cocaine and yield strong specificity, so that they can be used in a reasonably unambiguous manner by itself. These terms include *'cocaine', 'cocain', 'cocaina', 'coicaine', 'cokaine', 'concaine', 'coca leaf', 'coca leaves', 'kokaiini', 'kokaiinin', 'kokain', 'l-cocaine', 'methyl benzoylecgonine', 'neurocaine', 'beta-cocain', and 'benzoylmethylecgonine'*.

Category②: These search terms clearly have other meanings which are used on Internet, however, they are often used to refer to the drug. These terms include *'fishscale', 'fish scale', 'charlie', 'coke', 'coca', 'cola', 'nose candy', 'toot', 'white dust', and 'stardust'*. See below for discussion of treatment of these search terms.

Category③: These search terms are excluded because of different reasons: i) most of the Google search results of the term were not related to cocaine (e.g., *'bolivian', 'peruvian a++', 'peruvian axxx', 'white girl', 'snow', 'ski', 'blow', 'white', 'girl', 'C', 'crack', 'white lady', 'cocainum', 'methyl ester', '(-)-cocaine', and '(â˜')-cocaine'*); or ii) the search term is too scientific to be used by common Internet citizens. Such terms include:

[2] The Google search engine remembers or predicts user preferences from the user's activity history. As such, we use the private browsing mode to return results unbiased by our personal history of search and browsing activity.

DOI: 10.26190/r4dt-aj11

- '[1R-(exo,exo)]-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylic acid'
- '2-methyl-3beta-hydroxy-1alphaH,5alphaH-tropane-2beta-carboxylate benzoate (ester)'
- 'methyl [1R-(exo,exo)]-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate'
- '[1R-(Exo,exo)]-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylic acid'

Table 1. Exploratory analysis on popularity of potential search terms in category① using data from June 2022

Terms	**Google Trends** (GETH search probability X 10^7)	**Reddit** (count)[a]	**Twitter** (count)[b]
'cocaine'	258.01	1850	346007
'cocain'	30.56	13	3239
'cocaina'	94.37	5	916
'coicaine'	0.12	0	9
'cokaine'	0.91	1	84
'concaine'	0.38	0	25
'coca leaf'	0.97	3	458
'coca leaves'	2.72	13	445
'kokaiini'	0.77	0	3
'kokaiinin'	0.43	0	0
'kokain'	34.70	2	233
'l-cocaine'	0	1	0
'methyl benzoylecgonine'	0	0	0
'neurocaine'	0.23	0	0
'benzoylmethylecgonine'	0.28	0	19
'beta-cocain'	0	0	0

Note: [a] We used posts (i.e. submissions and comments) on the subreddit "r/cocaine" to obtain the count of Reddit posts.
[a] We used all worldwide tweets (i.e. including retweets) to obtain the count of tweets on Twitter.

In step 2), for each category① term, the column 'Google Trends' in Table 1 presents the average of daily Google search probability in the month of June 2022.

In step 3), we counted the sum number of submissions and comments which contained each category① term within the subreddit "r/cocaine", in the month of June 2022. This is presented in the column 'Reddit' of Table 1.

In step 4), for each category① term, we counted the number of all tweets containing the search term, in the month of June 2022, which is presented in the column 'Twitter' of Table 1.

DOI: 10.26190/r4dt-aj11

By investigating the three sources regarding our last reported month June 2022, we had the following observation: the numbers of *'cocaina'* and *'kokain'* in Google Trends and Twitter were oddly large, but those numbers in "r/cocaine" of Reddit were not particularly large. Considering that "r/cocaine" is a cocaine-specific discussion context, we further investigated the two terms *'cocaina'* and *'kokain'* on Google search engine and Twitter platform. The results showed that, *'cocaina'* is also a classic song of the Kosovo-Albanian singer Mozzik. It is also the Spanish, Portuguese and Italian word for cocaine; hence its use in Google searches appears to be primarily in the regions of the world where these languages are used, e.g. South America (see Google Trends website). *'Kokain'* is the German word for cocaine which is mainly used in Europe based on Google Trends website data. Thus, *'cocaina'* and *'kokain'* were moved to Category②.

After all the above exploratory analyses, our finalized search terms of cocaine can be named by the following two classes:

Accurate terms: *'cocaine', 'cocain', 'coicaine', 'cokaine', 'concaine', 'coca leaf', 'coca leaves', 'kokaiini', 'kokaiinin', 'l-cocaine', 'methyl benzoylecgonine', 'neurocaine' 'benzoylmethylecgonine', 'beta-cocain'*.

Ambiguous terms: *'kokain', 'cocaina', 'fishscale', 'fish scale', 'charlie', 'coke', 'coca', 'cola', 'nose candy', 'toot', 'white dust', 'stardust'*.

3.3. Definition of search criteria and scope of data

For all of our three online sources, we report on the data from 1st July 2021 to 30th June 2022, with each month being the unit for data aggregation as default. Specifically, each month has a single data point from each of Google Trends, Twitter and Reddit. The search terms and geography used by different online sources are introduced respectively as follows.

3.3.1. Google Trends

Geographical scope: Three geographic levels are provided by the GETH, which are Worldwide, Country, and Region. We used the worldwide level to report the global trends in Google searches for cocaine, and 'Australia' at the country level to report the Australia-specific trends in searches for cocaine.

Main cocaine search terms: Since Google Trends is based on the Google search engine which may not be used in cocaine-specific context, we only used the accurate terms of cocaine for Google Trends results. The GETH supports the Freebase Identifiers. By searching the Freebase Identifier of a topic, e.g., cocaine, the GETH will retrieve all

DOI: 10.26190/r4dt-aj11

searches which Google's AI algorithms have linked to that topic on its Knowledge Graph. Knowledge Graph is a technology that was released in 2012 and is based on the Freebase technology developed by Metaweb Technologies, which Google purchased in 2010. The Freebase Identifier of a topic can be found by searching the topic on Wikidata (https://www.wikidata.org/). The Freebase identifier of cocaine (/m/0256b) was added to the search terms when extracting data on Google searches. The main cocaine search string used for Google Trends data collection is: *'cocaine + cocain + coicaine + cokaine + concaine + "coca leaf" + "coca leaves" + kokaiini + kokaiinin + l-cocaine + "methyl benzoylecgonine" + neurocaine + benzoylmethylecgonine + beta-cocain + /m/0256b'*, where "+" is the logical operator OR to connect all our accurate terms plus the Freebase Identifier of cocaine. We have additionally searched on cocaine using the two main terms that yielded the majority of the cocaine-related Google search volume with the search string: '*cocaine + /m/0256b'.*

Terms for exclusion of news media from cocaine search: We have additionally searched on cocaine excluding related searches on news media that may obscure the trend in cocaine-related searches in Australia. A list of potential terms was obtained from the rising related searches for cocaine (as Drug, i.e. based on the freebase ID '*/m/0256b*', or as search term, '*cocaine*') in Australia in each of the 12 months using the Google Trends API. A short-list of these terms was then determined by Google searching for relevant news media using these terms and cocaine, e.g. '*nadia bartel cocaine'*. Because of the 30-term limit to GETH, we have only used the two main search terms for cocaine, i.e. *'cocaine'* and *'/m/0256b'.* The search string for excluding news media related searches from this process is: '*cocaine + /m/0256b - "nadia bartel" - /m/031tpn - "melbourne storm" - "storm cocaine" - /g/11dzswr4jq - "brandon smith" - /m/02qpwvy - /m/08h7qg - "bec judd" - "amber heard" - /m/075h8m - "tottie goldsmith" - "totti goldsmith" - /m/02wy12 - /g/11by_phg0t - "shane warne" - "geoff huegill" - /m/0941wt - /g/11h7nrgdq7 - "bailey smith" - /m/071ty - /m/04njq1 - "newcastle cocaine"'*, where *'-'* is the logical operator NOT for excluding news media related terms from our cocaine search using '*cocaine + /m/0256b'*.

3.3.2. Reddit

Geographical scope: In this bulletin, we only report the worldwide Reddit data without reporting the Australia-specific data. This is because neither the submission (comment) nor the user account location is accessible by the Reddit API due to privacy protection.

Subreddits scope: In Reddit, the "r/cocaine" is the primary subreddit people would choose to discuss about the substance cocaine. However, there exist many other subreddits involving non-negligible discussion about cocaine, e.g., "r/addiction" and "r/Stims". In

DOI: 10.26190/r4dt-aj11

order to report on Reddit as completely as possible, we investigated the following two summarized lists of subreddits regarding drug discussion:

1) The drug subreddits list summarized in the largest general drug subreddit "r/Drugs" (https://www.reddit.com/r/Drugs/wiki/subreddits/) which is maintained by the community members.
2) A list of drug related subreddits found by researchers: https://old.reddit.com/r/DrugMods/wiki/socialmediadrugresearchlist (The paper "Choosing Your Platform for Social Media Drug Research and Improving Your Keyword Filter List" can be found in https://doi.org/10.1177/0022042619833911).

After merging and deduplicating the two lists, there were 701 subreddits in total. We added three subreddits (r/Cocainegonewild2, r/CocaineRecovery and r/Cocaine3) by searching for communities on Reddit using the term *'cocaine'*, yielding a total of 704 subreddits.

For each of the subreddits, we counted the number of times either a submission or a comment mentioned the accurate term *'cocaine'* in June 2022. Of the 704 drug subreddits, we only included 120 subreddits with counts of *'cocaine'* higher than 10 in June 2022; these subreddits were used for data collection and analysis.

Cocaine search terms: Out of the 120 subreddits, only four subreddits were regarded as cocaine-specific subreddits: "r/cocaine", "r/cocainegonewild", "r/Cocainegonewild2/" and "r/CocaineRecovery". We used both our accurate terms and ambiguous terms to scrape these four subreddits; for all other subreddits, we only used the accurate terms.

3.3.3. Twitter

Geographical scope: In this bulletin, we only report the worldwide Twitter data, which means we collect all the tweets using the search terms without further filtering for Australia-specific data. This is because only 1~2% of tweets are geo-tagged according to the Twitter API document. See the bulletin for discussion of future potential work locating Australian-specific data.

Cocaine search terms: Since Twitter is a massive social platform which is not cocaine-specific in context, we only collect the tweets using our accurate terms of cocaine.

DOI: 10.26190/r4dt-aj11

3.4. Process of data collection, cleansing and collation

3.4.1. Google Trends

We used an Excel macro which was written by Raubenheimer (2021, https://github.com/TrueInsight/Google-Trends-Extraction-Tool) to access data from the GETH and the Google Trends API. A detailed user guide is available with the extraction tool downloaded. We used the geographical scope and cocaine search terms described in section 3.3.1 to query for data through the Excel macro from the GETH.

Google caches data requested through the API until midnight Pacific time, which means running the same query within the 24-hour day (according to Pacific time) will return the same results. However, using Raubenheimer's GETH Excel macro, we could obtain up to 12 samples in a single day using the unique sampling strategy described (Raubenheimer, 2022). To check the robustness of the results, we also ran the worldwide and Australia-specific queries over 3 days to check for consistency of data.

3.4.2. Reddit

For scraping the Reddit data, we used the Python Pushshift.io API wrapper (https://psaw.readthedocs.io/en/latest/; noted as PSAW in the bulletin) to access the Pushshift Reddit dataset (Baumgartner *et al.*, 2020), and then the Python Reddit API Wrapper (https://praw.readthedocs.io/en/stable/; noted as PRAW in the bulletin) to access more detailed fields of Reddit data. With a specified search term, e.g., *'cocaine'*, a targeted month, e.g., the June of 2022, and a targeted subreddit, e.g., "r/cocaine", the steps of collecting all the submissions and comments containing the string 'cocaine' from "r/cocaine" between 1st June 2022 and 30th June 2022 were as follows:

1) Input the search term, subreddit name and scraped time period into the function psaw.PushshiftAPI.search_submissions() of PSAW, then collect the returned ids as the submission identifiers.
2) Input the search term, subreddit name and scraped time period into the function psaw.PushshiftAPI.search_comments() of PSAW, then collect the returned ids as the comment identifiers.
3) By inputting each submission's identifier to the function praw.Reddit.submission() of PRAW, we could retrieve the following fields of each submission: [id, permalink, url, created_utc, author.name, title, selftext, upvote_ratio, num_comments, is_self, is_reddit_media_domain, distinguished]. Please refer to the documentation of PRAW for the exact meaning of each filed of a submission: https://praw.readthedocs.io/en/stable/code_overview/models/submission.html

DOI: 10.26190/r4dt-aj11

4) By inputting each comment's identifier to the function praw.Reddit.comment() of PRAW, we could retrieve the following fields of each comment: [id, link_id, author, body, created_utc, is_submitter, score, permalink, stickied, distinguished, edited]. Please refer to the documentation of PRAW for the exact meaning of each filed of a comment: https://praw.readthedocs.io/en/stable/code_overview/models/comment.html
5) We used the Python package pandas (https://pandas.pydata.org/) to store all the fields of all the submissions and comments into pandas.DataFrame, based on which we conducted the data cleansing and aggregating.

Process for data cleansing and collation:

Within a subreddit, since we used multiple search terms to scrape the submissions (resp. comments), and one submission (resp. comment) can contain multiple search terms at the same time, there existed duplications in our scraped submissions (resp. comments) after merging the submissions (resp. comments) scraped from all the different search terms into a single dataset. Thus, we used the unique identifier of each submission (resp. comment) to de-duplicate the scraped submissions (resp. comments) within each subreddit first. Then, for each of our targeted month, we collated the submissions (resp. comments) from all our studied subreddits. Note that, submissions and comments were treated equally and aggregated together for each month as the posts collection of a month for Reddit.

The PSAW returned the ids of posts which indeed contained the search term we input but some of the posts were deleted by users. The text contents of the deleted posts were invisible but other fields could be retrieved. We processed the deleted posts as follows. For the basic trend analysis of monthly counts of posts, deleted posts were still included. This is because those posts are ethically available for both PSAW and PRAW, and the posts were indeed part of the cocaine discussion back to their time periods. For all other analyses which need the actual text content, the deleted posts were excluded. Note that, the submissions in Reddit include two parts of texts, i.e., 'title' and 'content', and we concatenated the two parts together as the text content of each submission.

3.4.3. Twitter

Tweepy is an easy-to-use Python library for accessing the Twitter API (https://www.tweepy.org/). The function tweepy.Client.search_all_tweets is a wrapper for the full-archive search of the Academic Research access of the Twitter API, by which we could collect all the tweets satisfying certain standards. For example, to collect all the

DOI: 10.26190/r4dt-aj11

tweets containing the string ‘cocaine’ from the world within June of 2022, the steps are as follows.

1) Build the query clause: Since we only focus on the English written tweets, the query clause was ‘cocaine lang:en’.
2) Set the time window: The search function has two parameters ‘start_time’ and ‘end_time’. We set them as “start_time = ‘2022-07-01T00:00:00Z’” and “end_time = ‘2022-06-30T23:59:59Z’”.
3) Retrieve the specified fields of tweets: We use parameter ‘tweet_fields’ of the search function to choose the fields we wanted to collect from the following options: [attachments, author_id, context_annotations, conversation_id, created_at,entities, geo, id, in_reply_to_user_id, lang, non_public_metrics, organic_metrics, possibly_sensitive, promoted_metrics, public_metrics, referenced_tweets, reply_settings, source, text, withheld]. Please refer to the Twitter API documentation for the exact meaning of each filed of a tweet: https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/tweet

Due to the limitation of Twitter API, every request can only return 500 tweets. We utilised the ordering nature of retrieved tweets’ ids to solve the issue. Every time a request returned a collection of tweets, we recorded the last tweet’s id (the earliest tweet with the smallest id). For the next request, we set the parameter ‘until_id’ of the search function as the recorded smallest id, so that we could collect the next batch of tweets until we finished collecting all the tweets of the specified month.

Process for data cleansing and collation:

Because the tweets lookup function of Twitter API also returns the tweets when there is a username in the tweet containing our input search term, we manually implemented a cleansing function to remove such tweets. Then, for each month from July 2021 to June 2022, we collated all the tweets collected from different search terms. At last, via the identifiers of tweets, we de-duplicated the tweets retrieved more than once because they contained multiple search terms.

DOI: 10.26190/r4dt-aj11

4. Method of data analysis and presentation

4.1. Trend analysis of counts over time

For Google Trends, requested data were already aggregated into monthly search probability for worldwide and Australian searches as two separate time series (see section 3.3.1 on geographical scope). Further to that we have presented two additional time series for Australian searches: 1) using the search string 'cocaine + /m/0256b', and 2) using the search string for cocaine that excludes new media related (see section 3.3.1 for details). For Reddit, the number of posts were counted for each month after we cleansed and aggregated the scraped submissions and comments. For Twitter, the number of all tweets and number of original tweets (i.e. excluding retweets) were counted for each month after cleansing and presented as two separate time series. Retweeting is the user behaviour of forwarding an original tweet with or without adding extra comments. Since, the retweets mostly have much overlap with the original tweets, we present the two separate series. The aggregated monthly data from each of the three data sources are then used for the plot and analysis of monthly trend over time. Section 1 in the main bulletin presents the results of the trend analysis.

4.2. Topic modelling analysis

We used a topic modelling approach to divide the scraped posts from Reddit and Twitter (i.e. the text of the Reddit submissions, Reddit comments and tweets) into natural groups or topics so that we could describe how the discussed topics changed over time as well as study them in more detail (e.g. sentiment analysis for each of the topic). Topic modelling is a method for the classification of such posts, similar to the clustering of numeric data, which finds natural groups of items. Before the posts or text data can be analysed, data pre-processing is required as described in section 4.2.1. Several methods exist for topic modelling, and we use Latent Dirichlet Allocation (LDA) which is a popular method for fitting a topic model. It treats each post as a mixture of topics, and each topic as a mixture of words. This allows posts to "overlap" each other in terms of content, rather than being separated into discrete groups, in a way that mirrors typical use of natural language. We used Gensim (https://radimrehurek.com/gensim/#) for conducting the LDA model on our collected posts. It is an open-source library for unsupervised topic modelling, document indexing, retrieval by similarity, and other natural language processing (NLP) functionalities, using modern statistical machine learning methods. Gensim is implemented in Python and Cython for performance. It is designed to handle large text collections using data streaming and incremental online algorithms, which differentiates it from most other machine learning software packages that target only in-memory processing.

DOI: 10.26190/r4dt-aj11

We performed unsupervised LDA modelling of the topics for Reddit and Twitter separately because of the difference in their user base and context of the discussions on each of these two platforms. We found in our exploratory models that some of the topics from unsupervised LDA modelling were difficult to define. This is a commonly encountered issue in unsupervised LDA modelling[3]. Hence, we devised the following strategy for generating potential keywords for our domain experts to categorise into topics for each of Reddit and Twitter:

1) Data pre-processing (see section 4.2.1)
2) Generate multiple sets of outputs from unsupervised LDA modelling of topics (see section 4.2.2)
3) Compute keyword pair combinations from each topic in each of the LDA outputs and their frequency of occurrence, then generate a list of potential paired keywords ranked in descending order for their frequency of occurrence (see section 4.2.3)
4) Single-word frequency analysis (after the data pre-processing) and generate a list of potential single keywords ranked in descending order for their frequency of occurrence (see section 4.2.4)
5) Determine topic and associated single- and paired- keywords from a list of the top 1000 potential single keywords and a list of the potential paired keywords with a frequency of $\geqslant$5 (see section 4.2.5)

4.2.1. Data pre-processing

Before we fed the posts, i.e., all the cleansed and aggregated submissions and comments of Reddit (respectively tweets of Twitter) from July 2021 to June 2022 into the unsupervised LDA model and word frequency analysis, we performed the following pre-processing steps:

1) Tokenization: We used gensim.utils.simple_preprocess to split each post into words, and then lowercase the words into the Unicode-format strings.
2) Removing stop words: We only retained the tokens which did not belong to the stop word set of gensim.parsing.preprocessing.STOPWORDS and also were longer than one character.
3) Stemming: We used the English stemmer of nltk.stem.SnowballStemmer to retain the word stems of all the tokens.

[3] The course "Text as Data" by professor Chris Bail in Duke University claims the "Limitations of topic models" here: https://cbail.github.io/textasdata/topic-modeling/rmarkdown/Topic_Modeling.html

DOI: 10.26190/r4dt-aj11

In addition to Gensim, we used NLTK (https://www.nltk.org/) for data pre-processing in step 3 above. It is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning; wrappers for industrial-strength NLP libraries; and an active discussion forum.

4.2.2. Generating multiple sets of LDA model outputs

The unsupervised LDA modelling was performed on the whole year of data for Reddit and for Twitter. The module gensim.models.ldamodel.LdaModel provided us with a fast implementation of LDA. After transforming the pre-processed posts set into corpus and index format of the module, we set the parameters as following (see the details https://radimrehurek.com/gensim/models/ldamodel.html):

- num_topic = 10: The number of requested latent topics to be extracted from the training corpus.
- update_every = 1: The number of posts to be iterated through for each update.
- passes = 10: The number of passes through the corpus during training.

As the LDA model is based on probability, with the parameters varying, the output results also vary. While fixing the above parameters, for the next two parameters, we varied them in a range, so that we ran the LDA model 100 times in total:

- random_state: We fed the integers in range(50, 501, 50) as the 10 seeds to generate different random states.
- chunksize: The number of posts to be used in each training chunk. We varied it in the range(100, 1001, 100) as 10 different sizes.

4.2.3. Potential keyword pairs extraction

Each run of the LDA model on the posts output 10 subsets of keywords, where each subset contains the most representative keywords that the LDA model detected for each of the 10 topics. There exist overlaps among the 10 subsets generated within one LDA model output, as well as across the multiple LDA model outputs with different parameters for random_state and chunksize. We tallied the co-occurrence of two keywords within the 10-word set of each topic, by counting the frequency of each keyword pair in any of the 10-word sets across the 100 LDA model outputs. Note that, we excluded the keyword pairs if one keyword of them was from our search terms of cocaine, because the terms meaning cocaine itself are not helpful for topic modelling. For Reddit, we concatenated

DOI: 10.26190/r4dt-aj11

the comments together with the same submission they follow up, as one post. This is because, there is evidence (Vivek Kumar Rangarajan Sridhar. 2015) showing that, longer posts facilitate the detection of latent topics and their keywords in LDA models. A list of paired keywords ranked in descending order for their frequency of occurrence is generated for use by our domain experts to determine topics and keywords (see section 4.2.5).

4.2.4. Word frequency analysis

In addition to the keyword pairs from the LDA modelling outputs, we counted the frequencies of all the words (tokens) from all the pre-processed posts from Reddit. This is because some frequent keywords were representative and sufficiently unambiguous for us to determine that a post belonged to a certain topic as long as a post contained this keyword without further conditions. For example, when a submission contained the keyword '*addict*', it is fair to claim it is about the topic "negative effects". A list of the single words with their rank in descending order by their respective frequencies was generated for use in the following step to determine topics and keywords by our domain experts. Again, we excluded those words which were from our search terms of cocaine from this list.

4.2.5. Determination of topics and keywords

One of our domain experts (Nicola Man) in the drug and alcohol research field explored and determined the initial topics by browsing through the "r/cocaine" subreddit. The domain expert then determined the keywords to be categorised into topics. Where there was uncertainty in how a keyword might be used, the contextual use of the keyword was cross verified with a search through the posts dataset. The topics were further refined or added during the keyword categorisation process. A second domain expert (Amy Peacock) then cross-verified the definition of the topics and keywords. To reduce the workload on our domain experts, a short-list of: 1) the potential paired keywords with frequencies $\geqslant$5 (i.e. those that were <0.01% of the total tally were excluded[4]); and 2) the top 1000 potential single keywords, were used for categorisation into topics. Please note that only a small proportion of the keywords were categorised into topics because of the

[4] Given the number of potential keyword pair combinations in a 10-word set is 45, we obtain a total tally of 45,000 potential keyword pairs from 1,000 sets of topic keywords from the unsupervised LDA. This gives % of total tally for a co-occurring frequency of 4 as 4/45,000 = 0.009%.

DOI: 10.26190/r4dt-aj11

non-specific nature or ambiguity of the keywords. The determined topics at the broader (level 2) and narrower (level 1) categories and their definitions are shown in Table 2.

The single and paired keywords of each topic are listed in Table 3 for Reddit and in Table 4 for Twitter. As mentioned above (section 4.2.1), the keywords are processed by stemming and lemmatization so that all variants of a keyword can be captured. Note that the keywords used are different for Reddit and Twitter primarily because the top 1000 word frequency list used for each platform are different, i.e. the same terms (or term pairs) did not appear as frequently across the two platforms. To a lesser extent, some words were used in a different context across the two platforms, e.g. the word “question” would not be as specific to help-seeking on Twitter as it would on drug-related subreddits. Note that some of the terms identified may not directly relate to the actual topic. For example, many of the Twitter posts identified with terms on use of cocaine and sexual references were used as a means of political slandering.

DOI: 10.26190/r4dt-aj11

Table 2. Level 2 (broader) and Level 1 topics and their definition

Level 2 topic	Level 1 topic	Definitions
Direct effects	Negative effects	Terms referring broadly to possible acute or chronic negative effects of drugs (e.g., hangover, dependent, addict)
	Non-specific effects	Terms referring broadly to possible acute or chronic effects of drugs that don't have a specific valence (e.g., effect, feel)
	Positive effects	Terms referring broadly to possible positive effects of drugs (e.g., high)
Harm reduction & drug testing	Harm reduction & help-seeking	Terms referring to seeking information about the drug (e.g., help), seeking treatment or other services
	Drug testing	Terms referring to testing drugs (e.g., reagent) or biological samples (e.g., urine)
Poly-substance use	Use with alcohol	Terms referring to alcohol
	Use with other drugs	Terms referring to substances other than alcohol and cocaine
Use & markets	Drug forms	Terms referring to forms of cocaine (e.g., crack, powder), and includes making another form of/purifying the drug
	Drug markets	Terms referring to price (e.g., expensive), purity (e.g., pure), availability (easy), means of obtaining the drug (e.g. dealer) or large quantities (e.g. kg, tonne). This may include perception of the market (e.g. in comparison to fuel prices)
	Use behaviours	Terms referring to frequency, quantity of use or means of administration (e.g., nose, lines, gram)
Law & politics	Legal aspects	Terms referring to possible legal implications (e.g., police, seizure) and criminalised activities (e.g. launder, cartel, shipment)
	Politics	Terms referring to politics
Social context & impact	Sexual references	Terms referring to sex
	Social contexts of use	Terms referring to social impact (e.g. famili) or contexts of use (e.g., friends)
Miscellaneous	Coca cola	Terms referring to the drink Coca-cola
	Entertainment &media	Terms referring to celebrities, entertainment (e.g. movi, dance, parti), media channels (e.g. news) and videos/pics

DOI: 10.26190/r4dt-aj11

Table 3. Single keywords and keyword pairs for the topics on reddit

Level 1 topic	Single and paired keywords
Negative effects	*'heart', 'problem', 'anxieti', 'abus', 'comedown', 'die', 'pain', 'depress', 'withdraw', 'numb', 'crave', 'symptom', 'death', 'relaps', 'anxious', 'hospit', 'panic', 'paranoid', 'ill', 'psychosi', 'hangov', 'suicid', 'addict', ['feel', 'shit'], ['nose', 'burn'], ['nose', 'rate'], ['nose', 'drip'], ['nose', 'attack'], ['attack', 'caus'], ['caus', 'medic'], ['blood', 'rate'], ['blood', 'nasal'], ['blood', 'attack'], ['blood', 'pressur'], ['blood', 'nostril'], ['bodi', 'rate'], ['attack', 'bodi'], ['attack', 'rate'], ['attack', 'pressur'], ['nasal', 'burn'], ['nostril', 'burn'], ['rate', 'pressur'], ['nasal', 'nostril'], ['nasal', 'drip'], ['nostril', 'drip'], ['numb', 'burn'], ['nose', 'blood']*
Non-specific effects	*'effect', 'experi', 'sleep', 'dopamin', 'mental', 'toler', 'eye', 'serotonin', 'throat', 'cocaethylen', 'mood', 'chest', 'stomach', 'teeth', 'psycholog', 'sinus', 'liver', ['like', 'feel'], ['feel', 'dose'], ['feel', 'bodi'], ['caus', 'brain'], ['blood', 'hrs'], ['blood', 'bodi'], ['brain', 'dose'], ['brain', 'medic']*
Positive effects	*'high', 'happi', 'euphoria', 'amaz', 'energi', 'rush', 'euphor', 'kick', 'pleasur', ['feel', 'good'], ['feel', 'stimul'], ['stimul', 'brain']*
Harm reduction & help-seeking	*'help', 'wiki', 'water', 'quit', 'sober', 'danger', 'health', 'advic', 'harm', 'doctor', 'overdos', 'recoveri', 'receptor', 'rehab', 'treatment', 'healthi', 'sobrieti', 'recov', 'dosag', 'therapi', 'detox', 'harmreduct', 'aa', ['addict', 'year'], ['addict', 'anonym'], ['addict', 'stop'], ['addict', 'anon'], ['addict', 'start'], ['addict', 'al'], ['year', 'stop'], ['question', 'concern'], ['anonym', 'al'], ['anonym', 'na'], ['anon', 'al'], ['addict', 'medic'], ['nose', 'wash']*
Drug testing	*'test', 'urin', 'detect', 'saliva', 'drugtesthelp', 'kit', 'reagent', 'protestkit', 'reagenttest', 'drugtesthelp'*
Use with alcohol	*'alcohol', 'drunk', 'beer', 'booz', 'vodka', ['want', 'drink']*
Use with other drugs	*'meth', 'weed', 'heroin', 'amphetamin', 'mdma', 'ketamin', 'benzo', 'fentanyl', 'opiat', 'lsd', 'nicotin', 'xanax', 'pill', 'methamphetamin', 'speed', 'adderal', 'ecstasi', 'acid', 'benzodiazepin', 'caffein', 'opioid', 'psychedel', 'fent', 'cannabi', 'shroom', 'kratom', 'dmt', 'marijuana', 'coffe', 'molli', 'cigarett', 'mushroom', 'combo', 'thc', 'ritalin', 'dxm', 'morphin', 'prescript', 'downer', 'ket', 'phenibut', 'methylphenid', 'pcp', 'methadon', 'valium', 'pot', 'oxycodon', 'xan', 'codein'*
Drug forms	*'crack', 'powder', 'rock', 'white', 'freebas', 'crystal', 'fishscal', 'cook', 'hcl', 'snow'*
Drug markets	*'cut', 'buy', 'pure', 'money', 'product', 'dealer', 'sell', 'strong', 'wash', 'plug', 'street', 'scale', 'pay', 'expens', 'puriti', 'price', 'bought', 'aceton', 'order', 'lace', 'vendor', 'cost', 'cheap', 'batch', 'market', 'cheaper'*
Use behaviours	*'line', 'smoke', 'snort', 'blow', 'hit', 'bag', 'plate', 'home', 'trip', 'ski', 'bump', 'bing', 'ball', 'shot', 'shoot', 'crush', 'microwav', 'boof', 'sniff', 'inject', 'session', 'bender', 'gear', 'bathroom', 'sesh', 'iv', ['day', 'time'], ['use', 'day'], ['use', 'nose'], ['use', 'hot'], ['use', 'nasal'], ['use', 'stimul'], ['use', 'dose'], ['mg', 'mix'], ['hot', 'dri'], ['dri', 'heat'], ['use', 'nostril'], ['use', 'heat'], ['use', 'dri']*

DOI: 10.26190/r4dt-aj11

Legal aspects	*'legal', 'illeg', 'polic'*
Sexual references	*'dick', 'sex', 'horni', 'porn', 'cock', 'peni', ['shot', 'shoot']*
Social contexts of use	*'friend', 'social', 'chat', 'famili'*

Table 4. Single keywords and keyword pairs for the topics on Twitter

Level 1 topic	Single and paired keywords
Negative effects	*'addict', 'overdos', 'dead', 'death', 'danger', 'sinus', 'pain'*
Non-specific effects	*'brain', 'effect', 'bodi', 'mental'*
Positive effects	*'high', 'fun', 'energi', 'helluva', 'haven'*
Harm reduction & help-seeking	*'access', 'decrimin'*
Drug testing	*'test'*
Use with alcohol	*'alcohol', 'champagn', 'booz', 'wine', 'drunk', 'beer'*
Use with other drugs	*'heroin', 'fentanyl', 'meth', 'marijuana', 'weed', 'smoke', 'coffe', 'cannabi', 'pill', 'ecstasi', 'lsd', 'ketamin', 'caffein', 'xanax', 'mdma', 'methamphetamin', 'nespresso', 'mimosa', 'cigarett', 'adderal', 'opioid', 'oxycontin', 'mushroom', 'flower', 'ice', 'crystal', 'morphin', 'speed'*
Drug forms	*'crack', 'powder', 'dust', ['coca', 'leav'], ['gas', 'coca']*
Drug markets	*'worth', ' expens', 'sell', 'money', 'lace', 'gram', 'price', 'cheaper', 'dealer', 'pound', 'pure', 'sold', 'point', 'bag', 'suppli', 'kilo', 'cartel', 'import', 'colombia', 'freshest', 'pay', 'dollar', 'lbs', 'colombian', 'kilogram', 'columbia', 'cheap', 'shipment', 'estim', 'tonn', 'paid', 'ball', 'mexico', 'brick', 'bitcoin', 'credit', ['coke', 'gas'], ['buy', 'drug']*
Use behaviours	*'snort', 'line', 'snif', 'nose', 'inject', 'consum', 'dose', 'bump', 'ball'*

DOI: 10.26190/r4dt-aj11

Legal aspects	*'seiz', 'polic', 'arrest', 'court', 'legal', 'traffick', 'courtroom', 'illeg', 'possess', 'ndlea', 'ton', 'jail', 'charg', 'smuggl', 'bust', 'investig', 'border', 'kilo', 'law', 'cartel', 'crime', 'prison', 'custom', 'airport', 'crimin', 'murder', 'unit', 'cia', 'cop', 'dog', 'cbp', 'kilogram', 'intercept', 'corrupt', 'trial', 'guilti', 'ndlea_nigeria', 'shipment', 'judg', 'tonn', 'convict', 'testifi', 'narcot', 'brick', 'lawyer', 'launder', 'decrimin'*
Politics	*'republican', 'cawthorn', 'madison', 'gop', 'biden', 'hunter', 'trump', 'state', 'offic', 'donaldjtrumpjr', 'govern', 'donald', 'vaccin', 'russian', 'presid', 'mitch', 'war', 'nigeria', 'congress', 'vote', 'parliament', 'joe', 'mccarthi', 'kennedi', 'cawthornfornc', 'nation', 'joncoopertweet', 'offici', 'feder', 'democrat', 'ukrain', 'repswalwel', 'zelenski', 'lindyli', 'miller', 'patriottak', 'palin', 'conserv', 'politician', 'goplead', 'govt', 'potus', 'repcawthorn', 'madisoncawthorn', ['parti', 'run']*
Sexual references	*'orgi', 'sex', 'boob', 'hooker', 'sexual', 'dick', 'prostitut', 'stripper'*
Coca cola	*'cola'*
Entertainment & media	*'video', 'music', 'song', 'watch', 'slownewsdayshow', 'amber', 'news', 'abba', 'nba', 'cowboy', 'billmah', 'halftim', 'listen', 'grammi', 'escobar', 'club', 'pablo', 'uberfact', 'movi', 'photo', 'youtub', 'danc', 'clapton', 'tmz', 'patriottak', 'netflix', 'reuter', 'footbal', 'imag', 'foxnew', 'noliewithbtc', 'film', 'nypost', ['testimoni', 'heard'], ['testimoni', 'craziest']*

4.3. Trend and sentiment analysis of posts of each topic

With the detected topics and the determined keywords and keyword pairs of each topic, we firstly classified the posts into each topic, then we reported the monthly trend over each topic and also the sentiment of the classified posts.

4.3.1. Post classification by topic keywords

We used the basic string matching to classify the posts. In the bulletin, for a topic T, we denoted $\{T^s\}$ as its single keywords set and $\{T^p\}$ as its keyword pairs set. For a topic T, and each keyword $w \in \{T^s\}$, we collected the posts set $\{D^w\}$ where each post $D \in \{D^w\}$ contained the keyword w. Then, for each keyword pair$(w, w') \in \{T^p\}$, we collected the posts set $\{D^{(w,w')}\}$ where each post $D \in \{D^{(w,w')}\}$ contained both the keywords w and w'. At last, we de-duplicated the union post sets of all the retrieved $\{D^w\}$ and $\{D^{(w,w')}\}$ from all the single keywords and keyword pairs, ultimately arriving at a classified post collection of one topic. Since each post has a timestamp filed, it was easy to retrieve the classified posts of one topic in a certain month.

Note that, one post can be classified as multiple topics based on our method. This makes sense because different topics are never clearly separated and one can indeed talk about

DOI: 10.26190/r4dt-aj11

multiple topics in one post. For example, a user posted in the subreddit "r/2cb" on 29/07/2021 that, "Pro Tip for Snorting Just before snorting your 2cb, dip your finger in cocaine and line the inside of your nose with it. Let it sit there for a minute and try not to snort it. Then, snort your 2cb and you'll find that the pain is almost completely gone. Smooth comeup ahead :)". Because the submission contains the keyword 'snort', 'nose' and 'pain', this submission is considered as about the topic "Use of drugs" and "Negative effects" at the same time.

4.3.2. Trend analysis of each topic

In a similar manner to the trend analysis described in section 4.1, we counted the number of each topic's posts and reported the monthly trend of counts in the main bulletin. Please refer to the section 2 (Figures 4, 6 and 8) in the main bulletin for the detailed visualizations and analyses.

4.3.3. Sentiment analysis of each topic

The module nltk.sentiment.vader.SentimentIntensityAnalyzer in the VADER package provides us with an easy-to-use tool for analysing the sentiment of texts (Hutto et al., 2014). VADER is a lexicon and rule based sentiment analysis tool specifically calibrated to sentiments most commonly expressed on social media platforms. For each post, when calculating a polarity score, VADER outputs four metrics: negative, neutral, positive, and compound. Positive, negative, and neutral represent the proportions of the text that fall into these categories respectively, so that the values of these three metrics are within [0, 1]. The compound score calculates the sum of all lexicon ratings which is normalized between -1 (most negative) and +1 (most positive). With the computed four sentiment metrics of each post, we aggregated the sentiments for each categorised topic and also for each month. The results of the sentiment analysis is presented in the section 2 (Figures 5, 7, 9 and 10) in the main bulletin.

4.4. General caveats to interpretation of findings

- ***Identification of the drug (i.e. cocaine) may not be exhaustive.*** We cannot identify cocaine-related tweets and the submissions and comments in the non-cocaine-specific subreddits, using the ambiguous search terms. Also, misspelt terms are not identified unless they are specifically included in our list of search terms. Finally, we may not have exhaustively investigated all possible terms for cocaine. We plan to assess and/or include other sources of potential slang terms for a drug in

DOI: 10.26190/r4dt-aj11

our future work (e.g. https://urbanthesaurus.org/synonyms/cocaine, http://onlineslangdictionary.com/thesaurus/words+meaning+cocaine.html).

- ***Categorisation of data into topics may be subject to fallacy.*** The ambiguity or lack of specificity of some of the keywords may lead to misclassification of posts into topics.
- ***Censoring of data by the platform and/or the user of the platform.*** Some social media posts and/or comments may be deleted and unavailable at the time of data collection. For Twitter, the deleted tweets were not available for us to collect. For Reddit, if a deleted submission or comment satisfies our search term condition, it was included in the trend over counts. However, these deleted submissions and/or comments would not be available for topic modelling and sentiment analysis. Google's algorithms sort and extract the search probability data, so there is a degree of censoring or bias in the data obtained from GETH.
- ***Identification of geographical origin may not be available for all data.*** Some users may not disclose their geographical location, e.g. by not disclosing their location in their user account profile, or setting their preferences to not be tracked on their devices. Reddit also does not have a field for disclosure of location in the user profile or posts.
- ***Time series may be variable due to fluctuations unrelated to drug use, harms and markets.*** For example, the spike of tweets showing up in April 2022 was due to Elon Musk's tweet; "Next I'm buying Coca-Cola to put the cocaine back in", which has been retweeted more than 681,000 times so far. We have tried to account for or assess some of these fluctuations by excluding news media related Google searches for cocaine, as well as excluding retweets. However, it is difficult to account for all of the fluctuations or biases in trends from data that is unrelated to cocaine use, harms and markets.
- ***This is not a comprehensive scrape of all social media and search engine activity.*** In particular, while the majority of internet searches (>90%) are performed on Google, searches for an illicit drug such as cocaine particularly for the purpose of sourcing or buying may be performed on alternative search engines because of censorship (e.g. searching for "buy cocaine" would return more relevant results from Duckduckgo than from Google when the user is trying to purchase the drug). Further, our sample is limited to English language content, meaning we cannot generalise to those speaking other languages.

DOI: 10.26190/r4dt-aj11

Glossary

Term	Definition/Description
Post	An overarching term of social media for all three types of data we study, including the Reddit submissions, Reddit comments and tweets from Twitter.
Gensim	An open-source library for unsupervised topic modelling, document indexing, retrieval by similarity, and other natural language processing functionalities, using modern statistical machine learning. (https://radimrehurek.com/gensim/#)
GETH	The old Google Flu Trends (GFT) API, now named Google Extended Trends API for Health (GETH) provide the researchers with access to higher quality Google Trends data.
NLTK	A leading platform for building Python programs to work with human language data (https://www.nltk.org/)
PRAW	the Python Reddit API Wrapper (https://praw.readthedocs.io/en/stable/)
PSAW	the Python Pushshift.io API Wrapper (https://psaw.readthedocs.io/en/latest/)
Tweepy	An easy-to-use Python library for accessing the Twitter API (https://www.tweepy.org/)

Funding

The Drug Trends program is funded by the Australian Government Department of Health and Aged Care under the Drug and Alcohol Program.

Acknowledgements

We would like to acknowledge the following individuals for their contribution and/or advice to this project: Max Pedersen, Anant Mathur, Rajat Katyal, A/Prof Wayne Wobcke, A/Prof Yanan Fan and Prof Scott Sisson.

Recommended citation

Linghu, Q., Peacock, A., Sutherland, R., Bruno, R., Barratt, M. J., & Man, N. (2022). Trends in Google searches and social media discussions about cocaine, July 2021-June 2022: A pilot study. Drug Trends Bulletin Series. Sydney: National Drug and Alcohol Research Centre, UNSW Sydney. DOI: 10.26190/r4dt-aj11

DOI: 10.26190/r4dt-aj11

Related Links:

- Cryptomarket bulletin: https://ndarc.med.unsw.edu.au/resource-analytics/trends-cryptomarket-drug-listings-oct2021-sep2022
- Cryptomarket data visualisation: https://drugtrends.shinyapps.io/cryptomarkets
- For more research from the Drug Trends program go to: https://ndarc.med.unsw.edu.au/program/drug-trends

Contact us

Email: drugtrends@unsw.edu.au