COMMONWEALTH OF MASSACHUSETTS

Cannabis Control Commission's Metrc Data Dictionary and User Guide for the Laboratory Test Results Datasets for THC% and THCA%

Version 1.0

Massachusetts Cannabis Control Commission

Table of Contents

Last updated 11/5/2024

Our Mission

The mission of the Cannabis Control Commission is to honor the will of the voters of Massachusetts by safely, equitably, and effectively implementing and administering the laws enabling access to medical and adult use marijuana in the Commonwealth.

The Commission will foster the creation of a safely regulated industry that will create entrepreneurial and employment opportunities and incremental tax revenues in and to communities across the state and which will be a best practice model for other states. The industry will be characterized by participation by small and larger participants and with full and robust participation by minorities, women and veterans. We will develop policies and procedures to encourage and enable full participation in the marijuana industry by people from communities that have previously been disproportionately harmed by marijuana prohibition and enforcement and positively impact those communities.

About this Dataset

All testing results are entered by Independent Testing Laboratories (ITLs) into Metrc in one of three ways: manually, via a .csv upload, or via an API through a third-party Lab Information Management Systems (LIMS). To maintain consistency of the data reported, we included results from after April 2021, as testing requirements were modified in the first quarter.

Please note that the Commission releases testing data after a six-month lag period to prevent any potential interference with ongoing investigations.

Relevant tables were queried using Structured Query Languages (SQL) scripting, saved locally, identifying fields anonymized, and shared.

Test Types

The 2 test types included are:

- THC (%) Raw Plant Material
- THCA (%) Raw Plant Material

Dates included

- April 12, 2021-March 31, 2024

Products included

- Testing sample packages only

Data Dictionary

Fieldname	Format	Description	Cleaning Procedure	Inclusion Criteria
Product Category Type	Text/String	The types of products included in the dataset	None	All products from Flower
ID	Text/String	Anonymized package label from Metrc.	The Metrc label was scrambled using unique keys in Python to prevent third parties from tracking specific packages or products.	Sample packages with complete IDs.
Strain	Text/String	Anonymized strain name in the format of StrainT####.	Each strain was replaced by unique value for anonymization using a random key.	All strain names including null values.
Test Type Name	Text/String	The complete analyte's name as entered in Metrc.	None	11 test types. None include retests.
Test Result	Floating-point number	Result of the analyte test as entered by the ITLs directly into Metrc or via a third-party integrator. Please note that "0" for test results do not necessarily reflect an analyte test results. They were entered to satisfy	None	All test results entered in Metrc by the ITLs.

		requirements for data entry as set by Metrc. Refer to the "Test Comment" field for additional information on the analyte test.		
Testing Lab (ITL)	Text/String	Anonymized Lab names in the format of LabX where X is a letter used to differentiate lab names.	Each lab was replaced by unique value for anonymization using a random key.	Commercial Licensed Testing Labs operating in Massachusetts.
Test Performed Date	Date	Date value in the form of YYYY-MM-DD.	Time values in the datetime field were removed.	April 1, 2021-September 30, 2023
Test Comment	Text/String	Self-entered comments and notes about the analytes or samples, including sample adjustments and corrections.	None	All text included in the TestComment field in Metrc

Explanation of Lab Testing Status Often Seen in Test Comments

Lab Testing Status	Explanation of the Lab Testing Status
Not Required	The State does not require regulatory compliance testing of this type of item.
Not Submitted	The State requires regulatory compliance testing of this type of item, but it has not yet been tested.
Not Tested/ "Test Not Run"	Test result for this test type was not run at this lab/in this test session; see other results for this package. This is only a placeholder.
Submitted for Testing	A test sample has been taken from this package for purposes of state required testing but no test results have been reported.
Test Passed	The package passed regulatory compliance testing. (This should be verified via the Lab Results tab.)
Test Failed	The package failed regulatory compliance testing. (This can be verified via the Lab Results tab.)
Remediated	The facility has remediated the failed product and recorded the remediation in Metrc. The remediated package has not yet been tested. This cannot be transferred out of the facility.
Retest Passed	The package that had previously failed testing and was retested which all testing on the new sample(s) have passed.
Retest Failed	The original failed package has had at least one of the retest sample(s) fail retesting.
Awaiting Confirmation	When the state requires two retests after an initial fail and only a single retest has been done. The product will need to have one more test done to confirm the results of the first retest.
Testing in Progress	Sample(s) were taken from the package. Not all Test Results have been entered into Metrc from one or more test samples.

Ethics of Using Large Laboratory Test Result Datasets

Large datasets like laboratory test results offer tremendous potential for improving public health through research, trend analysis, and quality control initiatives. However, mishandling such data can have negative consequences. Here, we discuss some key ethical considerations for responsible use of this dataset:

1. **Privacy and Confidentiality**

    - While there are no Personally Identifiable Information (PII), this dataset contains identifiable licensees' information. Anonymization techniques like creating aliases for licensees, the labs' names, or the product Metrc tags or strains are crucial before sharing or analyzing the data.

2. **Algorithmic Bias and Fairness**

    - Algorithms used to analyze the data can perpetuate harms caused by the War on Drugs. It's important to be aware of potential biases in the data itself and to choose algorithms that mitigate these biases.
    - Validate and interpret results carefully to avoid drawing misleading or discriminatory conclusions.

3. **Transparency and Explainability**

    - Be transparent about the source, collection methods, and limitations of the data.
    - Document the rationale and methodology used for data analysis to ensure the results are verifiable and reproducible.

4. **Public Health Benefits vs. Commercial Interests**

    - Ensure the primary purpose of using the data aligns with understand the harmful effects of the War on Drugs in affected communities.
    - Be transparent about any potential commercial interests or conflicts of interest associated with the data analysis.

5. **Responsible Data Sharing**

    - Develop clear guidelines and data use agreements for sharing the anonymized dataset with researchers or institutions.
    - Encourage responsible secondary use of the data for further public health advancements.

By adhering to these ethical principles, we can harness the power of large datasets like this one to contribute meaningfully to public health research and improvement.

For Additional Information

For additional information on testing and analysis protocols, please visit
https://masscannabiscontrol.com/testing-and-analysis/.

(774) 415-0200
Commission@CCCMass.com

Questions?

If you have additional questions regarding our data, please contact the Commission at (774) 415-0200
Commission@CCCMass.com.