**SQL query in Hive**

Dataset from Data.gov: "Influenza, COVID-19, RSV, and Other Respiratory Virus Laboratory Surveillance”

URL: <https://catalog.data.gov/dataset/influenza-covid-19-rsv-and-other-respiratory-virus-laboratory-surveillance>

This dataset includes aggregated weekly respiratory virus laboratory data that the Chicago Department of Public Health (CDPH) uses to monitor influenza, COVID-19, respiratory syncytial virus (RSV), and other respiratory virus activity in Chicago. The data represents respiratory viral PCR tests performed by several hospital laboratories in Chicago as well as two commercial laboratories serving Chicago facilities. Data are voluntarily reported on a weekly basis and do not contain patient demographic or geographic information. The data reported represent both Chicago and non-Chicago residents tested by the reporting facility.

The respiratory viruses included are influenza, RSV, SARS-CoV-2, parainfluenza, rhinovirus/enterovirus, adenovirus, human metapneumovirus, and seasonal coronaviruses. Influenza laboratory data are available from 2010-2011 to present; for all other respiratory viruses data are available from 2019-2020 to present.

**Business question: What is the percentage positivity rate for each respiratory virus, based on the PCR tests performed by the hospital and commercial laboratories in Chicago?**

**SQL query and explanation:**

SELECT pathogen,

SUM(lab\_tot\_tested) AS Sum\_of\_Total\_Tested,

SUM(lab\_tot\_positive) AS Sum\_of\_Total\_Positive,

(SUM(lab\_tot\_positive) \* 100 / SUM(lab\_tot\_tested)) AS Sum\_of\_Total\_Percentage\_Positive

FROM respiratory\_virus\_laboratory\_surveillance

GROUP BY pathogen

ORDER by pathogen ASC;

Output:

A screenshot of a computer

Description automatically generated

pathogen: Represents the types of respiratory viruses.

sum\_of\_total\_tested: Represents the total number of tests conducted for each virus.

sum\_of\_total\_positive: Indicates the number of positive results for each virus.

sum\_of\_total\_percentage\_positive: Displays the percentage of positive tests for each virus.

The results tab provides a detailed breakdown of the total tests conducted, the number of positive results, and the percentage of positive results for various respiratory pathogens. The data provides insights into the prevalence and positivity rate of each pathogen based on laboratory tests.