Query: Track the number of passengers over time before and after COVID to see the effects that COVID may have had on people taking a taxi.

Another potential database query that can be used for our application would be tracking the number of passengers over time before and after the COVID-19 pandemic to observe the effects that the virus may have had on people taking a taxi.

Real-world relevance:

For the past 3 years, the world has been plagued by the fatal COVID-19 virus which is spread from either physical contact or inhalation. As a result of the deadly virus' high transmission rate, people were advised and forbidden from participating in activities that involved close contact with one another. Many public transportation services were shut down such as buses and trains since those involved people being contained in a space for a long time in proximity with other individuals. Due to the multiple safety procedures and bans set in place as well as fear of catching the virus, many people were afraid and unable to call and take taxis. Given the datasets provided to us, our application would be able to track the number of passengers that used certain NYC taxicabs from the years 2009-2022 which covers the time periods before the pandemic, during the pandemic, and after it.

Variables to consider:

Number of passengers:

One of the primary variables we will consider in this query is the number of passengers that took rides on the NYC taxicabs. With the data given to us, we can find trends in how the number of passengers changed due to potential factors such as the COVID-19 pandemic by fitting that information into a linear graph.

Number of COVID-19 cases in NYC :

Another potential variable we can consider is the number of COVID-19 cases in NYC for the time that the pandemic took place. By representing the number of cases in a linear graph, we can make connections between how the number of cases could affect peoples' reactions to the virus' severity and their unwillingness to take a taxicab causing a decrease in the number of passengers.

Number of shared/pool rides:

In addition to the previous parameters, this query shall also focus on the number of shared/pool rides that took place in our given dataset over the observed time. This shall be represented as well in a linear trend graph for visibility and clear analysis.

Time:

Time is by far the most important variable to consider for our query and data analysis because it helps us determine the trends that we can draw conclusions from. We will have three main time periods based on the data we have: before the pandemic, during the pandemic, and after the pandemic. The years 2009-2019 will comprise the "before the pandemic" timeframe, years 2020 and 2021 will represent "during the pandemic", and 2022 will be considered "after the pandemic". By representing both the datasets on the number of COVID cases in NYC and the number of passengers against these timeframes, we can make astute observations on any cause-effect relationships that may have occurred.