Phân tích câu hỏi nghiên cứu:

1. Số ca mắc Covid-19 có sự khác biệt giữa các châu lục hay không ? **(X: Continent, Total Cases)** **(Y: TotalCases)**

Với câu hỏi này, ta có thể tìm hiểu mối quan hệ giữa 2 biến **Continent** ( biến phân loại) và **TotalCases** ( biến số ) thông qua biểu đồ boxplot để có cái nhìn sơ lược về mối liên hệ giữa số ca mắc Covid-19 của các châu lục.

Sau đó, để khẳng định được số ca mắc có phụ thuộc vào từng châu lục hay không, ta xác định đây là suy luận: C 🡪 Q: nhiều nhóm độc lập, mỗi nhóm là 1 châu lục gồm nhiều nhiều nước. Tiếp theo, ta có thể thực hiện thống kê ANOVA F-test để tìm được p-value, sau đó có thể kết luận 2 biến có phụ thuộc nhau hay không?

1. Có phải các nước có nền kinh tế phát triển, thu nhập cao thì dịch bệnh Covid-19 sẽ ít nghiêm trọng hơn so với có nền kinh tế phát triển kém hơn, thu nhập thấp hơn không ? **(X: Continent, Location, TotalCases, TotalDeaths)** **(Y: True/False)**

Những nước có nền kinh tế phát triển thường sẽ có hệ thống y tế tốt hơn. Ta có thể nhận thấy những nước thuộc châu Âu (Anh, Đức….), châu Mỹ (Hoa Kỳ, Canada..) thường có nền kinh tế phát triển hơn các khu vực còn lại. Ta xem xét tình hình dịch Covid-19 tại châu Âu, Mỹ so với các châu lục còn lại, đồng thời so sánh, nhận xét về tình dịch (số ca mắc, tử vong) của các nước đứng đầu về nền kinh tế như Mỹ, Anh, Đức…

1. Dân số có phải là nguyên nhân dẫn tới việc gia tăng số ca mắc ở các nước châu Á ? **(X: Location, Population, TotalCases)** **(Y: TotalCases)**

**Dân số đông làm cho mọi người tiếp xúc với nhau gần hơn, điều đó có thể dẫn tới lây lan dịch Covid-19. Để xem xét sự phụ thuộc của dân số vào số ca mắc, ta sẽ xem xét mối quan hệ giữa hai biến: Q 🡪 Q. Dùng biểu đồ Scatter để thể hiện mối quan hệ giữa hai biến số, tính hệ số tương quan để có thể rút ra kết luận.**

1. Khi vaccine được phổ biến, điều đó có giúp ích cho việc chống đại dịch Covid-19 trên thế giới? **(X: Date, PeopleFullyVaccinated, NewCases, NewDeaths)** **(Y: True/False)**

Vaccine xuất hiện là một giải pháp tốt có thể giúp đẩy lùi đại dịch Covid-19 đang hoành hành. Để có thể đưa ra nhận xét vaccine có thật sự giúp ích cho đại dịch hay không, ta cần xác định được ngày bắt đầu triển khai tiêm vaccine, sau đó xem xét mối quan hệ giữa biến **PeopleFullyVaccinated, New Cases và PeopleFullyVaccinated, NewDeaths. Đó là các mối quan hệ giữa hai biến số (Q 🡪 Q), ta dựa vào biểu đồ scatter để đưa ra nhận xét, tính toán hệ số tương quan để nhận định mối quan hệ lần lượt giữa hai biến.**

1. Ở các châu lục, trước và sau khi tiêm vaccine, tỉ lệ tử vong của dịch bệnh Covid-19 như thế nào? **(X: Continent, Date, TotalDeaths, TotalCases)** **(Y: Tỉ lệ tử vong)**

**Để trả lời được câu hỏi, ta cần xác định được thời điểm trước và sau khi tiêm vaccine của các châu lục. Tiếp theo, ta vẽ biểu đồ đôi so sánh số ca mắc, tử vong của từng châu lục giữa hai thời điểm đó, tính toán tỉ lệ tử vong rồi đưa ra các kết luận**

1. Biến thể Omicron xuất hiện vào 24/11/2021 có gây nguy hiểm không ? So sánh với biến thể Delta xuất hiện trước đó ? **(X: Date, NewCases, NewDeaths)** **(Y: Số ca mắc và tử vong của hai loại biến thể theo từng ngày)**

Biến thể Omicron xuất hiện vào khoảng cuối năm 2021, được nhận định là biến thể có khả năng lây lan nhanh nhất. Ta có thể vẽ biểu đồ so sánh số ca mắc và tử vong mới mỗi ngày giữa hai loại biến thể, từ đó đưa ra nhận xét liệu số ca mắc tăng nhanh có đồng nghĩa số ca tử vong cũng tăng hay không ? Xét mối quan hệ giữa hai biến số Q 🡪 Q (NewCases, NewDeaths), dựa trên biểu đồ scatter rồi đưa ra kết luận dựa trên hệ số tương quan.

1. So với các bệnh như SARS, EBOLA, MERS, H1N1 thì dịch Covid-19 có nghiêm trọng hơn ? **(X: TotalDeaths, TotalCases)** **(Y: True/False)**

Vì các đại dịch trên đã kết thúc, ta có thể thu thập được số liệu về số ca mắc, tử vong của từng đại dịch. Ta tính toán tỉ lệ tử vong của từng đại dịch, rồi vẽ biểu đồ trực quan giữa các đại dịch đó. Cuối cùng, đưa ra kết luận mức độ nghiêm trọng của dịch Covid-19.