

# التقرير النهائي لتحليل البيانات " كوفيد-19"

## 1. جمع البيانات:

تم الاعتماد على مجموعة البيانات العالمية COVID-19 – Our World in Data المتاحة على منصة Kaggle، والتي تتضمن مؤشرات الحالات، الوفيات، التطعيمات، والعوامل الاقتصادية والديموغرافية عبر أكثر من 230 دولة خلال الفترة من 1 جانفي 2020 إلى 5 مارس 2022. تم تحميلها مباشرة عبر واجهة Kaggle API ملف تنظيف البيانات والمعالجة: `owid-covid-data.csv`.

## 2. نظيف البيانات (Data Cleaning) :

### التحقق الأولي

- فحص أنواع الأعمدة وعدد القيم المفقودة.
- التأكد من أن عمود `date` منسق كتواريخ.
- تحديد الفترة الزمنية وعدد المواقع (238 دولة).

### تنقية البيانات

- حذف الصفوف الخاصة بالتجميعات الإقليمية. (World, Africa, Europe ...)
- الاحتفاظ فقط بسجلات الدول.
- حذف الأعمدة التي تتجاوز نسبة القيم المفقودة فيها 70%.
- تعويض القيم المفقودة في الأعمدة الرقمية بالوسيط، وفي النصية بكلمة "Unknown".
- حذف التكرارات وتوحيد أسماء الأعمدة لتكون بصيغة `minuscule` ومن دون مسافات.

### النتيجة بعد التنظيف

- عدد الصفوف النهائية:  $\approx 158,000$ .
- عدد الأعمدة المحتفظ بها:  $\approx 50$ .
- خلو الجدول من التكرارات والقيم الناقصة المهمة.

## 3. التحليل الاستكشافي (Exploratory Data Analysis)

تم إجراء تحاليل مبدئية لتوضيح:

- تطور عدد الحالات والوفيات عبر الزمن.
- المقارنة بين القارات والدول.
- العلاقة بين التطعيمات وانتشار العدوى.
- الفوارق المرتبطة بالعوامل الاقتصادية.

## 4. Power BI Visualizations

بعد تجهيز الملف clean\_covid\_data.csv ، تم تحليلها والتوصل الى أبرز النقاط :

- متوسط العمر للمصابين كانوا من فئة الشباب.
- وصلت الإصابات ذروتها في حدود 2022 حيث كانت في تزايد المستمر قبل ذلك مع الأخذ بعين الاعتبار أن جمع البيانات توقف بتاريخ 5 مارس من نفس السنة، رغم أن عدد اللذين تلقوا اللقاح وصل ذروته قبل ذلك أي في حدود نهاية 2021 .
- الدول التي تتمتع بالقدرة الاقتصادية استطاعت تحسين البنية الصحية التحتية فتحكموا بذلك على عدد الوفيات على عكس الدول التي تعاني من هشاشة في احديهما أو كليهما بحيث ساهم الإهمال أو عدم القدرة على توفير الرعاية بتزايد الإصابات وحتى الوفيات .
- الدول التي تضم عدد كبير من المسنين غالبا ماكانت تسجل عدد وفيات اكبر وذلك بسبب الإصابة المسبقة للمسنين بالأمراض كالسكري وأمراض القلب، بحيث يكون تأثير الفايروس أقوى عليهم.