

**King Khalid University**

**Introduction to data science**



## **Report**

### **Implementing Python Basics for DS**

Areej khaled	442807629
Ashwag ali	441804462
Hanan ali	442807670
Mozun fayz	441802826

**Ms. Maram Alrajhi**

## INTRUCTION

This report contains data on cases and deaths from COVID-19 in the United States, and operations on this data on what we learned in class Introduction to Data Science

## Our notebook

### Step 1

From where getting our data? There is a COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University hosted on .Github

(..then imported by four librarys (pandas, numpy, matplotlib

### Step 2:

Data cleanup and transformation we used the data in California to make the data transformation run less intensively .because the US has duplicate county names

Then we added new data containing the population to compare it with our data

### Revisit step 2:

including rolling averages of new cases/deaths and data per million inhabitants

### Step 3 :

Step 3: Presentation and visualization In this step, we'll create a first couple of graphs. We want to visualize the data for the 4 counties nearest to California.

Then revisit step 2: including rolling averages of new cases/deaths and data per million inhabitants. In order to generate the data per million inhabitants, we'll first need to know the amount of inhabitants per county, so we add new data about the population.

## Sources

data1: [https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/time\\_series\\_covid19\\_confirmed\\_US.csv](https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv)

data2: [https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/time\\_series\\_covid19\\_deaths\\_US.csv](https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_US.csv)

data3: <https://gist.githubusercontent.com/NillsF/7923a8c7f27ca98ec75b7e1529f259bb/raw/3bedefbe2e242addba3fb47cbcd239fbed16cd54/california.csv>

## References

1- Classes of Introduction to data science.