



# Análisis de datos con Python

## Proyecto - Módulo 4

Equipo 10 - Integrantes:

- Carrillo Rivera Rafael Sebastián
- Espinosa Ávila Ana Yessica
- Velázquez Moreno Salvador





# Planteamiento del problema

Previamente, en el módulo 3 de ‘Procesamiento de datos con Python’, hicimos la exploración, transformación y análisis de datos sobre un dataset obtenido en Kaggle, actualizado constantemente por Gabriel Preda, el cual toma una “pequeña” porción de los datos originales recopilados por la asociación Our World in Data y muestra los datos referentes a la vacunación y al avance de esta a nivel mundial.

Enlace a página de inicio del dataset de Gabriel Preda:

<https://www.kaggle.com/gpreda/covid-world-vaccination-progress>



Dichos datos funcionaron a la perfección para hacer todo el procesamiento que deseábamos ejecutar sobre un dataset que fuera de nuestro interés para nosotros y para la mayoría de la gente, por los tiempos tan complicados que nos aquejan.

Sin embargo, en este caso decidimos expandir un poco más los datos con la que trabajaríamos, pues el objetivo era llevar a cabo un análisis un poco más exhaustivo, el cual implicaba correlaciones, exploración de variables categóricas, análisis multivariable, visualización de datos, etc.

Fue así que trabajamos esta vez con los datos originales de Our World in Data, el cual incluye datos relevantes referentes a las hospitalizaciones, fallecimientos, tests, esperanza de vida, índice de rigurosidad de los gobiernos, etc. Todo ello registrado por día y por país, y por supuesto, referente a la emergencia sanitaria del COVID-19.

# Data on COVID-19 (coronavirus) by *Our World in Data*

 Download our complete COVID-19 dataset : [CSV](#) | [XLSX](#) | [JSON](#)

Our complete COVID-19 dataset is a collection of the COVID-19 data maintained by [Our World in Data](#). We will update it daily throughout the duration of the COVID-19 pandemic. It includes the following data:

Metrics	Source	Updated	Countries
Vaccinations	Official data collated by the Our World in Data team	Daily	217
Tests & positivity	Official data collated by the Our World in Data team	Weekly	136
Hospital & ICU	Official data collated by the Our World in Data team	Weekly	37
Confirmed cases	JHU CSSE COVID-19 Data	Daily	194
Confirmed deaths	JHU CSSE COVID-19 Data	Daily	194
Reproduction rate	Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C	Daily	185
Policy responses	Oxford COVID-19 Government Response Tracker	Daily	186
Other variables of interest	International organizations (UN, World Bank, OECD, IHME...)	Fixed	240

Información sobre los datos del dataset de Our World in Data:

<https://github.com/owid/covid-19-data/blob/master/public/data/README.md>



# Análisis de datos

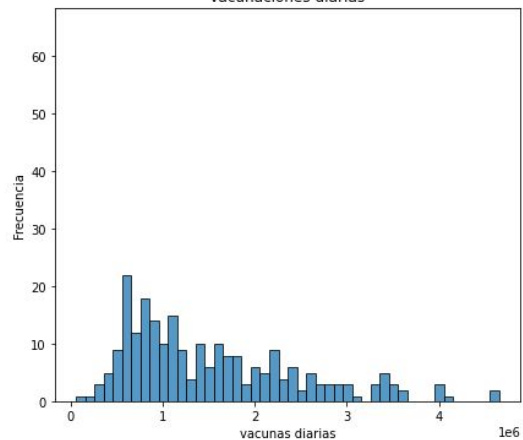
Es importante aclarar que el análisis realizado se basa en datos recopilados del 24 de febrero del 2020 hasta el 19 de septiembre del 2021.

Comenzamos realizando una función que recibía un argumento sencillo: el nombre en inglés de algún país de nuestro interés.

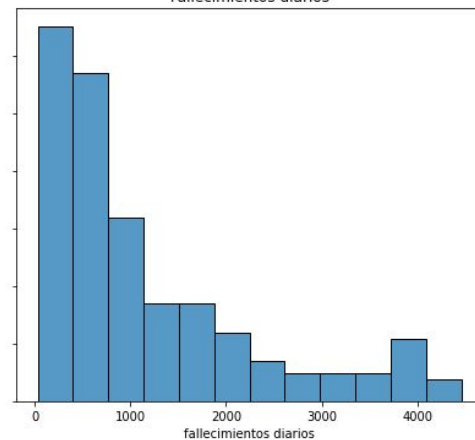
Dicha función nos entregaba datos relevantes como los histogramas de frecuencias de nuevos casos, nuevos fallecimientos, nuevos tests realizados y nuevas vacunaciones, así como un heatmap y scatterplot que permite ver la correlación entre estas variables

```
datos_covid_de('United States')
```

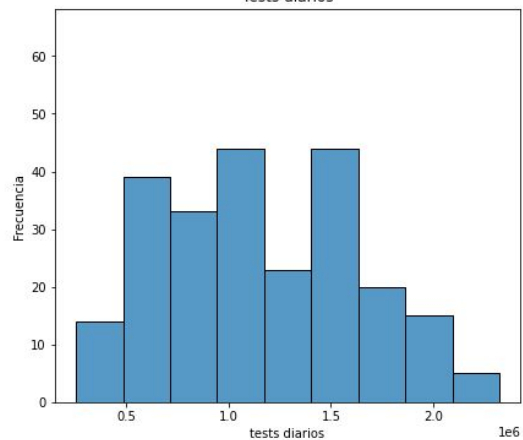
Vacunaciones diarias



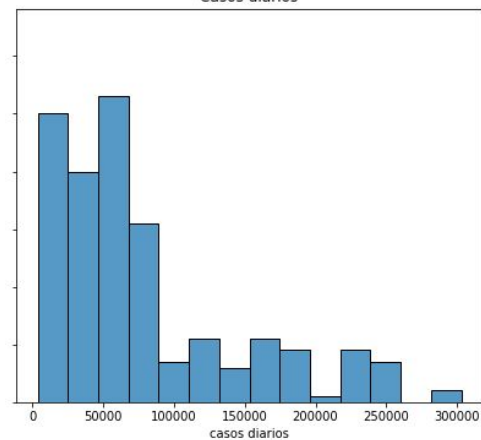
Fallecimientos diarios

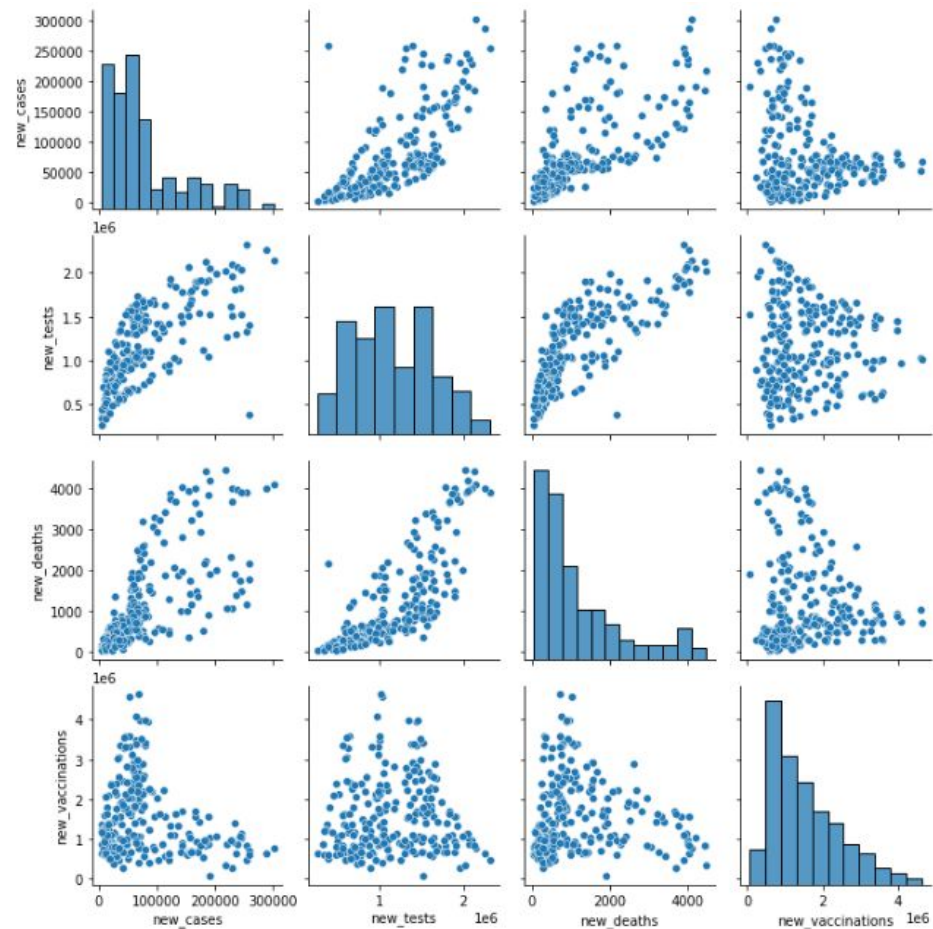


Tests diarios

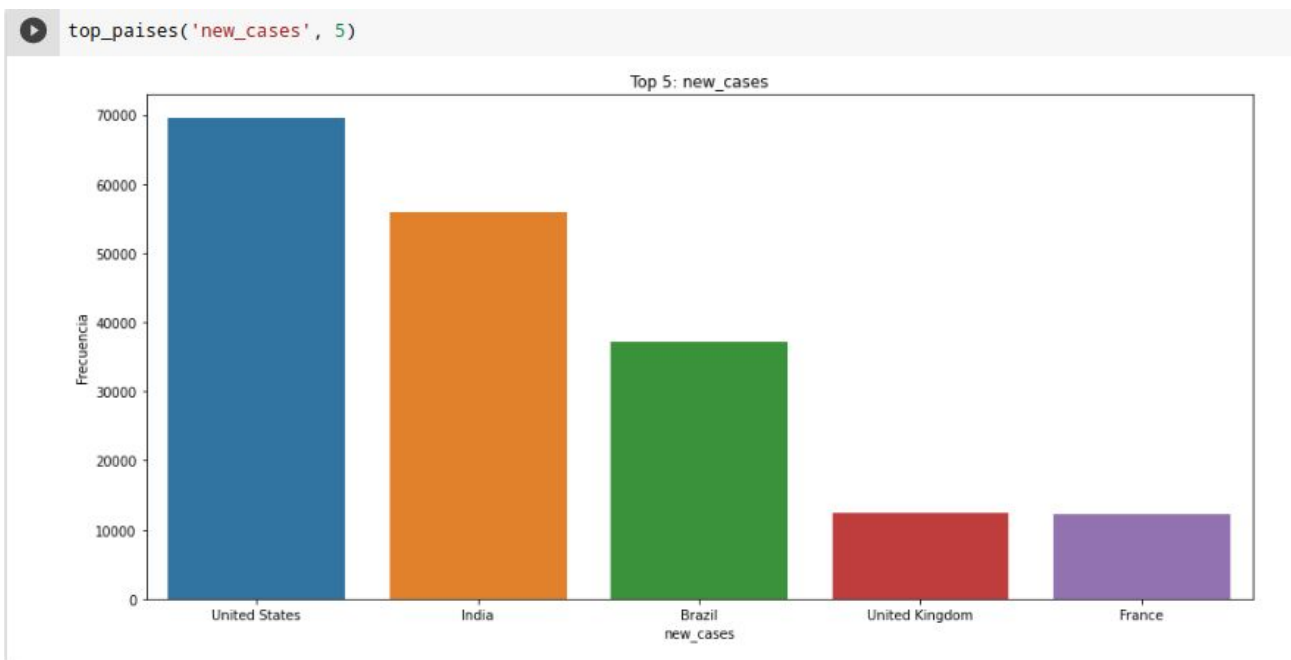


Casos diarios



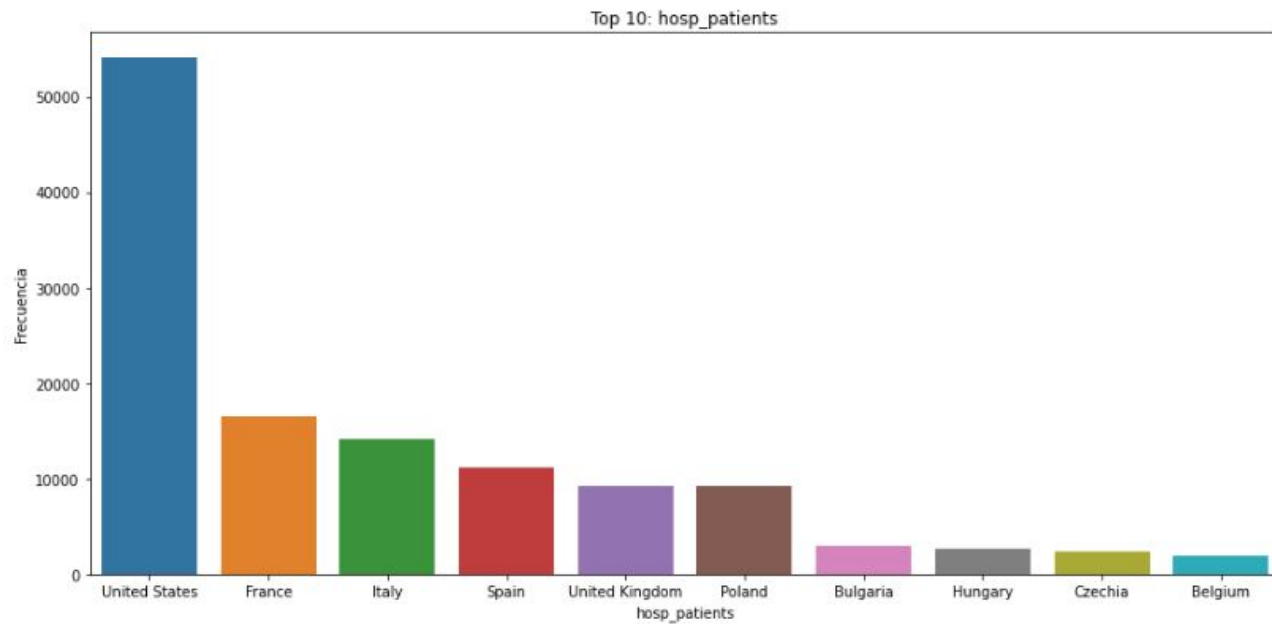


Proseguimos con otra función que igualmente recibía dos argumentos: un dato de interés, el cual nos permitiera conocer el TOP 'n' de los países con más casos, más hospitalizaciones, más vacunaciones, etc. (según el dato que quisiéramos conocer). El segundo argumento sería 'n'.

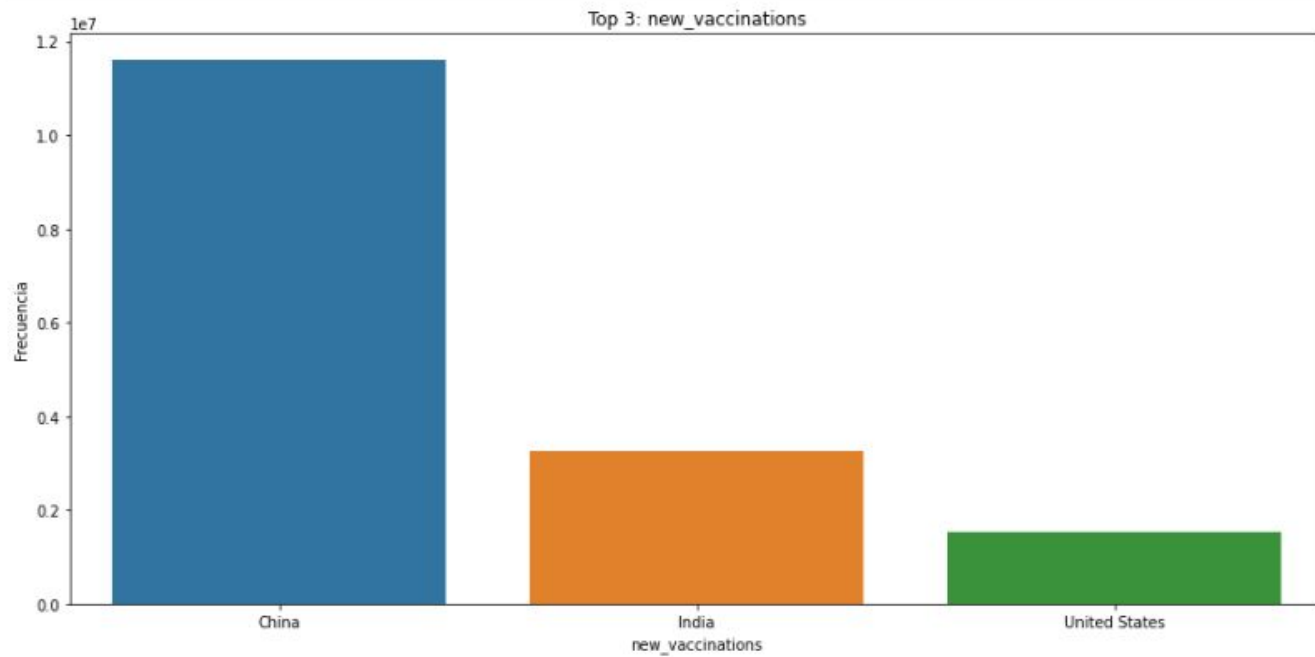




▶ `top_paises('hosp_patients', 10)`



```
top_paises('new_vaccinations', 3)
```





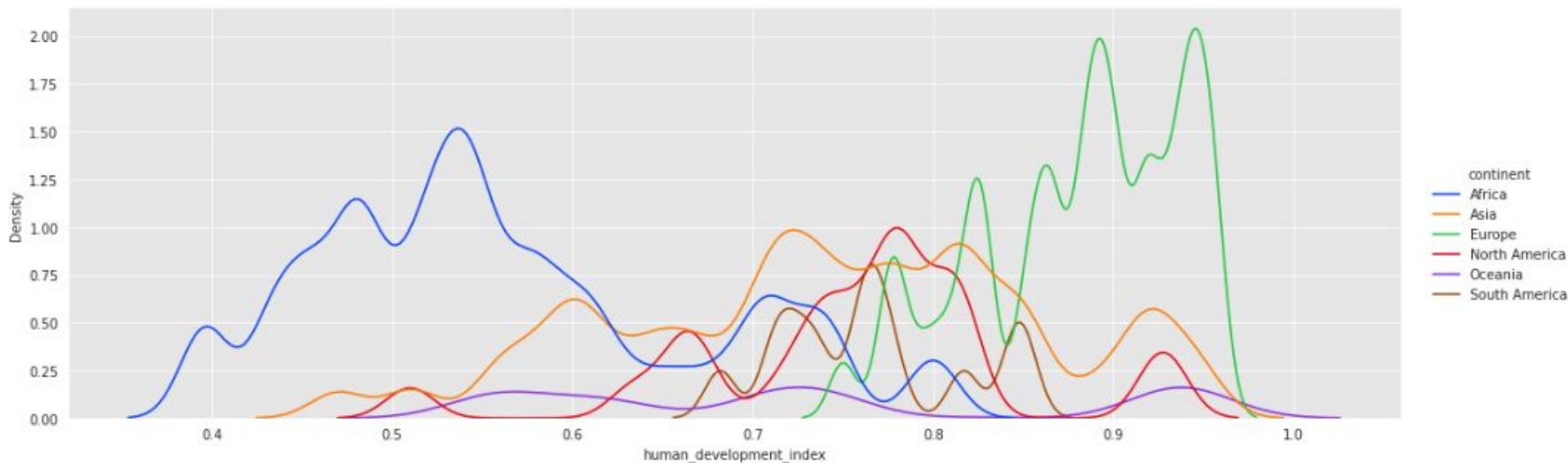
A continuación realizamos un heatmap y scatterplot para revisar la correlación de algunas variables, como en el primer paso de nuestro análisis, sólo que esta vez lo realizamos con todos los datos del dataset (a nivel mundial) y con algunas variables más que deseamos ver cómo se comportan entre ellas.



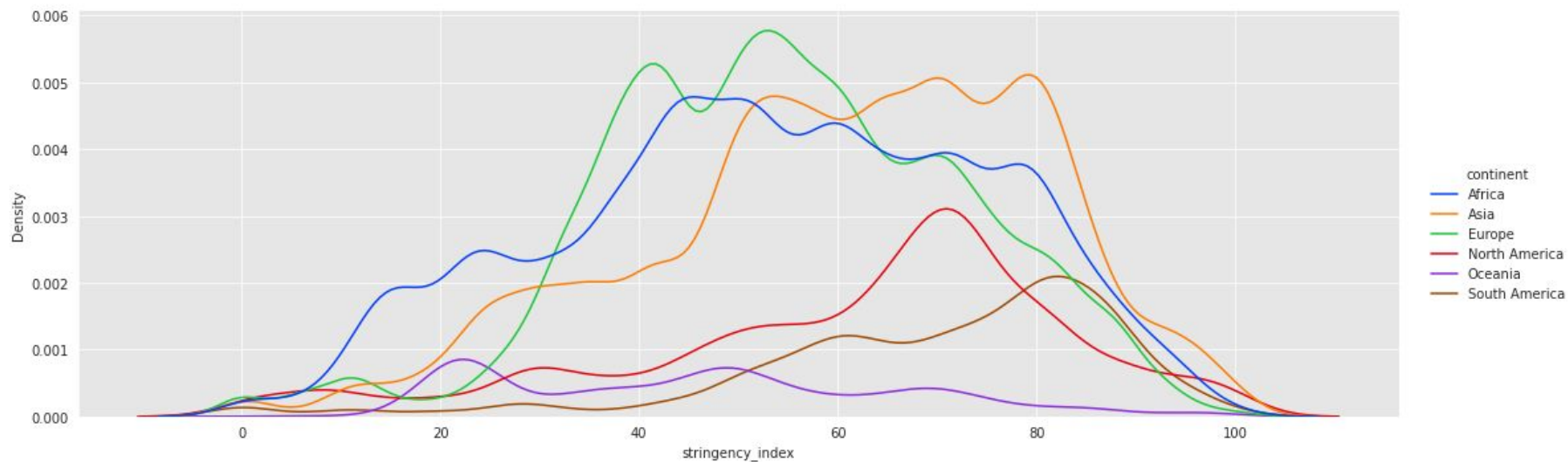


Posteriormente analizamos los datos esta vez por continente. Nos interesamos sobre datos en los que no habíamos profundizado tanto anteriormente: índice de desarrollo humano, índice de rigurosidad de los gobiernos y esperanza de vida.

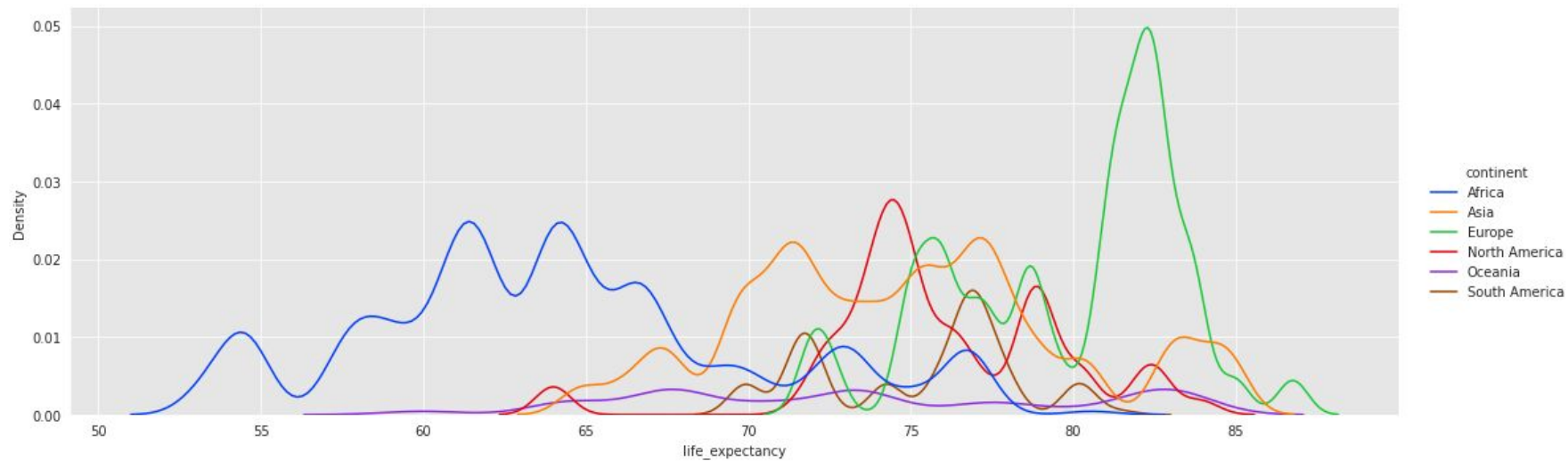
```
[ ] datos_por_continente_sobre('human_development_index')
```



```
[ ] datos_por_continente_sobre('stringency_index')
```



```
[ ] datos_por_continente_sobre('life_expectancy')
```





# Repositorio

- [https://github.com/salvador-vm/equipo\\_10-analisis\\_de\\_datos](https://github.com/salvador-vm/equipo_10-analisis_de_datos)

NOTA: Por motivos del tamaño de nuestro archivo original, no pudimos subirlo a nuestro repositorio. No obstante, el enlace mostrado a continuación permite obtener el dataset actualizado que utilizamos en nuestro proyecto -> <https://covid.ourworldindata.org/data/owid-covid-data.csv>