Librería: faker

Cosas a tener en cuenta para la creación de valores de ventas:

* Sus ventas globales: [Burger King's revenue 2022 | Statista](https://www.statista.com/statistics/266462/burger-king-revenue/)
  + 2018: 1,65
  + 2019: 1,78
  + 2020: 1,60
  + 2021: 1,81
  + 2022: 1,89

Convierto 2018 en índice base:

* + 2018: 1
  + 2019:1,78/1,65= 1,0787
  + 2020: 1,60/1,65= 0,9696
  + 2021: 1,81/1,65= 1,0969
  + 2022: 1,89/1,65= 1,1454
* Ventas nacionales: [Burger King Spain SLU: facturación anual 2014-2021 | Statista](https://es.statista.com/estadisticas/878626/valor-de-ventas-de-burger-king-spain-sl/)
* Índices estacionales:

Productos:

Precios:

* Big Mac Index <https://www.economist.com/big-mac-index>

<https://github.com/theeconomist/big-mac-data/releases/tag/2023-07>

<https://github.com/TheEconomist/big-mac-data>

GDP big mac:

* + Jan 2018: 38879,170318553
  + Jul 2018: 38095,5079972119
  + Jan 2019: 39939,26239
  + Jul 2019: 39645,59134
  + Jan 2020: 41397,06009
  + Jul 2020: 40512,0873041761
  + Jan 2021: 38808,437231472
  + Jul 2021: 38761,0785100852
  + Jan 2022: 41313,9087410665
  + Jul 2022: 40127,5176938427
  + Jan 2023: 42843,7289368498
  + Jul 2023: 41020,8592276786
* Ipc en España: <https://www.ine.es/prensa/ipc_tabla.htm>
* Scrappeo de: <https://www.ubereats.com/es/store/burger-king-plaza-espana/MxY7nva5QBabCHsELg6Qtg>

actua como un cientifico de datos que sabe programar en python. Generame el codigo que necesito para generar un numero aleatorio de venta\_euros con una media de 13,1526 y una desviación tipica de 9,7476:

import random

# Generar un número aleatorio con media de 13.1526 y desviación estándar de 9.7476

venta\_euros = random.normalvariate(13.1526, 9.7476)

print(venta\_euros)

Medidas de dispersión de venta\_euros:

* Media: 13.552564
* Desviación estándar: 9.976228
* Min: -11.86
* 25%: 6.2
* 50%: 11.28
* 75%: 19.465
* Max: 63.95

Quiero crear un dataframe en Python que contenga las columnas:

* id\_ticket: que empiece en el valor 1 y siga sumando 1 para cada valor.
* Canal\_venta: quiero que se componga de 4 valores únicos: kiosco, autoking, mostrador, comida de empleado. Para esta variable quiero que ‘kiosco’ tenga un peso entre 48,3366% y 49,7992%, ‘autoking’ tenga un peso entre 27,5930% y 24,4980%, ‘mostrador’ tenga un peso entre 21,3307% y 24,2972% y ‘comida de empleado’ tenga un peso entre 2,7397% y 1,4056%. Este peso se debe cumplir para el total del dataframe y aproximadamente para cada año.
* Fecha: quiero que sean fechas desde el 01-01-2018 hasta el 31-12-2022.
* Venta\_euros: Quiero que tengan como parámetros de dispersión estos valores: Media: 13.552564, Desviación estándar: 9.976228, Min: -11.86, 25%: 6.2, 50%: 11.28, 75%: 19.465, Max: 63.95. Además, quiero que respecto al canal de fecha se cumplan que las venta\_euros por fecha cumplan este índice (con año base 2018) de ingresos (ingresos=suma por año de venta\_euros) 2018: 1; 2019:1,0787; 2020: 0,9696; 2021: 1,0969, 2022: 1,1454. Además, quiero que sean estacionales estos ingresos por cada año, quiero que sigan un índice para la estacionalidad de los meses según el valor que tengan en columna ‘fecha’; el índice para la estacionalidad es la siguiente: enero: 1,17; febrero: 1,06; marzo: 1,01; abril: 1,06; mayo: 1,04; junio: 0,88; julio: 0,8; agosto:0,77; septiembre: 0,91; octubre:1; noviembre: 1,09; diciembre: 1,2.

Además, si siguen una distribución normal las columnas sería mejor.

siguientes columnas: id\_ticket, id\_producto, fecha, unidades, precio.

Primero crear las unidades, estas quiero que tengan las siguientes medidas de distribución: media: 1.4757, desviación estándar: 0.968754, min: -3.0, 25%: 1, 50%: 1, 75%: 2, max: 32.

Lo siguiente que quiero es la columna fechas: quiero que sean fechas desde el 01-01-2018 hasta el 31-12-2022.

El id\_producto quiero que contenga números aleatorios, siendo el valor máximo 162 y el mínimo 1.

Para crear la columna precio, según el id\_producto que tenga, cóger su precio correspondiente del dataframe llamado precios2023\_df de la columna precio. Precios2023\_df tiene una columna llamada id\_producto que se puede relacionar con la columna id\_producto que hemos creado en este df.

Para terminar, el id\_ticket, quiero que contenga todos los valores entre 1 y 10000 y se repitan. Se debe cumplir que la suma de los precios por cantidad de los id\_ticket iguales tengan como parámetros de dispersión estos valores: Media: 13.552564, Desviación estándar: 9.976228, Min: -11.86, 25%: 6.2, 50%: 11.28, 75%: 19.465, Max: 63.95. Además, quiero que respecto al canal de fecha se cumplan que la suma de los precios por cantidad de los id\_ticket iguales por fecha cumplan este índice (con año base 2018) de ingresos (ingresos=suma por año de venta\_euros) 2018: 1; 2019:1,0787; 2020: 0,9696; 2021: 1,0969, 2022: 1,1454. Además, quiero que sean estacionales estos ingresos (suma de los precios por cantidad de los id\_ticket iguales) por cada año, quiero que sigan un índice para la estacionalidad de los meses según el valor que tengan en columna ‘fecha’; el índice para la estacionalidad es la siguiente: enero: 1,17; febrero: 1,06; marzo: 1,01; abril: 1,06; mayo: 1,04; junio: 0,88; julio: 0,8; agosto:0,77; septiembre: 0,91; octubre:1; noviembre: 1,09; diciembre: 1,2.

Para hacer tabla clientes:

* fake.name()
* fake.address()
* fake.city()
* fake.country()
* fake.zipcode()
* fake.email()
* fake.phone\_number()