

# Procesamiento de Datos con Python 2020 - BEDU

ALUMNO: Abdiel Céspedes Jaén  
EXPERTO: José Ramón Arias  
GRUPO: data-analysis-gdl-20-06

**Reinterpretación de procesos  
de BBDD usando Python**



# Objetivos

## Objetivo del curso

*“Al finalizar el módulo serás capaz de utilizar Python para adquirir datos ya sea de archivos locales, APIs o Bases de Datos con el objetivo de limpiarlos, manejarlos, explorarlos para dejarlos de manera que queden en condiciones óptimas para su análisis y visualización.”*

## Objetivo del proyecto

Resolver las siguientes preguntas:

- ¿Qué clientes ya han sido satisfechos al haber registrado su estatus como “entregado”?
- ¿Cuántas veces ha aparecido el método de pago de tarjeta de crédito?
- ¿Cuántos son los pagos que están por encima del promedio de su propio valor?
- ¿Calcula la frecuencia de cada tipo de pago?

## Contexto

La base de datos proviene de una tienda ecommerce brasileña anónima, proveída por [www.olist.com](http://www.olist.com). Cuenta con información múltiples dimensiones:

- Estatus de orden.
- Precio.
- Pago.
- Atributos del producto.
- Locaciones del consumidor.
- Reseñas del consumidor.
- Etc.

The Olist logo consists of the word "olist" in a white, lowercase, sans-serif font, centered within a solid blue square.

olist



## Pregunta 1. Clientes con estatus de “entregado”

```
[90] a = join_customers_orders_dropped['order_status'] == 'delivered'  
a
```

```
0      True  
1      True  
2      True  
3      True  
4      True  
...  
995    False  
996    False  
997    False  
998    False  
999    False  
Name: order_status, Length: 1000, dtype: bool
```

```
[90] a.sum()
```

```
487
```

## Pregunta 2. Número de veces en que usaron tarjeta de crédito

```
[108] b = join_customers_payments['payment_type'] == 'credit_card'  
      b
```

```
0      True
```

```
1      True
```

```
2      True
```

```
3      True
```

```
4      True
```

```
...
```

```
995    False
```

```
996    False
```

```
997    False
```

```
998    False
```

```
999    False
```

```
Name: payment_type, Length: 1000, dtype: bool
```

```
[109] b.sum()
```

```
366
```

### Pregunta 3. Pagos por encima de la media del valor

```
[111] media = np.mean(payments['payment_value'])  
media
```

```
141.84510000000003
```

```
[112] c = payments['payment_value'] > media  
c
```

```
0      False  
1       True  
2       True  
3      False  
4       True
```

```
...
```

```
495    False  
496    False  
497    False  
498    False  
499    False
```

```
Name: payment_value, Length: 500, dtype: bool
```

```
[113] c.sum()
```

```
166
```

## Pregunta 4. Valor de pago por cada orden

```
c.drop(columns=['payment_sequential', 'payment_installments', 'payment_value'])
```

order_id	
payment_type	
boleto	97
credit_card	366
debit_card	8
not_defined	1
voucher	28



## Conclusión

Gracias al lenguaje de programación de Python y, en particular, a su paquetería “Pandas” orientada al análisis de BBDD, es que se pudo dar tratamiento y presentación al conjunto de datos que a su vez ya había sido gestionados por primera vez con un SGBD como MySQL y MongoDB. Así pues, puedo asegurar que Python me ha brindado las herramientas necesarias para recolectar datos precisos y para gestionar un buen análisis con el propósito de dar solución a una problemática.





# Créditos de las imágenes & de bases de datos

## Imágenes

- Desconocido. (1 de Junio de 2020). *Para el mercado, el PIB de Brasil caería 6,25%*. Recuperado el 30 de Octubre de 2020, de El Economista: <https://eleconomista.com.ar/2020-06-para-el-mercado-el-pib-de-brasil-caeria-625/>
- Desconocido. (s.f.). *[Imagen]*. Recuperado el 30 de Octubre de 2020, de <https://lh3.googleusercontent.com/leb2NQOJ7aqEyeyrsAtRpRURhAacviecKvWvUh8NIBOk3M8KT6Dr0zM0IXoW6WHYs78>

## Base de datos original

- kaggle. (s.f.). *Brazilian E-Commerce Public Dataset by Olist*. Recuperado el 27 de Octubre de 2020, de <https://www.kaggle.com/olistbr/brazilian-ecommerce>

## Base de datos utilizada

La base de datos utilizada para la realización de este proyecto se redujo de a 1000 registros/documentos para la primer tabla y a 500 para las demás, esto con el fin de facilitar el proceso de carga de datos a la base de datos local.