



Test técnico – Data Engineer

Importación de motocicletas

Se pide desarrollar un proceso para extracción, transformación y carga de vehículos. Se dispone de dos orígenes de datos:

Fuente de datos de vehículos (excel):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2	count	Fecha_Comp_id_ozon	pais	Semana	Ozocio o Re	Ozocio/Retomador	RFC	marca	modelo	año	cilindraje	Color	Valor Factura	num_motor	num_tarjeta_circ_o_Licencia	
3	1	7/21/2021 OCO001	colombia	6	Ozocio			Akt	nkdr 125	2021	124			157FMISE328496		1C
4	2	7/24/2021 OCO002	colombia	6	Ozocio			Akt	nkdr 125	2020	124			157FMISE225296		1C
5	3	6/16/2021 OMX001	mexico	1	Ozocio	Fernando		italika	ft150	2021	150	grafito		LC162FMIJTE100456		17
6	4	7/24/2021 OMX002	mexico	6	Ozocio	Rodrigo		vento	lithium 2.0	2020	150			KD162FMIJ20342416		4
7	5	8/13/2021 OMX003	mexico	9	Ozocio	Scarlett		vento	lithium 2.0	2020	150			KD162FMIJ20360266		4
8	6	8/16/2021 OMX004	mexico	10	Ozocio	Nancy		italika	ft150ts	2019	150			LC162FMIJSE126023		
9	7	8/16/2021 OMX005	mexico	10	Retomador	Eduardo		italika	dt150	2021	150			RW162FMIJ52100009817	NA	
10	8	8/16/2021 OMX006	mexico	10	Retomador	Eduardo		italika	ft125	2019	125			LC152FMIRE211412	NA	
11	9	8/21/2021 OMX007	mexico	10	Ozocio	Lalo		vento	ryder3.0	2021	150			KD162FMIJ20361040	NA	
12	10	9/22/2021 OMX008	mexico	15	Ozocio	Alfredo		italika	ft150	2021	150	grafito		LC162FMIJTE320423	NA	

Fuente de datos de marcas (csv):

1	id,createdAt,created_by,name,updatedAt,updated_by
2	61e791bc4192c77024f79180,2022-05-19T15:09:20.570Z,admin,VELOCI,2022-05-19T15:09:20.570Z,admin
3	61a09b616ce70c000afef618,2022-05-19T15:09:20.569Z,admin,ITALIKA,2022-05-19T15:09:20.569Z,admin
4	6290e37b261262f16831b02a,2022-05-19T15:09:20.569Z,admin,VENTO,2022-05-19T15:09:20.569Z,admin
5	62a24143cc8bdce214670ce1,2022-06-09T18:51:47.297Z,admin,KTM,2022-06-09T18:51:47.297Z,admin
6	62a2419ccc8bdce214670ce3,2022-06-09T18:53:16.632Z,admin,KURAZAI,2022-06-09T18:53:16.632Z,admin
7	62a26129cc8bdce214670d7f,2022-06-09T21:07:53.766Z,admin,BAJAJ,2022-06-09T21:07:53.766Z,admin
8	62a2615fcc8bdce214670d8e,2022-06-09T21:08:47.343Z,admin,BENELLI,2022-06-09T21:08:47.343Z,admin
9	62a2615fcc8bdce214670d92,2022-06-09T21:08:47.729Z,admin,CARABELA,2022-06-09T21:08:47.729Z,admin
10	62a26161cc8bdce214670da2,2022-06-09T21:08:49.298Z,admin,DINAMO,2022-06-09T21:08:49.298Z,admin
11	62a26162cc8bdce214670daa,2022-06-09T21:08:50.095Z,admin,HERO,2022-06-09T21:08:50.095Z,admin
12	62a26162cc8bdce214670db0,2022-06-09T21:08:50.668Z,admin,HONDA,2022-06-09T21:08:50.668Z,admin
13	62a2616acc8bdce214670e02,2022-06-09T21:08:58.597Z,admin,MB,2022-06-09T21:08:58.597Z,admin
14	62a2616bcc8bdce214670e0a,2022-06-09T21:08:59.342Z,admin,MOTOMEL,2022-06-09T21:08:59.342Z,admin

Construir un ETL en python que procese el Excel de vehículos y el CSV de Marcas y genere un JSON para poder insertar en batch en una base de datos MongoDB. El schema del JSON deberá ser el siguiente:



```
{
  "_id" : ObjectId("62aa54b9c66442472d27596d"),
  "salePrice" : 400,
  "details" : {
    "year" : 2021,
    "milage" : 8000
  },
  "oldPrice" : 500,
  "internalId" : "OMX001",
  "vehicleSN" : "3SCK2AEN0M1007310",
  "engineSN" : "LC162FMJTE100456",
  "purchaseCost" : 13000,
  "color" : "grafito",
  "cylindersCapacity" : 150,
  "brand" : ObjectId("61a09b616ce70c000afef618"),
  "createdAt" : ISODate("2022-06-15T21:52:57.364Z"),
  "updatedAt" : ISODate("2022-06-29T18:27:40.536Z"),
  "country" : 'mexico',
  "plate" : "1J9XT",
  "registrationCard" : "17950945837"
}
```

A continuación se detalla el mapeo de los campos:

	Excel	CSV	Observaciones
_id			Generar un ObjectId() para cada registro.
salePrice	cuota		En caso de existir "precio descuento", se deberá insertar la columna "cuota semanal descuento"
details.year	año		
details.milage	kilometraje_aprox		
oldPrice			En caso de existir "precio descuento", se deberá insertar la columna "cuota" sino 0.
internalId	id_ozon		
vehicleSN	serie_vehicular_o_num_chasis		
engineSN	num_motor		
purchaseCost	gasto_compra		
color	color		En caso de presentarse una combinación (ej. rojo/negro), se deberá tomar sólo el primero (ej. rojo).
cylindersCapacity	cilindraje		
brand		_id	Convertir a formato ObjectId() el "_id" del CSV matcheando por el texto del campo "marca" del Excel.
createdAt			Fecha en formato ISO de hoy.
updatedAt			Fecha en formato ISO de hoy.
country	país		
plate	placa		
registrationCard	num_tarjeta_circ		



La solución debe cumplir los siguientes requisitos:

- Generar el JSON a importar en la base de datos MongoDB sólo con las motos de México y que sean menores a la número 1000 ("OMX1000").
- Los valores que estén vacíos o no apliquen ("NA"), dejarlos en null.
- Subir la solución y el resultado a un repositorio en Git.