

Data and Schema Cleanup

Explanation and Setup

- There are several issues with the schema for `classicmodels`. Two of the issues are:
 - `customers.country`: Having programs or people enter country names is prone to errors.
 - `products.productCode` is clearly not an atomic value.
- The following SQL creates a schema with copies of the data. The SQL also loads a table of *ISO country codes*.

In [36]: `%sql create schema classicmodels_midterm;`

```
* mysql+pymysql://root:***@localhost
(pymysql.err.ProgrammingError) (1007, "Can't create database 'classicmodels_midterm';
database exists")
[SQL: create schema classicmodels_midterm;]
(Background on this error at: https://sqlalche.me/e/14/f405)
```

In [37]: `iso_df = pandas.read_csv('./wikipedia-iso-country-codes.csv')`
`iso_df.to_sql('country_codes', schema='classicmodels_midterm',`
`con=sql_engine, index=False, if_exists="replace")`

Out[37]: 246

In [38]: `%%sql`

```
use classicmodels_midterm;

alter table classicmodels_midterm.country_codes
change `English short name` lower case `short_name` text null;

alter table classicmodels_midterm.country_codes
change `Alpha-2 code` alpha_2_code text null;

alter table classicmodels_midterm.country_codes
change `Alpha-3 code` alpha_3_code text null;

alter table classicmodels_midterm.country_codes
change `Numeric code` numeric_code bigint null;

alter table classicmodels_midterm.country_codes
change `ISO 3166-2` iso_text text null;
```

```
* mysql+pymysql://root:***@localhost
0 rows affected.
0 rows affected.
0 rows affected.
0 rows affected.
0 rows affected.
0 rows affected.
```

Out[38]: []

In [39]: `%%sql`

```
use classicmodels_midterm;

drop table if exists customers;
create table customers as select * from classicmodels.customers;

drop table if exists products;
create table products as select * from classicmodels.products;
```

```
* mysql+pymysql://root:***@localhost
0 rows affected.
0 rows affected.
122 rows affected.
0 rows affected.
110 rows affected.
```

Out[39]: []

DE-1

Question

- There are four `country` values in `customers` that are not in `short_names` of `country_codes`.
- The four missing values are:

<code>country</code>
USA
Norway
UK
Russia

- Write an SQL query that returns the information about by querying `customers` and `country_codes`

Answer

In [40]: `%%sql`
`SELECT distinct(customers.country)`
`FROM classicmodels_midterm.customers`

```
LEFT JOIN classicmodels_midterm.country_codes ON customers.country = country_codes.short_name
WHERE country_codes.short_name IS NULL;
```

```
* mysql+pymysql://root:***@localhost
4 rows affected.
```

Out[40]: **country**

USA

Norway

UK

Russia

DE-2

Question

- Norway is on the list because there are spaces in the entry. The following query shows this fact.

In [41]: **%%sql**

```
select customerNumber, customerName, country
from customers where length(country) != length(trim(country));
```

```
* mysql+pymysql://root:***@localhost
2 rows affected.
```

Out[41]: **customerNumber** **customerName** **country**

167	Herkku Gifts	Norway
299	Norway Gifts By Mail, Co.	Norway

- The mapping of the other country names is:

customers.country	country_codes.short_name
USA	United States
UK	United Kingdom
Russia	Russian Federation

- Write a **single** **update** statement that corrects the values for **customers.country**.

Answer

In [42]: **%%sql**

```
UPDATE classicmodels_midterm.customers
SET country =
CASE
    WHEN country = 'USA' THEN 'United States'
    WHEN country = 'UK' THEN 'United Kingdom'
```

```

    WHEN country = 'Russia' THEN 'Russian Federation'
    WHEN country = 'Norway' THEN 'Norway'
    ELSE country
END;

```

```

* mysql+pymysql://root:***@localhost
122 rows affected.

```

Out[42]: []

DE-3

Question

- The final tasks are:
 - Add a column `iso_code` to `customers` that is the `alpha_2_code` from `country_codes`.
 - Create a foreign key relationship `customers.iso_code -> country_codes.alpha_2_code`.
 - Drop `country` from `customers`.
 - Create a view `customers_country` of the form `(customerNumber, customerName, country, iso_code)`.

Answer

In [43]: %%sql

```

/*
Update classicmodels_midterm.country_codes set alpha_2_code='NA' where short_name='Nan

ALTER TABLE classicmodels_midterm.customers ADD COLUMN iso_code VARCHAR(2);

UPDATE classicmodels_midterm.customers
SET customers.iso_code = (
    SELECT country_codes.alpha_2_code
    FROM country_codes
    WHERE country_codes.short_name = customers.country
);

ALTER TABLE customers ADD FOREIGN KEY(iso_code) references country_codes(alpha_2_code)

ALTER TABLE customers DROP COLUMN country;

create or replace view customers_country as
select
    customerNumber, customerName, country, iso_code
from
    customers;
*/

```

```

* mysql+pymysql://root:***@localhost
0 rows affected.

```

Out[43]: []

In [44]: **%%sql**

```
/* Write a SELECT that displays 25 customers sorted by customerName */
```

```
SELECT *  
FROM customers  
ORDER BY customerName  
LIMIT 25;
```

```
* mysql+pymysql://root:***@localhost  
25 rows affected.
```

Out[44]:

customerNumber	customerName	contactLastName	contactFirstName	phone	addressLine1	add
----------------	--------------	-----------------	------------------	-------	--------------	-----

242	Alpha Cognac	Roulet	Annette	61.77.6555	1 rue Alsace-Lorraine	
-----	--------------	--------	---------	------------	-----------------------	--

168	American Souvenirs Inc	Franco	Keith	2035557845	149 Spinnaker Dr.	
-----	------------------------	--------	-------	------------	-------------------	--

249	Amica Models & Co.	Accorti	Paolo	011-4988555	Via Monte Bianco 34	
-----	--------------------	---------	-------	-------------	---------------------	--

237	ANG Resellers	Camino	Alejandra	(91) 745 6555	Gran Vía, 1	
-----	---------------	--------	-----------	---------------	-------------	--

276	Anna's Decorations, Ltd	O'Hara	Anna	02 9936 8555	201 Miller Street	
-----	-------------------------	--------	------	--------------	-------------------	--

465	Anton Designs, Ltd.	Anton	Carmen	+34 913 728555	c/ Gobelas, 19-1 Urb. La Florida	
-----	---------------------	-------	--------	----------------	----------------------------------	--

206	Asian Shopping Network, Co	Walker	Brydey	+612 9411 1555	Suntec Tower Three	8
-----	----------------------------	--------	--------	----------------	--------------------	---

348	Asian Treasures, Inc.	McKenna	Patricia	2967 555	8 Johnstown Road	
-----	-----------------------	---------	----------	----------	------------------	--

103	Atelier graphique	Schmitt	Carine	40.32.2555	54, rue Royale	
-----	-------------------	---------	--------	------------	----------------	--

471	Australian Collectables, Ltd	Clenahan	Sean	61-9-3844-6555	7 Allen Street	
-----	------------------------------	----------	------	----------------	----------------	--

114	Australian Collectors, Co.	Ferguson	Peter	03 9520 4555	636 St Kilda Road	
-----	----------------------------	----------	-------	--------------	-------------------	--

333	Australian Gift Network, Co	Calaghan	Ben	61-7-3844-6555	31 Duncan St. West End	
-----	-----------------------------	----------	-----	----------------	------------------------	--

256	Auto Associés & Cie.	Tonini	Daniel	30.59.8555	67, avenue de l'Europe	
-----	----------------------	--------	--------	------------	------------------------	--

406	Auto Canal+ Petit	Perrier	Dominique	(1) 47.55.6555	25, rue Lauriston	
-----	-------------------	---------	-----------	----------------	-------------------	--

198	Auto-Moto Classics Inc.	Taylor	Leslie	6175558428	16780 Pompton St.	
-----	-------------------------	--------	--------	------------	-------------------	--

187	AV Stores, Co.	Ashworth	Rachel	(171) 555-1555	Fauntleroy Circus	
-----	----------------	----------	--------	----------------	-------------------	--

121	Baane Mini Imports	Bergulfsen	Jonas	07-98 9555	Erling Skakkes gate 78	
-----	--------------------	------------	-------	------------	------------------------	--

415	Bavarian Collectables Imports, Co.	Donnermeyer	Michael	+49 89 61 08 9555	Hansastr. 15	
-----	------------------------------------	-------------	---------	-------------------	--------------	--

293	BG&E Collectables	Harrison	Ed	+41 26 425 50 01	Rte des Arsenaux 41	
-----	-------------------	----------	----	------------------	---------------------	--

128	Blauer See Auto, Co.	Keitel	Roland	+49 69 66 90 2555	Lyonerstr. 34	
-----	----------------------	--------	--------	-------------------	---------------	--

219	Boards & Toys Co.	Young	Mary	3105552373	4097 Douglas Av.	
344	CAF Imports	Fernandez	Jesus	+34 913 728 555	Merchants House	^
173	Cambridge Collectables Co.	Tseng	Jerry	6175555555	4658 Baden Av.	
202	Canadian Gift Exchange Network	Tamuri	Yoshi	(604) 555-3392	1900 Oak St.	
339	Classic Gift	Cervantes	Francisco	2155554695	782 First	

DE-4

- *Just kidding.*
- *My first intent was to have you fix `products`.*
- *Then, I thought I would make this an extra credit question.*
- *Finally, I decided that all students get 5 points added to there score for this exam. Since I never "curve up," you all get a bonus on final grade for putting up with the class.*

```
In [45]: import pandas as pd
df = pd.read_csv("C:/Users/11139/Desktop/W4111-02/Midterm/pitching.csv")
a = []
for i in df.columns:
    a.append(i)
print(a)

['playerID', 'yearID', 'stint', 'teamID', 'lgID', 'W', 'L', 'G', 'GS', 'CG', 'SHO',
'SV', 'IPouts', 'H', 'ER', 'HR', 'BB', 'SO', 'BAOpp', 'ERA', 'IBB', 'WP', 'HBP', 'B
K', 'BFP', 'GF', 'R', 'SH', 'SF', 'GIDP']
```