Data cleaning about Lauren’s Furniture Store Customer Data Set

# Add a new customer’s info into data set#

Insert into customer\_data.customer\_address

(customer\_id, address, city, state, zipcode, country)

Values

(2332, ’333 queen Road’, ‘Jackson’, ‘MI’, 69930, ‘US’)

#Update Customer’s address #

Update customer\_data.customer\_address

Set address= ’45 Victoria St.’

Where customer\_id = 4654

#Create table if not exists#

#Drop table if exists#

#Get all customers’ ID from data set#

Select distinct customer\_ID

From customer\_data.customer\_address

#Make sure all country code has the same length#

Select length(country) as letters\_in\_country

From customer\_data.customer\_address

#Filtering those countries which has more than 2 letters#

Select length(country)

From customer\_data.customer\_address

Where length(country) > 2

#Query all customer ID from US/USA and remove all duplicates#

Select distinct customer\_id

From customer\_data.customer\_address

Where subtr(country,1,2) = ‘US’

#Eliminate spaces for consistency, find all customer IDs who live in Ohio#

Select distinct customer\_id

From customer\_data.customer\_address

Where trim(state) = ‘OH’

# order the purchase price by descending order#

Select purchase\_price

From customer\_data.customer\_address

Order by purchase\_price desc

#Converting data type correctly#

Select cast(purchase\_price as float)

From customer\_data.customer\_address

Order by cast (purchase\_price as float) desc

#Query purchases that occurred in December#

Select cast(date as date), purchase\_price

From customer\_data.customer\_purchase

Where date between ‘2020-12-01’ and ‘2020-12-31’

#Create a unique key to find which color couches are the most popular and order more #

 Select concat(product\_code, product\_color) as new\_product\_code

From customer\_data.customer\_purchase

Where product = ‘couch’

#If the product is not available would show the product code#

Select coalesce (product, product\_code) as product\_info

From customer\_data.customer\_purchase

#Correct misspelling customers’  names #

Select customer\_id,

Case

when first\_name = ‘Tnoy’ then ‘Tony’

when first\_name=’Tmo’ then ‘Tom’

Else first\_name

End as cleaned\_name

From

Customer\_data.customer\_name