Topic 03:
Data Alchemist
Data Aggregation

BDM3302: Data Management

Why Data Alchemist?

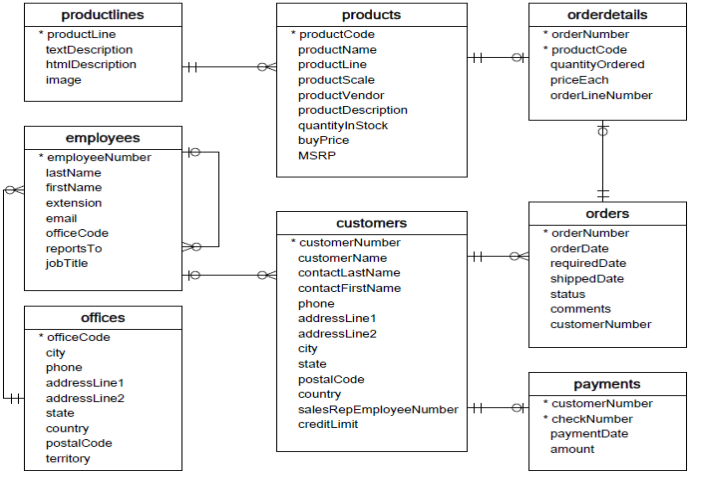
- The applied of querying data from what we have learned basic SQL
- Still use MySQL and Pop SQL as tools to continue
- In real world business, data alchemist is focusing on transforming big data with advanced analytics technique into a successful business formula (Source: Techsauce)
- Data alchemist will work seamlessly with other business units under the “Turning Data into Gold: the Ultimate of Advanced Analytics” (Source: Techsauce)
- Our class objective is to let you be in a novice level for data alchemist

Data Alchemist Key Method

- Sorting เรียงลำดับ order by
- Filtering การกรอง where
- Projection การฉายภาพ มากกว่า 1 filter

Remind your Tools for Data Alchemist

- MySQL
- PopSQL
- classicmodels database



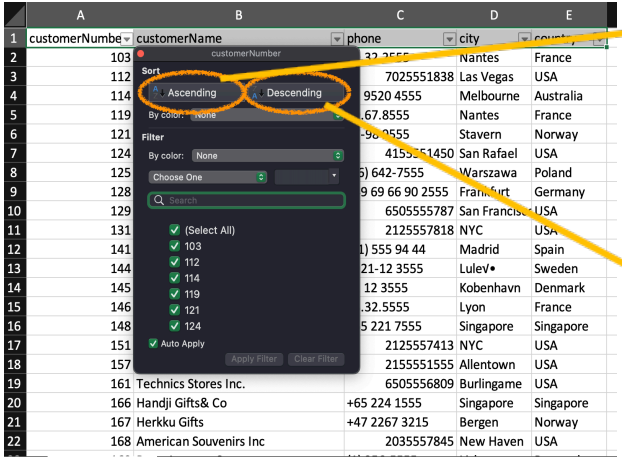
Data Alchemist - Sorting เรียงลำดับ

- Arranging the records in a specific way to make reported data more usable
- Sort records by choosing a specific field(s) within a record by which to sort
- For example, an alphabetical sort by the last name field will arrange text data in ascending alphabetical (A-Z) order
- If specified, the text fields can also be sorted in descending (Z-A) order

Data Alchemist - Sorting

Can you see the pain point of this?

- Sorting as you normally do in Excel (using Filter)



	A	B	C	D	E
1	customerNumbe	customerName	phone	city	country
2	103		32.2555	Nantes	France
3	112		7025551838	Las Vegas	USA
4	114		9520 4555	Melbourne	Australia
5	119		.67.8555	Nantes	France
6	121		-96 0555	Stavern	Norway
7	124		4155551450	San Rafael	USA
8	125		5) 642-7555	Warszawa	Poland
9	128		9 69 66 90 2555	Frankfurt	Germany
10	129		6505555787	San Francis	USA
11	131		2125557818	NYC	USA
12	141		1) 555 94 44	Madrid	Spain
13	144		21-12 3555	Lulev•	Sweden
14	145		12 3555	Kobenhavn	Denmark
15	146		.32.5555	Lyon	France
16	148		5 221 7555	Singapore	Singapore
17	151		2125557413	NYC	USA
18	157		2155551555	Allentown	USA
19	161	Technics Stores Inc.	6505556809	Burlingame	USA
20	166	Handji Gifts& Co	+65 224 1555	Singapore	Singapore
21	167	Herkku Gifts	+47 2267 3215	Bergen	Norway
22	168	American Souvenirs Inc	2035557845	New Haven	USA

customerNumbe	customerName	phone	city	country
103	Atelier graphique	40.32.2555	Nantes	France
112	Signal Gift Stores	7025551838	Las Vegas	USA
114	Australian Collectors, Co.	03 9520 4555	Melbourne	Australia
119	La Rochelle Gifts	40.67.8555	Nantes	France
121	Baane Mini Imports	07-98 9555	Stavern	Norway
124	Mini Gifts Distributors Ltd.	4155551450	San Rafael	USA
125	Havel & Zbyszek Co	(26) 642-7555	Warszawa	Poland
128	Blauer See Auto, Co.	+49 69 66 90 2555	Frankfurt	Germany

customerNumbe	customerName	phone	city	country
249	Amica Models & Co.	011-4988555	Torino	Italy
247	Messner Shopping Network	069-0555984	Frankfurt	Germany
242	Alpha Cognac	61.77.6555	Toulouse	France
240	giftsbymail.co.uk	(198) 555-8888	Cowes	UK
239	Collectable Mini Designs Co.	7605558146	San Diego	USA
237	ANG Resellers	(91) 745 6555	Madrid	Spain
233	Quv©bec Home Shopping Network	(514) 555-8054	Montrv©al	Canada
227	Heintze Collectables	86 21 3555	vÖrhus	Denmark
223	Natvºrlich Autos	0372-555188	Cunewalde	Germany
219	Boards & Toys Co.	3105552373	Glendale	USA
216	Enaco Distributors	(93) 203 4555	Barcelona	Spain
211	King Kong Collectables, Co.	+852 2251 1555	Central Hong	Hong Kong
209	Mini Caravy	88.60.1555	Strasbourg	France
206	Asian Shopping Network, Co	+612 9411 1555	Singapore	Singapore
205	Toys4GrownUps.com	6265557265	Pasadena	USA
204	Online Mini Collectables	6175557555	Brickhaven	USA
202	Canadian Gift Exchange Network	(604) 555-3392	Vancouver	Canada
201	UK Collectables, Ltd.	(171) 555-2282	Liverpool	UK

Data Alchemist - Sorting

- Sorting in SQL
  - **order by fieldname asc;** – Sorting texts or numbers in an ascending order (e.g., A-Z, 0-9)

```
select *
from customers
order by customerNumber asc;
```

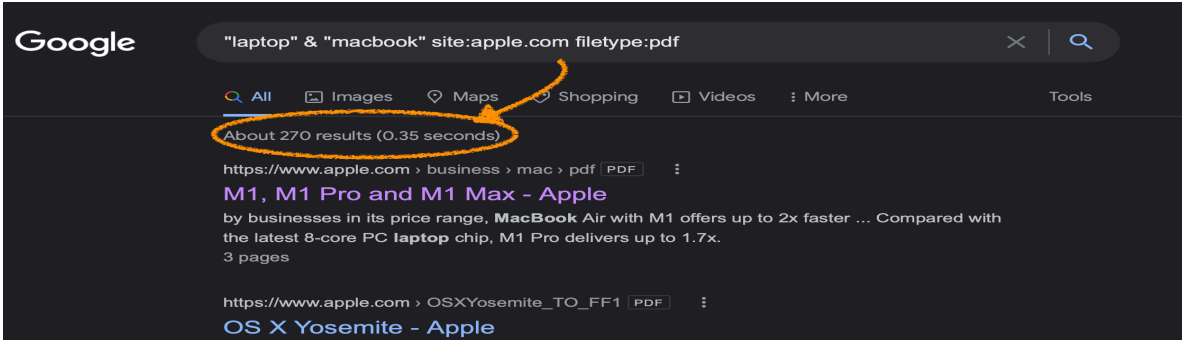
  - **order by fieldname desc;** – Sorting texts or numbers in a descending order (e.g., Z-A, 9-0)

```
select *
from customers
order by customerNumber desc;
```

1. What is customerNumber?
2. Can we change from customerNumber to another one for sorting? Try for it!

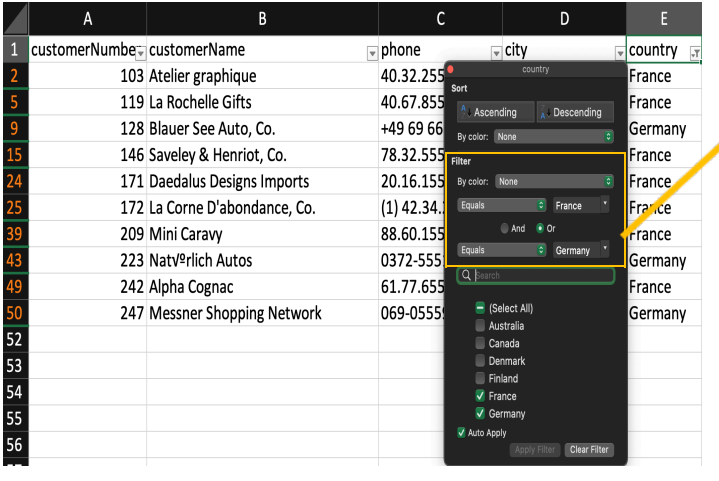
Data Alchemist - Filtering

- Filtering is the method that you use some commands with conditions to specify a subset of the data items
- The benefit is to help you to get the exactly result with less data records (like using the keyword to search command in Google)



Data Alchemist - Filtering

- Filtering as you normally do in Excel (using Filter)



	A	B	C	D	E
1	customerNumbe	customerName	phone	city	country
2	103	Atelier graphique	40.32.255		France
5	119	La Rochelle Gifts	40.67.855		France
9	128	Blauer See Auto, Co.	+49 69 66		Germany
15	146	Saveley & Henriot, Co.	78.32.555		France
24	171	Daedalus Designs Imports	20.16.155		France
25	172	La Corne D'abondance, Co.	(1) 42.34.		France
39	209	Mini Caravy	88.60.155		France
43	223	Natvºrlich Autos	0372-555		Germany
49	242	Alpha Cognac	61.77.655		France
50	247	Messner Shopping Network	069-0555		Germany
52					
53					
54					
55					
56					

customerNumbe	customerName	phone	city	country
103	Atelier graphique	40.32.2555	Nantes	France
119	La Rochelle Gifts	40.67.8555	Nantes	France
128	Blauer See Auto, Co.	+49 69 66 90 2555	Frankfurt	Germany
146	Saveley & Henriot, Co.	78.32.5555	Lyon	France
171	Daedalus Designs Imports	20.16.1555	Lille	France
172	La Corne D'abondance, Co.	(1) 42.34.2555	Paris	France
209	Mini Caravy	88.60.1555	Strasbourg	France
223	Natvºrlich Autos	0372-555188	Cunewalde	Germany
242	Alpha Cognac	61.77.6555	Toulouse	France
247	Messner Shopping Network	069-0555984	Frankfurt	Germany

Can you see the pain point of this?

Data Alchemist - Filtering

- Filtering in SQL
  - **where fieldname operator value;**
    - Specify one filter condition

```
select customerNumber, customerName,
contactFirstName, phone
from customers
where contactFirstName = "Rachel";
```

  - **where fieldname operator value [and | or] fieldname operator value;**
    - Specify two filter conditions

```
select customerNumber, customerName, contactFirstName, phone, country
from customers
where country = "Japan" or country = "UK";
```

Operator:

2. **> (greater than)**
3. **< (less than)**
4. **= (equal)**
5. **>= (greather than or equal)**
6. **<= (less than or equal)**
7. **!= (not equal)**
8. **like (equal)**

Make sure that two filter conditions should be in the same fieldname that can display the result …

Data Alchemist - Filtering

- Filtering in SQL
  - **where fieldname operator value [and | or] fieldname operator value [and | or] fieldname operator value;**
    - Specify three filter conditions

```
select customerNumber, customerName, contactFirstName, phone, country
from customers
where country = "France" and customerNumber >= 100 and customerNumber <= 200;
```

Try to change “and” to “or” or input parenthesis “(…)” in each condition and see the different results.

Data Alchemist - Projection

- Projection is defined as taking a vertical subset from the columns or fieldnames of a single table that retains the unique rows or records
- This kind of SELECT statement returns some of the columns or fieldnames and all the rows or records in a table
- Projection is also implemented through the projection list in the Projection clause of a SELECT statement

Data Alchemist - Projection

- Projection as you normally do in Excel

	A	B	C	D	E
1	customerNumbe	customerName	phone	city	country
2	103	Atelier graphique	40.32.2555	Nantes	France
3	112	Signal Gift Stores	7025551838	Las Vegas	USA
4	114	Australian Collectors, Co.	03 9520 4555	Melbourne	Australia
5	119	La Rochelle Gifts	40.67.8555	Nantes	France
6	121	Baane Mini Imports	07-98 9555	Stavern	Norway
7	124	Mini Gifts Distributors Ltd.	4155551450	San Rafael	USA
8	125	Havel & Zbyszek Co	(26) 642-7555	Warszawa	Poland
9	128	Blauer See Auto, Co.	+49 69 66 90 2555	Frankfurt	Germany
10	129	Mini Wheels Co.	6505555787	San Francisco	USA
11	131	Land of Toys Inc.	2125557818	NYC	USA
12	141	Euro+ Shopping Channel	(91) 555 94 44	Madrid	Spain
13	144	Volvo Model Replicas, Co	0921-12 3555	Luleå	Sweden
14	145	Danish Wholesale Imports	31 12 3555	Kobenhavn	Denmark
15	146	Saveley & Henriot, Co.	78.32.5555	Lyon	France
16	148	Dragon Souveniers, Ltd.	+65 221 7555	Singapore	Singapore
17	151	Muscle Machine Inc	2125557413	NYC	USA

customerNumber	customerName	country
103	Atelier graphique	France
112	Signal Gift Stores	USA
114	Australian Collectors, Co.	Australia
119	La Rochelle Gifts	France
121	Baane Mini Imports	Norway
124	Mini Gifts Distributors Ltd.	USA
125	Havel & Zbyszek Co	Poland
128	Blauer See Auto, Co.	Germany
129	Mini Wheels Co.	USA
131	Land of Toys Inc.	USA
141	Euro+ Shopping Channel	Spain
144	Volvo Model Replicas, Co	Sweden
145	Danish Wholesale Imports	Denmark
146	Saveley & Henriot, Co.	France
148	Dragon Souveniers, Ltd.	Singapore

Can you see the pain point of this?

TIPS: ALT + mouse click in each column, and then copy & paste in other blank cells or the new sheet.

Data Alchemist - Projection

- Projection in SQL
  - **select fieldname1, fieldname2, fieldname3, …**
    - Select some fieldnames from the table provided

```
select customerNumber, customerName, contactFirstName, phone
from customers;
```

Try to change some fieldnames and see the different results.

Data Alchemist – Combining

- Combine sorting, filtering, and projection
  - List some customers where locate in London or Dublin area with credit limit between $10,000 and $100,000, and display the top-most of credit limit

```
select customerNumber, creditLimit, customerName, contactFirstName, city, postalCode
from customers
where (city = "London" or city = "Dublin") and (creditLimit >= 10000 and creditLimit <= 100000)
order by creditLimit desc;
```

  - List some customers who is taken care by the sales representative employee number 1611, and display the contact first name in an ascending order

```
select salesRepEmployeeNumber, contactFirstName, phone
from customers
where salesRepEmployeeNumber = 1611
order by contactFirstName asc;
```

Try to apply in yours and see the different results.

Assignment 4 (Week 4-1)

- Form your group of 3 persons
- Download **rental.sql** from LMS and import database using MySQL
- Setup **rental** database that connect into PopSQL
- Create powerpoint to present the SQL and result follow the questions:
  - How many tables have you seen in rental database? List all table names and fields.
  - List one customer with customer id, name, and phone number whose customer id is 1003
  - List vehicles with vehicle registration number, brand, and daily rate which the daily rate is between 50 and 100, and display vehicle registration number in an ascending order
  - Find Mr. Kumar customer in all section fields
  - List vehicles with vehicle registration number, brand, and category which the category is referred to car
  - Try this SQL command to list all rental records (start date, end date) with vehicle's registration number, brand, and customer name, sorted by vehicle's categories followed by start date

```
select r.start_date as `Start Date`, r.end_date as `End Date`, r.veh_reg_no as `Vehicle No`, v.brand as `Vehicle Brand`, c.name as `Customer Name`
from rental_records as r inner join vehicles as v using (veh_reg_no) inner join customers AS c using (customer_id)
order by v.category, start_date;
```

- Submit into LMS through the link provided (Deadline also be noticed in LMS)
- Need one member per group to submit

Why Data Aggregation?

การรวมกลุ่ม

count sum max min

- The task of collecting a set of values to return a single value
- The basic functions that you should add more such as SUM, COUNT, MAX, MIN, and AVG (Average)
- It also can apply with basic calculation that return a result as an additional field
- Typically used in conjunction with grouping. The point is to help to generate a quick reports and insights from a database
- For example, an ecommerce company might want to see its highest spending customers over a given time period

Data Aggregation Basic Functions

- Basic calculation (like quantity * price as total)
- SUM (use to calculate the summation of all non-null values in a group)
- COUNT (use to count the number of row set including null values)
- MAX (use to return the highest value in a group)
- MIN (use to return the lowest value in a group)
- AVG (use to calculate the average of all values in a group)

group by - is the one of SQL command to arrange rows into groups (use for SUM, COUNT, MAX, MIN, and AVG)

Data Aggregation - Basic Calculation

- Basic calculation in SQL
  - List the order details including total of quantity ordered and price each as an additional value, and display order number in an ascending order

```
select orderNumber, productCode, quantityOrdered, priceEach, quantityOrdered * priceEach as total
from orderdetails
order by orderNumber asc;
```

Try to change some fieldnames and see the different results.

Data Aggregation - SUM

- SUM in SQL
  - Display the grand total of overall products in order details

```
select sum(quantityOrdered * priceEach) grandTotal
from orderdetails;
```

  - List the grand total of quantity ordered and price each in each product code in order details, and display product code in an ascending order

```
select productCode, sum (quantityOrdered * priceEach) grandTotal
from orderdetails
group by productCode
order by productCode asc;
```

Warning: the SUM, COUNT, MAX, MIN, and AVG function can display the result only two fieldnames from projection

Data Aggregation - COUNT

- COUNT in SQL
  - Display the overall customers

```
select count(customerName) totalCustomers
from customers;
```

  - List the total customers in each country, and display country in an ascending order

```
select country, count(country) totalCustomers
from customers
group by country
order by country asc;
```

Warning: the SUM, COUNT, MAX, MIN, and AVG function can display the result only two fieldnames from projection

Data Aggregation - MAX

- MAX in SQL
  - Display the maximum price each of order details

```
select max(priceEach) maxPriceEach
from orderdetails;
```

  - List the maximum price each of each product code in order details, and display the product code in an ascending order

```
select productCode, max(priceEach) maxPriceEach
from orderdetails
group by productCode
order by productCode asc;
```

Warning: the SUM, COUNT, MAX, MIN, and AVG function can display the result only two fieldnames from projection

Data Aggregation - MIN

- MIN in SQL
  - Display the minimum price each of order details

```
select min(priceEach) minPriceEach
from orderdetails;
```

  - List the minimum price each of each product code in order details, and display the product code in an ascending order

```
select productCode, min(priceEach) minPriceEach
from orderdetails
group by productCode
order by productCode asc;
```

Warning: the SUM, COUNT, MAX, MIN, and AVG function can display the result only two fieldnames from projection

Data Aggregation - AVG

- AVG in SQL
  - Display the average grand total of overall products in order details

```
select avg(quantityOrdered * priceEach) averageGrandTotal
from orderdetails;
```

  - List the average grand total of quantity ordered and price each in each product code in order details, and display product code in an ascending order

```
select productCode, avg(quantityOrdered * priceEach) averageGrandTotal
from orderdetails
group by productCode
order by productCode asc;
```

Warning: the SUM, COUNT, MAX, MIN, and AVG function can display the result only two fieldnames from projection

Assignment 5 (Week 4-2)

- Form your group of 3 persons
- Revise and setup **world** database into PopSQL
- Create powerpoint to present the SQL and result follow the questions:
  - List all the total number of each continent and display in an ascending order
  - Display the total number of Caribbean region
  - List the people with country code, name, language, is official, and percentage who can speak English and official (**Hint:** Check how to join country code)
  - Display the average percentage of people who can speak Chinese and non-official
  - Display the total population in Brazil (**Hint:** country code is BRA)
  - Display the maximum and minimum of country population
  - List all the countries population with the maximum to minimum. Which country has the most population and least population
  - Display the overall people's life expectancy average in the world
  - Display the maximum and minimum of country life expectancy
  - List all the countries life expectancy with the maximum to minimum. Which country has the most life expectancy and least life expectancy
- Submit into LMS through the link provided (Deadline also be noticed in LMS)