

Topic 01: Introduction to Data Management Business Data Formation and Data Storage

BDM3302: Data Management

Do you trust your data?

- Each digital touchpoint provides an opportunity to gain new insight that you can use to propel your business forward. But do you trust your data?
- If your business have so much data and it's impossible to tell what's important and what's not. Or your data may be stuck in different systems in inconsistent formats, which makes it difficult to trust or share with teams that need it. Or, worse yet, you have outdated and inaccurate data. How quickly and accurately you can resolve these challenges will determine whether your data is truly the asset your business needs to succeed.
- Most organizations, large or small, struggle with data management issues such as quality, speed, availability, and privacy. This is because data is commonly replicated across multiple silos with few or no data governance processes to manage or maintain it.
- you might have a variety of customer data sources, such as loyalty programs, sales records, service calls, or surveys. Those data sources reside in systems across the enterprise and extended value chain.

What is Data Management?

- Data Management refers to the development and execution of architectures, policies, practices and procedures, in order to manage the information lifecycle of an enterprise in an effective manner.
- There are eight subject areas (capabilities) in Data Management:

Meta Data Management

Data Quality Management

Data governance

Master/
Reference
Data

Data
Integration

Data
Analytics

Data Privacy

Data
Architecture
Privacy

Data Management Framework

- **People** refers to organizational aspects the roles and responsibilities required for each subject areas.
- **Process** refers to activities that are associated with the subject areas.
- **Technology** refers to technologies and tools required to support capability business processes.



People



Process



Technology

Meta Data Management

Data Integration

Data Quality Management

Data Analytics

Data governance

Data Privacy

Master/
Reference
Data

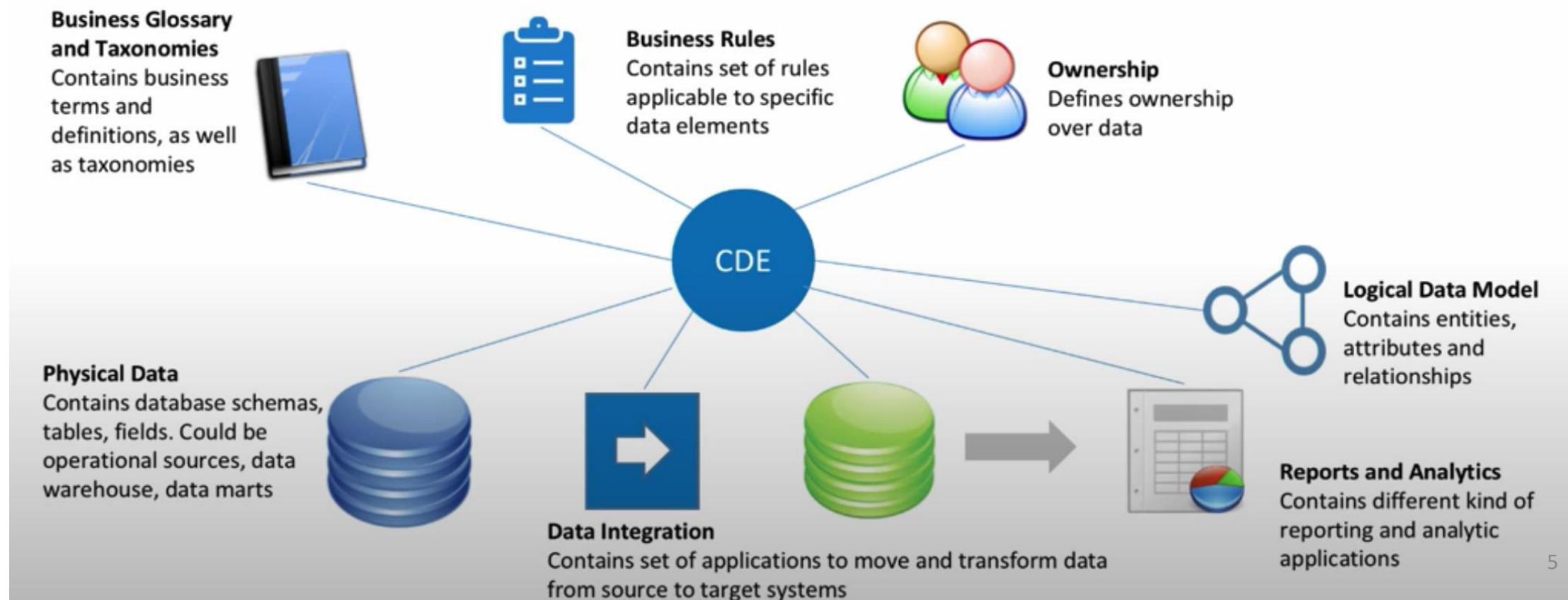
Data
Architecture
Privacy

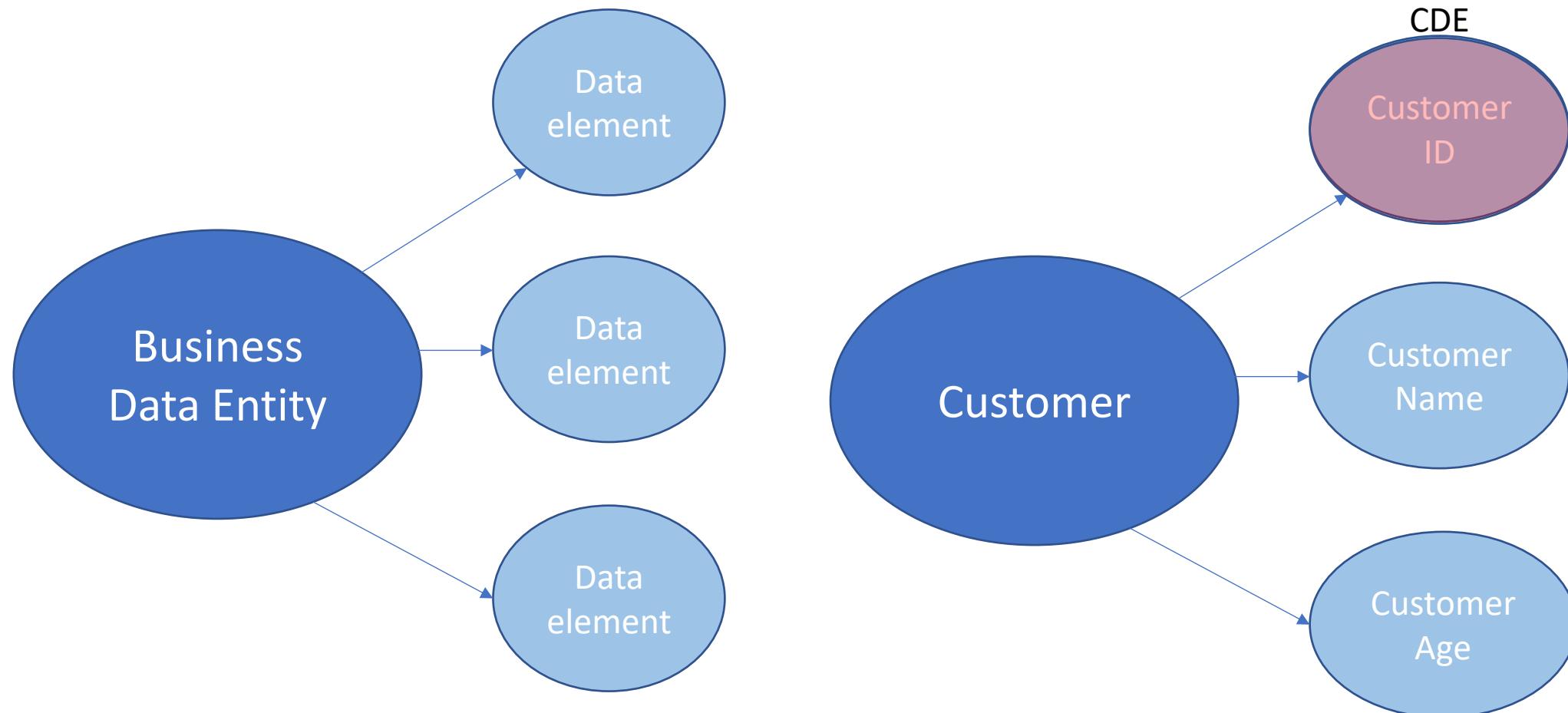
Metadata Management

- What is **Data Element**?
- **Data Element (DE)** is a unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes.
- **Critical Data Element (CDE)** is the data element that is “critical to success” in a specific business area or business process.
- Criteria for Data Element to become Critical
 - Business facts that are deemed critical to the organization
 - Support critical business processes across an organization and its components
 - Data used to derive values that appear in key reports
 - Unique identifiers of things important to the business (such as Customer ID)
 - Any data element that is required for the execution of a key business process is a Critical Data Element

What is Metadata Management?

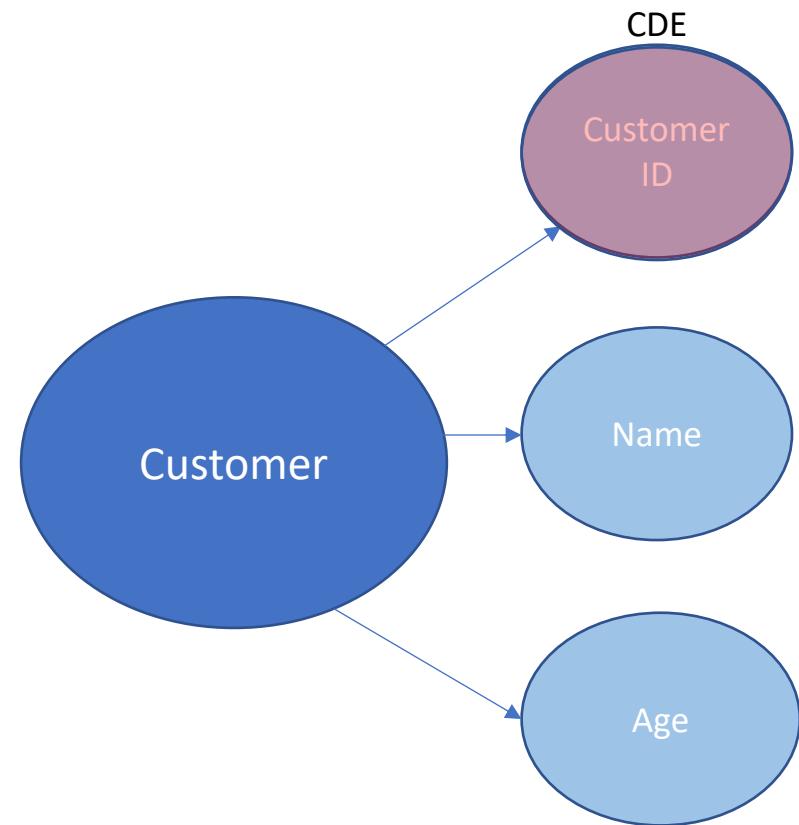
- Metadata Management involves managing data about other data, whereby this “other data” is generally referred to data models and structures, not the content. It includes managing information about data structures from different models and their mutual associations (like business terms in glossary, attributes in logical data model, or tables and columns in the database, as well as their associations).





Critical Data Element (CDE)

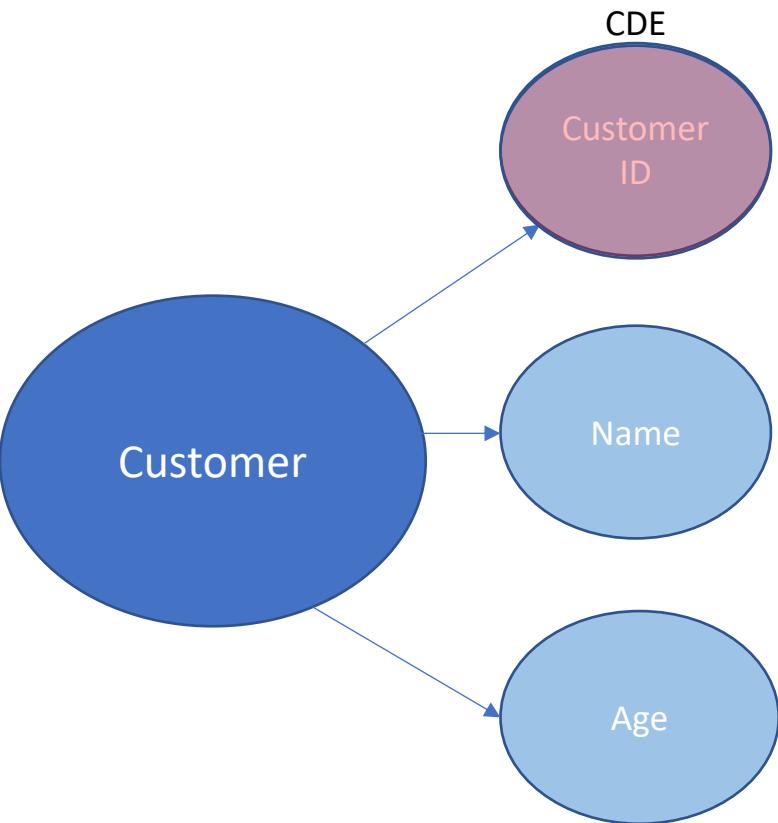
- Important fact for the success of organization/process/stakeholder
- Appears in Key report
- Important for a set of business processes



Data element characteristic

- DE : has size (What?) ✓
- DE : has data type (What?) ✓
- CDE : purpose of data serves (Why?)
- CDE : critical to whom/stakeholder (Who?)
- CDE : unique identification (text, or number) (What?) ✓
- CDE : time stamp (When?)
- CDE : Data source (Where?)

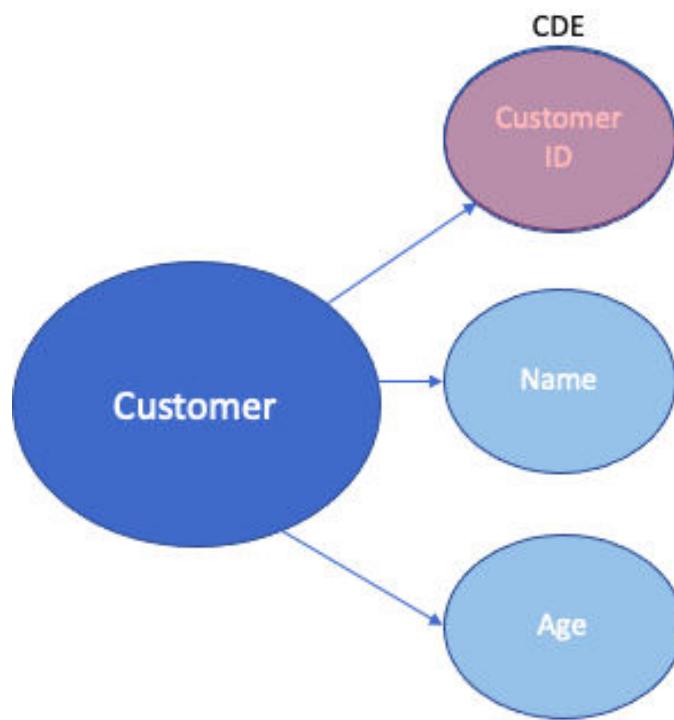
Data Name	Customer ID	Name	Age
type	int	Varchar(20)	int
role	PK,DE	DE	DE
sample	101	John	40



Data element characteristic

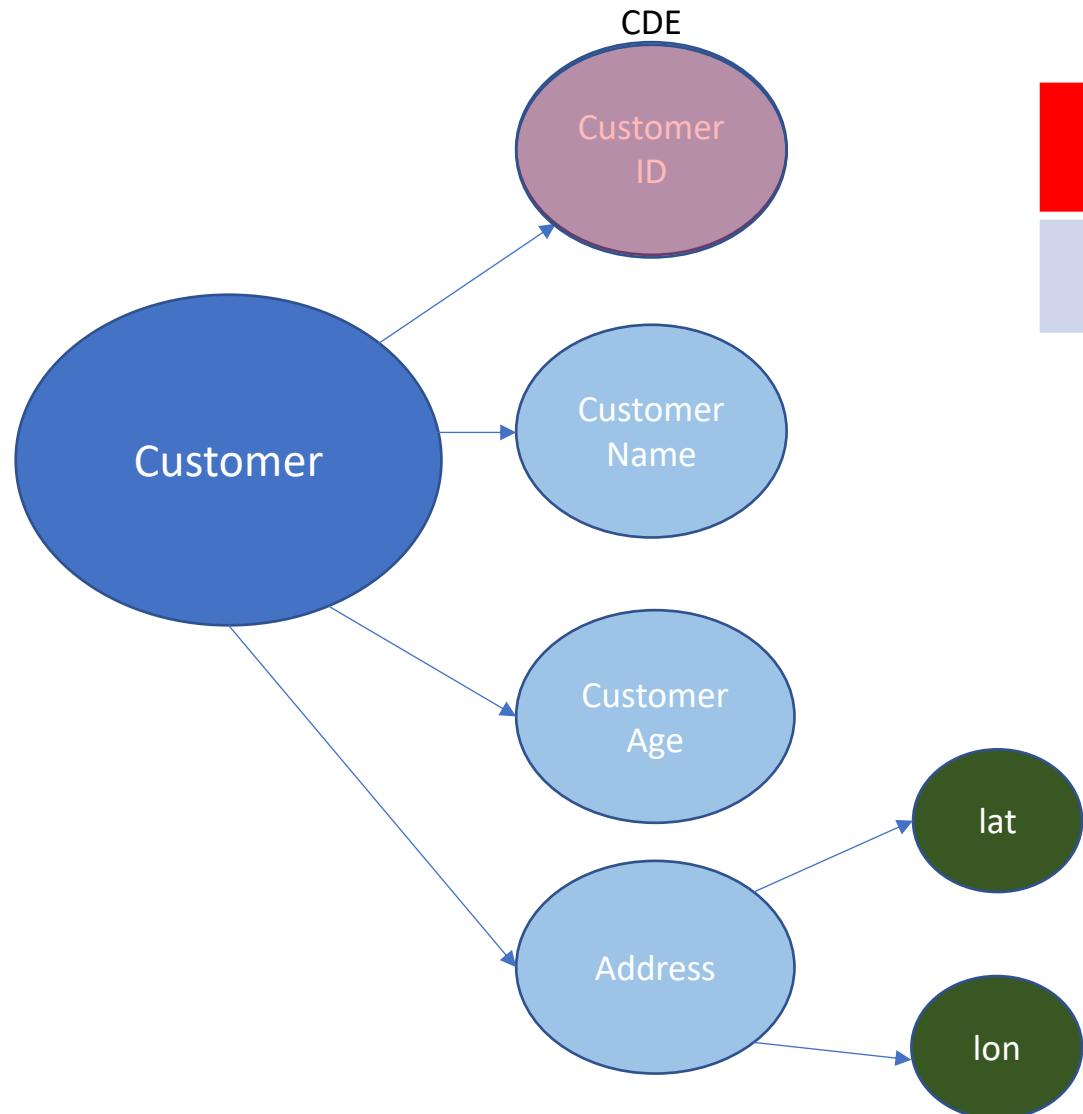
- DE : has size (What?) ✓
- DE : has data type (What?) ✓
- DE : purpose of data serves (Why?)
- DE : critical to whom/stakeholder (Who?) ✓
- DE : unique identification (text, or number) (What?) ✓
- DE : time stamp (When?) ✓
- DE : data source (Where?) ✓
- DE : constraints (PK, Nullity, Autoincrement, FK, etc)

Key processes	Product order, customer support, product enquiry, promotion etc.			
Table	Customer			
Data Name	Customer ID	Name	Age	Date
type	int	Varchar (20)	int	DateTime
constraint	PK, auto_inc	Not null	Not null	Data field
sample	101	John	40	15 Nov 2021



Key processes	Product order, customer support, product enquiry, promotion etc.			
Data Name	Customer ID	Name	Age	Date
type	int	Varchar (20)	int	DateTime
constraints	PK, DE	DE	DE	Data field
sample	101	John	40	15 Nov 2021

Data Name	Name
Description	represents customer full name
Business rule	Thai full name consist of First name and Last Name Chinese name has middle name in between
Data Owner/ Producer	Sale Team
Data Consumer	Sale Team, Marketing Team, Shipment Team, etc
Data Policy	Marketing Team can have access to the data prior to the year 2016
Report	Sale Report, Daily shipment Report, Promotion Report



Customer ID (int)	Name (varchar)	Age (int)	Address ?
101	John	40	???

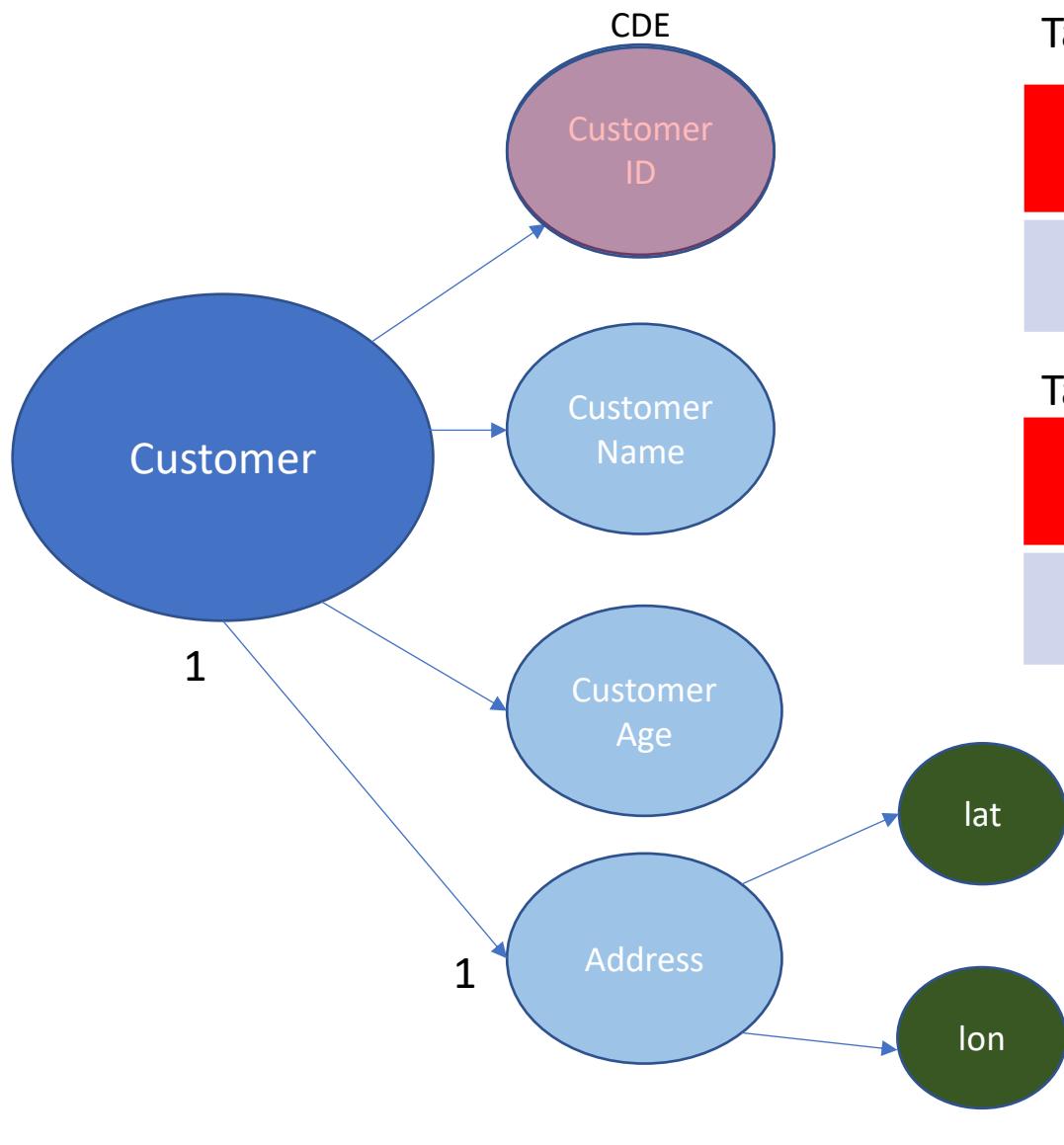


Table: Customer

Customer ID (int)	Name (varchar)	Age (int)	Address Id (int)
101	John	40	101

Table: Address

Address ID (int)	Lat (double)	Long (double)
101 ←	13.12345	101.1234

```
{                                         Object notation
  CustomerId : 101 ,  

  Name : "John" ,  

  Age : 40 ,  

  Address : { lat : 13.12345 ,  lon : 101.1234 }  

}  
}
```

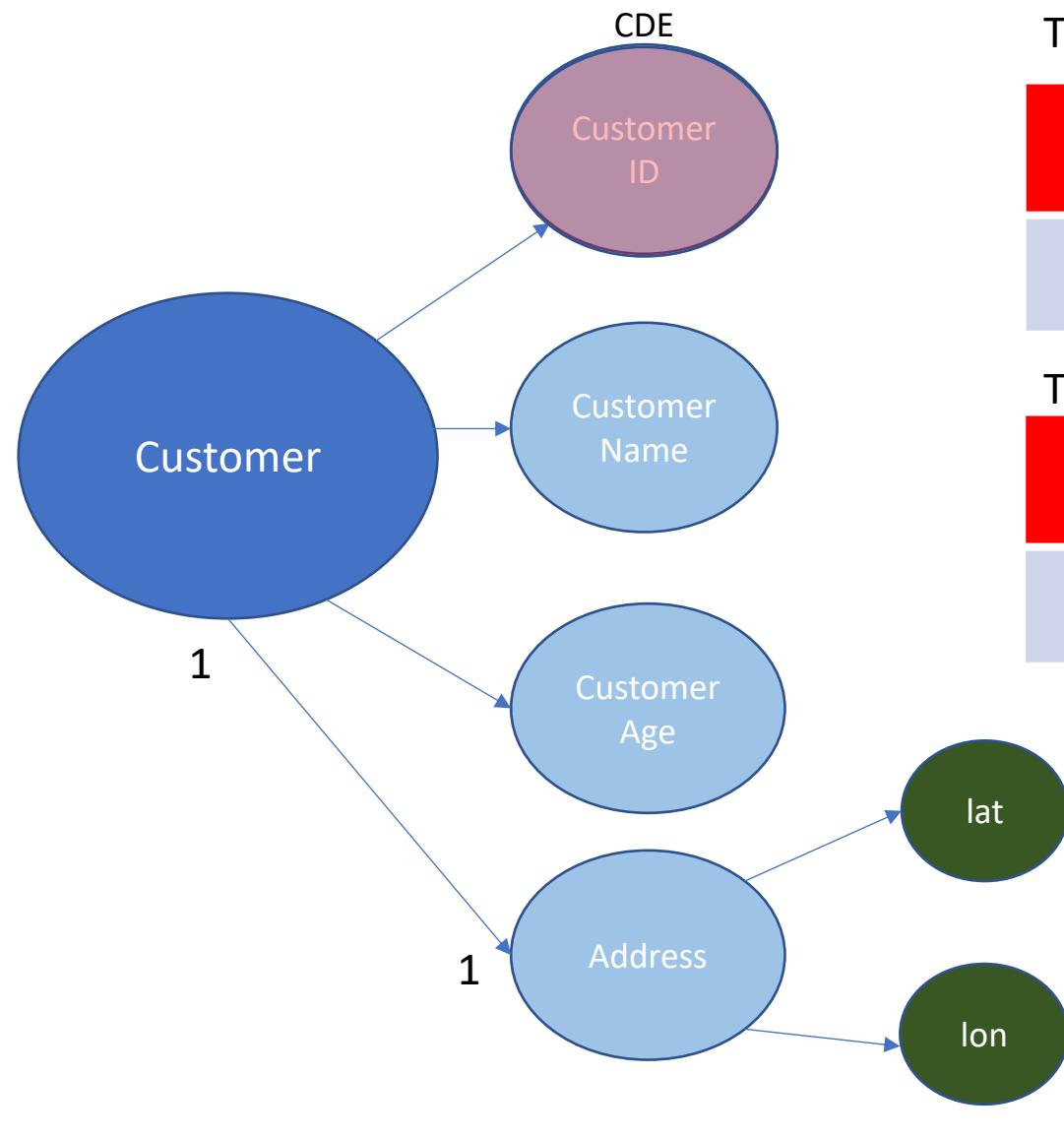


Table: Customer

Customer ID (int)	Name (varchar)	Age (int)	Address Id (int)
101	John	40	101

Table: Address

Address ID (int)	Lat (double)	Long (double)
101	13.12345	101.1234

{
 "CustomerId" : 101 ,
 "Name " : "John" ,
 "Age" : 40 ,
 "Address" :{ lat : 13.12345 , lon : 101.1234 }
 }

JSON format

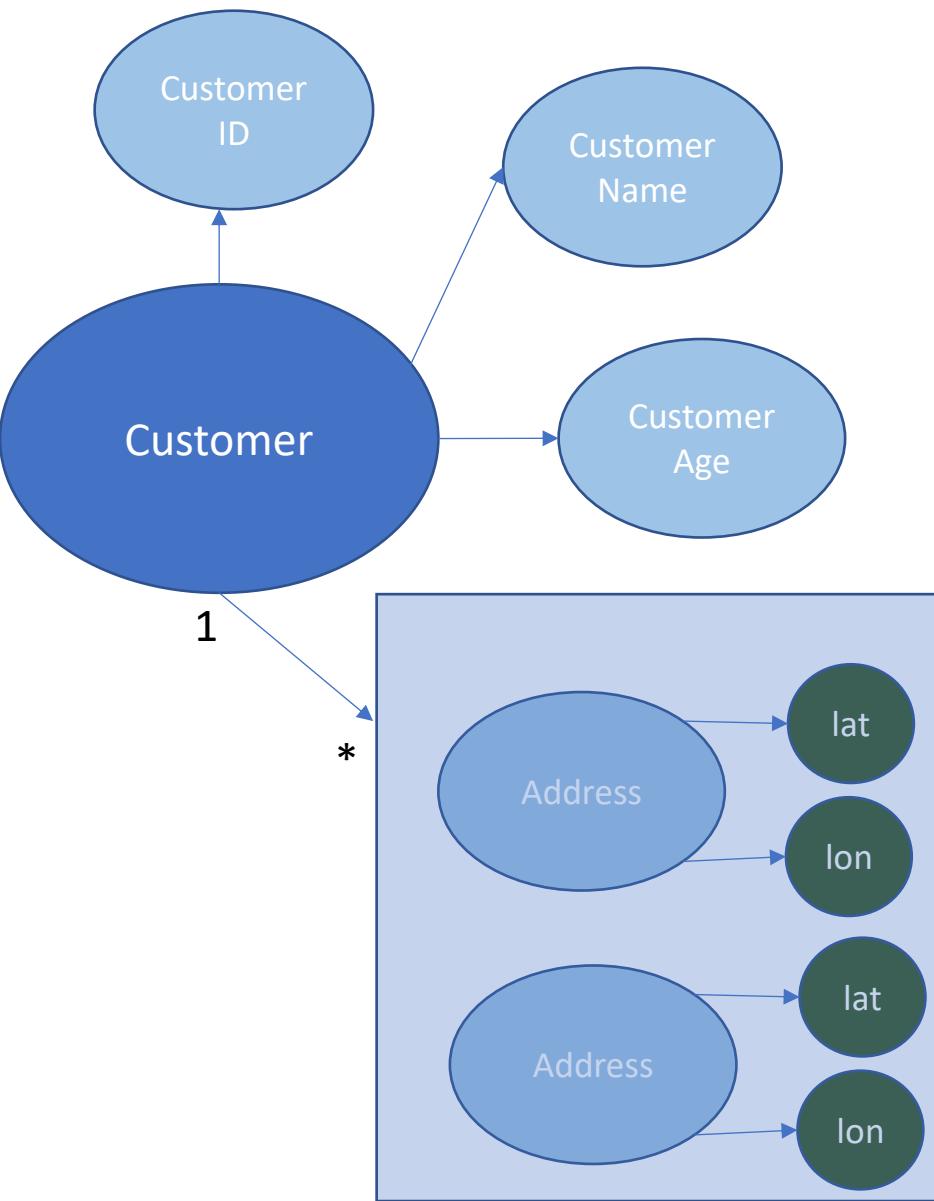


Table: Customer

Customer ID (int)	Customer (varchar)	Customer Age (int)	
101	John	40	
101	John	40	102
Address ID PK (int)	Lat (double)	Long (double)	Customer ID (int) FK
101	13.12345	101.1234	→ 101
102	13.22345	103.1234	→ 101

```
{
  CustomerId : 101 ,
  Name : "John" ,
  Age : 40 ,
  Addresses : [   { lat : 13.12345 , lon : 101.1234 } ,
                  { lat : 13.22345 , lon : 103.1234 } ]
}
```

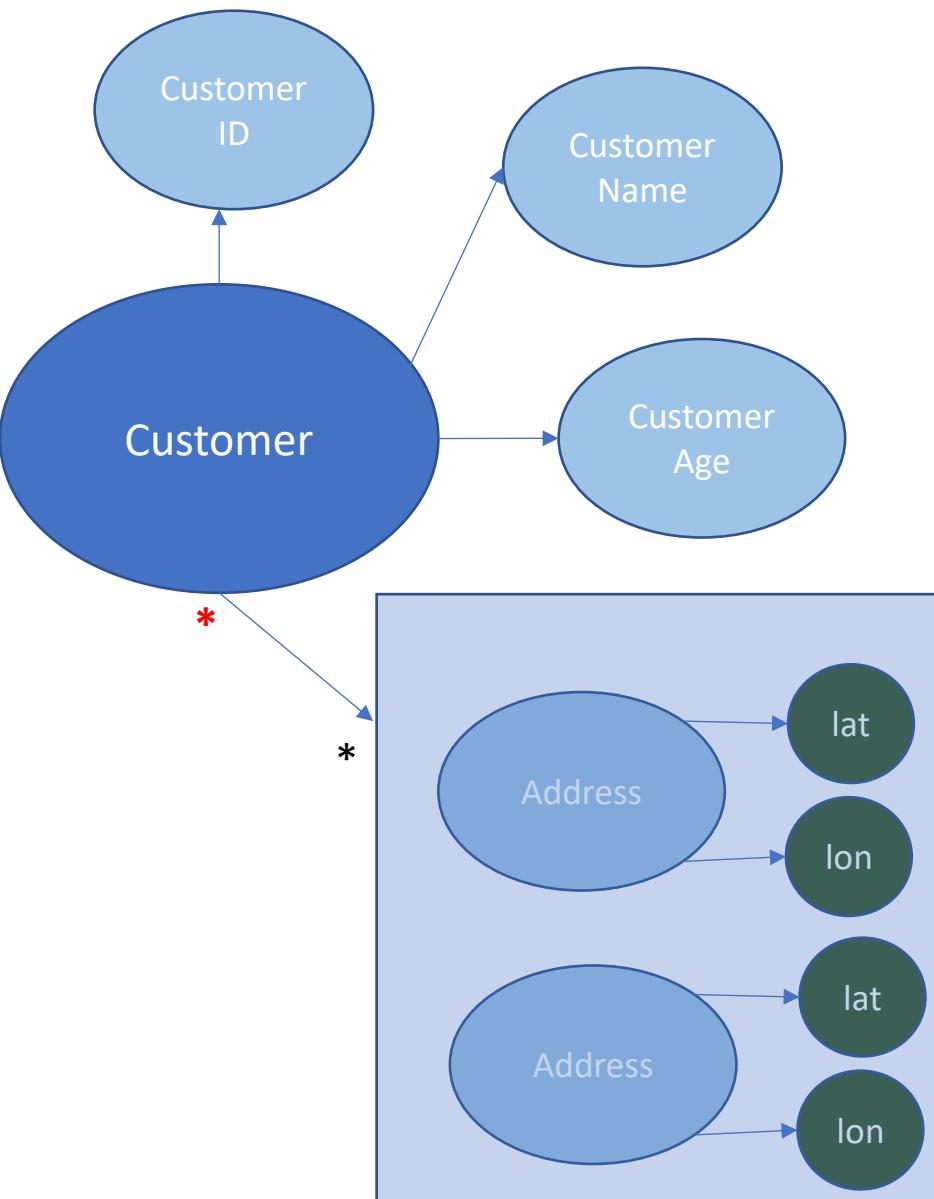


Table: Customer

Customer ID (int)	Customer (varchar)	Customer Age (int)
101	John	40
102	Sam	20

Table: Address

Address ID (int)	Lat (double)	Long (double)
101	13.12345	101.1234
102	13.22345	103.1234

Table: CustomerAddress

No	Customer ID (int)	Address ID (int)
1	101	101
2	101	102
3	102	101

Data Quality Management

- What is Data Quality (DQ)?
- Data Quality refers to the methodical approach, policies and processes by which an organization manages the **accuracy, validity, timeliness, completeness, uniqueness, and consistency** of its data in systems and data flows.
- Not all Data Quality Dimensions are applicable on particular CDE (e.g. Date of Birth can be assessed against Validity and Completeness only)

Data Quality Dimensions

- **Accuracy** refers to error-free records that can be used as a reliable source of information.
- **Validity** refers to information that does not conform to a specific format or does not follow business rules. (DOB format)
- **Timeliness** refers to the time expectation for accessibility and availability of information. It can be measured as the time between when information is expected and when it is readily available for use.
- **Completeness** refers to the degree to which all data in a data set is available. (all required fields like *first name* must have data.)
- **Uniqueness** refers to a measure of unwanted duplication existing within or across systems for a particular field, record, or data set.
- **Consistency** refers to data values that are the same for all instances of an applications. (an employee does not work anymore but still receiving a check or money transfer.)

What is Data Governance?

- Data Governance refers to a data management concept concerning the capability that enables an organization to ensure that high data quality exists throughout the complete lifecycle of the data, and data controls are implemented that support business objectives.
(Wikipedia.org)
- The key focus areas of data governance include availability, usability, consistency, data integrity and data security, standard compliance and includes establishing processes to ensure effective data management throughout the enterprise.

What is Master and Reference Data Management?

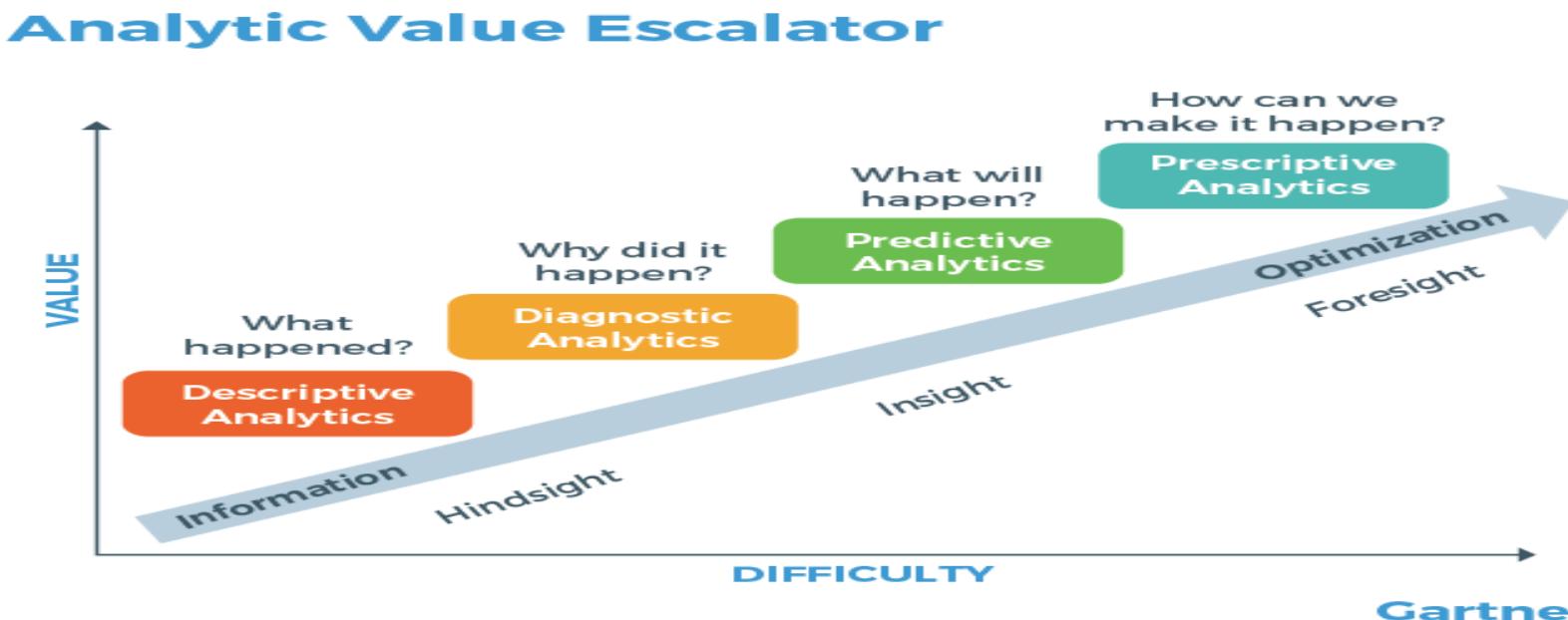
- Master data management (MDM) refers to a process that creates a uniform set of data on customers, products, suppliers and other business entities from different IT systems. One of the core disciplines in the overall data management process, MDM helps improve data quality by ensuring that identifiers and other key data elements about those entities are accurate and consistent enterprise-wide.
(TechTarget.com)
- Once created, this master data serves as a trusted view of business-critical data that can be managed and shared across the business to promote accurate reporting, reduce data errors, remove redundancy, and help workers make better-informed business decisions.
(Informatica.com)

What is Data Integration ?

- Data Integration (DI) refers to the process of combining data from different sources into a single, unified view. Integration begins with the ingestion process, and includes steps such as cleansing, ETL mapping, and transformation. Data integration ultimately enables analytics tools to produce effective, actionable business intelligence. (Talend.com)
- Extract, Transform, Load (ETL) is a process within data integration wherein data is taken from the source system and delivered into the warehouse. This is the ongoing process that data warehousing undertakes to transform multiple data sources into useful, consistent information for business intelligence and analytical efforts.
- Batch vs Real Time Data

What is Data Analytics?

- Data Analytics is the process of examining data sets in order to find trends and draw conclusions about the information they contain. Increasingly, data analytics is done with the aid of specialized systems and software. (TechTarget.com)
- Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions.

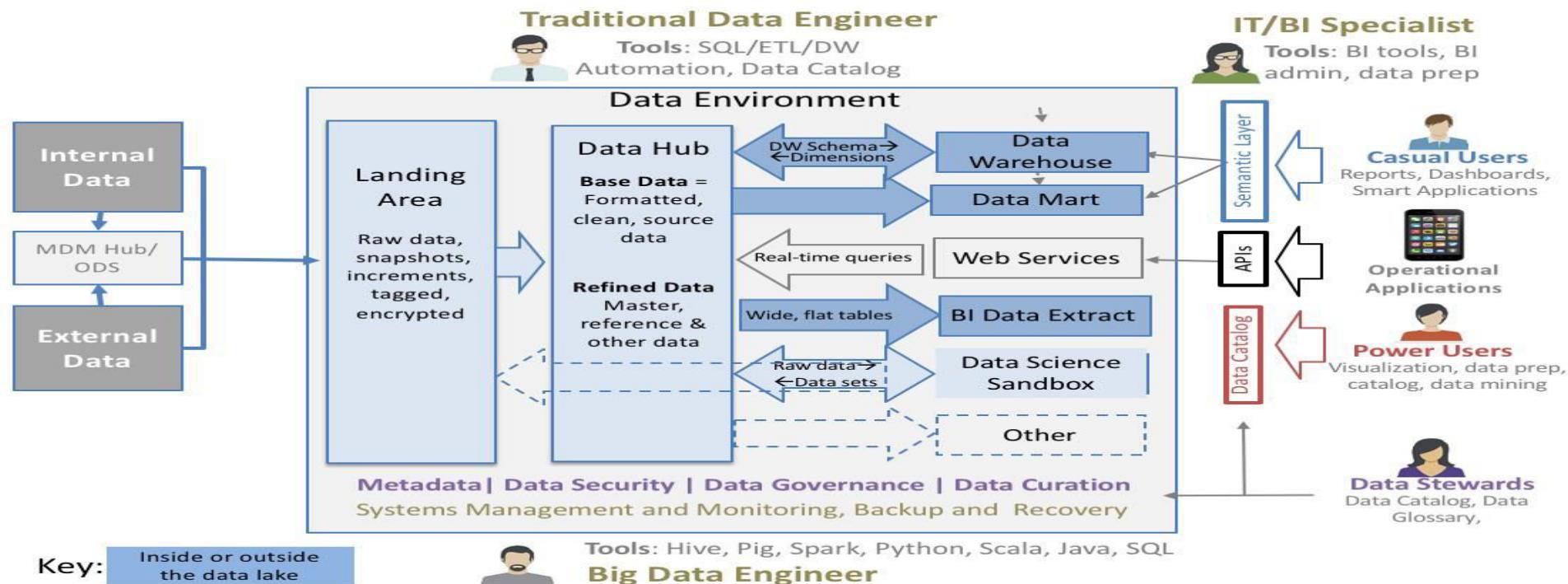


What is Data Privacy?

- Data privacy generally refers to the ability of a person to determine for themselves when, how, and to what extent personal information about them is shared with or communicated to others. This personal information can be one's name, location, contact information, or online or real-world behavior. (Cloudflare.com)
- Privacy is considered a fundamental human right, and data protection laws exist to guard that right. Data privacy is important because when individuals to be willing to engage online, they have to trust that their personal data will be handled with care.
- Organizations use data protection practices to demonstrate to their customers and users that they can be trusted with their personal data.
- General Data Protection Regulation (GDPR) - Regulates how the personal data of European Union (EU) data subjects, meaning individuals, can be collected, stored, and processed, and gives data subjects rights to control their personal data.
- California Consumer Privacy Act (CCPA) - Requires that consumers be made aware of what personal data is collected and gives consumers control over their personal data, including a right to tell organizations not to sell their personal data.

What is Data Architecture?

- Data Architecture is the models, policies, rules, and standards that govern which data is collected and how it is stored, arranged, integrated, and put to be used in data systems and in organizations. (Wikipedia.org)
- Data is usually one of several architecture domains that form the pillars of an enterprise architecture or solution architecture.



Data Formation

- Text: XML, PDF/A, HTML, ASCII, UTF-8 (not Word)
- Tabular Data: CSV (not Excel)
- MS Excel: XLS, XLSX
- Databases: XML, CSV

Multimedia formats:

- Still Images: TIFF, JPEG 2000, PDF, PNG, BMP (not GIF or JPG)
- Moving Images: MOV, MPEG, AVI, MXF (not Quicktime)
- Sounds: WAVE, AIFF, MP3, MXF

Statistics formats:

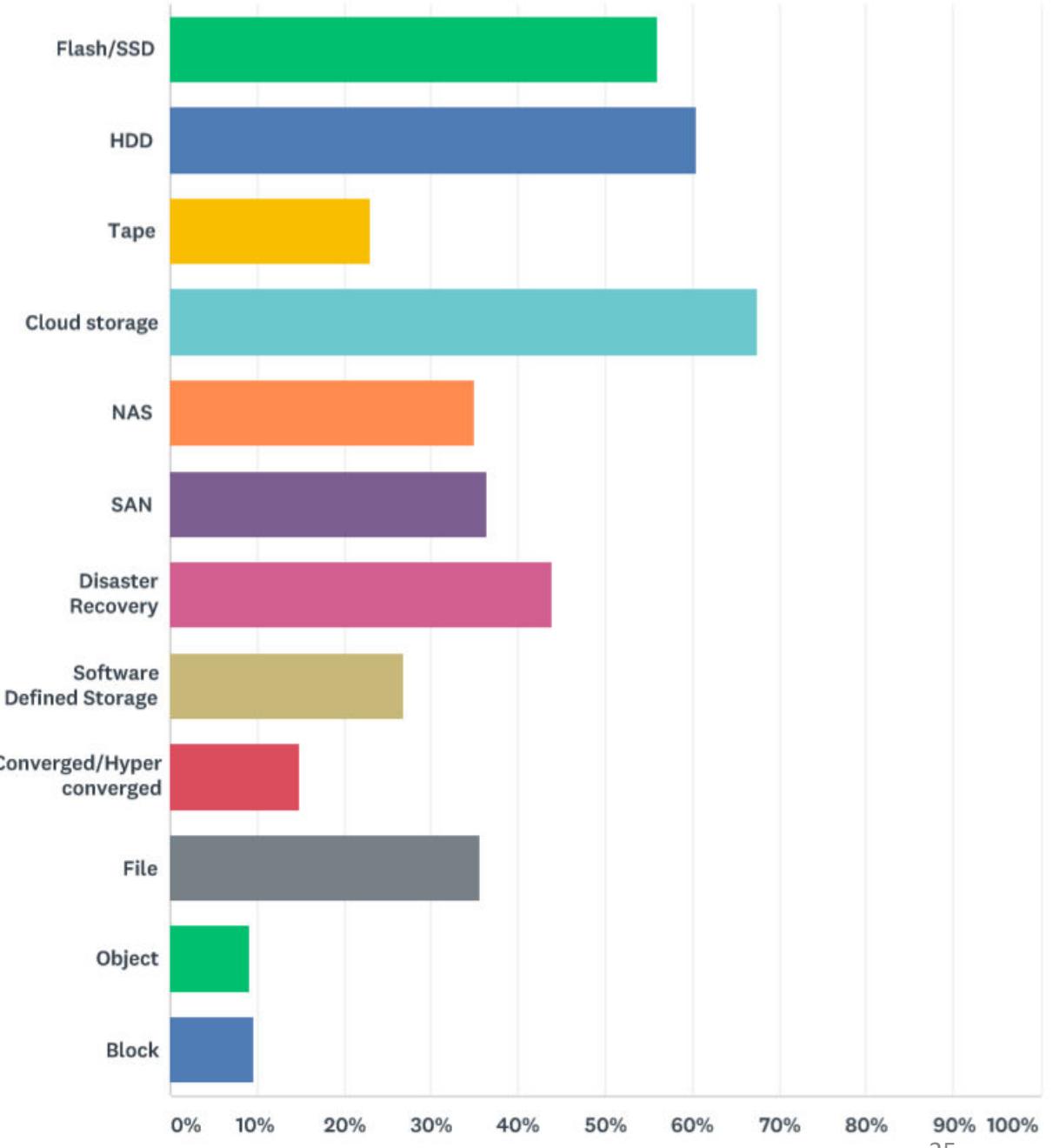
- Statistics: ASCII, SAS, SAV

Data Types

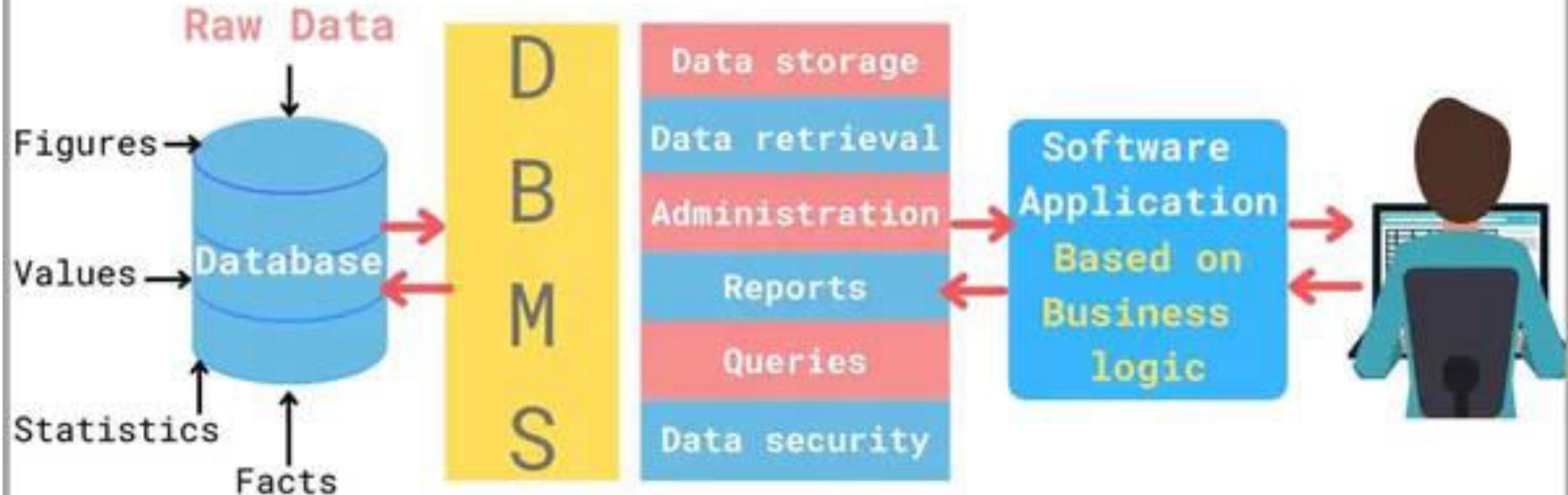
- String (text)
 - Varchar - last name
- Numeric (number)
 - Int – student ID
 - Double - weight
- Date
 - Date – enroll date
- Datetime
 - Clock-in/ Clock-out at workplace (Timestamp)

Data Storage

- Computer storage
 - Primary storage
 - Secondary storage
- Cloud storage - a cloud computing model that stores data on the Internet through a cloud computing provider who manages and operates data storage as a service. (AWS)



Database Management System



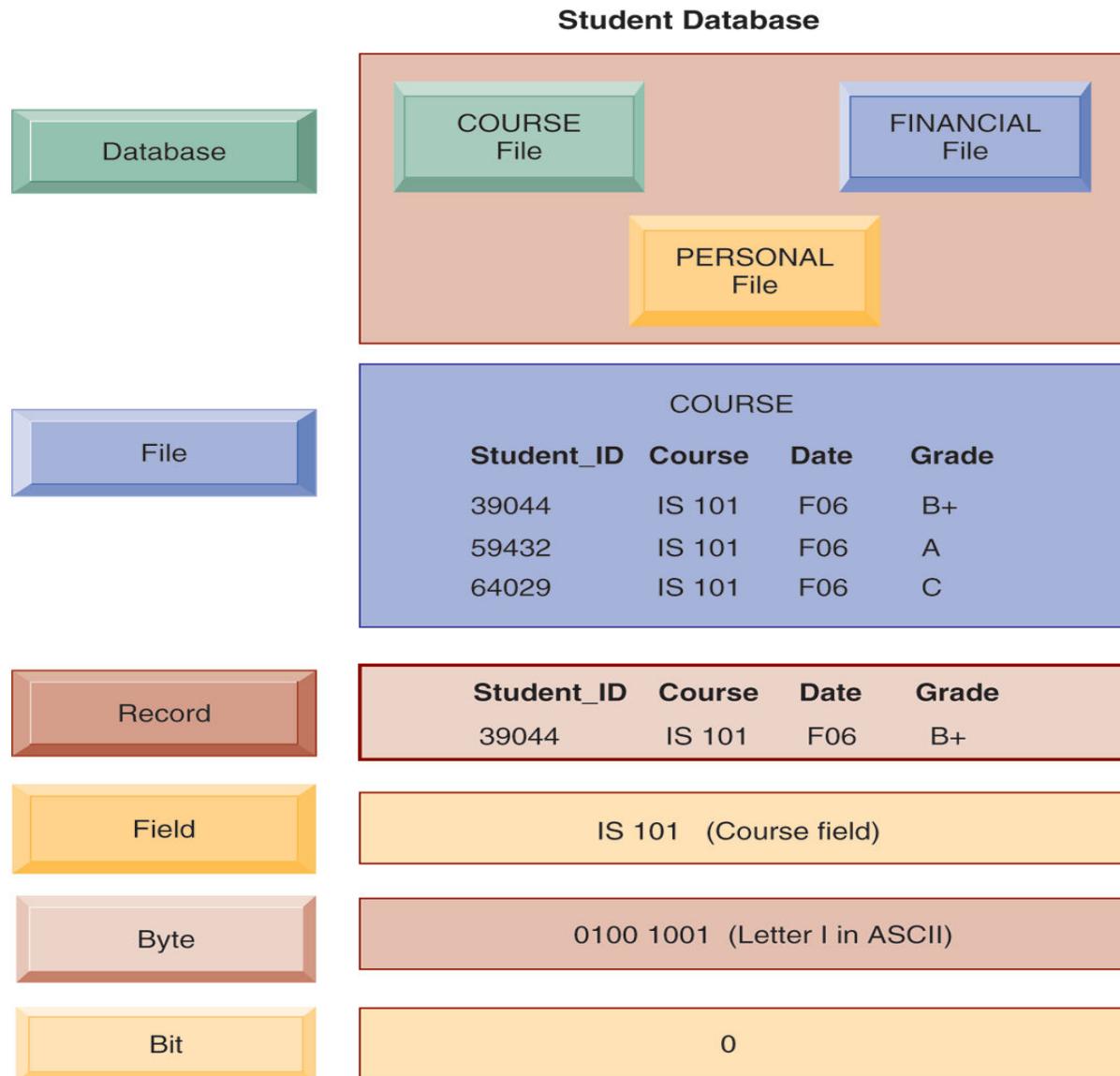
Organizing Data in a Traditional File Environment

- File organization concepts
 - Database: Group of related files
 - File: Group of records of same type
 - Record: Group of related fields
 - Field: Group of characters as word(s) or number
 - Describes an **entity** (person, place, thing on which we store information)
 - **Attribute:** Each characteristic, or quality, describing entity
 - E.g., Attributes Date or Grade belong to entity COURSE

Organizing Data in a Traditional File Environment

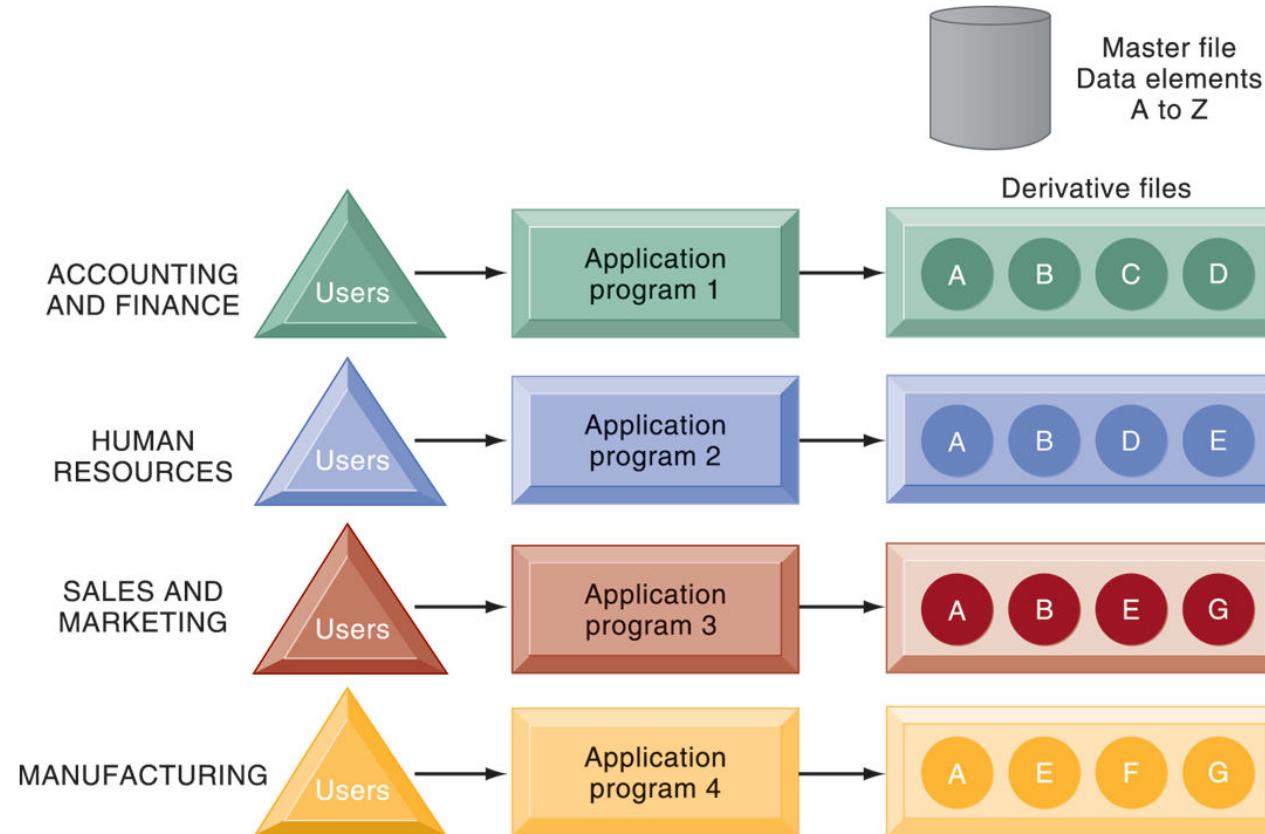
THE DATA HIERARCHY

A computer system organizes data in a hierarchy that starts with the bit, which represents either a 0 or a 1. Bits can be grouped to form a byte to represent one character, number, or symbol. Bytes can be grouped to form a field, and related fields can be grouped to form a record. Related records can be collected to form a file, and related files can be organized into a database.



Organizing Data in a Traditional File Environment

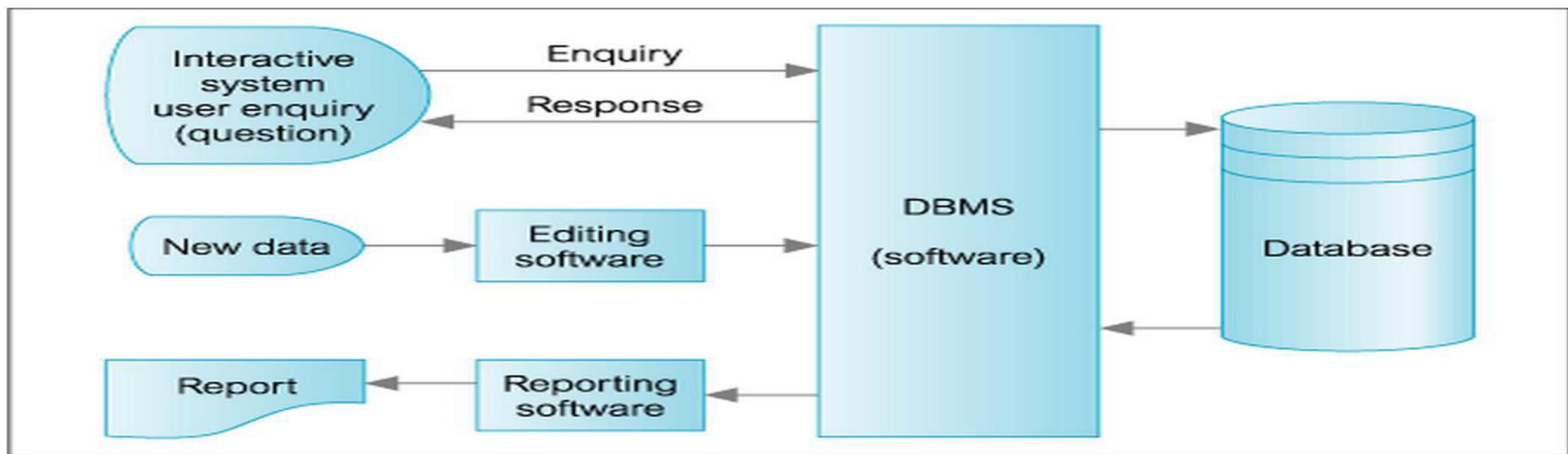
TRADITIONAL FILE PROCESSING



What kinds of data might be shared between sales and marketing and accounting?

Database Approach to Data Management

- Database
 - Eliminates many of the problems of traditional file organization by organizing data, centralizing data and controlling redundant data, and serve many applications and different groups at the same time.



Topic 02 - Basic SQL

BDM3302: Data Management

What is a Database?

- A collection of related information
 - Phone book
 - Name your 5 best friends
 - Facebook's user base
 - Shopping list
- Database can be stored in:
 - Paper
 - Your mind
 - Computer
 - Etc.

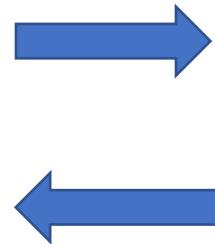
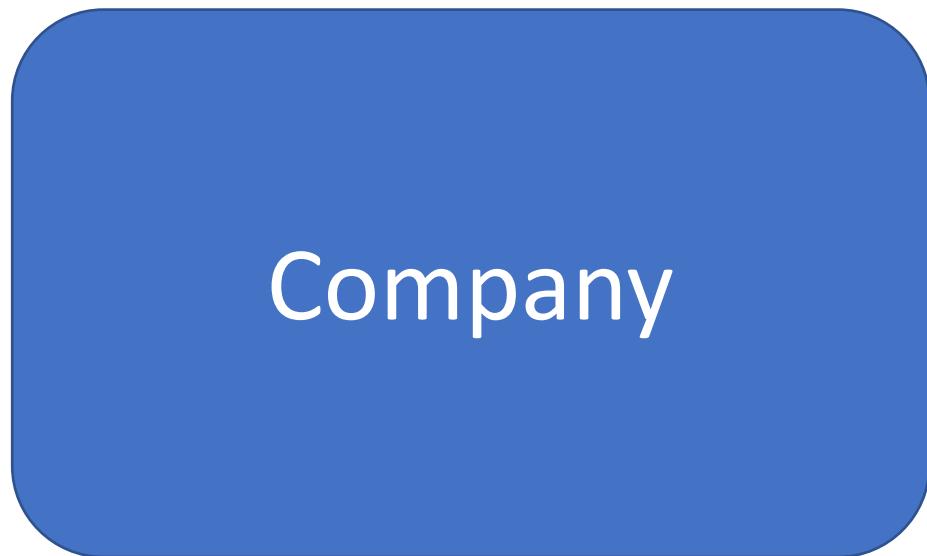
Database Management Systems (DBMS)

- A special software program that help users create and maintain a database.
- Makes it easy to manage large amounts of information
- Handles security
- Backups
- Import and export data
- Concurrency
- Interacts with other software applications

Database Management Systems (DBMS)

- Represent data as two-dimensional tables called relations or files
- Each table contains data on entity and attributes
- Table: grid of columns and rows
 - Rows (tuples): Records for different entities
 - Fields (columns): Represents attribute for entity
 - Key field: Field used to uniquely identify each record
 - Primary key: Field in table used for key fields
 - Foreign key: Primary key used in second table as look-up field to identify records from original table

A Company Database Diagram



A company interact with DBMS in order to create, read, update and delete information.

***** **C. R. U. D. operations** *****

Types of Databases

Relational Database (SQL)

- Organize data into one or more tables.
 - Each table has columns and rows
 - A unique key identifies each row

Non-Relational (noSQL / not just SQL)

- Organize data is anything but a traditional table
 - Key-value stores
 - Documents (JSON, XML, etc.)
 - Graphs
 - Flexible Tables

Relational Database (SQL)

- Relational Database Management Systems (RDBMS)
 - Help users create and maintain a relational database
 - mySQL, Oracle, PostgreSQL, MariaDB, etc.
- Structured Query Language (SQL)
 - Standardized language for interacting with RDBMS
 - Used to perform C.R.U.D. operations, and other administrative tasks (user management, security, backup, etc.)
 - Used to define tables and structures
 - SQL code used on one RDBMS is not always portable to another without modification.

RELATIONAL DATABASE TABLES

SUPPLIER

Columns (Attributes, Fields)					
Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5 th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Key Field
(Primary Key)

A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

RELATIONAL DATABASE TABLES

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key Foreign Key

A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

Non-Relational Databases (noSQL/ not just SQL)

- Non-Relational Database Management Systems (NRDBMS)
 - Help users create and maintain a non-relational database
 - mongoDB, DynamoDB, Apache Cassandra, firebase, etc.
- Implementation Specific
 - Any non-relational database falls under this category, so there's no set language standard.
 - Most NRDBMS will implement their own language for performing C.R.U.D. and administrative operation on the database.

Database Queries

- Queries are requests made to the database management system for specific information. (*basically, you are asking the DBMS for a specific information*)
- As the database's structure become more and more complex, it becomes more difficult to get the specific pieces of information we want.
- A google search is a query. (use English or other languages)
- SQL (use a specific language for database management systems)

Structured Query Language (SQL)

- A language used for interacting with RDBMS
- You can use SQL to get the RDBMS to do things for you.
 - CRUD data
 - Create and manage databases
 - Design and create database tables
 - Perform administrative tasks (user management, security, import/export, etc)

SQL

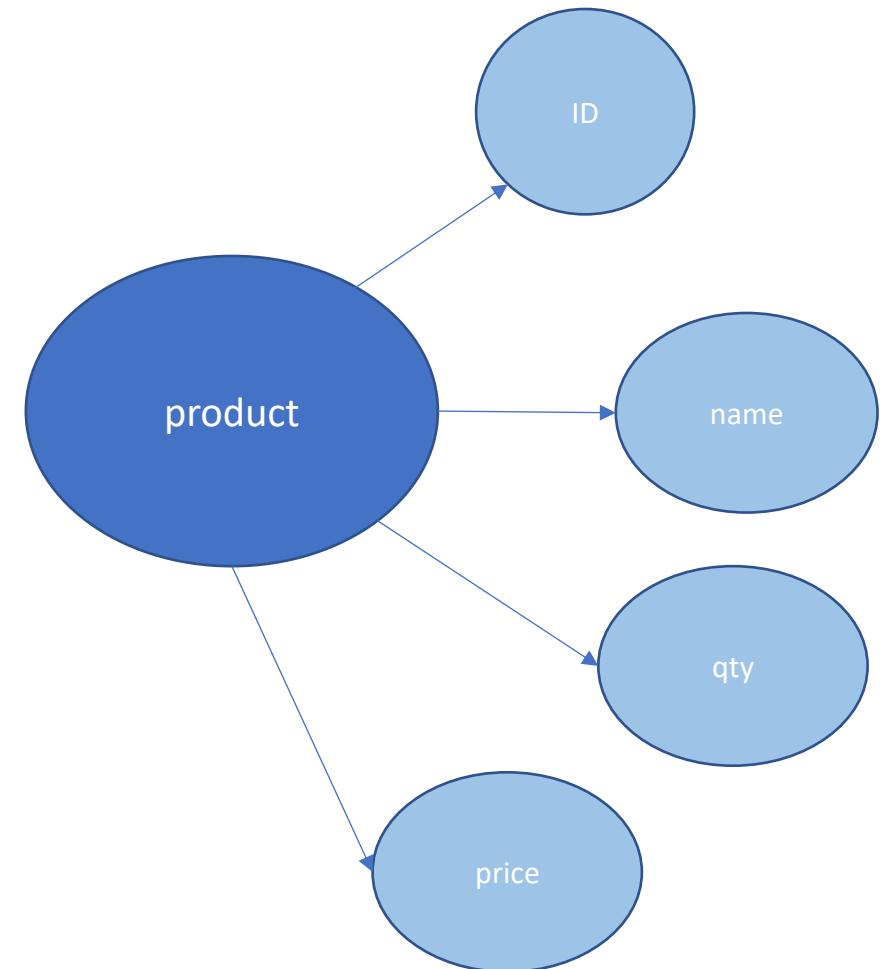
- Is actually a hybrid language, there are 3 types of languages in one.
- Data Definition Language (DDL)
 - Used for defining database schemas.
- Data Control Language (DCL)
 - Used for controlling access to the data in the database.
 - User and permissions management.
- Data Manipulation Language (DML)
 - Used for inserting, updating and deleting data from the database.

Queries

- A query is a set of instructions given to the RDBMS (written in SQL) that tell the RDBMS what information you want it to retrieve for you.
- TONS of data in a DB
- Often hidden in a complex schema
- Goal is to only get the data you need
 1. **SELECT** employee.name, employee.age
 2. **FROM** employee
 3. **WHERE** employee.salary > 30000

Table: Product

Row/Tuples/Records	Column/Data Field/Attribute			
	ID	name	qty	price
1	p1	10	100	
2	p2	20	200	
3	p3	30	300	

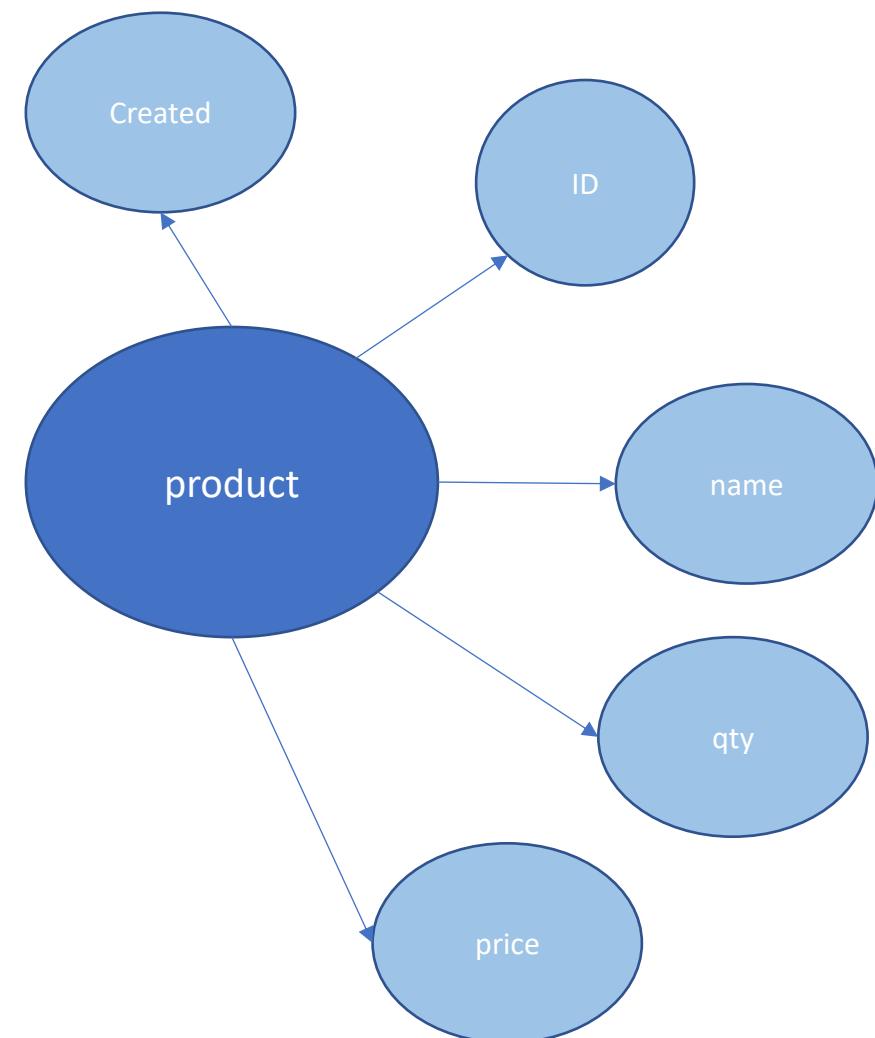


1. Define Primary Key
2. Define Column Schema

Users	Staff, Manager, Supplier, Account	
Process	Acquisition, Purchase, Inventory, Order	
Owner	Sale and Marketing Team	

ID	rule	Last reviewed on
id	Unique integer	10 Nov 2021
name	Max 200 character Unique name Lower case	20 Nov 2021
qty	Max 1000	10 Nov 2021
price	Numeric value	12 Nov 2021
Created	Date Time	14 Nov 2021

process	Data Query frequency	Backup policy
Acquisition	daily	daily
Order	sale transaction	daily
Inventory	daily	weekly
Purchase	daily	weekly



data

ID	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	30	300

```
create table product
(
    id      INT      NOT NULL PRIMARY KEY ,
    name   varchar(20) NOT NULL ,
    qty    integer    NOT NULL ,
    price  double    NOT NULL
)
```

data

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	30	300

insert **into**
product
(id , name , qty , price)
value
('1' , 'p1' , '10' , '100')

insert into product(id,name,qty,price) value('1','p1','10','100')

data

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

update product**set****qty = '20',****price='100'****where****name = 'p1'****update product set qty = '20', price='100' where name = 'p1'**

data

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

delete from product
where
name = 'p1'

delete from product where name = 'p1'

Data: product

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

```
select * from product
```

Relational Algebra

output

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

product



output

Data: product

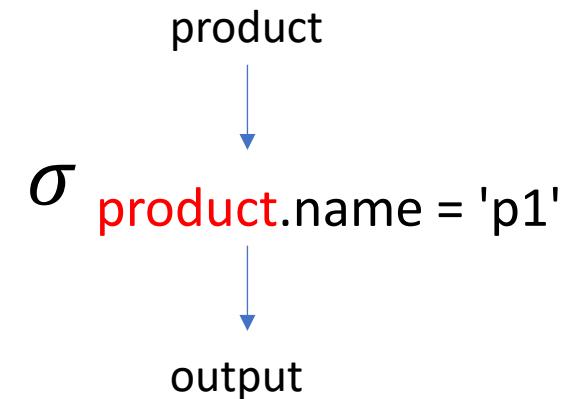
id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

output

id	name	qty	price
1	p1	10	100

select * from **product**
where
name = 'p1'

Relational Algebra



Data: product

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

output

id	name	qty	price
3	p3	10	300

select * from product

where

qty >= 10

and

price > 200

Relational Algebra

product

$\sigma_{\text{product.qty} \geq 10 \text{ AND } \text{product.price} > 200}$

output

Data: product

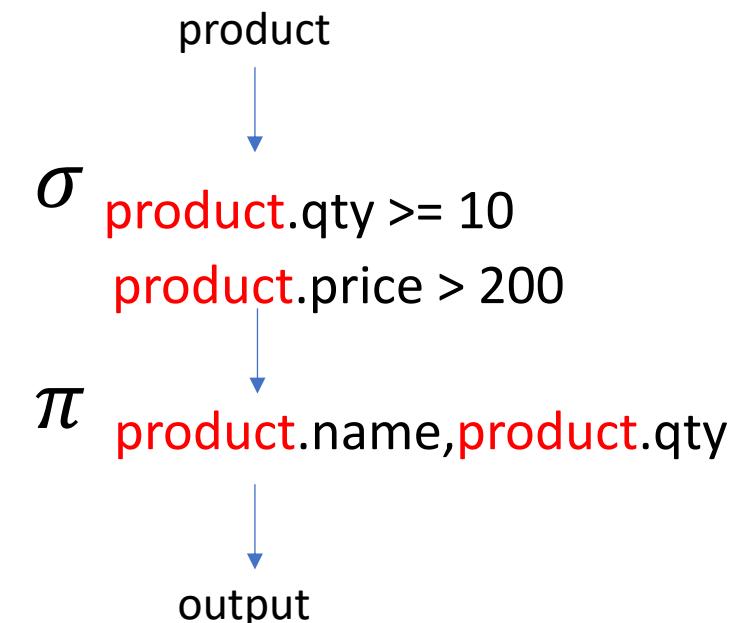
id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

select **name, qty** **from** **product**
where
 qty >= 10
and
 price > 200

Relational Algebra

output

name	qty
p3	10



Data: product

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

output

product name	quantity
p3	10

select **name as `product name`**, **qty as quantity**
from **product**
where
 qty >= 10
 and
 price > 200

Relational Algebra

product

σ **product.qty >= 10**
product.price > 200

π **product.name as `product name`**,
product.qty as quantity

output

Data: product

id	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

output

product name	quantity	total
p3	10	3000

select name as `product name`,
 qty as quantity,
 qty*price as total

from **product**

where

qty >= 10

and

price > 200

Relational Algreba

product

σ **product.qty >= 10**
 product.price > 200

π **product.name as product name ,**
 product.qty as quantity,
 product.qty*product.price as total

output

Data: product

pid	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

Data: Supplier virtual Stock

no	sid	pid	qty	price
1	s1	1	10	50
2	s2	1	20	60
3	s1	2	30	70
4	s2	2	40	40
5	s1	3	0	100

output

name	qty	sqty	Total
p1	10	30	40
p2	20	70	90
P3	10	0	10

- Locate product with empty virtual stock
- Display total stock for each product

Data Management > SQL > CRUD > Sub Query for a stock condition process

Data: product

pid	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

Data: VirtualStock

no	sid	pid	qty	price
1	s1	1	10	50
2	s2	1	20	60
3	s1	2	30	70
4	s2	2	40	40
5	s1	3	0	100

```

select
    p.name, p.qty
    ( select sum(v.qty)
        from VirtualStock as v
        where p.productId = v.productId
    ) as sqty
    ,
    p.qty + ( select sum(v.qty)
        from VirtualStock as v
        where v.productId = p.productId
    ) as Total
from Product as p

```

output

name	qty	sqty	Total
p1	10	30	40
p2	20	70	90
P3	10	0	10

```

select
    p.productName,
    p.qty,
    (select sum(vs.qty)
        from VirtualStock as vs
        where vs.productId = p.productId
    ) as sqty
    ,
    p.qty+(select sum(vs.qty)
        from VirtualStock as vs
        where vs.productId = p.productId
    ) as `total`
from
    Product as p

```

Data: Product

pid	name	qty	price
1	p1	10	100
2	p2	20	200
3	p3	10	300

Data: VirtualStock

no	sid	pid	qty	price
1	s1	1	10	50
2	s2	1	20	60
3	s1	2	30	70
4	s2	2	40	40
5	s1	3	0	100

Data: Acquisition

no	sid	pid	qty	price	total	status	commit
1	s1	1	10	50	500	yes	no
2	s1	1	10	60	600	yes	no
3	s3	3	20	100	2000	Back-order	no

- Locate product with qty <= 10
- Reorder with qty = 20,
go with lower price first
- If virtual qty =0, put back-order status
- Create stock acquisition log
- Set review field as no

```

select
v.supplierId as sid,
p.productId as pid,
v.qty,
v.price,
"no" as commit

```

```

from
Product as p,
VirtualStock as v

```

```

where
p.productId = v.productId
and
p.qty <=10

```

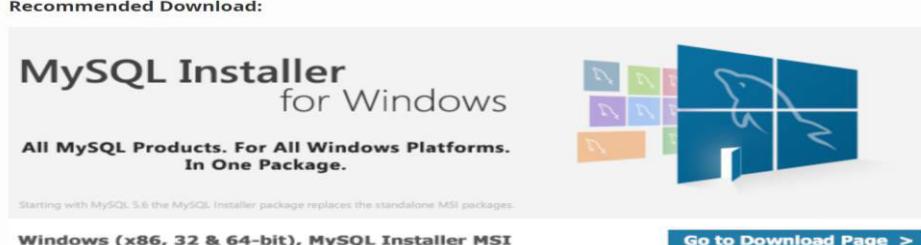
sid	pid	qty	price	commit
s1	1	10	50	no
s2	1	20	60	no
s1	3	0	100	no

SQL Lab Practical class – Let's try this !!!

- Tools & files requirements in our lab practical class
 - **MySQL** – The database management software that requires to import database
 - **PopSQL** – The ease-of-use application which manages database using SQL commands to display the result
 - **mysqlsampledatabase.sql** – The database file that uses to import in MySQL. This file is available in LMS

SQL Lab Practical class – Let's try this !!!

- MySQL installations:
 - Windows
 - Google search: "mysql community server"
 - Click "[**Download MySQL Community Server**](#)"
 - Click this picture below
 - MySQL installations:
 - macOS
 - Google search: "mysql community server"
 - Click "[**Download MySQL Community Server**](#)"
 - Click Operating System: macOS
 - Choose 1 download button from 2 of these (check your MacBook compatible):

macOS 12 (ARM, 64-bit), DMG Archive	8.0.29	445.6M	Download
(mysql-8.0.29-macos12-arm64.dmg)	MD5: da05e15230aad1b35a010b1048b76b85 Signature		
macOS 12 (x86, 64-bit), DMG Archive	8.0.29	452.0M	Download
(mysql-8.0.29-macos12-x86_64.dmg)	MD5: 45b72487279018bf8091a3fc7d278bb8 Signature		
- Recommended Download:**
- 
- MySQL Installer for Windows
All MySQL Products. For All Windows Platforms. In One Package.
- Windows (x86, 32 & 64-bit), MySQL Installer MSI
- Go to Download Page >
- Click Operating System: Microsoft Windows
 - Click "[**Windows \(x86, 64-bit\), MSI Installer**](#)" download button
- 
- Windows (x86, 32-bit), MSI Installer
- 8.0.29 2.3M [Download](#)
- (mysql-installer-web-community-8.0.29.0.msi)
- MD5: 4f735569267527dec28d9e8d977f33d1 | Signature
- Select "**No thanks, just start my download**" (See next page)
- Note for macOS - If your MacBook uses M1 chip, you have to choose macOS 12 (ARM, 64-bit), DMG Archive instead.**

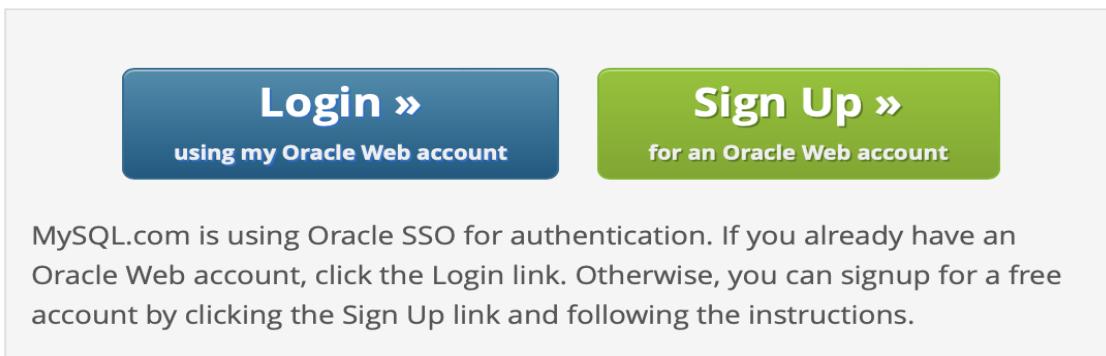
SQL Lab Practical class – Let's try this !!!

④ MySQL Community Downloads

[Login Now](#) or [Sign Up](#) for a free account.

An Oracle Web Account provides you with the following advantages:

- Fast access to MySQL software downloads
- Download technical White Papers and Presentations
- Post messages in the MySQL Discussion Forums
- Report and track bugs in the MySQL bug system



The image shows a snippet of the MySQL.com website. It features two main buttons: a blue 'Login »' button with the subtext 'using my Oracle Web account' and a green 'Sign Up »' button with the subtext 'for an Oracle Web account'. Below these buttons is a text box containing instructions about Oracle SSO authentication. At the bottom of the snippet, there is a link 'No thanks, just start my download.' which is circled in orange.

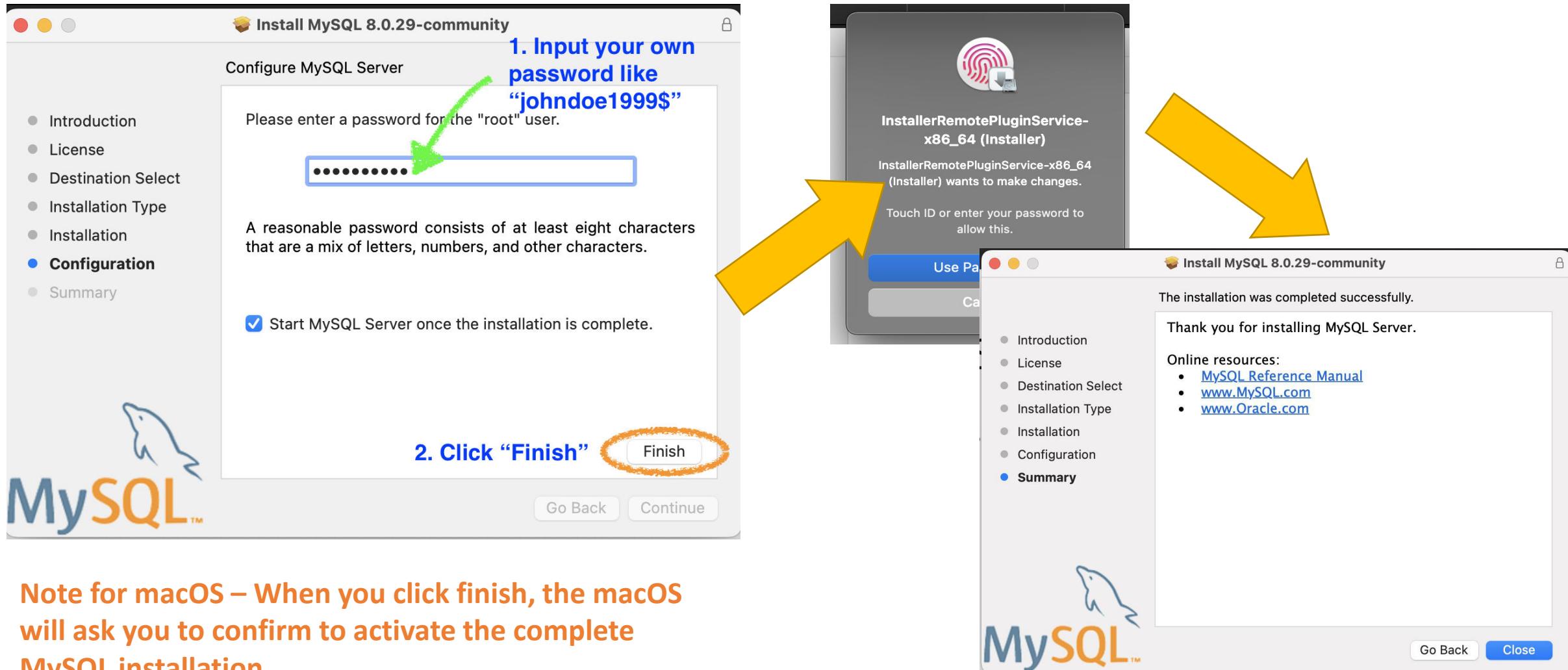
MySQL.com is using Oracle SSO for authentication. If you already have an Oracle Web account, click the Login link. Otherwise, you can signup for a free account by clicking the Sign Up link and following the instructions.

When the MySQL file is complete download, please double click to install and follow the instructions.

SQL Lab Practical class – Let's try this !!!

- MySQL installation setup – macOS
 - Double click installation file “**mysql-8.0.29-macos12-x86_64.dmg**”
 - Double click “**mysql-8.0.29-macos12-x86_64.pkg**”
 - When the setup screen appear, Click “**Allow**” button
 - Click “**Continue**” button and then click “**Agree**” button
 - Click “**Install**” button and input your MacBook password or fingerprint to ready install
 - During installation, the screen will ask you to create the MySQL password for security. Choose “**Use Strong Password Encryption**” and then click “**Next**”
 - Input your MySQL appropriate password including alphabets, numbers, and symbols at least 8 characters (e.g., johndoe1999\$) (See next page)

SQL Lab Practical class – Let's try this !!!

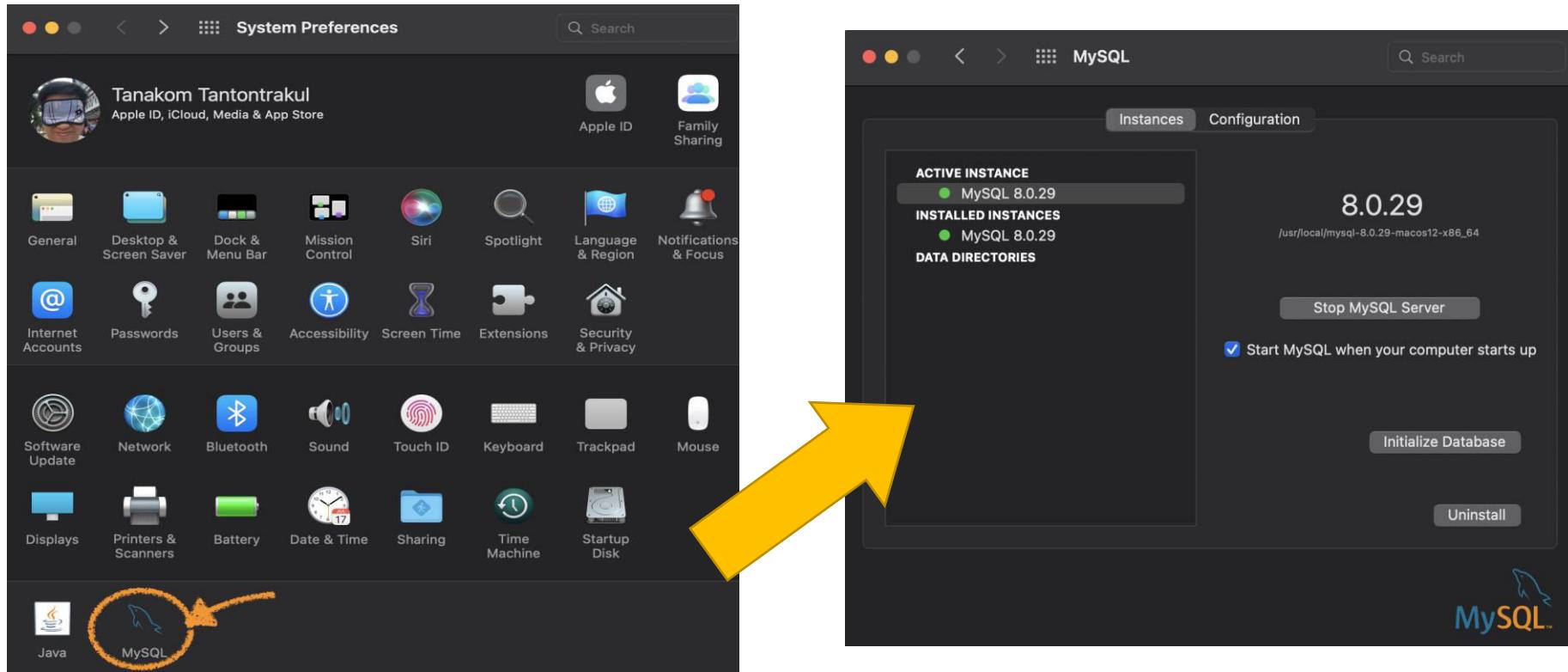


Note for macOS – When you click finish, the macOS will ask you to confirm to activate the complete MySQL installation.

SQL Lab Practical class – Let's try this !!!

- macOS

- When you complete MySQL instatallation, you can check how MySQL activate by type and choose “System Preferences” in Spotlight Search to open. If you see like these pictures below, please close it and follow next step.



SQL Lab Practical class – Let's try this !!!

- Import database into MySQL – macOS
 - Open “Terminal” app in your macOS by typing “Terminal” and choose in Spotlight Search
 - Need some unix commands that allow to enter “mysql” in Terminal by typing and press enter key following like these below:

```
Prompt$ echo 'export PATH=/usr/local/mysql/bin:$PATH' >> ~/.bash_profile  
Prompt$ . ~/.bash_profile
```

- Then you can run mysql by typing and press enter key following like these below (See next page):

```
Prompt$ mysql -u root -p
```

SQL Lab Practical class – Let's try this !!!

```
Tanakoms-MacBook-Pro:~ tanakomtantontrakul$ mysql -u root -p
Enter password:  2. Enter your MySQL password like "johndoe1999$"
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 12
Server version: 8.0.29 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

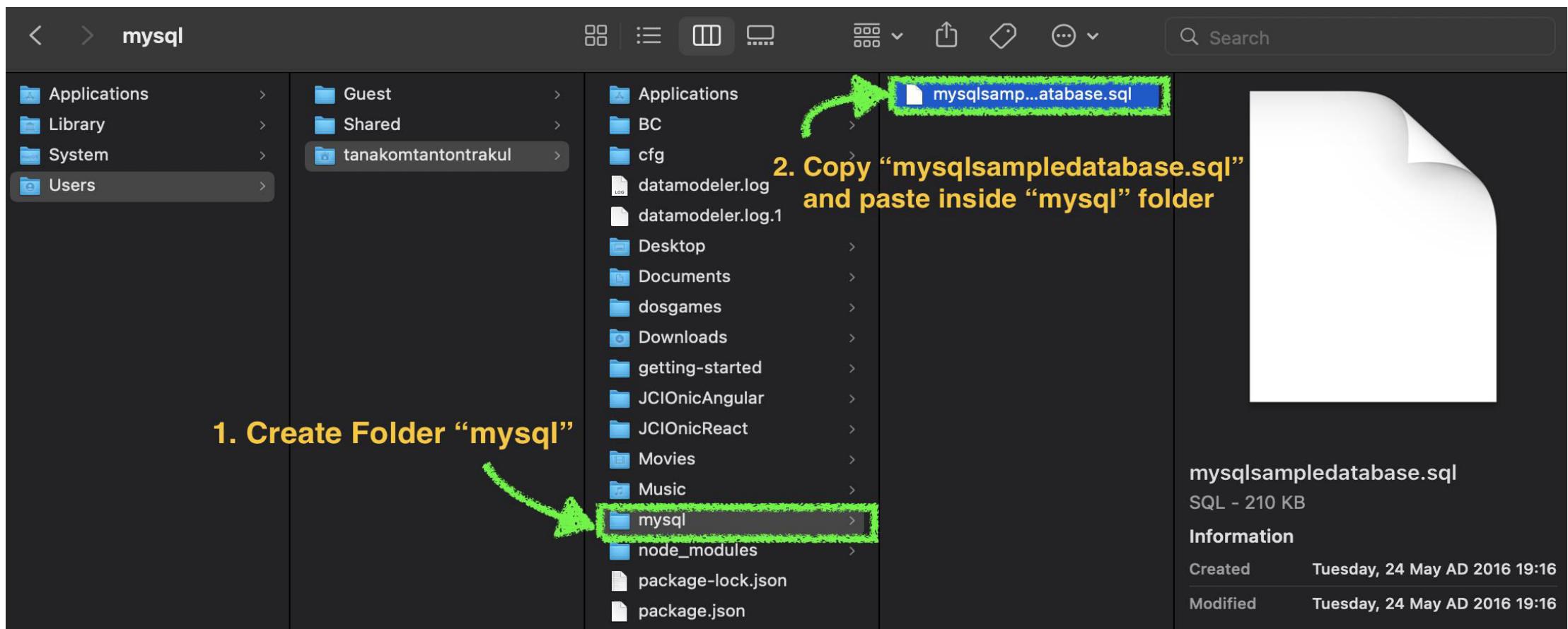
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>  3. If you see this prompt, congratulations and ready to import database file
```

SQL Lab Practical class – Let's try this !!!

- Importing database file – macOS



SQL Lab Practical class – Let's try this !!!

- Importing database file – macOS
 - When you complete input database file into an appropriate folder, please follow these commands to import database file into MySQL

```
mysql> source ./mysql/mysqlsampledatabase.sql
```

← Need to put sql file inside mysql folder of root mac.

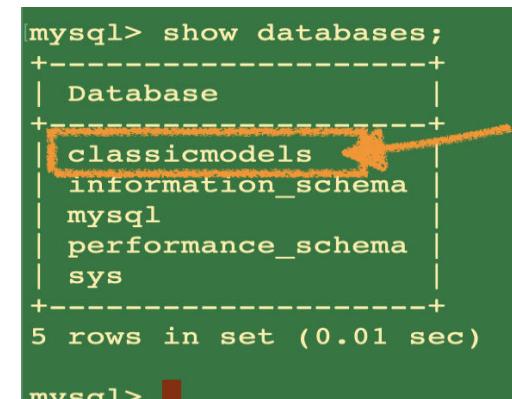
```
mysql> show databases;
```

← Check how many tables inside your database.

```
mysql> use classicmodels;
```

← To select database and ready for query.

Please note that our database name is “classicmodels”. It is inside “mysqlsampledatabase.sql” file.



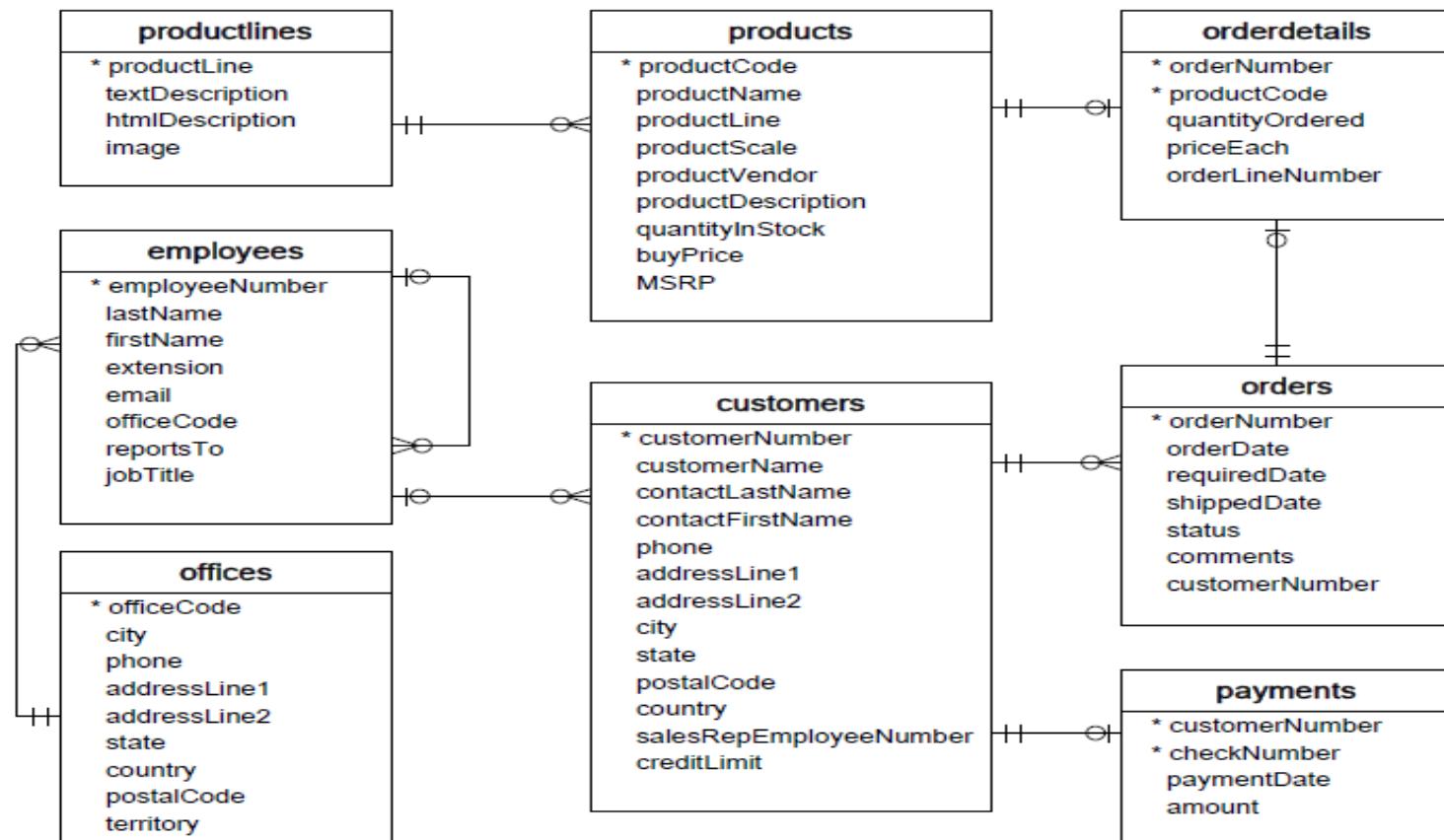
```
mysql> show databases;
+-----+
| Database |
+-----+
| classicmodels |
| information_schema |
| mysql |
| performance_schema |
| sys |
+-----+
5 rows in set (0.01 sec)

mysql>
```

```
mysql> use classicmodels;
Database changed
mysql>
```

SQL Lab Practical class – Let's try this !!!

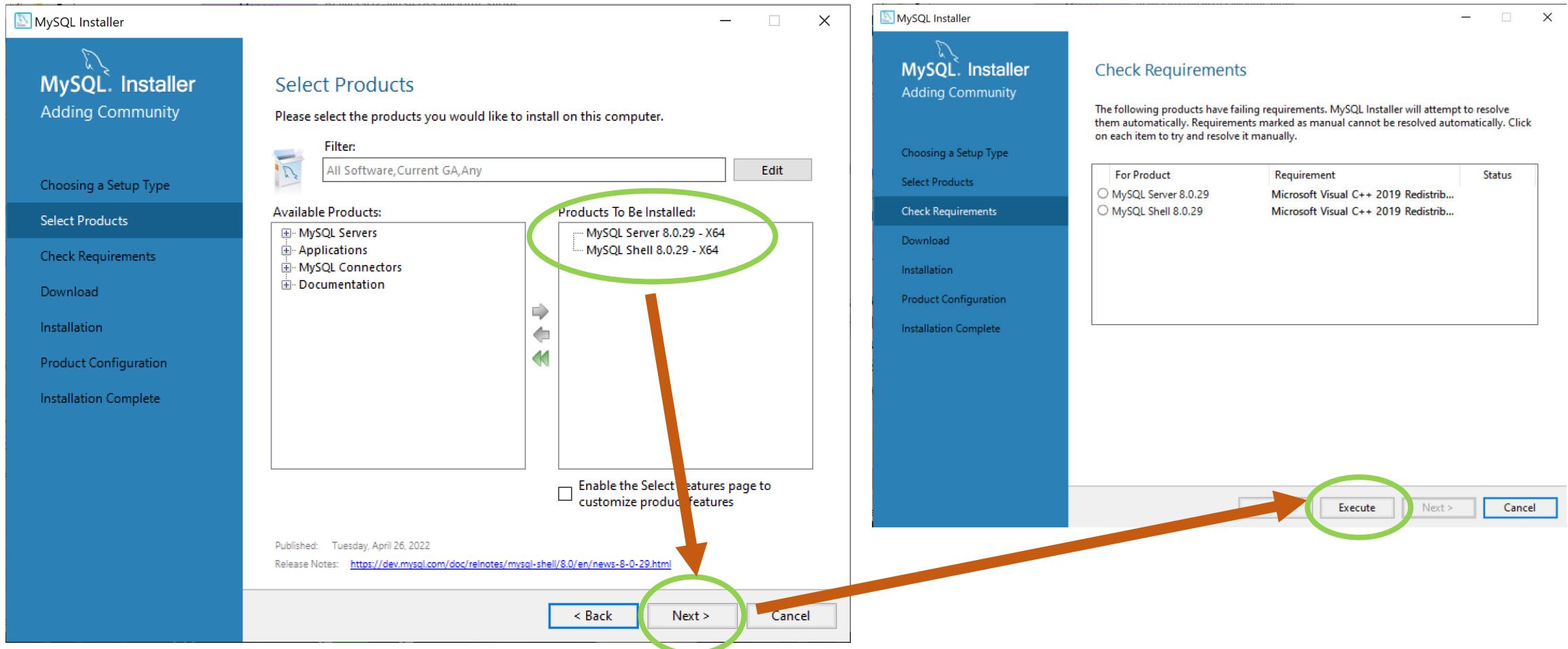
- “classicmodels” Entity Relationship Diagram



SQL Lab Practical class – Let's try this !!!

- MySQL installation setup – Windows
 - Double click “**mysql-installer-web-community-8.0.29.0.msi**”
 - When the setup screen appear, choose “**Custom**” and then click “**Next**”
 - Choose and add **MySQL Server** → **MySQL Server 8.0** → **MySQL Server 8.0.29 – X64**
 - Choose and add **Applications** → **MySQL Shell** → **MySQL Shell 8.0**
(See next page)

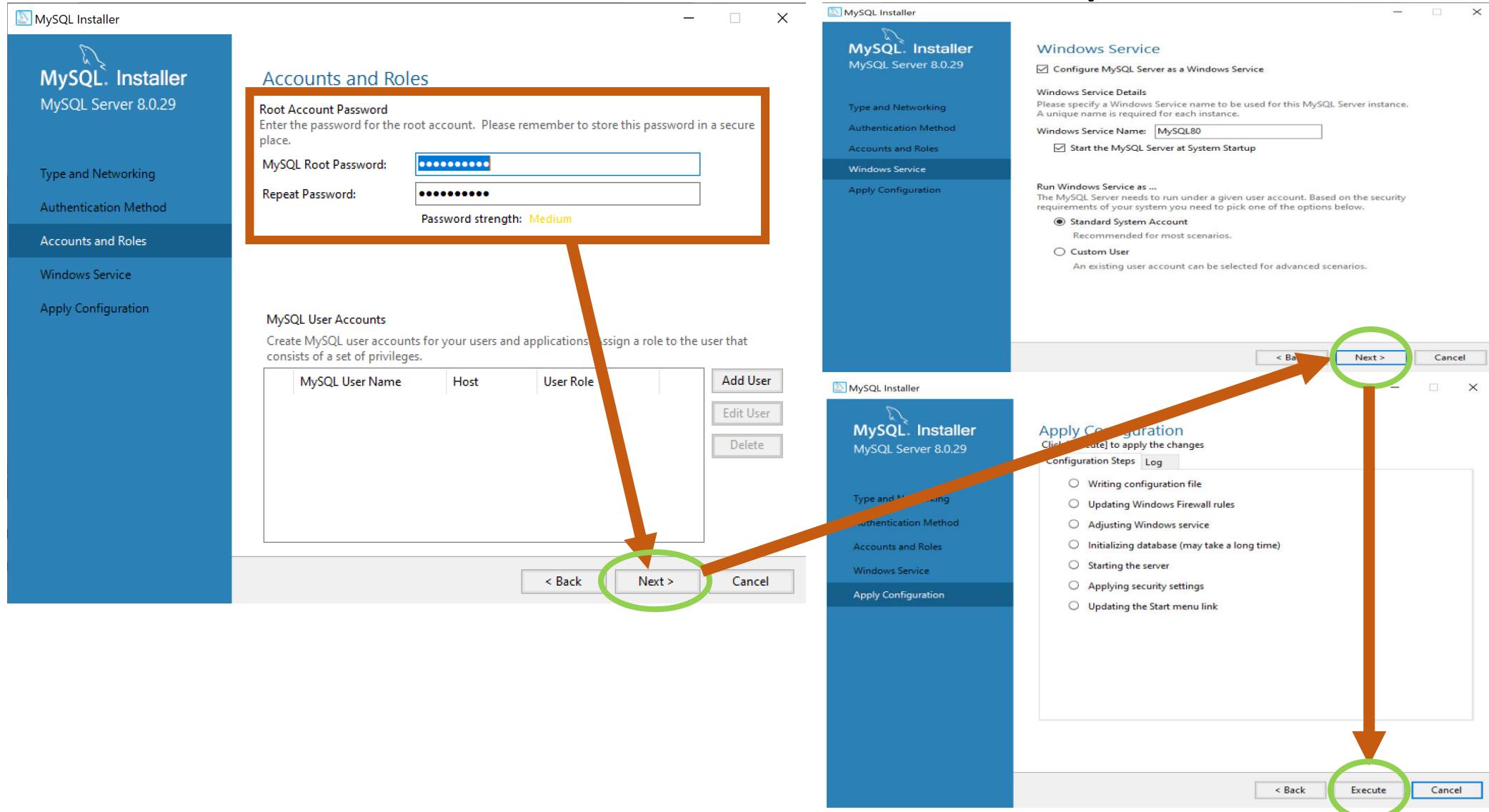
SQL Lab Practical class – Let's try this !!!



SQL Lab Practical class – Let's try this !!!

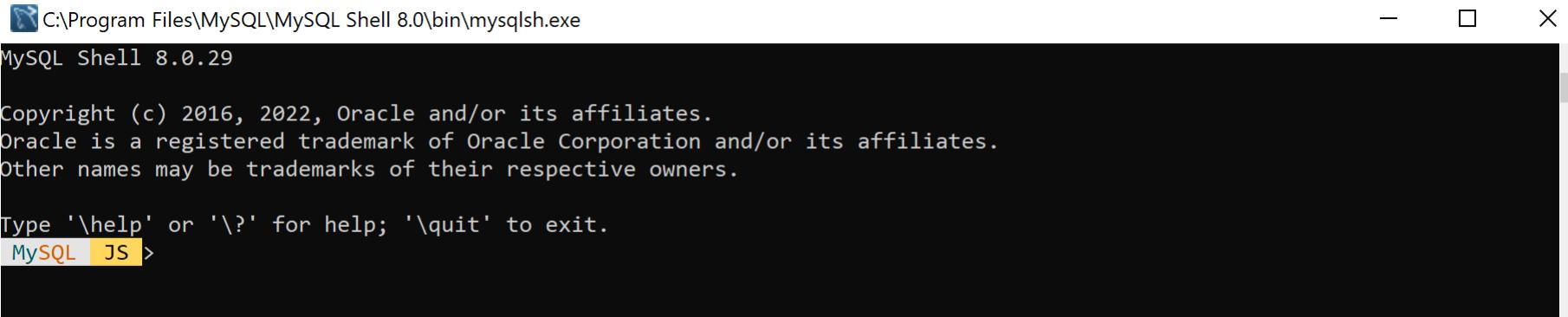
- MySQL installation setup – Windows
 - During download, there are something ask you to download some additional prerequisite tools. Please allow them to install all also
 - When download is completed, click “**Next**” to continue
 - Click “**Execute**” again to install
 - When install is completed, click “**Next**” to continue
 - Click “**Next**” to continue for **Product Configuration**
 - Choose Config Type: “**Development Computer**” and then click “**Next**”
 - In Authentication Method, choose “**Use Strong Password Encryption for Authentication (RECOMMENDED)**” and then click “**Next**”
 - Input your MySQL appropriate password including alphabets, numbers, and symbols at least 8 characters (e.g., johndoe1999\$) and type repeat password, and then click next (See next page)

SQL Lab Practical class – Let's try this !!!



SQL Lab Practical class – Let's try this !!!

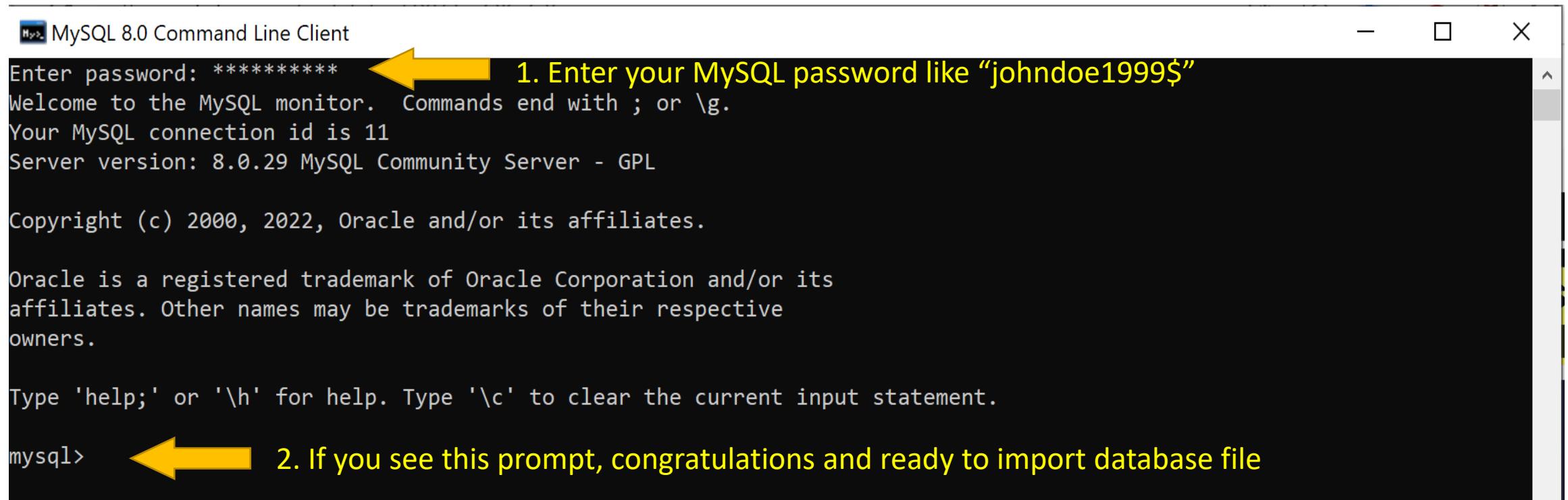
- Windows
 - When you complete MySQL instatallation, you can click “Finish” button at the rest. The **MySQL Shell** will auto open to display like this. **Please close it:**



The screenshot shows a terminal window titled "C:\Program Files\MySQL\MySQL Shell 8.0\bin\mysqlsh.exe". The title bar also displays "MySQL Shell 8.0.29". The window content is a black terminal screen with white text. It starts with copyright information: "Copyright (c) 2016, 2022, Oracle and/or its affiliates. Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners." Below that, it says "Type '\help' or '\?' for help; '\quit' to exit." At the bottom left, there is a prompt "MySQL JS >" followed by a large black rectangular redaction box.

- Choose “**MySQL 8.0 Command Line Client**” directly from **Windows Start Menu**  by typing “**MySQL 8.0 Command Line Client**” and click (See next page)

SQL Lab Practical class – Let's try this !!!



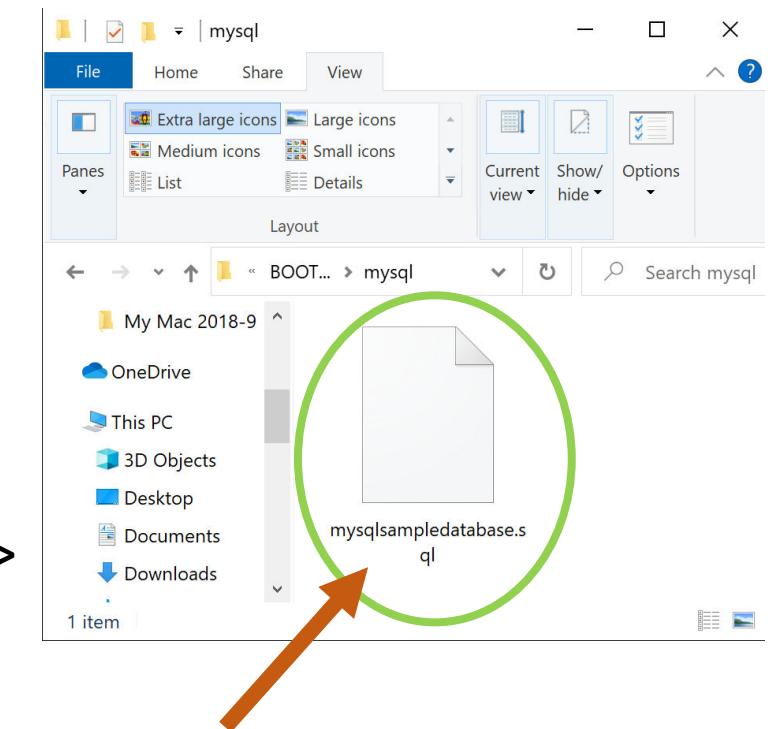
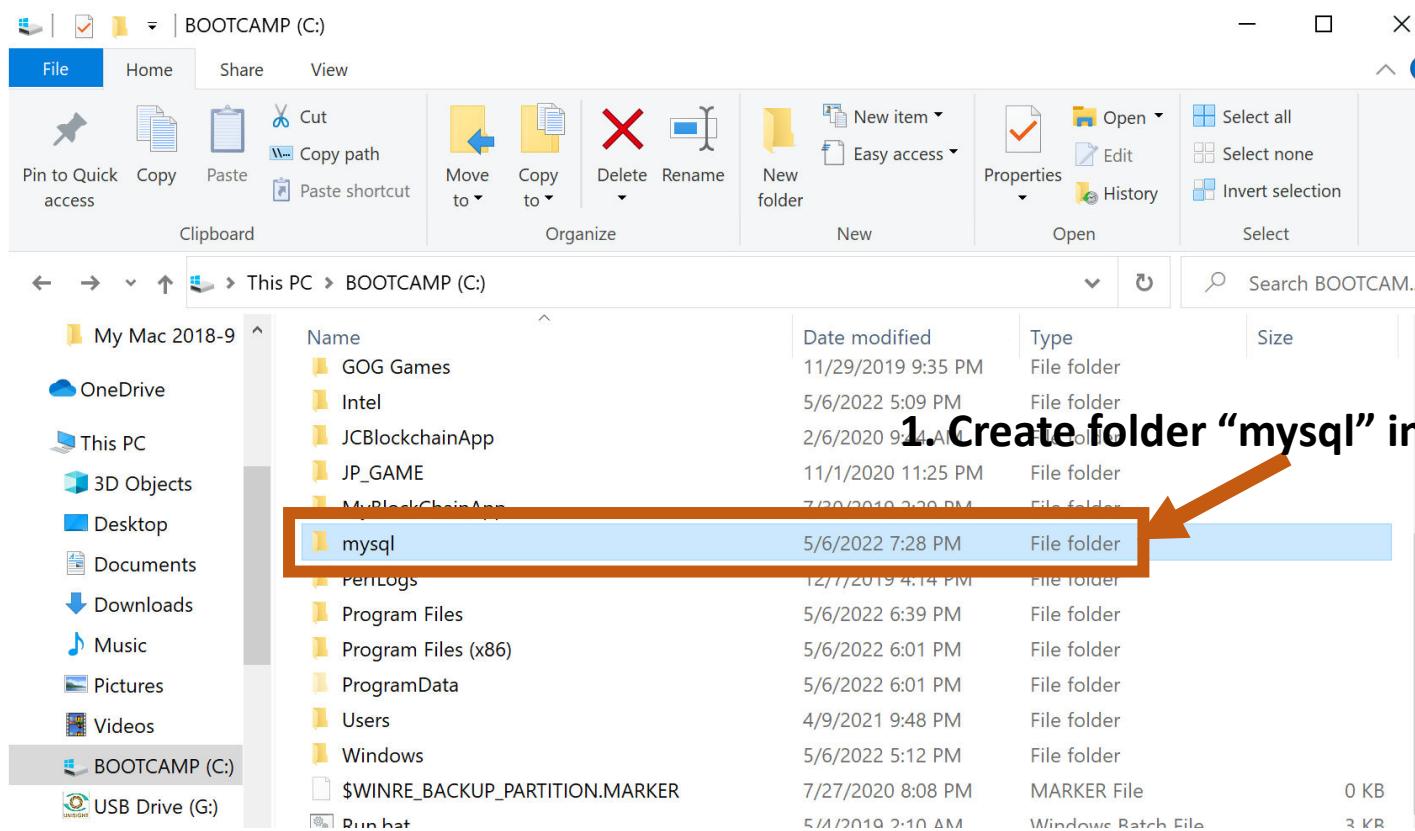
MySQL 8.0 Command Line Client

```
Enter password: *****  
1. Enter your MySQL password like "johndoe1999$"  
Welcome to the MySQL monitor. Commands end with ; or \g.  
Your MySQL connection id is 11  
Server version: 8.0.29 MySQL Community Server - GPL  
  
Copyright (c) 2000, 2022, Oracle and/or its affiliates.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
mysql> 2. If you see this prompt, congratulations and ready to import database file
```

A screenshot of the MySQL 8.0 Command Line Client window. The title bar says "MySQL 8.0 Command Line Client". The main area shows the MySQL welcome message, connection information, server version, copyright notice, and a trademark notice. At the bottom, it provides help instructions and a prompt "mysql>". Two yellow arrows point to specific parts of the text: one points to the password entry field with the instruction "1. Enter your MySQL password like \"johndoe1999\$\"", and another points to the "mysql>" prompt with the instruction "2. If you see this prompt, congratulations and ready to import database file".

SQL Lab Practical class – Let's try this !!!

- Importing database file – Windows



SQL Lab Practical class – Let's try this !!!

- Importing database file – Windows
 - When you complete input database file into an appropriate folder, please follow these commands to import database file into MySQL

```
mysql> source c:\mysql\mysqlsampledatabase.sql
```

← Need to put sql file inside mysql folder of root as C:\>.

```
mysql> show databases;
```

← Check how many tables inside your database.

```
mysql> use classicmodels;
```

← To select database and ready for query.

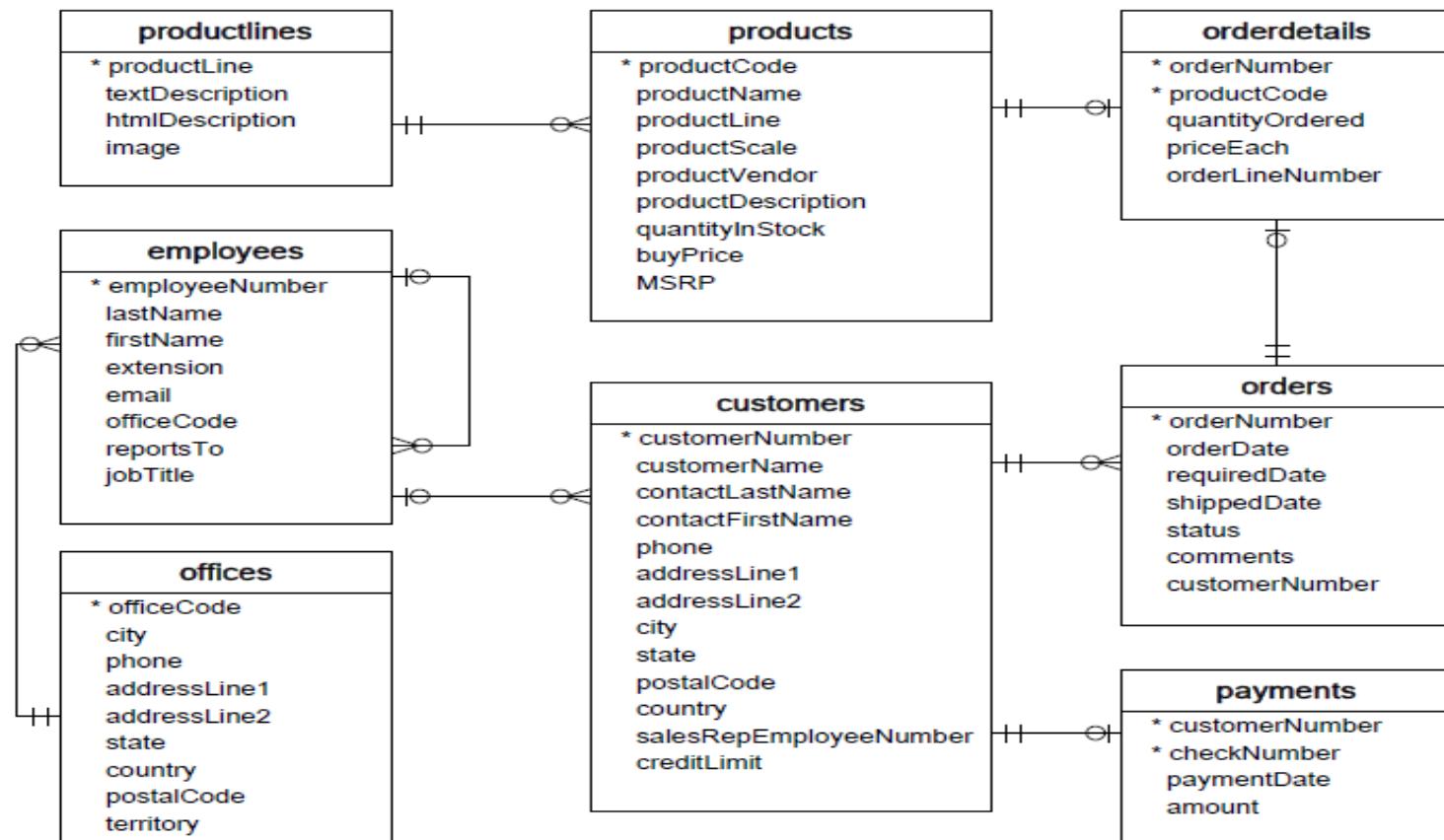
Please note that our database name is “classicmodels”. It is inside “mysqlsampledatabase.sql” file.

```
+-----+  
| Database          |  
+-----+  
| classicmodels    |  
| information_schema|  
| mysql             |  
| performance_schema|  
| sys               |  
+-----+  
5 rows in set (0.00 sec)  
  
mysql>
```

```
mysql> use classicmodels;  
Database changed  
mysql>
```

SQL Lab Practical class – Let's try this !!!

- “classicmodels” Entity Relationship Diagram



SQL Lab (Let's Try This!)

- Form your group of 3 persons
- Provide the video clip of demonstration how to setup MySQL and import sql file from your computer
 - **Hint:** You can meet and discuss to do in your own **private MS Teams chat** and save video into **mp4** file format
- You can choose which OS you would like to do (Windows, macOS)

SQL Lab Practical class – Let's try this !!!

- PopSQL Installations and Setup

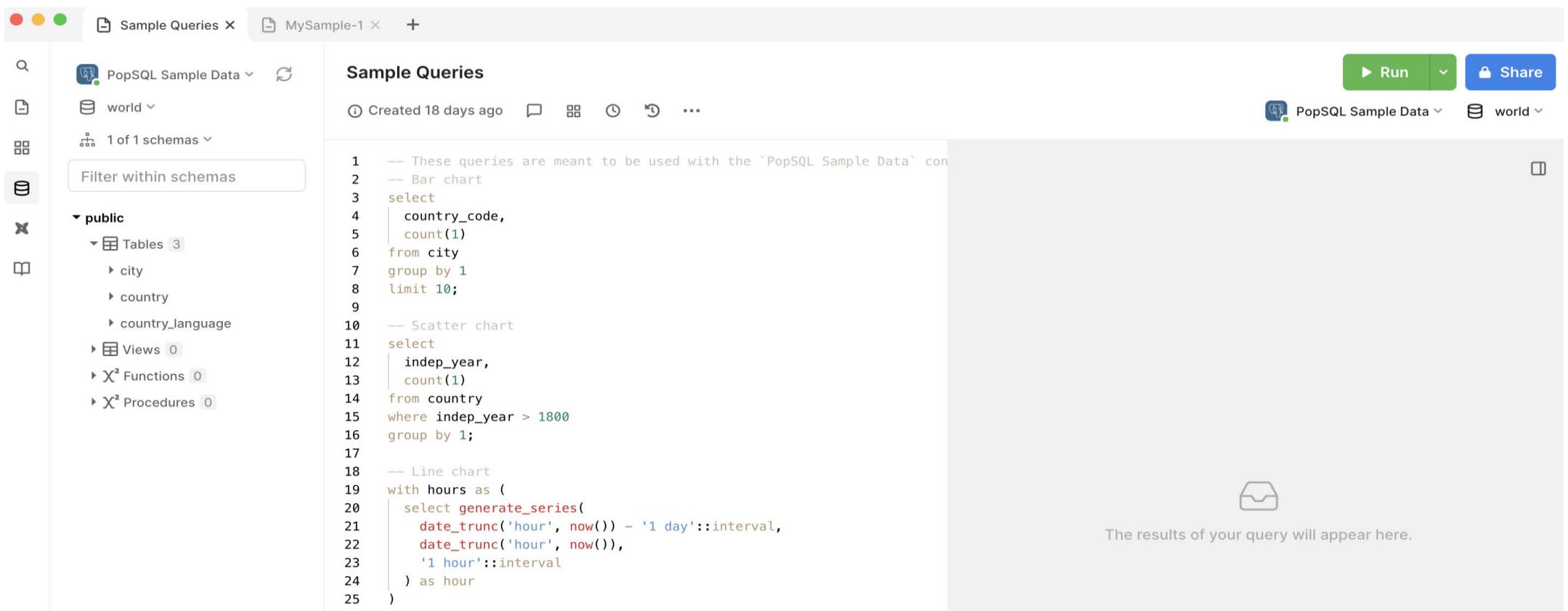
- Go to <https://popsql.com>
- Click “**Get started for Free**”
- Sign up your account and remember account to access any time
- Choose “**Download PopSQL for desktop**”



- Double click PopSQL to install and follow the instructions

SQL Lab Practical class – Let's try this !!!

- When complete install PopSQL and open, the screen will display like this:



The screenshot shows the PopSQL interface. On the left, there is a sidebar with icons for search, file, grid, and table. It displays the 'PopSQL Sample Data' connection, the 'world' schema, and a message indicating '1 of 1 schemas'. A 'Filter within schemas' button is present. Below this, the 'public' schema is expanded, showing 'Tables' (city, country, country_language), 'Views' (0), 'X² Functions' (0), and 'X² Procedures' (0). The main area is titled 'Sample Queries' and shows three SQL scripts. The first script is a bar chart query, the second is a scatter chart query, and the third is a line chart query using a generated series of hours. The right side of the interface has a 'Run' button, a 'Share' button, and a results panel with a placeholder message: 'The results of your query will appear here.'

```
1 -- These queries are meant to be used with the `PopSQL Sample Data` con
2 -- Bar chart
3 select
4   | country_code,
5   | count(1)
6   from city
7   group by 1
8   limit 10;
9
10 -- Scatter chart
11 select
12   | indep_year,
13   | count(1)
14   from country
15   where indep_year > 1800
16   group by 1;
17
18 -- Line chart
19 with hours as (
20   select generate_series(
21     date_trunc('hour', now()) - '1 day'::interval,
22     date_trunc('hour', now()),
23     '1 hour'::interval
24   ) as hour
25 )
26
```

SQL Lab Practical class – Let's try this !!!

- Setup imported MySQL database into PopSQL
 - Click “**PopSQL Sample Data**” and then choose “**Manage connections**”
 - Click “**Add new connection**” and then choose “**MySQL**”
 - Input information like this picture below, and click “**Test**” to make sure that your imported database is connected and then click “**Connect**”
(See next page)

SQL Lab Practical class – Let's try this !!!

<  Edit My Classic Model

Connection name

Hostname/Port

Database

Username/Password

Credentials will be securely stored in your computer's keychain.

Connection Type **Connect directly from my computer**
Direct: Use your computer's network to connect to your database.

Who has access?

 Tanakom Tanto... Owner
Grant granular permissions to teams and specific users.
[Talk to sales](#) to add this to your team.

Advanced Options **Connect over SSH**
Useful to connect to a private network.

SSL options
SSL is used when available

Test **Save** **Connect**

Looking for help?

-  [MySQL connection guide](#)
-  [Connection Types](#)
-  [Sharing Connections](#)

Whitelist IPs

If you maintain an IP whitelist, you'll need to whitelist PopSQL static IPs:

**23.20.131.72,
54.211.234.135**

SQL Lab Practical class – Let's try this !!!

- PopSQL Screen Overview

The image shows the PopSQL interface. On the left, there's a sidebar with a 'classicmodels' schema expanded, showing tables like customers, employees, and offices. A red box highlights this sidebar with the text 'Display database table'. In the center, a query editor window titled 'MySample-1' shows several SQL commands. A red box highlights this window with the text 'Type SQL commands in this area, select one part, and click Run button'. The first command is 'select * from customers;'. To the right, a results viewer window titled 'Data result area' shows a table of customer data with columns for customerNumber, customerName, and contactLastName. A red box highlights the 'Run' button in the top right of the query editor and the 'Data' tab in the results viewer. Red text above the 'Run' button says 'Click Run to show the data result'.

Sample Queries × MySample-1 × +

My Classic Model classicmodels 1 of 1 schemas Filter within schemas

classicmodels

- Tables 8
 - customers
 - employees
 - offices
 - orderdetails
 - orders
 - payments
 - productlines
 - products
- Views 0
- X² Functions 0
- X² Procedures 0

MySample-1

Created 17 days ago

```
1 --Show all customer data
2 select *
3   from customers;
4
5 --Show some customer data
6 select customerNumber, customerName, phone, city, country
7   from customers
8   limit 50;
9
10 --Display total customers per country data
11 select country, count(1)
12   from customers
13   group by 1;
```

Run Share

My Classic Model classicmodels

Clear all

MySample-1

Success (122 rows) 0.2 s 8:50 PM

Explore SQL Data Chart Export

customerNumber	customerName	contactLastName
103	Atelier graphique	Schmitt
112	Signal Gift Stores	King
114	Australian Collectors, Co.	Ferguson
119	La Rochelle Gifts	Labrune
121	Baane Mini Imports	Bergulfsen
124	Mini Gifts Distributors Ltd.	Nelson
125	Havel & Zbyszek Co	Piestrzaniewi
128	Blauer See Auto, Co.	Keitel
129	Mini Wheels Co.	Murphy
131	Land of Toys Inc.	Lee

Type SQL commands in this area, select one part, and click Run button

Click Run to show the data result

1 to 10 of 122 |< < Page 1 of 154 >|

SQL Lab Practical class – Let's try this !!!

- Exporting file into CSV, JSON

The screenshot shows the SQL Lab interface with a database named 'MySample-1'. A table named 'Customer' is displayed with 122 rows. The 'Export' button in the top navigation bar is highlighted with an orange box. A context menu is open over the table, listing options: 'Copy TSV', 'Copy CSV', 'Copy JSON', 'Export CSV' (which is highlighted with a blue box), and 'Export JSON'. Another context menu is open on the right side of the screen, titled 'Export MySample-1', with fields for 'Save As:' (set to 'MySampleData.csv'), 'Where:' (set to 'Desktop'), and 'Format:' (set to 'CSV'). The 'Export' button in this menu is also highlighted with an orange box.

customerNumber	customerName	contactLastN	contactFirstN	phone	addressLine1	addressLine2	city	state	postalCode	country	salesRepEmpl	creditLimit
103	Atelier graphique	Schmitt							44000	France	1370	21000
112	Signal Gift Stores	King						Nantes				
114	Australian Collectors, Co.	Ferguson						Las Vegas	83030	USA	1166	71800
119	La Rochelle Gifts	Labrune						Melbourne	3004	Australia	1611	117300
121	Baane Mini Imports	Beraulfsen						Victoria				
124	Mini Gifts Distributors Ltd.								44000	France	1370	118200
125	Havel & Zbyszek Co								4110	Norway	1504	81700
128	Blauer See Auto, Co.								97562	USA	1165	210500
129	Mini Wheels Co.								01-012	Poland		0
131	Land of Toys Inc.								60528	Germany	1504	59700

1 to 10 of 122 |< < >|>|

55

Basic SQL Commands

- Try to input SQL commands in PopSQL

- Display all fields in products table

```
select *  
from products;
```

- Display some fields in products table with limit by 10

```
select productCode, productName, quantityInStock, buyPrice  
from products  
limit 10;
```

- Try some more using “select” command in other tables

Basic SQL Commands

- Try to input SQL commands in PopSQL
 - Sort all product name in an ascending order

```
select *  
from products  
order by productName asc;
```

- Sort all product name in a descending order

```
select *  
from products  
order by productName desc;
```

- Try some more using an enhance command in other tables (Check **classicmodels** entity relationship diagram carefully)

Basic SQL Commands

- Try to input SQL commands in PopSQL
 - Display all fields where productCode is ‘S12_1666’ in products table

```
select *  
from products  
where productCode = 'S12_1666';
```

- Display some fields where quantityInStock >= 300 and buyPrice > 50 in products table

```
select productCode, productName, quantityInStock, buyPrice  
from products  
where quantityInStock >= 300 and buyPrice > 50;
```

- Try some more using “select” command in other tables

Basic SQL Commands

- Try to input SQL commands in PopSQL
 - Display some fields with adding “total” field where quantityInStock > 300 and buyPrice > 50 in products table

```
select productCode, productName, quantityInStock, buyPrice, quantityInStock * buyPrice as total  
from products  
where quantityInStock > 300 and buyPrice > 50;
```

- Try some more using “select” command in other tables

Basic SQL Commands

- Some more SQL commands with PopSQL
 - Insert new item in products table

```
insert into products (productCode, productName, productLine, productScale, productVendor,  
productDescription, quantityInStock, buyPrice, MSRP)  
value ('S11_1050', '2022 Diamond Supra', 'Classic Cars', '1:15', 'Autocart Studio Design', 'Upgrade classic  
mode for enhancing', '1200', '75.25', '99.99');
```

- Update current item in products table

```
update products  
set buyPrice = '85.78', MSRP = '120.12'  
where productCode = 'S11_1050';
```

- Delete current item in products table

```
delete from products  
where productCode = 'S11_1050';
```

- Try some more using “insert”, “update”, “delete” command in other tables

Basic SQL Commands

- Some more SQL commands with PopSQL
 - Enhancing two tables between “products” and “orderDetails”

```
select p.productCode as pid, o.orderNumber as oid, p.quantityInStock as qtyStock, o.quantityOrdered as qtyOrdered  
from products as p, orderDetails as o  
where p.productCode = o.productCode;
```

```
select p.productCode as pid, o.orderNumber as oid, p.quantityInStock as qtyStock, o.quantityOrdered as qtyOrdered  
from products as p, orderDetails as o  
where p.productCode = o.productCode and p.quantityInStock >= 300;
```

- Display all tables in database

```
show full tables;
```
- Try some more using an enhance command in other tables (Check **classicmodels** entity relationship diagram carefully)

SQL Lab Assignment 2

- Form your group of 3 persons
- Download **world.sql** from LMS and import database using MySQL
- Setup **world** database that connect into PopSQL
- Create powerpoint to present the SQL and result follow the questions:
 - How many tables have you seen in world database? List all table names and fields.
 - List first 20 countries's name, continent, and population
 - List all cities's name, country code, and distinct only in Australia (**Hint:** check country code)
 - List the first 50 countries's name, surface area, population, and life expectancy with greater than 500,000 people of population and less than 60 years of life expectancy
 - List all city's name, country's continent, country's name, and country's population in the North America. Display city's name in an ascending order. How many city name in the list? (**Hint:** Check how to join country code)
- Submit into LMS through the link provided (Deadline also be noticed in LMS)
- Need one member per group to submit

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

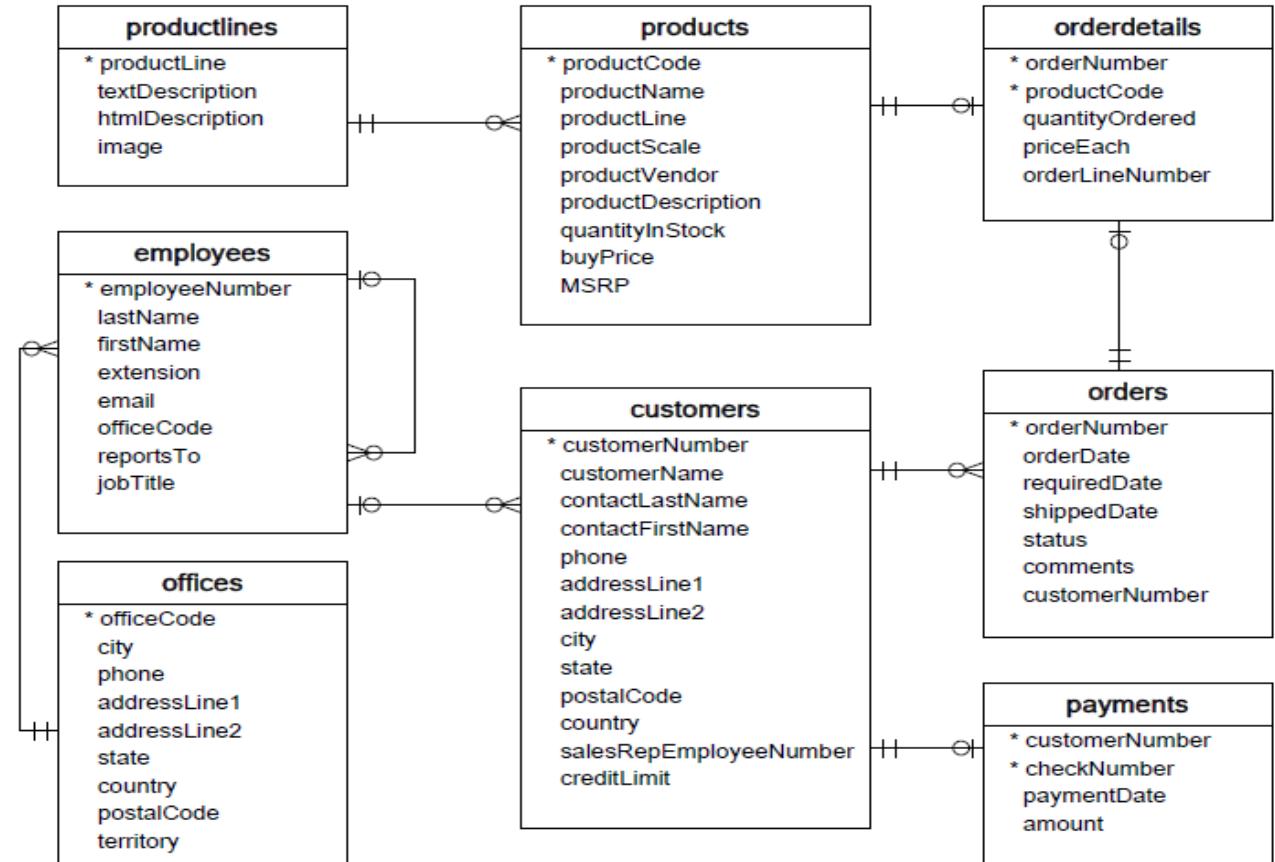
- Sorting
- Filtering
- Projection

Remind your Tools for Data Alchemist

- MySQL
- PopSQL
- classicmodels database



POPSQL



Data Alchemist - Sorting

- 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	customerNumber	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	Rlaier See Auto Co	+49 69 66 90 2555	Frankfurt	Germany
10	129				
11	131	(Select All)			
12	141	✓ 103			
13	144	✓ 112			
14	145	✓ 114			
15	146	✓ 119			
16	148	✓ 121			
17	151	✓ 124			
18	157	✓ Auto Apply			
19	161	Technics Stores Inc.			
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

customerNumber	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	Rlaier See Auto Co	+49 69 66 90 2555	Frankfurt	Germany
customerNumber	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	Montral	Canada
227	Heintze Collectables	86 21 3555	Vrhus	Denmark
223	Natrlich 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 Kong	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;
```

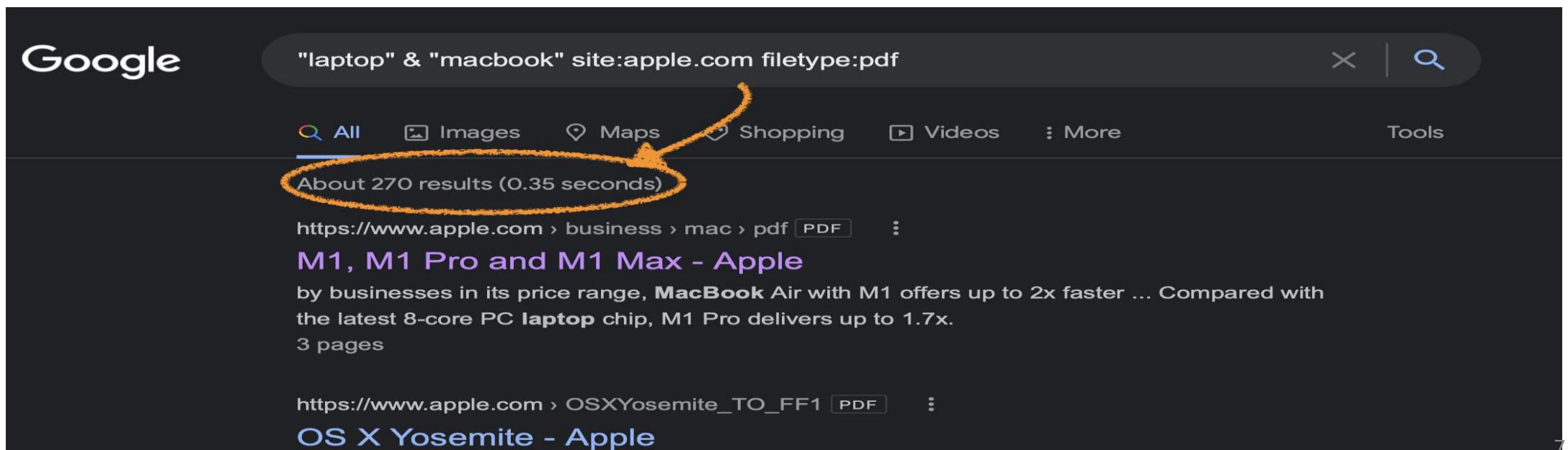
- **order by fieldname desc;** – Sorting texts or numbers in a descending order (e.g., Z-A, 9-0)

```
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

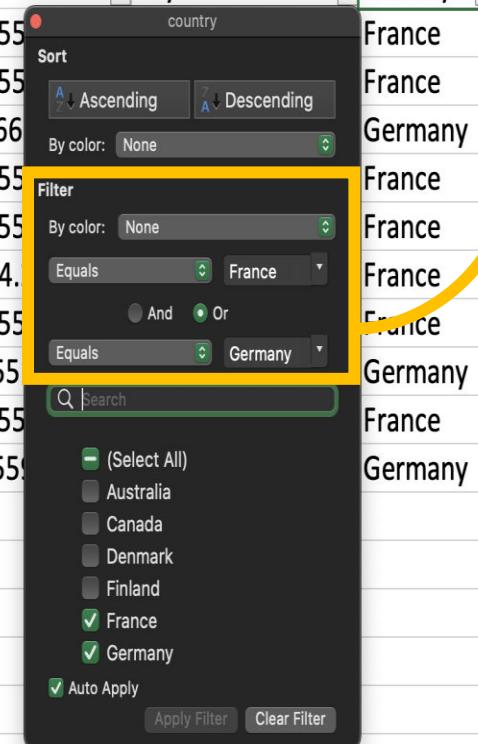
- Filtering is the method that you use some commands with conditions to specify a subset of the data items
- The benefit is to help you to get the exactly result with less data records (like using the keyword to search command in Google)



Data Alchemist - Filtering

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

A	B	C	D	E
customerNumber	customerName	phone	city	country
103	Atelier graphique	40.32.2555		France
119	La Rochelle Gifts	40.67.8555		France
128	Blauer See Auto, Co.	+49 69 66		Germany
146	Saveley & Henriot, Co.	78.32.5555		France
171	Daedalus Designs Imports	20.16.1555		France
172	La Corne D'abondance, Co.	(1) 42.34.		France
209	Mini Caravy	88.60.1555		France
223	Natv&rlieh Autos	0372-5555		Germany
242	Alpha Cognac	61.77.6555		France
247	Messner Shopping Network	069-0555		Germany



customerNumber	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	Natv&rlieh 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:

1. > (greater than)
2. < (less than)
3. = (equal)
4. >= (greater than or equal)
5. <= (less than or equal)
6. != (not equal)
7. 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
customerNumber	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
129	Mini Wheels Co.	6505555787	San Francisco	USA
131	Land of Toys Inc.	2125557818	NYC	USA
141	Euro+ Shopping Channel	(91) 555 94 44	Madrid	Spain
144	Volvo Model Replicas, Co	0921-12 3555	Lulev•	Sweden
145	Danish Wholesale Imports	31 12 3555	Kobenhavn	Denmark
146	Saveley & Henriot, Co.	78.32.5555	Lyon	France
148	Dragon Souveniers, Ltd.	+65 221 7555	Singapore	Singapore
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 3

- 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?

- 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 4

- 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)
- Need one member per group to submit

Topic 04: Data Lifecycle and Data Quality

BDM3302: Data Management

What is Data Lifecycle Management?

- Data lifecycle management (DLM) is an approach to managing data throughout its lifecycle, from data entry to data destruction. Data is separated into phases based on different criteria, and it moves through these stages as it completes different tasks or meets certain requirements. (IBM)



Phases of Data Lifecycle Management



Phases of Data Lifecycle Management

- A data lifecycle consists of a series of phases over the course its useful life. Each phase is governed by a set of policies that maximizes the data's value during each stage of the lifecycle. DLM becomes increasingly important as the volume of data that is incorporated into business workstreams grows.

Phase 1: Data Creation

- Data creation occurs throughout organizations. This stage involves acquiring data from internal sources like customer data. Most data is created by business functions such as finance, marketing, sales, and human resources rather than the IT department whose typically responsible for managing the data once it's been created.
- Some data can be collected from external like market research companies. Once the data is collected, it must be cleansed and organized so that it can be used effectively. Data collection can be a time-consuming and costly process, but it is essential for businesses to have accurate and up-to-date data.



Phases of Data Lifecycle Management



Phase 2: Data Storage

- Once data has been created, it is typically stored on a computer hard drive or in a secure location. Maintaining data so it can be accessible and usable. Data must be backed up regularly to prevent loss in disaster like earthquake, fire, tsunami. Storage also involves backups that makes the data remains protected. Once data is stored, responsibility for its management typically falls to the IT or security team.
- Storage protections include access control around who can read and overwrite the data, device control such as data encryption, backups to protect the data from loss, plus security measures to protect the backups themselves.



Phases of Data Lifecycle Management

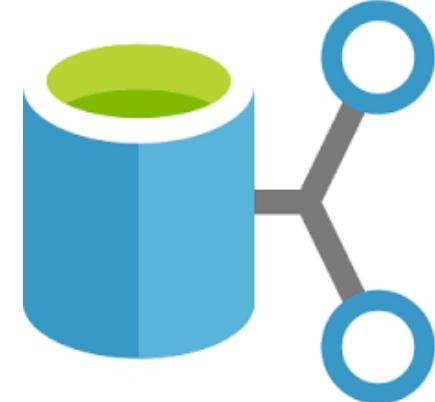


Phase 3: Data Usage

- This stage involves using the data to make decisions and take action. Data can be used for marketing, product development, or any other purpose. During this stage, data is accessed, viewed or processed. Protecting data during this phase will usually fall equally between the lines of business and the IT department.
- Protections during data usage include access control, encryption, data rights management for copyrighted information and data loss prevention, which involves software and business rules to prevent unauthorized access to sensitive information.

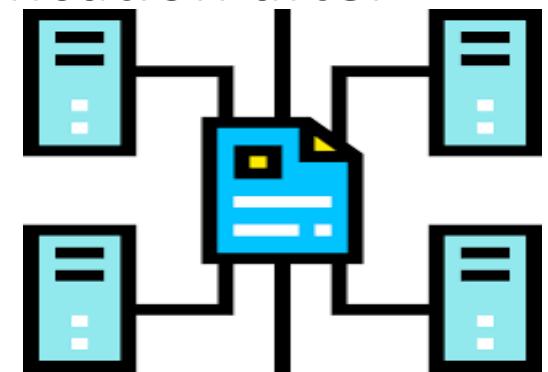


Phases of Data Lifecycle Management



Phase 4: Data Sharing

- Data is often shared amongst internal employees and to corporate partners outside of the organization. Data sharing can occur through the network, via removable media, or across the internet via transfer sites (FTP) or email. When data is shared, it can be beneficial to business, but it also carries some risks.
- Data sharing safeguards involve access control, encryption, network security (firewalls/intrusion detection) and data loss prevention. When organizations are dealing with third-party vendors, they should have clear measures in place for data removal and verification after services have ceased.



Phases of Data Lifecycle Management



Phase 5: Data Archiving

- This phase involves storing data in a safe place for future use. Data can be stored on servers, in the cloud, or physical media. For short-term data protection, all data must be backed up regularly, either onsite or offsite. When an organization needs to retain data for the long term, it can be archived to tape or disk media and placed in remote, secure locations.
- An organization's operations team would usually take responsibility for archiving as opposed to IT or the lines of business. Protecting archived data include access control and encryption.

Phases of Data Lifecycle Management



Phase 6: Data Destroy

- When an organization's data reaches the end of its life, it must be permanently erased. Determining which data is erased, how it's erased and how that erasure is verified depends on several factors, such as content type, usage needs and regulatory requirements.
- When it's considered at all, the "Destroy" phase is most often addressed by the operations team. But when managed properly, end-of-life data destruction is truly the responsibility of all stakeholders, from IT to the lines of business.
- However, in practice, organization would like to have one more column to identify that a record (customer) data is active or inactive. Because when you deleted that record from database, it will be disappeared.

Reasons Why NOT To Delete Data

- Need to keep records for historical purposes
 - Accounting, reporting, legal, etc.
- Need to go back and look up information
- Legally required to keep information
 - Accounting data (Tax info, employee withholdings)
 - Regulatory authorities (stock trades, etc.)
 - Discovery for lawsuits





Data Quality Management

- Critical business decisions and allocation of resources are made based on what is found in the data.
- Prices are changed, marketing campaigns created, customers are communicated with, and daily operations evolve around whatever data points are churned out by an organization's various systems.
- The data that serves as the foundation of these systems *must be good quality data*. Otherwise, we fail before we ever begin. It doesn't matter how pretty the screens are, how intuitive the interfaces are, how high the performance rockets, how automated the processes are, how innovative the methodology is, and how far-reaching the access to the system is, *if the data are bad—the systems fail*. And if the systems fail, or at the very least provide inaccurate information, every process, decision, resource allocation, communication, or interaction with the system will have a damaging, if not disastrous impact on the business itself.

Data Quality Management

- Data quality management (DQM) is a business principle that relies on the skillsets of people, processes, and technologies to ensure high-quality data throughout your organization.
- The goal of DQM extends beyond maintaining and improving data quality. In the long term, well managed data is the key to making educated business decisions and attaining your desired business outcomes. Teams must be able to trust that the data they're accessing is correct, current, and consistent across the board.



Data Quality Dimensions

- Data Quality Dimensions refers to an aspect or feature of information that can be assessed and used to determine quality of data.
- There are 6 keys for Data Quality Dimensions:
- **Accuracy** – Data accurately represent the “real-world” values. (incorrect of customer names or addresses) Data must be both accurate and precise enough for their intended use. Knowing sales accurately is important, but for many decisions, knowing sales only to the nearest \$1000 per month for each product is sufficient.
- **Validity** – Data conforms to the syntax (format, type, range) of its definition. (incorrect customer type)
- **Timeliness** – means meeting the expectation for the time between when data are expected and when they are readily available for use. (changing customer address which is effective on June 1st is entered into the system on June 15th)
- **Completeness** – Data are complete in terms of required potential of data or Not Null. (customer address **missing** postal code or zip code)
- **Uniqueness** – Data are properly identified and recorded only once. (one customer is recorded twice in database, with different identifiers)
- **Consistency** - Data are represented consistently across the data set. (customer account is closed, but there is a new order associated to that account)



Why is Data Quality Management Important?

- The benefits of DQM ripple across every department in your organization. Reliable data reveals trends, informs proactive strategies, and boosts team performance and efficiency. Here are a few more benefits you can expect from data quality management:
 - Correct data creates more efficient data processes and more informed business decisions
 - Quality analytics offer a better view of customers, prospects, vendors, partners, etc.
 - Departmental alignment increases when analytical insights can be shared across teams
 - Businesses save money in the long term by avoiding initiatives that are unsupported by their data
 - Proper data quality helps streamline data governance procedures as well
 - Satisfaction and trust improve when a business can reference comprehensive customer data

Phases of Data Quality Management





1. Data Profiling

- The first step towards proper data quality, businesses review their data in detail to uncover issues as they relate to their quality goals. Data profiling looks at whether the **format and content of the data** matches its metadata and whether it is accurate, complete, and valid. Are there blank values? Duplicate data? Strange patterns? These types of questions can be answered with proper data profiling.
- To perform profiling, consider these three aspects:
- **Data structure discovery:** How is the data formatted? Is that consistent with the data standards the business is envisioning? Does the data pass mathematical checks such as the ability to accurately calculate sums?
- **Data content discovery:** Are there any fields or values that are incorrect or null? Do any specific rows in a table have issues? Are there any recurring patterns of concern? For example, are most phone numbers missing area codes?
- **Data relationship discovery:** How is the data interrelated? Does the metadata align?



2. Data Rules

- It's no use going through the process of data profiling if you're not going to maintain the new ideals going forward. That's why businesses doing DQM must create a set of technical and business rules for standardizing how the data should be formatted and accessed going forward. These rules can lay out protocols for properly formatting numbers and dates (ex: DD/MM/YY vs. MM-DD-YY) or email addresses.
- One key piece of advice on data rules: don't overdo it. Too many rules can be just as bad as no rules at all. Get together with stakeholders from different departments in the company to understand the data that is most valued to each of these departments. You can work together to build data rules that fulfill organization's standards.

3. Data Monitoring



- Once your data rules are in place, data monitoring ensures data is periodically checked against those standards to maintain its integrity. Any deviations from the data rules should be reported through an automated process.
- Business intelligence software can often assist with this aspect of DQM by capturing the irregularities before they infiltrate your data sets and sending alerts to management. Monitoring the entire of data can run the risk of information overload, it's best practice to only monitor the data that drives business decisions. Any information considered vital or sensitive should be monitored through your automated processes.



4. Data Remediation

- Monitoring the data is necessary to reveal the deviations of your data away from your set rules. Data remediation initiates a process for resolving the data quality concerns that come from that monitoring. One of the most important aspects of data remediation is examining the root cause of the defective data. **Was it human error? or Processing issues?**
- If the data is found to be corrupt or inaccurate, your organization should decide whether that “bad data” should be modified, or more simply, deleted. Who will be in charge of this? Should stakeholders and team members be notified before the modifications/deletions take place? An agreed upon approach for how to handle data remediation is part of your long game for data quality management.
- Review data quality rules again to determine if anything should be adjusted or updated. If you uncover any data processes that are affected by the bad data, you’ll need to re-initiate them and align them with the new adjustments you’ve set in place.

5. Data Quality Reporting



- Data quality reporting helps teams easily keep track of the information gathered during the data quality management process. One of the easiest ways to keep data quality top-of-mind is to create and share data quality reports with your team. CRM dashboards can be utilized to show this data across departments, or quarterly scorecards can be created and emailed to the entire organization to keep everyone looped in.
- However, the way you choose to approach it, creating a standardized method for **cataloging data quality reports** is essential. It keeps everyone thinking about data quality and the role they play in maintaining it. It's also a tangible way to reflect on your data quality progress and share that progress with stakeholders in the organization.



6. Data Discovery

- The final step of the DQM lifecycle is the data discovery process. Data discovery deepens DQM by gathering, analyzing, and reporting on the metadata associated to your data.
- As a simple definition, metadata is just data about data. It adds context to your data to reveal deeper insights on where data is located, what it means, how it's being used, and more.
- Data discovery is a process of exploring data through visual tools that can help non-technical business leaders find new patterns and outliers to help an organization better understand the insights their data has to offer. (Oracle NetSuite)

Why do we need to manage our data?

- Impacted by government which require PDPA rules for every year
- Increasing pressure on departments or agencies to manage these data properly

Benefit of Good Data Management

- Benefits to Data Suppliers
 - Increased confidence and trust their data that will be used according to their agreed conditions of use
 - Clear understanding of the subsequent use of their data, documented in a formal Memorandum of Agreement signed by supplier and user
 - Fair return for the use of their data

Benefit of Good Data Management

- Benefits to Data Brokers/Intermediaries
 - Better quality
 - Better care of the data holdings
 - Better control over the data
 - Improved knowledge and understanding of data holdings
 - Improved business processes, including better and more efficient use and re-use of data
 - Increased confidence that the organization complies with statutory and non-statutory obligations
 - Better control over access to data
 - More sensible and consistent data charges and conditions of use
 - Increased trust and confidence of customers in the quality and reliability of outputs

Benefit of Good Data Management

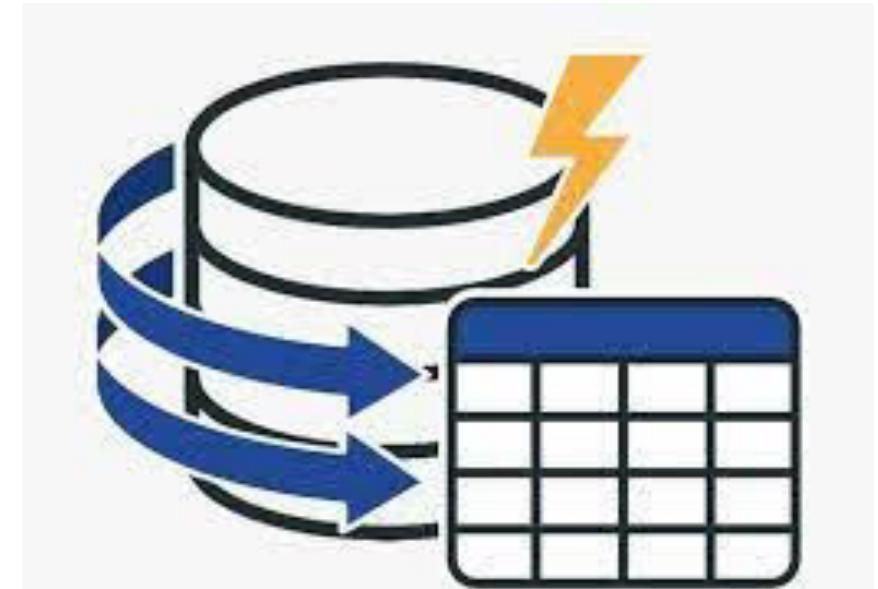
- Benefits to users and customers
 - Improved awareness and understanding of what data are available for current and future use
 - Improved access to data, free from unnecessary obstacles, safeguarded from disclosure of personal information or infringement of legal and contractual obligations
 - Better quality and timely information i.e., access to the right information at the right time
 - Better value for money, resulting from clear, fair and consistent data charges and conditions of use
 - Better exploitation of data generally, enabled by easier data exchange and integration with other harmonized data
 - Efficiency gains resulting from easier use of better-quality data

Topic 05: Data Model Diagram

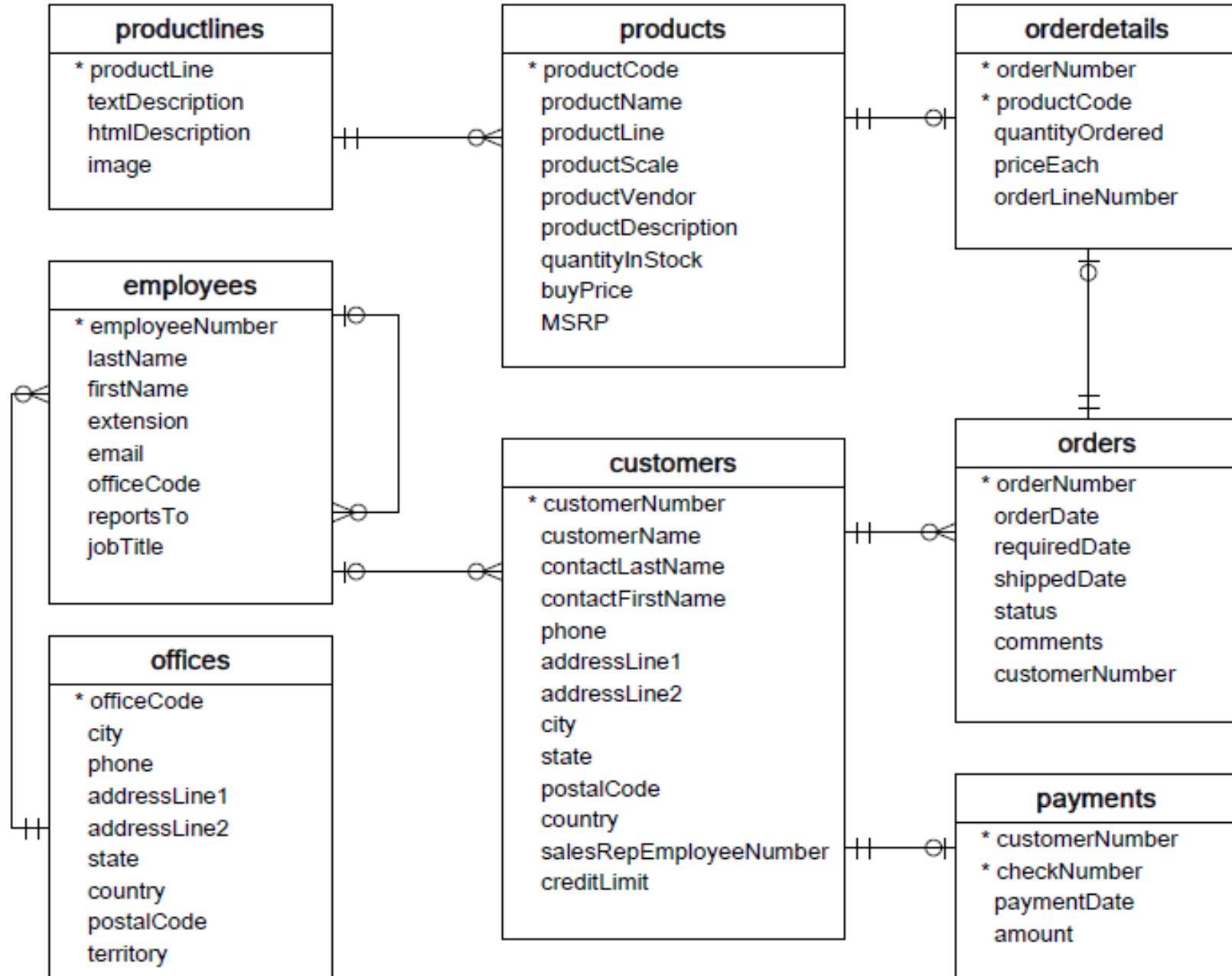
BDM3302: Data Management

Data Models

- There are three main popular data models which are
 - Entity Relationship Diagrams (E-R Diagrams)
 - Unified Modeling Language (UML)
 - Data Dictionary (will be used in DBM Tool Lab Class)

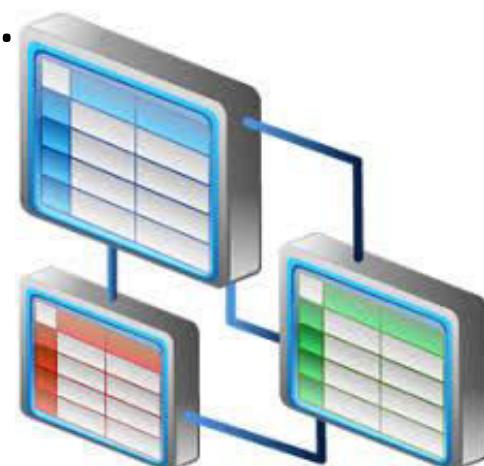


Entity Relationship Diagram (Classic Models)

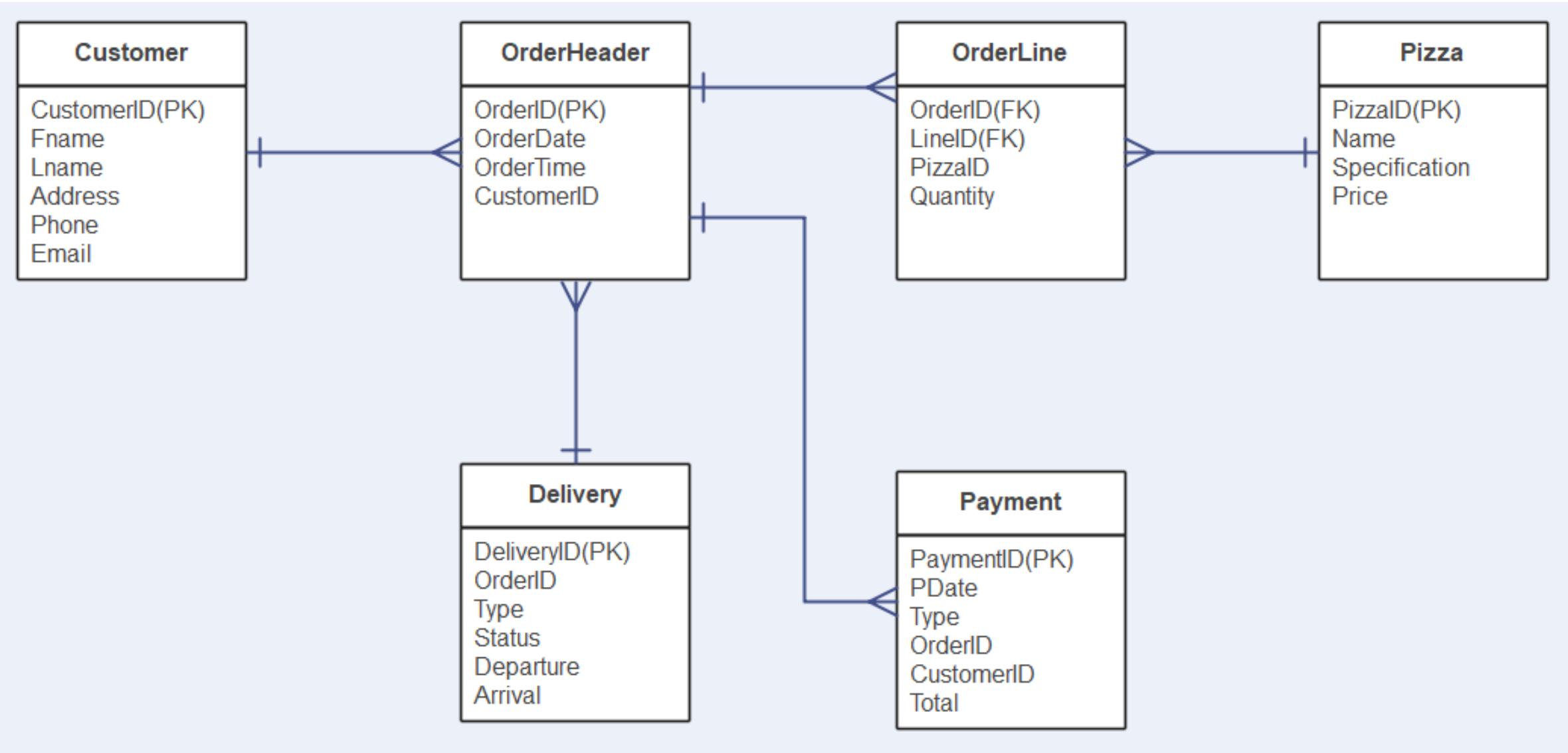


Entity Relationship Diagrams (E-R Diagrams)

- Entity-Relationship modeling is a default technique for modeling and the design of relational (traditional) databases. In this notation architect identifies:
- **Entities** representing objects (or tables in relational database),
- **Attributes** of entities including **data type**,
- **Relationships** between entities/objects (or foreign keys in a database).
- E-R Diagrams work well if you want to design a relational (classic) database, Excel databases or CSV files (tabular data). They work well for visualization of database schemas and communication of top-level view of data.



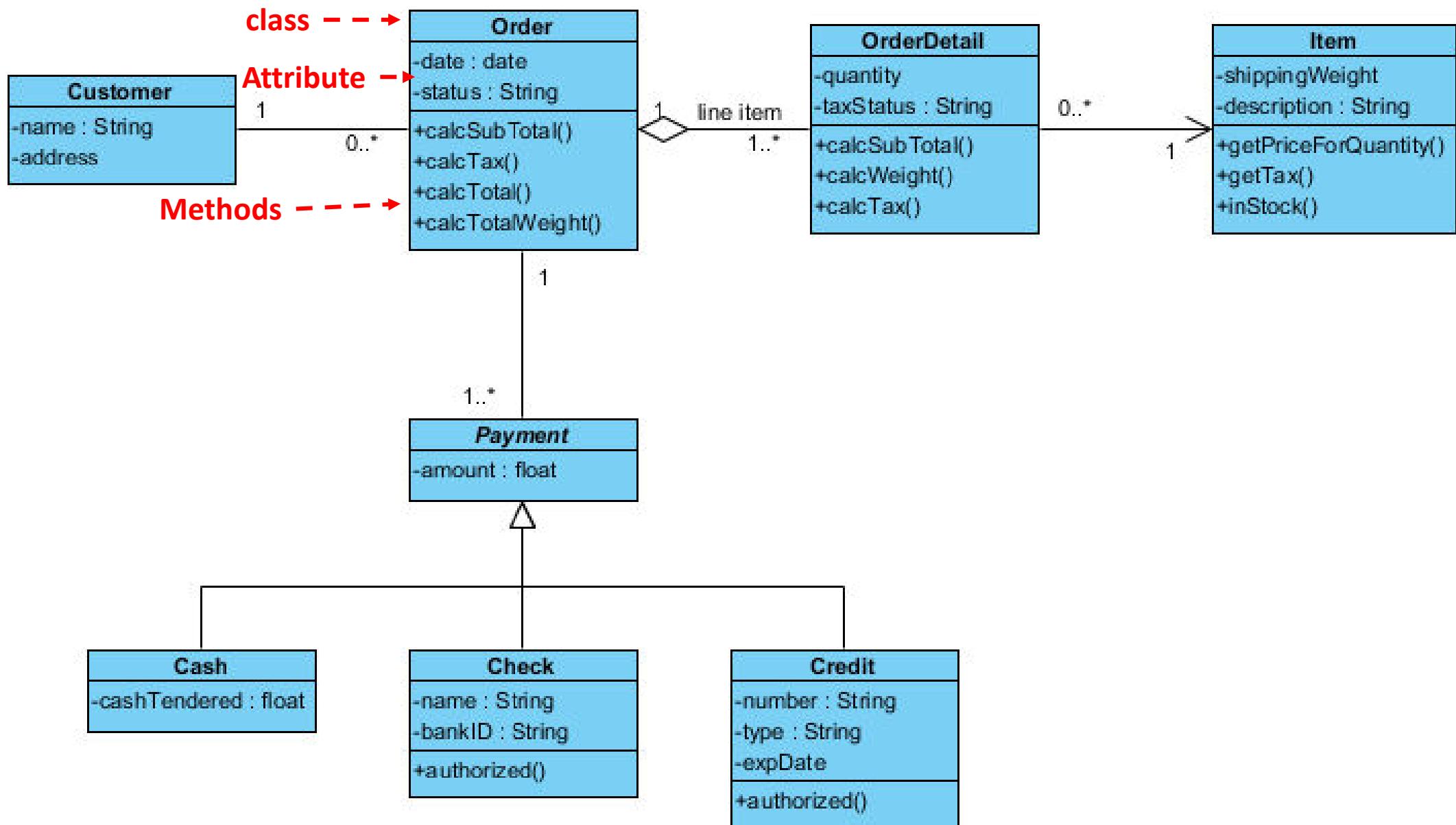
Entity Relationship Diagrams (E-R Diagrams)



UML Class Diagrams

- **UML** (Unified Modeling Language) is a standardized family of notations for modeling and design of information systems. It was derived from various existing notations to provide a standard for *software engineering or software developers*.
- Class diagrams are equivalent of ERDs in relational world and are mostly used to design classes in object-oriented programming languages (such as Java or C#).
- In class diagrams architects define:
- **Classes** (equivalent of entity in relational world),
- **Attributes** of a class (same as in an ERD) including **data type**,
- **Methods** associated to specific class, representing its behavior (in relational world those would be stored procedures),
- **Relationships** grouped into two categories:
 - **Relationships between objects** (instances of Classes) differentiated into Dependency, Association, Aggregation and Composition (equivalent to relationships in an ERD),
 - **Relationships between classes** of two kinds Generalization/Inheritance and Realization/Implementation (this has no equivalent in relational world).

UMLDiagrams



Business Document Sample (Invoice)



Company Name

Oklahoma City
Postal Address #34512 - HO
Telephone: 1547 000 220
Email: papilshipment@gmail.com

INVOICE

DATE:
INVOICE NO.
DUE DATE: 15 JUNE 2021

INVOICE TO:

Oklahoma City
Postal Address #34512 - HO
Telephone: 1547 000 220
Email: papilshipment@gmail.com

INVOICE TO:

PP Celestine Logistic.
Nii Wulomei Street - Cape forth
Villa-ville State City
Phone: 549 1458 7596

NO.	ITEM DESCRIPTION	QTY.	UNIT PRICE	AMOUNT
1	A5 flyers for Fall Beach Resort	1	0.00	0.00
2	Lorem Lorem Lorem	7	0.00	0.00
3	Pull up printings for TV Adverts	3	0.00	0.00
4	Car Branding for MO consults	5	0.00	0.00
5	Lorem Lorem Lorem Lorem Lorem	8	0.00	0.00
6	Lorem Lorem Lorem	9	0.00	0.00
7	Lorem Lorem Lorem Lorem Lorem	8	0.00	0.00
8	Lorem Lorem Lorem	6	0.00	0.00

NOTE:
Payment must be payed by the end od the month
All banking transfer will be included

Account Name:
Bank Name:
Bank Branch:



Sub Total: \$46.00
Discount: \$10.00
Tax Rate: 15.5%
Total \$ 16,000.00

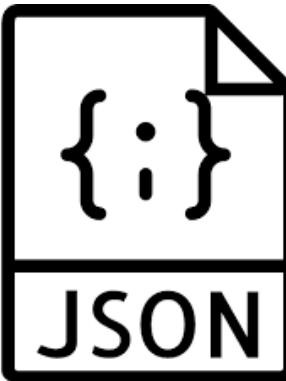
Authorized Sign:

Client Sign:

TERM & CONDITION APPLYS

Lorem Lorem Lorem Lorem Lorem Lorem Lorem Lorem

Thank You Customer for Chosing Us



Data Modeling Tool

- DBM Tool v.1.2.2 by Dr. Anan Osothsip (Download from LMS and unzip file).
- It is the tool that introduces how data can be constructed and converted into the SQL and database tables.
- The input can be JSON only that process the output into SQL and database tables.
- This tool also provides catalog, departments, roles, positions, users, processes, and assets in order to identify what kind of these data to support your business (**will use in Data Accessibility lab class later**).
- The Internet connection is also required for this tool.

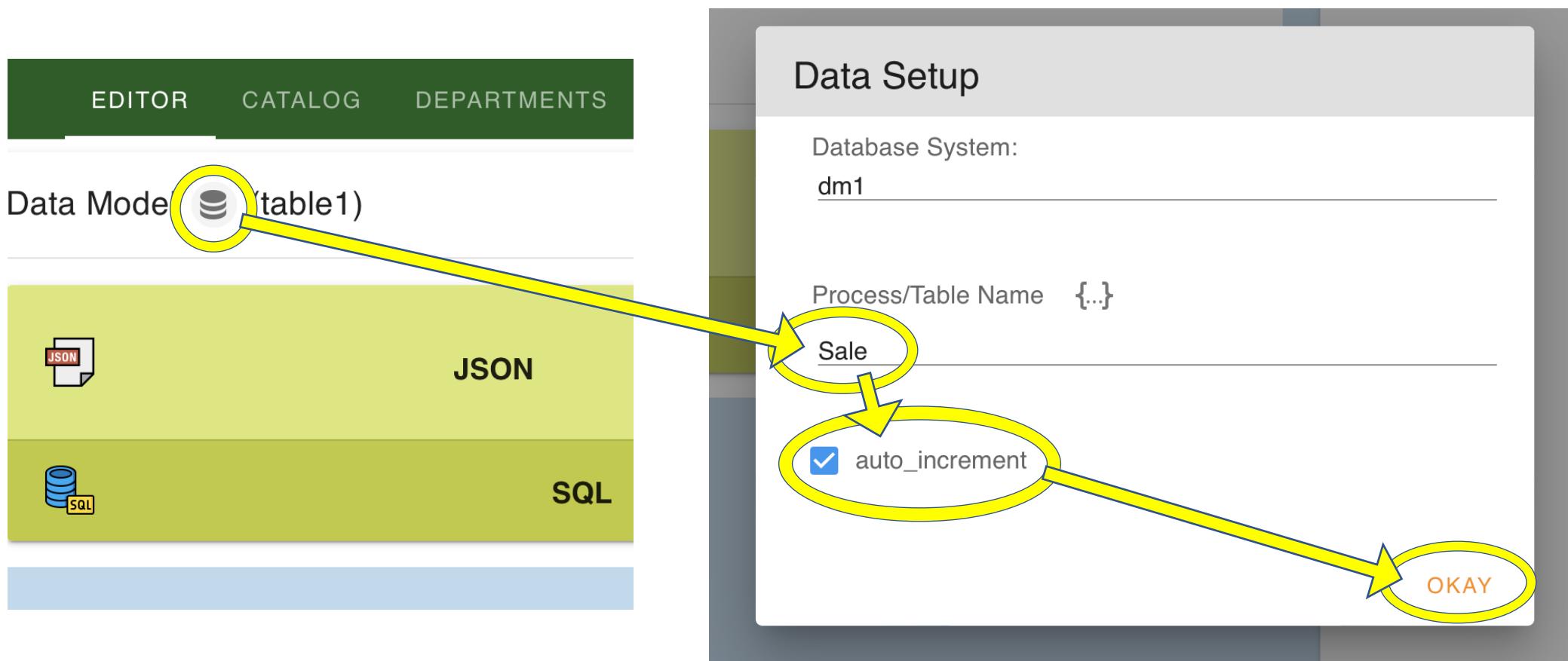
Data Modeling Tool Overview

The screenshot shows the jsontool.html interface with several highlighted features:

- File Tab:** "jsontool.html" is highlighted with a blue bar and a yellow arrow pointing to it.
- Menu Selection:** A green bar labeled "Menu Selection" contains links: EDITOR, CATALOG, DEPARTMENTS, ROLES, POSITIONS, USERS, and a dropdown arrow.
- Data Model Name:** "Data Model" and "(table1)" are highlighted with a yellow oval.
- Toolbar:** A black bar at the top includes checkboxes for "json editor" (checked), "c# class editor" (unchecked), "sql editor" (checked), "code editor" (checked), "table editor" (checked), and "diagram" (unchecked). A red note says "Please uncheck this check box first!" next to the "c# class editor" checkbox.
- Import/Export file:** A yellow circle highlights the three-dot menu icon in the top right, which is also annotated with "Import/Export file". A dropdown menu shows options: import, export, darcula, and version.
- Convert button:** A yellow circle highlights the "CONVERT" button.
- JSON Input Area:** A yellow box highlights the area where JSON input is shown, containing a "JSON" tab and a toolbar with icons for quotes, copy, paste, and others.
- SQL Output Area:** A blue box highlights the area where SQL output is shown, containing a "SQL" tab and a toolbar with icons for copy and dropdown.
- Database Table Output Area:** A large orange box highlights the main workspace labeled "DATABASE TABLE".

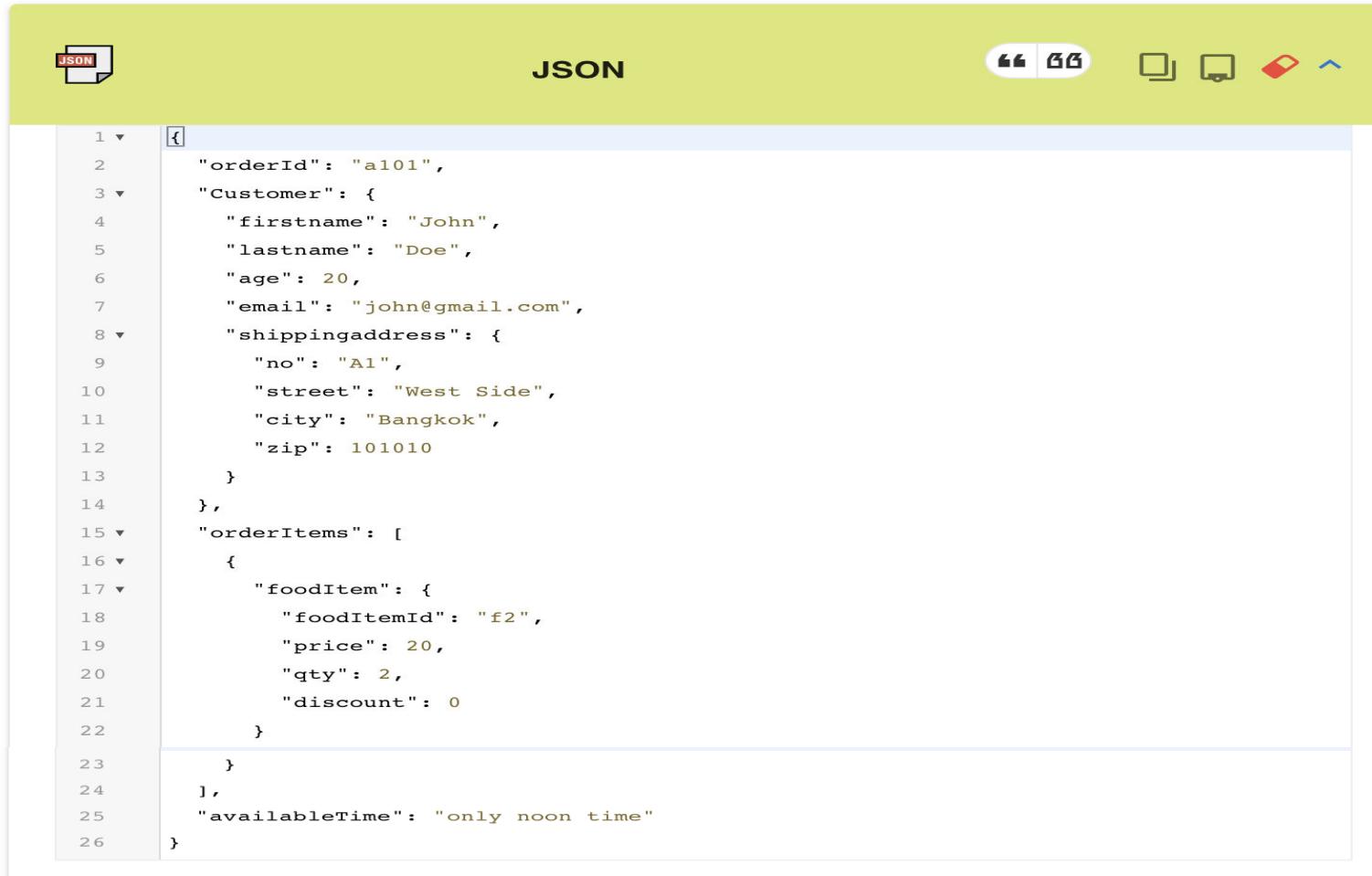
Data Modeling Tool (Lab for beginners)

- First, let's try to create Data Setup from "table1" to "Sale".



Data Modeling Tool (Lab for beginners)

- Second, try to input sample JSON how customer orders food.



The screenshot shows a JSON editor window with a yellow header bar. The header contains the word "JSON" and several small icons for file operations like copy, paste, and save. The main area displays a JSON object with line numbers on the left:

```
1  {
2   "orderId": "a101",
3   "Customer": {
4     "firstname": "John",
5     "lastname": "Doe",
6     "age": 20,
7     "email": "john@gmail.com",
8     "shippingaddress": {
9       "no": "A1",
10      "street": "West Side",
11      "city": "Bangkok",
12      "zip": 101010
13    }
14  },
15  "orderItems": [
16    {
17      "foodItem": {
18        "foodItemId": "f2",
19        "price": 20,
20        "qty": 2,
21        "discount": 0
22      }
23    }
24  ],
25  "availableTime": "only noon time"
26 }
```

Data Modeling Tool (Lab for beginners)

- JSON structure explains ...
 - Stands for JavaScript Object Notation
 - It is a text format for storing and transporting data
 - It is a “self-describing” and easy to understand

- The example looks like this:

```
{  
  "name": "John",  
  "age": 30,  
  "car": "Toyota"  
}
```

*“Mr. John with age of 30 years old
has got a Toyota car.”*

It identifies with 3 properties:
1. name = John (text)
2. age = 30 (integer)
3. car = Toyota (text)

*Each property has a value

Data Modeling Tool (Lab for beginners)

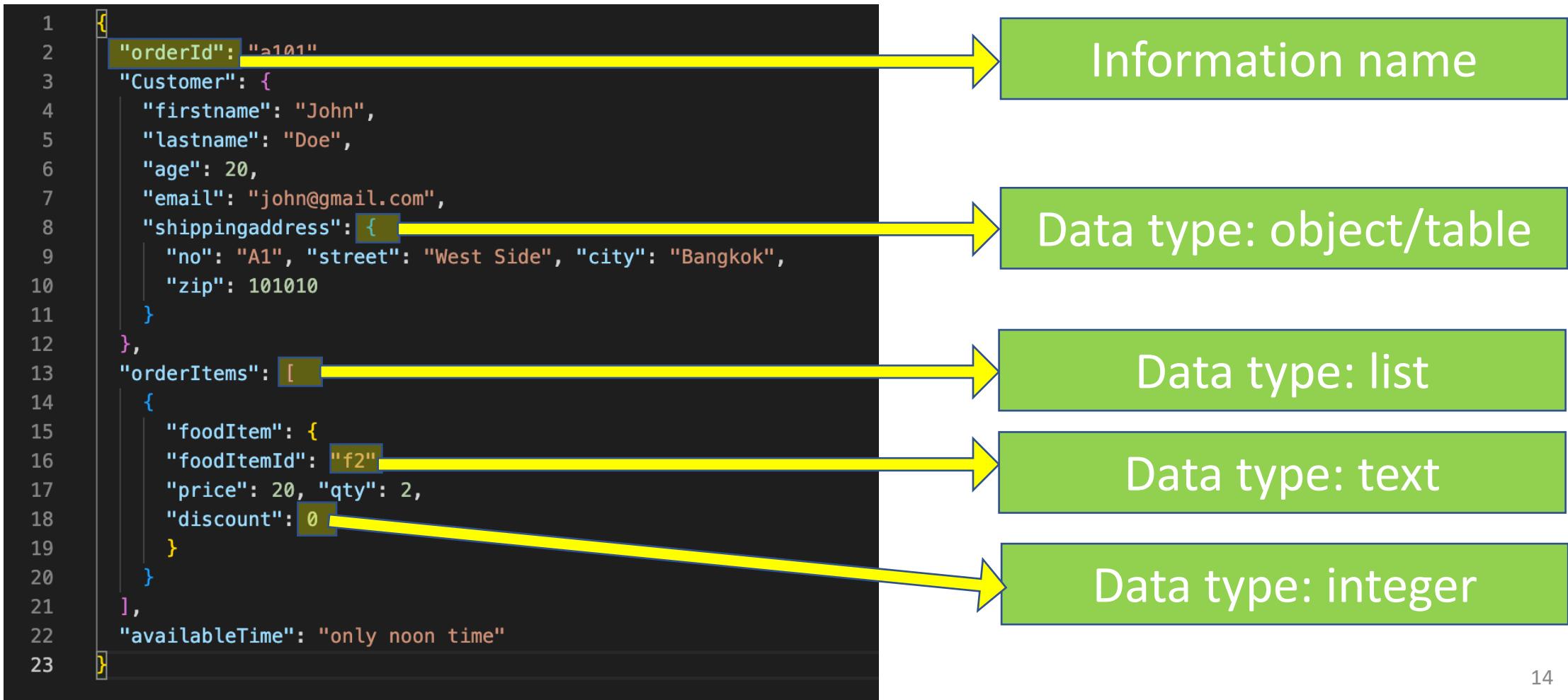
- JSON structure explains ...

```
1  {
2      "orderId": "a101",
3      "Customer": {
4          "firstname": "John",
5          "lastname": "Doe",
6          "age": 20,
7          "email": "john@gmail.com",
8          "shippingaddress": {
9              "no": "A1", "street": "West Side", "city": "Bangkok",
10             "zip": 101010
11         }
12     },
13     "orderItems": [
14         {
15             "foodItem": {
16                 "foodItemId": "f2",
17                 "price": 20, "qty": 2,
18                 "discount": 0
19             }
20         }
21     ],
22     "availableTime": "only noon time"
23 }
```

“Mr. John Doe would like to order two f2 foods for price \$20 each and deliver to his home only noon time.”

Data Modeling Tool (Lab for beginners)

- JSON structure explains ...



Data Modeling Tool (Lab for beginners)

- Third, try to click convert button and see the result.

The screenshot shows a Data Modeling Tool interface. On the left, there's a dark header bar with tabs: EDITOR, CATALOG, DEPARTMENTS, ROLES, POSITIONS, USERS, and a right-pointing arrow. Below this, the main area has a title "Data Model" followed by a small icon and "(Sale)". To the right of the title is a green "CONVERT" button, which is circled in yellow and has a yellow arrow pointing to it from the top right. Below the title is a green toolbar with icons for JSON (selected), XML, CSV, and others. The main content area contains a JSON code editor with the following content:

```
1 {  
2   "orderId": "a101",  
3   "Customer": {  
4     "firstname": "John",  
5     "lastname": "Doe",  
6     "age": 20,  
7     "email": "john@gmail.com",  
8     "shippingaddress": {  
9       "no": "A1",  
10      "street": "West Side",  
11      "city": "Bangkok",  
12      "zip": 101010
```

To the right of the JSON editor is a "DATABASE TABLE" panel with a blue border. It lists five tables with their respective counts: shippingaddress (5), Customer (6), foodItem (5), orderItem (3), and Sale (5). Each table entry has up and down arrows, a refresh icon, and a details icon. At the top right of the panel is a blue "EXPAND" button.

Data Modeling Tool (Lab for beginners)

- Somehow, you will see SQL appear when JSON is converted. You may bring SQL to create .sql file and try to import into “mysql” also (Empty).



SQL

```
1 drop database IF EXISTS dml;
2 ▼ CREATE DATABASE dml /*!40100 DEFAULT CHARACTER SET utf8mb4 */;
3 USE dml;
4 create table `Shippingaddress` (
5     shippingaddressId int(11) not null auto_increment,
6     no longtext default null,
7     street longtext default null,
8     city longtext default null,
9     zip int(11) not null,
10    primary key (shippingaddressId)
11 ) engine=InnoDB default charset=utf8mb4;
12
13
14 create table `Customer` (
15     CustomerId int(11) not null auto_increment,
16     firstname longtext default null,
17     lastname longtext default null,
18     age int(11) not null,
19     email longtext default null,
20     shippingaddressId int(11) not null,
21     primary key (CustomerId),
22     KEY IX_Customer_shippingaddressId (shippingaddressId),
23     CONSTRAINT FK_Customer_shippingaddressId FOREIGN KEY (shippingaddressId)
24         REFERENCES Shippingaddress (shippingaddressId)
```

```
[mysql]> use dm1;
Database changed
[mysql]> show tables;
+-----+
| Tables_in_dm1 |
+-----+
| Customer      |
| Sale          |
| Shippingaddress |
+-----+
3 rows in set (0.00 sec)

mysql>
```

Data Modeling Tool (Lab for beginners)

- Four, let's try to click Expand button. You will see each table displays data dictionary for each field. (e.g., shippingaddress)

The screenshot shows a data modeling interface with two main panels. On the left, under the heading 'DATABASE TABLE', there is a list of tables: 'shippingaddress (5)', 'Customer (6)', 'foodItem (5)', 'orderItem (3)', and 'Sale (5)'. Each table entry has three icons: a green arrow pointing up, a red arrow pointing down, and a blue grid icon. An 'EXPAND' button is located at the top right of this list, highlighted with a yellow oval and a large yellow arrow pointing towards it. On the right, a detailed view of the 'shippingaddress' table is shown. The title bar says 'shippingaddress (5)'. Below it is a search bar and a table with columns: name, type, constraint, and term. The table contains five rows, each representing a field in the 'shippingaddress' table:

name	type	constraint	term
shippingaddressId	int	PK	Shippinga
no	varchar		No
street	varchar		Street
city	varchar		City
zip	int		Zip

At the bottom of the right panel, there are buttons for 'Rows per page:' (set to 10), a page number '1-5 of 5', and navigation arrows.

Data Modeling Tool (Lab for beginners)

- Five, try to click name item to open Data Catalogs which displays in Catalog menu.

Click this item to open Data Catalogs

name	type	constraint	term
shippingaddressId	int	PK	ShippingaddressId
no	varchar	No	No
street	varchar	Street	Street
city	varchar	City	City
zip	int	Zip	Zip

Primary Key

Term

8 "shippingaddress": {
9 "no": "A1", "street": "West Side", "city": "Bangkok",
10 "zip": 10100 }

shippingaddress (5)

PK ShippingaddressId

PK Primary Key

Term Term

Data Modeling Tool (Lab for beginners)

- Display Data Catalogs for “**shippingaddressId**”.

Data Element Catalogs> **shippingaddress**

General Information

Name
shippingaddressId

Term
Shippingaddress

Status  **Use to type the description**

Type
int

Constraints

PK

**Try to click other name items
in each table to open Data
Catalogs and see the
different result ...**

Data Modeling Tool (Lab for beginners)

- Six, let's see inside the Customer table (one-to-one relationship).

name	type	constraint	term
CustomerId	int	PK	Customer
firstname	varchar		Firstname
lastname	varchar		Lastname
age	int		Age
email	varchar		Email
shippingaddressId	int	FK	Shippingaddress

```
2 "orderId": "a101",
3 "Customer": {
4   "firstname": "John",
5   "lastname": "Doe",
6   "age": 20,
7   "email": "john@gmail.com",
8   "shippingaddress": {
9     "no": "A1", "street": "West Side", "city": "Bangkok",
10    "zip": 101010
11  }
12 },
```

Why does “shippingaddressId” appear as a foreign key inside Customer table?

Data Modeling Tool (Lab for beginners)

- Seven, let's see inside the foodItem table (subset of orderItems).

Up Down grid foodItem (5) ^

Search 🔍

name	type	constraint	term
foodItemId	int	PK	Food Item
foodItemId	varchar		Food Item Id
price	int		Price
qty	int		Qty
discount			Discount

See the colors inside each layer
which corresponds to JSON

```
1  {
2    "orderId": "a101",
3    "Customer": {
4      "firstname": "John",
5      "lastname": "Doe",
6      "age": 20,
7      "email": "john@gmail.com",
8      "shippingaddress": {
9        "no": "A1", "street": "West Side", "city": "Bangkok",
10       "zip": 101010
11     }
12   },
13   "orderItems": [
14     {
15       "foodItem": {
16         "foodItemId": "f2",
17         "price": 20, "qty": 2,
18         "discount": 0
19       }
20     }
21   ],
22   "availableTime": "only noon time"
23 }
```

Data Modeling Tool (Lab for beginners)

- Eight, let's see inside the orderItems table (many-to-many relationship).

The screenshot shows a data modeling interface with a header "orderItem (3)". Below it is a search bar and a table with four columns: name, type, constraint, and term. The table contains three rows:

name	type	constraint	term
orderItemId	int	PK	Order Item
foodItemId	int	FK	Food Item
saleId	int	FK	Sale

**See the colors inside each layer
which corresponds to JSON**

```
1  {
2    "orderId": "a101",
3    "Customer": {
4      "firstname": "John",
5      "lastname": "Doe",
6      "age": 20,
7      "email": "john@gmail.com",
8      "shippingaddress": {
9        "no": "A1", "street": "West Side", "city": "Bangkok",
10       "zip": 101010
11     }
12   },
13   "orderItems": [
14     {
15       "foodItem": {
16         "foodItemId": "f2",
17         "price": 20, "qty": 2,
18         "discount": 0
19       }
20     }
21   ],
22   "availableTime": "only noon time"
23 }
```

Data Modeling Tool (Lab for beginners)

- Nine, let's see inside the Sale table (one-to-many relationship).

The screenshot shows a data modeling interface with a header "Sale (5)". Below it is a search bar and a table with columns: name, type, constraint, and term. The table rows are:

name	type	constraint	term
SaleId	int	PK	Sale
orderId	varchar		Order Id
CustomerId	int	FK	Customer
orderItems	list		Order Items
availableTime	varchar		Available Time

**See the colors inside each layer
which corresponds to JSON**

```
1  {
2    "orderId": "a101",
3    "Customer": {
4      "firstname": "John",
5      "lastname": "Doe",
6      "age": 20,
7      "email": "john@gmail.com",
8      "shippingaddress": {
9        "no": "A1", "street": "West Side", "city": "Bangkok",
10       "zip": 101010
11     }
12   },
13   "orderItems": [
14     {
15       "foodItem": {
16         "foodItemId": "f2",
17         "price": 20, "qty": 2,
18         "discount": 0
19       }
20     }
21   ],
22   "availableTime": "only noon time"
23 }
```

Data Modeling Tool (Lab for beginners)

- Moreover, you can see the SQL view with an empty record.

The screenshot shows a data modeling interface with a table named "shippingaddress (5)". The table has five columns: name, type, constraint, and term. The columns are: shippingaddressId (int, PK, Shippinga), no (varchar, No), street (varchar, Street), city (varchar, City), and zip (int, Zip). The "Shippinga" in the term column is likely a typo for "Shipping address".

name	type	constraint	term
shippingaddressId	int	PK	Shippinga
no	varchar		No
street	varchar		Street
city	varchar		City
zip	int		Zip

The screenshot shows an SQL terminal with a green header bar labeled "SQL". It displays the SQL code for creating the "Shippingaddress" table and its description.

```
1 create table `Shippingaddress` (
2     shippingaddressId int(11) not null auto_increment,
3     no longtext default null,
4     street longtext default null,
5     city longtext default null,
6     zip int(11) not null,
7     primary key (shippingaddressId)
8 ) engine=InnoDB default charset=utf8mb4;
```

```
mysql> describe shippingaddress;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| shippingaddressId | int | NO | PRI | NULL | auto_increment |
| no | longtext | YES | | NULL |
| street | longtext | YES | | NULL |
| city | longtext | YES | | NULL |
| zip | int | NO | | NULL |
+-----+-----+-----+-----+-----+
```

5 rows in set (0.00 sec)

```
mysql> 
```

How is JSON different from SQL?

Data Modeling Tool (Lab for beginners)

- Moreover, you can see the SQL view with an empty record.

Customer (6)

Search 

name	type	constraint	term
CustomerId	int	PK	Customer
firstname	varchar		Firstname
lastname	varchar		Lastname
age	int		Age
email	varchar		Email
shippingaddressId	int	FK	Shippingaddres

How is JSON different from SQL?

SQL 

```
12  create table `Customer` (
13      CustomerId int(11) not null auto_increment,
14      firstname longtext default null,
15      lastname longtext default null,
16      age int(11) not null,
17      email longtext default null,
18      shippingaddressId int(11) not null,
19      primary key (CustomerId),
20      KEY IX_Customer_shippingaddressId (shippingaddressId),
21      CONSTRAINT FK_Customer_Shippingaddress_shippingaddressId FOREIGN KEY
22      (shippingaddressId) REFERENCES `Shippingaddress` (shippingaddressId) ON DELETE CASCADE
23      ) engine=InnoDB default charset=utf8mb4;
```

```
mysql> describe Customer;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| CustomerId | int | NO | PRI | NULL | auto_increment |
| firstname | longtext | YES | | NULL |
| lastname | longtext | YES | | NULL |
| age | int | NO | | NULL |
| email | longtext | YES | | NULL |
| shippingaddressId | int | NO | MUL | NULL |
+-----+-----+-----+-----+-----+
6 rows in set (0.01 sec)

mysql> 
```

Data Modeling Tool (Lab for beginners)

- Moreover, you can see the SQL view with an empty record.

The screenshot shows a data modeling interface with a light blue header bar containing icons for sorting and filtering, and the text "foodItem (5)". Below this is a search bar labeled "Search". A table displays the following columns: name, type, constraint, and term. The rows are:

name	type	constraint	term
foodItemId	int	PK	Food Item
foodItemId	varchar		Food Item Id
price	int		Price
qty	int		Qty
discount			Discount

How is JSON different from SQL?

The screenshot shows a green header bar with a database icon and the word "SQL". Below it is a code editor window displaying the following SQL code:

```
25  create table `FoodItem` (
26    foodItemId int(11) not null auto_increment,
27    foodItemId longtext default null,
28    price int(11) not null,
29    qty int(11) not null,
30    primary key (foodItemId)
31  ) engine=InnoDB default charset=utf8mb4;
```

At the bottom of the code editor, there is a green box containing the following error messages:

ERROR 1060 (42S21): Duplicate column name 'foodItemId'
ERROR 1824 (HY000): Failed to open the referenced table 'FoodItem'
Query OK, 0 rows affected, 2 warnings (0.01 sec)

foodItem table cannot be include in mysql (Need fixing a bit) ...

Data Modeling Tool (Lab for beginners)

- Moreover, you can see the SQL view with an empty record.

The screenshot shows a data modeling interface with a sidebar on the left containing icons for up, down, and refresh, followed by the text "orderItem (3)". Below this is a search bar with a magnifying glass icon. The main area displays a table with four columns: name, type, constraint, and term. The table has three rows:

name	type	constraint	term
orderItemId	int	PK	Order Item
foodItemId	int	FK	Food Item
saleId	int	FK	Sale

How is JSON different from SQL?

SQL

```
34 create table `OrderItem` (
35     orderItemId int(11) not null auto_increment,
36     foodItemId int(11) not null,
37     saleId int(11) not null,
38     primary key (orderItemId),
39     KEY IX_OrderItem_foodItemId (foodItemId),
40     KEY IX_OrderItem_saleId (saleId),
41     CONSTRAINT FK_OrderItem_FoodItem_foodItemId FOREIGN KEY (foodItemId) REFERENCES
42     `FoodItem` (foodItemId) ON DELETE CASCADE,
43     CONSTRAINT FK_OrderItem_Sale_saleId FOREIGN KEY (saleId) REFERENCES `Sale` (saleId) ON
        DELETE CASCADE
    ) engine=InnoDB default charset=utf8mb4;
```

ERROR 1060 (42S21): Duplicate column name 'foodItemId'
ERROR 1824 (HY000): Failed to open the referenced table 'FoodItem'
Query OK, 0 rows affected, 2 warnings (0.01 sec)

**orderItem table cannot be include
in mysql (Need fixing a bit) ...**

Data Modeling Tool (Lab for beginners)

- Moreover, you can see the SQL view with an empty record.

Sale (5)

Search

name	type	constraint	term
SaleId	int	PK	Sale
orderId	varchar		Order Id
CustomerId	int	FK	Customer
orderItems	list		Order Items
availableTime	varchar		Available Time

How is JSON different from SQL?

```
46  create table `Sale` (
47    SaleId int(11) not null auto_increment,
48    orderId longtext default null,
49    CustomerId int(11) not null,
50    availableTime longtext default null,
51    primary key (SaleId),
52    KEY IX_Sale_CustomerId (CustomerId),
53    CONSTRAINT FK_Sale_Customer_CustomerId FOREIGN KEY (CustomerId) REFERENCES `Customer`(
54      CustomerId) ON DELETE CASCADE
) engine=InnoDB default charset=utf8mb4;
```

```
mysql> describe Sale;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| SaleId | int | NO | PRI | NULL | auto_increment |
| orderId | longtext | YES | MUL | NULL |
| CustomerId | int | NO | MUL | NULL |
| availableTime | longtext | YES | NULL |
+-----+-----+-----+-----+
```

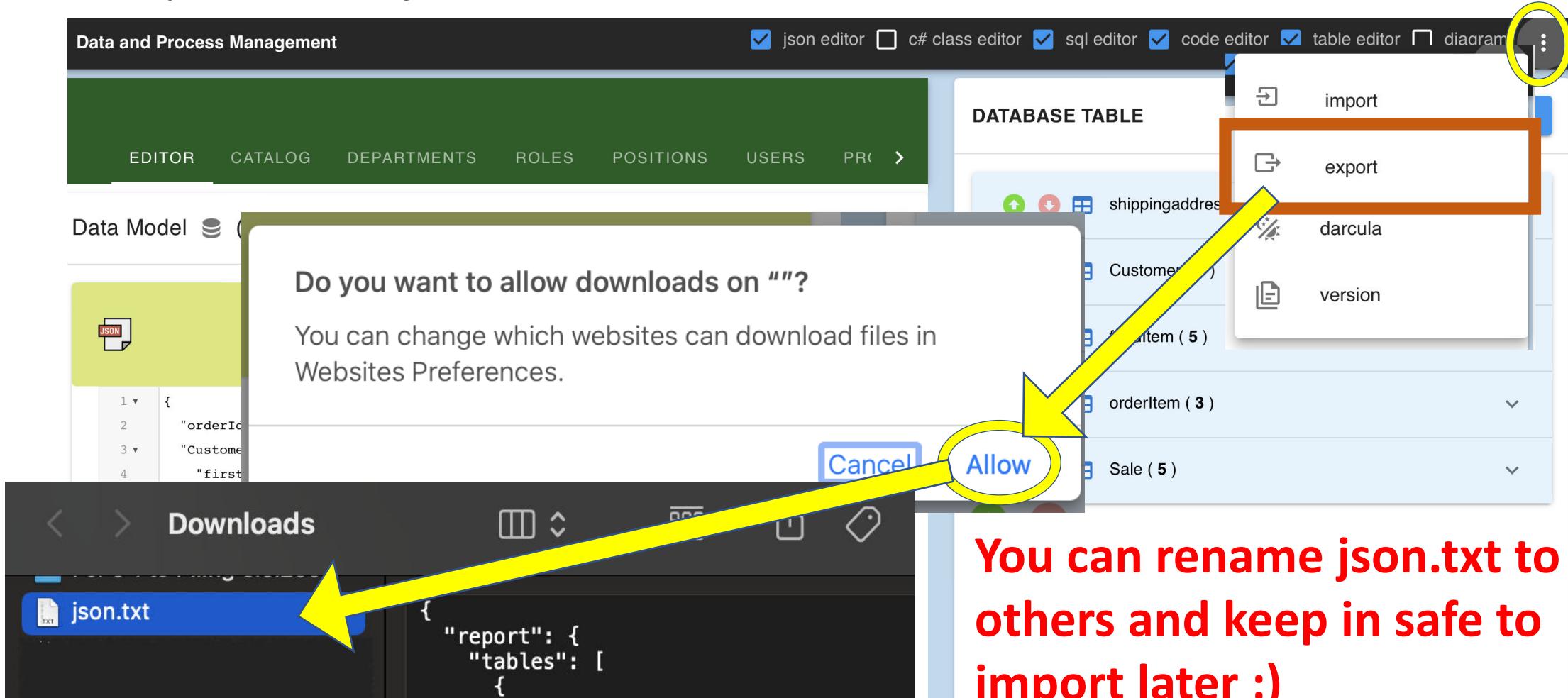
4 rows in set (0.00 sec)

```
mysql> ■
```

Data Modeling Tool (Lab for beginners)

- Export file to json.txt

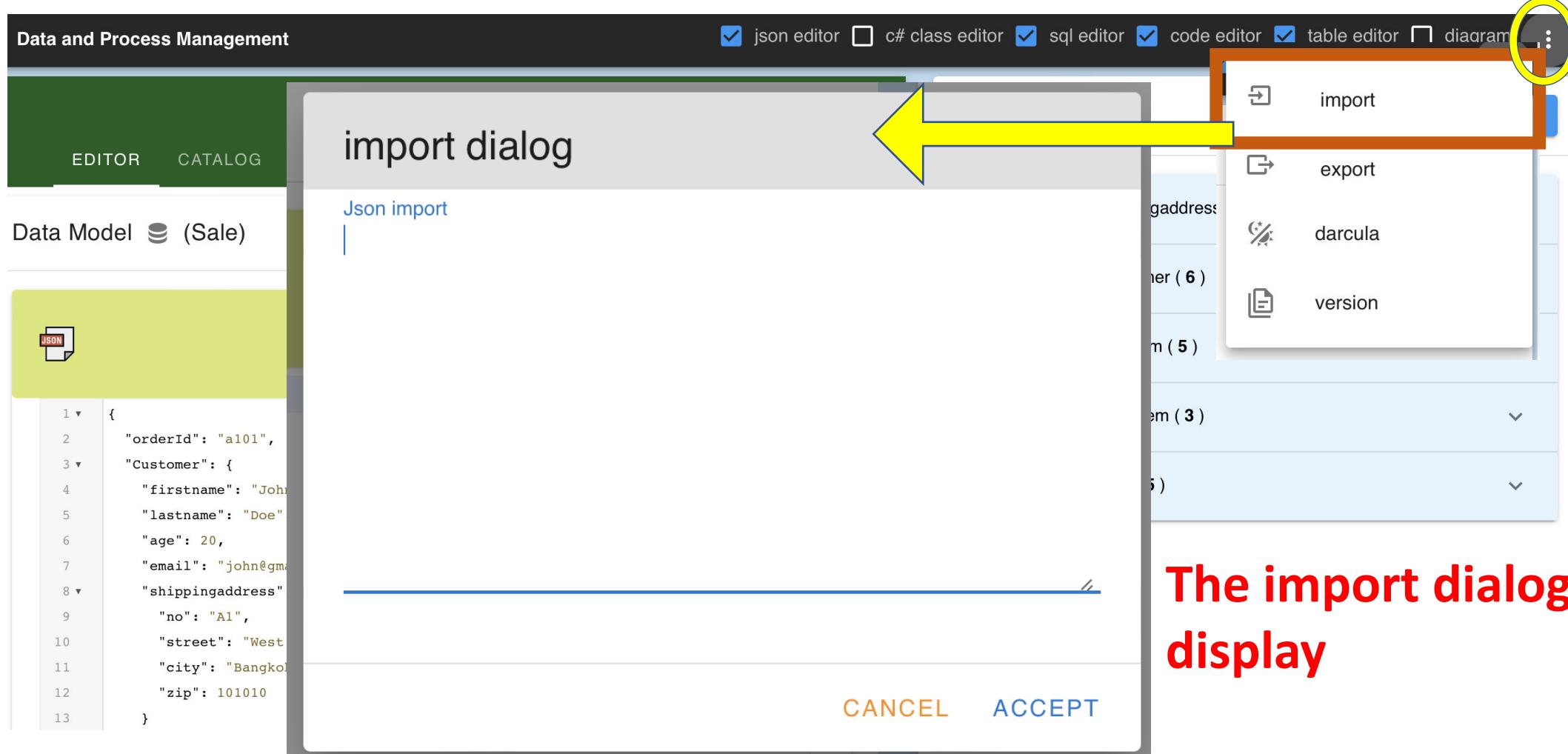
Click this circle button and choose export



Data Modeling Tool (Lab for beginners)

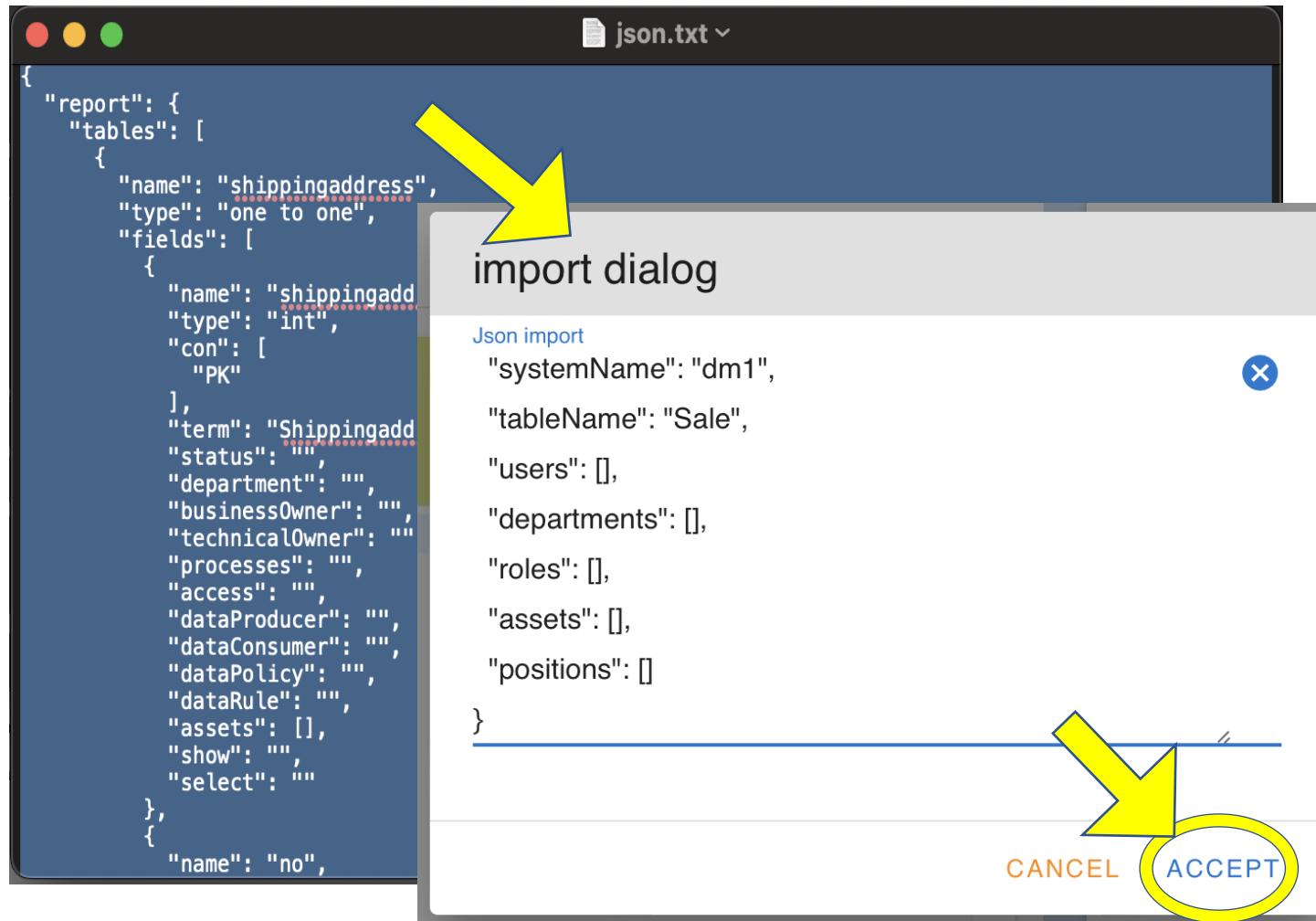
- Import file into DBM Tool.

Click this circle button and choose import



Data Modeling Tool (Lab for beginners)

- Import file into DBM Tool.



Step for importing file:

1. Open **json.txt** into **Notepad** or **TextEdit** App
2. Use hotkey "**Ctrl + A**" to select all texts
3. Use hotkey "**Ctrl + C**" to copy
4. Click in the **text area of import dialog**
5. Use hotkey "**Ctrl + V**" to paste and click **ACCEPT**

Data Modeling Tool (Lab for beginners)

- Let's try this (In class).
 - Interpret this business communication phrase into JSON. Then use DBM Tool to convert and see the different result:

“Ms. Evangeline McDowell would like to rent a Honda Civic, plate 2AB1820, year 2016 with dark silver color for 2,000 baht a day to drive in Phuket for 3 days. (Hint: rentId is “R1301”)

She is 24 years old and lives in 1135 Ladphrao 81 Wangthonglang Wangthonglang Bangkok 10310, and phone number is 0954220896”

Topic 06 – Data Sharing, Data Consolidation and Scheduling

BDM3302: Data Management

Why we need data sharing?

- An organization acquires important insight by analyzing data they share with other departments.
- finance teams need sales data to forecast future financial performance.
- Product management teams require marketing data to determine future products and services.
- Executive management needs up-to-the minute dashboards, fueled by data from many parts of the enterprise, to make timely, data-driven business decisions.

Why we need data sharing?

- Outside an organization, retailers share sales data with their vendors to manage inventory and supply chains.
- Software-as-a-service (SaaS) providers share the data they collect with their customers to provide them with deeper insights into their business.
- Healthcare providers securely share patient data with vendors that provide ancillary products and with other business partners that analyze that data to help improve patient services.
- Data has become more than something to collect and analyze. It's an asset you can easily and securely make available inside and outside your organization to streamline operations, swiftly deliver more-personalized customer experiences, and open up new market opportunities.

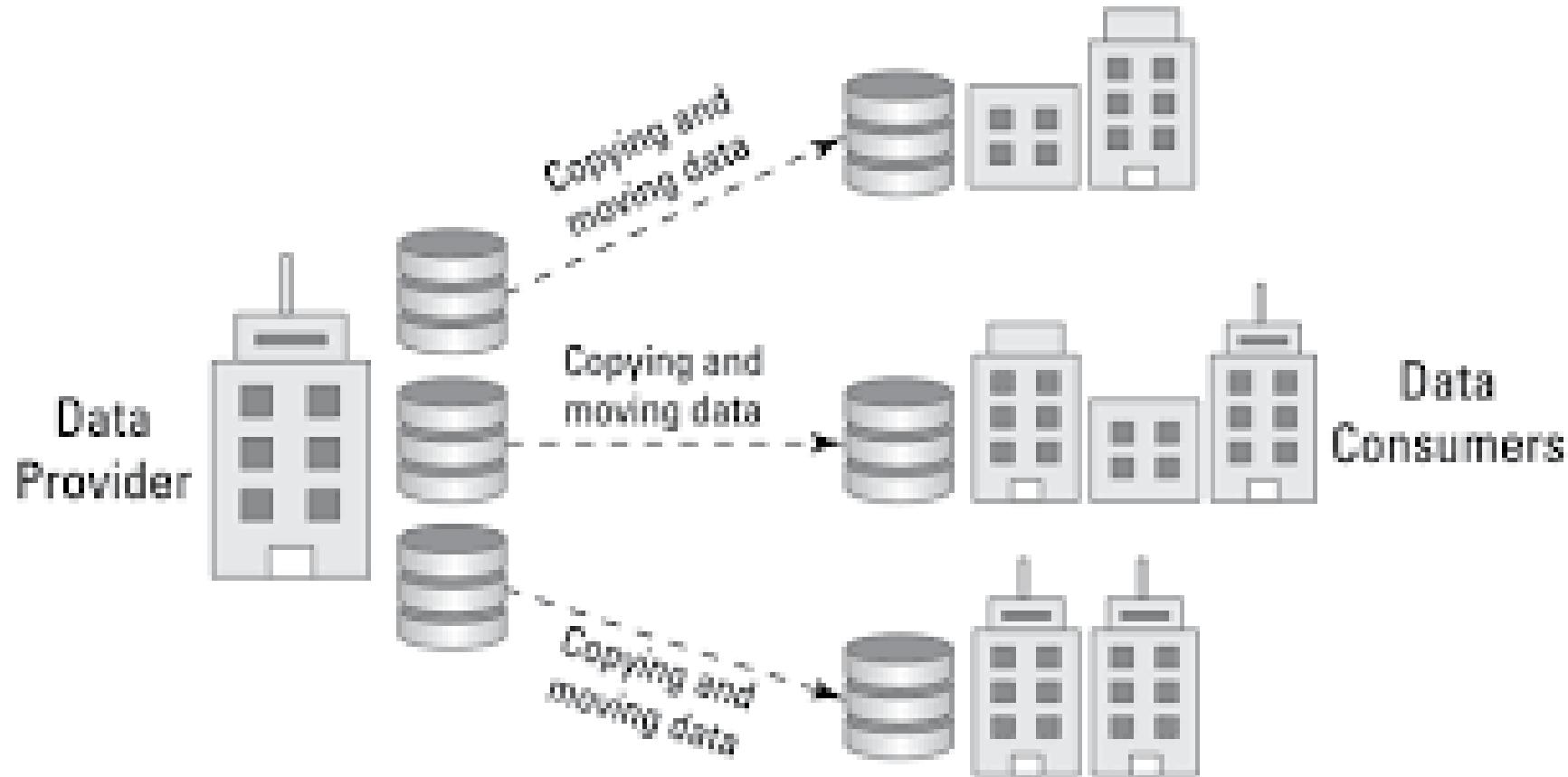
What is Data Sharing?

- Data can originate from the many software applications an enterprise uses to run its business, from the constant activity of visitors engaging a website, from a mobile phone, or from an Internet of Things (IoT) device attached to the refrigerator and other electronic appliances in your home.
- An organization can use data to track business results, make decisions, engage customers, define and create products, forecast trends, and more.
- Data is also a resource used and consumed between organizations, internal and external to one another, to collaborate on business plans, mutual initiatives, or joint opportunities.

What is Data Sharing?

- Many enterprises have come to realize they could enhance their business operations if they had access to data outside their organizations. Enterprises also recognize it is not easy to access data they don't generate themselves.
- ***Data sharing*** is the act of providing access to data between business units inside the organization, or between organizations external to each other.
- The organization that shares its data is called a “*data provider*”. The organization that wants to utilize shared data is called the “*data consumer*”. Any organization can be a data provider, data consumer, or both.

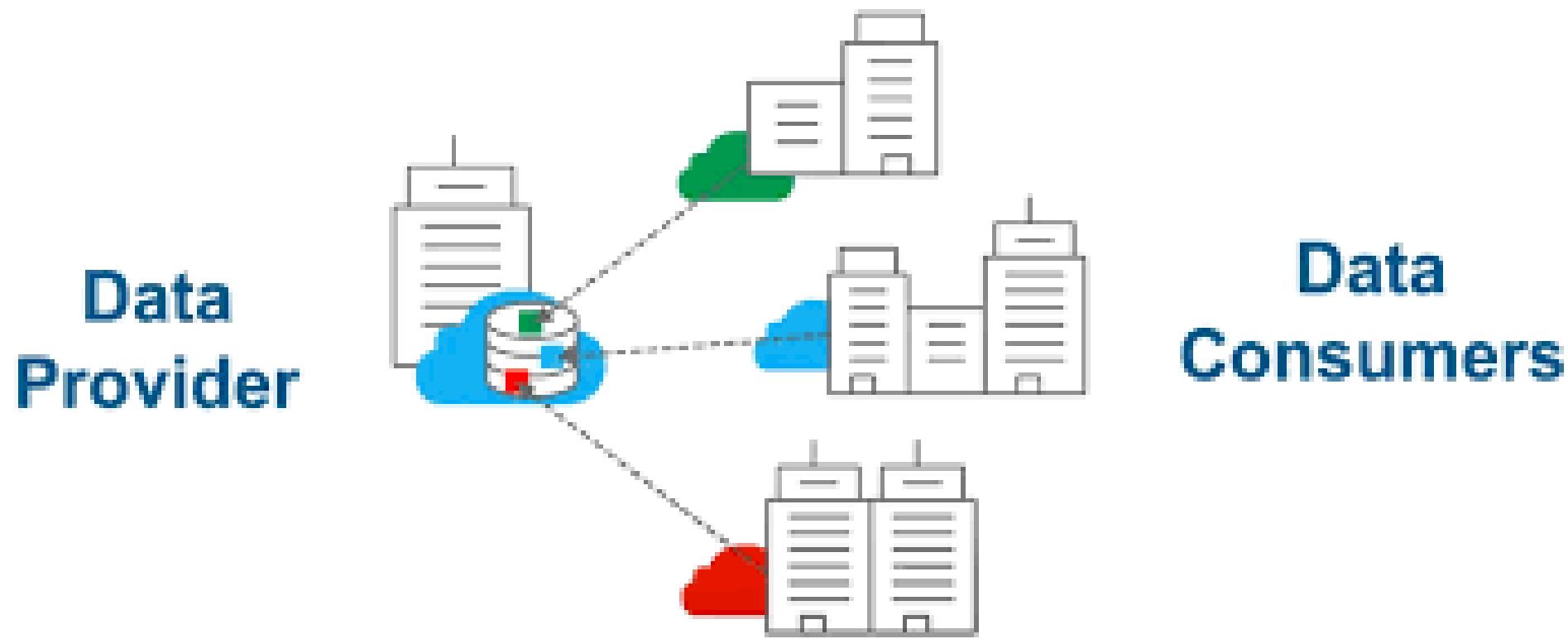
Traditional Data Sharing



Traditional Data Sharing

- Some organizations have traditionally shared data by making a copy of the shared data and sending it to their data consumers.
- The data consumers then download the data to analyze or combine that data with their existing data for deeper insights into who their customers are, how efficiently their business operates, and into which new industries their business is heading.
- This process is slow, cumbersome, and costly and only allows for moving limited amounts of shared data.

Modern Data Sharing



Modern Data Sharing

- A modern data sharing happens without moving data.
- A data provider makes available live, fast, secure, and read-only copies of data to its data consumers via modern cloud data sharing.
- New business opportunities modern cloud data sharing makes possible:
 - **Data sharing to eliminate data silos** – develop a single database and share among business units within an organization
 - **Data sharing for business efficiencies** - share live data with your business partners
 - **Data sharing as a product** – provide live and direct access to data as a monetized service.
 - **Data sharing as a product differentiator** – SaaS providers can offer direct access to data generated from B2B subscriber's activity and perform deeper analysis.

How Organizations Share Data

- Email
- File Transfer Protocol (FTP)
- Extract, Transfer, Load (ETL)
- Online File Sharing Services
- Cloud Storage
- Application Programming Interfaces (APIs)

Advantages of Data Sharing

- **Data sharing improves company outcomes:** Firms that support data sharing will outperform their counterparts on the majority of business value criteria.
- **Innovative Technologies:** implement AI/ML for accurate predictive data analysis.
- **Unified View:** Create a unified source of the truth across all departments.
- **Transparency:** Develop openness and transparency among stakeholders for improved decision-making.
- **Speed:** Widen the area of analysis for an organization's data consumers.

Advantages of Data Sharing

- **Flexibility:** Create new analytic reports swiftly and with fewer interruptions.
- **Synergy:** Data sharing has intrinsic advantages for the researcher and research sponsor. Making the data accessible to their colleagues and the general public encourages academics to better manage and assure the quality of their data.
- **Improving Science and Decision-making:** Sharing data enhances data circulation and usage within the scientific community by fostering more openness, facilitating the reproducibility of discoveries, and enlightening the scientific community at large.
- **Collaboration:** Sharing data stimulates more interaction and cooperation among academics, which may lead to significant new discoveries in the area.

Disadvantages of Data Sharing

- **The more data openness the higher digital security risks:** data sharing needs information systems to be open so that data may be viewed and shared. This may expose further portions of an organization to digital security risks that may result in events that compromise the availability, integrity, or confidentiality of data and information systems upon which economic and social activities depend.
- **Higher impact of personal data breaches:** Where data is accessible and shared, personal data breaches are potential. They will not only create damage from personal data loss. They may also result in substantial economic losses for the afflicted firm, such as a loss of competitiveness and reputation.
- **Violation of privacy, intellectual property rights, and other interests:** The hazards of increased access and sharing extend beyond breaches of digital security and personal data. They include risks of breaking contractual and socially agreed-upon conditions of data re-use, and therefore risks of acting against the reasonable expectations of users.

Disadvantages of Data Sharing

- **Data reuse in violation of agreed terms and expectations:** Even when people and organizations agree to and consent to particular conditions for data sharing and data re-use, including the reasons for which the data should be re-used, there is a high danger that a third party would deliberately or accidentally use the data differently.
- **Loss of authority over data and the significance of permission:** Unless certain data stewardship and processing rules are in place, once data are accessed or shared, they will leave the information system of the original data holder and therefore be out of his/her control. The same is true for those who supply their data and agree to its reuse and distribution. Then, data owners and people lose control over how their data are repurposed.
- **The limitations of anonymization and the growing potential of data analytics:** Once connected to sufficient additional information, it is possible to forecast an individual's chance of possessing certain qualities in order to construct a profile. Even if the inferences are valid, there is a possibility that they might be exploited against an individual's best interests, desires, or expectations.

Data Sharing for Business-to-Business (B2B)

Whether sharing data to external organizations or receiving shared data from them, if organizations cannot collaborate on data, they are less efficient and run the risk of operating at a higher cost and lower productivity.

Examples of B2B Data Sharing:

- » A hotel booking website shares reservation patterns and trends with hotel properties to develop promotional and pricing programs.
- » A grocery chain provides store sales data to suppliers to ensure shelves are adequately stocked to meet demand.
- » Retailers share in-store sales data for merchandising, so the hot items are always available.

Business Value of Data Sharing for Organizations

- Data sharing across and beyond an organization consists of four basic workflows:
 - **Across lines of business (LOBs):** Sharing data between business units within the same organization
 - **Between organizations:** Outbound data sharing to another, separate organization to benefit your business
 - **Between organizations:** Receiving inbound data shared from another organization to benefit your business
 - **Monetizing data:** Sharing live data as a service so data consumers can enrich their own, existing data

Sharing Data Across Line of Business (LOB)

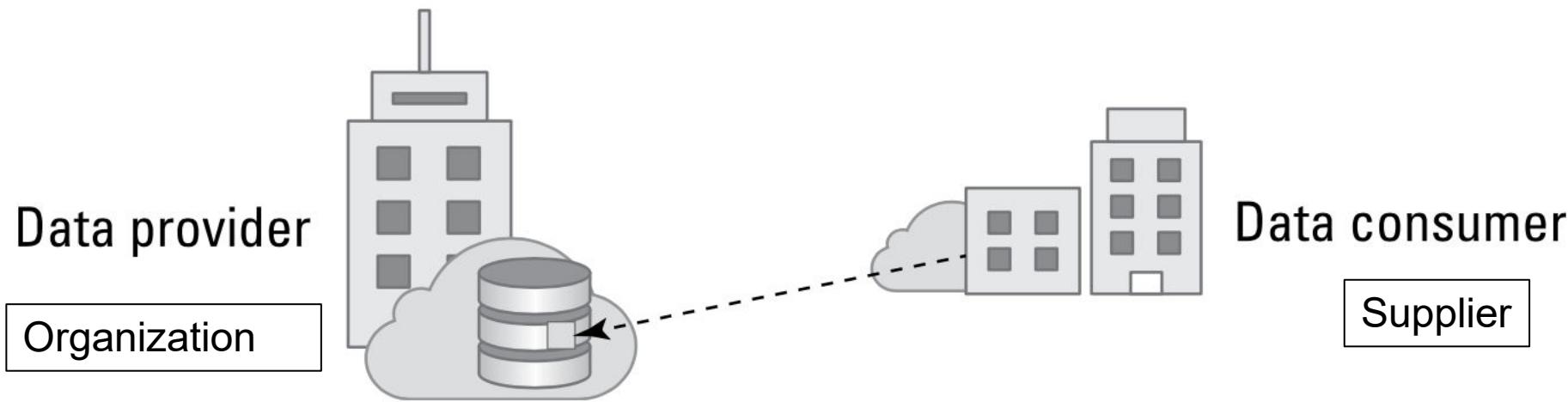
- Within the same organization, business units depend on email, spreadsheets, shared network drives, application programming interfaces (APIs), and other methods for communicating and for sharing data. Sharing data across an organization enables and fosters increased levels of business intelligence and drives timely and informed business decisions.
- Within an organization, however, data is often locked in silos. Mergers or acquisitions, firewall restrictions, or other business or technology barriers often restrict an organization from easily sharing data across its business units. These physical or logical separations of infrastructure can prevent two or more business units from accessing all available data within an organization to deliver all-inclusive, data-driven insights. These data silos emerge when an organization relies on a traditional, on-premises data warehouse or a traditional data warehouse ported to the cloud.

Sharing Between Organizations: Outbound

- External data sharing takes place all the time. A vendor-supplier relationship, a partner relationship, a developer-producer relationship — or any number of other business relationships — all require two or more organizations to collaborate with data to drive business. The primary organization is sharing data, outbound, to the partner organization.
- For example, in a vendor-supplier relationship with data sharing, a supplier knows in advance when to replenish the stock of a particular item. Well-managed inventory also prevents overstocking, minimizing the need to significantly reduce prices.

Sharing Between Organizations: Outbound

Outbound Modern Data Sharing

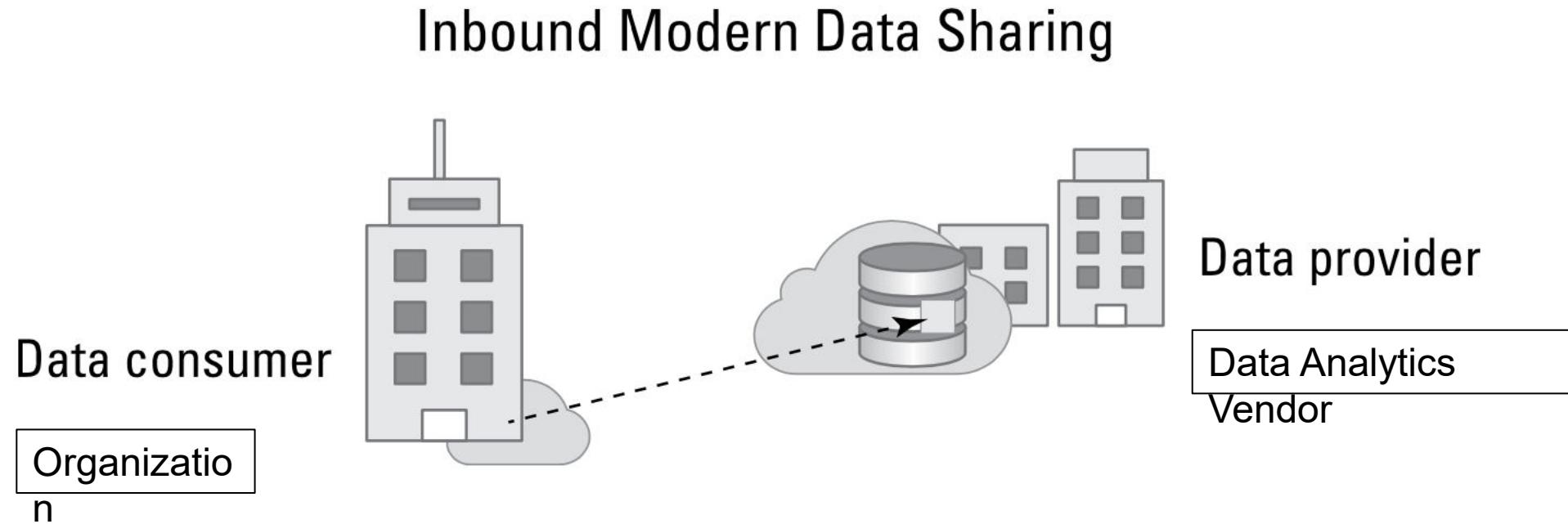


An organization, acting as the data provider, shares data with its supplier, the data consumer.

Sharing Between Organizations: Inbound

- Increasingly, organizations engage outside service companies. These contracted companies can specialize in logistics, shipping, marketing services, or sales operations, just to name a few. For example, a large retailer would collect massive amounts of demographic data about its target customers. The retailer would then share this data as a data provider to a data analytics company. The analytics company would analyze the data for the retailer.

Sharing Between Organizations: Inbound



The organization is the data consumer, accessing the data from its outside data analytics vendor, which is the data provider.

Sharing Between Organizations: Inbound

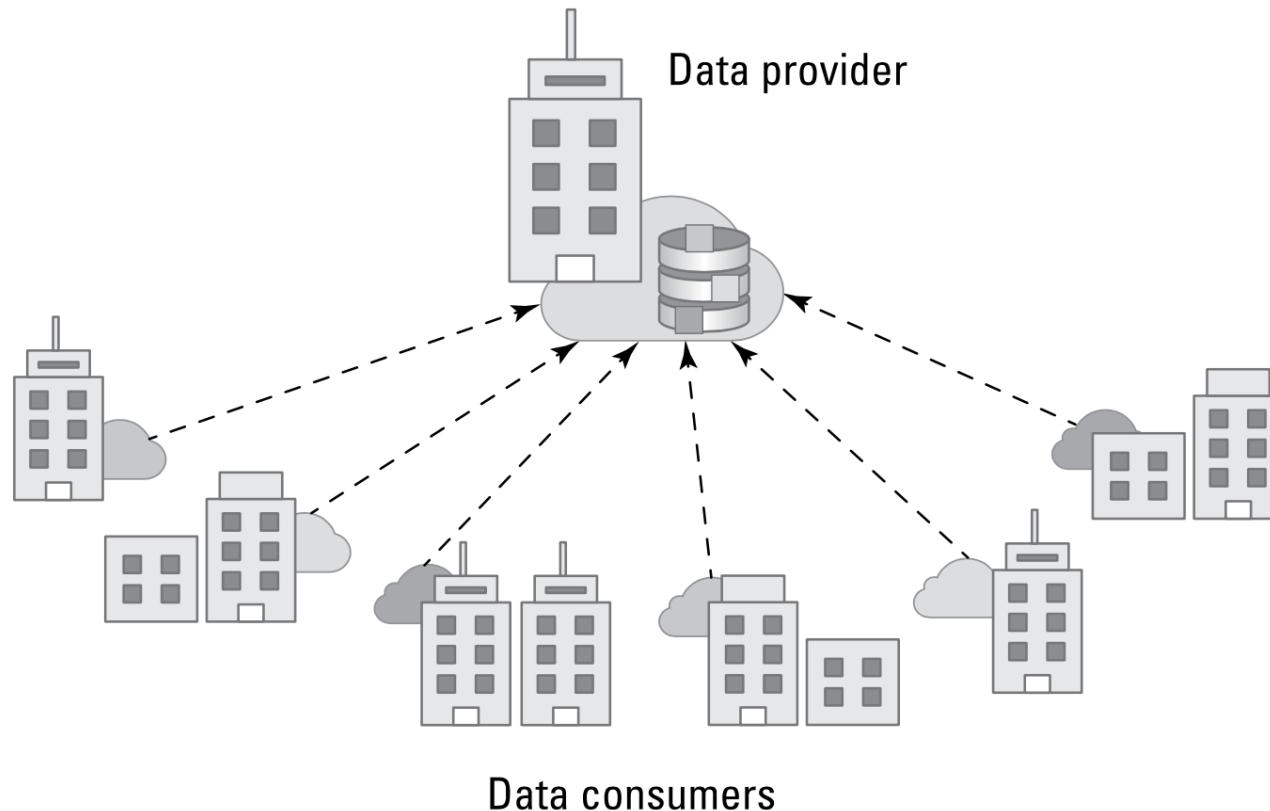
- The organization contracts a service provider to perform a function the organization chooses not to perform in- house. In turn, the service provider generates data as a result of that service — data that belongs to the organization, which is the service provider's customer. With inbound data sharing between organizations, the data generated by the service provider is shared with its customer. The customer then executes additional analytics to develop deeper insights and value from additional data generated outside its data center but within its business ecosystem.

Monetizing Data

- Data can also take on more significance today than just day- to-day collaboration. Data is a business asset — a currency. As such, data can offer different types of value depending on the organization that wants to consume that data. Thus, as with any asset, data has value. To monetize the value of its data, a provider can sell data to consumers that can then use the data to advance their own business objectives.
- Data consumers can use shared data without having to capture and collect it themselves. They can benefit directly from analyzing that data or combining it with other data to enhance its value.

Monetizing Data

Monetized Cloud Data Sharing



An enterprise data provider creates new revenue opportunities by sharing data with other organizations acting as data consumers.

Monetizing Data

- Data monetization is a process by which a data provider charges data consumers a fee to gain access to the provider's data or data services, so a data consumer can enrich its existing data sets to benefit its business and its customers. Some providers offer access to the data itself while others offer services, such as data modeling, data enrichment, and data analytics.
- Nielsen is a pioneer and offers service for more than 90 years, the company has collected, analyzed, and sold consumer data to media companies, advertising firms, retail organizations, and many other industries.
- Nielsen's proven business model has inspired numerous niche data sharing opportunities. For example, in the financial services industry, some companies collect stock market data, package it, and sell it to brokers and hedge funds.

Data Consolidation

- Many companies are looking for ways to increase their competitiveness, their efficiency and effectiveness and their adaptability to unexpected changes. Data consolidation enables business to retrieve data from all their departments in a comprehensive way and prepares data to make predictions, detect errors and make decisions.
- However, many times the data different sources that companies store is made up of data from diverse origins, with various formats and with different purposes. This can result in duplication, different data types and formats, or problems with data security.
- In addition, this great diversity makes it difficult to quickly and efficiently analyze the data, so it must be treated to make it more uniform, eliminate errors and duplicates, and for it to be in the right place, such as a data warehouse or data lake.
- This process is also known as data integration and can be carried out using various methods that allow different types of data to be manipulated from a single location to become corporate insight data that will lead to better decision making.

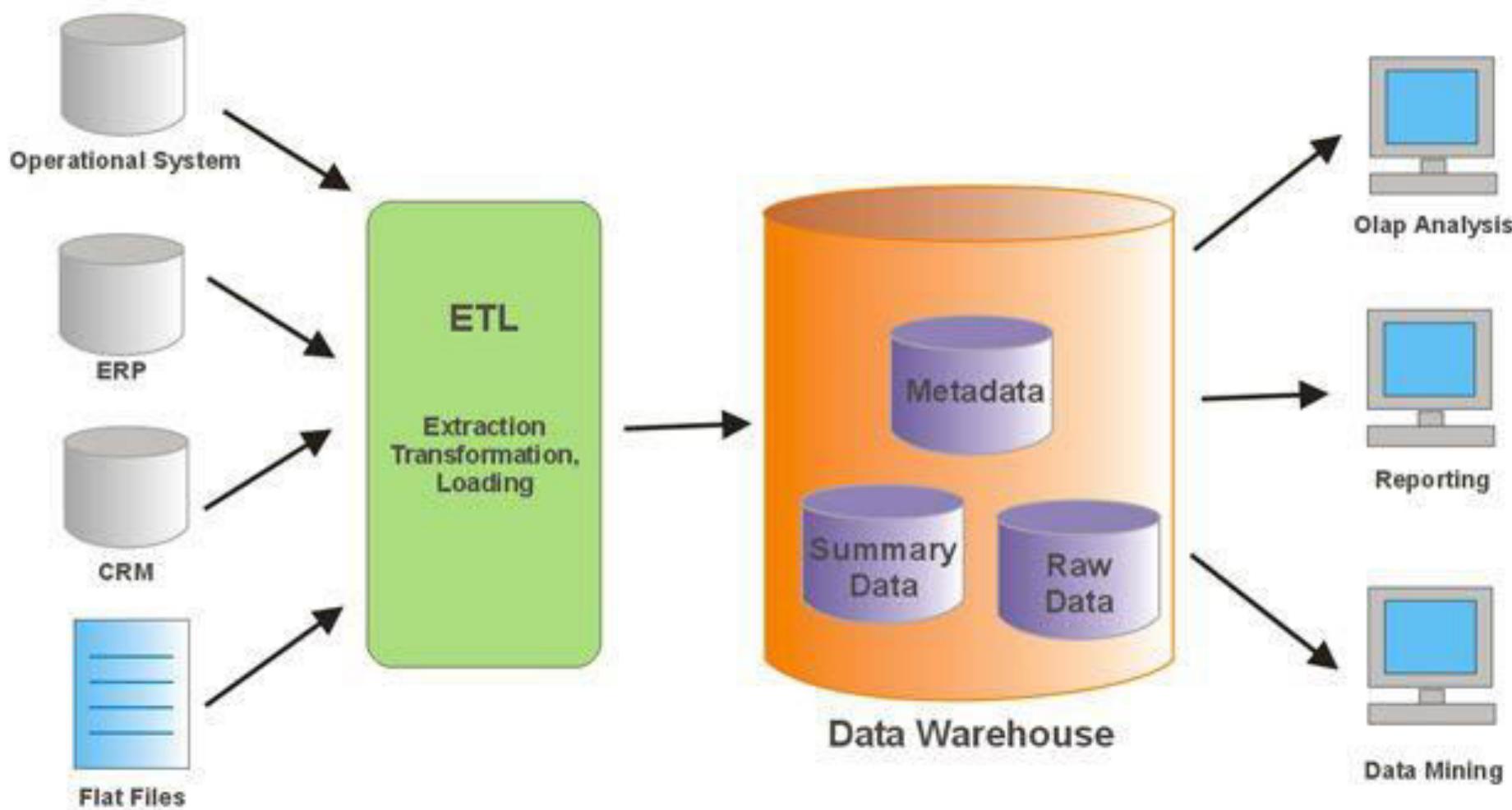
Data Consolidation

- Data consolidation is the process of combining data from multiple sources, cleaning and verifying it by removing errors, and storing it in a single location, such as a data warehouse or database. Data is produced from various sources and in multiple formats in every business. The data consolidation process makes it easier to unify that data.
- Consolidating data enables companies to efficiently plan, implement, and execute business processes and disaster recovery solutions. This is done because all critical data in one place grants users a 360-degree view of all their business assets. It improves data quality, fast-tracks process execution, and simplifies information access.
- This process ensures that quality and accurate data is available, making it quicker and easier to process and treat this data. Data consolidation eliminates disparities before data is used, saving time, improving efficiency and adding value to the company's analytical operations.

Centralized Implementation

- The centralized Implementation enables you to efficiently create and manage master data attributes in one place. It supports the validation and verification of data at the origin and always guarantees high-quality data, across all business systems. Once the data is curated, it's automatically available to other systems through integrations. All data is maintained in the platform, and other systems are kept up to date with one-way integrations from the platform.
- The real advantage of centralized implementation is that your master data is always accurate and complete. The system also supports security and visibility policies at a data attribute level. With this approach, you gain a core set of master data for one or more domains. Centralized is the ideal implementation style, with the most benefits for business.

Data Consolidation



Data Scheduling

- Data Integration – In business process, you may require to get some data from other departments like sales, marketing, finance, accounting, etc. You can schedule your data transfer jobs to run at any frequency. It can be hourly, daily, weekly, monthly, or yearly. You can set execution times for data load rules by scheduling them. For example, you may want actual sales ordering data to be imported into corporate systems everyday at 12.00 am.
- Schedule for data collection and consolidation (7-11)
- Schedule jobs for consolidation tasks (SAP)

Topic 07: Data Accessibility

BDM3302: Data Management

What is Data Accessibility?

- Data accessibility is the degree to which people in an organization can access and use data.
- Data accessibility determines the extent to which data consumers in an organization can access and utilize data to achieve the organizational goals, increase productivity and efficiency without requiring advanced know-how and experience in working with data.
- Data is not just available, but it is also usable. It's not enough to know that the data lives in some platforms. If someone in your organization really needed it, they could submit a request for a very specific report.

What is Data Accessibility?

- Data accessibility is the on-demand, authorized capability to retrieve, modify, copy, or move data from IT systems. With data access, users can perform these activities from any location.
- Data access is the means by which users can access this data in an authenticated manner approved by the organization that *owns* the data.
- Data accessibility is one of the main outputs of an effective data governance program. Ideally, organizations should think through structured approaches to grant data access to various stakeholders, both internal and external to the organization.

User and Role Management

The screenshot shows the WordPress admin dashboard with the 'Users' page selected. The page displays a list of users with columns for Username, Name, Email, Role, and Posts. A red callout points to the 'Capabilities' link next to the 'Edit' link for the user 'admin'. The callout contains the text: 'Click to assign capabilities directly to the selected user'.

Username	Name	Email	Role	Posts
admin	Bill Bell	admin@nomail.my	Administrator	29
baron		baron@mail.co	Widgets Manager	1
barry		barry@mail.com	Shop Manager	4
bololo		bololo@mail.co.co	Customer	1
born	Born Dow	born@mole.co.ro	Subscriber	0

Permission

Form1

ID	Full Name	Email	Phone Number	Language	Country	Gender	Image Path
	Johnathan Maggy	JohnathanMaggy@emailextension.com	+19999999999	English	United States Of America	Male	C:\Users\Public\Pictures\Sample Pictures\Chrysanthemum.jpg

Export
 Insert

Import
 Update

Print
 Clear

Delete
 Refresh DGV

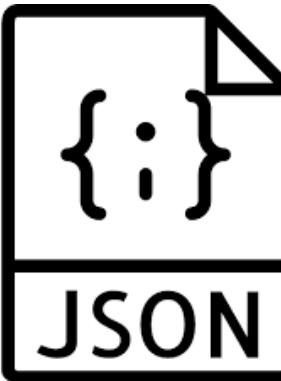
CHOOSE IMAGE

View Selected Rows On New DGV

View Selected Row On New Form

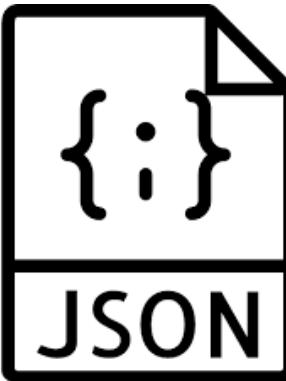
ID	Full Name	Email	Phone Number	Language	Country	Gender	Image Path	Image
10	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	
11	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	
12	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	
13	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	
14	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	
15	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	
16	hokj	chkhg	cghk	cghkk	cghkc	gchk	C:\Users\Public\...	

	Permission	<<	Create	View/read	Update	Delete	Import	Export
Permission to create/view/update/delete assessments rules for a survey	Assessments	<input checked="" type="checkbox"/>						
Permission to create/view/update/delete quota rules for a survey	Quotas	<input checked="" type="checkbox"/>						
Permission to create(data entry)/view/update/delete/import/export responses	Responses	<input checked="" type="checkbox"/>						
Permission to view statistics	Statistics	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>				
Permission on survey (delete). Read permission is used to give access to this group.	Survey	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>



Data Modeling Tool (Reminding)

- DBM Tool v.1.2.2 by Dr. Anan Osothsip (Download from LMS and unzip file)
- It is the tool that introduces how data can be constructed and converted into the SQL and database tables
- The input can be JSON only that process the output into SQL and database tables
- This tool also provides catalog, departments, roles, positions, users, processes, and assets in order to identify what kind of these data to support your business (**will use in Data Accessibility lab class later**)
- The Internet connection is also required for this tool



Data Modeling Tool (Reminding)

- For Data Accessibility, there are the menus provided below:
 - Catalog
 - Data Element Catalogs
 - General Information
 - Business Detail
 - Business Assets
 - UI
 - Departments
 - Roles
 - Positions
 - Users
 - ~~Processes~~ [NO NEED]
 - Assets

Data Modeling Tool Overview (Reminding)

The screenshot shows the jsontool.html interface with several highlighted features:

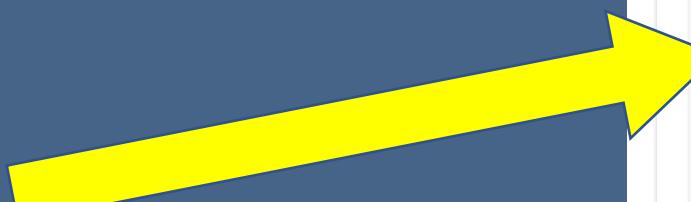
- File Tab:** "jsontool.html" is highlighted with a blue bar and a yellow arrow pointing to it.
- Menu Selection:** A green bar labeled "Menu Selection" contains links: EDITOR, CATALOG, DEPARTMENTS, ROLES, POSITIONS, USERS, and a dropdown arrow.
- Data Model Name:** "Data Model" and "(table1)" are highlighted with a yellow oval.
- Toolbar:** A black bar at the top includes "Data and Process Management" and several editor checkboxes:
 - json editor
 - c# class editor
 - sql editor
 - code editor
 - table editor
 - diagramA yellow arrow points to the "c# class editor" checkbox with the text "Double click to open". A red annotation above the toolbar says "Please uncheck this check box first!".
- Import/Export file:** A yellow circle highlights the three-dot menu icon in the top right, which is also annotated with "Import/Export file". The dropdown menu shows options: import, export, darcula, and version.
- Convert button:** A yellow circle highlights the "CONVERT" button in the center of the interface.
- JSON Input Area:** A yellow box highlights the top section where JSON input is shown. It includes a "JSON" label, a file icon, and a set of icons for copy/paste, delete, and dropdown.
- SQL Output Area:** A yellow box highlights the bottom section where SQL output is shown. It includes a "SQL" label and a dropdown icon.
- Database Table Output Area:** A large orange box highlights the right side of the interface, showing a table structure with columns and rows under the heading "DATABASE TABLE".

Data Modeling Tool (Data Accessibility Lab)

- First, let's try to import the “CustomerOrder-Sample-Real.json” file from LMS **(Open file in Note Pad or any Text Editor and then copy and paste into JSON Input Area)**

```
CustomerOrder-Sample-Real.json

{
  "Customer": {
    "firstname": "John",
    "lastname": "Doe",
    "age": 20,
    "email": "john@gmail.com",
    "shippingaddress": {
      "no": "A1",
      "street": "West Side",
      "city": "Bangkok",
      "zip": 101010
    }
  },
  "orderItems": {
    "foodItem": {
      "foodName": "Noodle",
      "price": 20,
      "qty": 2,
      "discount": 0
    },
    "orderRequest": "None"
  },
  "availableTime": "only noon time"
}
```

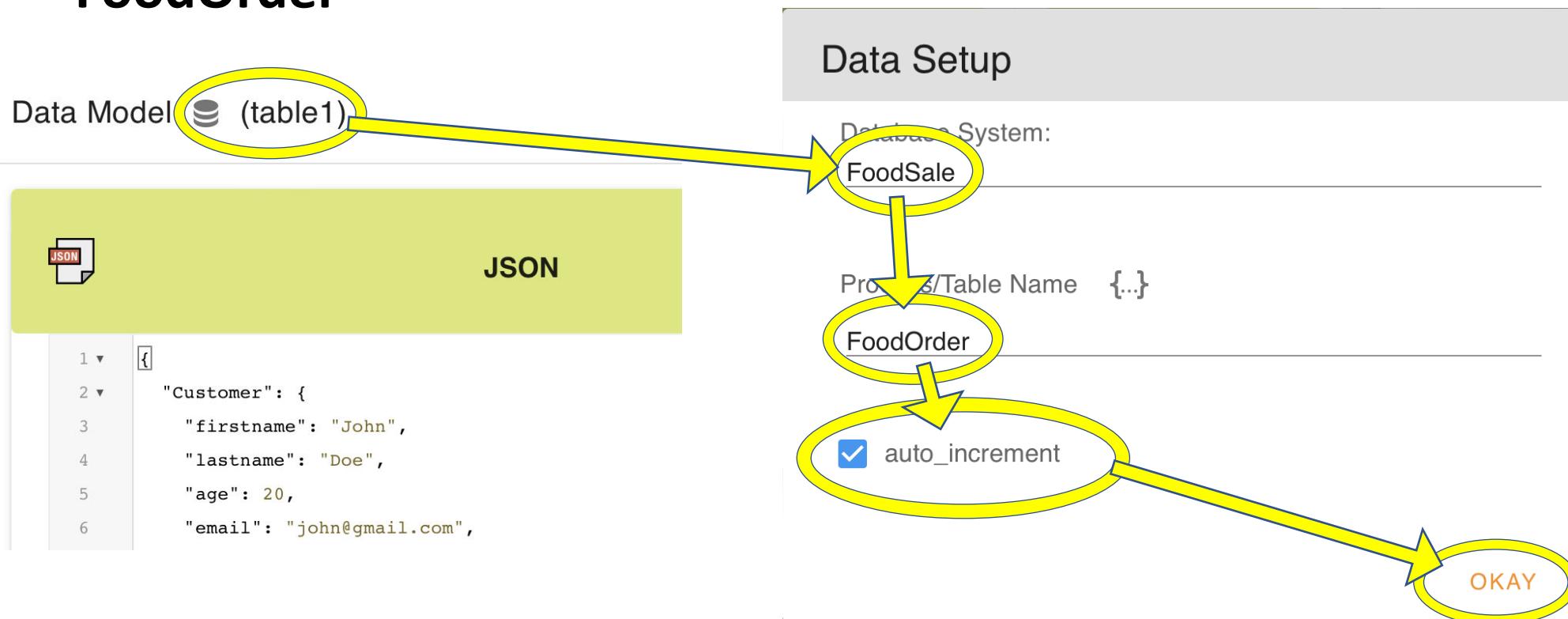


```
JSON

{
  "Customer": {
    "firstname": "John",
    "lastname": "Doe",
    "age": 20,
    "email": "john@gmail.com",
    "shippingaddress": {
      "no": "A1",
      "street": "West Side",
      "city": "Bangkok",
      "zip": 101010
    }
  },
  "orderItems": {
    "foodItem": {
      "foodName": "Noodle",
      "price": 20,
      "qty": 2,
      "discount": 0
    },
    "orderRequest": "None"
  },
  "availableTime": "only noon time"
}
```

Data Modeling Tool (Data Accessibility Lab)

- Second, create Data Setup to “FoodSale” and Process/Table Name to “FoodOrder”



Data Modeling Tool (Data Accessibility Lab)

- Third, click convert button to provide database table

The screenshot shows a data modeling interface with a sidebar on the left and a main content area on the right.

Left Sidebar:

- A dark green header bar with the text "PROCESS >".
- A yellow card-like panel with three small icons: a document, a magnifying glass, and a lock.
- A green button labeled "CONVERT" with a yellow oval and arrow pointing to it.

Main Content Area:

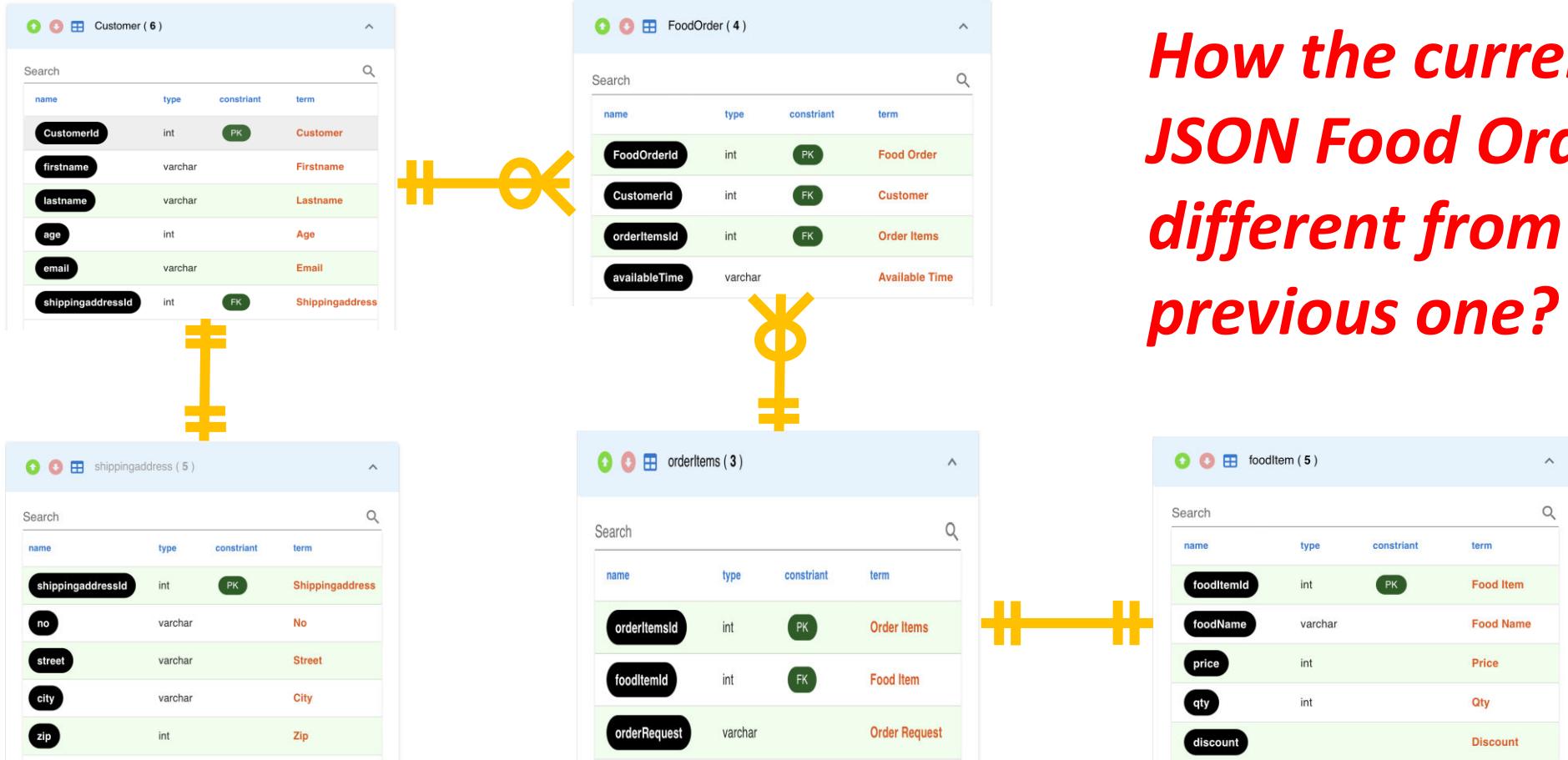
- A title bar labeled "DATABASE TABLE" with an "EXPAND" button.
- A table titled "Customer (6)" showing the structure of the database table.
- The table has columns: name, type, constraint, and term.
- The rows are:

name	type	constraint	term
CustomerId	int	PK	Customer
firstname	varchar		Firstname
lastname	varchar		Lastname
age	int		Age
email	varchar		Email
shippingaddressId	int	FK	Shippingaddress

- At the bottom, there is a pagination control: "Rows per page: 10" and "1-6 of 6".

Data Modeling Tool (Data Accessibility Lab)

- Food Order Table List and Relationship (In DBM Tool)



How the currently JSON Food Order is different from the previous one?

Data Modeling Tool (Data Accessibility Lab)

- When we bring SQL converted and try to import this into mySQL ...

```
mysql> show tables;
+-----+
| Tables_in_foodorder |
+-----+
| Customer
| FoodItem
| FoodOrder
| OrderItems
| Shippingaddress
+-----+
5 rows in set (0.00 sec)
```

```
mysql> describe Customer;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| CustomerId | int | NO | PRI | NULL | auto_increment |
| firstname | longtext | YES | NULL |
| lastname | longtext | YES | NULL |
| age | int | NO | NULL |
| email | longtext | YES | NULL |
| shippingaddressId | int | NO | MUL | NULL |
+-----+-----+-----+-----+-----+-----+
6 rows in set (0.02 sec)

mysql> describe Shippingaddress;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| shippingaddressId | int | NO | PRI | NULL | auto_increment |
| no | longtext | YES | NULL |
| street | longtext | YES | NULL |
| city | longtext | YES | NULL |
| zip | int | NO | NULL |
+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

```
mysql> describe OrderItems;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| orderItemsId | int | NO | PRI | NULL | auto_increment |
| foodItemId | int | NO | MUL | NULL |
| orderRequest | longtext | YES | NULL |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.00 sec)
```

```
mysql> describe FoodItem;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| foodItemId | int | NO | PRI | NULL | auto_increment |
| foodName | longtext | YES | NULL |
| price | int | NO | NULL |
| qty | int | NO | NULL |
+-----+-----+-----+-----+-----+
4 rows in set (0.01 sec)
```

How the currently JSON Food Order is different from the previous one?

```
mysql> describe FoodOrder;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| foodOrderId | int | NO | PRI | NULL | auto_increment |
| CustomerId | int | NO | MUL | NULL |
| orderItemsId | int | NO | MUL | NULL |
| availableTime | longtext | YES | NULL |
+-----+-----+-----+-----+-----+
4 rows in set (0.01 sec)
```

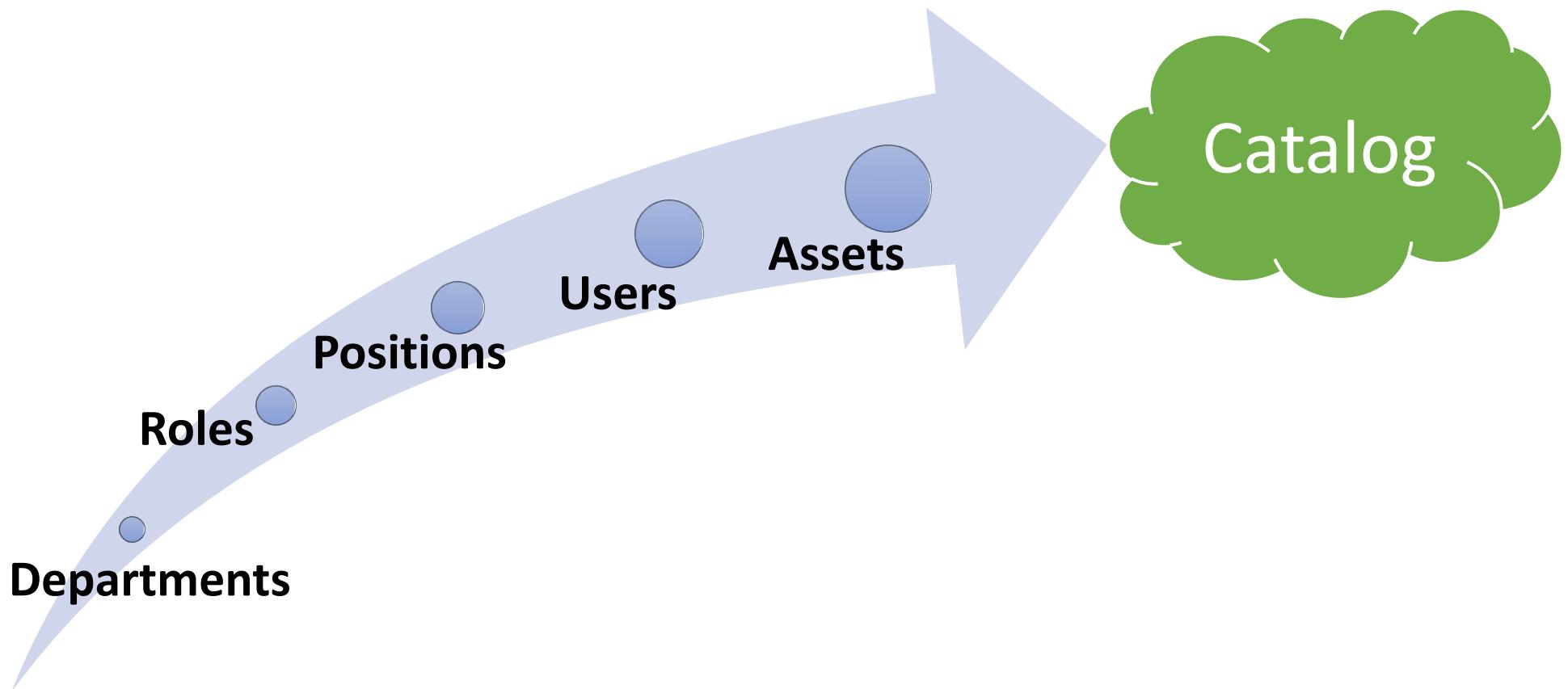
Data Modeling Tool (Data Accessibility Lab)

- Forth, let's check the data accessibility inside Food Sale
 - What are the departments inside food sale?
 - Who has the roles inside food sale?
 - Who has the positions inside food sale?
 - Who are the users inside food sale?
 - What are the assets (report/document/dashboard) inside food sale?
 - What are data element catalogs (in each field) including general information, business detail, business assets, and UI inside food sale?

*Try to questions and answers by yourself ...
(giving for 20 minutes to list the answers)*

Data Modeling Tool (Data Accessibility Lab)

- The process of input data accessibility would be look like ...



Data Modeling Tool (Data Accessibility Lab)

- Fifth, let's try to input departments inside Food Order

Departments			
	no	code	name
	1	MGT	Management
	2	SLM	Sales & Marketing
	3	CSH	Cashier
	4	KIT	Kitchen

Data Modeling Tool (Data Accessibility Lab)

- Sixth, let's try to input roles inside Food Order

ROLES



Search

no	name
1	Decision Maker
2	Marketing Plan
3	Sell Food
4	Collect Money and Bill
5	Food Provider

Data Modeling Tool (Data Accessibility Lab)

- Seventh, let's try to input positions inside Food Order

POSITIONS

The screenshot shows a data modeling interface with a table titled "POSITIONS". The table has two columns: "no" and "name". There is a green "+" button to add new rows. A search bar is at the top left. The table contains five rows of data:

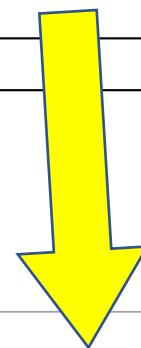
no	name
1	Manager
2	Marketer
3	Salesperson
4	Cashier
5	Chef

Data Modeling Tool (Data Accessibility Lab)

- Eighth, let's try to input users inside Food Order

no	name	role	department	position	email	phone
1	Ms.Yor Forger	Decision Maker	Management	Manager	yorforger@gmail.com	0825224222
2	Mr. Loid Forger	Marketing Plan, Collect Money and Bill	Sales & Marketing, Cashier	Marketer, Cashier	loidforger@gmail.com	0852233444
3	Ms. Anya Forger	Sell Food	Sales & Marketing	Salesperon	anyaforger@gmail.com	0943224422
4	Mr. Yuri Briar	Food Provider	Kitchen	Chef	yuribriar@gmail.com	0871223344

Users



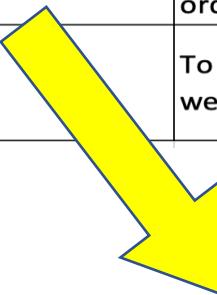
no	name	role	department	position	email	phone
1	Ms.Yor Forger	Decision Maker	Management	Manager	yorforger@gmail.com	0825224222
2	Mr. Loid Forger	Marketing Plan Collect Money and Bill	Sales & Marketing	Cashier	loidforger@gmail.com	0852233444
3	Ms. Anya Forger	Sell Food	Sales & Marketing	Salesperson	anyaforger@gmail.com	0943224422
4	Mr. Yuri Briar	Food Provider	Kitchen	Chef	yuribriar@gmail.com	0871223344

Data Modeling Tool (Data Accessibility Lab)

- Ninth, let's try to input assets (report/document/dashboard) inside Food Order

no	name	type	description	department	access
1	Top 10 Food Order	report	To provide the top 10 food order and report in weekly	Sales & Marketing, Kitchen, Management	Sell Food, Food Provider, Marketing Plan, Decision Maker
2	Top 10 Customer Order	report	To provide the top 10 customer order and report in weekly	Sales & Marketing, Management	Marketing Plan, Sell Food, Decision Maker
3	Food Sales Report	report	To provide food sales report in weekly, monthly	Management, Cashier, Kitchen	Collect Money and Bill, Food Provider, Decision Maker

assets (Report/Document/Dashboard)



no	name	type	description	department ↑	access
3	Food Sales Rep	report	To provide food	Management, Cashier, Kitchen	Collect Money and Bill, Food Provider, Decision Maker
1	Top 10 Food Or	report	To provide the t	Sales & Marketing, Kitchen, Management	Sell Food, Food Provider, Marketing Plan, Decision Maker
2	Top 10 Custom	report	To provide the t	Sales & Marketing, Management	Marketing Plan, Sell Food, Decision Maker

Data Modeling Tool (Data Accessibility Lab)

- Tenth, let's try to check Data Catalog for each field in every tables and input the summary inside General Information, Business Detail, Business Assets, and UI (e.g., click **CustomerId** to display data element catalogs and input information)

The screenshot shows two panels. On the left is a table titled "Customer (6)" with columns: name, type, constraint, and term. The table contains six rows with fields: CustomerId (int, PK, Customer), firstname (varchar, Firstname), lastname (varchar, Lastname), age (int, Age), email (varchar, Email), and shippingaddressId (int, FK, Shippingaddress). The "CustomerId" row is highlighted with a yellow oval. On the right is a panel titled "Data Element Catalogs > Customer" under "General Information". It shows a list of fields: Name (CustomerId), Term (Customer), Status (Null), Type (int), and Constraints (PK). The "Customer" entry in the term list is circled in yellow, and a large yellow arrow points from the "CustomerId" field in the table to this circled entry.

Data Modeling Tool (Data Accessibility Lab)

- Summary inside **CustomerId**'s Data Catalog (1)

Data Element Catalogs > Customer

General Information	
Name	CustomerId
Term	Customer
Status	Null
Type	int
Constraints	PK

*Let's try to check other
Data Catalogs and
input the summary ...*

Data Modeling Tool (Data Accessibility Lab)

- Summary inside CustomerId's Data Catalog (2)

Business Detail

Department
Sales & Marketing

Business Owner
Ms. Yor Forger

Technical Owner
Mr. Loid Forger

Data Producer
Mr. Loid Forger

Data Consumer
Ms. Yor Forger **Ms. Anya Forger**

Data Policy
To provide an unique value for customer in order not to duplicate data records

Data Rule
Require only 4 digit numbers started from 1001

Processes
1. Auto generate 4 digit numbers started from 1001
2. Not duplicate value

Access
Sell Food

*Let's try to check other
Data Catalogs and
input the summary ...*

Business Assets

Asset

report:Top 10 Customer Order

UI

Show

Inside Customer table

Select

Customer table

Data Modeling Tool (Data Accessibility Lab)

- When you complete to input one data catalog for **CustomerId**, the field color is changed from black to orange

name	type	constraint	term
CustomerId	int	PK	Customer
firstname	varchar		Firstname
lastname	varchar		Lastname
age	int		Age
email	varchar		Email
shippingaddressId	int	FK	Shippingaddress

*Let's try to check other
Data Catalogs and
input the summary ...*

Data Modeling Tool (Data Accessibility Lab)

- When you finish the summary ...
 - Export file to “**json.txt**” and try to rename to another name (e.g., **FoodOrder-json.txt**)
 - In the future, you can import this file again by **copy & paste** the syntax into **import dialog**
 - Review **Data Model Diagram** slide no. 29-31 again

Data Modeling Tool (Data Accessibility Lab)

- Let's try this (in class)
 - Review the JSON syntax in what you have done and save from the car rental business in Data Modeling Diagram using DBM Tool
 - Provide the Data Accessibility inside the car rental business follow these questions:
 - What are the departments inside car rental?
 - Who has the roles inside car rental?
 - Who has the positions inside car rental?
 - Who are the users inside car rental?
 - What are the assets (report/document/dashboard) inside car rental?
 - What are data element catalogs (in each field) including general information, business detail, business assets, and UI inside car rental?
 - Export file into “**json.txt**” and rename file to “**CarRental-json.txt**”

Topic 08: Data Visualization

BDM3302: Data Management

What is Data Visualization?



- The representation of data through use of common graphics, such as charts, plots, infographics, and even animations
- These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand
- Data visualization can be utilized for a variety of purposes, and it is important to note that it is not only reserved for use by data teams
- Management also leverages it to convey organizational structure and hierarchy while data analysts and data scientists use it to discover and explain patterns and trends

Why Data Visualization?



- [Harvard Business Review](#) categorizes data visualization into four key purposes:
 - Idea generation
 - Idea illustration
 - Visual discovery
 - Everyday dataviz

Why Data Visualization?



- Idea generation
 - Commonly used to spur idea generation across teams
 - Frequently leverage during brainstorming or design thinking at the start of a project by supporting the collection of different perspectives and highlighting the common concerns of the collective
 - While these visualizations are usually unpolished and unrefined, they help set the foundation within the project to ensure that the team is aligned on the problem that they're looking to address for key stakeholders

Why Data Visualization?



- Idea illustration
 - Assists in conveying an idea, such as a tactic or process
 - Commonly used in learning settings, such as tutorials, certification courses, centers of excellence, but it can also be used to represent organization structures or processes, facilitating communication between the right individuals for specific tasks
 - Project managers frequently use Gantt charts and waterfall charts to illustrate workflows

Why Data Visualization?



- Visual discovery and Every dataviz
 - While visual discovery helps data analysts, data scientists, and other data professionals identify patterns and trends within a dataset, every day dataviz supports the subsequent storytelling after a new insight has been found
 - Data visualization is a critical step in the data science process, helping teams and individuals convey data more effectively to colleagues and decision makers
 - However, it's important to remember that it is a skillset that can and should extend beyond your core analytics team

Type of Data Visualizations



- **Tables:** This consists of rows and columns used to compare variables. Tables can show a great deal of information in a structured way, but they can also overwhelm users that are simply looking for high-level trends
- **Pie charts and stacked bar charts:** These graphs are divided into sections that represent parts of a whole. They provide a simple way to organize data and compare the size of each component to one other
- **Line graphs and area charts:** These visuals show change in one or more quantities by plotting a series of data points over time. Line graphs utilize lines to demonstrate these changes while area charts connect data points with line segments, stacking variables on top of one another and using color to distinguish between variables

Type of Data Visualizations



- **Histograms:** This graph plots a distribution of numbers using a bar chart (with no spaces between the bars), representing the quantity of data that falls within a particular range. This visual makes it easy for an end user to identify outliers within a given dataset
- **Scatter plots:** These visuals are beneficial in revealing the relationship between two variables, and they are commonly used within regression data analysis. However, these can sometimes be confused with bubble charts, which are used to visualize three variables via the x-axis, the y-axis, and the size of the bubble
- **Heat maps:** These graphical displays are helpful in visualizing behavioral data by location. This can be a location on a map, or even a webpage
- **Tree maps:** Display hierarchical data as a set of nested shapes, typically rectangles. Tree maps are great for comparing the proportions between categories via their area size

Type of Data Visualizations



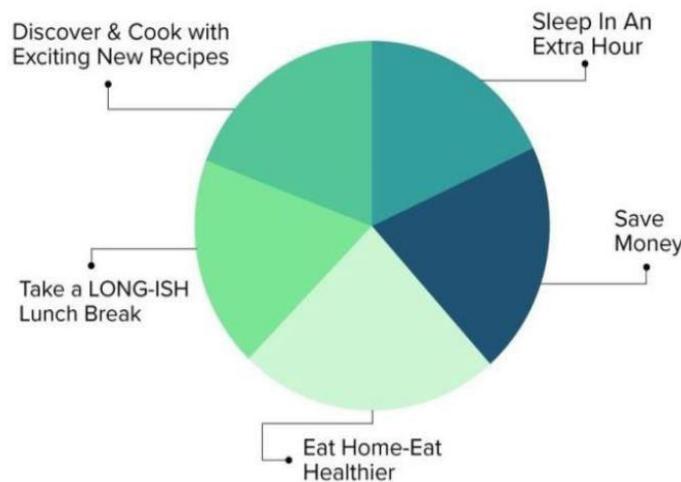
Table 3. Energy production by major source from 1960 to 1980 ¹					
Year	Total production (quad Btu)	Percent production			
		Coal	Petroleum	Natural gas	Other ²
1960	41.5	26.1	36.0	34.0	3.9
1970	62.1	23.5	32.9	38.9	4.7
1980	64.8	28.7	28.2	34.2	8.9

¹Source: U.S. Energy Information Administration, *Annual Energy Review*

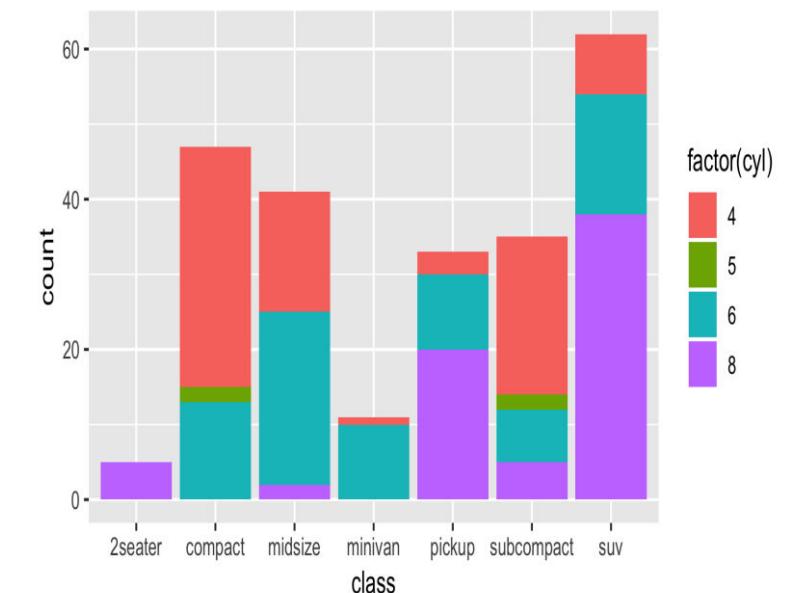
²Includes hydropower, nuclear power, geothermal power, and others.

Tables

Advantages of Working From Home



Pie Charts

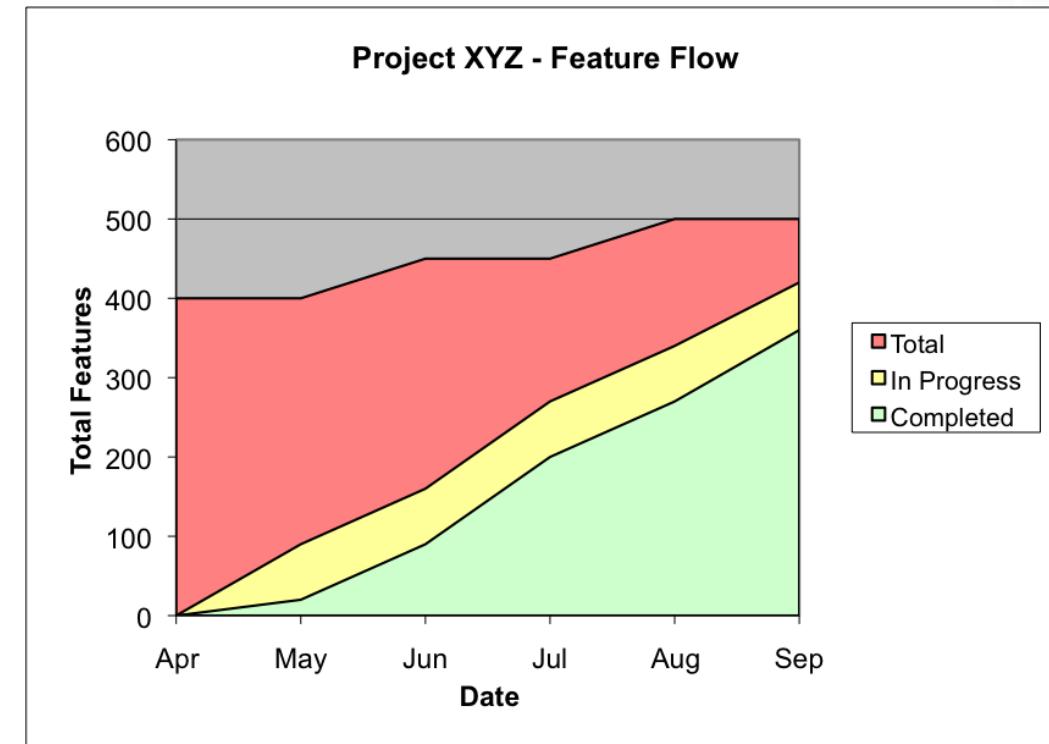


Stack Bar Charts

Type of Data Visualizations

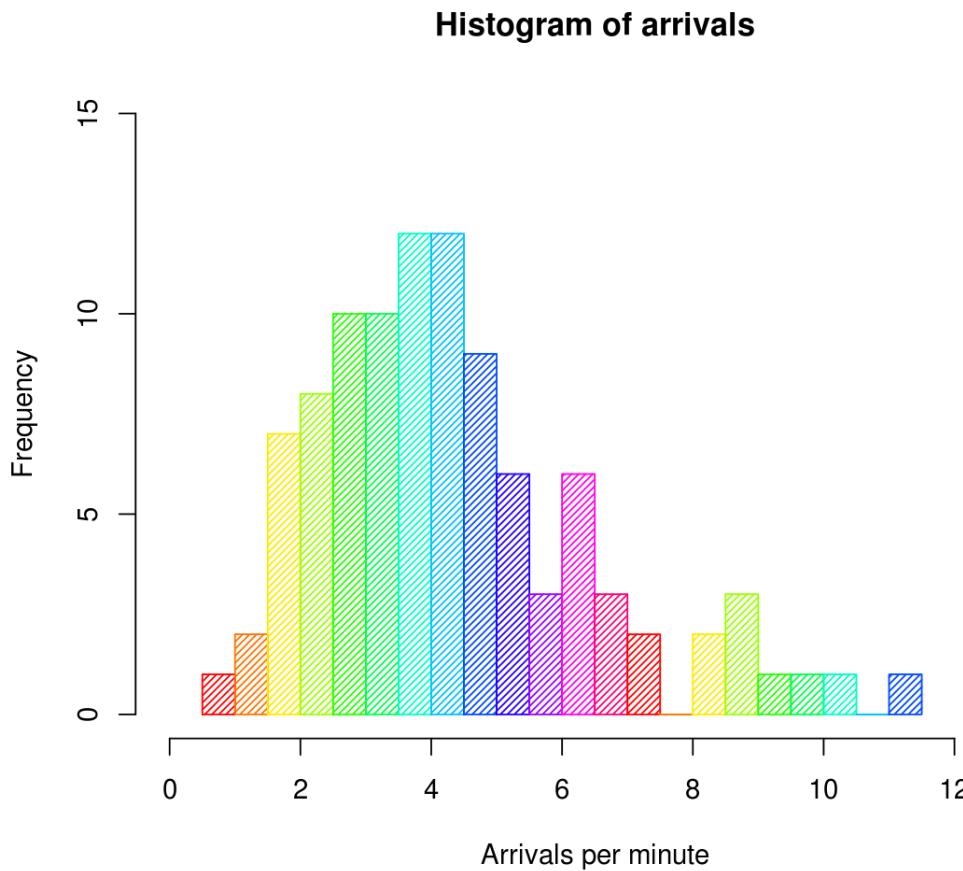


Line Graphs

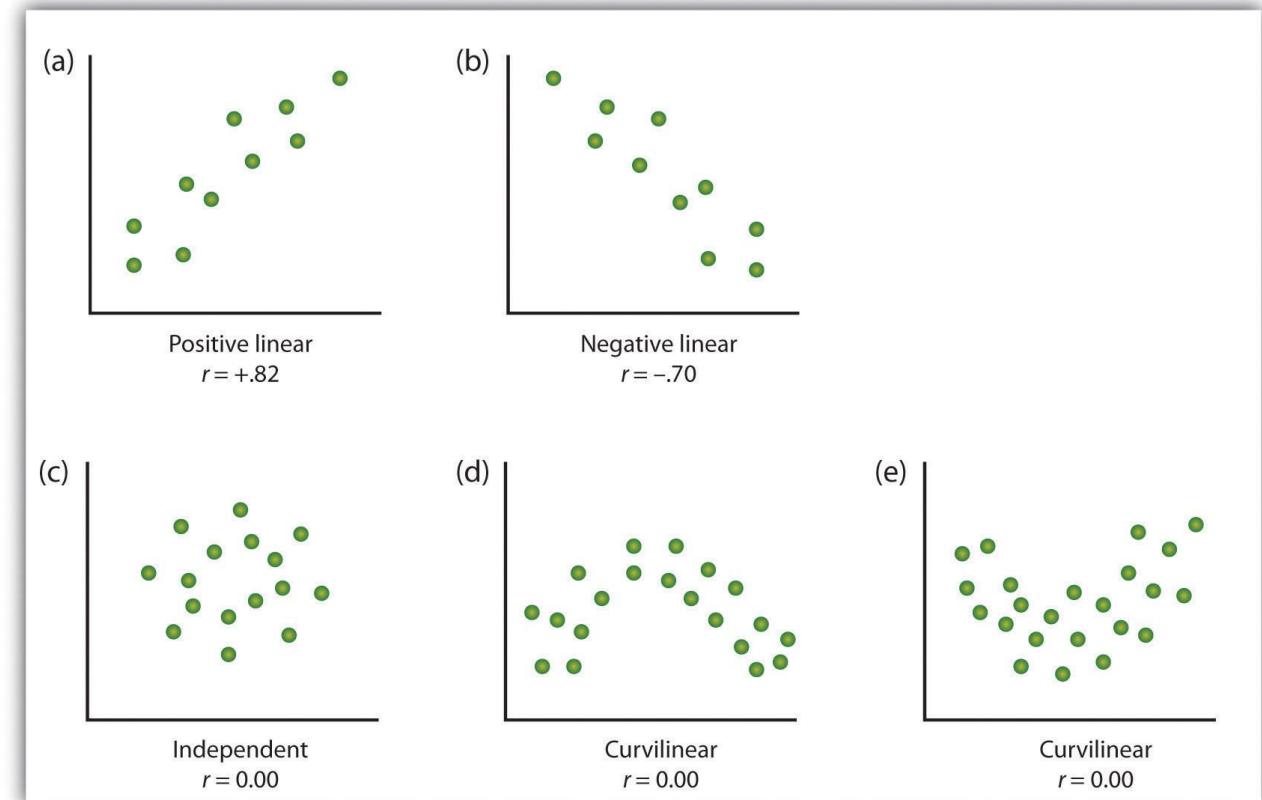


Area Charts

Type of Data Visualizations

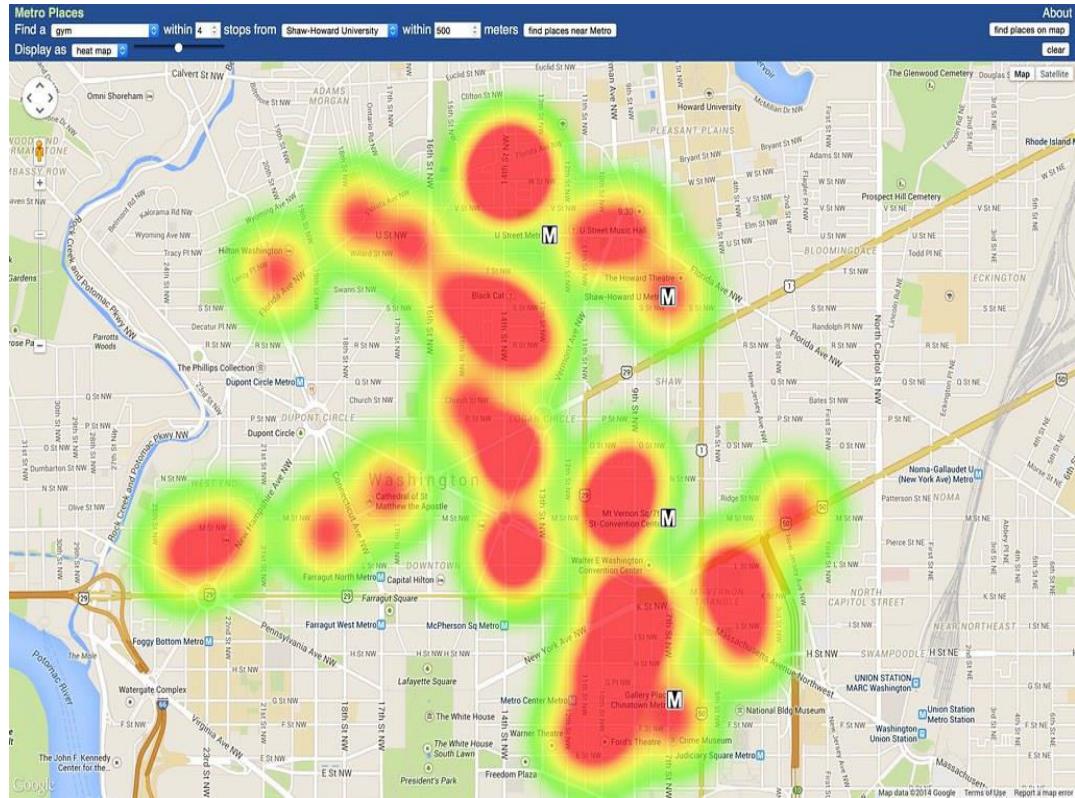


Histograms

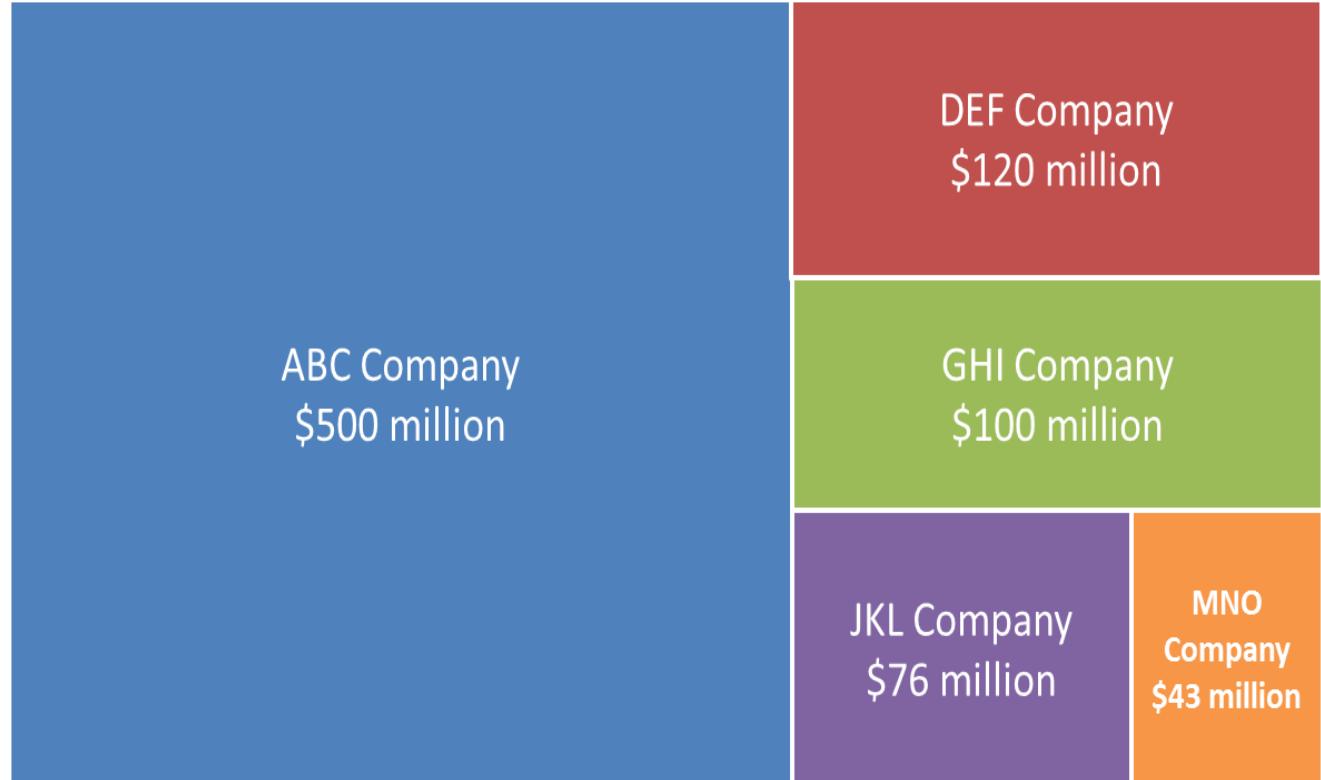


Scatter Plots

Type of Data Visualizations



Heat Maps

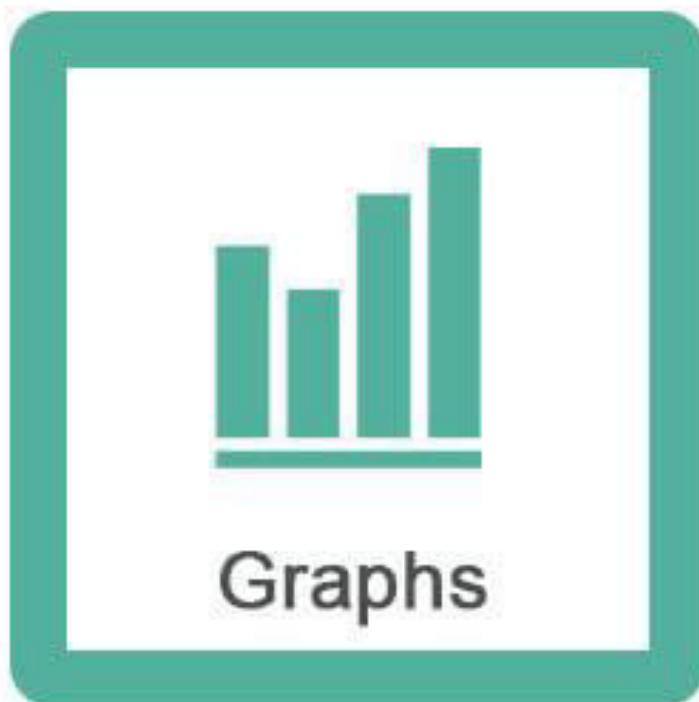


Tree Maps

Graphs vs Charts



Plot number in x,y axis



Graphs

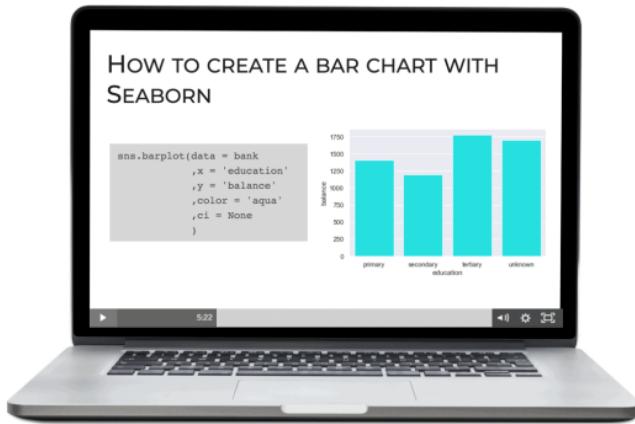
Compare and run trends

vs



Charts

How to create Data Visualization?



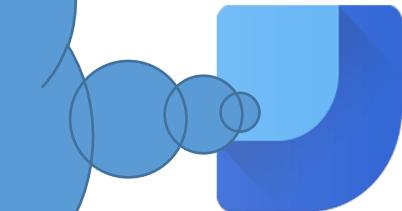
- In this era of big data, businesses need a data visualization tool that would solve all their visualization needs
- The effectiveness of any data visualization tool lies in the richness of data visualization capabilities that it offers
- There are four basic data visualization tools you should know: Excel, Tableau, Google Data Studio, Power BI

Data Visualization Tools

Basic spreadsheet tool which is ease of use to collect data but cannot enhance in huge amount of data.



Simple and nice with collaboration. Huge amount of data can be accepted and free of charge but based on web browser.



Google Data Studio



Create interactive graphs, stunning maps, and live dashboards in minutes. Support both cloud and public desktop (free).

Easily connect to, model, and visualize your data. Get fast, AI-powered answers to your business questions. But require paid subscription version when you would like to share for collaboration.



+a**bleau**

Which tool do we use in lab class?

- We decide to choose **Tableau** due to free of charge for education use (Power BI requires paid subscription before sharing)
- There is the **Public Desktop** as a free version for you to download in your own desktop or laptop at home (**This version can save file in Tableau Public Cloud only**)
- It can be easily to share the link to the audiences that open from cloud directly
- Inside computer lab room SC0501 & SC0502, there is the full usage educational license for **Professional Desktop** version (apply year by year)



+ableau

Public Desktop vs Professional Desktop

- Public Desktop

- Essentially a free version of Tableau visualization tool
- It allows you to use most of the tool functions
- You can create visualizations and connect to CSV, Text and Excel documents
- The largest difference is that Tableau Public does not allow you to save your workbooks locally

- Professional Desktop

- The professional version of this can transform, process and store huge volumes of data which is responsible for all the data-driven decision making of an organization
- You can create visualizations for full edition and can save in desktop or laptop computer
- Require paid subscription
- Computer lab room SC0501 & SC0502 has already set for the educational license



tableau

Set up Tableau Public Desktop at home (Quick Install)

- Go to this [website](#) and click orange button “**DOWNLOAD THE APP**” and then click “**DOWNLOAD THE APP**” again
- Register your personal information and then click “**DOWNLOAD THE APP**” again to download the app
- When the app is downloaded, double click file to install and follow the instruction until complete



tableau

Set up Tableau Public Desktop at home (Quick Install)

You'll be exploring in minutes

Create interactive graphs, stunning maps, and live dashboards in minutes. Save your viz to your Tableau Public profile, and share it anywhere on the web. Anyone can do it, it's that easy—and it's free.



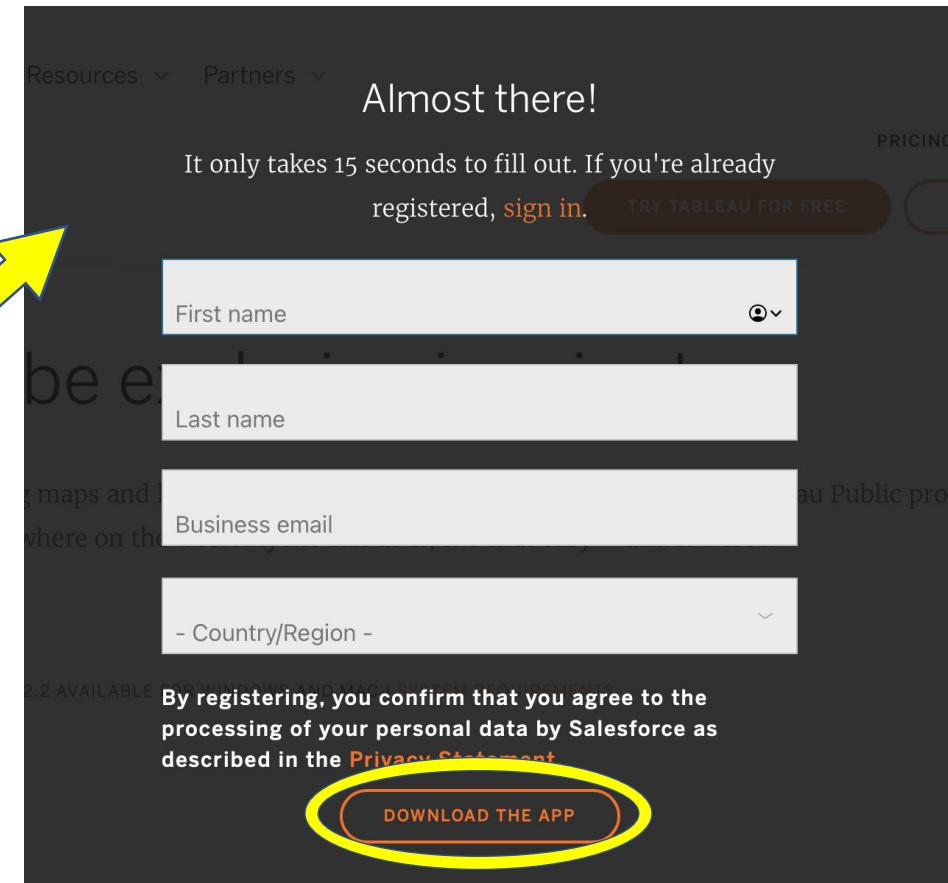
2022.2 Available for Windows and Mac | System Requirements

You'll be exploring in minutes

Create interactive graphs, stunning maps, and live dashboards in minutes. Save your viz to your Tableau Public profile, and share it anywhere on the web. Anyone can do it, it's that easy—and it's free.



2022.2 AVAILABLE FOR WINDOWS AND MAC | SYSTEM REQUIREMENTS



Input your personal information, and then click “DOWNLOAD THE APP”



tableau

Set up Tableau Public Desktop at home (Quick Register)

The image shows the Tableau Public website interface. At the top right, there are 'SIGN UP' and 'SIGN IN' buttons, both highlighted with yellow circles. A large yellow arrow points from the 'SIGN IN' button down to a sample profile page for 'Tanakom'. The profile page displays basic user information like 'Vizzes 1', 'Favourites 0', 'Following 0', and 'Followers 0'. Below this, there are two small visualizations: a bar chart titled 'City' and a map titled 'Map of Country with Unit Sold'. A second yellow arrow points from the 'Create My Profile' button on the sign-up page down to the same area on the profile page.

Create a Profile

Name
Real names are important to us and build a sense of community.

Email
Use your email to sign in to Tableau Public. Nobody sees this but us. We promise never to rent, sell or barter your personal information to anyone.

Password
Must be a minimum of 8 characters and contain alphabetic, numeric and special characters.

Confirm

Country
Select a country

I agree to the [Terms](#), [Data Policy](#) and [Community Code](#).
By registering, you confirm that you agree to the processing of your personal data by Salesforce as described in the [Privacy Statement](#).

I'm not a robot

Create My Profile

When you sign in, the screen will display like this ...

Discover Blog Resources About SIGN UP SIGN IN

Shar with the +tableau public

tanakom

Vizzes 1 Favourites 0 Following 0 Followers 0

Create a Viz

City

Compare Top 3 Products

Map of Country with Unit Sold

© 2022 Mapbox © OpenStreetMap

Year of Profit

Product

Country

Profit

117,319,471,014

Regist your personal information, and then click "Create My Profile"



tableau

Tableau Overview

Tableau Public - SampleKevinCookies-1

Dashboard Layout Default Phone Device Preview

Size min 420x560 - max 650x8...

Sheets Map of Country wit... Compare Top 3 ... Year of Profit

Object Selection

Objects Horizontal Blank Vertical Navigation Text Download Image Extension Web Page Ask Data Data Story Tiled Floating Show dashboard title

Input Object Area

Compare Top 3 Products

Product Chocolate Chip White Chocolate Macada.. Snickerdoodle

Units Sold 0K 50K 100K 150K 200K 250K 300K

Product Chocolate C.. Snickerdood.. White Choco..

Country Canada France Germany Mexico United States

Profit 117,319 471,014

Map of Country with Unit Sold

Year of Profit

2019 2019 2019 2019 2019

Menu Selection

Show Me

Select view

Tab Selection

Data Source My Dashboard Map of Country with Unit Sold Compare Top 3 Products Year of Profit + + +

Tanakom



tableau

Tableau Lab Requirements

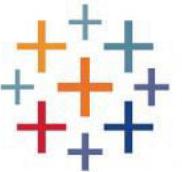
- Tableau **Public Desktop or Professional Desktop**
- **Country.csv** (Download from LMS)



tableau

Before using Tableau

- Source usage (CSV, JSON, XLSX)
 - Data Schema / Data Table
 - Feature / Information / Column / Attribute / Property
 - Data Type (Number, Text)
 - Measure (Continuous Data) and Dimension (Discrete Data)
 - Continuous data (1,2,3,...) uses for aggregation (sum, average, count, max, min)
 - Discrete data (China, USA) uses for comparison and segmentation
- Step to Provide
 - Data Selection
 - Visualization
 - Analytic



+tableau

Tableau Lab (Let's Try this)

- First, let's try to import **Country.csv** into Tableau

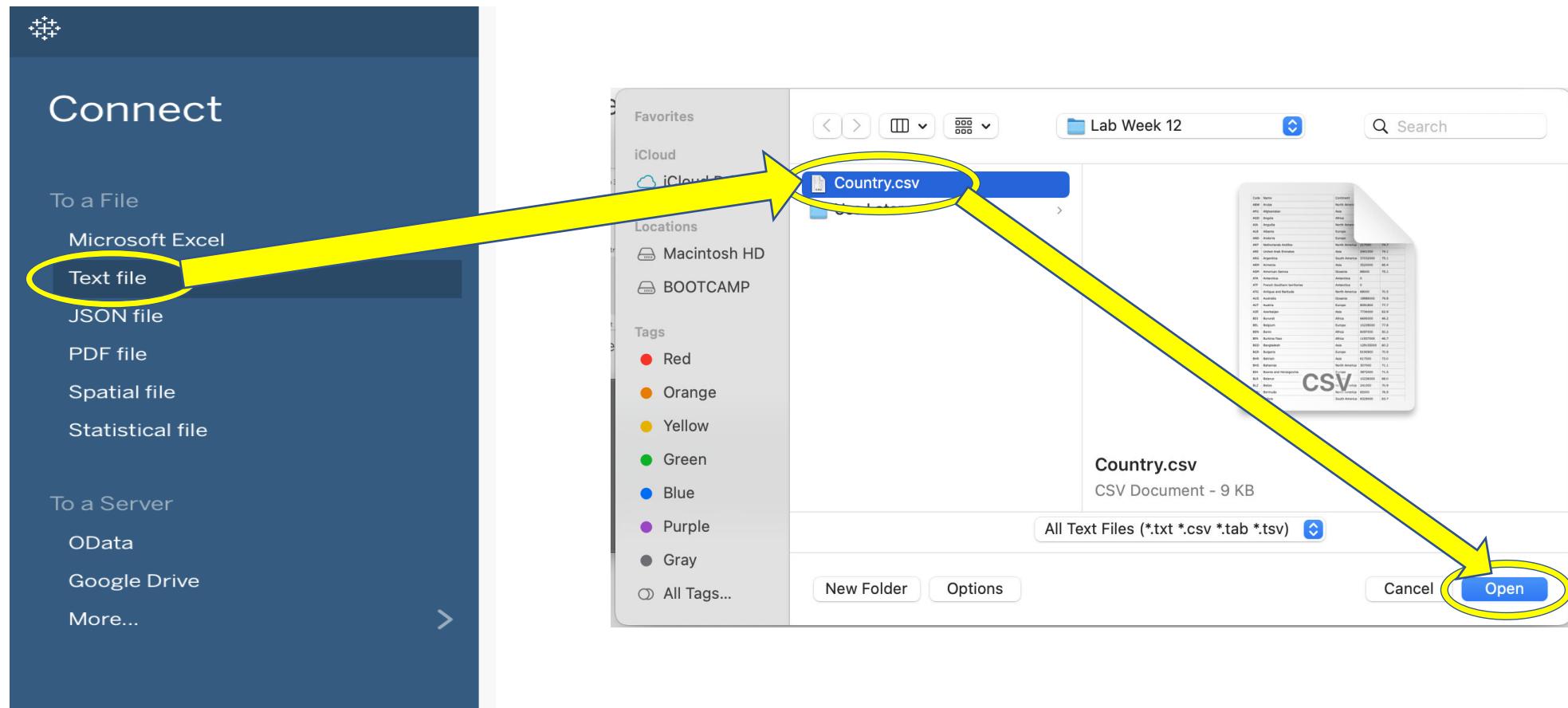


Tableau Lab (Let's Try this)



+tableau

- Second, when you have imported data, check and clean data first before use

Country

Country
Text file

Connections Add

Files

Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.

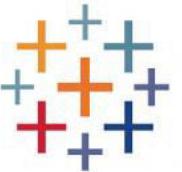
Country.csv

Country.csv

Need more data?
Drag tables here to relate them. [Learn more](#)

Clean data this area before click "Sheet1"

Code	Name	Continent	Population	Life Expectancy
ABW	Country.csv.Code	North America	103,000	78.4000
AFG	Afghanistan	Asia	22,720,000	45.9000
AGO	Angola	Africa	12,878,000	38.3000
AIA	Anguilla	North America	8,000	76.1000
ALB	Albania	Europe	3,401,200	71.6000
AND	Andorra	Europe	78,000	83.5000
ANT	Netherlands Antilles	North America	217,000	74.7000



+a**b**eau

Tableau Lab (Let's Try this)

- Third, when you are in “Sheet1”, try to see the data on the left-hand side and select it to drag drop into **Rows** or **Columns**

The screenshot shows the Tableau interface with the following elements:

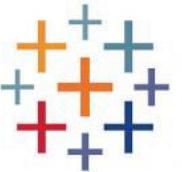
- Data Pane:** On the left, under "Tables", "Country" is selected. The "Measure Names" field is highlighted with a blue selection bar.
- Columns/Rows Shelf:** At the top right, there are two shelves: "Columns" (with a "Rows" button) and "Rows". A large yellow arrow points from the "Measure Names" field in the Data pane towards the "Columns" shelf.
- Marks Shelf:** On the right, under "Marks", "Automatic" is selected, with options for Color, Size, Text, Detail, and Tooltip.
- Sheet 1:** The main workspace is labeled "Sheet 1".

Annotations:

- A red box labeled "Discrete Data" highlights the "Country" field in the Data pane.
- A red box labeled "Continuous Data" highlights the "Measure Names" field in the Data pane.
- A red box highlights the "Columns" shelf.

Suppose that you would like to know the total population in each continent.

Which data would you like to display in rows and column?



+a**b**|eau

Tableau Lab (Let's Try this)

- Four, follow step 3 that you would like to display total population in each continent. Try to drag drop **Population** into **Rows** and **Continent** into **Columns**. The chart will display like this

Data Analytics < Pages

Country

Search Filters

Tables

- Code
- Continent
- Name
- Measure Names
- Life Expectancy
- Population
- Country.csv(Count)
- Measure Values

Columns

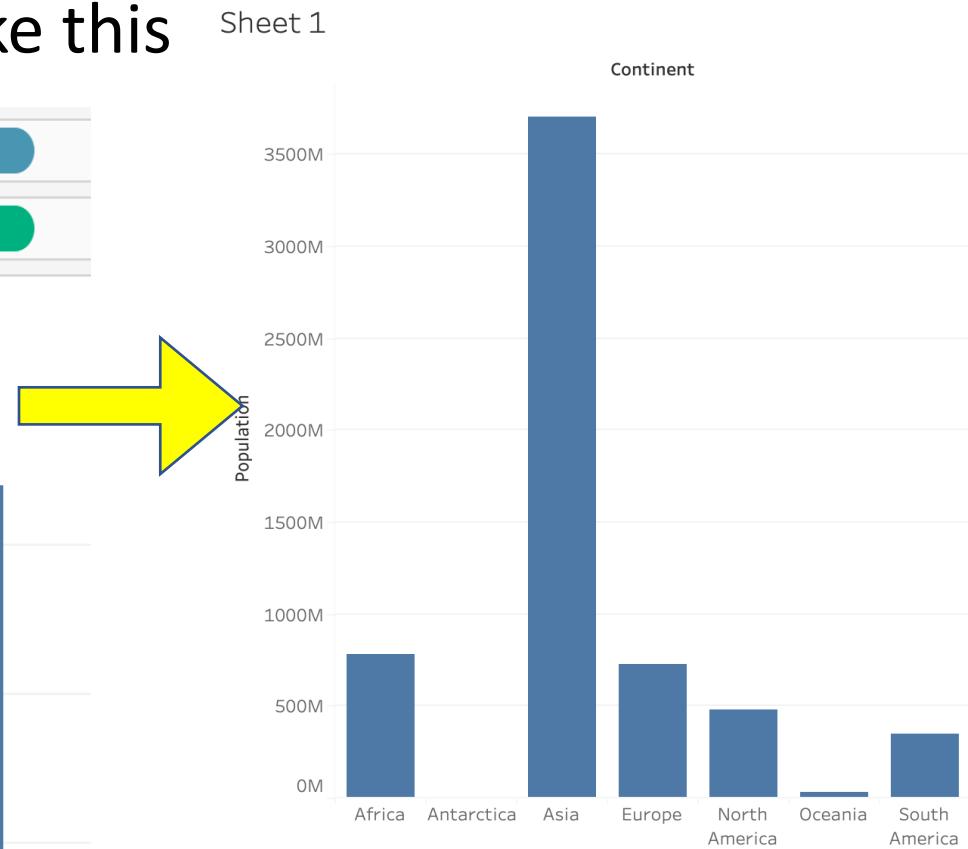
ROWS

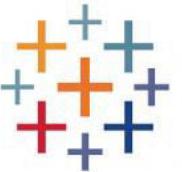
Sheet 1

Mark Type: Automatic

Color Size Label

Detail Tooltip

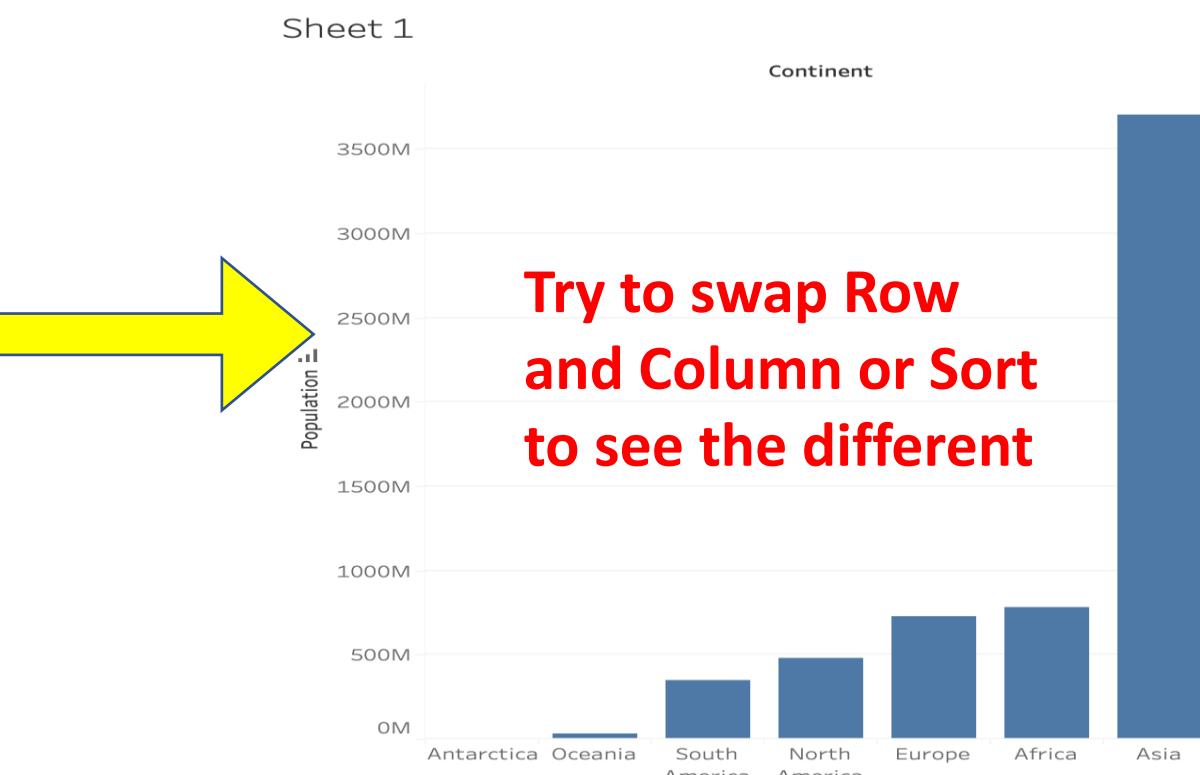
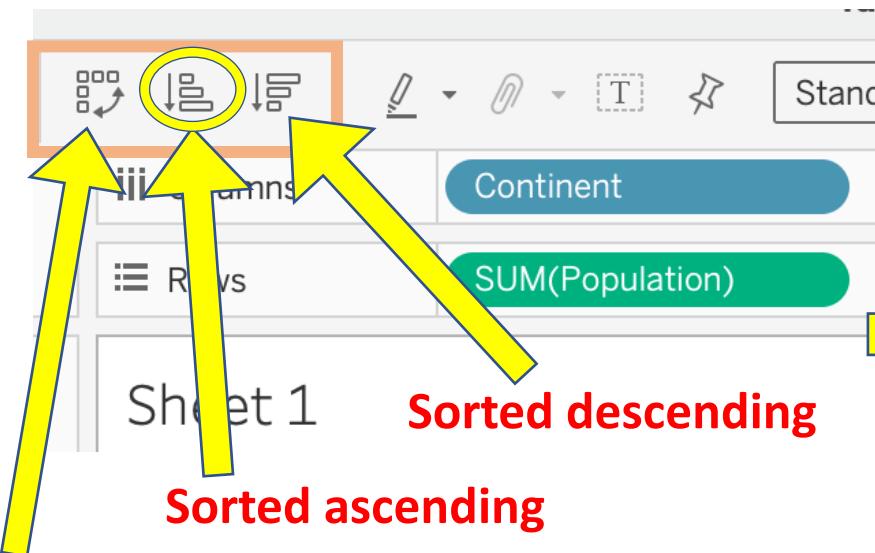


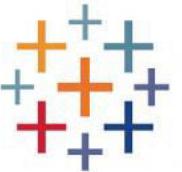


+a**bleau**

Tableau Lab (Let's Try this)

- Five, check the menu above to **Swap Rows and Columns** or **Sort** in an **Ascending** or **Descending** (e.g., Sort in an Ascending)

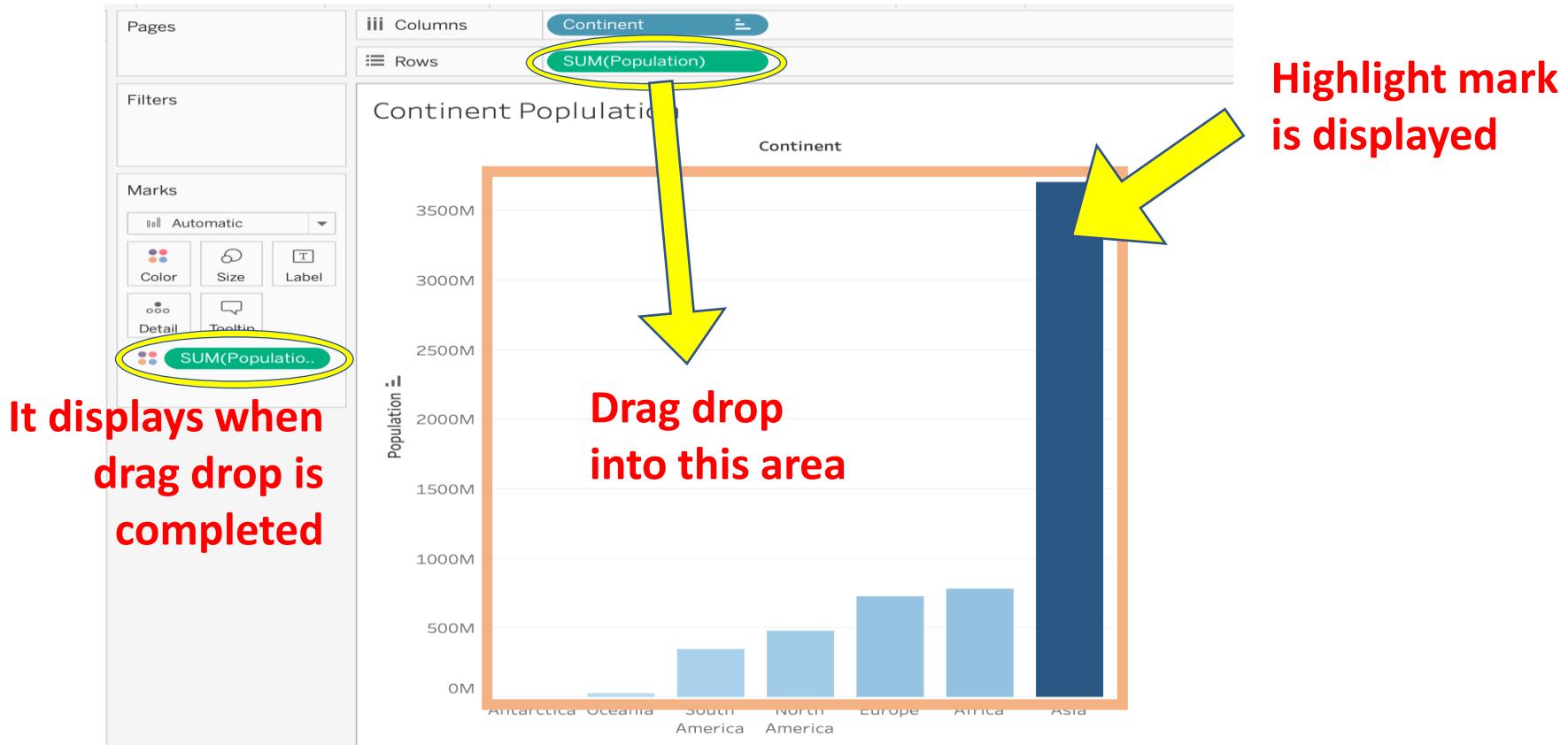


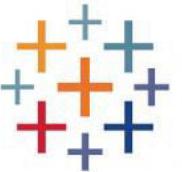


+a**bleau**

Tableau Lab (Let's Try this)

- Six, try to drag drop **Population** from **Rows** into chart area. The highlight mark of maximum population will display



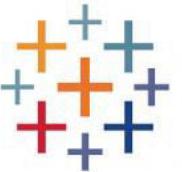


+a**bleau**

Tableau Lab (Let's Try this)

- Seven, rename **Sheet1** tab to **Continent Population** and right click to duplicate this sheet. Then rename the duplicate sheet to **Population Number**

The screenshot shows a Tableau worksheet titled "Continent Population". The bottom navigation bar displays the tabs "Continent Population" and "Population Number". A yellow arrow points from the "Population Number" tab to a right-click context menu. The menu includes options like "New Worksheet", "New Dashboard", "New Story", "Paste", "Rename", "Delete", "Duplicate" (which is highlighted in blue), "Duplicate as Crosstab", "Hide All Sheets", "Unhide All Sheets", "Copy Formatting", "Paste Formatting", and "Color". At the bottom of the screen, the status bar indicates "7 columns SUM(Population): 6,078,749,450".



+a**bleau**

Tableau Lab (Let's Try this)

- Eight, go to **Population Number** and try to remove **Population** by right click and choose Remove or drag drop to some areas outside

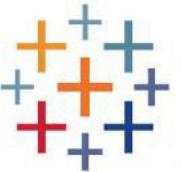
The screenshot shows the Tableau interface with a data view and a sidebar.

Top Bar: Shows "Columns" and "Continent" selected in the dropdown menus. There is also a "Rows" button.

View Area: The title "Population Number" is displayed above a table. The table has "Continent" as the column header and rows for "Antarct..", "Oceania", "South America", "North America", "Europe", "Africa", and "Asia". Each row contains the value "Abc".

Sidebar: On the left side, there is a vertical toolbar with icons for "Text" and a dropdown arrow.

	Continent						
	Antarct..	Oceania	South America	North America	Europe	Africa	Asia
	Abc	Abc	Abc	Abc	Abc	Abc	Abc



+tableau

Tableau Lab (Let's Try this)

- Nine, Drag drop **Population** into “Abc” area and see the result

The screenshot shows the Tableau interface with a data view and a calculated field editor.

Data View:

- Top navigation: Data, Analytics, Country.
- Search bar.
- Tables list:
 - Code
 - Continent
 - Name
 - Measure Names
 - Life Expectancy
 - Population** (highlighted with a yellow oval)
 - Country.csv (Count)
 - Measure Values

Calculated Field Editor:

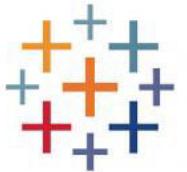
- Columns: Continent (selected).
- Rows: Population Number.
- Content pane:

Population Number

Continent	Population Number
Antarctica	Abc
Oceania	Abc
South America	Abc
North America	Abc
Europe	Abc
Africa	Abc
Asia	Abc

Data Table:

Continent	Population Number
Antarctica	0
Oceania	30,401,150
South America	345,780,000
North America	482,993,000
Europe	730,074,600
Africa	784,475,000
Asia	3,705,025,700



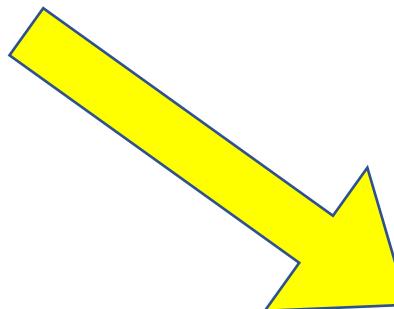
+a**bleau**

Tableau Lab (Let's Try this)

- Ten, try to **Swap Rows and Columns and Sorted descending**

Population Number

Continent						
Antarctica	Oceania	South America	North America	Europe	Africa	Asia
0	30,401,150	345,780,000	482,993,000	730,074,600	784,475,000	3,705,025,700



The screenshot shows a Tableau interface with the following components:

- Toolbar:** Includes icons for refresh, columns, rows, and standard mode.
- Rows Panel:** Shows "Rows" and "Continent" selected.
- Population Number:** A list of continents sorted by population in descending order.
- Data:** The sorted data is as follows:

Continent	Population
Asia	3,705,025,700
Africa	784,475,000
Europe	730,074,600
North America	482,993,000
South America	345,780,000
Oceania	30,401,150
Antarctica	0



+a**b**eau

Tableau Lab (Let's Try this)

- Eleven, try to hold **CTRL** or **command** key and click to drag drop **Population** into **Color**. The result will display like this

The screenshot shows two views of the Tableau interface. On the left, the 'Marks' shelf is open, showing options for Automatic, Color, Size, Text, Detail, and Tooltip. A yellow circle highlights the 'Color' button, and another yellow circle highlights the 'SUM(Population)' button at the bottom. A large yellow arrow points from the left view to the right view. In the right view, the 'Color' button is highlighted with a yellow oval. The data table shows population counts for continents, with Asia being colored red.

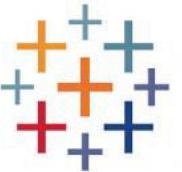
Continent	Population Number
Asia	3,705,025,700
Africa	784,475,000
Europe	730,074,600
North America	482,993,000
South America	345,780,000
Oceania	30,401,150
Antarctica	0

Hold CTRL or command key and click to drag drop into Color.

Continent	Population Number
Asia	3,705,025,700
Africa	784,475,000
Europe	730,074,600
North America	482,993,000
South America	345,780,000
Oceania	30,401,150
Antarctica	0

Highlight mark is displayed

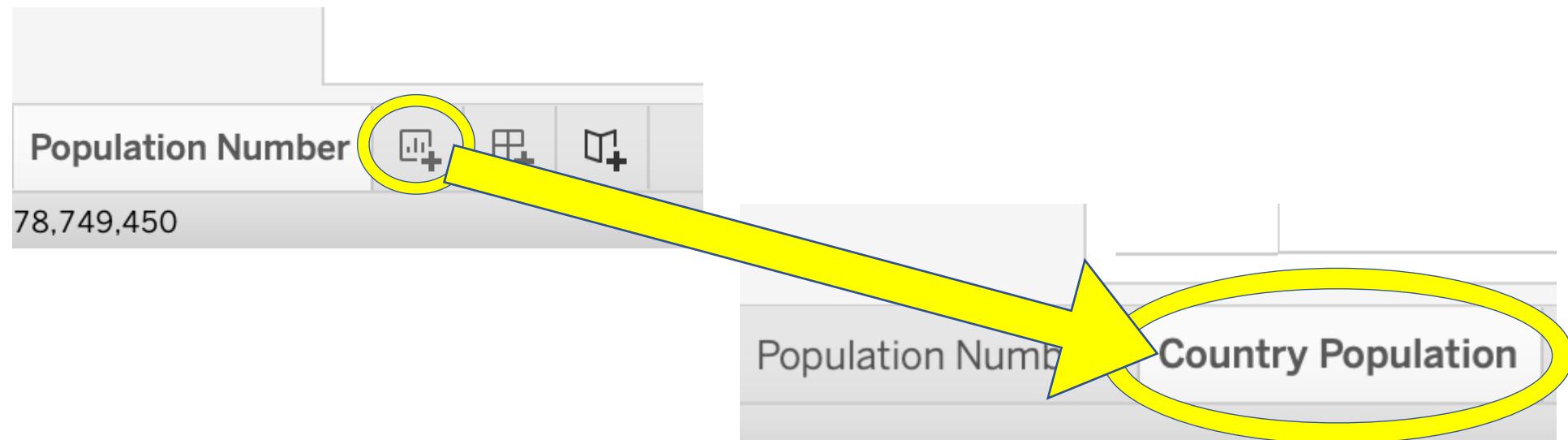
This will display when drag drop is completed



tableau

Tableau Lab (Let's Try this)

- Twelve, try to click **New Worksheet** icon to and rename to **Country Population**



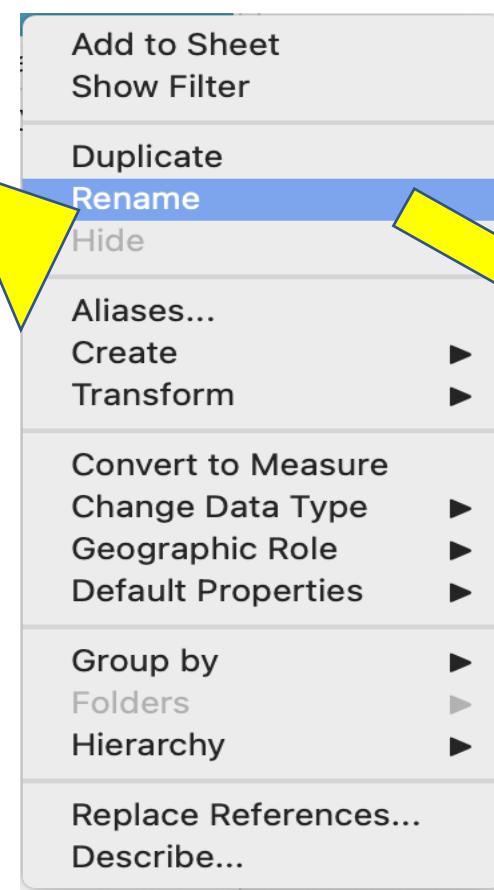


+tableau

Tableau Lab (Let's Try this)

- Thirteen, try to rename one discrete data from **Name** to **Country**

Data Analytics <
Country
Search
Tables
Abc Code
Abc Continent
Abc **Name** (highlighted)
Abc Measure Names
Life Expectancy
Population
Country.csv (Count)
Measure Values



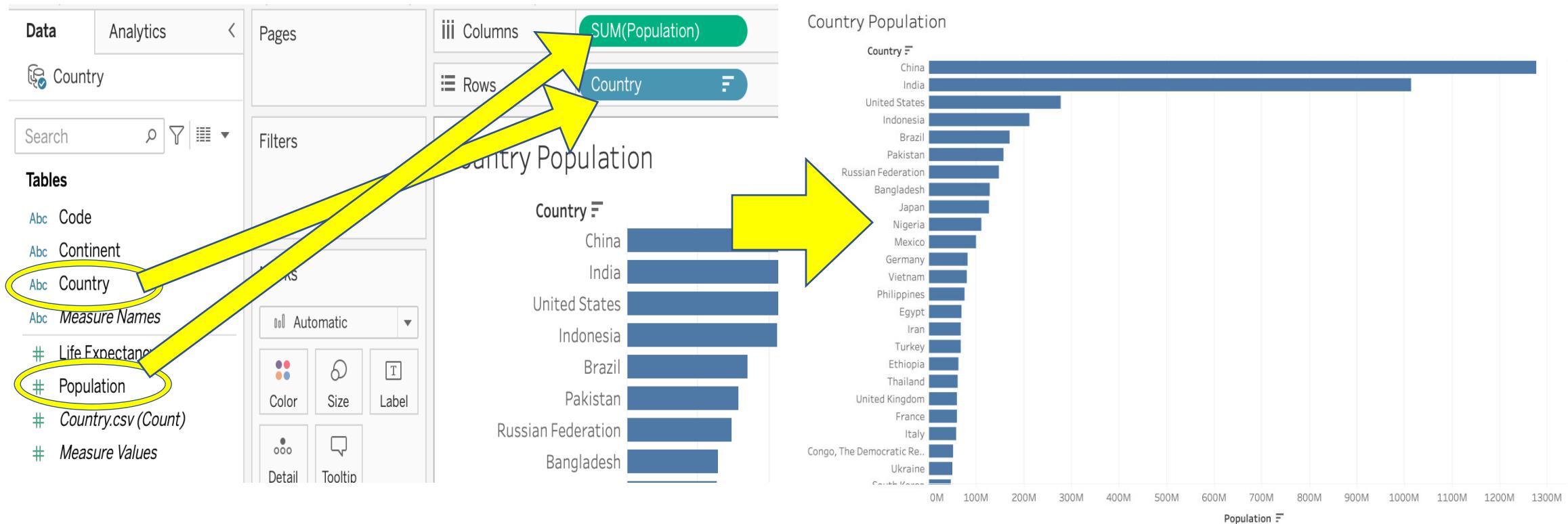
Data Analytics <
Country
Search
Tables
Code
Continent
Country (highlighted)
Measure Names
Life Expectancy
Population
Country.csv (Count)
Measure Values

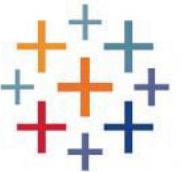


+a**b**|eau

Tableau Lab (Let's Try this)

- Fourteen, try to drag drop **Country** into **Rows** and **Population** into **Columns** and **Sorted descending**. The result will look like this





+a**b**|eau

Tableau Lab (Let's Try this)

- Fifteen, from step 14, you will see that there are so many countries. It is difficult to read. You must filter some significant countries (e.g., Top 10 Population ASEAN Countries)

The screenshot shows the Tableau interface with the 'Data' tab selected. In the 'Tables' section, the 'Country' table is highlighted with a blue selection bar. A large yellow arrow points from this selection bar to a callout window titled 'Filter [Country]'. The callout window displays a list of countries under the 'Custom value list' tab. The list includes Singapore, Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam. The 'OK' button at the bottom right of the callout is also highlighted with a yellow arrow.

Filter [Country]

General Wildcard Condition Top

Select from list Custom value list Use all

Sing
Singapore

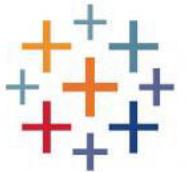
Brunei
Cambodia
Indonesia
Laos
Malaysia
Myanmar
Philippines
Singapore
Thailand
Vietnam

Clear List Include all values when empty Exclude

Summary

Field: [Country]
Selection: Selected 10 values
Wildcard: All
Condition: None
Limit: None

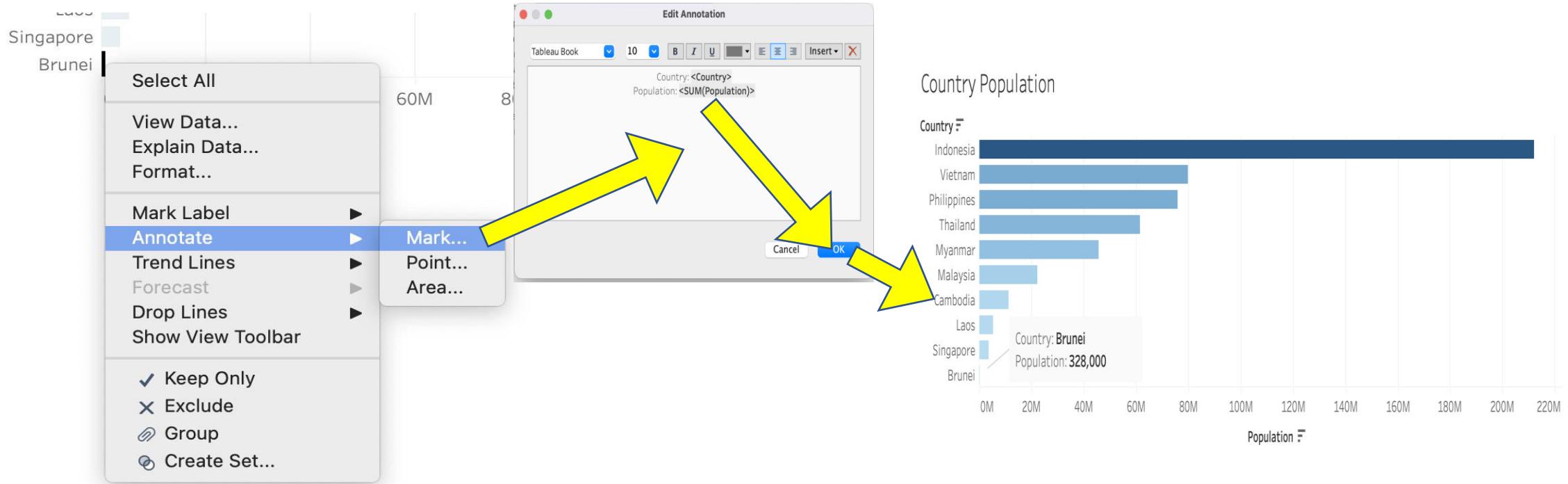
Reset Apply Cancel OK



+a**b**|eau

Tableau Lab (Let's Try this)

- Sixteen, try to highlight according to step 6 and right click to Brunei's chart, select **Annotate** → **Mark ...**, and click OK. The result will look like this

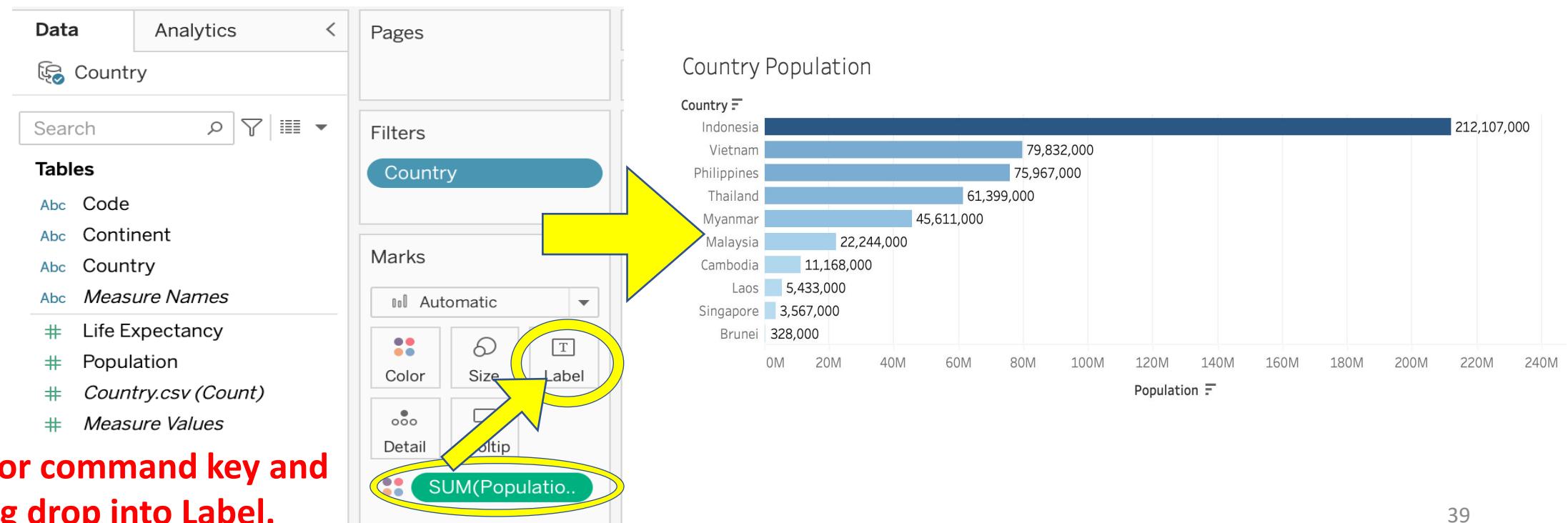


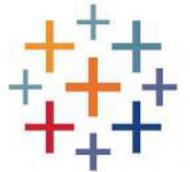


+a**b**|eau

Tableau Lab (Let's Try this)

- Seventeen, (**Another Method**) try to hold **CTRL** or **command** key and click to drag drop **Population** into **Label** (Remove mark). The result will display like this

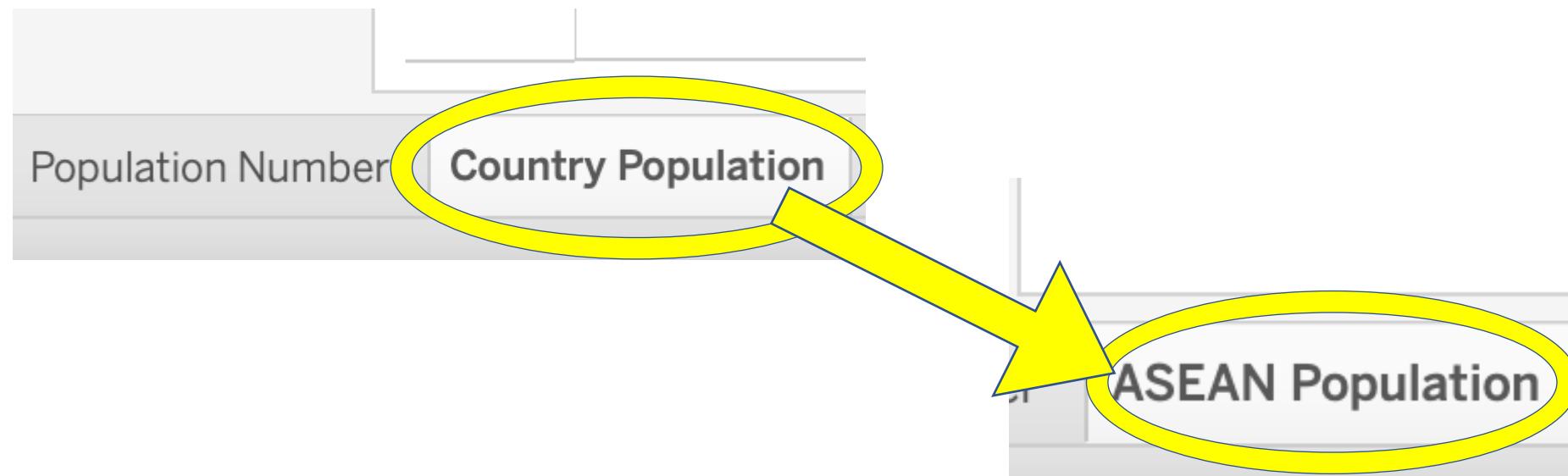


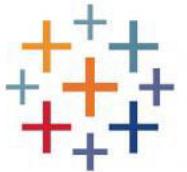


+tableau

Tableau Lab (Let's Try this)

- Eighteen, try to rename **Country Population** sheet name again to **ASEAN Population**





+a**b**|eau

Tableau Lab (Let's Try this)

- Nineteen, try to create new sheet and then change name into **World Population**. This way we would like to display it in **packed bubble chart**

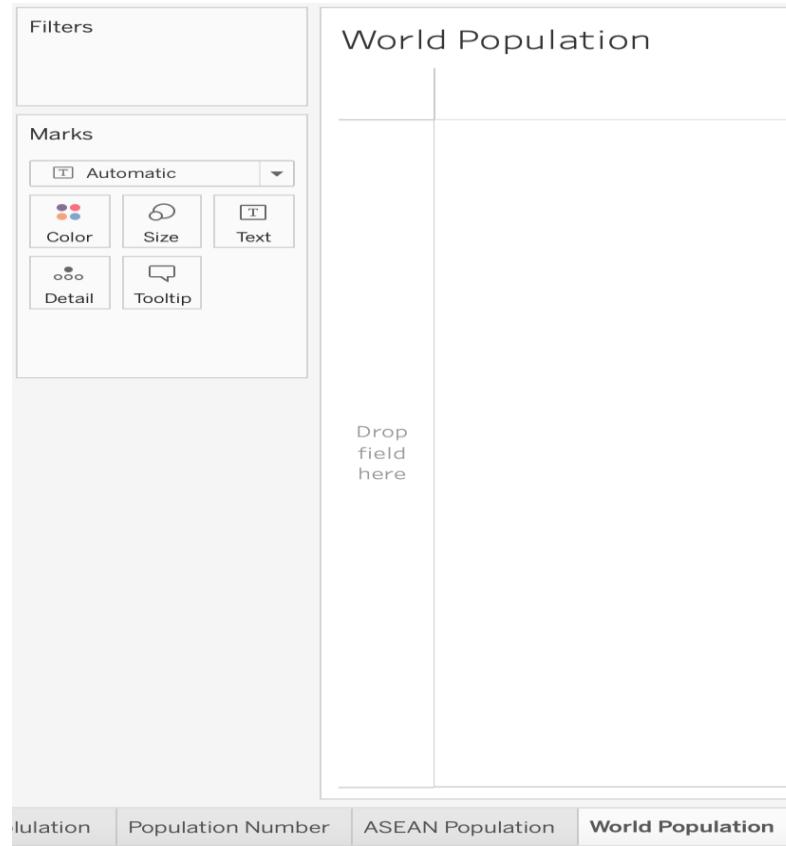


Tableau Lab (Let's Try this)



+tableau

- Twenty, try to drag drop **Country** into Rows and **Population** into **Columns** again, and choose **packed bubbles chart**

The screenshot shows a data visualization interface with several panels:

- Data Panel:** Contains a search bar, a filter icon, and a dropdown menu. Below it, a list of tables includes "Country" (circled in yellow), "Code", "Continent", "Measure Names", "Life Expectancy", "Population" (circled in yellow), "Country.csv (Count)", and "Measure Values".
- Analytics Panel:** Shows "Pages" and "Filters" sections.
- Chart Area:** A bar chart titled "Country Population" displays the population of various countries. The y-axis is labeled "Country" and has a dropdown arrow pointing to the chart. The x-axis represents population values. The bars are blue, and the chart area has a light gray background.
- Legend/Color Panel:** Shows a color palette with a gradient from light blue to dark blue, labeled "Country". Below it, a legend indicates "Automatic" for the color mapping.
- Formatting Panel:** Includes options for "Color", "Size", "Label", "Detail", and "Tooltip".

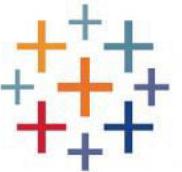
Yellow arrows highlight the "Country" table in the Data panel and the "Population" measure in the Analytics panel. The chart itself is also highlighted with yellow arrows.

Country	Population
China	1400
India	1300
United States	300
Indonesia	250
Brazil	200
Pakistan	180
Russian Federation	150
Bangladesh	120

For packed bubbles try

1 or more Dimensions

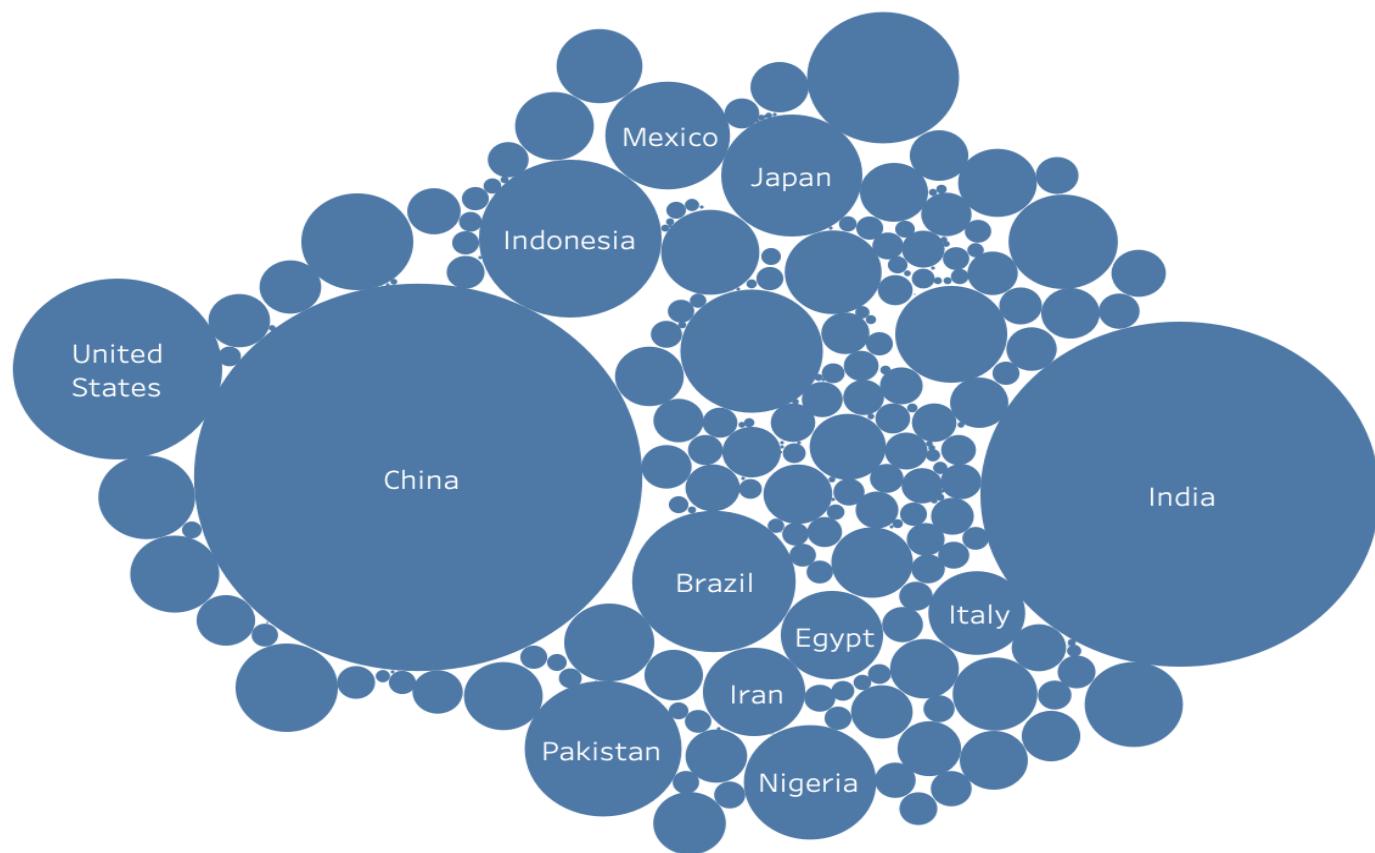
1 or 2 Measures

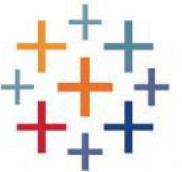


+a**b|e|a**u

Tableau Lab (Let's Try this)

- Twenty-one, when you click **packed bubble chart**, the result will look like this



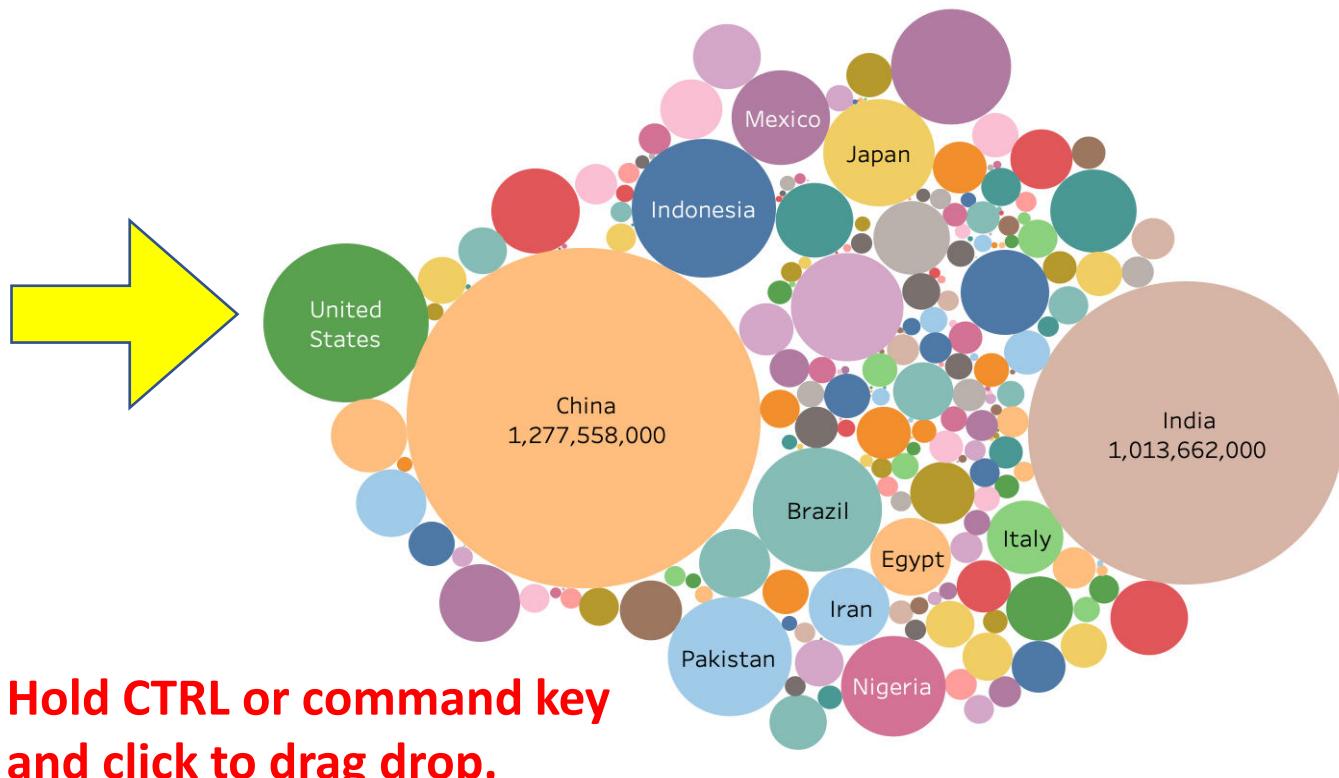


+tableau

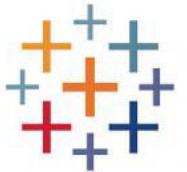
Tableau Lab (Let's Try this)

- Twenty-two, remind step 11 and 17 to hold CTRL or command key and drag **Country** into **Color** and **Population** into **Label**

The screenshot shows the Tableau interface with the 'Marks' card open. The 'Color' button and the 'Label' button are highlighted with yellow circles and arrows pointing to them from the text instructions below. Other buttons like 'Size', 'Tooltip', and 'Detail' are also visible.



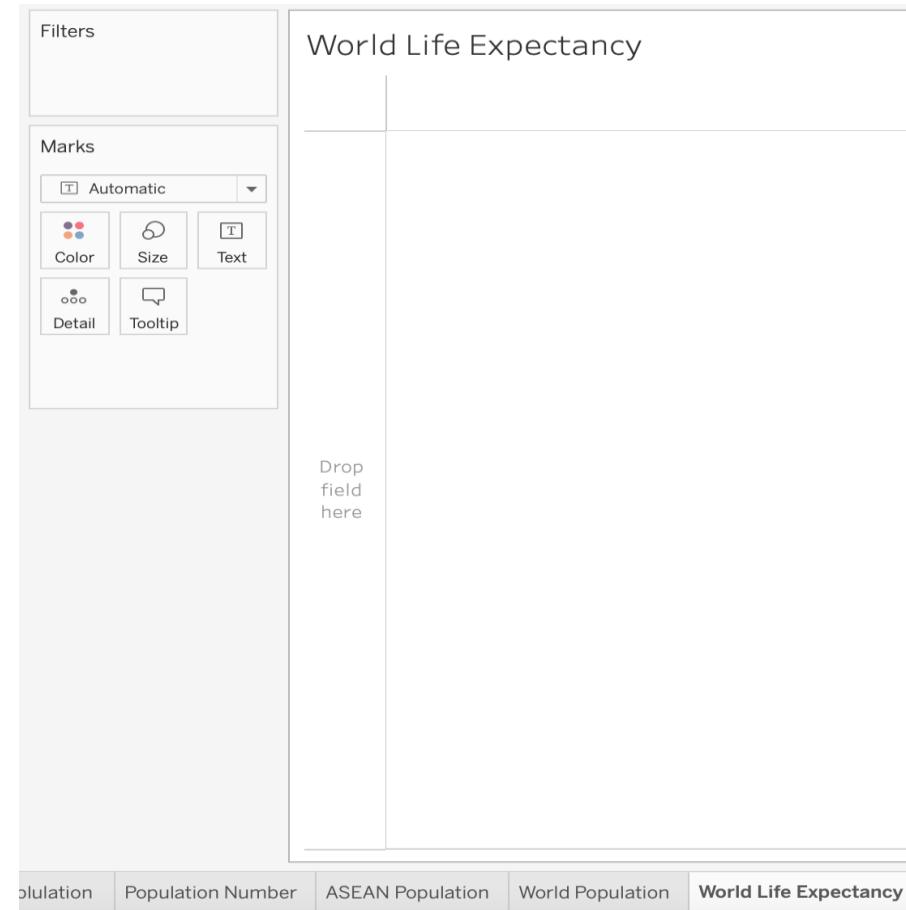
Hold CTRL or command key
and click to drag drop.



+a**b**|eau

Tableau Lab (Let's Try this)

- Twenty-three, try to create new sheet and then change name into **World Life Expectancy**.

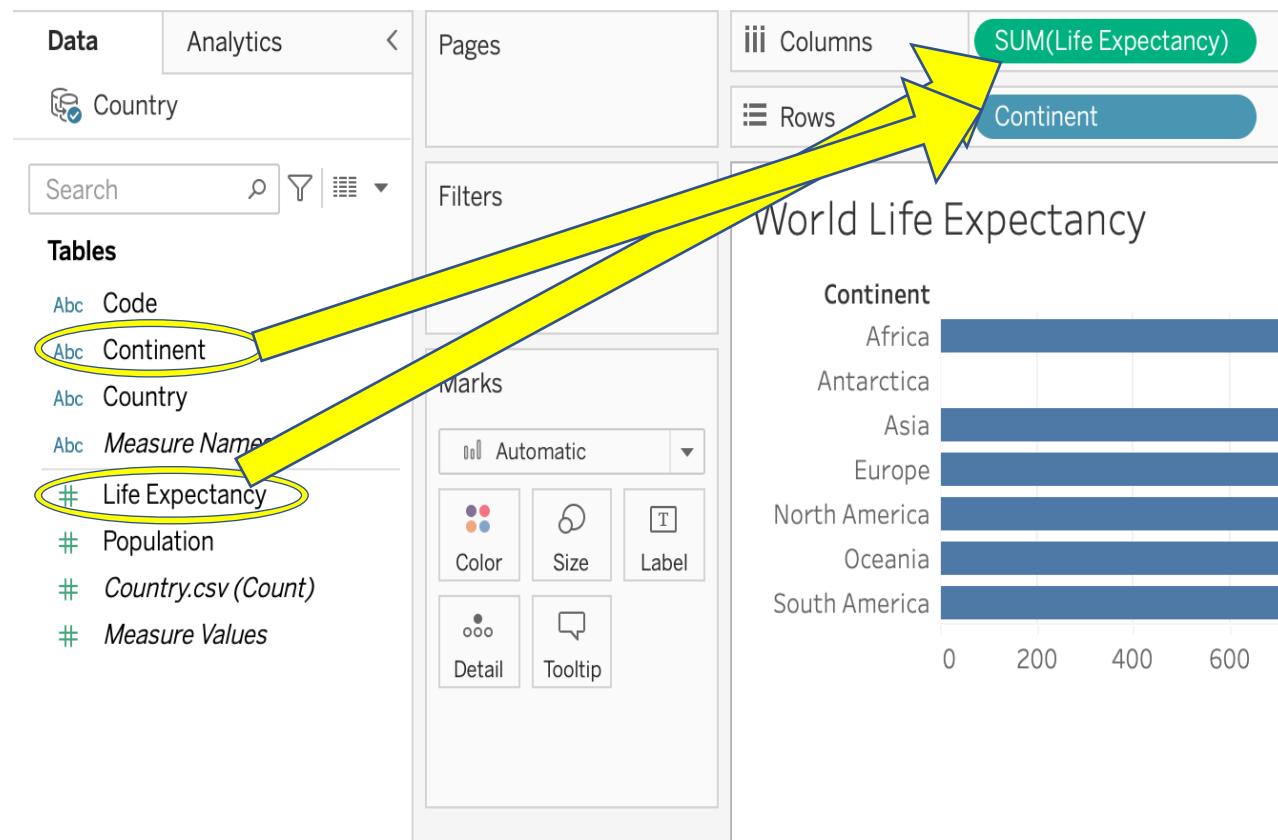




+a**b**|eau

Tableau Lab (Let's Try this)

- Twenty-four, try to drag drop **Continent** into Rows and **Life Expectancy** into **Columns**



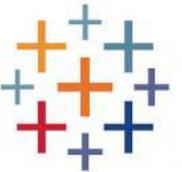


+a**b**eau

Tableau Lab (Let's Try this)

- Twenty-five, try to change the calculation of Life Expectancy from **SUM (Summation)** to **AVG (Average)**. It is impossible to display life age in summation

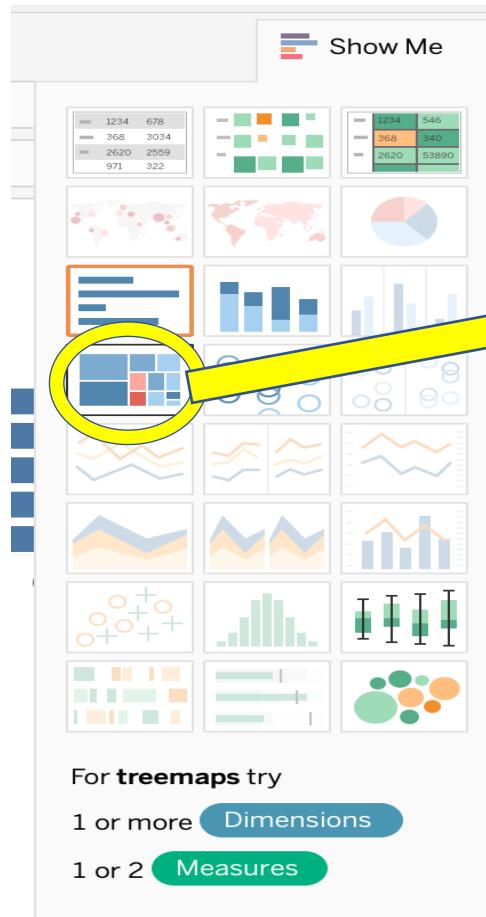
The screenshot shows a Tableau interface with a bar chart titled "World Life Expectancy" on the left. The chart displays life expectancy values for continents: Africa (~54), Antarctica (~78), Asia (~70), Europe (~75), North America (~78), Oceania (~78), and South America (~70). A context menu is open over the "Life Expectancy" measure, specifically the dropdown menu for "Measure (Sum)". The menu includes options like "Filter...", "Show Filter", "Apply to Worksheets", "Format...", "Show Header" (checked), "Include in Tooltip" (checked), "Dimension", "Attribute", "Measure (Sum)" (selected), "Discrete", "Continuous", "Edit in Shelf", "Add Table Calculation...", "Quick Table Calculation", and "Remove". A yellow arrow points from the "Average" option in the "Measure (Sum)" submenu to a green box labeled "AVG(Life Expectancy)", indicating the target for changing the aggregation method.

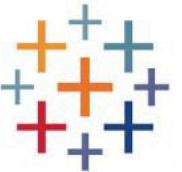


+a**b**|eau

Tableau Lab (Let's Try this)

- Twenty-six, try to change the chart into **treemaps**. It displays look like this





+a|bleau

Tableau Lab (Let's Try this)

Children: 0 – 13 years
Teenager: 14 – 24 years
Working Age: 25 – 39 years
Middle Age: 40 – 59 years
Senior: 60 – 100 years

- Twenty-seven, let's create the new calculation with syntax condition named **Life Age Condition** in Analysis → Create Calculated Field..., input this syntax condition, and then click OK

The screenshot shows the Tableau interface with the 'Analysis' menu open. The 'Create Calculated Field...' option is highlighted with a yellow oval. The main pane displays a calculated field named 'Life Age Condition' with the following syntax:

```
IF ([Life Expectancy] >= 60 AND [Life Expectancy] <= 100.99) THEN
    "Senior"
ELSEIF ([Life Expectancy] >= 40 AND [Life Expectancy] <= 59.99) THEN
    "Middle Age"
ELSEIF ([Life Expectancy] >= 25 AND [Life Expectancy] <= 39.99) THEN
    "Working Age"
ELSEIF ([Life Expectancy] >= 14 AND [Life Expectancy] <= 24.99) THEN
    "Teenager"
ELSEIF ([Life Expectancy] >= 0 AND [Life Expectancy] <= 13.99) THEN
    "Children"
ELSE
    "N/A"
END
```

A yellow arrow points from the 'Create Calculated Field...' menu option to the calculated field definition. Another yellow arrow points from the 'OK' button in the bottom right corner of the dialog to the 'Measure Names' section in the sidebar.

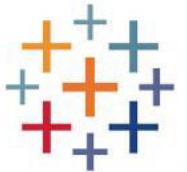
Tables

- Abc Code
- Abc Continent
- Abc Country
- =Abc Life Age Condition

Measure Names

- # Life Expectancy
- # Population
- # Country.csv (Count)
- # Measure Values

This condition syntax is to display the life age in category: Children, Teenager, Working Age, Middle Age, Senior



+a**bleau**

Tableau Lab (Let's Try this)

- Twenty-eight, try to drag drop **Life Age Condition** into **Marks** area and see the result

Data Analytics <

Country

Search

Tables

- Abc Code
- Abc Continent
- Abc Country
- =Abc Life Age Condition
- Abc Measure Names
- # Life Expectancy
- # Population
- # Country.csv (Count)
- # Measure Values

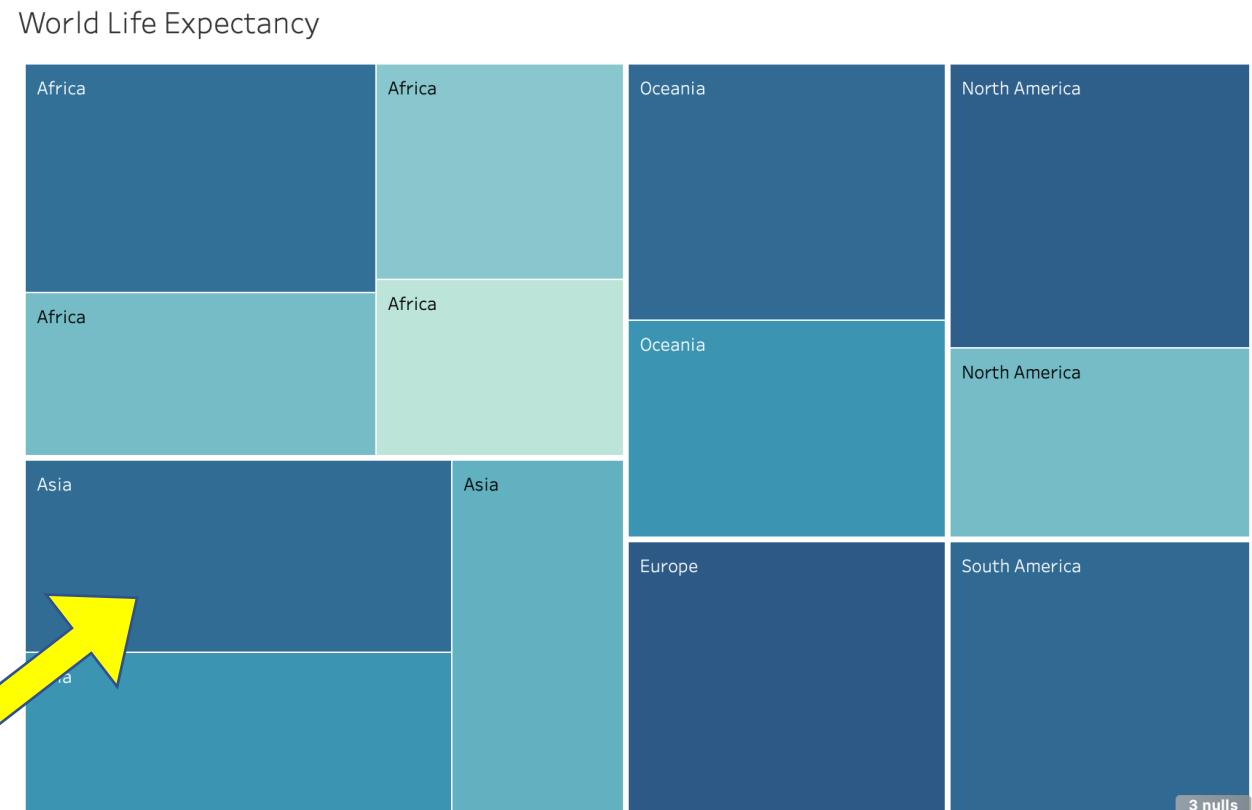
Pages

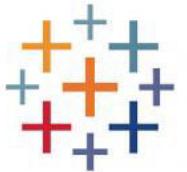
Filters

Marks

- Automatic
- Color
- Size
- Label
- Detail
- Tooltip

AVG(Life Expe...
Continent
Life Age Cond...





+tableau

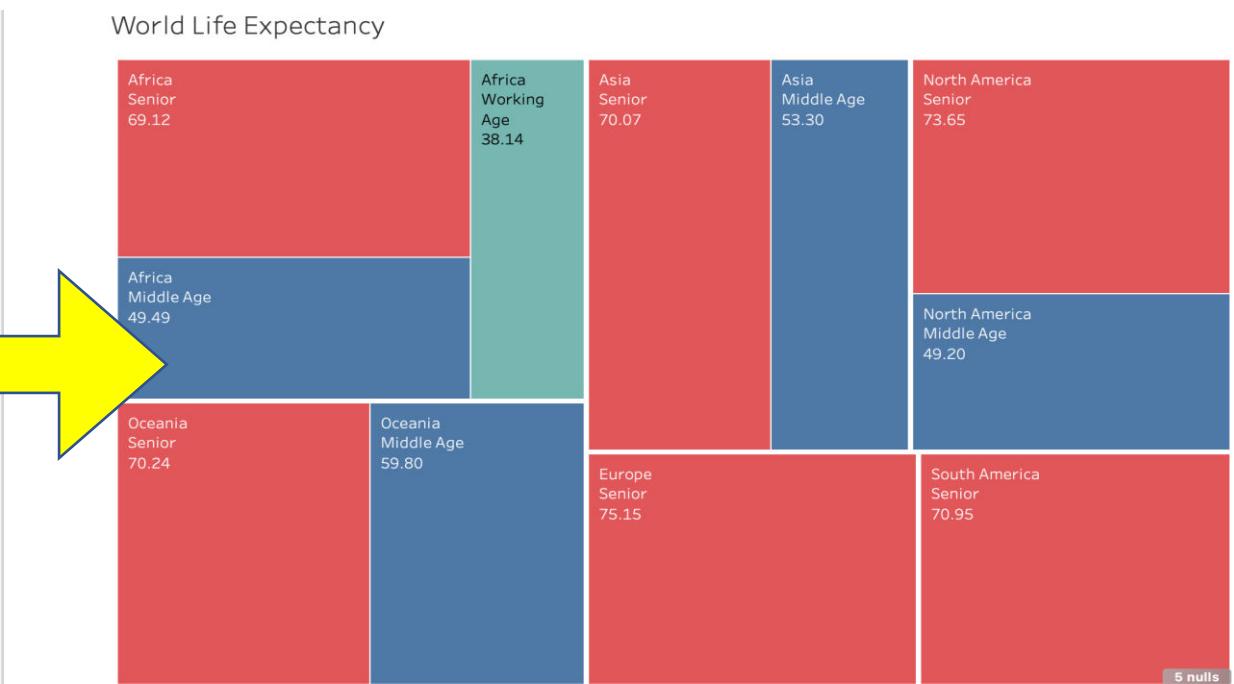
Tableau Lab (Let's Try this)

- Twenty-nine, try to hold **CTRL** or **command** key and click to drag drop **Life Age Condition** into **Color** and **Label**, and **Life Expectancy** into **Label**. The result will display like this

Marks

- Automatic
- Color
- Size
- Label
- Detail
- Tooltip

AVG(Life Expe..)
Life Age Co..
Continent
Life Age Co..
AVG(Life Expe..)
Life Age Condit..



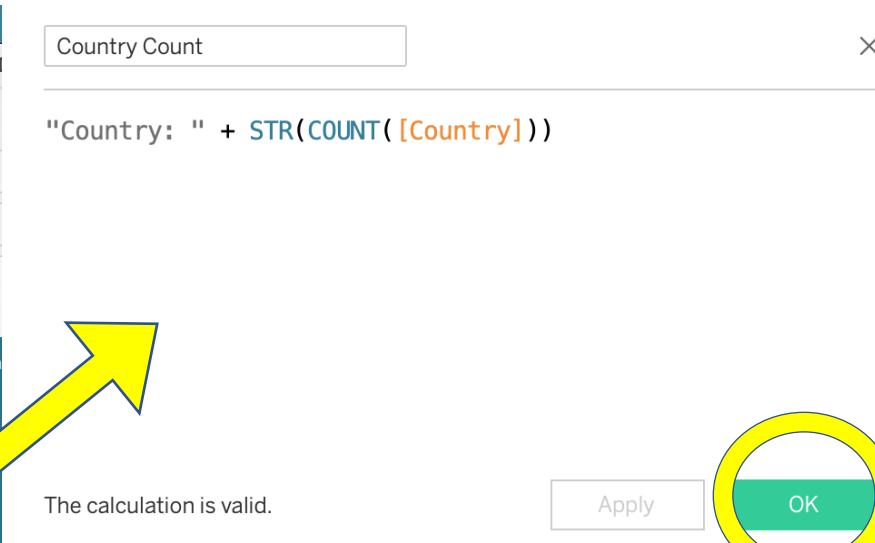
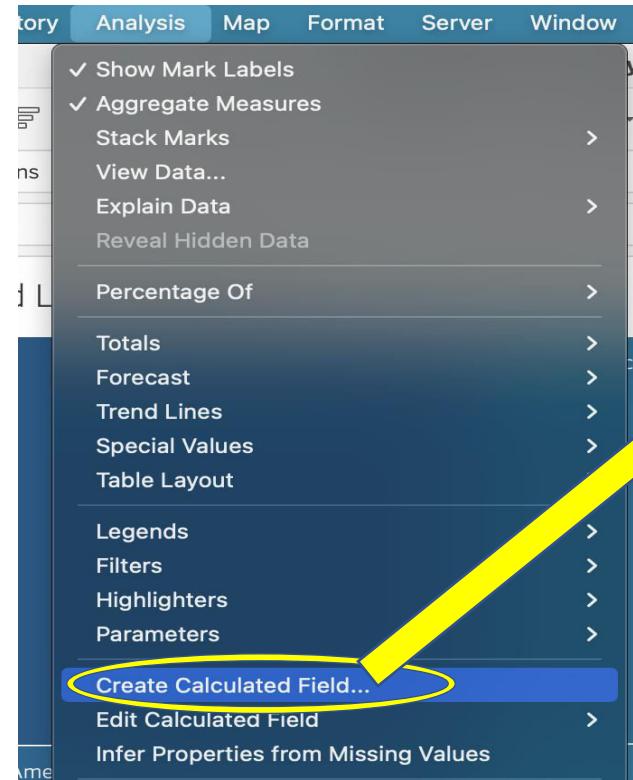
Hold **CTRL** or **command** key and click to drag drop. 51



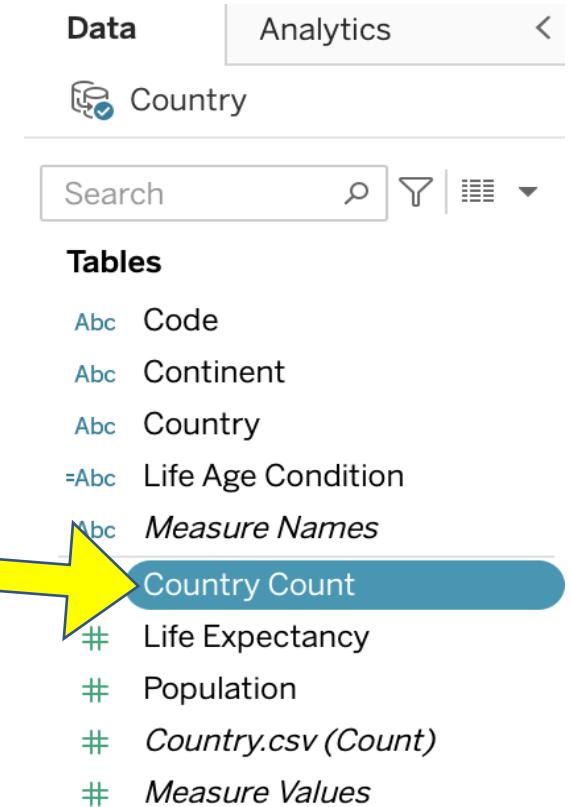
+a**b**eau

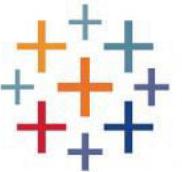
Tableau Lab (Let's Try this)

- Thirty, let's create the new calculation again with syntax named **Country Count** in Analysis → Create Calculated Field..., input this syntax condition, and then click OK



This syntax is to display number of country in each continent.





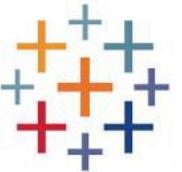
+a**b**eau

Tableau Lab (Let's Try this)

- Thirty-one, try to drag drop **Country Count** into **Marks** area and hold **CTRL** or **command** key and click to drag drop **Country Count** into **Label**

The screenshot shows the Tableau interface with the following components:

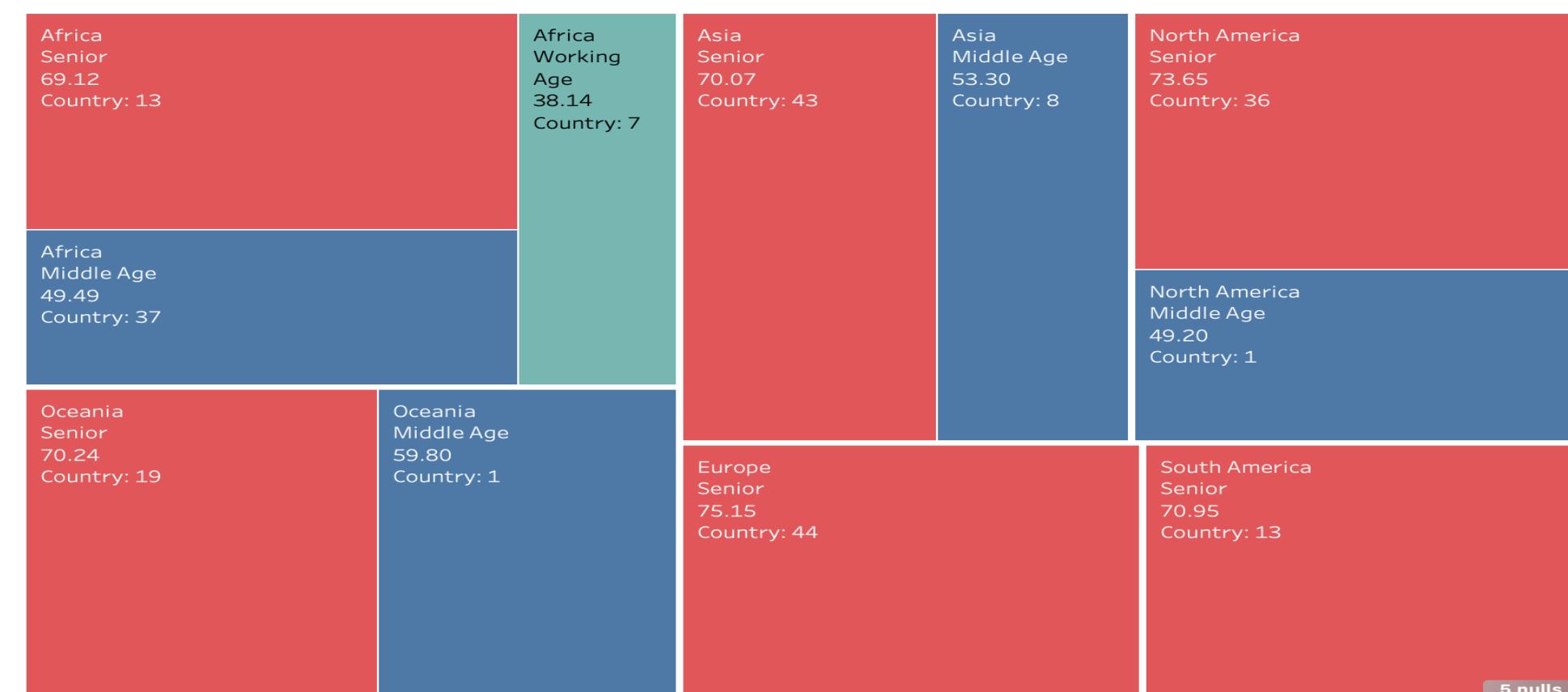
- Data Pane:** Shows a list of tables and measures. **Country Count** is highlighted with a blue bar.
- Analytics Pane:** Shows a search bar and filter options.
- Marks Shelf:** Shows the **Marks** section with buttons for Color, Size, Label, Detail, and Tooltip. The **Label** button is circled in yellow.
- Marks List:** A list of items including:
 - AVG(Life Expe..)
 - Life Age Condit..
 - Continent
 - Life Age Condit..
 - AVG(Life Expe..)
 - AGG(Country ..)
 - Life Age Condit..
 - AGG(Country ..)



+tableau

Tableau Lab (Let's Try this)

- Thirty-two, when you complete from step 31, the result will look like this

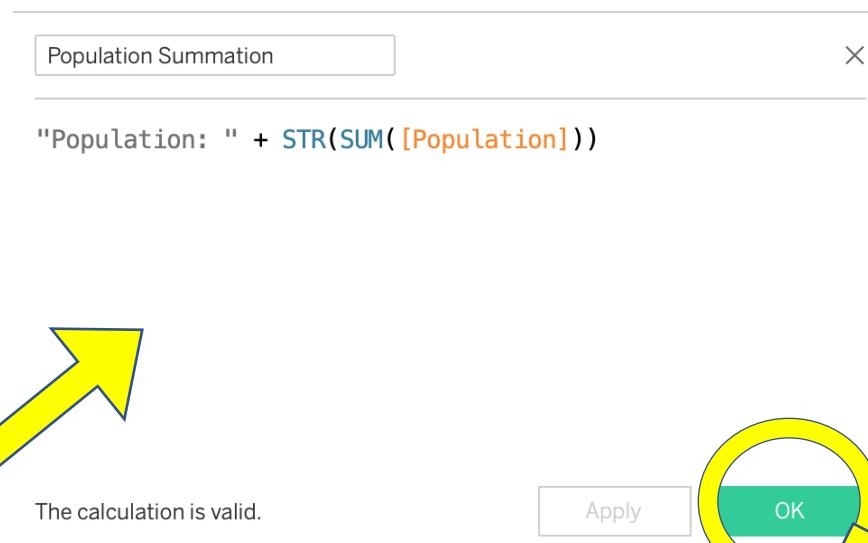
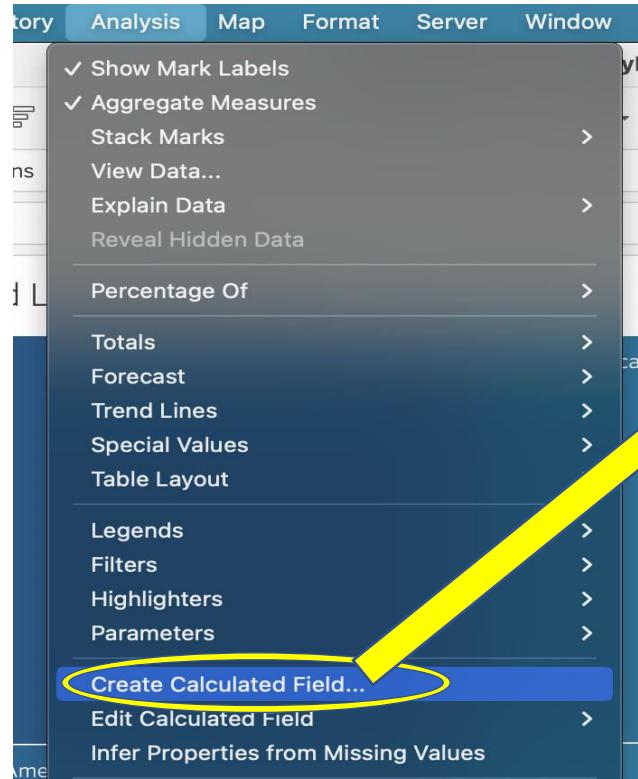




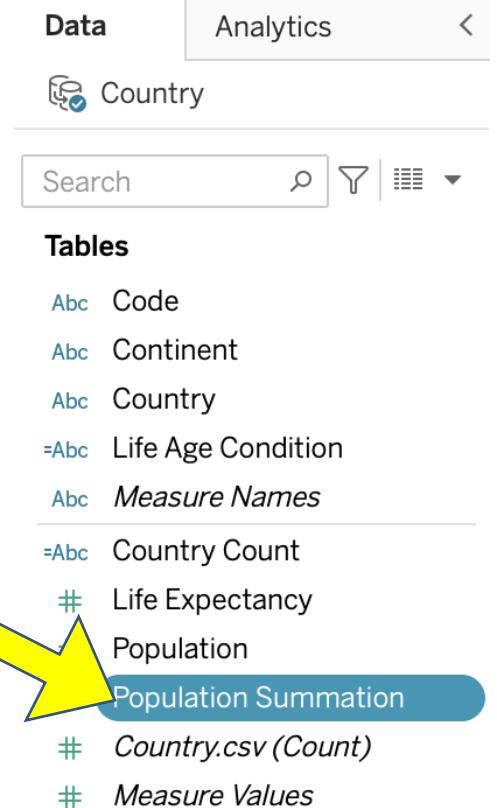
+a**b**eau

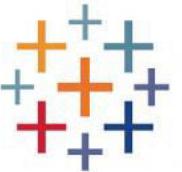
Tableau Lab (Let's Try this)

- Thirty-three, let's create the new calculation again with syntax named **Population Summation** in **Analysis** → **Create Calculated Field...**, input this syntax condition, and then click OK



This syntax is to display summation of population in each continent.





+a**b**eau

Tableau Lab (Let's Try this)

- Thirty-four, try to drag drop **Population Summation** into **Marks** area and hold **CTRL** or **command** key and click to drag drop **Population Summation** into **Label**

The screenshot shows the Tableau interface with the following components:

- Data Pane:** Shows a list of tables and measures. The 'Population Summation' measure is highlighted with a blue oval.
- Analytics Pane:** Shows a search bar and filter controls.
- Marks Shelf:** Shows the 'Marks' section with four buttons: Color, Size, Label, Detail, and Tooltip. The 'Label' button is highlighted with a yellow circle.
- Marks List:** Shows a list of marks, each represented by a colored button. The first mark is 'AVG(Life Expe..)' (green). Other visible marks include 'Life Age Condit...', 'Continent', 'Life Age Condit...', 'AVG(Life Expe..)', 'AGG(Country ..)', 'AGG(Populatio...', 'Life Age Condit...', 'AGG(Country ..)', and 'AGG(Populatio...'.

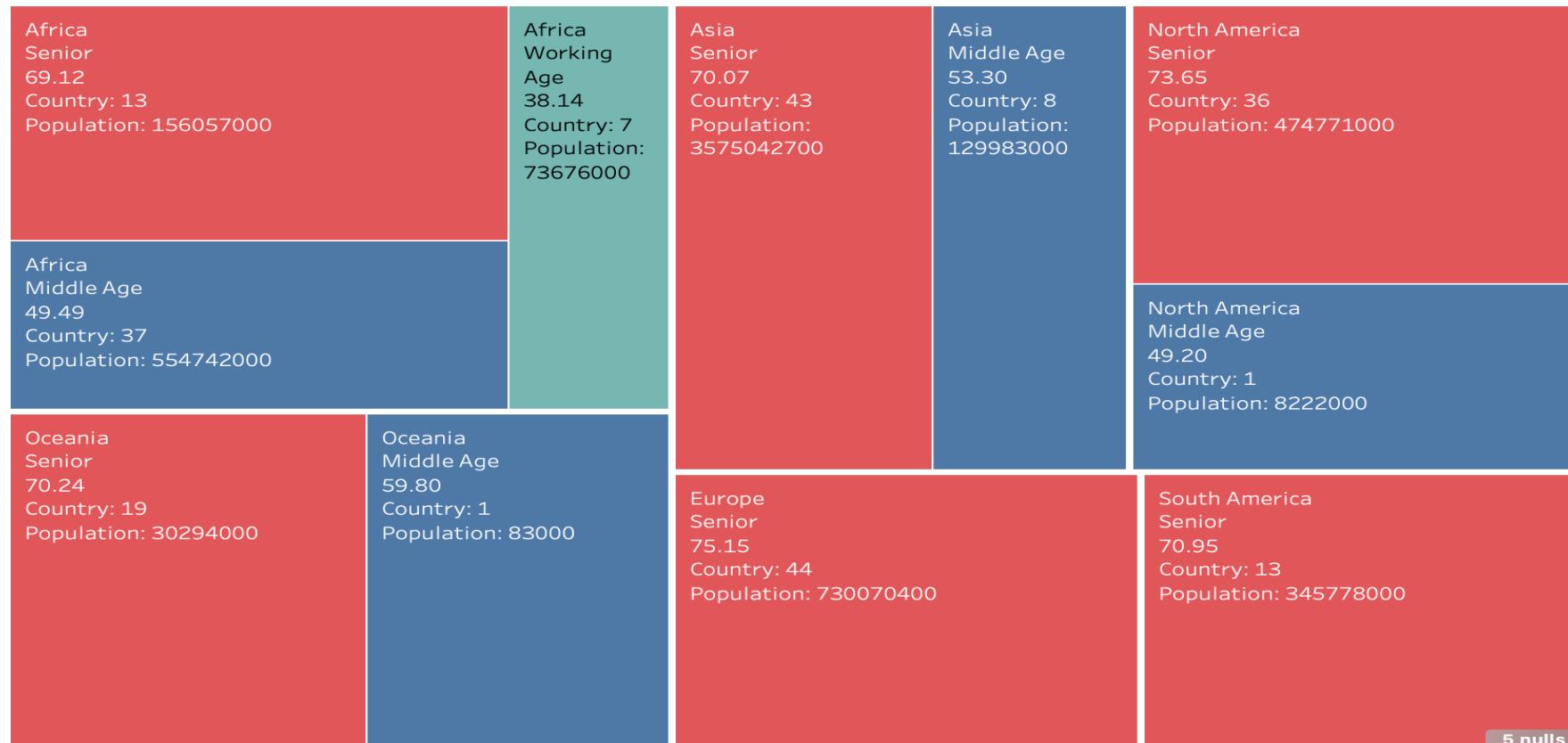


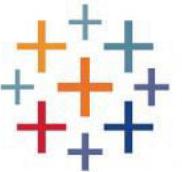
+tableau

Tableau Lab (Let's Try this)

- Thirty-five, when you complete from step 34, the result will look like this

World Life Expectancy





+a|bleau

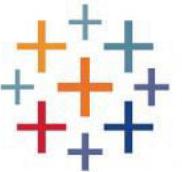
Tableau Lab (Let's Try this)

- Thirty-six, try to create new dashboard and then change name into **World DataViz**

The screenshot shows the Tableau Dashboard editor. On the left, there is a sidebar with various settings:

- Dashboard**: Shows "Default" and "Phone" options, with "Device Preview" button.
- Size**: Set to "min 420x560 - max 650x8...".
- Sheets**: A list of five sheets: "Continent ...", "Population Number", "ASEAN Population", "World Population", and "World Life ...".
- Objects**: A list of objects with icons: Horizontal, Vertical, Text, Image, Web Page, Data Story, Blank, Navigation, Download, Extension, Ask Data.
- Layout**: Buttons for "Tiled" (selected) and "Floating".
- Show dashboard title.

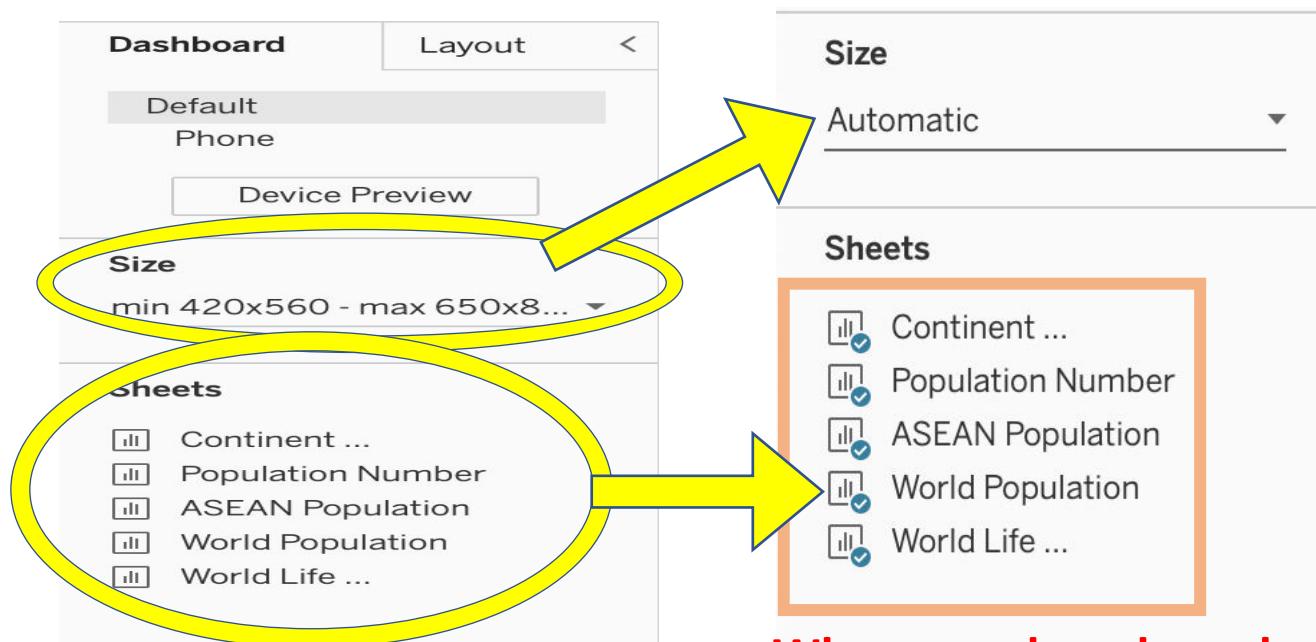
In the center, there is a large white area with the placeholder text "Drop sheets here". At the bottom of the dashboard editor, there is a navigation bar with tabs: Data Source, Continent Population, Population Number, ASEAN Population, World Population, World Life Expectancy, and a tab labeled "World DataViz" which is currently active.



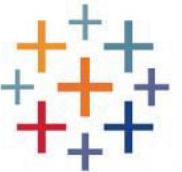
+a**bleau**

Tableau Lab (Let's Try this)

- Thirty-seven, try to change size to **Automatic** and drag drop and adjust whole **Sheets** into your favorite style



When you drag drop sheets into dashboard, the status displays like this

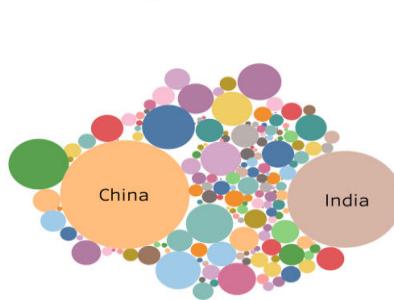


+tableau

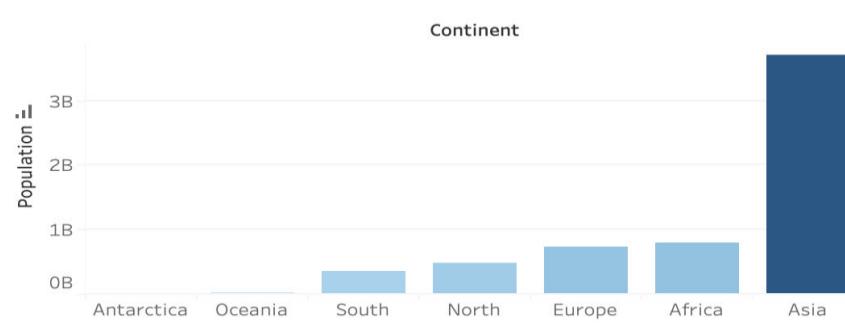
Tableau Lab (Let's Try this)

- Thirty-eight, when you complete from step 37, the result will look like this

World Population



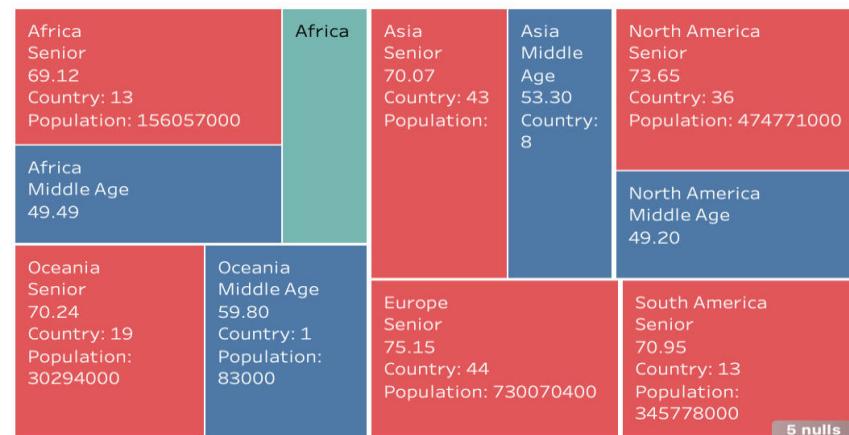
Continent Population



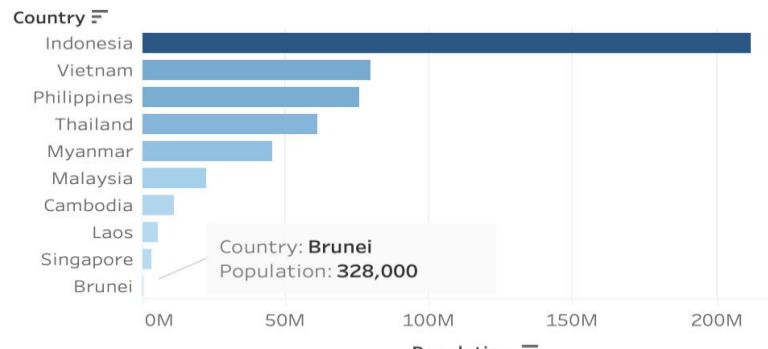
Population Number



World Life Expectancy



ASEAN Population



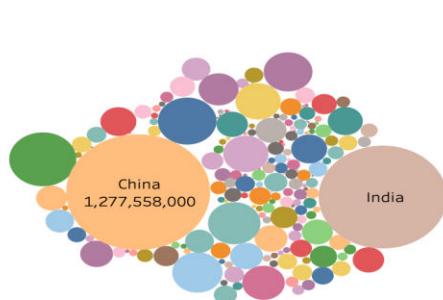


+tableau

Tableau Lab (Let's Try this)

- Thirty-nine, congratulations! You can click  icon to display Presentation

World Population



Continent Population



Population Number

Continent	Population
Asia	3,705,025,700
Africa	784,475,000
Europe	730,074,600
North America	482,993,000
South America	345,780,000
Oceania	30,401,150
Antarctica	0

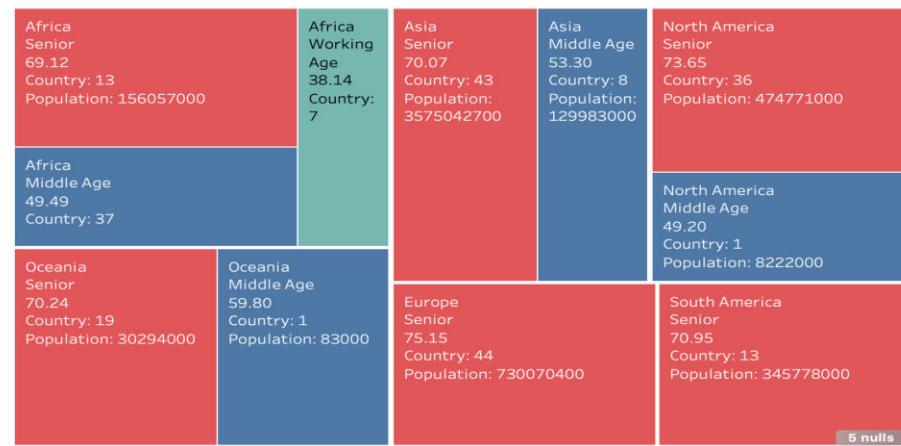
Population

0  4B

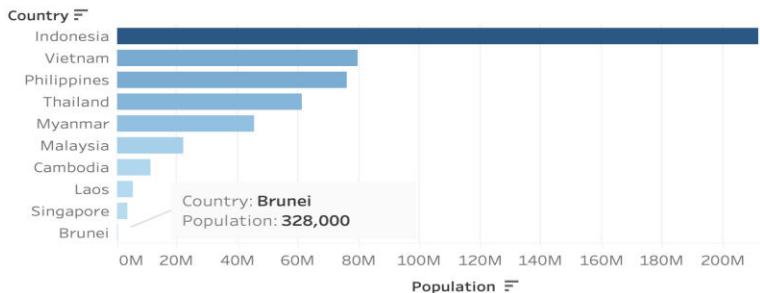
Country

Afghanistan
Albania
Algeria
American Sam..
Andorra
Angola
Anguilla
Antarctica
Antigua and Ba..
Argentina
Armenia
Aruba
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Barbados
Belarus
Belgium
Belize
Benin
Bermuda
Bhutan
Bolivia
Bosnia and Her..
Botswana
Bouvet Island
Brazil
British Indian ..
Brunei
Bulgaria
Life Age Condition
Middle Age
N/A
Senior
Working Age

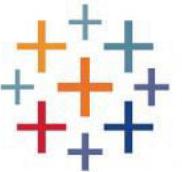
World Life Expectancy



ASEAN Population



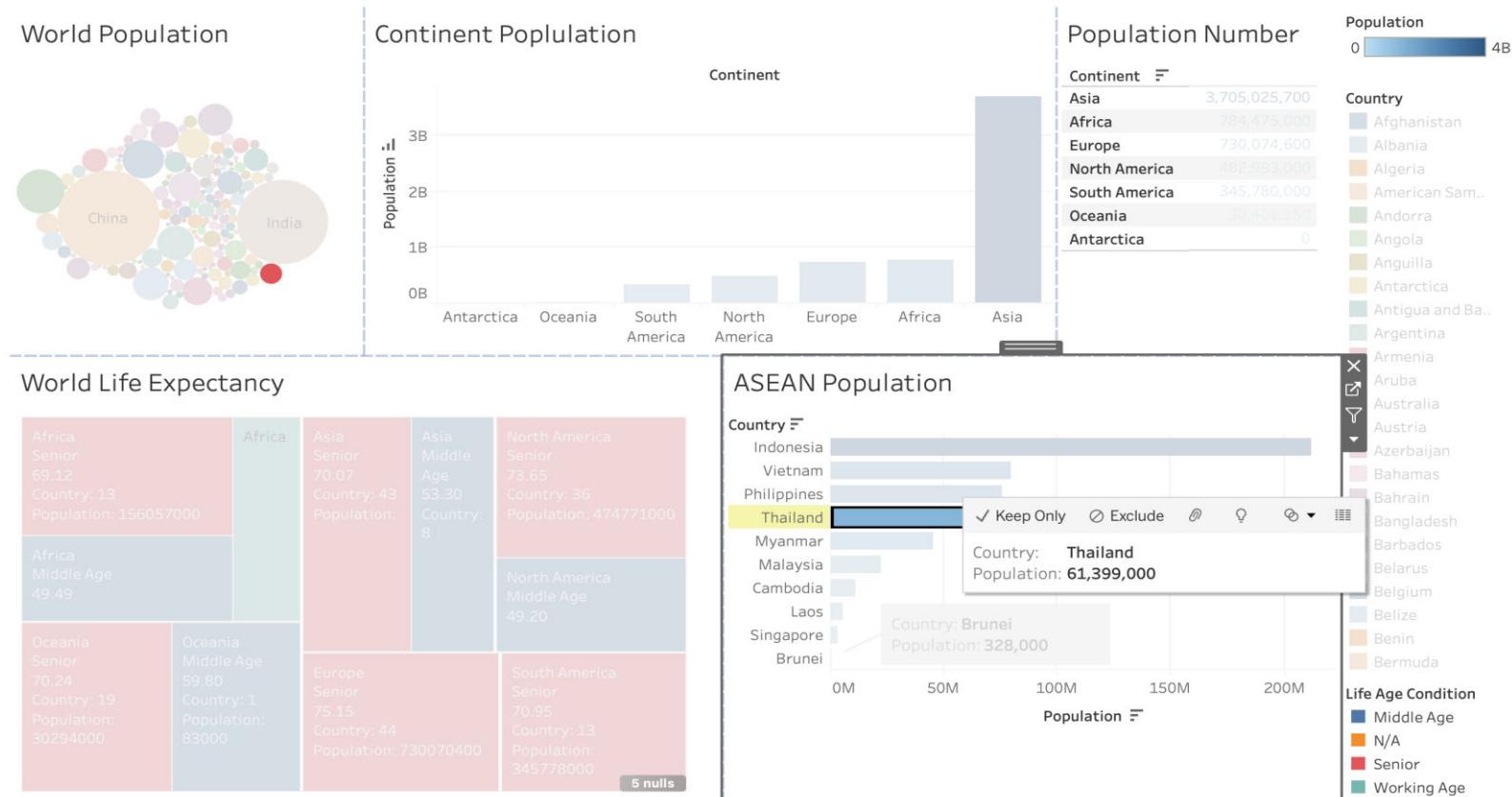
Press ESC key to exit
Presentation Mode



+tableau

Tableau Lab (Let's Try this)

- Forty, try to click chart in some countries in ASEAN like Thailand. The result will focus only country with information you select



Try to move or
click your mouse
cursor to other
areas and see
the result!

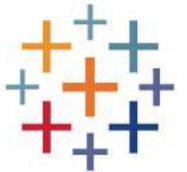


Tableau Lab (Let's Try this)

- Forty-one, do not forget to save file by click **File → Save to Tableau Public**. This will let you save into **Tableau Public Cloud** directly

The screenshot illustrates the process of saving a Tableau Public workbook. On the left, the Tableau Public interface shows the 'File' menu open, with the option 'Save to Tableau Public...' highlighted. A yellow arrow points from this menu item to a central dialog box titled 'Saving workbook to Tableau Public...'. Another yellow arrow points from this dialog box to a screenshot of a web browser window. The browser window displays a dashboard titled 'CountryDataViz' by 'Tanakom'. The dashboard includes several visualizations: a bubble chart of World Population centered on China and India; a bar chart of Continent Population showing Asia as the largest continent; a treemap of World Life Expectancy by continent and age group; and a bar chart of ASEAN Population with Indonesia at the top. The browser's address bar shows 'tableau-public', and the top navigation bar includes links for 'DISCOVER', 'BLOG', 'RESOURCES', and 'ABOUT'.

When the save file is completed, it opens your cloud save location file automatically in your web browser.

Disadvantage:
It shows your data visualization file in the public, and anyone can see and use it!



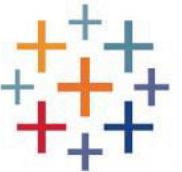
+a**bleau**

Tableau Lab (Let's Try this)

- Forty-two, from the **Tableau Public Cloud** in web browser, let's try to download the file and choose file type into your device

The image shows a composite of three screens illustrating the process of downloading a Tableau workbook.

- Tableau Public Cloud Interface:** On the left, a screenshot of the Tableau Public Cloud interface shows a "Download" button highlighted with a yellow circle. A large yellow arrow points from this button to a "Tableau Workbook" option in a dropdown menu.
- Download Dialog Box:** In the center, a "Download Tableau Workbook" dialog box is open. It contains instructions: "Select a version to download the workbook as. Keep the workbook in its current version or select a different version. If you downgrade the workbook to a previous version, some functionality may be lost." Below this is a "Version" dropdown set to "Current". At the bottom is a blue "Download" button, which is also highlighted with a yellow circle.
- File Explorer Window:** On the right, a "Downloads" folder is shown in a file explorer. The "CountryDataViz.twbx" file is selected and highlighted with a yellow circle. The file details are visible at the bottom: "CountryDataViz.twbx", "Tableau Packaged Workbook - 20 KB", and "Information".



+a**b**leau

Tableau Lab (Let's Try this)

- Forty-three, from the **Tableau Public Cloud** in web browser, let's try to share the file and copy link to paste into any collaboration software or social media or generate QR code to display in Smartphone

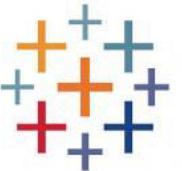
The link will look like this ...

https://public.tableau.com/views/CountryDataViz_16578783818970/WorldDataViz?:language=en-GB&:display_count=n&:origin=viz_share_link

TANAKOM TANTONTRAKUL 18:19
https://public.tableau.com/views/CountryDataViz_16578783818970/WorldDataViz?:language=en-GB&:display_count=n&:origin=viz_share_link

OR





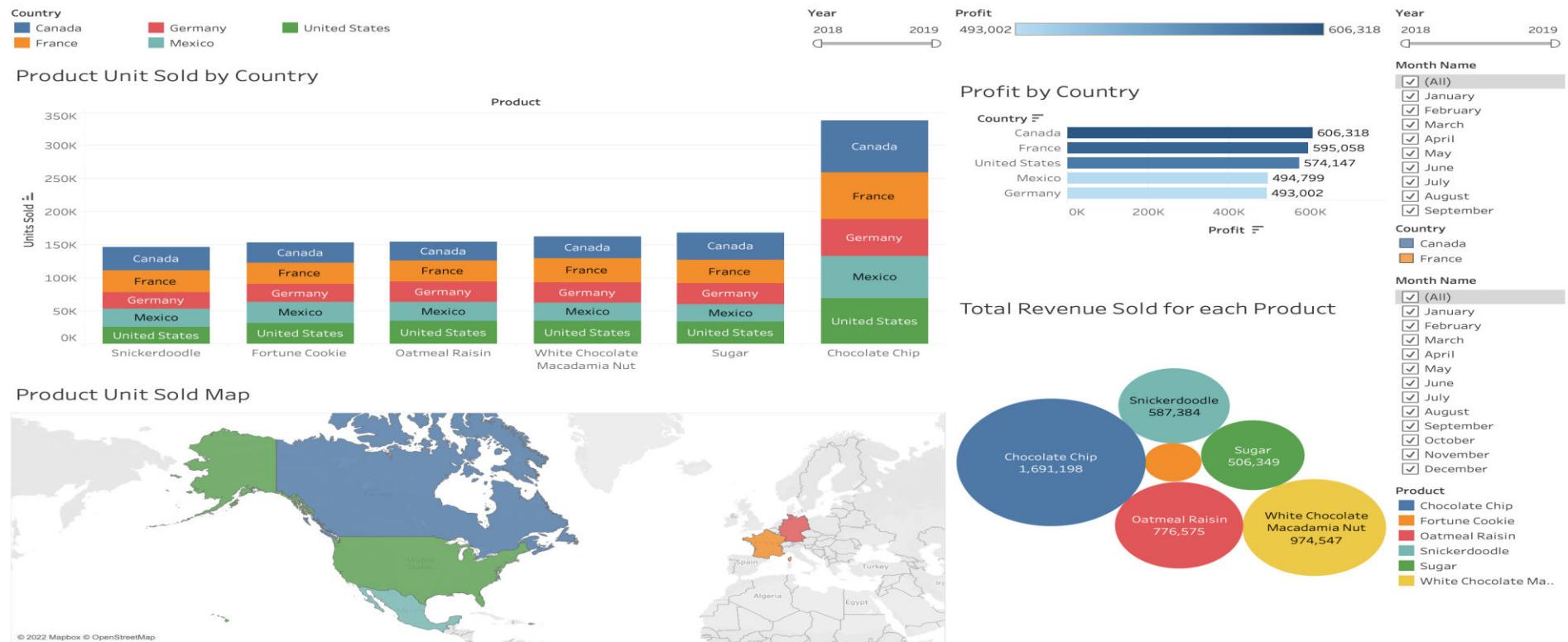
+tableau

Tableau Lab (Let's Try this)

- Let's try this (in class)

Hint: Total Revenue Sold = Units Sold * Revenue per cookie

- Download **Kevin Cookies Financial-Revised.csv** from LMS and try to create Data Visualization like this picture below:



Topic 09

Low Code for Data Management Data and Web Service Integration

BDM3302: Data Management

What is Low Code?

- Low code is a visual approach to software development that enables faster delivery of applications through minimal hand-coding
- Most low code tools include the graphical user interface (GUI) and the drag-and-drop features
- The point is to automate aspects of the development process, eliminating dependencies on traditional computer programming approaches
- Low code platforms democratize app development, particularly for the “citizen” developers
 - i.e. business users with little formal coding experience, such as business analysts or project managers

Why Low Code?

- These low code tools enable less technical employees to make a larger business impact in numerous ways
- Such as relieving IT department backlogs, reducing shadow IT, and taking more ownership over business process management (BPM) workstreams
- Low code development platforms also aid more seasoned programmers. Since they require little to no coding experience, they allow for more flexibility in a developer's coding background
- For example, some business applications require knowledge around a specific programming language, narrowing the selection of developers further
- By eliminating this bottleneck, low-code platforms shorten the application development lifecycle, enabling them to accomplish more in a less time

Low Code for Data Management

- There are so many low code tools for you to choose but most are in the paid subscription per month or year, and not relate for data management 100% (**Most are for web/mobile app development**)
- Here are 4 basic free low code tools below:
 - Pop SQL (Scripting Language and drag drop)
 - DBM Tool for JSON (Scripting Language and convert to table)
 - Python (100% Pure Scripting Language)
 - Node-RED (Drag drop)

Pop SQL



- The ease-of-use application which manages database using SQL commands to display the result
- Easy to setup the database environment
- Include export features to CSV or JSON format
- Free versions can accept the maximum of draft file up to five files, and one setup database environment
- Require MySQL installation (Review Basic SQL Week 2 slide)

Pop SQL

Edit My Classic Model

Connection name: My Classic Model

Hostname/Port: localhost 3306

Database: classicmodels

Username/Password: root *****

Credentials will be securely stored in your computer's keychain.

Connection Type: Connect directly from my computer
Direct: Use your computer's network to connect to your database.

Who has access? Add users or teams

Tanakom Tanto... Owner
Grant granular permissions to teams and specific users.
Talk to sales to add this to your team.

Advanced Options: Connect over SSH
Useful to connect to a private network.
 SSL options
SSL is mandatory

Test **Save** **Connect**

Sample Queries X MySample-1 X +

My Classic Model classicmodels 1 of 1 schemas Filter within schemas

MySample-1 Created 17 days ago Run Share

My Classic Model classicmodels

MySample-1

1 --Show all customer data
2 select *
3 from customers;
4
5 --Show some customer data
6 select customerNumber, customerName, phone, city, country
7 from customers
8 limit 50;
9
10 --Display total customers per country data
11 select country, count(1)
12 from customers
13 group by 1;

Clear all

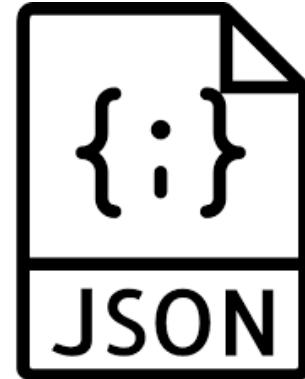
MySample-1 Success (122 rows) 0.2 s 8:50 PM

Explore SQL Data Chart Export

customerNumber	customerName	contactLast
103	Atelier graphique	Schmitt
112	Signal Gift Stores	King
114	Australian Collectors, Co.	Ferguson
119	La Rochelle Gifts	Labrune
121	Baane Mini Imports	Bergulfsen
124	Mini Gifts Distributors Ltd.	Nelson
125	Havel & Zbyszek Co	Piestrzewiewi
128	Blauer See Auto, Co.	Keitel
129	Mini Wheels Co.	Murphy
131	Land of Toys Inc.	Lee

1 to 10 of 122 |<|>| Page 1 of 13 |>|

DBM Tool for JSON



- DBM Tool v.1.2.2 by Dr. Anan Osothsip
- The tool that introduces how data can be constructed and converted into the SQL and database tables
- The input can be JSON only that process the output into SQL and database tables
- This tool also provides catalog, departments, roles, positions, users, processes, and assets in order to identify what kind of these data to support your business
- The Internet connection is also required for this tool (Review Data Model Diagram Week 6 Slide)

DBM Tool for JSON

- JSON structure explains ...
 - Stands for JavaScript Object Notation
 - It is a text format for storing and transporting data
 - It is a “self-describing” and easy to understand

- The example looks like this:

```
{  
    "name": "John",  
    "age": 30,  
    "car": "Toyota"  
}
```

*“Mr. John with age of 30 years old
has got a Toyota car.”*

It identifies with 3 properties:
1. name = John (text)
2. age = 30 (integer)
3. car = Toyota (text)

*Each property has a value

DBM Tool for JSON

Data
Model
Name

The screenshot shows the jsontool.html interface with several highlighted and annotated sections:

- Top Bar:** A blue bar contains the title "jsontool.html". Below it is a toolbar with the text "Data and Process Management". On the right side of the toolbar are several checkboxes: "json editor" (checked), "c# class editor" (unchecked), "sql editor" (checked), "code editor" (checked), "table editor" (checked), and "diagram" (unchecked). A yellow arrow points to the "Double click to open" text above the toolbar.
- Left Sidebar:** Labeled "Menu Selection", it includes tabs for "EDITOR", "CATALOG", "DEPARTMENTS", "ROLES", "POSITIONS", and "USERS". A yellow oval highlights the "Data Model" section, which shows "(table1)" and has a yellow arrow pointing to it from the left.
- Center Area:** A "Convert button" (green button labeled "CONVERT") is highlighted with a yellow oval. To its left is a "JSON Input Area" containing a JSON editor with a "JSON" tab and a "SQL" tab below it. The "SQL" tab is also highlighted with a yellow oval.
- Right Area:** A "DATABASE TABLE" output area is shown, which is also highlighted with a yellow oval. Above this area is a message: "Please uncheck this check box first!" followed by "Import/Export file". A yellow circle highlights the "c# class editor" checkbox in the top toolbar, and another yellow circle highlights the three-dot menu icon in the top right corner of the interface.
- Bottom Area:** A "SQL Output Area" is labeled in red at the bottom left of the interface.

Python

- Python is a high-level programming language developed by Guido Van Rossum
- It is an interpreted language and supports multiple programming paradigms such as imperative, functional, and object-oriented
- Python has a design philosophy that emphasizes code readability, and this makes it the most favorite programming language for beginners
- Data management can be a challenging task for beginners because it requires sophisticated and complex knowledge
- However, with python, you can automate data management processes and manage your data more efficiently
- Python is a 100% pure scripting language that not quite a low code but can be less code for beginners to development more
- Example: Pandas Profiling

Python

The screenshot shows a Jupyter Notebook interface with a red header bar. The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, and Python 3. Below the menu is a toolbar with icons for file operations like Open, Save, Run, and Cell. A code editor tab is open, showing a single line of code: `# Or use the HTML report in an iframe`. The main area displays a table titled "Last rows" from a pandas DataFrame. The table has 10 columns: Index, Age, Cabin, Embarked, Fare, Name, Parch, PassengerId, Pclass, Sex, SibSp, Survived, and Ticket. The data consists of 10 rows, each representing a passenger record. The "Name" column contains full names, and the "Ticket" column contains ticket numbers.

	Age	Cabin	Embarked	Fare	Name	Parch	PassengerId	Pclass	Sex	SibSp	Survived	Ticket
881	33.0	NaN	S	7.8958	Markun, Mr. Johann	0	882	3	male	0	0	349257
882	22.0	NaN	S	10.5167	Dahlberg, Miss. Gerda Ulrika	0	883	3	female	0	0	7552
883	28.0	NaN	S	10.5000	Banfield, Mr. Frederick James	0	884	2	male	0	0	C.A/SOTON 34068
884	25.0	NaN	S	7.0500	Suttehall, Mr. Henry Jr	0	885	3	male	0	0	SOTON/OQ 392076
885	39.0	NaN	Q	29.1250	Rice, Mrs. William (Margaret Norton)	5	886	3	female	0	0	382652
886	27.0	NaN	S	13.0000	Montvila, Rev. Juozas	0	887	2	male	0	0	211536
887	19.0	B42	S	30.0000	Graham, Miss. Margaret Edith	0	888	1	female	0	1	112053
888	NaN	NaN	S	23.4500	Johnston, Miss. Catherine Helen "Carrie"	2	889	3	female	1	0	W/J.C. 6607
889	26.0	C148	C	30.0000	Behr, Mr. Karl Howell	0	890	1	male	0	1	111369
890	32.0	NaN	Q	7.7500	Dooley, Mr. Patrick	0	891	3	male	0	0	370376

Report generated with [pandas-profiling](#).

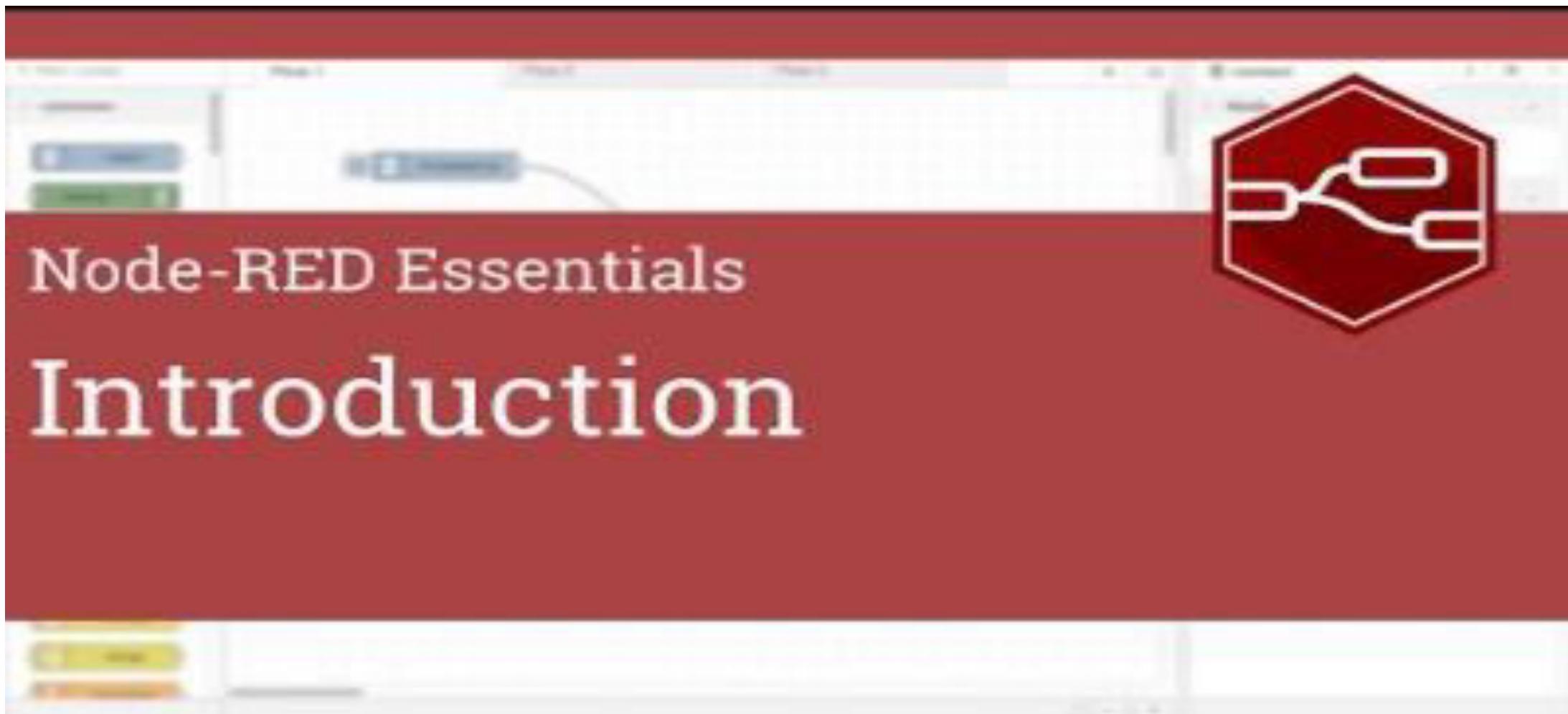
Out[6]:

```
In [ ]: # Or use the HTML report in an iframe
profile.to_notebook_iframe()
```

Node-RED

- An open-source low code tool, for connecting hardware devices, APIs and online services creatively and easily.
- Primarily, it is a visual tool designed for the Internet of Things, but can also be used for other applications to very quickly assemble flows of various services
- Node-RED provides a browser-based flow editor that makes it easy to wire together flows using the wide range of nodes in the palette. Flows can be then deployed to the runtime in a single-click

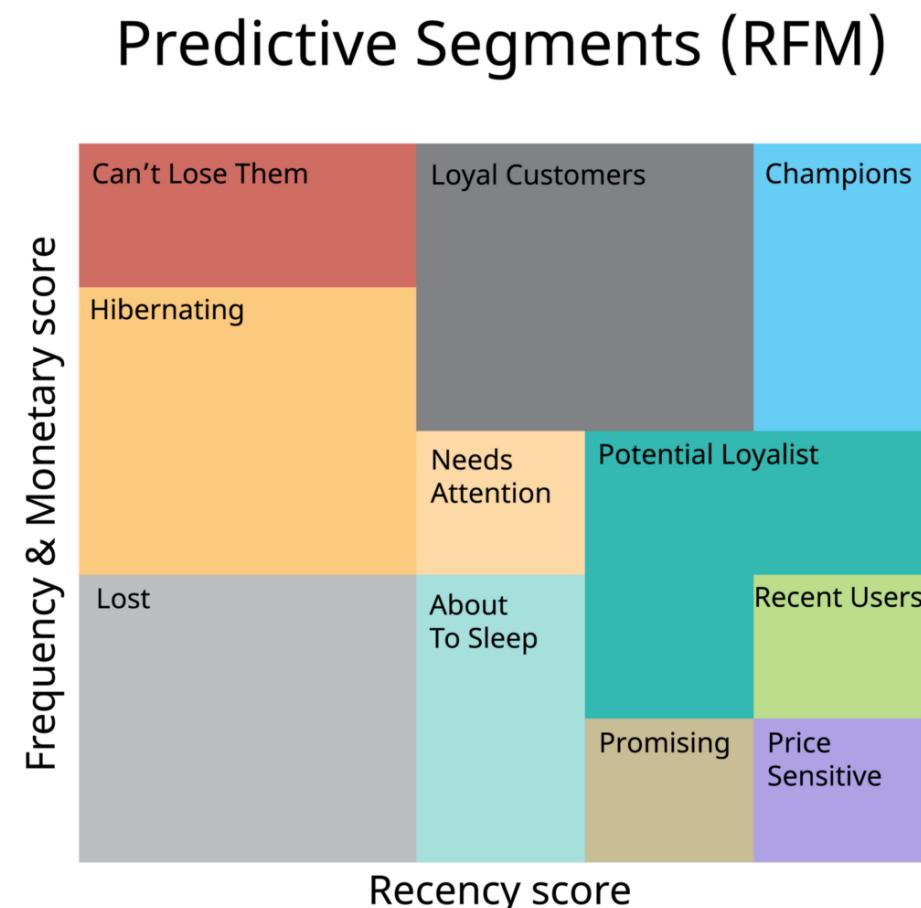
Node-RED



Source for this video: <https://youtu.be/ksGeUD26Mw0>

Web Service Integration using RFM Analysis

- A classification technique to help segment customers into a group of customer segmentation
- Use Big Sale Data (customer info, order detail, and date time)
- Based on Sale User Story
- Goals
 - ✓ Increased customer retention
 - ✓ Increased response rate
 - ✓ Increased conversion rate
 - ✓ Increased revenue





customer1



customer2



customer3

monetary

1000\$

500\$

100\$



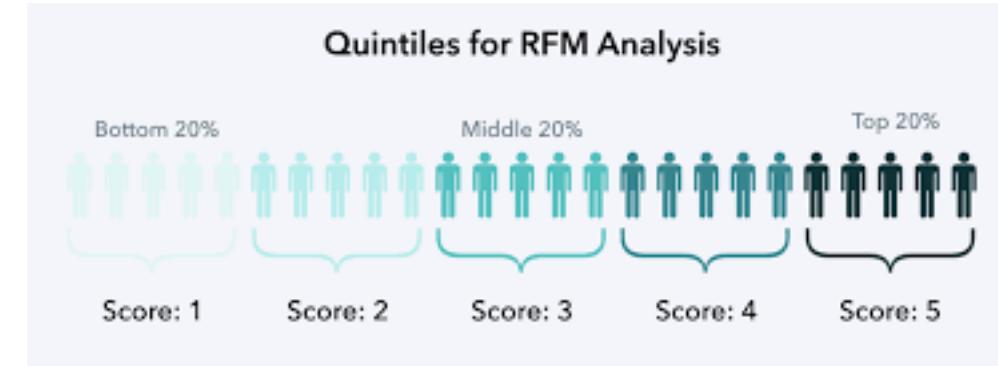
frequency

recency

2 years ago

Last 12 months

10 days



Quartile ranking

monetary	5	3	1	1000	500	100
frequency	1	5	1	1	4	1
recency	1	3	5	650 days	365 days	10 days
rfm score	115	353	511			

Data Management > RFM Score Matching

Segment	Scores
<i>Champions</i>	555, 554, 544, 545, 454, 455, 445
<i>Loyal Customers</i>	543, 444, 435, 355, 354, 345, 344, 335
<i>Potential Loyalist</i>	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
<i>Recent Customers</i>	512, 511, 422, 421, 412, 411, 311
<i>Promising</i>	525, 524, 523, 522, 521, 515, 514, 513, 425, 424, 413, 414, 415, 315, 314, 313
<i>Customers Needing Attention</i>	535, 534, 443, 434, 343, 334, 325, 324
<i>About To Sleep</i>	331, 321, 312, 221, 213
<i>At Risk</i>	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
<i>Can't Lose Them</i>	155, 154, 144, 214, 215, 115, 114, 113
<i>Hibernating</i>	332, 322, 231, 241, 251, 233, 232, 223, 222, 132, 123, 122, 212, 211
<i>Lost</i>	111, 112, 121, 131, 141, 151

Data Management > RFM Marketing action per segmentation

Customer Segment	Activity	Actionable Tip
Champions	Bought recently, buy often and spend the most!	Reward them. They can be early adopters for new products. Will promote your brand.
Loyal Customers	Often spend good money buying your products. Responsive to promotions.	Upsell higher value products. Ask for reviews. Engage them.
Potential Loyalist	Recent customers, but spent a good amount and bought more than once.	Offer membership/loyalty programs and recommend other products.
Recent Customers	Bought most recently, but not often.	Provide onboarding support, give them early success, and start building relationships.
Promising	Recent shoppers, but haven't spent much.	Create brand awareness, offer free trials.
Customers Needing Attention	Above-average recency, frequency and monetary values. They may not have bought very recently though.	Make limited-time offers. Recommendations based on past purchases. Reactivate them.

Customer Segment	Activity	Actionable Tip
About To Sleep	Below average recency, frequency, and monetary values. Will lose them if not reactivated.	Share valuable resources. Recommend popular products/renewals at discount. Reconnect with them.
At Risk	They spent big money and purchased often. But the last purchase was a long time ago. Need to bring them back!	Send personalized emails to reconnect, offer renewals, provide helpful resources.
Can't Lose Them	Often made the biggest purchases but they haven't returned for a long time.	Win them back via renewals or newer products. Don't lose them to competition, talk to them.
Hibernating	The last purchase was long ago. Low spenders with a low number of orders.	Offer other relevant products and special discounts. Recreate brand value.
Lost	Lowest recency, frequency, and monetary scores.	Revive interest with a reach-out campaign, ignore otherwise.

1: purchase (the majority)

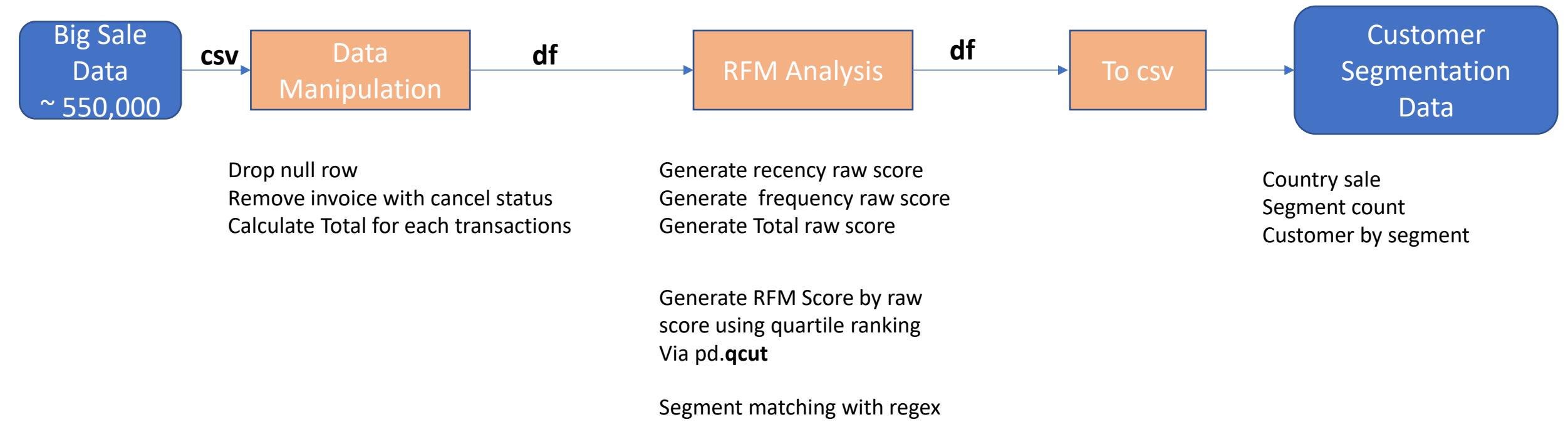
2 : purchases (those who had overcome the psychological barrier and came again, knowing how the store works)

3-4 : purchases (in spite of everything, they continue to buy, so it is necessary to keep them at any costs)

5-15 : purchases (believed us, they are our main hope. The upper limit here can be very different)

more than 15 purchases : (freaks, often resellers, partners, etc. They need to be reviewed separately)

Data Management > Data Management workflow for RFM analytic



Customer ID	recency	frequency	monetary	recency_sc	frequency_sc	monetary_sc	rfm_score	segment
12346	326	1	77183.6	5	1	5	515	new customer big bill
12347	40	7	4310	2	5	5	255	loyal customer
12348	76	4	1797.24	3	4	4	344	loyal customer
12349	19	1	1757.55	2	1	4	214	about to sleep