

DATABASE, DATA WAREHOUSE, AND BUSINESS INTELLIGENCE

EGCO103 INFORMATION TECHNOLOGY IN THE DAILY LIFE



KANAT POOLSAWASD
DEPARTMENT OF COMPUTER ENGINEERING
MAHIDOL UNIVERSITY

WHY DO WE NEED DATABASE? (1)

- You want to store your employee phone numbers.
- What will you do?
- What if?
 - You have a company with 100000+ employees across the globe.
 - You want to add these information
 - Job name
 - Job description
 - Min/max salary for the job
 - Working location
 - Location telephone

WHY DO WE NEED DATABASE? (2)

- When we have more complicate question to ask.
- When there are a lot of information and there are many duplicate information.

TYPE OF DATABASE

- Three Types of Database
 - Relational
 - Multidimensional
 - NoSQL
- Relational databases is most common

BIG PLAYERS

- IBM
 - DB2
- Oracle
 - Oracle DB = Leading in the Market.
 - MySQL = Free
- SAP
- Microsoft
 - Access = Personal/Small Data.
 - Microsoft SQL Server = Enterprise/Big Data.

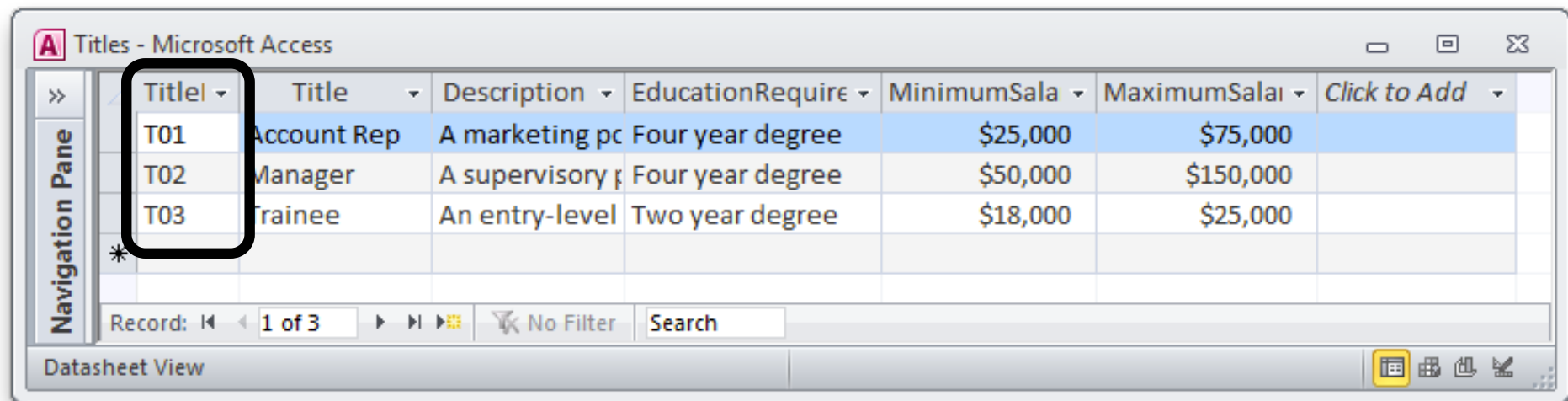
RELATIONAL DATABASES

- Data in tables
- Define relation between tables.
- Primary key is unique for each record (row)

DATABASE TERMINOLOGIES

- Databases have three main components;
 - Fields
 - Store each category of information
 - Displayed in columns
 - Records
 - Group of related fields
 - Tables (or files)
 - Group of related records

FIELD (COLUMN)



The screenshot shows the Microsoft Access interface for a table named 'Titles'. The table is displayed in Datasheet View. The 'Title' field in the first row is highlighted with a black box. The table has the following data:

Title	Description	EducationRequire	MinimumSala	MaximumSala	Click to Add
T01	Account Rep	Four year degree	\$25,000	\$75,000	
T02	Manager	Four year degree	\$50,000	\$150,000	
T03	Trainee	Two year degree	\$18,000	\$25,000	
*					

The interface includes a 'Navigation Pane' on the left, a status bar at the bottom showing 'Record: 1 of 3', and a search bar.

RECORD (ROW)

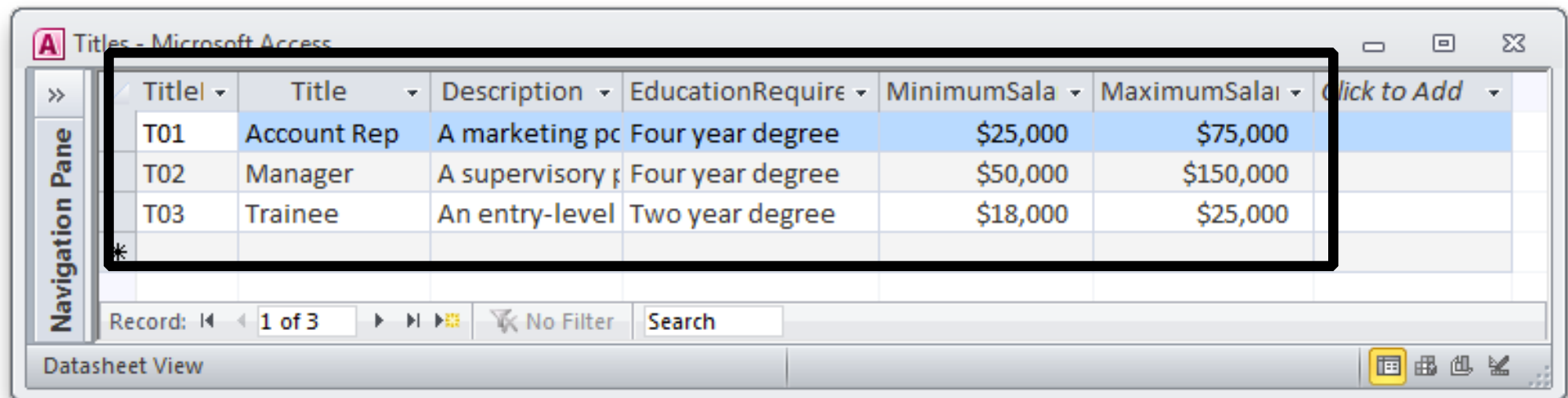
Navigation Pane

Title	Title	Description	EducationRequire	MinimumSala	MaximumSala	Click to Add
T01	Account Rep	A marketing pc	Four year degree	\$25,000	\$75,000	
T02	Manager	A supervisory p	Four year degree	\$50,000	\$150,000	
T03	Trainee	An entry-level	Two year degree	\$18,000	\$25,000	
*						

Record: 1 of 3 No Filter Search

Datasheet View

TABLE (ENTITY)



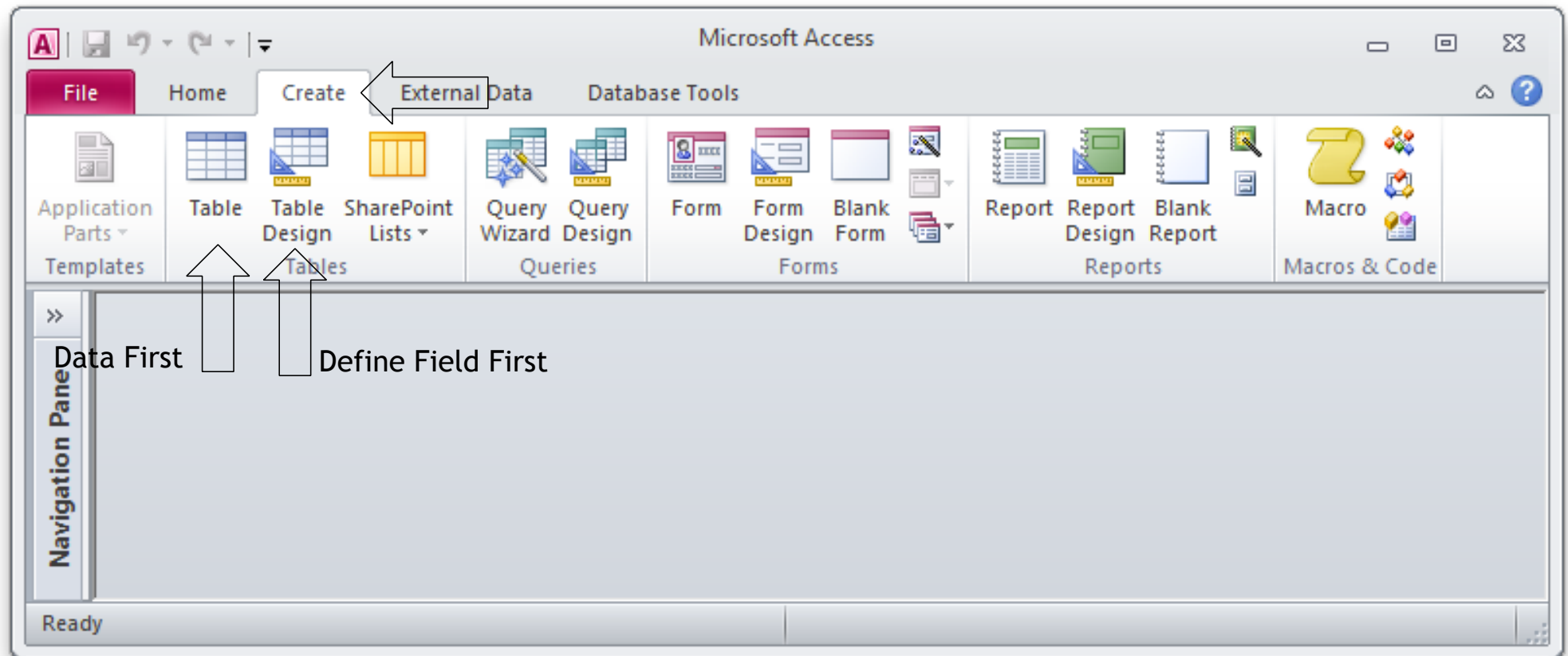
Navigation Pane

Title	Description	EducationRequired	MinimumSalary	MaximumSalary	Click to Add
T01	Account Rep	A marketing position	\$25,000	\$75,000	
T02	Manager	A supervisory position	\$50,000	\$150,000	
T03	Trainee	An entry-level position	\$18,000	\$25,000	
*					

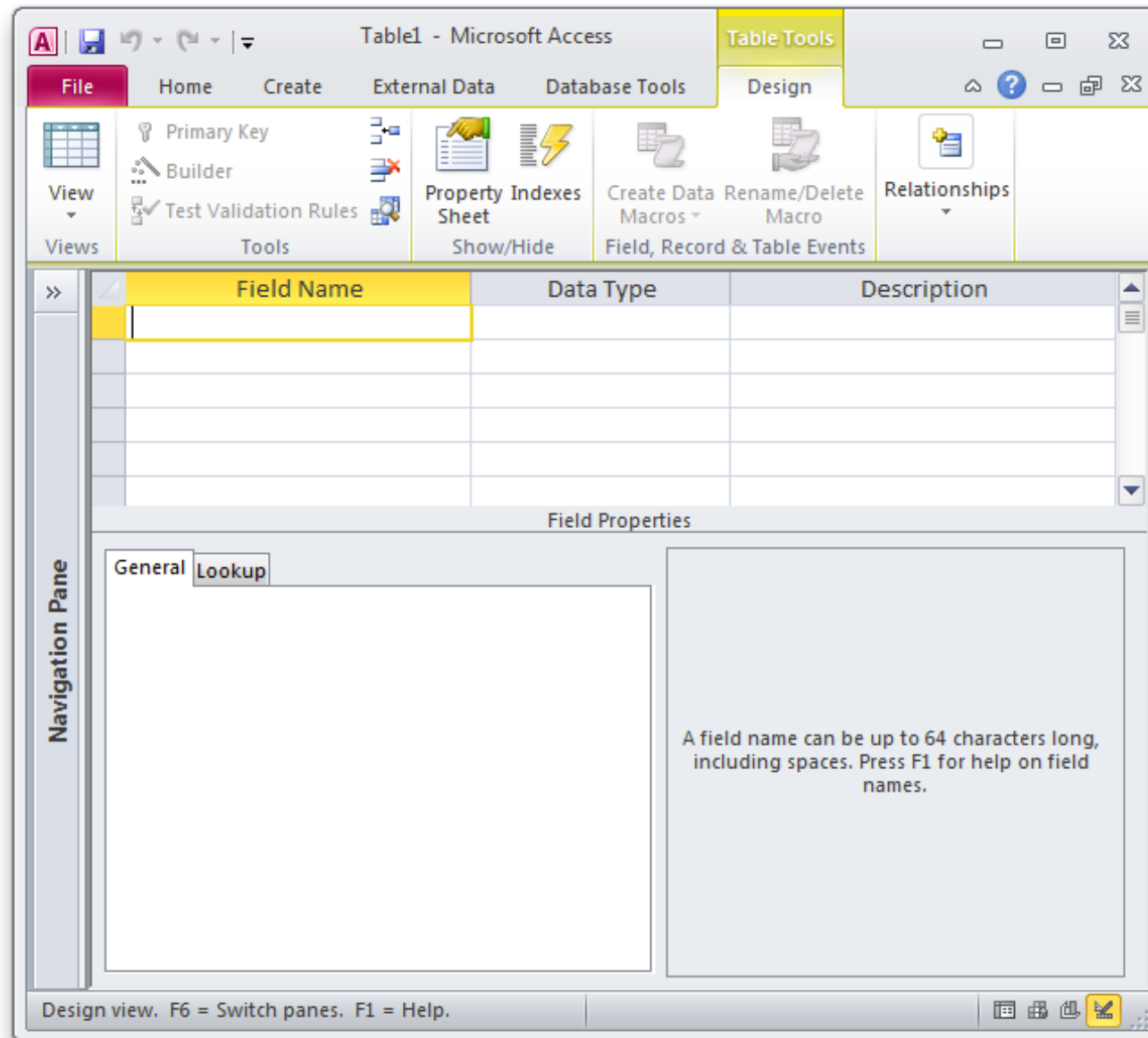
Record: 1 of 3 No Filter Search

Datasheet View

NEW TABLE IN MICROSOFT ACCESS



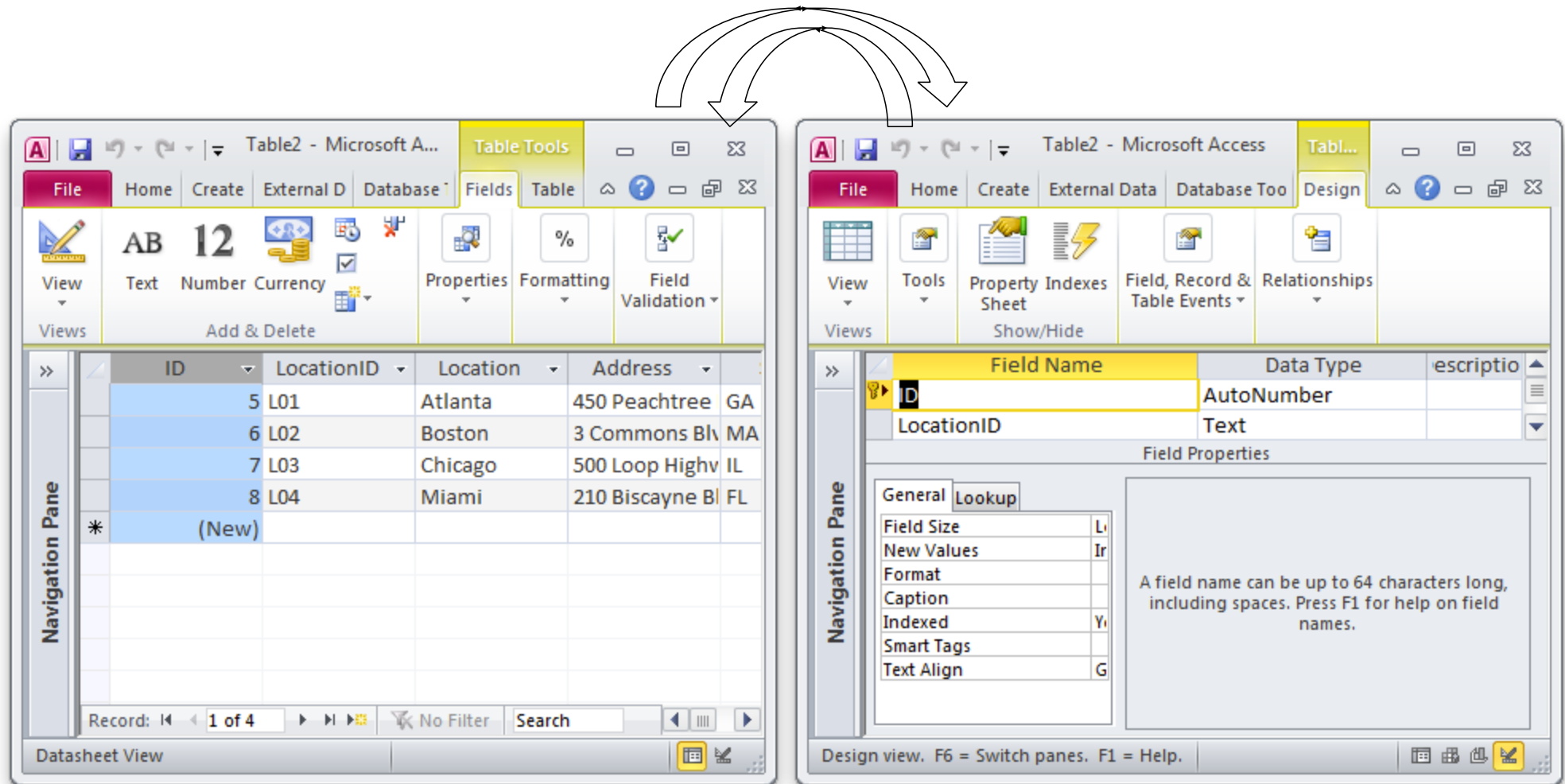
DEFINE FILED



COMMON DATA TYPES

Data Type	Used to Store	Examples
Text	Short text with maximum length limit	<i>John Doe</i>
Number	Numbers	<i>2901 or 3.499</i>
Yes/No	Boolean	<i>False/True or 0/1</i>
Date & Time	Dates	<i>2/21/2016</i>
Memo	Long text with no limit	<i>I want to finish this exercise so I can go home.</i>
Calculated	Formula	<i>Grade * Credit</i>
Attachment	Files	<i>Document, Picture</i>
Hyperlink	Hyperlink to a Web page	<i>Google.com</i>

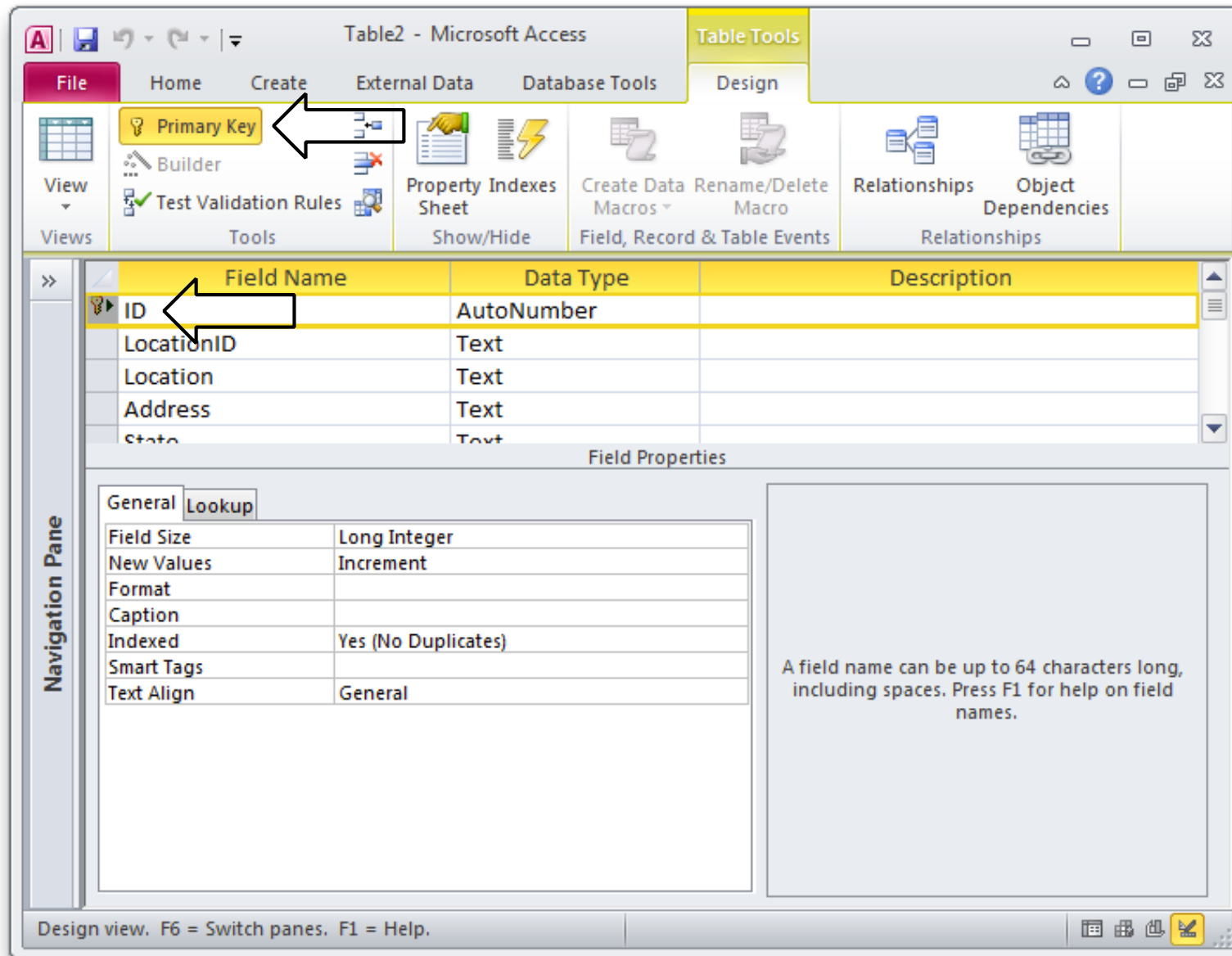
SWITCH BETWEEN DESIGN AND VIEW



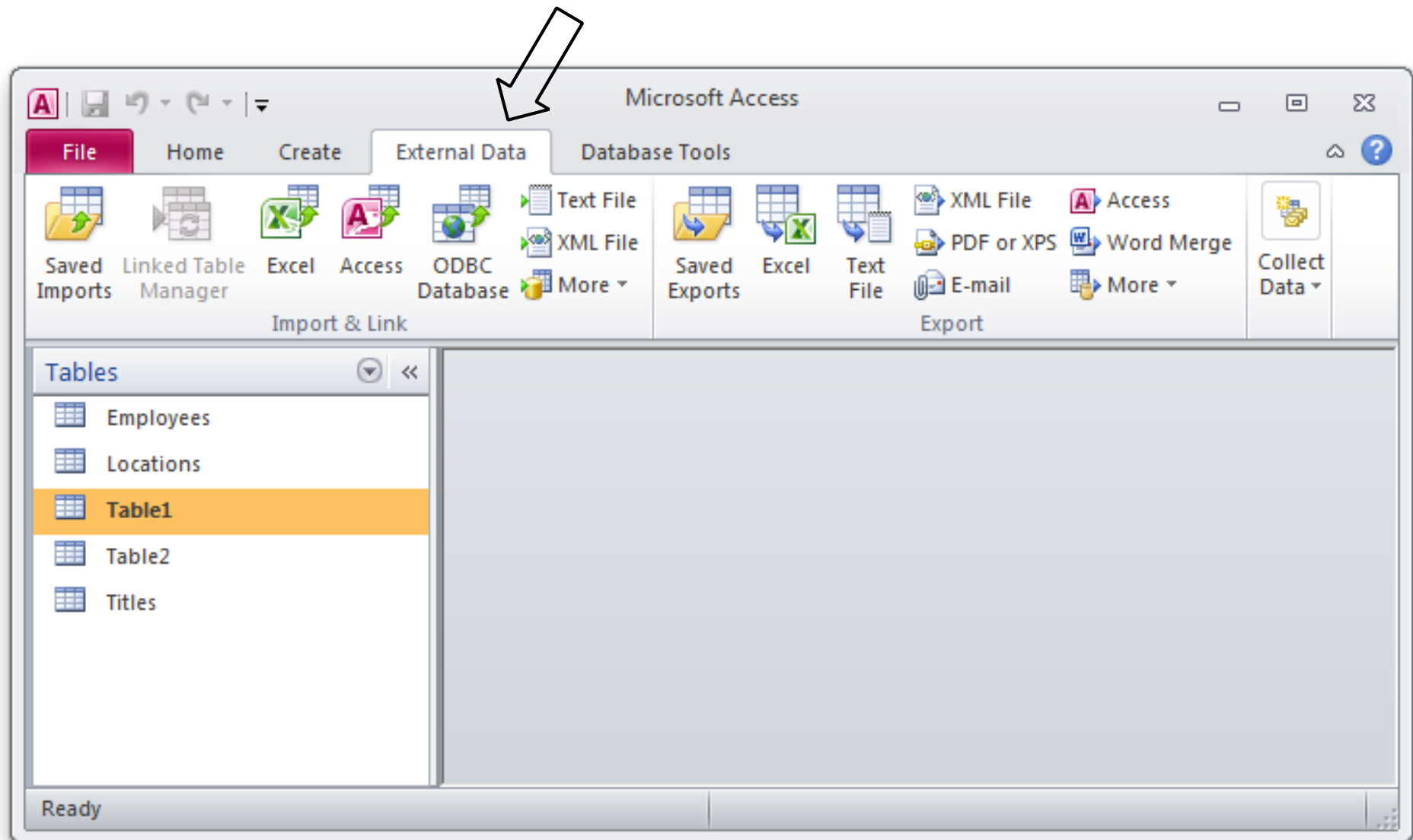
PRIMARY KEY (PK)

- Row must be unique!
- Database must be able to tell rows apart quickly to be fast
- Find fields (columns) that together will be unique and set them to be primary key.
- Can be set in Design mode

SET PK IN MICROSOFT ACCESS



IMPORT & EXPORT DATA (1)



IMPORT & EXPORT DATA (2)

Get External Data - Excel Spreadsheet

Select the source and destination of the data

Specify the source of the data.

File name: C:\Users\Ruj\Documents\ Browse...

Specify how and where you want to store the data in the current database.

☒ **Import the source data into a new table in the current database.**
If the specified table does not exist, Access will create it. If the specified table already exists, Access might overwrite its contents with the imported data. Changes made to the source data will not be reflected in the database.

☐ **Append a copy of the records to the table:** Locations
If the specified table exists, Access will add the records to the table. If the table does not exist, Access will create it. Changes made to the source data will not be reflected in the database.

☐ **Link to the data source by creating a linked table.**
Access will create a table that will maintain a link to the source data in Excel. Changes made to the source data in Excel will be reflected in the linked table. However, the source data cannot be changed from within Access.

OK Cancel

IMPORT & EXPORT DATA (3)

Import Spreadsheet Wizard

Your spreadsheet file contains more than one worksheet or range. Which worksheet or range would you like?

☒ Show Worksheets
☐ Show Named Ranges

Employees

Sample data for worksheet 'Employees'.

	EmployeeID	LastName	FirstName	LocationID	TitleID	Salary	Gender	Performance
1	10000	Milgrom	Pamela	L02	T02	57,500.00	F	Average
2	11111	Adams	Jennifer	L01	T03	19,500.00	F	Average
3	20000	Johnson	James	L03	T01	47,500.00	M	Good
4	22222	Coulter	Tracey	L01	T02	100,000.00	F	Good
5	30000	Marlin	Billy	L04	T02	125,000.00	M	Good
6	33333	Smith	Mark	L03	T01	42,500.00	M	Average
7	40000	Manin	Ann	L02	T01	49,500.00	F	Average
8	44444	Smith	Francine	L01	T01	65,000.00	F	Good
9	50000	Brown	Mark	L01	T03	18,500.00	M	Poor
10	55555	Frank	Vernon	L04	T01	75,000.00	M	Good
11	60000	Rubin	Patricia	L02	T01	45,000.00	F	Average
12	66666	Charles	Kenneth	L02	T01	40,000.00	M	Poor
13	70000	Adamson	David	L03	T02	52,000.00	M	Poor

Cancel < Back Next > Finish

IMPORT & EXPORT DATA (4)

Import Spreadsheet Wizard

Microsoft Access can use your column headings as field names for your table. Does the first row specified contain column headings?

☒ First Row Contains Column Headings

	EmployeeID	LastName	FirstName	LocationID	TitleID	Salary	Gender	Performance
1	10000	Milgrom	Pamela	L02	T02	57,500.00	F	Average
2	11111	Adams	Jennifer	L01	T03	19,500.00	F	Average
3	20000	Johnson	James	L03	T01	47,500.00	M	Good
4	22222	Coulter	Tracey	L01	T02	100,000.00	F	Good
5	30000	Marlin	Billy	L04	T02	125,000.00	M	Good
6	33333	Smith	Mark	L03	T01	42,500.00	M	Average
7	40000	Manin	Ann	L02	T01	49,500.00	F	Average
8	44444	Smith	Francine	L01	T01	65,000.00	F	Good
9	50000	Brown	Mark	L01	T03	18,500.00	M	Poor
10	55555	Frank	Vernon	L04	T01	75,000.00	M	Good
11	60000	Rubin	Patricia	L02	T01	45,000.00	F	Average
12	66666	Charles	Kenneth	L02	T01	40,000.00	M	Poor
13	70000	Adamson	David	L03	T02	52,000.00	M	Poor
14	77777	Marder	Kelly	L03	T01	38,500.00	F	Average

Cancel < Back Next > Finish

IMPORT & EXPORT DATA (5)

Import Spreadsheet Wizard

You can specify information about each of the fields you are importing. Select fields in the area below. You can then modify field information in the 'Field Options' area.

Field Options

Field Name: Data Type:

Indexed: ☐ Do not import field (Skip)

	EmployeeID	LastName	FirstName	LocationID	TitleID	Salary	Gender	Performance
1	10000	Milgrom	Pamela	L02	T02	57,500.00	F	Average
2	11111	Adams	Jennifer	L01	T03	19,500.00	F	Average
3	20000	Johnson	James	L03	T01	47,500.00	M	Good
4	22222	Coulter	Tracey	L01	T02	100,000.00	F	Good
5	30000	Marlin	Billy	L04	T02	125,000.00	M	Good
6	33333	Smith	Mark	L03	T01	42,500.00	M	Average
7	40000	Manin	Ann	L02	T01	49,500.00	F	Average
8	44444	Smith	Francine	L01	T01	65,000.00	F	Good
9	50000	Brown	Mark	L01	T03	18,500.00	M	Poor
10	55555	Frank	Vernon	L04	T01	75,000.00	M	Good
11	60000	Rubin	Patricia	L02	T01	45,000.00	F	Average
12	66666	Charles	Kenneth	L02	T01	40,000.00	M	Poor
13	70000	Adamson	David	L03	T02	52,000.00	M	Poor
14	77777	Marder	Kelly	L03	T01	38,500.00	F	Average

Cancel < Back Next > Finish

IMPORT & EXPORT DATA (6)

Import Spreadsheet Wizard

Microsoft Access recommends that you define a primary key for your new table. A primary key is used to uniquely identify each record in your table. It allows you to retrieve data more quickly.

☐ Let Access add primary key.

☒ Choose my own primary key. EmployeeID

☐ No primary key.

	EmployeeID	LastName	FirstName	LocationID	TitleID	Salary	Gender	Performance
1	10000	Milgrom	Pamela	L02	T02	57,500.00	F	Average
2	11111	Adams	Jennifer	L01	T03	19,500.00	F	Average
3	20000	Johnson	James	L03	T01	47,500.00	M	Good
4	22222	Coulter	Tracey	L01	T02	100,000.00	F	Good
5	30000	Marlin	Billy	L04	T02	125,000.00	M	Good
6	33333	Smith	Mark	L03	T01	42,500.00	M	Average
7	40000	Manin	Ann	L02	T01	49,500.00	F	Average
8	44444	Smith	Francine	L01	T01	65,000.00	F	Good
9	50000	Brown	Mark	L01	T03	18,500.00	M	Poor
10	55555	Frank	Vernon	L04	T01	75,000.00	M	Good
11	60000	Rubin	Patricia	L02	T01	45,000.00	F	Average
12	66666	Charles	Kenneth	L02	T01	40,000.00	M	Poor
13	70000	Adamson	David	L03	T02	52,000.00	M	Poor
14	77777	Marder	Kelly	L03	T01	38,500.00	F	Average

Cancel < Back Next > Finish

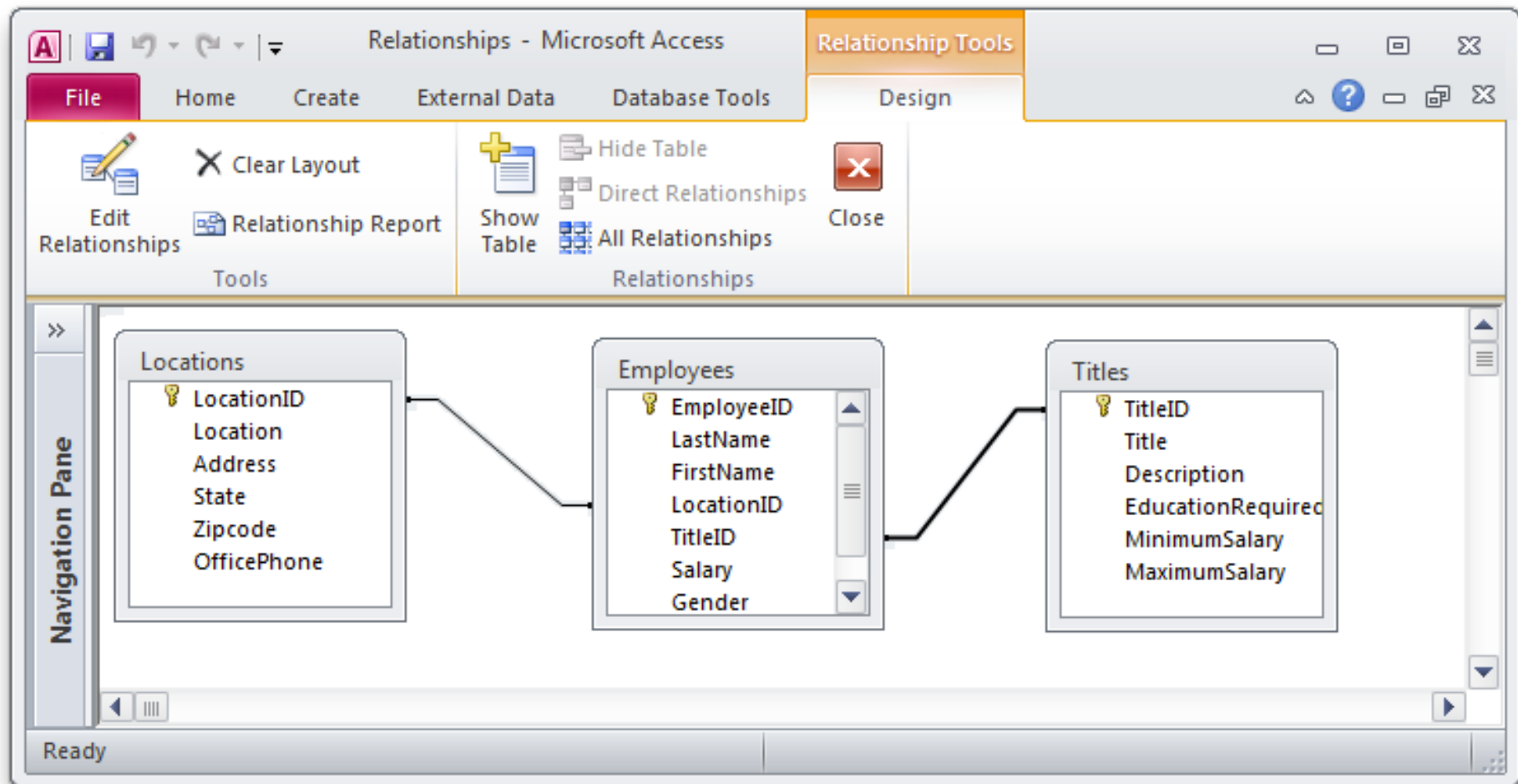
RELATIONSHIP

- Define relationship to make querying easier
- Tell how database can look up more info from other tables

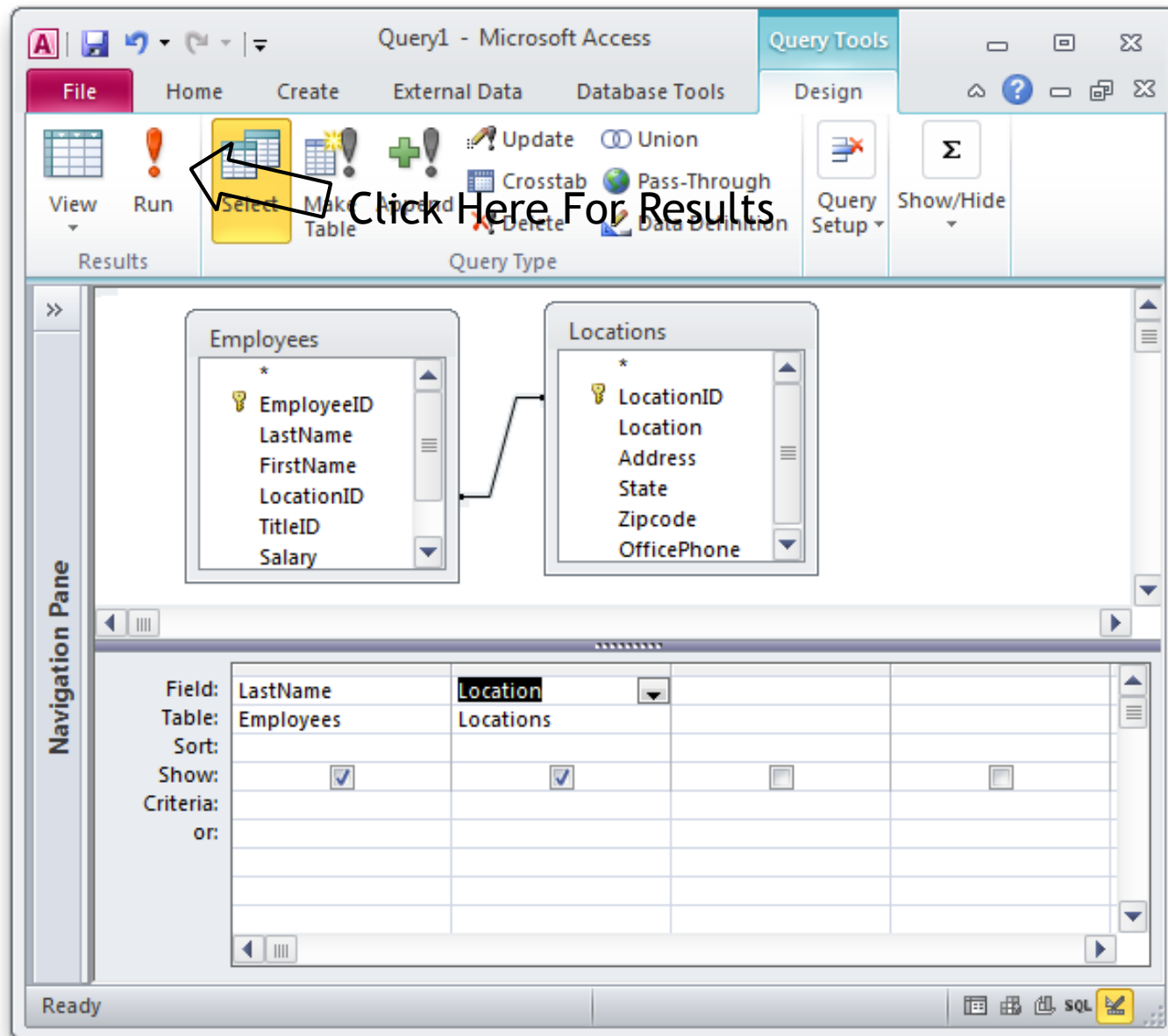
TYPE OF RELATIONSHIPS

- One-to-one
 - For each record in a table, there is only one corresponding record in a related table
- One-to-many
 - Only one instance of a record in one table; many instances in a related table
- Many-to-many
 - Records in one table related to multiple records in another

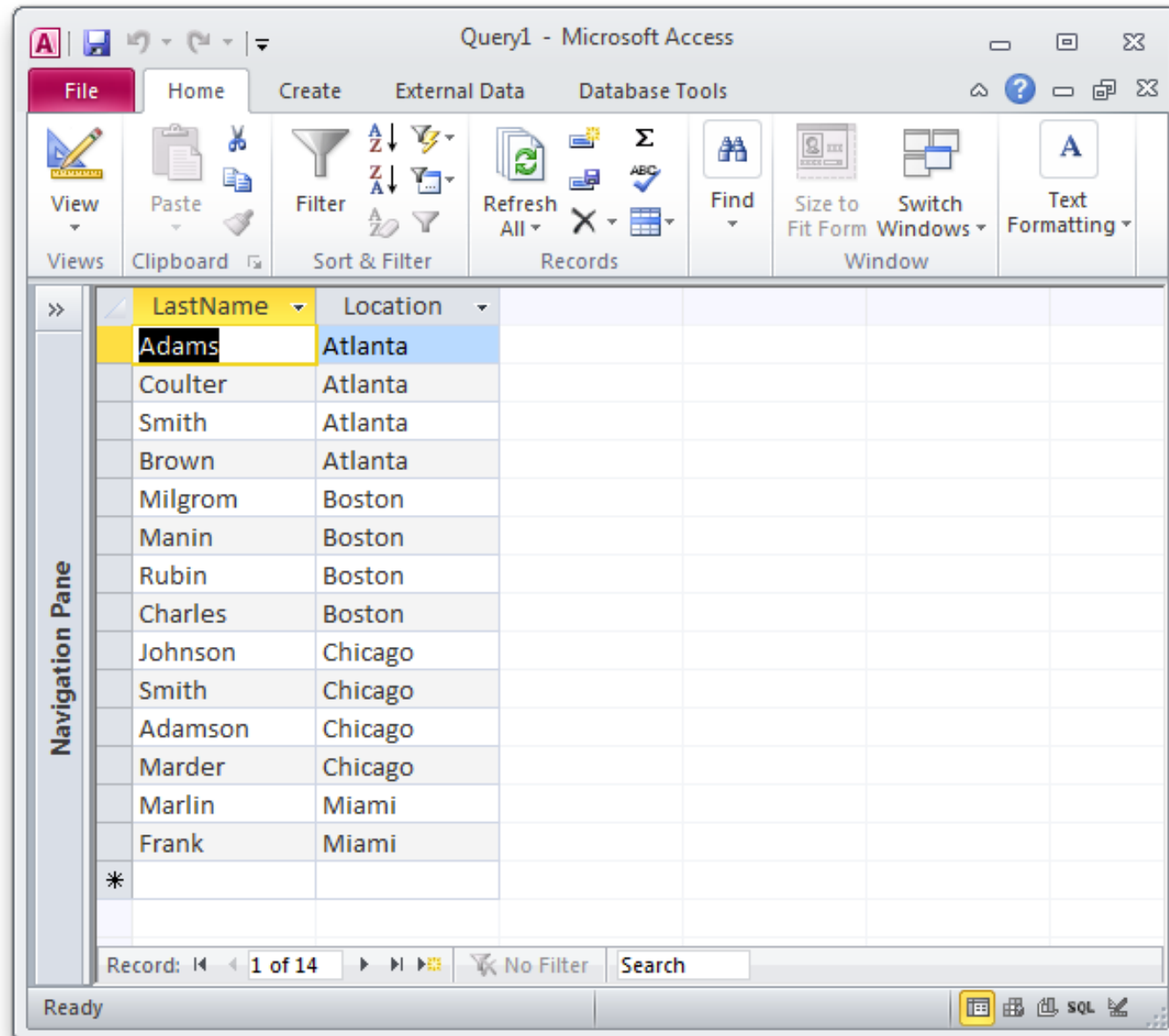
DEFINE RELATIONSHIP



QUERYING (1)



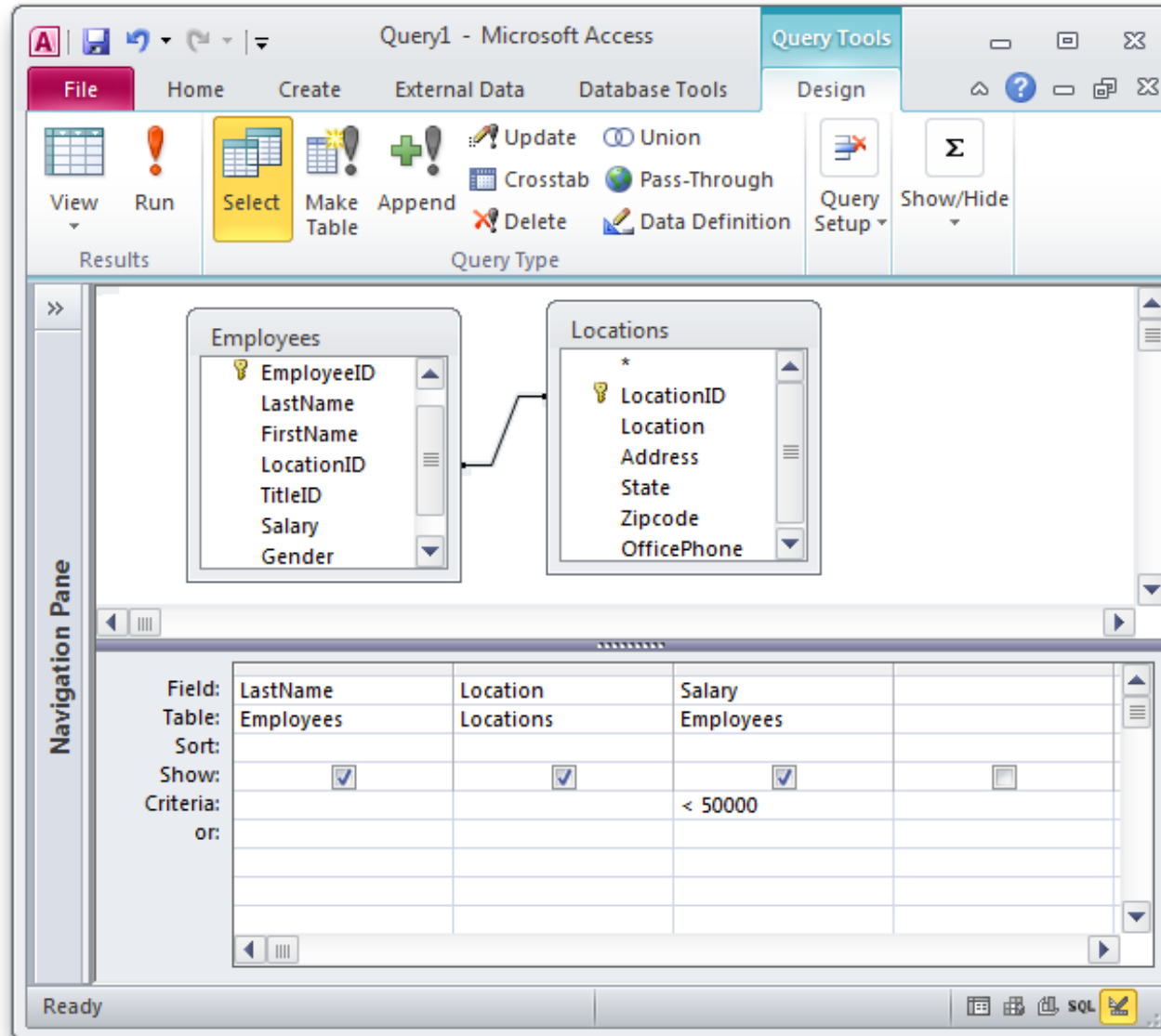
QUERYING (2)



QUERYING (3)

- You can set condition
 - *Salary < 50,000*
- Same row means "and"
- Different row means "or"
- Try putting "Boston" into different rows

QUERYING (4)



QUERYING (5)

- Using SQL command for querying data.

```
SELECT Lastname, Locaton, Salary  
FROM Employees JOIN Locations  
ON Employees.LocatonID = Locations.LocationID  
WHERE Salary < 50000
```

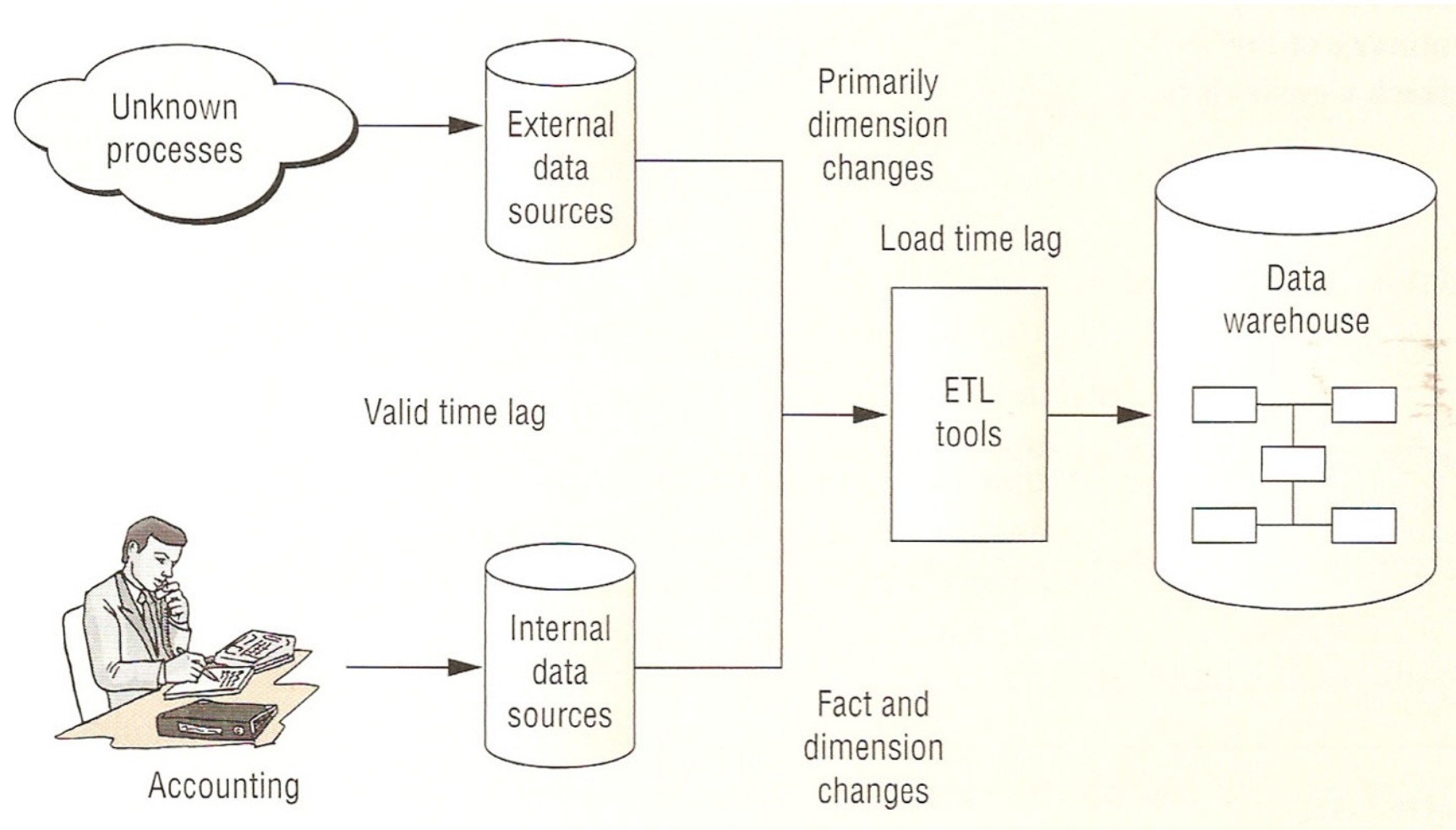
DATA WAREHOUSE

- Data warehouse is a central repository for summarized and integrated data from operational databases and external data sources.
- The processing requirement of decision support applications have led to four distinguishing characteristics for data warehouses, as described in the following:
 - Subject-Oriented
 - Integrated
 - Time-Variant
 - Nonvolatile

COMPARISON OF OPERATIONAL DATABASES AND DATA WAREHOUSES

Characteristic	Operational Database	Data Warehouse
Currency	Current	Historical
Detail level	Individual	Individual and summary
Orientation	Process orientation	Subject orientation
Number of records processed	Few	Thousands
Normalization level	Mostly normalized	Frequent violations of BCNF
Update level	Volatile	Nonvolatile (refreshed)
Data model	Relational	Relational model with star schemas and multidimensional model with data cubes

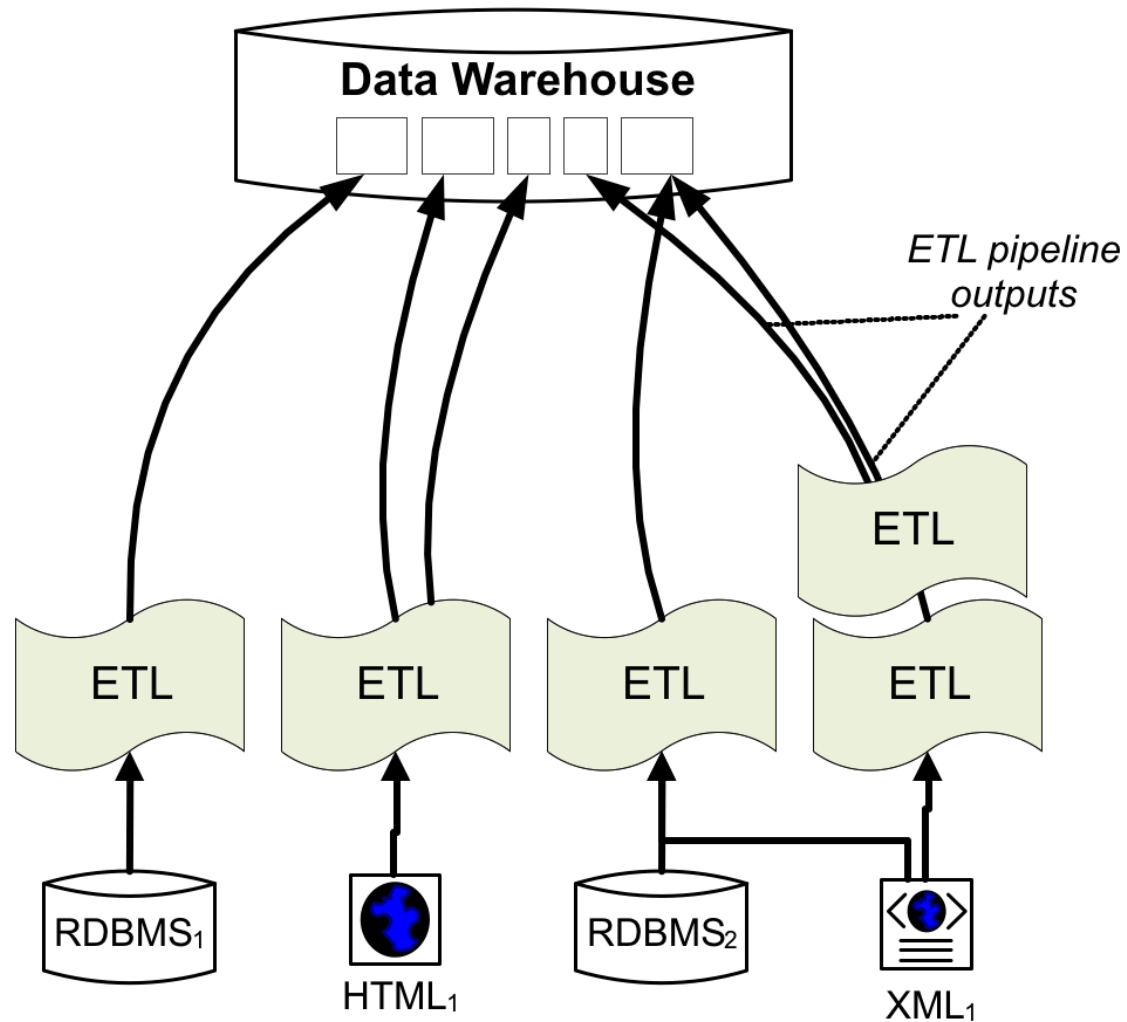
OVERVIEW OF THE DATA WAREHOUSE REFRESH PROCESS



ETL TOOLS (1)

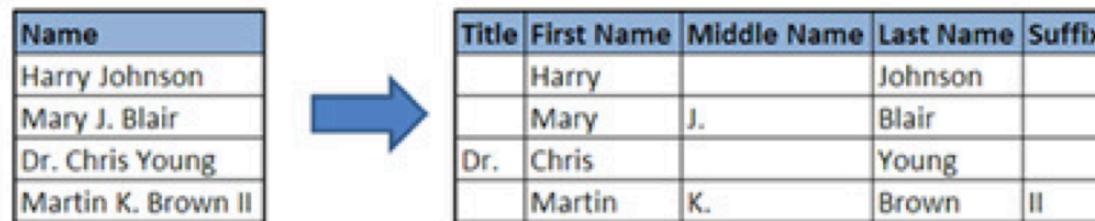
- To support complexity of data warehouse maintenance, software product known as Extraction, Transformation, and Load (ETL) tools have been developed.
- ETL tools are software tools for extraction, transformation, and loading of change data from data sources to a data warehouse.
- ETL tools eliminate the need to write custom coding for many data warehouse maintenance tasks.

ETL TOOLS (2)




DATA CLEANING - PARSING

- Parsing: locates and identifies individual data elements in the source file and then isolates these data elements in the target files.

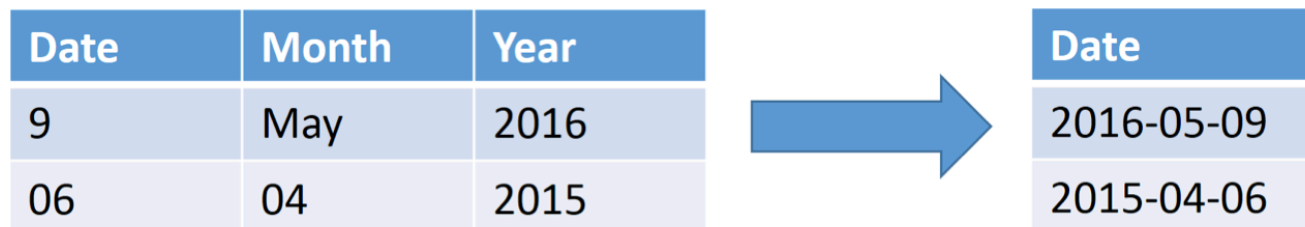


Name
Harry Johnson
Mary J. Blair
Dr. Chris Young
Martin K. Brown II




Title	First Name	Middle Name	Last Name	Suffix
	Harry		Johnson	
	Mary	J.	Blair	
Dr.	Chris		Young	
	Martin	K.	Brown	II

- Combing: locates and identifies individual data elements in the source file and then combines these data elements in the target files.



Date	Month	Year
9	May	2016
06	04	2015



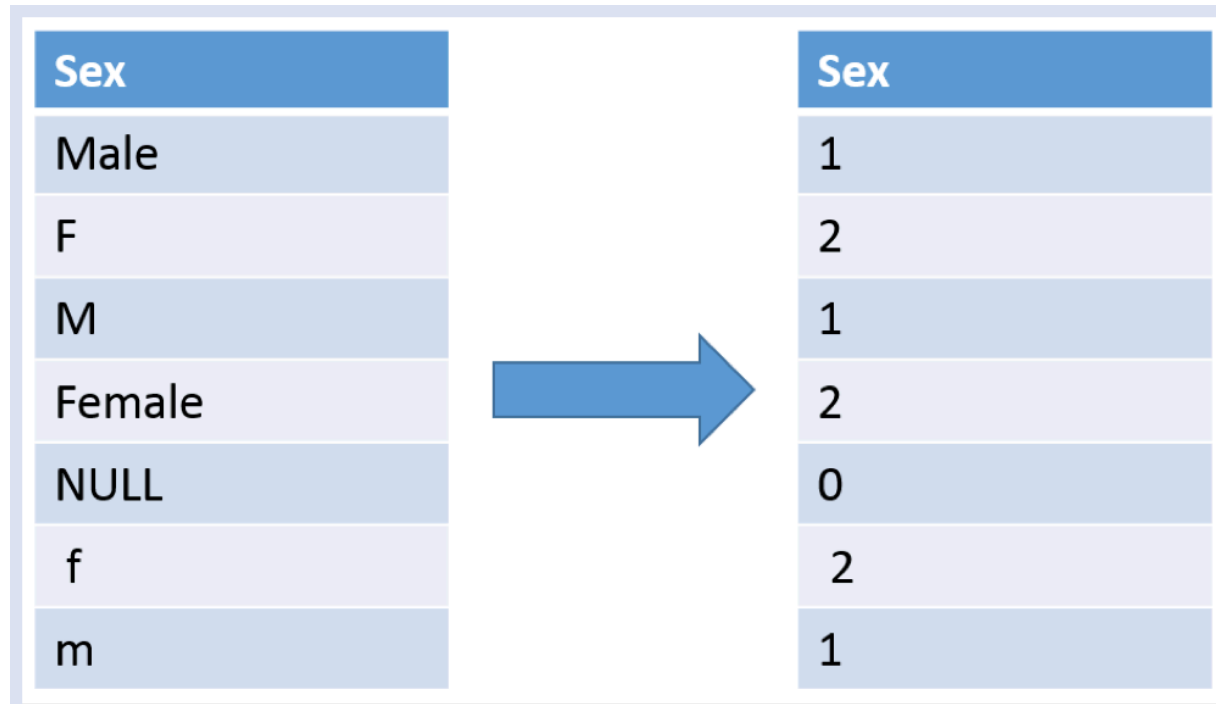
Date
2016-05-09
2015-04-06

DATA CLEANING - CORRECTING

- Correcting parsing individual data components using sophisticated data algorithms and secondary data sources
- Correct data according to data rules
- Example includes converting the combined date into a standard date format.

DATA CLEANING - STANDARDIZING

- Standardizing: applies conversion routines to transform data into its preferred (and consistent) format using standard and custom data rules.



The diagram illustrates the process of standardizing sex data. It shows two tables connected by a large blue arrow pointing from left to right. The left table lists various input formats for sex, and the right table shows the corresponding standardized numerical values.

Sex		Sex
Male		1
F		2
M		1
Female		2
NULL		0
f		2
m		1

DATA CLEANING - MATCHING

- Searching and matching records within and across the parsed, corrected and standardized data based on predefined data rules to eliminate duplications, sequences.

Pregnancy Number	Outcome Date	Outcome
1	2011-06-07	Twin


DoB	Name
2011-06-07	Child1
2011-06-07	Child2



Pregnancy Number	Birth Date	Name
1	2011-06-07	Child1
1	2011-06-07	Child2

DATA CLEANING - CONSOLIDATING

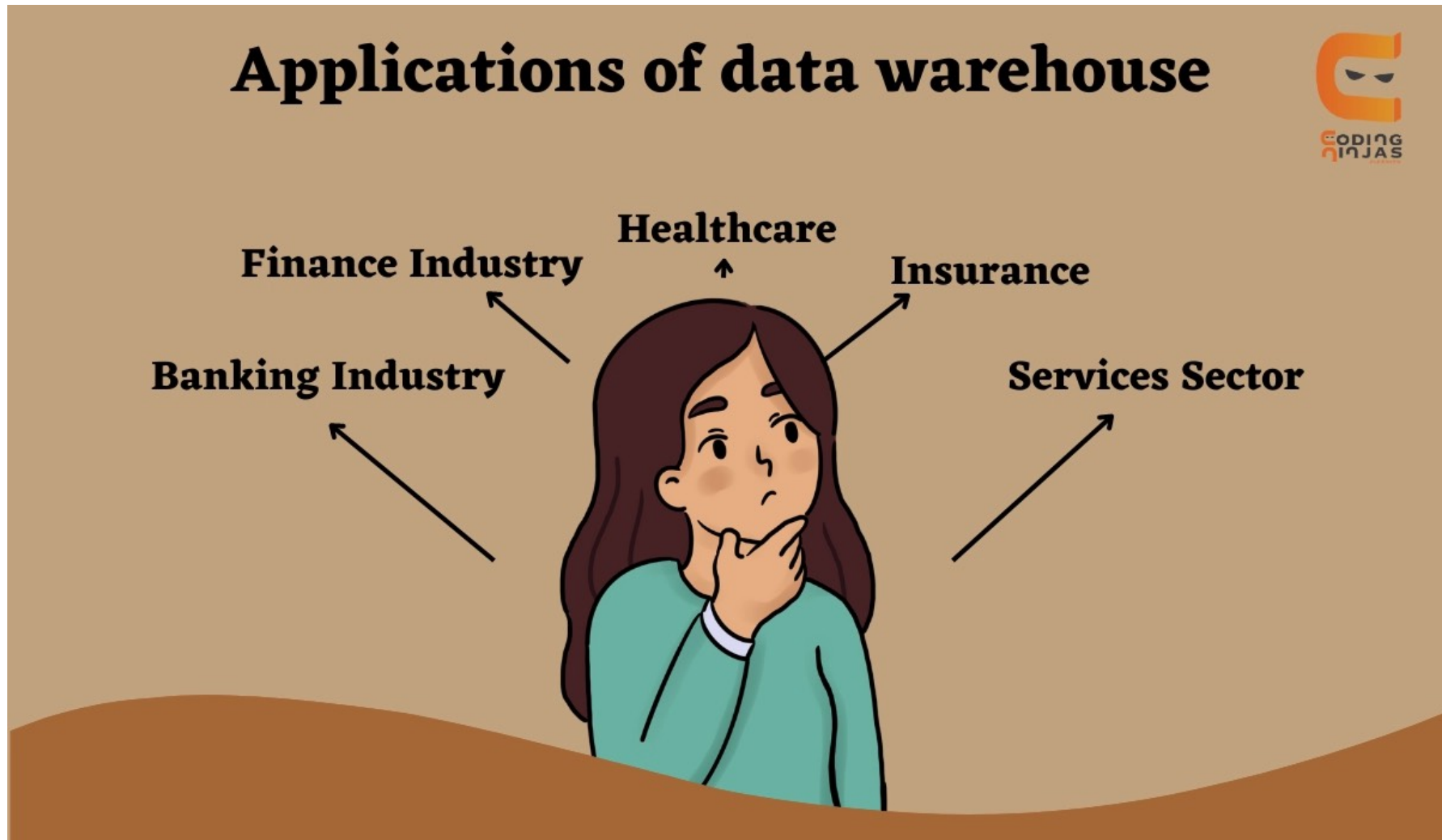
- Analyzing and identifying relationships between matched records and consolidating / merging them into correct representation.

Migration Dates		Sequence	Event	Date
2006-05-09		1	<u>Inmigration</u>	1995-06-06
1995-06-06		2	Outmigration	2006-05-09

DATA MART

- Data mart is a subset or view of a data warehouse, typically at a department or functional level, that contains all data required for decision support tasks of that department.

APPLICATION OF DATA WAREHOUSES



MULTIDIMENSIONAL DATA

- The multidimensional data model supports data representation and operations specifically tailored for decision support processing in data warehouses.

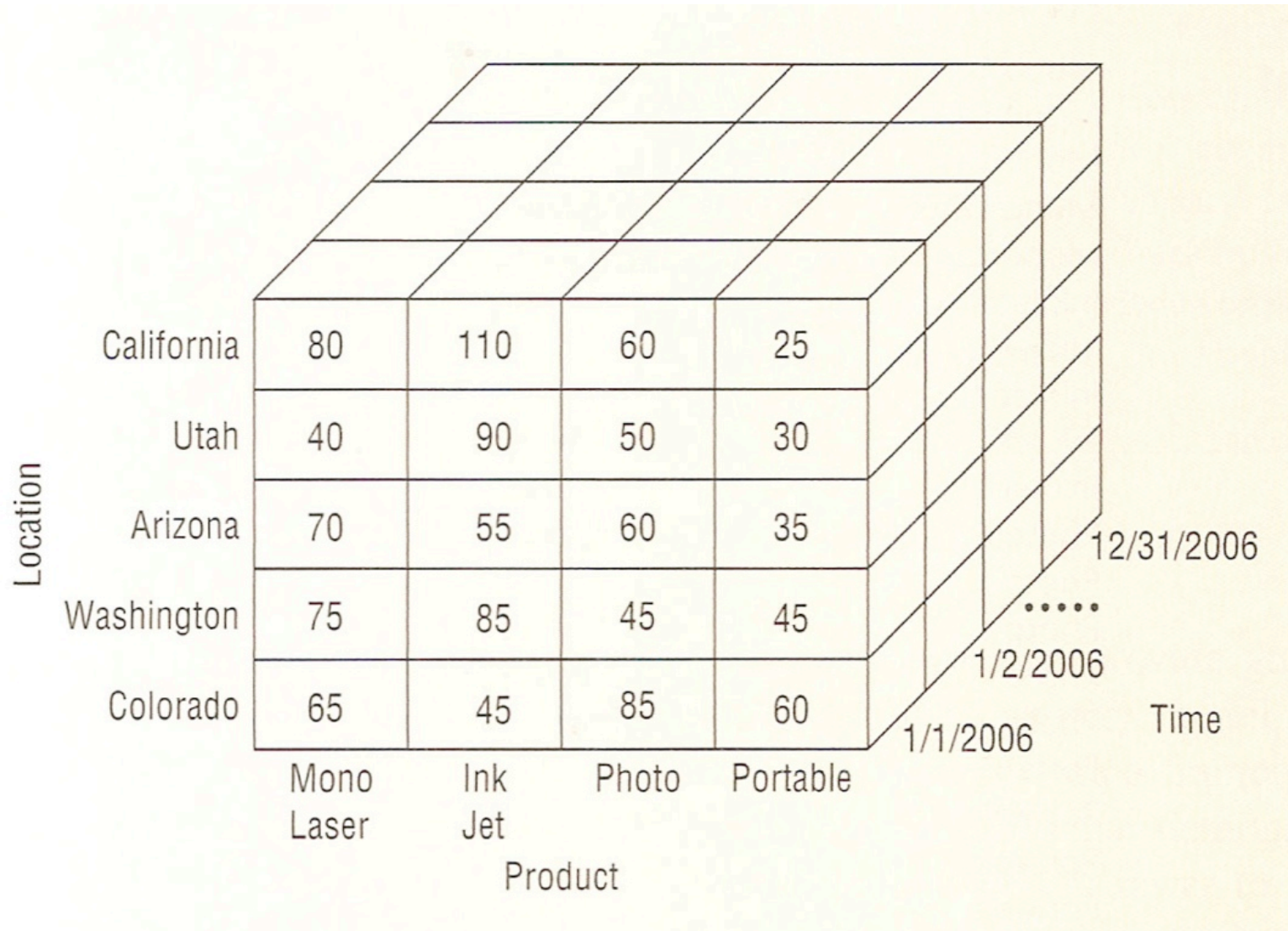
RELATIONAL REPRESENTATION OF SALES DATA

Product	Location	Sales
Mono Laser	California	80
Mono Laser	Utah	40
Mono Laser	Arizona	70
Mono Laser	Washington	75
Mono Laser	Colorado	65
Ink Jet	California	110
Ink Jet	Utah	90
Ink Jet	Arizona	55
Ink Jet	Washington	85
Ink Jet	Colorado	45
Photo	California	60
Photo	Utah	50
Photo	Arizona	60
Photo	Washington	45
Photo	Colorado	85
Portable	California	25
Portable	Utah	30
Portable	Arizona	35
Portable	Washington	45
Portable	Colorado	60

MULTIDIMENSIONAL REPRESENTATION OF SALES DATA

Location	Product			
	Mono Laser	Ink Jet	Photo	Portable
California	80	110	60	25
Utah	40	90	50	30
Arizona	70	55	60	35
Washington	75	85	45	45
Colorado	65	45	85	60

A THREE-DIMENSIONAL DATA CUBE



MULTIDIMENSIONAL REPRESENTATION OF SALES DATA WITH ROW TOTALS

Location	Product				Totals
	Mono Laser	Ink Jet	Photo	Portable	
California	80	110	60	25	275
Utah	40	90	50	30	210
Arizona	70	55	60	35	220
Washington	75	85	45	45	250
Colorado	65	45	85	60	255

DATA CUBE

- Data cube is a multidimensional format in which cells contain numeric data called measures organised by subjects called dimensions.
- A data cube is sometimes known as a hypercube because conceptually it can have an unlimited number of dimensions.

DATA CUBE OPERATIONS (1)

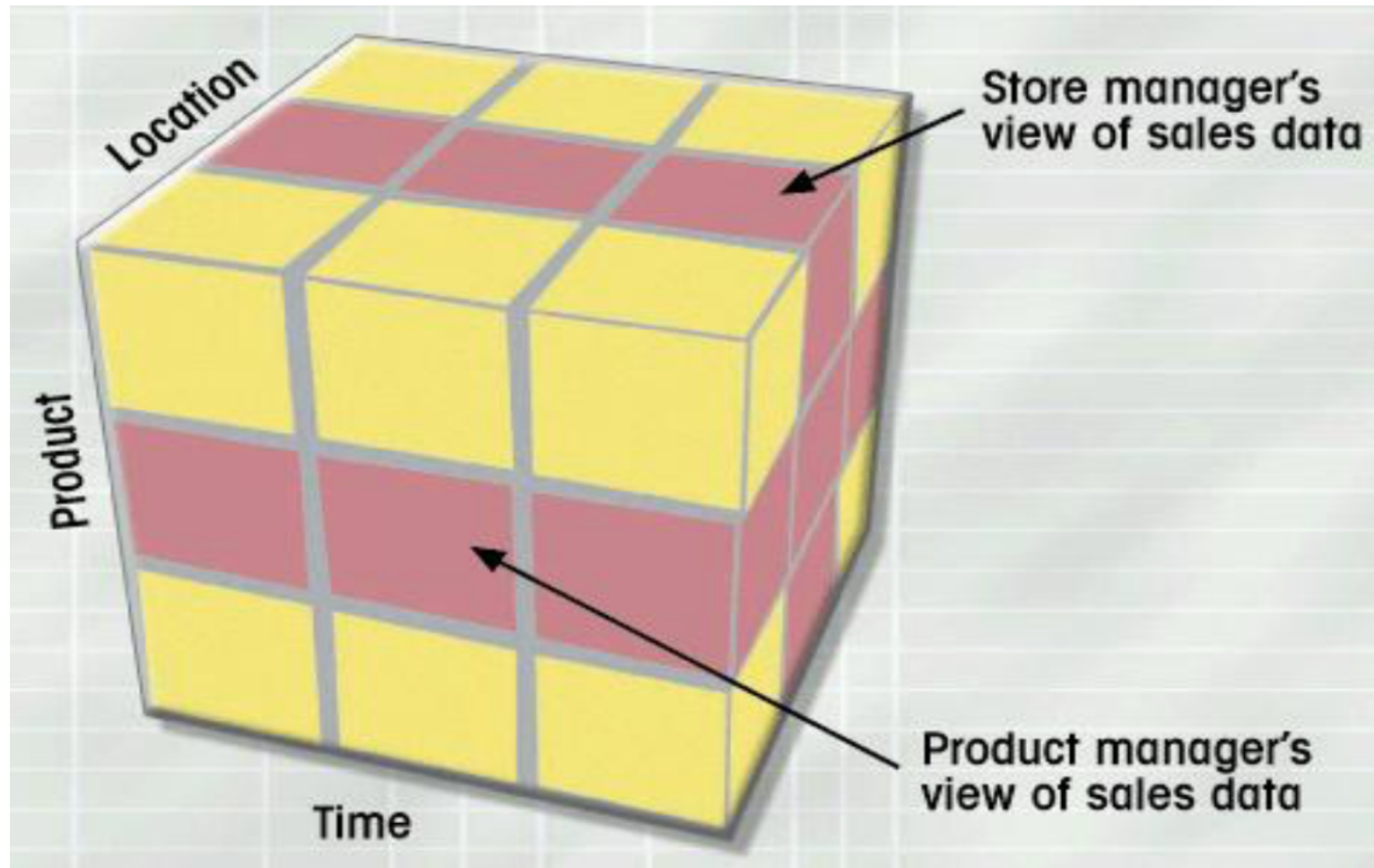
- **Slice**

Because a data cube can contain a large number of dimensions, users often need to focus on a subset of the dimensions to gain insights.

- **Dice**

Because individual dimensions can contain a large number of members, users need to focus on a subset of members to gain insights.

DATA CUBE OPERATIONS (2)



DATA CUBE OPERATIONS (3)

- **Drill-Down**

Users often want to navigate among the levels of hierarchical dimensions. The drill-down operator allows users to navigate from a more general level to a more specific level.

- **Roll-Up**

Roll-up (also called drill-up) is the opposite of drill-down. Roll-up involves moving from a specific level to a more general level of a hierarchical dimension.

- **Pivot**

The pivot operator supports rearrangement of the dimensions in a data cube.

EXAMPLE SLICE AND DICE OPERATION

Location	Product			
	Mono Laser	Ink Jet	Photo	Portable
California	80	110	60	25
Utah	40	90	50	30
Arizona	70	55	60	35
Washington	75	85	45	45
Colorado	65	45	85	60

Location	Time			Total Sales
	1/1/2006	1/2/2006	...	
California	400	670	...	16,250
Utah	340	190	...	11,107
Arizona	270	255	...	21,500
Washington	175	285	...	20,900
Colorado	165	245	...	21,336

Location	Utah	40	90	50	30
		Mono Laser	Ink Jet	Photo	Portable

DRILL-DOWN OPERATION FOR THE STATE OF UTAH

Location	Product			
	Mono Laser	Ink Jet	Photo	Portable
California + Utah	80	110	60	25
Salt Lake	20	20	10	15
Park City	5	30	10	5
Ogden	15	40	30	10
Arizona	70	55	60	35
Washington	75	85	45	45
Colorado	65	45	85	60

SUMMARY OF THE DATA CUBE OPERATIONS

Operator	Purpose	Description
Slice	Focus attention on a subset of dimensions	Replace a dimension with a single member value or with a summary of its measure values
Dice	Focus attention on a subset of member values	Replace a dimension with a subset of members
Drill-down	Obtain more detail about a dimension	Navigate from a more general level to a more specific level of a hierarchical dimension
Roll-up	Summarize details about a dimension	Navigate from a more specific level to a more general level of a hierarchical dimension
Pivot	Allow a data cube to be presented in a visually appealing order	Rearrange the dimensions in a data cube

BUSINESS INTELLIGENCE (1)

"BI results when organizational culture, business processes and technologies are designed and implemented with the goal of improving the strategic and operational decision-making capabilities of a wide range of internal and external stakeholders."

International Data Corporation (IDC)

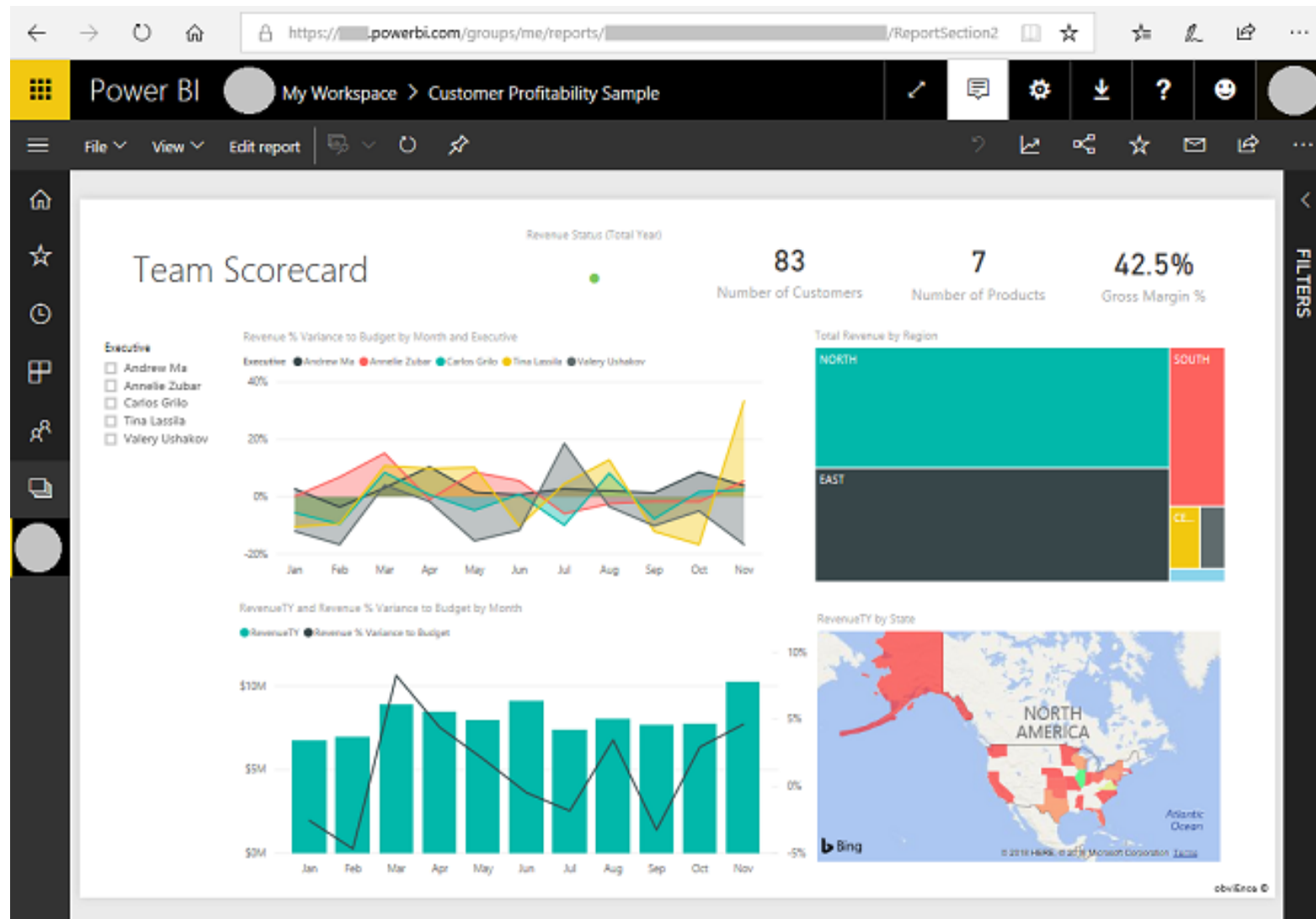
BUSINESS INTELLIGENCE (2)

Culture, processes and technologies to improve decision making for stakeholders.

WHAT IS POWER BI ?

- Power BI is a cloud-based business analysis and intelligence service by Microsoft. It is a collection of business intelligence and data visualization tools such as software services, apps and data connectors.
- We can use the datasets imported in Power BI for data visualization and analysis by making sharable reports, dashboards, and apps. Power BI is a user-friendly tool offering impressive drag-and-drop features and self-service capabilities.

MICROSOFT POWER BI



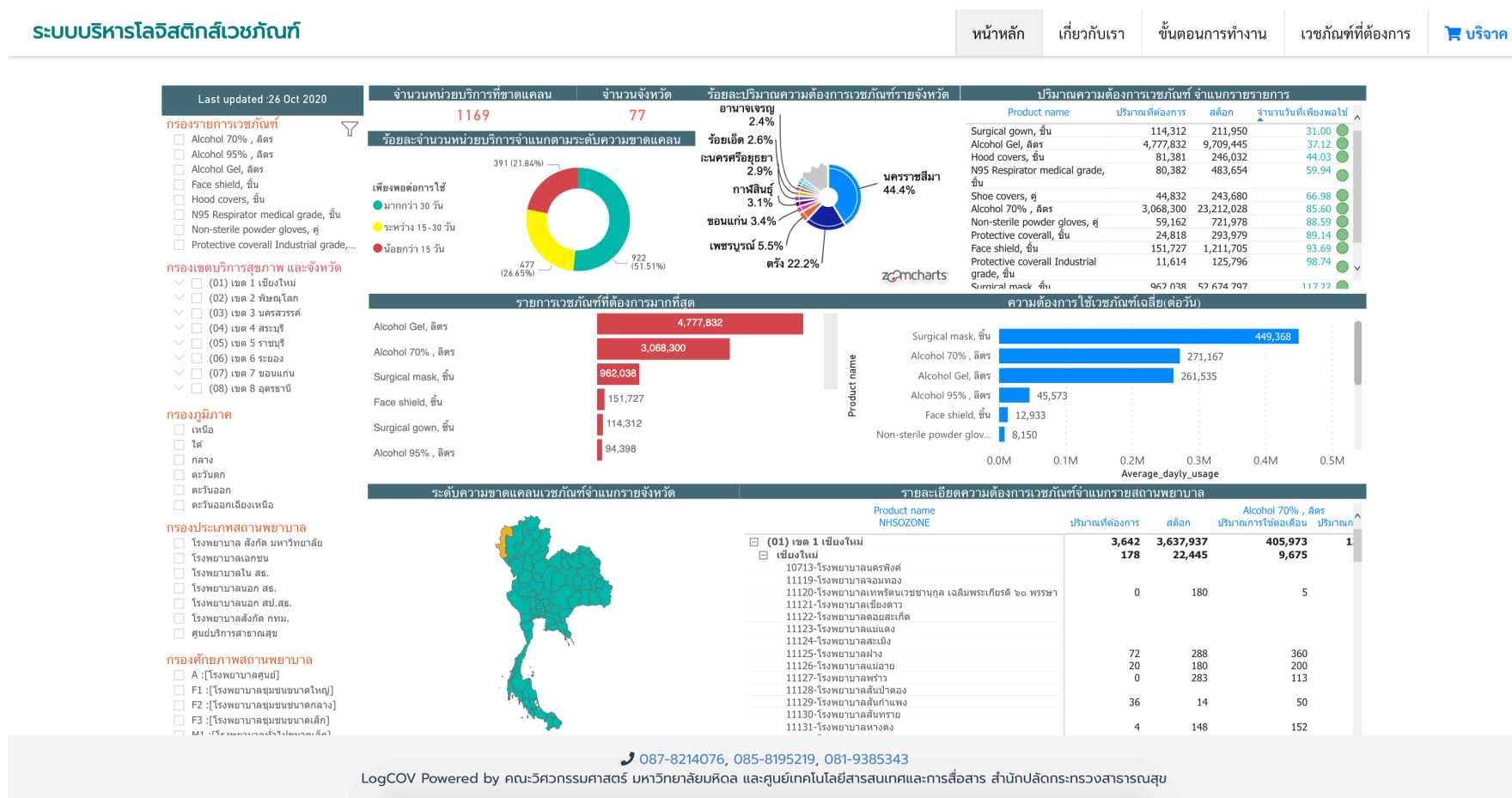
<https://powerbi.microsoft.com/en-au/getting-started-with-power-bi/>

TYPE OF POWER BI

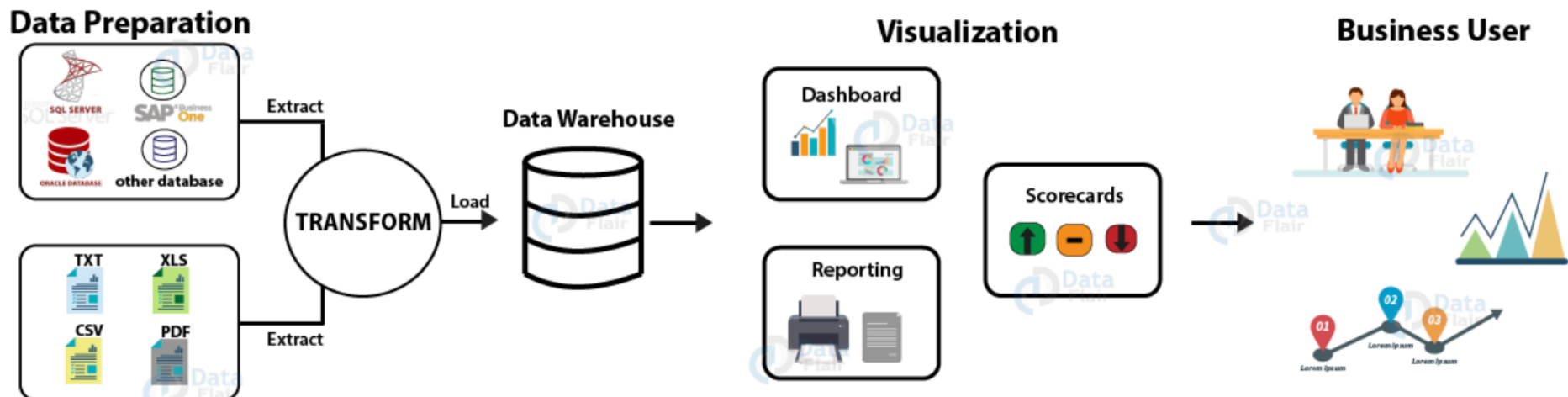
- Microsoft offers three types of Power BI platforms:
 - Power BI Desktop (Desktop Application)
 - Power BI Service (Software as a Service)
 - Power BI Mobile (For iOS and Android devices)

EXAMPLE OF BI

- SmartMedSupply for COVID-19
- <http://smartmedsupply.moph.go.th/newhomepage>

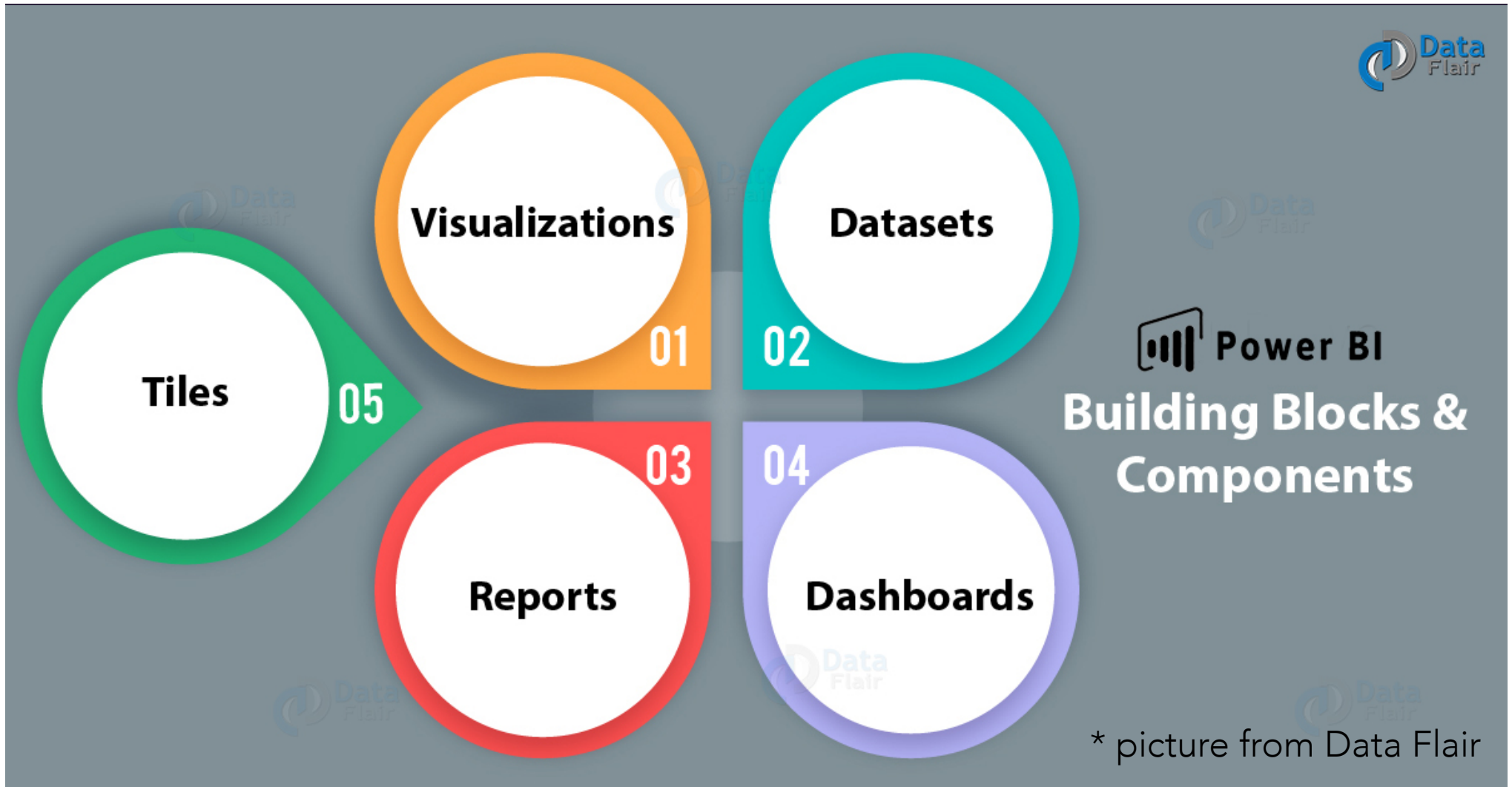


POWER BI ARCHITECTURE

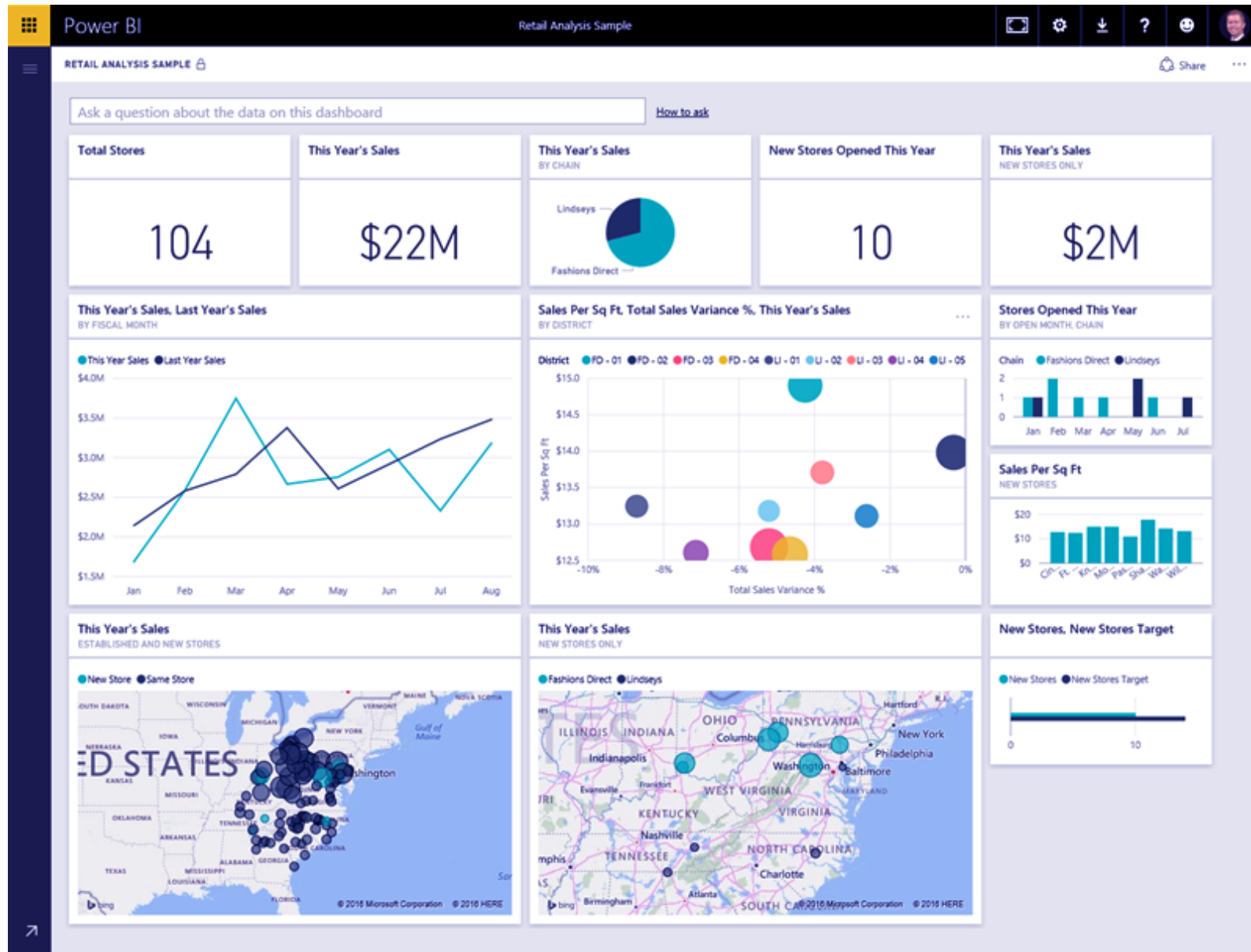


* picture from Data Flair

POWER BI BUILDING BLOCKS



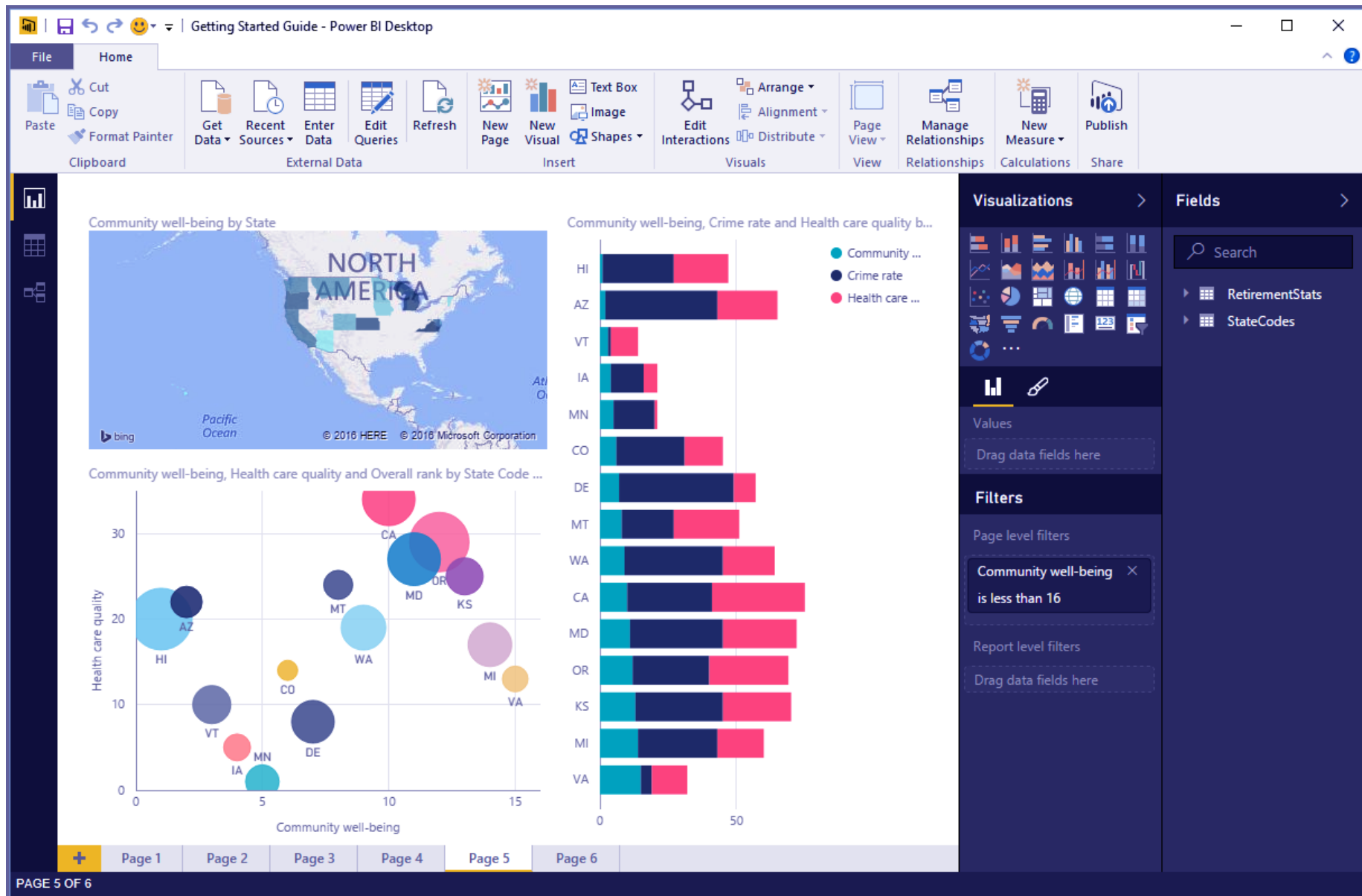
VISUALISATION



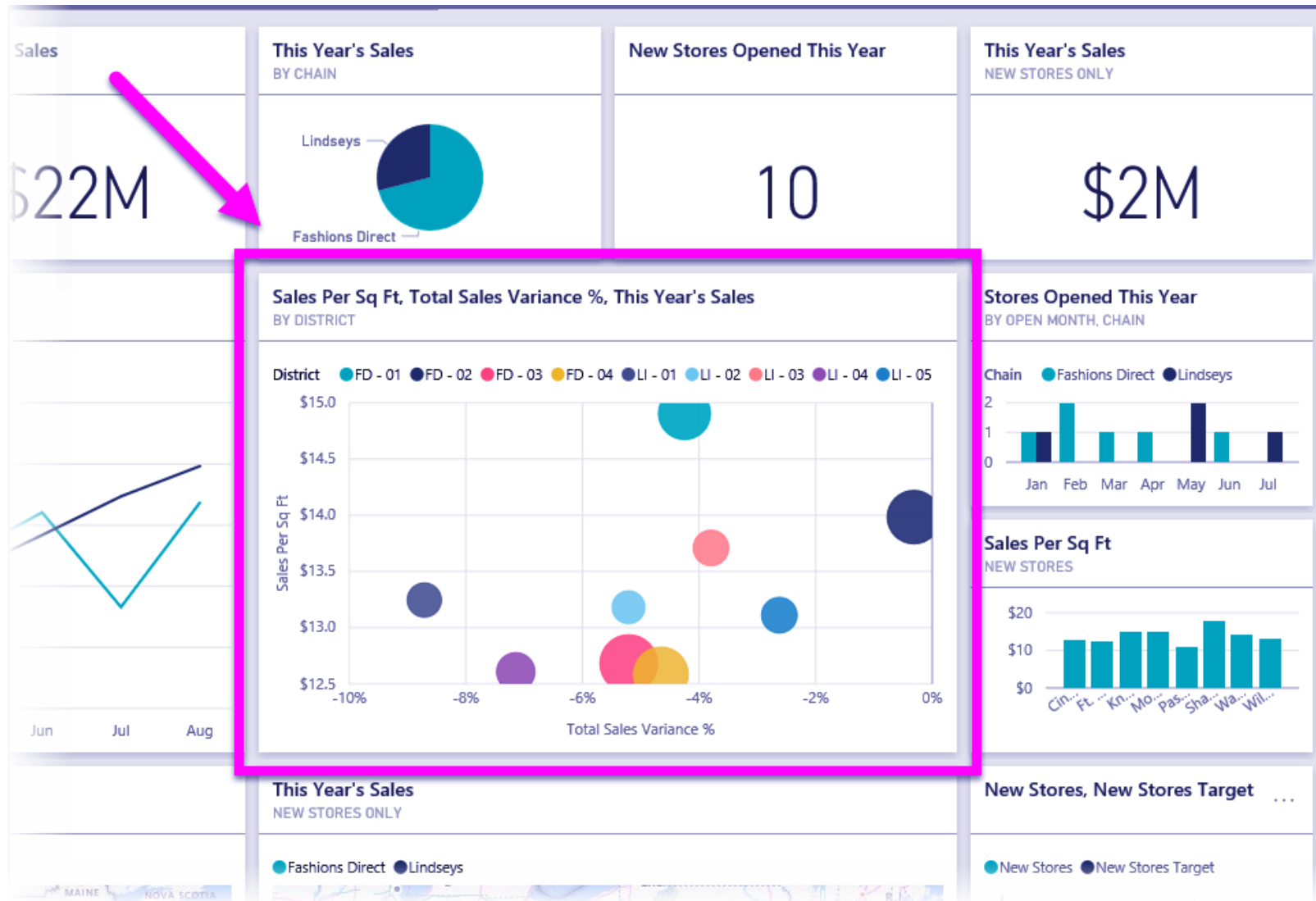
DATASETS

C2132								2
	B	C	D	E	F	G	H	
1	Year	Month	Month Name	Calendar Month	Births	Births Per Day	Births (Normalized)	
2119	2004	1	January	1/1/2004	2,937	94.7	2842	
2120	2004	2	February	2/1/2004	2,824	97.4	2921	
2121	2004	3	March	3/1/2004	3,128	100.9	3027	
2122	2004	4	April	4/1/2004	2,896	96.5	2896	
2123	2004	5	May	5/1/2004	3,008	97.0	2911	
2124	2004	6	June	6/1/2004	3,047	101.6	3047	
2125	2004	7	July	7/1/2004	2,981	96.2	2885	
2126	2004	8	August	8/1/2004	3,079	99.3	2980	
2127	2004	9	September	9/1/2004	3,219	107.3	3219	
2128	2004	10	October	10/1/2004	3,547	114.4	3433	
2129	2004	11	November	11/1/2004	3,365	112.2	3365	
2130	2004	12	December	12/1/2004	3,143	101.4	3042	
2131	2005	1	January	1/1/2005	2,921	94.2	2827	
2132	2005	2	February	2/1/2005	2,699	96.4	2892	
2133	2005	3	March	3/1/2005	3,024	97.5	2926	
2134	2005	4	April	4/1/2005	3,037	101.2	3037	
2135	2005	5	May	5/1/2005	3,231	104.2	3127	
2136	2005	6	June	6/1/2005	3,163	105.4	3163	
2137	2005	7	July	7/1/2005	3,119	100.6	3018	
2138	2005	8	August	8/1/2005	3,156	101.8	3054	
2139	2005	9	September	9/1/2005	3,439	114.6	3439	

REPORTS



TILES



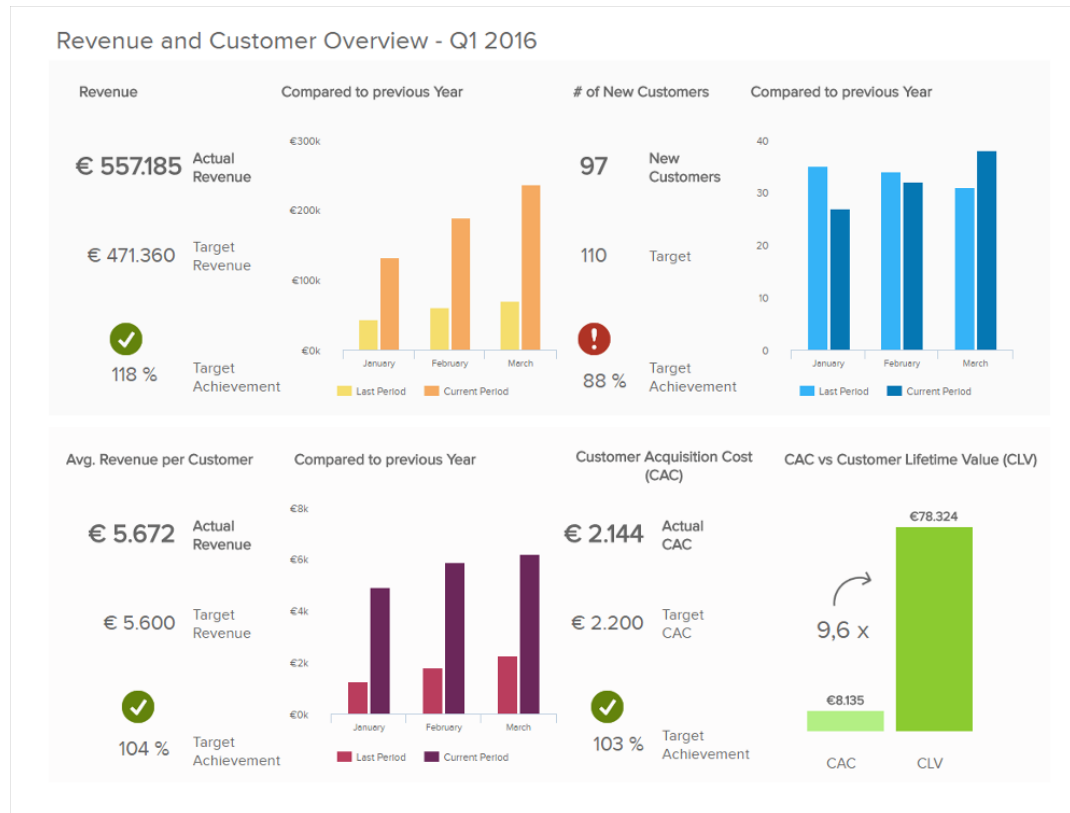
DASHBOARDS (1)

- Power BI dashboard is a gathering of visuals from a solitary page that you can impart to others. Frequently, it's a chosen gathering of visuals that give snappy understanding into the information or story you are attempting to exhibit.
- Dashboard needs to fit on a solitary page, frequently called a canvas. Consider it like the canvas that a craftsman or painter utilizes.
- Workspace where you make, consolidate, and adjust fascinating and convincing visuals. You can impart dashboards to different clients or gatherings, who would then be able to communicate with your dashboard when they're in Power BI benefit, or on their cell phone.

DASHBOARDS (2)

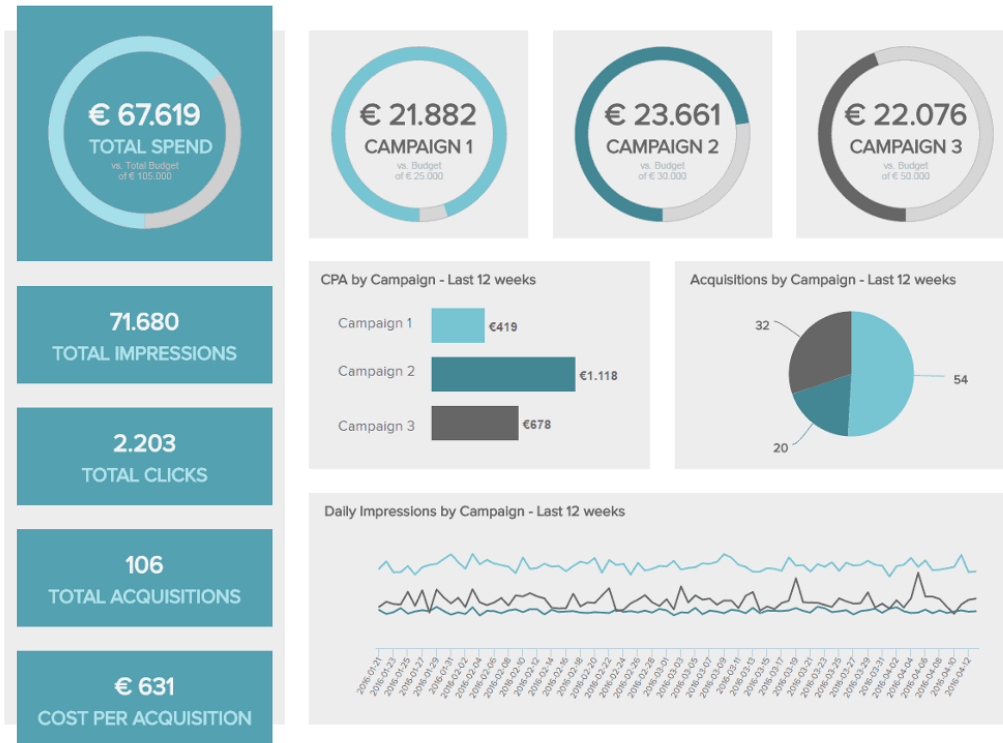
- There are several types of dashboards. Based on their purpose, they can be categorized into three main types:
 - Strategic Dashboard
 - Operational Dashboard
 - Analytical Dashboard

STRATEGIC DASHBOARD



- Dashboard for strategic planning
- Decision making at all levels.
- Focuses on displaying key organizational indicators or KPIs, making it easy to understand and see the organization's overall performance.

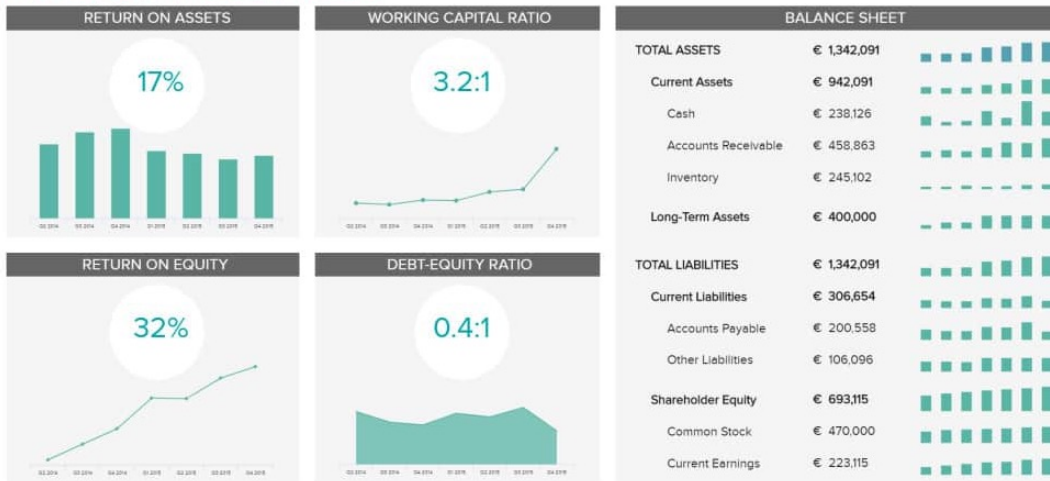
OPERATIONAL DASHBOARD



- Dashboard for management and team lead level.
- Monitors and controls task progress.
- Displays a current overview of an area or product line, and shows real-time results when issues arise.
- Typically, detailed data drilling is not required.

ANALYTICAL DASHBOARD

FINANCIAL PERFORMANCE DASHBOARD



BALANCE SHEET		
TOTAL ASSETS	€ 1,342,091	
Current Assets	€ 942,091	
Cash	€ 238,126	
Accounts Receivable	€ 458,863	
Inventory	€ 245,102	
Long-Term Assets	€ 400,000	
TOTAL LIABILITIES	€ 1,342,091	
Current Liabilities	€ 306,654	
Accounts Payable	€ 200,558	
Other Liabilities	€ 106,096	
Shareholder Equity	€ 693,115	
Common Stock	€ 470,000	
Current Earnings	€ 223,115	

- Dashboard for analysis.
- Provides data from multiple perspectives, often without specific indicators.
- It includes a variety of data, such as historical data that can be filtered and compared, and emphasizes viewing the data from different analytical perspectives

SCORECARD

- A Scorecard in Power BI is a tool within the Power BI Service that allows organizations to track and measure key performance indicators (KPIs) and goals.
- Using the Power BI Goals feature, which includes scorecards, users can create interactive visualizations of organizational metrics, set targets, and monitor progress toward specific business objectives over time.
- Scorecards help teams focus on outcomes, provide insights into performance gaps, and drive data-informed decision-making.

POWER BI PROCESS

- Connect to Data Source (Get Data)
- Clean Up (ETL: Power Query)
- Import
- Report (Visualization)
- Publish & Share

GETTING DATA

- Power BI can connect to wide variety of data sources, including on-premises databases, Azure storage, Excel worksheets and a large number of 3rd party services.
 - Clean and transform data with the Query Editor
 - Connect to advanced data sources and create transformations
 - “Massage” irregularly formatted data

POWER QUERY (1)

- Power Query is a data transformation and preparation tool (ETL) available in Microsoft Excel and Power BI.
- It allows users to connect to various data sources, clean, transform, and reshape data before analysis.
- Power Query is designed to automate repetitive tasks, such as data extraction, filtering, merging, and sorting, without needing advanced programming skills.
- In Excel, go to the Data tab and select "Get Data" (Power Query) to open Power Query.

POWER QUERY (2)

Power Query Editor

Home Transform Add column View Help

Close & load Get data Enter data Options Manage parameters Refresh Properties Advanced editor Choose columns Remove columns Keep rows Remove rows Filter rows Sort Transform Combine

Queries [1] country_vaccinations_b...

Table.TransformColumnTypes("#Promoted headers", {{"location", type

	locat...	d...	vaccine	total_vaccinati...
1	Argentina	12/29/20...	Moderna	2
2	Argentina	12/29/20...	Oxford/AstraZen...	3
3	Argentina	12/29/20...	Sinopharm/Beijing	1
4	Argentina	12/29/20...	Sputnik V	20481
5	Argentina	12/30/20...	Moderna	2
6	Argentina	12/30/20...	Oxford/AstraZen...	3
7	Argentina	12/30/20...	Sinopharm/Beijing	1
8	Argentina	12/30/20...	Sputnik V	40583
9	Argentina	12/31/2020	Moderna	2
10	Argentina	12/31/2020	Oxford/AstraZen...	3
11	Argentina	12/31/2020	Sinopharm/Beijing	1
12	Argentina	12/31/2020	Sputnik V	43388
13	Argentina	1/1/2021	Moderna	2
14	Argentina	1/1/2021	Oxford/AstraZen...	5
15	Argentina	1/1/2021	Sinopharm/Beijing	1
16	Argentina	1/1/2021	Sputnik V	43513
17	Argentina	1/2/2021	Moderna	2
18	Argentina	1/2/2021	Oxford/AstraZen...	6
19	Argentina	1/2/2021	Sinopharm/Beijing	1
20	Argentina	1/2/2021	Sputnik V	46824
21	Argentina	1/3/2021	Moderna	2
22	Argentina	1/3/2021	Oxford/AstraZen...	6

Query settings

Properties

Name

country_vaccinations_...

Applied steps

Source Promoted head... Changed column...

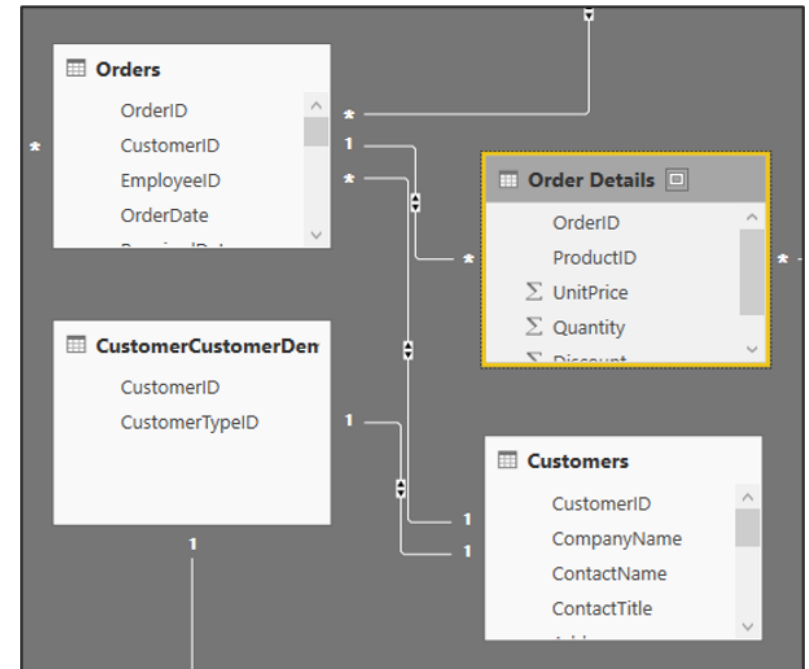
Completed (0.13 s) Columns: 4 Rows: 99+

ONELAKE DATA HUB

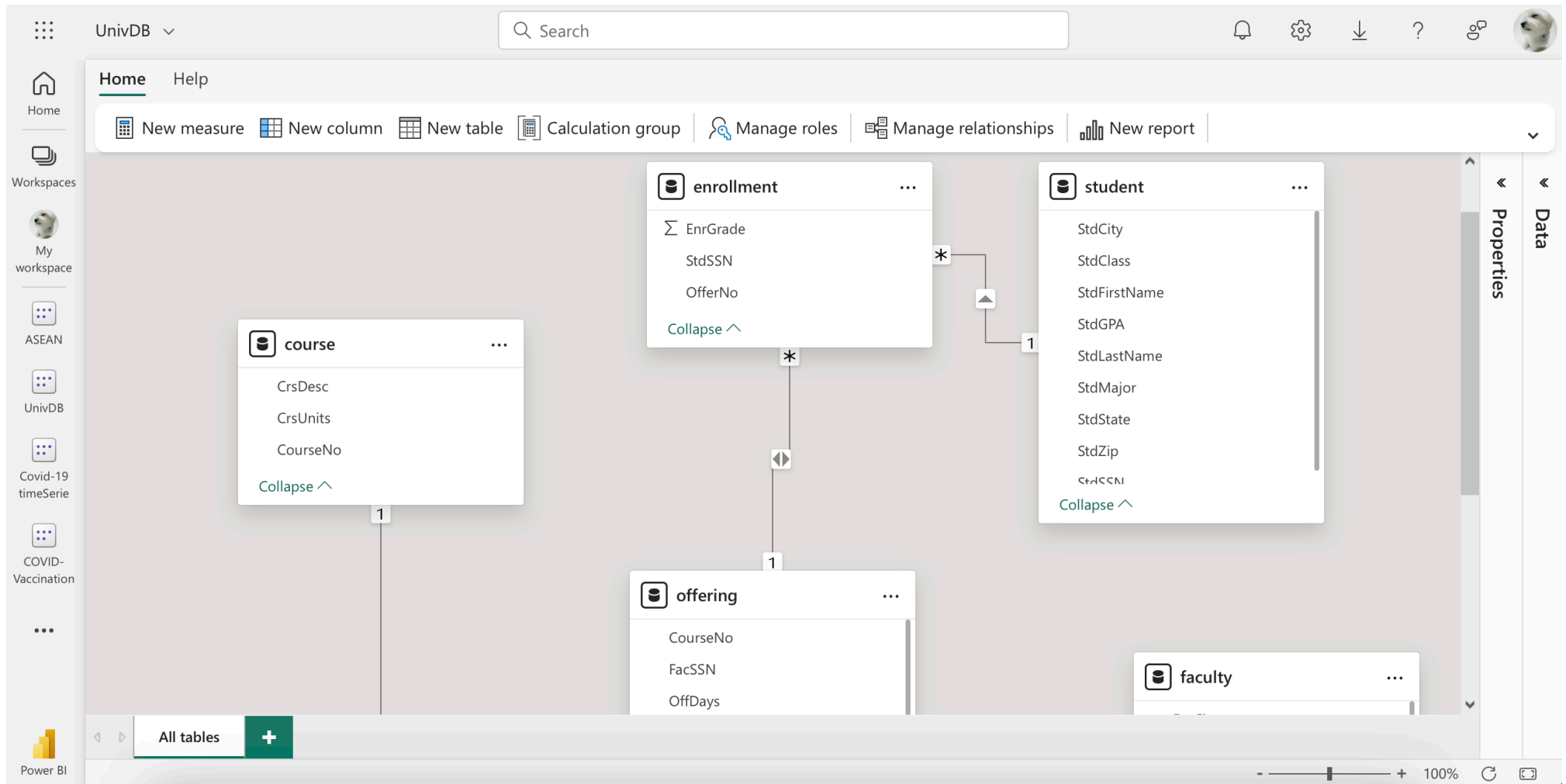
- Data Lake in Power BI typically refers to integrating Power BI with a data lake storage, commonly using Azure Data Lake Storage (ADLS).
- This allows users to centralize large datasets from multiple sources into a single, scalable storage.
- By connecting Power BI with a data lake, users gain robust capabilities to explore, transform, and visualize massive datasets without the need to import all data directly into Power BI.

DATA MODELING (1)

- Modeling is the technique of creating a logical connections and relationships between data sources.
- Create Calculated Columns
- Optimize data models for better visuals
- Create Measures and work with time-based functions
- Create Calculated Tables

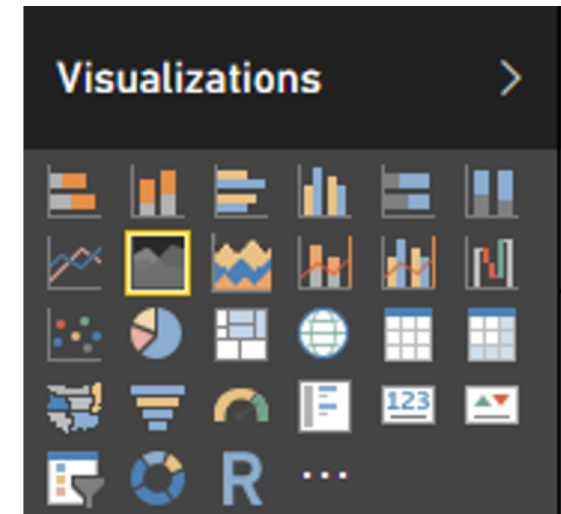


DATA MODELING (2)



VISUALISATION

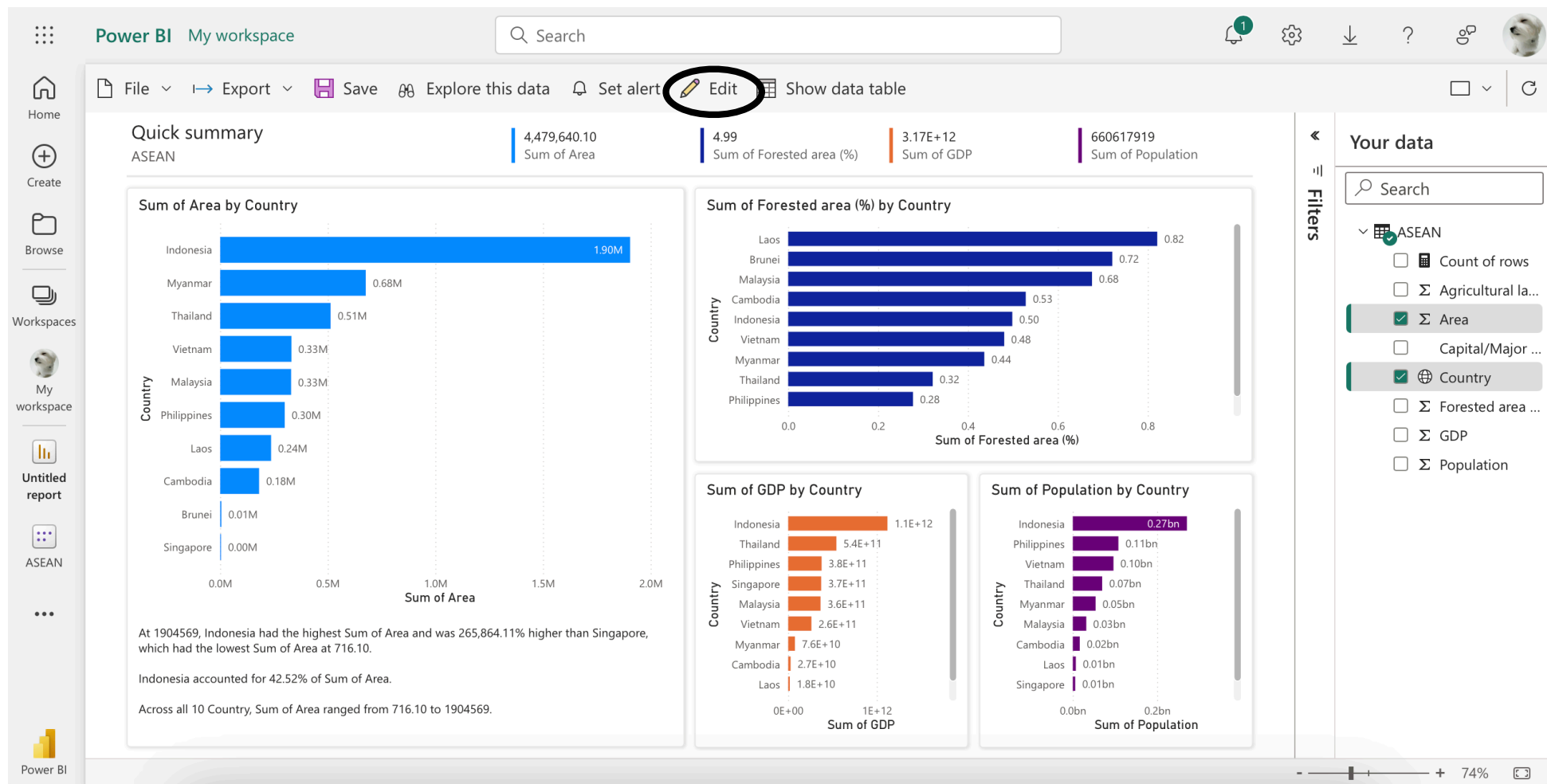
- A visualization is a visual representation of data, such as a chart, graph, map, or other “visual” representations of data.
 - Line, bar, pie, stacked
 - Matrix/pivots
 - Key performance
 - Tree maps
 - Geo and filled maps
 - Slicers/filters
 - Custom visualizations



DATA ANALYSIS EXPRESSIONS

- DAX is a collection of functions, operators, and constants that can be used in a formula, or expression, to calculate and return one or more values.
 - **Measure**
Calculation on a collection of rows AFTER aggregation.
 - **Calculated Column**
Calculation on each individual row BEFORE aggregation.

VISUALIZE DATA



CREATE PIVOT (MATRIX) TABLE

OneLake data hub



Pick a semantic model to use in your report

All

My data

Endorsed in your org

Filter by keyword

Explorer >>	Name	Owner	Refreshed	Location	Enc
	ASEAN	Kanat Poolsawasd	24/10/21, 20:46:09	My Workspace	—
	UnivDB	Kanat Poolsawasd	05/11/24, 23:13:24	My Workspace	—
	country_vaccinations_by_manufact...	Kanat Poolsawasd	04/11/24, 15:57:28	My Workspace	—
	SalesData	Kanat Poolsawasd	05/11/23, 18:48:30	My Workspace	—
	Covid-19 timeSerie	SUPHAVIT NORKAEW	22/04/20, 07:19:40	BI project	—
	COVID-Vaccination	Kanat Poolsawasd	15/11/23, 08:52:12	My Workspace	—
	Population_Example	Kanat Poolsawasd	21/11/21, 22:15:04	My Workspace	—
	ASEAN	Kanat Poolsawasd	05/11/23, 14:10		—

Auto-create report

Create a blank report

Auto-create report



Cancel

CREATE PIVOT (MATRIX) TABLE

The screenshot displays the Microsoft Power BI workspace interface. The top navigation bar includes the 'Power BI My workspace' header, a search bar, and various utility icons. The left-hand navigation pane lists options such as Home, Create, Browse, OneLake data hub, Apps, Metrics, Workspaces, My workspace, and SalesData. The central canvas area is titled 'Build visuals with your data' and prompts the user to 'Select or drag fields from the Data pane onto the report canvas.' The right-hand pane is divided into three sections: 'Filters', 'Visualizations', and 'Data'. The 'Visualizations' section is active, showing a grid of visualization icons. A black circle highlights the 'Pivot table' icon, which is a grid with a small square in the top-left corner. Below the visualization icons are sections for 'Values' (with an 'Add data fields here' box), 'Drill through' (with 'Cross-report' set to 'Off' and 'Keep all filters' set to 'On'), and another 'Add drill-through fields here' box. The 'Data' pane on the far right shows a search bar and a list containing 'SalesData'. The bottom status bar indicates 'Page 1' and includes a green plus icon for adding new pages.

CREATE PIVOT (MATRIX) TABLE

The screenshot displays the Microsoft Power BI interface. The top navigation bar includes the Power BI logo, 'My workspace', a search bar, and various utility icons. The left sidebar contains navigation options: Home, Create, Browse, OneLake data hub, Workspaces, My workspace, SalesData, and a Power BI icon at the bottom.

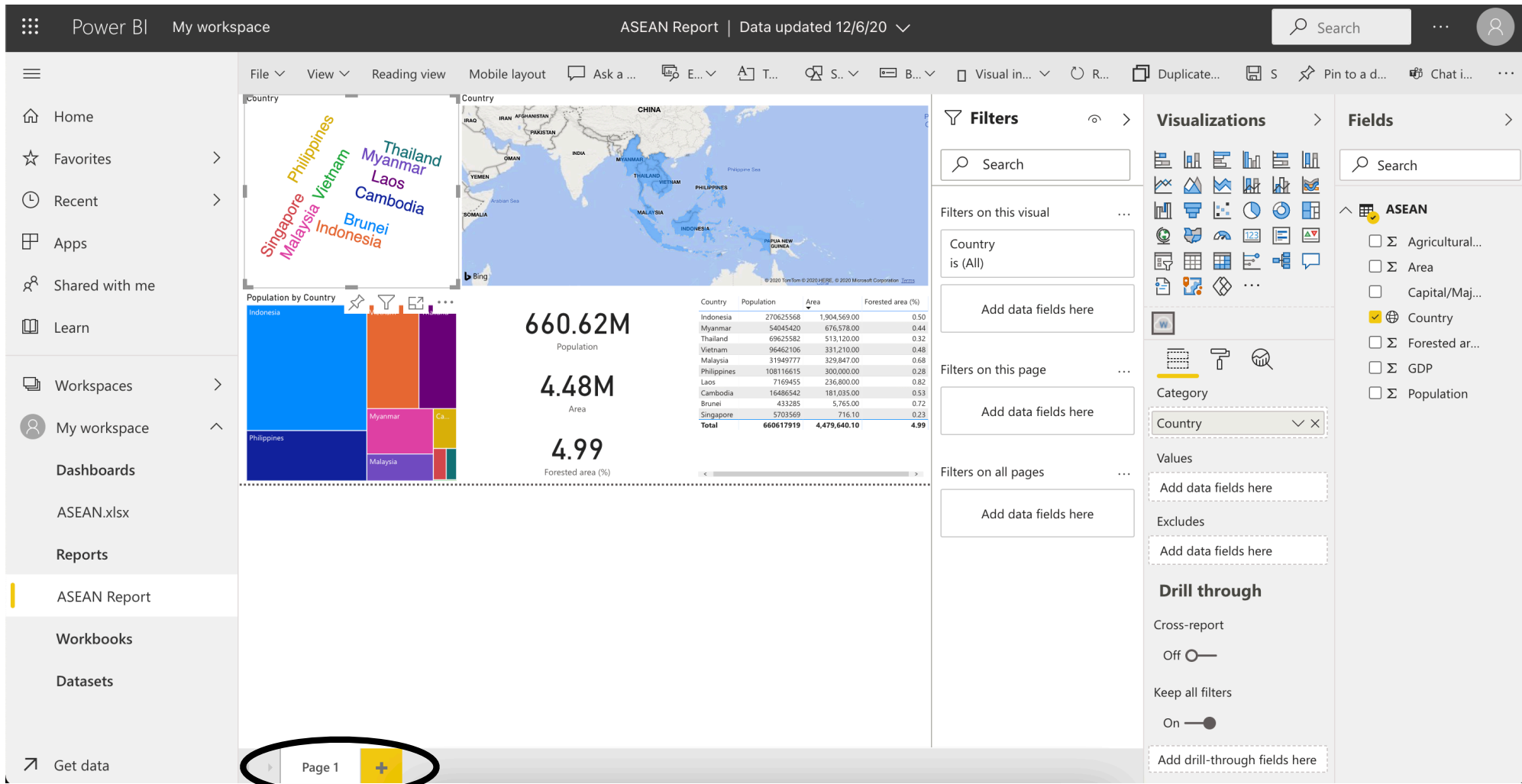
The main content area is divided into four panes:

- Table View:** Displays a Pivot table with 'Product' as the row header and 'Sum of Sales' as the column header. The data is as follows:

Product	Sum of Sales
Ink Jet	385
Arizona	55
California	110
Colorado	45
Utah	90
Washington	85
Mono Laser	330
Photo	300
Portable	195
Total	1210

- Filters Pane:** Contains a search bar and sections for 'Filters on this visual', 'Filters on this page', and 'Filters on all pages'. Each section has a search bar and a list of filterable fields (Location, Product, Sum of Sales) with a dropdown menu set to 'is (All)'. There are also 'Add data fields here' buttons.
- Visualizations Pane:** Features a 'Build visual' section with a grid of visualization icons. Below this are sections for 'Rows' (containing 'Product' and 'Location' dropdowns), 'Columns' (with an 'Add data fields here' button), and 'Values'.
- Data Pane:** Shows the 'SalesData' table with columns 'Location', 'Product', and 'Sales' all checked.

POWER BI - REPORT



PUBLISHING & SHARING (1)

- Publish reports from Power BI Desktop to the Power BI Service
- Print and export dashboards
- Manually republish and refresh data
- Create and connect to content packs
- Integrate OneDrive for Business

PUBLISHING & SHARING (2)

Success! Your report is ready to share

Link you can send in email

<https://app.powerbi.com/view?r=eyJrljoiYz>

Copy

HTML you can paste into a website

`<iframe title="ASEAN" width="600" height="`

Copy

Size 600 x 373.5 px

Placeholder image

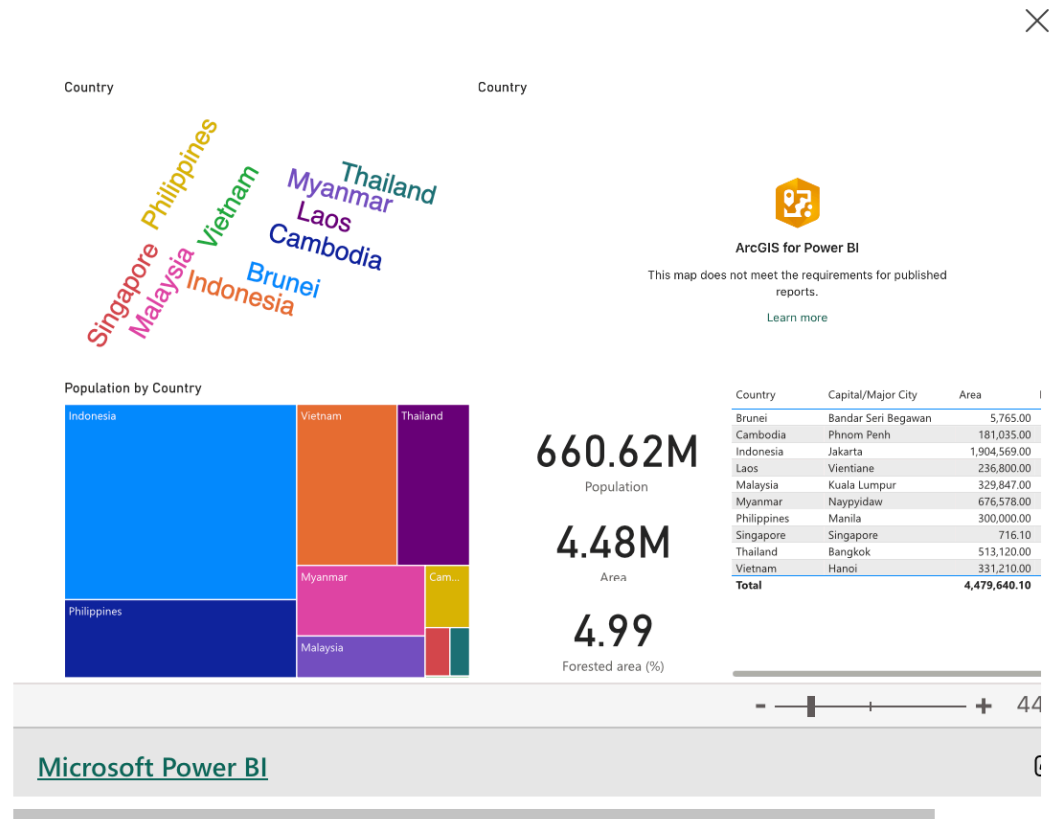


Upload

Delete

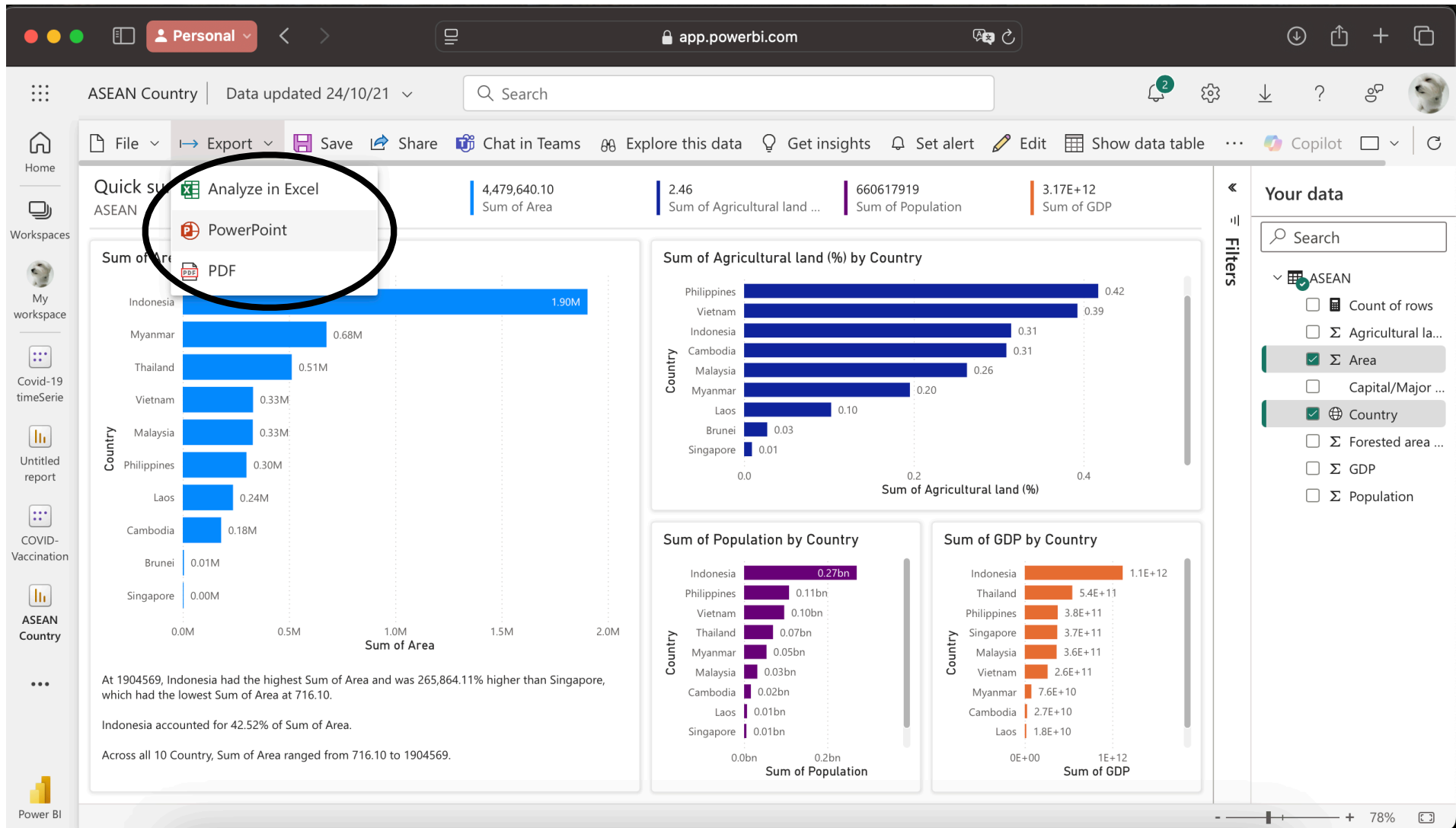
Default Page

Default



Close

EXPORT REPORT (1)



EXPORT REPORT (2)

Export to PowerPoint

Choose how to export:

Embed live data

Copy the report URL and paste it into your existing presentation or select Open in PowerPoint to create a new presentation.

☒ Embed report with the data filters you selected

Report page link:

<https://app.powerbi.com/groups/me/reports/a447fcda-6295-4...> **Copy**

Open in PowerPoint **Cancel**

Quick summary

ASEAN

Metric	Value
Sum of Area	4,479,640.10
Sum of Agricultural land ...	2.46
Sum of Population	660617919
Sum of GDP	3.17E+12

Sum of Area by Country

Country	Sum of Area
Indonesia	0.68M
Myanmar	0.51M
Thailand	0.33M
Vietnam	0.33M
Malaysia	0.30M
Philippines	0.24M
Laos	0.18M
Cambodia	0.01M
Brunei	0.00M
Singapore	0.00M

At 1904569, Indonesia had the highest Sum of Area and was 2.46bn higher than Singapore, which had the lowest Sum of Area at 716.10.

Indonesia accounted for 42.52% of Sum of Area.

Across all 10 Country, Sum of Area ranged from 716.10 to 1904569.

Your data

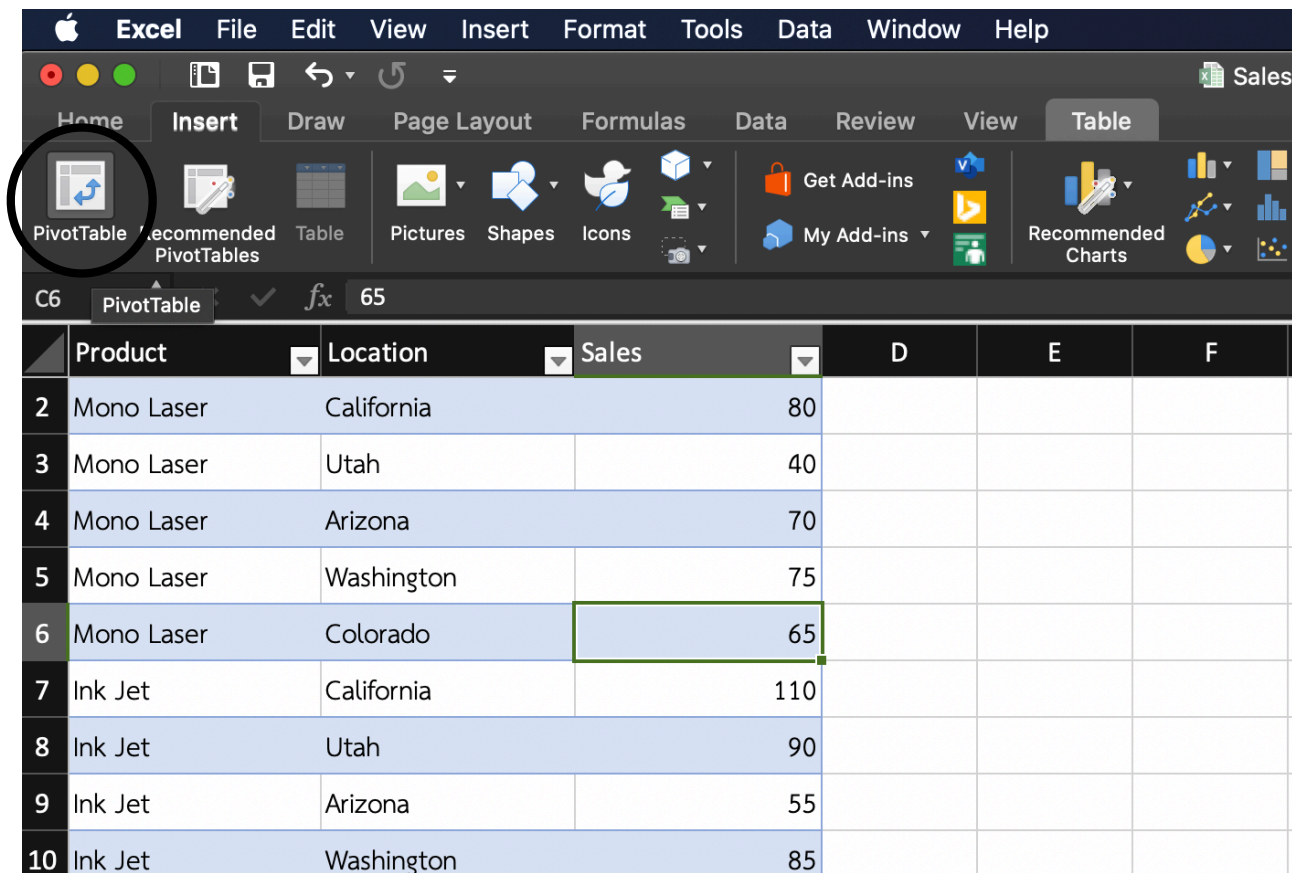
- ASEAN
 - ☐ Count of rows
 - ☐ Σ Agricultural la...
 - ☒ Σ Area
 - ☐ Capital/Major ...
 - ☒ Country
 - ☐ Σ Forested area ...
 - ☐ Σ GDP
 - ☐ Σ Population

PIVOT TABLES IN EXCEL (1)

- Pivot tables are one of Excel's most powerful features. A pivot table allows you to extract the significance from a large, detailed data set.

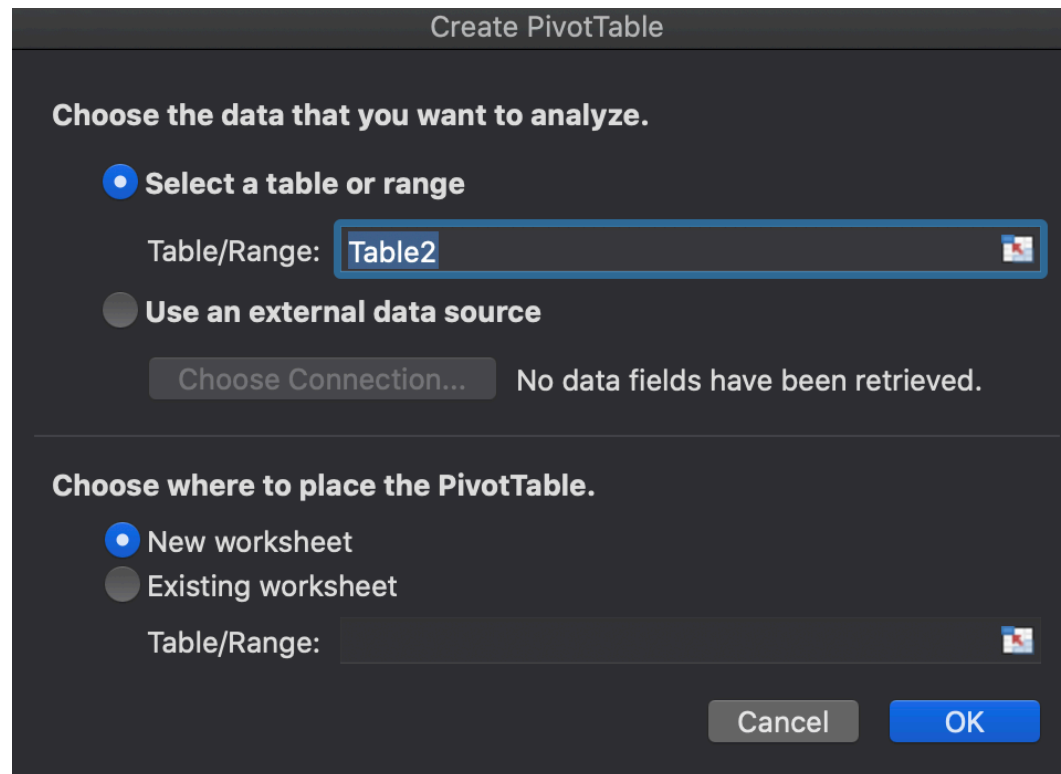
PIVOT TABLES IN EXCEL (2)

- Click any single cell inside the data set.
- On the Insert tab, in the Tables group, click PivotTable.



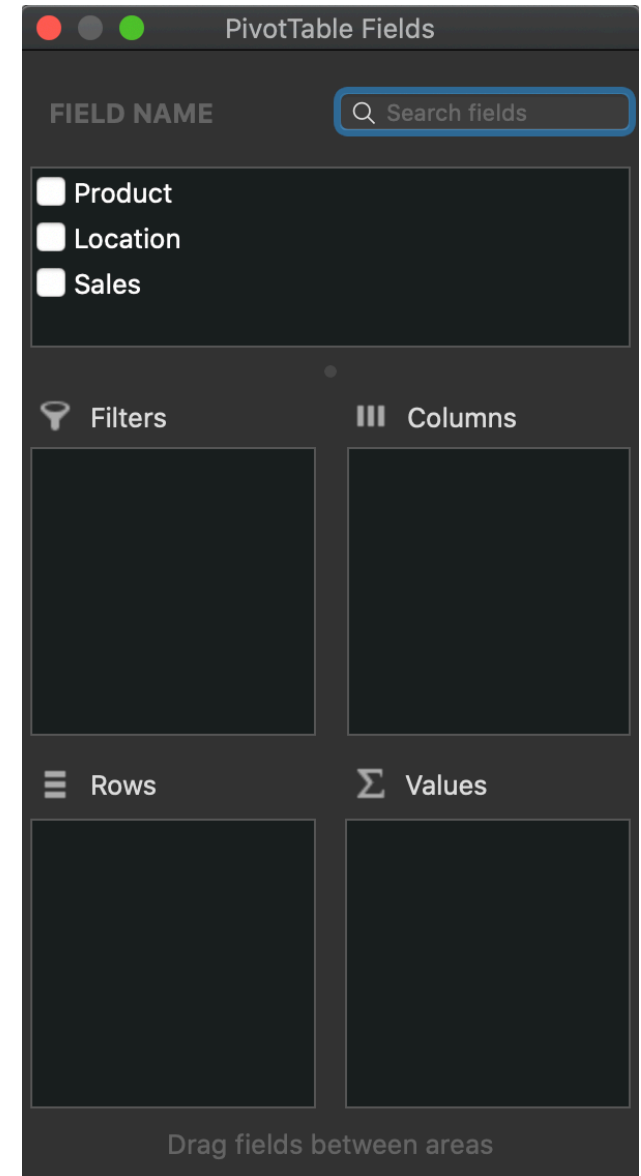
PIVOT TABLES IN EXCEL (3)

- The following dialog box appears. Excel automatically selects the data for you. The default location for a new pivot table is New Worksheet, and click OK



PIVOT TABLES IN EXCEL (4)

- The PivotTable Fields pane appears. To get the total amount exported of each product, drag the following fields to the different areas.



PIVOT TABLES IN EXCEL (5)

The screenshot displays the Microsoft Excel interface with a PivotTable and the PivotTable Fields task pane. The PivotTable is located in the range G3:G9, with the following data:

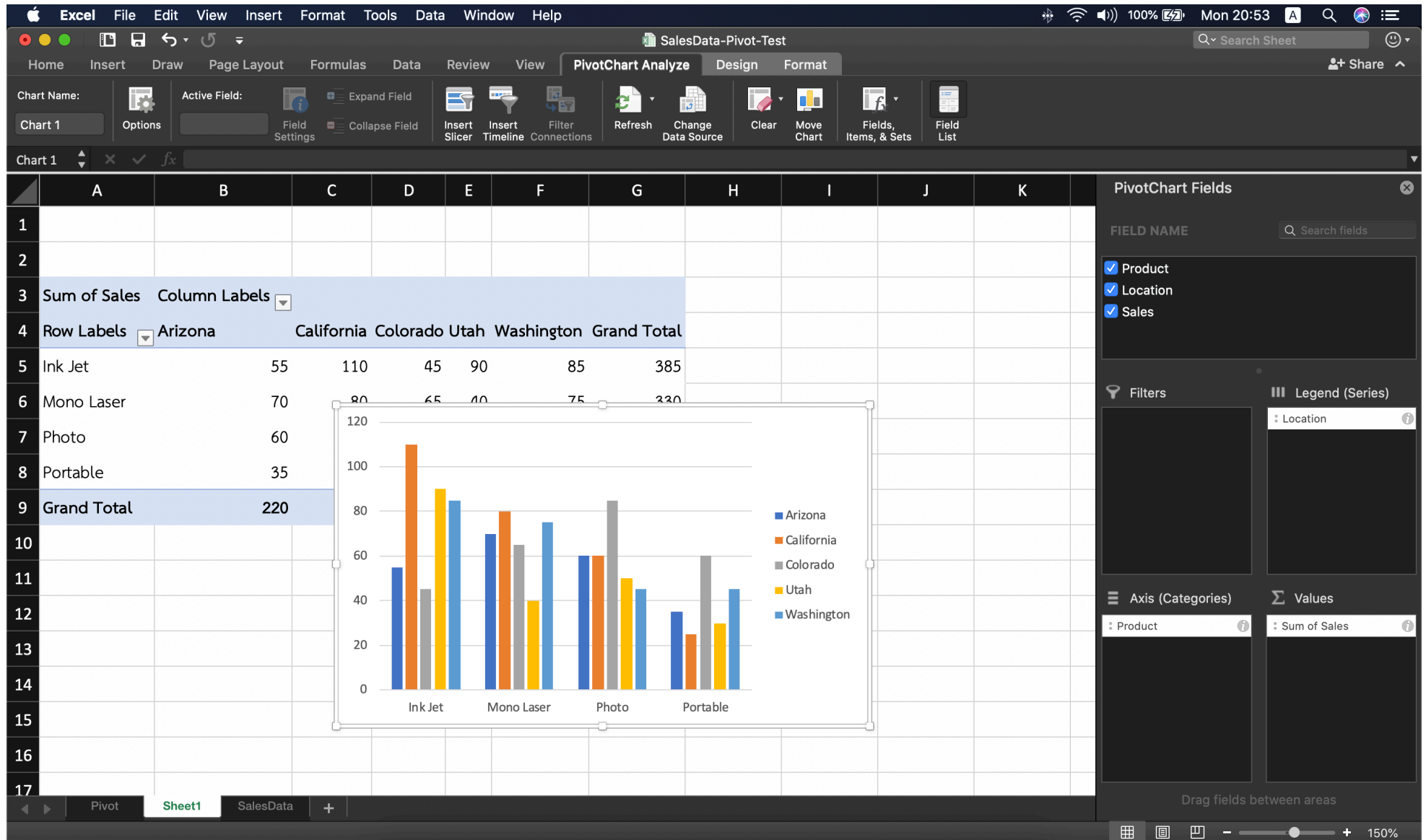
Row Labels	Arizona	California	Colorado	Utah	Washington	Grand Total
Ink Jet	55	110	45	90	85	385
Mono Laser	70	80	65	40	75	330
Photo	60	60	85	50	45	300
Portable	35	25	60	30	45	195
Grand Total	220	275	255	210	250	1210

The PivotTable Fields task pane on the right shows the following configuration:

- Field Name:** Search fields
- Fields:** Product, Location, Sales (all checked)
- Filters:** (Empty)
- Columns:** Location
- Rows:** Product
- Values:** Sum of Sales

The status bar at the bottom indicates the current sheet is "Sheet1" and the PivotTable is selected.

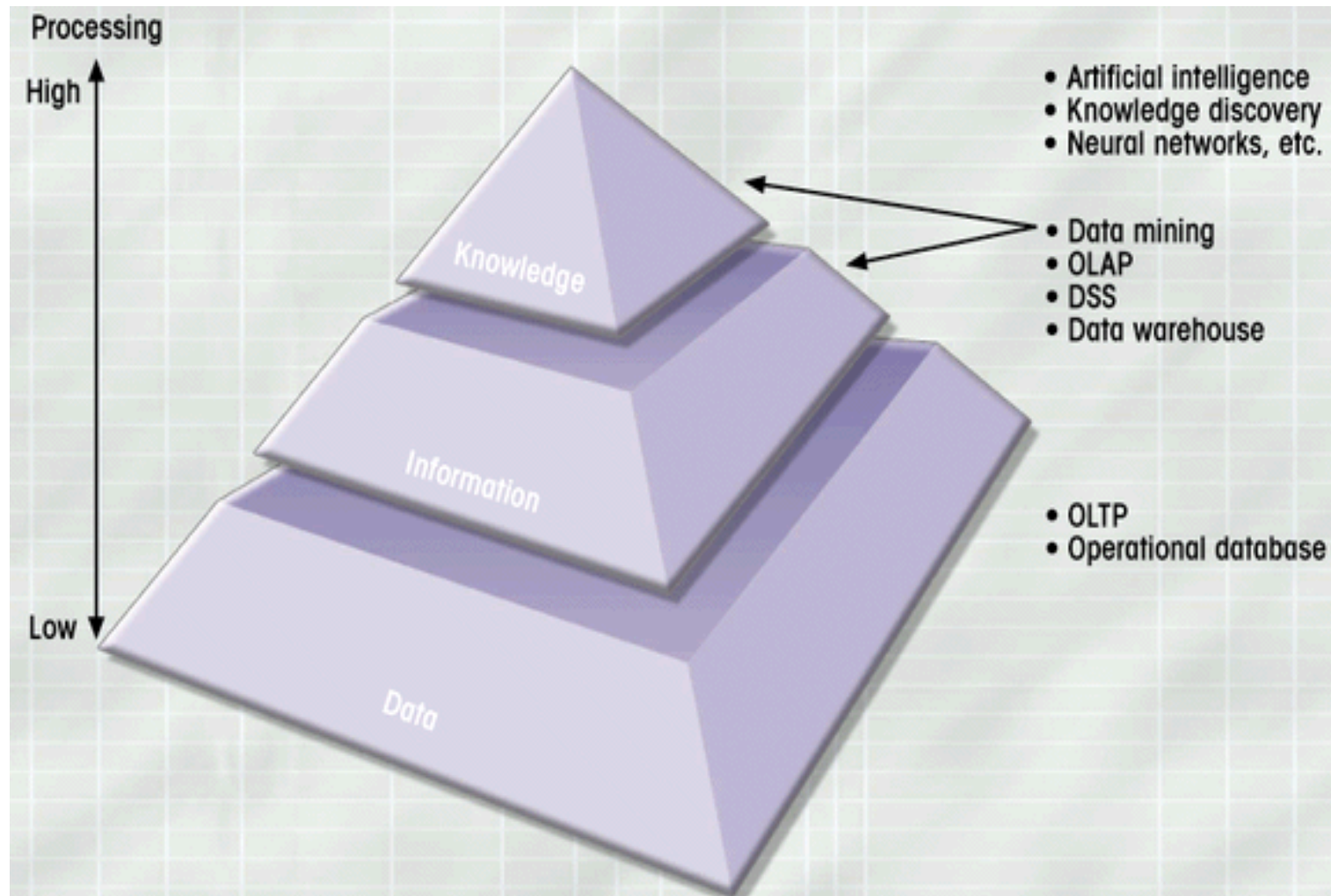
PIVOT TABLES IN EXCEL (6)



DATA MINING

- Seeks to discover unknown data characteristics
- Automatically searches data for anomalies and relationships
- Data mining tools
 - Analyse data
 - Uncover problems or opportunities
 - Form computer models based on findings
 - Predict business behaviour with models
 - Require minimal end-user intervention

EXTRACTION OF KNOWLEDGE FROM DATA



ASSIGNMENT 7

- ให้ดาวน์โหลดไฟล์ SampleData.xlsx เพื่อใช้เป็นข้อมูลในการตอบคำถามต่อไปนี้
 - ถ้าบริษัทต้องการที่จะปิดสาขาเพื่อลดค่าใช้จ่าย คิดว่าควรปิดสาขาในภูมิภาค (Region) อะไร
 - ถ้าบริษัทต้องการปิดสายการผลิต คิดว่าควรปิดสายการผลิตสินค้า สี (Color) อะไร
 - ไตรมาสไหนที่มียอดขายต่ำสุด และถ้าต้องการเพิ่มยอดขายในไตรมาสนั้น คิดว่าควรทำโปรโมชั่นที่สินค้าสี (Color) อะไร
- ตอบคำถามพร้อมส่งไฟล์ SampleData.xlsx ที่สร้าง Pivot Table สำหรับตอบคำถามข้างต้น พร้อมพิมพ์คำตอบลงในไฟล์ Excel แล้วกลับมาที่อีเมล kanat@egco.org