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.

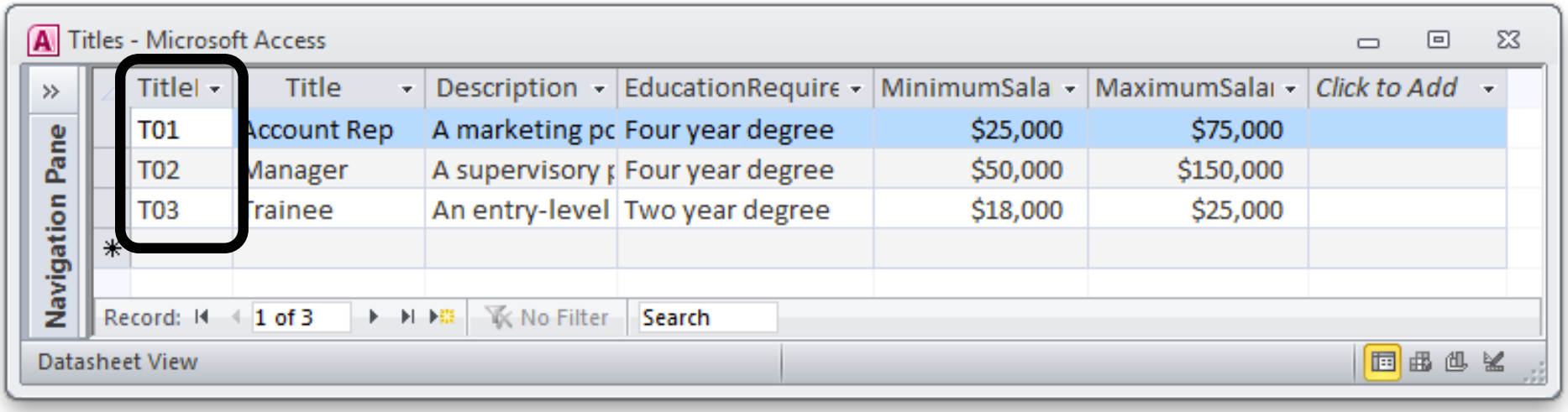
RELATIONAL DATABASES

- 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)



RECORD (ROW)

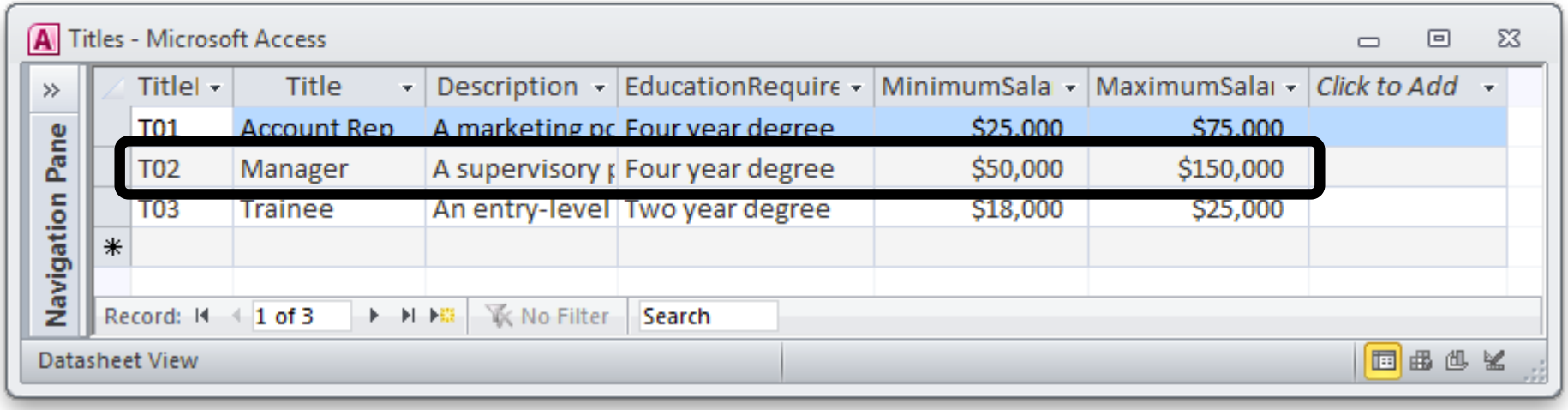
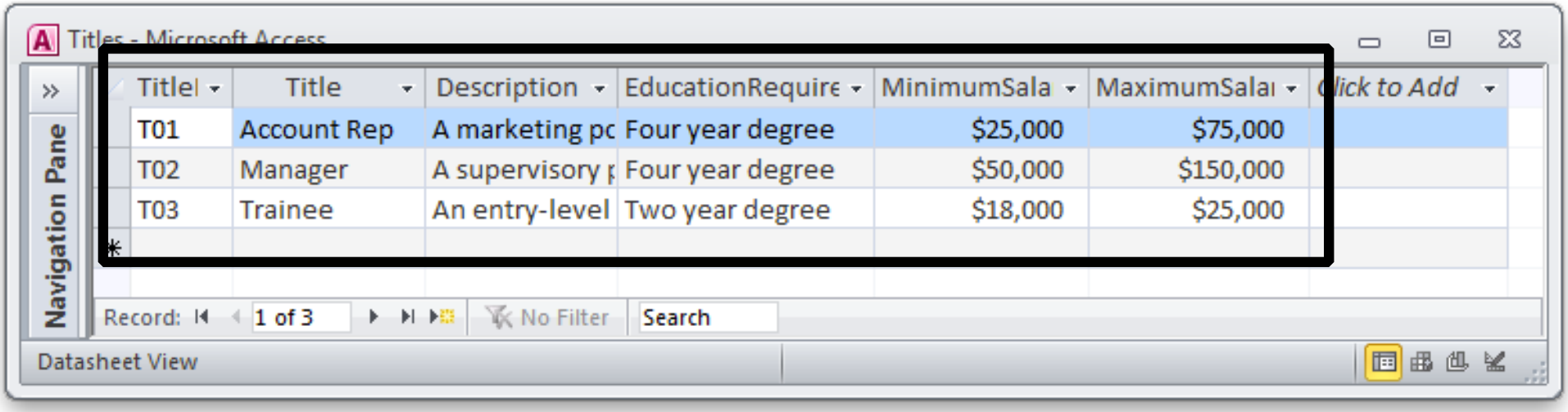
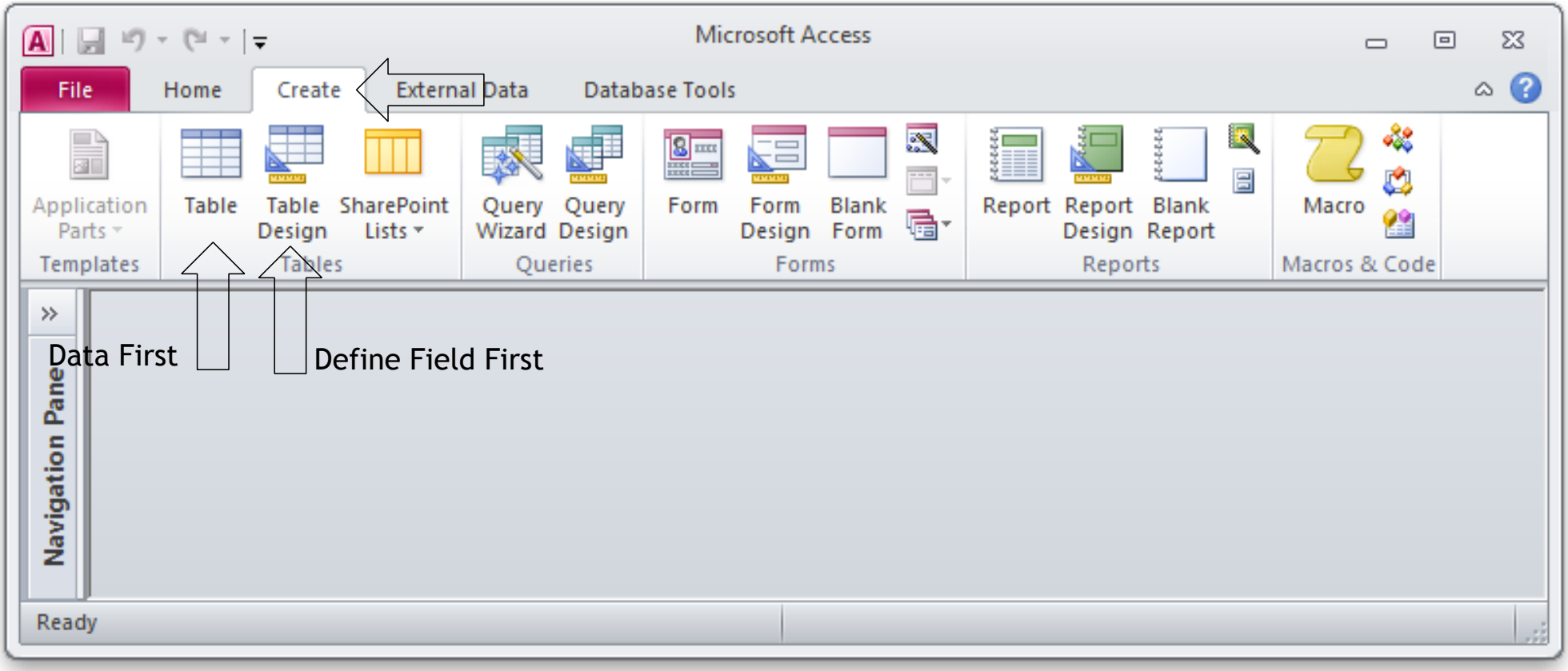

Titles - Microsoft Access
Navigation Pane
TitleI
Title
Description
EducationRequire
MinimumSala
MaximumSalai
Click to Add
T01
Account Rep
A marketing pc
Four year degree
$25,000
$75,000
T02
Manager
A supervisory p
Four year degree
$50,000
$150,000
T03
Trainee
An entry-level
Two year degree
$18,000
$25,000
Record: 1 of 3
No Filter
Search
Datasheet View

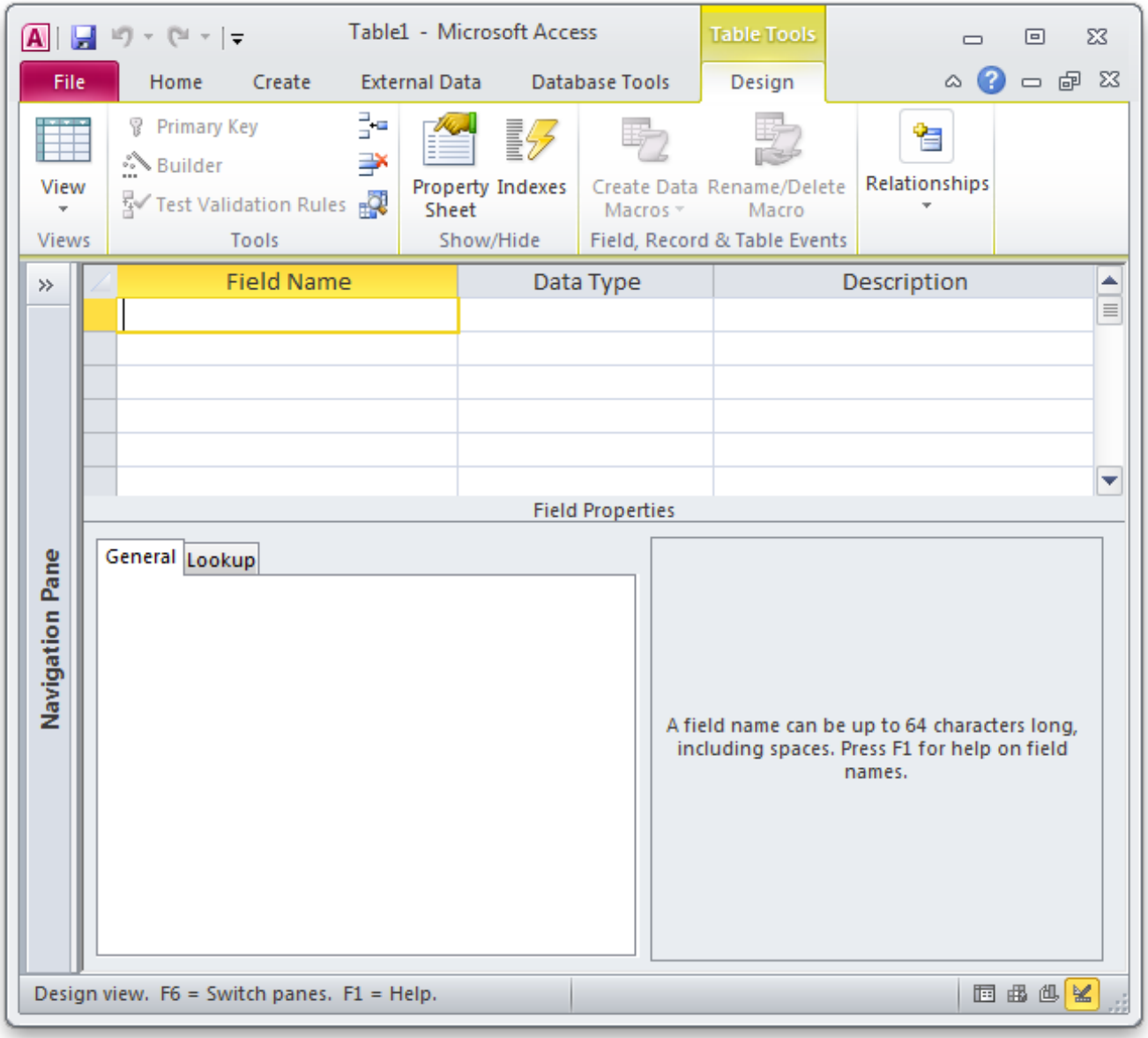
TABLE (ENTITY)



NEW TABLE IN MICROSOFT ACCESS



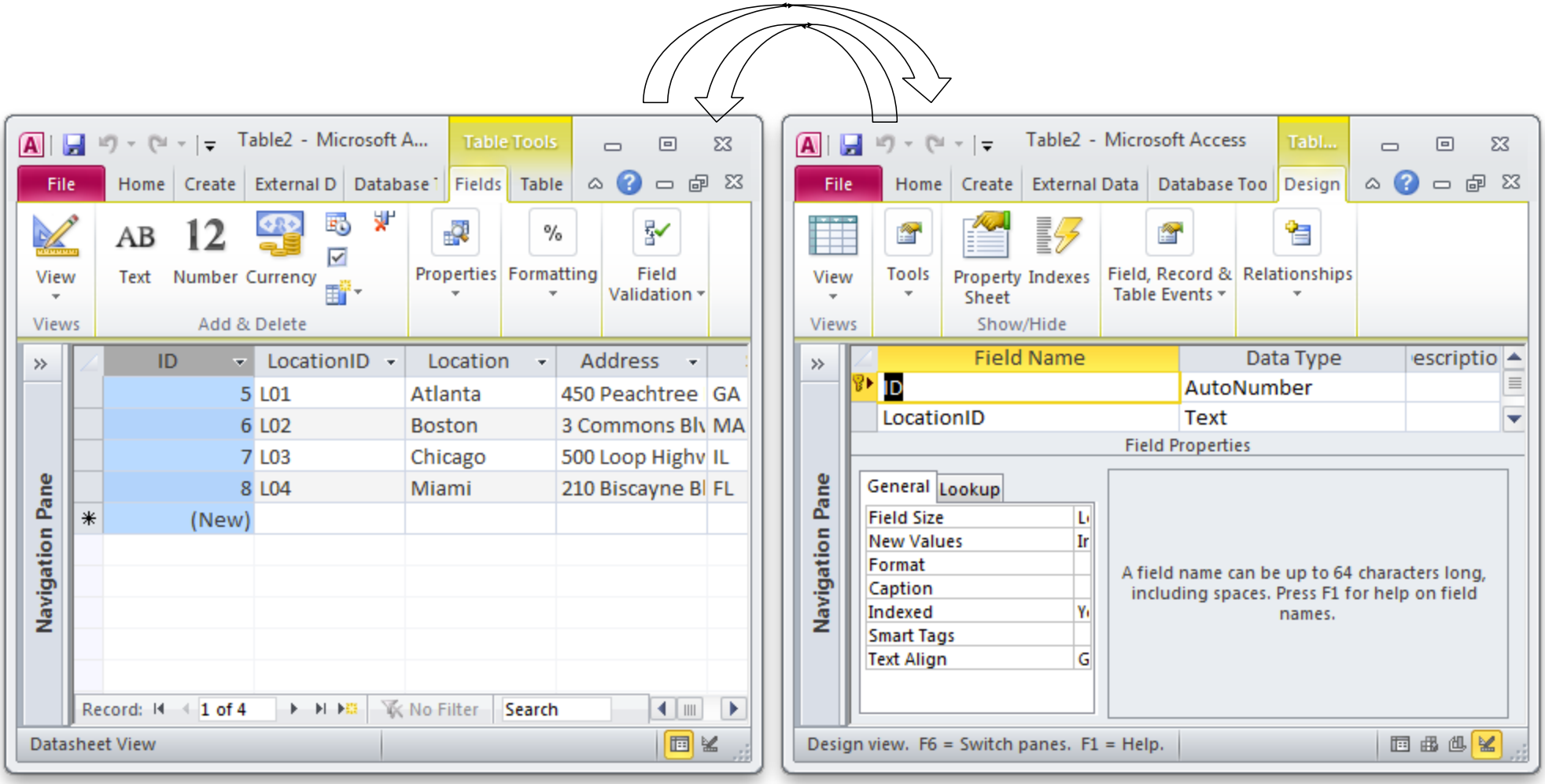
DEFINE FILED



COMMON DATA TYPES

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

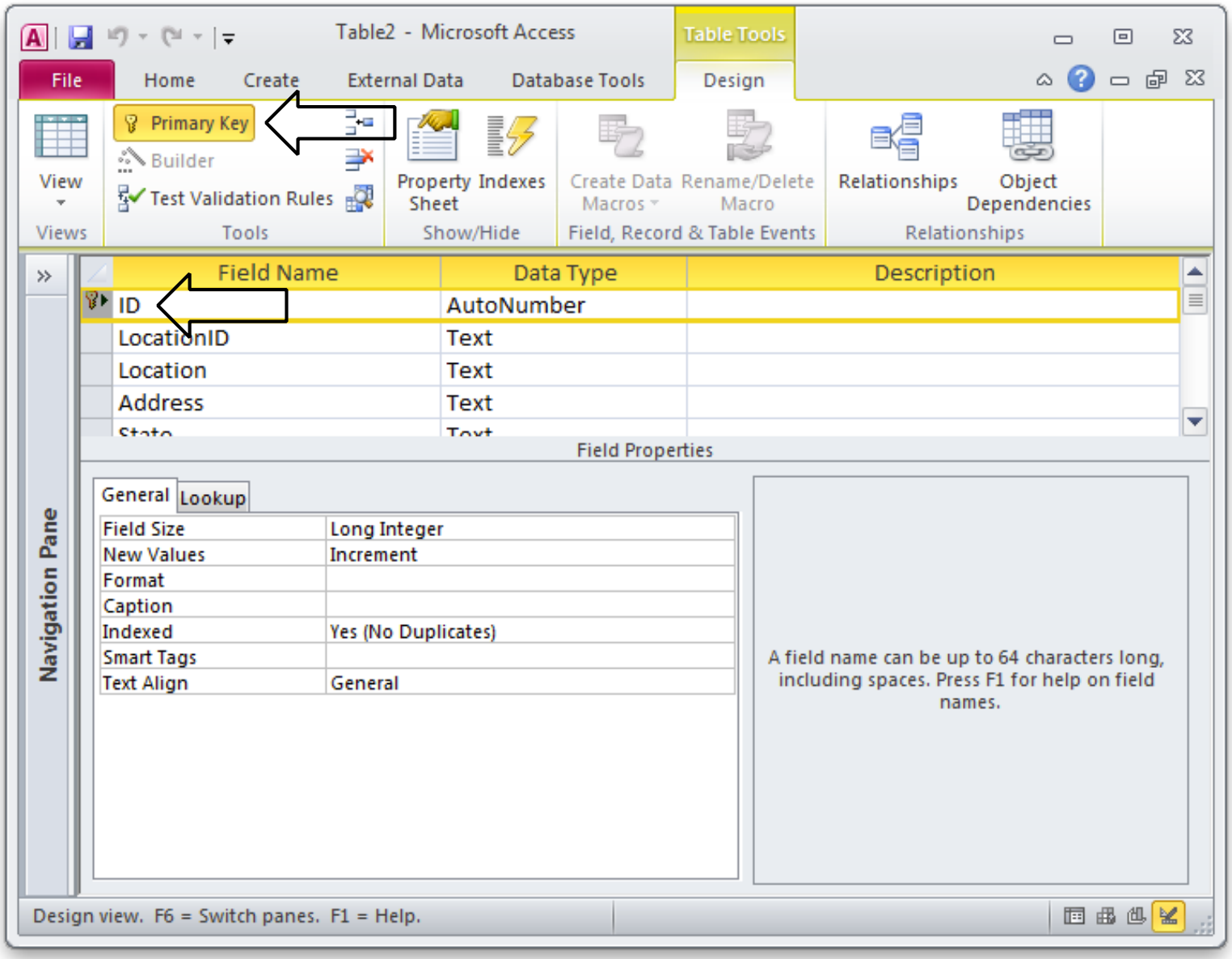
SWITCH BETWEEN DESIGN AND VIEW



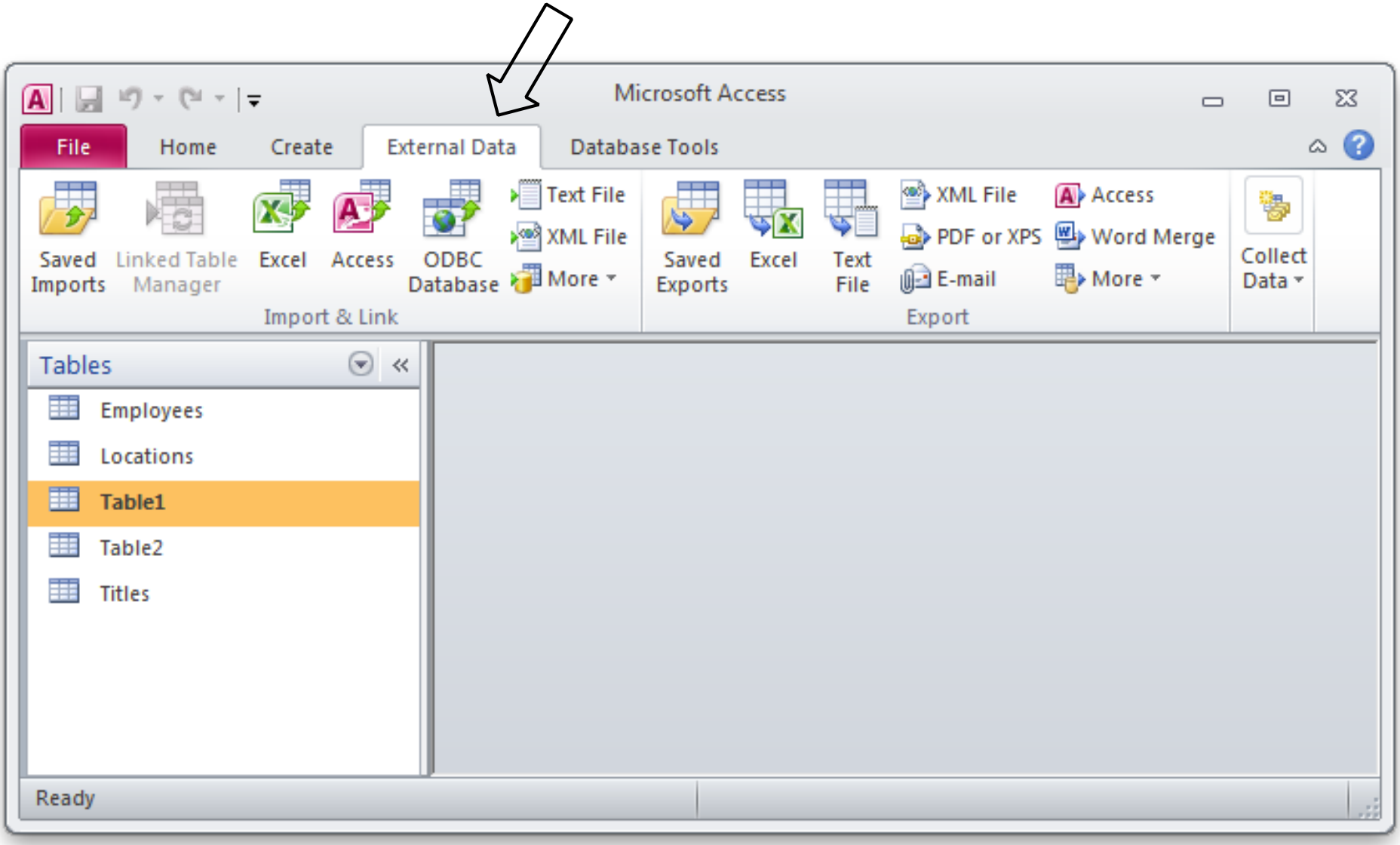
PRIMARY KEY (PK)

- 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

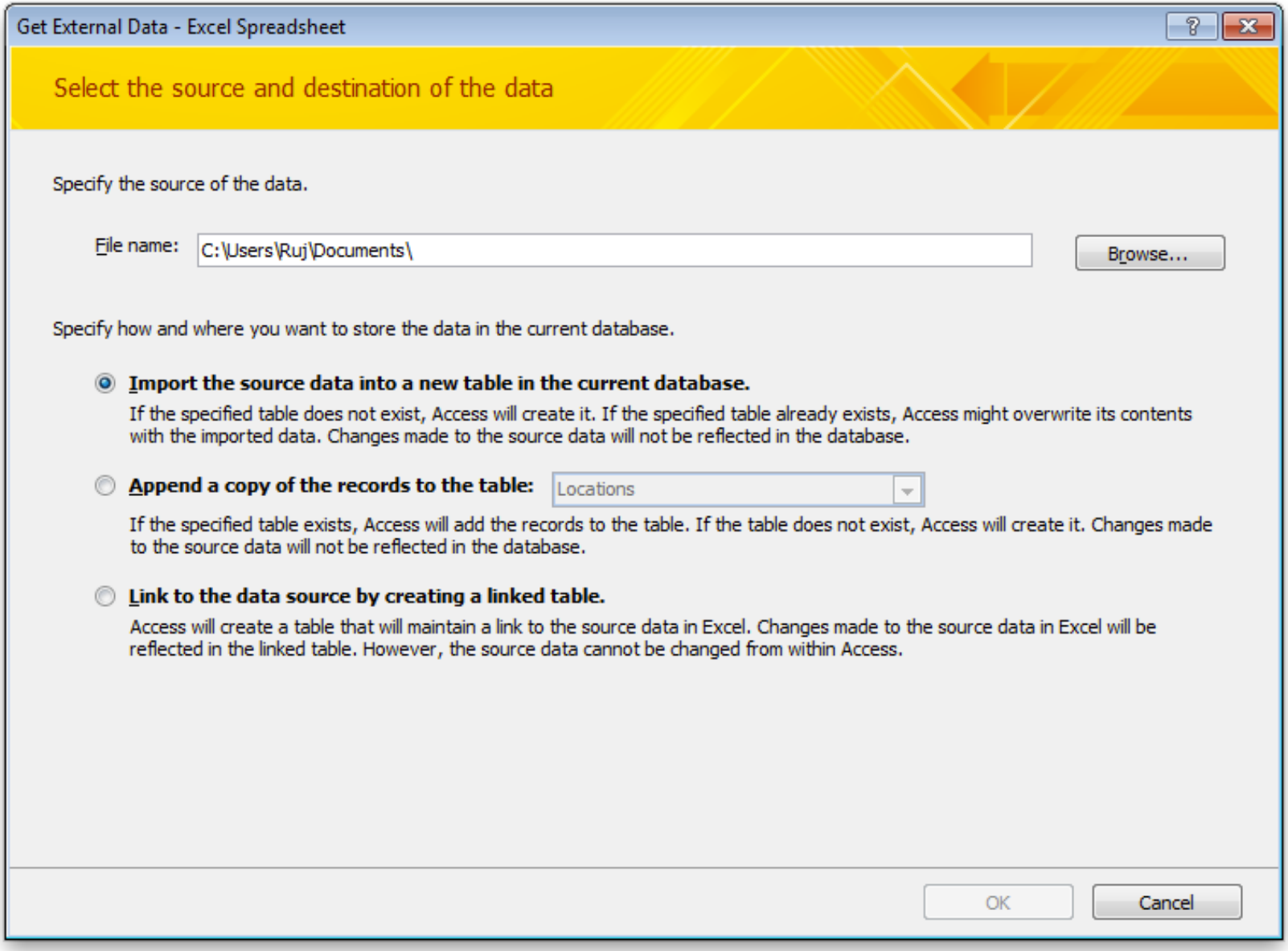
SET PK IN MICROSOFT ACCESS



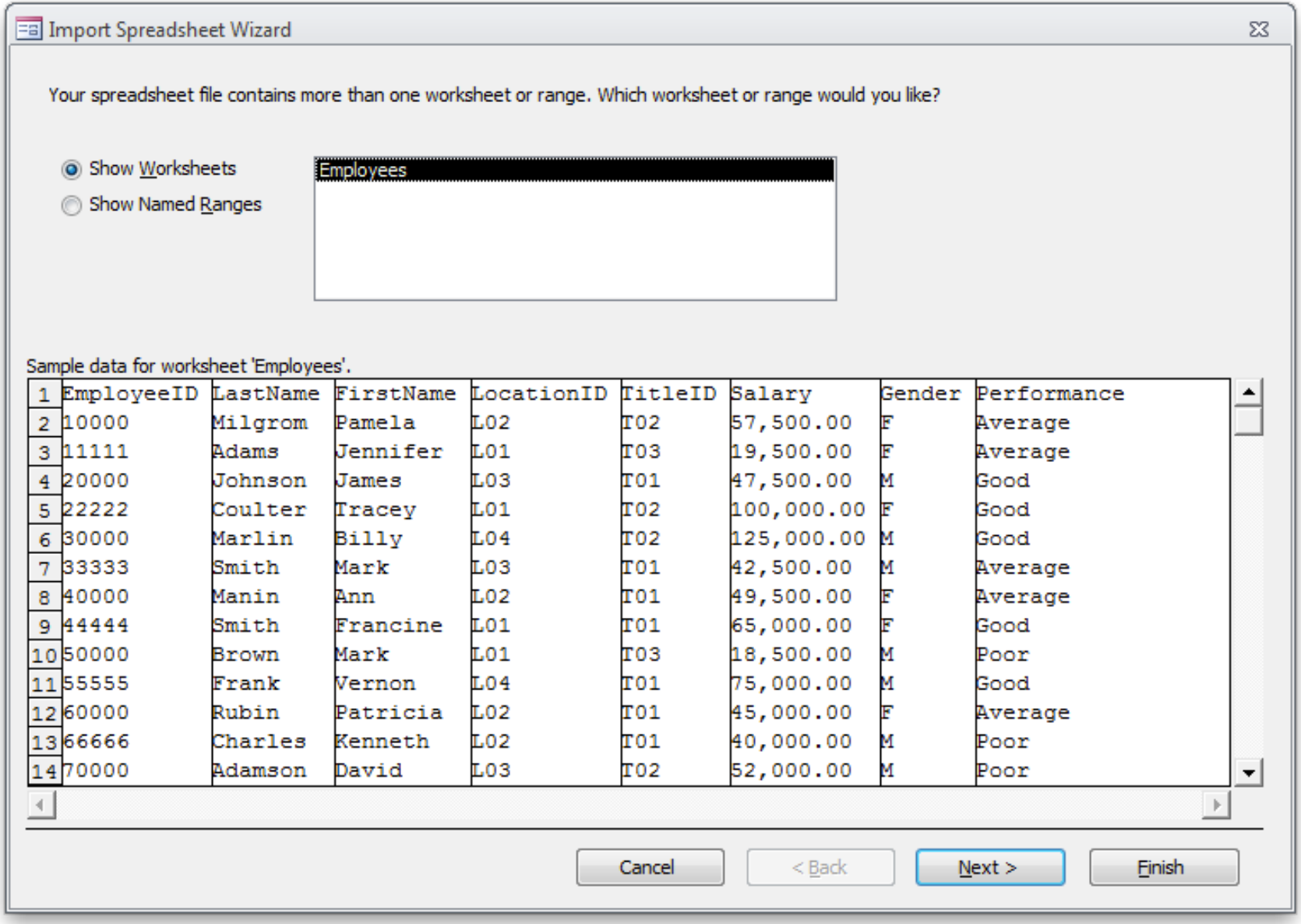
IMPORT & EXPORT DATA (1)



IMPORT & EXPORT DATA (2)

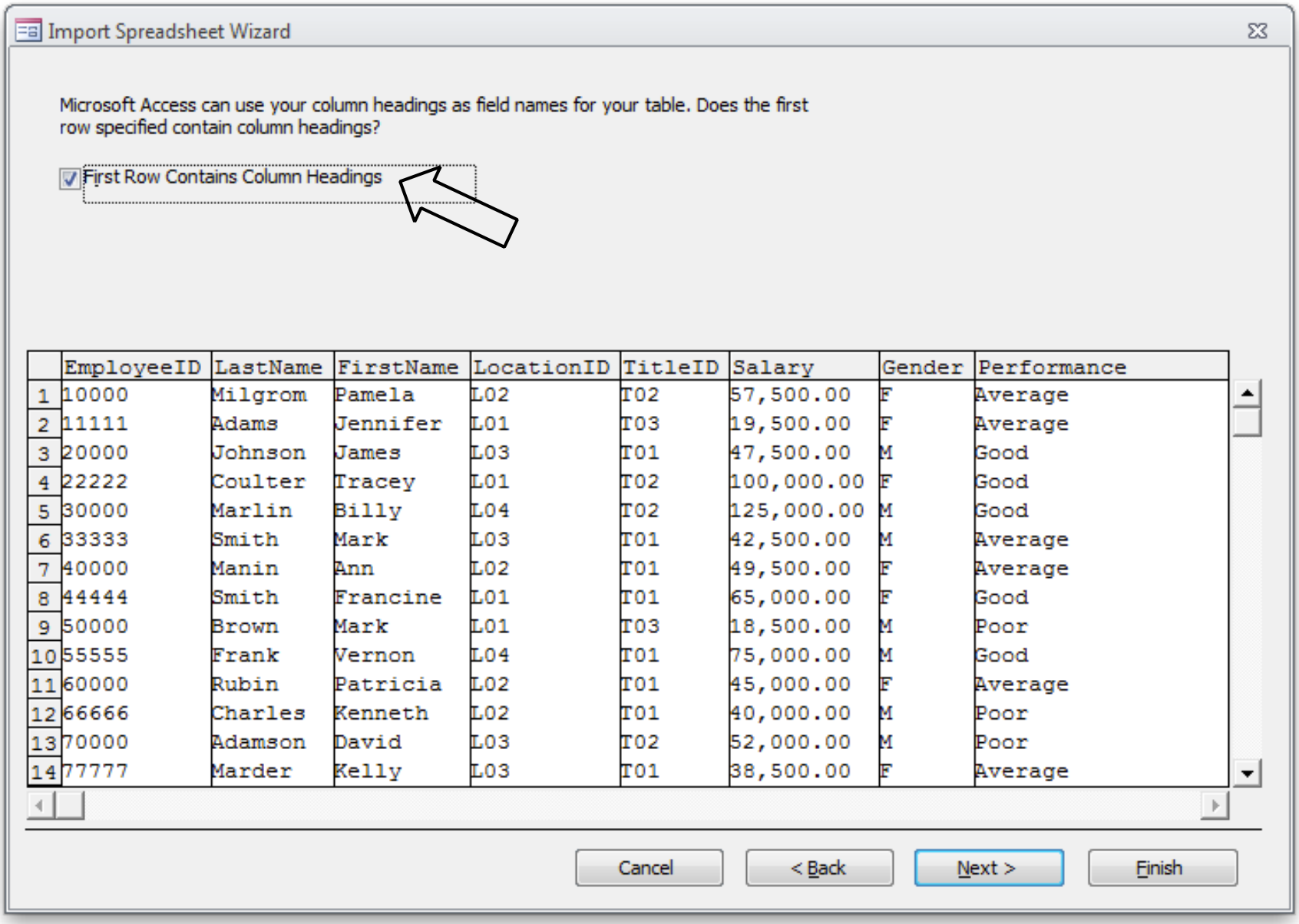


IMPORT & EXPORT DATA (3)



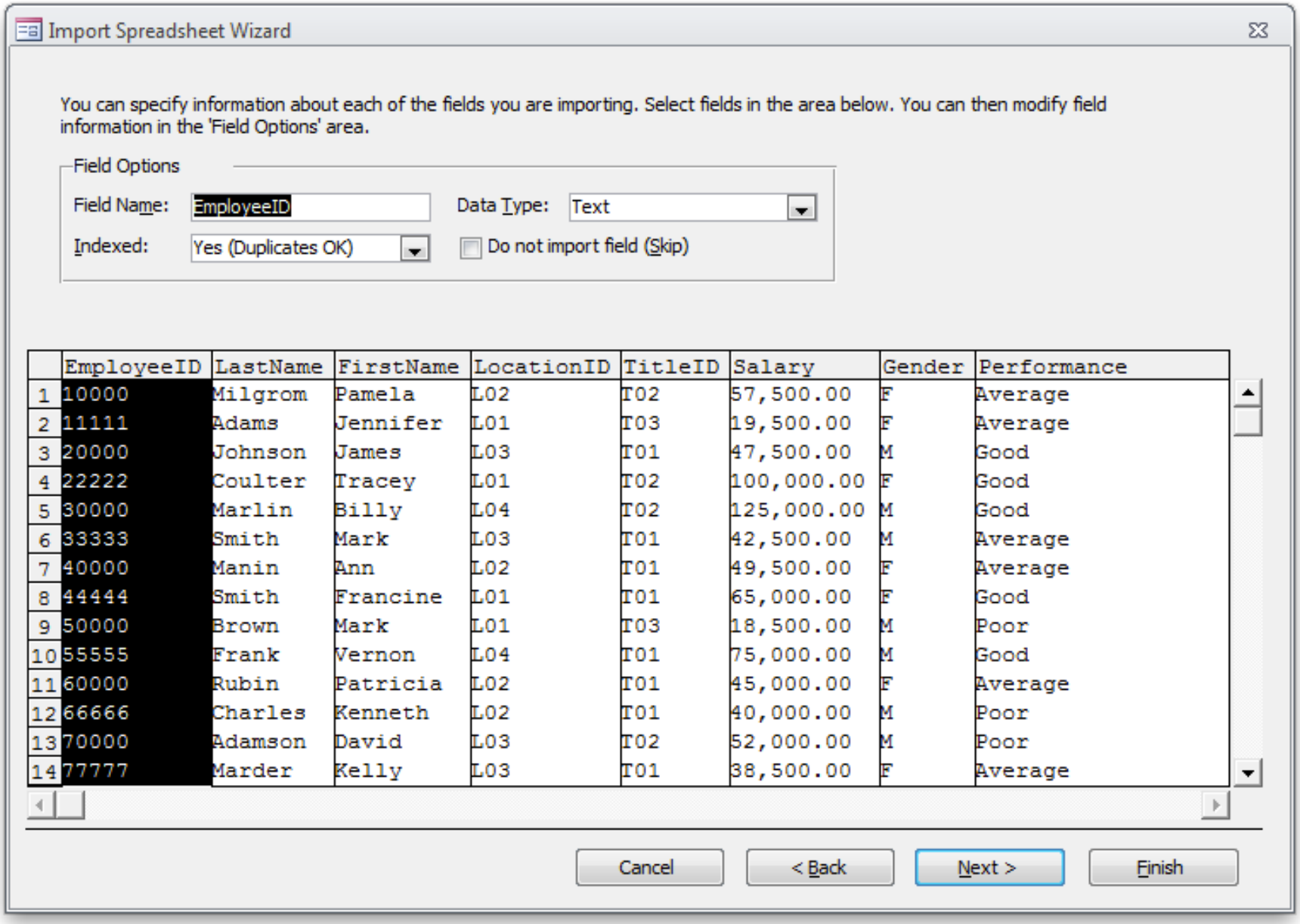
	EmployeeID	LastName	FirstName	LocationID	TitleID	Salary	Gender	Performance
2	10000	Milgrom	Pamela	L02	T02	57,500.00	F	Average
3	11111	Adams	Jennifer	L01	T03	19,500.00	F	Average
4	20000	Johnson	James	L03	T01	47,500.00	M	Good
5	22222	Coulter	Tracey	L01	T02	100,000.00	F	Good
6	30000	Marlin	Billy	L04	T02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

IMPORT & EXPORT DATA (4)



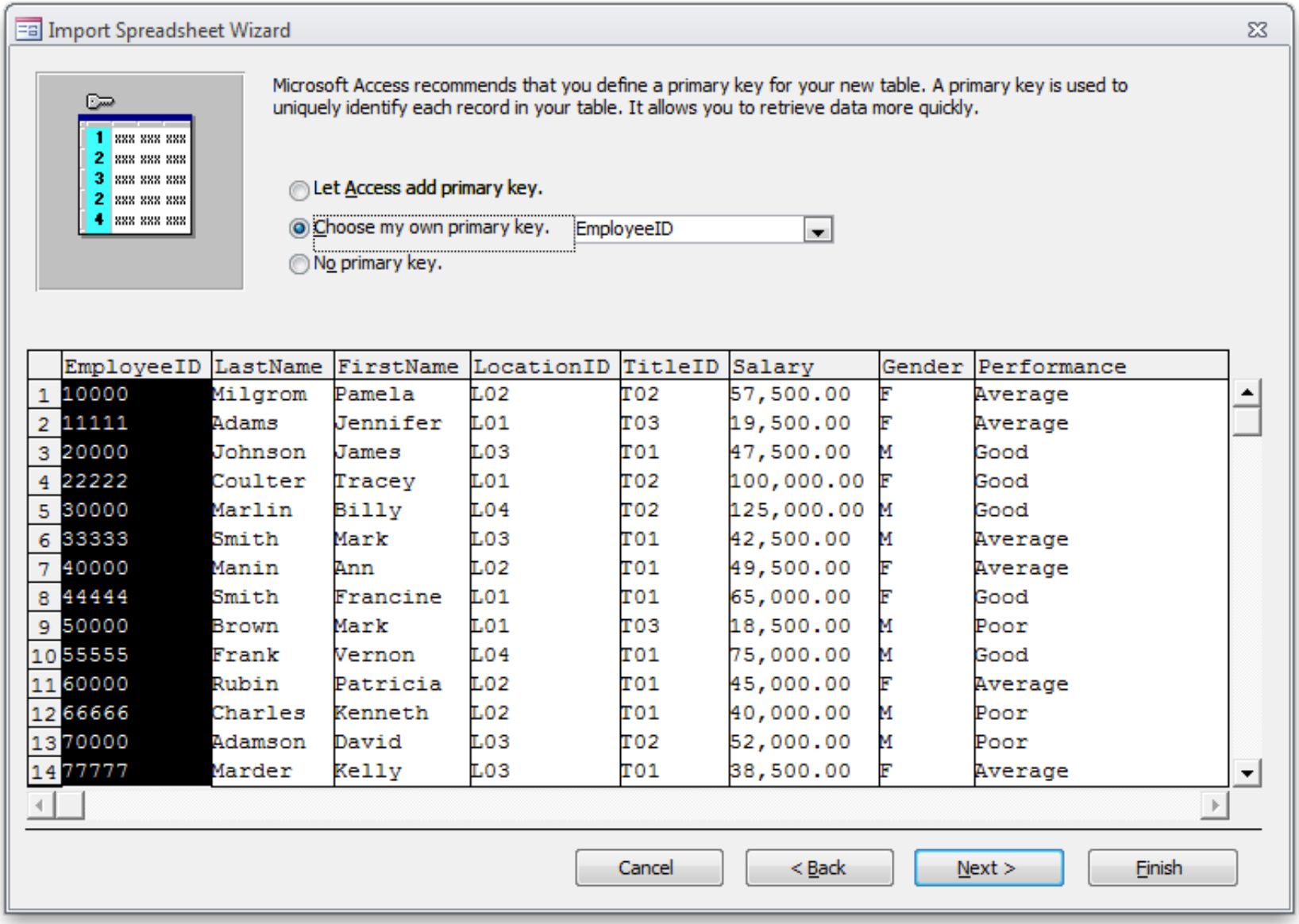
	EmployeeID	LastName	FirstName	LocationID	TitleID	Salary	Gender	Performance
1	10000	Milgrom	Pamela	L02	T02	57,500.00	F	Average
2	11111	Adams	Jennifer	L01	T03	19,500.00	F	Average
3	20000	Johnson	James	L03	T01	47,500.00	M	Good
4	22222	Coulter	Tracey	L01	T02	100,000.00	F	Good
5	30000	Marlin	Billy	L04	T02	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

IMPORT & EXPORT DATA (5)



	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

IMPORT & EXPORT DATA (6)



	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

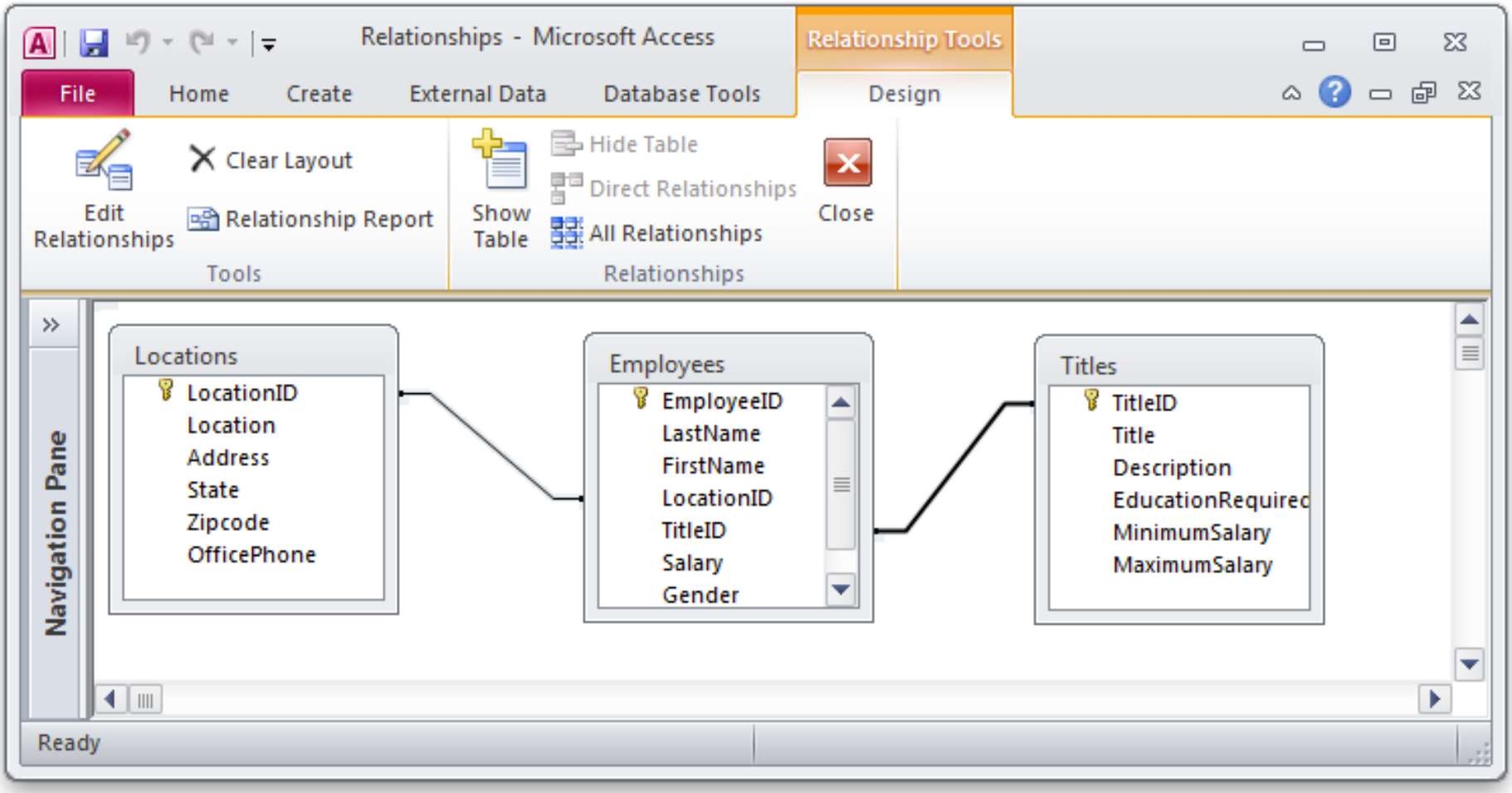
RELATIONSHIP

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

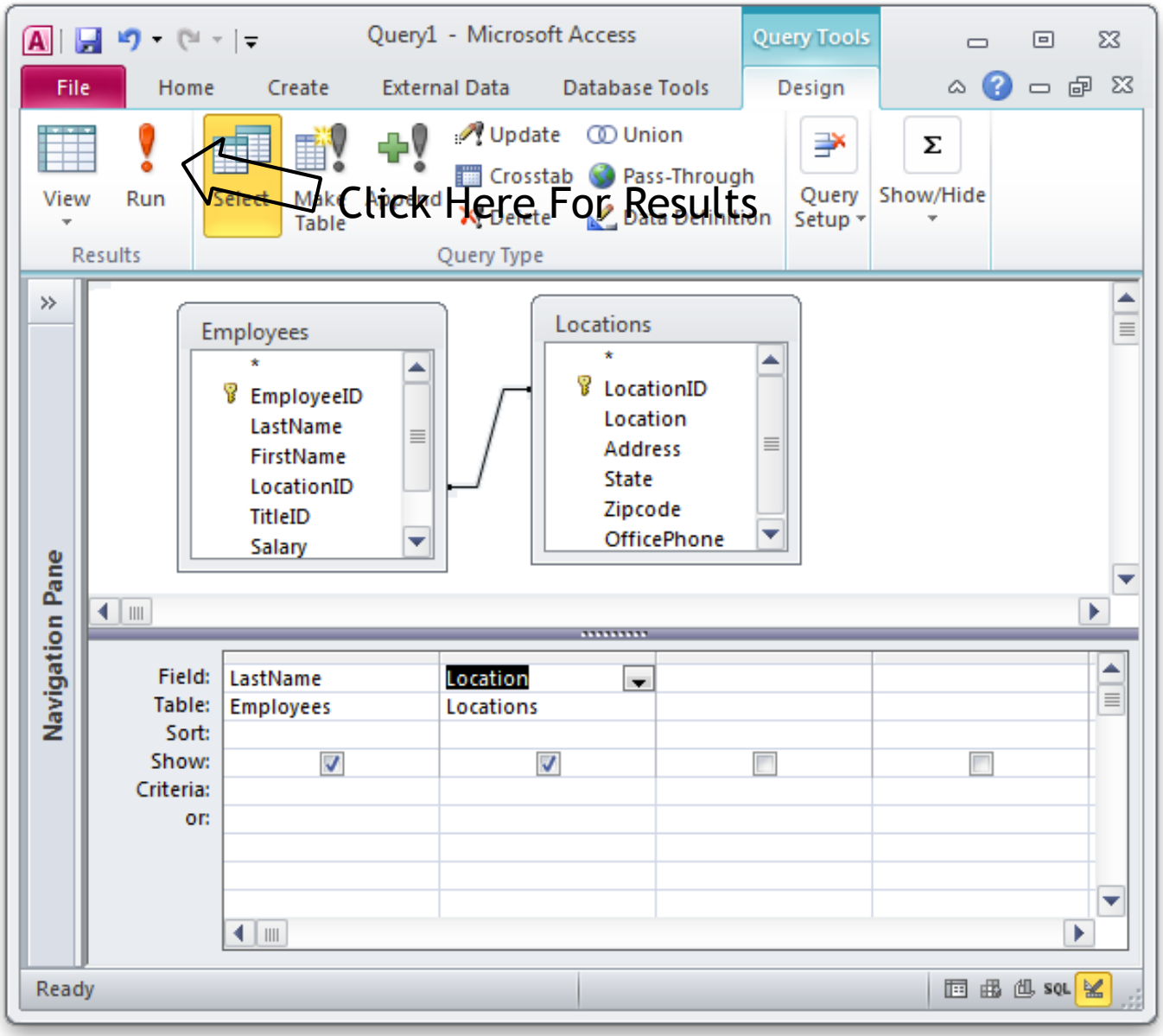
TYPE OF RELATIONSHIPS

- 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

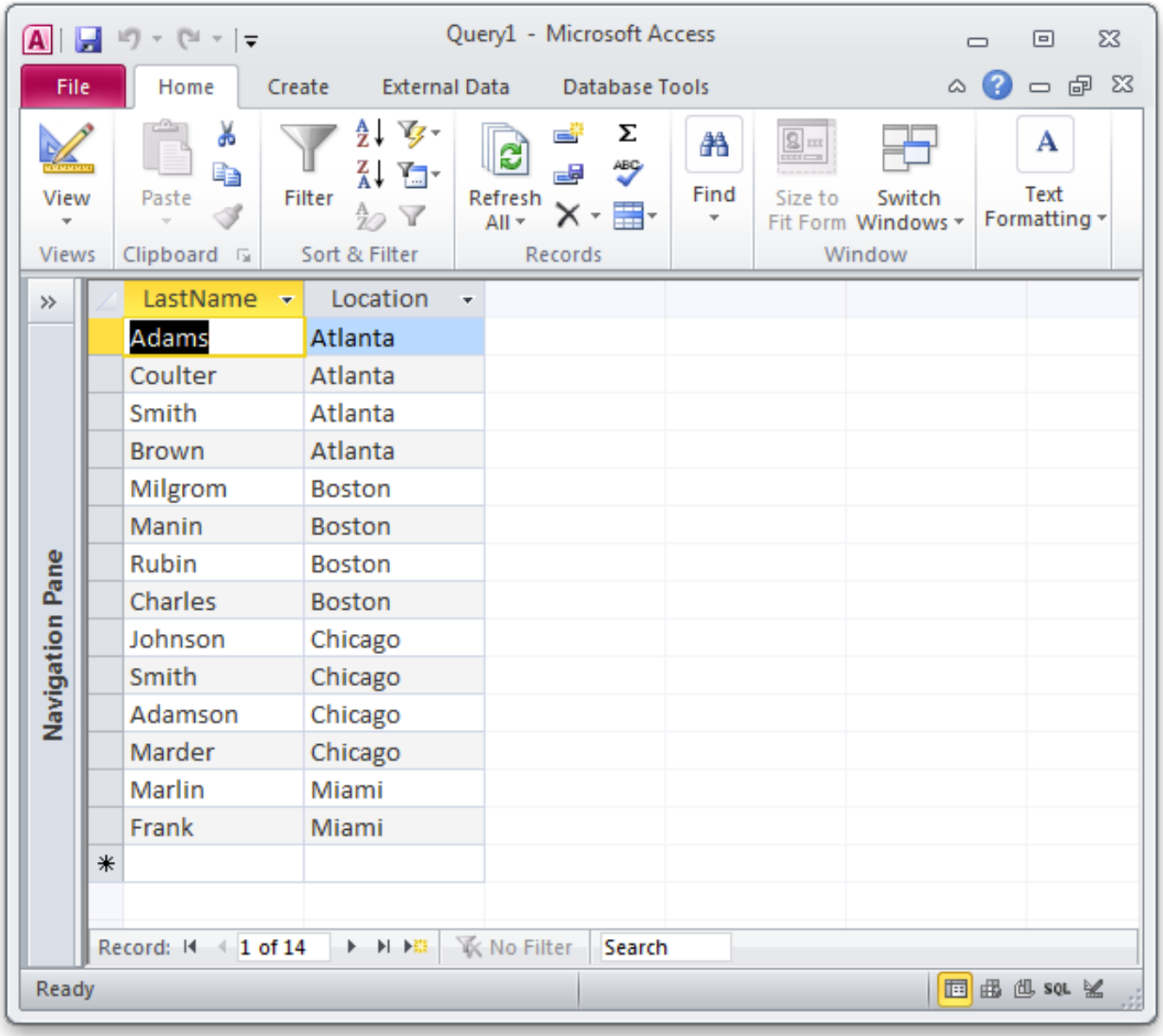
DEFINE RELATIONSHIP



QUERYING (1)


Query1 - Microsoft Access
Query Tools
File
Home
Create
External Data
Database Tools
Design
View
Run
Select
Make Table
Update
Union
Crosstab
Pass-Through
Delete
Data Definition
Query Setup
Show/Hide
Results
Query Type
Click Here For Results
Navigation Pane
Employees
*
EmployeeID
LastName
FirstName
LocationID
TitleID
Salary
Locations
*
LocationID
Location
Address
State
Zipcode
OfficePhone
Field:
Table:
Sort:
Show:
Criteria:
or:
LastName
Location
Employees
Locations
Ready
SQL

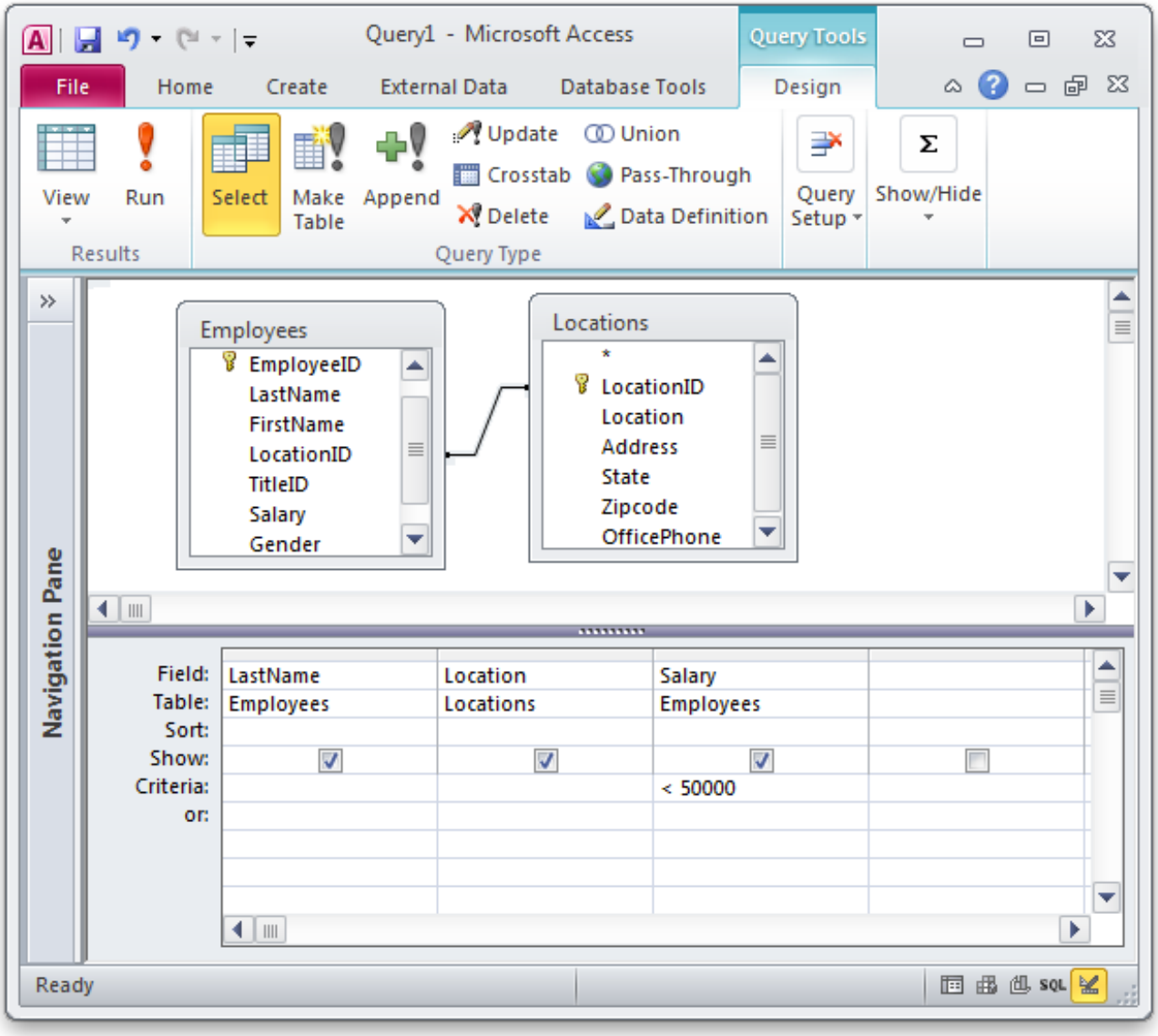
QUERYING (2)


Query1 - Microsoft Access
File
Home
Create
External Data
Database Tools
View
Views
Paste
Clipboard
Filter
Sort & Filter
Refresh All
Records
Find
Size to Fit Form
Switch Windows
Window
Text Formatting
Navigation Pane
LastName
Location
Adams Atlanta
Coulter Atlanta
Smith Atlanta
Brown Atlanta
Milgrom Boston
Manin Boston
Rubin Boston
Charles Boston
Johnson Chicago
Smith Chicago
Adamson Chicago
Marder Chicago
Marlin Miami
Frank Miami
Record: 1 of 14
No Filter
Search
Ready
SQL

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)


Query1 - Microsoft Access
Query Tools
File
Home
Create
External Data
Database Tools
Design
View
Run
Select
Make Table
Append
Update
Union
Crosstab
Pass-Through
Delete
Data Definition
Query Setup
Show/Hide
Results
Query Type
Navigation Pane
Employees
EmployeeID
LastName
FirstName
LocationID
TitleID
Salary
Gender
Locations
*
LocationID
Location
Address
State
Zipcode
OfficePhone
Field:
Table:
Sort:
Show:
Criteria:
or:
LastName
Location
Salary
Employees
Locations
Employees
< 50000
Ready
SQL

QUERYING (5)

- Using SQL command for querying data.

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

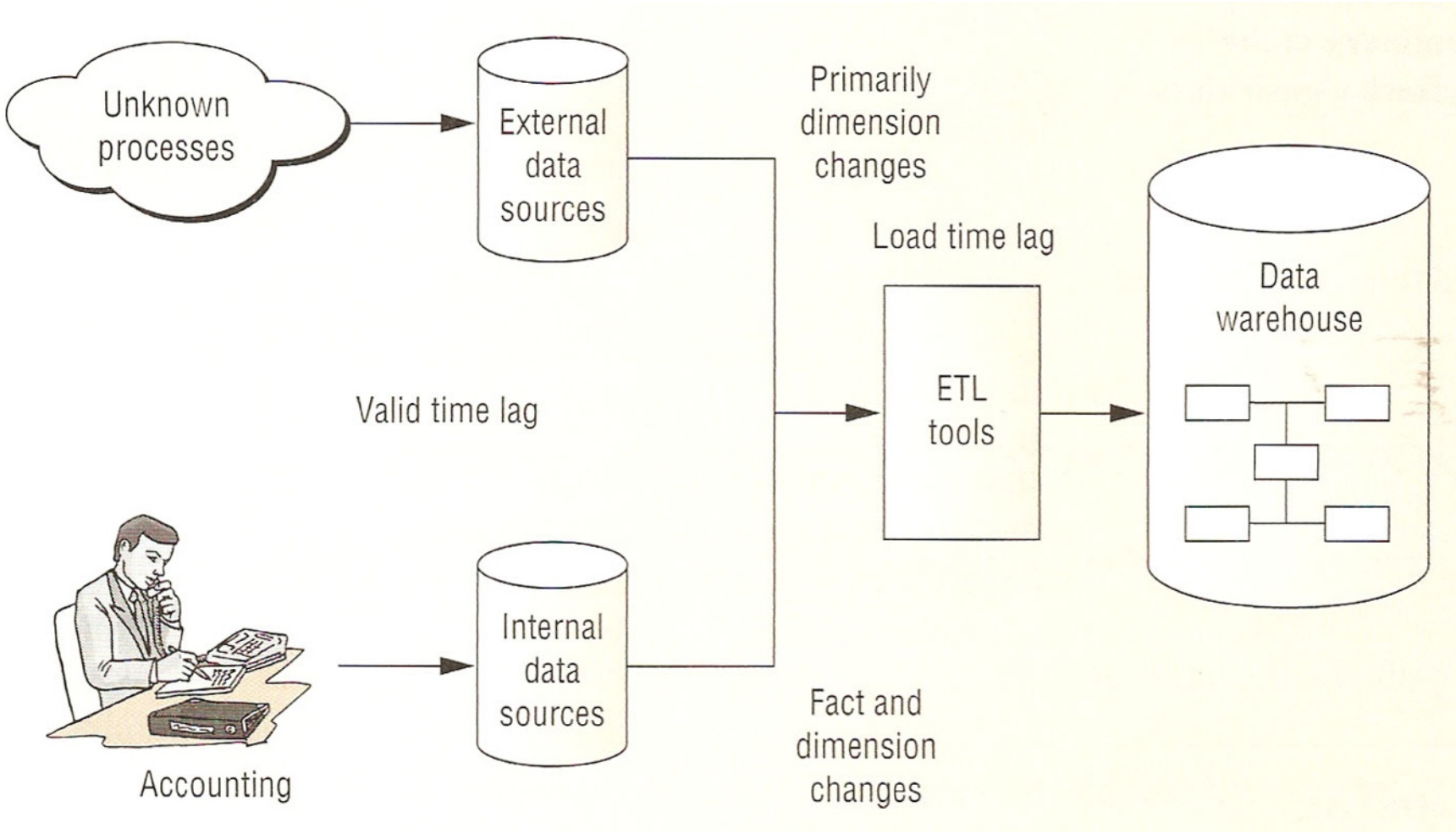
DATA WAREHOUSE

- 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

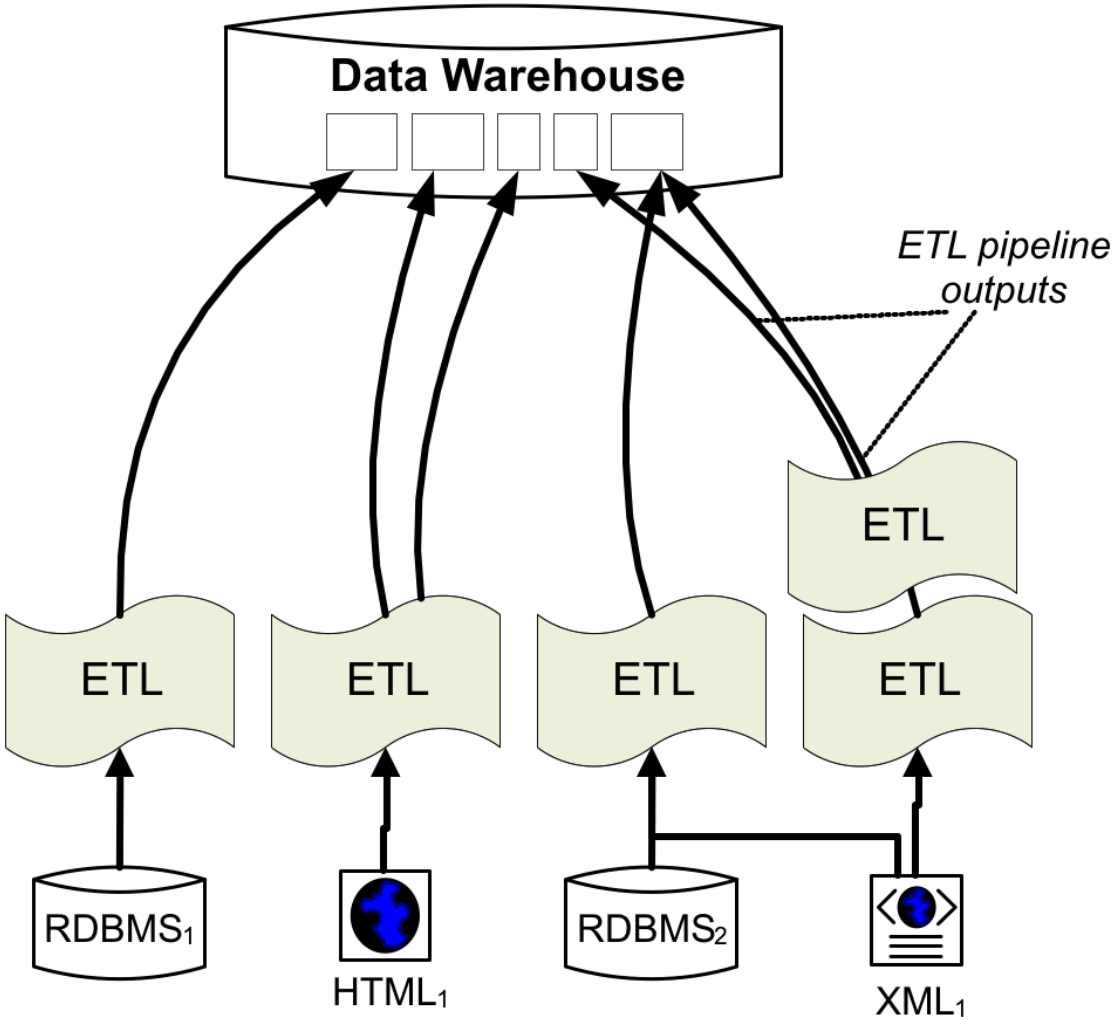
OVERVIEW OF THE DATA WAREHOUSE REFRESH PROCESS



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 Warehouse
ETL pipeline outputs
ETL
ETL
ETL
ETL
ETL
RDBMS1
HTML1
RDBMS2
XML1

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.

Sex
Male
F
M
Female
NULL
f
m

Sex
1
2
1
2
0
2
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

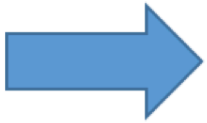
DoB	Name
2011-06-07	Child1
2011-06-07	Child2

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
2006-05-09
1995-06-06

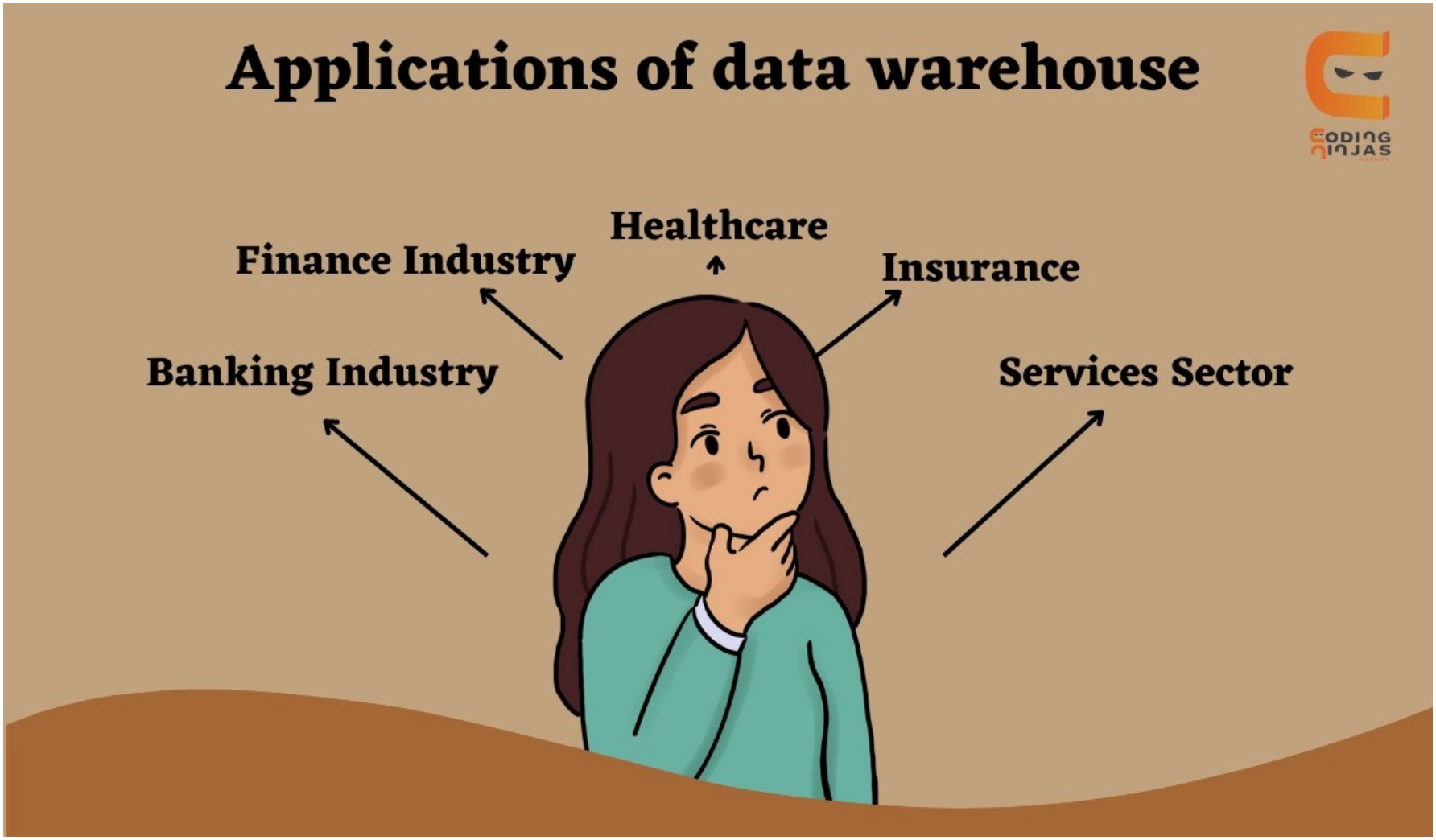


Sequence	Event	Date
1	Inmigration	1995-06-06
2	Outmigration	2006-05-09

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



* https://www.codingninjas.com/studio/library/applications-of-data-warehousing

MULTIDIMENSIONAL DATA

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

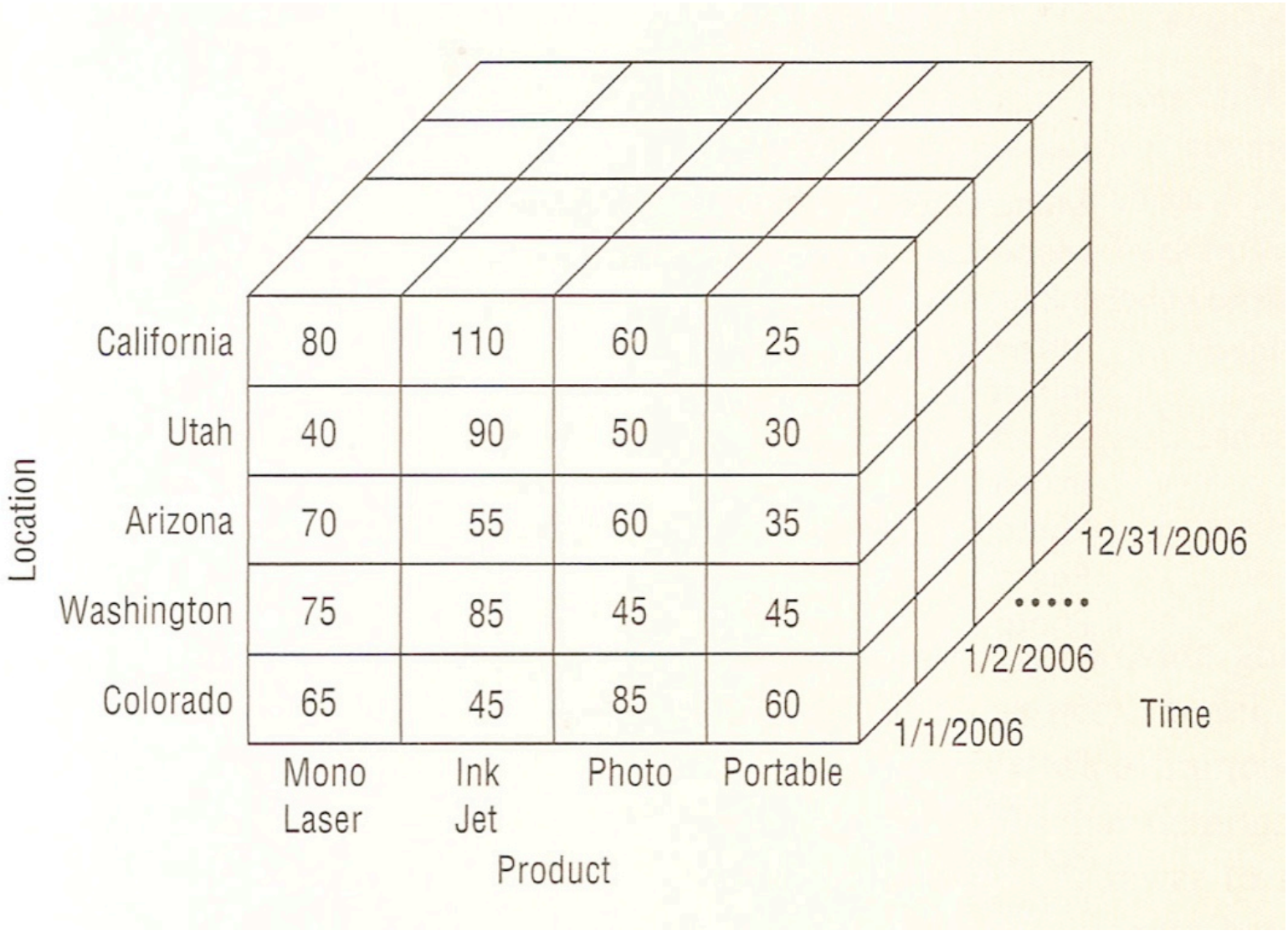
RELATIONAL REPRESENTATION OF SALES DATA

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

	Product			
Location	**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

	Product				
Location	Mono Laser	Ink Jet	Photo	Portable	Totals
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

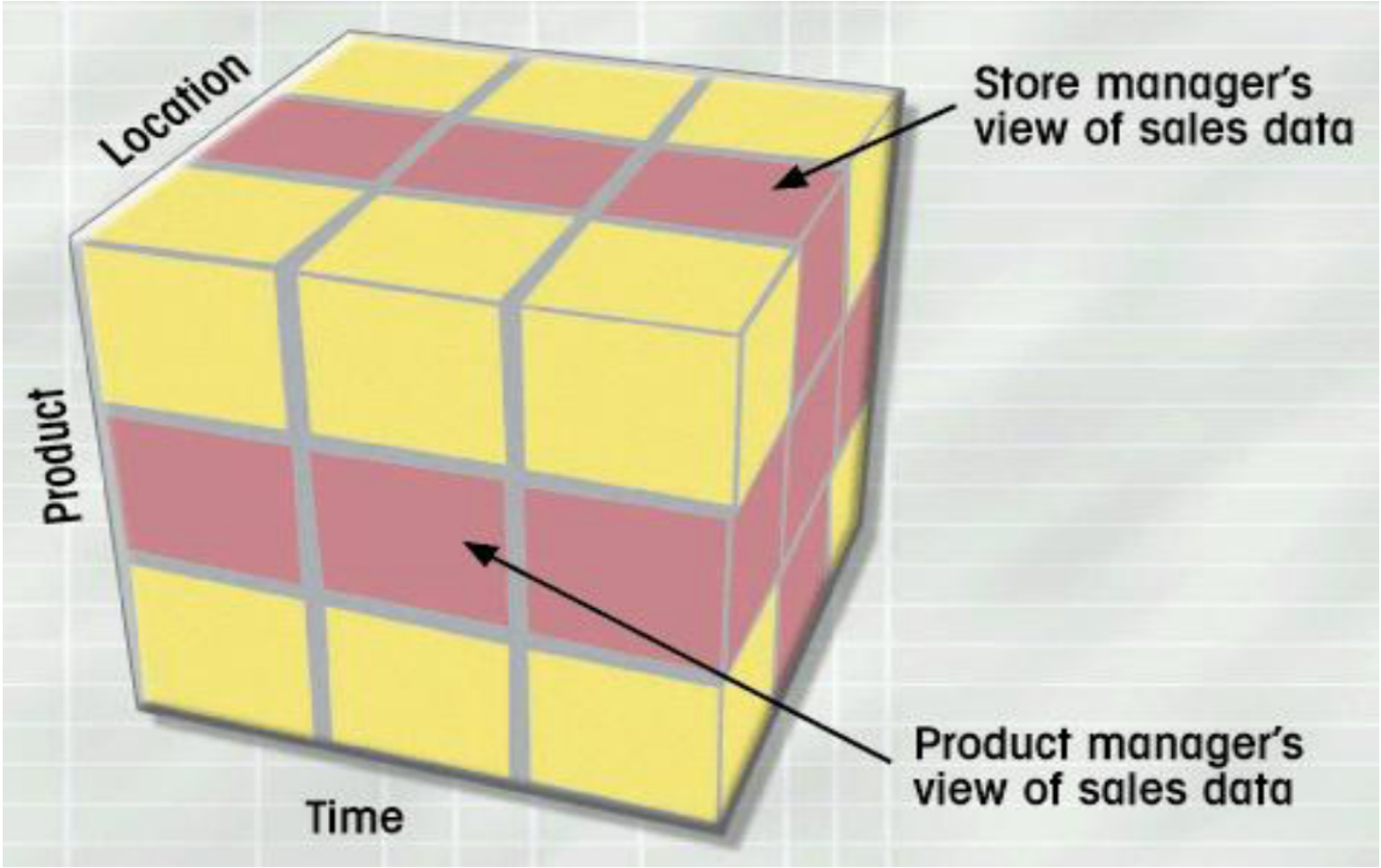
DATA CUBE

- 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, uses often need to focus on a subset of the dimensions to gain insights.
- **Dice**
  Because individual dimensions can contain a large number of members, uses need to focus on a subset of members to gain insights.

DATA CUBE OPERATIONS (2)



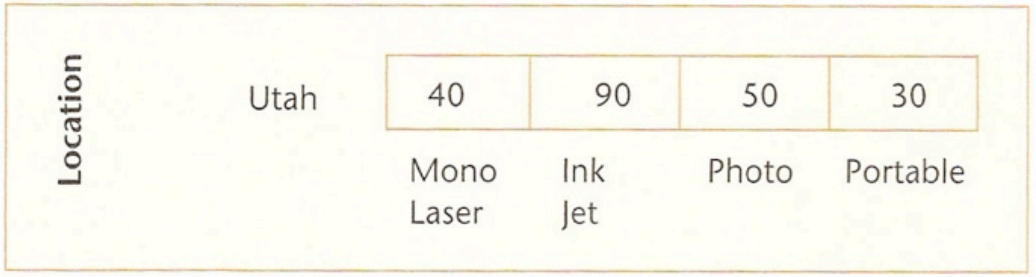
DATA CUBE OPERATIONS (3)

- **Drill-Down**
  Uses often want to navigate among the levels of hierarchical dimensions. The drill-down operator allows uses 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

	Product			
Location	**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

	Time			
Location	**1/1/2006**	**1/2/2006**	**. . .**	**Total Sales**
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



DRILL-DOWN OPERATION FOR THE STATE OF UTAH

	Product			
Location	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)

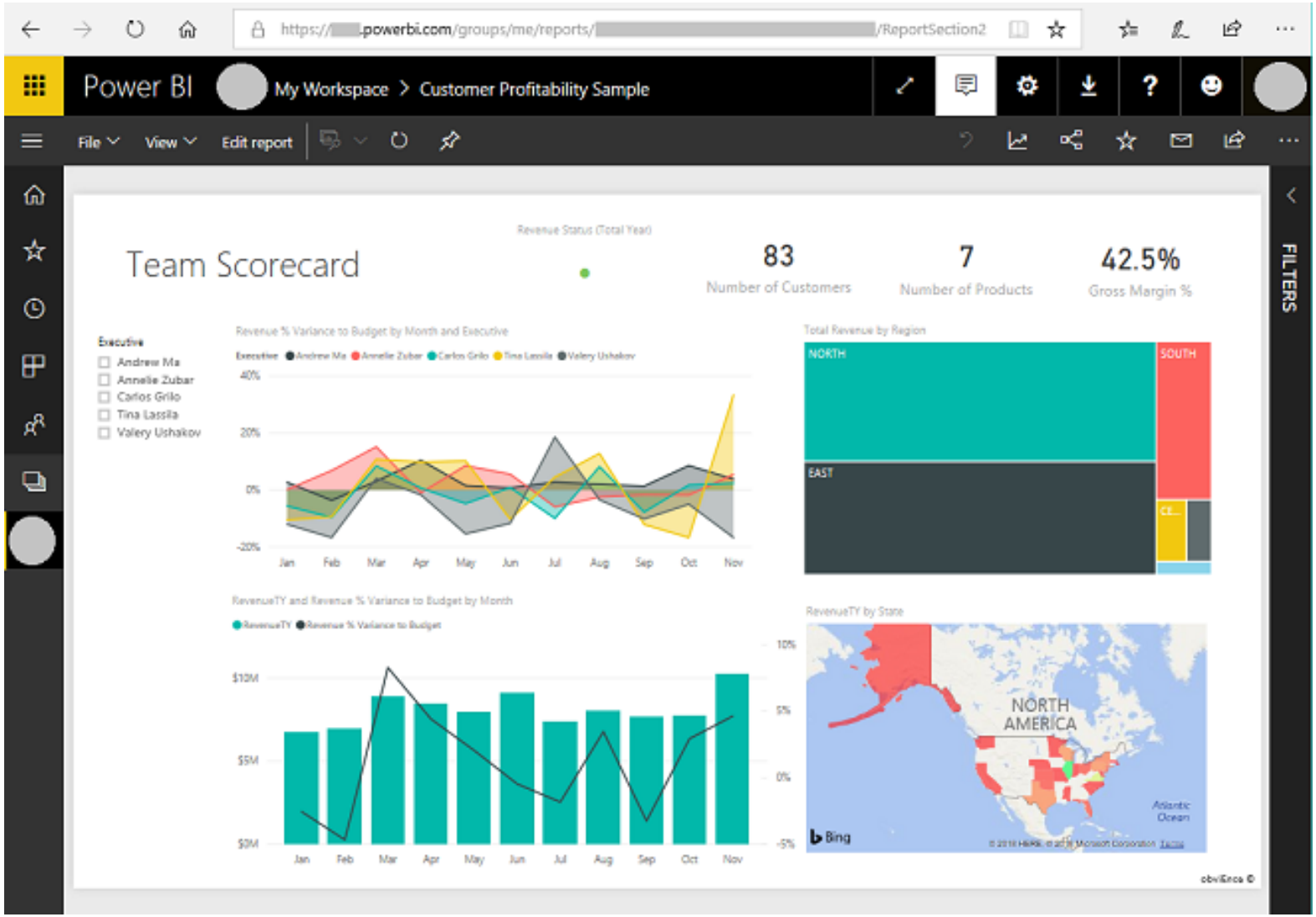
BUSINESS INTELLIGENCE (2)

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



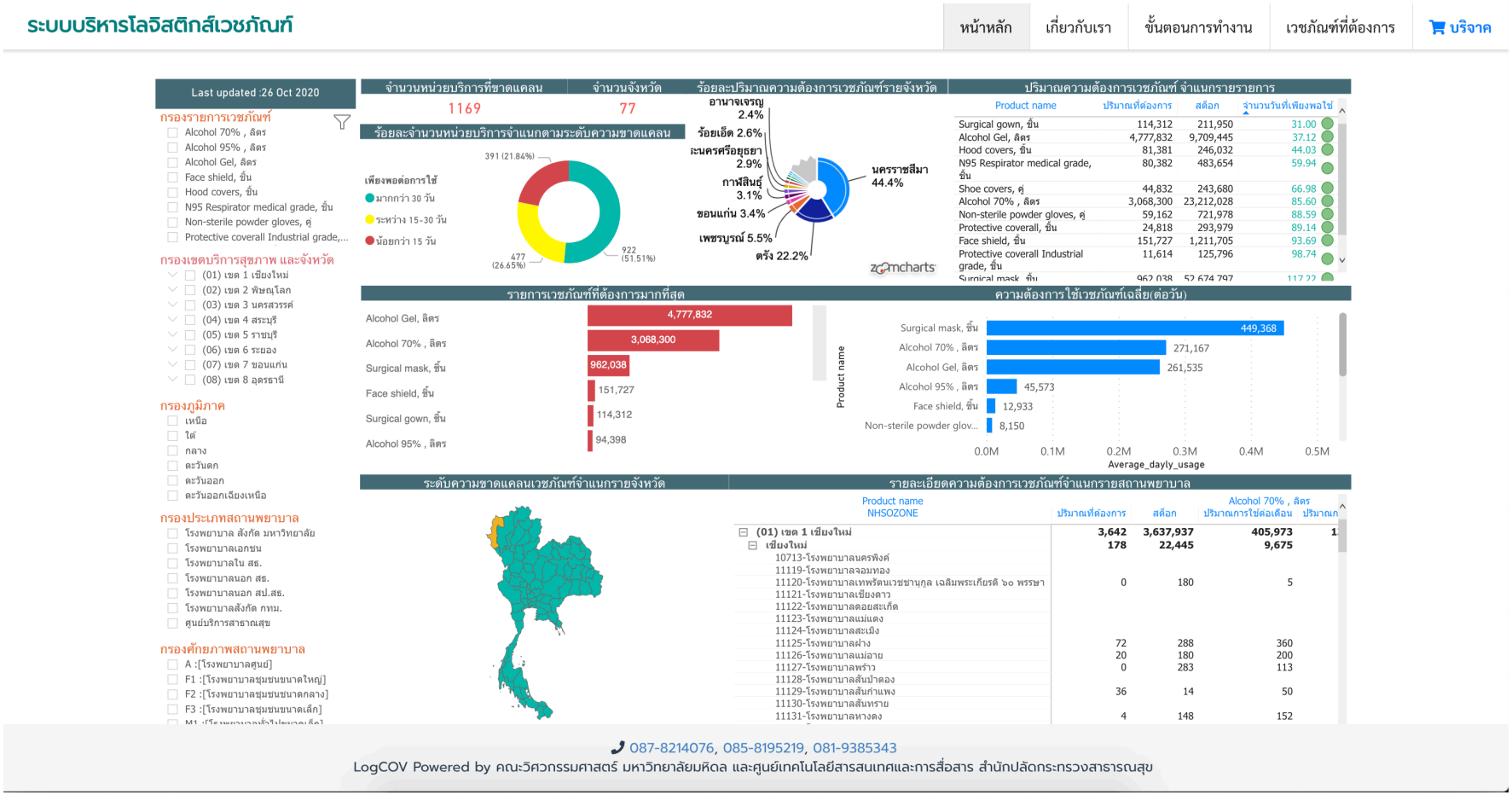
https://powerbi.microsoft.com/en-au/getting-started-with-power-bi/

TYPE OF POWER BI

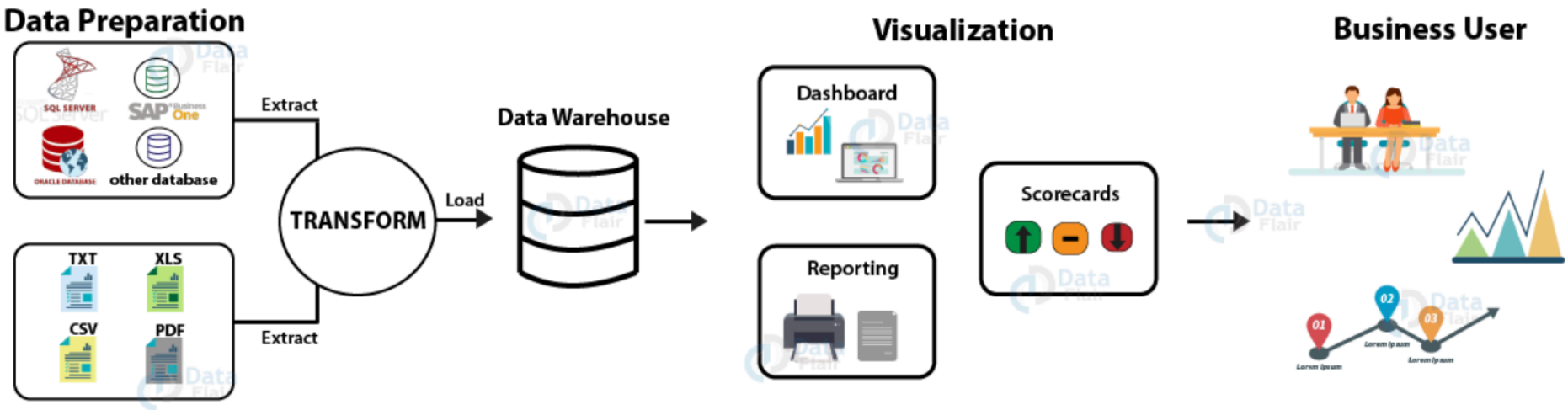
- Microsoft offers three types of Power BI platforms:
  - Power BI Desktop (Desktop Application)
  - Power BI Service (Software as a Service)
  - Power BI Mobile (For iOS and Android devices)

EXAMPLE OF BI

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

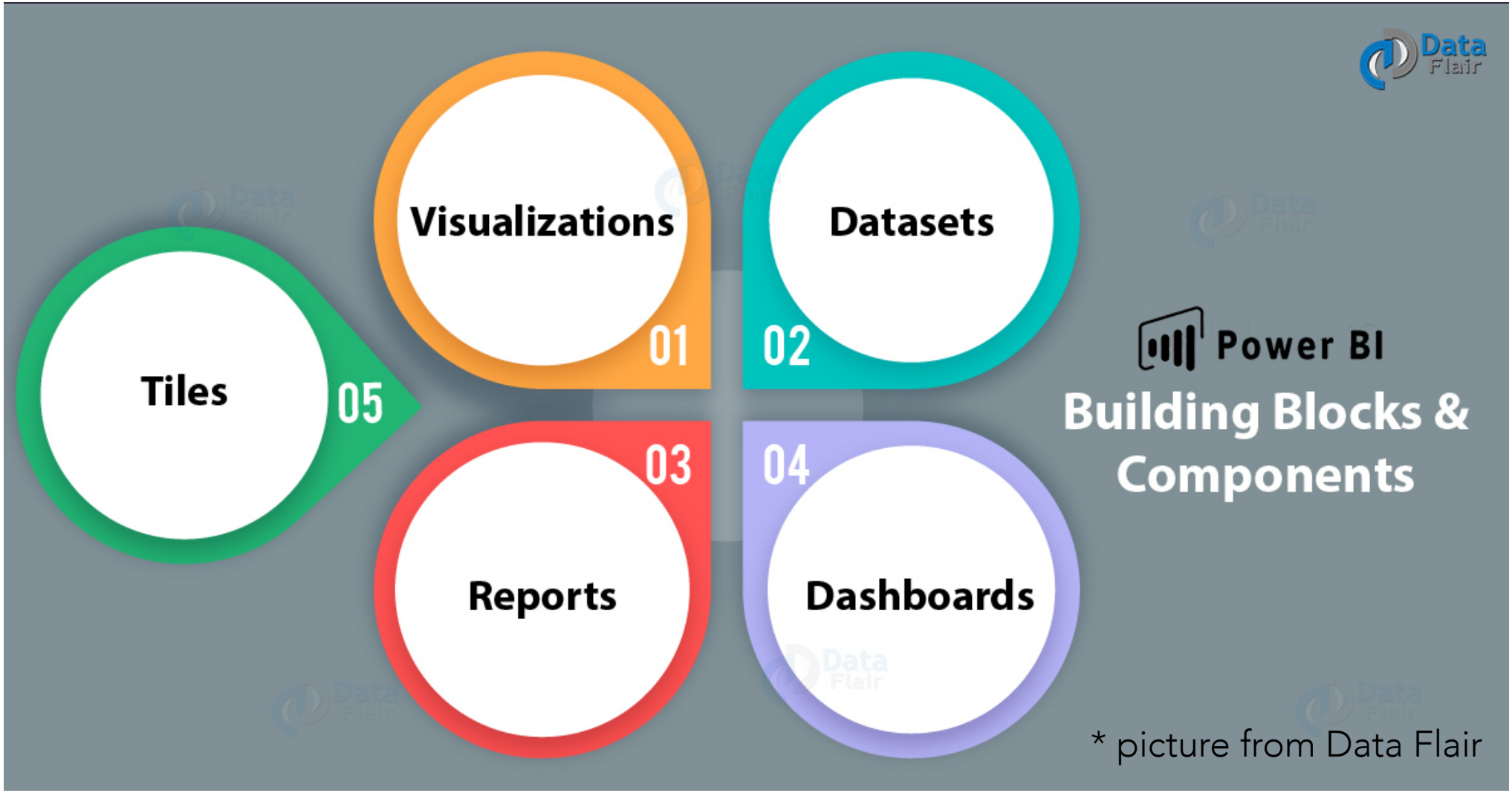


POWER BI ARCHITECTURE

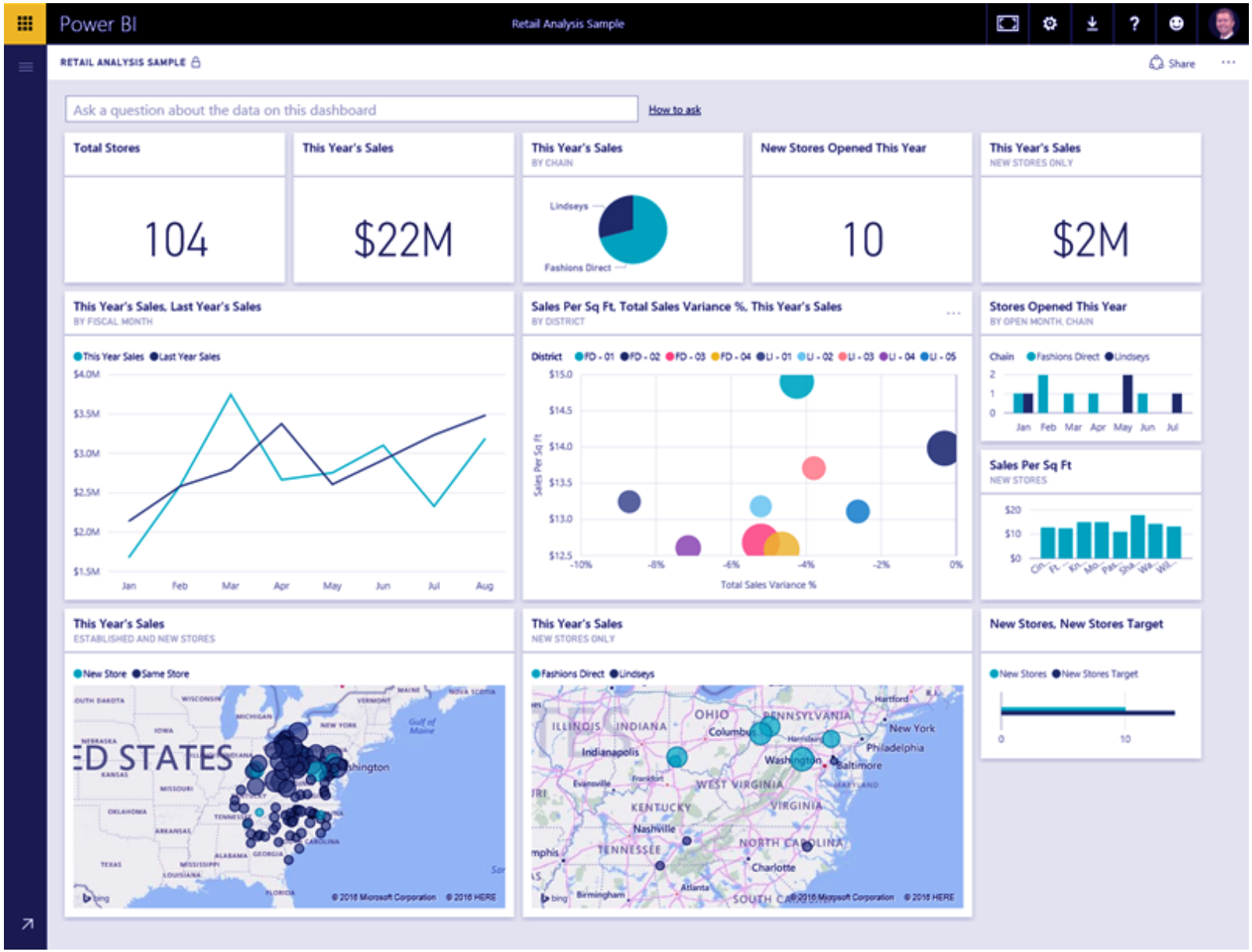


* picture from Data Flair

POWER BI BUILDING BLOCKS



VISUALISATION

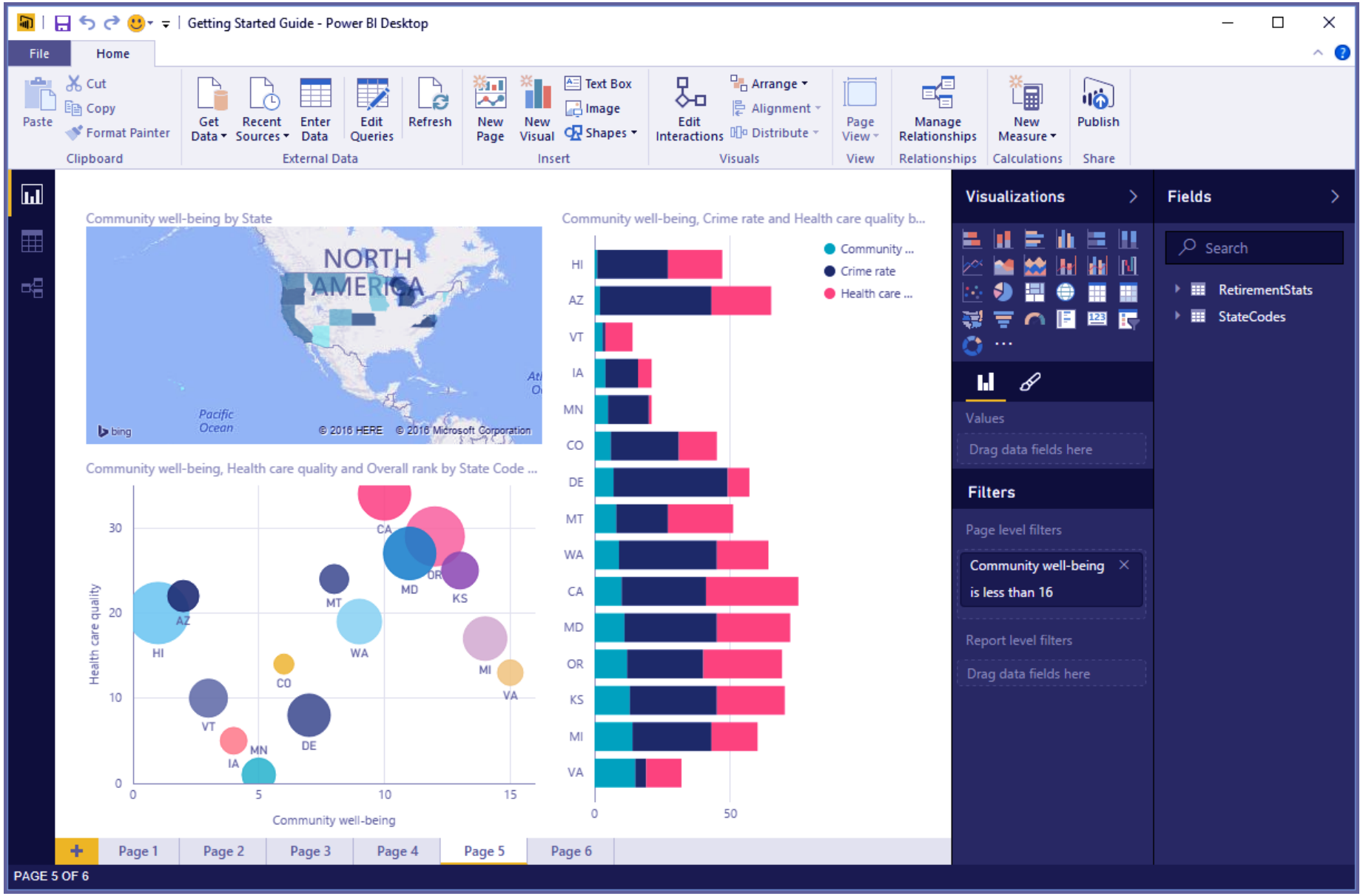


DATASETS

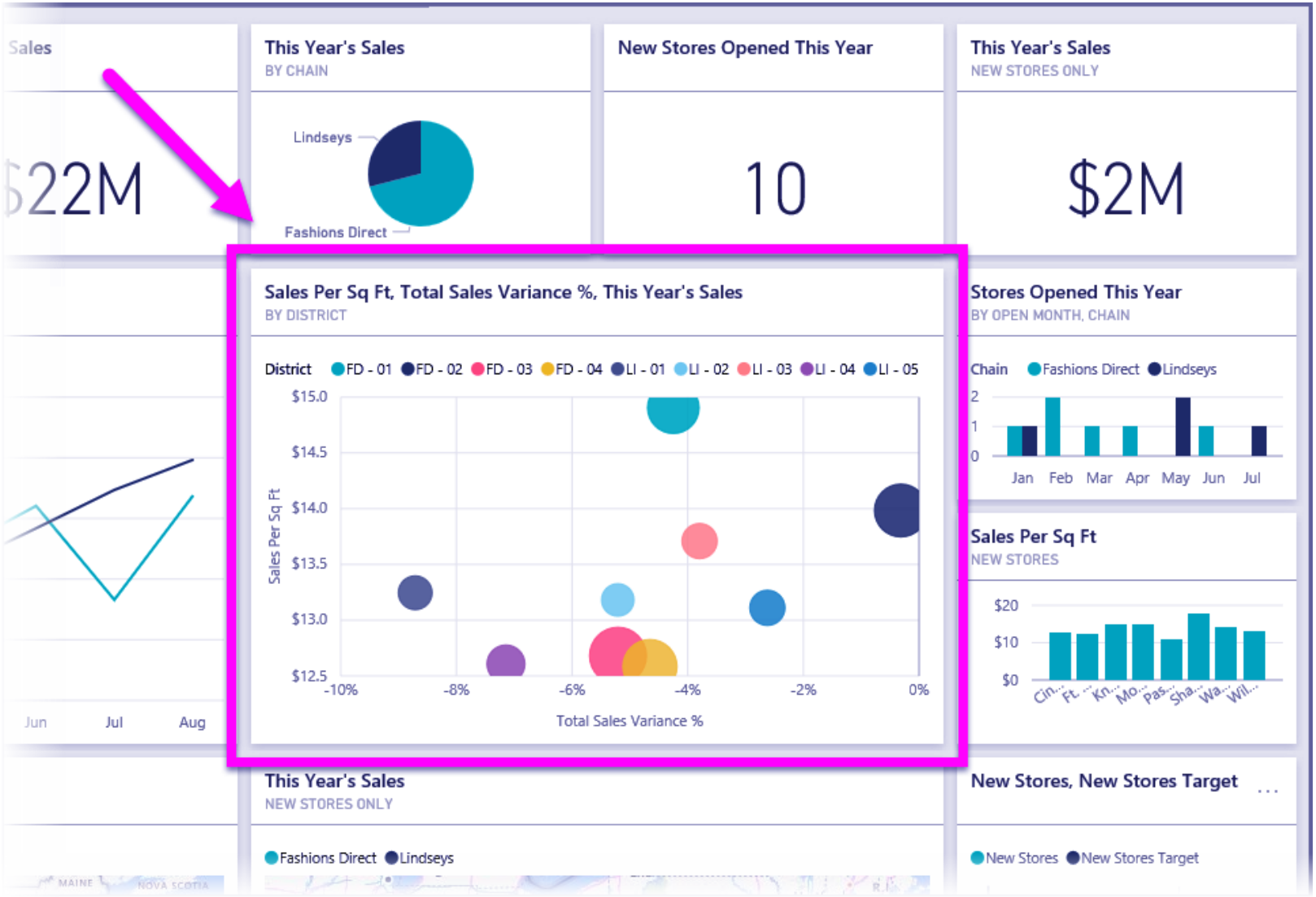
C2132 | 2

	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


Getting Started Guide - Power BI Desktop
File
Home
Paste
Cut
Copy
Format Painter
Clipboard
Get Data
Recent Sources
Enter Data
Edit Queries
Refresh
External Data
New Page
New Visual
Text Box
Image
Shapes
Insert
Edit Interactions
Arrange
Alignment
Distribute
Visuals
Page View
View
Manage Relationships
Relationships
New Measure
Calculations
Publish
Share
Community well-being by State
NORTH AMERICA
Pacific Ocean
© 2016 HERE © 2016 Microsoft Corporation
Community well-being, Crime rate and Health care quality b...
Community ...
Crime rate
Health care ...
HI
AZ
VT
IA
MN
CO
DE
MT
WA
CA
MD
OR
KS
MI
VA
0
50
Community well-being, Health care quality and Overall rank by State Code ...
Health care quality
Community well-being
0
10
20
30
5
15
Visualizations
Values
Drag data fields here
Filters
Page level filters
Community well-being
is less than 16
Report level filters
Drag data fields here
Fields
Search
RetirementStats
StateCodes
Page 1
Page 2
Page 3
Page 4
Page 5
Page 6
PAGE 5 OF 6

TILES


Sales
$22M
This Year's Sales
BY CHAIN
Lindseys
Fashions Direct
New Stores Opened This Year
10
This Year's Sales
NEW STORES ONLY
$2M
Jun
Jul
Aug
Sales Per Sq Ft, Total Sales Variance %, This Year's Sales
BY DISTRICT
District
FD - 01
FD - 02
FD - 03
FD - 04
LI - 01
LI - 02
LI - 03
LI - 04
LI - 05
$15.0
$14.5
$14.0
$13.5
$13.0
$12.5
Sales Per Sq Ft
-10%
-8%
-6%
-4%
-2%
0%
Total Sales Variance %
Stores Opened This Year
BY OPEN MONTH, CHAIN
Chain
Fashions Direct
Lindseys
2
1
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Sales Per Sq Ft
NEW STORES
$20
$10
$0
This Year's Sales
NEW STORES ONLY
Fashions Direct
Lindseys
New Stores, New Stores Target
New Stores
New Stores Target

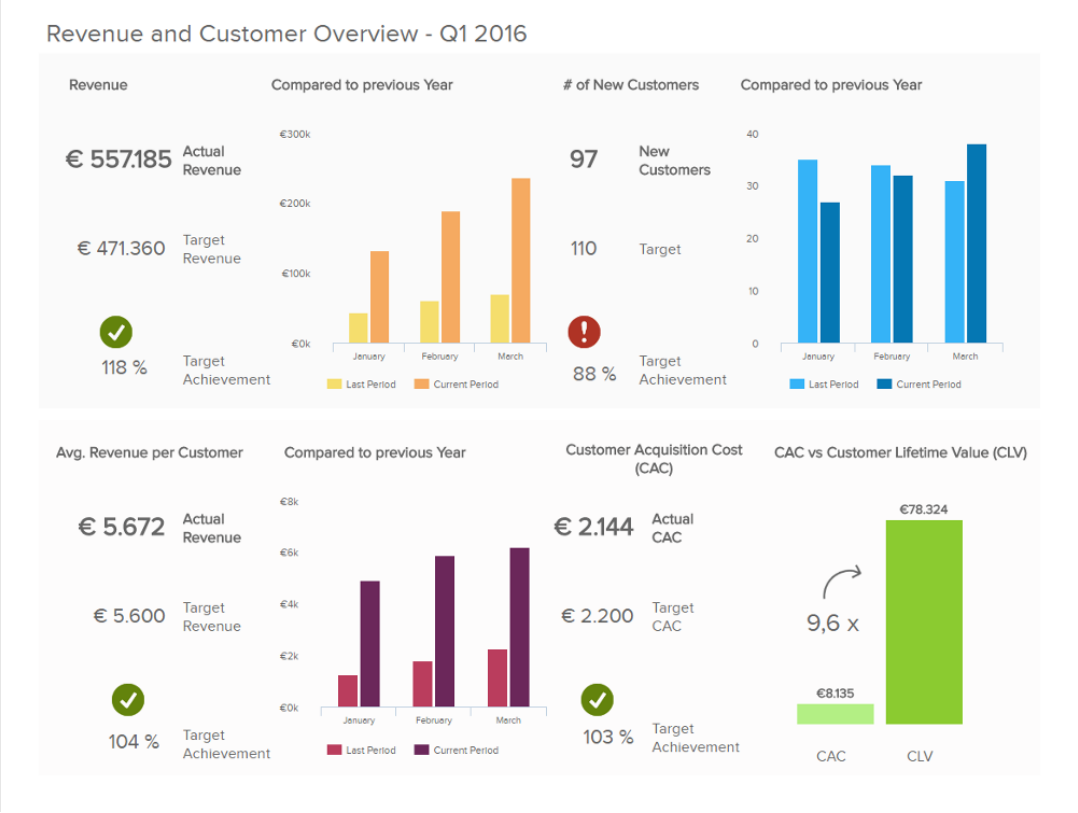
DASHBOARDS (1)

- 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 able to communicate with your dashboard when they're in Power BI benefit, or on their cell phone.

DASHBOARDS (2)

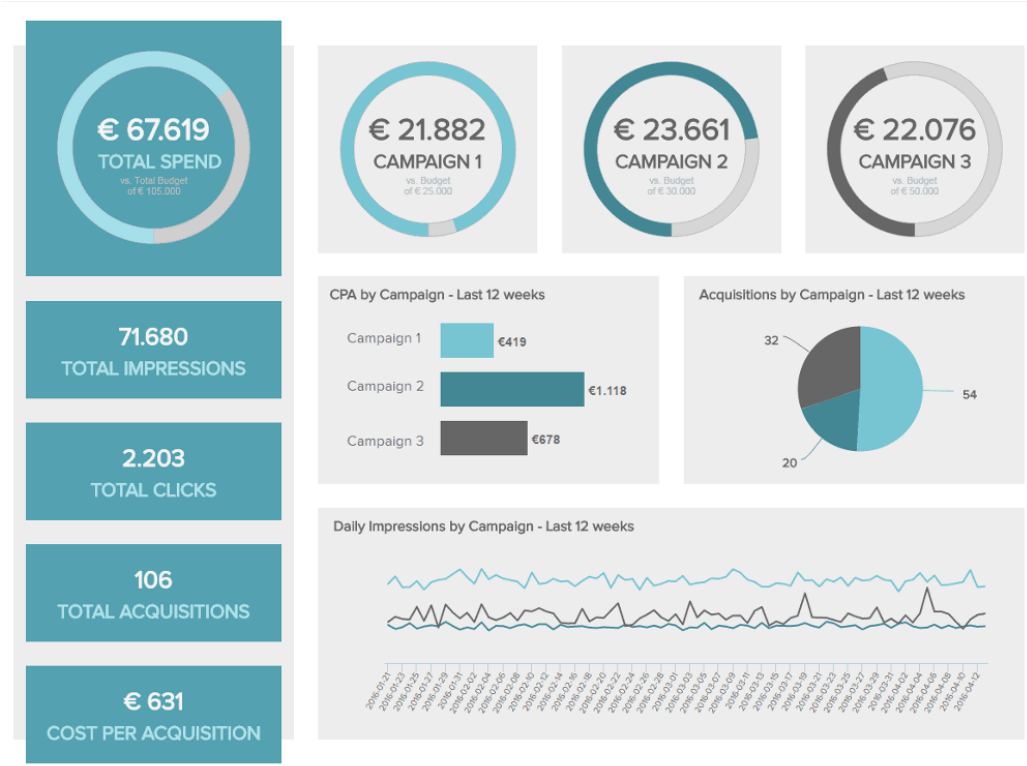
- There are several types of dashboards. Based on their purpose, they can be categorized into three main types:
  - Strategic Dashboard
  - Operational Dashboard
  - Analytical Dashboard

STRATEGIC DASHBOARD



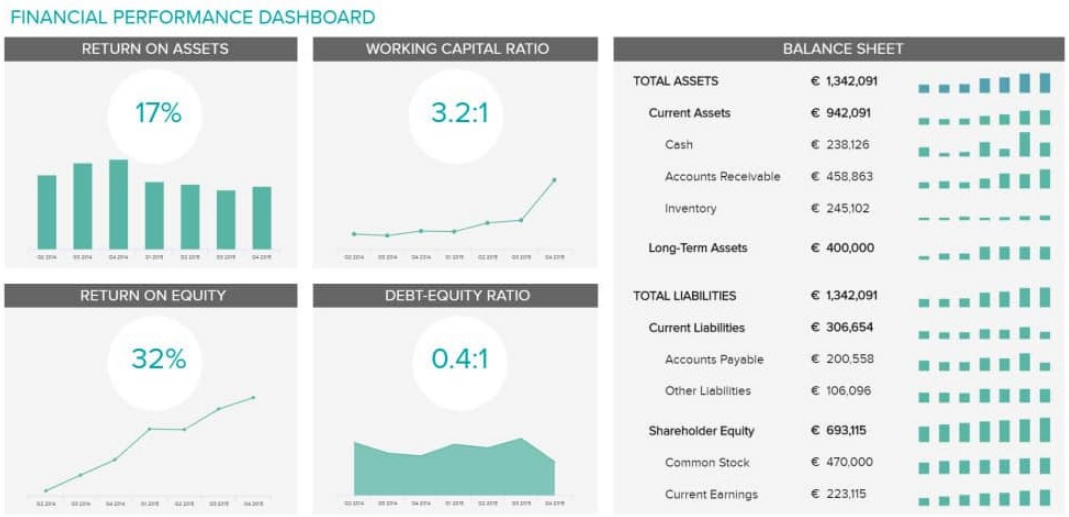
- Dashboard for strategic planning
- Decision making at all levels.
- Focuses on displaying key organizational indicators or KPIs, making it easy to understand and see the organization's overall performance.

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



- 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

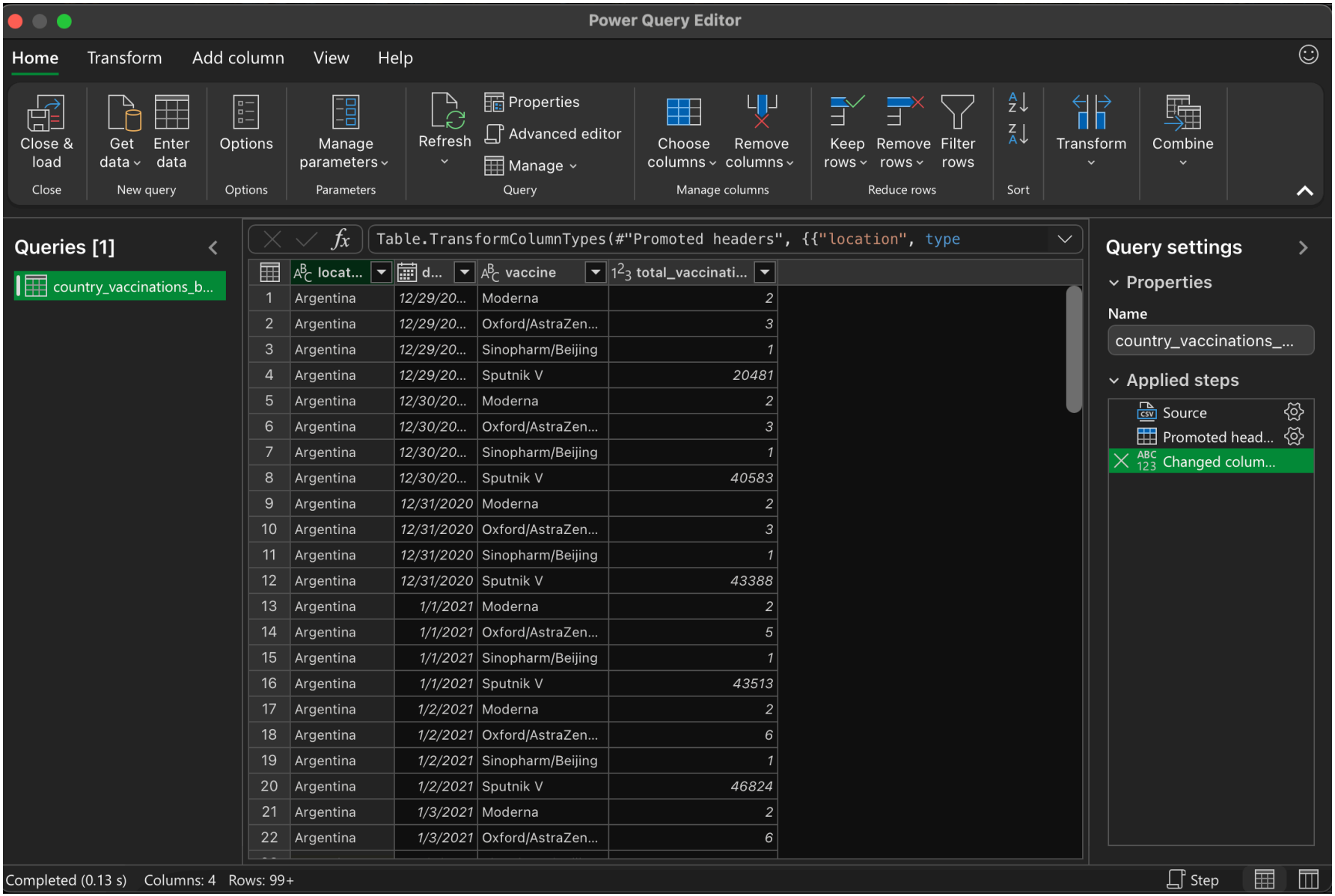
GETTING DATA

- 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)

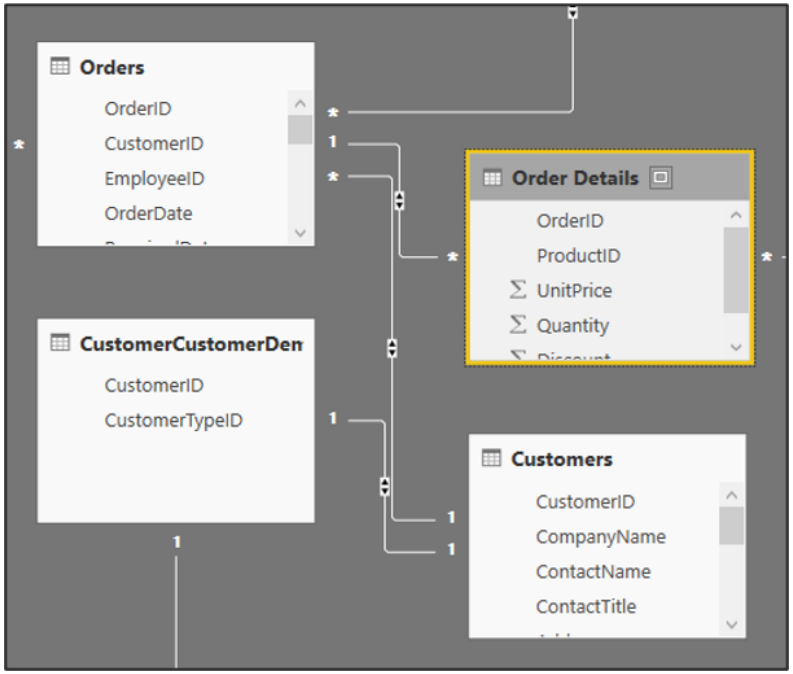


ONELAKE DATA HUB

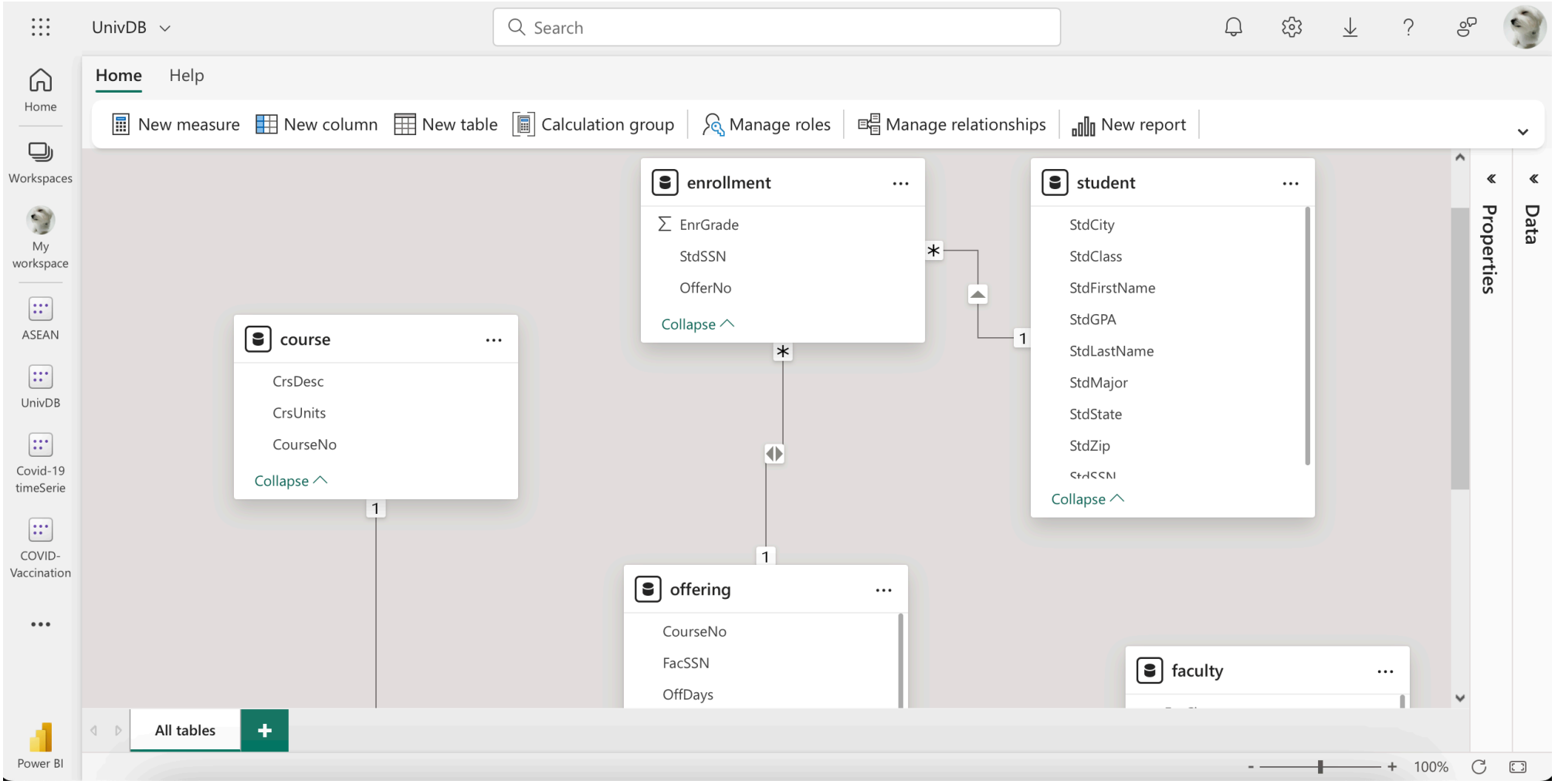
- Data Lake in Power BI typically refers to integrating Power BI with a data lake storage, commonly using Azure Data Lake Storage (ADLS).
- This allows users to centralize large datasets from multiple sources into a single, scalable storage.
- By connecting Power BI with a data lake, users gain robust 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

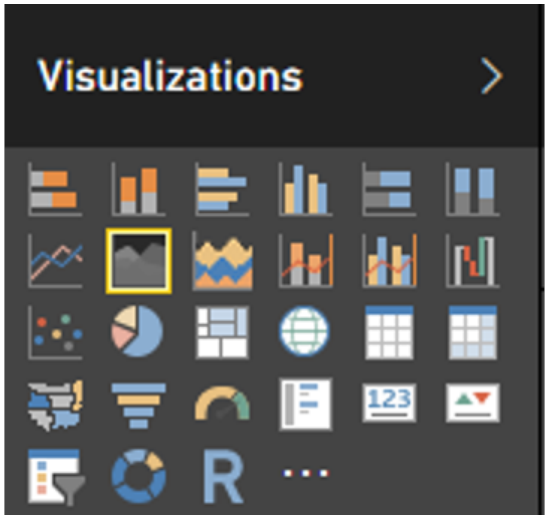


DATA MODELING (2)



VISUALISATION

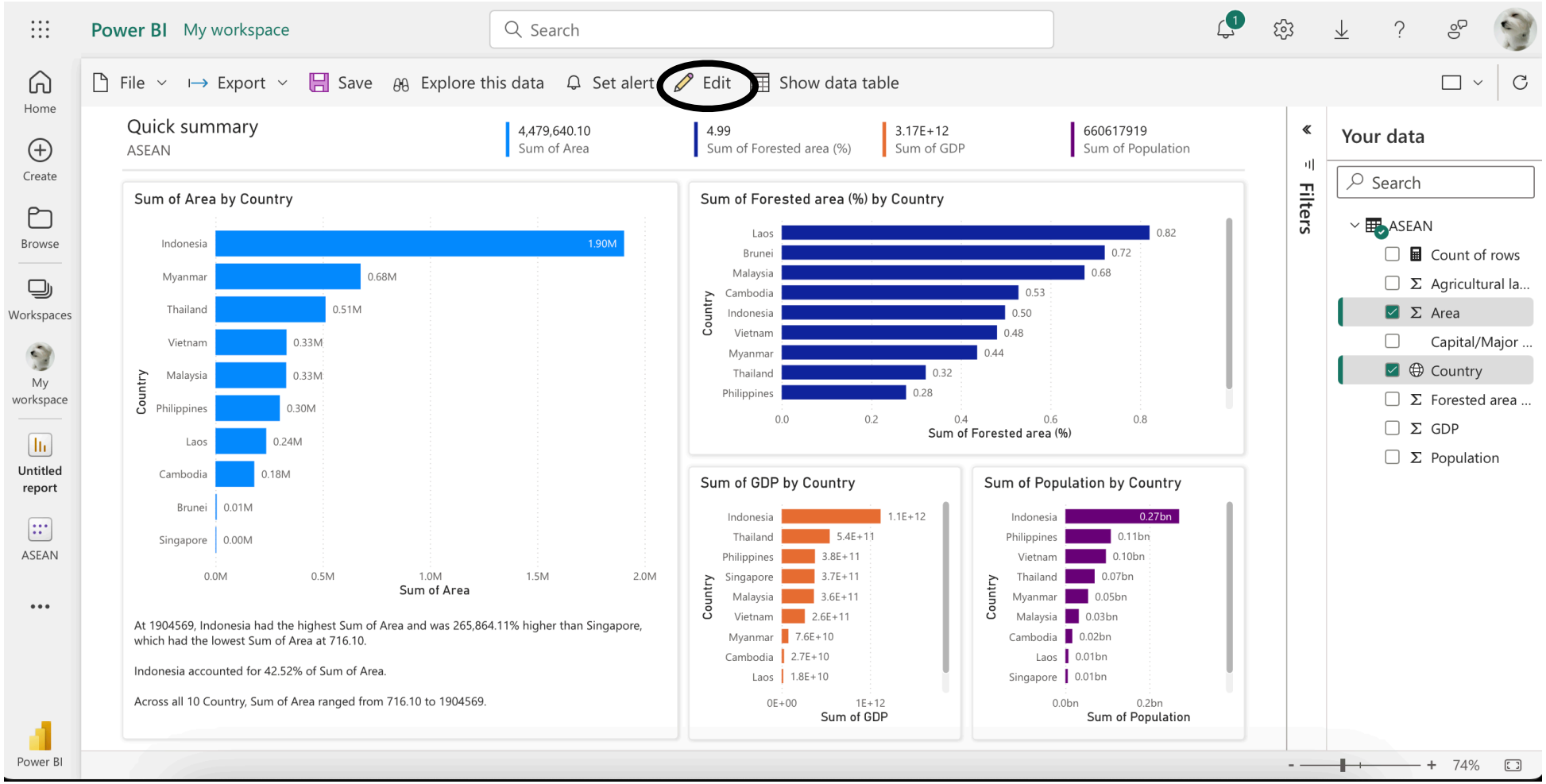
- A visualization is a visual representation of data, such as a chart, graph, map, or other “visual” representations of data.
  - Line, bar, pie, stacked
  - Matrix/pivots
  - Key performance
  - Tree maps
  - Geo and filled maps
  - Slicers/filters
  - Custom visualizations



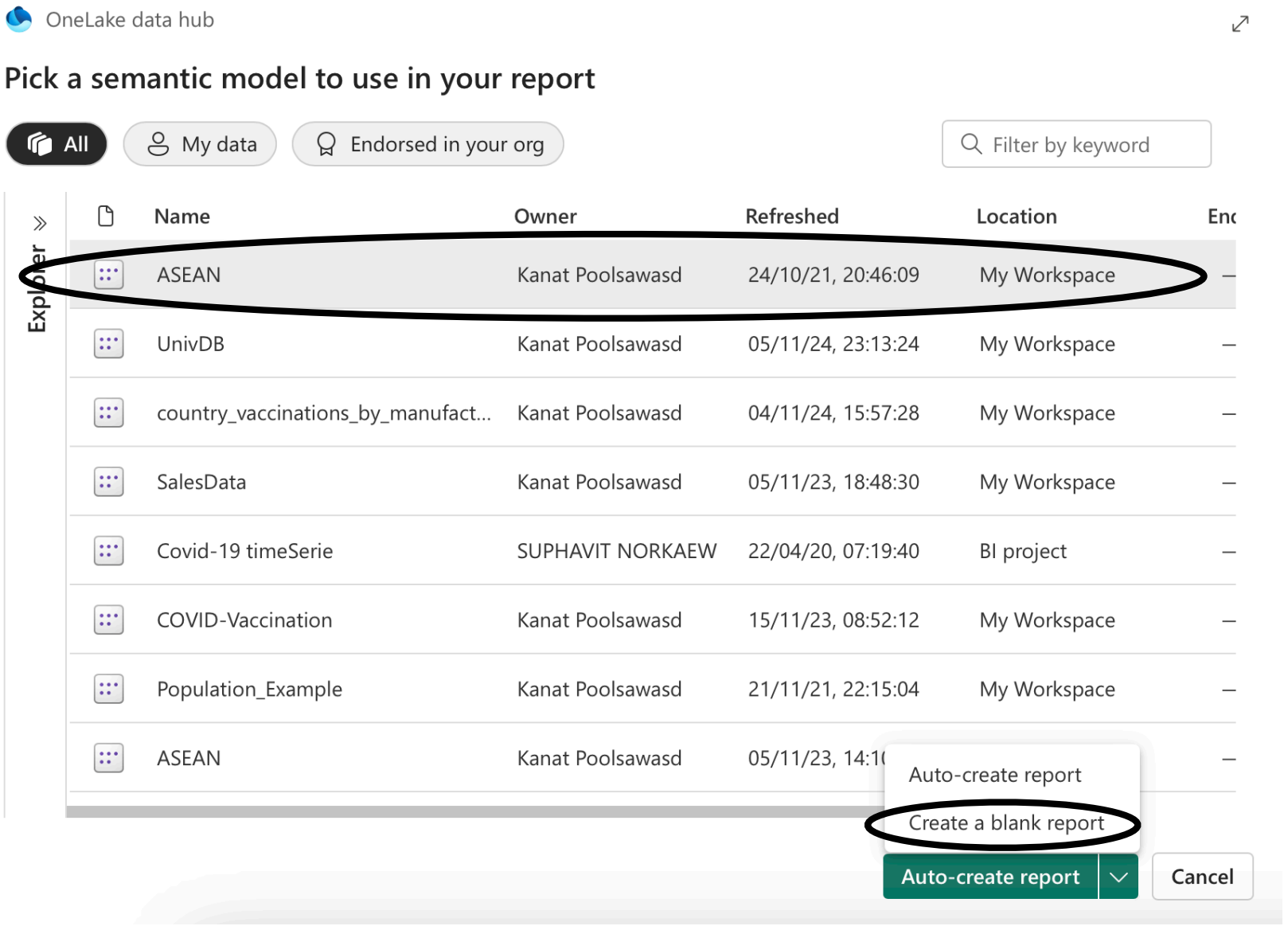
DATA ANALYSIS EXPRESSIONS

- 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.

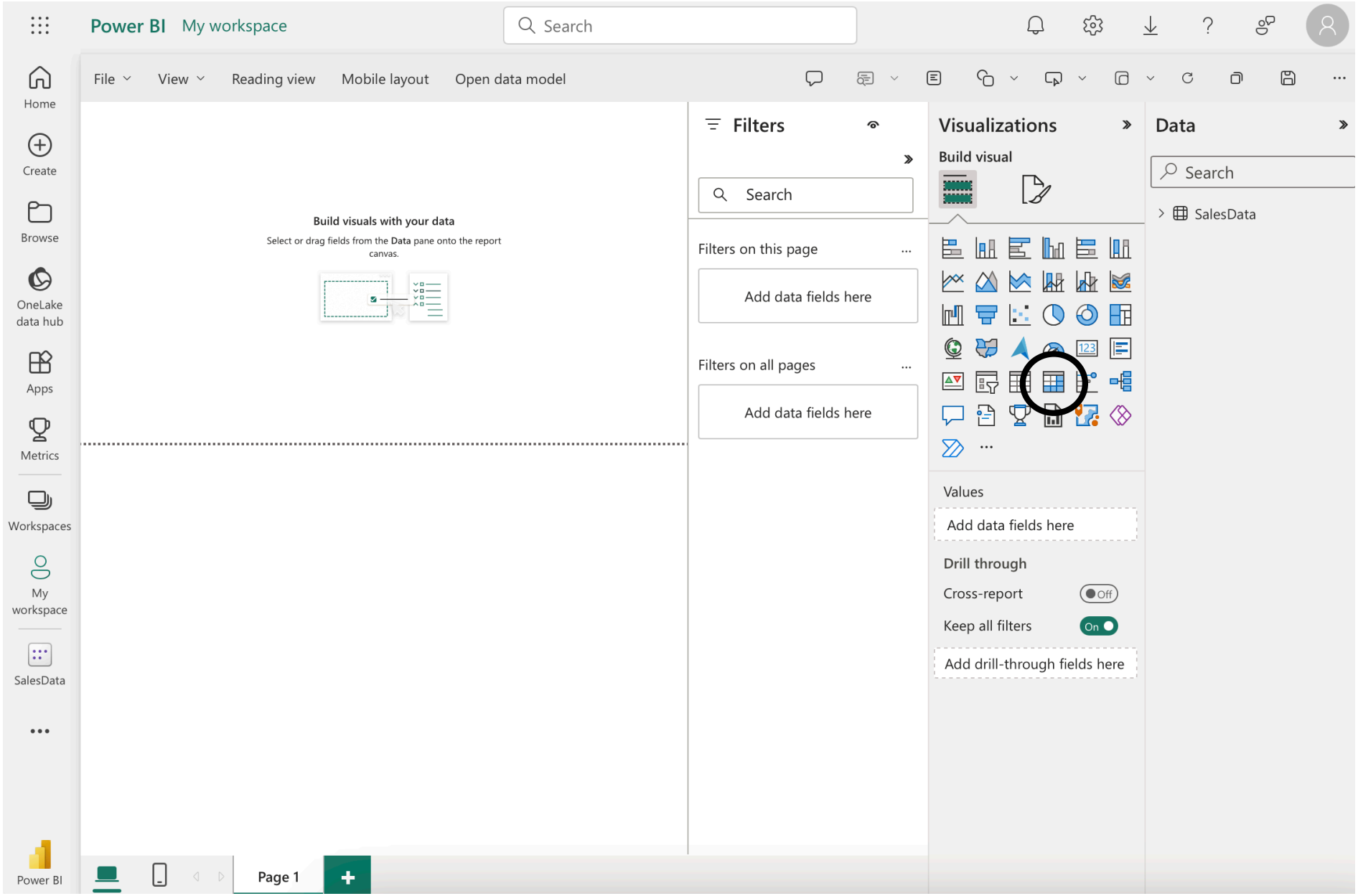
VISUALIZE DATA



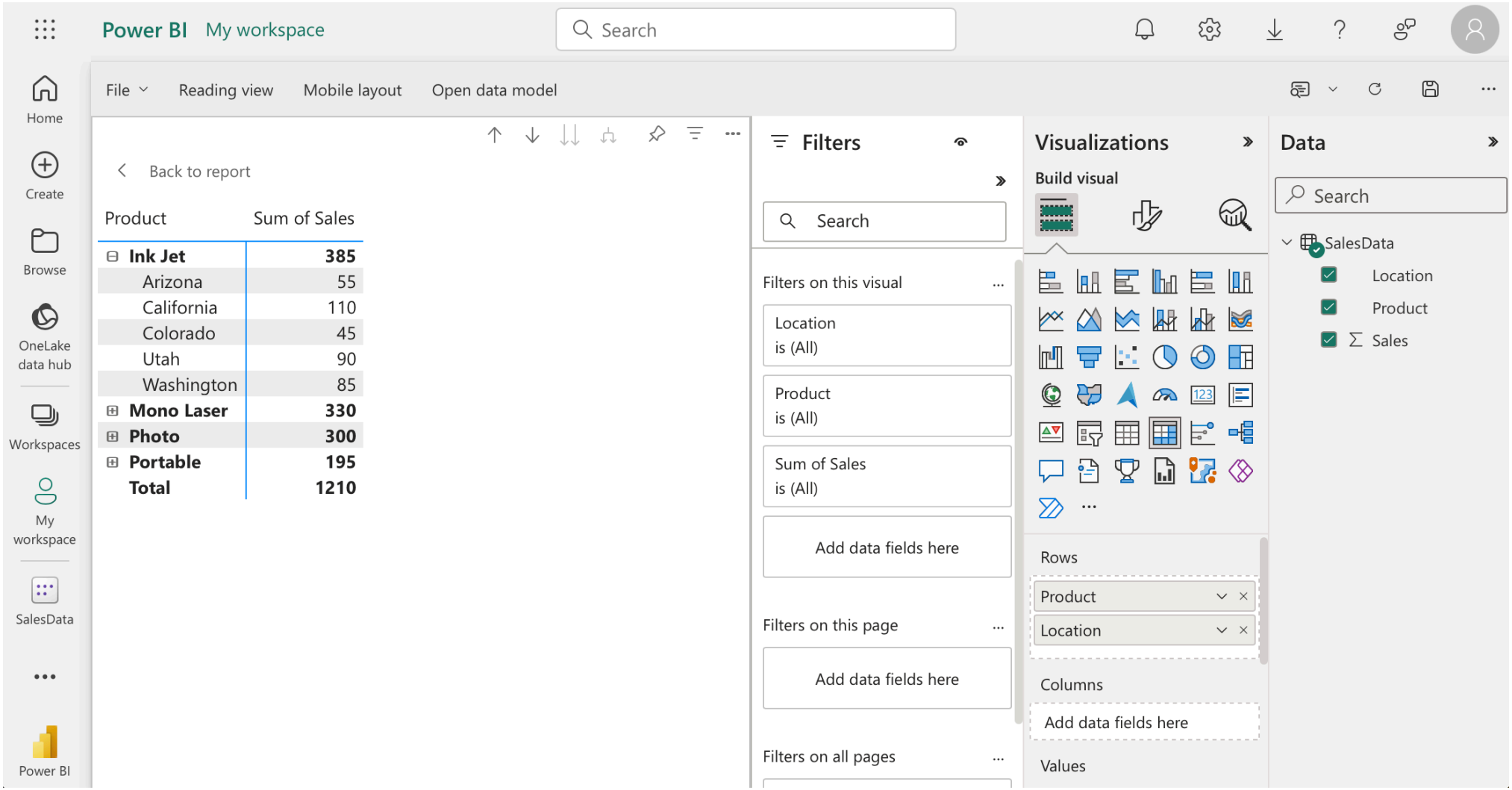
CREATE PIVOT (MATRIX) TABLE



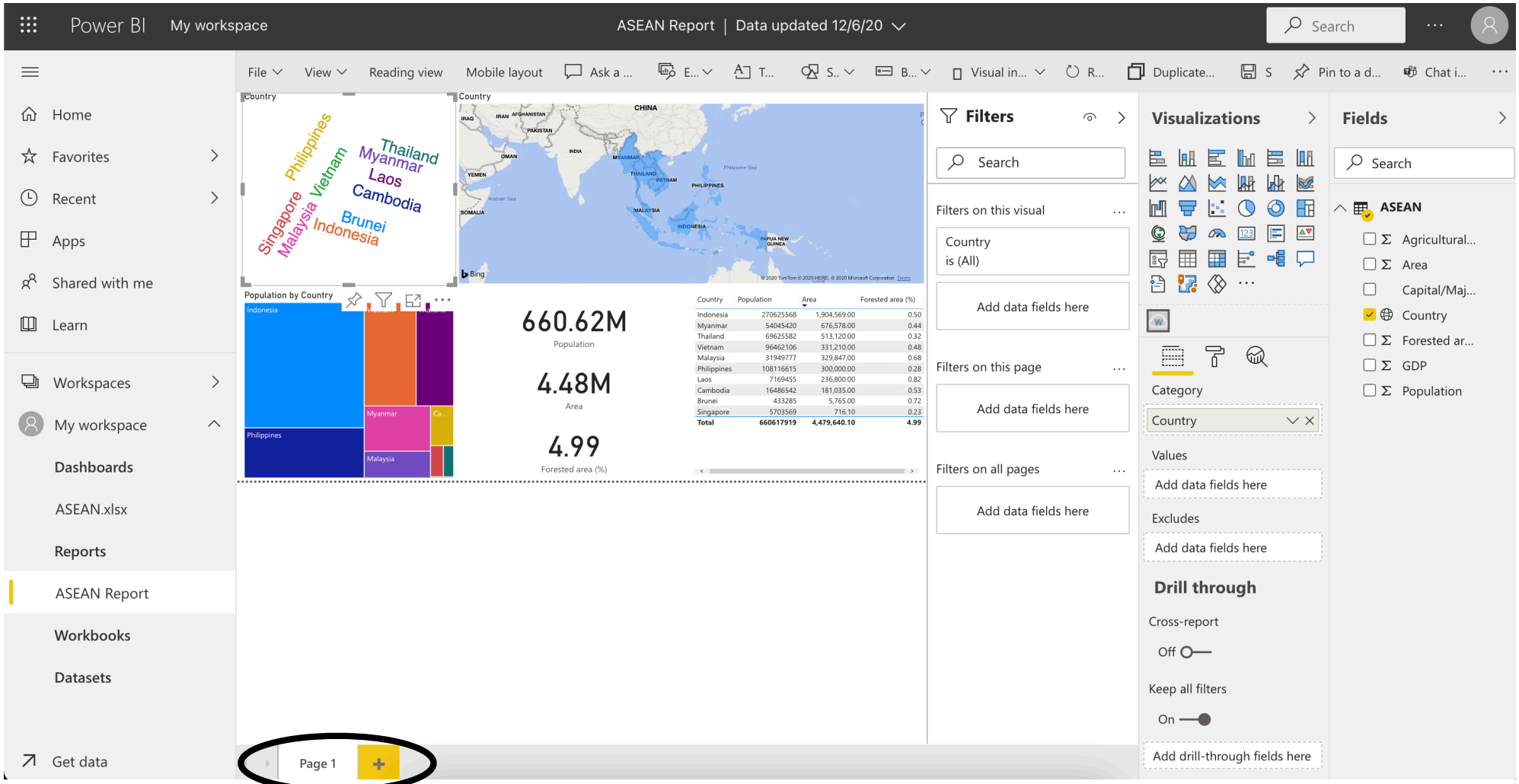
CREATE PIVOT (MATRIX) TABLE



CREATE PIVOT (MATRIX) TABLE



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=eyJrIjoiYz Copy

HTML you can paste into a website

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

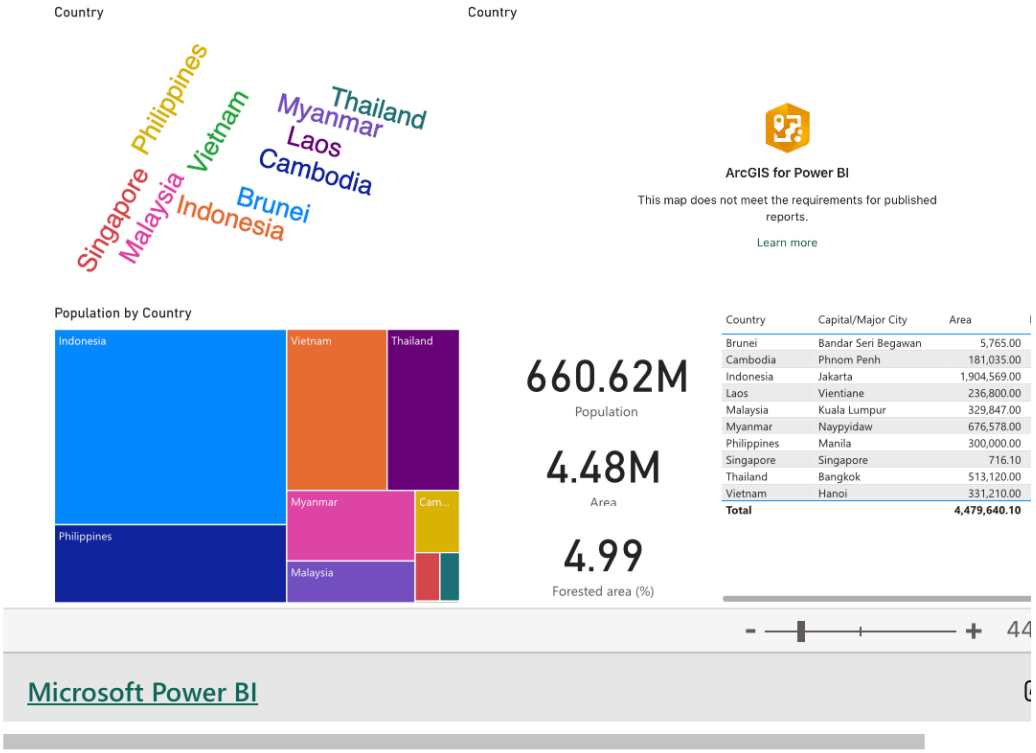
Size 600 x 373.5 px

Placeholder image

Upload

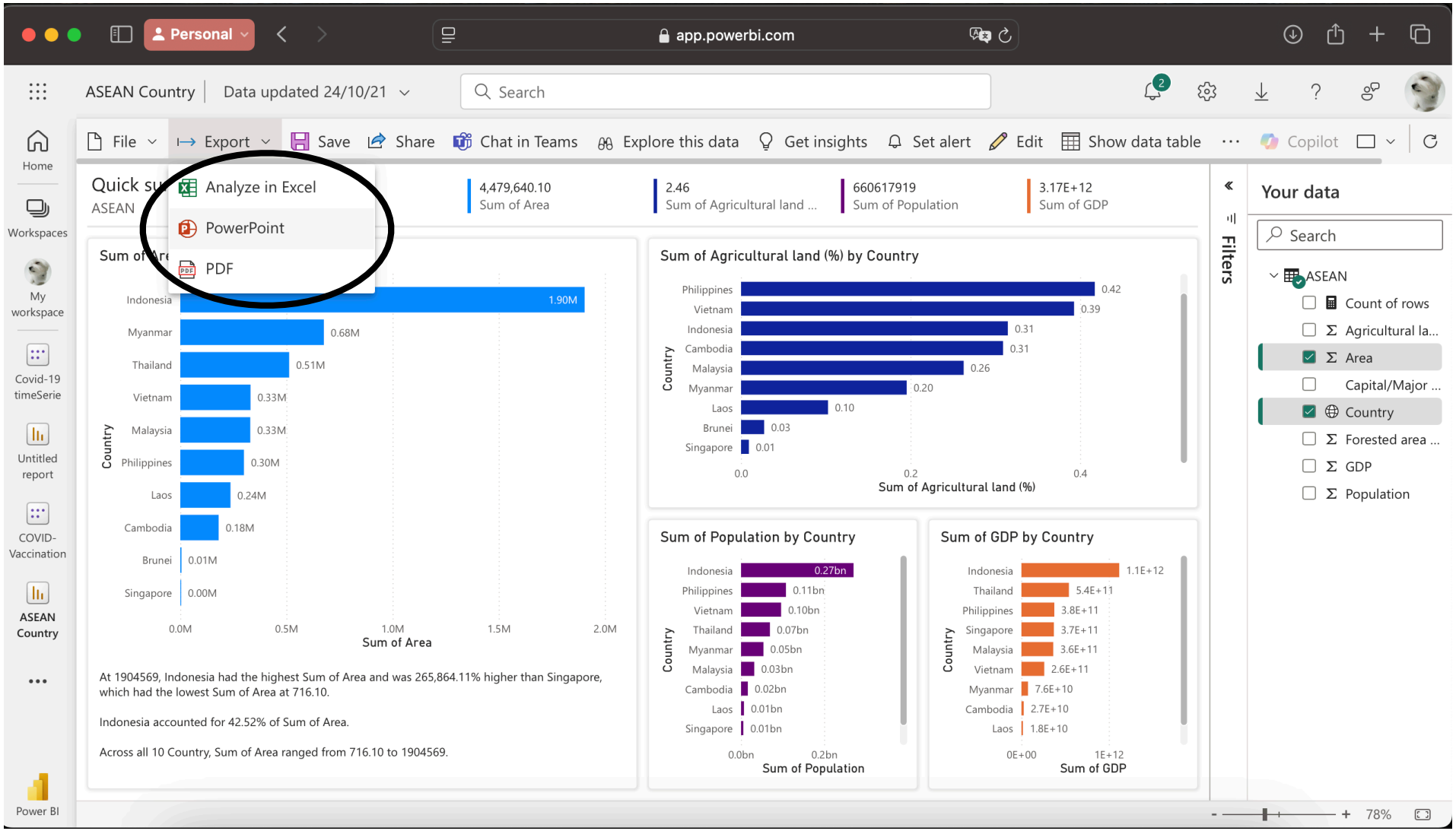
Delete

Default Page Default

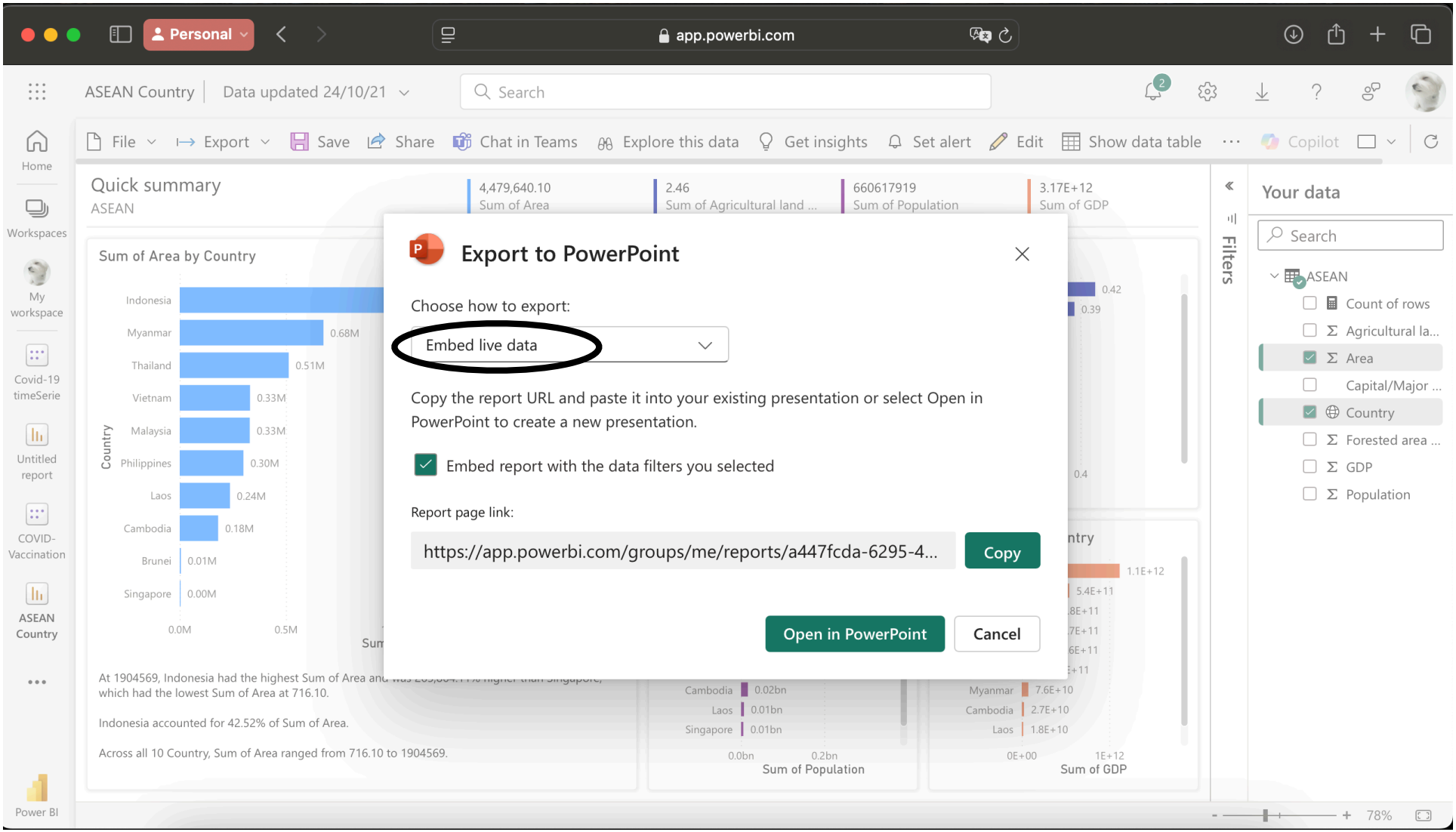


Close

EXPORT REPORT (1)



EXPORT REPORT (2)

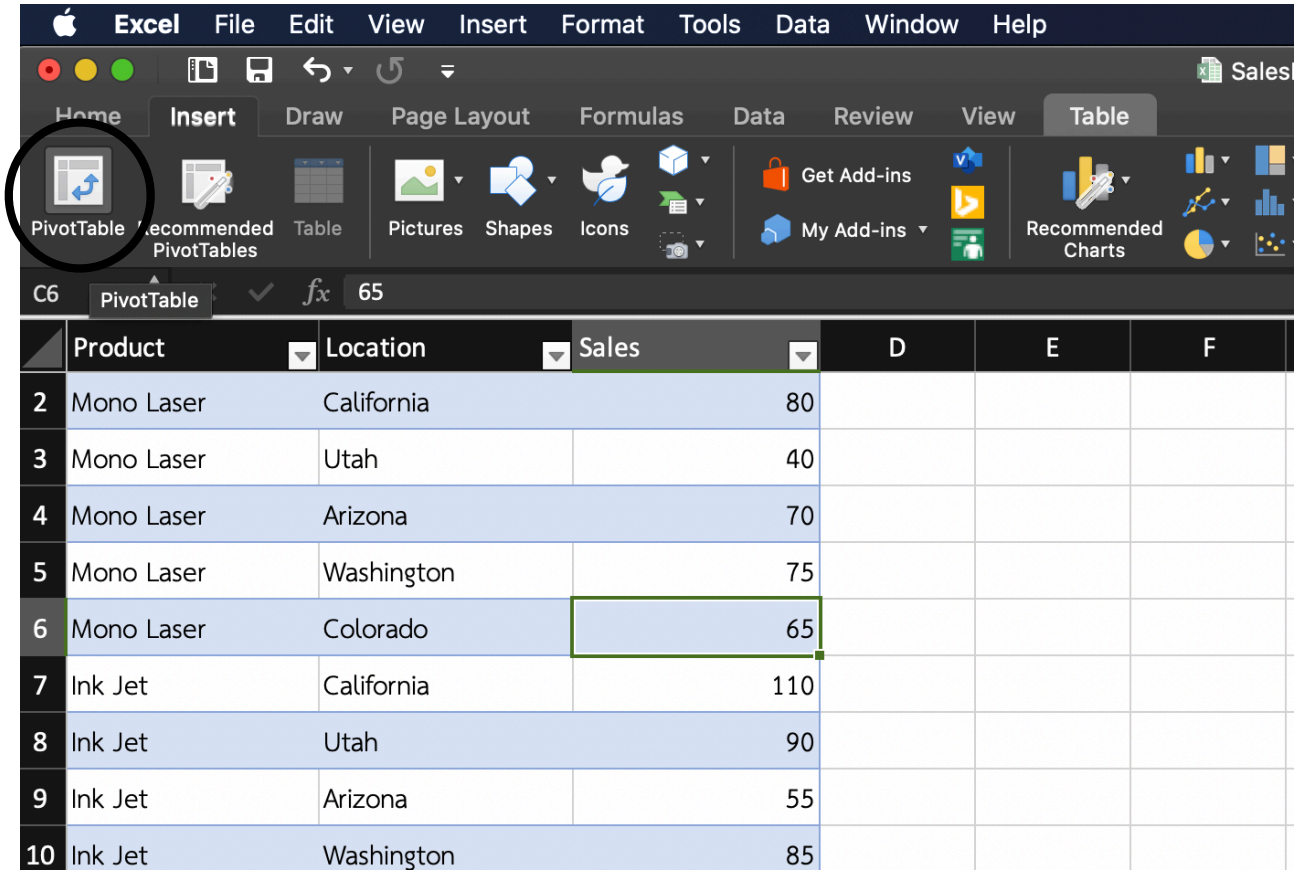


PIVOT TABLES IN EXCEL (1)

- 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)

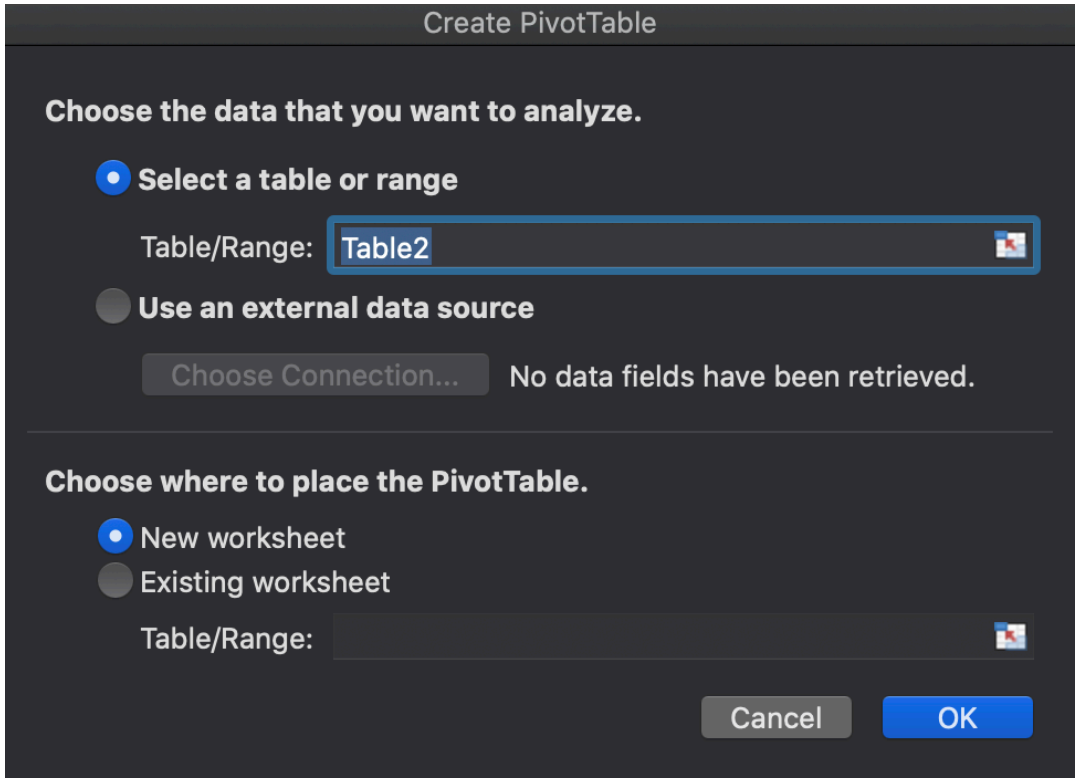
- Click any single cell inside the data set.
- On the Insert tab, in the Tables group, click PivotTable.



	Product	Location	Sales	D	E	F
2	Mono Laser	California	80			
3	Mono Laser	Utah	40			
4	Mono Laser	Arizona	70			
5	Mono Laser	Washington	75			
6	Mono Laser	Colorado	65			
7	Ink Jet	California	110			
8	Ink Jet	Utah	90			
9	Ink Jet	Arizona	55			
10	Ink Jet	Washington	85			

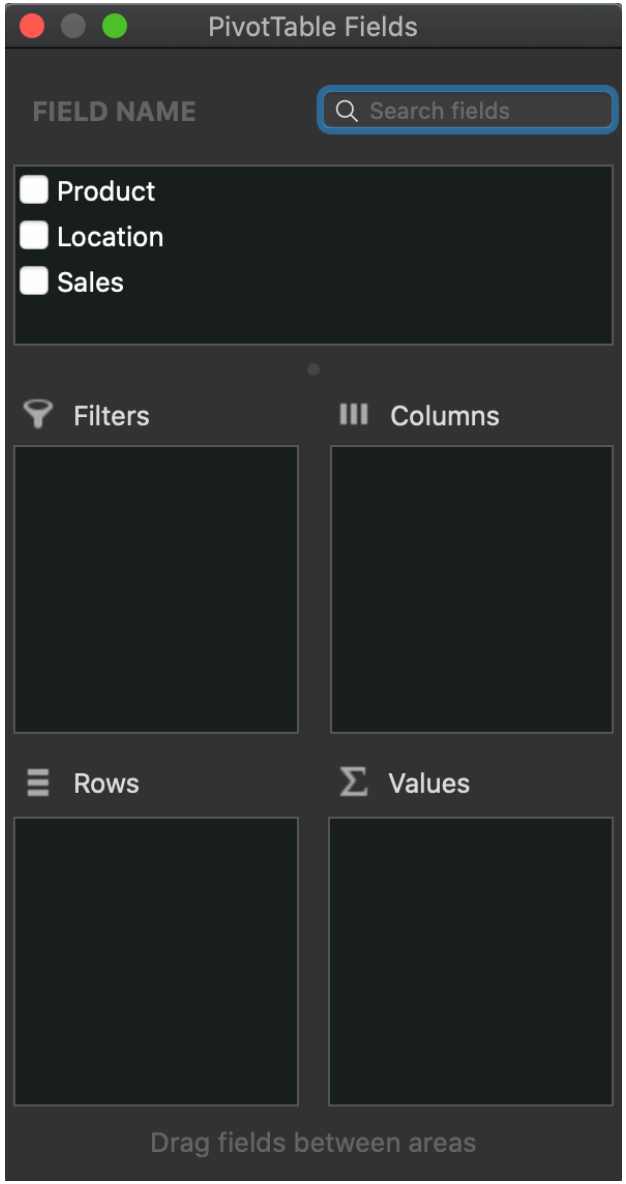
PIVOT TABLES IN EXCEL (3)

- 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

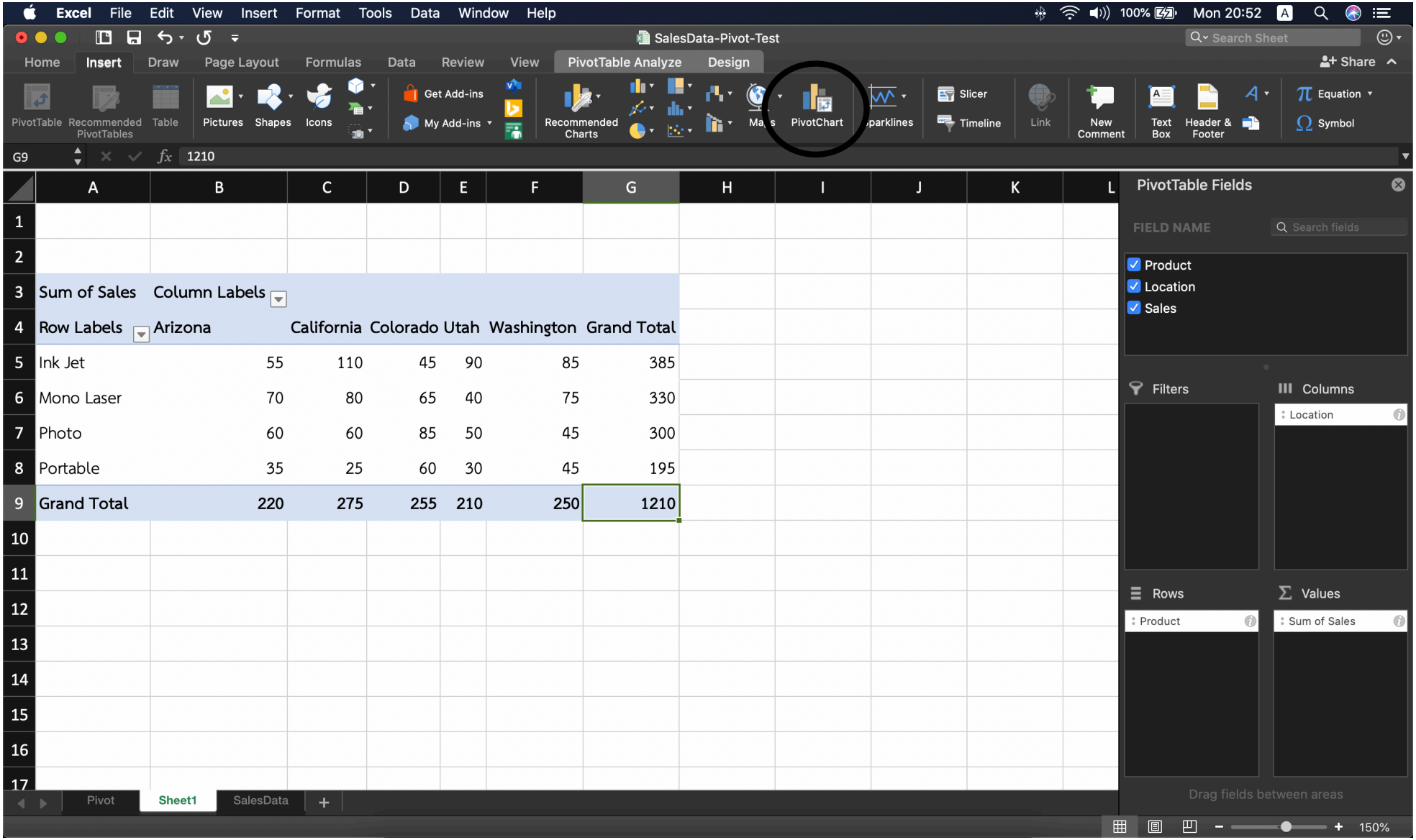


PIVOT TABLES IN EXCEL (4)

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

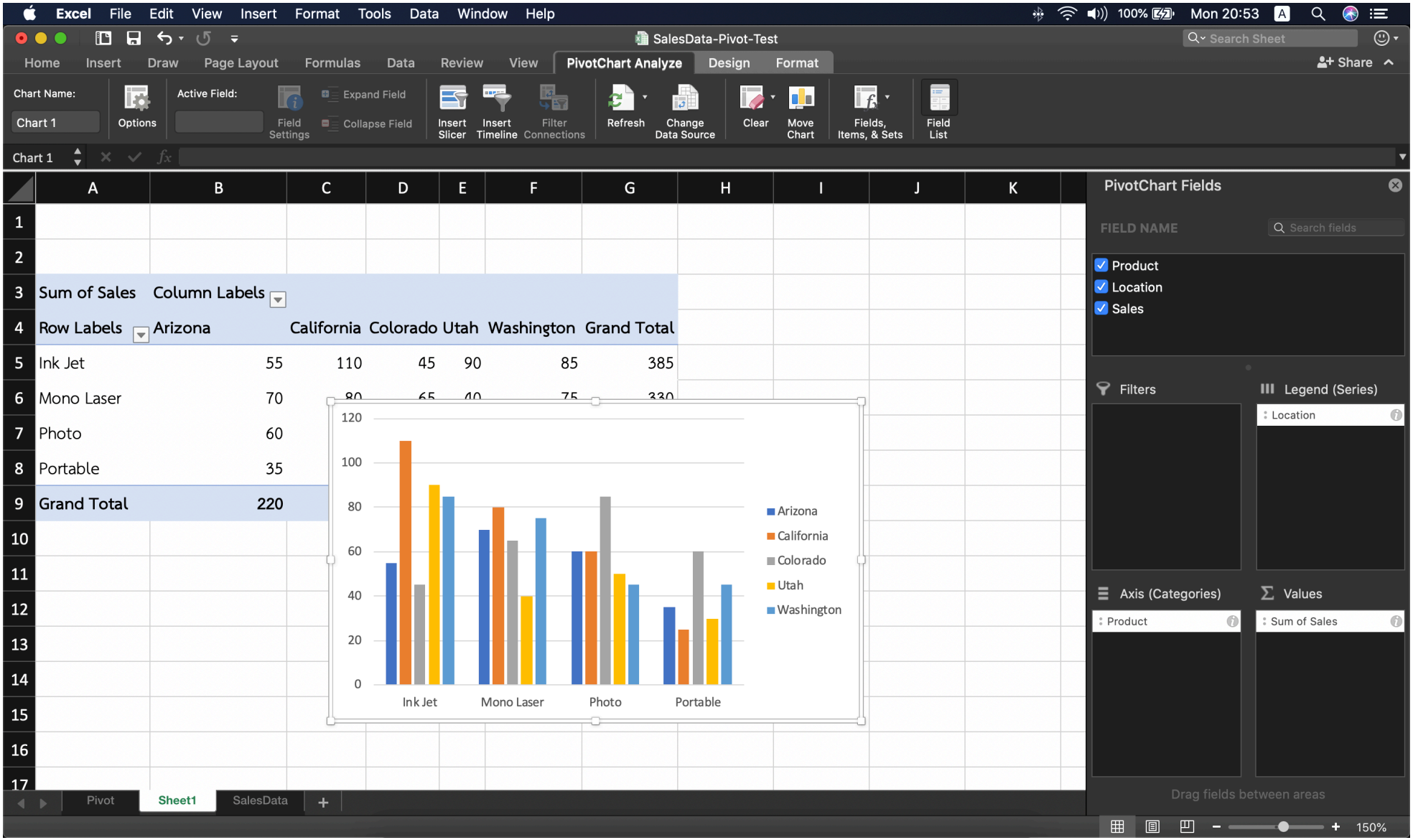


PIVOT TABLES IN EXCEL (5)



Sum of Sales	Column Labels					
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

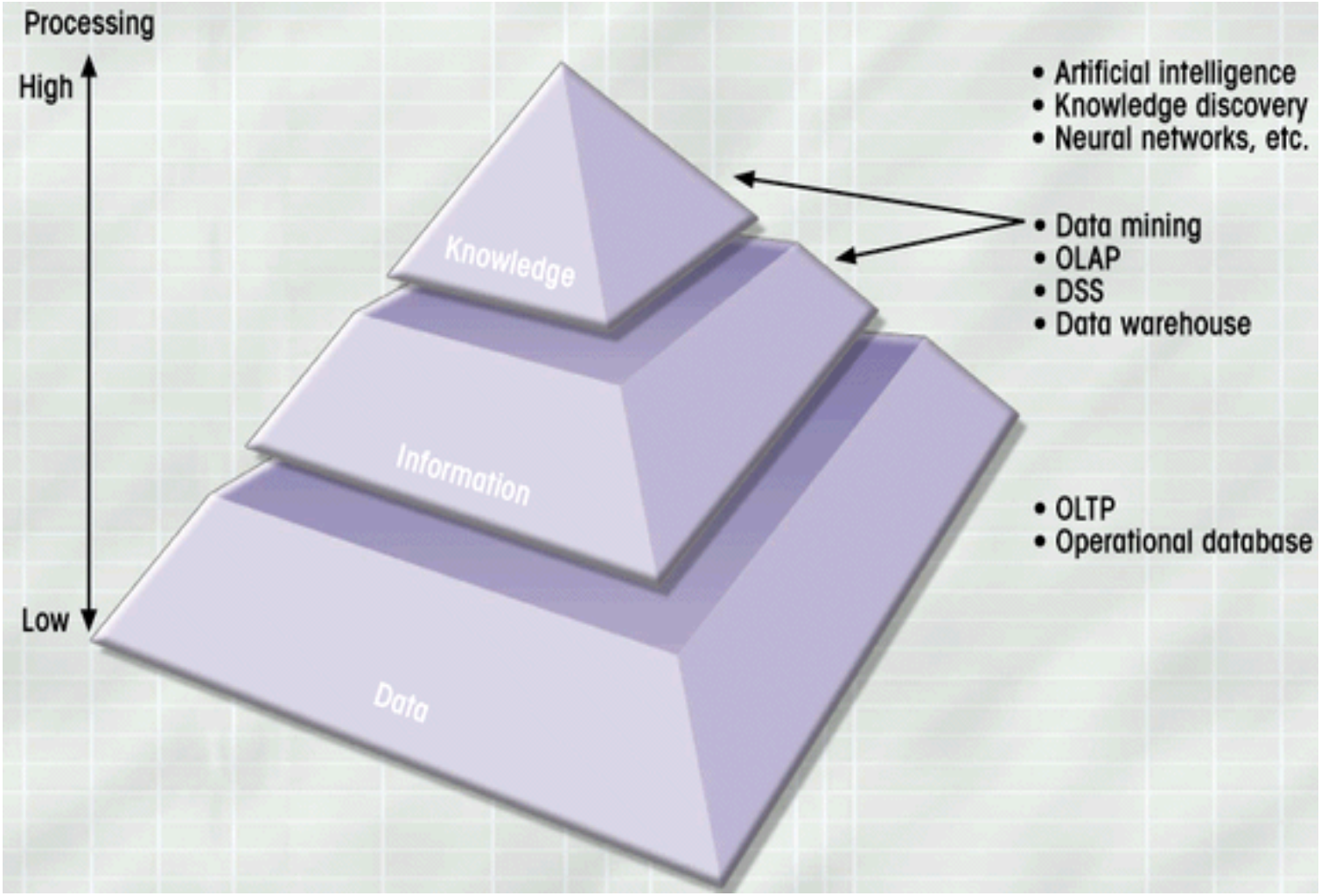
PIVOT TABLES IN EXCEL (6)



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 คิดว่าควรทำโปรโมชันที่สินค้าสี (Color) อะไร
- ตอบคำถามพร้อมส่งไฟล์ SampleData.xlsx ที่สร้าง Pivot Table สำหรับตอบคำถามข้างต้น พร้อมพิมพ์คำตอบลงในไฟล์ Excel แล้วกลับมาที่อีเมล kanat@egco.org