Microsoft Power Tools for Data Analysis #02:

Data Analysis & Business Intelligence Terms

Notes from Video:

Table of Contents:

1.	Data Analysis	2
2.	Business Intelligence	2
3.	Goal of Data Analysis and Business Intelligence:	2
4.	Query	3
5.	Raw Data	3
6.	Proper Data Set	4
7.	Field Names / Column Names / Header Names	4
8.	Naming	5
9.	Data Type	5
10.	Excel Table feature	5
11.	Grain or Granularity	6
12.	Keys	7
13.	Fact Table / Transaction Table	8
14.	Dimension Table / Lookup Table / Entity Tables	8
15.	Is a number a fact or an attribute?	9
16.	History	9
17.	Relationships	10
18.	Data Models	
ii.	Star Schema Data Model	11
iii.	Snow Flake Data Model	12
19.	Relational Database = Normalized Database	13
20.	Columnar Database	14
21.	Clean Raw Data	15
22.	Transform Data Sets	15
23.	Import Data	15
24.	ETL	15
25.	Data Warehousing	15
26.	SQL	15
27.	Concluding Picture	16

Data Analysis and Business Intelligence Terms:

- 1. Data Analysis
 - Convert Data into Useful Information for Decision Makers
- 2. Business Intelligence
 - Convert Data into Useful/Actionable/Refreshable Information for Decision Makers in a Business Situation
- 3. Goal of Data Analysis and Business Intelligence:
 - Create useful, updateable, actionable information for decision makers
 - Three examples of Data Analysis / Business Intelligence:





- 4. Query
 - Ask a question.
 - When we change criteria or filters in a Slicer and drag and drop a field in a PivotTable, we are changing the question / query.

5. <u>Raw Data</u>

- Data in its smallest form that allows Excel Data Analysis features & Power Tools to work.
- Example:

Raw Data = Data in its smallest form that allows Excel Data Analysis features & Power Tools to work.

Yes:

NI	0	٠
1.1	U	

Addresses
313 173rd Blvd, Kent, WA 98121
316 66th Blvd, Kent, WA 98124
4358 23rd St, Kent, WA 98122
965 151st St, Kent, WA 98116
7900 173rd Lane, Kent, WA 98126
4047 15th Ave, Kent, WA 98122
4907 13th Ave, Kent, WA 98123
3789 4th Blvd, Seattle, WA 98115
2977 66th Lane, Seattle, WA 98117
3392 23rd St, Seattle, WA 98113

Address	City	State	Zip
313 173rd Blvd	Kent	WA	98121
316 66th Blvd	Kent	WA	98124
4358 23rd St	Kent	WA	98122
965 151st St	Kent	WA	98116
7900 173rd Lane	Kent	WA	98126
4047 15th Ave	Kent	WA	98122
4907 13th Ave	Kent	WA	98123
3789 4th Blvd	Seattle	WA	98115
2977 66th Lane	Seattle	WA	98117
3392 23rd St	Seattle	WA	98113

Why?

Because it is easier to analyze data when it is stored in its smallest parts.

6. Proper Data Set

- Data Table with Field Names in first row and Records in subsequent rows.
- In Excel you must have empty cells or Column/Row headers around the Proper Data Set in order for the Excel & Power Tools to work.
- 7. Field Names / Column Names / Header Names
 - Fields define what type of data goes into the column
 - Synonyms for Field:
 - i. Field
 - ii. Column
 - iii. Header
 - Fields can be:

7

- i. Numbers we need to aggregate
- ii. Categories / criteria / filters / labels used as categories for analysis
- iii. Lookup values
- iv. Helper columns
- Example of Proper Data Set in Excel:

Example	of Proper	Data Set ir	n a Text Fi	le:
	••••••••••••••••••••••••••••••••••••••			

8	Example 1:							
9					1		Fields = Column =	
10	Date	Units	SalesRepKey Pro	oductKey	Sales		Headers	
11	10/20/17	44	1	4	\$620.95		in first rour	
12	10/20/17	37	1	3	\$484.31		in first row	$TransactionID \rightarrow Date \rightarrow Website \rightarrow Product \rightarrow Quantity \rightarrow RevenueDiscount$
13	10/20/17	38	2	5	\$376.69			→ NetStandardCost → CountryCode¶
14	10/20/17	82	2	1	\$1,141.96	R	ecords in each row	7742562 → $1/1/2018$ → boomerangs.com → Sunshine → 19 → $0.19 \rightarrow 1.01 \rightarrow JPN\P$
15	10/20/17	56	3	2	\$725.45	R R	ecords in each row	$7742563 \rightarrow 1/1/2018 \rightarrow \text{boomerangs.com} \rightarrow \text{Carlota} \rightarrow 3 \rightarrow 0 \rightarrow 1.01 \rightarrow \text{BRA}$
16	10/20/17	25	3	2	\$222.67			$7742565 \rightarrow 1/1/2018 \rightarrow \text{amazon.com} \text{Manu MTA} \rightarrow 2 \rightarrow 0 \rightarrow 1 \rightarrow \text{FRA}$
17	10/20/17	130	4	3	\$1,038.10			7742566 → 1/1/2018 → boomerangs.com → Yanaki → 2 → 0 → 1.015+NOR¶
18	10/21/17	10	3	1	\$154.41			7742567 → 1/1/2018 → amazon.com→Aspen+48 → 0.43→1.01→CoL¶
19	10/24/17	82	5	5	\$1,236.00			$7/42568 \rightarrow 1/1/2018 \rightarrow \text{amazon.com} \oplus \text{Quad} \rightarrow 3 \rightarrow 0 \rightarrow 0.99 \rightarrow \text{DEU}$ $7/42569 \rightarrow 1/1/2018 \rightarrow \text{cel-boomerang comMetestic Beaut } 24 \rightarrow 0.3475$
20	10/24/17	53	3	4	\$684.48	8		→ 1.015+PSE¶
21	10/24/17	12	5	6	\$127.10			7742570 → 1/1/2018 → gel-boomerang.comFun·Fly → 6 → 0.105+1.01→FRA¶
22	10/24/17	22	4	5	\$269.52			$7742571 \rightarrow 1/1/2018 \rightarrow \text{target.com} \rightarrow \text{Fun Fly} \rightarrow 3 \rightarrow 0 \rightarrow 1.015 + \text{UKR}$
23	10/25/17	82	3	7	\$739.69			$7742573 \rightarrow 1/1/2018 \rightarrow \text{phomerangs.com} \rightarrow \text{Carlota} \rightarrow 72 \rightarrow 0.46541 \rightarrow \text{HUN}$ $7742573 \rightarrow 1/1/2018 \rightarrow \text{phay.com} \rightarrow \text{Ouad} \rightarrow 4 \rightarrow 0 \rightarrow 1.0154\text{CaN}$
24	10/25/17	172	1	7	\$1,201.01	In Exc	cel, empty cells or	$7742574 \rightarrow 1/1/2018 \rightarrow \text{amazon.com} Fun Fly \rightarrow 3 \rightarrow 0 \rightarrow 0.99 \rightarrow \text{VNM}$
25	10/25/17	78	4	6	\$546.06	row/c	olumn headers all	7742575 → 1/1/2018 → amazon.com→Eagle+12 → 0.2 → 1.015+SYR¶
26	10/25/17	12	3	1	\$162.45	ar	ound data set	$7742576 \rightarrow 1/1/2018 \rightarrow \text{boomerangs.com} \rightarrow \text{Quad} \rightarrow 3 \rightarrow 0 \rightarrow 1.01 \rightarrow \text{ITA}$
27			. L.		1. 31			\rightarrow GBR¶
28								7742578 → 1/1/2018 → ebay.com → Fun·Fly → 108 → 0.47→1.015+VEN¶

8. Naming

- Name all tables, fields, queries and other items with smart and consistent names.
- Example of Data Model with smart and consistent table and column names. Notice the f in the name of the Fact Table and d in the name of the Dimension Tables:



9. <u>Data Type</u>

- What type of data is it? Date, Number, Text and so on?
- Traditionally, in Excel we do not have Data Types. But with our Power Tools we will have to define Data Types such as Date, Currency, Whole Number and so on if we want our analysis to work correctly.
- Most of the time we will define Data Types using Power Query.

10. Excel Table feature

- If the source of the Proper Data Set is an Excel Spreadsheet and we want to import it into Power Query or Power Pivot from the current workbook, then we must convert the Proper Data Set to an Excel Table.
- Why Excel Table feature for your Proper Data Set?
 - i. Because the table has dynamic ranges.
 - ii. Any Formula, Chart, PivotTable and other Excel feature will update when you add/delete new records or fields.
 - iii. Convert to Table: Ctrl + T
 - iv. Always name your Table! Keyboard: Alt, J, T, A

11. Grain or Granularity

- Size or level of detail.
 - i. What is the grain or granularity of the sales table?
 - 1. Is the grain of the sales table at the invoice line item product level?
 - 2. Is the grain of the sales table at the Total Invoice Level?
 - ii. Is the granularity of the 1) helper column or 2) iteration of the DAX Formula or 3) of the report at the day, month, year level?
- More or Less Granularity?
 - i. The more granular, the smaller the size, like a grain of sand is small.
 - ii. The less granularity, the more aggregated the number is (the bigger the number is), like moving from adding total sales from daily sales to monthly sales to year sales and so on.
- Granularity will be important when designing a data model and creating formulas in both Power Query and the Data Model.
- Example of Grain or Granularity:

Grain or Granularity = Size or level of detail.							For Criteri	ia in Rep	oort ==>>	Granularity is decre	asing				
Grain = Invoice Line or Product Level The more granular , the smaller the size.					Grain = Invoi The less gra	ce Level anularity, t	he more aggre	gated the nu	umber is (bigger number)	Day Grain		Month Grain		Year Grain	
Date	Invoice #	Units	SalesRepKey	ProductKey	Sales	Date	Invoice #	SalesRepKey	Sales	1	Date	Day	Month	MonthNumber	Year
10/20/17	27002	44	1	4	\$620.95	10/20/17	27002	1	\$1,105.26		10/1/17	1	Oct	10	2017
10/20/17	27002	37	1	3	\$484.31	10/20/17	27003	2	\$1,518.65		10/2/17	2	Oct	10	2017
10/20/17	27003	38	2	5	\$376.69	10/20/17	27004	3	\$948.12		10/3/17	3	Oct	10	2017
10/20/17	27003	82	2	1	\$1,141.96	10/20/17	27005	4	\$1,038.10		10/4/17	4	Oct	10	2017
10/20/17	27004	56	3	2	\$725.45	10/21/17	27006	3	\$154.41		10/5/17	5	Oct	10	2017
10/20/17	27004	25	3	2	\$222.67	10/24/17	27007	5	\$1,236.00		10/6/17	6	Oct	10	2017
10/20/17	27005	130	4	3	\$1,038.10	10/24/17	27008	3	\$684.48		10/7/17	7	Oct	10	2017
10/21/17	27006	10	3	1	\$154.41	10/24/17	27009	5	\$127.10		10/8/17	8	Oct	10	2017
10/24/17	27007	82	5	5	\$1,236.00	10/24/17	27010	4	\$269.52		10/9/17	9	Oct	10	2017
10/24/17	27008	53	3	4	\$684.48	10/25/17	27011	3	\$739.69		10/10/17	10	Oct	10	2017
10/24/17	27009	12	5	6	\$127.10	10/25/17	27012	5	\$1,909.52		10/11/17	11	Oct	10	2017
10/24/17	27010	22	4	5	\$269.52						10/12/17	12	Oct	10	2017
10/25/17	27011	82	3	7	\$739.69						10/13/17	13	Oct	10	2017
10/25/17	27012	172	5	7	\$1,201.01						10/14/17	14	Oct	10	2017
10/25/17	27012	78	5	6	\$546.06						10/15/17	15	Oct	10	2017
10/25/17	27012	12	5	1	\$162.45						10/16/17	16	Oct	10	2017
(A)											10/17/17	17	Oct	10	2017

12. <u>Keys</u>

- Primary Key
 - i. Unique Identifier for each row in table so there are no duplicates (like no duplicate Product Names or Sales Reps).
 - ii. Primary Keys are the first column in Dimension / Lookup Tables and are used on the one-side of a one-to-many-relationship.
 - iii. If you are coming from the Excel side, you already know what a Primary Key is because all lookup tables are required to have a Unique List in the first column, which is, of course, a Primary Key.
- Foreign Key
 - i. A Column that is allowed duplicates and is used on the many-side of a one-to-many-relationship.
- Primary & Foreign Keys are used together with:
 - i. VLOOKUP in Excel
 - ii. Relationships in the Data Model
 - iii. Merges in Power Query

Foreign Key	Fact	Foreign Key	Foreign Key	Fact
			×	
Date	Units	SalesRepKey	ProductKey	Sales
10/20/17	44	1	4	\$620.95
10/20/17	37	1	3	\$484.31
10/20/17	38	2	5	\$376.69
10/20/17	82	2	1	\$1,141.96
10/20/17	56	3	2	\$725.45
10/20/17	25	3	2	\$222.67
10/20/17	130	4	3	\$1,038.10
10/21/17	10	3	1	\$154.41
10/24/17	82	5	5	\$1,236.00
10/24/17	53	3	4	\$684.48
10/24/17	12	5	6	\$127.10
10/24/17	22	4	5	\$269.52
10/25/17	82	3	7	\$739.69
10/25/17	172	1	7	\$1,201.01
10/25/17	78	4	6	\$546.06
10/25/17	12	3	1	\$162.45

Primary Key	Attributes For Filtering	Attributes For Filtering
SalesRepKey	Sales Rep	Region
1	Gigi	USA
2	Freddy	Canada
3	Chin	Mexico
4	June	USA
5	Tyrone	Europe

Primary Key Attributes For Filtering Values used in Calculations Values used in Calculations

ProductKey	Product	Price	SCost
1	Fun Fly	9.99	6.49
2	Sunshine	19.99	8.8
3	Fast Catch	23.45	14.54
4	Yanaki	24.95	15.22
5	Carlota	25.99	17.15
6	Crested Beaut	32.99	14.19
7	Quad	43.69	22.72

13. Fact Table / Transaction Table

- Table that has:
 - i. Numbers we need to summarize (like Sales or Units)
 - ii. Foreign Key columns that we will use in relationships (like Date, SalesRepKey, ProductKey)
 - iii. After we complete our Data Model, the DAX Measures (Formulas for reports) will usually be listed in the Fact Table.
- "Fact" = measurements of business activities (like amount of sales, or how many units sold, or clicks on web links)
 - i. Now we know where the DAX word "Measure" came from!
- Fact Tables are usually large. Dimension Tables are comparatively small.
- Grain of Fact Table
 - i. The grain of the Fact Table defines what dimensions you can use to filter the calculation.
 - 1. If you have a Fact Table with Invoice totals, but no product totals, you can't filter a report by product.
 - ii. In general, you want the Fact Table grain to be the smallest size that has not been aggregated and will allow the most dimensionality
 - 1. More dimension tables = more filtering and varied types of queries.
 - iii. The grain of the Fact Table usually comes from the Measured Business Event, like the measurements of a line item sold: sales amount, price, quantity.
 - iv. Data at its most granular or atomic level (not aggregated) is the foundation of every Fact Table.
 - 1. Example: If you sell pallets of plywood (one product) to construction companies and the wood comes from more than one supplier, you might have to break the Product Level Sales Amount Fact Table up into a smaller grain to show sales amounts by supplier.

14. Dimension Table / Lookup Table / Entity Tables

- Table that has:
 - i. First column is a Primary Key (Unique Identifier) for the Entity (Product, Sales Rep and so on) used in relationship with Fact Table.
 - ii. Remaining columns are attributes that we can use as:
 - 1. Criteria / Filters / Categories / Report Labels for our Reports & Dashboards.
 - 2. Values we can lookup (like Price).
 - 3. Helper Columns (like Sort Helper Columns or Intermediate calculations).
- Fact Tables are usually large. Dimension Tables are comparatively small.
 - i. An Insurance company might have a large Customer Dimension Table.
- Date Table
 - i. Type of Dimension table required by MS Power Tools for DAX Time Intelligence Functions.
 - ii. We use Data Tables rather than the Group By Date feature in a PivotTable.
- Synonyms for Filtering columns:
 - i. Attribute, Criteria, Filters, Category, Categorical attributes, Constraints, Groupings, Report labels for the reports and analytics.
- When you choose a column as a filter for a report, you are choosing the grain of the calculation. For example: Day, Month or Year columns.

• Example of Fact and Dimension Tables:

Fact Table =

Table that has numbers we need to summarize (like Sales and Units) and

Foreign Keys that we will use in relationships (like Date, SalesRepKey, ProductKey) "Fact" = measurements of business activities (like amount of sales or how many units)

Foreign Key Fact Foreign Key Foreign Key Fact

Date	Units	SalesRepKey	ProductKey	Sales
10/20/17	44	1	4	\$620.95
10/20/17	37	1	3	\$484.31
10/20/17	38	2	5	\$376.69
10/20/17	82	2	1	\$1,141.96
10/20/17	56	3	2	\$725.45
10/20/17	25	3	2	\$222.67
10/20/17	130	4	3	\$1,038.10
10/21/17	10	3	1	\$154.41
10/24/17	82	5	5	\$1,236.00
10/24/17	53	3	4	\$684.48
10/24/17	12	5	6	\$127.10
10/24/17	22	4	5	\$269.52
10/25/17	82	3	7	\$739.69
10/25/17	172	1	7	\$1,201.01
10/25/17	78	4	6	\$546.06
10/25/17	12	3	1	\$162.45

Dimension Table / Lookup Table = Tables with:

1) First column is a Primary Key (Unique Identifier) used in relationship with Fact Table

- 2) Remaining columns are attributes that we can use as:
 - Criteria or Filters for our Reports & Dashboards Values we can lookup Helper Columns.

Primary Key Attributes For Filtering Attributes For Filtering

SalesRepKey	Sales Rep	Region	
1	Gigi	USA	
2	Freddy	Canada	
3	Chin	Mexico	
4	June	USA	
5	Tyrone	Europe	

Primary Key Attributes For Filtering Values used in Calculations

ations Values used in Calculations

ProductKey	Product	Price	SCost
1	Fun Fly	9.99	6.49
2	Sunshine	19.99	8.8
3	Fast Catch	23.45	14.54
4	Yanaki	24.95	15.22
5	Carlota	25.99	17.15
6	Crested Beaut	32.99	14.19
7	Quad	43.69	22.72

15. Is a number a fact or an attribute?

- If Price or Cost does not change much and is discrete, it seems like a dimension
- If Price or Cost changes often or for each transaction, then it seems more like a fact.

16. <u>History</u>

• Terms "Fact" and "Dimension" go back to 1960s (General Mills and Dartmouth University), 1970s (AC Nielsen & IRI), but was popularized by Ralph Kimball in the 1980s.

17. Relationships

- Connects tables so we can build solutions based on fields from multiple tables.
- Relationships between tables allows us to:
 - i. Use Criteria from Dimension tables to filter Fact Table Calculations (Measures).
 - 1. The Arrow Direction on the Relationship Arrow indicates the direction of the filter (From Dimension to Fact Table)
 - 2. What makes the filtering of a Fact Table so efficient, is that a Measure (Formula) will then have to make its calculation over fewer rows. If you are an Excel person, compare this to Array Formulas that must calculate over all rows in a column even if there are rows that do not match the criteria. Array Formulas are notoriously slow to calculate. DAX Measures that can work over a filtered Fact Table are usually fast to complete the calculations.
 - ii. Lookup values
 - iii. For Excel people, it helps avoid VLOOKUP driven solutions that may have many spreadsheet formulas that slows down query time (change criteria on report)
- Types of Relationships:
 - i. One-To-Many
 - 1. Example: There is one unique product in each row of a Product Dimension / Lookup Table, but there can be many products in the Sales Fact Table. The Product Dimension Table is on the One-Side and the Sales Fact Table is on the Many-Side.
 - 2. Excel users are used to a One-To-Many Relationships from using the VLOOKUP Function. When you lookup a product price in the sales table, the sales side is the Many-Side and the product lookup table is the One-Side.
 - ii. Many-To-Many
 - 1. Example: you may have an Order Table and an Invoice Table, where orders can be partially invoiced resulting in one order with multiple invoices; and multiple orders may be filled in one invoice.
 - iii. One-To-One
 - 1. Example: If you have a Sales Rep Table and a Auto Table and each Sale Rep is issued only one car.

18. Data Models

- Structure of Data Model must come from the type of Measures & Reports that the Business Decision Maker requires.
- Requirements for good Data Model:
 - i. Contains necessary data for end solution
 - ii. Easy for the decision maker to use
 - iii. Allows fast queries (change criteria and solution updates quickly)
 - iv. Is easy to update when new data is available
 - v. Easy to update structurally, if needed
- Data Models can be stored in Power Pivot & Power BI Desktop
- Data Model can Contain:
 - i. Fact Table / Tables
 - ii. Dimension Tables

- iii. Relationships
- iv. Helper Columns / Tables
- v. DAX Calculated Columns and DAX Measures
- vi. Hidden / Not Hidden Tables and Columns
- vii. Formatting
- Data Model is also known as: "Semantic Model", which is defined as a model that represents the real word relationships between Facts and Entities and contains the data, relationships, calculated columns and other complexities together as a single model and allows us to create DAX Queries / Measures (formulas, calculations) that can be used in multiple reports or end solutions. This is in comparison to storing data in a relational or object-oriented database that require that you redefine the relationships and other complexities each time you create a Query or Measure.
- Some types of models:
 - i. Flat Table = one table with all columns necessary to make report. Facts and Dimensions were mixed together in one table.
 - 1. Before Power Pivot, we had to create one Flat Table when we wanted to make PivotTable Reports.
 - 2. If we created the flat table with formulas, it could slow down the calculation speed of the workbook.
 - ii. Star Schema Data Model = Dimensional Model
 - 1. Fact Table with most atomic granular data surrounded by Dimension Tables using One-To-Many Relationships
 - 2. Star Schema Models tend to:
 - Be easier to use than other models.
 - Allow faster queries than other types of models.
 - Any Criteria or Filter from Dimension Tables can accurately filter the Measures (calculations) in the final reports.
 - Create One Formula that can be used in many reports
 - 3. Power Pivot & Power BI Desktop are designed to work efficiently with the Star Schema Data Model.
 - 4. Example of simple Star Schema Data Model and PivotTable Report:



Total Sales	Manufacturer 💌			
Sales Rep 💌	Colorado	Darnell	Gel Booms	Grand Total
Chin	\$684.48	\$316.86	\$1,687.81	\$2,689.15
Freddy		\$1,141.96	\$376.69	\$1,518.65
Gigi	\$620.95	\$484.31	\$1,201.01	\$2,306.27
June	\$546.06	\$1,038.10	\$269.52	\$1,853.68
Tyrone	\$127.10		\$1,236.00	\$1,363.10
Grand Total	\$1,978.59	\$2,981.23	\$4,771.03	\$9,730.85

iii. <u>Snow Flake Data Model</u> =

- 1. Follows Database Normalization rules where columns in tables should not have duplicate values and must be broken apart into separate tables.
- 2. Example: Rather than have a Manufacturer Column in a Product Table Snow Flake Schema would have a Product Table and Manufacturer Table with a relationship between the two tables.
- 3. In this class if we encounter a Snow Flake Schema, we will use Power Query to transform it to Star Schema.
- 4. Why we will avoid Snow Flake:
 - i. Columnar Database will store Dimension Tables with duplicates in an efficient way. No need to complicate model with many tables.
 - ii. According to tests done by Marco Russo and Alberto Ferrari, for larger models the extra relationships used in a Snow Flake Schema can be less efficient than a single Dimension table with Duplicate Values (for exact models you must test, however).
 - iii. Hierarchies cannot be made from Snow Flake Model.
- 5. Example of Snow Flake Model and PivotTable Report:



iv. Other Models

1. There may be cases where we have to build a Data Model that is not Star Schema because of business requirements.

19. <u>Relational Database = Normalized Database</u>

- Database that stores company's raw data with the goals of 1) no-redundancy and 2) accuracy.
- Relational databases can have many tables and many relationships between tables.
- In general, values are entered into one and only one location. There are usually not columns with duplicate values.
- Relational databases serve the goal of storing raw data with accuracy, but they can be complicated to navigate, and queries may not perform as fast as is necessary for end users creating reports.
- In this class when we connect to a Relational Database, we will use Power Query to extract the necessary data and then transform it and load it into the Star Schema Data Model.

20. Columnar Database

- Synonyms for Columnar Database:
 - i. XVelocity analytics engine
 - ii. VertiPaq
- In Power Pivot and Power BI Desktop, the data you import is stored in an In-RAM-Memory Columnar Database.
- When you import a table into the Data Model, each field in the imported table is stored separately with a unique list of values for that field. There is a sort of "map" that allows the database, Data Model and DAX Formulas to reconstruct the original table with the required records for any given calculation.
- The Columnar Database is a behind the scenes In-Memory (RAM) Database.
 - i. RAM = Random Access Memory.
 - ii. The number of unique values in any one field (and other factors) determine the amount of RAM that is used.
- Benefits of Columnar Database:
 - i. The Columnar Database allows you to import large data sets (millions of rows) that would not fit in an Excel sheet. You can safely handle 100 million rows or more.
 - ii. For an Excel user storing data in Excel, The Columnar Database stores data efficiently and can dramatically reduce file size.

Sales Rep

Jo Nina

Kip Gigi

- iii. The Columnar Database is designed to work with DAX Formulas to calculate quickly on Big Data.
- Example of Columnar Database, where each field is stored in a separate column with a unique list of values only:

Power Pivot & Power BI Desktop Columnar Database:

Proper Data Set

Import into Data Model ==>>

Columnar Database:

Sales	Sales Rep	Region	
\$54.00	Jo	West	
\$26.00	Nina	East	
\$54.00	Jo	South	
\$57.00	Кір	West	
\$22.00	Gigi	West	
\$59.00	Gigi	South	
\$95.00	Кір	East	
\$99.00	Кір	South	
\$51.00	Nina	South	
\$49.00	Nina	East	
\$12.00	Jo	East	
\$30.00	Jo	East	
\$20.00	Nina	West	
\$92.00	Кір	West	
\$73.00	Gigi	South	

ales	
\$54.00)
\$26.00)
\$57.00)
\$22.00)
\$59.00)
\$95.00)
\$99.00	D
\$51.00	C
\$49.00	D
\$12.00)
\$30.00	D
\$20.00	D
\$92.00)
\$73.00)

Why Columnar Database?

Region

West

East South

1) Reduces File Size

2) Allows us to import 100 million of rows of data or more

3) Works with DAX Formulas to calculate quickly on Big Data.

21. Clean Raw Data

- Fix unusable raw data so that it can be used to perform data analysis
- Examples:
 - i. Split data apart into desired data, like "Gel Boomerangs CA" ==>> "Gel Boomerangs", "CA"
 - ii. Join data together to get desired data, like 2018 and Quarter 1 ==>> "2018 Quarter 01"
 - iii. Other cleaning goals.

22. Transform Data Sets

- Fix unusable data set so that it can be used to perform data analysis
- Examples:
 - i. Add, remove or filter columns in data sets
 - ii. Combine, merge, append or unpivot data sets
 - iii. Taking Relational Database with many tables and converting it into a Star Schema using Power Query
 - iv. Other data transformational goals

23. Import Data

- Import data from external sources (single or multiple sources) into Excel or Power Pivot's Data Model or Power BI Desktop;
- Optimally, the import will allow refreshes so Reports and Dashboards update when source data changes.

24. <u>ETL</u>

• Extract, Transform and Load Data (term from Data Warehousing)

25. Data Warehousing

- Term that means get data from original source and form it so that you can make useful reports and dashboards.
- Term made famous by Ralph Kimball.
- We will be doing a version of Data Warehousing when we build Star Schema Data Models with MS Power Tools.

26. <u>SQL</u>

- SQL = Structured Query Language
- Used to build, maintain and query databases
- We will not write SQL in this class, but much of what we to is designed to replace SQL with an easier tool
- Some of the terms we use like "Grouping" come from SQL



Define: Data Analysis / Business Intelligence: