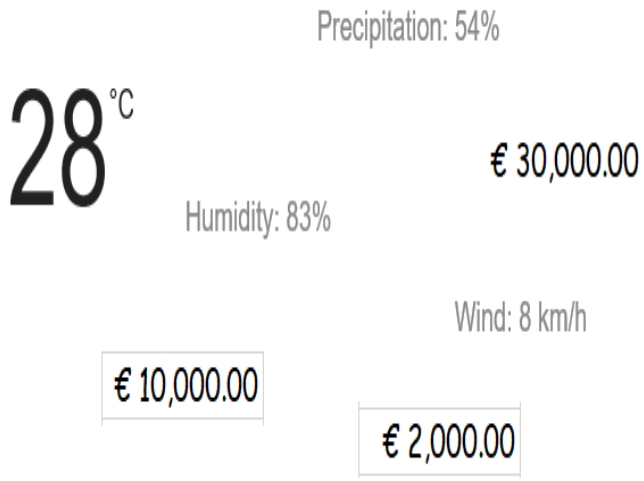
ADVANCED DATA VISUALIZATION USING TABLEAU

ADVANCED DATA VISUALIZATION USING TABLEAU
COURSE OUTLINE

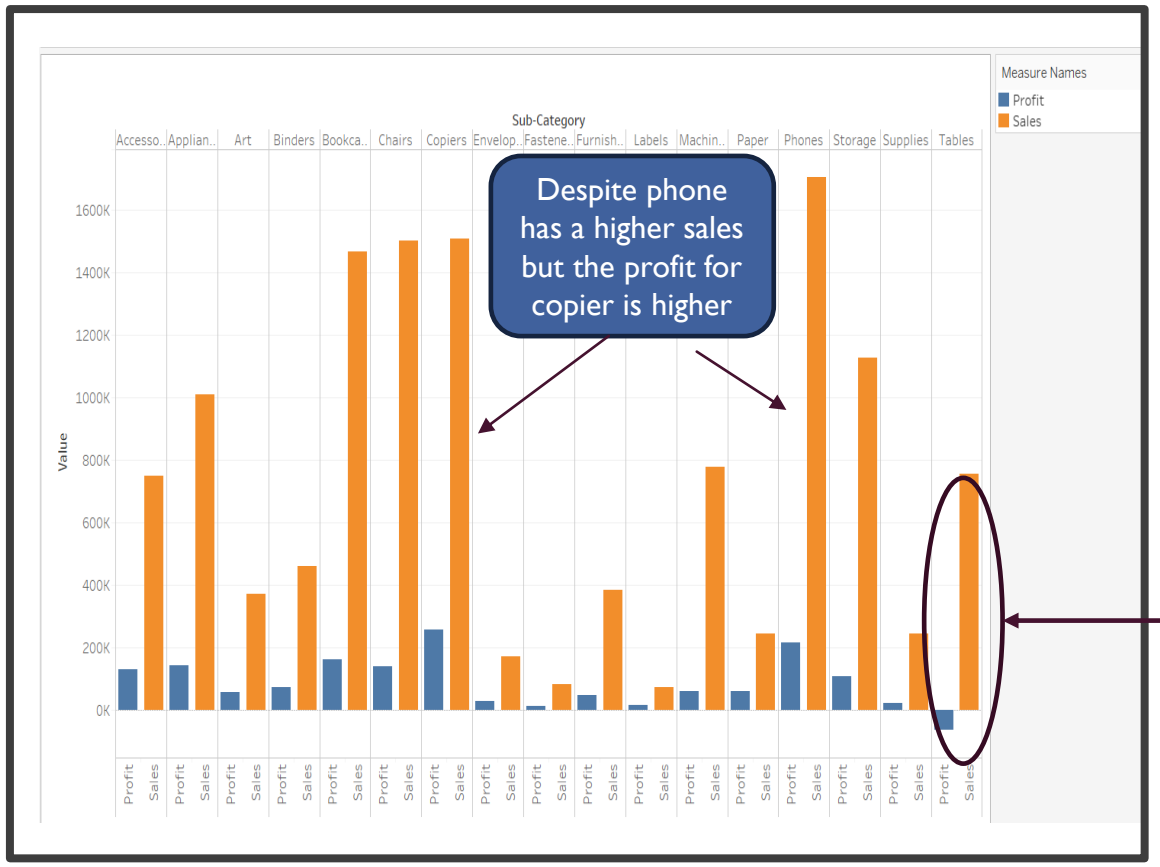
- In this course, we will cover the following areas:
  - Data Management
  - Advance Data Computation
  - Data Analysis
  - Statistics and Forecasting

ADVANCED DATA VISUALIZATION USING TABLEAU
WHAT IS DATA?

- Data has grown exponentially in the last 2 decades with the introduction of Smart Devices, Sensors and Transaction Systems (e.g. Sales, Accounting, Inventory)
- Data from these systems are raw and does not provide meaningful insights on it own.
- Data need to be manipulated and integrated to create insights and analysis



ANALYSING THE DATA



- Using different data variables, we are able to generate the various insights
- For example, the performance of Sub-Category sales vs profit
- From the chart, we notice despite the high sales for table, it is making a loss

CHALLENGES FACED BY MOST DATA SCIENTISTS

- Accuracy of Data
  - Data comes from different data sources. The format of the data may be different
  - There are missing data and incorrect data found in the data sources
- Understanding requirements
  - What are the expectations from the analysis?
  - Do we have the required data?
- Aligning requirements with the Data Visualisation Tool
  - Is it possible to align to requirements to the tools?
  - Does the tools provide the required visualization tool?

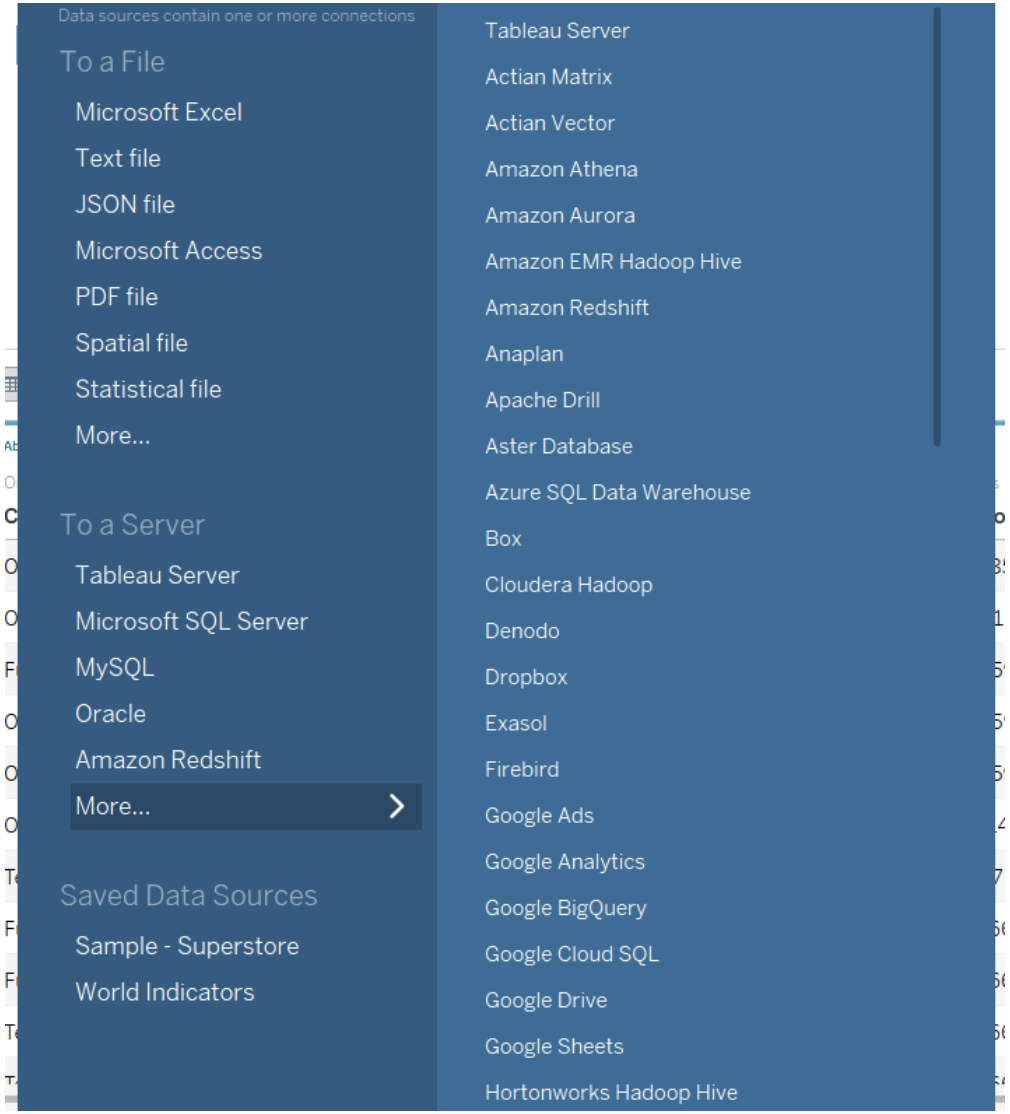
STARTING THE JOURNEY

- It is important to know where are the data sources
- You need to know the relationship between the different data sources (E.g. Is table A related table B by a particular field)
- Type of connectivity - Do you need to integrate the different data sources?
- More importantly, do you need to ‘clean’ the data? The data you have gathered, may consists of errors (e.g. characters instead of numbers), different format (e.g. Date Field – ’15/05/2019’ or 05/15/2019’ or missing values
- Data Preparation normally take up at least 70% of the project. Tools such as Tableau Prep are used to clean the data before visualisation and analysis

DATA MANAGEMENT

ACCESSING THE DATA

- In Tableau, data can be acquired from multiple data sources

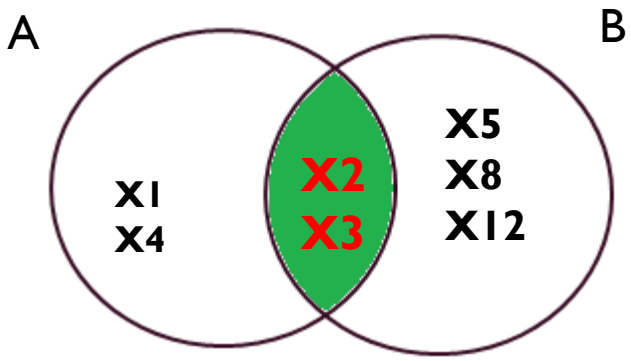


CONCEPTS OF JOIN

- As data comes from different sources, you may be required to join the different data sources for analysis
- Typically, there are the following type of join:
  - Inner Join
  - Left/Right Join
  - Full Join

INNER JOIN

- Inner join is used when you require the data that are found in both sources.

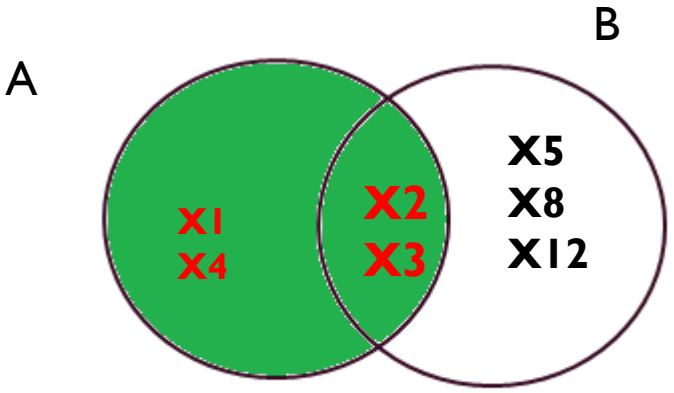


e.g
A = {**X1, X2, X3, X4**}
B = {**X2, X3, X5, X8,X12**}

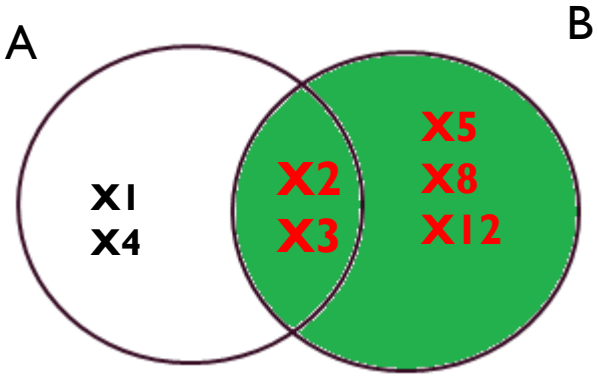
Output from Inner Join:
{X2, X3}

LEFT/RIGHT OUTER JOIN

- Left or Right Join - all the data from one of the sources is collected regardless whether it exist in the other source



e.g
A = {**X1, X2, X3, X4**}
B = {**X2, X3, X5, X8,X12**}

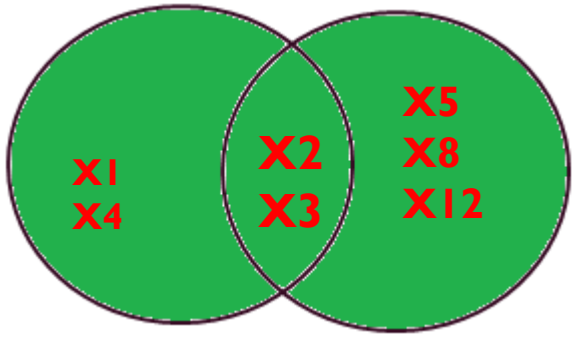


Output from a Left Join:
{X1, X2, X3, X4}

Output from a Right Join:
{X2,X3,X5,X8,X12}

FULL OUTER JOIN

- Data from BOTH sources are retrieved



e.g
A = {**X1, X2, X3, X4**}
B = {**X2, X3, X5, X8,X12**}

Output from Full Outer Join:
{X1, X2, X3, X4, X5, X8,X12}

DATA JOIN IN REAL LIFE

Student

StudentID	Name	Gender	DOB	Race
190000001	Loh Ah Seng	M	15 Jul 2000	Chinese
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese
190000003	Ravi Chandran	M	05 Apr 2000	Indian
190000004	Mohammad Firdaus	M	11 Feb 2000	Malay

StudentID	Module	Mid Term	Final Exam
190000002	Chemistry	40	80
190000002	Mathematics	70	90
190000003	Chemistry	90	95
190000004	Chemistry	80	60
190000005	Physics	88	88

Student Grade

EXAMPLE OF INNER JOIN

StudentID	Name	Gender	DOB	Race
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese
190000003	Ravi Chandran	M	05 Apr 2000	Indian
190000004	Mohammad Firdaus	M	11 Feb 2000	Malay

StudentID	Module	Mid Term	Final Exam
190000002	Chemistry	40	80
190000002	Mathematics	70	90
190000003	Chemistry	90	95
190000004	Chemistry	80	60

Student Inner Join Student Grade by using Student ID

EXAMPLE OF LEFT OUTER JOIN

StudentID 190000001 does not have a record in StudentGrade

StudentID	Name	Gender	DOB	Race	*StudentID*	*Module*	*MidTerm*	*Final Exam*
190000001	Loh Ah Seng	M	15 Jul 2000	Chinese	***Null***	***Null***	***Null***	***null***
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese	190000002	Chemistry	40	80
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese	190000002	Mathematics	70	90
190000003	Ravi Chandran	M	05 Apr 2000	Indian	190000003	Chemistry	90	95
190000004	Mohammad Firdaus	M	11 Feb 2000	Malay	190000004	Chemistry	80	60

Student LEFT Join Student Grade by using Student ID

StudentID 190000002 is repeated based on the number of occurrance in the StudentGrade

EXAMPLE OF RIGHT OUTER JOIN

StudentID	Name	Gender	DOB	Race	StudentID	Module	Mid Term	Final Exam
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese	190000002	Chemistry	40	80
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese	190000002	Mathematics	70	90
190000003	Ravi Chandran	M	05 Apr 2000	Indian	190000003	Chemistry	90	95
190000004	Mohammad Firdaus	M	11 Feb 2000	Malay	190000004	Chemistry	80	60
Null	*Null*	*Null*	*Null*	*Null*	190000005	Physics	88	88

Student RIGHT Join Student Grade by using Student ID

StudentID 190000005 is not found in the student

StudentID 190000002 is repeated based on the number of occurrance in the StudentGrade

REPUBLIC POLYTECHNIC
DISCOVER. TRANSFORM. ACHIEVE

EXAMPLE OF FULL OUTER JOIN

StudentID	Name	Gender	DOB	Race	StudentID	Module	Mid Term	Final Exam
190000001	Loh Ah Seng	M	15 Jul 2000	Chinese	*Null*	*Null*	*Null*	*Null*
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese	190000002	Chemistry	40	80
190000002	Wong Siew Heng	F	18 Sep 2000	Chinese	190000002	Mathematics	70	90
190000003	Ravi Chandran	M	05 Apr 2000	Indian	190000003	Chemistry	90	95
190000004	Mohammad Firdaus	M	11 Feb 2000	Malay	190000004	Chemistry	80	60
Null	*Null*	*Null*	*Null*	*Null*	190000005	Physics	88	88

Student FULL Join Student Grade by using Student ID

DATA BLENDING

- The concepts of Data Blending is quite like LEFT-OUTER join
- However, the way the data are integrated are different

Difference between Data Blending and Data Join

Data Blending	Data Join
Data are from different data sources	Data are normally from the same data source
Data may have different data granularity (Different level of data details)	Data must be from the same granuality

WHEN TO USE DATA BLENDING

- Data source not supported by cross database join (e.g. Cubes)
- Data are at different level of details
- Join causes duplicated data
- Data need cleaning
- Large amount of Data

DATA BLENDING
DIFFERENT LEVEL OF DETAILS

State	City	Tourist Number
Hawaii	Honolulu	1,000,000
Hawaii	Pearl City	500,000
Arizona	Globe	10,000
Arizona	Eloy	10,000
Michigan	Singapore	300

State Tourism Number

State	Expected Tourist No.
Hawaii	1,500,000
Arizona	1,000,00
Michigan	800,000

City Expected Tourism Number

Tourism in United States

- The two data sources have different level of detail
- In the City Expected Tourism Number, the highest level is State
- In the State Tourism Number data source, it has included another level called city

REPUBLIC POLYTECHNIC
DISCOVER. TRANSFORM. ACHIEVE

DATA BLENDING
- CASE OF JOIN WITH DIFFERENT LEVEL OF DETAILS

State	City	Tourist Number
Hawaii	Honolulu	1,000,000
Hawaii	Pearl City	500,000
Arizona	Globe	10,000
Arizona	Eloy	10,000
Michigan	Singapore	300

State Tourism Number

State	Expected Tourist No.
Hawaii	1,500,000
Hawaii	1,500,000
Arizona	1,000,00
Arizona	1,000,00
Michigan	800,000

City Expected Tourism Number

- Although using a join is possible, the 'Expected Tourist No.' will be duplicated in all rows. This will give an inaccurate number of expected tourist. In this case, Data Blending is a better option

DATA BLENDING
-CASE OF USING DATA BLENDING

State	City	Tourist Number
Hawaii	Honolulu	1,000,000
Hawaii	Pearl City	500,000
Arizona	Globe	10,000
Arizona	Eloy	10,000
Michigan	Singapore	300

State Tourism Number

State	Expected Tourist No.
Hawaii	1,500,000
Arizona	1,000,00
Michigan	800,000

City Expected Tourism Number

- Basically, the integrity of both data sources are maintained.
- It is difficult to illustrate how blending work without the exercise

State	City	Tourist Number	State	Expected Tourist No.
Hawaii	Honolulu	1,000,000	Hawaii	1,500,000
Hawaii	Pearl City	500,000		
Arizona	Globe	10,000	Arizona	1,000,00
Arizona	Eloy	10,000		
Michigan	Singapore	300	Michigan	800,000

REPUBLIC POLYTECHNIC
DISCOVER. TRANSFORM. ACHIEVE

THERE ARE SOMETHING ABOUT UNION

If you have different data sources with the same file structure and you need to combine from these sources into a single table?

- Join has the 'Union' features that 'merge' different data source together.
- However, the columns are added to the row

Category	Region	Sales	Business Unit(Q2)	Region(Q2)	Sales(Q2)
Consumer	East	2500.00	Consumer	East	3359.00
Consumer	South	3548.00	Consumer	East	
Enterprise	East	5548.00	Enterprise	South	

This may not be suitable in some cases of analysis

Sales (Q1)

Category	Region	Sales
Consumer	East	2500.00
Consumer	South	3548.00
Enterprise	East	5548.00

Sales (Q2)

Business Unit	Region	Sales
Consumer	East	3359.00
Consumer	East	4578.00
Enterprise	South	4444.00

Category	Region	Sales
Consumer	East	2500.00
Consumer	South	3548.00
Enterprise	East	5548.00
Consumer	East	3359.00
Consumer	East	4578.00
Enterprise	South	4444.00

- There is a features that allow you to append records into the same data sources
- Ideally, the structure of the data sources must be the same

PERFORM EXERCISE 1 TO 3 (20 MINUTES)

DATA SETS AND GROUP

- The are two features that can manage data:
  - Group
  - Set
    - Static
    - Dynamic
- The purpose of these two features is to allow the analyst to view from different perspective and insights
- Both are able to boost the efficiency of using Tableau if it is used correctly

DIFFERENCE BETWEEN GROUP AND SETS

- Although both GROUP and SET has some similarity, they are totally different
- About GROUP
  - Grouping does not change the data but changes how the data is perceived
  - Creating GROUP is changing the metadata, when data category are grouped together (e.g. Construction and Environment Services), tableau will consolidate the different data category into a single Group category (e.g. Construction and Environment Services -> Building and Environment)
  - So, each of data category belong to a particular group, it will be recognized as that grouped category -> data that belong to ‘Construction’ OR ‘Environment Service’ category will be recognized as ‘Building and Environment’ category
- About SETS
  - It is creating a subset of the data. Each data category will remain in their respective category.
  - It can be used to emphasize specific data subset

SIMPLE ANALOGY OF GROUP AND SETS

Group
When two categories are group together, they are classified as a single category

Set
When two categories are put in a set, they remain as the same two categories.

SCENARIO ONE: INVESTMENT IN STARTUPS

You are to review a list of start-ups to determine which is the best start-ups to invest.

Key criterion for selecting investments are based on a combination of:

- High Revenue (2015) and Low Revenue (2015)
- Top Growth (2015)

You are required to determine which business has the best investment opportunities.

*BEFORE YOU **TAKE A JAB AT SCENARIO***

- Ask yourself how you will handle the scenario
- Open the data source '1000-startup.xlsx'
- It is important to ask yourself some questions. For Example:
  - What are the data you need?
  - What does it mean by Top Growth? Does it refer to top growth across the industry?
  - How many companies are you looking for Top Growth? Etc.
- If you are unsure or not clear, you may need to consult the by business partner that has asked you for the analysis.
- For this challenge, we assume that the 'Top Growth' is across the industry

DATA EXTRACT

- Data Extract is a user-defined data subset from a data source
- Data Extract consists of
  - All records
  - Records based on filter or limits
- Data Extract can also be used to aggregate data source
- Refreshing of data can either be:
  - All refresh (All record are updated)
  - Incremental refresh (Only new rows included from the last refresh)

WHEN TO USE DATA EXTRACT

- Use Data Extract when:
  - Managing and Limit Data Access especially for security reasons
  - Making the workbook portable
  - Improve performance
- Avoid Data Extract when:
  - You wish to have control where data reside (May cause security issues)

PERFORM EXERCISE 4
(60 MINUTES)

USING CALCULATION

TABLEAU CALCULATION

- The following are example of calculations processed in Tableau:
  - Calculated Field
  - Table Calculation
- What are the difference between the two?

WHAT ARE THE DIFFERENCES BETWEEN CALCULATED FIELDS AND TABLE CALCULATION?

- They are similar as they uses formula for computing, but the difference is where the data is processed and where it will reside

Table Calculation	Calculated Fields
Computations are done locally	Computations are done at the data source (e.g. Excel, MySQL)
Created as view and stay locally in the worksheet	Created at data level and stored as a new column in the data source
It cannot be sent back to the data source to be re-used (unless you convert it to Calculated Fields) for different worksheet	Can be re-used for different analysis for different worksheet
It is simpler and has limited scope	Has more diverse scope and can be used for deeper analysis

USAGE OF CALCULATED FIELDS

- Calculated fields are used in the following situations:
  - Mathematical operations
  - Logic statement (e.g. IF/ELSE, CASE)
  - Aggregating Data (e.g. Average)
  - Manipulating Strings
  - Date Formula (e.g. Computing Age)

CALCULATION AND AGGREGATION

- Aggregation affect the order of the operations -> It will affect the return results could be different and potentially incorrect.
- There are situations where you to perform the overall aggregation before computing the ratio
- There are situations where aggregation should not be applied when computing line-item calculations

AGGREGATING DIMENSIONS IN CALCULATION

- Aggregating a dimension will create a new temporary new measure column
- An aggregated dimension is needed when
  - Blending multiple data sources which the dimensions do not have a consistent level of detail
  - Requesting that a dimension member to be returned when aggregating elsewhere in the dimension

Functions for Aggregating Dimensions
Minimum
Maximum
Count
Count (Distinct)
Attr

LOST?? PERFORM EXERCISE 5 TO UNDERSTAND (20 MINUTES)

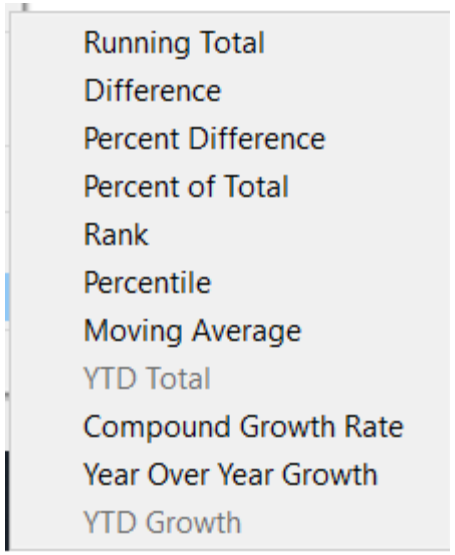
TABLE CALCULATION

- Tables Calculation are computation made locally
  - Applies to values in the internal table
  - Depends on the table structure
- Level of Control
  - Quick Table Calculation
  - Add Table Calculation
  - Edit Table Calculation
  - Custom Table Calculation
- Scope and Directions – Define the direction is performed

TABLE CALCULATION

- Tableau has the following list of table calculation:

For this course, we will not explore the functionality of the different table calculation



SCOPE AND DIRECTIONS

- Table calculation allows to control the order of calculation based on specific direction and area

		Order Date				
Category	Sub-Catego..	2012	2013	2014	2015	Grand To..
Furniture	Bookcases	11.48%	11.88%	11.04%	11.94%	11.60%
	Chairs	12.65%	11.02%	12.55%	11.47%	11.88%
	Furnishings	2.83%	3.06%	3.28%	2.98%	3.05%
	Tables	6.51%	6.13%	5.94%	5.66%	5.99%
Office Supplies	Appliances	7.67%	8.33%	7.49%	8.37%	8.00%
	Art	2.84%	3.08%	2.88%	2.97%	2.94%
	Binders	3.85%	3.49%	3.56%	3.73%	3.65%
	Envelopes	1.24%	1.42%	1.49%	1.26%	1.35%
	Fasteners	0.60%	0.73%	0.63%	0.66%	0.66%
	Labels	0.60%	0.58%	0.54%	0.60%	0.58%
	Paper	1.89%	1.92%	2.07%	1.85%	1.93%
	Storage	9.10%	8.54%	9.09%	8.92%	8.92%
	Supplies	2.11%	1.62%	1.94%	2.01%	1.92%
Technology	Accessories	5.02%	6.44%	6.16%	5.90%	5.93%
	Copiers	9.58%	12.22%	12.20%	12.80%	11.94%
	Machines	7.11%	5.97%	5.82%	6.05%	6.16%
	Phones	14.93%	13.60%	13.32%	12.84%	13.50%
Grand Total		100.00%	100.00%	100.00%	100.00%	100.00%

Table Across

Table Down

Category	Sub-Catego..	2012	2013	2014	2015	Grand To..
Furniture	Bookcases	11.48%	11.88%	11.04%	11.94%	11.60%
	Chairs	12.65%	11.02%	12.55%	11.47%	11.88%
	Furnishings	2.83%	3.06%	3.28%	2.98%	3.05%
	Tables	6.51%	6.13%	5.94%	5.66%	5.99%
Office Supplies	Appliances	7.67%	8.33%	7.49%	8.37%	8.00%
	Art	2.84%	3.08%	2.88%	2.97%	2.94%
	Binders	3.85%	3.49%	3.56%	3.73%	3.65%
	Envelopes	1.24%	1.42%	1.49%	1.26%	1.35%
	Fasteners	0.60%	0.73%	0.63%	0.66%	0.66%
	Labels	0.60%	0.58%	0.54%	0.60%	0.58%
	Paper	1.89%	1.92%	2.07%	1.85%	1.93%
	Storage	9.10%	8.54%	9.09%	8.92%	8.92%
	Supplies	2.11%	1.62%	1.94%	2.01%	1.92%
Technology	Accessories	5.02%	6.44%	6.16%	5.90%	5.93%
	Copiers	9.58%	12.22%	12.20%	12.80%	11.94%
	Machines	7.11%	5.97%	5.82%	6.05%	6.16%
	Phones	14.93%	13.60%	13.32%	12.84%	13.50%
Grand Total		100.00%	100.00%	100.00%	100.00%	100.00%

Pane (Down)

Pane

OTHER SCOPE AND DIRECTIONS

Category	Sub-Catego..	Order Date 2012	2013	2014	2015
Furniture	Bookcases	11.48%	11.88%	11.04%	11.94%
	Chairs	12.65%	11.02%	12.55%	11.47%
	Furnishings	2.83%	3.06%	3.28%	2.98%
	Tables	6.51%	6.13%	5.94%	5.66%
Office Supplies	Appliances	7.67%	8.33%	7.49%	8.37%
	Art	2.84%	3.08%	2.88%	2.97%
	Binders	3.85%	3.49%	3.56%	3.73%
	Envelopes	1.24%	1.42%	1.49%	1.26%
	Fasteners	0.60%	0.73%	0.63%	0.66%
	Labels	0.60%	0.58%	0.54%	0.60%
	Paper	1.89%	1.92%	2.07%	1.85%
	Storage	9.10%	8.54%	9.09%	8.92%
	Supplies	2.11%	1.62%	1.94%	2.01%
Technology	Accessories	5.02%	6.44%	6.16%	5.90%
	Copiers	9.58%	12.22%	12.20%	12.80%
	Machines	7.11%	5.97%	5.82%	6.05%
	Phones	14.93%	13.60%	13.32%	12.84%
Grand Total		100.00%	100.00%	100.00%	100.00%

Table across than down

Category	Sub-Catego..	Order Date 2012	2013	2014	2015
Furniture	Bookcases	11.48%	11.88%	11.04%	11.94%
	Chairs	12.65%	11.02%	12.55%	11.47%
	Furnishings	2.83%	3.06%	3.28%	2.98%
	Tables	6.51%	6.13%	5.94%	5.66%
Office Supplies	Appliances	7.67%	8.33%	7.49%	8.37%
	Art	2.84%	3.08%	2.88%	2.97%
	Binders	3.85%	3.49%	3.56%	3.73%
	Envelopes	1.24%	1.42%	1.49%	1.26%
	Fasteners	0.60%	0.73%	0.63%	0.66%
	Labels	0.60%	0.58%	0.54%	0.60%
	Paper	1.89%	1.92%	2.07%	1.85%
	Storage	9.10%	8.54%	9.09%	8.92%
	Supplies	2.11%	1.62%	1.94%	2.01%
Technology	Accessories	5.02%	6.44%	6.16%	5.90%
	Copiers	9.58%	12.22%	12.20%	12.80%
	Machines	7.11%	5.97%	5.82%	6.05%
	Phones	14.93%	13.60%	13.32%	12.84%
Grand Total		100.00%	100.00%	100.00%	100.00%

Table down than across

Pane across than down

Category	Sub-Catego..	Order Date 2012	2013	2014	2015
Furniture	Bookcases	11.48%	11.88%	11.04%	11.94%
	Chairs	12.65%	11.02%	12.55%	11.47%
	Furnishings	2.83%	3.06%	3.28%	2.98%
	Tables	6.51%	6.13%	5.94%	5.66%
Office Supplies	Appliances	7.67%	8.33%	7.49%	8.37%
	Art	2.84%	3.08%	2.88%	2.97%
	Binders	3.85%	3.49%	3.56%	3.73%
	Envelopes	1.24%	1.42%	1.49%	1.26%
	Fasteners	0.60%	0.73%	0.63%	0.66%
	Labels	0.60%	0.58%	0.54%	0.60%
	Paper	1.89%	1.92%	2.07%	1.85%
	Storage	9.10%	8.54%	9.09%	8.92%
	Supplies	2.11%	1.62%	1.94%	2.01%
Technology	Accessories	5.02%	6.44%	6.16%	5.90%
	Copiers	9.58%	12.22%	12.20%	12.80%
	Machines	7.11%	5.97%	5.82%	6.05%
	Phones	14.93%	13.60%	13.32%	12.84%
Grand Total		100.00%	100.00%	100.00%	100.00%

Pane down than across

PERFORM EXERCISE 7 AND 8 (45 MINUTES)

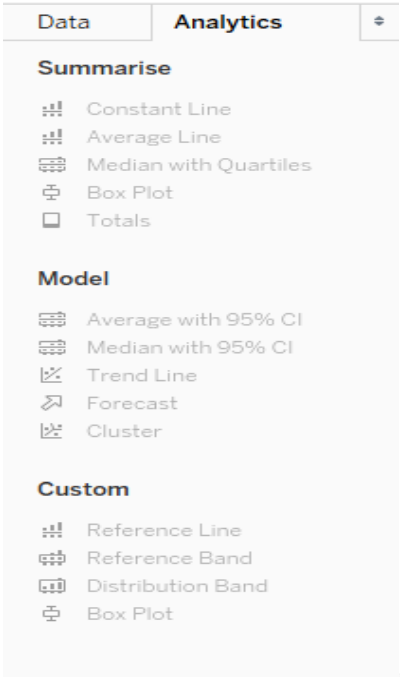
STATISTICS AND FORECASTING

STATISTICS AND FORECASTING

- Tableau offers a range of power analytics tool
- This section will cover
  - Viewing Distribution
  - Boxplot
  - Trend lines
  - Forecasting

USING ANALYTICS PANE

- Tableau consists a variety of common Data Analytics tools in the Analytic Pane
- It allows for quick drag and drop for your analytic work



PERFORM EXERCISE 9 TO 10 (45 MINUTES)

END OF TRAINING