Introduction to

1

Power BI

Agenda

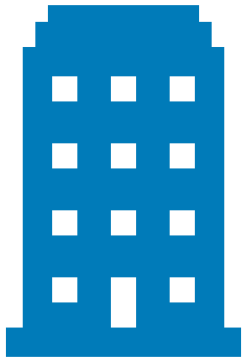
1. Introduction to Power BI
    - Components
    - Architecture
    - Product Portfolio
    - Life Hack: Guide to install Pro
2. Desktop Features
3. Power BI Services and Integration with Various Apps
4. Power Query Editor: The Heart of Power BI
5. Understanding DAX
6. Power BI Functions
7. Power BI Visuals
8. Power BI Charts
9. Power BI KPIs
10. Administration Options
11. Data Visualization
12. Exploratory Data Analysis
13. Project: Subscriber Churn

Power BI

- Business Analytics Solution that lets you visualize the data
- **Share insights** to stakeholders and business owners
- Components
  - Power BI Desktop
  - Power BI service (SaaS –Software as a Service)
  - Power BI Mobile Apps
- Common Workflow
  - Begins by connecting to data sources and building a report in Power BI Desktop
  - Publish report from **Power BI Desktop** to the Power BI service
  - Share it to end users with the **Power BI Service**
  - **Mobile Devices** can view and interact with the report

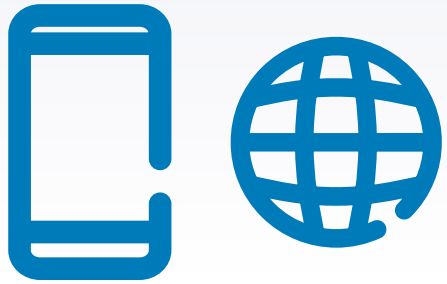
Architecture

Cloud Services



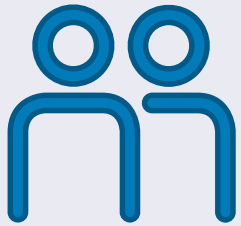
On Premises

- Out of the box SaaS content packs
- Real time dashboards & interactive reports
- Natural Language query
- Custom visualizations
- Native Office 365 integration



Mobile, Web, Excel, Cortana

Product Portfolio


Desktop

Author

Free data analysis and reporting authoring tool

Service

Share & Collaborate

Cloud based modern business analytics solution

Premium

Large Scale Deployments

Dedicated capacity for increased performance

Report Server

Share & Collaborate

On-premises report server

Embedded

App Dev

Visual analytics embedded in your applications

Life Hack

- Download Power BI Desktop:
  - https://powerbi.microsoft.com/en-us/downloads/
- How to sign up for Power BI without a work email?
  - Use incognito browser
  - Log in to office.com
  - Enterprise ? Plans & Pricing ? E3 or E5 Account
  - Try for free!

2

Desktop Features

Power BI Desktop

Get Data

Analyze

Visualize

Publish

Collaborate

Power BI Desktop

Get Data

Easily connect, clean, and mashup data

- Connect to **80+ data sources**, both on-premises and cloud
- **Shape, transform, and clean data** for analysis
- **Live connectivity** to on-premises and cloud data sources
- Extend with **custom data connectors** for any data source
- Prep your data using the familiar **Power Query experience** on the web
- Get started quickly with a **common data model**
- Extend self-service prep to **Azure Data Lake Storage**

Power BI Desktop

Analyze

Build powerful models and flexible measures

- **Automatically create model** when connecting to data
- **High performance, in-memory engine**
- **Point and click analysis** with Quick measures, clustering & binning
- Create **powerful measures with familiar DAX** (Data Analysis Expressions) formulas

Power BI Desktop

Visualize

Create stunning interactive reports

- Author reports using **150+ visuals via a drag-drop canvas**
- Explore data across **multiple interactive visualizations**
- **Provide insights** in the context of the business with Custom Visuals
- Visualize data story with **bookmarks and customer navigation**

Power BI Desktop

Publish

Share insights with others

- Publish directly to the **cloud or on-premises**
- **Automatic data refresh**, so the reports are always up to date
- Package your reports in apps for **easy consumption and control**
- Manage analytics content with **admin and governance tools**

Power BI Desktop

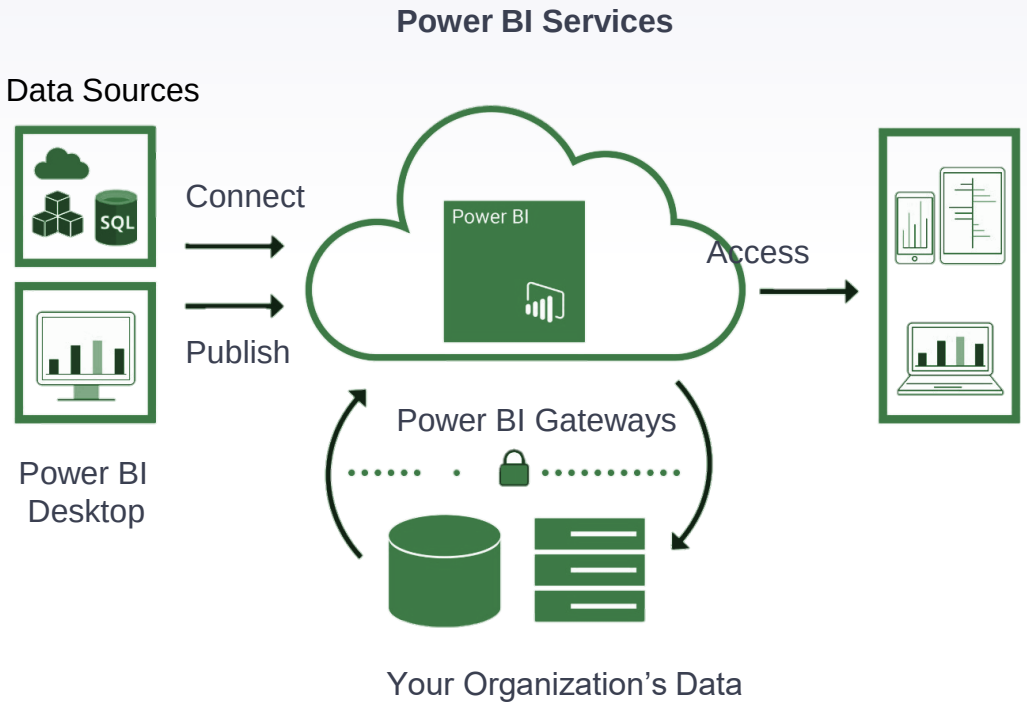
Collaborate

Empower your organization with self-service analytics

3

Power BI Services & Integration with other Apps

Power BI Services



Secure, live connection to the data sources on-premises and in the cloud

1. Keep data anywhere
2. Keep data fresh

Integration with Power BI

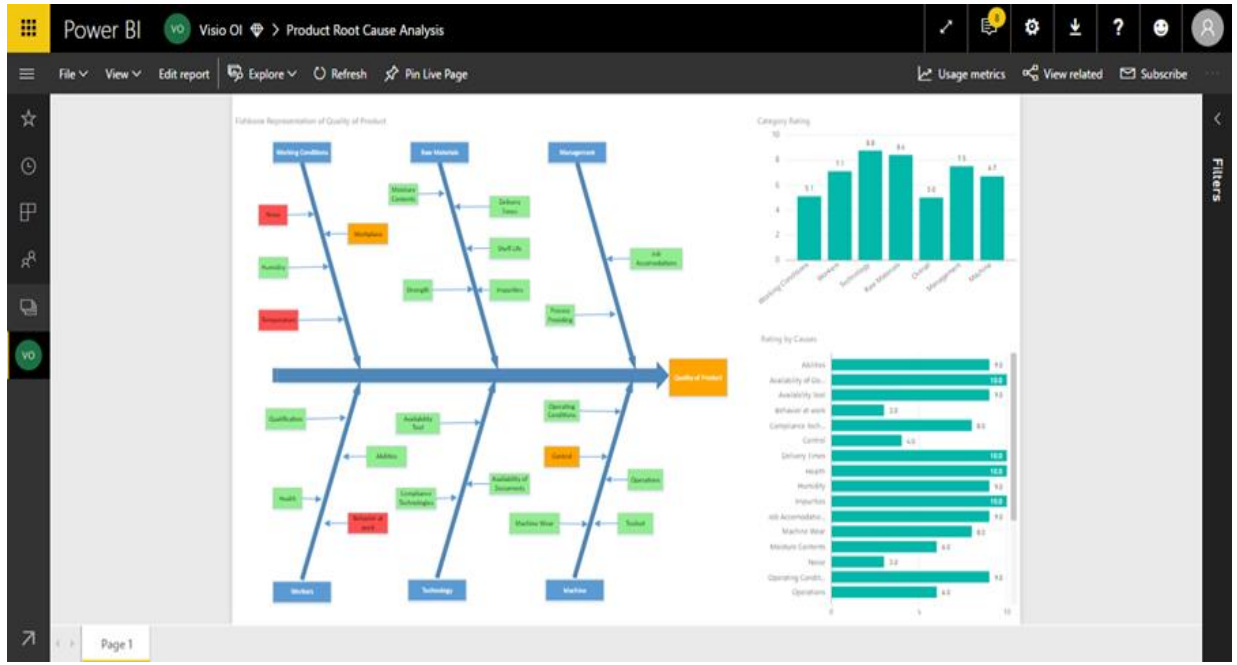
Deliver insights through other services

1. **Collaborate and share insights** with teams in your organization using existing services
2. Fully **interactive reports** integrated into the service

Excel and Power BI

Easily aggregate objects from multiple Excel files on the same dashboard in Power BI

- Analyze in Excel
- Use Excel to view and interact with a dataset you have in Power BI
- Import Excel data into Power BI
- Connect to the data in your workbook so you can create Power BI report and dashboards
- Upload your Excel file to Power BI
- Bring your Excel file into Power BI to view and interact with it just as you would in Excel Online. Pin ranges to Dashboards

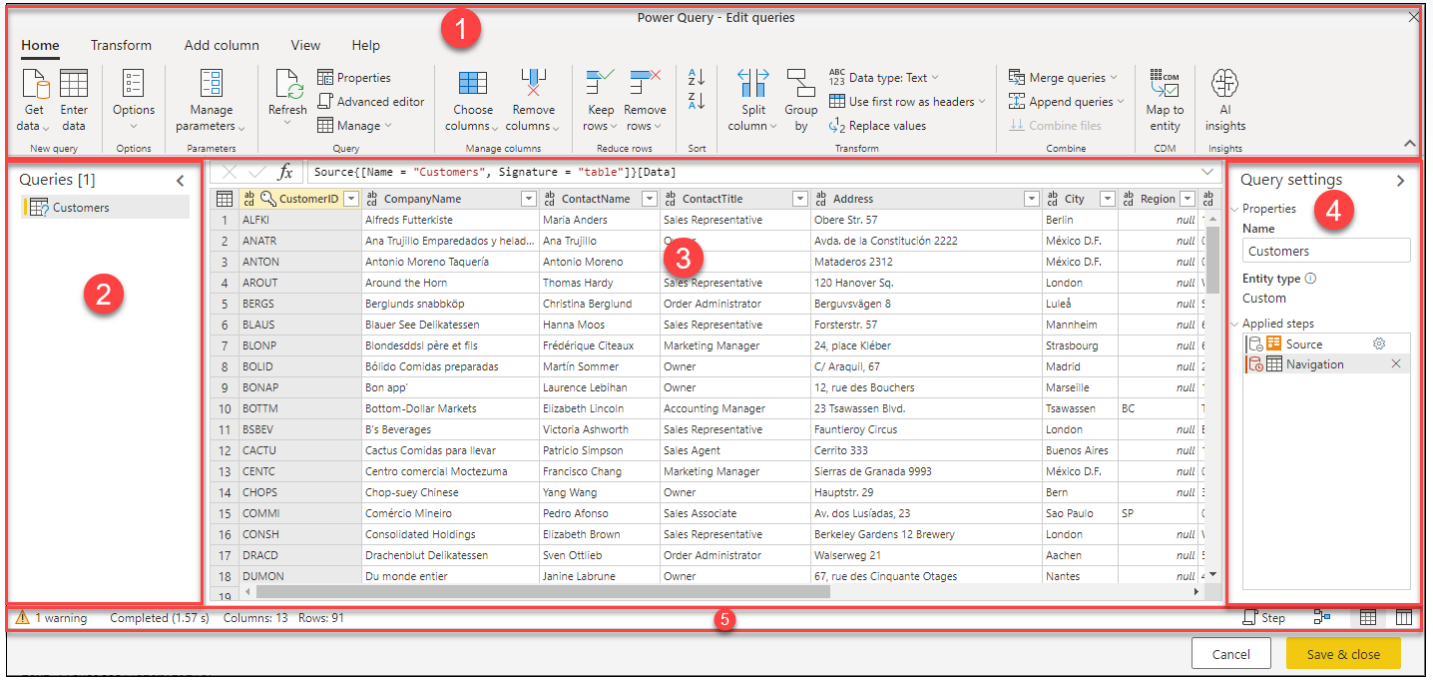


4

Power Query Editor: The Heart of Power BI

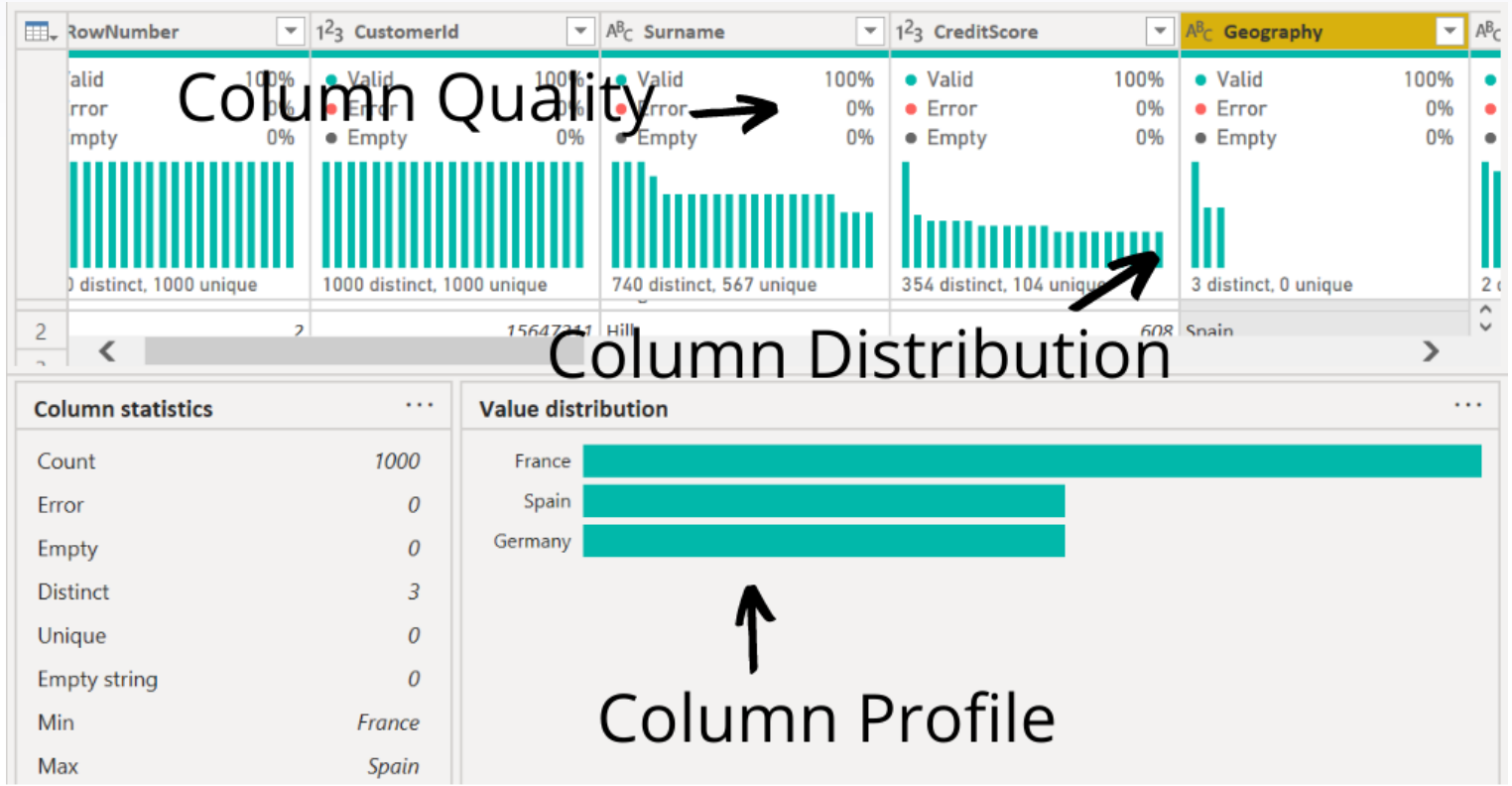
User Experience

- Power Query Editor represents the **user interface**
- **Modify or Add Queries**
- **Manage Queries** by grouping or adding descriptions to query steps
- **Visualize queries** and their structure
- **Five Distinct Components**

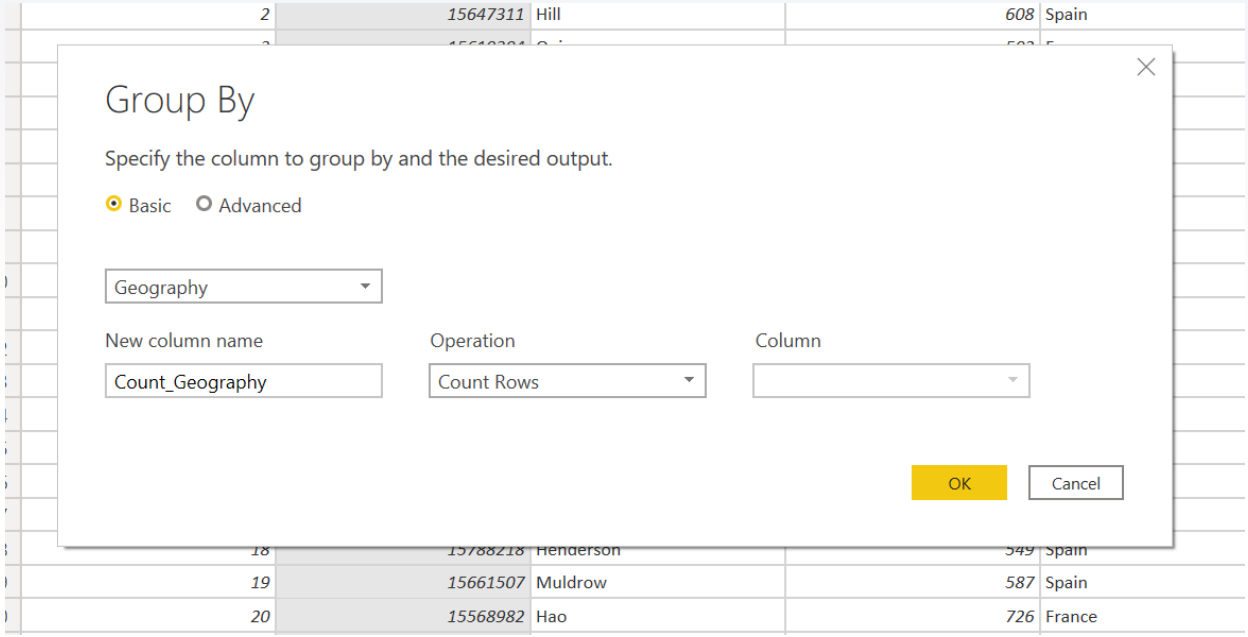


Data Profiling Tools

- Provide new and intuitive ways to **clean, transform, and understand data**
- Includes:
  - **Column Quality**
  - **Column Distribution**
  - **Column Profile**



Group By Dialog



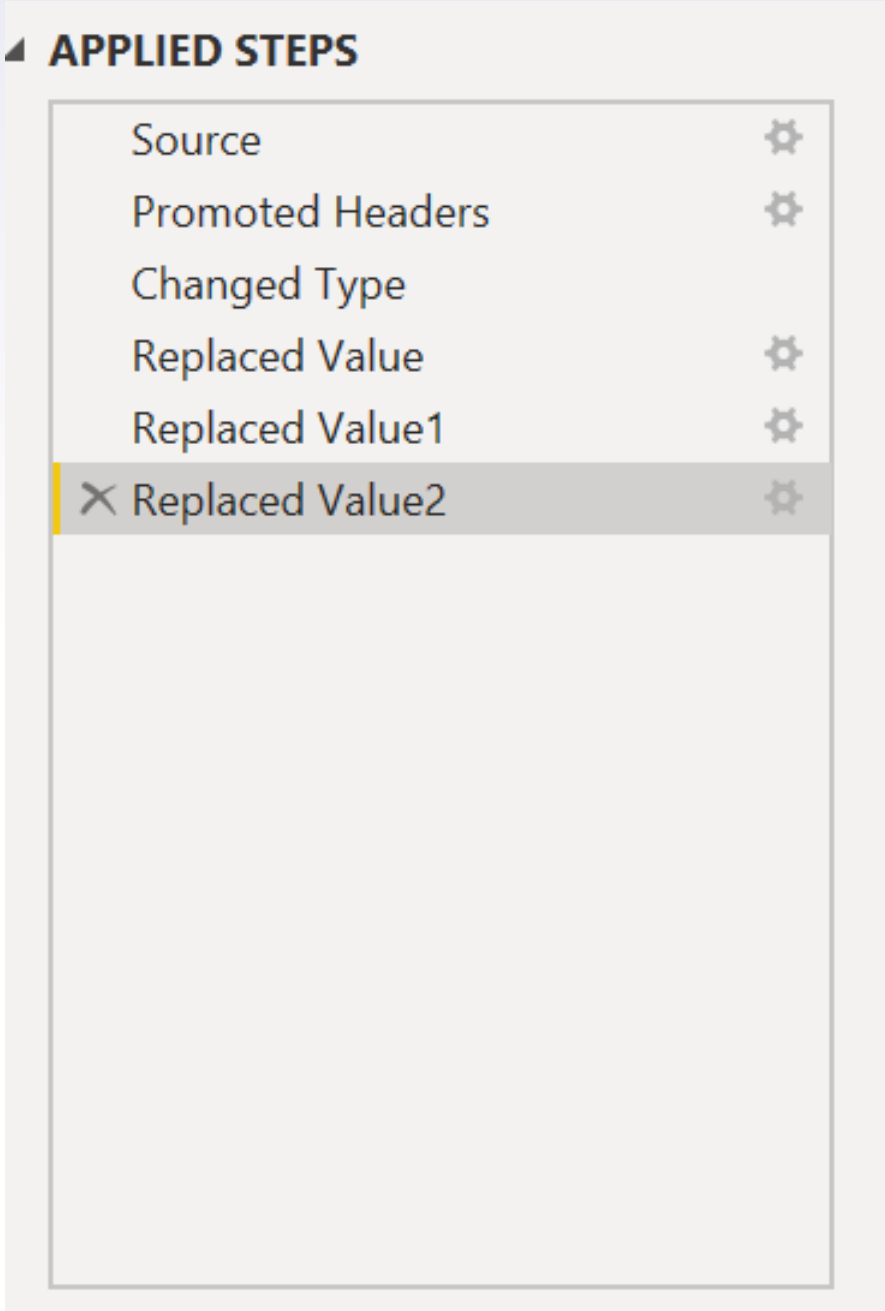
Set the **Group By operation** to:

- Group by the Geography
- Count the number of supplier rows per Geography

	Geography	Count_Geography
1	France	5014
2	Spain	2477
3	Germany	2509

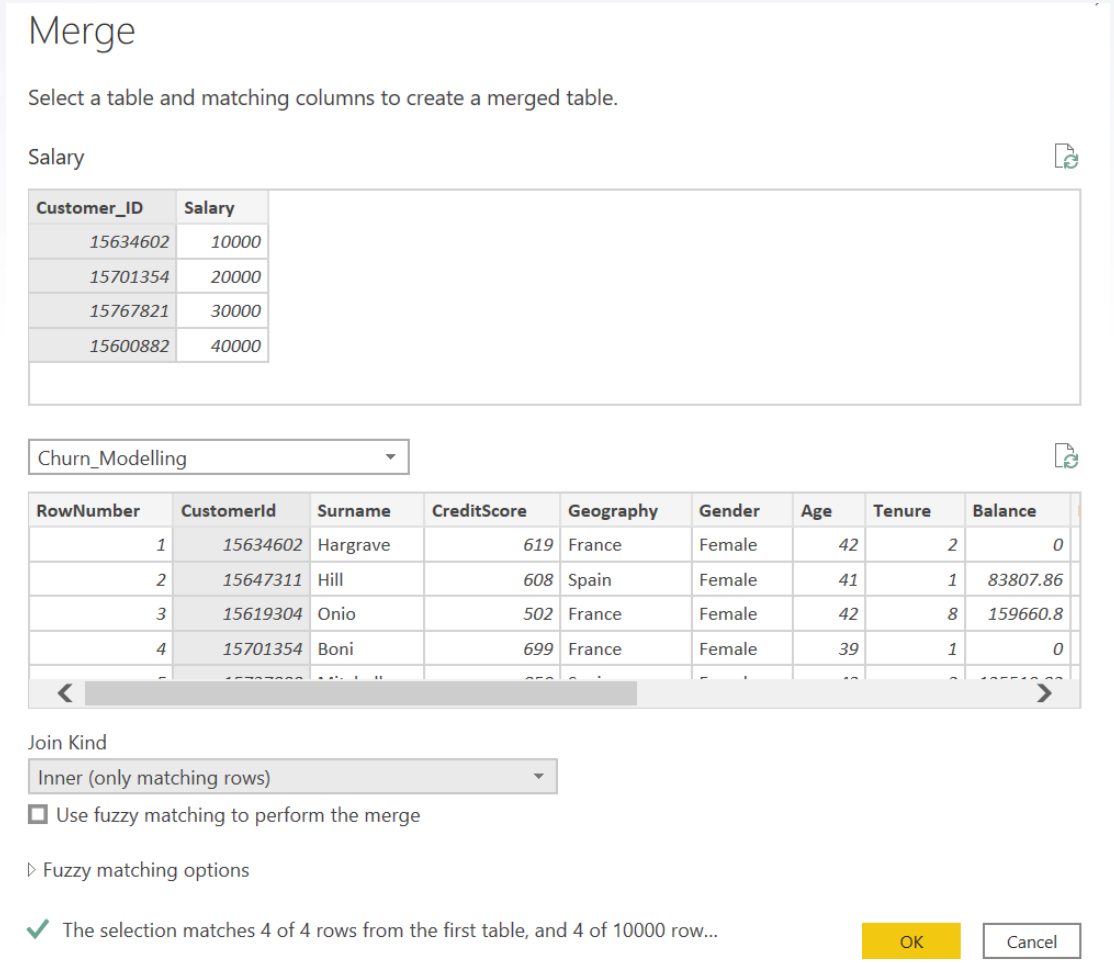
Applied Steps

- Any steps performed in Power BI is logged under the **Applied Steps**
- Steps can be **added or deleted anytime** during the process



Appending vs Merging

- **Merging**
  - When you have one or more columns that you'd like to add to another query
- **Appending**
  - When you have additional rows of data that you'd like to add to an existing query



5

Understanding Data Analysis Expression

What do you need to know?

- Contexts
  - Row Context
  - Filter Context
- Formatting
- Best Practice
- X vs non-X functions
- Time Intelligence functions

- Functions
  - SUM
  - AVERAGE
  - MIN
  - MAX
  - COUNT
  - COUNTROWS
  - CALCULATE
  - FILTER, etc.

DAX: Data Analysis Expression

- Two Business Logics
  - **Measures**
  - **Calculated Columns**
- Difference?
  - Context of Evaluation
  - Measures
    - Evaluated in the **context of the cell** evaluated in a report or in a DAX query
  - Calculated Column
    - Computed at the **row level** within the table it belongs to

Measures

- Represents a single value per data model
- Computed at run time
- Dynamic results, based on filters
- Filter Context
- Not attached to any specific table

```
TotalQuantity := SUM(Sales[Quantity])
```

Calculated Columns

- Represents a single value per row
- Computed at compile time
- Dynamic Results, based on Rows
- Row Context

```
Tenure_Months := Churn[Tenure]*12
```

Implicit Measures

If we use a calculated column as a value/result, it creates an ***implicit measure***.

- For example:
  - If we have columns such as:
    - Tenure in years,
    - Monthly average usage
  - Goal: to create the overall average usage for that customer

```
Churn[Tenure_Months] = Churn[Tenure]*12
```

  - Total usage would be:

```
Churn[Total Usage] = Churn[Tenure_Months] * Churn[Monthly_Average_Usage]
```

  - Change in the Primitive Column, i.e. Tenure, will impact the change in the Total Usage column

DAX is great at two things in particular

Aggregations & Filtering

Aggregations: Combining a group of values into one value

Examples: Sum, Average, Min, Max, Distinct Count

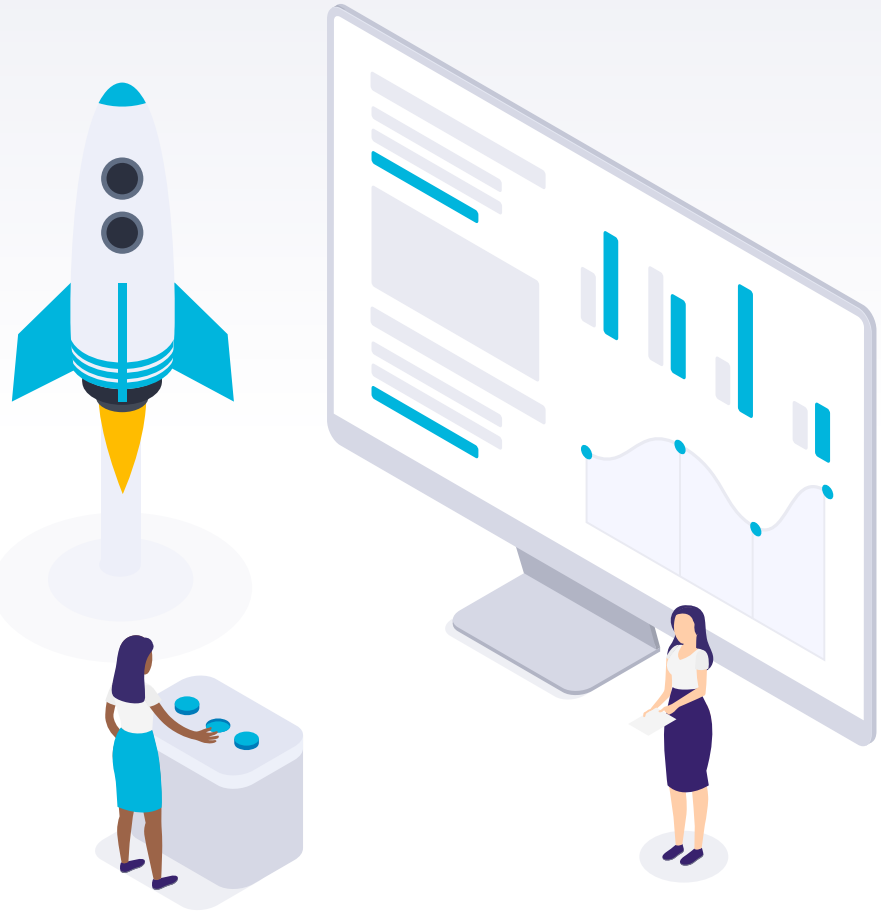
Power BI: Example

Let's do a SUM(Column)? How to check if it's correct?

If you are using select SUM(quantity) from tablename;

6

Power BI Functions



Power BI: Functions

SUM

AVERAGE

MIN

MAX

COUNT

COUNTROWS

DATEDIFF

DATEADD

Average and Datediff

Probation Period = DATEDIFF(column1, column2, DAY)

Average = AVERAGE(column)

Calculated Table

Dates = CALENDAR(range)

- Creates a dates table with a date per day between the specified range
- Also creates a Date Hierarchy

Contexts

- Two different contexts: **1. Row context, 2. Filter context**
- We've been using it for all our calculated columns so far, let's revisit our first DAX

```
Tenure in Years = ROUND(Churn_Modelling[Tenure]/12,2)
```

- Notice we **expect a value per row** in a table
- This runs at import and gets stored
- Might increase file size

Filter Context

- Easy to show with measures

Calculate: Breaking out of the filter context

```
Total Sales - Beverages =
CALCULATE(sum('Sales OrderDetails
'[Order Line Total]), 'Production
Categories'[categoryname] = "Beverages")
```

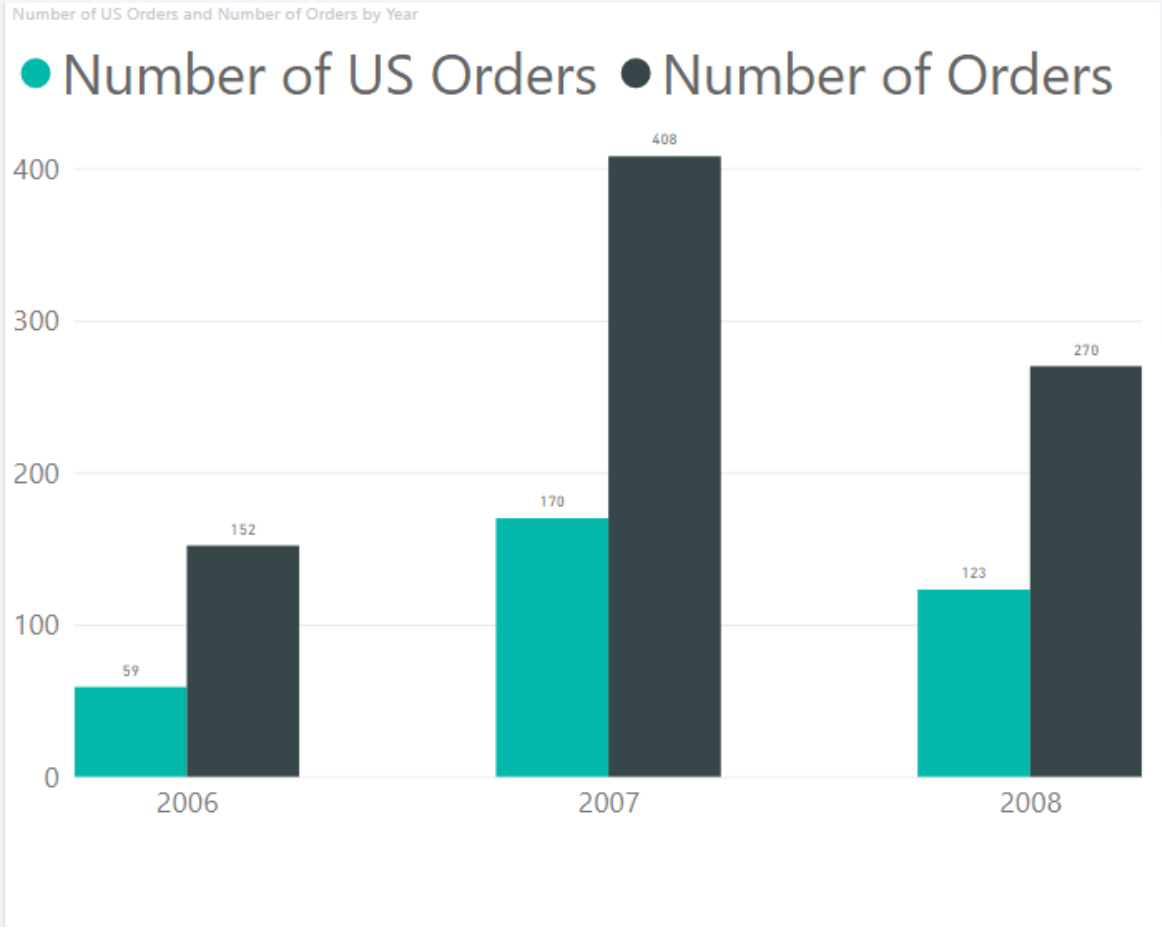
Year	Total Sales	Total Sales - Beverages
2006	$2,26,298.5	$53,879.2
2007	$6,58,388.75	$1,10,424
2008	$4,69,771.34	$1,22,223.75
Total	**$13,54,458.59**	**$2,86,526.95**

categoryname	Meat/Poultry		Produce		Seafood		**Total**	
Year	Sales	Total Sales - Beverages	Total Sales	Total Sales - Beverages	Total Sales	Total Sales - Beverages	**Total Sales**	**Total Sales - Beverages**
2006	),292.2	$53,879.2	$15,134.2	$53,879.2	$21,589.6	$53,879.2	**$2,26,298.5**	**$53,879.2**
2007	621.03	$1,10,424	$57,718.55	$1,10,424	$71,320.65	$1,10,424	**$6,58,388.75**	**$1,10,424**
2008	275.57	$1,22,223.75	$32,415.85	$1,22,223.75	$48,712.84	$1,22,223.75	**$4,69,771.34**	**$1,22,223.75**
Total	**,188.8**	**$2,86,526.95**	**$1,05,268.6**	**$2,86,526.95**	**$1,41,623.09**	**$2,86,526.95**	**$13,54,458.59**	**$2,86,526.95**

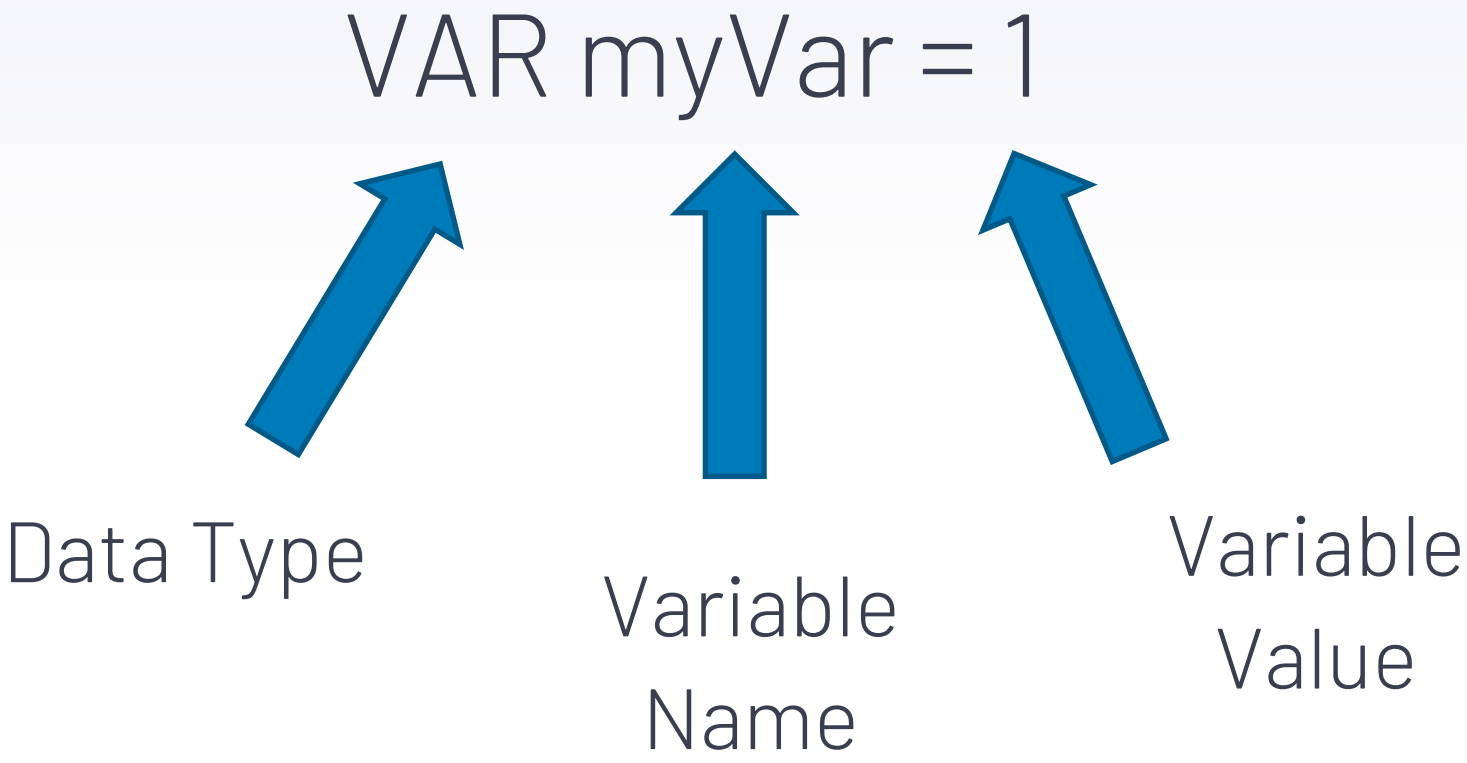
Filter

```
Number of US Orders =
CALCULATE(COUNT(
'SalesOrderDetails'[orderid]),
FILTER( 'Sales Customers' , 'Sales
Customers'[country] = "USA" ))
```

```
Number of Orders = COUNT('Sales
Orders'[orderid])
```



Variables



RETURN myVar + 25

If-Else and Nested If Blocks

- Similar concepts like other programming languages.

```
Age_Bins = IF(Churn_Modelling[ Age ]>=60,
"Above 60", "Below 60")
```

Time Intelligence Functions

- Enables user to **manipulate data using time periods** such as years, quarters, months, and days
- Creating calculations over those time periods
- **Most common time periods:**
  - Year– to – Date
  - Quarter – to – Date
  - Month – to – Date
  - Last Year
  - Full Year
  - Rolling 12 Months

Time Intelligence: TOTALYTD

```
YTD Total Sales =
TOTALYTD(SUM('Sales OrderDetails
'[Order Line Total]),
Dates[Date].[Date])
```

Month	2006	2007	2008	**Total**
January		$66,692.8	$1,00,854.72	
February		$1,07,900	$2,05,416.67	
March		$1,47,879.9	$3,15,242.12	
April		$2,03,579.29	$4,49,872.68	
May		$2,60,402.99	$4,69,771.34	
June		$2,99,490.99	$4,69,771.34	
July	$30,192.1	$3,54,955.92	$4,69,771.34	
August	$56,801.5	$4,04,937.61	$4,69,771.34	
September	$84,437.5	$4,64,670.63	$4,69,771.34	
October	$1,25,641.1	$5,34,999.13	$4,69,771.34	
November	$1,75,345.1	$5,80,912.49	$4,69,771.34	
December	$2,26,298.5	$6,58,388.75	$4,69,771.34	
Total	**$2,26,298.5**	**$6,58,388.75**	**$4,69,771.34**	

Time Intelligence: PREVIOUSMONTH

```
Total Sales Previous Month =
CALCULATE(sum('Sales OrderDetails'[Order Line Total]),
PREVIOUSMONTH(Dates[Date]))
```

Year	2006		2007		2008		Total	
Month	Total Sales	Total Sales Previous Month	Total Sales	Total Sales Previous Month	Total Sales	Total Sales Previous Month	**Total Sales**	**Total Sales Pre**
December	$50,953.4	$49,704	$77,476.26	$45,913.36			**$1,28,429.66**	
November	$49,704	$41,203.6	$45,913.36	$70,328.5			**$95,617.36**	
October	$41,203.6	$27,636	$70,328.5	$59,733.02			**$1,11,532.1**	
September	$27,636	$26,609.4	$59,733.02	$49,981.69			**$87,369.02**	
August	$26,609.4	$30,192.1	$49,981.69	$55,464.93			**$76,591.09**	
July	$30,192.1		$55,464.93	$39,088			**$85,657.03**	
June			$39,088	$56,823.7		$19,898.66	**$39,088**	
May			$56,823.7	$55,699.39	$19,898.66	$1,34,630.56	**$76,722.36**	
April			$55,699.39	$39,979.9	$1,34,630.56	$1,09,825.45	**$1,90,329.95**	
March			$39,979.9	$41,207.2	$1,09,825.45	$1,04,561.95	**$1,49,805.35**	
February			$41,207.2	$66,692.8	$1,04,561.95	$1,00,854.72	**$1,45,769.15**	
January			$66,692.8	$50,953.4	$1,00,854.72	$77,476.26	**$1,67,547.52**	
Total	**$2,26,298.5**		**$6,58,388.75**	**$50,953.4**	**$4,69,771.34**	**$77,476.26**	$13,54,458.59	

non-X vs X functions (SUM vs SUMX)

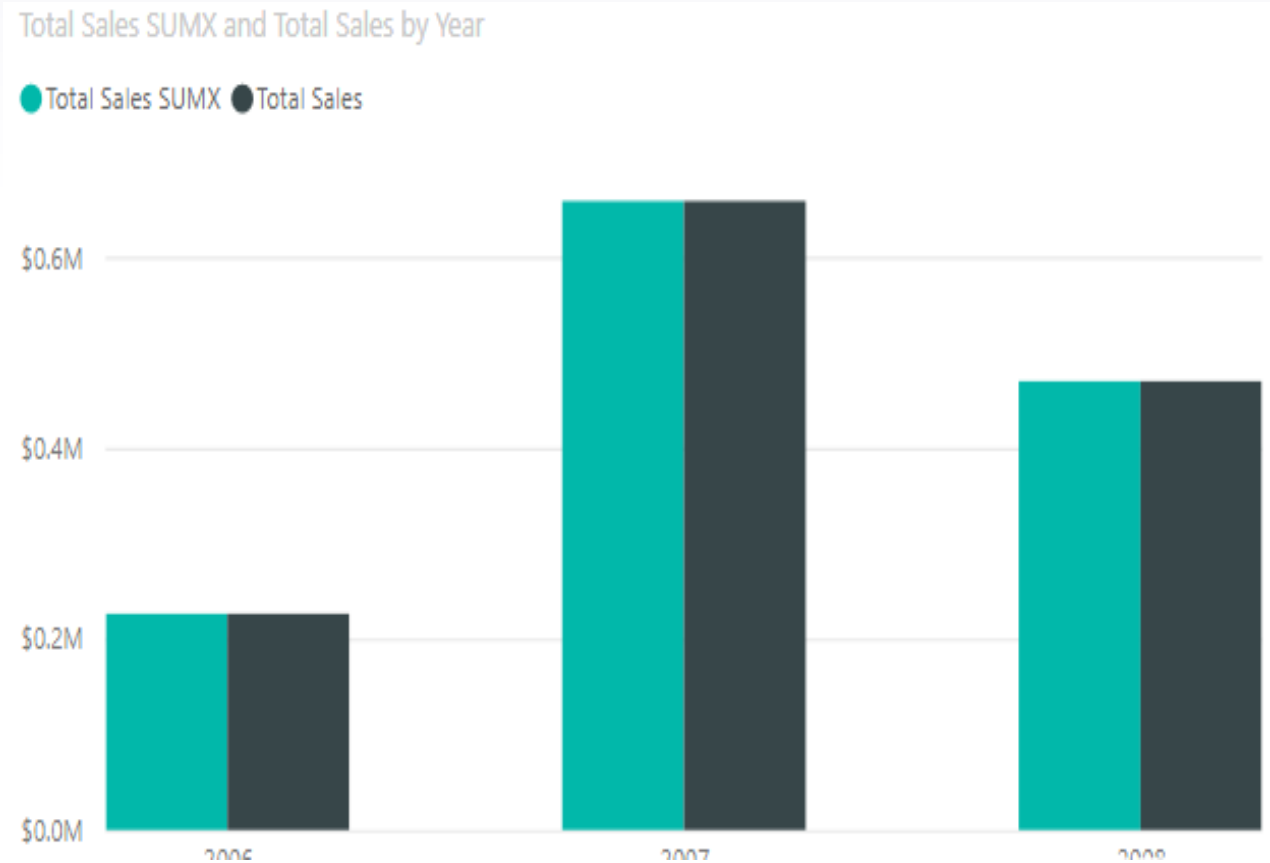
SUM is an **aggregator function**.
It works like a measure, **calculating based on the current filter context**.

SUMX is an **iterator function**. It works row by row. SUMX has awareness of rows in a table, and can **reference the intersection of each row with any columns** in the table.

non-X vs X functions (SUM vs SUMX) – An Example

```
Total Sales SUMX =
SUMX('Sales OrderDetails',
'Sales OrderDetails'[qty]*
'Sales OrderDetails'[unitprice])
```

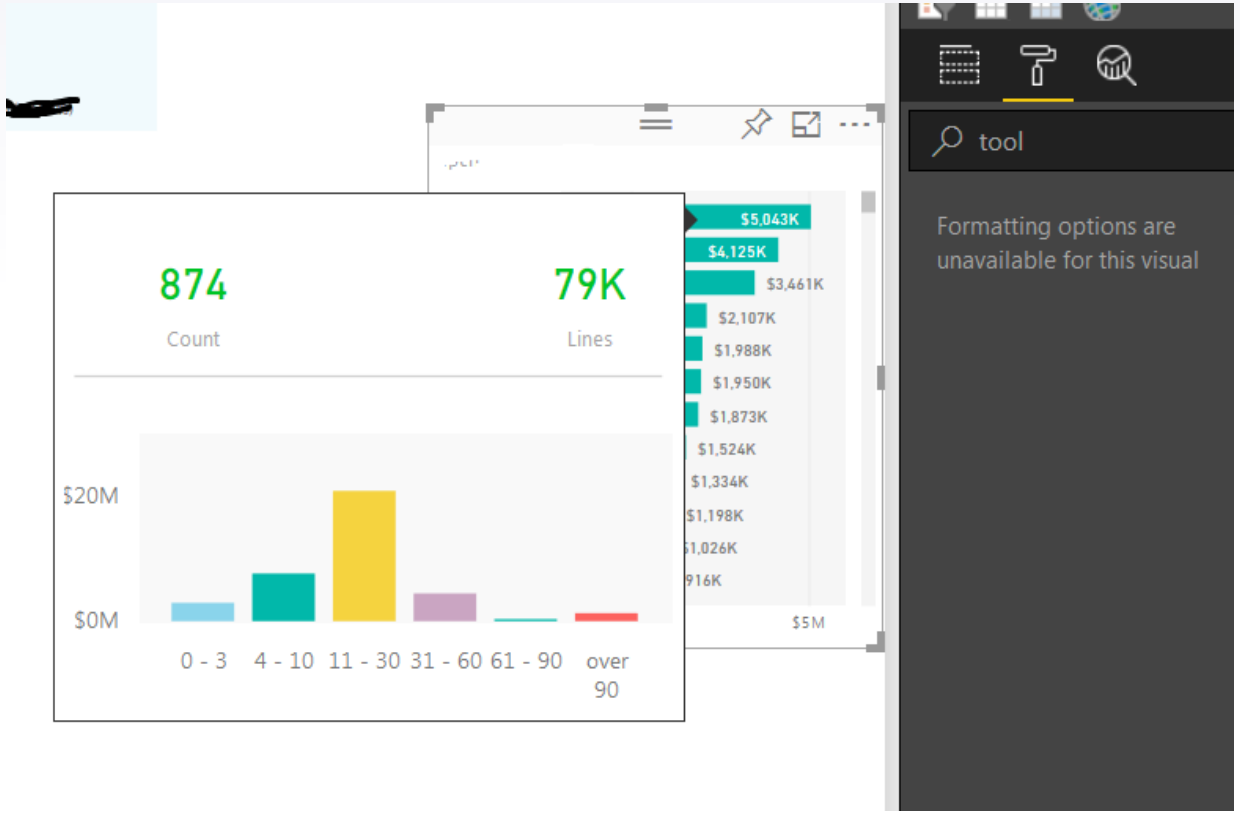
```
Total Sales =
sum('Sales OrderDetails'
[Order Line Total])
```

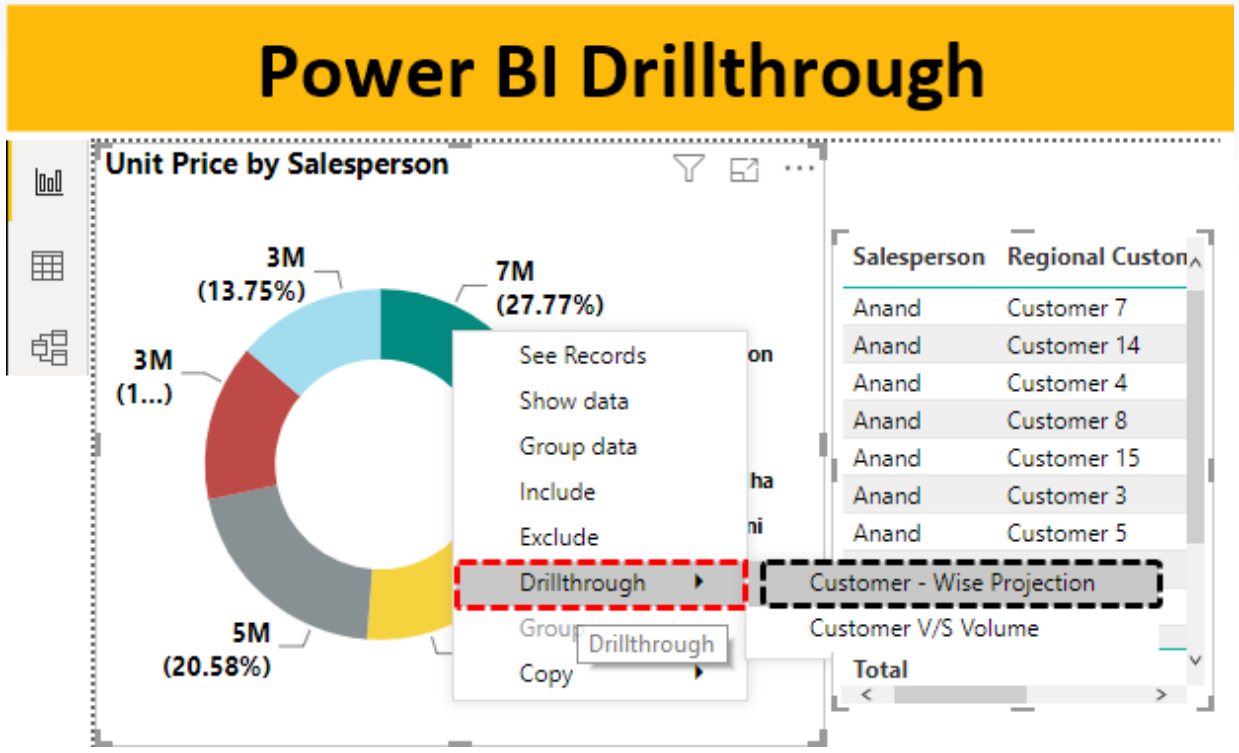


AVERAGE, AVERAGEA, AVERAGE X

- **AVERAGE** → Averages out the data
- **AVERAGEA** → Considers non-integer values as null
- **AVERAGEX** → Creates In memory measure
  - Also an iterator function
  - Works row by row
  - Has awareness of rows in a table
  - Can reference the intersection of each row with any columns in the table

Tool Tips and Drill Throughs





Best Practice: Organize your Code

- Create a **separate table for measures**
- **Limit Visuals:** As visuals interact with each other, if we have more visuals, it might take a lot of time to refresh.
  Tool tips & Drill through can be used.
- **Process as much data** as required in the original source
- **Certified Visuals** are recommended
- Use a **lighter background**

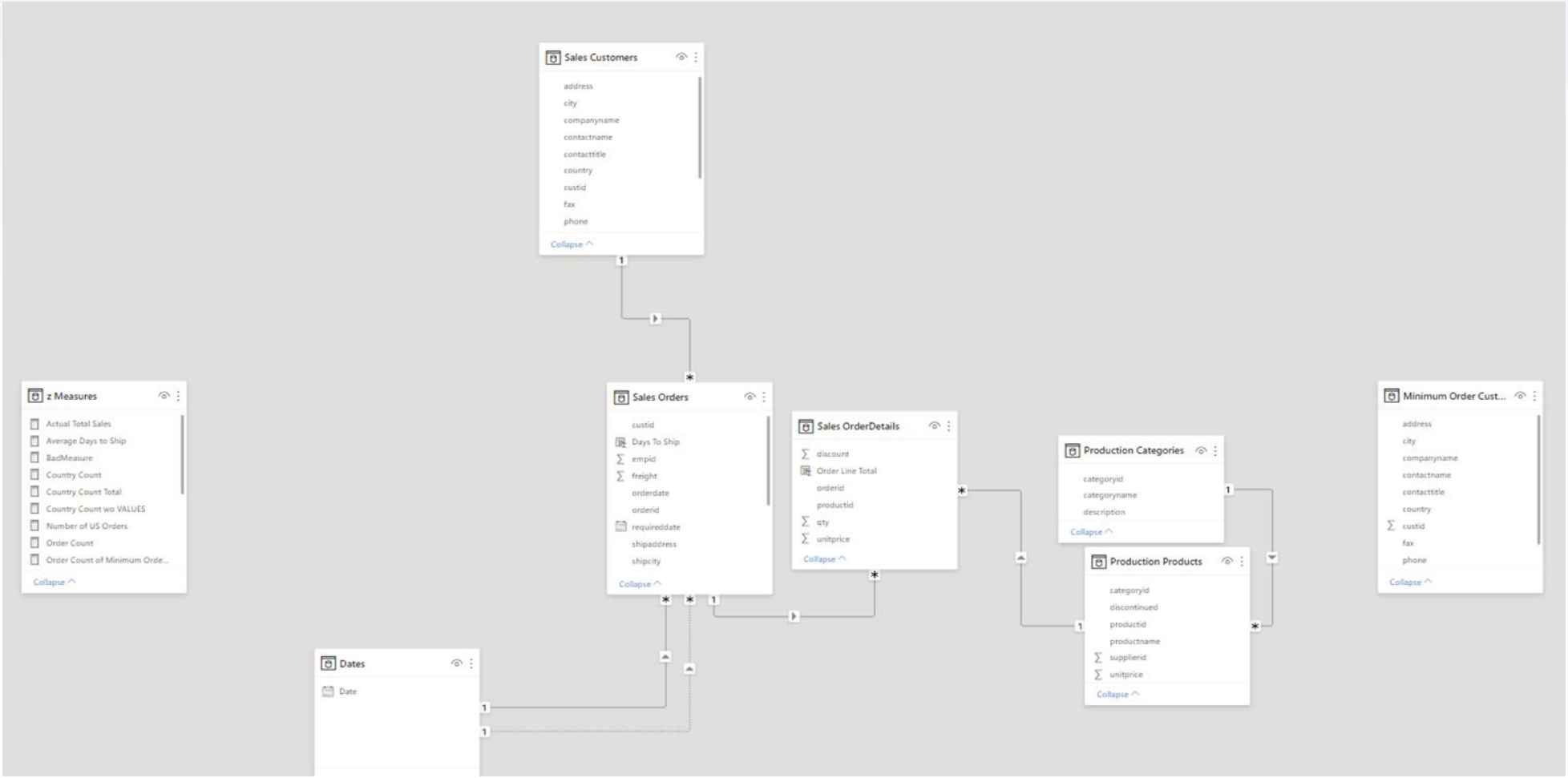
Data Types

- Numeric
- String
- Bool
- DateTime

Data Modding Trends for 2019 and Beyond

- If a function is expecting a numeric, but gets a string, it won't work. Clean up the model and **watch it start working!**
- Uses **less space and memory** with your model
- **Improves performance**

Relationship



Manipulating the Relationship

```
Total Sales By Ship Year =
CALCULATE(SUM('Sales OrderDetails'
[Order Line Total]),USERELATIONSHIP
('Sales Orders'[shippeddate], Dates[Date]))
```

7

Power BI Visuals

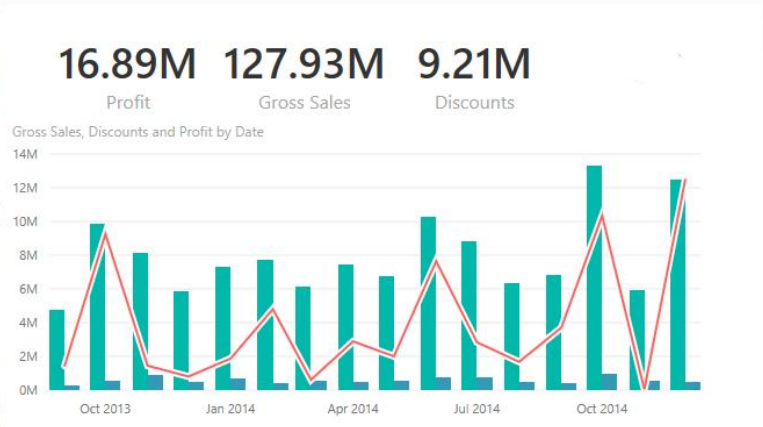
Building Blocks of Power BI

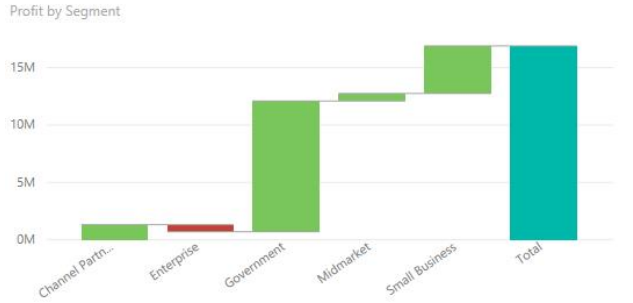
Building Blocks of Power BI

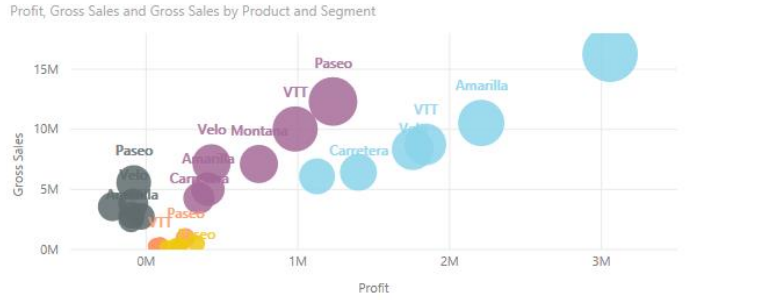
Visualizations | Datasets | Reports | Dashboards | Tiles

A **visual representation of data** is called visualization.

For example, a chart, or a graph can be used to represent data visually.







Building Blocks of Power BI

Visualizations | **Datasets** | Reports | Dashboards | Tiles

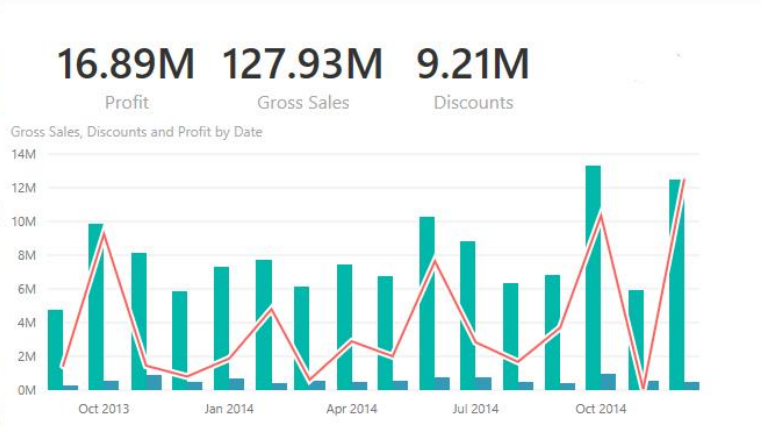
A dataset is a **collection of data or information.**

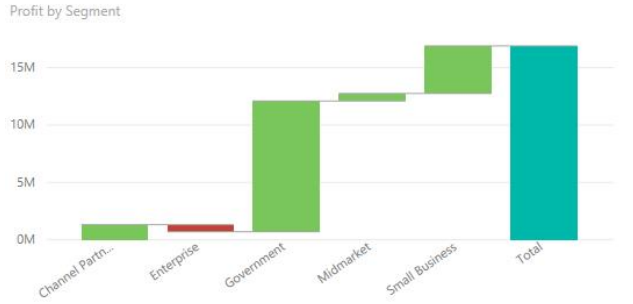
RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance
1	15634602	Hargrave	619	France	Female	42	2	0
2	15647311	Hill	608	Spain	Female	41	1	83807.86
3	15619304	Onio	502	France	Female	42	8	159660.8
4	15701354	Boni	699	France	Female	39	1	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.82
6	15574012	Chu	645	Spain	Male	44	8	113755.78
7	15592531	Bartlett	822	France	Male	50	7	0
8	15656148	Obinna	376	Germany	Female	29	4	115046.74
9	15792365	He	501	France	Male	44	4	142051.07
10	15592389	H?	684	France	Male	27	2	134603.88
11	15767821	Bearce	528	France	Male	31	6	102016.72
12	15737173	Andrews	497	Spain	Male	24	3	0
13	15632264	Kay	476	France	Female	34	10	0
14	15691483	Chin	549	France	Female	25	5	0
15	15600882	Scott	635	Spain	Female	35	7	0
16	15643966	Goforth	616	Germany	Male	45	3	143129.41
17	15737452	Romeo	653	Germany	Male	58	1	132602.88
18	15788218	Henderson	549	Spain	Female	24	9	0
19	15661507	Muldrow	587	Spain	Male	45	6	0
20	15568982	Hao	726	France	Female	24	6	0
21	15577657	McDonald	732	France	Male	41	8	0
22	15597945	Dellucci	636	Spain	Female	32	8	0
23	15699309	Gerasimov	510	Spain	Female	38	4	0
24	15725737	Mosman	669	France	Male	46	3	0
25	15625047	Yen	846	France	Female	38	5	0
26	15738191	Maclean	577	France	Male	25	3	0
27	15736816	Young	756	Germany	Male	36	2	136815.64
28	15700772	Nebechi	571	France	Male	44	9	0
29	15728693	McWilliams	574	Germany	Female	43	3	141349.43
30	15656300	Lucciano	411	France	Male	29	0	59697.17
31	15589475	Azikiwe	591	Spain	Female	39	3	0
32	15706552	Odinakachukwu	533	France	Male	36	7	85311.7
33	15750181	Sanderson	553	Germany	Male	41	9	110112.54
34	15659428	Maggard	520	Spain	Female	42	6	0

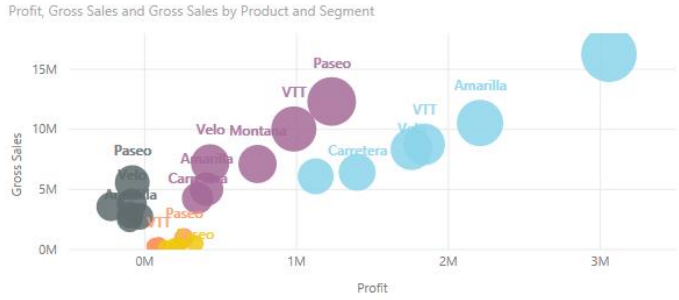
Building Blocks of Power BI

Visualizations | Datasets | **Reports** | Dashboards | Tiles

A **collection of visualizations** that appear together on one or more pages. It is a **collection of items that have common motive**.



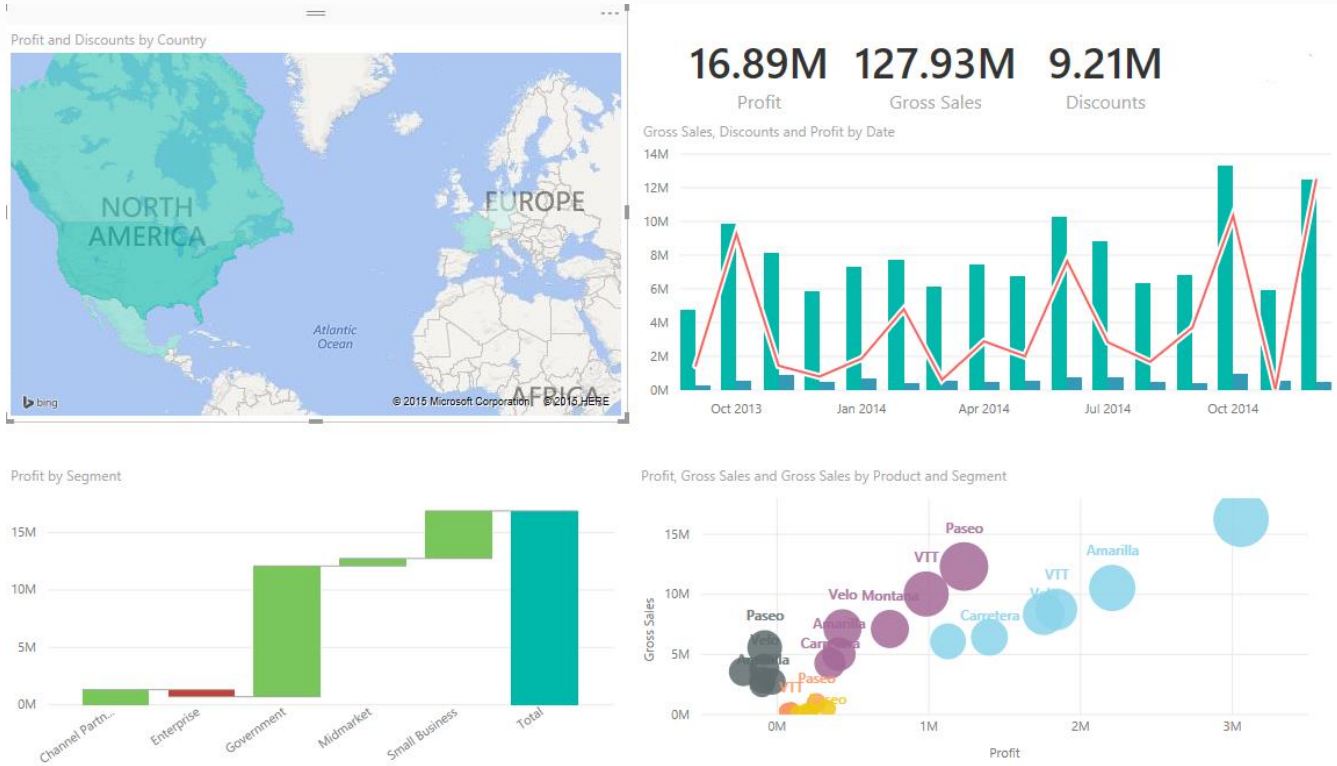




Building Blocks of Power BI

Visualizations | *Datasets* | *Reports* | ***Dashboards*** | *Tiles*

A **single page interface** that uses the most **important elements** of a report to tell a story.



Building Blocks of Power BI

Visualizations → *Datasets* → *Reports* → *Dashboards* → ***Tiles***

A **tile is a single visualization** found in a report or on a dashboard.

8

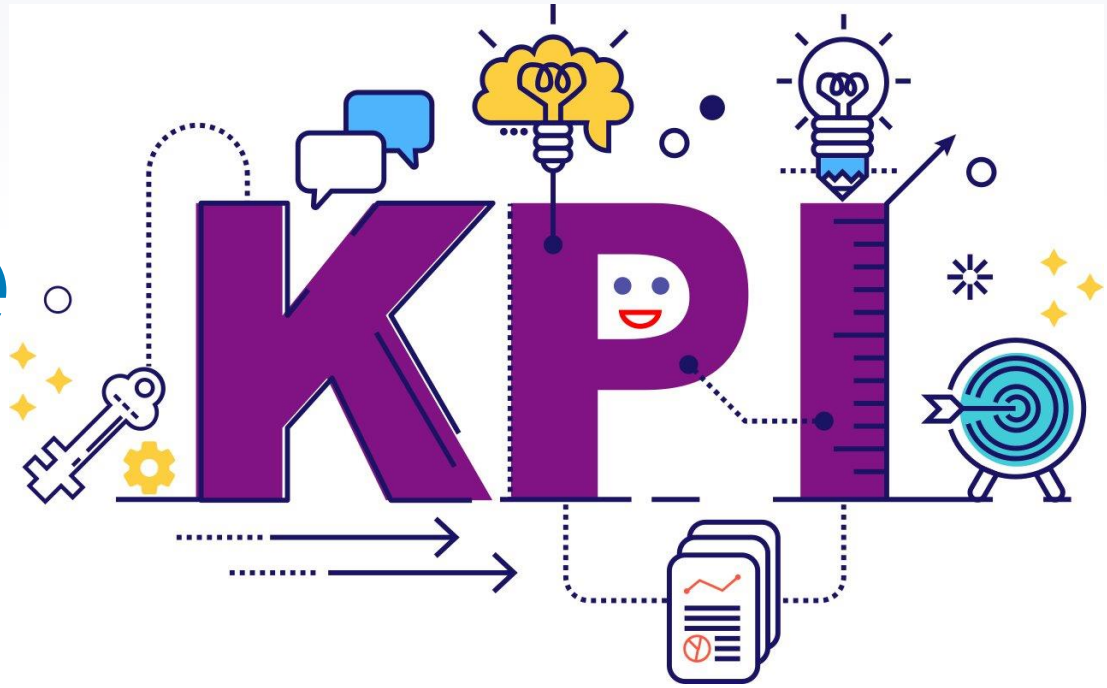
Power BI Charts

Different Charts in Power BI

1. Bar, Column, Line & Area Charts
2. Combination Charts
3. Pie-Charts, Doughnut Charts
4. Maps, Funnel Charts
5. Gauge, Cards, Tables & Matrices

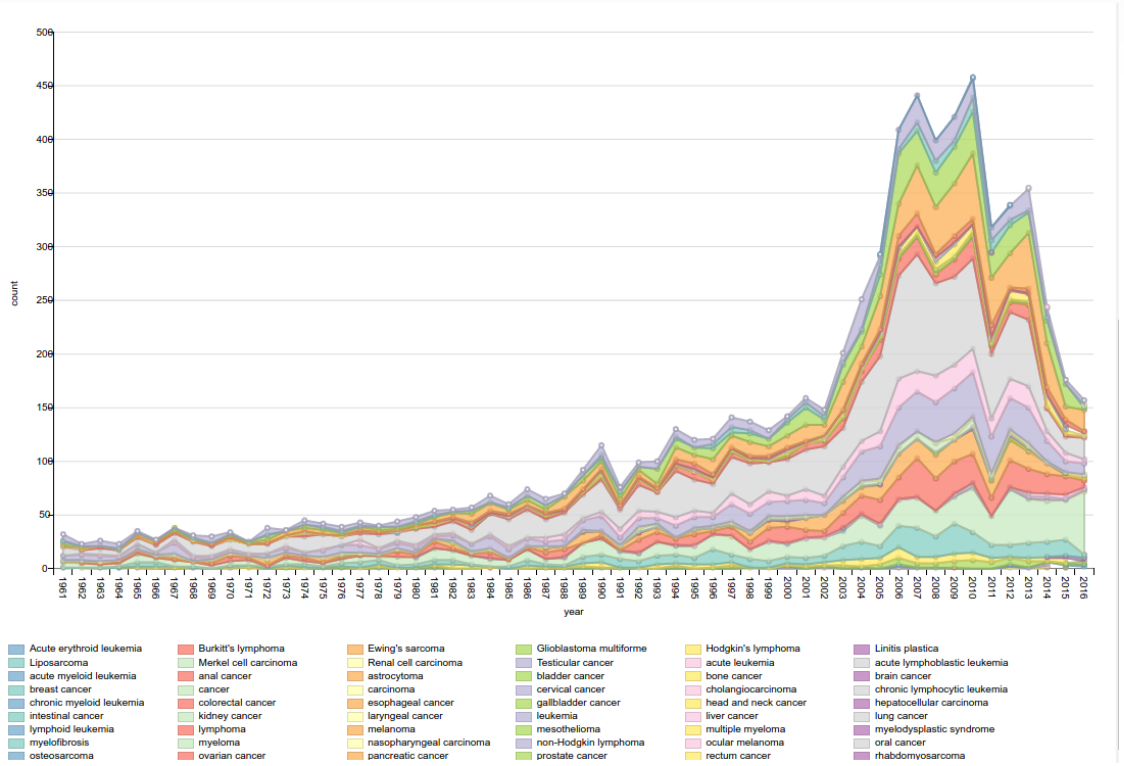
9

Key Performance Indicators



What is a KPI?

A key performance indicator (KPI) is a **visual cue that communicates the amount of progress made toward a target.**



When should we use KPIs?

TARGET

TREND

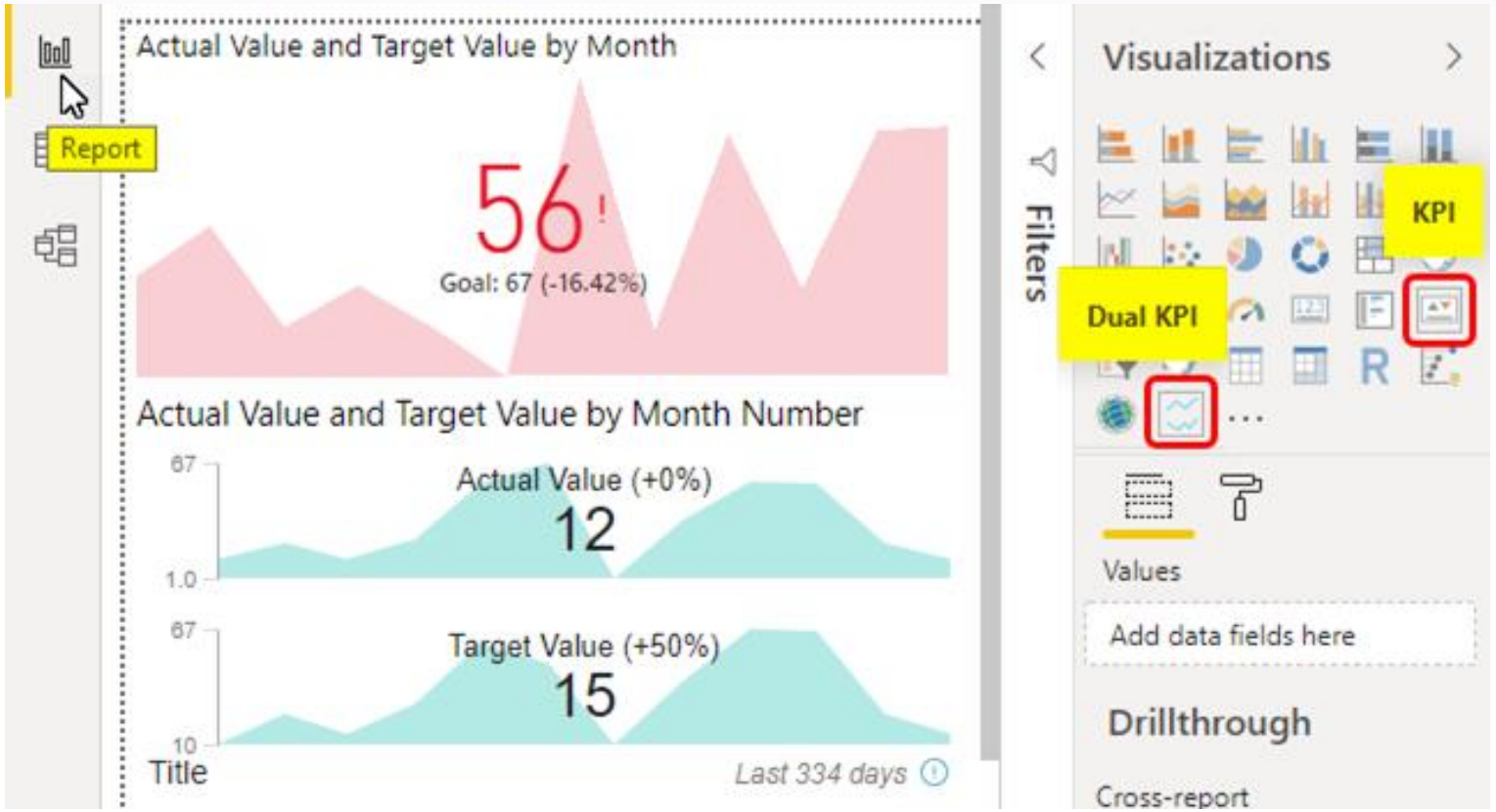
Requirements for KPIs

BASE MEASURE

TARGET MEASURE

THRESHOLD

KPI Visualizations



10

Edit Interactions

11

Formatting Options

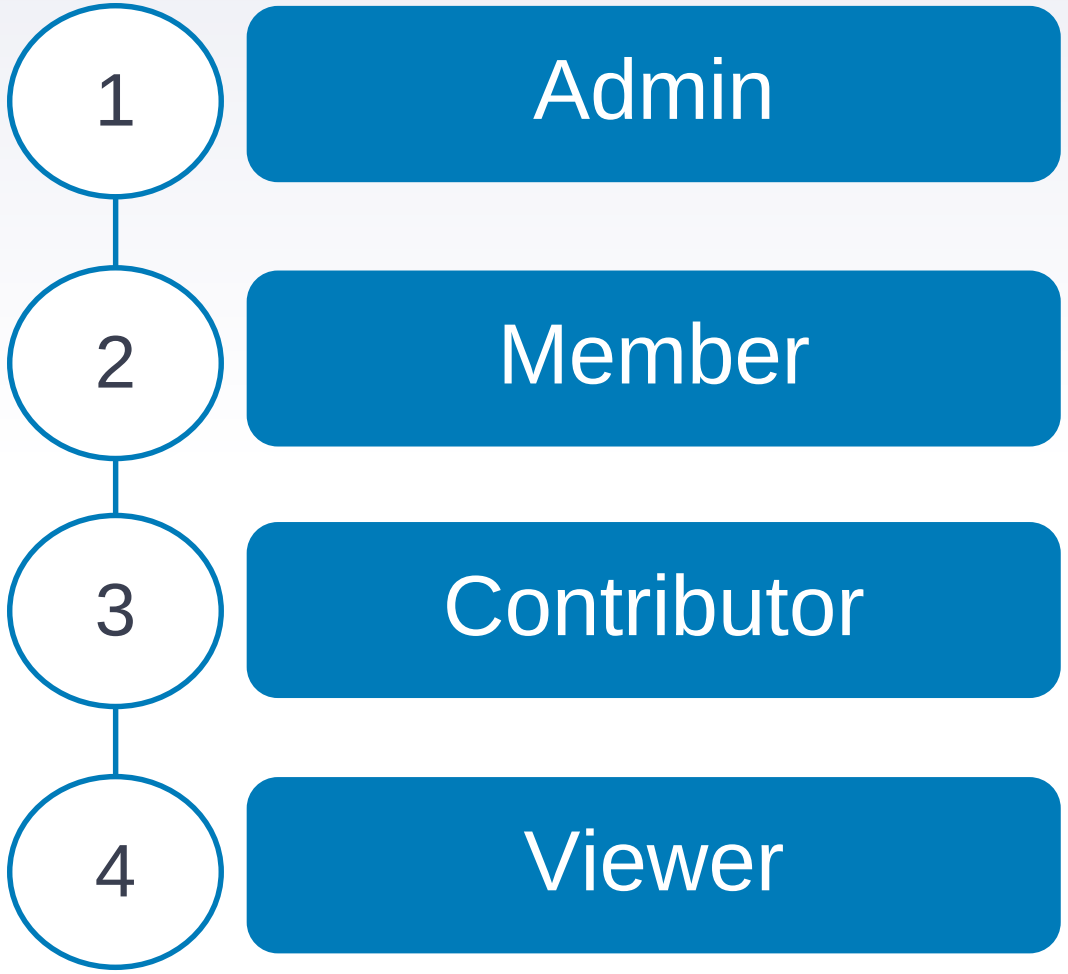
12

Security in Power BI

13

Administration Options

Different Roles in Power BI



Different Roles in Power BI

Link:

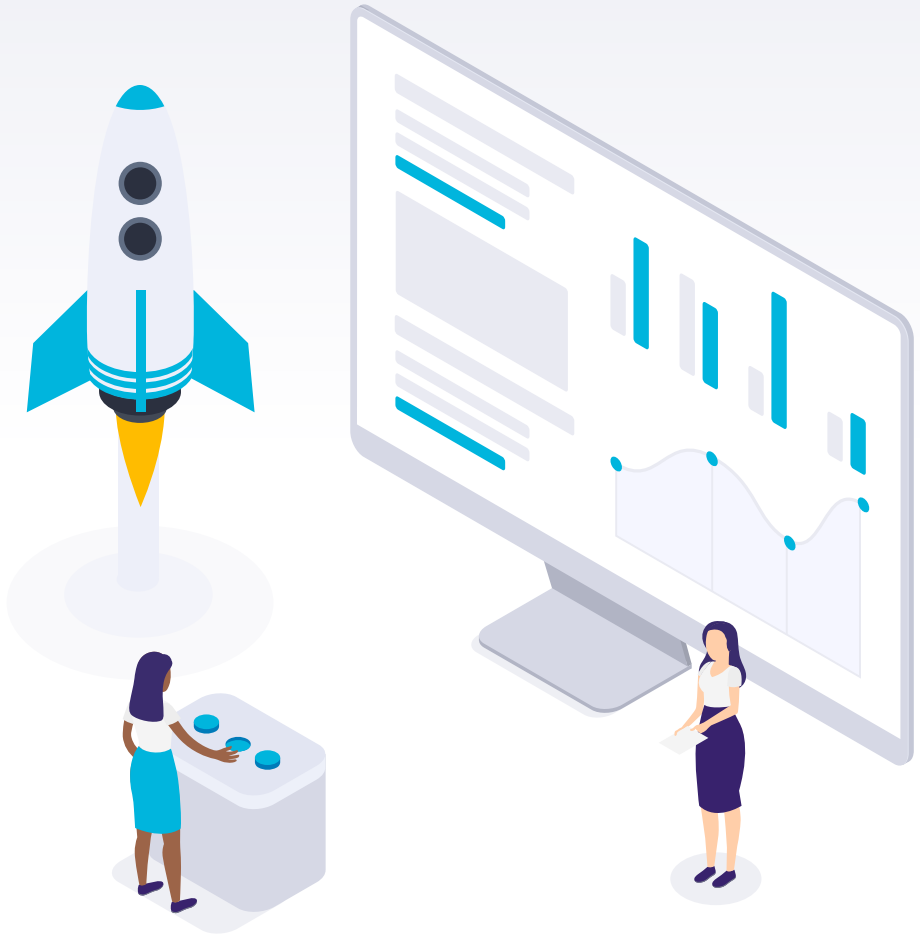
https://docs.microsoft.com/en-us/power-bi/collaborate-share/service-roles-new-workspaces

Workspace roles

Capability	Admin	Member	Contributor	Viewer
Update and delete the workspace.	✓			
Add/remove people, including other admins.	✓			
Allow Contributors to update the app for the workspace	✓			
Add members or others with lower permissions.	✓	✓		
Publish, unpublish, and change permissions for an app	✓	✓		
Update an app.	✓	✓	If allowed [1]	
Share an item or share an app.[2]	✓	✓		
Allow others to reshare items.[2]	✓	✓		
Feature apps on colleagues' Home	✓	✓		
Manage dataset permissions.[3]	✓	✓		
Feature dashboards and reports on colleagues' Home	✓	✓	✓	
Create, edit, and delete content in the workspace.	✓	✓	✓	
Publish reports to the workspace, delete content.	✓	✓	✓	
Create a report in another workspace based on a dataset in this workspace.[3]	✓	✓	✓	
Copy a report.[3]	✓	✓	✓	
Create goals based on a dataset in the workspace.[3]	✓	✓	✓	
Schedule data refreshes via the on-premises gateway.[4]	✓	✓	✓	
Modify gateway connection settings.[4]	✓	✓	✓	
View and interact with an item.[5]	✓	✓	✓	✓
Read data stored in workspace dataflows	✓	✓	✓	✓

14

Data Visualization



Data visualization helps to ***bridge the gap*** between ***numbers and words***

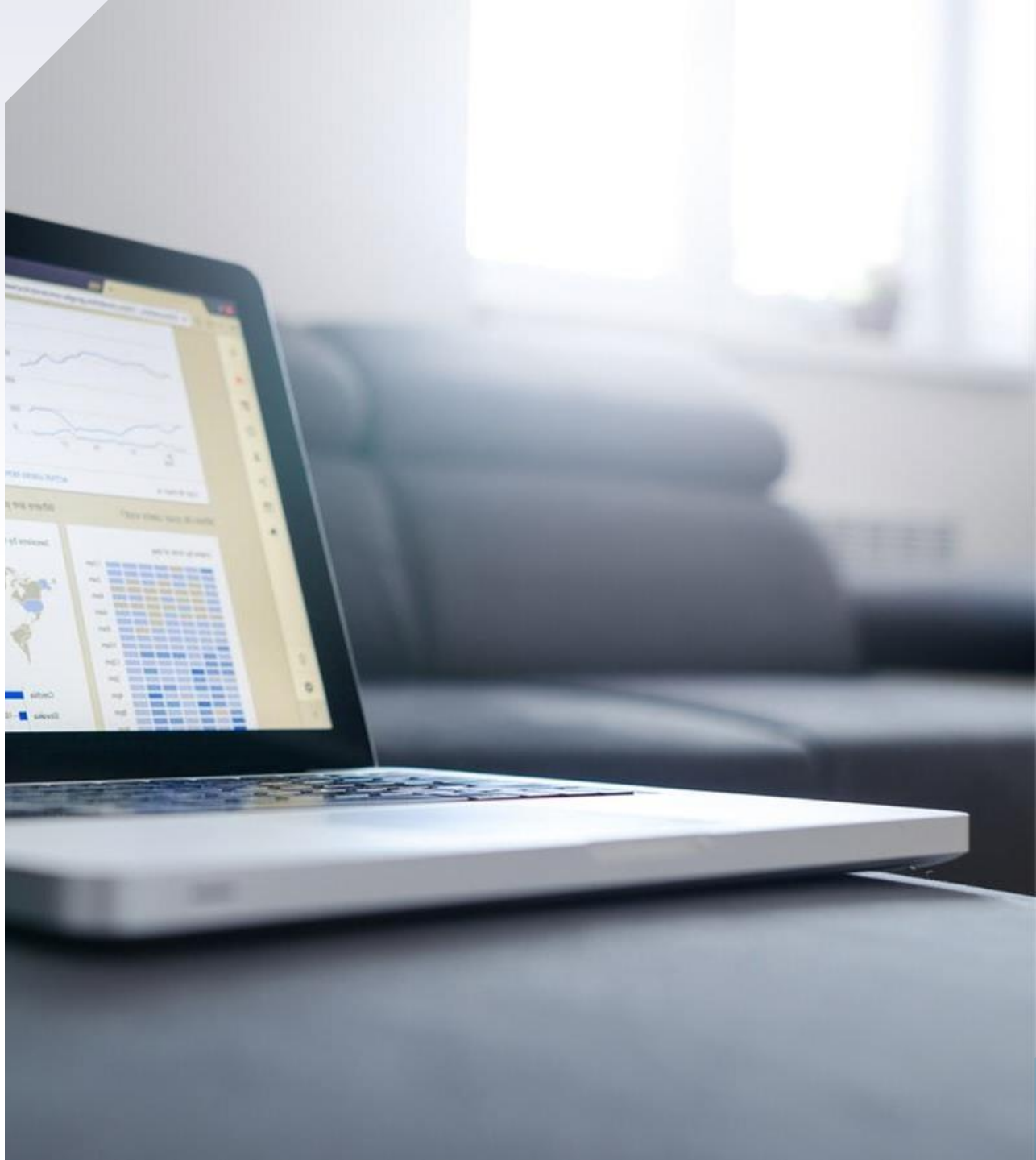
— Brie E. Anderson, Digital Marketer and Data Scientist at BEAST Analytics

Data Visualization

Giving **visual context** to information to help **identify and infer trends, patterns, and outliers in data sets**

A picture is worth a thousand words

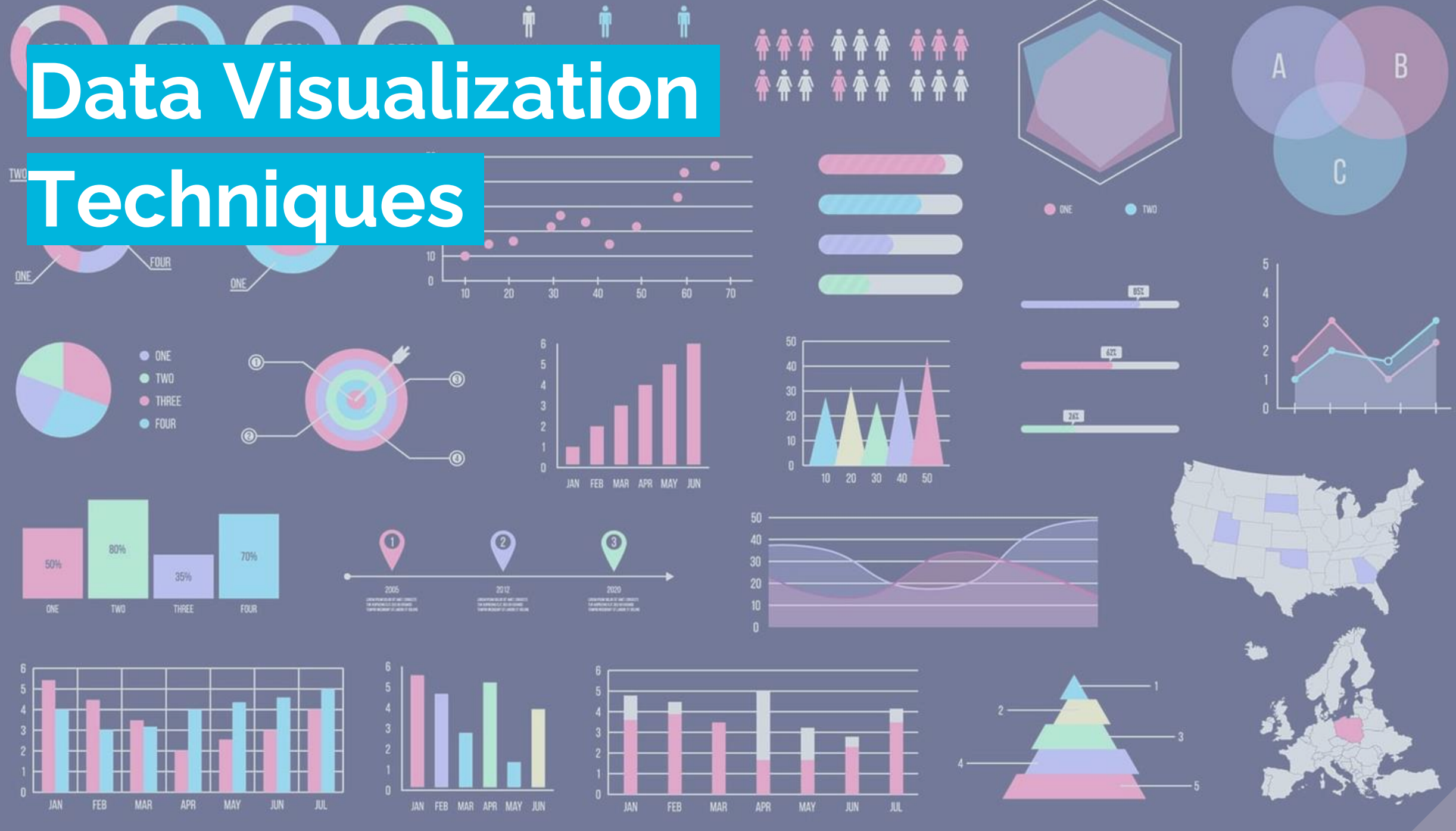
A **complex idea** can be conveyed with just a **single still image**, namely making it possible to **absorb large amounts of data** quickly



Importance of Data Visualization

- Data is only useful **if we can learn from it**
- It delivers data with **efficiency, clarity and effectiveness**
- Can identify patterns, e.g.
  - Correlations
  - Trends over time
  - Frequency
- Analyze **large data sets** and have **data-driven decision management**

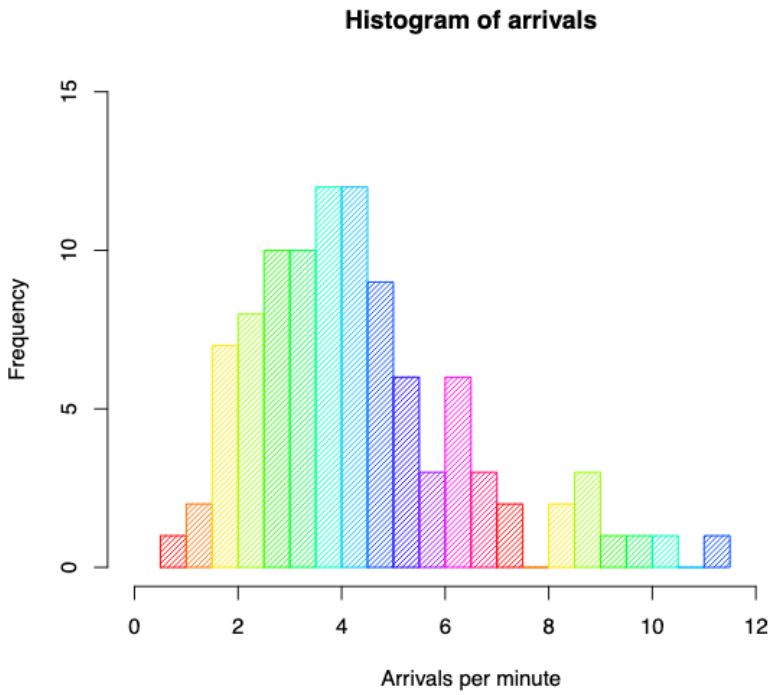
Data Visualization Techniques



Data Visualization Techniques

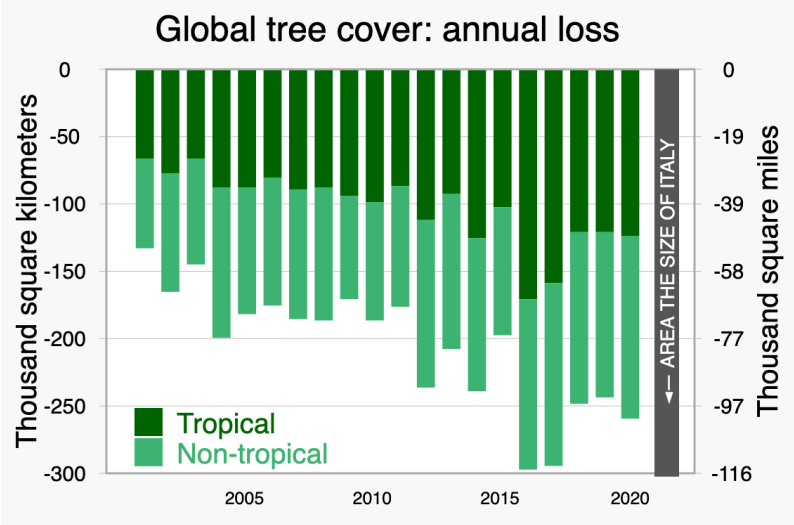
Histograms

Measure frequency distribution of data



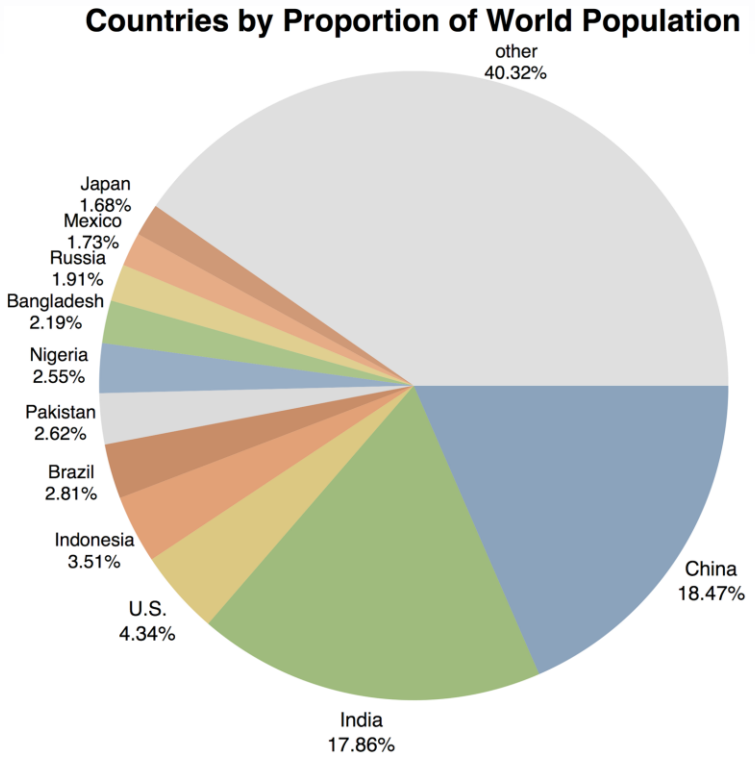
Area / Bar Charts

Represent no. of observations for different categories



Pie Charts

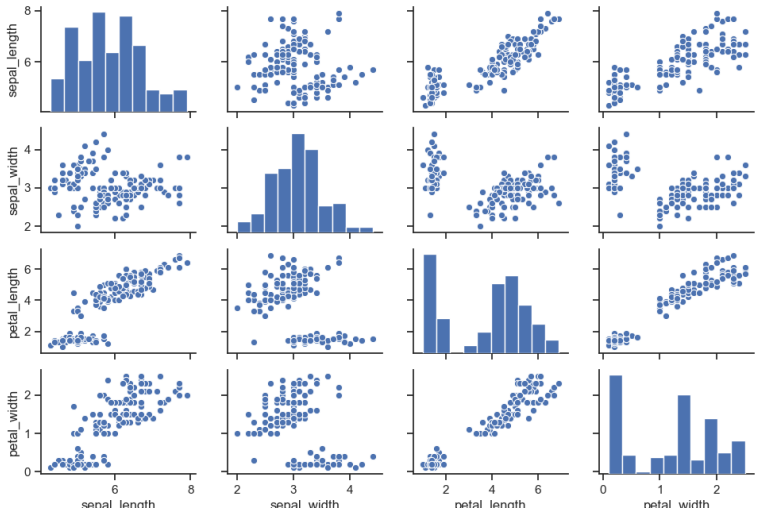
Represent the percentage of data by each category



Data Visualization Techniques

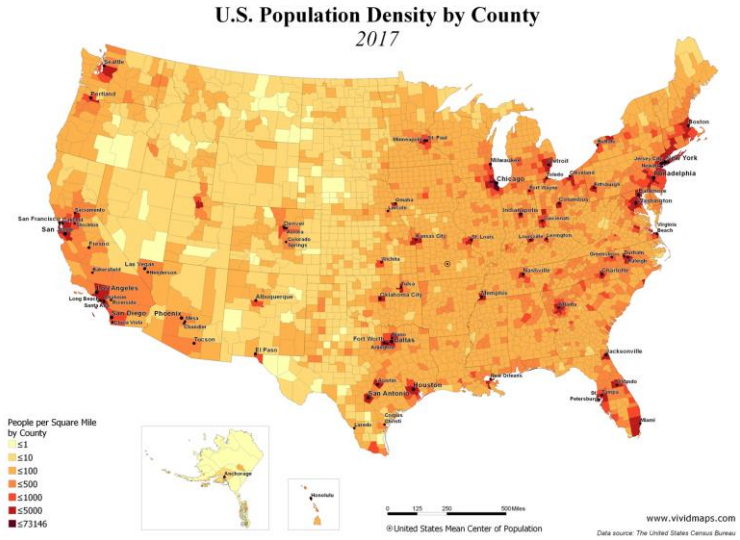
Pair-Plots

Bivariate distribution of datasets. Shows the pairwise relationship between variables



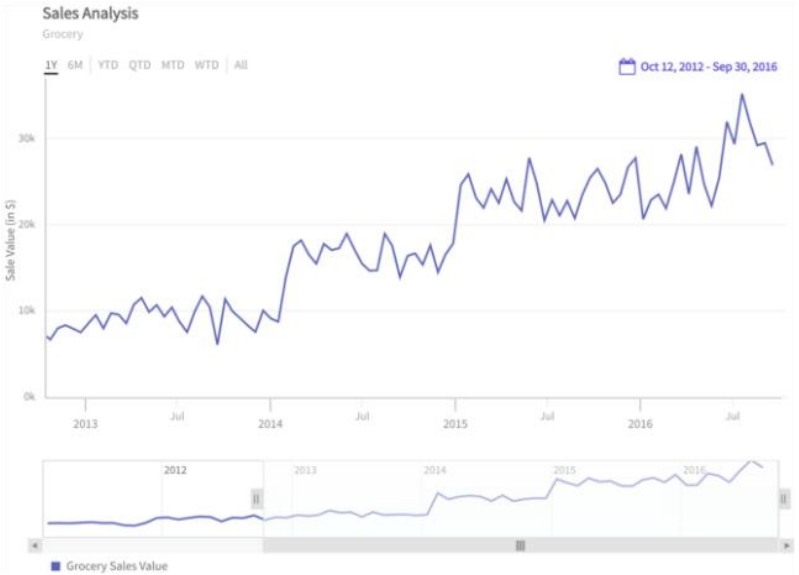
Heatmaps

Light/warm colours to indicate low- and high-value points. Humans interpret colour better than numbers



Fever Chart

Time-Series chart for change of data over a period of time



12

Exploratory Data Analysis

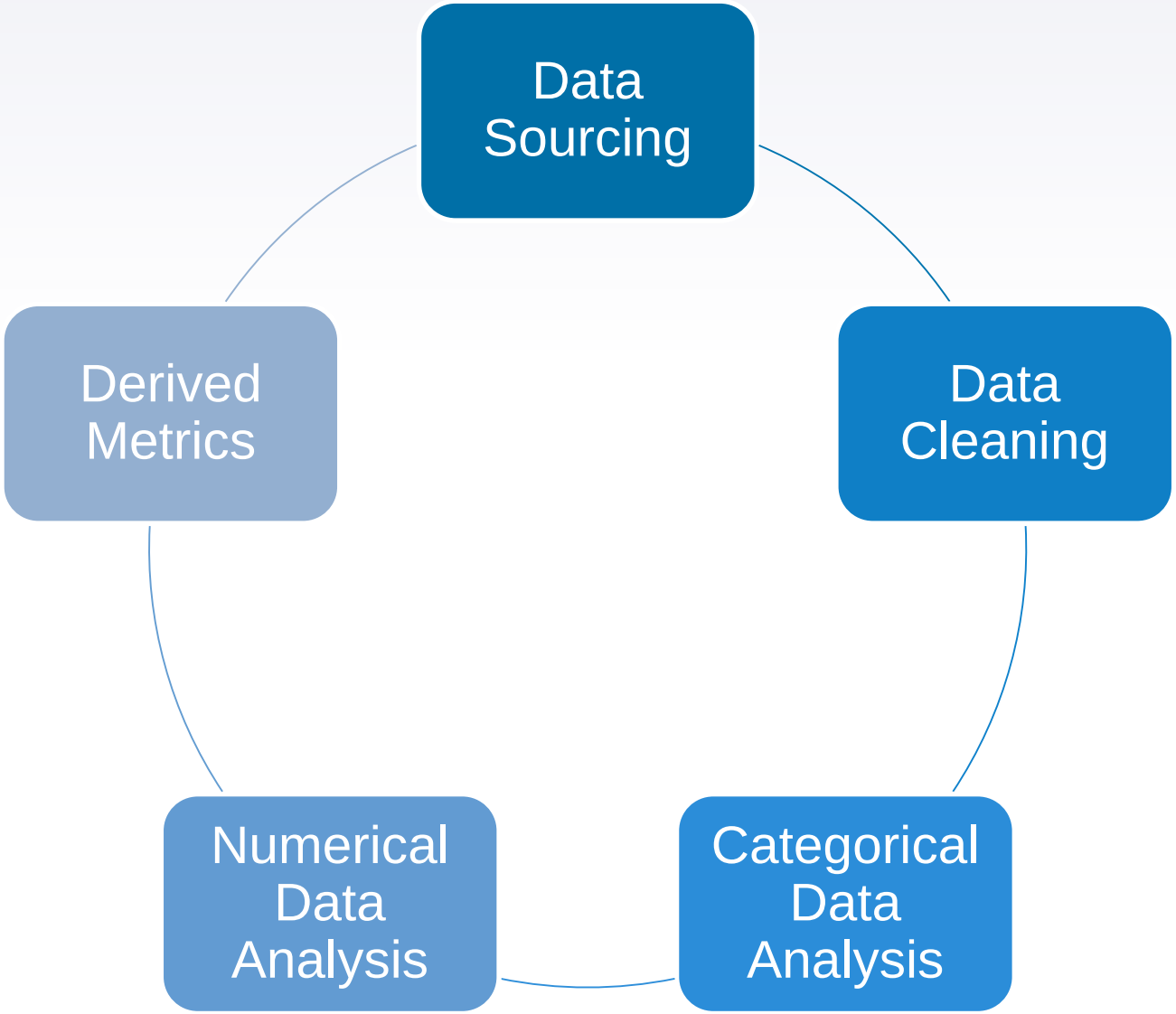
What is EDA?

- Deals with the process of performing initial investigations on data with the help of summary statistics and graphical representations
  - To discover patterns
  - Spot anomalies
  - Test hypotheses
  - Check assumptions

Outline of Performing EDA

1. What **question(s)** are you trying to solve (or prove wrong)?
2. What **kind of data** do you have and how do you treat different types?
3. What's **missing** from the data and how do you deal with it?
4. Where are the **outliers** and why should you care about them?
5. How can you **add, change or remove features** to get more out of your data?

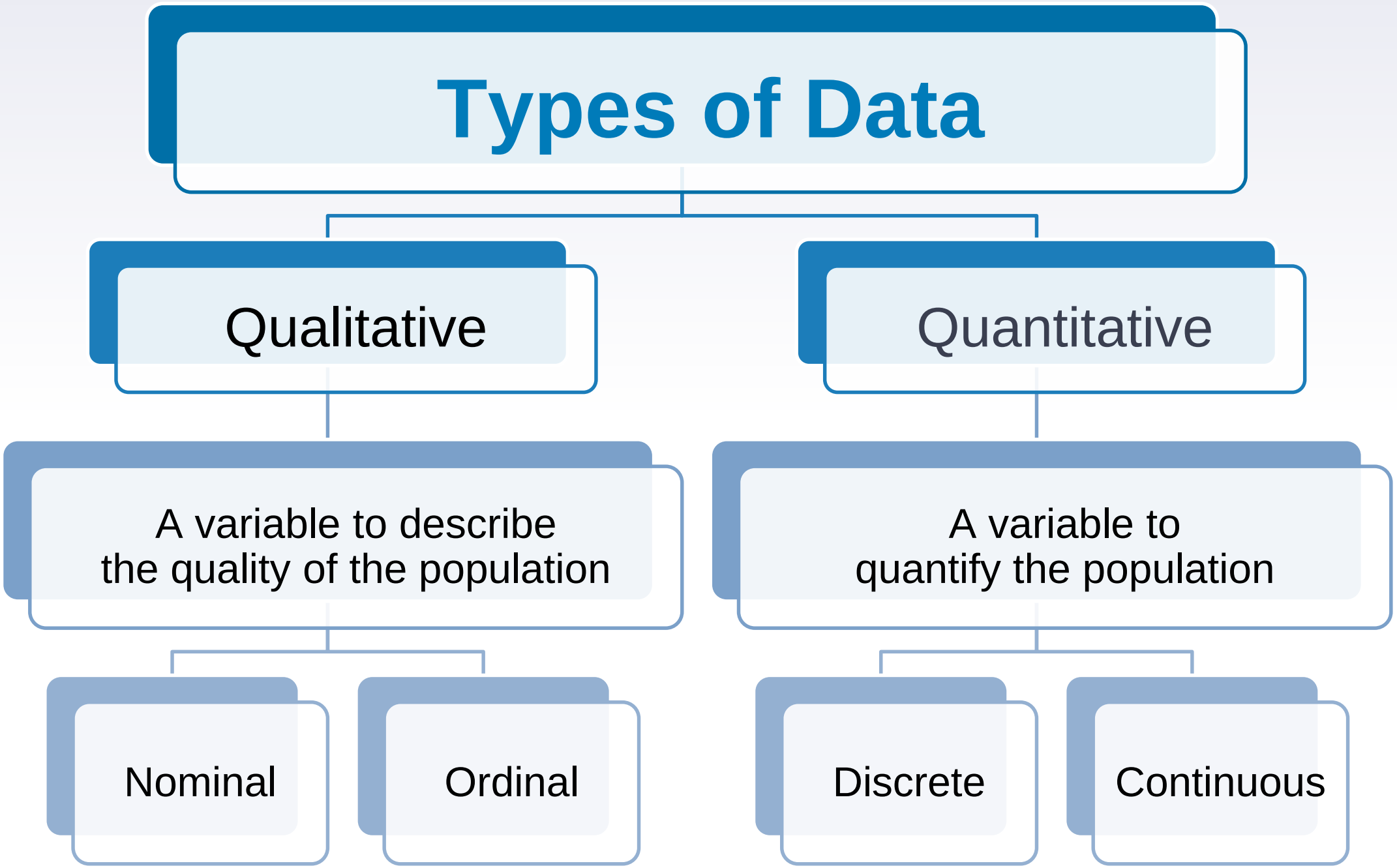
Steps Involved in EDA



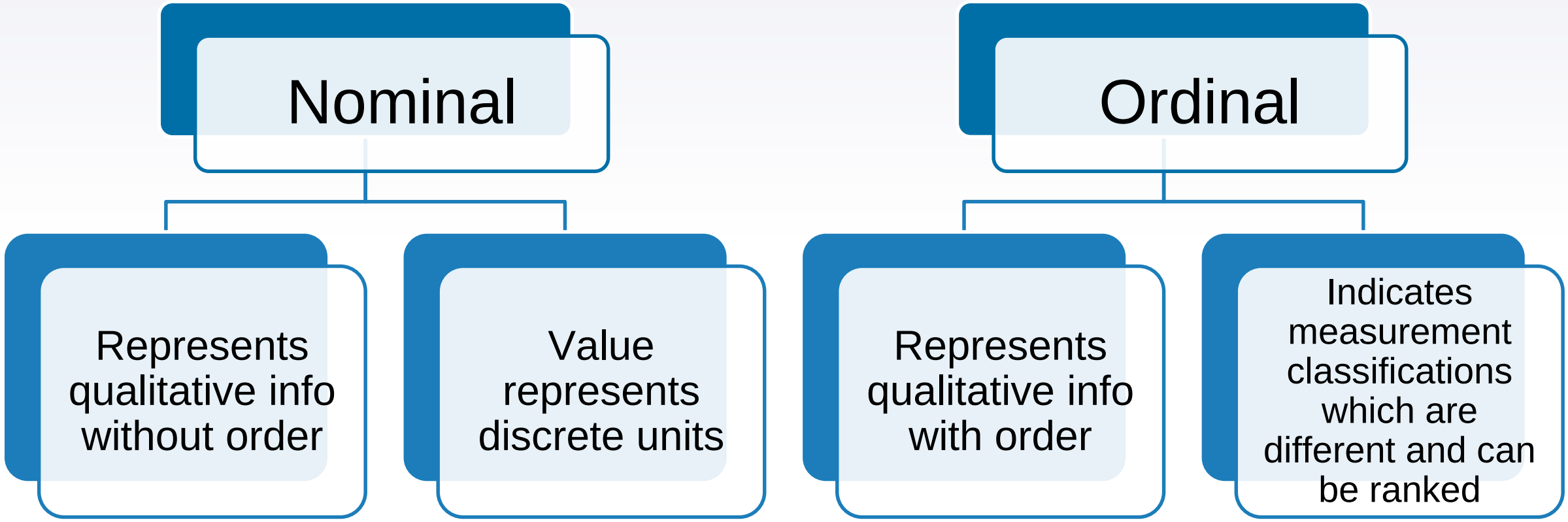
Data Cleaning: Handling Missing Values

1. **Delete rows/columns**
   - Rows: can be deleted if it has an insignificant no. of missing values
   - Columns: can be deleted if it >75%
     of missing values
2. **Replace with mean/median/mode**
   - Can be used on an independent variable when it has numerical variables
   - Categorical features: Apply mode method
3. **Algorithm Imputation**
   - Machine learning algorithms e.g. KNN, Naïve Bayes, Random Forest
4. **Predicting the missing values**
   - Training set: Data set with no missing values
   - Testing set: Data set with missing values
   - Target variable: Missing values

Types of Data



Qualitative



e.g. Gender: M/F

e.g. Economic Status: Low/Medium/High

Quantitative

Discrete

Only takes counted values, not decimal values

e.g. Number of students in a class

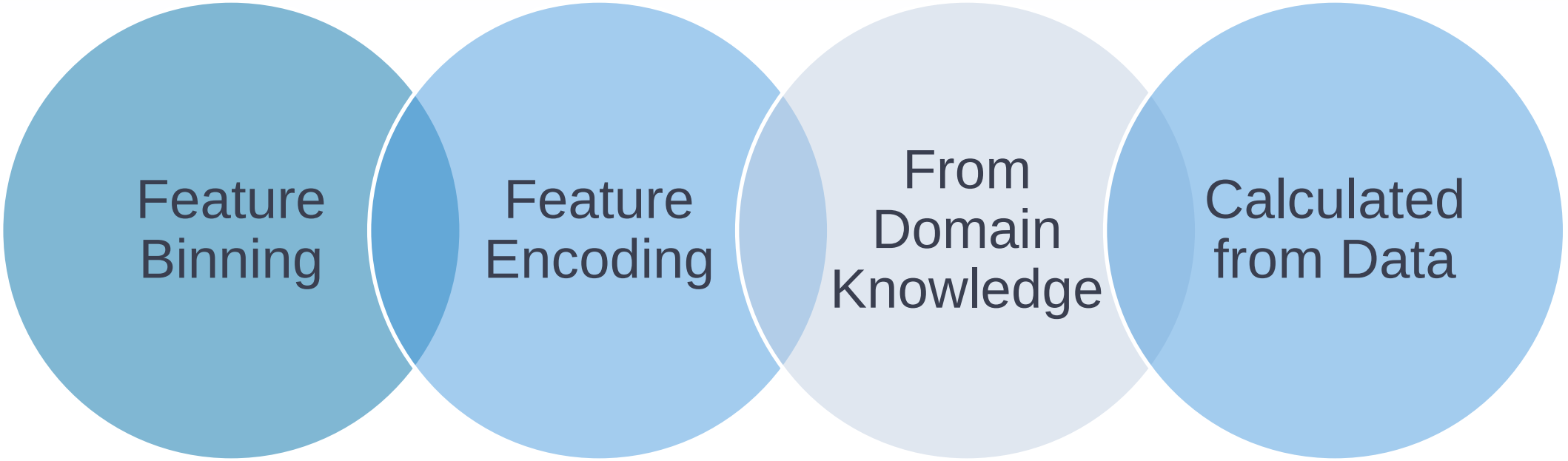
Continuous

Numbers within a range of values

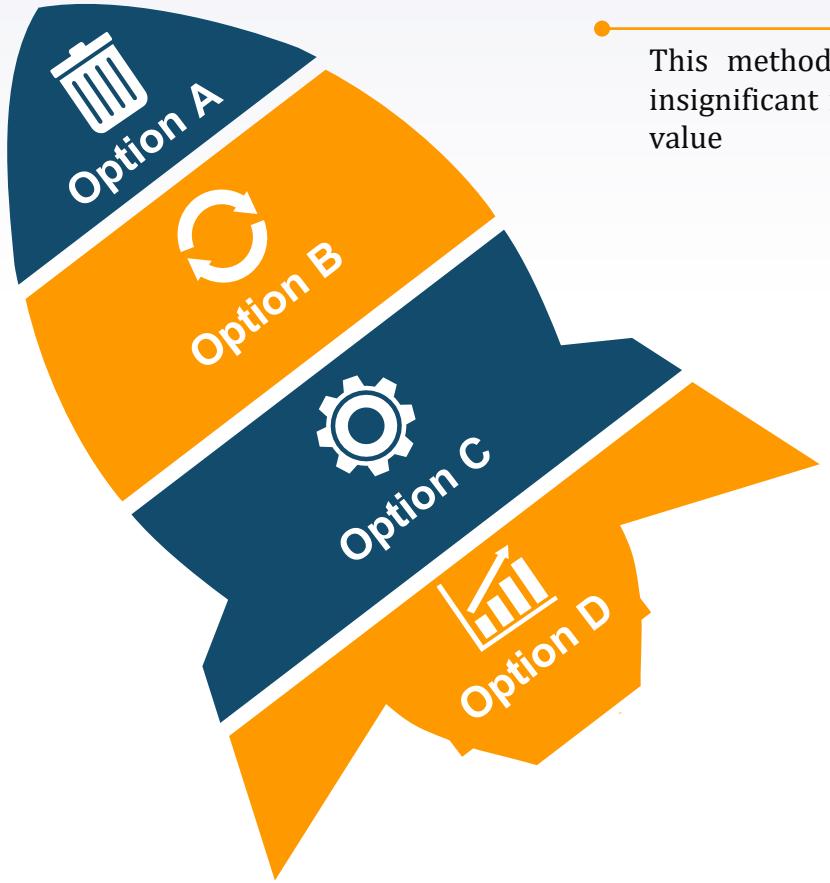
e.g. Height

Derived Metrics

Create a new variable from the existing variables to get insightful information from the data



Handle Missing Value



Delete Rows/Columns

This method we commonly used to handle missing values. Rows can be deleted if it has insignificant number of missing value Columns can be delete if it has more than 75% of missing value

Replacing with mean/median/mode

This method can be used on independent variable when it has numerical variables. On categorical feature we apply **mode** method to fill the missing value.

.

Algorithm Imputation

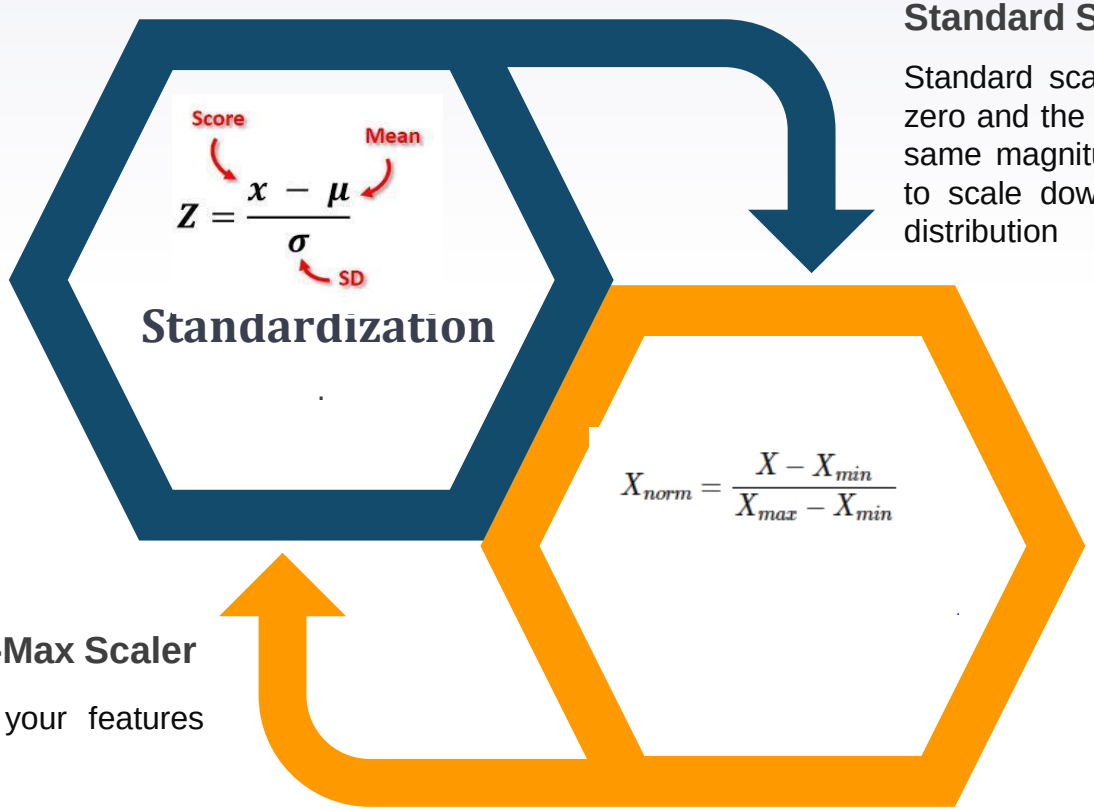
Some machine learning algorithm supports to handle missing value in the datasets. Like KNN, Naïve Bayes, Random forest.

.

Predicting the missing values

Prediction model is one of the advanced method to handle missing values. In this method dataset with no missing value become training set and dataset with missing value become the test set and the missing values is treated as target variable.

Feature Scaling Technique



Standard Scaler

Standard scaler ensures that for each feature, the mean is zero and the standard deviation is 1,bringing all feature to the same magnitude. In simple words Standardization helps you to scale down your feature based on the standard normal distribution

Min-Max Scaler

Normalization helps you to scale down your features between a range 0 to 1

Outlier Treatment

Outliers are the most extremes values in the data. It is a abnormal observations that deviate from the norm. Outliers do not fit in the normal behaviour of the data.

Detect Outliers using following methods:

1.Boxplot **2.** Histogram **3.** Scatter plot
4.Z-score
5. Interquartile range(values out of 1.5 time of IQR)

Handle Outlier using following methods:-

1.Remove the outliers.
2.Replace outlier with suitable values by using following methods:-

- Quantile method
- Interquartile range

3.Use that ML model which are not sensitive to outliers
Like:-KNN,Decision Tree,SVM,NaïveBayes,Ensemble methods

Handle Invalid Value

- **Encode Unicode properly:-**In case the data is being read as junk characters, try to change encoding, E.g. CP1252 instead of UTF-8.

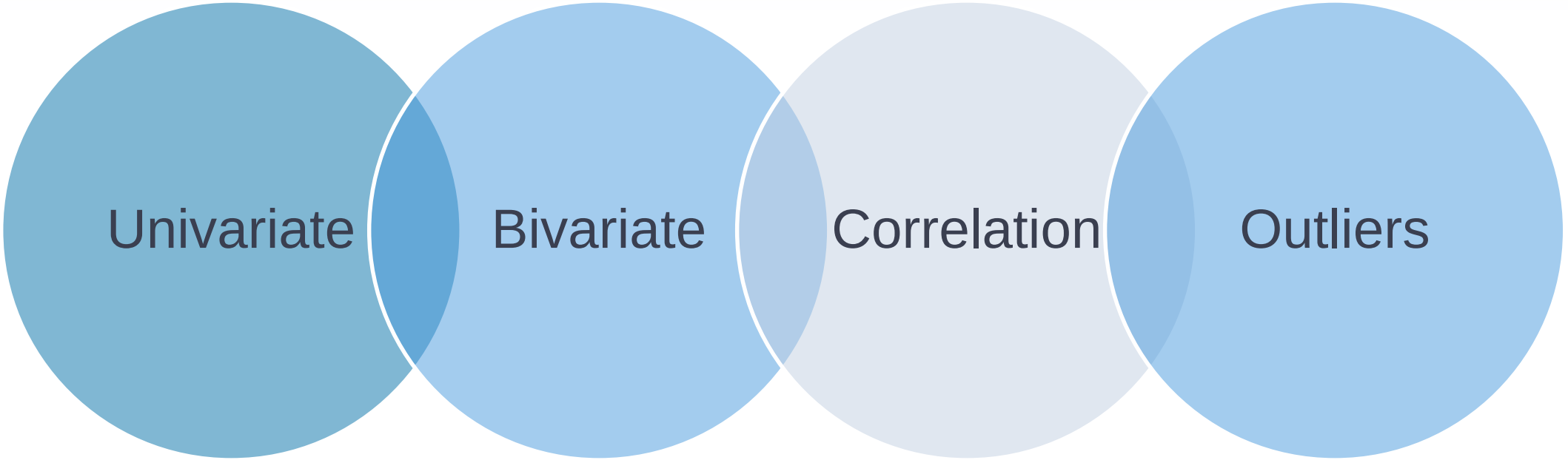
- **Convert incorrect data types:-** Correct the incorrect data types to the correct data types for ease of analysis. E.g. if numeric values are stored as strings, it would not be possible to calculate metrics such as mean, median, etc. Some of the common data type corrections are — string to number: "12,300" to “12300”; string to date: "2013-Aug" to “2013/08”; number to string: “PIN Code 110001” to "110001"; etc.

- **Correct values that go beyond range:-** If some of the values are beyond logical range, e.g. temperature less than -273° C (0° K), you would need to correct them as required. A close look would help you check if there is scope for correction, or if the value needs to be removed.

- **Correct wrong structure:-** Values that don’t follow a defined structure can be removed. E.g. In a data set containing pin codes of Indian cities, a pin code of 12 digits would be an invalid value and needs to be removed. Similarly, a phone number of 12 digits would be an invalid value

Analysis

EDA is evolving around these 4 concepts



Uses Cases

Basically EDA is important in every business problem, it's a first crucial step in data analysis process.
Some of the use cases where we use EDA is:-

- **Cancer Data Analysis :-** In this data set we have to predict who are suffering from cancer and who's not.

- **Fraud Data Analysis in E-commerce Transactions :-** In this dataset we have to detect the fraud in a E-commerce transaction.

13

Project: Subscriber Churn

Subscriber Churn can be in different forms and not just exit from the base

Tariff Plan Churn (e.g. €50 to €30 monthly)

Service Churn (e.g. Weekly/ Monthly Subscription)

Different Churn Scenarios

Leads to Subscriber churn

Subscriber Churn (e.g. Port Out to Competition)

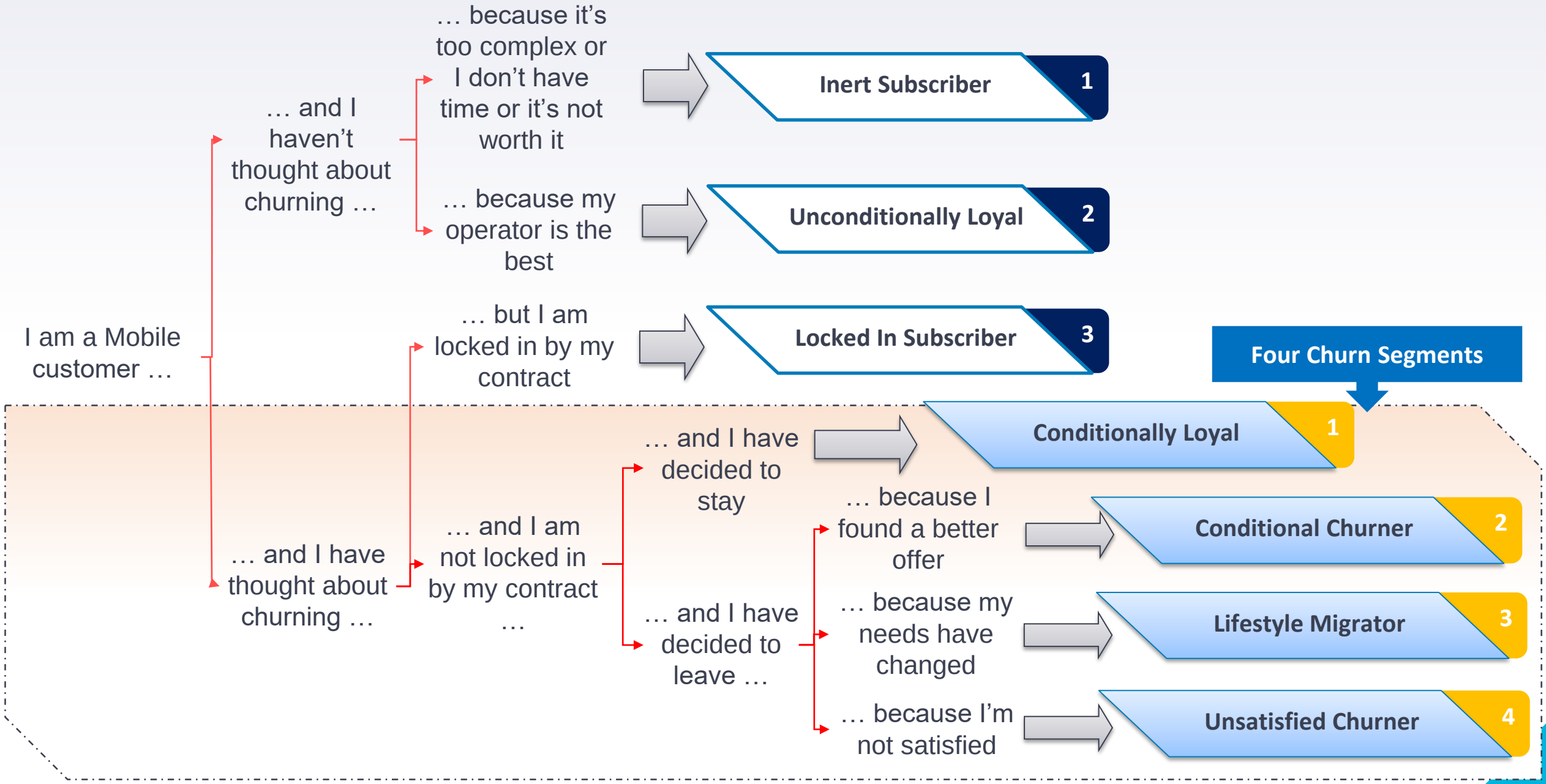
Provider A

Provider B

Product Churn (e.g. Postpaid to Prepaid)

Usage Churn (e.g. Inactive or Zero Usage)

Decision Cycle of a Subscriber: Changes as per needs and/or experiences



Four Churn Segments: Loyalty drivers for each segment

Conditionally Loyal 1

Conditional Churner 2

- Frequently re-evaluates purchase decisions
- Choose brand on (semi) rational basis
- Likes to change to test new products

Lifestyle Migrator 3

- Uncontrollable change in needs / usage behavior
- Involuntary Churn

Unsatisfied Churner 4

- Unsatisfied subscriber

Loyalty Drivers

Key drivers that Influence Churn

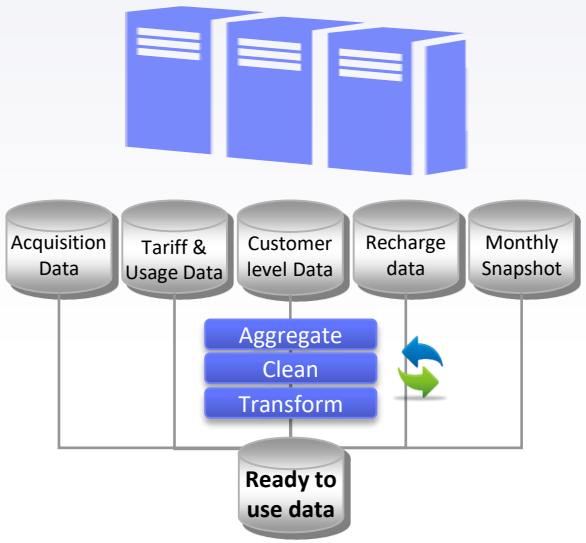
1. Handset Loss/Upgrade
2. Cost of Service / Competitor pricing
3. Network Quality
4. Customer Care Quality

Key drivers for Subscriber loyalty

1. Offers and services
2. Price
3. Quality of products and services
4. Quality of customer service
5. Length of contract period
6. Perception of telecom brand
7. Marketing programs and campaigns

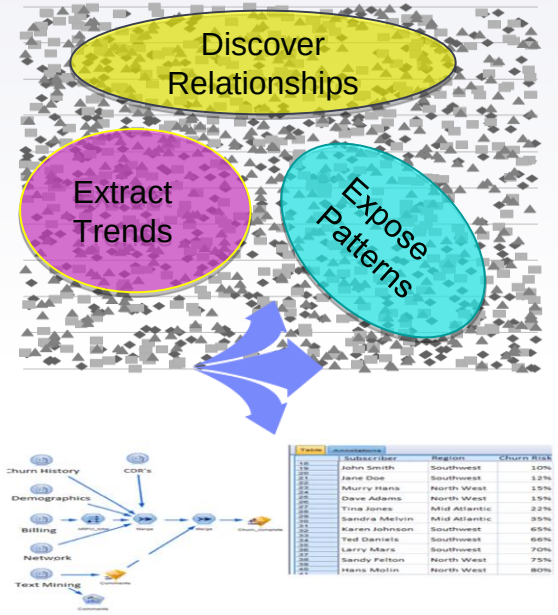
High level Overview of a Data Science led approach to Manage Churn

Capture & Analyze



- Business Understanding
- Identify data requirements and explore data availability
- Request and extract data required to build a model
- Aggregate, Clean and Standardize data in desired format for model

Report & Predict



- Business Analysis of standardized data
- Predictive model design
- Development and Implementation of Predictive model

Engage & Act

- List of churn drivers / KPIs for tracking and monitoring
- A generated list of recommended subscribers for targeted churn campaigns
- Recommendations on monthly churn initiatives

Thank you!

Any questions?

Resources

- Power BI Documentation
  - https://docs.microsoft.com/en-us/power-bi/
- Power BI Guided Learning
  - https://docs.microsoft.com/en-us/power-bi/guided-learning/
  - https://www.youtube.com/playlist?list=PL1N57mwBHtN0JFoKSR0n-tBkUJHeMP2cP
- Power BI White Paper
  - https://docs.microsoft.com/en-us/power-bi/guidance/whitepapers
- Power BI Blogs
  - https://powerbi.microsoft.com/en-us/blog/