Actionable Analytics and Dashboards

25 June 2022

Today's Session

I. Uses of Analytics

II. Business Improvement Opportunities

III. Actionable Analytics
    1. Regression Analysis
    2. Logistic Regression
    3. Clustering
    4. Event Analysis
    5. Network Analysis

IV. Dashboard Principles

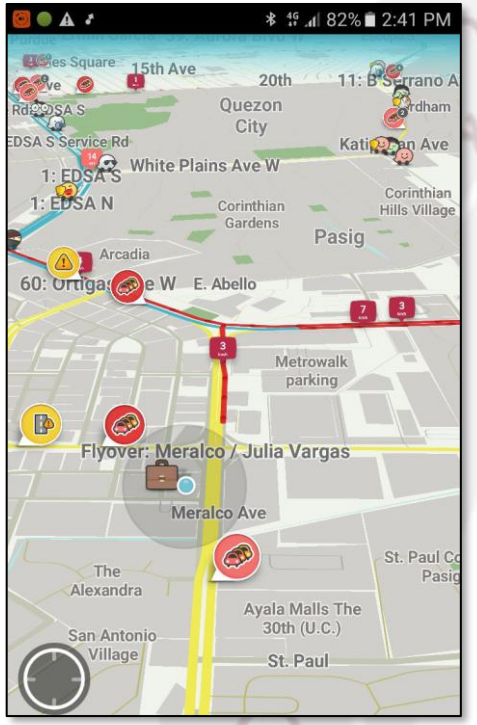
3 uses for Analytics

Data-driven Benchmarking (Descriptive)

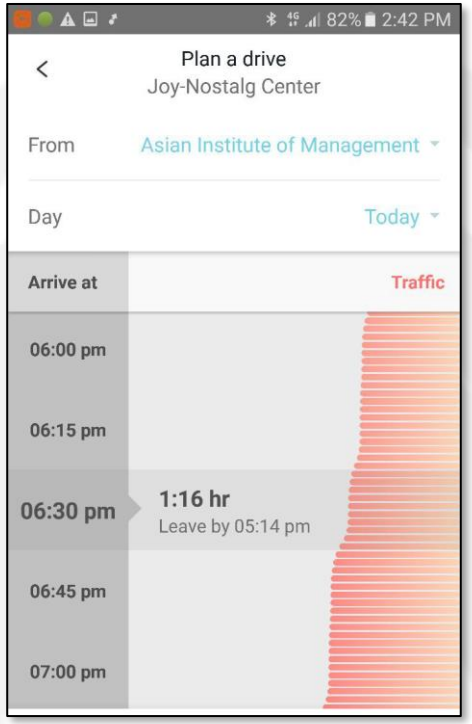
Data-Driven Prediction (Predictive)

Data-driven Recommendation (Prescriptive)

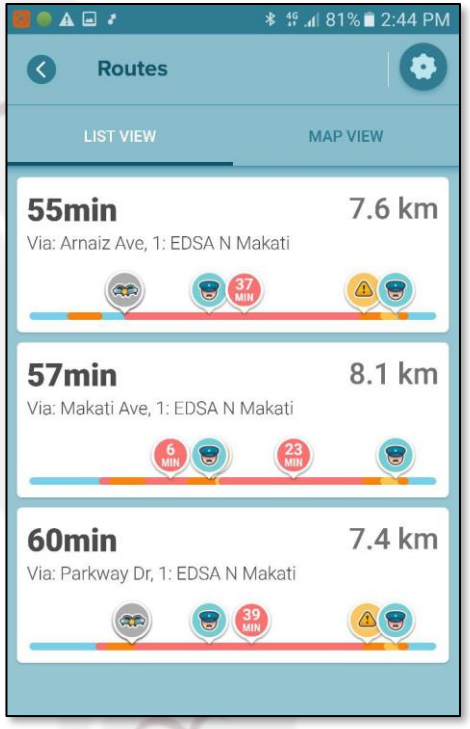
Waze App



How is the traffic? (Descriptive)

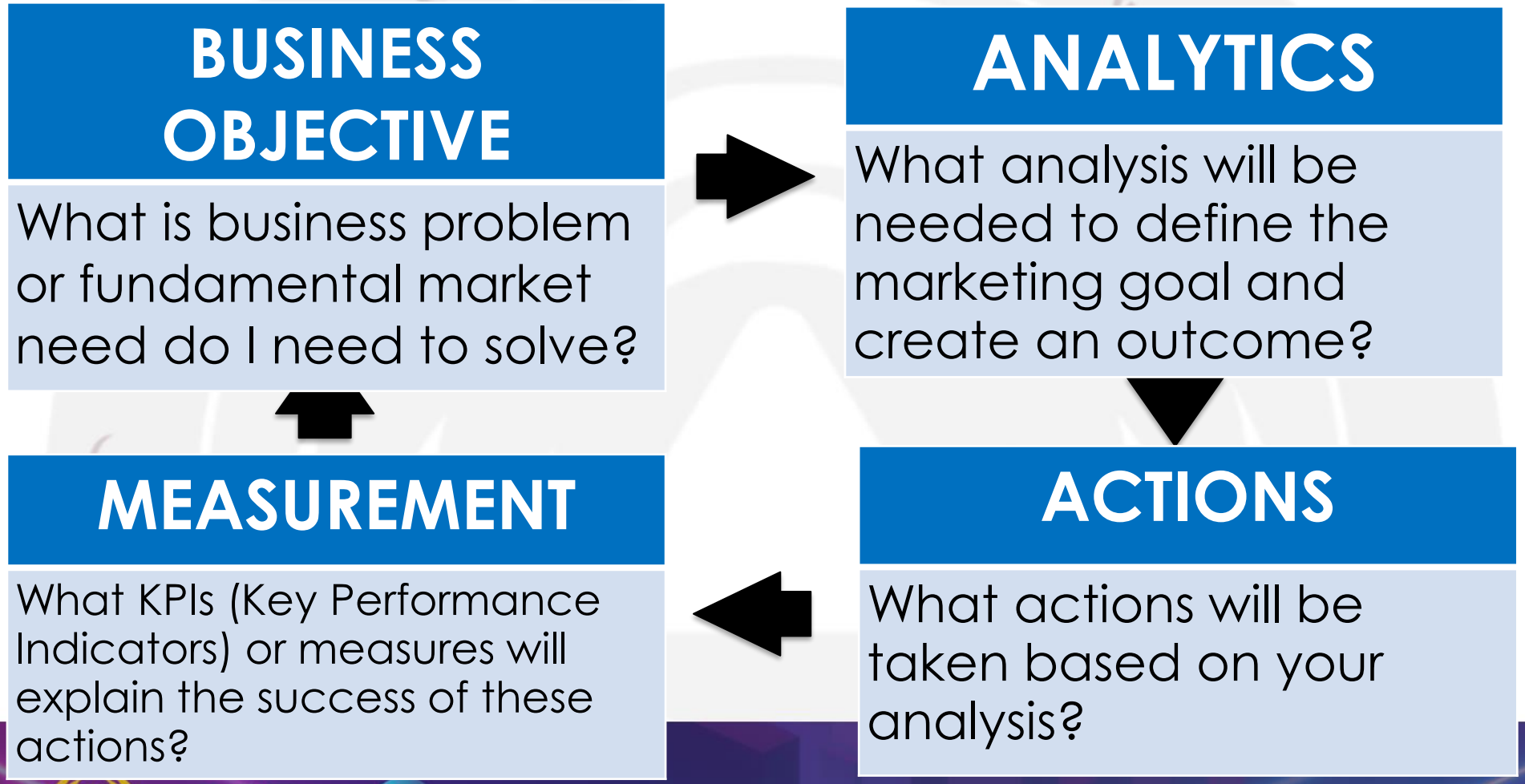


How long will my travel be? (Predictive)

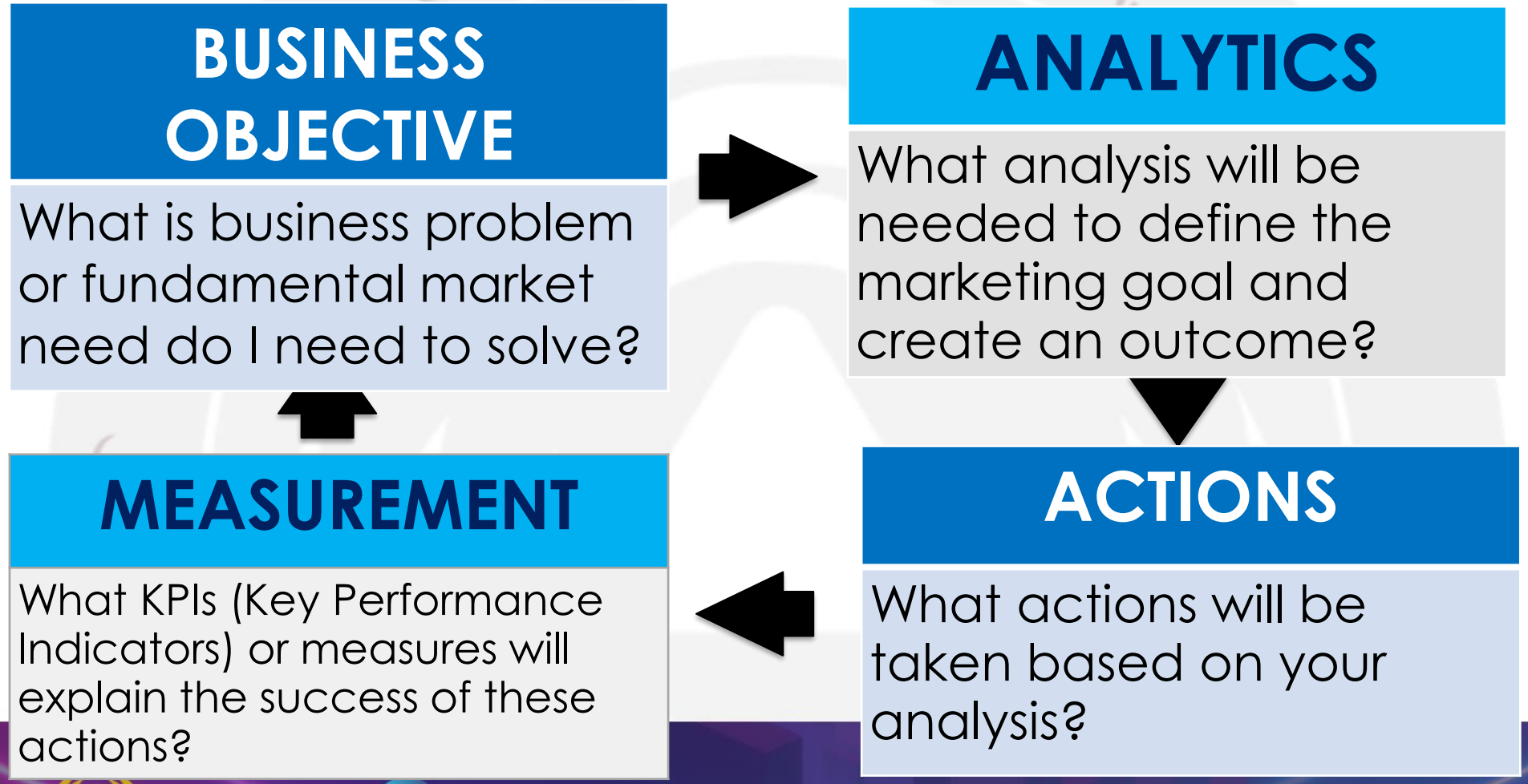


Which is the best route to my destination? (Prescriptive)

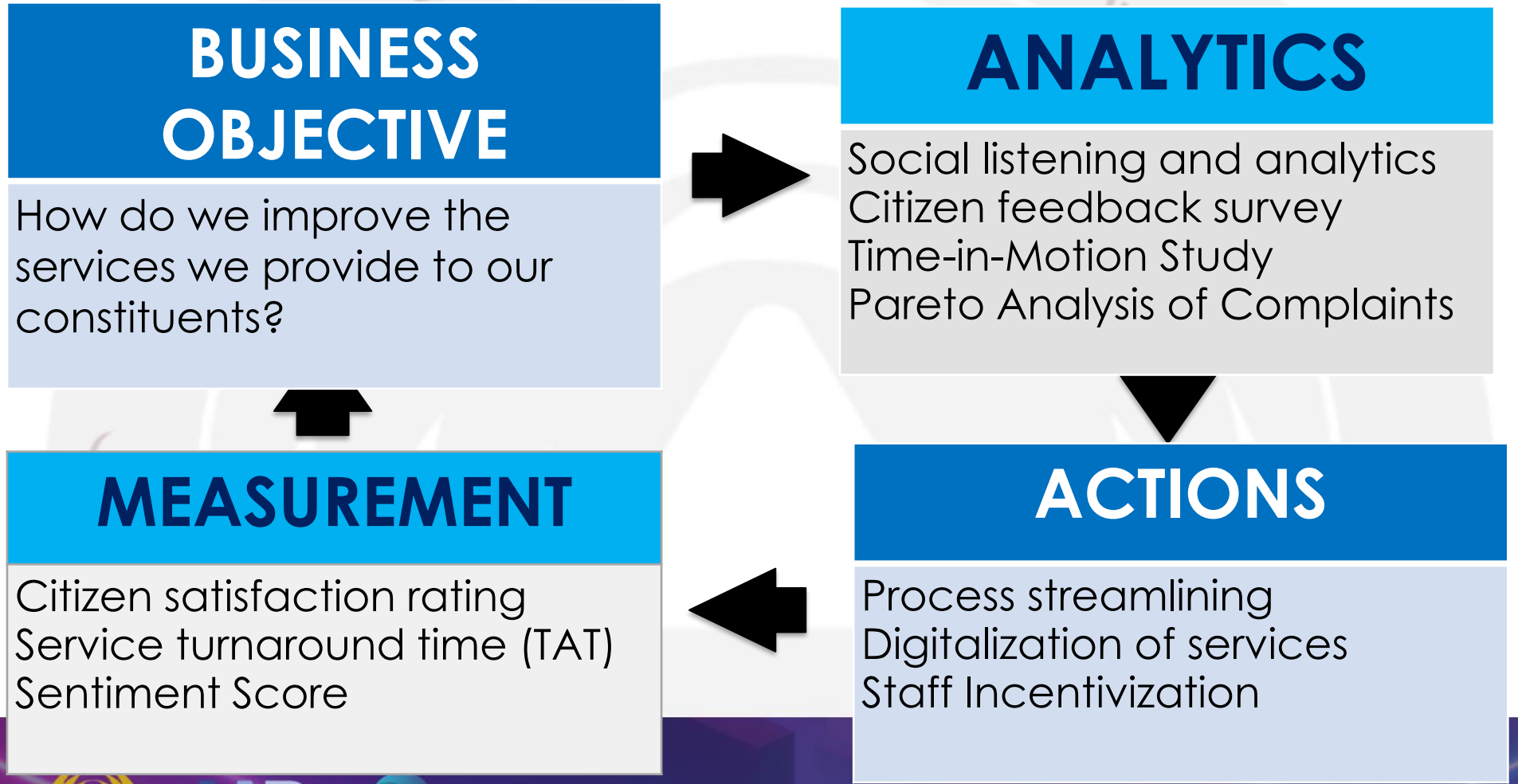
Data and Analytics Scenarios



Data and Analytics Scenarios



Data and Analytics Scenarios



The essence of analytics:

What question are you asking?

Keys To Actionable Analytics

What is it?

What data do you need?

How do you visualize it?

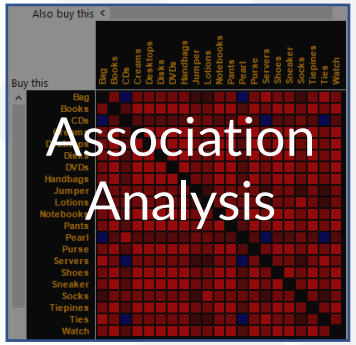
What questions does it answer?

Actionable Analytics
Asking Questions of your Data

Measuring magnitude and trend.

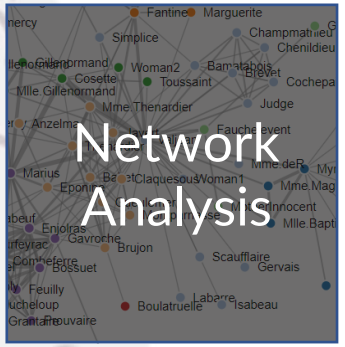
Measuring probability or likelihood.

Determining similarities and groupings.



Determining the items, events, or individuals that go together.

Understanding sequences of complex events.

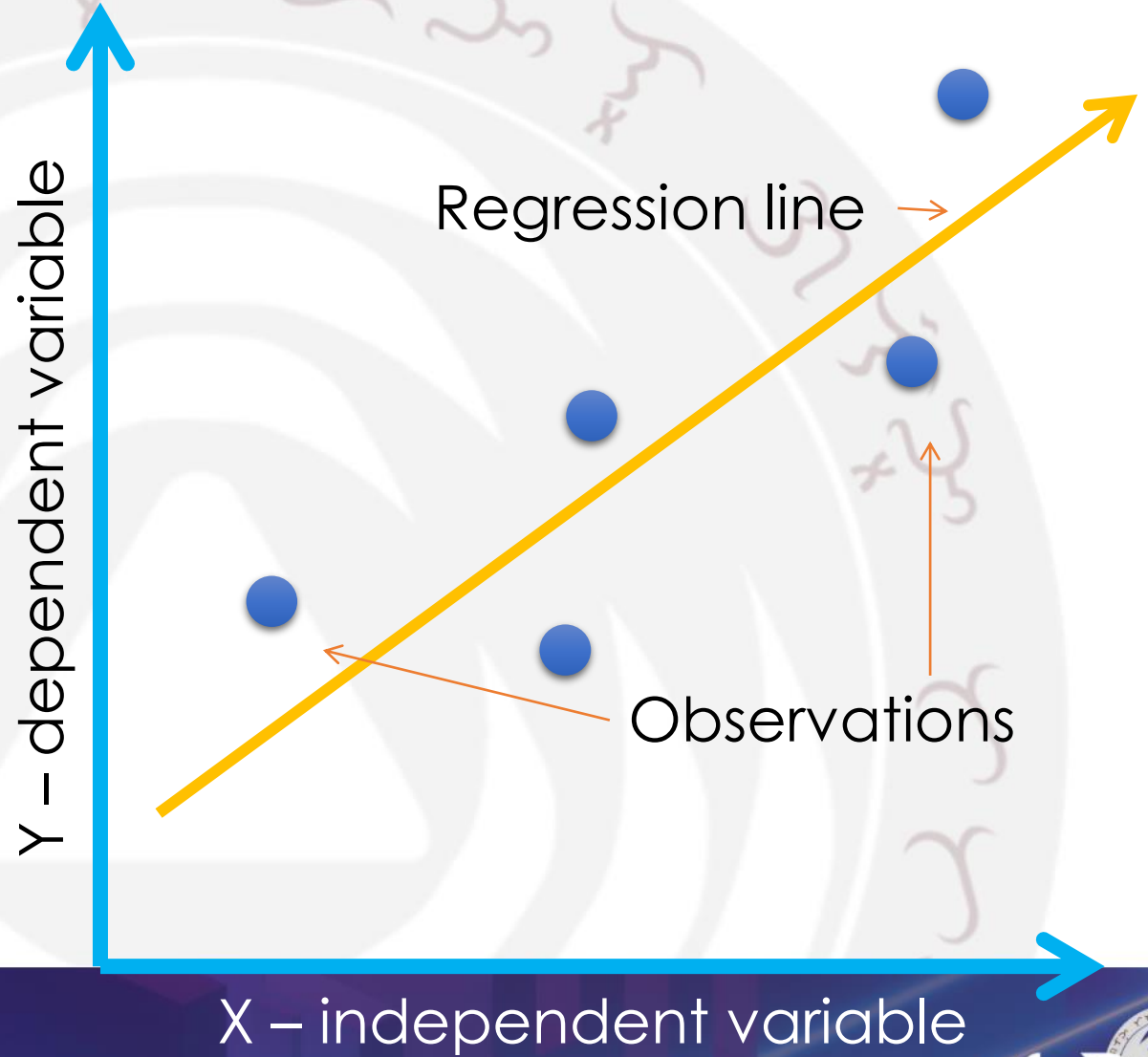


Determining the connections between entities.

https://github.com/fstayco/sparta-butuan-incubation-workshop

Linear Regression

$$y = \beta x + \alpha$$



Simple Linear Regression Model

Dependent Variable

Sales

Beta coefficient

Sales Sensitivity

$$y = \beta x + \alpha + \varepsilon$$

Independent Variable

Ad Spending

Alpha intercept

Baseline Sales

Error term

What is the **effect** of my **ad spending** to my **sales**?

Multiple Linear Regression

$$y = \beta x + \alpha + \varepsilon$$

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \alpha + \varepsilon$$

You can have more than one predictor variable (x1 – xn).

Linear Regression

Use-case: Understand the factors that affect the length of defendant's sentence from the International Crime Tribunal.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.778601141
R Square	0.606219737
Adjusted R Square	0.590593536
Standard Error	127.8313226
Observations	132

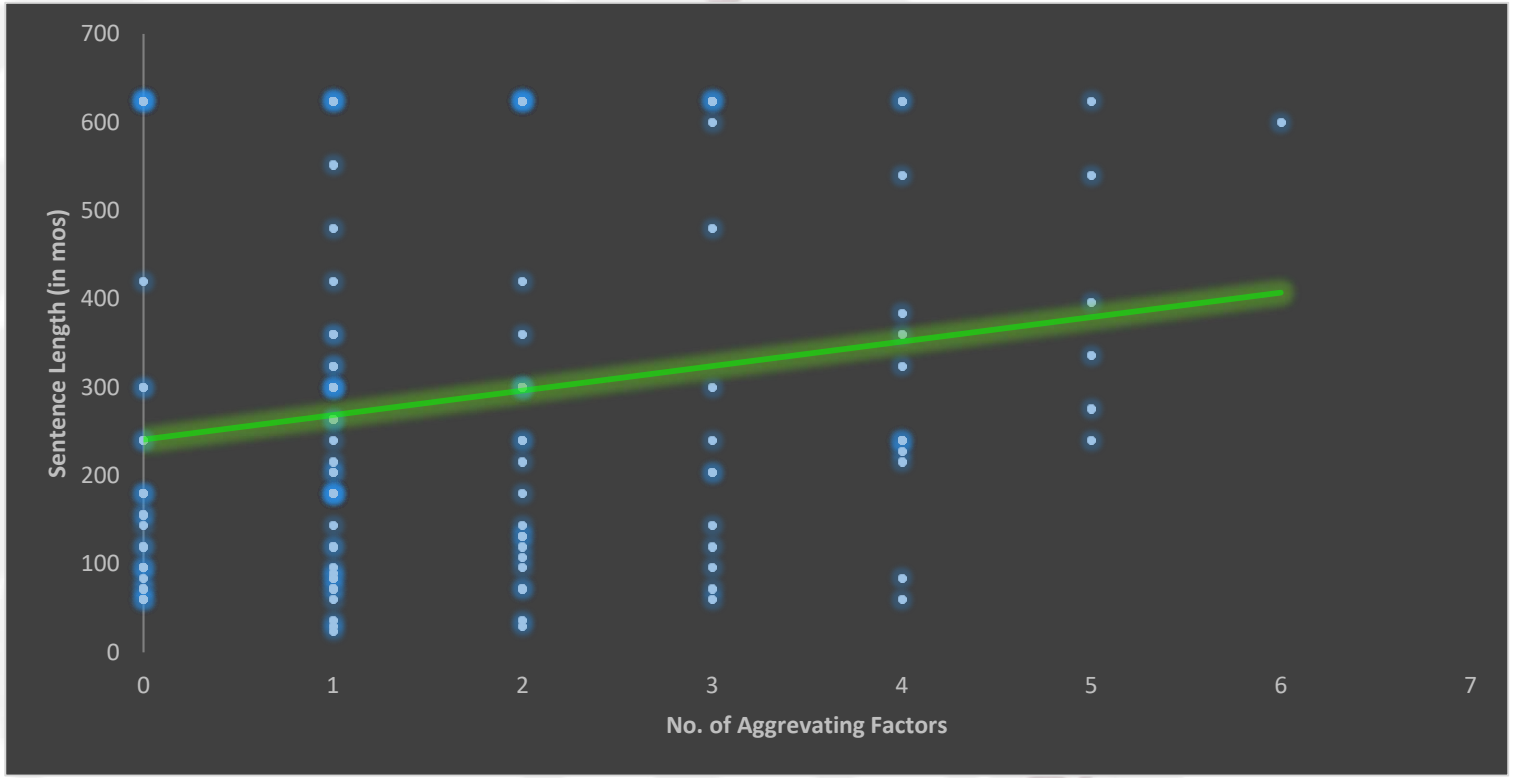
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	3169722.456	633944.4913	38.79508144	5.88441E-24
Residual	126	2058946.726	16340.84703		
Total	131	5228669.182			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	63.33614341	37.61999357	1.683576668	0.094739509	-11.1127179	137.7850047	-11.1127179	137.7850047
numGuil	14.23268286	2.573693834	5.530060596	1.76686E-07	9.139418423	19.3259473	9.139418423	19.3259473
mfTotal	-19.04232356	5.056890237	-3.765619317	0.000253715	-29.04976068	-9.034886444	-29.04976068	-9.034886444
afTotal	18.09961427	7.873504904	2.298800153	0.023163008	2.518179535	33.68104901	2.518179535	33.68104901
genocide	214.7432536	25.33288965	8.476855841	5.22716E-14	164.6102097	264.8762975	164.6102097	264.8762975
crimAg	112.2060653	30.62807359	3.663503843	0.000364992	51.59400807	172.8181225	51.59400807	172.8181225

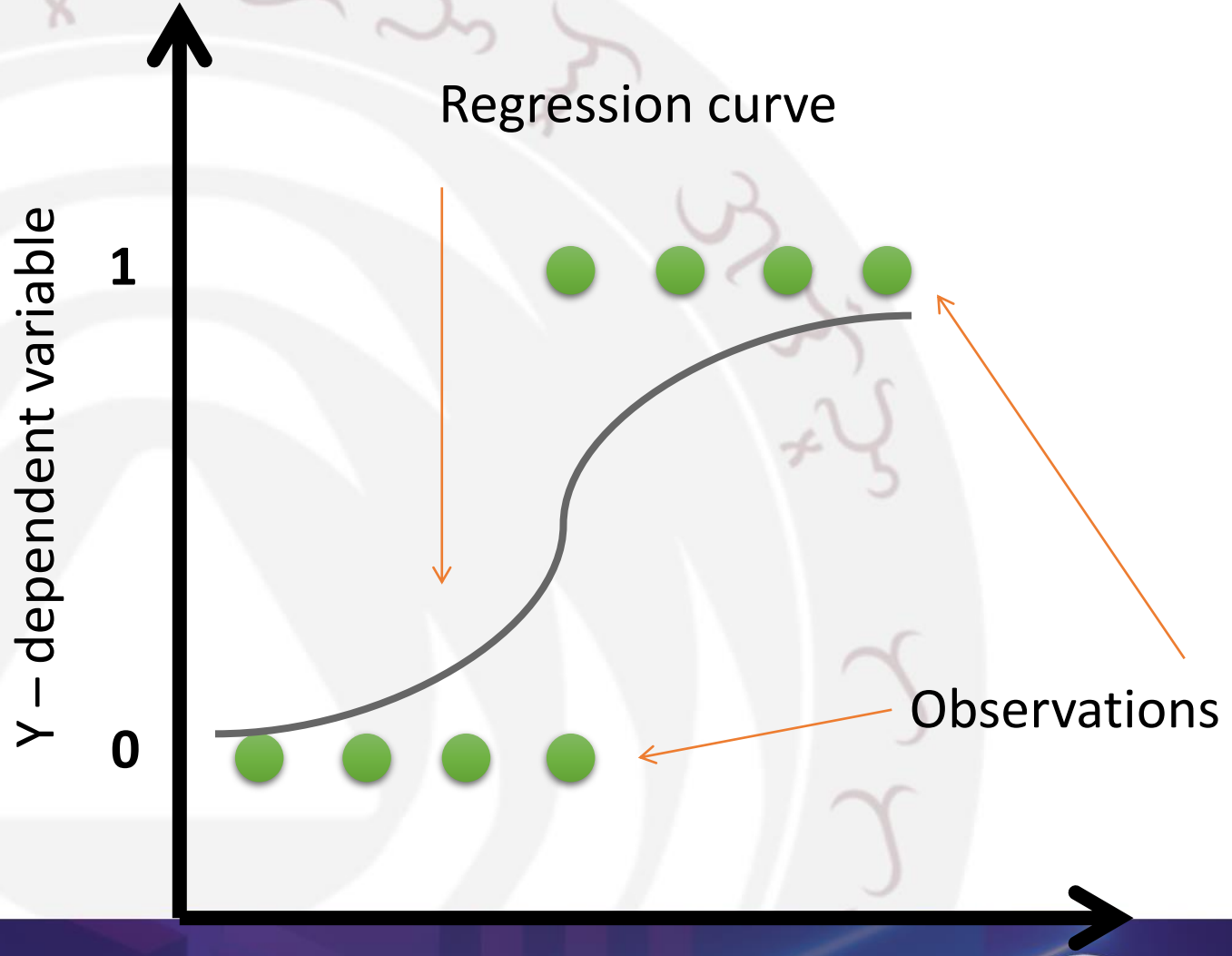
Linear Regression

Use-case: Understand the factors that affect the length of defendant's sentence from the International Crime Tribunal.



Logistic Regression

$$y = \frac{1}{1 + e^{-(\beta x + \alpha)}}$$



Multiple Logistic Regression

$$p = \frac{1}{1 + e^{-(\beta x + \alpha)}}$$

$$p = \frac{1}{1 + e^{-(\beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \alpha)}}$$

You can have more than one predictor variable (x1 – xn).

Logistic Regression

- **Use-case:** Credit scoring model to predict likelihood of bankruptcy.

Coefficients	Variables		Estimates	Odds Ratio	Interpretation
b0	Intercept	-	2.8850		
b1	STD_SALES	-	0.3367	0.7141	The odds of bankruptcy decreases by 29% for every one standard deviation increase in sales.
b2	STD_ROCE	-	1.3359	0.2629	The odds of bankruptcy decreases by 74% for every one standard deviation increase in profit before tax to capital employed.
b3	STD_FFTL	-	1.5883	0.2043	The odds of bankruptcy decreases by 80% for every one standard deviation increase in funds flow (earnings before interest, tax & depreciation) to total liabilities.
b4	STD_GEAR		2.7299	15.3311	The odds of bankruptcy increases by 1,433% for every one standard deviation increase in (current liabilities + long-term debt) to total assets.
b5	STD_CLTA	-	2.3730	0.0932	The odds of bankruptcy decreases by 91% for every one standard deviation increase in current liabilities to total assets.
b6	STD_CACL	-	2.6506	0.0706	The odds of bankruptcy decreases by 93% for every one standard deviation increase in current assets to current liabilities.
b7	STD_QACL	-	4.2539	0.0142	The odds of bankruptcy decreases by 99% for every one standard deviation increase in (current assets - stock) to current liabilites.
b8	STD_WCTA		3.1433	23.1799	The odds of bankruptcy increases by 2,318% for every one standard deviation increase in (current assets - current liabilities) to total assets.
b9	STD_AGE	-	0.3420	0.7104	The odds of bankruptcy decreases by 29% for every one standard deviation increase in number of years the company has been operating since incorporation date.
b10	CHAUD		2.5977	13.4335	The odds of bankruptcy increases by 1,243% if the company changed auditor in the previous 3 years.

Logistic Regression

- **Use-case:** Credit scoring model to predict likelihood of bankruptcy.

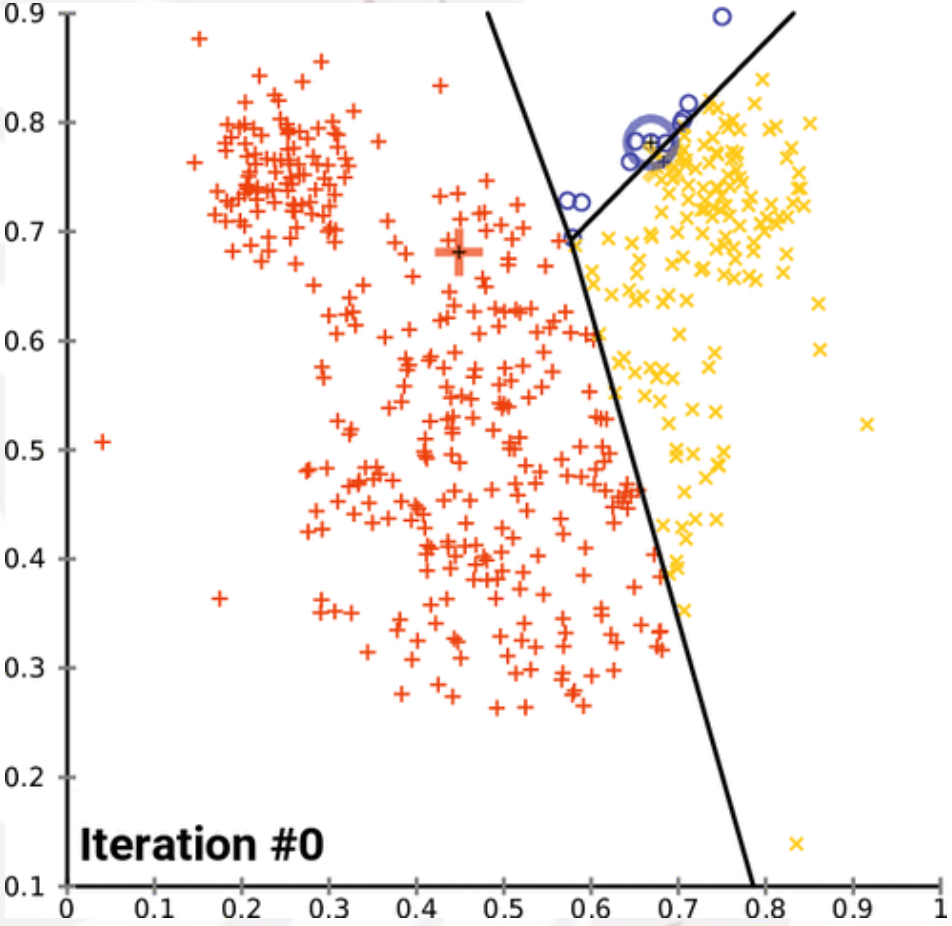
Optimal Cut-off:
p = 45.9%
Approval Rate = 50%
Bankruptcy Rate = 5.8%

Alternative Cut-off:
p = 82.5%
Approval Rate = 71%
Bankruptcy Rate = 25%

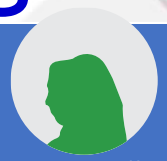
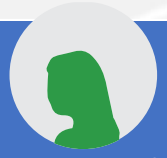
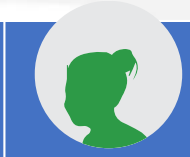
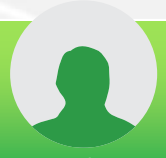
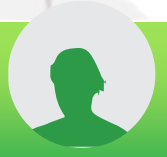
Count of Firm	% to Total		Bankruptcy Rate		Cum % by Decile		
Decile	Marginal	Cumulative	Marginal	Cumulative	Non-Bankrupt	Bankrupt	Spread
0	0%	0%	0.00%	0.00%	0%	0%	0%
1	12%	12%	0.00%	0.00%	21%	0%	21%
2	9%	21%	0.00%	0.00%	37%	0%	37%
3	9%	29%	0.00%	0.00%	53%	0%	53%
4	12%	41%	0.00%	0.00%	74%	0%	74%
5	9%	50%	33.33%	5.88%	84%	7%	78%
6	9%	59%	66.67%	15.00%	89%	20%	69%
7	12%	71%	75.00%	25.00%	95%	40%	55%
8	9%	79%	66.67%	29.63%	100%	53%	47%
9	9%	88%	100.00%	36.67%	100%	73%	27%
10	12%	100%	100.00%	44.12%	100%	100%	0%
Grand Total			44.12%				

Clustering

- Determine the likely clusters or **groupings within a population**. Also helps determine membership in a group.

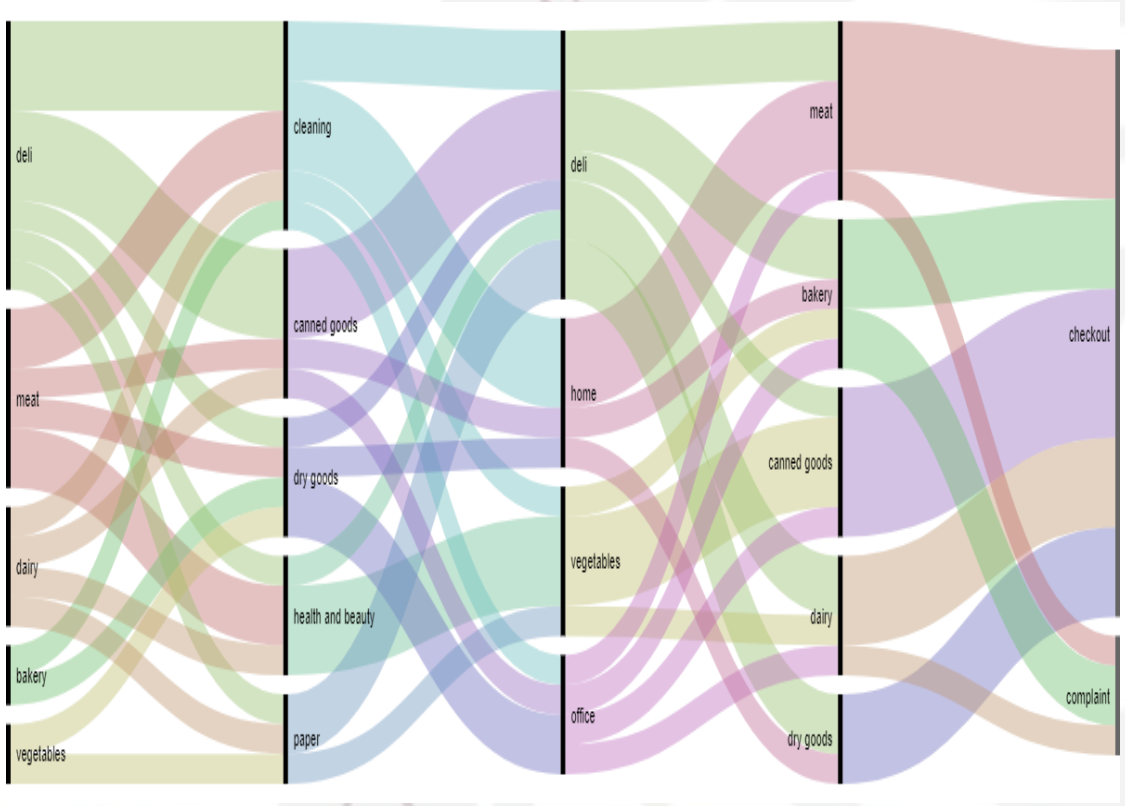


Clustering

	 Easy Sell	 Concerned	 Supporter	 Pragmatist	 Complacent	 Obligated	 Skeptic
Percentage of Population	27%	22%	16%	5%	11%	28%	9%
Hesitancy	1%	2%	6%	9%	17%	25%	82%
Demographics	25-54yo, Married, Higher income, Highly Educated	25-54yo, Married, Mid income	25-44yo, Married, Male, Mid income	18-44yo, Female, Low income, Vocational	25-44yo, Married, Male, Mid income	25-44yo, Married, Highly educated	25-44yo, Married, Female, Mid income
Geographic	NCR+	VisMin	Provincial	NCR+	VisMin	LuzMin	VisMin
Reasons for getting vaccinated	• Risk of COVID • Protect one's self and family	• Risk of COVID • Protect one's self and family • Trusts experts	• Risk of COVID • Protect one's self and family • Vaccination process	• Risk of COVID • Protect one's self and family • Less concerned about side effects	• Protect one's self and family • Less concerned about side effects	• Risk of COVID • Protect one's self and family • **Work-related**	
Reasons for avoiding vaccination	• Lack of supply • Vaccination process	• Side effects • Comorbidities • Vaccination process	• Side effects • Vaccine access	• Lack of supply • Vaccination process	• **Does not believe COVID is a risk** • Vaccination process	• Side effects • **Lack of trust in experts** • Vaccination process	• Side effects • **Negative news** • **Does not believe COVID is a risk** • **Lack of trust in experts**
Information and Media habits	Social Media	TV Social Media	DOH media Social media	DOH media	DOH media	TV	TV

Reference: https://www.undp.org/sites/g/files/zskgke326/files/migration/ph/FINAL---Vaccine-Acceptance-Report_Aug30_signed.pdf

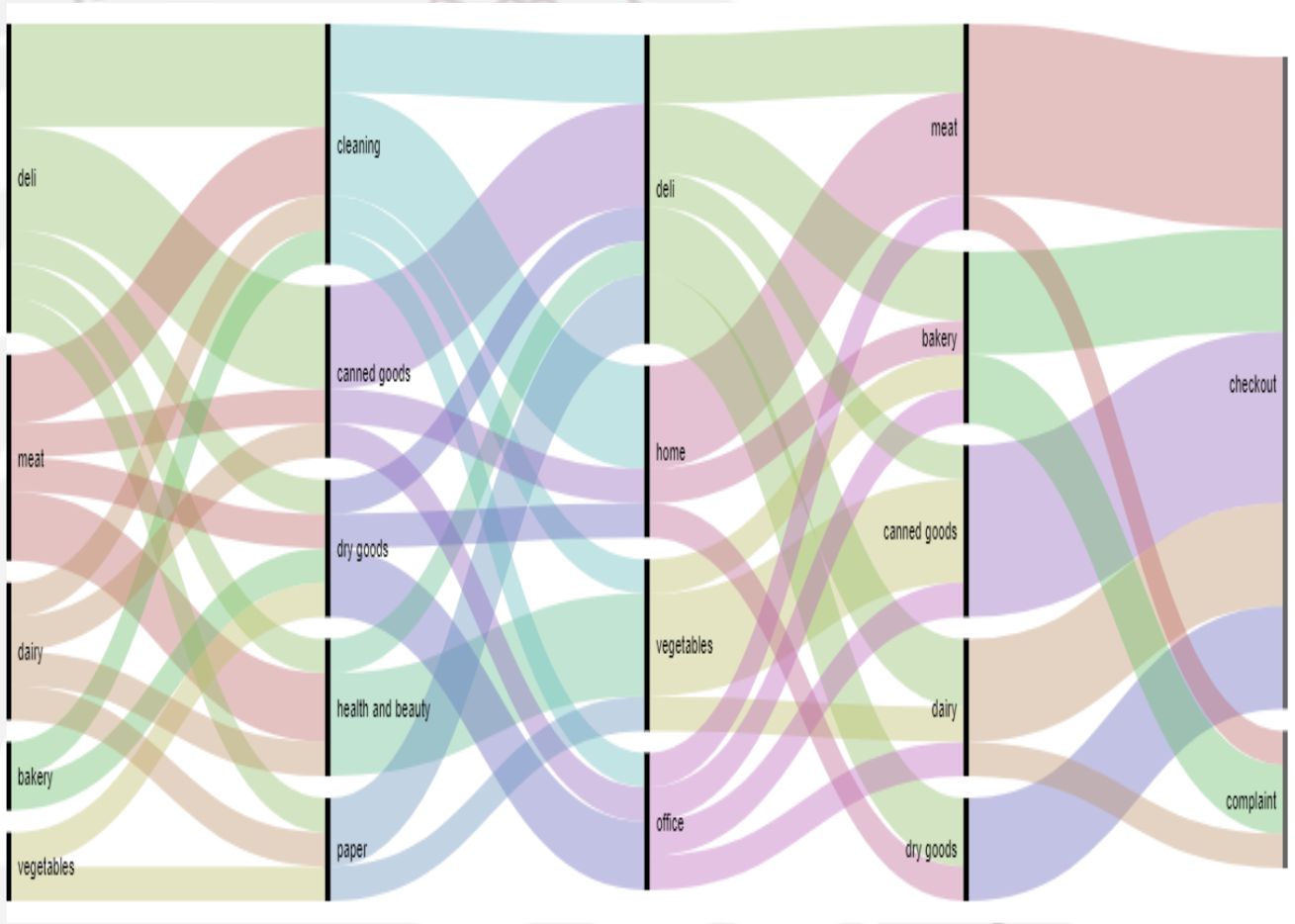
Event Analysis



- Used to understand sequences of events – what comes before what.
- What were the actions that led to customer complaints?
- What is the next best product to offer a customer?
- What is the pattern of occurrences before our systems break down?

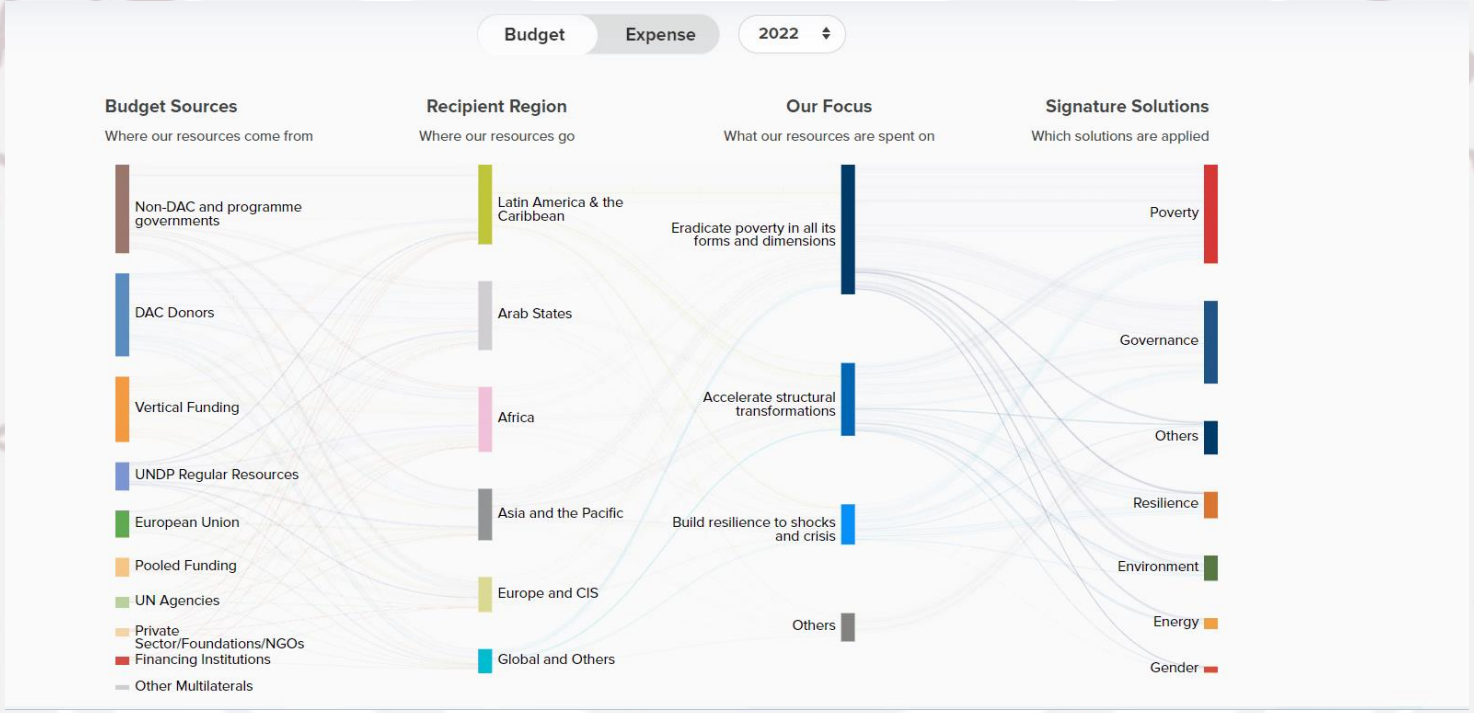
Event Analysis

- Used to understand **sequences of events** – what comes before what.
- **Sankey diagram** - graphic illustration of **flows from one set of values to another** as a series of paths, where **width** of each flow represents the **quantity**



Event Analysis

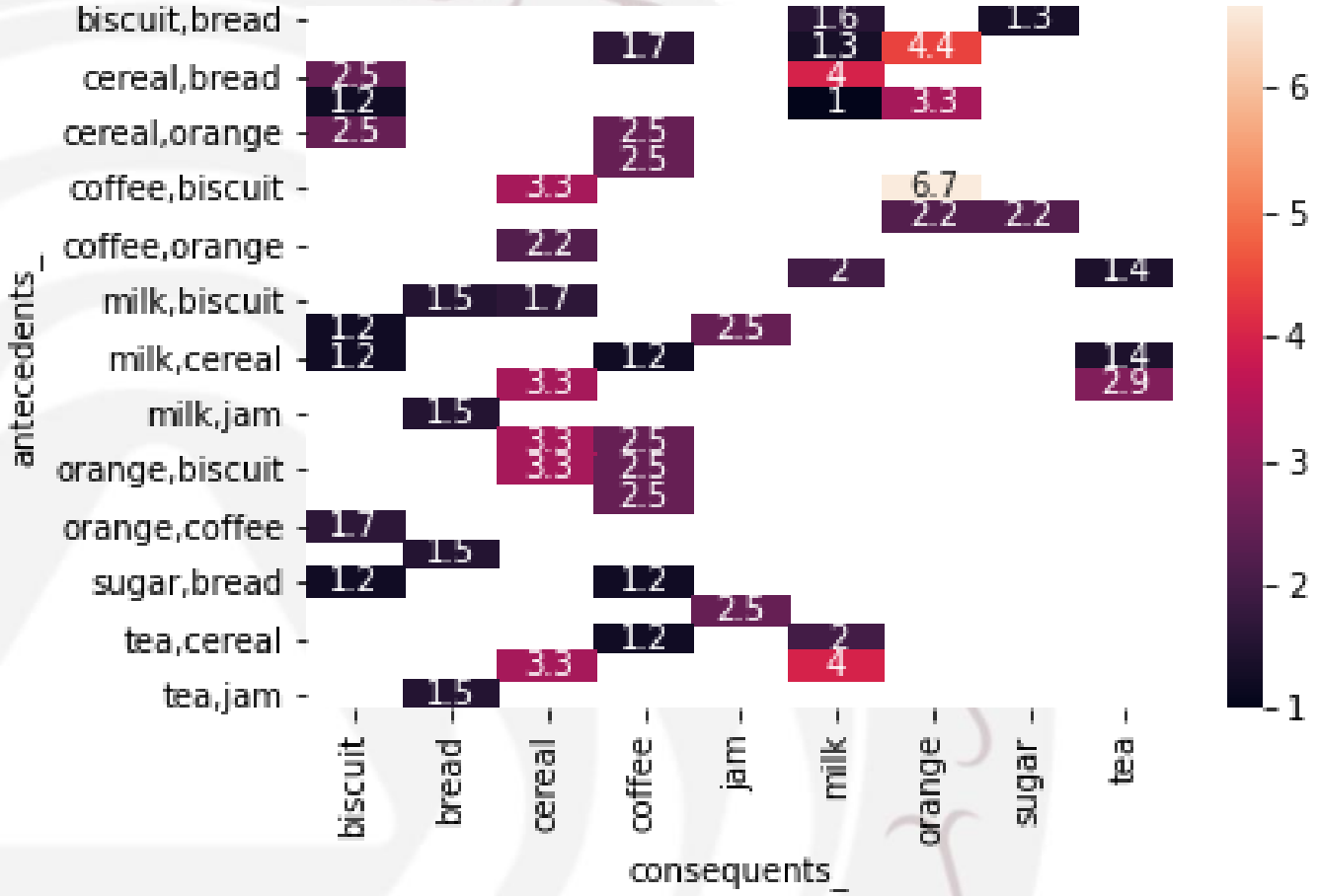
- **Use-case:** Understand what development causes and geographies are supported by various funders.



Source: https://open.undp.org/

Association Analysis

- Detect **relationships or associations** between specific values of **categorical variables in large data sets**
  - **Antecedent (IF):** This is an item/group of items that are typically found in the Itemsets or Datasets.
  - **Consequent (THEN):** This comes along as an item with an Antecedent/group of Antecedents.



Association Analysis

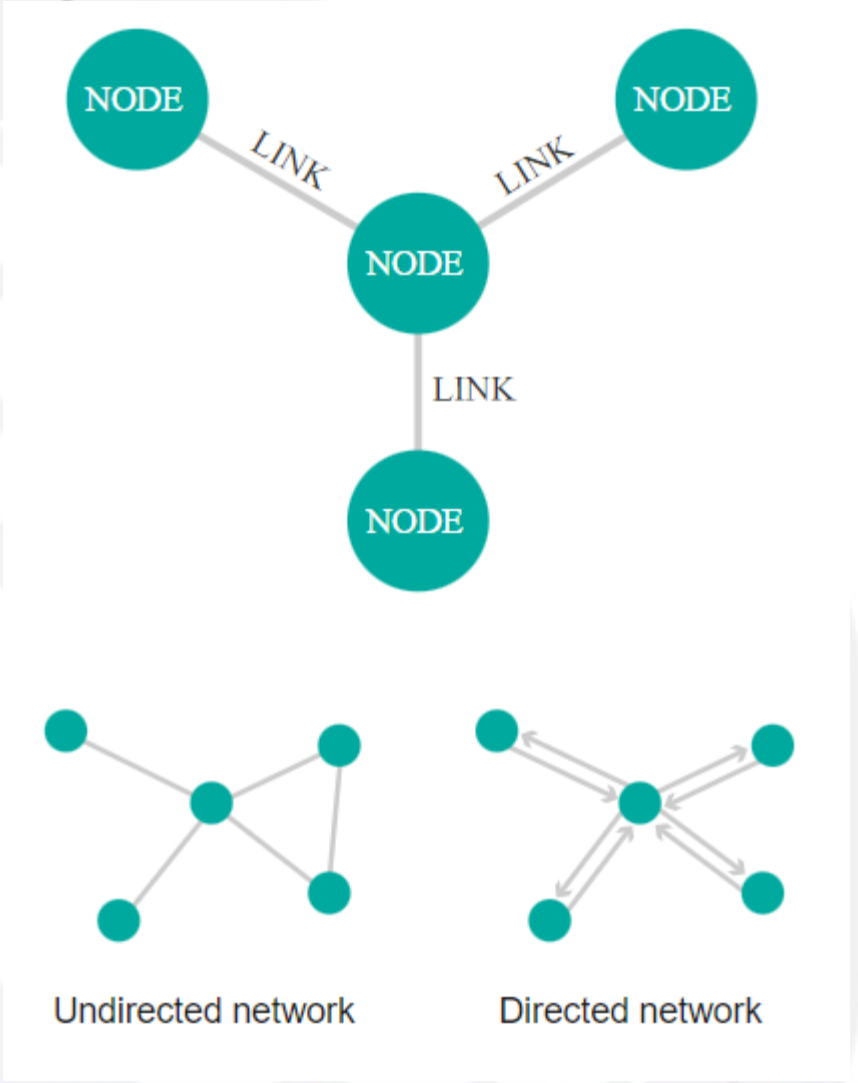
- **Use-case:** Create microspecializations based on learners' preferred SPARTA courses to enroll in.

Itemsets	Specialization Names	Support			
		Training	Test	Z-stat	p-Value
'Data Management Fundamentals', 'Data Visualization Fundamentals', 'Essential Excel Skills for Data Preparation and Analysis', 'Getting Grounded on Analytics'	Data Analysis Non-Programming Specialization	1	0.26506	1.12369	0.261145
'Computing in Python', 'Data Visualization using Tableau and Python', 'Statistical Analysis and Modeling using Excel', 'Statistical Analysis and Modeling using SQL and Python'	Statistical Programming Specialization	1	0.253012	1.313265	0.189094
'Dashboards and Drill-Down Analytics', 'Data Management Fundamentals', 'Data Visualization Fundamentals', 'Getting Grounded on Analytics'	Data Visualization Speicalization	0.330645	0.253012	1.195094	0.23205
'Computing in Python', 'Data Visualization Fundamentals', 'Data Visualization using Tableau and Python', 'Statistical Analysis and Modeling using Excel'	Data Analysis Programming Specialization (Visualization Variant)	0.314516	0.253012	0.955883	0.339131
'Computing in Python', 'Data Visualization using Tableau and Python', 'Design Thinking for Analytics', 'SQL for Business Users'	Data Analysis Programming Specialization (Database Variant)	0.298387	0.253012	0.712435	0.476196
'Data Management Fundamentals', 'Data Science and Analytics Project Management', 'Data Visualization Fundamentals', 'Getting Grounded on Analytics'	Analysis Project Lead Sepcialization (Data Viz Variant)	0.290323	0.253012	0.588957	0.55589

Support: % of all records which contain the itemset

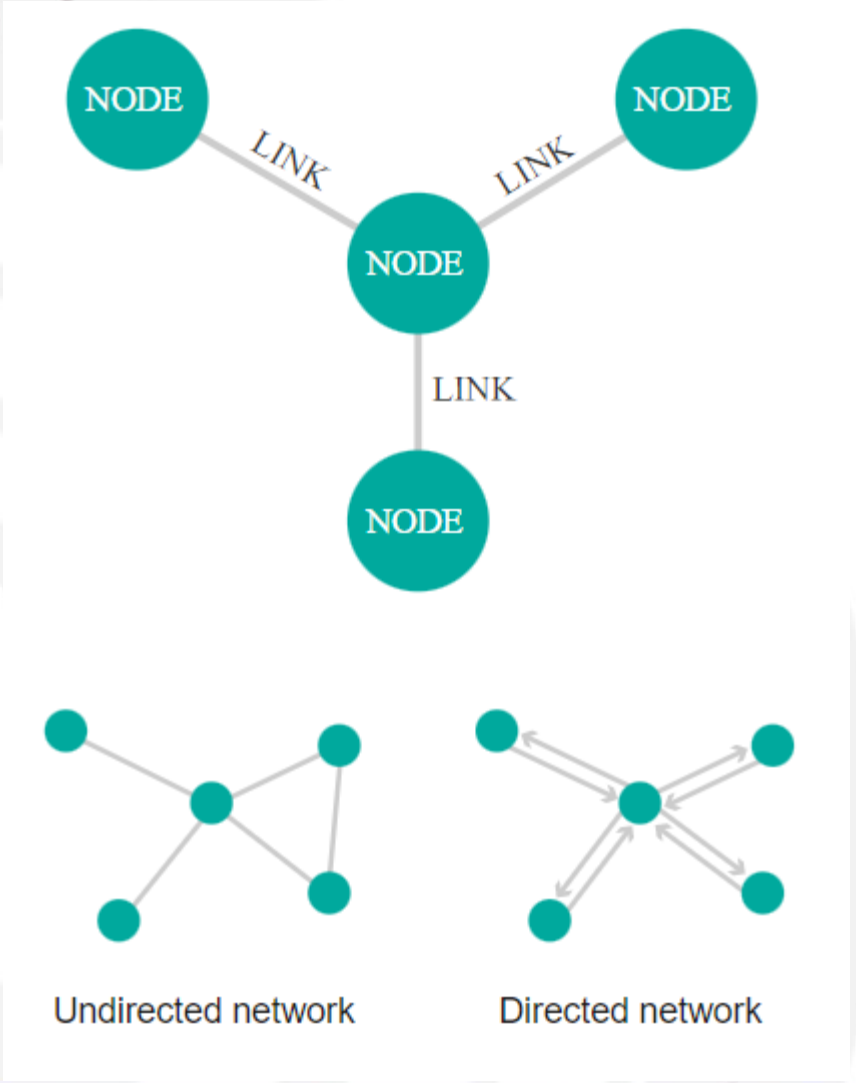
Network Analysis

- A social network consists of both **nodes** (vertices) and **edges. Nodes** represent the **objects of interest**, while the **edges** represent the **connection between them**.
- Network graph shows interconnection between a set of entities.



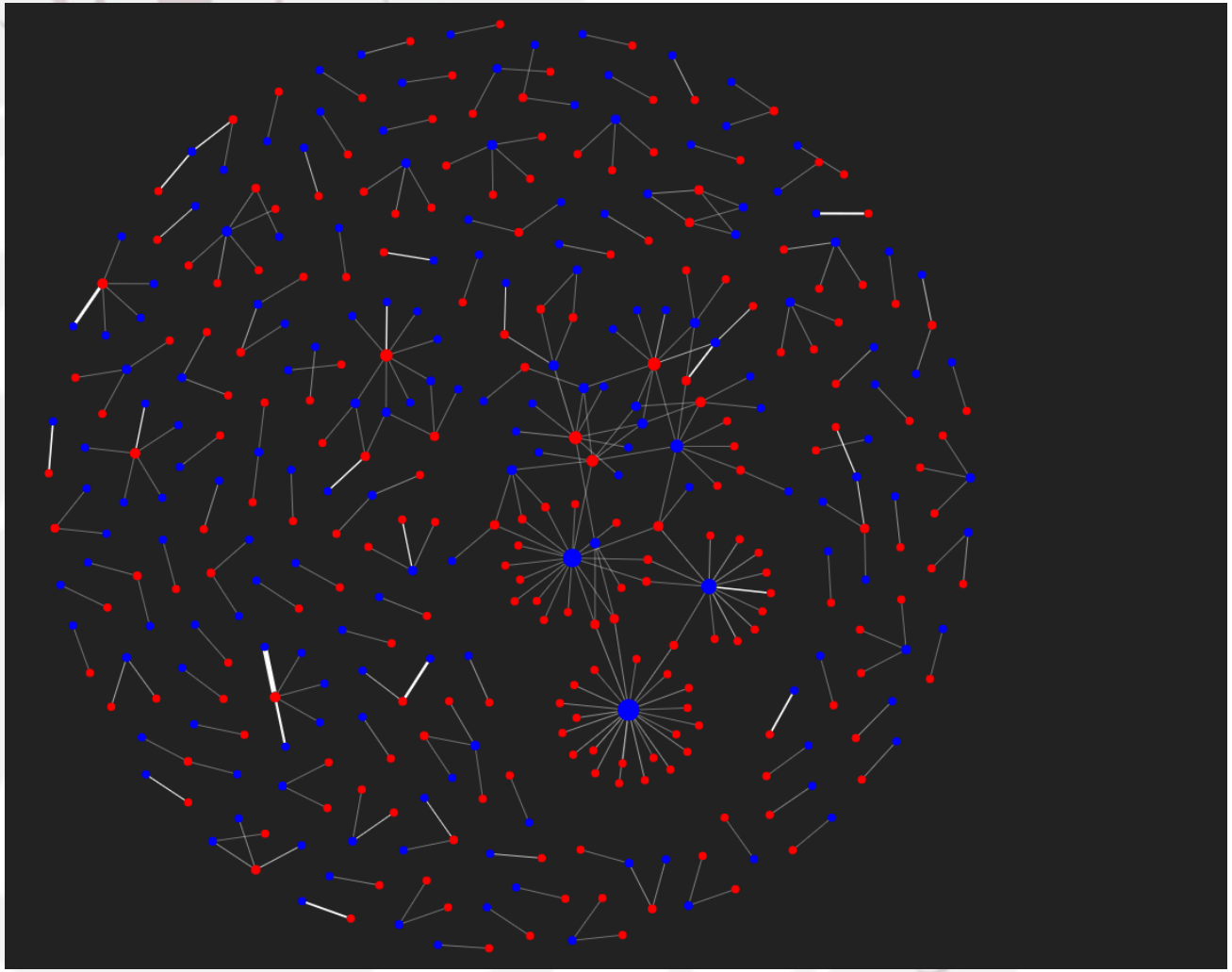
Network Analysis

- Use-cases:
  - Network-based diffusion analysis
  - Community detection
  - Influential node detection



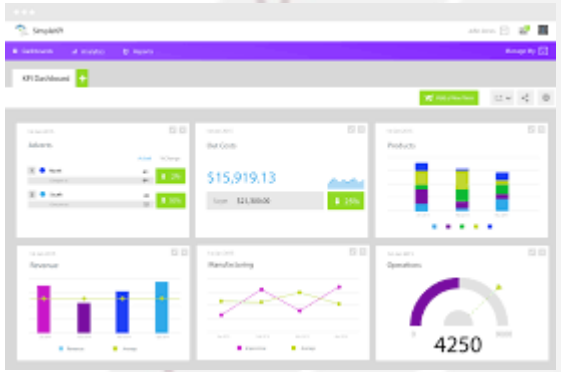
Network Analysis

Use-case: Detection of coordinated sharing of incriminating news / issues on Food and Drug Authority (FDA) by cigarette, heated tobacco products (HTPs), and vape groups and pages in Facebook.



What are dashboards?

Uses of Dashboards

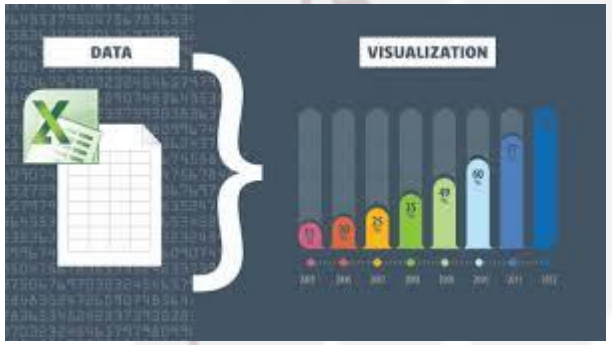


Easily Present KPIs and Measurements

Top Level Summary at a glance

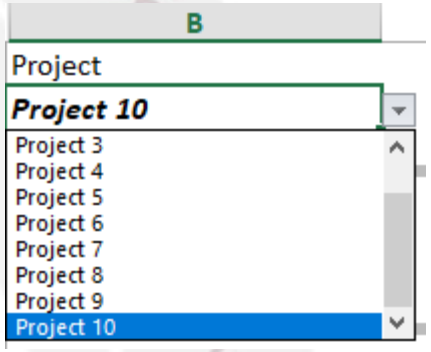
Leads to further analysis

Dashboarding Principles



What is the story?

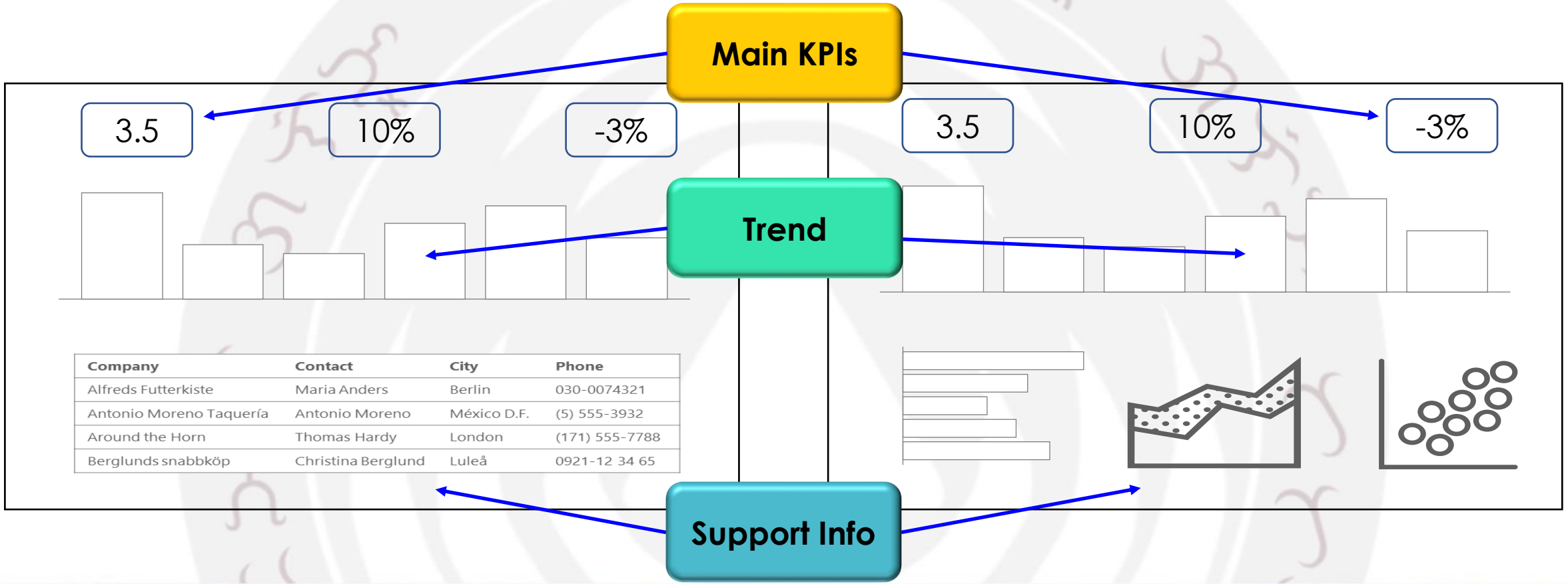
Clean layout and design



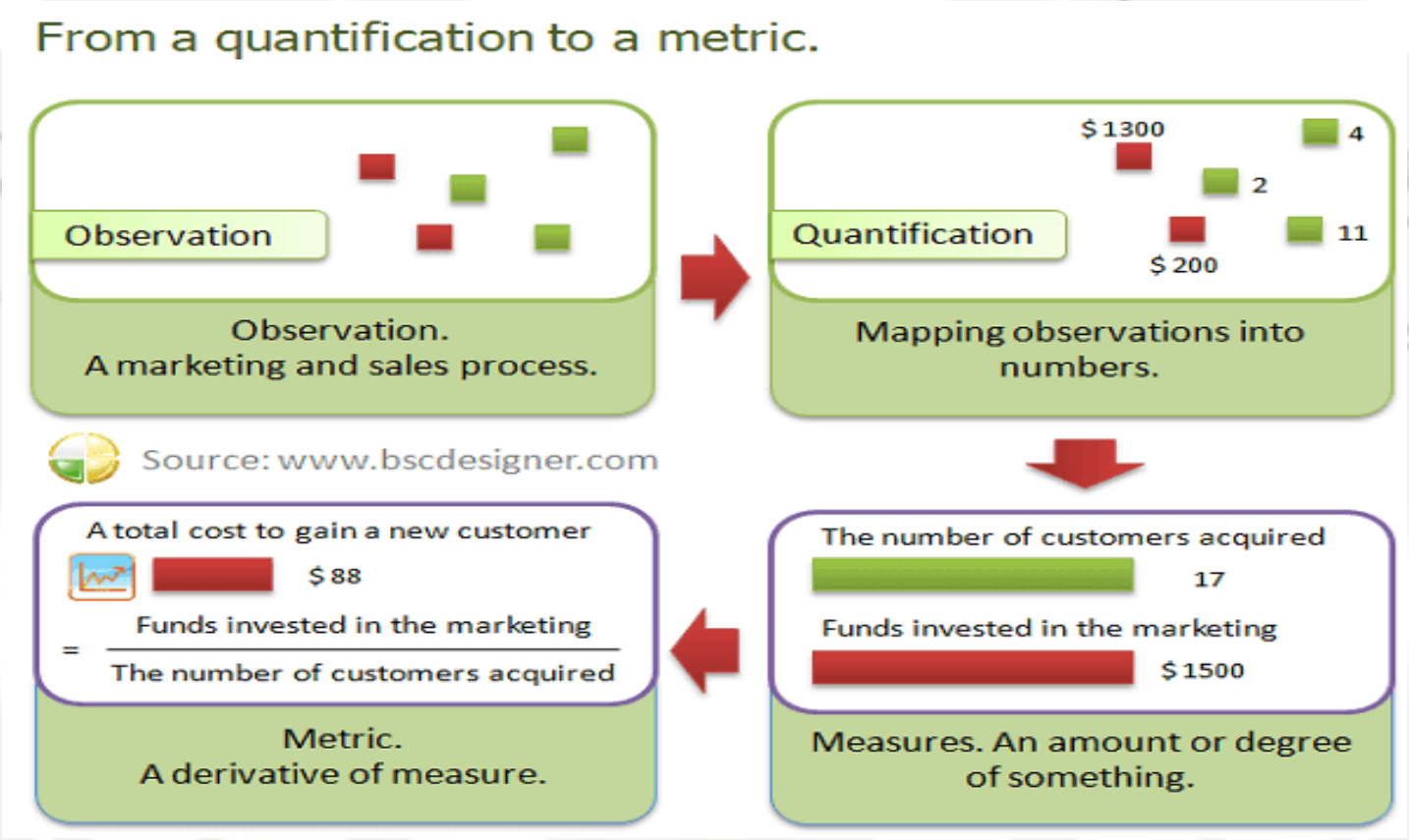
Interactivity and automation

Which is better?

Clean Dashboard Design



What is a metric?

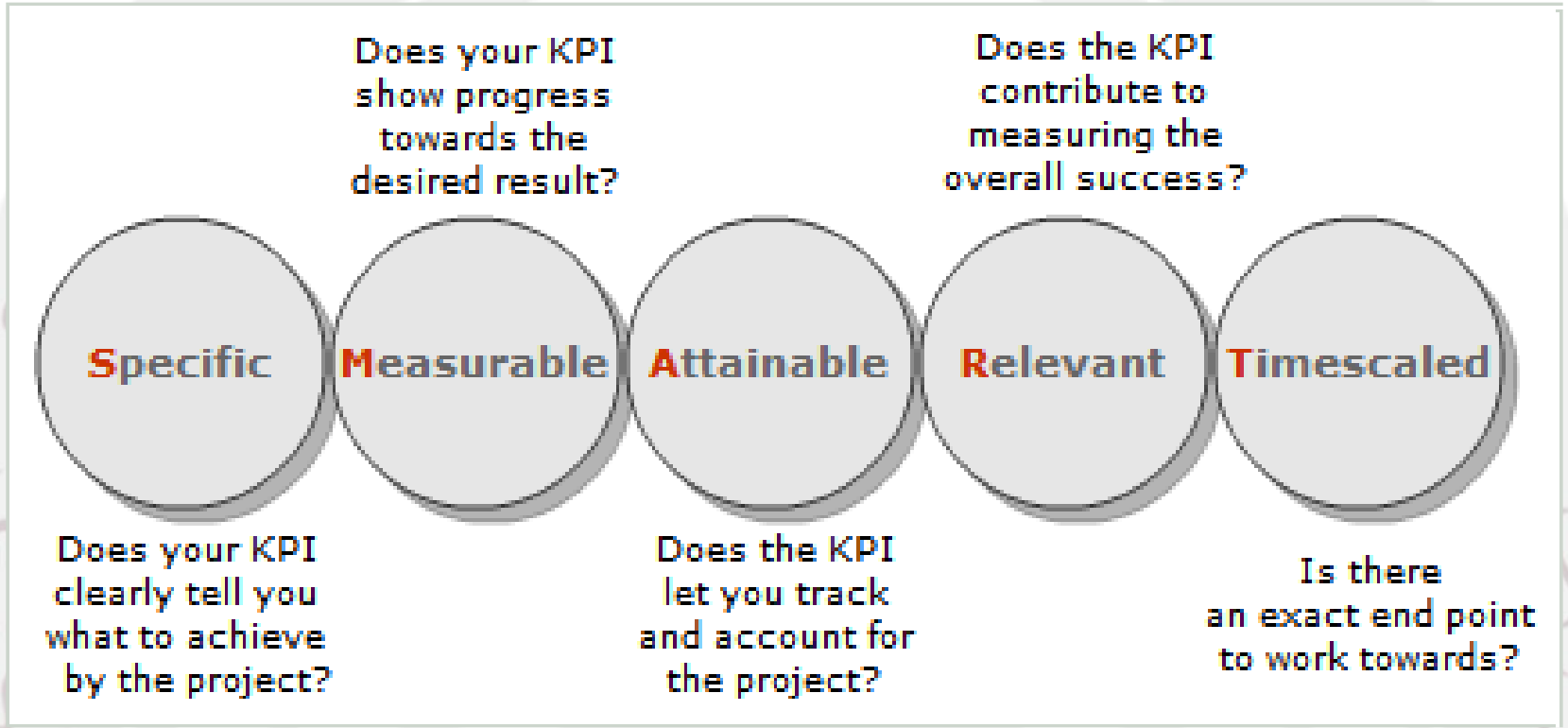


What are Key Performance Indicators?

ACTIONABLE INSIGHTS

What are Key Performance Indicators?

What are Key Performance Indicators?



Example of KPIs for Disaster Management

Natural Disaster Monitoring and Funding

- Number of early warning models developed
- Regular Space remote sensing
- % increase in government funding
- % increase in NGO's and direct public funding

Natural Disaster Risk Reduction

- Natural Disaster risk maps produced
- Number of vulnerable area identified
- Dissemination of Early Warnings
- % of natural disaster risks prioritized.

Natural Disaster Preparedness

- % area and building inspections
- Percentage increase in relief measures included
- Number of Damage assessment initiatives
- Number of Rehabilitation and reconstruction activities

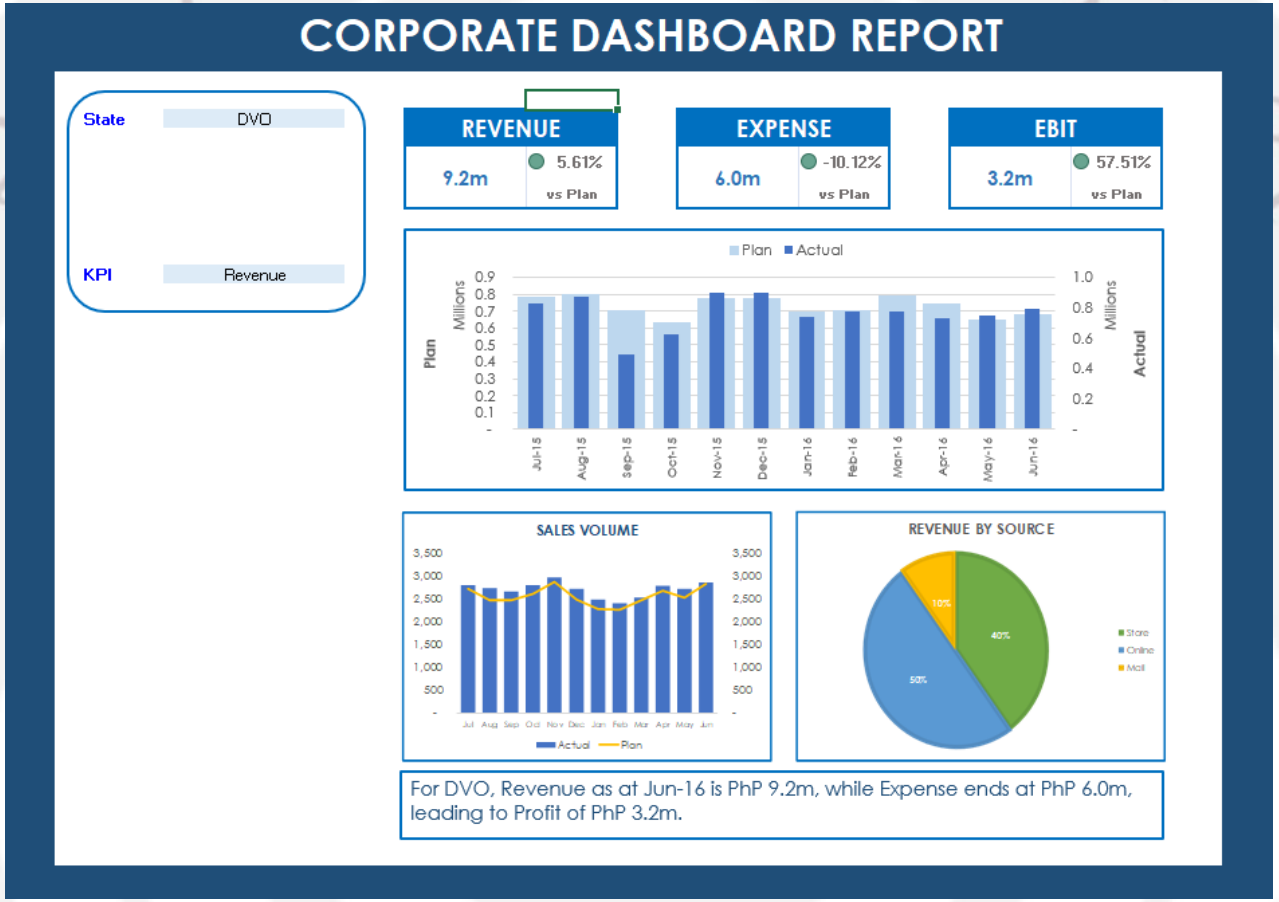
Education and training

- Number of training sessions of safety management
- Psychological strength development training sessions
- Number of medicinal training programs conducted
- Number of on-field consultations provided

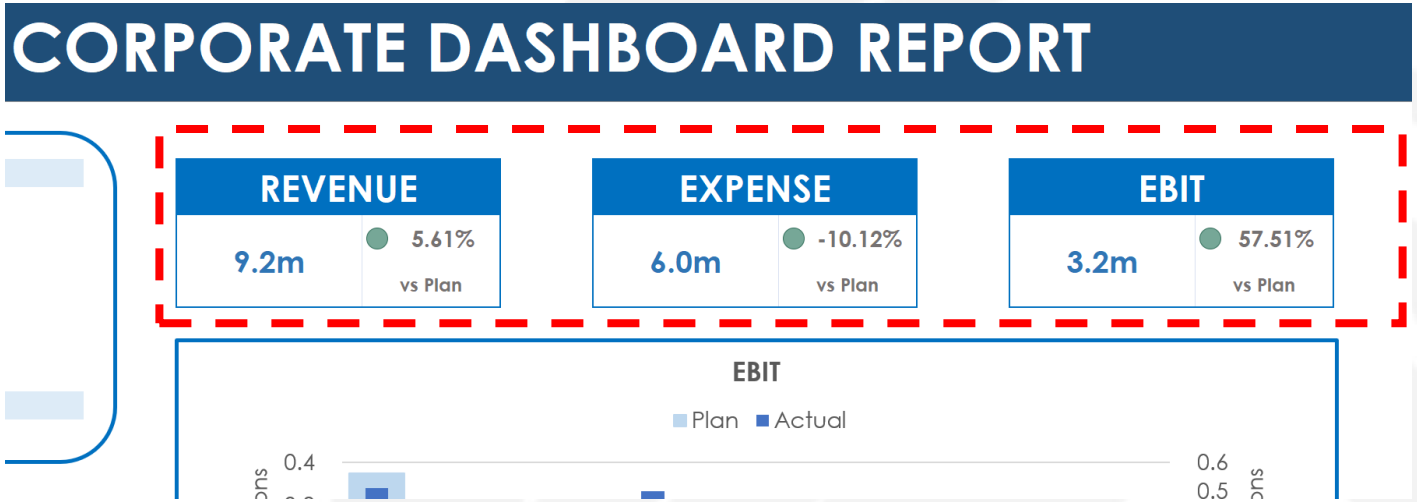
Dashboard Creation Process

- Highlight
- Trend
- Comparison
- Commentary
- Composition
- Interactivity

Example of a Finance Dashboard

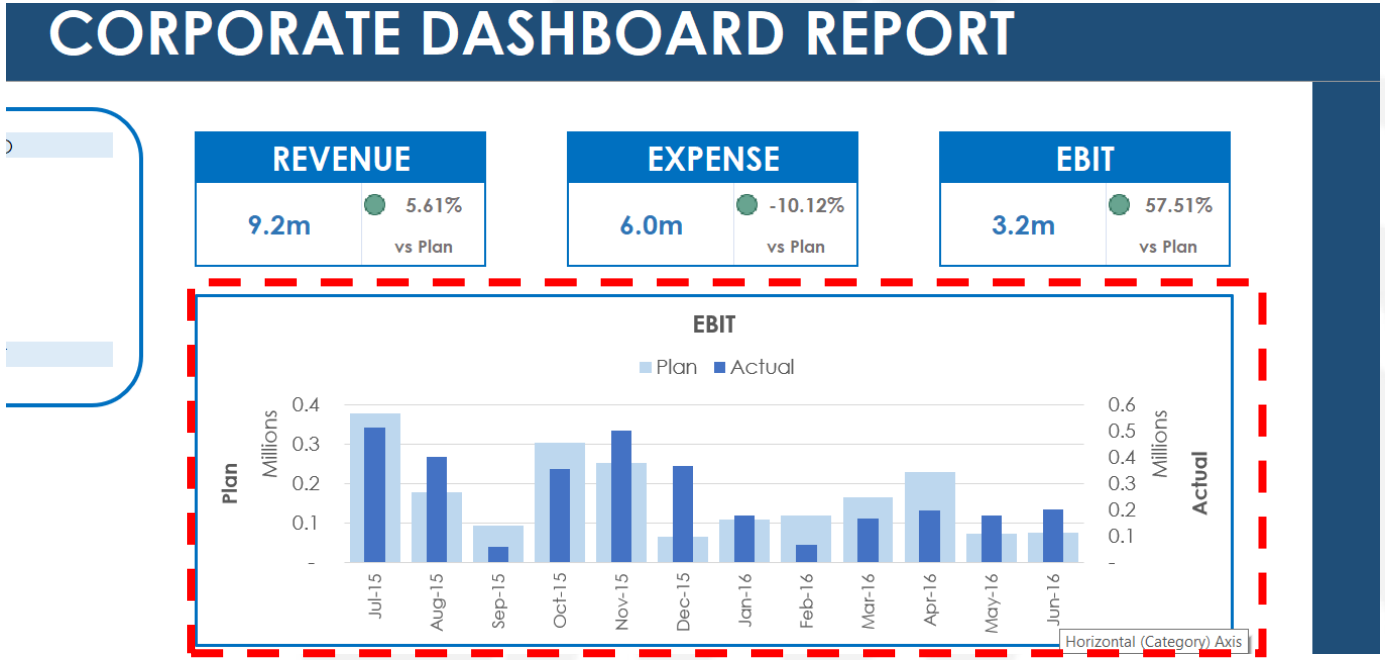


Highlight



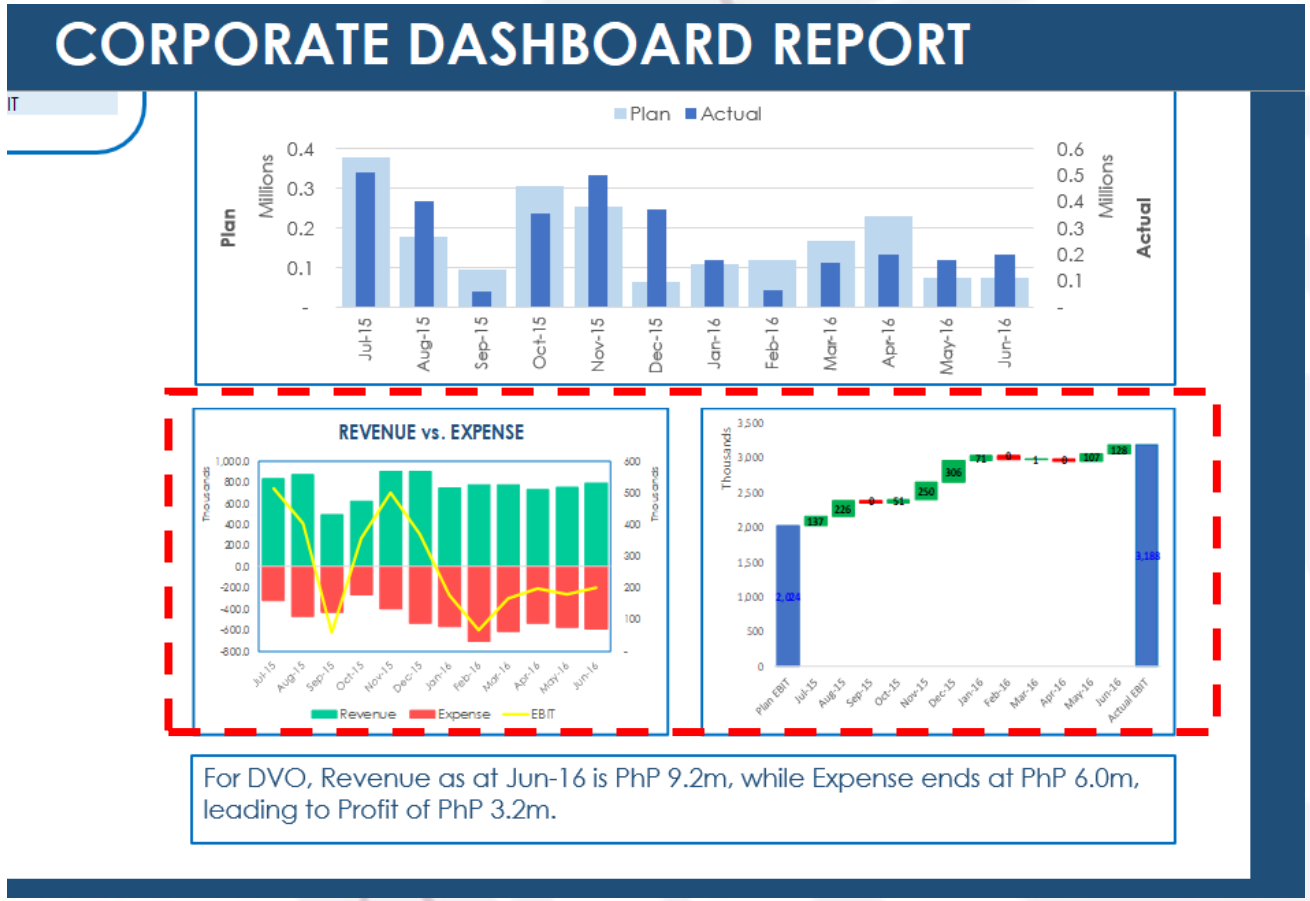
- Top level automatically draws attention to the story being discussed.

Trend



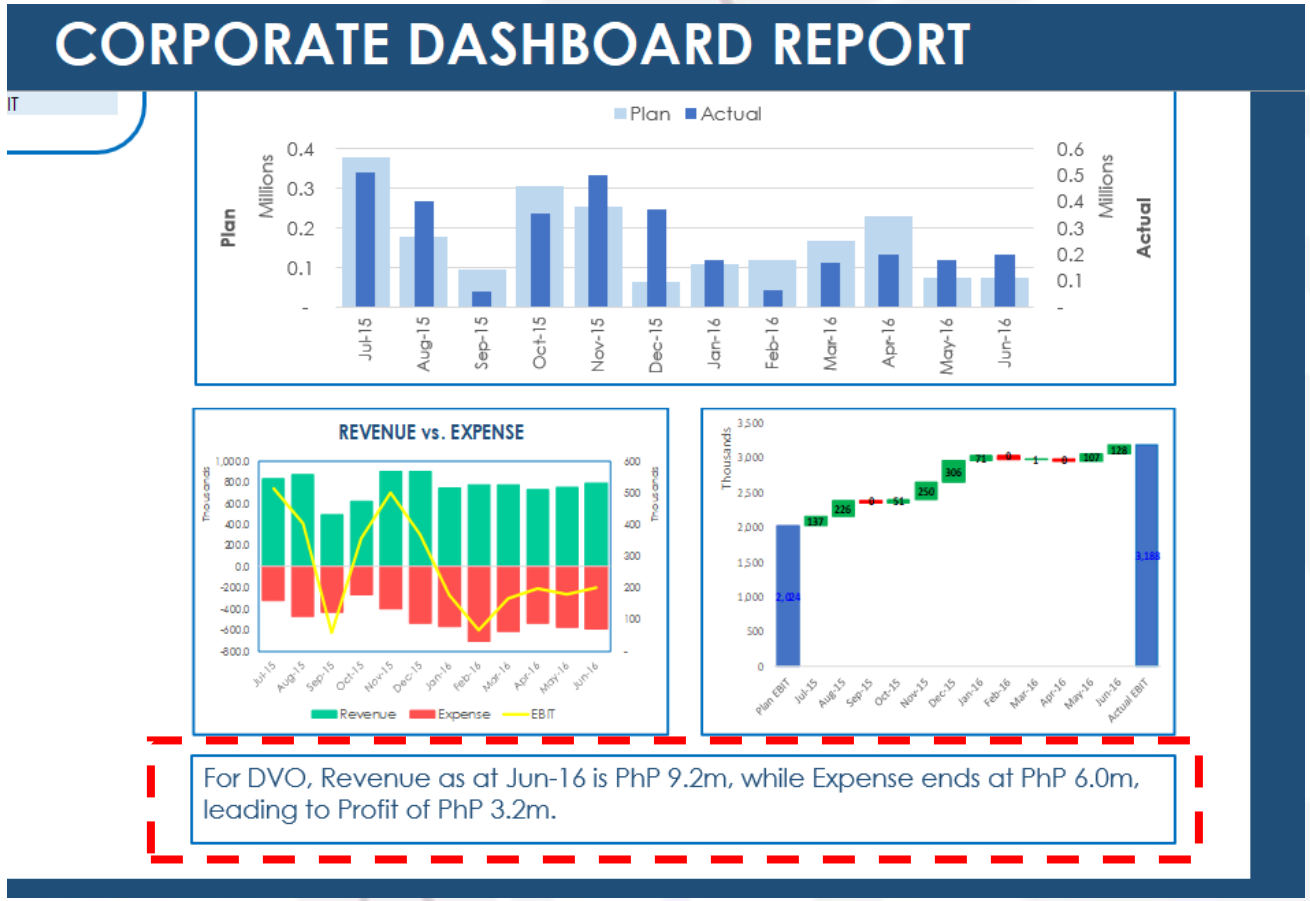
- Provide context by showing change over time.

Composition



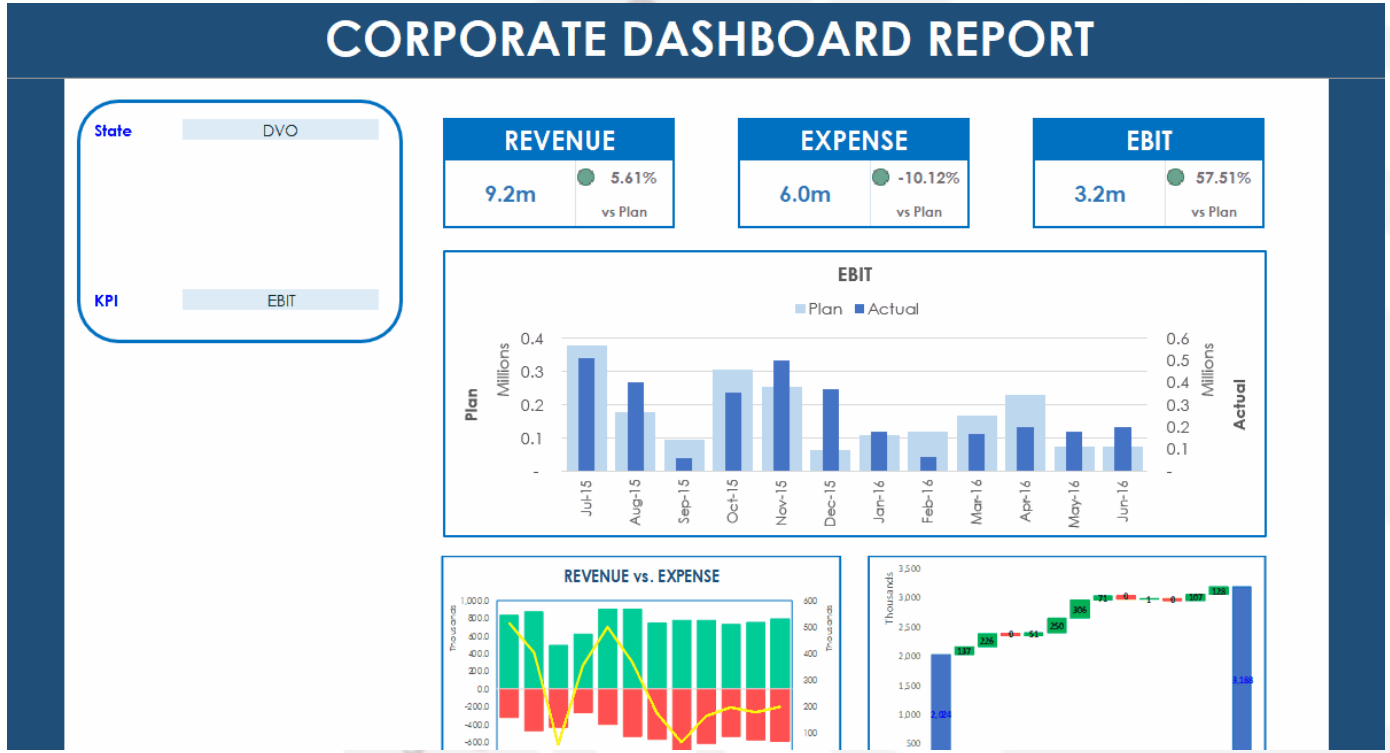
- This could be finer summarizations (e.g. tables) or visualizations, that answer immediate questions arising from observations in the first 2 sections

Commentary



- Reinforce insight by verbalizing trend.

Interactivity



- Allow users to safely interact with data by providing controls.

THANK YOU!

E-mail: francesclaire.tayco@cirrolytix.com

Website: https://cirrolytix.com

LinkedIn: https://www.linkedin.com/in/claire-san-juan-tayco/