86th Annual SYMPOSIUM

ncaCPA®

Become a Data Analysis Expert

Developed by Bryan L. Smith, CPA.CITP, CISA

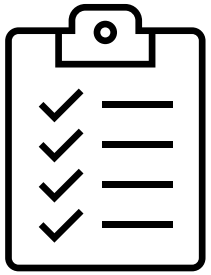
November 18, 2025 – 1:00-2:15 PM

Bryan L. Smith, CPA.CITP, CISA

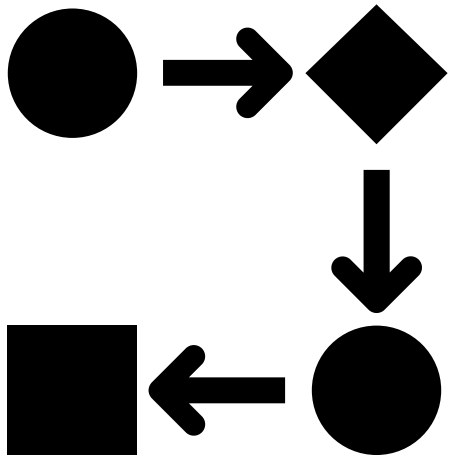
- Owner DataSmithPro, LLC
- Co-founder of CPA Crossings, LLC
- Strategic technology advisor to CPAs w/ 40+ years of experience
- Nationally recognized author and presenter on CPA technology
- Extensive knowledge and experience in proper Excel design techniques as well as helping CPAs build spreadsheets that are accurate
- As a Certified Information Systems Auditor (CISA), spent a considerable amount of time auditing and validating thousands of Excel spreadsheets
- Developed various data analysis and Microsoft Power BI courses and has worked with multiple organizations to train and coach staff on proper data analysis techniques
- Masters of Science in Information Assurance (MSIA)

Presentation Outline

Click links to navigate to specific section ---------- >



Data Analysis Process Flow



Data Analysis Defined

- **Data analysis** is a process of *inspecting, cleansing, transforming, and modeling data* with the goal of *discovering useful information, suggesting conclusions, and supporting decision-making*.

Source: Wikipedia

Why Perform Data Analysis?

Data Analysis Process Flow

Reality

Collect Raw Data

Process Data

Clean Data

Exploratory Data Analysis

Build Data Model & Measures

Share Data Product

Communicate/ Visualize Report

Make Insightful Decisions

Data science process flowchart from "Doing Data Science", Cathy O'Neil and Rachel Schutt, 2013

Reality

- Your Organization
- People performing various activities
- Activities generate data
- No two organizations have identical data
- Determine your organization's data

Collect Raw Data

Collect Raw Data

Build “pipelines” to your data

- Data connectors
- Files (Excel, csv, text, report, databases)

Where is it?

How do we get it?

Is it complete?

Process Data

- Data is processed through the pipelines
- Wherever possible automate this
- Many tools on the market to help with this
- Different tools for different data
- Sometimes left to the technical staff

Clean Data

Clean Data

Get data into a suitable format

May have several data tables

- Sales, customers, products, sales person

Filter, extract, combine, merge, eliminate, match, etc.

Create calculated columns for missing data

Exploratory Data Analysis

Do you have the right data for what you want to know?

Is our data clean?

- Duplicates, missing values, absurd outliers, incorrect, etc.

Garbage In, Garbage Out

May need to go back to previous steps

Build Data Model & Measures

Build Data Model & Measures

Put the raw data into a useable form

Connect tables through relationships

Allows access to all cleansed data

Build measures/metrics for what you want to know

- Dependent on who is using the data model

Share Data Product

Share Data Product

- Data Product = Data Model & Measures
- Goal is to eliminate siloing of data
- No ability to change the data, the model or the measures
- Provides consistent analysis across the organization
- Otherwise, people just copy & paste into Excel!!

Communicate/ Visualize Report

Communicate/Visualize Report

- Tells the story of our data
- Pictures are worth a 1,000 words
- Limited tables (rows & columns of data)
- Graphics, pictures, charts, etc.
- Interactive visualizations
- Take your audience into consideration

Make Insightful Decisions

Make Insightful Decisions

The ultimate goal of Data Analysis

What is my data telling me?

- Trends, patterns, anomalies, etc.

Data backed decisions

Alternative is gut-based decisions

The Importance of the Data Analysis Process Flow

Each step is an important as another

- Don't discount any step

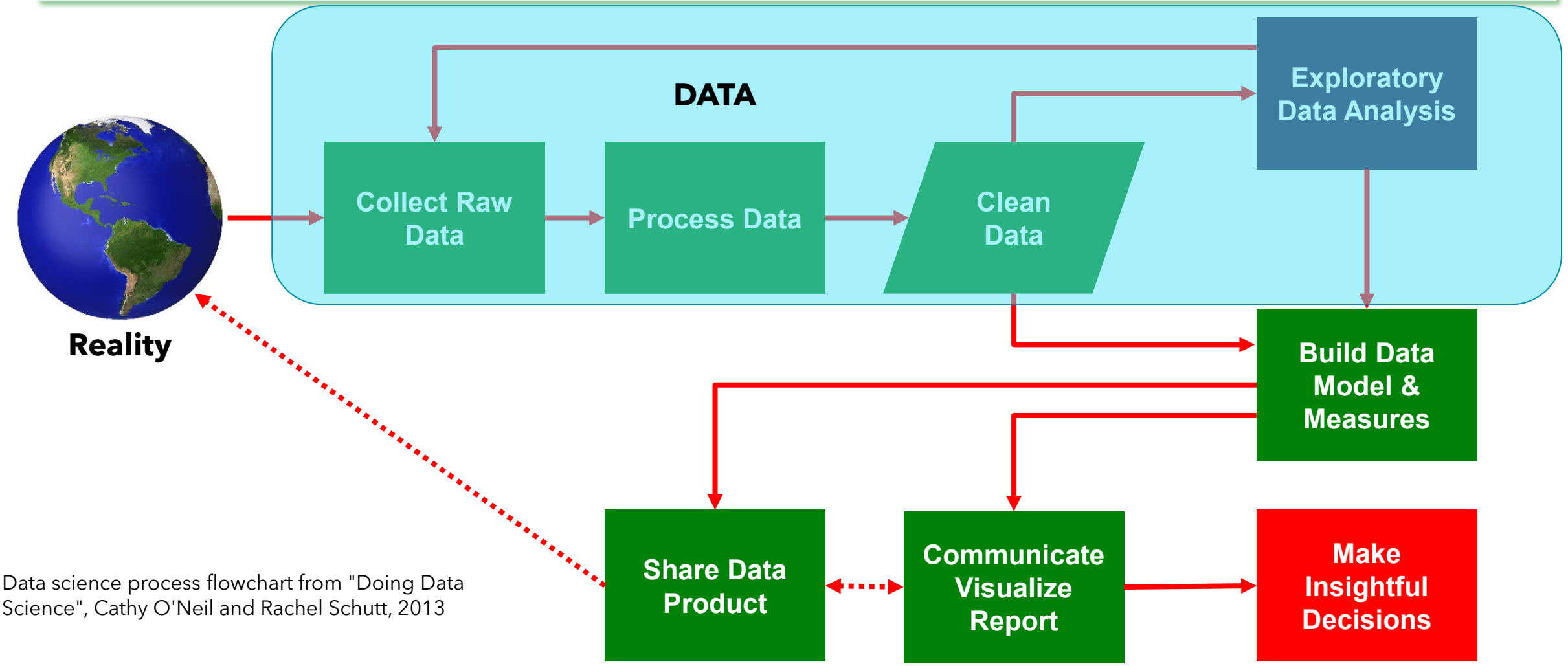
Following from beginning to end = greatest success

- Don't skip steps

No magic button to push

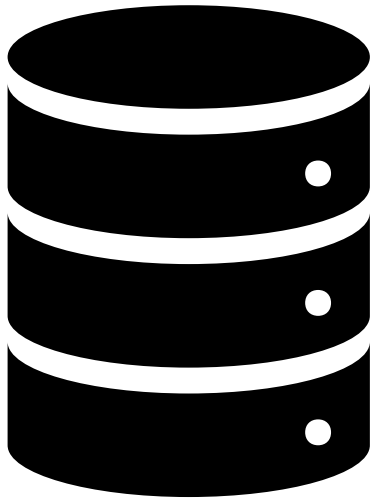
Team effort to eliminate data silos

Data Analysis Process Flow



Data science process flowchart from "Doing Data Science", Cathy O'Neil and Rachel Schutt, 2013

Data



Data Example

Determine what data is needed	What are you trying to analyze?	Want to determine profitability between product categories
		Like to see how it compares between locations
		Like to see if there is any difference between salespeople
	What do we need to accomplish this?	Sales data with product categories, locations, salespeople
		Profitability will require cost data
		Calculations for profitability

Data Example – Where's Our Data?

Sales Data

- In order table
  - Line by line – Qty, unit price, discount, cost

Product Category Data

- In product table

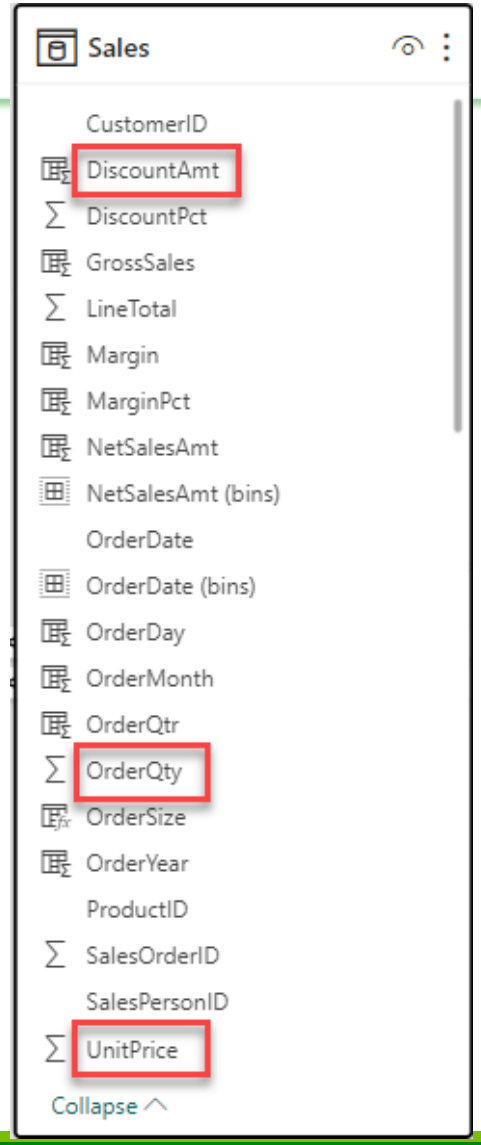
Location Data

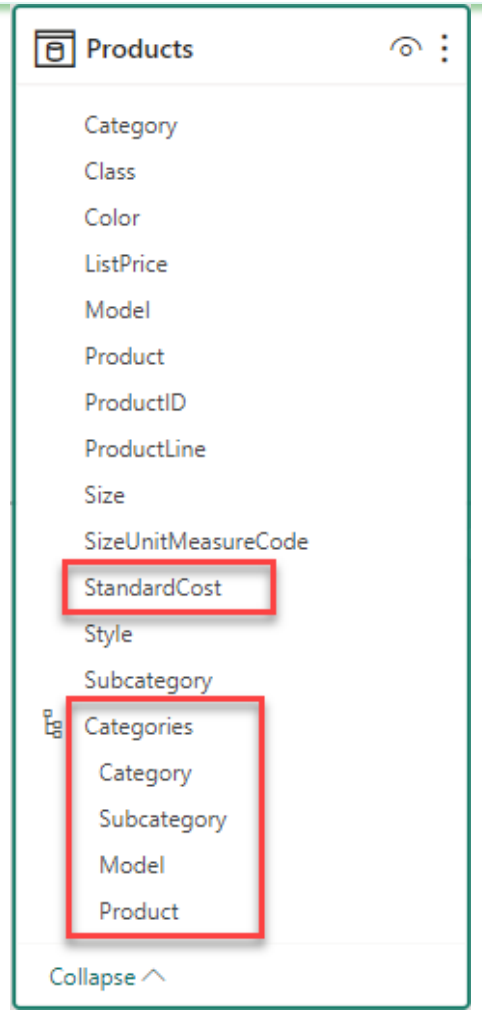
- In customer table

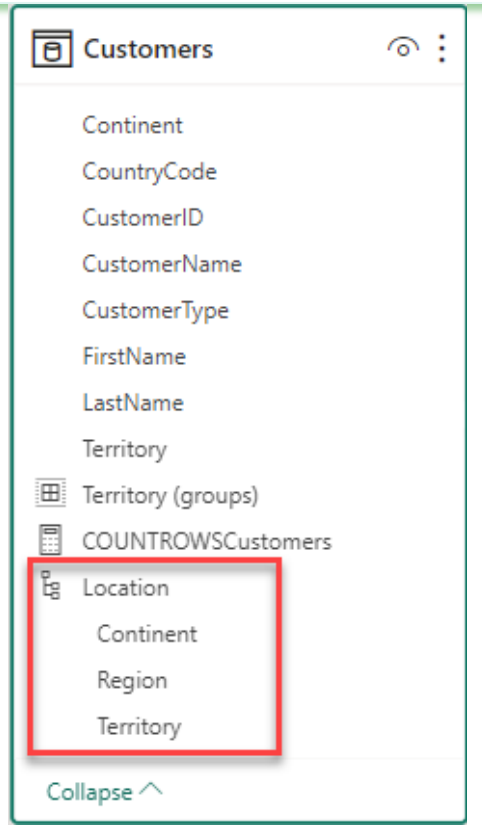
Salesperson Data

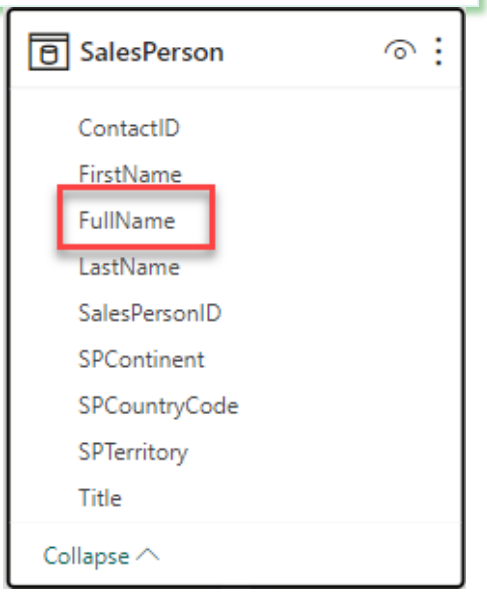
- In salesperson table

Data Example – Our Data Tables



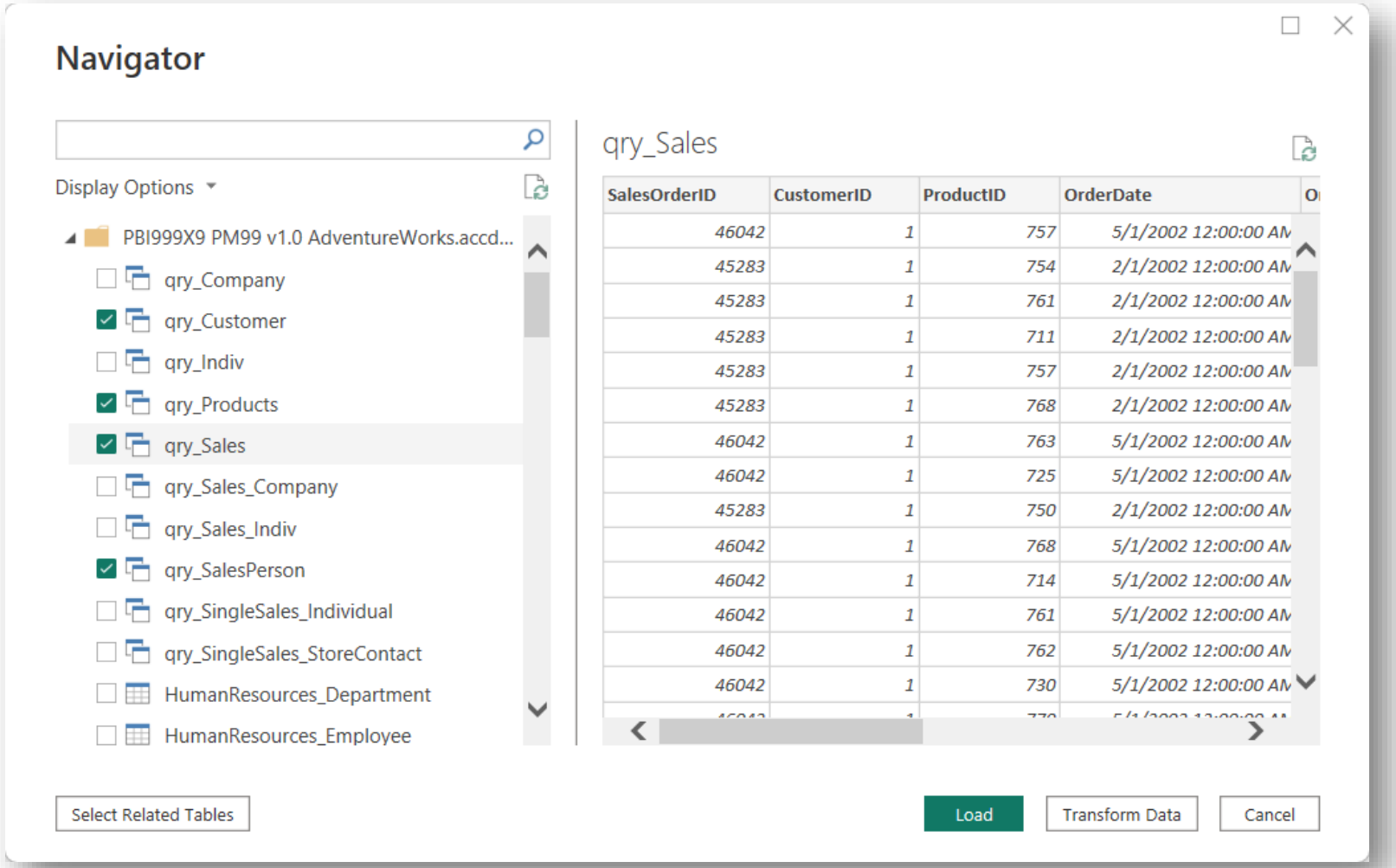






Transaction/Detail table
vs.
Dimension/Lookup table

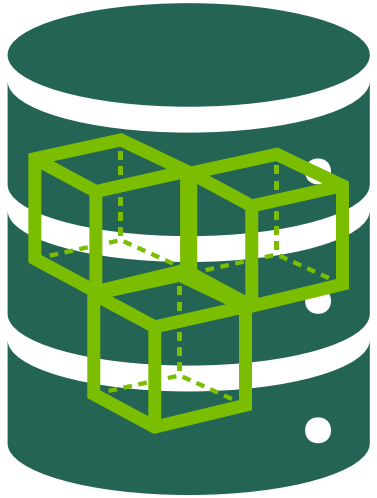
Use a Tool to Collect Raw Data



Clean Data

- Rename fields
- Remove fields
- Calculated columns
- Change data types
- Split data into multiple fields
- Normalize data into lookup tables

Data Modeling



Data Modeling in Data Analysis

Starts with cleansed data

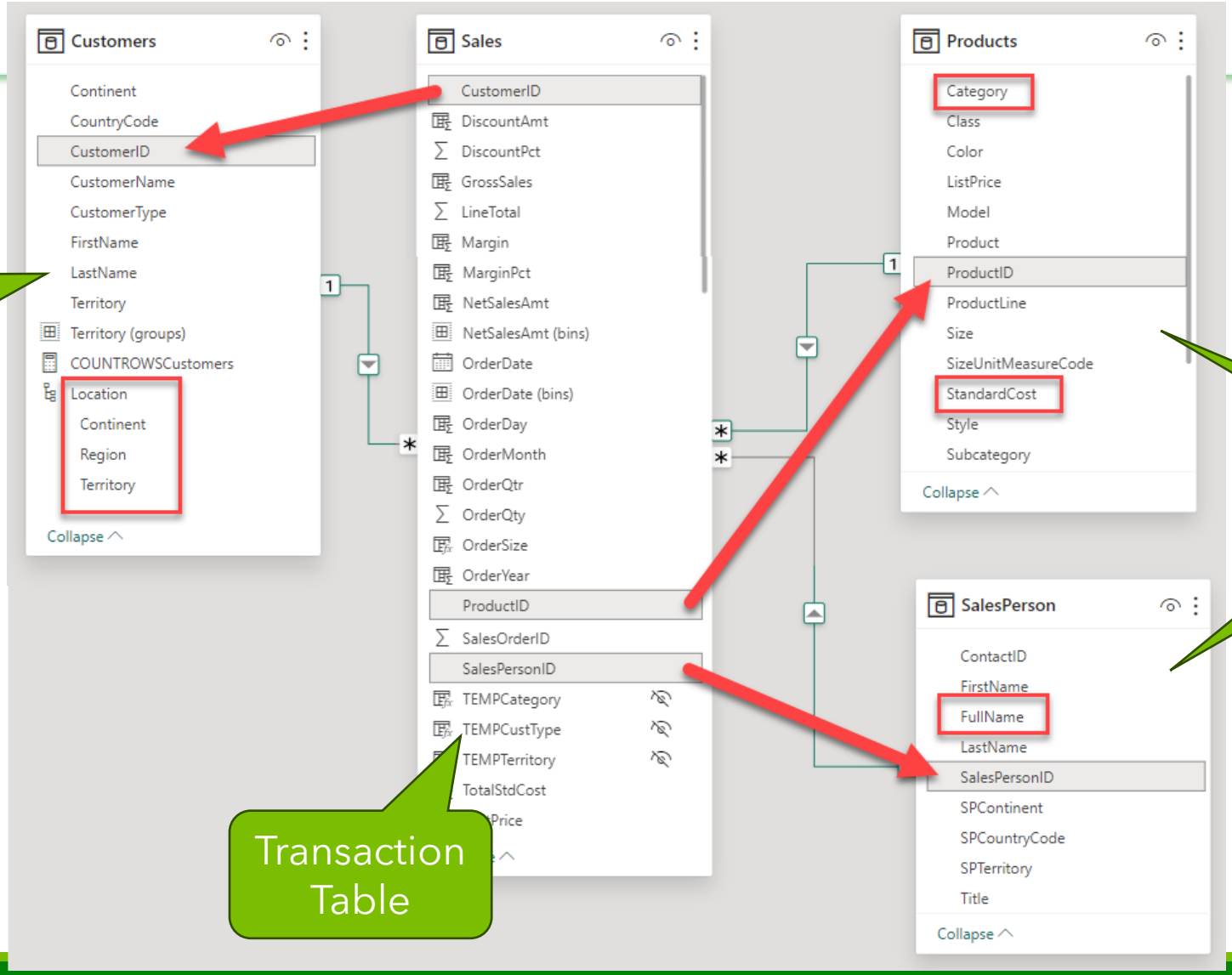
Build relationships between tables

- Allows access to all fields in all tables
- Excel users do this with VLOOKP

Create measures

- Metrics by which you want to analyze

Relationships Between Tables



What Relationships Provide

Salesperson - *Name*

Customer - *Continent*

Product - *Category*

Sales – *Net Sales*

FullName	Europe	North America	Pacific	Total
Ito, Shu		**$6,427,006**		**$6,427,006**
Bikes		$5,365,754		**$5,365,754**
Components		$915,730		**$915,730**
Clothing		$108,683		**$108,683**
Accessories		$36,839		**$36,839**
Saraiva, José	**$3,837,927**	**$2,088,491**		**$5,926,419**
Bikes	$3,050,029	$1,893,519		**$4,943,547**
Components	$640,208	$177,821		
Clothing	$107,827	$10,773		**$118,599**
Accessories	$39,863	$6,379		**$46,242**
Varkey Chudukatil, Ranjit	**$4,509,889**			**$4,509,889**
Bikes	$3,477,638			**$3,477,638**
Components	$860,159			**$860,159**
Clothing	$125,403			**$125,403**
Accessories	$46,688			**$46,688**
Campbell, David		**$3,729,945**		**$3,729,945**
Bikes		$3,111,694		**$3,111,694**
Components		$544,478		**$544,478**
Total	**$10,174,882**	**$66,893,601**	**$1,421,810**	**$78,490,294**

Measures

Build measures based upon analysis needs

Measure:

- A calculation based upon the data in our table, so we can perform analysis based on the calculation

Example:

- Sales Per (order, customer, product)
- Sales Based On “X” (product type, trans type, location)
- Time intelligence (MTD, YTD, LY, LYMTD, LYTD)

Measure Example (Average Order Per Customer)

CustomerName	Sum of NetSalesAmt	Average of NetSalesAmt
A Bike Store	$85,177	$1,167
A Great Bicycle Company	$9,055	$647
A Typical Bike Shop	$83,457	$1,113
Total	**$177,690**	**$1,097**

Implicit Measure calculates avg based on ALL items in the file

Average Line Item Per Customer

CustomerName	Sum of NetSalesAmt	Count of SalesOrderID	Average of NetSalesAmt
⊟ **A Bike Store**	**$85,177**	**73**	**$1,167**
45283	$28,395	18	$1,578
46042	$26,119	23	$1,136
44501	$19,669	20	$983
43860	$10,993	12	$916
⊞ **A Typical Bike Shop**	**$83,457**	**75**	**$1,113**
⊞ **A Great Bicycle Company**	**$9,055**	**14**	**$647**
Total	**$177,690**	**162**	**$1,097**

# of Line Items	**=**	73
Sum of NetSalesAmt	**=**	**$85,177**
Average per Line Items	**=**	**$1,167**

# of Sales Orders	**=**	**4**
Sum of NetSalesAmt	**=**	**$85,177**
Average per Sales Order	**=**	**$21,294**

Actual Average per Order

Average Order Per Customer using Measure

```
SumSales = SUM([NetSalesAmt])
Orders = DISTINCTCOUNT( [SalesOrderID])
SalesPerOrder = [SumSales]/[Orders]
```

Measures used to correctly calculate the value we want

CustomerName	Sum of NetSalesAmt	Average of NetSalesAmt	SalesPerOrder
A Bike Store	**$85,177**	**$1,167**	**$21,294.28**
43860	$10,993	$916	$10,993.40
44501	$19,669	$983	$19,669.42
45283	$28,395	$1,578	$28,395.22
46042	$26,119	$1,136	$26,119.08
A Great Bicycle Company	**$9,055**	**$647**	**$1,509.22**
A Typical Bike Shop	**$83,457**	**$1,113**	**$20,864.29**
Total	**$177,690**	**$1,097**	**$12,692.11**

Data Example (Repeat)

Determine what data is needed	What are you trying to analyze?	Want to determine **profitability between product categories**
		Like to see how it **compares between locations**
		Like to see if there is **any difference between salespeople**
	What do we need to accomplish this?	Sales data with product categories, locations, salespeople
		Profitability will require cost data
		Calculations for profitability

Performing our Analysis

Profitability

- Margin = SumSales – SumCost
- MarginPct = Margin / SumSales

Create table by Category

Create table with Location by Category

Create matrix with Location then Salesperson by Category

Results

Category	MarginMeasure	MarginPctM
Bikes	$7,936,387	8.38%
Accessories	$636,318	50.02%
Components	$490,226	4.15%
Clothing	$308,891	14.57%
Total	**$9,371,822**	**8.53%**

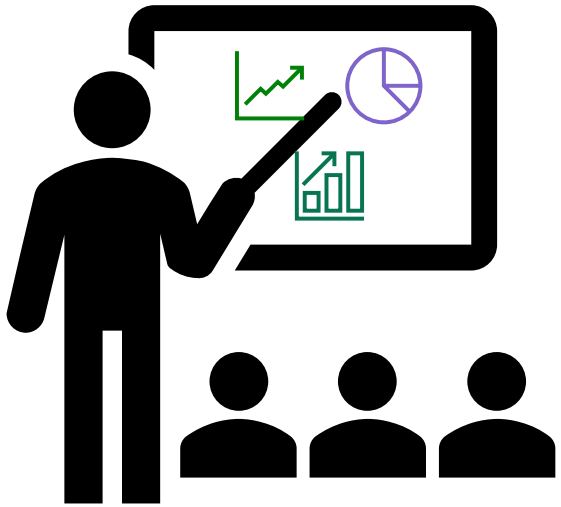
Category Continent	Accessories MarginMeasure	 MarginPctM	Bikes MarginMeasure	 MarginPctM	Clothing MarginMeasure	 MarginPctM	Components MarginMeasure	 MarginPctM	**Total** **MarginMeasure**	 **MarginPctM**
Pacific	$95,782	58.89%	$3,297,244	32.40%	$27,518	24.31%	$13,700	6.73%	**$3,434,244**	**32.23%**
Europe	$169,636	51.72%	$2,857,434	16.63%	$52,665	13.03%	$88,264	4.60%	**$3,167,999**	**15.97%**
North America	$370,900	47.46%	$1,781,709	2.65%	$228,708	14.27%	$388,261	4.01%	**$2,769,578**	**3.49%**
Total	**$636,318**	**50.02%**	**$7,936,387**	**8.38%**	**$308,891**	**14.57%**	**$490,226**	**4.15%**	**$9,371,821**	**8.53%**

Category Continent	Accessories MarginMeasure	 MarginPctM	Bikes MarginMeasure	 MarginPctM	Clothing MarginMeasure	 MarginPctM	Components MarginMeasure	 MarginPctM	**Total** **MarginMeasure**	 **MarginPc**
⊟ **Pacific**	**$95,782**	**58.89%**	**$3,297,244**	**32.40%**	**$27,518**	**24.31%**	**$13,700**	**6.73%**	$3,434,244	32.2
	$88,790	61.64%	$3,404,815	37.88%	$26,900	33.84%	$1,693	8.26%	**$3,522,198**	**38.1**
Tsoflias	$6,991	37.59%	($107,571)	-9.07%	$618	1.83%	$12,007	6.56%	**($87,954)**	**-6.1**
⊟ **Europe**	**$169,636**	**51.72%**	**$2,857,434**	**16.63%**	**$52,665**	**13.03%**	**$88,264**	**4.60%**	$3,167,999	15.9
	$128,606	61.81%	$3,342,774	36.14%	$30,847	30.88%	$6,176	5.84%	**$3,508,402**	**36.3**
Valdez	$11,600	34.75%	($115,271)	-8.19%	$396	0.56%	$13,699	4.36%	**($89,577)**	**-4.9**
Saraiva	$13,770	34.54%	($158,762)	-5.21%	$9,341	8.66%	$30,582	4.78%	**($105,069)**	**-2.7**
Varkey Chudukatil	$15,660	33.54%	($211,307)	-6.08%	$12,082	9.63%	$37,808	4.40%	**($145,758)**	**-3.2**
⊟ **North America**	**$370,900**	**47.46%**	**$1,781,709**	**2.65%**	**$228,708**	**14.27%**	**$388,261**	**4.01%**	$2,769,578	3.4
	$228,595	61.91%	$4,198,694	35.79%	$83,845	39.47%	$9,333	6.40%	**$4,520,467**	**36.2**
Ansman-Wolfe	$4,624	34.53%	($4,142)	-0.15%	$6,626	11.90%	$28,496	6.33%	**$35,604**	**1.0**
Campbell	$5,054	34.47%	($84,422)	-2.71%	$7,862	13.30%	$30,825	5.66%	**($40,681)**	**-1.0**
Saraiva	$2,244	35.17%	($49,926)	-2.64%	($3,203)	-29.73%	($2,814)	-1.58%	**($53,699)**	**-2.5**
Total	**$636,318**	**50.02%**	**$7,936,387**	**8.38%**	**$308,891**	**14.57%**	**$490,226**	**4.15%**	$9,371,821	8.5

What Would Help You to Make Better Decisions?

- Compare MTD this Year vs Last Year
- Sales based on product rating
- Average sales per capita distribution
- Sales per square feet
- Revenue vs forecast and variance
- YoY Revenue growth and variance

Data Visualizations



Things to Know about Visualizations

At its most basic, a visualization is a chart

- Built using data in the model

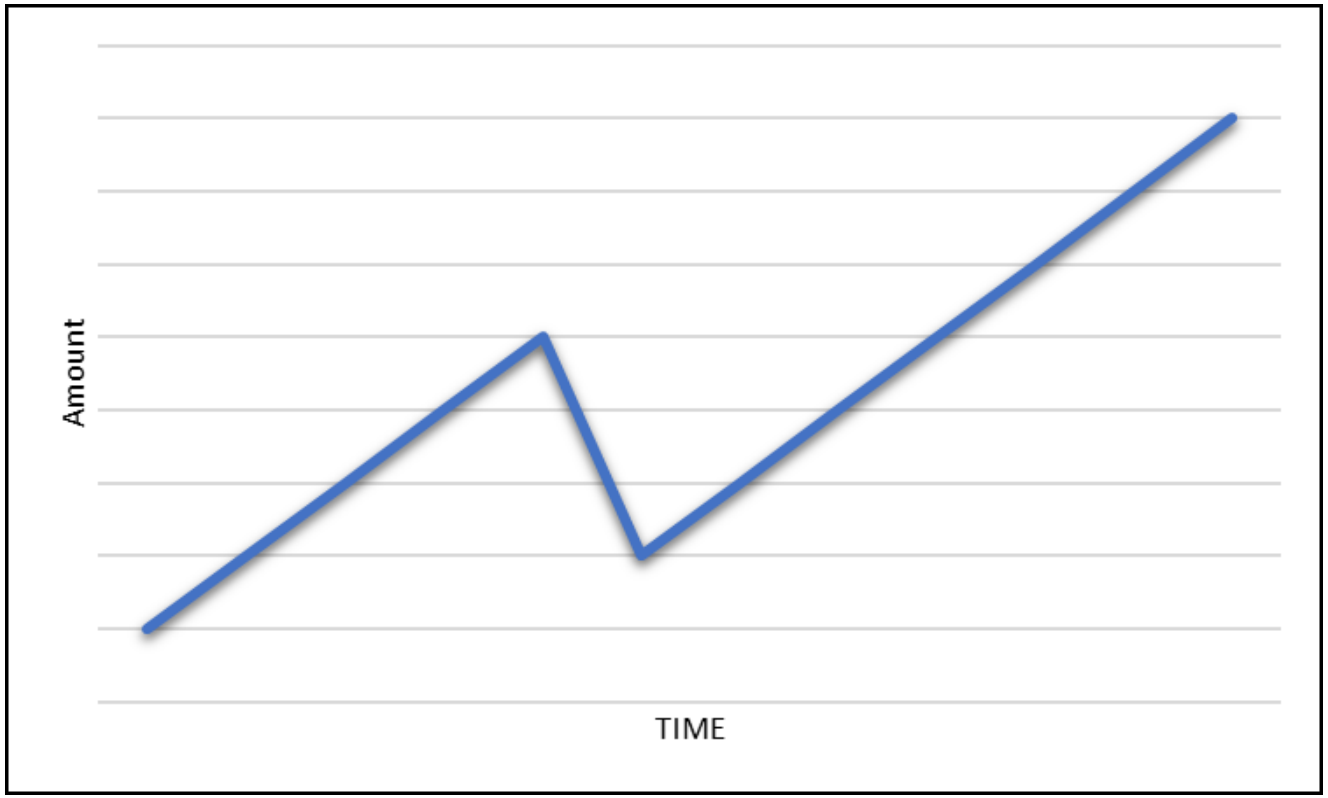
Graphical representation of information

Communicates insights from your data

Quickest way to spot trends and anomalies

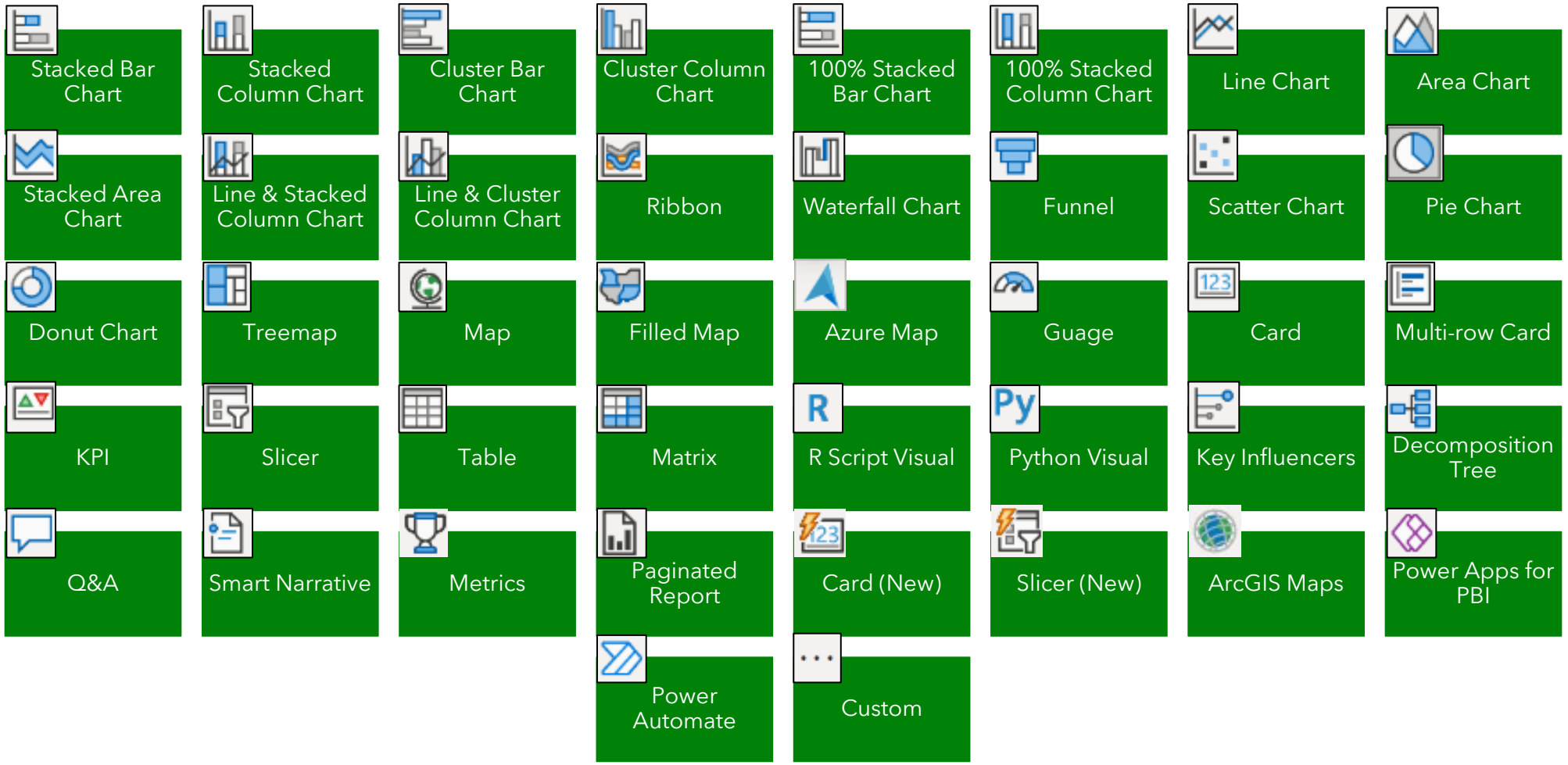
Visualizations – The New Paradigm

Time	Amount
Jan	100
Feb	200
Mar	300
Apr	400
May	500
Jun	200
Jul	300
Aug	400
Sep	500
Oct	600
Nov	700
Dec	800



Visualizations

(New ones being added regularly)

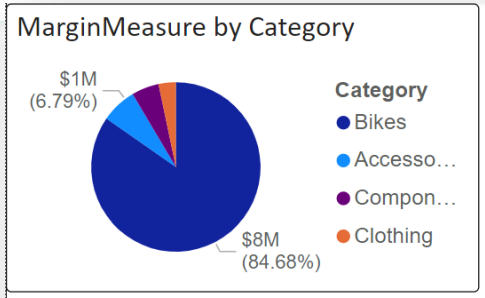


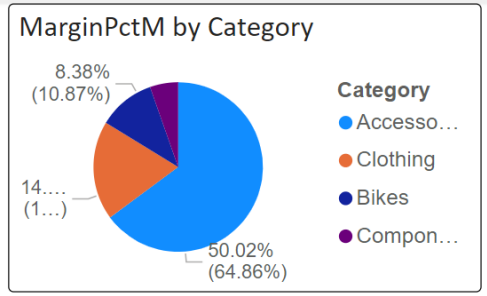
Visualizations and Profitability Example

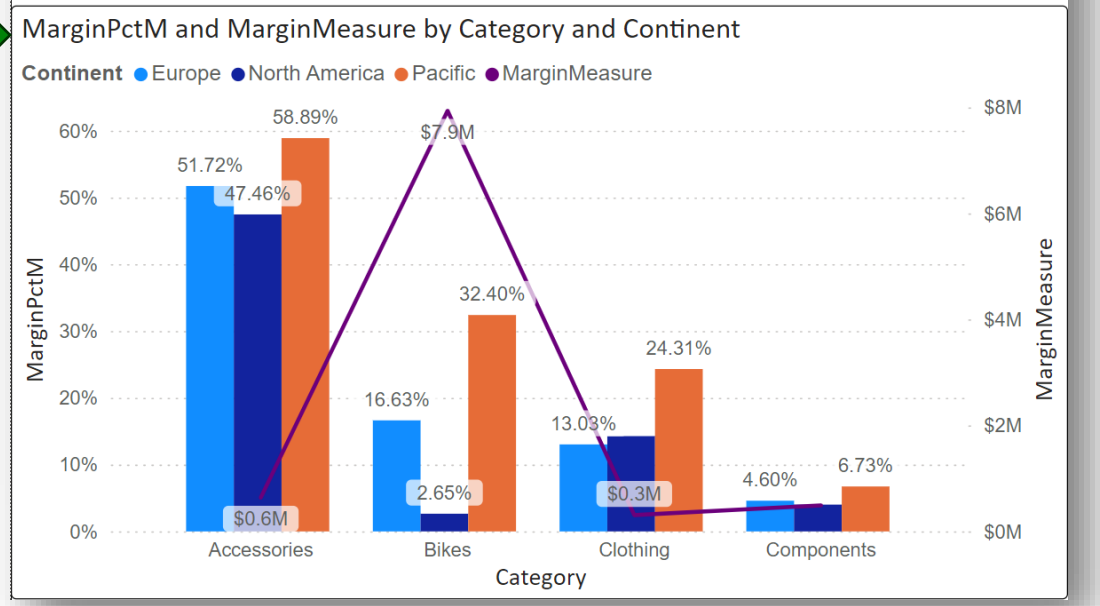
Category	MarginMeasure	MarginPctM
Bikes	$7,936,387	8.38%
Accessories	$636,318	50.02%
Components	$490,226	4.15%
Clothing	$308,891	14.57%
Total	**$9,371,822**	**8.53%**

Category / Continent	Accessories MarginMeasure	Accessories MarginPctM	Bikes MarginMeasure	Bikes MarginPctM	Clothing MarginMeasure	Clothing MarginPctM	Components MarginMeasure	Components MarginPctM	Total MarginMeasure	Total MarginPctM
Pacific	$95,782	58.89%	$3,297,244	32.40%	$27,518	24.31%	$13,700	6.73%	$3,434,244	32.23%
Europe	$169,636	51.72%	$2,857,434	16.63%	$52,665	13.03%	$88,264	4.60%		
North America	$370,900	47.46%	$1,781,709	2.65%	$228,708	14.27%	$388,261	4.01%		
Total	**$636,318**	**50.02%**	**$7,936,387**	**8.38%**	**$308,891**	**14.57%**	**$490,226**	**4.15%**		

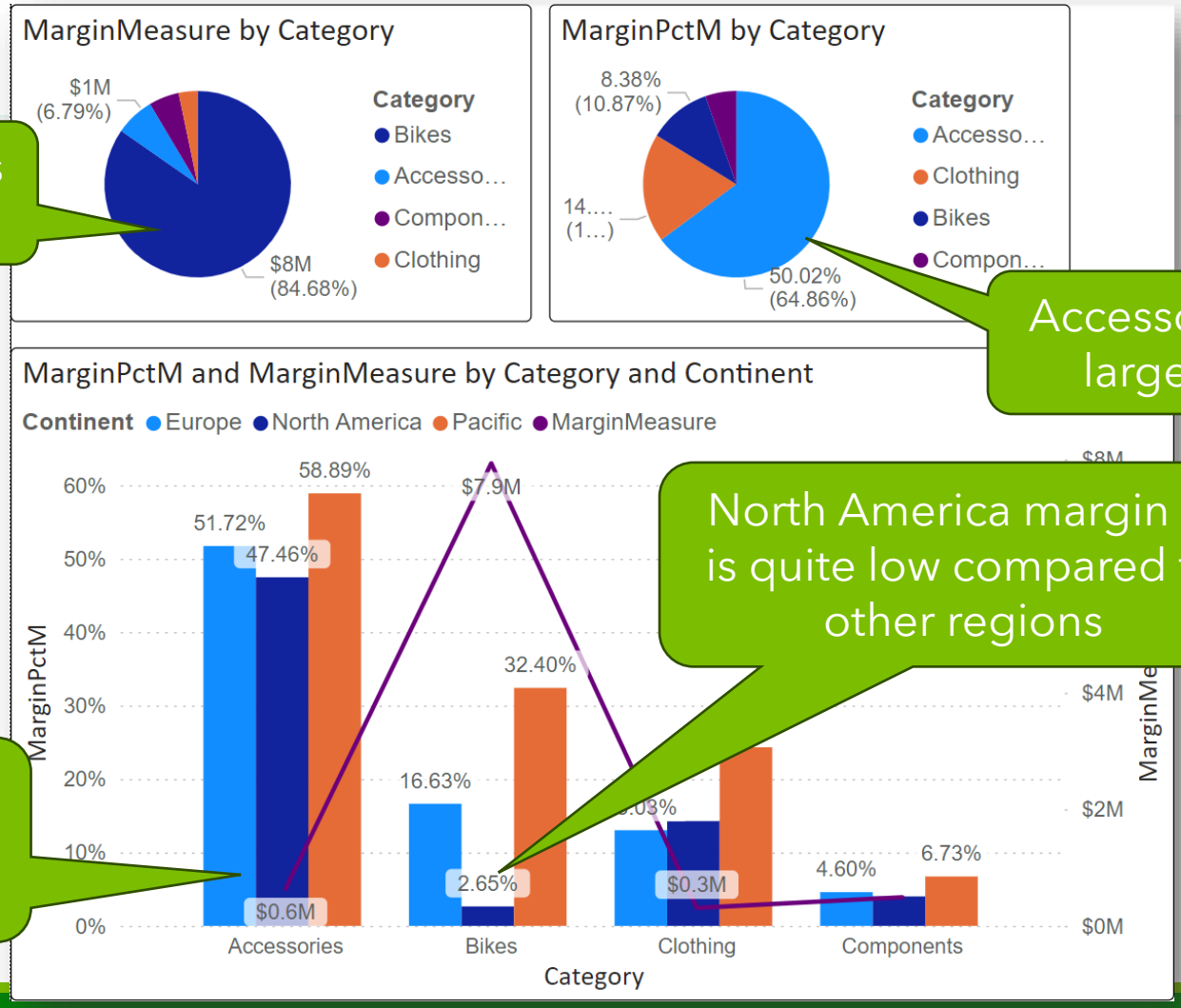
Category / Continent	Accessories MarginMeasure	Accessories MarginPctM	Bikes MarginMeasure	Bikes MarginPctM	Clothing MarginMeasure	Clothing MarginPctM	Components MarginMeasure	Components MarginPctM	Total MarginMeasure	Total MarginPc
Pacific	**$95,782**	**58.89%**	**$3,297,244**	**32.40%**	**$27,518**	**24.31%**	**$13,700**	**6.73%**	**$3,434,244**	**32.2**
	$88,790	61.64%	$3,404,815	37.88%	$26,900	33.84%	$1,693	8.26%	**$3,522,198**	**38.1**
Tsoflias	$6,991	37.59%	($107,571)	-9.07%	$618	1.83%	$12,007	6.56%	**($87,954)**	**-6.1**
Europe	**$169,636**	**51.72%**	**$2,857,434**	**16.63%**	**$52,665**	**13.03%**	**$88,264**	**4.60%**	**$3,167,999**	**15.9**
	$128,606	61.81%	$3,342,774	36.14%	$30,847	30.88%	$6,176	5.84%	**$3,508,402**	**36.3**
Valdez	$11,600	34.75%	($115,271)	-8.19%	$396	0.56%	$13,699	4.36%	**($89,577)**	**-4.9**
Saraiva	$13,770	34.54%	($158,762)	-5.21%	$9,341	8.66%	$30,582	4.78%	**($105,069)**	**-2.7**
Varkey Chudukatil	$15,660	33.54%	($211,307)	-6.08%	$12,082	9.63%	$37,808	4.40%	**($145,758)**	**-3.2**
North America	**$370,900**	**47.46%**	**$1,781,709**	**2.65%**	**$228,708**	**14.27%**	**$388,261**	**4.01%**	**$2,769,578**	**3.4**
	$228,595	61.91%	$4,198,694	35.79%	$83,845	39.47%	$9,333	6.40%	**$4,520,467**	**36.2**
Ansman-Wolfe	$4,624	34.53%	($4,142)	-0.15%	$6,626	11.90%	$28,496	6.33%	**$35,604**	**1.0**
Campbell	$5,054	34.47%	($84,422)	-2.71%	$7,862	13.30%	$30,825	5.66%	**($40,681)**	**-1.0**
Saraiva	$2,244	35.17%	($49,926)	-2.64%	($3,203)	-29.73%	($2,814)	-1.58%	**($53,699)**	**-2.5**
Total	**$636,318**	**50.02%**	**$7,936,387**	**8.38%**	**$308,891**	**14.57%**	**$490,226**	**4.15%**	**$9,371,821**	**8.5**







Our Profitability Example



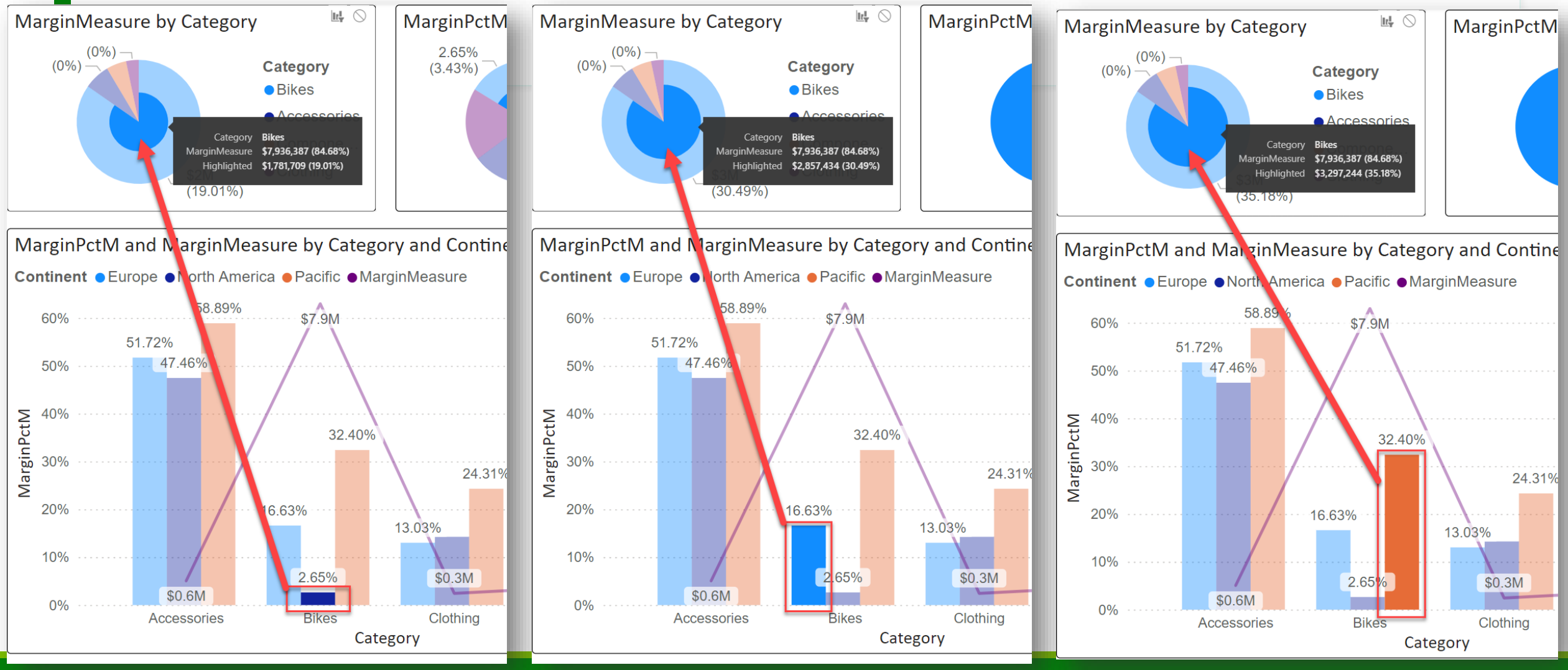
Bikes represents largest dollars

Accessories represent largest margin %

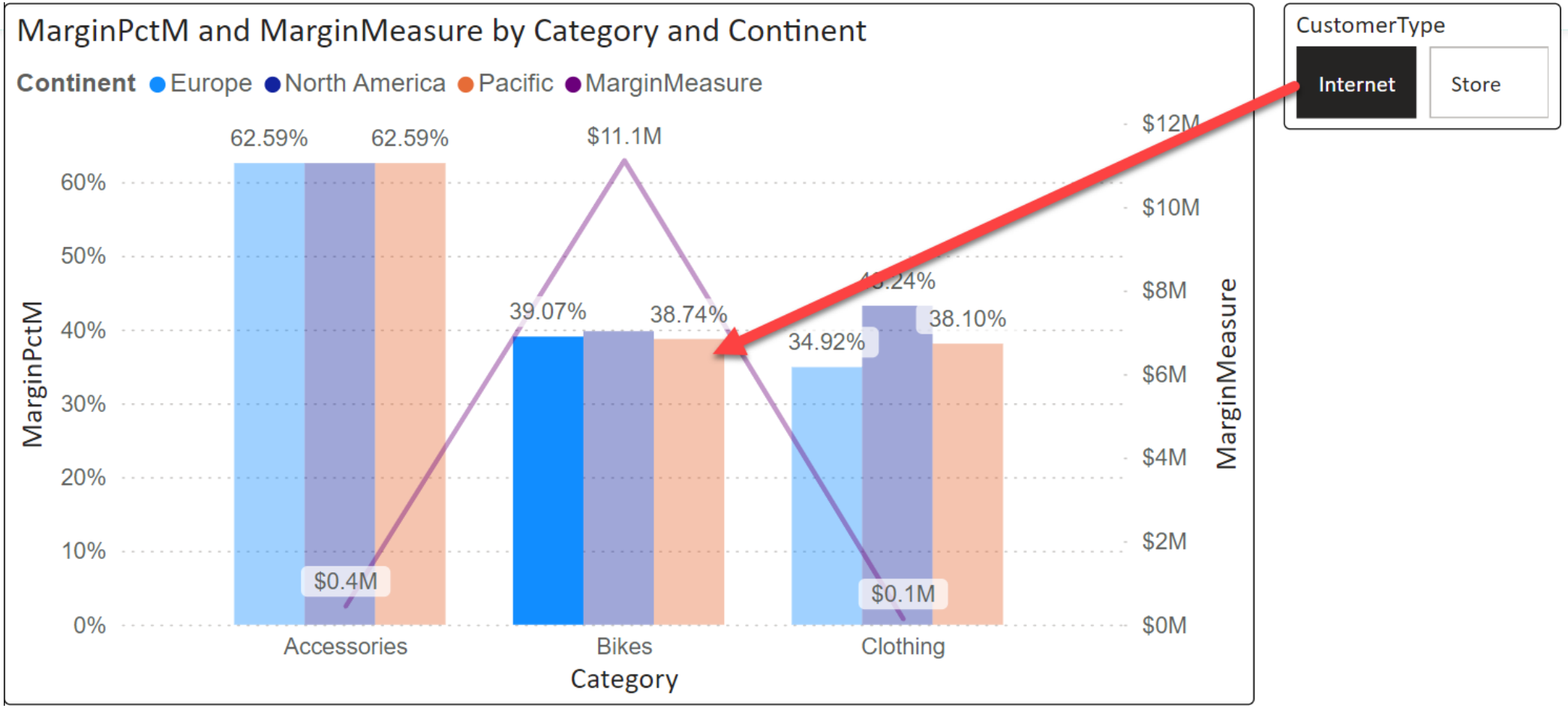
North America margin % is quite low compared to other regions

Shows the high % doesn't represent a lot of $

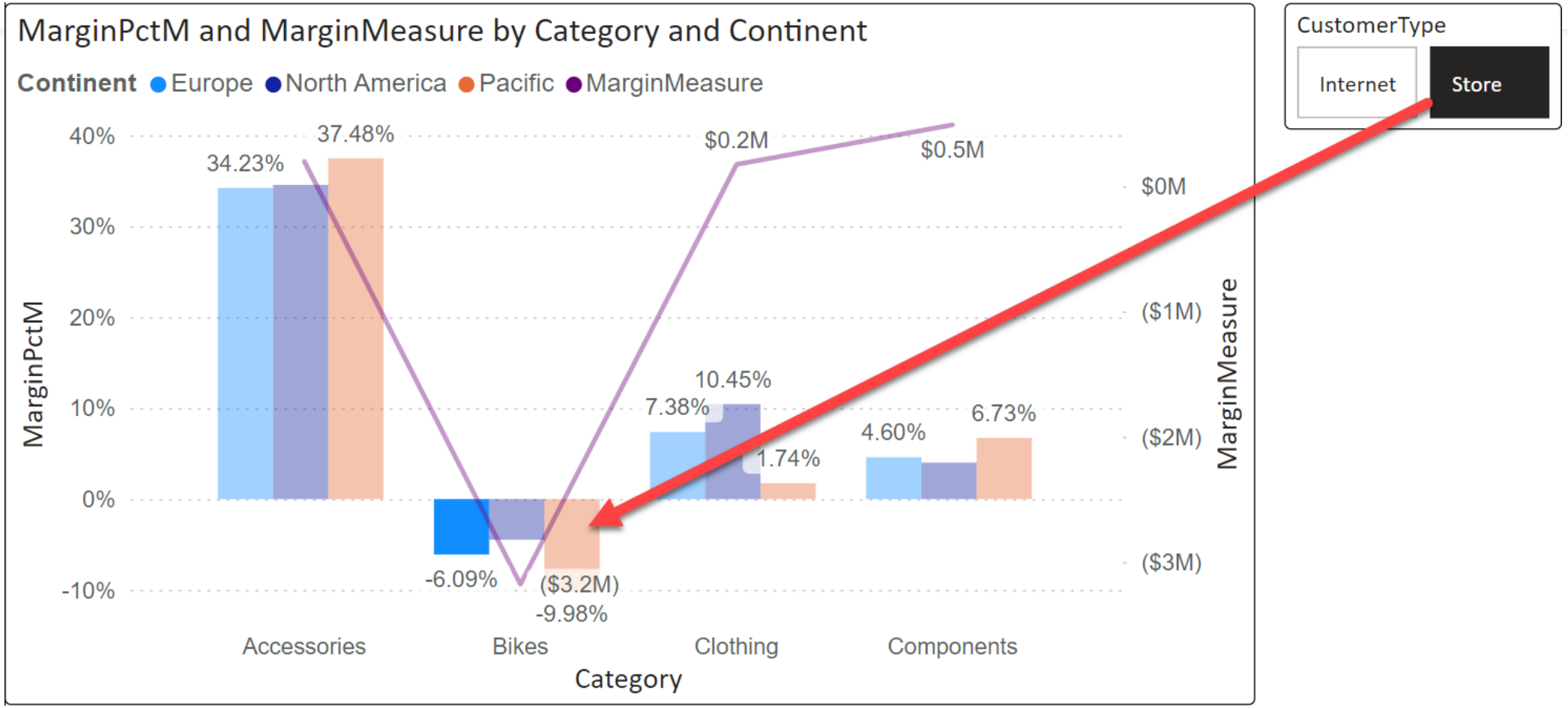
Interactivity & Tooltips



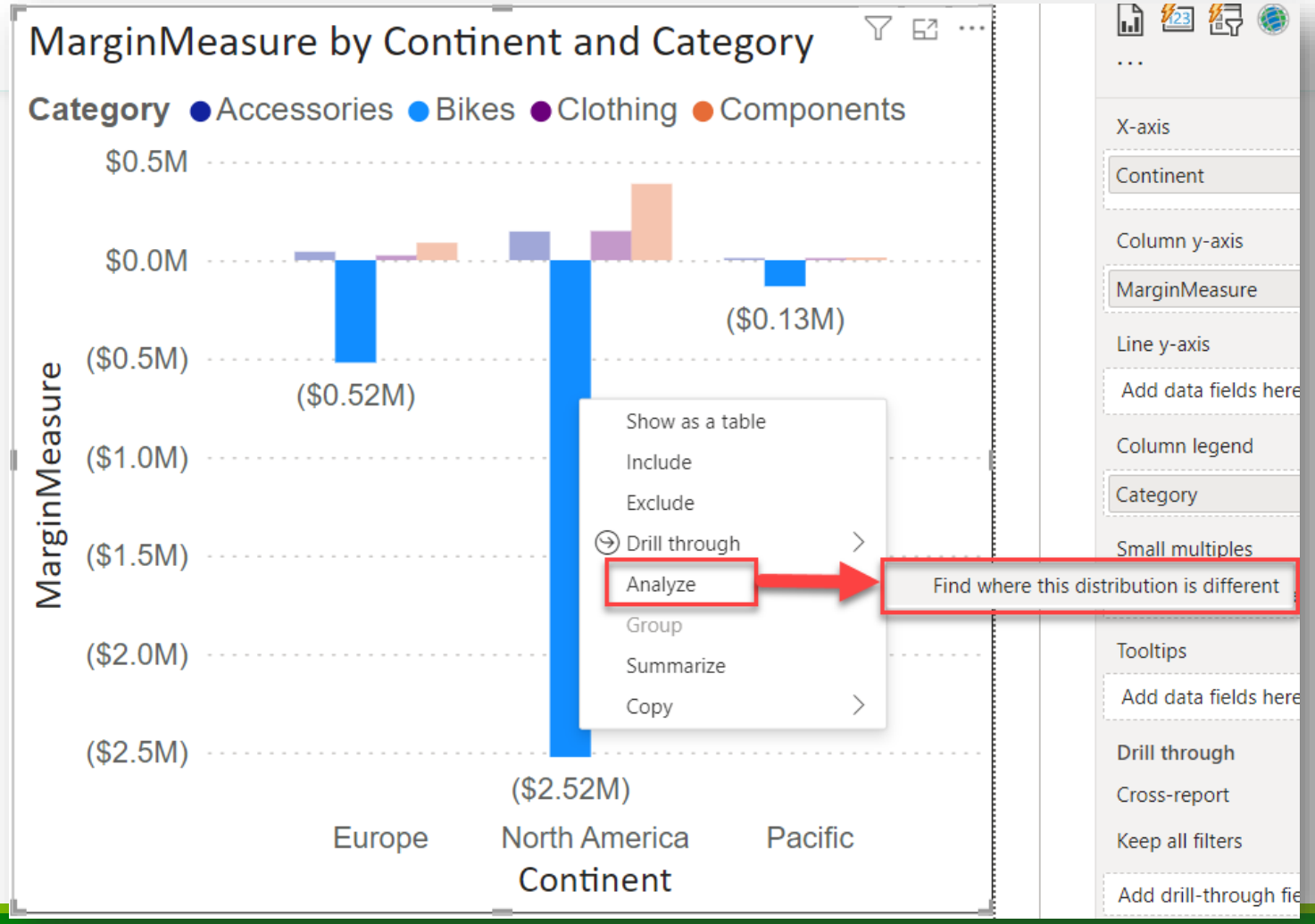
Using Slicers – Internet Customers



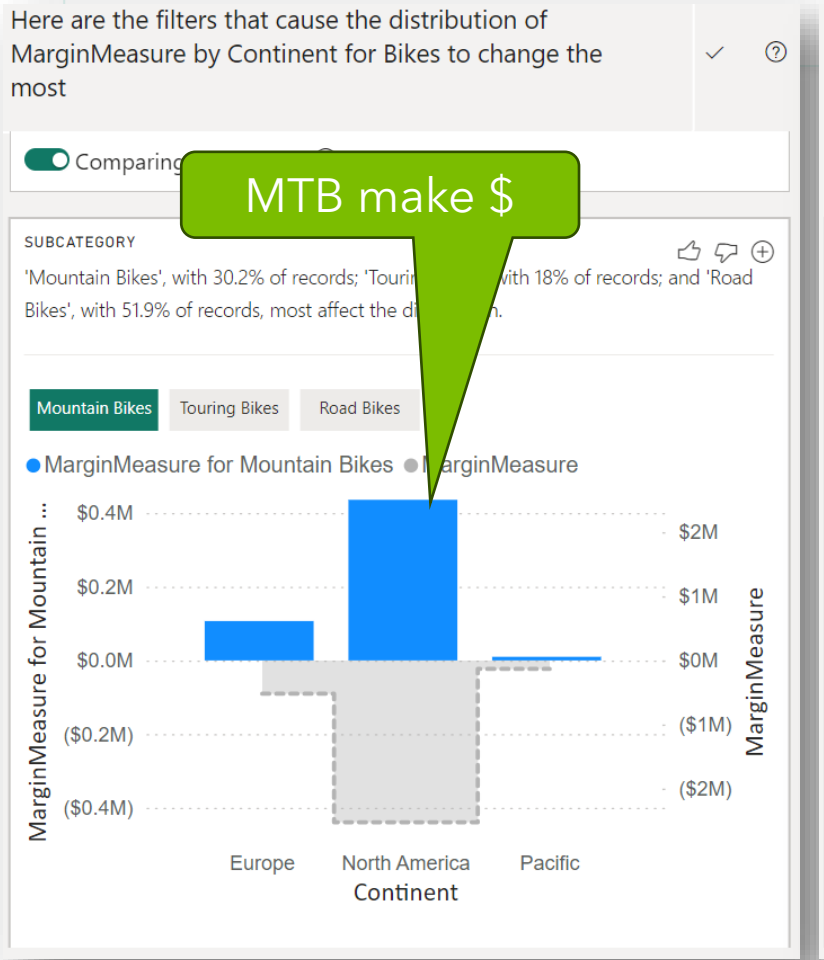
Using Slicers – Store Customers

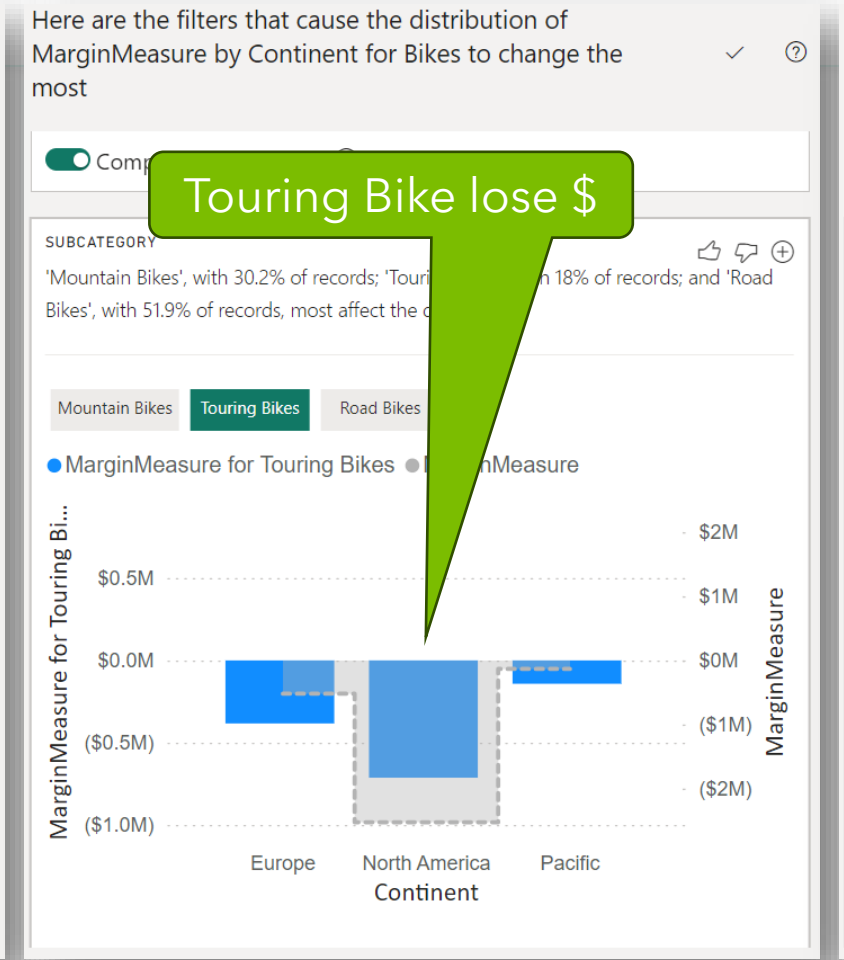


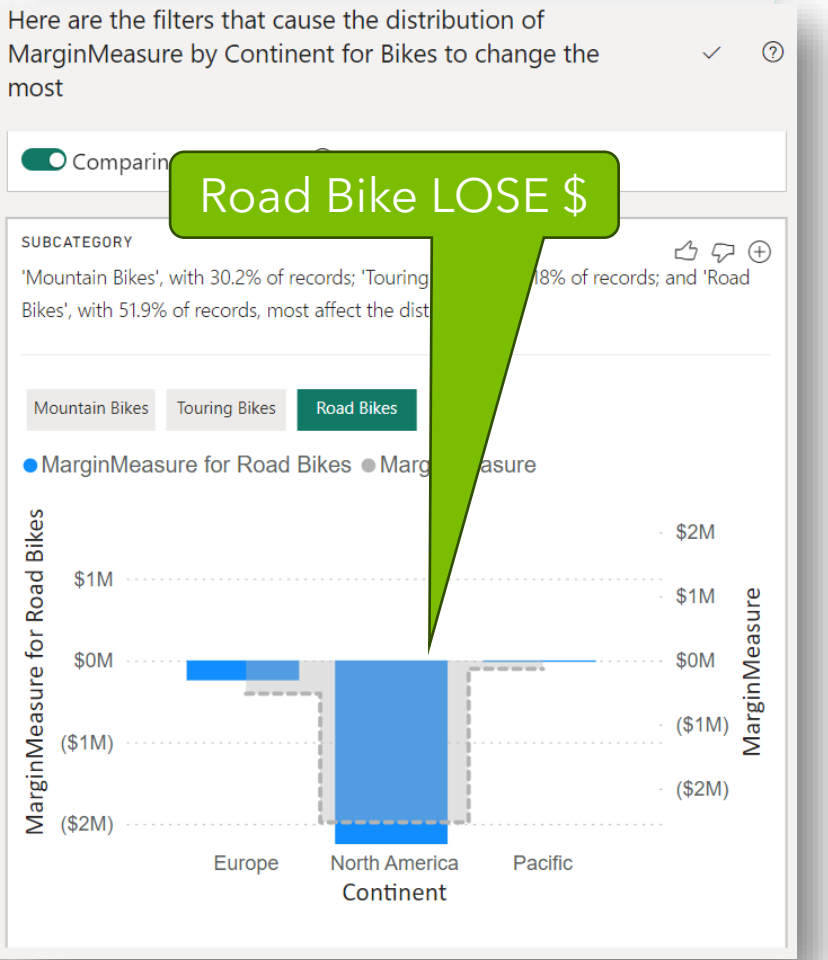
Use the Tools Built-In Analysis



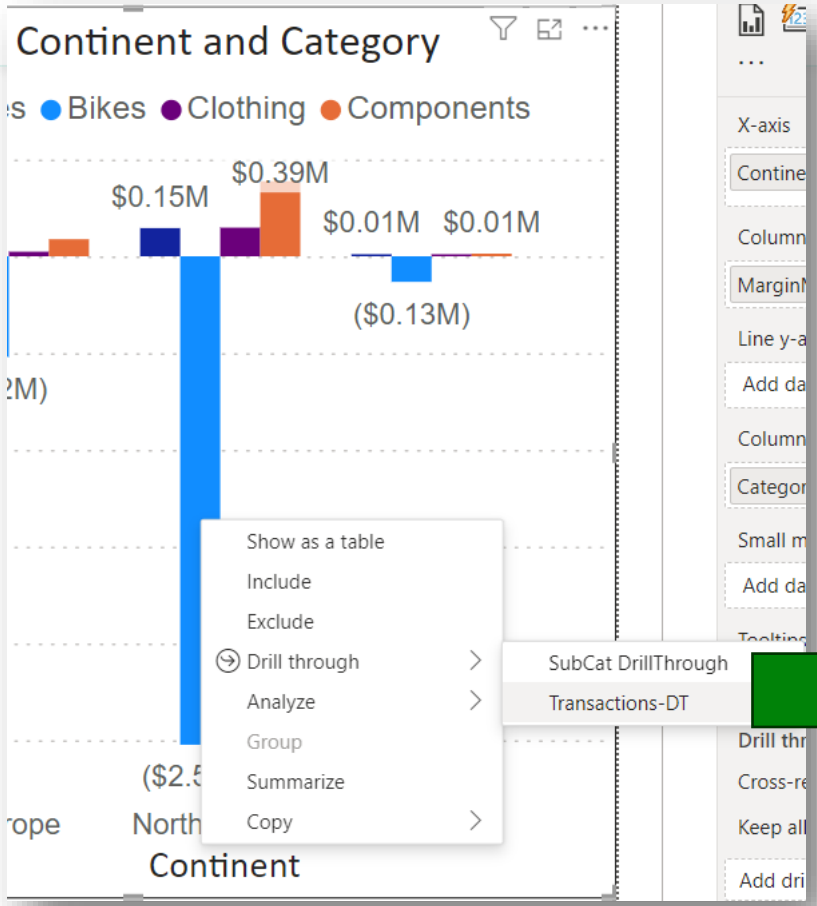
Analysis Results







Drill Through to Details



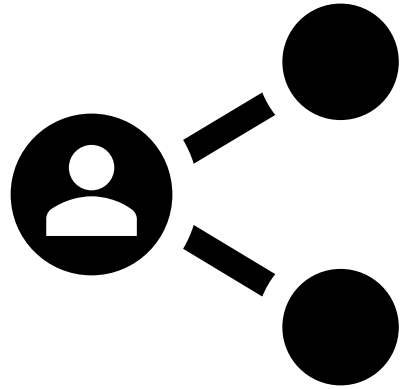
Transaction Details

CustomerID	CustomerName	Category	Subcategory	Order Qty	UnitPrice	GrossSales	DiscountAmt	NetSalesAmt	MarginMeasure	MarginPctM
676	Better Bike Shop	Bikes	Mountain Bikes	10	14,234.96	$20,324.92	$0.00	$20,325	$1,274	6.27%
117	Pedals Warehouse	Bikes	Road Bikes	2	1,294.25	$1,294.25	$0.00	$1,294	($77)	-5.96%
442	Original Bicycle Supply Company	Bikes	Mountain Bikes	13	10,139.97	$26,354.93	$0.00	$26,355	$1,652	6.27%
227	Health Spa, Limited	Bikes	Road Bikes	41	11,148.77	$26,116.84	$0.00	$26,117	($1,983)	-7.59%
510	World Bike Discount Store	Bikes	Road Bikes	1	419.46	$419.46	$0.00	$419	($67)	-16.03%
397	Capable Sales and Service	Bikes	Mountain Bikes	12	12,194.96	$24,374.92	$0.00	$24,375	$1,527	6.27%
146	Latest Sports Equipment	Bikes	Mountain Bikes	7	10,139.97	$14,189.95	$0.00	$14,190	$889	6.27%
511	Wheel Gallery	Bikes	Road Bikes	6	4,280.13	$4,699.59	$0.00	$4,700	($303)	-6.45%
646	Yellow Bicycle Company	Bikes	Mountain Bikes	3	6,089.98	$6,089.97	$0.00	$6,090	$382	6.27%
514	Retail Mall	Bikes	Road Bikes	47	9,421.27	$29,544.27	$0.00	$29,544	($2,272)	-7.69%
504	Historic Bicycle Sales	Bikes	Mountain Bikes	3	4,064.99	$6,104.98	$0.00	$6,105	$383	6.27%
200	Basic Bike Company	Bikes	Road Bikes	10	5,154.93	$7,288.10	$0.00	$7,288	($458)	-6.28%
119	Red Bicycle Company	Bikes	Mountain Bikes	3	4,064.99	$6,089.98	$0.00	$6,090	$382	6.27%
618	Seventh Bike Store	Bikes	Road Bikes	6	2,133.17	$2,972.09	$0.00	$2,972	($346)	-11.65%
8	Requisite Part Supply	Bikes	Road Bikes	3	874.79	$2,624.38	$0.00	$2,624	($30)	-1.13%
	st Bike Store	Bikes	Road Bikes	9	2,588.51	$5,141.14	$0.00	$5,141	($433)	-8.43%
	d Catalog Store	Bikes	Mountain Bikes	7	6,074.98	$14,174.96	$0.00	$14,175	$888	6.27%
	perb Sales and Repair	Bikes	Road Bikes	11	5,574.39	$7,252.22	$0.00	$7,252	($582)	-8.03%
20	Separate Parts Corporation	Bikes	Road Bikes	13	5,596.47	$7,729.64	$0.00	$7,730	($588)	-7.60%
480	General Bike Corporation	Bikes	Road Bikes	3	838.92	$1,258.38	$0.00	$1,258	($202)	-16.03%
491	Area Bike Accessories	Bikes	Road Bikes	13	5,574.39	$10,273.98	$0.00	$10,274	($616)	-6.00%
423	Bike Rims Company	Bikes	Road Bikes	20	6,413.30	$11,938.02	$0.00	$11,938	($1,073)	-8.99%
				63221	18,621,945.11	$56,840,296.14	$371,163.66	$56,469,132	($2,524,601)	-4.47%

What We Learned From Visualizations

- Tells the story of our data
- Easier to view vs. a table of data
- Provides ability to look at data without adding visuals
- Built-In Analysis features provide quick insights
- Ability to look at detail data helps find issues

Share Data Product



Benefit of Sharing the Data Product

- Provides other users access to data model
- No need to recreate data model
- No need to copy & paste into Excel
- Ability to create your own visuals
- Reduces the silos of data in an organization

What is Needed to Share Data Product

Location where others can access

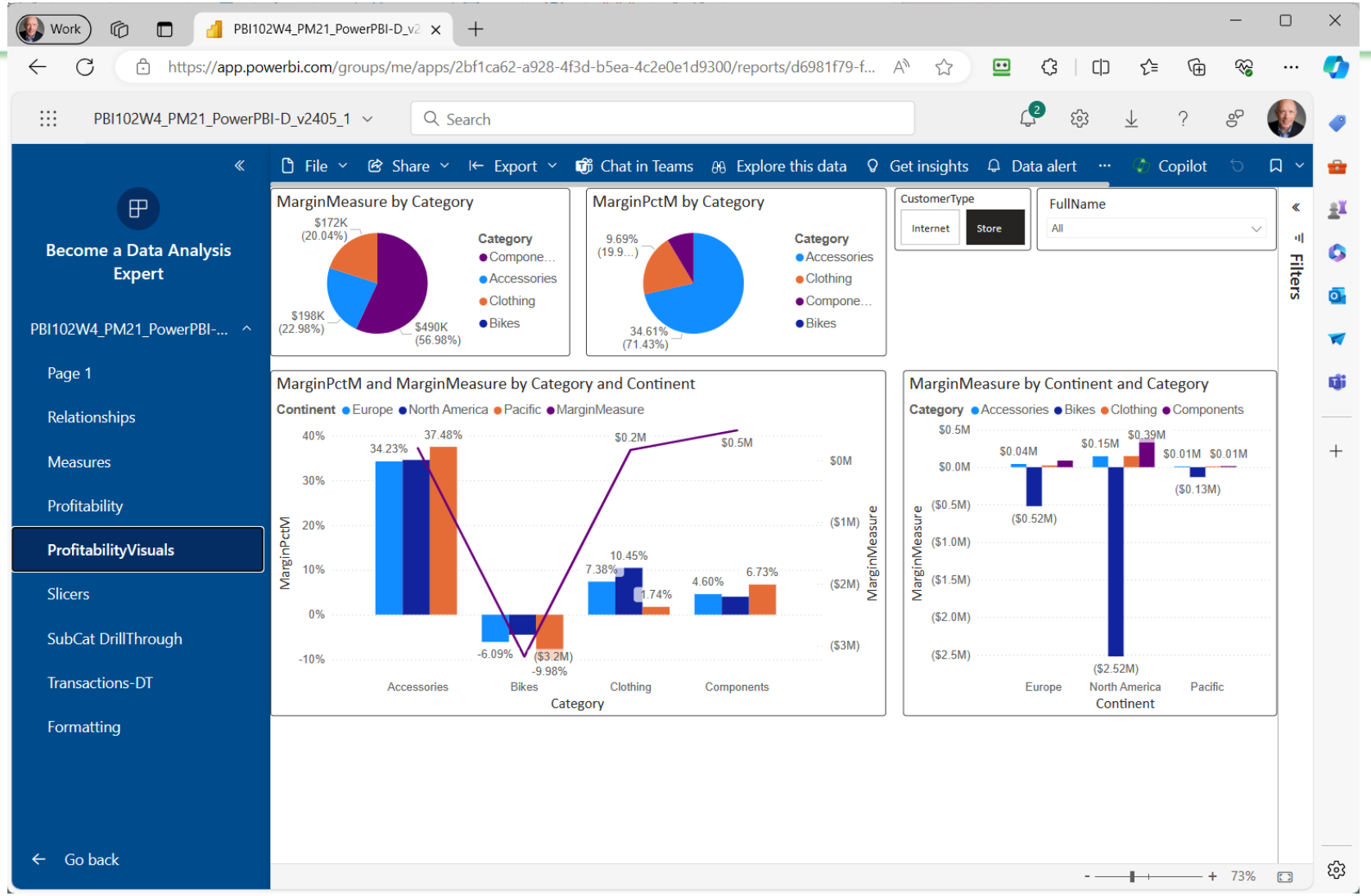
- Internal web service
- External web service

Setup security parameters

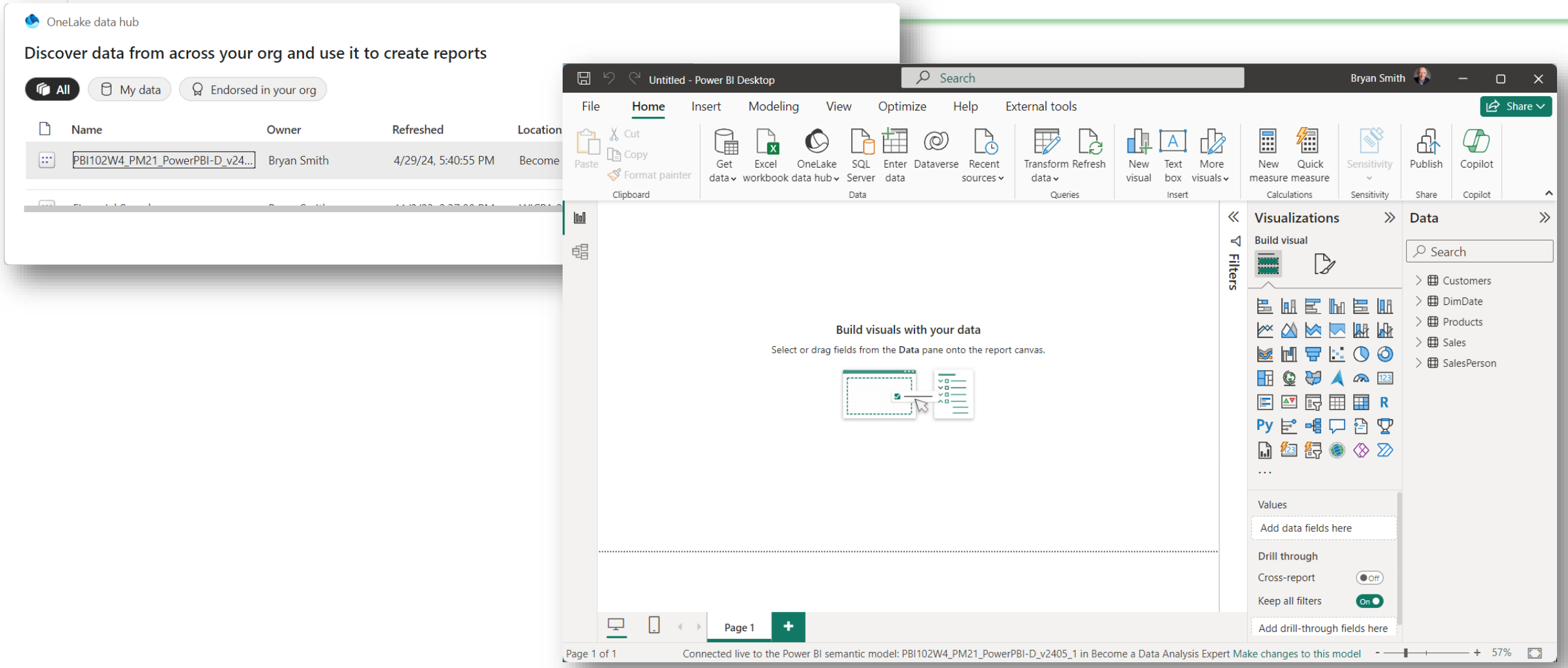
- Location security
- Table security
- Row level security

App to access the shared data product

Shared Report



Connect to Data and Build Your Own Visuals



Wrap-Up

Presentation Wrap-Up

- Data Analysis is a *process*
- The Data Analysis Process Flow provides a roadmap to successful analysis
- Data is a BIG / Important step
- Data modeling builds relationships between tables and creates measures
- Visualizations communicate insights into our data
- Sharing our data product allows everyone to take advantage of our work

Questions?

Thank You!!

Bryan L. Smith, CPA.CITP, CISA

bsmith@datasmithpro.com

This course and Bryan L. Smith are independent of Microsoft and are neither affiliated with, nor authorized, sponsored, or approved by Microsoft Corporation.

All trademarks are the property of their respective owners.

ONE FINAL NOTE