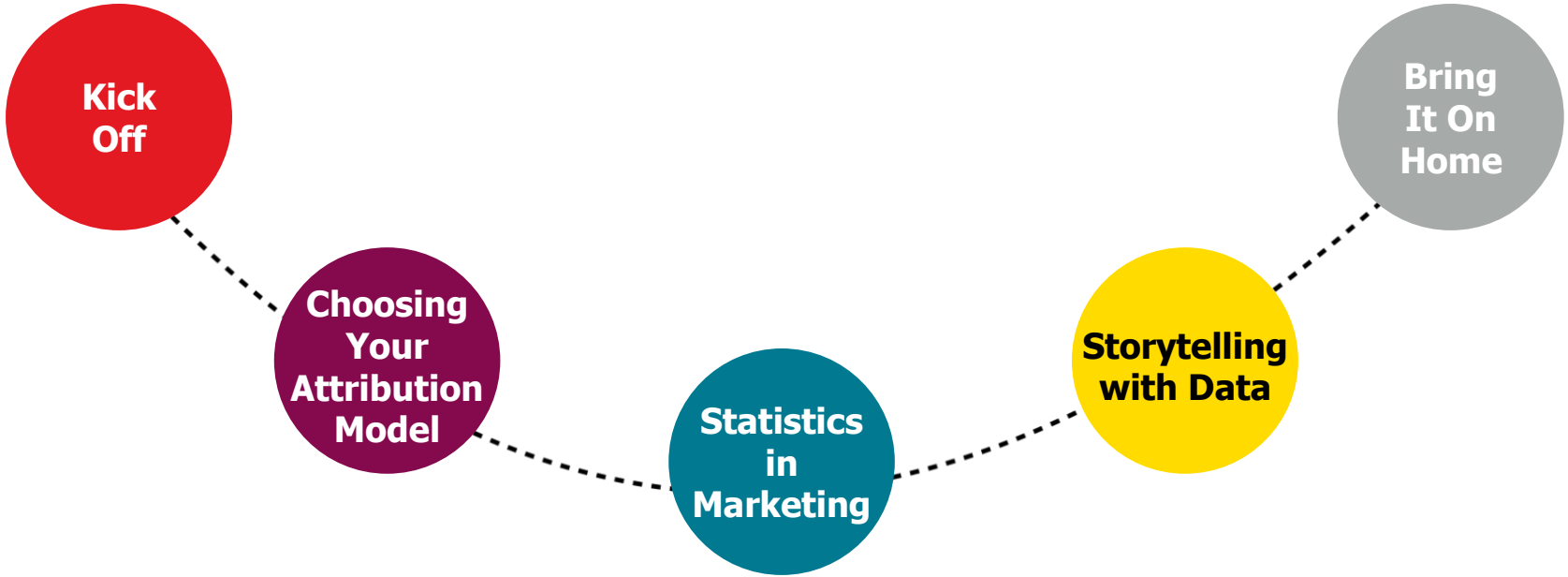
Advanced Marketing Analytics

Instructor: Alex Sierra, CEO Sigma Ridge

Our Roadmap



Alexander Sierra

CEO, Sigma Ridge

Alexander Sierra has a diverse background. He has a bachelor's degree in Aerospace Engineering from the University of Florida, A master's in Finance from Harvard University, and an MBA from the University of Massachusetts. Over 23 years of experience as a leader in Consulting, Sales & Marketing. Alex has managed teams of over 130 direct reports in 13 different countries and Marketing budgets of over 25 million dollars a year. He has been able to achieve year-over-year growth for his clients from 25% to 120%.

His consulting practice **-Sigma Ridge-** was a spinoff from the Harvard University Consulting club where he works with companies like Cisco, Dell, and some of the largest fashion companies in the US.

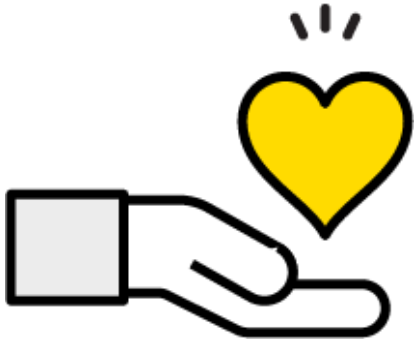
Rules of the Road

- Be here now and share your experiences.
- Don’t wait for “Q&A” - our time together is a facilitated discussion.
- You’ll take away as much as you put in… so put the work away! There will be breaks throughout the day.
- Have fun!

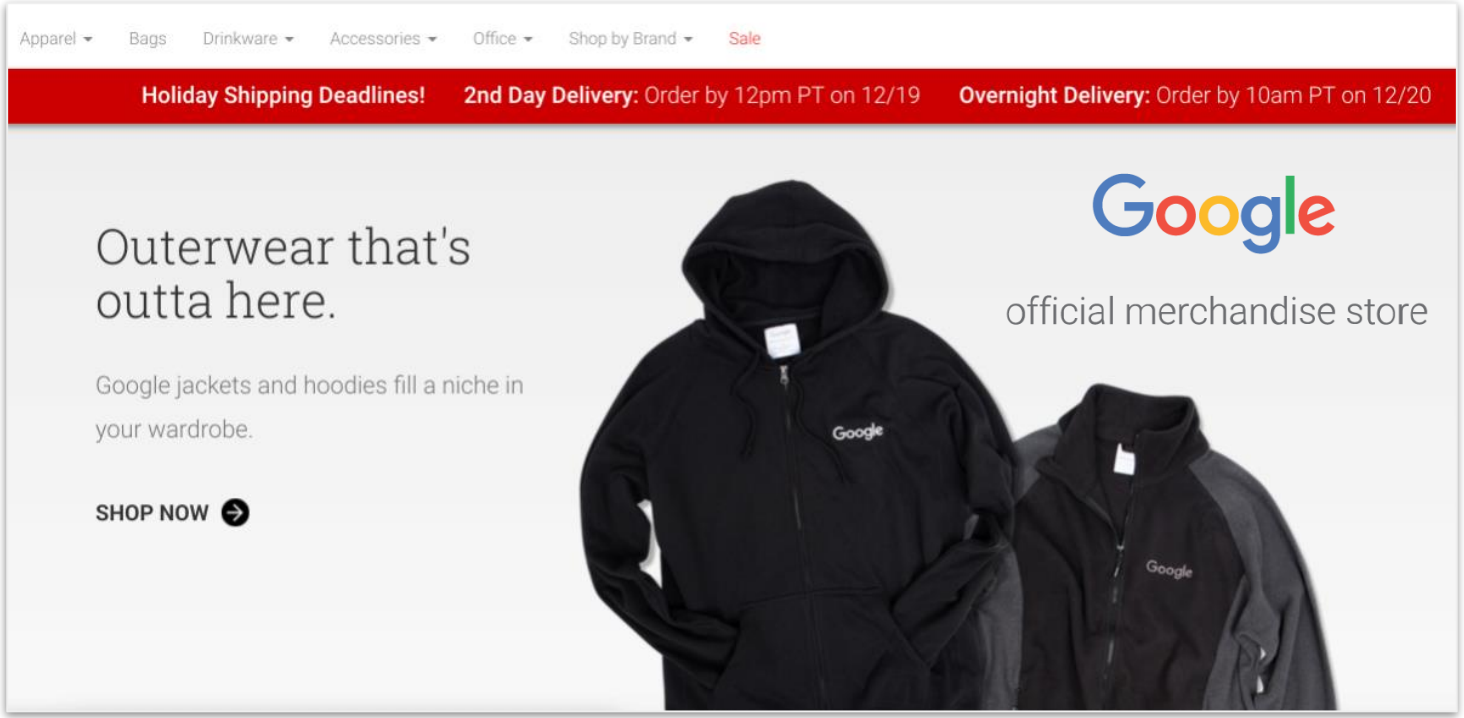
Who’s in the room?

Let’s get to know each other!

- Your name
- Your experience as a marketer
- One thing you’re hoping to learn about during this course



Your mission



Your stakeholders

Claire, Owner

Show me needles in the haystack on how to improve my business.

Maggie, Marketer

I have a monthly budget and need to use it to acquire as many customers as possible.

Clark, Product Manager

I develop quality merchandise and to make sure there is enough supply for the demand.

Tony, Developer

I keep the site optimized to ensure it's easy to browse products and complete purchases.

Choosing Your Attribution Model

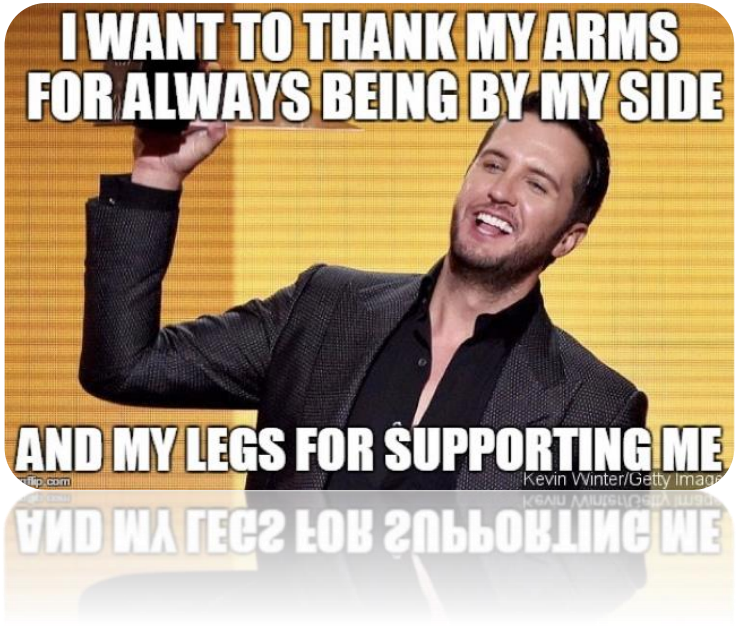
Solo Exercise:

Attributing Your Success

10 minutes

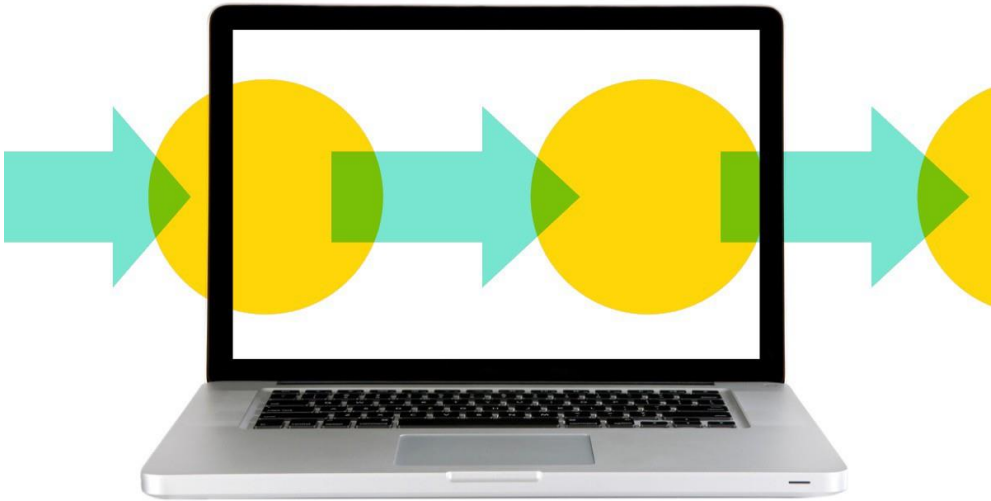
- Part 1: What, in your life, led you to your current position? Think about the *people* and *institutions* that led you there from childhood to recent times, and draw it out on a rough timeline.

- Part 2: Due to a bizarre new regulation, you have to distribute $10,000 of your salary to the parties that led to your current employment. Looking at your timeline, how would you distribute this income? Write the % you would give to each party underneath them on the timeline.



Attribution

- The rule, or set of rules, that determines how credit for conversions is assigned to various touchpoints before a conversion
- Without some form of attribution model, you are not taking into account the full value of your marketing efforts



Attribution: Who Gets The Credit?

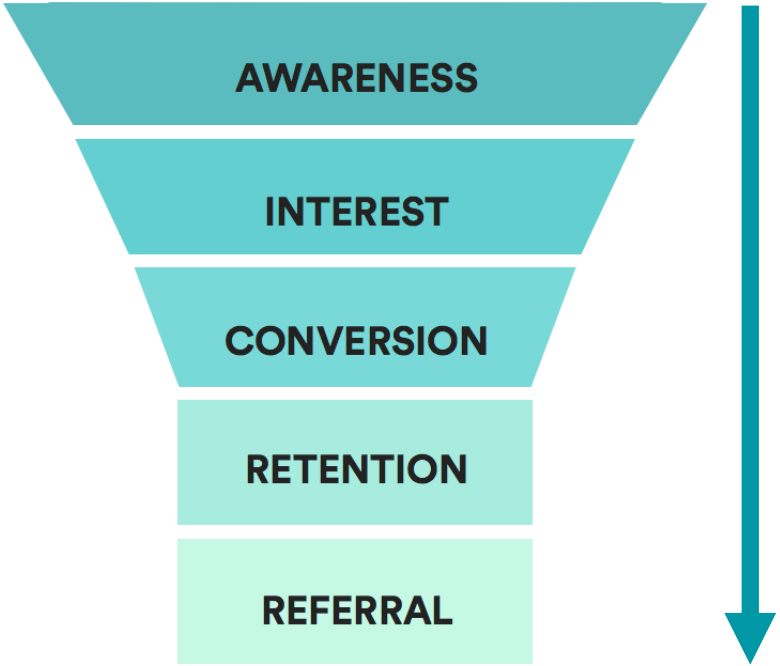
SMS

Web Retargeting

Social

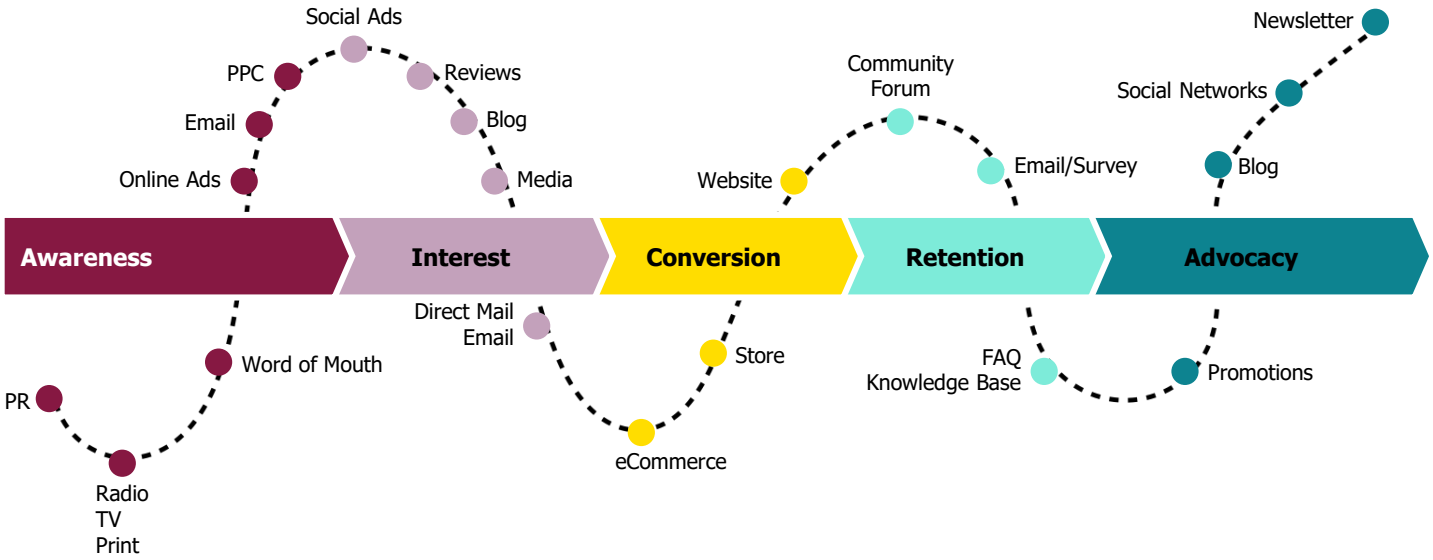
Email Cart Reminder

Marketing objectives are typically rooted in the funnel.



- Brand Marketing:
  "I want to make people aware and excited about what we have to offer!"
- Customer Acquisition:
  "I want to grow my bottom line by gaining new customers."
- Retention and Loyalty:
  "I want to engage my customers and get them to buy more!"

Attribution

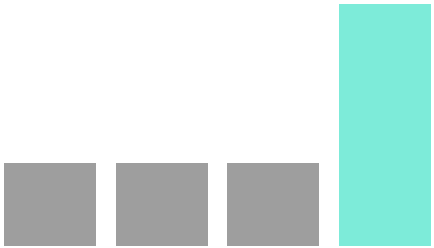


Four things to know about attribution

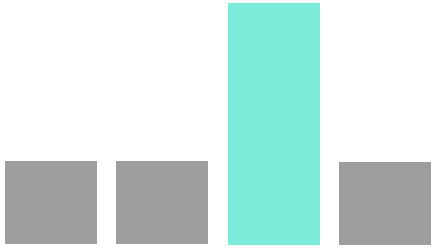
1. No channel is **independent**
2. There is **no perfect** attribution model
3. Attribution is art **& Science**
4. There is **no complete** data set

Bring Out The Models

First interaction



Last interaction



Last non-direct interaction

Position-based

Time decay

Linear

Group Exercise:

Choose the Model

30 minutes

Visit: http://bit.ly/MarketingModels
or
https://www.sigmaridge.com/general-assembly/advanced-marketing-analytics

Review the handout for your assigned attribution model.

Maggie, Marketer

"Why would I use each of those attribution models?"

Group Exercise:

Choose the Model

10 minutes

Visit: http://bit.ly/MarketingModels
or
https://www.sigmaridge.com/general-assembly/advanced-marketing-analytics

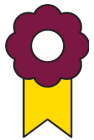
Prepare a 3-minute presentation that covers:

- A quick overview of how that model works
- Why the Google Merchandise store would (or would not) want to use that model

Maggie, Marketer

"Why would I use each of those attribution models?"

Choosing your model



Campaign goal

Awareness

- First interaction
- Last interaction
- Position-based

Demand generation

- First interaction
- Linear
- Position-based

Conversions

- Last interaction
- Last non-direct Interaction
- Time decay

Engagement

- First interaction
- Position-based

Choosing your model

Sales cycle

Last interaction

Linear

Position-based

Shorter — **Longer**

First interaction

Last non-direct

Time decay

Choosing your model

Tech stack

Less complex

First interaction

Linear

Position-based

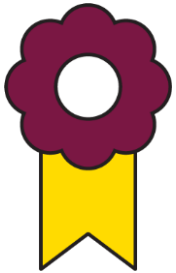
Last interaction

Last non-direct

Time decay

More complex

Choosing your model



Campaign goal

What am I trying to measure and accomplish?

- Awareness/Demand, New Product or Brand, single point campaigns: FI
- Conversion, easy & default for Google Analytics: LI
- Interest marketing: LN-DI
- Awareness & Conversion: PB
- Promotion: TD
- Awareness TO conversion: Linear

Sales cycle

How long is my sales cycle?

- Short Sales Cycle: LI, PB (U), FI
- Long sales Cycle: LN-DI, TD, Linear

Tech stack

What data am I able to gather and analyze?

- Not a lot of data: FI, LI
- A lot of Data: LN-DI, PB (U), TD, Linear

Trivia!

Which model is best?

You market computing infrastructure to large enterprises. You spend a long time nurturing leads over a variety of touchpoints - email, paid search, events.

A. Position-based

B. First interaction

C. Time decay

D. Linear

+10 points!

Which model is best?

You market computing infrastructure to large enterprises. You spend a long time nurturing leads over a variety of touchpoints - email, paid search, events.

C. Time decay

+10 points!

Which model is best?

You run an e-commerce site that sells high-quality organic makeup at a low price. Quickly getting customers to convert is your top priority.

A. Last interaction

B. Position-based

C. Time decay

D. Linear

+10 points!

Which model is best?

You run an e-commerce site that sells high-quality organic makeup at a low price. Quickly getting customers to convert is your top priority.

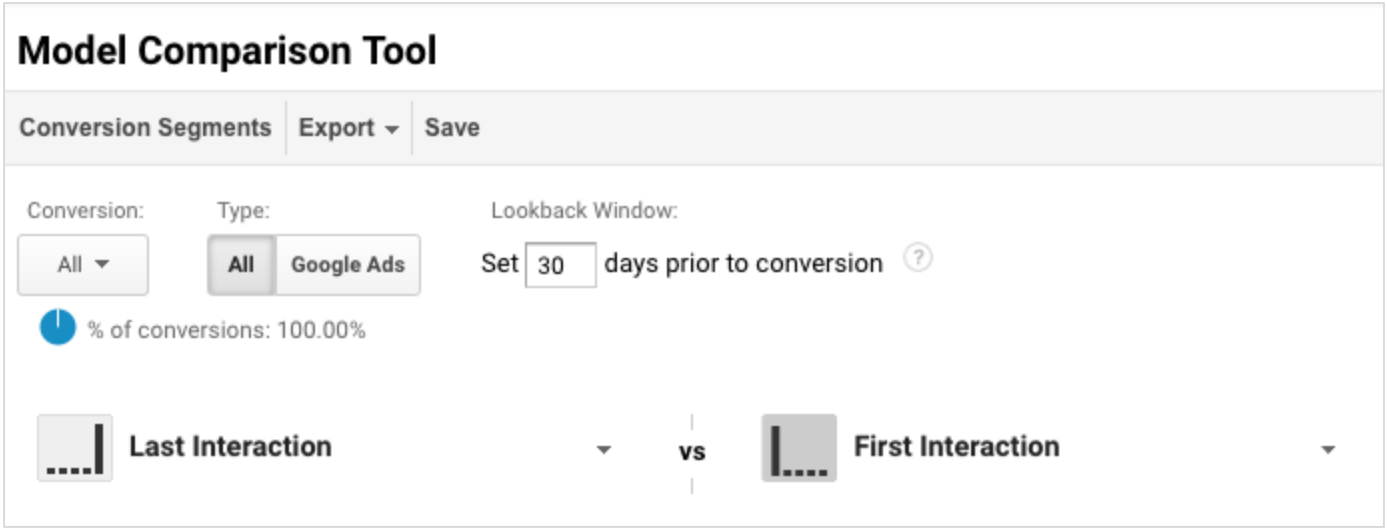
A. Last interaction

Real Cases:
Case study

5 minutes

Share a real-life example of how a company chose an attribution model for a given campaign, or to use at a high level. Explain why they chose to use that model and the benefits they got from using it.

The Model Hack



Guided Walk-Through:
Google Analytics Model Comparison Tool

10 minutes

Let's review the key features of the Google Analytics Model Comparison Tool.
Visit:
https://analytics.google.com/analytics.
https://analytics.google.com/analytics/web/demoAccount

We'll use the Google Analytics demo account data for this exercise.
Already have a Google Analytics account? You're all set!
Don't have a Google Analytics account? Create one now! (It's free and you can connect it to an existing Google account.)

Guided Walk-Through:

Google Analytics Model Comparison Tool

10 minutes

Access the demo account

To access the demo account, click the *ACCESS DEMO ACCOUNT* link at the end of this section. When you click the link:

- If you already have a Google account, you are prompted to log in to that account.
- If you do not have a Google account, you are prompted to create an account and then log in.

When you click the *ACCESS DEMO ACCOUNT* link below, you agree to let Google perform one of two actions related to your Google account:

- If you already have a Google Analytics account, we will add the demo account to your Analytics account.
- If you do not have a Google Analytics account, we will create one for you in association with your Google account, and then add the demo account to your new Analytics account.

The demo account is available from the universal picker in Analytics where you select organization and account links.

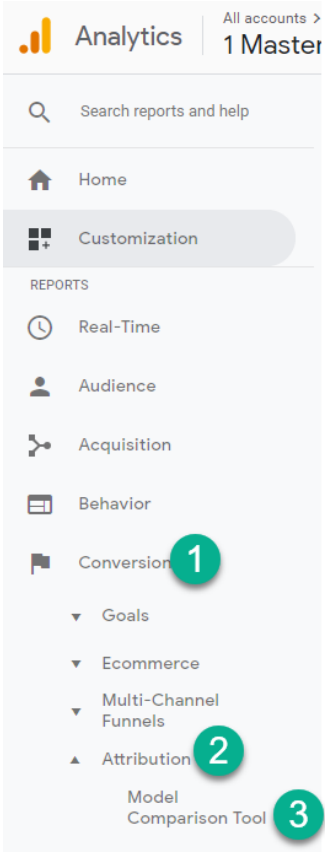
The demo account counts against the maximum number of Analytics accounts you are permitted to create under a single Google account. The current maximum for Google Analytics Standard is 100 Analytics accounts per Google account.

You can remove the demo account at any time.

ACCESS DEMO ACCOUNT

Guided Walk-Through:
Google Analytics Model Comparison Tool

10 minutes



In the left hand navigation, select:

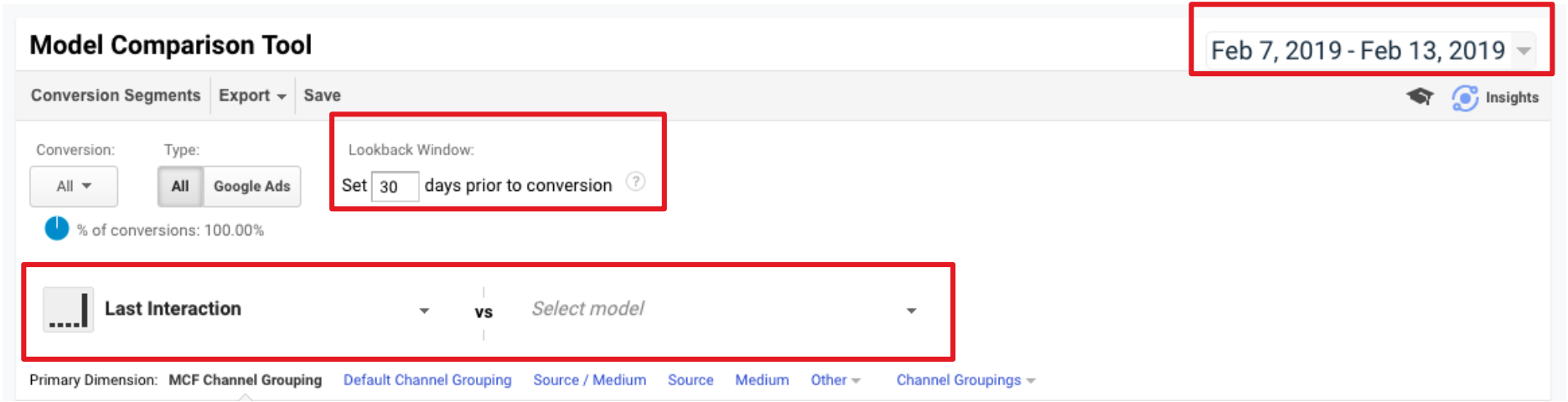
1-Conversions,

2- Attribution,

3- Model Comparison Tool.

Guided Walk-Through:
Google Analytics Model Comparison Tool

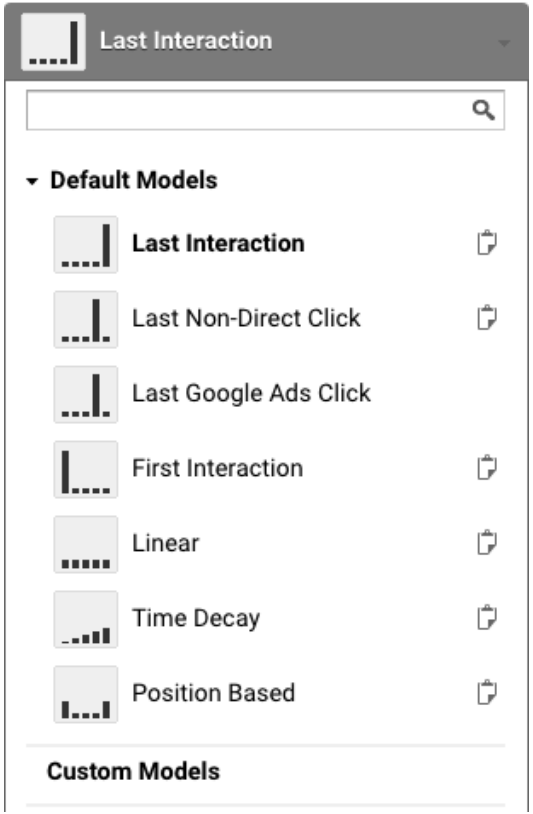
10 minutes



Guided Walk-Through:
Google Analytics Model Comparison Tool

10 minutes

- Compare up to three models at once
- Choose from 7 different models or a custom model



Guided Walk-Through:

Google Analytics Model Comparison Tool

10 minutes

Conversions & CPA

% change in Conversions (from Last Interaction)

MCF Channel Grouping	Spend (for selected time range)	Last Interaction		Linear		Position Based		% change in Conversions (from Last Interaction)	
		Conversions	CPA	Conversions	CPA	Conversions ↓	CPA	Linear	Position Based
1. Direct	–	1,297.00 (45.19%)	–	1,162.06 (40.49%)	–	**1,143.44** (39.84%)	–	-10.40%	-11.84%
2. Organic Search	–	846.00 (29.48%)	–	906.71 (31.59%)	–	**913.84** (31.84%)	–	7.18%	8.02%
3. Referral	–	350.00 (12.20%)	–	384.98 (13.41%)	–	**393.17** (13.70%)	–	9.99%	12.34%
4. Paid Search	$381.70	187.00 (6.52%)	$2.04	207.66 (7.24%)	$1.84	**210.30** (7.33%)	$1.82	11.05%	12.46%
5. (Other)	–	89.00 (3.10%)	–	98.43 (3.43%)	–	**98.35** (3.43%)	–	10.59%	10.51%
6. Other Advertising	–	50.00 (1.74%)	–	52.67 (1.84%)	–	**52.90** (1.84%)	–	5.33%	5.80%
7. Social Network	–	38.00 (1.32%)	–	43.83 (1.53%)	–	**44.20** (1.54%)	–	15.35%	16.32%
8. Display	–	13.00 (0.45%)	–	13.67 (0.48%)	–	**13.80** (0.48%)	–	5.13%	6.15%

Partner Exercise:

Make Your Case

5 minutes

Using Google Analytics, answer:

- Does organic search generate more conversion value when conversions are attributed on the **last interaction** or **first interaction model**?

Maggie, Marketer

"I want to use attribution modeling to understand how the SEO work, we recently launched, is paying off in terms of conversions."

Guided Walk-Through:

Google Analytics Model Comparison Tool

10 minutes

MCF Channel Grouping	Spend (for selected time range)	Conversions & CPA: Last Interaction		First Interaction		Position Based		% change in Conversions (from Last Interaction)	
		Conversions	CPA	Conversions	CPA	Conversions	CPA	First Interaction	Position Based
1. Direct	–	1,297.00 (45.19%)	–	961.00 (33.48%)	–	1,143.44 (39.84%)	–	-25.91%	-11.84%
2. Organic Search	–	846.00 (29.48%)	–	994.00 (34.63%)	–	913.84 (31.84%)	–	17.49%	8.02%
3. Referral	–	350.00 (12.20%)	–	447.00 (15.57%)	–	393.17 (13.70%)	–	27.71%	12.34%
4. Paid Search	$381.70	187.00 (6.52%)	$2.04	238.00 (8.29%)	$1.60	210.30 (7.33%)	$1.82	27.27%	12.46%
5. (Other)	–	89.00 (3.10%)	–	107.00 (3.73%)	–	98.35 (3.43%)	–	20.22%	10.51%
6. Other Advertising	–	50.00 (1.74%)	–	57.00 (1.99%)	–	52.90 (1.84%)	–	14.00%	5.80%
7. Social Network	–	38.00 (1.32%)	–	51.00 (1.78%)	–	44.20 (1.54%)	–	34.21%	16.32%
8. Display	–	13.00 (0.45%)	–	15.00 (0.52%)	–	13.80 (0.48%)	–	15.38%	6.15%

Show rows: 10 Go to: 1 1 - 8 of 8

Partner Exercise:

Make Your Case

5 minutes

Using Google Analytics, answer:

- Does organic search generate more **conversion value** when conversions are attributed on the **last interaction** or **first interaction model**?

Maggie, Marketer

"I want to use attribution modeling to understand how the SEO work we recently launched is paying off in terms of conversions."

Guided Walk-Through:

Google Analytics Model Comparison Tool BEFORE

Conversions & CPA ▾ | % change in Conversions ▾ (from Last Interaction)

MCF Channel Grouping	Spend (for selected time range)	Last Interaction		First Interaction		Position Based		% change in Conversions (from Last Interaction)	
		Conversions ↓	CPA	Conversions	CPA	Conversions	CPA	First Interaction	Position Based
1. Direct	–	1,297.00 (45.19%)	–	961.00 (33.48%)	–	1,143.44 (39.84%)	–	-25.91% ↓	-11.84% ↓
2. Organic Search	–	846.00 (29.48%)	–	994.00 (34.63%)	–	913.84 (31.84%)	–	17.49% ↑	8.02% ↑
3. Referral	–	350.00 (12.20%)	–	447.00 (15.57%)	–	393.17 (13.70%)	–	27.71% ↑	12.34% ↑
4. Paid Search	$381.70	187.00 (6.52%)	$2.04	238.00 (8.29%)	$1.60	210.30 (7.33%)	$1.82	27.27% ↑	12.46% ↑
5. (Other)	–	89.00 (3.10%)	–	107.00 (3.73%)	–	98.35 (3.43%)	–	20.22% ↑	10.51% ↑
6. Other Advertising	–	50.00 (1.74%)	–	57.00 (1.99%)	–	52.90 (1.84%)	–	14.00% ↑	5.80% ↑
7. Social Network	–	38.00 (1.32%)	–	51.00 (1.78%)	–	44.20 (1.54%)	–	34.21% ↑	16.32% ↑
8. Display	–	13.00 (0.45%)	–	15.00 (0.52%)	–	13.80 (0.48%)	–	15.38% ↑	6.15% ↑

Show rows: 10 ▾ Go to: 1 1 - 8 of 8 ‹ ›

Guided Walk-Through:

Google Analytics Model Comparison Tool AFTER

Model Comparison Tool

Mar 20, 2019 - Mar 26, 2019

Conversion Segments | Export | Save | Insights(11)

Conversion: All | Type: All | Google Ads | Lookback Window: Set 30 days prior to conversion

% of conversions: 100.00%

Last Interaction vs First Interaction vs Position Based

Primary Dimension: MCF Channel Grouping | Default Channel Grouping | Source / Medium | Source | Medium | Other | Channel Groupings

Secondary dimension | advanced

Conversion Value & ROAS

MCF Channel Grouping	Spend (for selected time range)	Last Interaction		First Interaction		Position Based		% change in Conversion Value (from Last Interaction)	
		Conversion Value	ROAS	Conversion Value	ROAS	Conversion Value	ROAS	First Interaction	Position Based
1. Organic Search	–	$440.75	–	$390.23	–	$415.49	–	-11.46%	-5.73%
2. Direct	–	$252.69	–	$270.23	–	$262.12	–	6.94%	3.73%
3. Paid Search	$212.17	$115.33	54.36%	$148.31	69.90%	$131.16	61.82%	28.60%	13.73%
4. (Other)	–	–	–	–	–	–	–	–	–
5. Display	–	–	–	–	–	–	–	–	–
6. Other Advertising	–	–	–	–	–	–	–	–	–
7. Referral	–	–	–	–	–	–	–	–	–
8. Social Network	–	–	–	–	–	–	–	–	–

Show rows: 10 | Go to: 1 | 1 - 8 of 8

Guided Walk-Through:
Google Analytics Model Comparison Tool AFTER 2.0

Last Interaction vs First Interaction vs Position Based

Primary Dimension: MCF Channel Grouping | Default Channel Grouping | Source / Medium | Source | Medium | Other | Channel Groupings

Secondary dimension | advanced

Conversion Value & ROAS

MCF Channel Grouping	Spend (for selected time range)	Last Interaction		First Interaction		Position Based		% change in Conversion Value (from Last Interaction)	
		Conversion Value	ROAS	Conversion Value	ROAS	Conversion Value	ROAS	First Interaction	Position Based
1. Direct	–	**$479.61**	–	$209.86	–	$360.92	–	-56.24%	-24.75%
2. Organic Search	–	**$292.08**	–	$292.08	–	$292.08	–	0.00%	0.00%
3. Paid Search	$381.70	**$35.97**	9.42%	$197.85	51.83%	$100.72	26.39%	450.04%	180.02%
4. (Other)	–	–	–	$107.87	–	$53.94	–	–	–
5. Display	–	–	–	–	–	–	–	–	–
6. Other Advertising	–	–	–	–	–	–	–	–	–
7. Referral	–	–	–	–	–	–	–	–	–
8. Social Network	–	–	–	–	–	–	–	–	–

Show rows: 10 Go to: 1 1 - 8 of 8

Partner Exercise:

Make Your Case

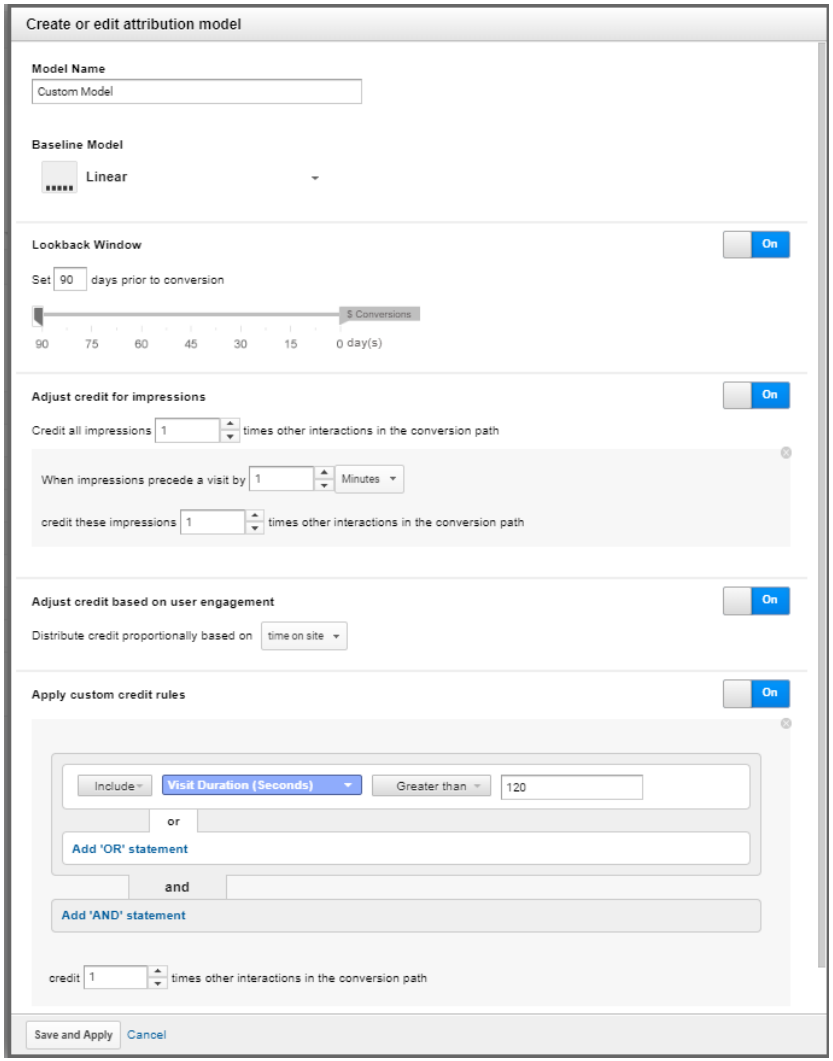
5 minutes

Create a custom position-based model for Maggie, keeping these items in mind:

1. Focus on understanding how new SEO work is paying off
2. New SEO work launched 45 days ago
3. Last touch should be weighted more heavily than first touch
4. Give more credit to users who spend more time on the site

Maggie, Marketer

"Thanks for the insights! Now, can you create a **custom position-based model** for me?"



5 minutes

Maggie, Marketer

"Thanks for the insights! Now, can you create a **custom position-based model** for me?"

5 minutes

Create or edit attribution model

Model Name

Custom Model

Baseline Model

First Interaction

Lookback Window

On

Set 44 days prior to conversion

$ Conversions

90 75 60 45 30 15 0 day(s)

Specify which interaction will receive 100% of credit

On

Include | Interaction Type | Exactly matching | Direct Visit

or

Add 'OR' statement

and

Add 'AND' statement

If no interactions satisfy the above conditions, the absolute first interaction will receive 100% credit. Change

Maggie, Marketer

"Thanks for the insights! Now, can you create a **custom position-based model** for me?"

5 minutes

Create or edit attribution model

Model Name

Custom Model

Baseline Model

First Interaction & Last Interaction

Lookback Window

On

Set 44 days prior to conversion

$ Conversions

90 75 60 45 30 15 0 day(s)

Specify which interaction will receive 100% of credit

On

Include | Interaction Type | Exactly matching | Direct Visit

or

Add 'OR' statement

and

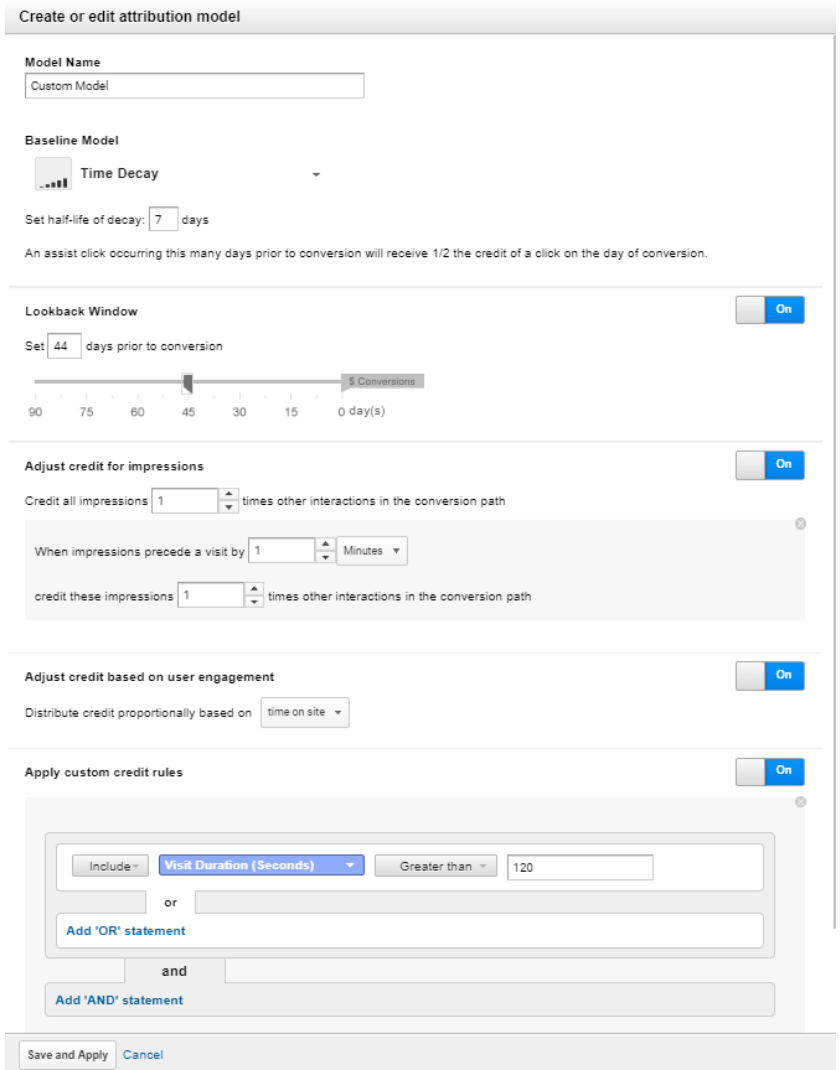
Add 'AND' statement

If no interactions satisfy the above conditions, the absolute first interaction will receive 100% credit. Change

Maggie, Marketer

"Thanks for the insights! Now, can you create a **custom position-based model** for me?"

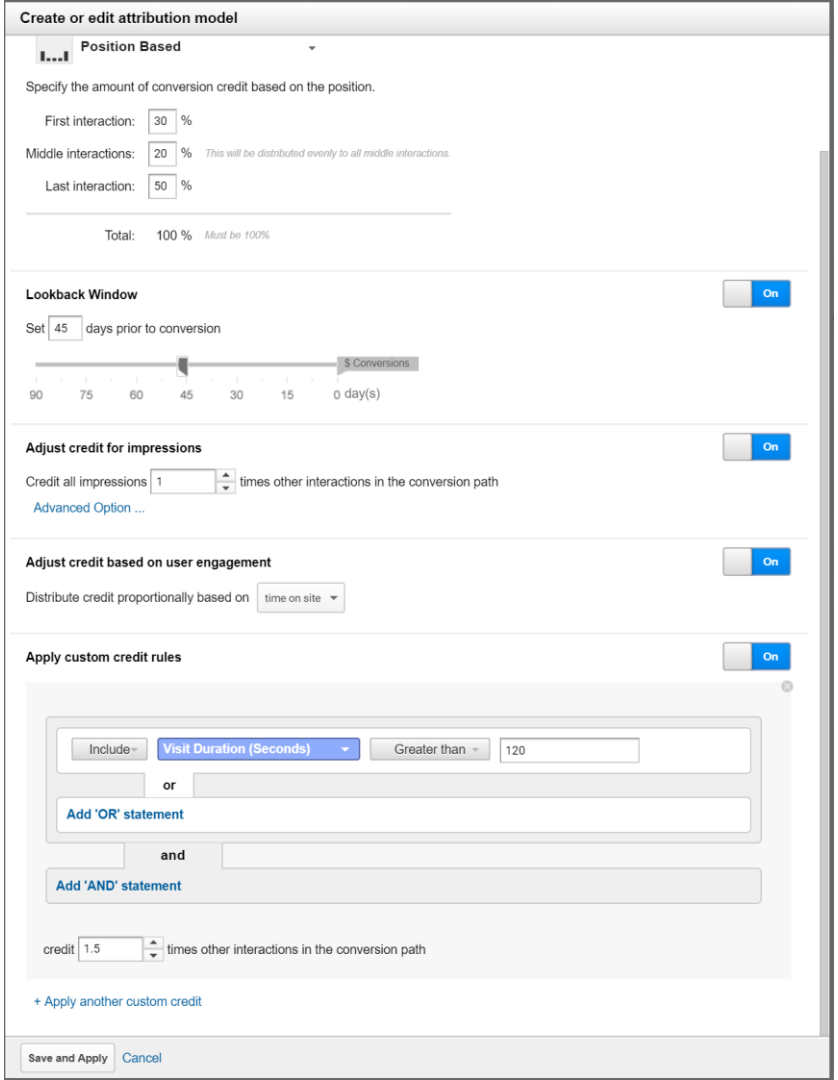
5 minutes



Maggie, Marketer

"Thanks for the insights! Now, can you create a **custom position-based model** for me?"

5 minutes



Maggie, Marketer

"Thanks for the insights! Now, can you create a **custom position-based model** for me?"

Partner Exercise:

Make Your Case

5 minutes

Which is the best model under the following needs?

1. Focus on understanding how new SEO work is paying off
2. New SEO work launched 45 days ago
3. Last touch should be weighted more heavily than first touch
4. Give more credit to users who spend more time on the site

Maggie, Marketer

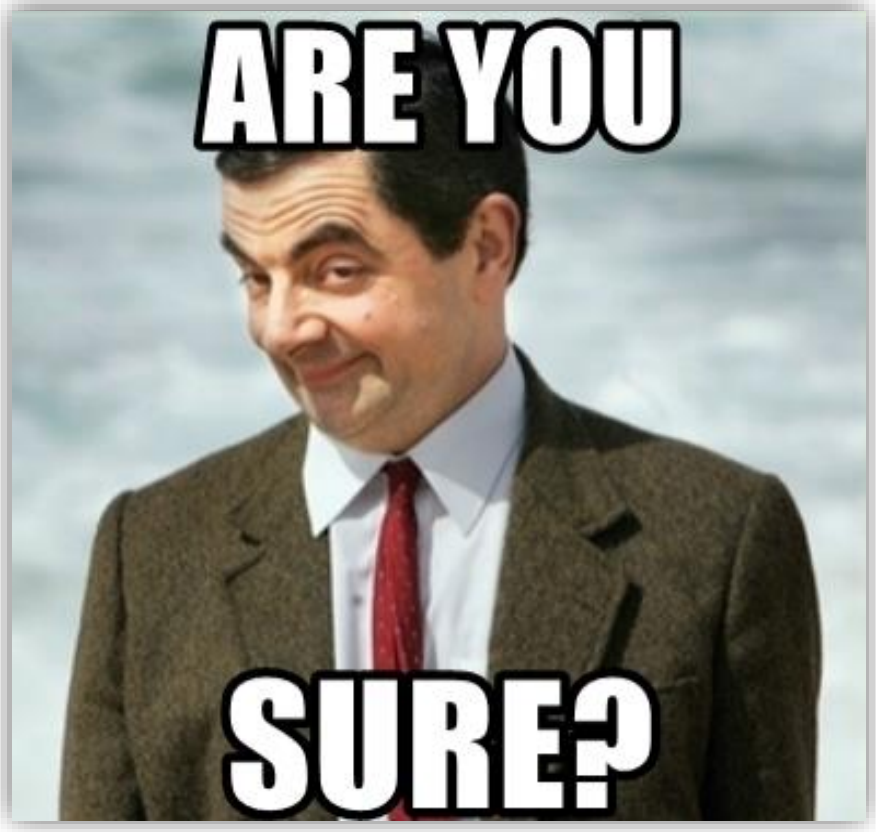
"Thanks for the insights! Now, can you create a **custom position-based model** for me?"

Statistics in Marketing

What to ask your data

- Should I take action based on these results?
- Does *x* really affect *y*?

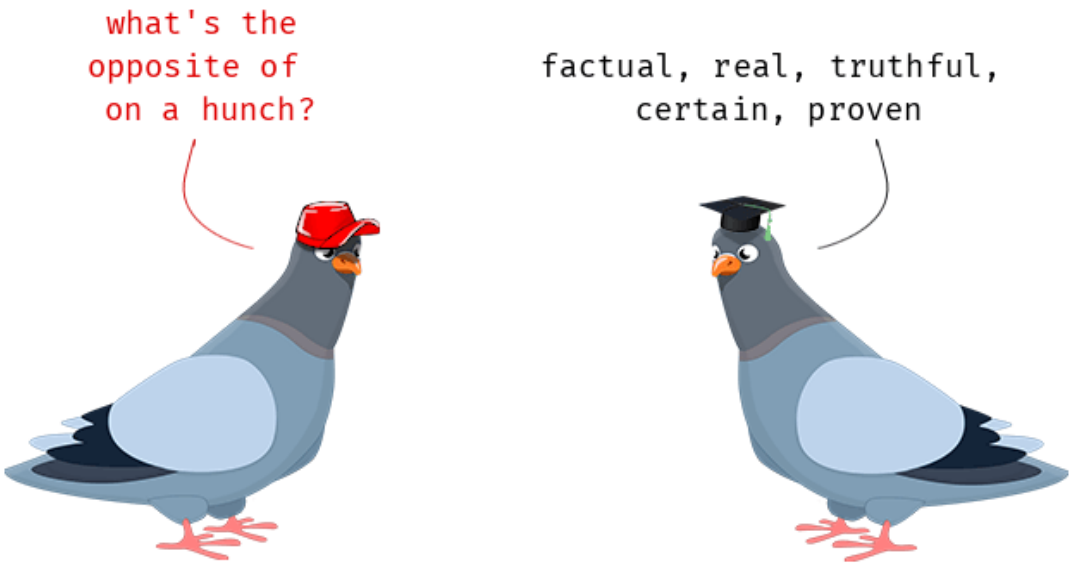
Statistical Significance



Real Cases:
Case study

5 minutes

For Sigma Ridge Data significance is the difference between a hunch and a fact.
Between getting the job or not.



Statistical Significance

https://www.youtube.com/watch?v=R13BD8qKeTg

Use significance to review your insights

Null hypothesis

- What we currently have
- What we assume to be true
- Status quo

H_0

Alternative hypothesis

- The new guess, theory
- Opposite to what we had
- The "new"

$H_1 \, or \, H_A$

Use significance to review your insights

A Dominos pizza manager is worried that his cheese distribution machine for each pizza is giving clients extra cheese for free!! -Anything more than 6 ounces of cheese is free cheese to the clients. The manager doesn't like that!

He makes **30** pizzas to see if the average weight of cheese is over 6 ounces.

Null hypothesis

H_0: μ = 6 ounces

Alternative hypothesis

H_1: μ > 6 ounces

Use significance to review your insights

The National Sleep Foundation recommends that teenagers aged 14 to 17 years old get at least 8 hours of sleep per night for proper health and wellness.

A statistics class at a large high school suspects that students at their school are getting less than 8 hours of sleep on average. To test their theory, they randomly sample 42 of these students and ask them how many hours of sleep they get per night. The mean from the sample is $\bar{x} = 7.5$ hours.

Here's their alternative hypothesis:

H_a : The average amount of sleep students at their school get per night is...

What is an appropriate ending to their alternative hypothesis?

Null hypothesis

H_0: μ ≥ 8 hours

Alternative hypothesis

H_a: μ < 8 Hours

Source, Khan Academy

Use significance to review your insights

Step 1: Form two hypotheses to test

Null hypothesis

Email B is getting a higher conversion rate than Email A due to **random chance**.

Alternative hypothesis

Email B is getting a higher conversion rate because it's **more effective** than Email A.

H_0: CR (B) = CR (A)

H_A: CR (B) > CR (A)

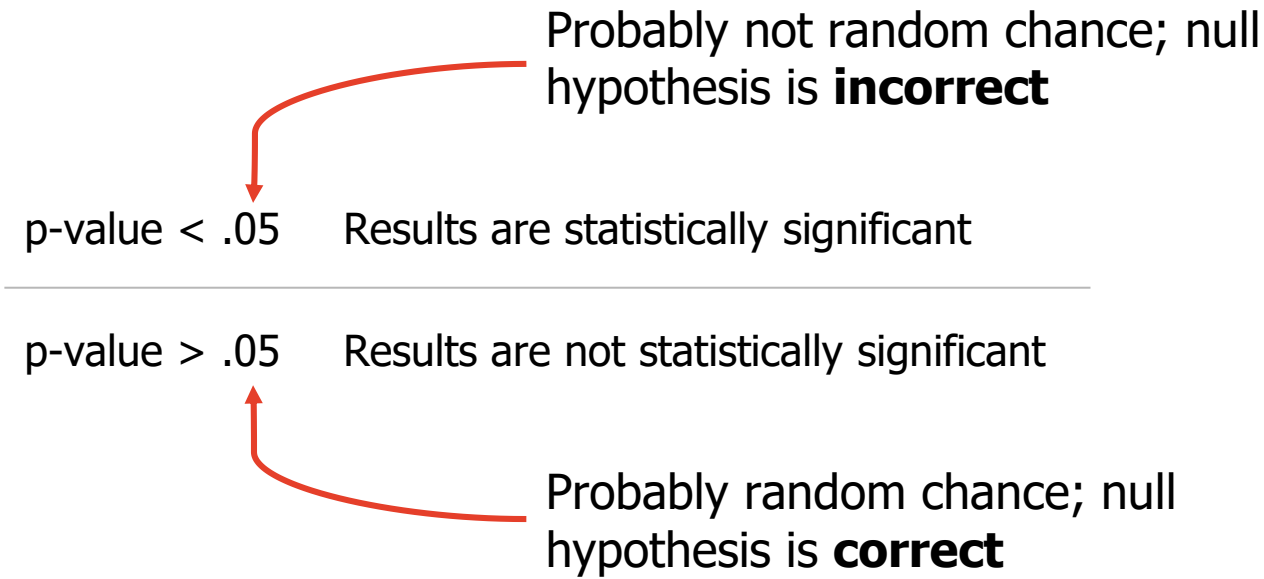
Use significance to review your insights

Step 2: Run your test, get your data

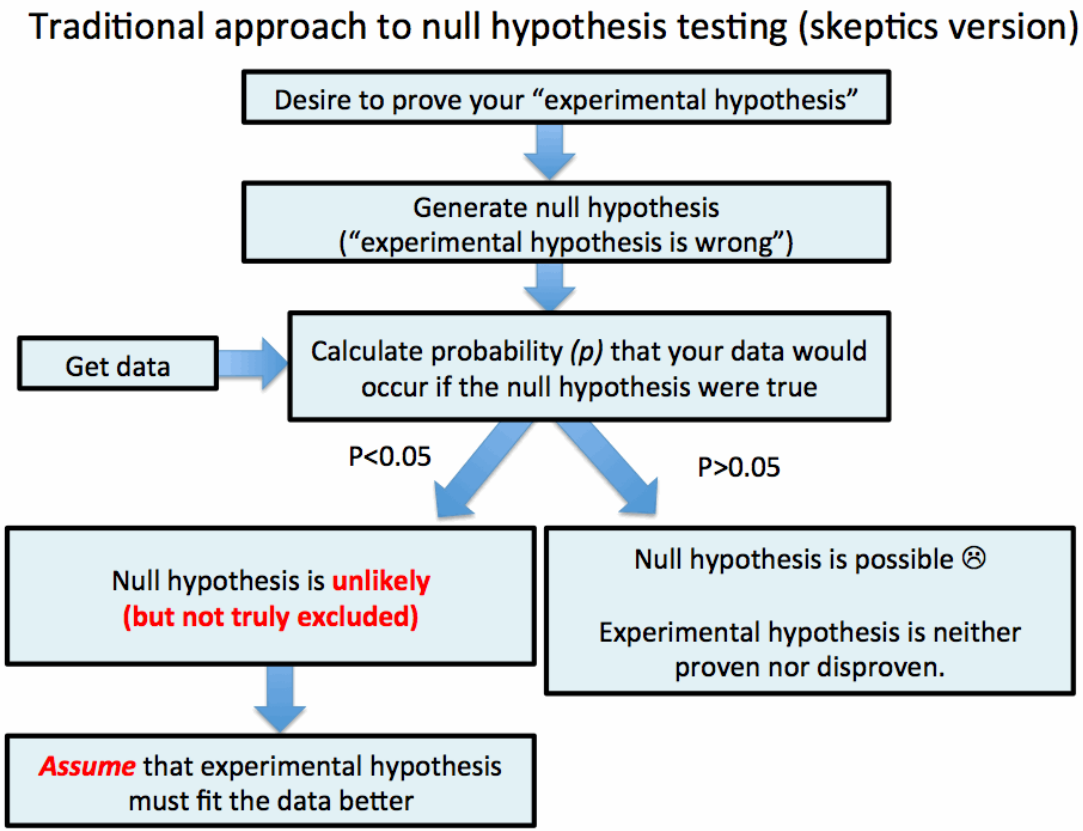
Email options	Email A	Email B
Visits	5,300	5,200
Orders	46	55
Conversion rate	0.87%	1.06%

Use significance to review your insights

Step 3: Calculate the p-value



Use significance to review your insights



Source: https://emcrit.org/pulmcrit/demystifying-the-p-value/

Which one won?

p-value = .15 > 0.05

Null hypothesis

Email B is getting a higher conversion rate than Email A due to **random chance**.

Alternative hypothesis

Email B is getting a higher conversion rate because it's **more effective** than Email A.

Guided Walk-Through:
No Math Needed

5 minutes

Use AB Testguide's calculator to determine the winner and p-value for the A/B test.

https://abtestguide.com/calc/

Email options	Email A	Email B
Visits	5,300	5,200
Orders	46	55
Conversion rate	0.87%	1.06%

Maggie, Marketer

"I want to know which of the two email options performed better."

AB Testguide

Is your test result significant? Does it have enough power?
Play with the controls and get a better feel for how a lower confidence level will boost the power or how an increase in test size can make a small CR-difference significant!

Pre-test calculation or post-test evaluation?

- ○ Pre-test analysis
- ◉ Test evaluation

Test data

Visitors A	Conversions A
5300	46
Visitors B	**Conversions B**
5200	55

Apply changes

Settings

Hypothesis (?)

- ◉ One-sided
- ○ Two-sided

Confidence (?)

- ○ 90%
- ◉ 95%
- ○ 99%

Test result

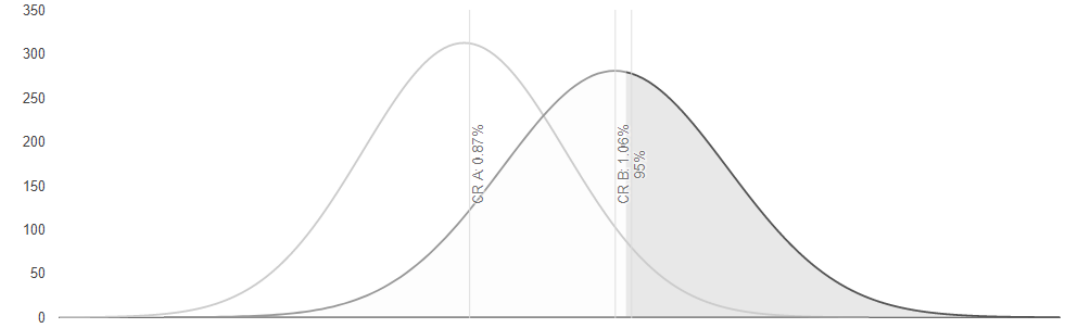
The test result is *not* significant.

The observed difference in conversion rate (21.86%) isn't big enough to declare a significant winner. There is no real difference in performance between A and B or you need to collect more data.

A 0.87%

B 1.06%

The expected distributions of variation A and B.



Conversion Rate Control	Conversion Rate B	Relative uplift in Conversion Rate
Conversions A / Visitors A	Conversions B / Visitors B	$CR_B - CR_A / CR_A$
0.87%	1.06%	21.86%
Observed Power	**p value**	**Z-score**
		$(CR_B - CR_A) / SE_{difference}$
44.45%	0.1598	0.9952
Standard error A	**Standard error B**	**Std. Error of difference**
$(CR_A * (1-CR_A) / Visitors_A)^{1/2}$	$(CR_B * (1-CR_B) / Visitors_B)^{1/2}$	$SE_{difference} = (SE_A^2 + SE_B^2)^{1/2}$
0.001274	0.001419	0.001907

Use significance to review your insights

Step 4: Do something

p-value < .05 → Take action!

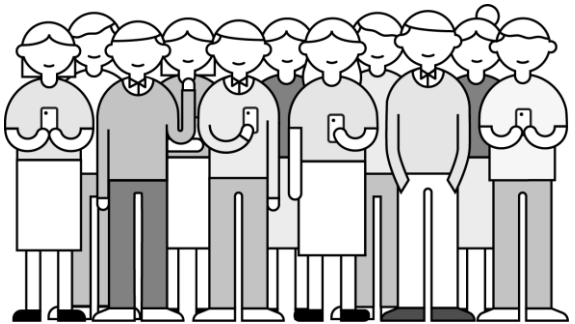
p-value > .05 → More testing needed

What do I do if it's not significant?

Step 4: Do something

Wait!
Run the test longer

Get more people to test



Partner Exercise:

Make Your Case

10 minutes

1. Use AB Testguide to determine if the results from the second test are statistically significant.

Email options	Email A	Email B
Visits	7,150	6,800
Orders	60	76
Conversion rate	0.84%	1.12%

2. What next steps will you recommend for Maggie to take?

Maggie, Marketer

"I ran the A/B test again, for a longer time period, and have new results to share. Can you let me know if they're statistically significant?"

AB Testguide

Is your test result significant? Does it have enough power?
Play with the controls and get a better feel for how a lower confidence level will boost the power or how an increase in test size can make a small CR-difference significant!

Pre-test calculation or post-test evaluation?

- ○ Pre-test analysis
- ◉ Test evaluation

Test data

Visitors A	Conversions A
7150	60
Visitors B	**Conversions B**
6800	76

Apply changes

Settings

Hypothesis (?)

- ◉ One-sided
- ○ Two-sided

Confidence (?)

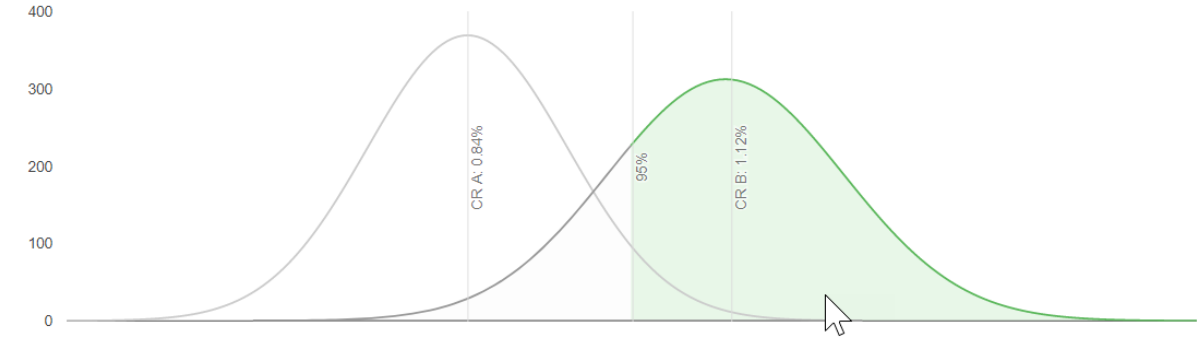
- ○ 90%
- ◉ 95%
- ○ 99%

Test result

Significant test result!
Variation B's observed conversion rate (1.12%) was 33.19% higher than variation A's conversion rate (0.84%). You can be 95% confident that this result is a consequence of the changes you made and not a result of random chance.

A 0.84%
B 1.12%

The expected distributions of variation A and B.



Conversion Rate Control Conversions A / Visitors A	Conversion Rate B Conversions B / Visitors B	Relative uplift in Conversion Rate $CR_B - CR_A / CR_A$
0.84%	1.12%	33.19%
Observed Power	**p value**	**Z-score** $(CR_B - CR_A) / SE_{difference}$
78.60%	0.0477	1.6675
Standard error A $(CR_A * (1-CR_A) / Visitors_A)^{1/2}$	**Standard error B** $(CR_B * (1-CR_B) / Visitors_B)^{1/2}$	**Std. Error of difference** $SE_{difference} = (SE_A^2 + SE_B^2)^{1/2}$
0.001079	0.001275	0.00167

AB Testguide

Is your test result significant? Does it have enough power?
Play with the controls and get a better feel for how a lower confidence level will boost the power or how an increase in test size can make a small CR-difference significant!

Pre-test calculation or post-test evaluation?

- ○ Pre-test analysis
- ◉ Test evaluation

Test data

Visitors A	Conversions A
6800	76
Visitors B	**Conversions B**
7150	60

Apply changes

Settings

Hypothesis (?)

- ◉ One-sided
- ○ Two-sided

Confidence (?)

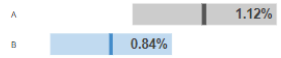
- ○ 90%
- ◉ 95%
- ○ 99%

save & share url

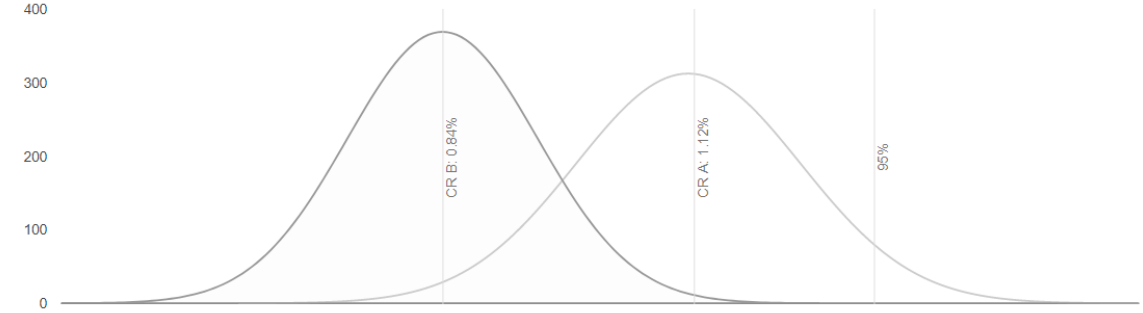
Test result

The test result is *not* significant.

The observed difference in conversion rate (-24.92%) isn't big enough to declare a significant winner. There is no real difference in performance between A and B or you need to collect more data.



The expected distributions of variation A and B.



Conversion Rate Control	Conversion Rate B	Relative uplift in Conversion Rate
Conversions A / Visitors A	Conversions B / Visitors B	$CR_B - CR_A / CR_A$
1.12%	0.84%	-24.92%
Observed Power	**p value**	**Z-score**
		$(CR_B - CR_A) / SE_{difference}$
0.00%	0.9523	-1.6675
Standard error A	**Standard error B**	**Std. Error of difference**
$(CR_A * (1-CR_A) / Visitors_A)^{1/2}$	$(CR_B * (1-CR_B) / Visitors_B)^{1/2}$	$SE_{difference} = (SE_A^2 + SE_B^2)^{1/2}$
0.001275	0.001079	0.00167

Use significance to review your insights

Since our sample usually only contains a subset of the data in the population, we **cannot be absolutely** certain as to whether the null hypothesis is true or not. We can merely gather information (via statistical tests) to determine whether it is likely or not. We therefore speak about **rejecting** or **not rejecting** (aka **retaining**) the null hypothesis on the basis of some test, but not of **accepting** the null hypothesis or the alternative hypothesis. Often in an experiment we are actually testing the validity of the alternative hypothesis by testing whether to reject the null hypothesis.

Source: http://www.real-statistics.com/hypothesis-testing/null-hypothesis/

Use significance to review your insights

When performing such tests, there is some chance that we will reach the wrong conclusion. There are two types of **errors**:

- **Type I** – H_0 is rejected even though it is true (**false positive**) - The acceptable level of a Type I error is designated by **alpha** (α)
- **Type II** – H_0 is not rejected even though it is false (**false negative**) - the acceptable level of a Type II error is designated **beta** (β).

Source: http://www.real-statistics.com/hypothesis-testing/null-hypothesis/

1. **Significance level** is the acceptable level of type I error, denoted α. Typically, a significance level of $\alpha = .05$ is used (although sometimes other levels such as $\alpha = .01$ may be employed). This means that we are willing to tolerate up to 5% of type I errors, i.e. we are willing to accept the fact that in 1 out of every 20 samples we reject the null hypothesis even though it is true.

2. **P-value** (the **probability value**) is the value p of the statistic used to test the null hypothesis. If $p < \alpha$ then we reject the null hypothesis.

3. **Statistical power** is $1 - \beta$. Thus power is the probability that you find an effect when one exists, i.e. the probability of correctly rejecting a false null hypothesis. While a significance level for type I error of $\alpha = .05$ is typically used, generally the target for β is .20 or .10, and so .80 or .90 is used as the target value for power.

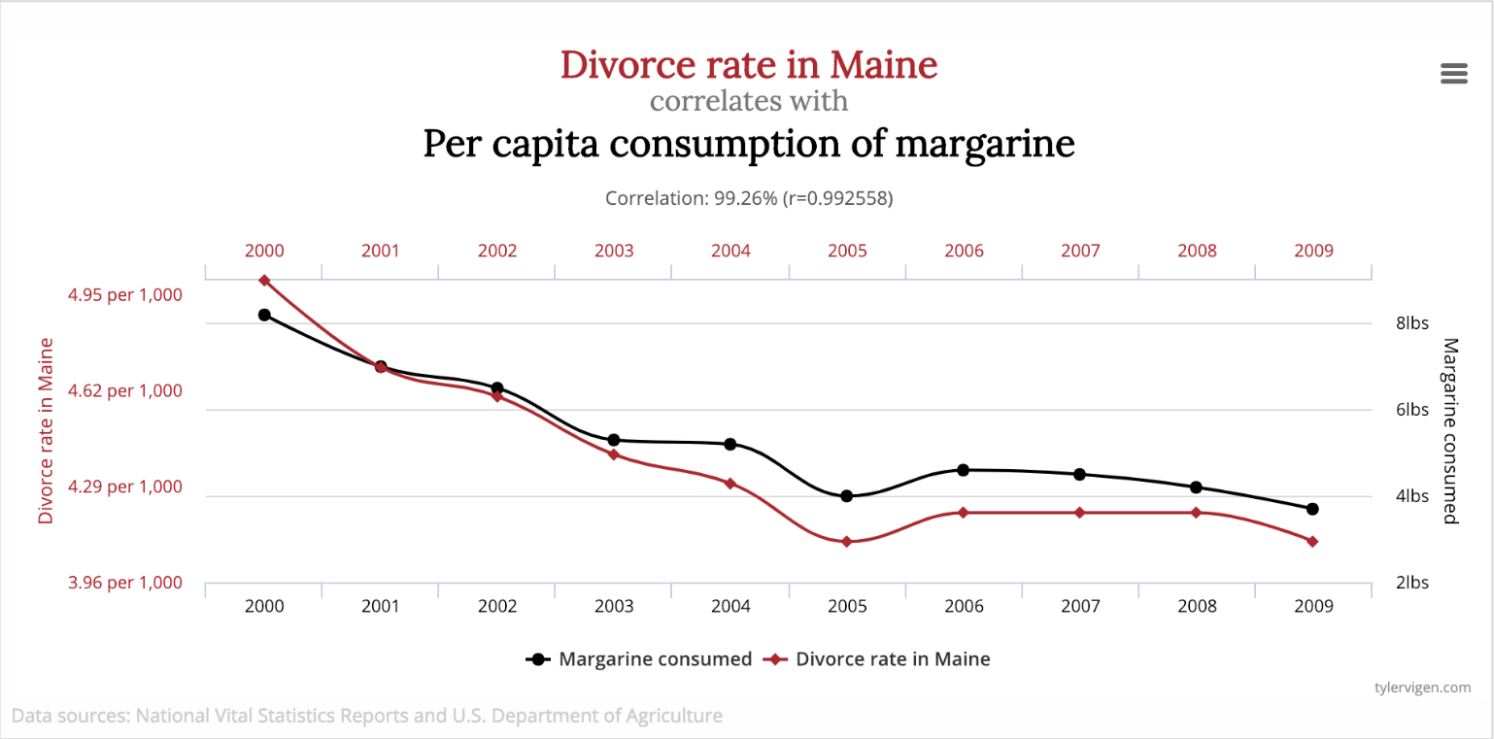
Source: http://www.real-statistics.com/hypothesis-testing/null-hypothesis/

Real Cases:
Case study

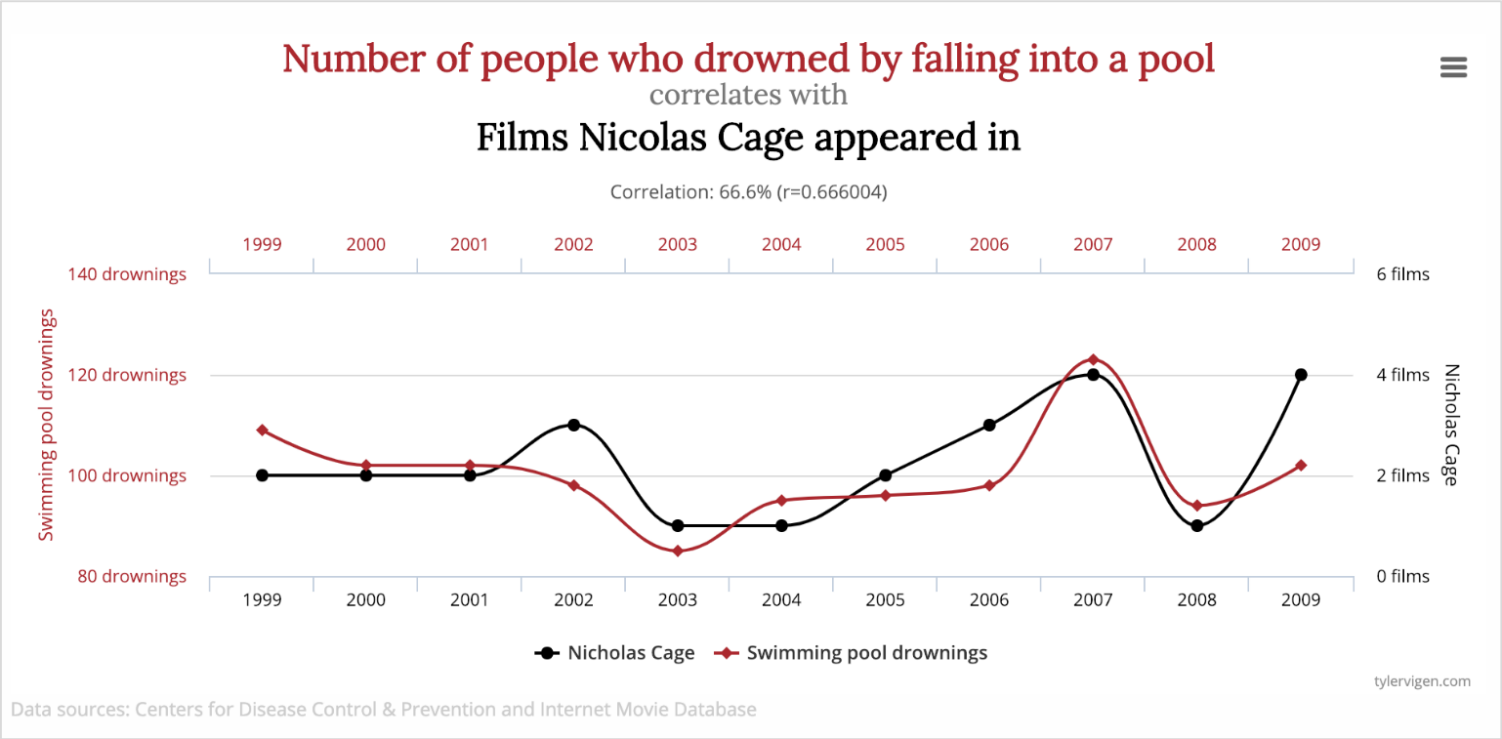
5 minutes

Can you, yes you! Give us an example of how a company has made a bad decision based on insignificant test data, or avoided making a bad decision by looking at the significance of test data. Basically - how has a company used test data in order to make an informed decision?

Causation or Correlation?



Causation or Correlation?



Regression Analysis

A way to measure the relationship between one variable and another

Guided Walk-Through:
Linear Regression in Excel

15 minutes

Let's use Excel to regress **ad spend** and **revenue** to understand if an increased advertising budget will actually pay off.

How is ad spend affecting revenue?

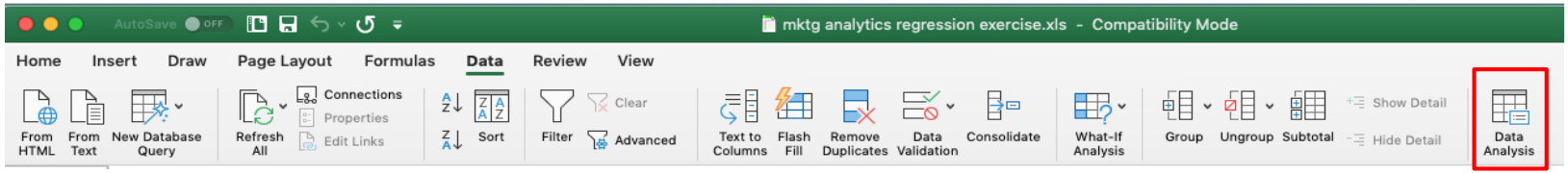
Claire, Owner

"Maggie requested a bigger advertising budget for next quarter, but I'm not sure if I should approve it."

Guided Walk-Through:
Linear Regression in Excel

15 minutes

Visit: https://www.sigmaridge.com/general-assembly/advanced-marketing-analytics to download the Excel file. Open the file in Excel.



Make sure you have the Data Analysis ToolPak installed.

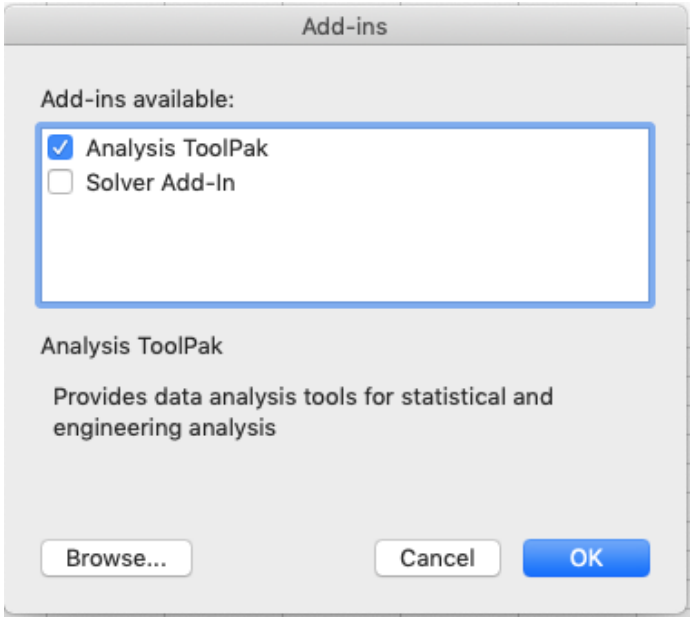
Guided Walk-Through:

Linear Regression in Excel

15 minutes

Installing the Data Analysis ToolPak

- Click "Tools" in the top menu
- Click "Excel Add-ins"
- In the "Add-ins available" box that pops up, select the "Analysis ToolPak" box and click OK
- Quit and restart Excel



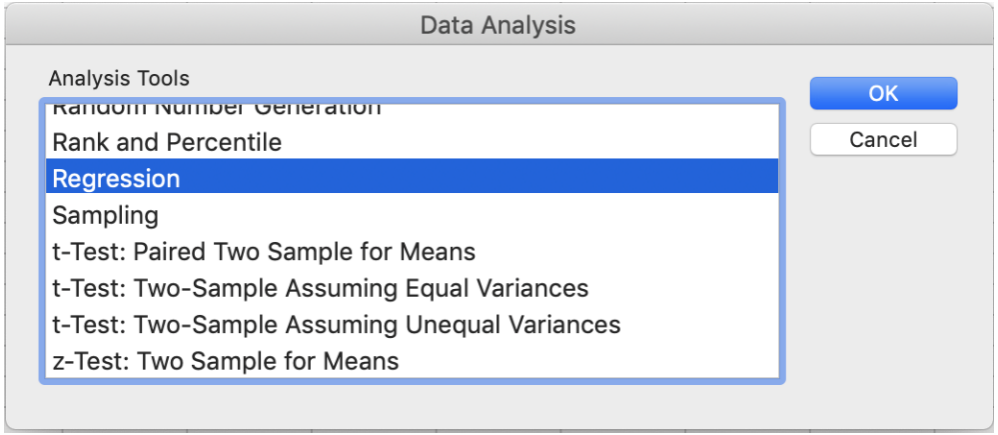
Guided Walk-Through:

Linear Regression in Excel

15 minutes

Creating the regression

Click “Data Analysis” and then “Regression”.



Guided Walk-Through:
Linear Regression in Excel

15 minutes

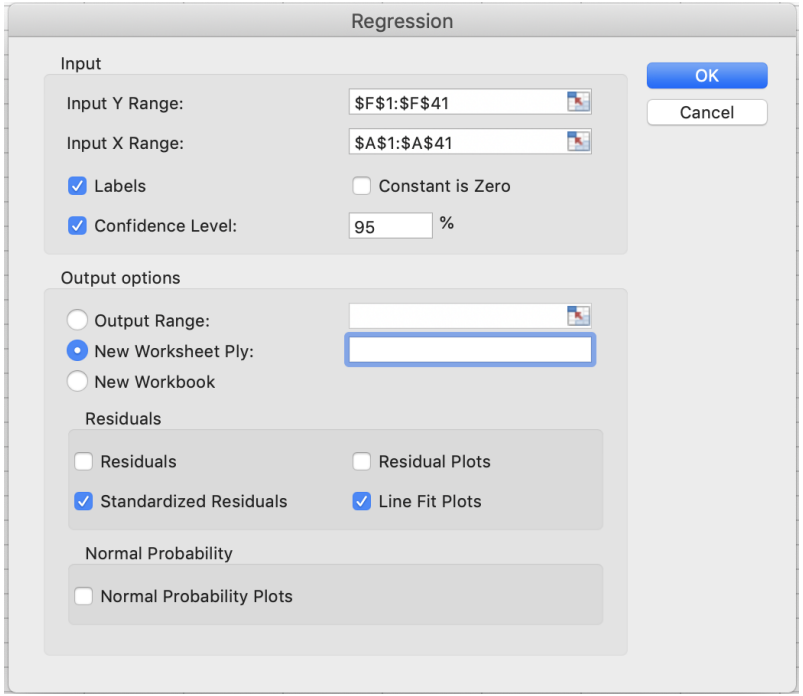
Creating the regression

Select the data we want to regress.

- Y Range: what's being affected?
- X Range: what's the constant?

Check the other boxes as they appear here. Click okay!

How is revenue affected by spend?



Guided Walk-Through:
Linear Regression in Excel

15 minutes

We did it!

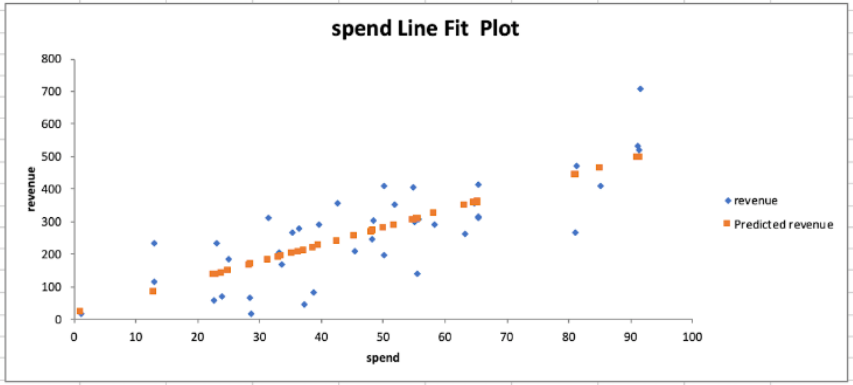
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.789455027
R Square	**0.62323924**
Adjusted R Square	0.613324483
Standard Error	93.81829768
Observations	40

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	553283.6274	553283.627	62.8597605	1.41504E-09
Residual	38	334471.1732	8801.87298		
Total	39	887754.8006			

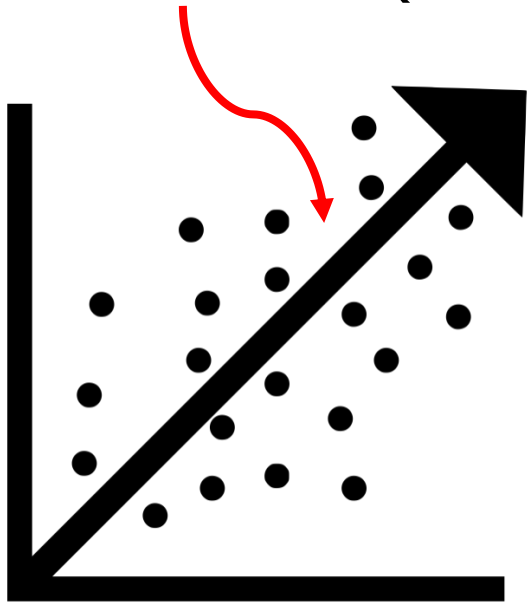
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	15.70580455	35.17265777	0.44653448	0.65774566	-55.49751857	86.90912766	-55.4975186	86.90912766
spend	**5.251744867**	0.662395325	7.92841475	**1.415E-09**	3.910795637	6.592694098	3.910795637	6.592694098



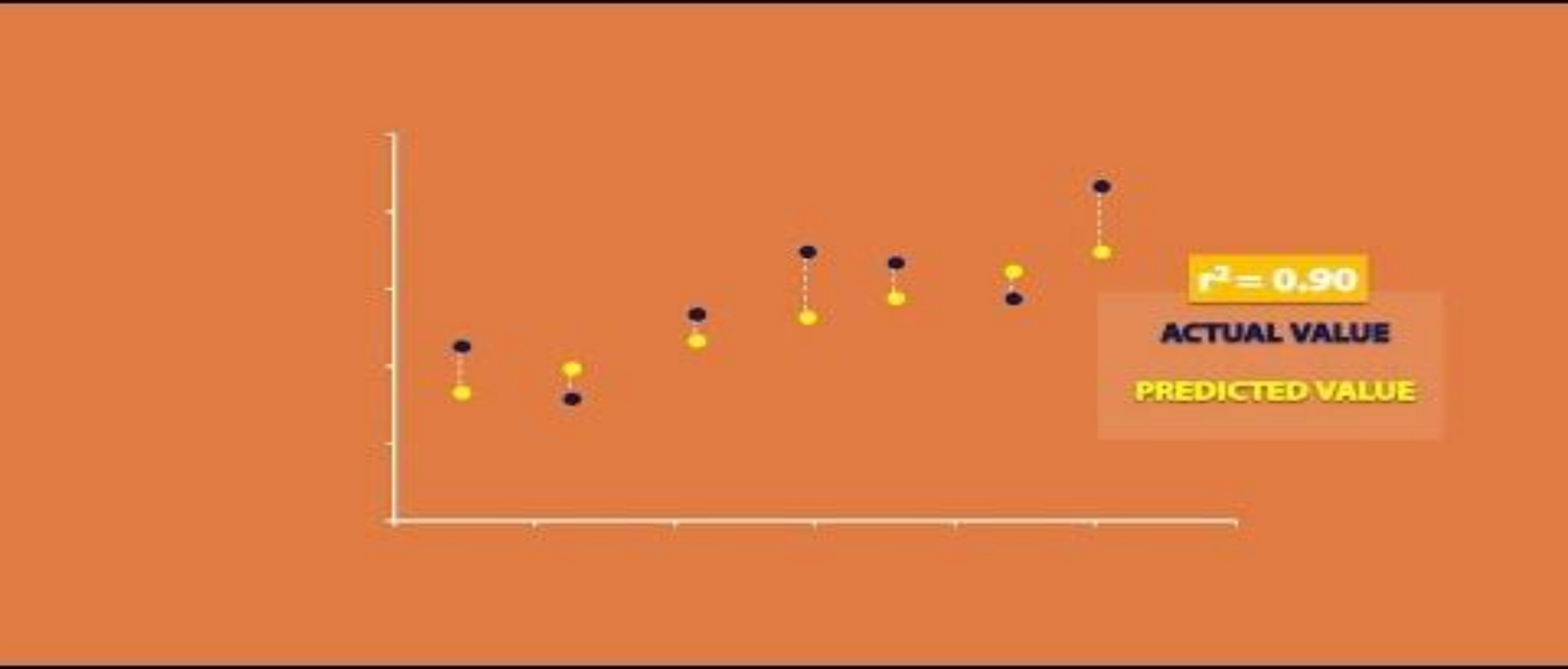
Let's dive into what this data means.

What you get from a regression

The Coefficient (AKA the slope)



"When we increase *x* by 1, we increase *y* by this much"



https://www.youtube.com/watch?v=Q-TtIPF0fCU

What you get from a regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.789455
R Square	0.623239
Adjusted R	0.613324
Standard E	93.8183
Observatic	40

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressior	1	553283.6	553283.6	62.85976	1.42E-09
Residual	38	334471.2	8801.873		
Total	39	887754.8			

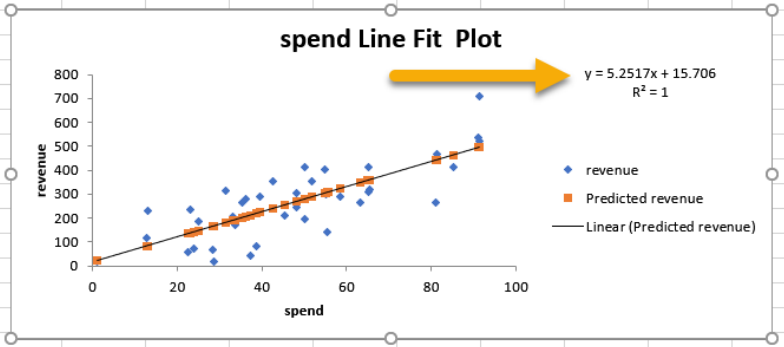
	Coefficients	*andard Erı*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	15.7058	35.17266	0.446534	0.657746	-55.4975	86.90913	-55.4975	86.90913
spend	5.251745	0.662395	7.928415	1.42E-09	3.910796	6.592694	3.910796	6.592694

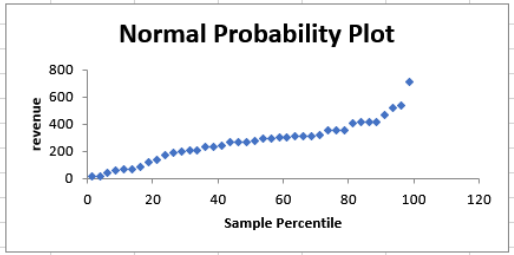
RESIDUAL OUTPUT

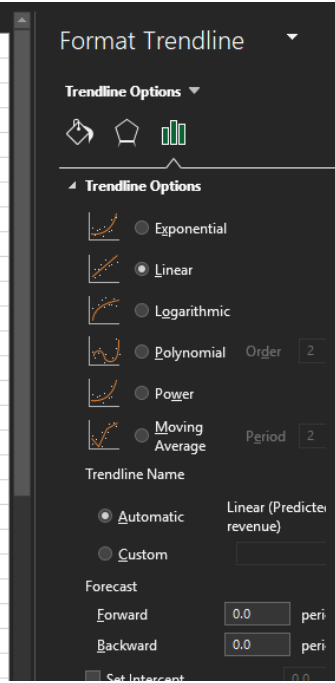
Observatior	*dicted reve*	*Residuals*	*dard Residuals*
1	134.4478	-75.5678	-0.816
2	211.4909	-166.571	-1.79867

PROBABILITY OUTPUT

Percentile	*revenue*
1.25	16.16
3.75	17.44







https://www.youtube.com/watch?v=Q-TtIPF0fCU

What you get from a regression

R square **< 85%**

"The output graph and coefficient **don't match** our data very well."

R square **> 85%**

"The output graph and coefficient **match** our data very well."

What you get from a regression

R square **< 85??%**

Are Low R-squared Values Always a Problem?

No! Regression models with low R-squared values can be perfectly good models for several reasons.

Some fields of study have an inherently greater amount of unexplainable variation. In these areas, your R^2 values are bound to be lower. **For example, studies that try to explain human behavior generally have R^2 values less than 50%.** People are just harder to predict than things like physical processes.

Fortunately, if you have a low R-squared value but the independent variables are statistically significant, you can still draw important conclusions about the relationships between the variables. Statistically significant coefficients continue to represent the mean change in the dependent variable given a one-unit shift in the independent variable. Clearly, being able to draw conclusions like this is vital.

Partner Exercise:

Make Your Case

5 minutes

Based on our regression analysis…

- When we increase ad spend, how much does revenue increase by?
- How well does this coefficient (Slope) explain the data?
- How sure are we that these results are accurate?
- Should Claire approve the larger budget?

"Maggie requested a bigger advertising budget for next quarter, but I'm not sure if I should approve it."

Real Cases:
Case study

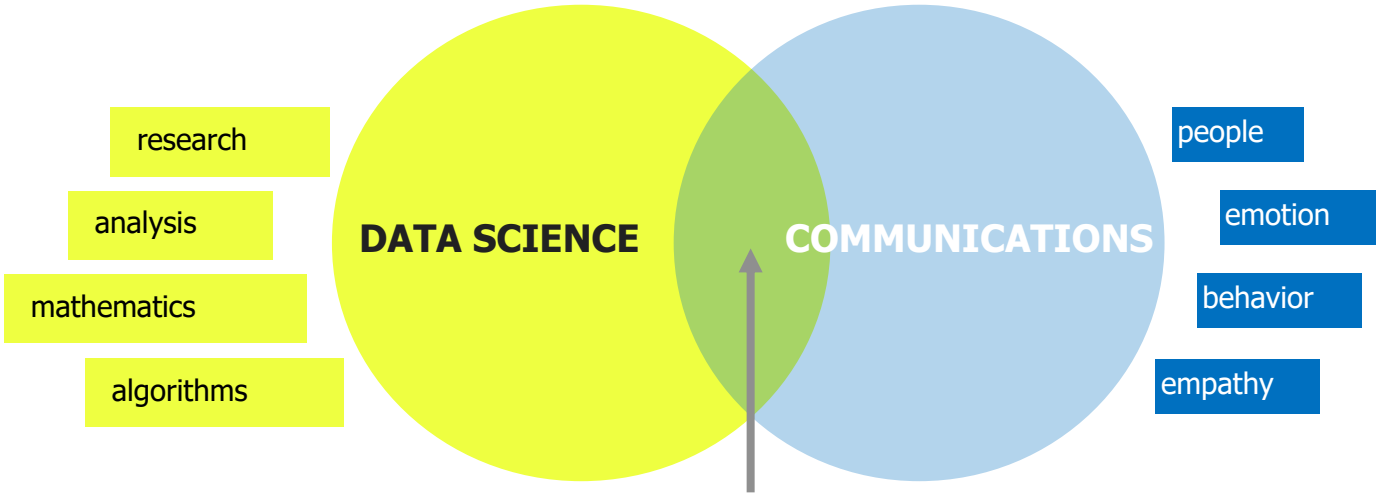
5 minutes

Share a real-life example of how a company has made a bad decision based on a correlation that didn't really exist, or avoided making a bad decision by using regression. Basically - how has a company used test data in order to make an informed decision?

Storytelling with Data

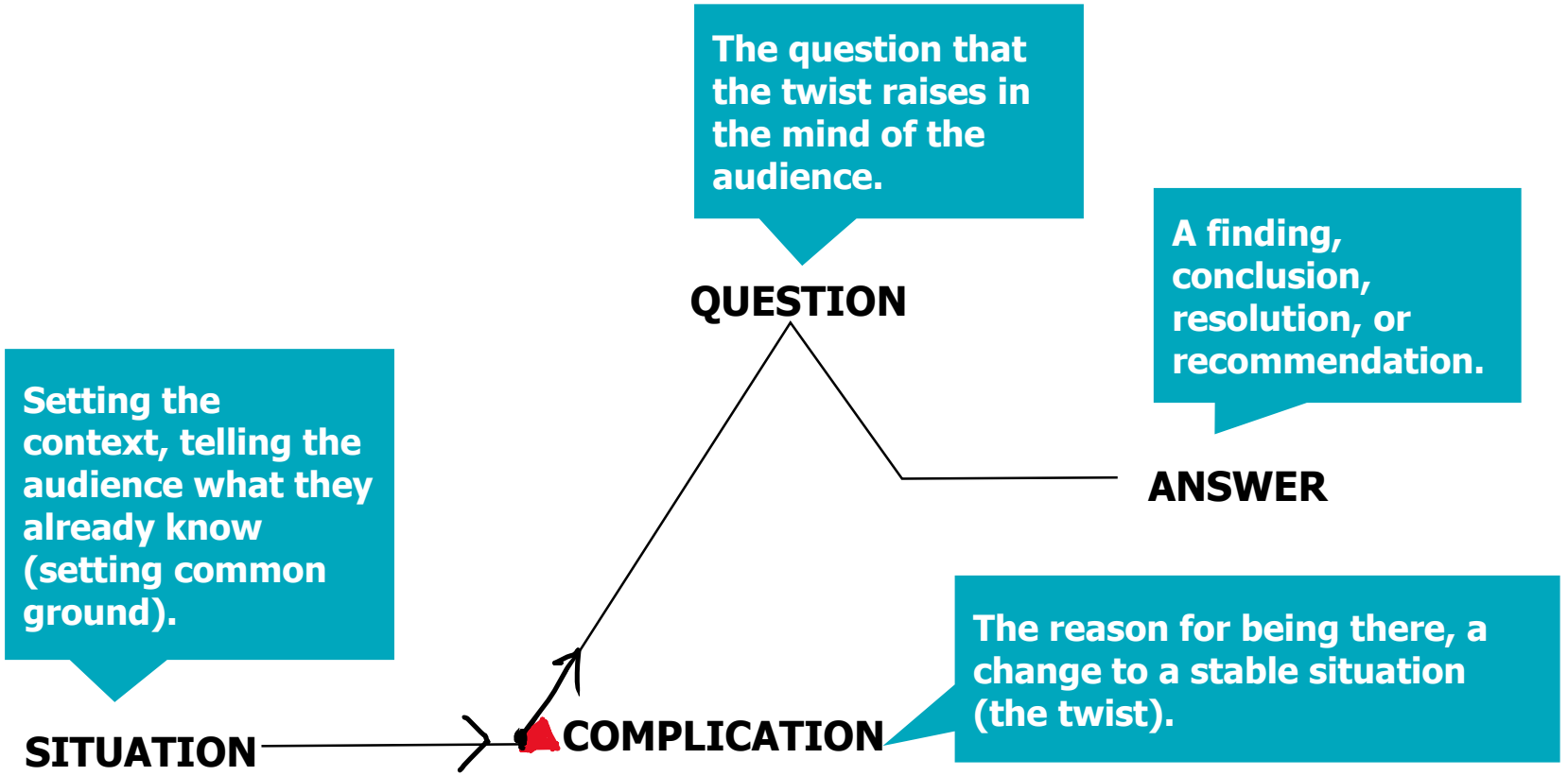
Storytelling with Data

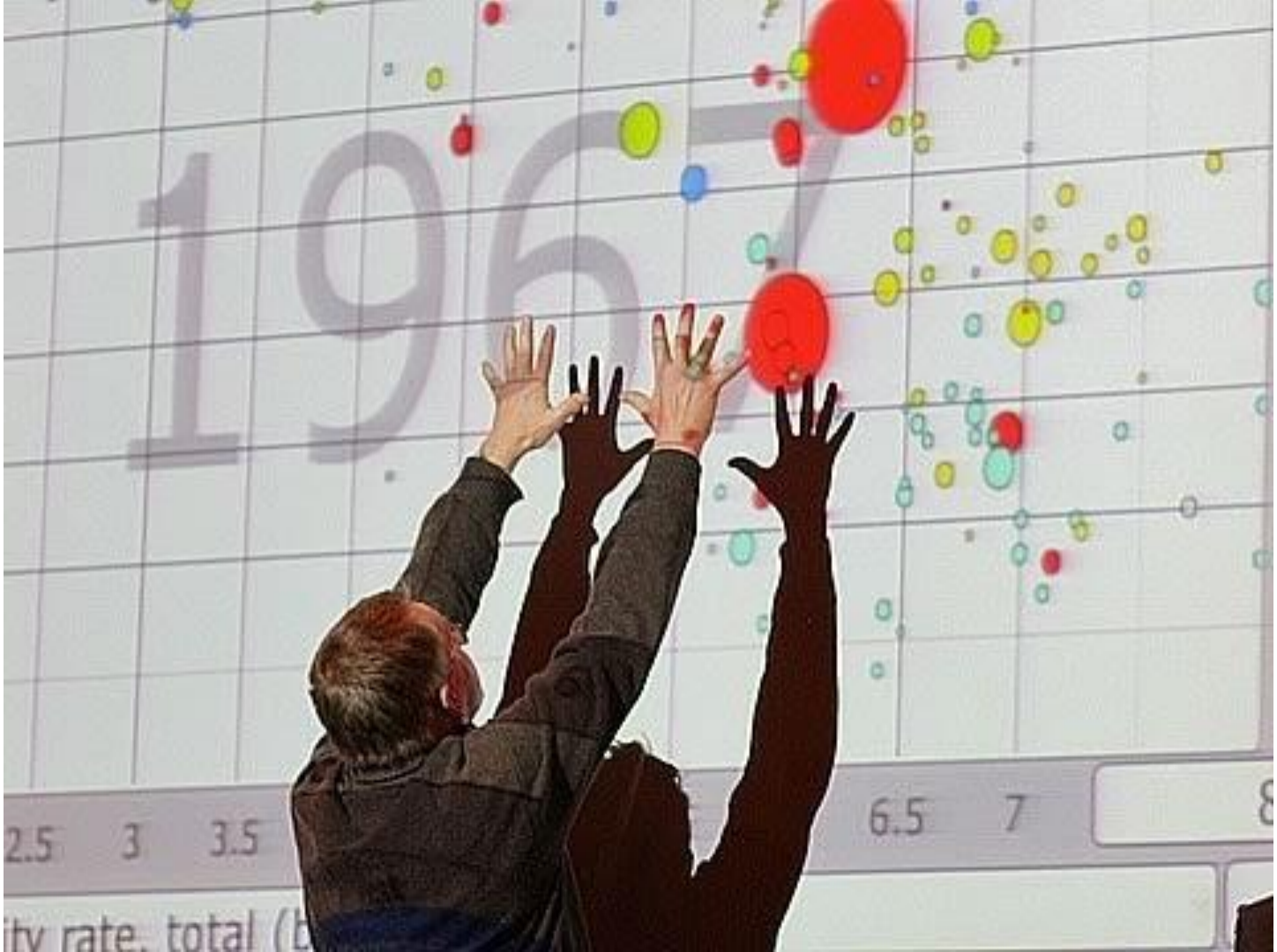
DATA MAKES IT RELIABLE, STORIES MAKE IT RELATABLE.



A POWERFUL DATA-DRIVEN STORY

The Shape of a Story




1967
2.5
3
3.5
6.5
7

Storytelling with Data

SITUATION	COMPLICATION	QUESTION	ANSWER
There is a lot of rich data about family size and life expectancy around the world.	Many people believe that family size and life expectancy differ widely between the first world and third world, but don't seem to know the exact facts.	Is the world still really that different when it comes to family size and life expectancy?	A lot has changed since the 1960s and the world is not as different as we think.

Know your audience and objective

WHO

What does your audience care about?
What questions would they have related to this content?

WHAT

What do you want this person to think, feel, and/or do because of this information?

WHY

How will sharing this information with them help the business?

Google Data Studio

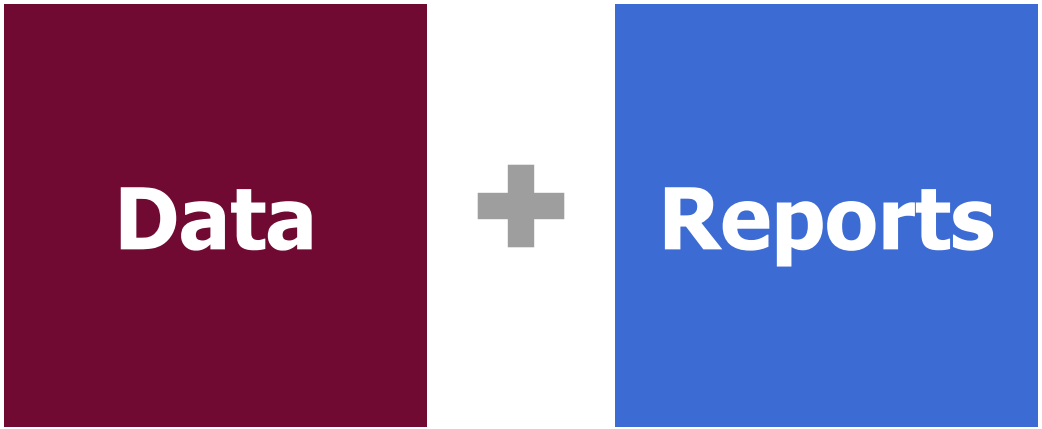
GENERAL ASSEMBLY

Google Data Studio

- Web-based reports, connected to live data
- Visual control with many different chart types
- New features rolling out frequently
- Connects to many different data sources
- It’s free

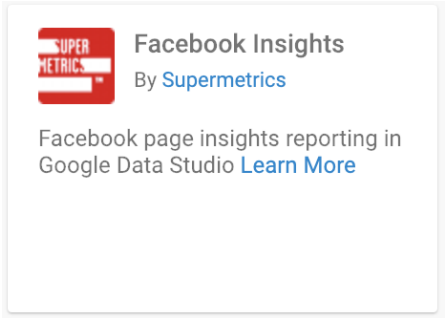
Google Data Studio

Google Data Studio



Where Does the Data Come From?

Google data



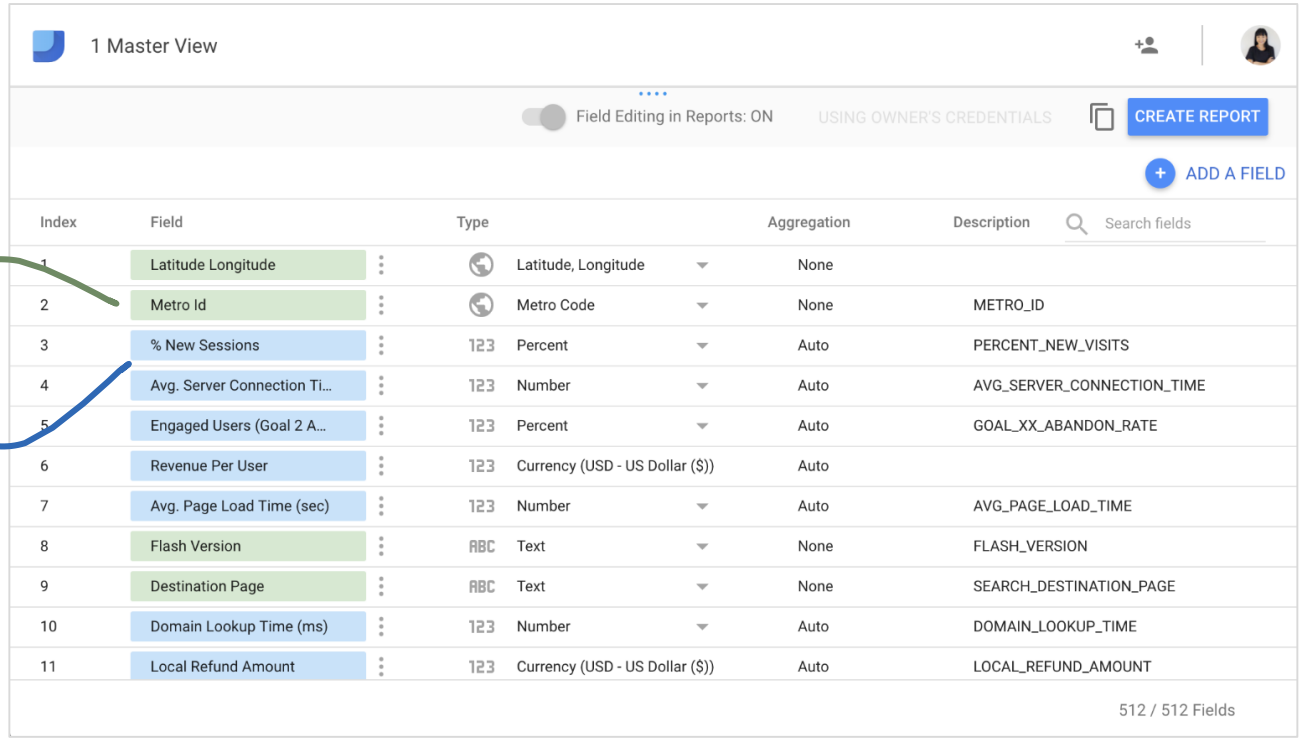
Partner Connections

Data you upload

Importing Data

Dimensions (describe data)

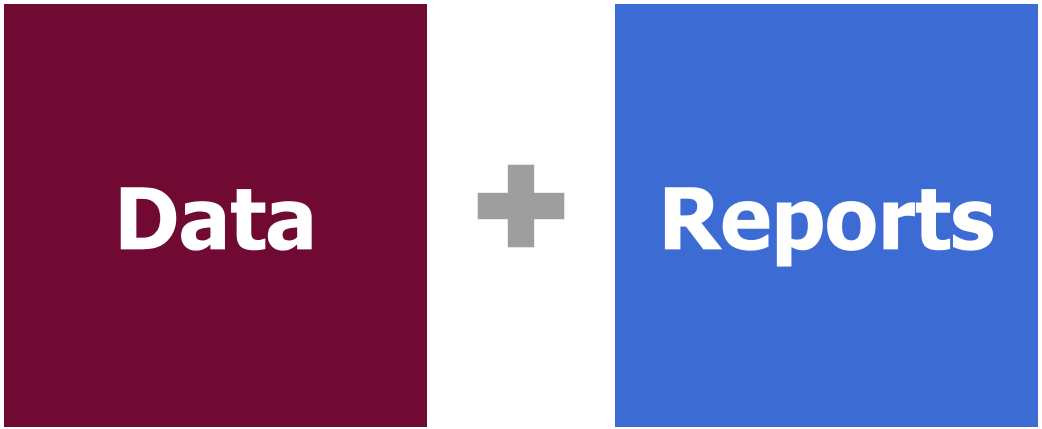
Metrics (measure data)



Index	Field	Type	Aggregation	Description
1	Latitude Longitude	Latitude, Longitude	None	
2	Metro Id	Metro Code	None	METRO_ID
3	% New Sessions	Percent	Auto	PERCENT_NEW_VISITS
4	Avg. Server Connection Ti...	Number	Auto	AVG_SERVER_CONNECTION_TIME
5	Engaged Users (Goal 2 A...	Percent	Auto	GOAL_XX_ABANDON_RATE
6	Revenue Per User	Currency (USD - US Dollar ($))	Auto	
7	Avg. Page Load Time (sec)	Number	Auto	AVG_PAGE_LOAD_TIME
8	Flash Version	Text	None	FLASH_VERSION
9	Destination Page	Text	None	SEARCH_DESTINATION_PAGE
10	Domain Lookup Time (ms)	Number	Auto	DOMAIN_LOOKUP_TIME
11	Local Refund Amount	Currency (USD - US Dollar ($))	Auto	LOCAL_REFUND_AMOUNT

512 / 512 Fields

Google Data Studio

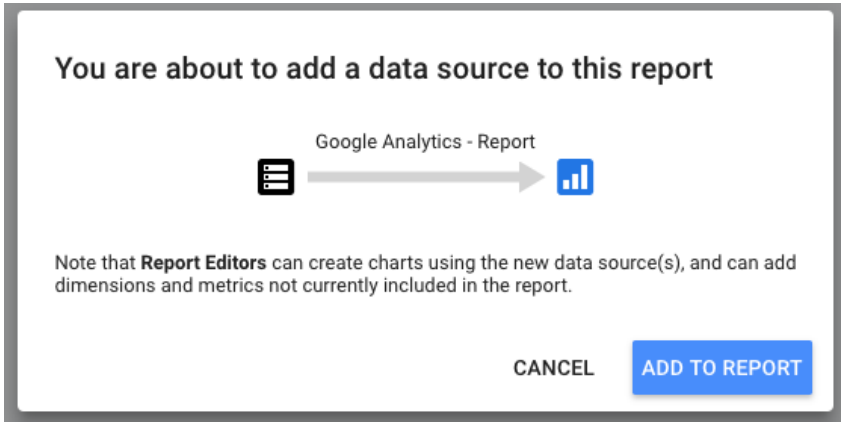


Guided Walk-Through:
Create a Report

10 minutes

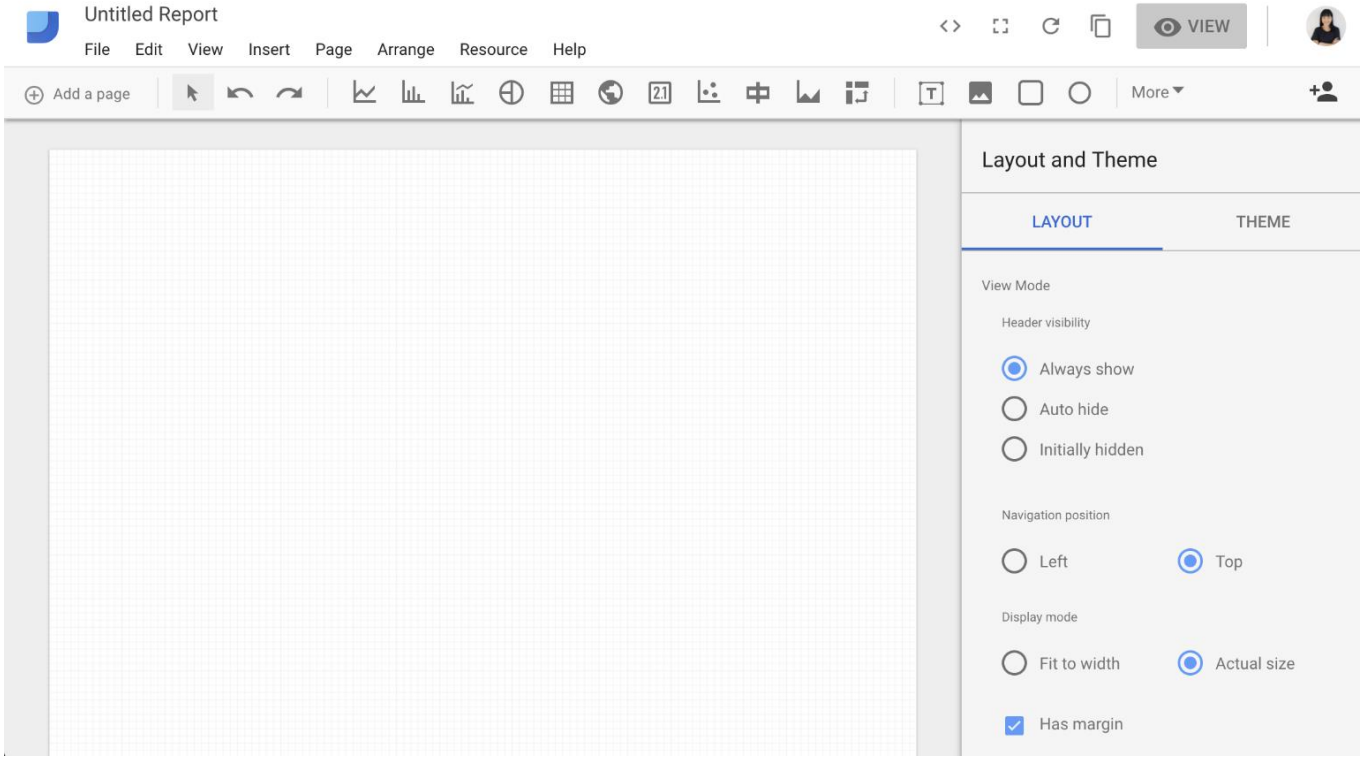
Let's play with some data.

- Visit https://goo.gl/1Tnfvr
- Click "Create Report" button
- Agree to terms (if asked)
- Click "Add to Report" button



Guided Walk-Through:
Create a Report

10 minutes



Guided Walk-Through:
Report Header

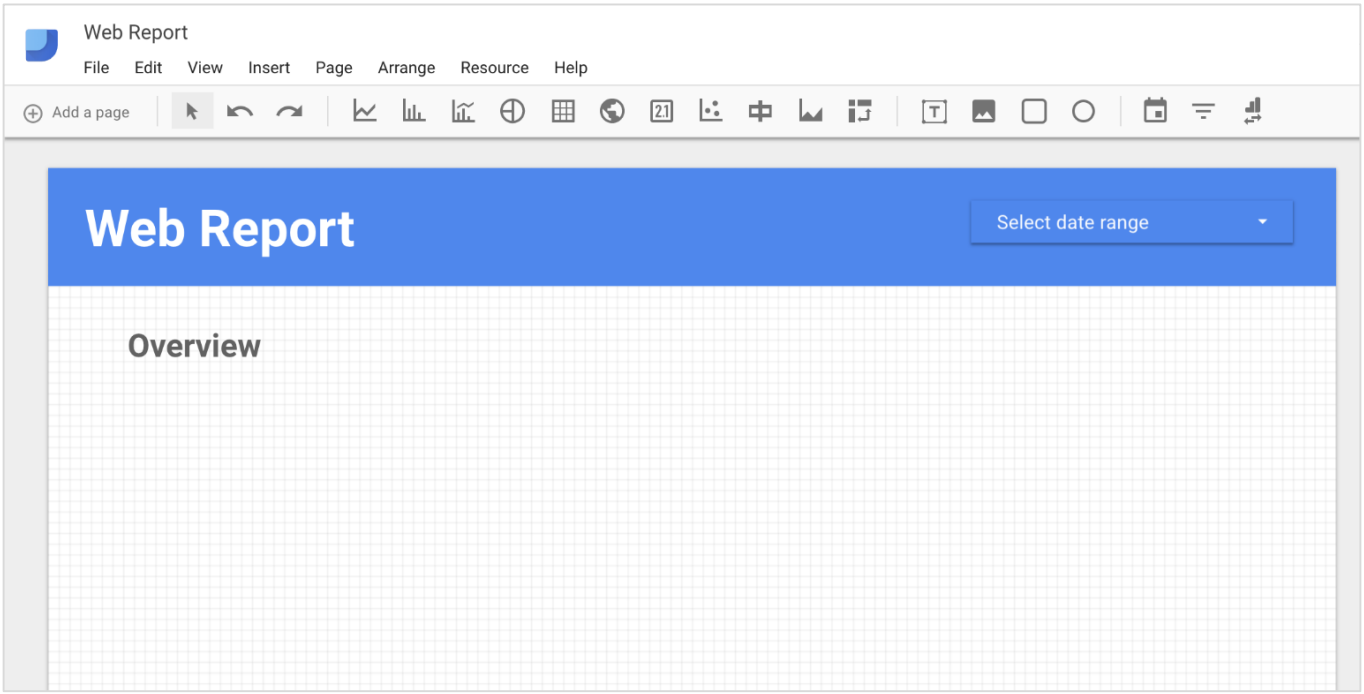
10 minutes

Let's make the report look nice.

1. Click the "Rectangle" icon.
2. Format the shape so that it creates a header across the top of the page.
3. Click the "Text" icon.
4. Write in a title and format the text.
5. Click the "Calendar" icon.
6. Set the Default Date Range as "Last 30 days".
7. Format the date selector.

Guided Walk-Through:
Report Header

10 minutes



Guided Walk-Through:
Make Report-Level

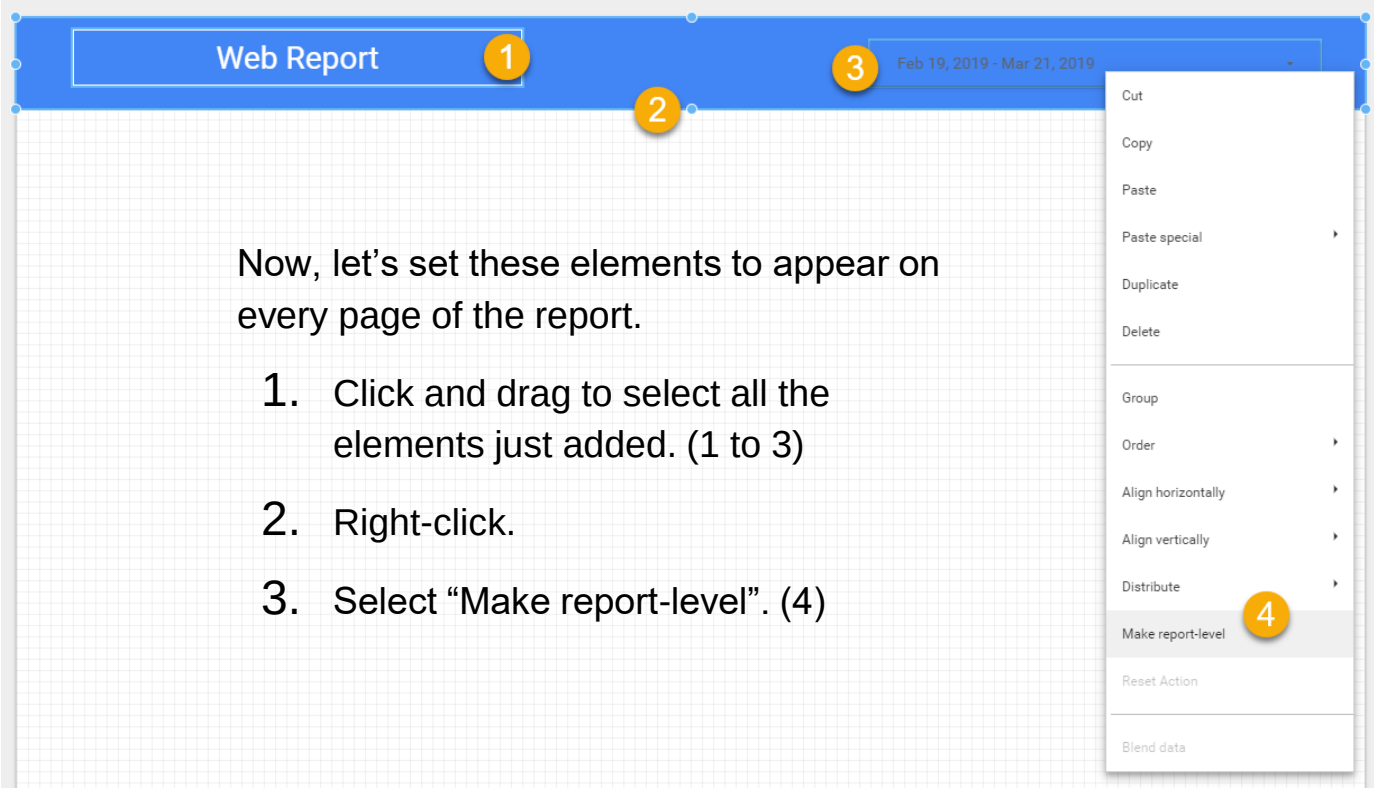
10 minutes

Now, let's set these elements to appear on every page of the report.

1. Click and drag to select all the elements just added.
2. Right-click.
3. Select "Make report-level".

Guided Walk-Through:
Make Report-Level

10 minutes



Now, let's set these elements to appear on every page of the report.

1. Click and drag to select all the elements just added. (1 to 3)
2. Right-click.
3. Select "Make report-level". (4)

Guided Walk-Through:
Data Time

10 minutes

Let's add some data!

- Click "Add a chart" and add the following charts:
  - Time series
  - Scorecard
  - Table
  - Stacked bar chart

Solo Exercise:

Data Time

15 minutes

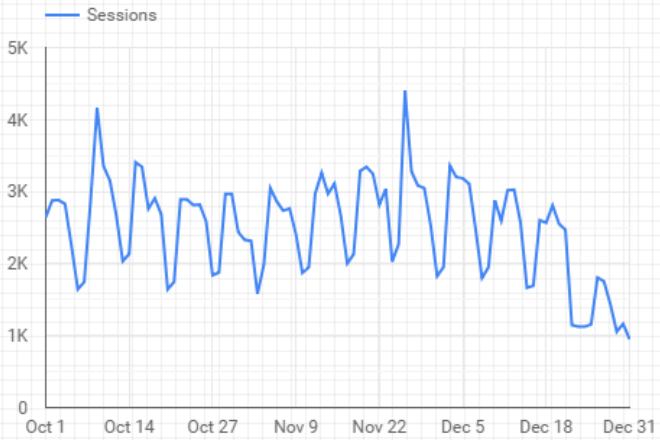
Edit each of the charts to show…

Time Series	Scorecard	Table	Stacked Bar Chart
Sessions on the site in Q4 2018	Transactions and revenue and change from the previous year (hint: add another scorecard)	Number of transactions that acquisition channels are generating	Daily traffic by country of origin

(and don’t forget - make ‘em look nice!)

Web Report

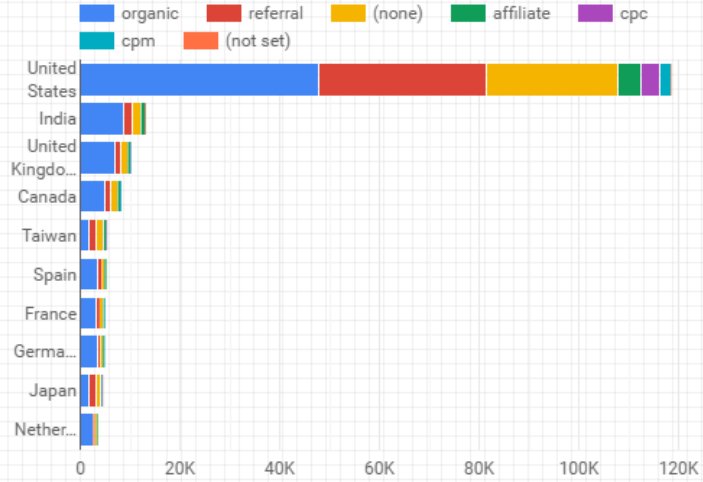
Oct 1, 2018 - Dec 31, 2018



	Source	Transactions
1.	google	365
2.	mall.googleplex.com	259
3.	(direct)	167
4.	sites.google.com	57
5.	Partners	17
6.	yahoo	5
7.	dfa	5

1 - 15 / 15

2018	2017
Transactions	Transactions
17,545	25,226
Revenue	Revenue
$2,967,171.33	$4,745,000.17



Which chart works?

Time Series	Change in one variable (or more) over a period of time
Combo Chart	Compare three data sets to visualize a correlation
Scorecard	A running performance tally (e.g. revenue, transactions)
Bar Chart	Comparison of different items, or comparison of items over time
Stacked Bar Chart	Demonstrate the composition of an item being compared
Pie Chart	How different categories represent parts of a whole
Scatter Chart	Demonstrate the relationship between two variables or distribution trends
Table	Comparison of different items

You want to show new user growth over the course of a quarter. Use a...

A. Bar chart

B. Scatter chart

C. Scorecard

D. Line chart

You want to show new user growth over the course of a quarter. Use a…

D. Line chart

You're demonstrating the relationship between email opens and website visitors during a campaign that ran for 2 weeks. Use a…

A. Time series

B. Scatter chart

C. Table

D. Scorecard

+10 points

You're demonstrating the relationship between email opens and website visitors during a campaign that ran for 2 weeks. Use a...

B. Scatter chart

Group Exercise:

Data Studio Dashboard

30 minutes

You have been asked to build a Dashboard for either Claire, Maggie, Clark or Tony.

You must build a dashboard that is most relevant to each person. Think beyond the needs they have requested and identify other reports that could be useful for them.

Claire, Owner

Clark, Product Manager

Maggie, Marketer

Tony, Developer

Group Exercise:

Data Studio Dashboard

30 minutes

Claire, Owner

- Learn more about the customers using the site
- Get deeper insights per segment
- Demographics Overview turned on
- Event Tracking implemented

Clark, Product Manager

- Can only produce 1,000 units per product per month
- Know when he's getting close to the limit and if he needs support
- Know what products are selling well to plan resources

Maggie, Marketer

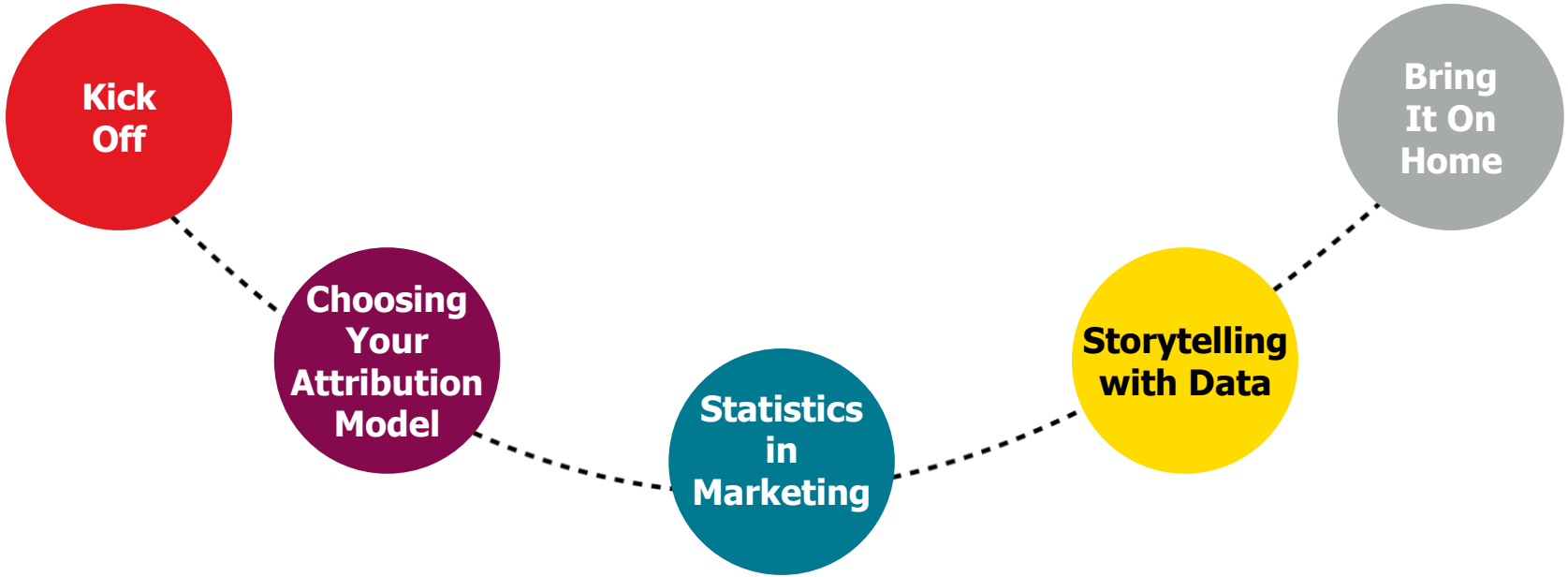
- Understand how long people take to convert on the site
- Identify channels people visit before converting
- Know how Organic and Paid Search contribute to conversions

Tony, Developer

- Reduce page load time
- Justify improving the search feature
- Insights on check out process to make it easier to buy online
- Where in the funnel he's losing visitors

Bring it on Home

Where We've Been



Solo Exercise:

Reflection Time

5 minutes

Let's take the time to reflect on what you learned today.

Complete the answer to 3 of the following statements (in an email to yourself, in your notebook… this is just for you).

I don't want to forget…

The best thing I got out of today was…

When I get back to work, I will…

I'm going to start…

I want to learn more about…

One More Thing...

Fill out an exit ticket!

(add link to exit ticket)