

# INTRO TO DATA ANALYTICS

---

HELLO!

---

## NIKOLAI CHOWDHURY



- Director, Customer Success @ App Annie
- Founder @Morescape - Technology Consultancy
  
- Previously: Mobile Engineer, Product Manager & Owner, R&D innovation leader
- MBA, Columbia Business School

---

## **OPENING**

---

# **LEARNING OBJECTIVES**

- ▶ Explain the relevance of data analysis across industries and job functions.
- ▶ Evaluate the quality and structure of a dataset.
- ▶ Use Google Sheets to perform descriptive and exploratory analysis on datasets.
- ▶ Use data analytics to inform business recommendations.

---

**DATA ANALYTICS 101**

---

# PRE-WORK

---

# PRE-WORK

---

- Are you able to access Google Sheets?
  - If yes, awesome!
  - If not, create a free Google account by visiting: [accounts.google.com](https://accounts.google.com)

---

**DATA ANALYTICS 101**

---

# **INTRODUCTION: DEFINING DATA**

---

## DEFINING DATA

---

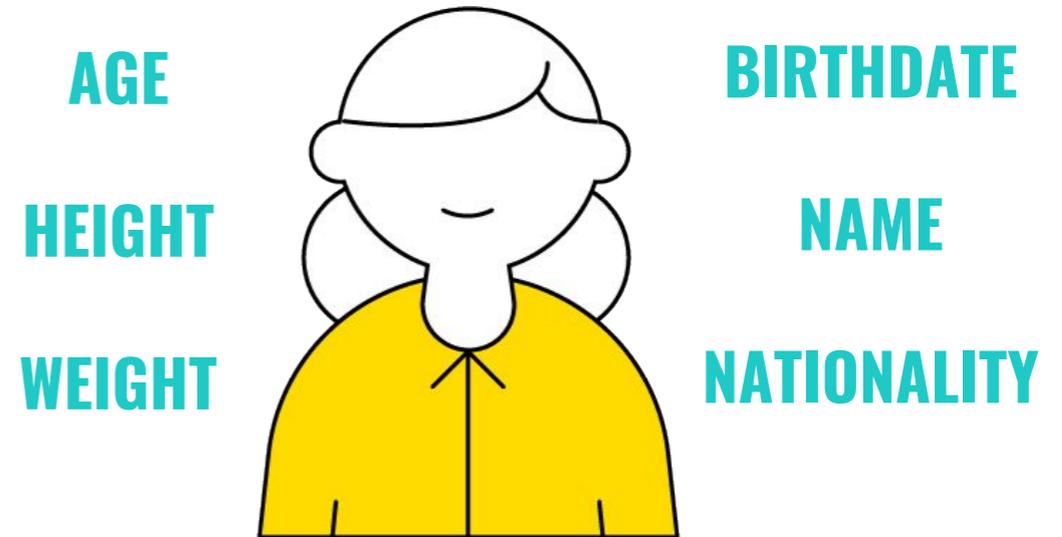
- ▶ We live in a world filled with data, it's all around us. But, if it's all around us, what is it?
- ▶ Data is simply *information* that exists in a variety of formats and sizes.
- ▶ Data can be thought of as an “object” but it's not necessarily seen at first glance.

---

# DEFINING DATA

---

- ▶ Even you have lots of data attached to you.
- ▶ You have an age, a height, a weight, a birthdate, a name, and a nationality.
- ▶ Each piece of information is considered data!

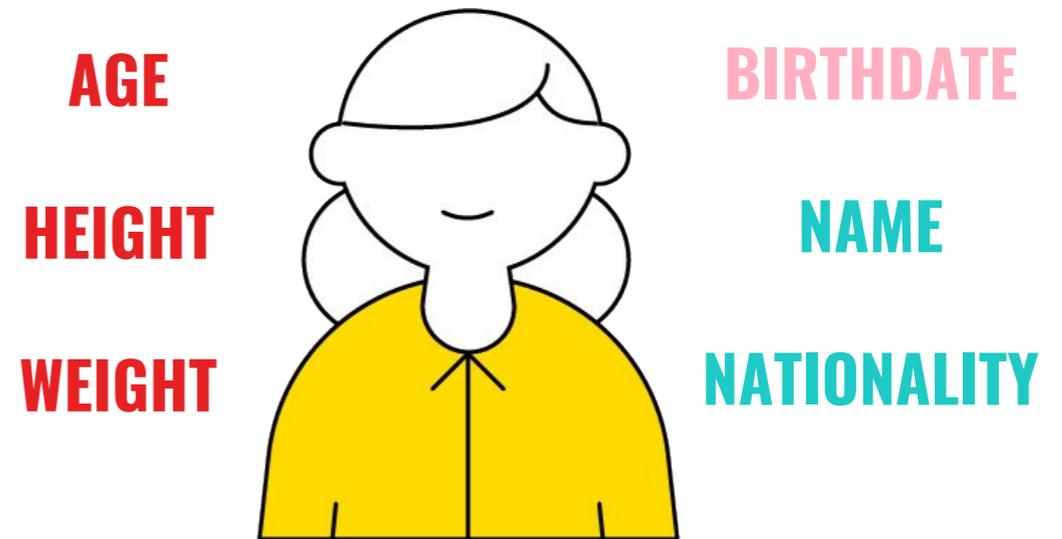


---

# DEFINING DATA

---

- ▶ This data, only about you, represents a variety of data types.
- ▶ Your age, height, and weight are all **numbers**.
- ▶ Your birthday is a **date**.
- ▶ Your name and nationality are text (or **strings**, as we call them in data analytics).



---

## DEFINING DATA

---

- Not only do you have data attached to you, but you're creating data with every turn.



- Say hello to your neighbor!
- Work with your neighbor to generate a list of 3-5 ways you have you created data today.

---

# DEFINING DATA

---

- ▶ You've created data in many ways today, perhaps by:
  - ▶ Logging into a website
  - ▶ Making a purchase with a credit/debit card
  - ▶ Taking public transportation
  - ▶ Making a phone call
  - ▶ Registering for a class at General Assembly

---

## **DEFINING DATA**

---

- Data Analytics is the process of transforming raw data, perhaps like the types you just mentioned, with the purpose of drawing conclusions about that information.

**LESSON: HOW DATA  
ANALYST THINK ABOUT  
DATA**

---

## HOW DATA ANALYSTS THINK ABOUT DATA

---

- ▶ The rate at which data is being created is rapidly accelerating. The key for successful data analysis is taking data and forming actionable conclusions and insights.
- ▶ Data Analysts do this by using a “workflow” to guide to them through the process.
- ▶ What might be some of the steps in this process?

# HOW DATA ANALYSTS THINK ABOUT DATA

- Identify the Problem
  - Before you begin working with any data, you must understand the problem that you're trying to answer.



# HOW DATA ANALYSTS THINK ABOUT DATA

- ▶ Obtain the Data
  - ▶ To work with the data, you first have to find it or collect it, and it has to be the right data to help you answer the question.



# HOW DATA ANALYSTS THINK ABOUT DATA

- Understand the Data
  - Then you need to ensure you can correctly interpret the results and trust the data.



# HOW DATA ANALYSTS THINK ABOUT DATA

- Prepare the Data

- You should making sure the data doesn't contain incorrect or missing values.



# HOW DATA ANALYSTS THINK ABOUT DATA

- Analyze the Data
  - Now, you are ready to uncover the answer to your question, assuming you haven't ended up at a prior step due to missing data or a poorly understood question.



# HOW DATA ANALYSTS THINK ABOUT DATA

- Present the Results
  - Assuming you find what you are looking for, and it seems compelling enough to share with others, you need to determine the best way to share your results with others.



---

## HOW DATA ANALYSTS THINK ABOUT DATA

---

- ▶ These general steps are necessary for each and every data analysis you do; but like a snowflake, each time will be a little different, as well.
- ▶ One of the most important things you'll notice is that the workflow is not strictly linear. Even though you begin at the top and end at the bottom, you will revisit various steps along the way as needed.
- ▶ Instead of thinking of this as step-by-step instructions for doing data analysis, you should think about these stages as the guiding principles for analyzing data. Experience in navigating these competing factors is what separates a good data analyst from a great one.

**DATA ANALYTICS 101**

---

# **LESSON: TOOLS OF THE DATA ANALYST**

---

## **TOOLS OF THE DATA ANALYST**

---

- ▶ In order to proceed through the workflow, a data analyst uses a suite of tools to assist along the way.
- ▶ While data analysts can apply a specialized suite of tools, the analyst's judgement and intuition is the most important tool.
  - ▶ Is there a tool that will “Identify the Question” or “Understand the Data” for you? Likely not; the analyst themselves become the best tool in this scenario.
- ▶ A Data Analyst may use tools like: Google Sheets, Excel, SQL, or Tableau.

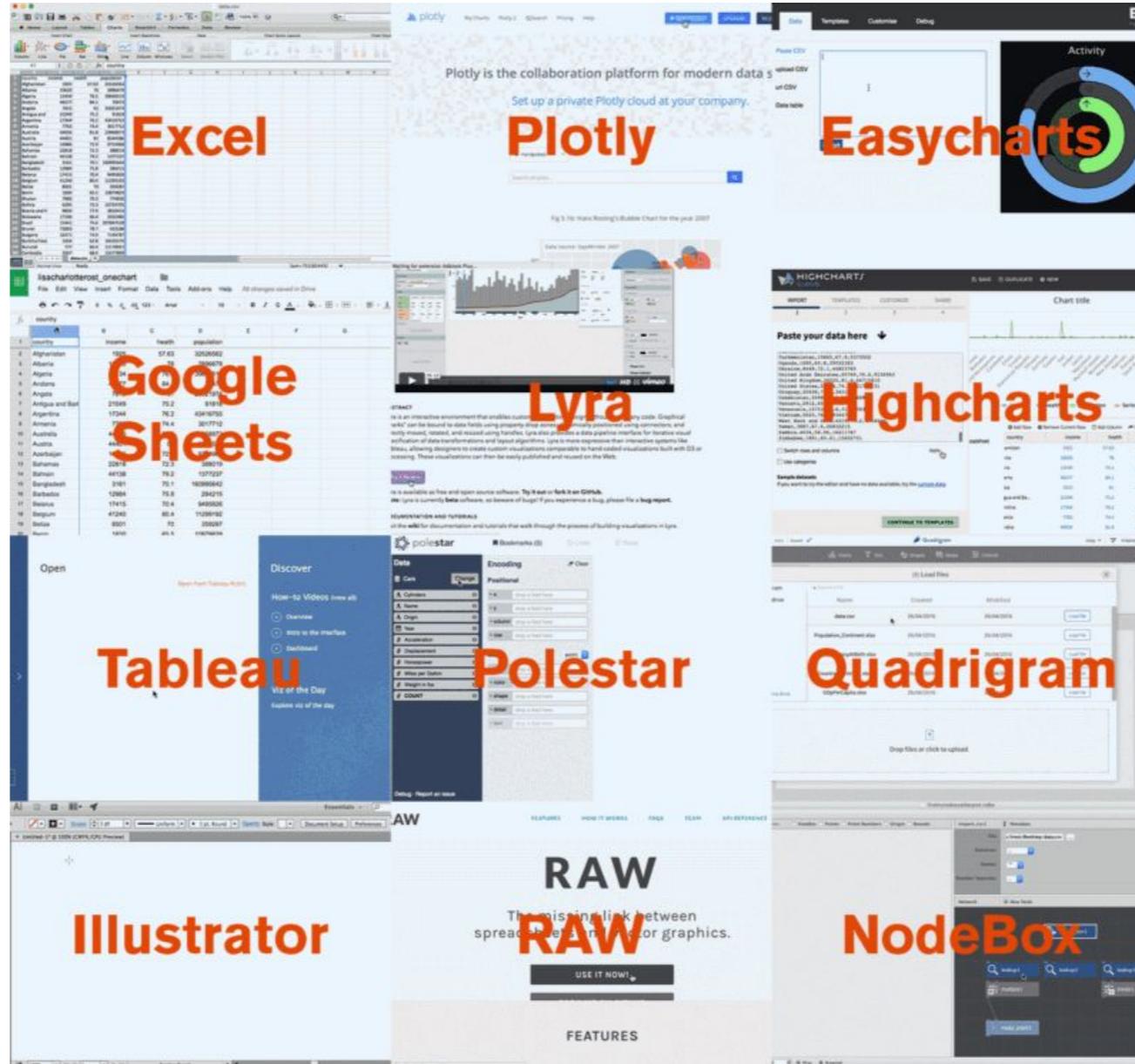
---

## **TOOLS OF THE DATA ANALYST**

---

- One thing is for sure: there is no shortage of data tools to use for purposes that range from analysis to visualization.
- Lisa Charlotte Rost, a fellow at National Public Radio, set out to make the same chart, using the same data, but by using multiple data tools. Let's look at some of the graphs she created!

# TOOLS OF THE DATA ANALYST



---

## TOOLS OF THE DATA ANALYST

---

- Rather than saying that one tool was superior, Lisa distilled a few lessons about data tools:
  - There are no perfect tools, just good tools for people with certain *goals*.
  - There are no perfect tools, just good tools for people with certain *mindsets*.
  - Every tool forces you down a path.

**DATA ANALYTICS 101**

---

# **INTRODUCTION: USING DATA TO INFORM BUSINESS DECISIONS**

---

# USING DATA TO INFORM BUSINESS DECISIONS

---

- ▶ It's time to get our hands dirty working with real data. It looks like a job for you:

*Chipotle Founder, Steve Ells, has hired a team of General Assembly (GA) Analysts to help him assess the current state of Chipotle to help plan for future growth of the business. Steve has task the GA Analysts with helping him document Chipotle's current growth patterns across geographies and to better understand the "sides" purchased by customers at Chipotle restaurants.*

*The GA Analysts will provide insights for Chipotle's strategy within the next fiscal year related to (1) restaurant expansion and (2) menu changes.*

- ▶ Before we jump into any analysis, let's [watch a clip](#) to better understand Chipotle's business.

---

# USING DATA TO INFORM BUSINESS DECISIONS

---

- We will want to consider Chipotle's strategy in two directions: expansion plans and menu alterations.
- We have a clear direction on which "business" questions to focus our analysis on, but we will want to provide insights based on "data" questions.



- Working in pairs, consider what types of data we might need to gather to answer these questions.
  - Consider the "information" we would want to gather.
  - What sources would gather this from?

**GUIDED PRACTICE:  
ANALYZING PROPERTY  
GROWTH AT CHIPOTLE**

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- ▶ Let's access the Google Sheet we will be using for the remainder of today.
  - ▶ Visit this link: <http://nikolai.mobi/2fVSc4m>
  - ▶ Next, you'll need to click on "File" and "Make a copy..." to create a version of the file on your own Google Drive.
    - ▶ This Google Sheet is in "View only" mode, so you can't edit it.
- ▶ When you copy the file to your Google Drive, you'll have full edit access to the copied version!

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- Let's talk through some important vocabulary related to the Google Sheets environment:
  - **Workbook:** A collection of worksheets.
  - **Worksheet:** The area where data is arranged and calculations are performed.
  - **Column:** A vertical collection of cells.
  - **Row:** A horizontal collection of cells.
  - **Cell:** The intersection of a column and row on a worksheet.
  - **Array:** A collection of cells in a row, column, or across rows or columns.
  - **Function:** Instructions for calculations.

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- › We are getting ready to start **exploratory data analysis** (EDA), where we summarize and display data in such a way that interesting features become apparent.
  - › Data is often “rectangular” (i.e. defined by row and columns); EDA uses graphical and non-graphical methods to break down this “rectangle” using univariate (one variable), bivariate (two variables), or multivariate methods (2+ variables).

---

## ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- › How many total properties did Chipotle operate?
  - › Syntax: =SUM(number1, [number2],...)
  - › Solution in J3:=SUM(B2, B3, ..., B52) or =SUM(B2:B52)= 2010
- › How many Chipotle properties did Chipotle operate in the US?
  - › Solution in J4:=SUM(B2:B48) =1987
- › How many Chipotle properties did Chipotle outside the US?
  - › Solution in J5:=SUM(B49:B52) or =J3-J4 =23

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- Which geography had the most Chipotle properties?
  - Syntax: `=MAX(number1, [number2],...)`
  - Solution in J6: `=MAX(B2:B52)= 351`
  - Note: 351 is the largest number of properties in a geography, but does not indicate the geography itself. California has the most Chipotle properties.
- Which geography had the least Chipotle properties?
  - Syntax: `=MIN(number1, [number2],...)`
  - Solution in J7: `=MIN(B2:B52)=1`
  - Note: There are multiple geographies with one Chipotle property, North Dakota, Vermont, Wyoming, Germany.

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- Now that we've done some analysis of Chipotle's 2015 property data, let's turn this analysis into a recommendation.
  - Running the analysis and summarizing the result is only part of the process, drawing a recommendation or insight from the data is where analytics can add value to Chipotle's business.
- We'll form recommendations in three steps:
  - 1) Summarize (stating the results)
  - 2) Recommend (interpret the results and draw a conclusion)
    - What is this data saying?
    - What can we infer from this data?
  - 3) Future Analysis (consider the shortcomings of the data and plan for future analysis)

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- We can form a recommendation from the question: Which geography had the most Chipotle properties?
  - 1) **Summarize:**
    - California is the geography with the largest number of Chipotle properties.
  - 2) **Recommend:**
    - Since there are so many properties in this geography, adding additional properties in 2016 could cannibalize store sales. However, this could also indicate that the California geography has a growing taste for Chipotle's products.
  - 3) **Future Analysis:**
    - Further research should be done to see how these properties are distributed across California. Are most locations in the same part of the geography or are they distributed throughout?
- Note that recommendations and insights from data could request that more analysis be conducted on the existing dataset or that a different dataset be analyzed.

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- ▶ When first approaching a new dataset, it's helpful to understand the data's structure and look for any anomalies.
  - ▶ This data includes multiple years, so we have introduced time as a variable to consider.
  - ▶ Most importantly, there are missing data points for some rows of data.
- ▶ How do we address missing data?
  - ▶ Step 1: Research the reason for the omission of data: Is there actually no data available? Was it a data entry error?
  - ▶ Step 2: Make a decision on how to correct the omission: Leave it, delete it, change it (edit to correct value, use imputation, etc.)

---

## **ANALYZING PROPERTY GROWTH AT CHIPOTLE**

---

- ▶ Empty cells in the dataset represent years where there were no properties in the geography.
- ▶ The missing data represent years in which there were zero properties, and 0 can safely be inserted in empty cells.

---

# ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- › To build the scatter plot, start by summing the total number of properties for each year to complete the table:

Year	Total Properties
2015	=SUM(B2:B52) =2010
2014	=SUM(C2:C52) =1783
2013	=SUM(D2:D52) =1595
2012	=SUM(E2:E52) =1450
2011	=SUM(F2:F52) =1230
2010	=SUM(G2:G52) =1084
2009	=SUM(H2:H52) =956

---

## ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- › Which geography has the largest percent change from 2009 to 2015?
- › The **percent change formula** calculates the extent of change between a measurement taken at two different times: “initial/old/before” and “final/new/after”

$$\textit{Percent Change} = \frac{\textit{New Value} - \textit{Old Value}}{|\textit{Old Value}|} \times 100$$

---

## ANALYZING PROPERTY GROWTH AT CHIPOTLE

---

- › Relating to the Chipotle data, our percent change formula looks like:

$$\textit{Percent Change for Properties} = \frac{\# \textit{ of 2015 Properties} - \# \textit{ of 2009 Properties}}{|\# \textit{ of 2009 Properties}|} \times 100$$

- › Define column I as “% Change”
- › I2= ((B2-H2)/H2)\*100
- › Drag the formula from I2 to I52 using the “square” in the right hand corner of the cell.
- › The largest percent change occurs in South Carolina, where growth from 2009 to 2015 was 1500%.

**DATA ANALYTICS 101**

---

# **DEMO: STRUCTURING SALES DATA AT CHIPOTLE**

---

# **STRUCTURING SALES DATA AT CHIPOTLE**

---

- Let's take a look at a new dataset that describes sales data for Chipotle.
- We'll notice that the size of this dataset is much bigger than our property data! No need to fear!

---

# STRUCTURING SALES DATA AT CHIPOTLE

---

- ▶ First, what do we know about this data?
  - ▶ The data contains approximately 3,000 meals in about 1,800 Grubhub orders from July to December 2012.
    - ▶ The dataset assumes an individual's eat only one salad, burrito bowl, burrito or set of tacos at a time. When more than one of these items appears in an order, this is counted as a second individual's order.
    - ▶ Side items, like chips, are assumed to be split among all individuals in an order.
  - ▶ Most of the sales data comes from two Chipotle restaurants, one in Washington, D.C., and another in East Lansing, Michigan.
    - ▶ Ordering behavior might vary around the country!
  - ▶ Some menu items, like sofritas and brown rice, have been introduced nationwide since this data was collected.

**DATA ANALYTICS 101**

---

# **GUIDED PRACTICE: ANALYZING SALES DATA AT CHIPOTLE**

---

## ANALYZING SALES DATA AT CHIPOTLE

---

- A **frequency distribution** shows the number of instances a variable takes on a specific value.
  - In our dataset, we are broadly interested in the variable “Chips” and all the different combinations of chips that are purchases. The frequency is related to the “Quantity” purchased of each “Item\_Name”.
- Find the frequency of each type of chip (using the Quantity and Item\_Name columns).
  - Hint: Use the =SUM(number1, [number 2],...) syntax to calculate this value.

---

# ANALYZING SALES DATA AT CHIPOTLE

---

Type of Chips	Frequency
Chips	=SUM(B2:B212) =230
Chips and Fresh Tomato Salsa	=SUM(B213:B322) =130
Chips and Guacamole	=SUM(B323:B801) =506
Chips and Mild Fresh Tomato Salsa	=SUM(B802) =1
Chips and Roasted Chili Corn Salsa	=SUM(B803:B842) =41
Chips and Tomatillo-Green Chili Salsa	=SUM(B843:B916) =78
Chips and Tomatillo-Red Chili Salsa	=SUM(B917:B984) =75
Side of Chips	=SUM(B985:B1085) =110

---

# ANALYZING SALES DATA AT CHIPOTLE

---

▶ Reference [Chipotle's menu](#) for clarity on internal naming conventions:

+		<b>CHIPS</b>
+		<b>CHIPS AND GUACAMOLE</b>
+		<b>CHIPS &amp; FRESH TOMATO SALSA</b>
+		<b>CHIPS &amp; ROASTED CHILI- CORN SALSA</b>
+		<b>CHIPS &amp; TOMATILLO GREEN-CHILI SALSA</b>
+		<b>CHIPS &amp; TOMATILLO RED- CHILI SALSA</b>

---

## **ANALYZING SALES DATA AT CHIPOTLE**

---

- ▶ Chipotle's menu lists six different types of chips, however, our data file contains eight types of chips.
  - ▶ “Chips” and “Side of Chips” can be combined into “Chips”
  - ▶ “Chips and Fresh Tomato Salsa” and “Chips and Mild Fresh Tomato Salsa” can be combined into “Chips and Fresh Tomato Salsa”

---

## ANALYZING SALES DATA AT CHIPOTLE

---

Type of Chips	Frequency
Chips	$=H_5+H_{12} = 340$
Chips and Fresh Tomato Salsa	$=H_6+H_8 = 131$
Chips and Guacamole	$=H_7 = 506$
Chips and Roasted Chili Corn Salsa	$=H_9 = 41$
Chips and Tomatillo-Green Chili Salsa	$=H_{10} = 78$
Chips and Tomatillo-Red Chili Salsa	$=H_{11} = 75$

**INDEPENDENT  
PRACTICE:  
SUMMARIZING THE  
FINDINGS**

---

# SUMMARIZING FINDINGS

---



## ACTIVITY

The GA Analysts have been invited to Chipotle’s Headquarters in Denver, CO to share the results of their analysis. While preparing for your presentation to senior managers, you bump into Steve Ells, Chipotle’s founder in the hallway. He’s excited and asks, “I’m on my way to another meeting, but what did you find?”

Working in pairs, and using our format for a recommendation based on data, what one recommendation do you share with him? Remember, you’ll only have his attention for a minute, at most!

1. Summarize (stating the results)
2. Recommend (interpret the results and draw a conclusion)
  - a. What is this data saying?
  - b. What can we infer from this data?
3. Future Analysis (consider the shortcomings of the data and plan for future analysis)

---

**DATA ANALYTICS 101**

---

# CONCLUSION

---

## **CONCLUSION**

---

- ▶ Today, we've seen the Data Analytics workflow in action!
- ▶ We've identified questions to answer with data; we took the necessarily steps (obtaining, understanding, and preparing our data) prior to starting our analysis; then, we performed analysis. Finally, we've used our analysis to form actionable insights from our data.

---

# CONCLUSION

---

- Want more help?
  - [Google Sheets Help Documentation](#) by Google
- Want to continue your data journey at GA?
  - SQL Bootcamp
  - Data Analysis Circuit ([Part-Time Online](#))
  - Data Analytics ([Part-Time or 1-Week Accelerated](#))
  - Data Science ([Part-Time](#) or [Immersive](#))

# DATA ANALYTICS 101

---

# Q&A

---

THANKS!

---

## NIKOLAI CHOWDHURY

- LinkedIn: <http://linkedin.com/in/nikolaic>
- Email: [n@nikolai.in](mailto:n@nikolai.in)
- Website: <http://www.nikolai.in>
- Twitter: [@chowdhury](https://twitter.com/chowdhury)

# CITATIONS

- › Chipotle Mexican Grill, Inc., Annual Reports, 2009 - 2015:  
<http://ir.chipotle.com/phoenix.zhtml?c=194775&p=irol-reportsAnnual>
- › New York Times, “At Chipotle, How Many Calories Do People Really Eat?”:  
<http://www.nytimes.com/interactive/2015/02/17/upshot/what-do-people-actually-order-at-chipotle.html>