

Advanced Data Analytics

Agenda

1. Approach a Business Task
2. Different Viewpoints
 - Data Analysis Process (Udacity)
 - Data Analytics Life Cycle (EMC2)
 - Data Analysis Life Cycle (Google)
 - Data Analysis Process (Role of Looker)
3. Questions

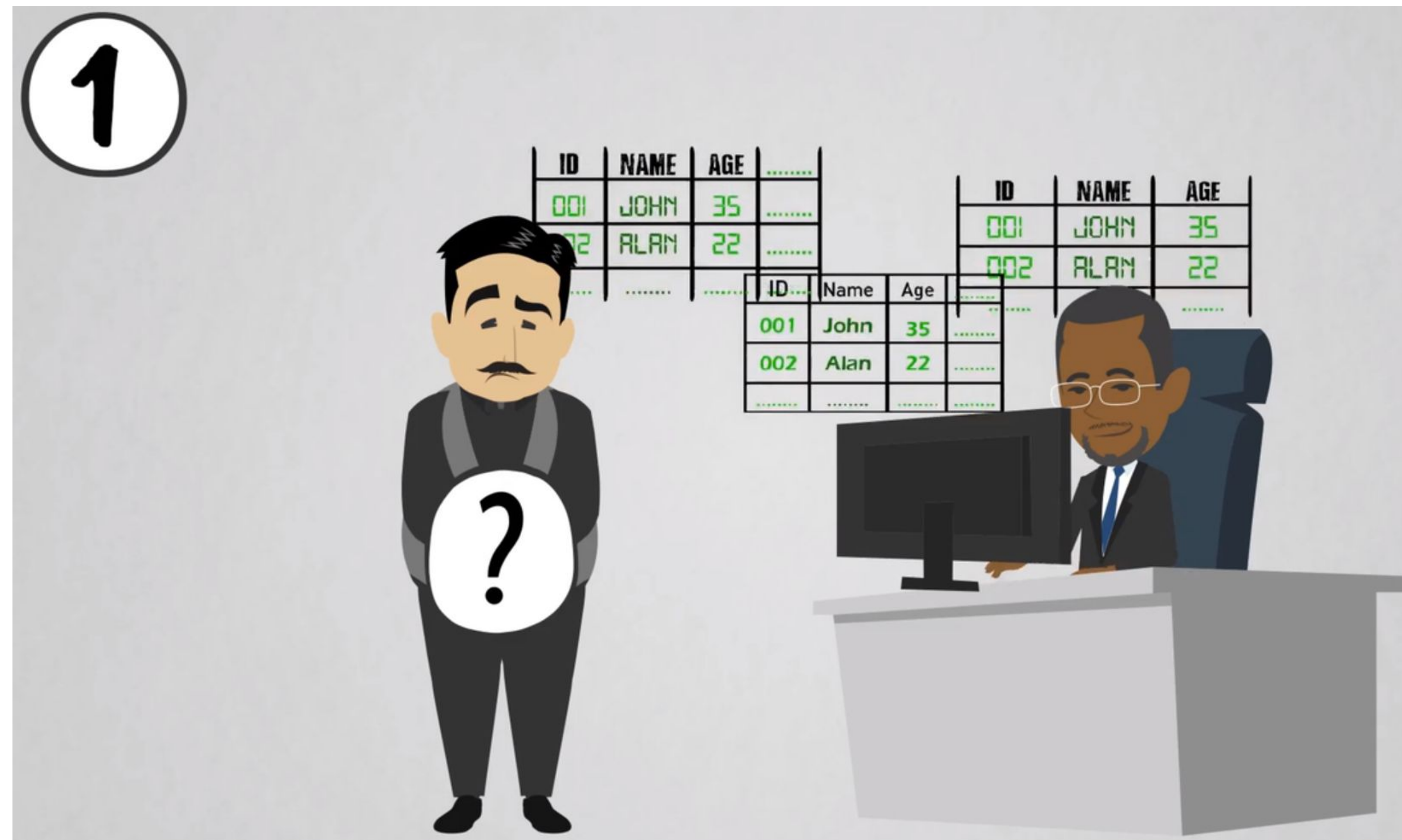


Approach a Business Task

Scenario 1

The boss

- has read the reports/dashboards
- want you to make some predictions for the firm's outgoing costs over the next year



Scenario 1

The logical way to approach this problem is to:

- gather some relevant data
- then prepare it for analysis



Scenario 2

The boss says

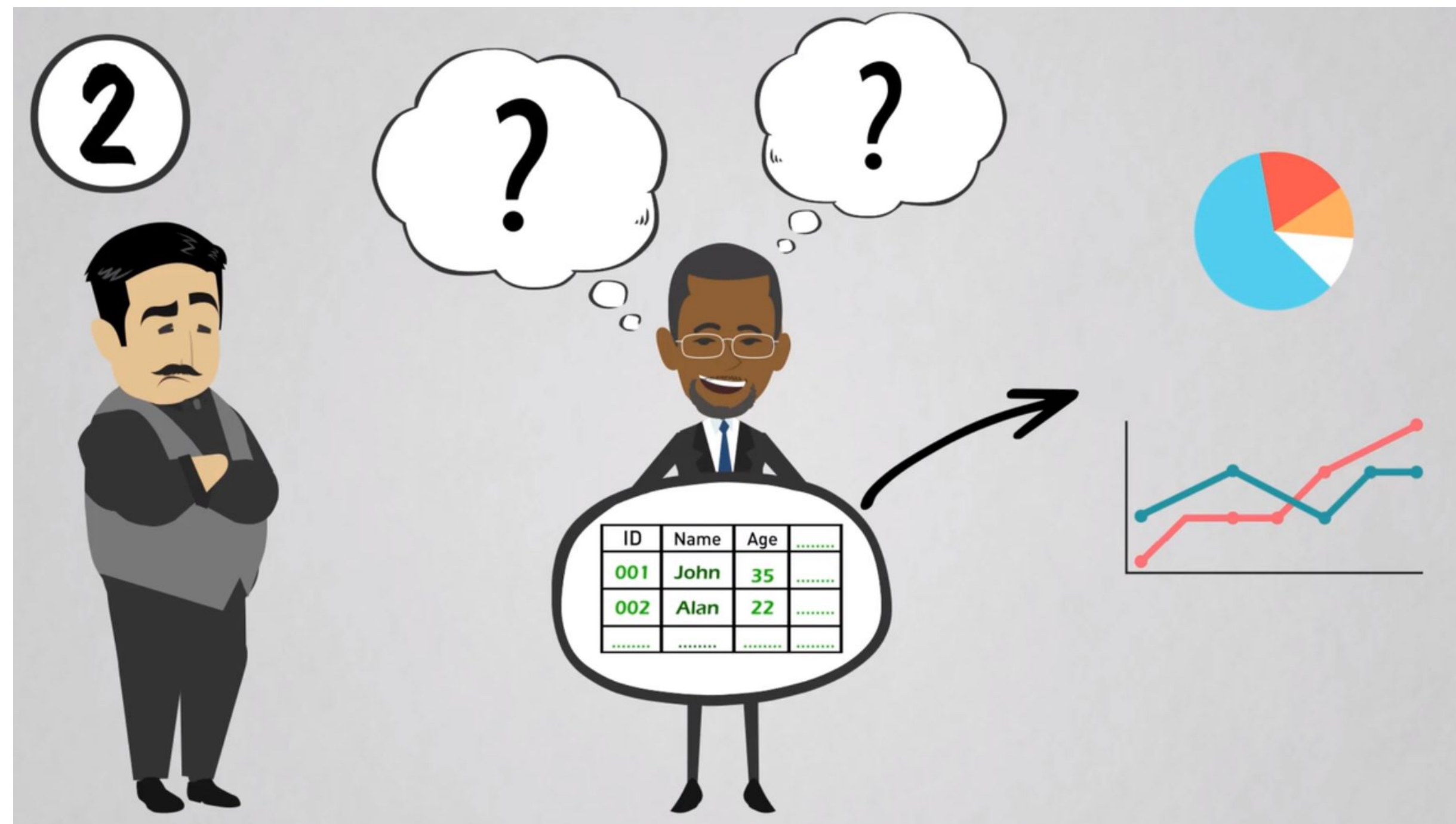
- We have an enormous amount of data
- We don't know what we could do with it but it must be useful
- Can you do something with it, such as:
 - Tell us how we could increase our profit for next year



Scenario 2

In this case

- Having the dataset is the starting point
- You don't need to collect data to answer a business question
- You can analyse it and apply different analytics tools to extract insights and make forecasts



In both Scenario 1 & Scenario 2

- The solution to any task begins with having a proper dataset
- This must be first on the to-do list
- Only then, we can proceed with
 - further analysis
 - and forecasting



Different Viewpoints



Data Analysis Process

Data Analysis Process (From Udacity Nanodegree)

Lesson 1:
The Data Analysis Process

SEARCH

RESOURCES

CONCEPTS

✓ 1. Handoff to Juno Lee

✓ 2. Lesson Overview

✓ 3. Problems Solved by Data Analysts

✓ 4. Setting Up Your Programming E...

✓ 5. Data Analysis Process Overview

✓ 6. Data Analysis Process Quiz

✓ 7. Packages Overview

✓ 8. Packages Overview Quiz

✓ 9. Asking Questions

✓ 10. Questions for a Dataset

✓ 11. Data Wrangling and EDA

≡

Data Analysis Process Overview

visualiz 1/3 ^ v x

Step 1: Ask questions

Either you're given data and ask questions based on it, or you ask questions first and gather data based on that later. In both cases, great questions help you focus on relevant parts of your data and direct your analysis towards meaningful insights.

Step 2: Wrangle data

You get the data you need in a form you can work with in three steps: gather, assess, clean. You gather the data you need to answer your questions, assess your data to identify any problems in your data's quality or structure, and clean your data by modifying, replacing, or removing data to ensure that your dataset is of the highest quality and as well-structured as possible.

Step 3: Perform EDA (Exploratory Data Analysis)

You explore and then augment your data to maximize the potential of your analyses, visualizations, and models. Exploring involves finding patterns in your data, visualizing relationships in your data, and building intuition about what you're working with. After exploring, you can do things like remove outliers and create better features from your data, also known as feature engineering.

Step 4: Draw conclusions (or even make predictions)

This step is typically approached with machine learning or inferential statistics that are beyond the scope of this course, which will focus on drawing conclusions with descriptive statistics.

More on machine learning: [Machine Learning Engineer Nanodegree](#)

Step 5: Communicate your results

You often need to justify and convey meaning in the insights you've found. Or, if your end goal is to build a system, you usually need to share what you've built, explain how you reached design decisions, and report how well it performs. There are many ways to communicate your results: reports, slide decks, blog posts, emails, presentations, or even conversations. Data visualization will always be very valuable.

Before walking through each of these steps with real datasets using Python, let's build a bit of

Data Analysis Process (From Udacity Nanodegree)

1. Question

2. Wrangle

3. Explore

4. Draw Conclusions

5. Communicate

Data Analysis Process (From Udacity Nanodegree)

Step 1: Ask Questions

- Given data then ask questions, or
- Ask questions then **gather** data

Step 2: Wrangle Data

- a. **Gather** data to answer question
- b. **Assess** data to identify any problems in your data's quality or structure
- c. **Clean** data by modifying, replacing, or removing data

Data Analysis Process (From Udacity Nanodegree)

Step 3: Perform Exploratory Data Analysis (EDA)

- **Explore then augment** data to maximize the potential of:
 - analyses & visualizations & models
- **Exploring** involves:
 - finding **patterns** in data
 - **visualizing** relationships in data
 - building **intuition** about what you're working with
- **After Exploring (optional)**
 - **Remove Outliers:**
 - **Feature Engineering:** create better features from data

Data Analysis Process (From Udacity Nanodegree)

Step 4: Draw Conclusions (or even make **predictions**)

- typically approached with **inferential statistics** or **ML**

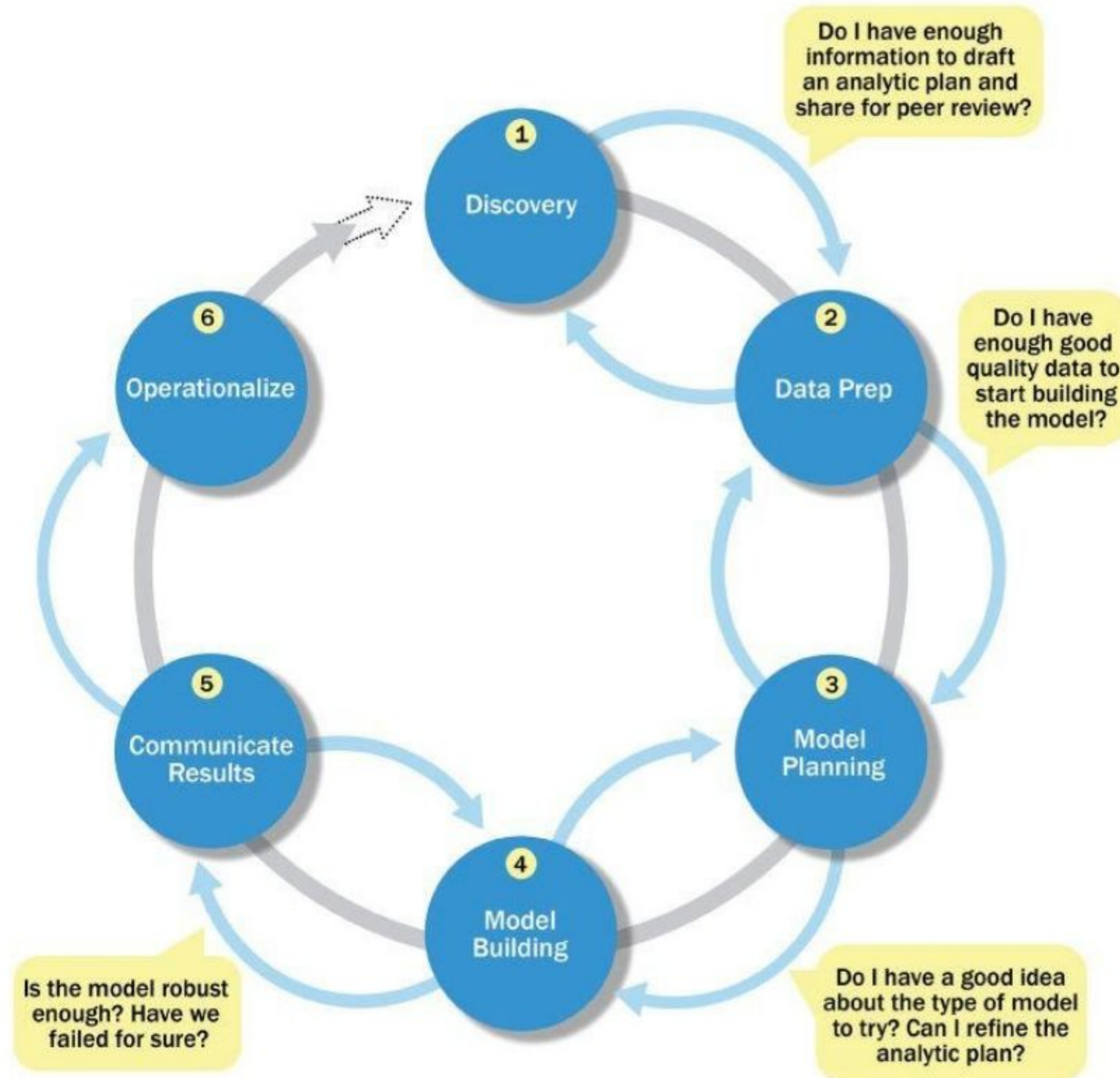
Step 5: Communicate Results

- often need to **justify** and **convey** meaning in the insights
- if your end goal is to build a system, you usually need to:
 - **share** what you've built
 - **explain** how you reached design decisions
 - **report** how well it performs
- communicate results by: report | slides | presentation | post | email | conversation
- **Data Visualization** will always be very valuable



Data Analytics Life Cycle

Data Analytics Life Cycle (from EMC2)



Data Analytics Life Cycle (from EMC2)

Phase 1: Discovery

- team **learns** the **business** domain
- team **assesses** the **resources** available to support the project
- **framing** the **business problem** as an **analytics challenge**
- **formulating** initial **hypotheses** to test and begin learning the data.

Data Analytics Life Cycle (from EMC2)

Phase 2: Data Preparation

- presence of an **analytic sandbox**
- Execute ELT or ETL to get data into the **sandbox**
 - Extract, Transform and Load (**ETL**)
 - Extract, Load, and Transform (**ELT**)
 - Data should be **transformed** so the team can work with it and analyze it
- team also needs to familiarize itself with the data thoroughly
- team may perform data **visualizations** to help understand the data,
 - including its trends, outliers, and relationships among data variables

Data Analytics Life Cycle (from EMC2)

Phase 3: Model Planning

- team **determines** the **methods**, **techniques**, and **workflow** it intends to follow
- team **explores** the **data** to learn about the relationships between variables
- Objective of the **data exploration** in this phase
 - understand relationships among variables to inform selection of the variables
 - A common way to conduct this step is to perform data **visualizations**

Phase 4: Model Building

- team **develops datasets** for testing, training, and production purposes
- team **builds/executes models** based on the work done in Model Planning
- team **considers** whether its existing **tools** will suffice for running the models

Data Analytics Life Cycle (from EMC2)

Phase 5: Communicate Results

- team **determines** if the **results** of the project are a **success** or a failure
- team **identify** key **findings**
- team **quantify** the **business value**
- team **develop** a **narrative** to summarize and convey findings to stakeholders
- The deliverable of this phase will be the **most visible** portion of the process to the outside stakeholders and sponsors

Data Analytics Life Cycle (from EMC2)

Phase 6: Operationalize

- team **delivers** final reports, briefings, code, and technical documents
- team may **run** a **pilot** project to implement the models in production
- Presentation for project sponsors:
 - contains high-level takeaways for executive level stakeholders,
 - with a few key messages to aid their decision-making process.
 - Focus on clean/easy **visuals** for presenter to explain and for the viewer to grasp
- Use imagery or data **visualization** when possible.
 - Although it may take more time to develop imagery,
 - people remember mental pictures to demonstrate a point more than long lists



Data Analysis Life Cycle

Data Analysis Life Cycle (From Google Data Analytics Professional Certificate)

coursera



Search in course

Search

Foundations: Data, Data, Every... > Week 1 > Origins of the data analysis process

Understanding the data ecosystem

- ✓ **Video:** What is the data ecosystem?
4 min
- ✓ **Video:** How data informs better decisions
4 min
- ✓ **Reading:** Data and gut instinct
10 min
- ✓ **Reading:** Origins of the data analysis process
20 min
- 📋 **Practice Quiz:** Test your knowledge on the data ecosystem
4 questions

Program expectations and proper use of the discussion forum

It is time to enter the **data analysis life cycle**—the process of going from data to decision. Data goes through several phases as it gets created, consumed, tested, processed, and reused. With a life cycle model, all key team members can drive success by planning work both up front and at the end of the data analysis process. While the data analysis life cycle is well known among experts, there isn't a single defined structure of those phases. There might not be one single architecture that's uniformly followed by every data analysis expert, but there are some shared fundamentals in every data analysis process. This reading provides an overview of several, starting with the process that forms the foundation of the Google Data Analytics Certificate.

The process presented as part of the Google Data Analytics Certificate is one that will be valuable to you as you keep moving forward in your career:

1. **Ask:** Business Challenge/Objective/Question
2. **Prepare:** Data generation, collection, storage, and data management
3. **Process:** Data cleaning/data integrity
4. **Analyze:** Data exploration, visualization, and analysis
5. **Share:** Communicating and interpreting results
6. **Act:** Putting your insights to work to solve the problem

Understanding this process—and all of the iterations that helped make it popular—will be a big part of guiding your own analysis and your work in this program. Let's go over a few other variations of the data analysis life cycle.

Data Analysis Life Cycle

(From Google Data Analytics Professional Certificate)

1. Ask

2. Prepare

3. Process

4. Analyse

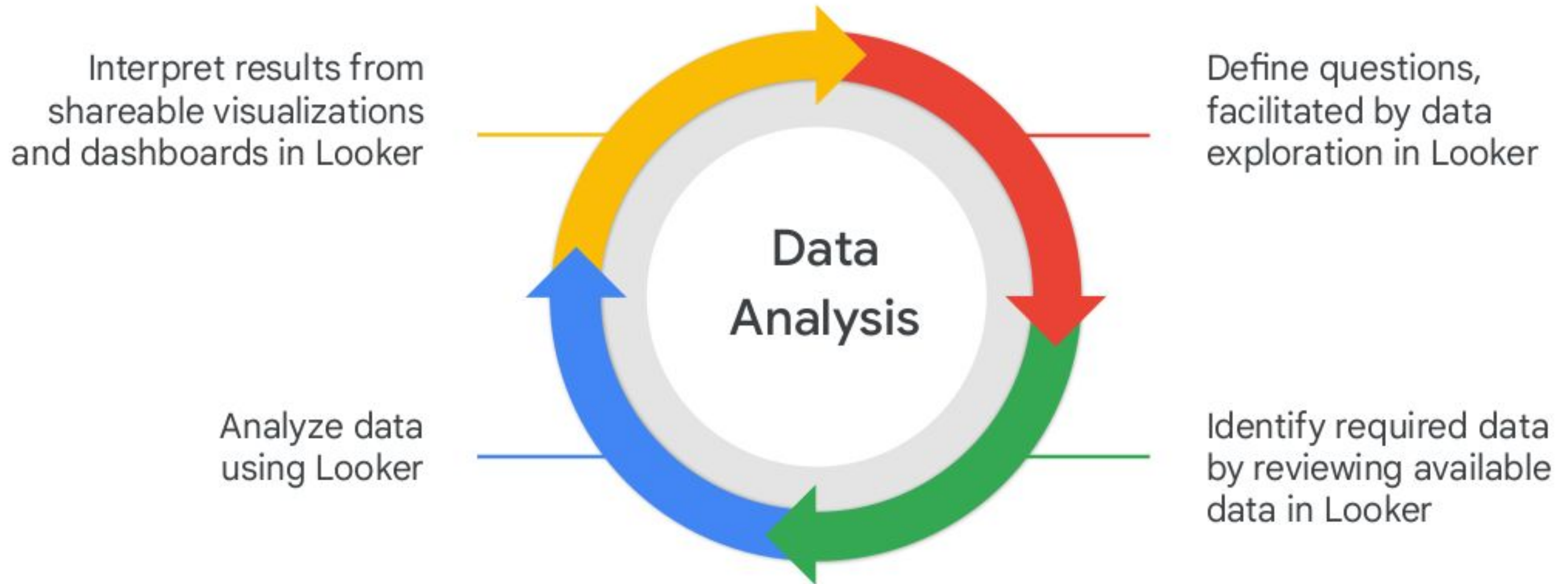
5. Share

6. Act

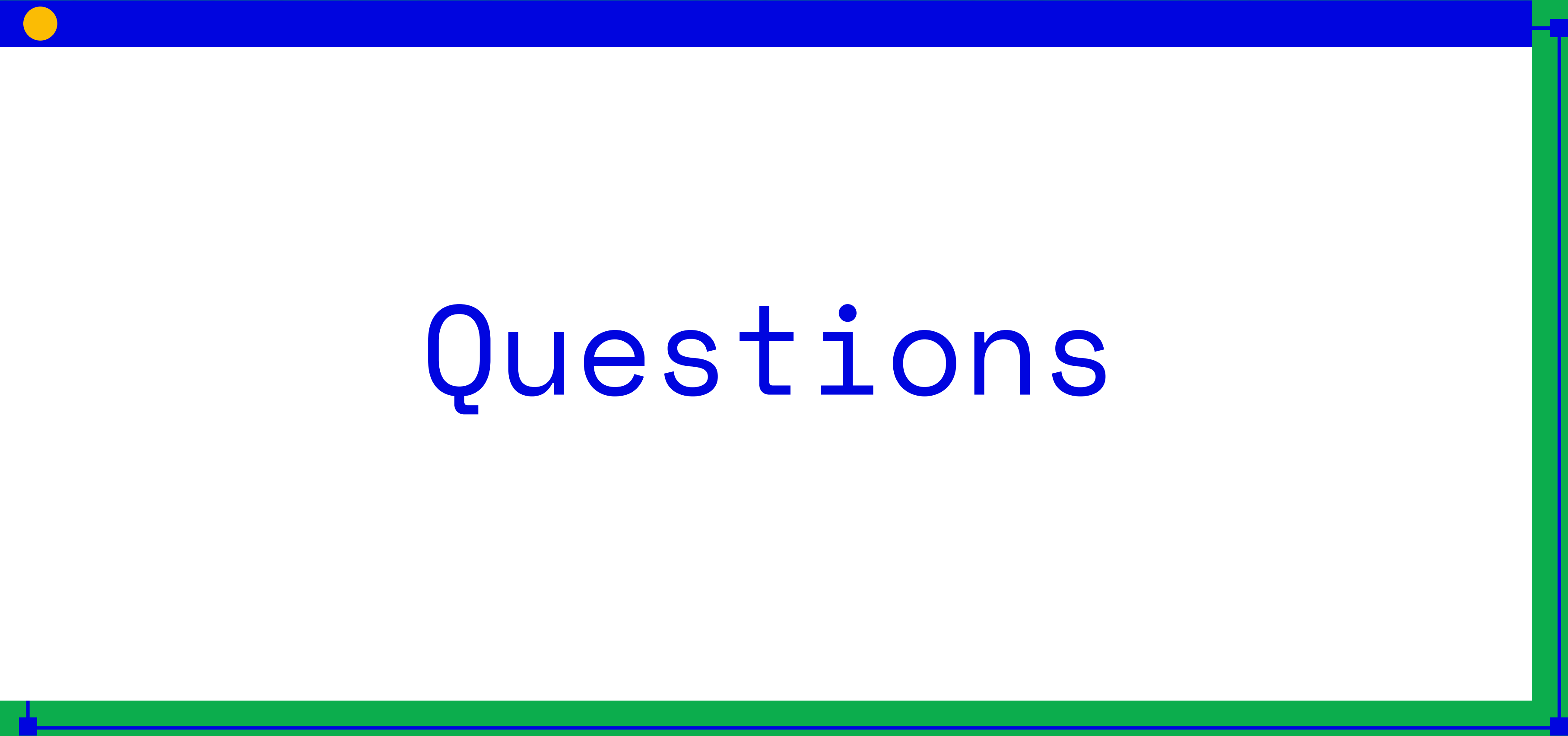


Data Analysis Process

Role of Looker in the Data Analysis Process



Looker: Modern BI Platform



Questions

Links

<https://github.com/FCAI-B/da>