Data Visualization

Agenda

1. Average Salaries in Egypt

2. Data Analysis Steps

3. Why build visuals and stream?

4. Questions

Average Salaries

Data Analyst Average Salaries: Egypt 16/02/23

Experience

| All years of Experience ⌄ |

Industry

| All industries ⌄ |

EGP 25,000 /mo

Very High Confidence

Total Pay

EGP 7,000 /mo
Base Pay

EGP 25,000 /mo
▼

EGP 15K   EGP 46K

EGP 18,000 /mo
Additional Pay

EGP 8K   EGP 110K

■ Most Likely Range   ■ Possible Range

According to GlassDoor

Data Analyst Average Salaries: Egypt 26/09/23

Experience

| All years of Experience ⌄ |

Industry

| All industries ⌄ |

EGP 27,000 /mo

📊 Very High Confidence

Total Pay

EGP 9,000 / mo
Base Pay

EGP 18,000 / mo
Additional Pay

EGP 27,000 / mo
▼

EGP 16K - EGP 45K

EGP 7K   EGP 124K

■ Most Likely Range   ■ Possible Range

Data Scientist Average Salaries: Egypt 24/09/2023

Experience

All years of Experience ⌄

Industry

All industries ⌄

EGP 35,010 / mo

Total Pay

🟩 Very High Confidence

EGP 12,000 / mo
Base Pay

EGP 23,010 / mo
Additional Pay

EGP 35,010 / mo
▼

EGP 18K   EGP 63K

EGP 12K   EGP 95K

⬛ Most Likely Range   🟦 Possible Range

Data Analysis Steps

Data Analysis Process Overview

visualiz   1/3   ∧ ∨ ✕

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

1. **Question**

2. **Wrangle**

3. **Explore**

4. **Draw Conclusions**

5. **Communicate**

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

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

Step 4: Draw Conclusions (or even make predictions)

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

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 Analysis Life Cycle (From EMC) - 1

EMC's data analysis life cycle

EMC Corporation's data analytics life cycle is cyclical with six steps:

1. Discovery

2. Pre-processing data

3. Model planning

4. Model building

5. Communicate results

6. Operationalize

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of real-world projects. The phases aren't static milestones; each step connects and leads to the next, and eventually repeats. Ke questions help analysts test whether they have accomplished enough to move forward and ensure that teams hav spent enough time on each of the phases and don't start modeling before the data is ready. It is a little different fro the data analysis life cycle this program is based on, but it has some core ideas in common: the first phase is intere in discovering and asking questions; data has to be prepared before it can be analyzed and used; and then finding should be shared and acted on.

For more information, refer to this e-book, Data Science & Big Data Analytics.

Data Analysis Life Cycle (From EMC) - 2
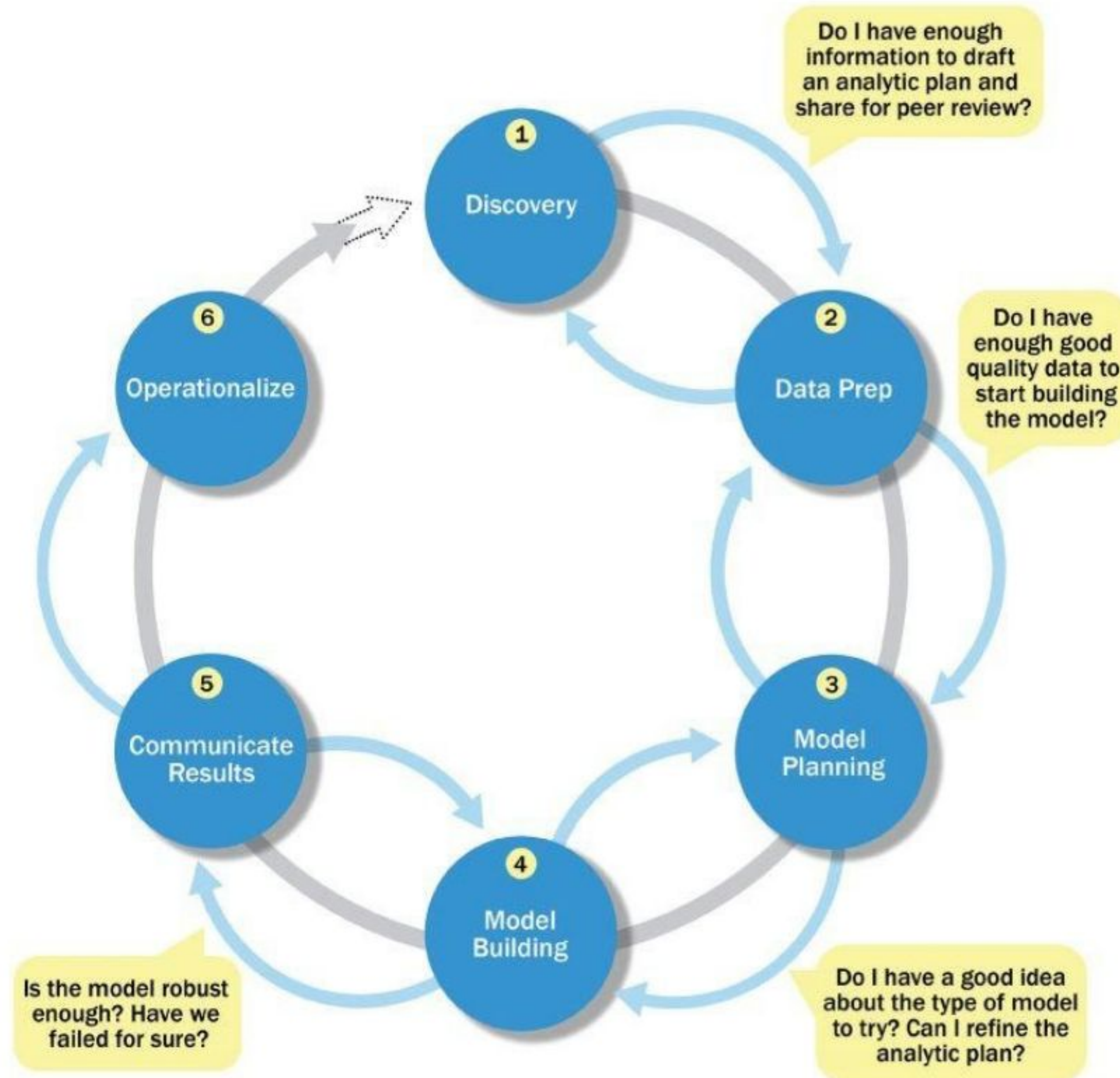
Phase 1: Discovery

- 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.

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

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 <u>testing</u>, <u>training</u>, and <u>production</u> purposes

- team **builds/executes models** based on the work done in Model Planning

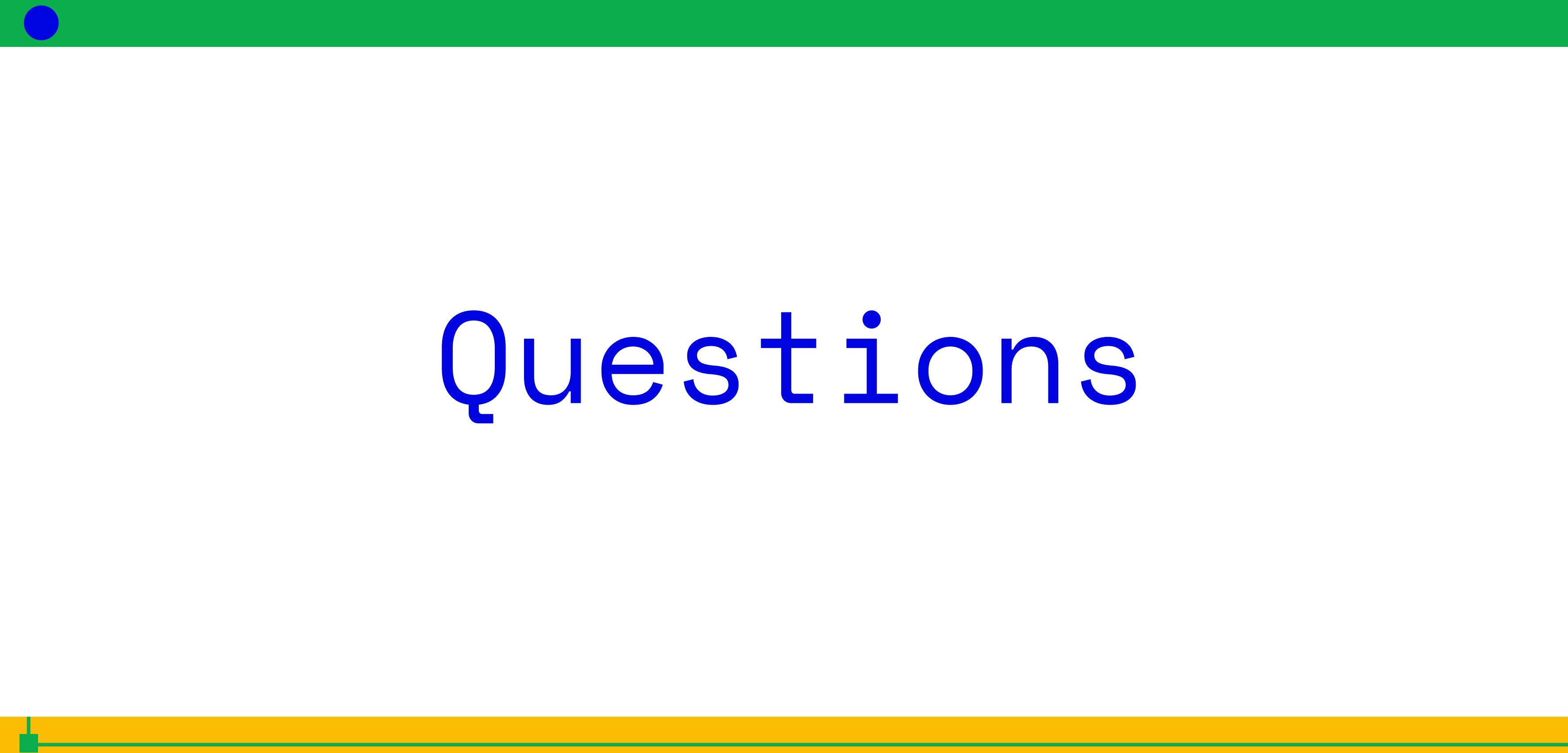- team **considers** whether its existing **tools** will suffice for running the models

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

Phase 6: Operationalize

- team **delivers** final <u>reports</u>, <u>briefings</u>, <u>code</u>, and <u>technical documents</u>
- 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

Questions

Links

https://github.com/fcai-b/dv

References

1. https://www.udacity.com/course/data-analyst-nanodegree--nd002
   - Udacity Nanodegree
2. https://www.coursera.org/learn/foundations-data
   - Google Data Analytics Professional Certificate - 1st Course
3. https://cloud.google.com/bigquery/docs/sandbox
   - BigQuery Sandbox