Data Analytics

Agenda

1. Data Analysis Steps
2. Analysis vs Analytics
3. Average Salaries (Egypt)
4. 365 Data Science Infographic

Data Analysis Steps

Data Analysis Steps (From Udacity ND) - 1

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 Steps (From Udacity ND) - 2

1. Question
2. Wrangle
3. Explore
4. Draw Conclusions
5. Communicate

Data Analysis Steps (From Udacity ND) - 3

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 Steps (From Udacity ND) - 4

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 Steps (From Udacity ND) - 5

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 Google Data Analytics Certificate) -1

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 Certificate) -2

1. Ask
2. Prepare
3. Process
4. Analyse
5. Share
6. Act

Data Analysis Life Cycle (From EMC)

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

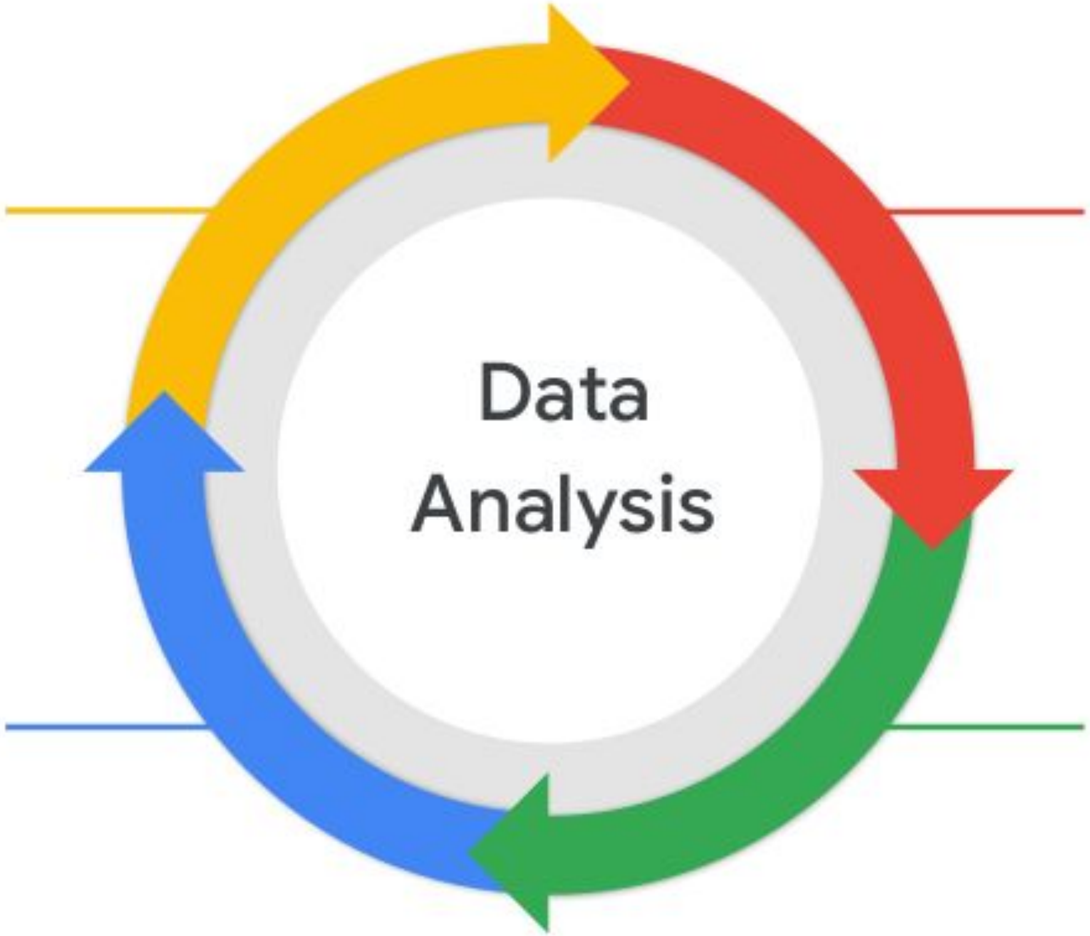
EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of real-world projects. The phases aren't static milestones; each step connects and leads to the next, and eventually repeats. Ke questions help analysts test whether they have accomplished enough to 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 fr 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.

Role of Looker in the Data Analysis Process

Define questions, facilitated by data exploration in Looker

Identify required data by reviewing available data in Looker

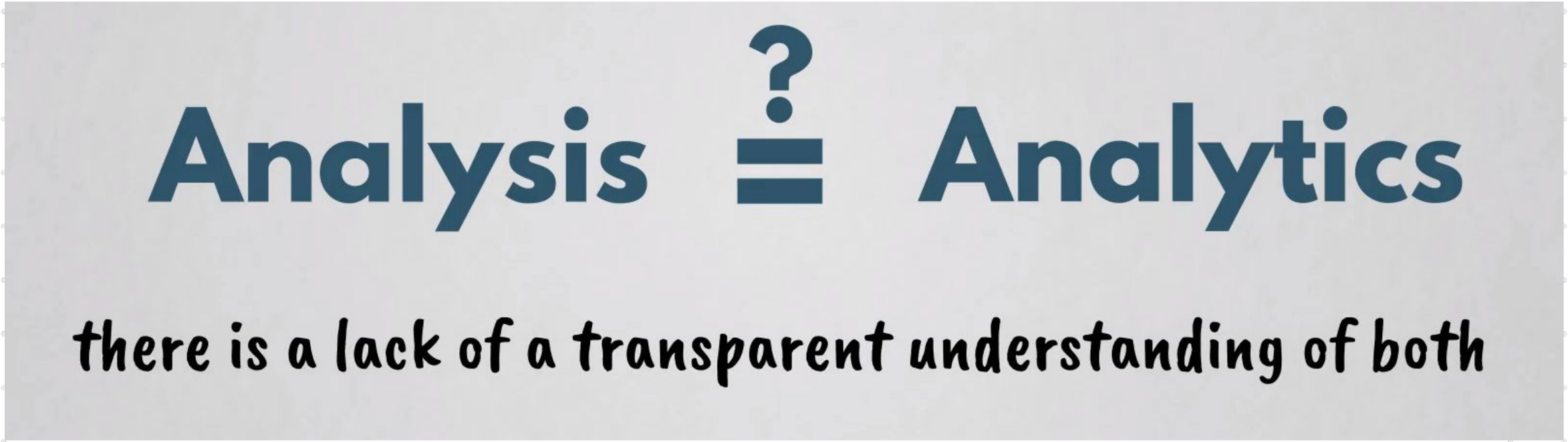


Analyze data using Looker

Interpret results from shareable visualizations and dashboards in Looker

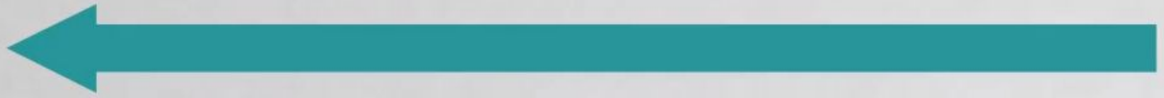
Analysis vs Analytics

Analysis vs Analytics



Data Analysis - 1

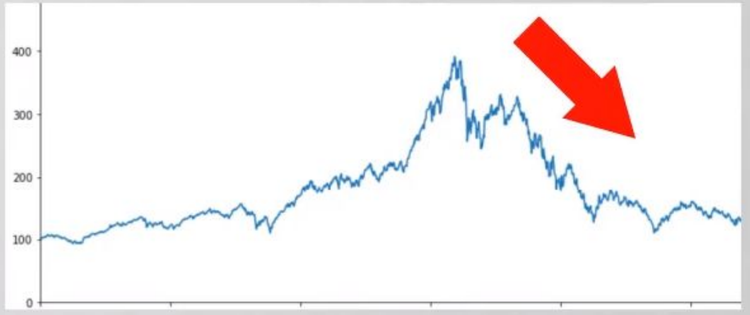
Analysis



Past

Explain

How? Why?



Data Analysis - 2

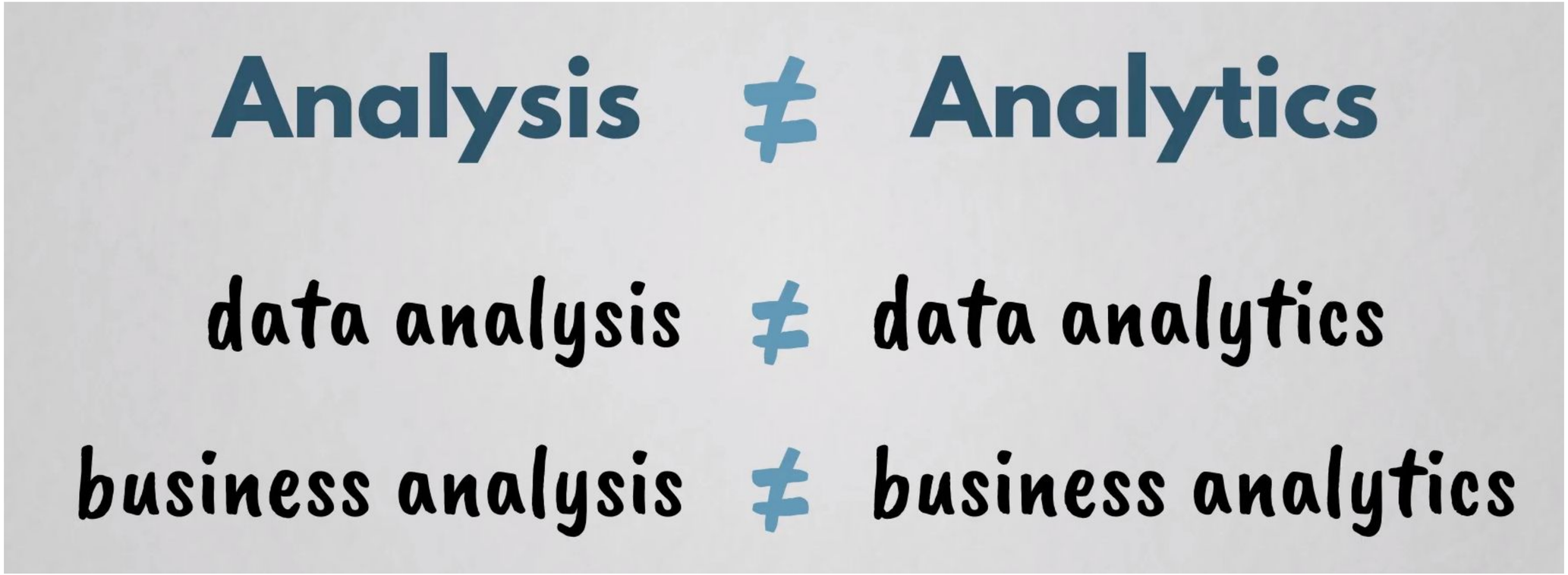
- Dividing data into **digestible components** that are easier to understand
- Examining how **different parts** relate to each other
- Performed on **past** data, explaining ‘how’ and ‘why’:
  - the story ended in the way that it did
  - something happened

Data Analytics - 1

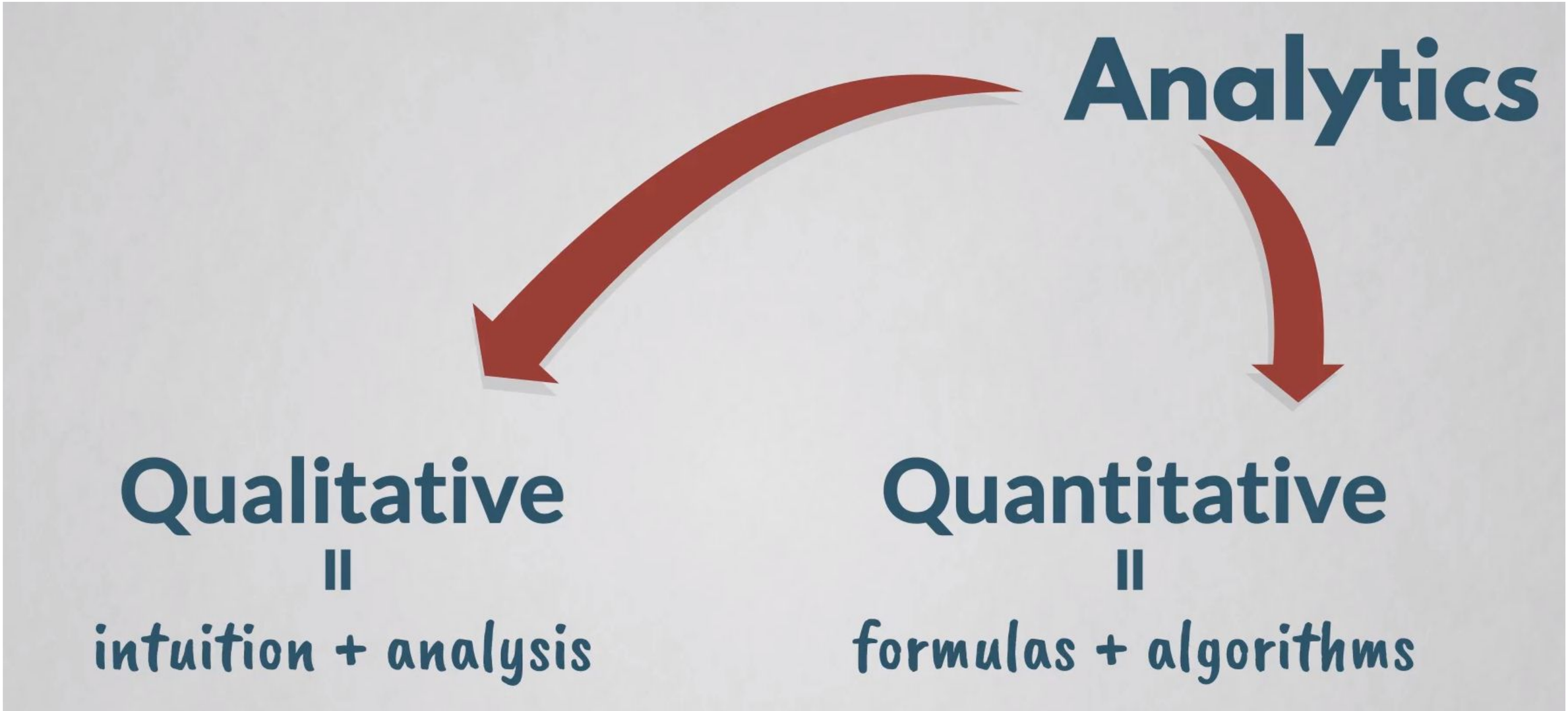
Data Analytics - 2

- The application of logical and computational reasoning to the **component parts** obtained in an **analysis**
- We look for patterns and explore what we can do with **them** in the **future**

Analysis vs Analytics



Quantitative Analytics vs Qualitative Analytics



Quantitative Analytics vs Qualitative Analytics

`Quantitative Analytics`

- Applying **formulas** & **algorithms** to **numbers** we have gathered from **analysis**

`Qualitative Analytics`

- Using **intuition** & **experience** + **analysis** to plan next business move

Quantitative Analytics Example

- When to introduce the new collection?
  - Relying on past sales data and UX data,
  - we could predict in which month it would be best to do that

Qualitative Analytics Example

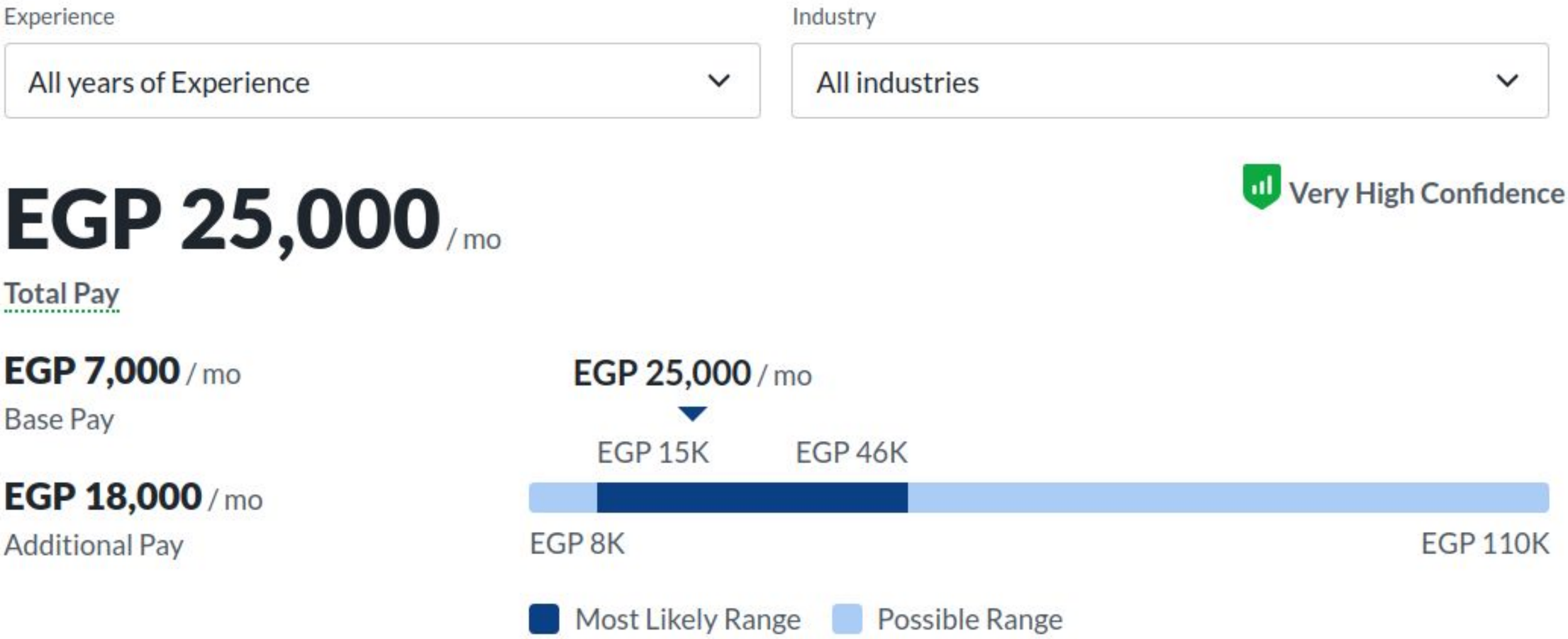
- An owner of an online clothing store
  - understand greatly customers' needs and wants
  - perform a very details analysis of women's clothing articles
  - feel sure about which fashion trends to follow
  - use this intuition to decide on which styles of clothing to start selling

Data Analytics - Note

- **Analytics** has become a term comprising both:
  - **Analysis**
  - **Analytics**
- which is something confusing in practice

Average Salaries

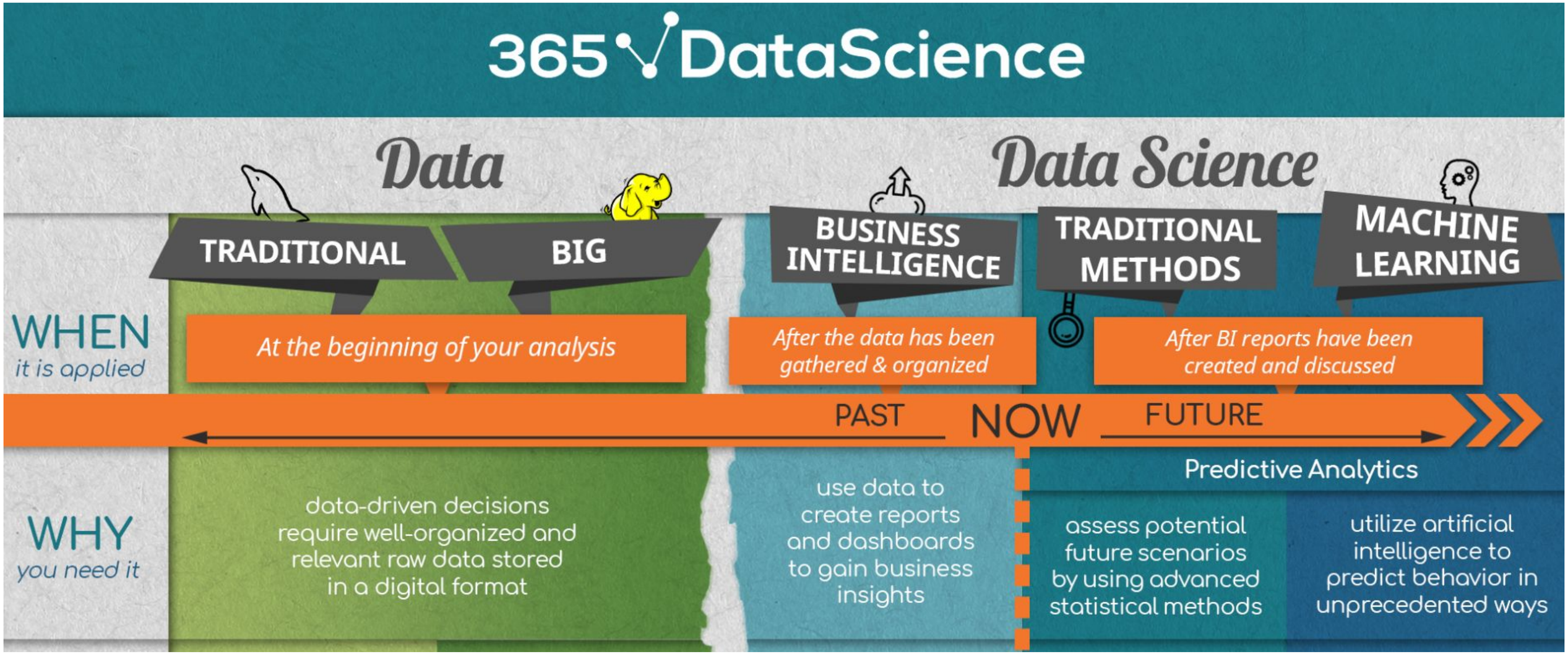
Data Analyst Average Salaries: Egypt 16/2/23



https://www.glassdoor.com/Salaries/cairo-data-analyst-salary-SRCH_IL.0,5_IM1175_KO6,18.htm

365 Data Science Infographic

Infographic (When & Why)



Questions

Links

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

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://learn.365datascience.com/courses/intro-to-data-and-data-science
    - 365 Data Science - Introduction to Data and Data Science