FROM CURIOSITY TO INSIGHT
My Data Learning
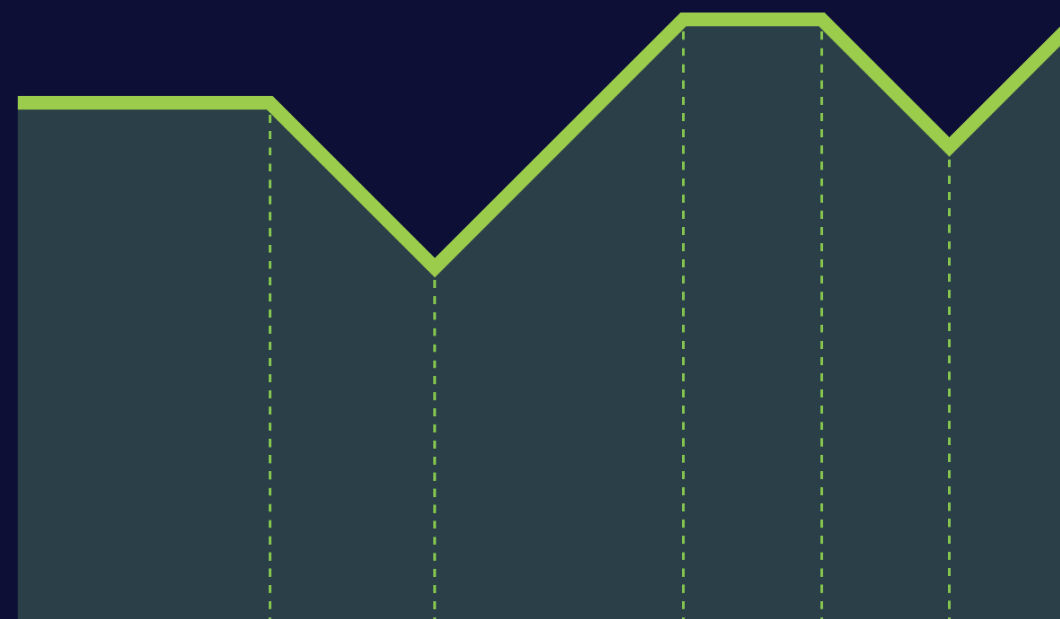
DATA ANALYSIS GOOGLE CERTIFICATE
DOCUMENT

TABLE OF CONTENTS

SELF-DICTIONARY

1. Adequately: một cách đầy đủ.
2. Dig: đào.
3. Quirk: kỳ quặc.
4. Naunce: tinh tế, phức tạp.
5. Anticipate: dự đoán.
6. Intui: trực giác
7. Interpret: giải thích.
8. Interventions: can thiệp.

Nên sắp lại theo thứ tự sau khi xong
Gợi ý: Dùng excel

COURSE 1: FOUNDATIONS: DATA, DATA, EVERYWHERE

MODULE 1:

INTRODUCING DATA ANALYTICS AND ANALYTICAL THINKING

1. **GET STARTED**
2. **TRANSFORM DATA INTO INSIGHTS**
3. **UNDERSTAND THE DATA ECOSYSTEM**
4. **EMBRACE YOUR DATA ANALYST SKILLS**

4.1.Key data analysis skills

1. **Curiosity**
- Ask and answer a question about your dataset.
2. Understanding the context
- Understand when things are not in the context and remove them, E.g., when you shuffle the deck of cards. You don't need Jockers. That means the Jockers are not in the context.
3. **Technical mindset**
- Divided a single task into smaller steps. E.g., when you want to pay a bill, you need to sort them and check your balance to see whether you can pay. Then, you'd pay. This is a Technical mindset.
4. **Data design**
- This is how you organize information. E.g., this is how you manage your contacts: maybe you list your contacts by the first or last name, or you can use email instead. That

makes your contact list clear and logical and lets you call or text a contact quickly.

5. Data strategy

- The management of people, processes, and tools used in data analysis. Let's break it down, you manage people to make sure they know how to use the right data to find the solutions to the problem you are working on. For processes, it's about making sure the path to that solution is clear and accessible. For tools, you make sure the right technology is being used for the job.

CASE-STUDY/SCENARIO

Use data to create better movies.
The movies **Mega-Pik** has released recently aren't having the impact they used to. Five of their last six releases barely broke even at the box office, and the sixth film lost a lot of money. The lead executives at **Mega-Pik** have noticed that their competitors went through a similar slump, but recovered when they started producing remakes of past successes and marketing them to a new audience.

Mega-Pik is interested in following this trend. They want to do this based on data-driven strategies, so they hire your analytics company to help them make popular movies again. Specifically, they ask for exploratory data analysis (*EDA*) to help them understand what audiences have liked in the past and determine if the successes of those films can be replicated.

You and your team develop the following objectives for Mega-Pik's EDA:

- Identify key factors that contribute to a movie's opening weekend success.
- Understand the relationship between a movie's budget and its revenue.
- Determine which genres are most successful.
-

The right dataset

Your company collects, cleans, and organizes the following relevant information into a dataset:

- Movie name.
- Release date.
- Opening night revenue.
- Opening weekend revenue.
- Budget (cost to create).
- Marketing costs.
- Ratings.
- Genre.

Now, use your skills

- *Curiosity*

What kinds of questions would you ask based on the data and how it relates to the objectives of *EDA*?

You might wonder:

- Is there a relationship between a movie's budget and the revenue it generates on opening night or over the opening weekend?
- You might be curious about combining columns to make the new metrics, such as which genres tend to perform better on opening weekend-both overall and in the season in which the movies were released.
- You might ask if there should be additional columns of data you don't already have, such as audience demographics?

- **Understanding context**
- Analysts determine context by looking for patterns or anomalies in a dataset. It also helps to understand the entertainment industry, which provides a whole other set of contextual clues. For example, family films typically generate more revenue when children are on vacation from school. This provides important context about the relationship between genre and revenue over a short timeframe.

- **Technical mindset**
- This means how you approach the problems (and datasets) systematically and logically. This starts with the way you clean, organize, and prepare your data. It can also guide the tools or software you use to break down data and help you identify and fix incorrect data that can skew your analysis.

- **Data design**
- For example, you might organize the data by revenue and then by genre, which could reveal that comedies are more profitable than dramas. How you choose to structure your data makes analysis easier and more insightful.

- **Data strategy**
- The data strategy you select should be based on the dataset and the deliverables. Think about a data strategy as a kind of resource allocation—the tools, time, and effort that you put into a project will vary based on what you need to accomplish.

5. ANALYTICAL THINKING FOR EFFECTIVE OUTCOMES

5.2. Explore core analytical skills

- Five aspects of analytical thinking
- Visualization.
- Strategy.
- Problem-orientation.
- Correlation.
- Big-picture and detail-oriented thinking.

5.3. Use these five whys for the root cause analysis

- The big question is "What is the root cause of the problem?".
- By identifying and eliminating the root cause, the data professional can stop those problems from occurring again.
- E.g.:
  - ❖ **WHY #1:** "Customers are complaining about poor grocery deliveries. Why?".
  - ❖ **WHY #2:** "Products are arriving damaged. Why?".
  - ❖ **WHY #3:** "Products are not packaged properly. Why?".
  - ❖ **WHY #4:** "Grocery packers are not adequately trained. Why?"
  - ❖ **WHY #5:** "Packers have not completed the required training. Why?"

KEY TAKEAWAYS

- The five whys can be used to analyse problems in any industry, helping organizations of all kinds identify and fix business problems.

5.3. Data-driven decision-making

This means using facts to guide business strategy.

5.4. Witness data magic

- Visuals are extremely helpful when trying to understand a problem or concept.

6. MODULE 1 CHALLENGE

6.1. What to expect moving forward

MODULE 1 CHALLENGE: 82.50%

MODULE 2: THE WONDERFUL WORLD OF DATA

1. FOLLOW THE DATA LIFE CYCLE

1.1. Phases of data analysis

- We have 6 data phases:
  1. Ask
  2. Prepare
  3. Process
  4. Analyze
  5. Share
  7. Act

1.2. Stages of the data life cycle

1. Plan

2. Capture
3. Manage
4. Analyze
5. Archive
6. Destroy

PLAN:
- Business decides what kind of data it needs, how it will be managed throughout its life cycle, who will be responsible for it, and the optional outcomes.

~ DATABASE: A collection of data stored in a computer system ~

CAPTURE:
- Collect or bring data from a variety of different sources.

MANAGE:
- How do we care for our data?
- How and where it's stored, and the tools used to keep it safe and secure.
- And the actions taken to make sure that it's maintained properly.
  - ⇨ This is where data analysts really shine. Data is used to solve problems, make great decisions, and support business goals.

ARCHIVE:
- This means storing data in a place where it's available, but may not be used again.

 # DESTROY:
- The company should use secure data erasure software.

2. OUTLINE THE DATA ANALYSIS PROCESS

2.1. The phase of data analysis and this program

THE ASK PHASE
- We will do two things:
+ Define the problem to be solved.
+ Make sure that we fully understand stakeholder expectations.

~ STAKEHOLDERS: Those who have invested time and resources into a project and are interested in the outcome.

~ Defining the problem means you look at the current state and identify how it's different from the ideal state.

? Determine who the stakeholders are?

THE PREPARE PHASE
- How to identify which kinds of data are most useful for solving a particular problem.
- Why is this data important?
- Any decisions should be based on facts and be fair and impartial.

THE PROCESS PHASE
- In this phase, the analyst must find and eliminate errors.

- This usually means cleaning data, transforming it into a more useful format, combining two or more datasets to make information more complete, and removing outliers.
  (Outliers could skew the information.)

THE ANALYZE PHASE

- Analyzing the data collected involves using tools to transform and organize that information.

THE SHARE PHASE

- How data analysts interpret results and share them with others.
- In this place, visualization is a data analyst's best friend.

THE ACT PHASE

- The business will take all of the insights you, the data analyst, have provided and put them to work in order to solve the original problem.

EXAMPLE OF THE DATA ANALYSIS PROCESS

By Molly (Head of People, Analytics)

THE ASK PHASE

- Ask all of the right questions at the beginning of the engagement so that you better understand what your leader and stakeholders need from this analysis.

THE PREPARE PHASE

- The general types of questions:

+ What is the problem that we're trying to solve?
+ What is the purpose of this analysis?
+ What are we hoping to learn from it?
+ What kind of data do we need to solve?

THE PROCESS PHASE

- This phase begins with cleaning data.
- This is where you get a chance to understand its structure, its quirks, its naunces.
- E.g.:
  + Are we missing data at random, or is it missing in a systematic way?
  + Are there any outliers?

THE ANALYZE PHASE

- This phase is also a chance for us to step back.

THE SHARE PHASE

- Sharing the high-level findings with the executive team.
- We want them to have a landscape view of how the organization is feeling.
- To make sure there are no surprises as they dig deeper and deeper into the data.

THE ACT PHASE

- This is where we use all of the data-driven insights to decide what types of interventions we want to introduce.

3. THE DATA ANALYSIS TOOLBOX

4. MODULE 2 CHALLENGE