



BEFORE BED

EP03 - Data Analyst Study Guide 2020

10 July 2020 Time: 8-9PM

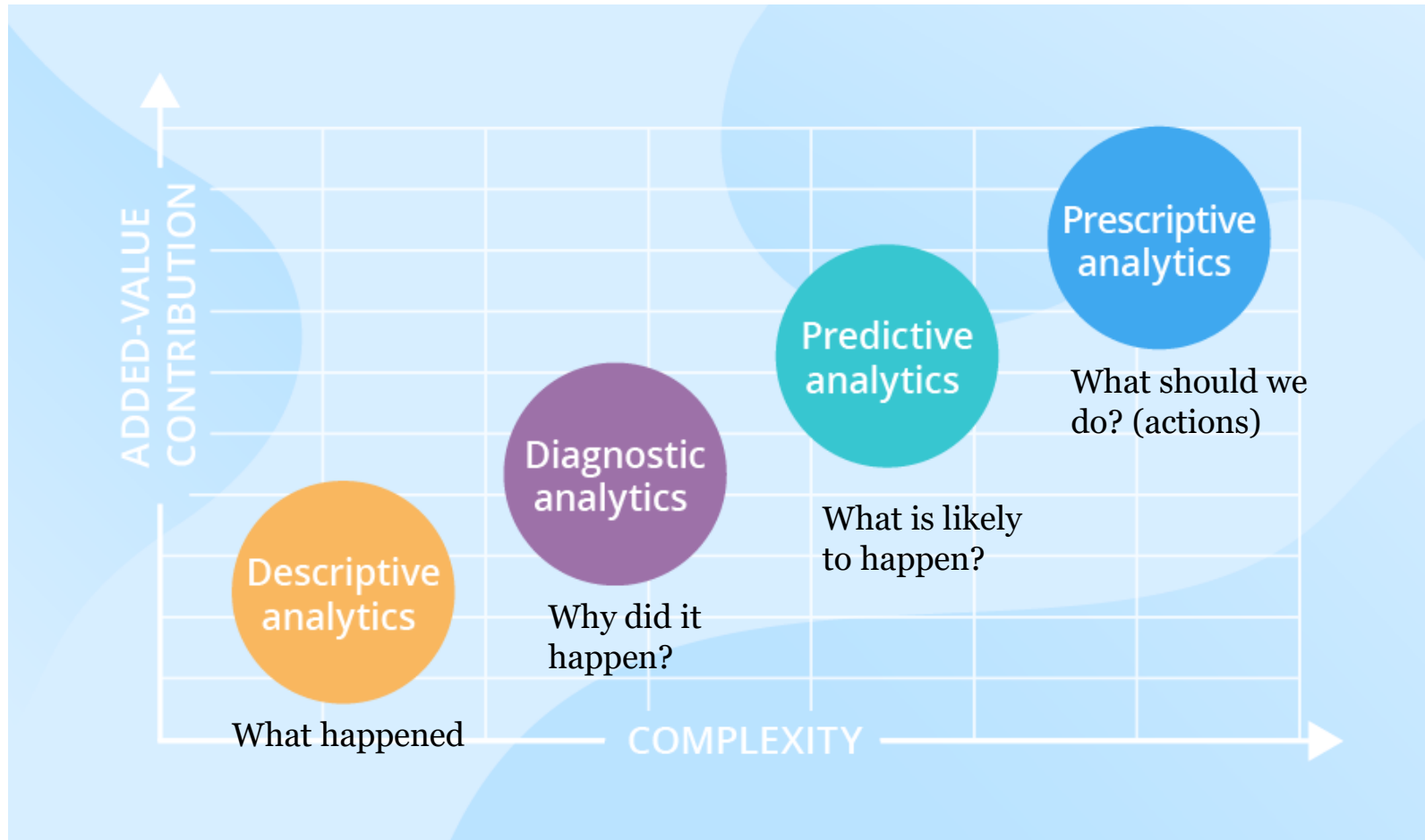


Today Agenda

- Understand 4 phases of analytics and where data analyst fits in?
- Understand job description
- Understand required skills/ tools
- Recommended Online Courses
- Talking about Certificates

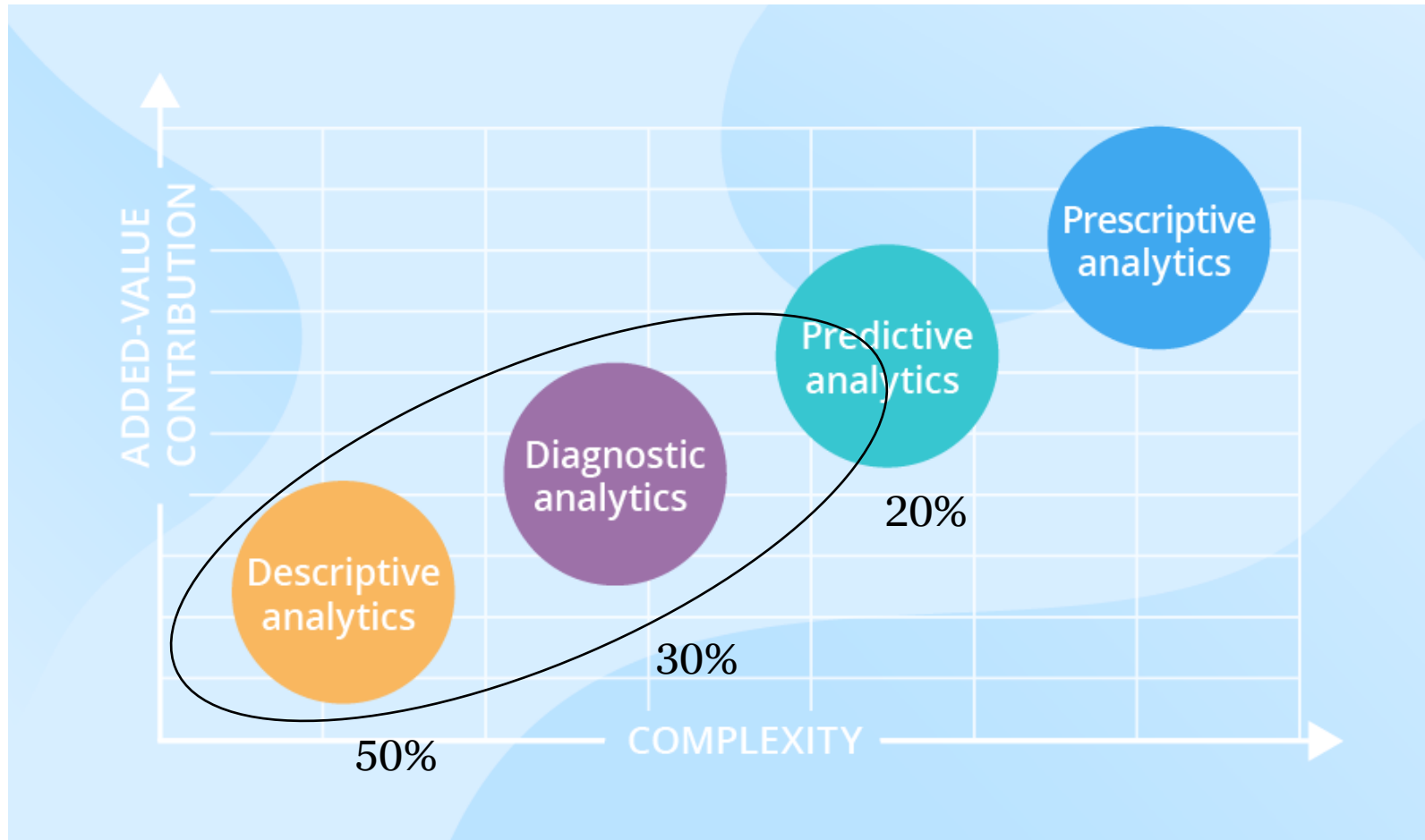
Four Phases of Analytics

<https://www.scnsoft.com/blog/4-types-of-data-analytics>

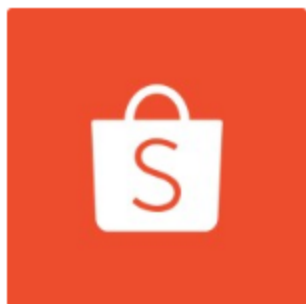


Industry needs to standardize job description

<https://www.scnsoft.com/blog/4-types-of-data-analytics>



* Estimated numbers, varied by company



BI Data Analyst

Shopee · Bangkok, TH

Posted 2 months ago · 1,828 views

Save

Apply

Job Description

- Work closely with Seller management team to support data and analysis and better support the Campaigns team in strategizing and executing various campaigns on the platform
- Establish processes and identify appropriate indicators to monitor performance and progress
- Collaborate with different stakeholders to identify the business problems, understand project requirements, define the project use cases and project scope, as well as develop predictive analytics solutions
- Deep dive into big data by applying advanced statistical methodologies for warehouse

Requirements

- Degree in computing or technical discipline (e.g., Computer Science, Engineering, Math, IT) preferred
- 3-5 years of experience in business intelligence, data analytics or marketing analytics preferred
- Proficiency in Excel and SQL required.
- Passionate about data and solving problems
- Systematic thinker with ability to understand the holistic picture and complex considerations



Data Analyst Specialist

CP ALL Plc · 📍 Bang Rak, TH

Posted 3 weeks ago · 107 views

Save

Apply

Data analyst responsibilities include conducting full lifecycle analysis to include requirements, activities and design. Data analysts will develop analysis and reporting capabilities. They will also monitor performance and quality control plans to identify improvements and support business decision-makers.

- Interpret data, analyze results using statistical techniques and provide ongoing reports
- Develop and implement databases, data collection systems, data analytics and other strategies that optimize statistical efficiency and quality
- Acquire data from primary or secondary data sources and maintain databases/data systems
- Identify, analyze, and interpret trends or patterns in complex data sets
- Work with management to prioritize business and information needs
- Present analysis result to executives and stakeholders
- Locate and define new process improvement opportunities

Seniority Level

Entry level

Employment Type

Full-time

Industry

Information Technology & Services ,
Computer Software , Internet

Job Functions

Information Technology



Data Analyst

Robert Walters · Bangkok, TH

Posted 1 month ago · 755 views

Save

Apply

Key Responsibilities

- Acquire data from multiple data sources by utilising business intelligence tools and SQL programming
- Manipulates, analyses, and summarises data to derive trends, patterns and summarises into key findings and recommendations
- Create presentation and present key findings to cross functional and management team

To succeed in this role, you must have the ability to query data by using SQL and experience in data analyst.

Key Requirements

- Bachelor's degree or higher in Statistics, Marketing or any related fields
- Over five years of working experience in marketing, sales data analysis, from FMCG would be preferred
- Sense in Business/Marketing
- Able to write SQL Query to acquire data
- Experience with Tableau, Siebel Marketing is a plus
- Proficient in using MS Word, Excel and PowerPoint
- Good analytical, communication and presentation skills
- Self motivated and able to work independently
- Good command of English
- This role is open for Thai nationality only due to nature of business

This nutrition firm needs no introduction and is regarded as the leader in its field of expertise.

Required Skills

- Database
- Spreadsheets
- Visualization
- Statistical Analysis

Foundations for Data Analyst

Coding



Data Visualization: Power BI & Tableau



Everyday Tools: Excel & SQL

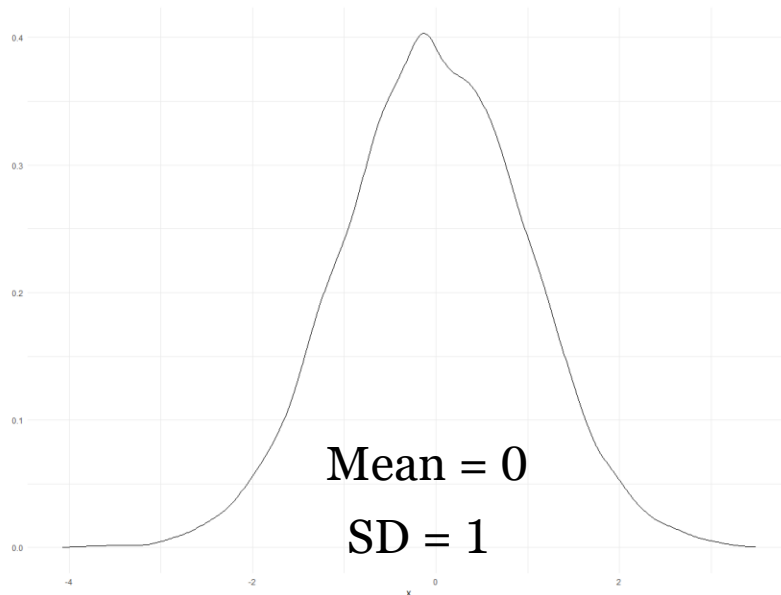


Statistics & Domain Knowledge (Business/ Marketing)

Difference between Stats vs. Math

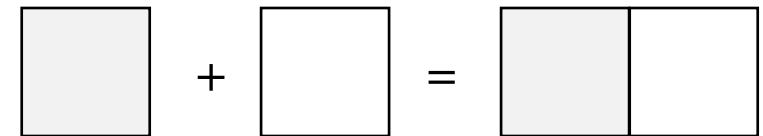
Stats

Uncertainty



Math

Certainty



$$1 + 1 = 2$$
$$\text{Exp}(2) = 7.38$$

The most widely used statistics for data analyst role



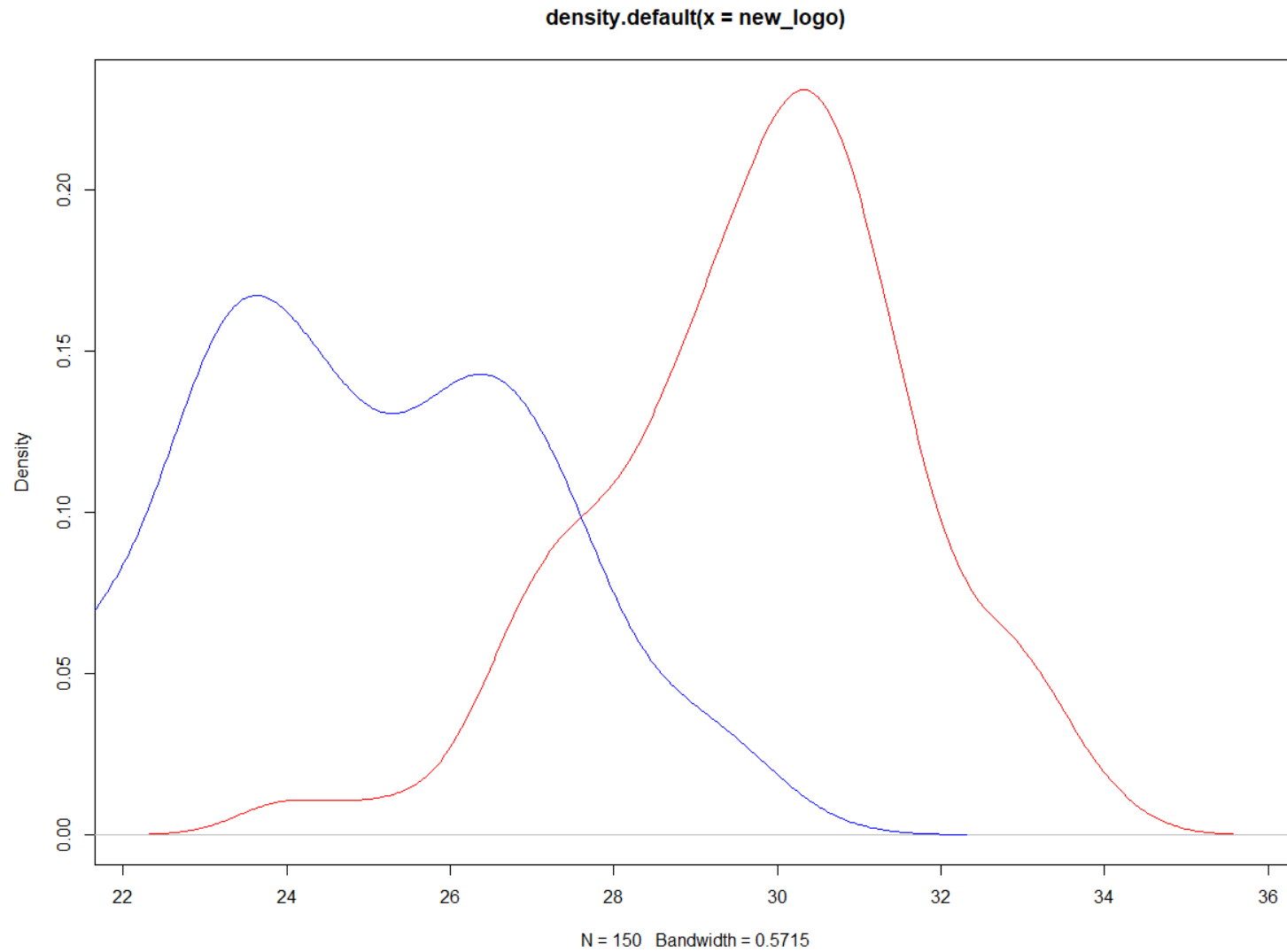
AB testing to test the impact of logo change in sales, % subscription etc.

Statistics:

Sales = dependent variable or output (y)

Logo = independent variable or input (x)

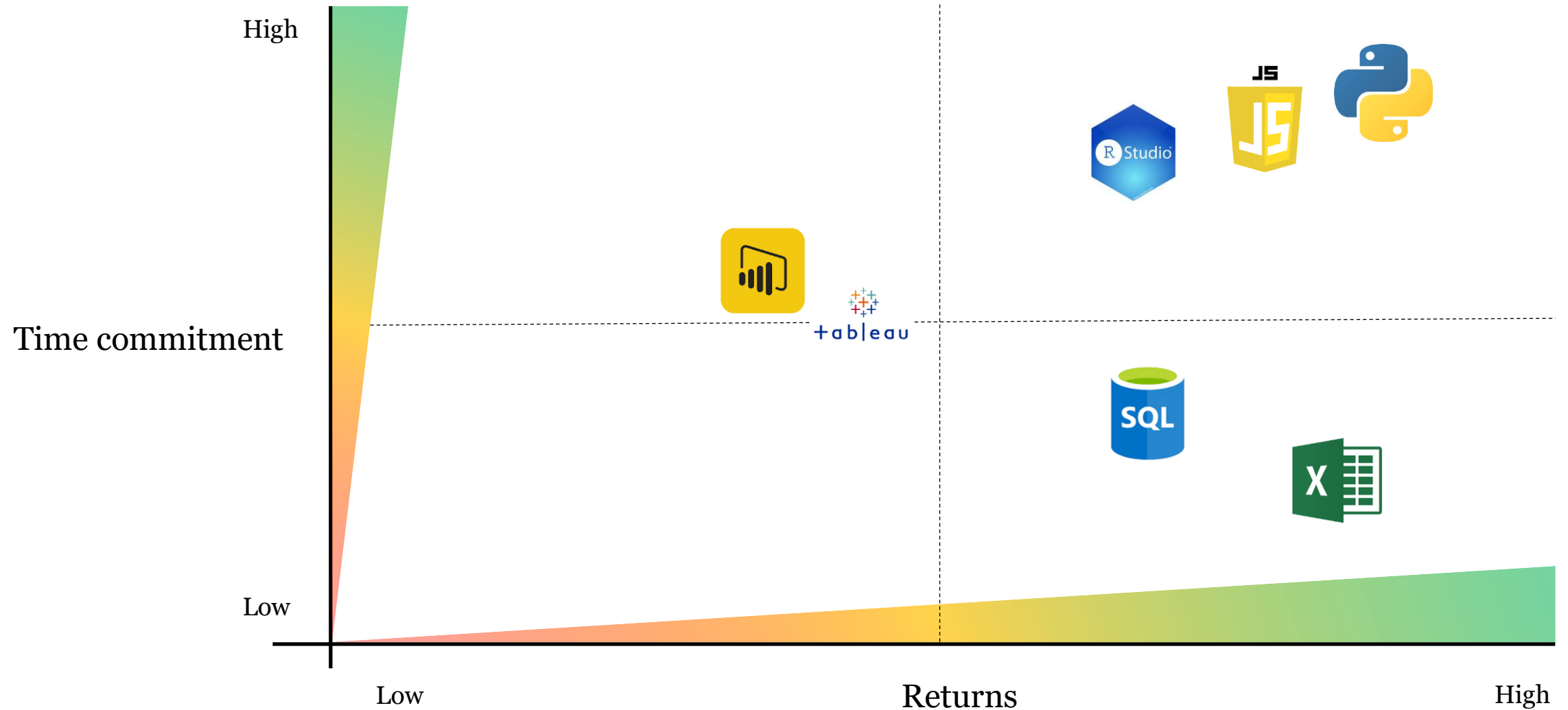
Test the means between two groups



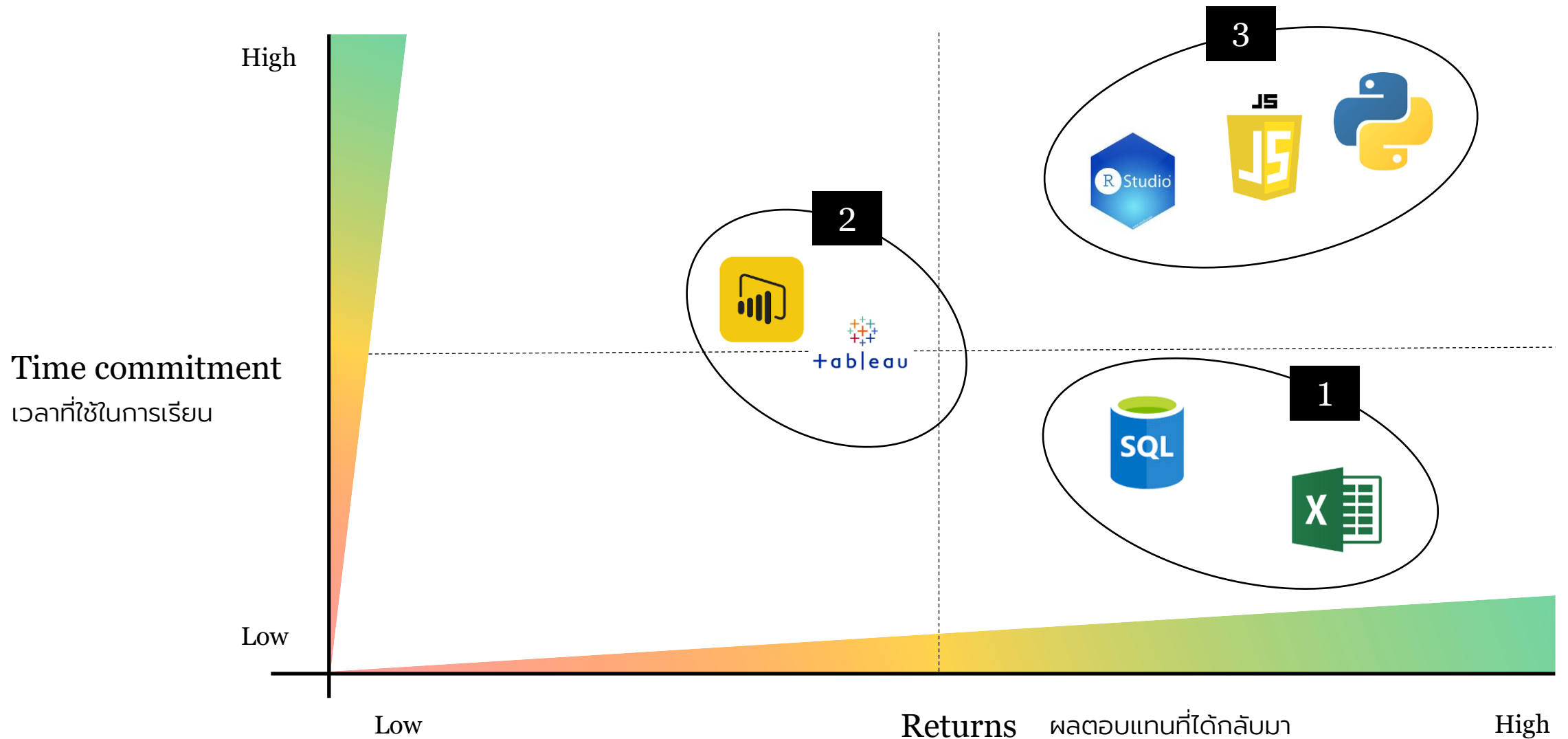
A top-down photograph of various vintage tools and objects arranged on a dark, vertically-grained wooden surface. The tools include two axes with wooden handles and metal heads, a claw hammer with a wooden handle, a mallet with a metal head and handle, a pair of large, dark metal shears, a pair of worn, tan leather work gloves, a circular metal object with a logo, a small folding knife, a utility knife, a curved metal tool, and a metal mug. The word "Tools" is written in a large, white, cursive font across the right side of the image.

Tools

How to choose which tools to learn?



How to choose which tools to learn?



SQL is the language of database

Get data from table customer

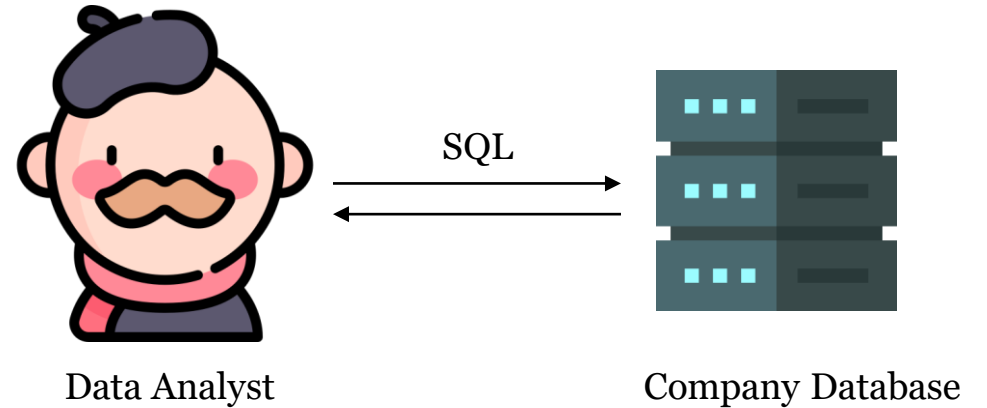
```
SELECT * FROM customer;
```

Get only Thai customers from table customer

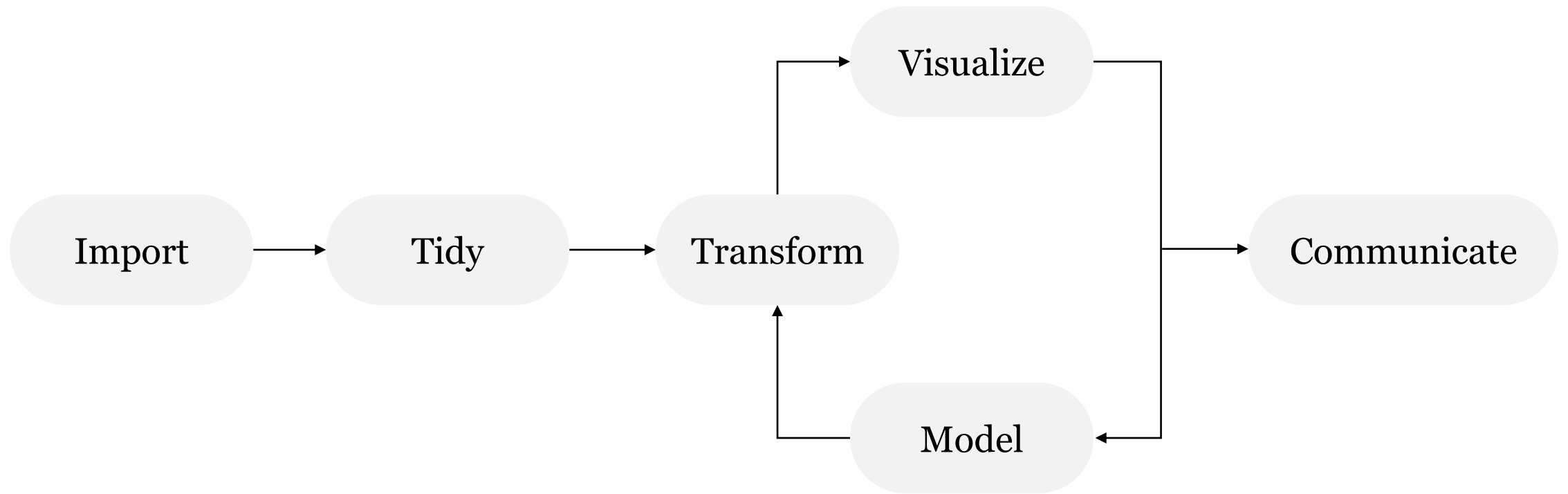
```
SELECT * FROM customer  
WHERE country = 'Thailand';
```

Get summary statistics for revenue in year 2020 from table invoice

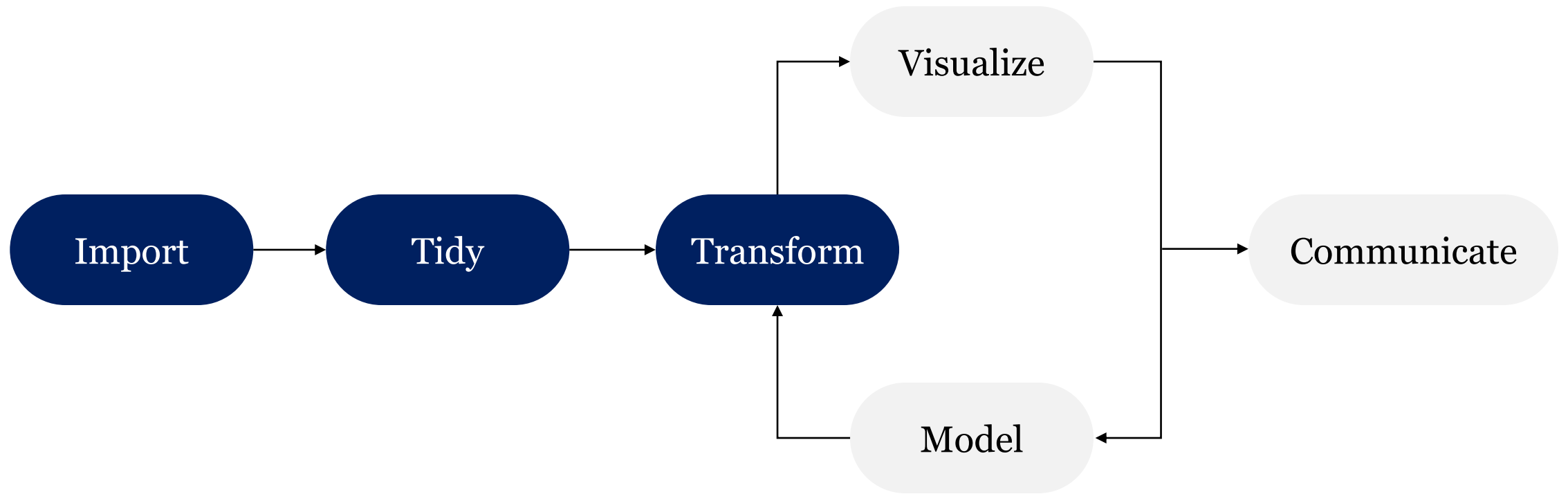
```
SELECT  
    SUM(revenue),  
    AVG(revenue),  
    COUNT(revenue)  
FROM invoice  
WHERE year = 2020;
```



Data Analysis is Iterative Process



We spent 50-80% of our time on wrangling part



Data Transformation (Wrangling)

Data Transformation with dplyr :: CHEAT SHEET



dplyr functions work with pipes and expect **tidy data**. In tidy data:



Each **variable** is in its own **column**

&



Each **observation**, or **case**, is in its own **row**



pipes

x %>% f(y) becomes **f(x, y)**

Summarise Cases

These apply **summary functions** to columns to create a new table of summary statistics. Summary functions take vectors as input and return one value (see back).

summary function



summarise(.data, ...)
Compute table of summaries.
summarise(mtcars, avg = mean(mpg))



count(x, ..., wt = NULL, sort = FALSE)
Count number of rows in each group defined by the variables in ... Also **tally()**.
count(iris, Species)

VARIATIONS

summarise_all() - Apply funs to every column.
summarise_at() - Apply funs to specific columns.
summarise_if() - Apply funs to all cols of one type.

Group Cases

Use **group_by()** to create a "grouped" copy of a table. dplyr functions will manipulate each "group" separately and then combine the results.



*mtcars %>%
group_by(cyl) %>%*

Manipulate Cases

EXTRACT CASES

Row functions return a subset of rows as a new table.



filter(.data, ...) Extract rows that meet logical criteria. *filter(iris, Sepal.Length > 7)*



distinct(.data, ..., .keep_all = FALSE) Remove rows with duplicate values.
distinct(iris, Species)



sample_frac(tbl, size = 1, replace = FALSE, weight = NULL, .env = parent.frame()) Randomly select fraction of rows.
sample_frac(iris, 0.5, replace = TRUE)



sample_n(tbl, size, replace = FALSE, weight = NULL, .env = parent.frame()) Randomly select size rows. *sample_n(iris, 10, replace = TRUE)*



slice(.data, ...) Select rows by position.
slice(iris, 10:15)



top_n(x, n, wt) Select and order top n entries (by group if grouped data). *top_n(iris, 5, Sepal.Width)*

Logical and boolean operators to use with filter()

<	<=	is.na()	%in%		xor()
>	>=	!is.na()	!	&	

See **?base::logic** and **?Comparison** for help.

ARRANGE CASES



arrange(.data, ...) Order rows by values of a column or columns (low to high), use with **desc()** to order from high to low.
arrange(mtcars, mpg)

Manipulate Variables

EXTRACT VARIABLES

Column functions return a set of columns as a new vector or table.



pull(.data, var = -1) Extract column values as a vector. Choose by name or index.
pull(iris, Sepal.Length)



select(.data, ...)
Extract columns as a table. Also **select_if()**.
select(iris, Sepal.Length, Species)

Use these helpers with **select()**,
e.g. *select(iris, starts_with("Sepal"))*

contains(match)	num_range(prefix, range) : e.g. <i>mpg:cyl</i>
ends_with(match)	one_of(...) : e.g. <i>-Species</i>
matches(match)	starts_with(match)

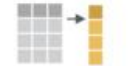
MAKE NEW VARIABLES

These apply **vectorized functions** to columns. Vectorized funs take vectors as input and return vectors of the same length as output (see back).

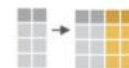
vectorized function



mutate(.data, ...)
Compute new column(s).
mutate(mtcars, gpm = 1/mpg)



transmute(.data, ...)
Compute new column(s), drop others.
transmute(mtcars, gpm = 1/mpg)



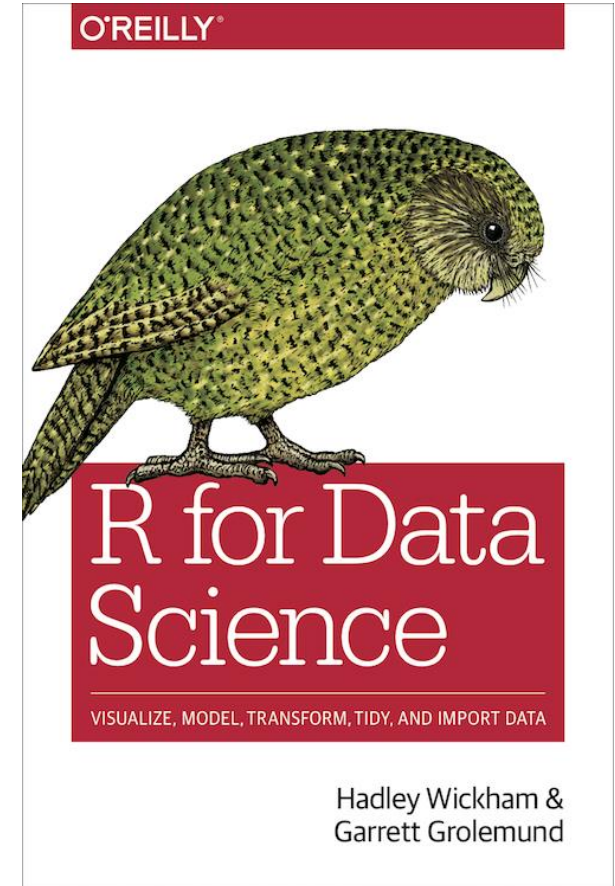
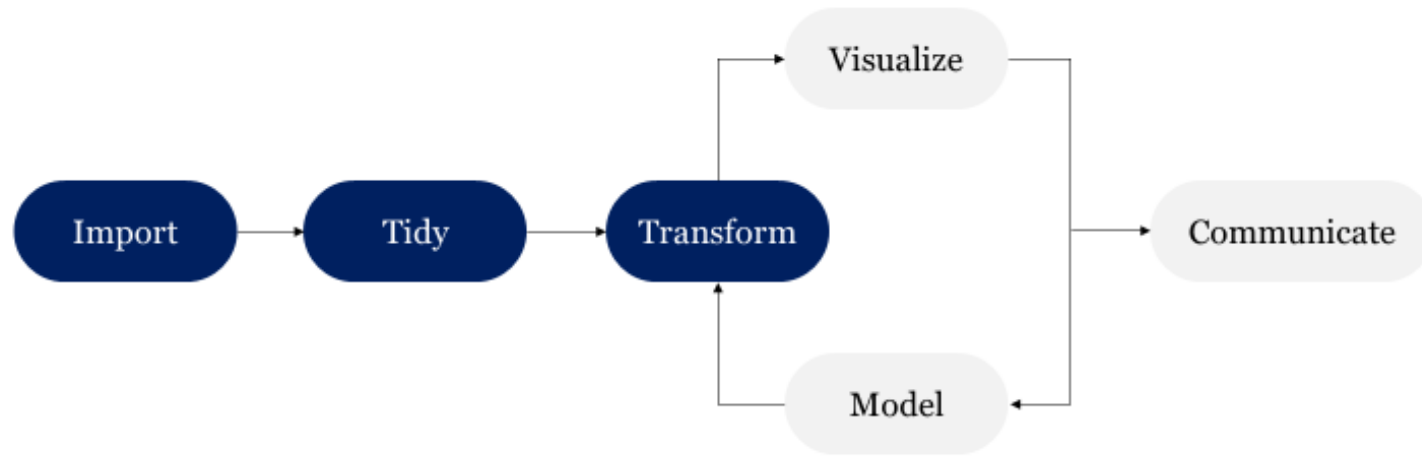
mutate_all(.tbl, .funs, ...) Apply funs to every column. Use with **funs()**. Also **mutate_if()**.
mutate_all(faithful, funs(log(.), log2(.)))
mutate_if(iris, is.numeric, funs(log(.)))



mutate_at(.tbl, .cols, .funs, ...) Apply funs to specific columns. Use with **funs()**, **vars()** and

Excel is good, R is great

If you want to learn programming for data analysis, R is by far a perfect choice for starter.



<https://r4ds.had.co.nz/>

Quick Demo Using RStudio

<https://rstudio.cloud/>

≡ Your Workspace / Data Science with R

⚙️ ⋮ 👤 Kasidis Satangmongkol

File · Edit · Code · View · Plots · Session · Build · Debug · Profile · Tools · Help

📁 📄 📦 📄 📄 📄 📄 Go to file/function 📄 Addins

R 4.0.0

Untitled1*

```
1 install.packages("gapminder")
2
```

Run Source

2:1 (Top Level)

R Script

Console Terminal Jobs

/cloud/project/

```
> print("hello world")
[1] "hello world"
>
```

Environment History Connections

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

	Name	Size	Modified
📁	..		
📄	.Rhistory	0 B	Jun 23, 2020, 10:01 AM
📄	myNotebook.html	678.7 KB	Jul 5, 2020, 9:03 AM
📄	myNotebook.nb.html	655.2 KB	Jul 5, 2020, 9:03 AM
📄	myNotebook.Rmd	890 B	Jul 5, 2020, 9:03 AM
📄	project.Rproj	205 B	Jul 10, 2020, 5:53 PM

Online Courses



Free LinkedIn Learning Path

Study here <https://www.linkedin.com/learning/paths/become-a-data-analyst>



1.40 Learning Data Analytics



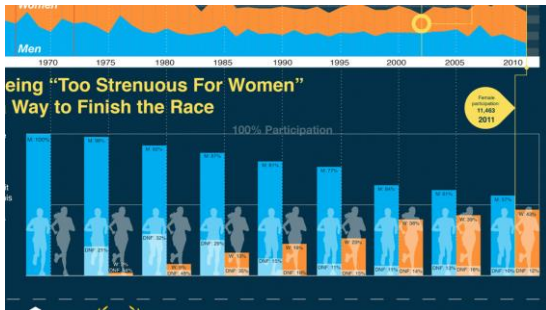
4.20 Data Fluency



3.40 Excel Statistics



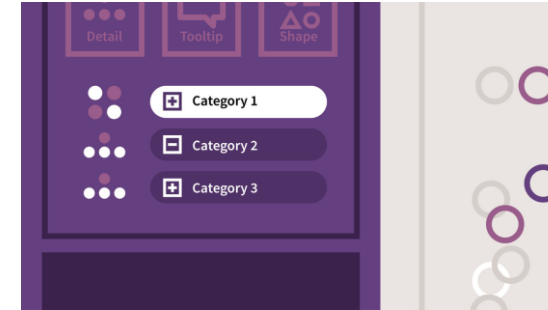
2.00 Excel Data Analysis



3.50 Data Visualization



3.20 Power BI



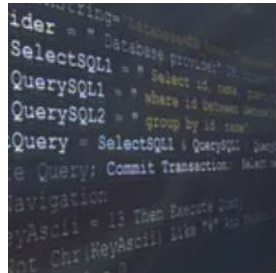
4.30 Tableau 2020.1

20 hours



Curated Data Analyst Learning Path

Study here <https://www.coursera.org/>



SQL for Data Science

University of California, Davis

COURSE

★★★★☆ 4.6 (5,467) | 180K students

Beginner



Statistics with R

Duke University

SPECIALIZATION

★★★★☆ 4.7 (5,055) | 240K students

Beginner



Excel Skills for Business

Macquarie University

SPECIALIZATION

★★★★☆ 4.9 (21,419) | 330K students

Beginner



From Data to Insights with Google Cloud Platform

Google Cloud

SPECIALIZATION

★★★★☆ 4.6 (2,176) | 30K students

Beginner



Data Visualization and Communication with Tableau

Duke University

COURSE

★★★★☆ 4.7 (2,417) | 140K students

Mixed



Digital Marketing

University of Illinois at Urbana-Champaign

SPECIALIZATION

★★★★☆ 4.6 (12,395) | 420K students

Beginner

60 hours



Good Free Resources to learn Cloud

<https://docs.microsoft.com/en-us/learn/browse/?roles=data-analyst>

Filter

Clear all

☐ Hide completed

Products

- ☐ Azure
- ☐ Dynamics
- ☐ Power Platform

Roles

Find a role

- ☐ Data Engineer
- ☐ App Maker
- ☐ Business Analyst
- ☐ Business Owner
- ☐ Business User
- ☒ Data Analyst
- ☐ Data Engineer
- ☐ Data Scientist

Search

Search

Data Analyst ✕

39 results



MODULE

Get data with Power BI Desktop

1 hr 15 min ★★★★★ 4.7 (11K)

Power Platform Business Analyst
Intermediate

COMPLETED ✓



LEARNING PATH

Create and use analytics reports with Power BI

6 hr 22 min

Power Platform Business User Beginner

89%



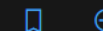
MODULE

Model data in Power BI

1 hr 7 min ★★★★★ 4.7 (4K)

Power Platform Business Analyst
Intermediate

COMPLETED ✓



MODULE



MODULE



MODULE

Certificates



03/11/2019

KASIDIS SATANGMONGKOL

has successfully completed the online, non-credit Specialization

From Data to Insights with Google Cloud Platform

This four-course accelerated online specialization teaches course participants how to derive insights through data analysis and visualization using the Google Cloud Platform. The courses feature interactive scenarios and hands-on labs where participants explore, mine, load, visualize, and extract insights from diverse Google BigQuery datasets. The courses cover data

1. Free Certificates
2. Paid Certificates
 - MOOC
 - University
 - Business/ Organization
3. Professional
 - AWS
 - Google Cloud
 - Microsoft Azure
4. Mini Bootcamp
5. Full Bootcamp

Google Cloud Training

Free Certificates

- Google Analytics <https://analytics.google.com/analytics/academy/>
- Google Fundamental of Digital Marketing
<https://learndigital.withgoogle.com/digitalgarage/course/digital-marketing>
- Free Code Camp
<https://www.freecodecamp.org/>



Free Certificates

<https://www.freecodecamp.org/>

- ▶ Responsive Web Design Certification (300 hours)
- ▶ JavaScript Algorithms and Data Structures Certification (300 hours)
- ▶ Front End Libraries Certification (300 hours)
- ▶ Data Visualization Certification (300 hours)
- ▶ APIs and Microservices Certification (300 hours)
- ▶ Quality Assurance Certification (300 hours)
- ▶ Scientific Computing with Python Certification (300 hours)
- ▶ Data Analysis with Python Certification (300 hours)
- ▶ Information Security Certification (300 hours)
- ▶ Machine Learning with Python Certification (300 hours)
- ▶ Coding Interview Prep (Thousands of hours of challenges)

Python for Everybody - Introduction: Why Program? ✓



01 - Intro A - Python for Everybody Course

Watch later Share

What is Code? Software? A Program?

- A sequence of stored instructions
 - It is a little piece of our intelligence in the computer
 - We figure something out and then we encode it and then give it to someone else to save them the time and energy of figuring it out
- A piece of creative art - particularly when we do a good job on user experience



[Help improve or add subtitles.](#)

Paid Certificates

Cornell University



This is to acknowledge that

Kasidis Satangmongkol

has completed the certificate program entitled

Business Statistics

This Twenty-Fifth Day of May, 2020

through



Dyana P. Wooten

Dean

Dyson School of Applied Economics and Management
Cornell University

Many options/ prices

- MOOC typically 49USD – 99USD for most certificates
- High end university offers such courses at very high prices

Closing Remarks



Professional Certificates

Leading companies and cloud providers

- Cost varies between 50-300 USD per certificate
- Need to recertify every 1-2 year depends on level



AWS has highest market share in cloud market, Azure is rising in popularity. GCP is strong for ML/AI.

Should you go for college degree?

“ **Education** is what, when and why to do things.
Training is how to do it.

Richard Hamming. (1915-1998)

Plan your study

1. Define your goal
2. Do the calculation

There are always alternatives for our every choice



Economics – We always trade off



250,000 – 300,000 THB
(2 years)

=

coursera PLUS

Unlimited access to **3,000+** courses,
Specializations, and Professional
Certificates

Try Coursera Plus

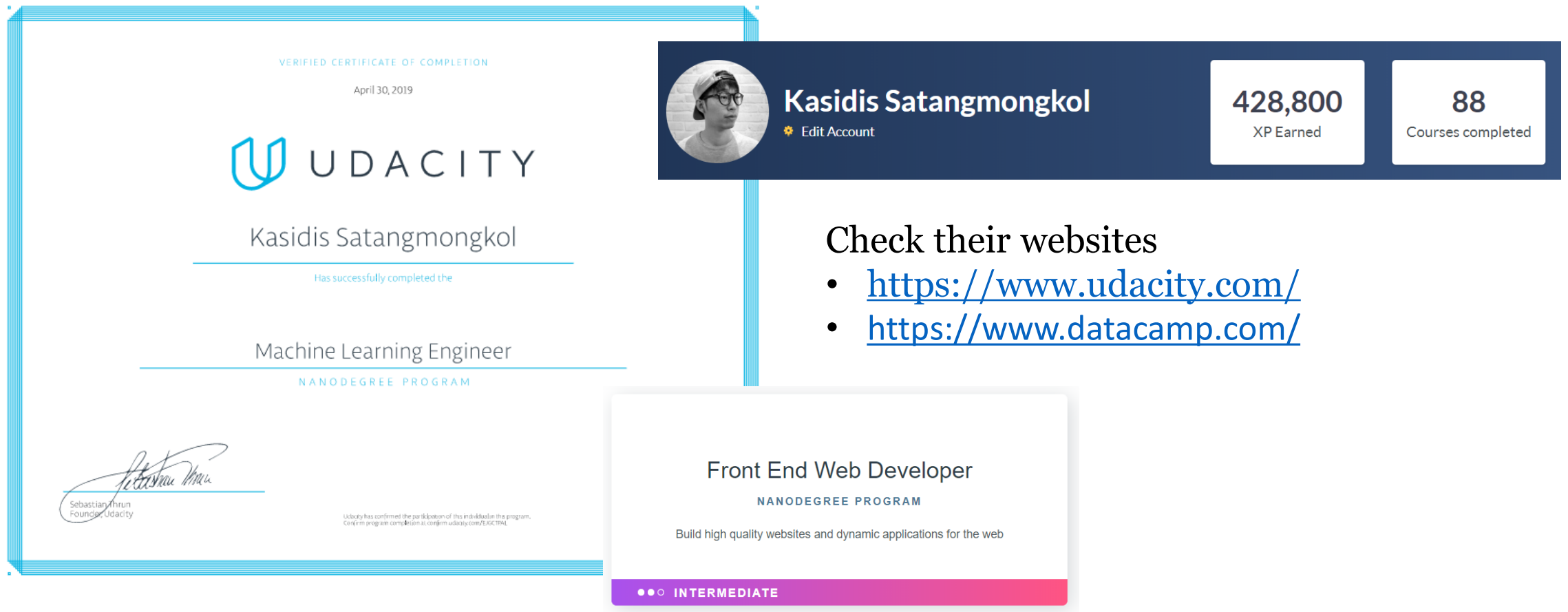
\$399 / year

14-day money-back guarantee

15-20 years on Coursera

Mini Bootcamp

- Udacity is a good example for this type of certificate
- Datacamp and Dataquest would fit into this category too



The image displays a Udacity 'Verified Certificate of Completion' for Kasidis Satangmongkol, dated April 30, 2019. The certificate is for the 'Machine Learning Engineer' role, part of the 'NANODEGREE PROGRAM'. It features the Udacity logo, a signature from Sebastian Thrun, and a confirmation of participation. To the right, a dark blue profile bar shows a circular profile picture of Kasidis, his name, an 'Edit Account' link, and statistics: 428,800 XP Earned and 88 Courses completed. Below the certificate, a white box highlights the 'Front End Web Developer' role, also part of the 'NANODEGREE PROGRAM', with a description: 'Build high quality websites and dynamic applications for the web'. At the bottom of this box, a pink bar indicates the 'INTERMEDIATE' level.

VERIFIED CERTIFICATE OF COMPLETION

April 30, 2019

UDACITY

Kasidis Satangmongkol

Has successfully completed the

Machine Learning Engineer

NANODEGREE PROGRAM

Sebastian Thrun
Founder, Udacity

Udacity has confirmed the participation of this individual in the program.
Confirm program completion at confirm.udacity.com/EJGKTR9L

Kasidis Satangmongkol
✦ Edit Account

428,800
XP Earned

88
Courses completed

Check their websites

- <https://www.udacity.com/>
- <https://www.datacamp.com/>

Front End Web Developer

NANODEGREE PROGRAM

Build high quality websites and dynamic applications for the web

●●○ INTERMEDIATE

Bootcamp

High end + High quality

Similar to taking classes in University, focus on building skills and applications

<https://generalassemb.ly/education/data-science-immersive/singapore>

Black lives matter. We do not tolerate racism or racial harassment of any kind — and we never will. Read our full statement [here](#).

GENERAL ASSEMBLY

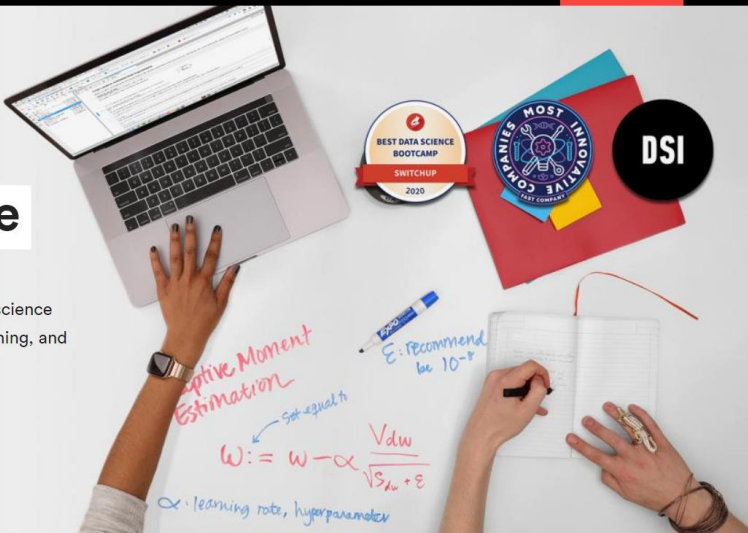
On Campus * Online * Locations * For Companies * About * Find Your Course Sign In

A-LIST HIRING PARTNERS

Data Science Immersive: Singapore

Your best course for career transformation. This full-time data science bootcamp features expert instruction, one-on-one career coaching, and connections to top employers to get you hired.

[Apply Now](#) [Request Syllabus](#)



19K+ Premier Hiring Partners From Around the World

Overview

Connected Campus

walkme CHASE Cerner IndigoSlate Freelock AT&T

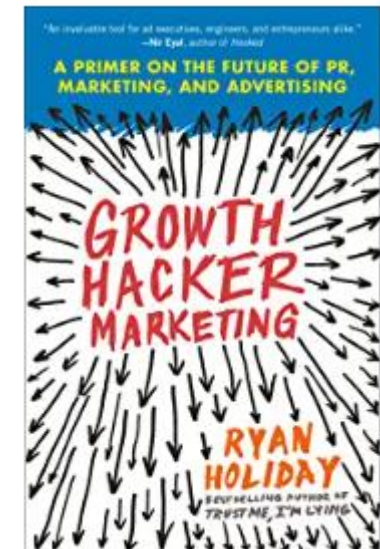
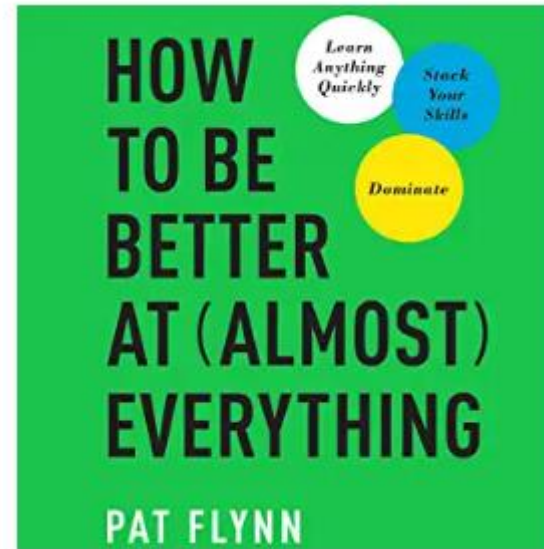
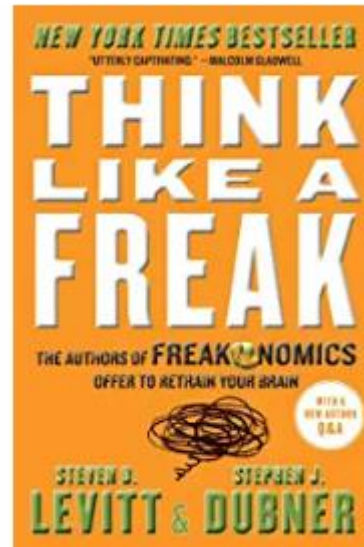
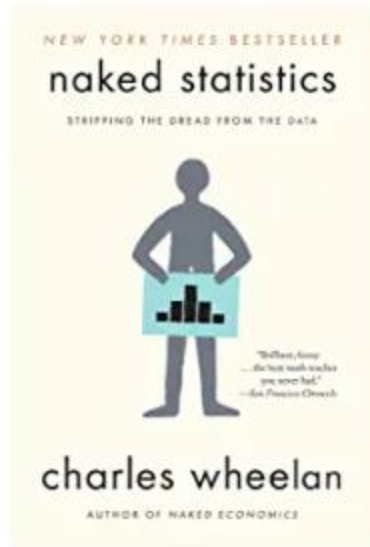
300,000-500,000 THB

How to learn effectively?



- Find your best time to learn
- Consistency > Quantity
- Learning is like riding a bicycle
- Find learning partners/ friends
- Your mind determines your success

Books for Starter :D



Highly recommended for everyone who work with data/ business

Join Our Groups and Learn Together

<https://web.facebook.com/groups/databootcamp>



Data Analyst Accelerated Bootcamp

Public group · 3.7K members



+ Invite



About

Discussion

Announcements

Members

Events

Media





Q&A