MySkill

Intro to Data Analysis

MySkill

Ada yang pernah pakai SQL?

- Pernah
- Kadang-kadang
- Sering
- Baru dengar

MySkill

Ada yang pernah pakai Excel/Google Sheet?

- Pernah
- Kadang-kadang
- Sering
- Baru dengar

MySkill

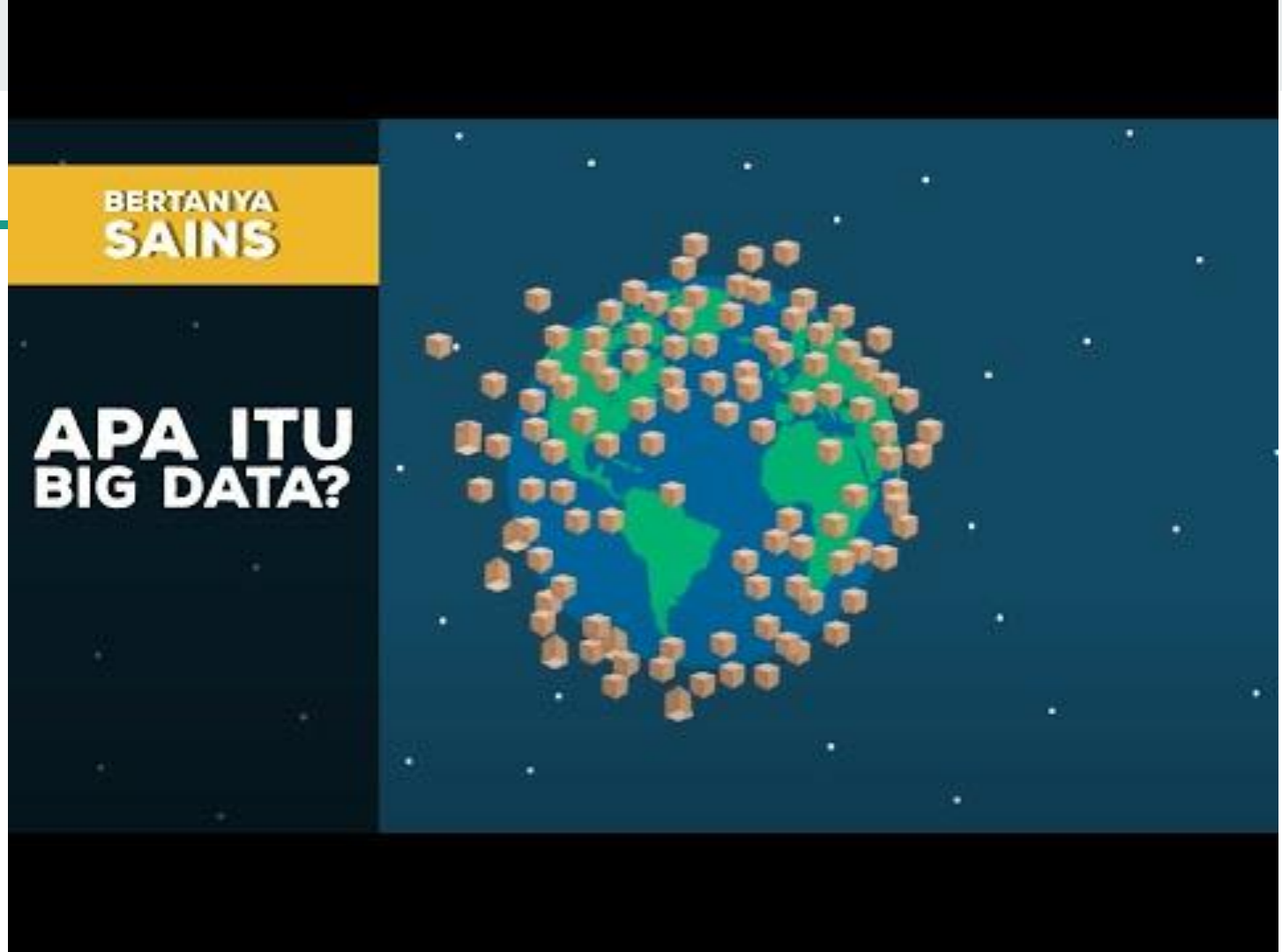
Ada yang pernah pakai Google Datastudio?

- Pernah
- Kadang-kadang
- Sering
- Baru dengar

MySkill

Data_

MySkill



MySkill

The Economist

Menu | Weekly edition | Search | Subscribe

Leaders

May 6th 2017 edition ›

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

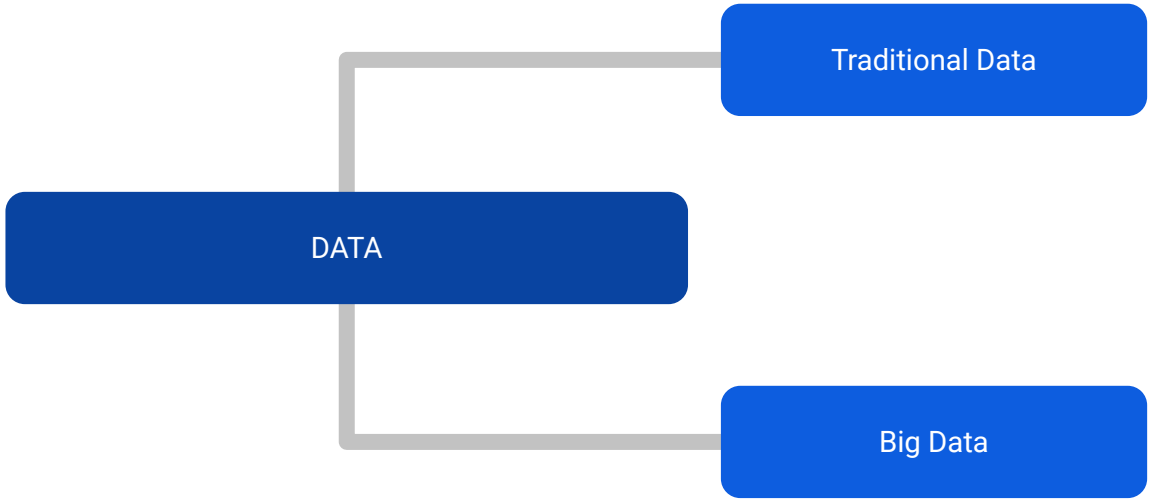
The data economy demands a new approach to antitrust rules

Forbes

RANK	NAME	NET WORTH	AGE	COUNTRY / TERRITORY	SOURCE
1.	Elon Musk	$219 B ▲	50	United States	Tesla, SpaceX
2.	Jeff Bezos	$171 B ▼	58	United States	Amazon
3.	Bernard Arnault & family	$158 B ▲	73	France	LVMH
4.	Bill Gates	$129 B ▲	66	United States	Microsoft
5.	Warren Buffett	$118 B ▲	91	United States	Berkshire Hathaway
6.	Larry Page	$111 B ▲	49	United States	Google
7.	Sergey Brin	$107 B ▲	48	United States	Google

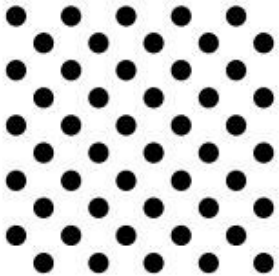
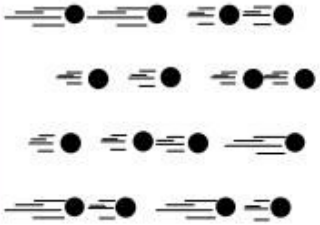
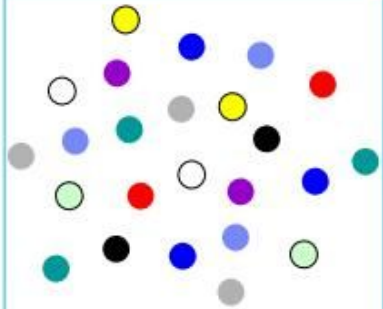
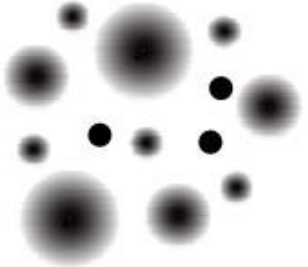
Big Data vs Traditional Data

MySkill



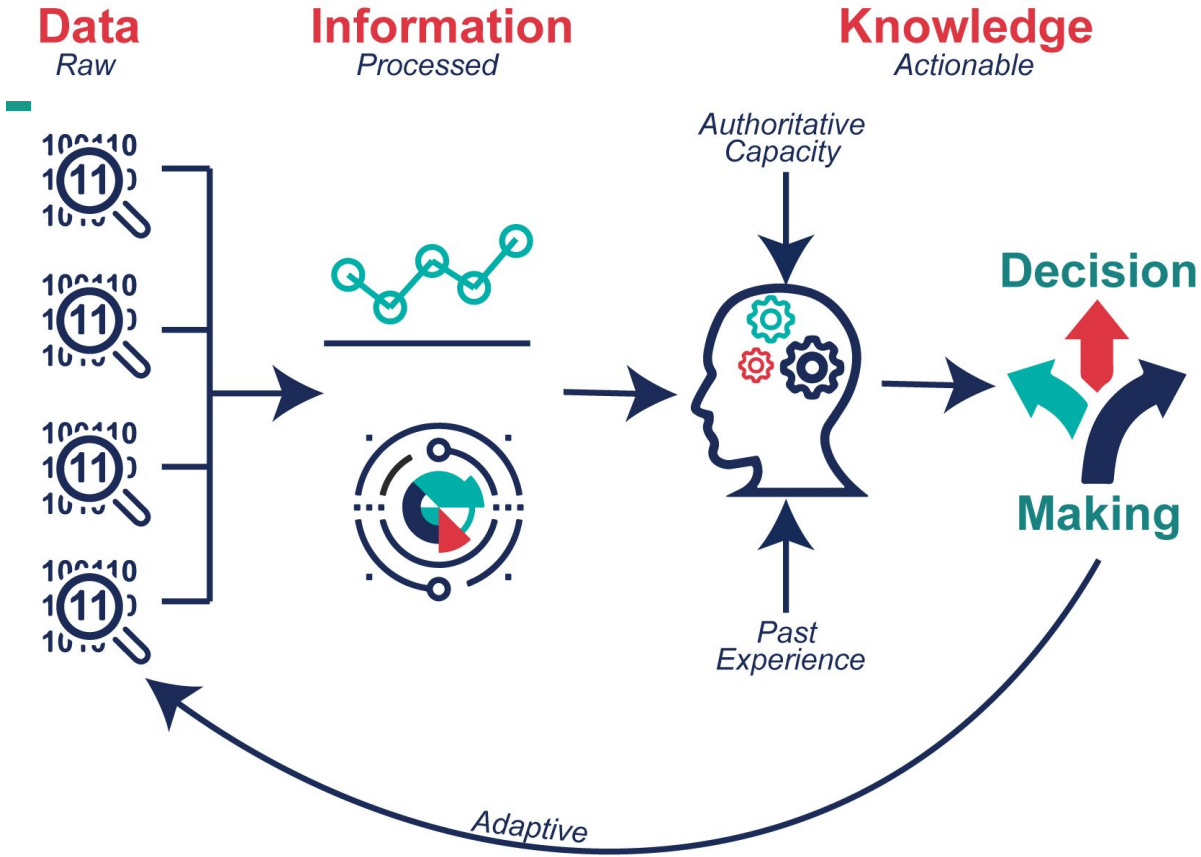
Big Data vs Traditional Data

MySkill

Volume	Velocity	Variety	Veracity*
			
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

https://www.datasciencecentral.com/profiles/blogs/data-veracity

MySkill



https://internetofwater.org/valuing-data/what-are-data-information-and-knowledge/

MySkill

Data	Information	Knowledge
Amount of precipitation in rain gauge	Assess whether annual precipitation is increasing, decreasing, or staying the same	Prioritize improvement of irrigation system given increases in precipitation over last 20 years

MySkill

Data Analysis

MySkill

What is Data Analysis

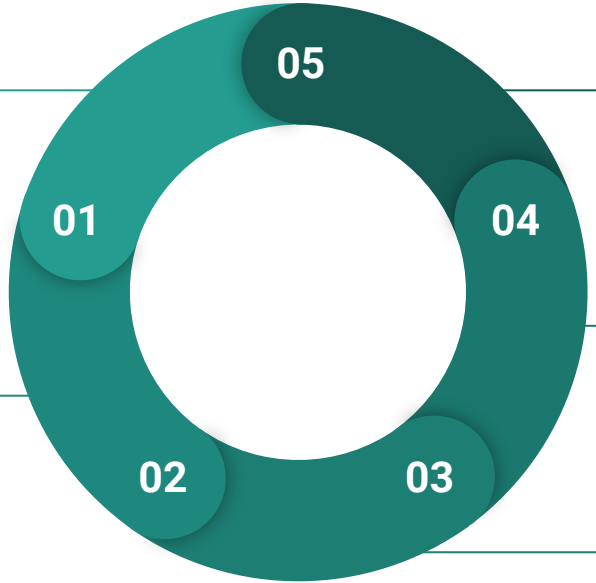
"Data analysis is the practice of working with data to glean useful information, which can then be used to make informed decisions."

Importance of Data Analysis

MySkill

Track performance

Optimize operations



Spot market trends

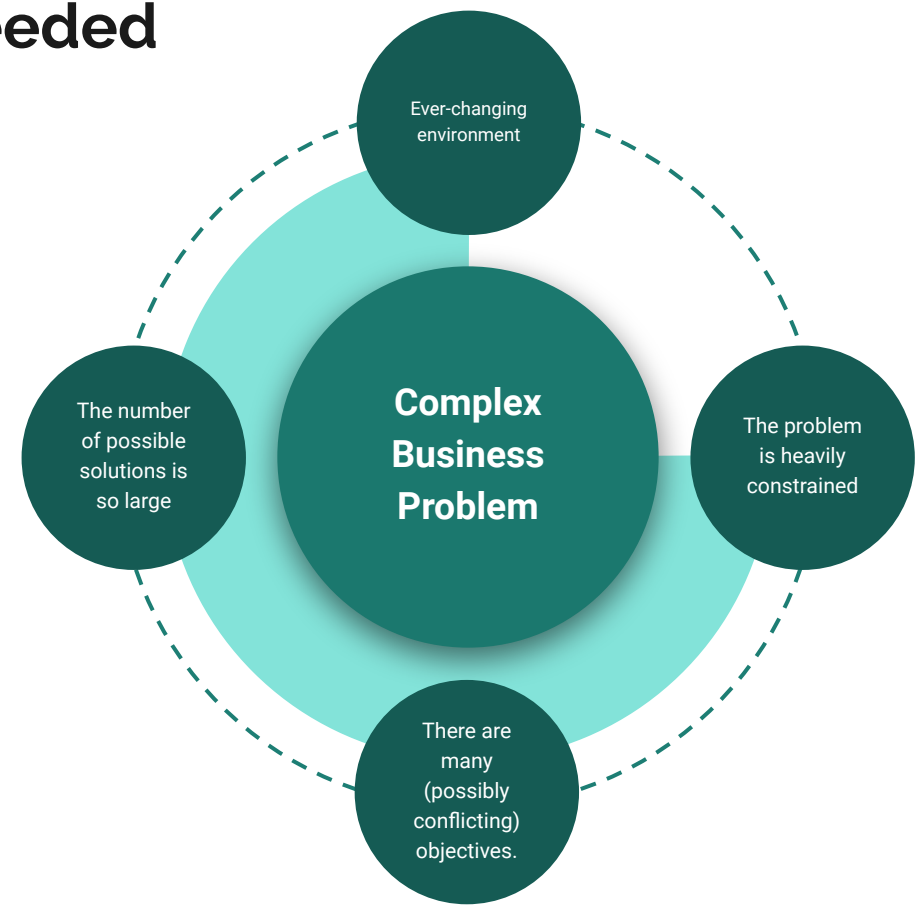
Faster decision making

Identify ways to increase profit

https://www.tableau.com/learn/articles/business-intelligence

Why DA needed

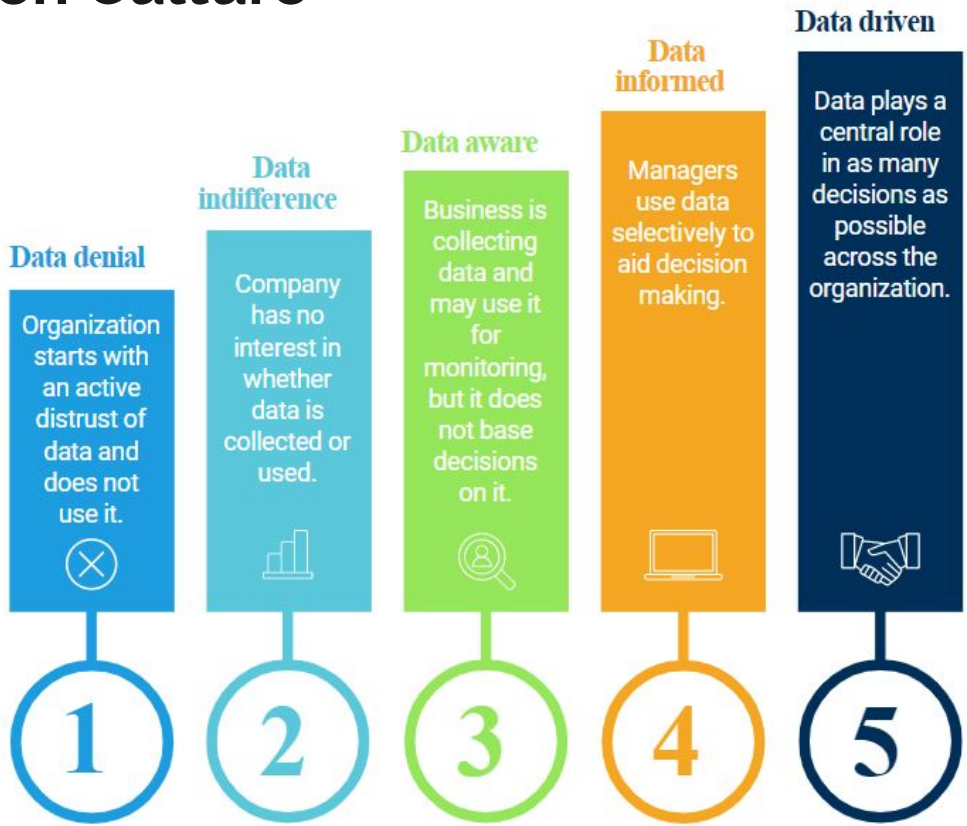
MySkill



https://link.springer.com/chapter/10.1007/978-3-540-32929-9_2

Data Driven Culture

MySkill



https://www.smartsheet.com/data-driven-decision-making-management

WORLD ECONOMIC FORUM

Top 10 skills of 2025

MySkill

Source: Future of Jobs Report 2020, World Economic Forum.

MySkill

What a Data Analyst does?

What a DA does

https://www.g2.com/articles/data-analysis-process

MySkill

Data Scientist

uses statistics and machine learning to make predictions and answer key business questions

Skills - Math, Programming, Statistics


Tech - SQL, Python, R, Cloud

Data Engineer

build and optimize the systems that allow data scientists and analysts to perform their work

Skills - Programming, BigData & Cloud


Tech - SQL, Python, Cloud, Distributed Computing

Data Analyst

deliver value by taking data, communicating the results to help make business decisions

Skills - Communication, Business Knowledge


Tech - SQL, Excel, Tableau

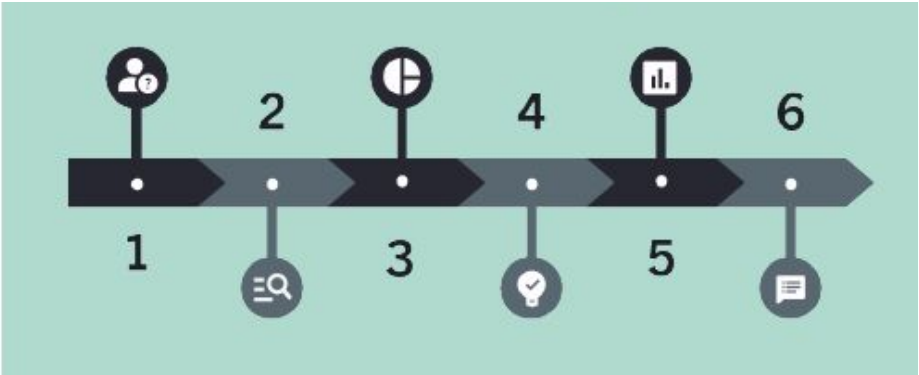
MySkill

Data-driven decision making

using facts, metrics, and data to guide strategic business decisions that align with your goals, objectives, and initiatives

MySkill

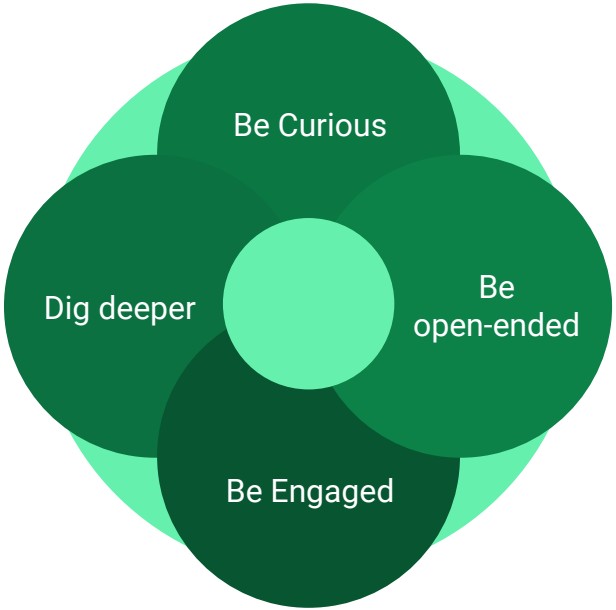
Data-driven decision making framework



1. **Understand the business problem:** What are you looking to understand or accomplish?
2. **Wrangle data:** Clean, validate, and organize the data.
3. **Create visualizations:** Present the data in a way that shows trends and relationships of interest.
4. **Generate hypotheses:** Formulate predictions based on emerging trends.
5. **Conduct analysis:** Run statistical tests to determine if your hypotheses are correct.
6. **Communicate results:** Present your findings in the context of the original business problem.

MySkill

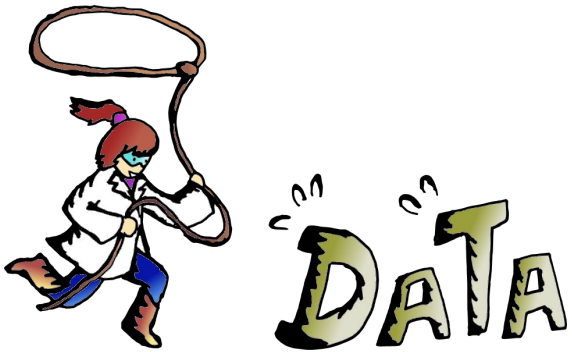
Understanding Business Problem



MySkill

Data Wrangling

Data wrangling is the process of cleaning raw data in preparation for analysis. It involves identifying and resolving mistakes, filling in missing data, and organizing and transferring it into an easily understandable format.



MySkill

Data Visualizations

Data **visualization** gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets.

MySkill

Data Visualizations

PUBG Finish Placement Prediction (Kernels Only)

Can you predict the battle royale finish of PUBG Players?

Last Updated: 3 years ago

About this Competition

In a PUBG game, up to 100 players start in each match (matchId). Players can be on teams (groupId) which get ranked at the end of the game (winPlacePerc) based on how many other teams are still alive when they are eliminated. In game, players can pick up different munitions, revive downed-but-not-out (knocked) teammates, drive vehicles, swim, run, shoot, and experience all of the consequences -- such as falling too far or running themselves over and eliminating themselves.

You are provided with a large number of anonymized PUBG game stats, formatted so that each row contains one player's post-game stats. The data comes from matches of all types: solos, duos, squads, and custom; there is no guarantee of there being 100 players per match, nor at most 4 player per group.

MySkill

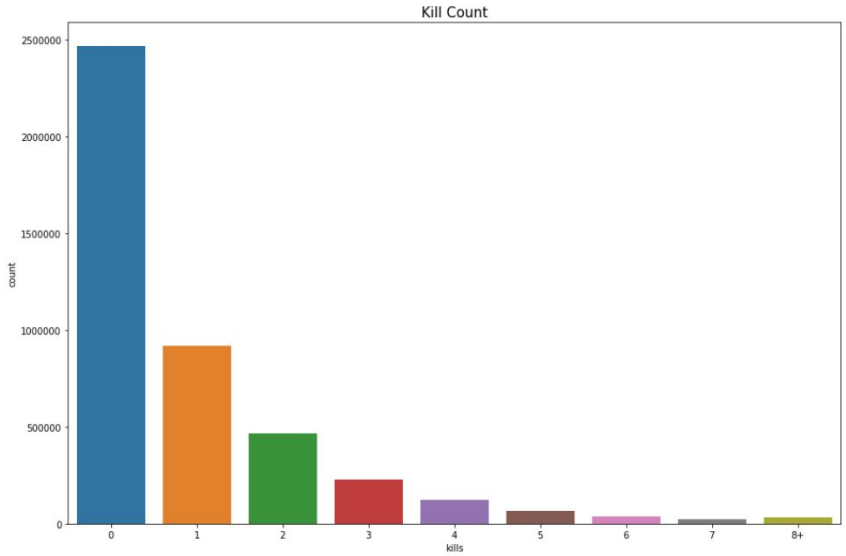
Data Visualizations

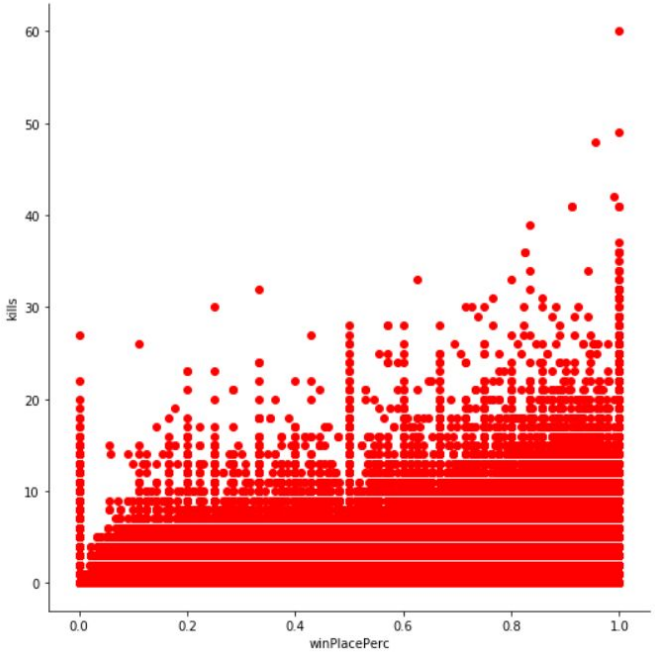
Raw Data

Id	groupId	matchId	assists	boosts	damageDealt	DBNOs	headshotKills	heals	killPlace	killPoints	kills	killStreaks	longestKill	maxPlace	numGroups	revives	rideDistance	roadKills	swimDistance	teamKills	vehicleDestroys	walkDistance	weaponsAcquired	winPoints	winPlacePerc
0	0	24	0	0	5	247.3	2	0	4	17	1050	2	1	65.32	29	28	1	591.3	0	0	0	0	782.4	4	1458
1	1	440875	1	1	0	37.65	1	1	0	45	1072	1	1	13.55	26	23	0	0	0	0	0	0	119.6	3	1511
2	2	878242	2	0	1	93.73	1	0	2	54	1404	0	0	0	28	28	1	0	0	0	0	0	3248	5	1583
3	3	1319841	3	0	0	95.88	0	0	0	86	1069	0	0	0	97	94	0	0	0	0	0	0	21.49	1	1489
4	4	1757883	4	0	1	0	0	0	1	58	1034	0	0	0	47	41	0	0	0	0	0	0	640.8	4	1475

MySkill

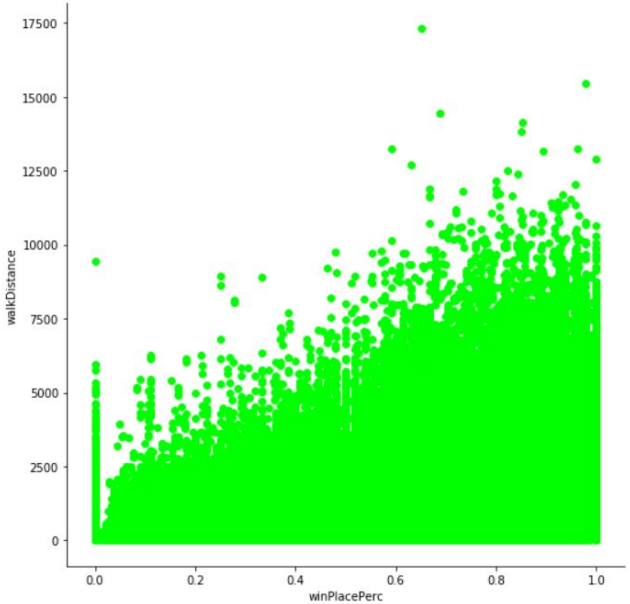
Data Visualizations

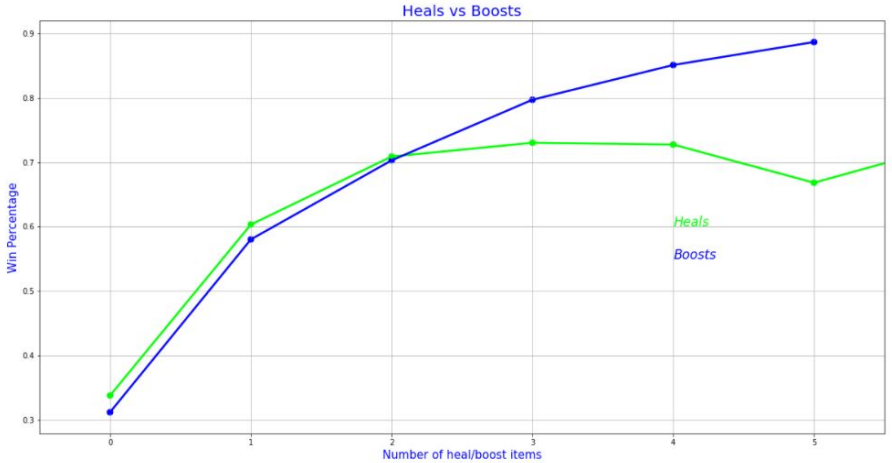




MySkill

Data Visualizations



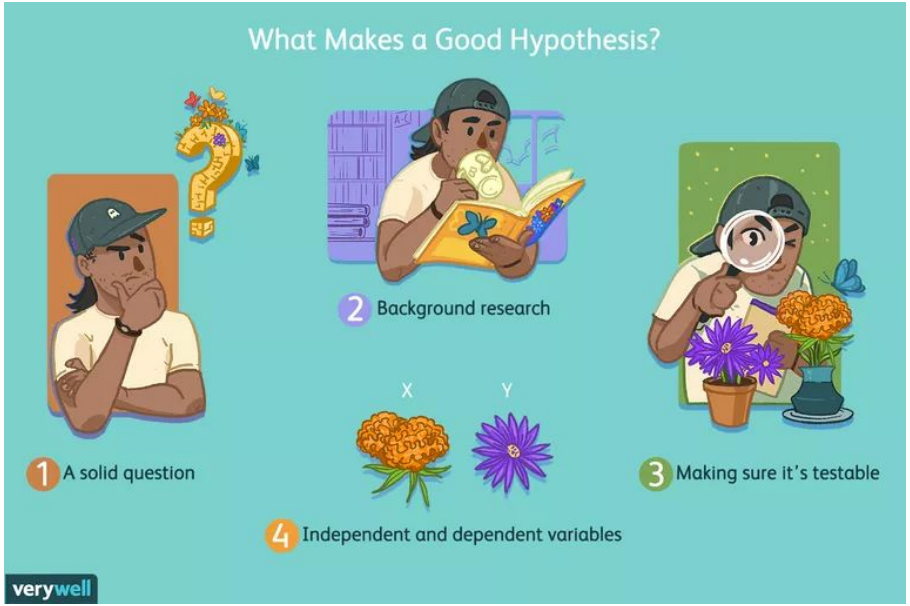


MySkill

Data Visualizations

MySkill

Generate Hypothesis



MySkill

Conduct Analysis

Performing investigations on data so as to **discover patterns,to spot anomalies,to test hypothesis and to check assumptions, and gather insights**.

MySkill

Type of Analysis

- Descriptive Analysis: **summarizes** the data at hand and presents your data in a nice way.
- Exploratory Analysis: discover **correlations and relationships** between variables in your data
- Inferential Analysis: **generalizing** the larger population with a smaller sample size of data
- Predictive Analysis: make **predictions** about the future with data
- Prescriptive Analysis: offers specific recommendations for **changing the future**
- Causal Analysis: **finding the cause** of a correlation between variables

https://towardsdatascience.com/the-six-types-of-data-analysis-75517ba7ea61

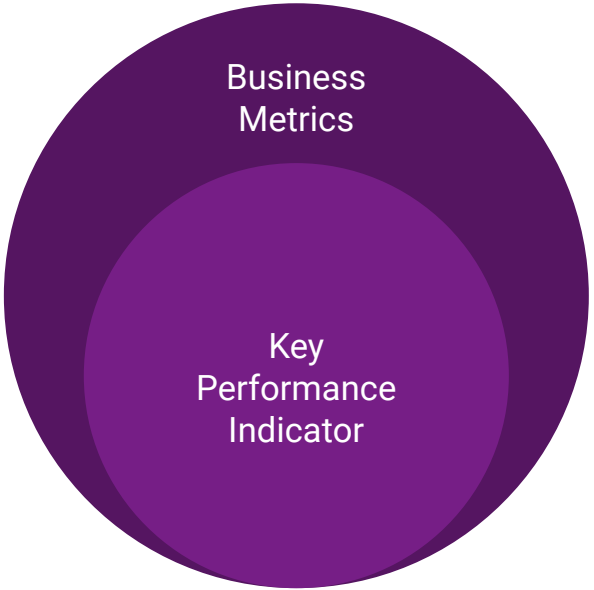
MySkill

Business Metrics

A business metric is a **quantifiable measure businesses use to track, monitor and assess** the success or failure of various business processes.

MySkill

Measure whats matters



Business Metrics: Example

MySkill

- Daily/Monthly Active Users
- Churn Rate/ Retention Rate
- Adoption Rate
- Completion Rate
- Conversion Rate
- Unit Economics

MySkill

Marketing Funnel p30d

May 2016

Metric	Value	Conversion Rate
Visitors	152,485	
Interactions	77,113	50.57%
Leads	3,923	5.09%
New Wins	975	24.85%

MySkill

Web Traffic Targets (This Month)

Channel	Progress	Sessions	Conversion Rate
Organic		243,737	5.63%
Display		112,378	1.98%
Direct		47,324	8.53%
Referral		13,790	13.09%
Social		12,023	26.95%
Email		3,864	5.61%
Paid Search		3,549	13.01%
Other		1,943	11.89%
		438,606	**5.91%**

MySkill

11 KEY HR METRICS

AIHR
BLOG & ACADEMY

MySkill

Communicate Result

Present the findings along with action/suggestion that need to be done to the stakeholders.

Data storytelling: context, problem, finding, action!

Don’t share too many details

Highlight the key point

MySkill

MySkill Store

Oct 1, 2021 - Oct 31, 2021

Device Category | Country | Source / Medium | User Type

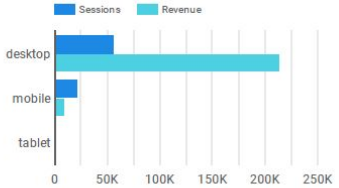
Sessions
78.4K

Product Detail Views
27.8K

Product Adds To Cart
No data

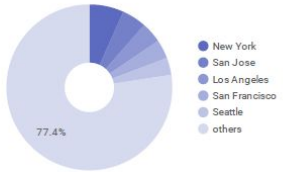
Product Checkouts
No data

Product	Product Detail Views	Product Adds To Cart	Product Revenue	Product Revenue per Purchase	Cart-to-Detail Rate
Google Youth Hero Tee Maroon	22	0	$96	$24	0%
Google Large Pet Leash (Blue/Gr...	23	0	$56	$56	0%
Google Pen Grey	34	0	$375.55	$10.15	0%
Google Black Eco Zip Hoodie	871	0	$8,640	$55.38	0%
Google Soft Frisbee Yellow	0	0	$34	$4.86	0%



Default Channel Grouping	Product Detail Views	Revenue	Ecommerce Conversion Rate
Direct	26,974	$213,178.44	3.07%
Affiliates	68	$351.5	2.61%
Paid Search	710	$8,247.02	2.09%
Display	2	$0	0%
Referral	0	$0	0%

Revenue by City



What we gonna learn

MySkill

1. Intro to Data Analysis
2. Basic Statistics
3. Business Metrics: Measure What Matters
4. SQL Basic 1: Basic Clause
5. SQL Basic 2: Working with Multiple Data Sources
6. SQL Basic 3: Analyzing Business Data
7. Data Analysis with Python 1: Basic Python + Project 1
8. Data Analysis with Python 2: Work with Numpy and Pandas
9. Data Analysis with Python 3: Study Cases
10. Data Visualisation 1
11. Data Visualisation 2: Google Data Studio
12. Secure Your First Data Analyst Job

MySkill

Thank you