

# Business Statistics

## Overview

**Zhanrui Cai**

Assistant Professor in Analytics and Innovation

# Zhanrui Cai

---



**Zhanrui Cai**

Having fun in statistics



HK



Github



Google Scholar

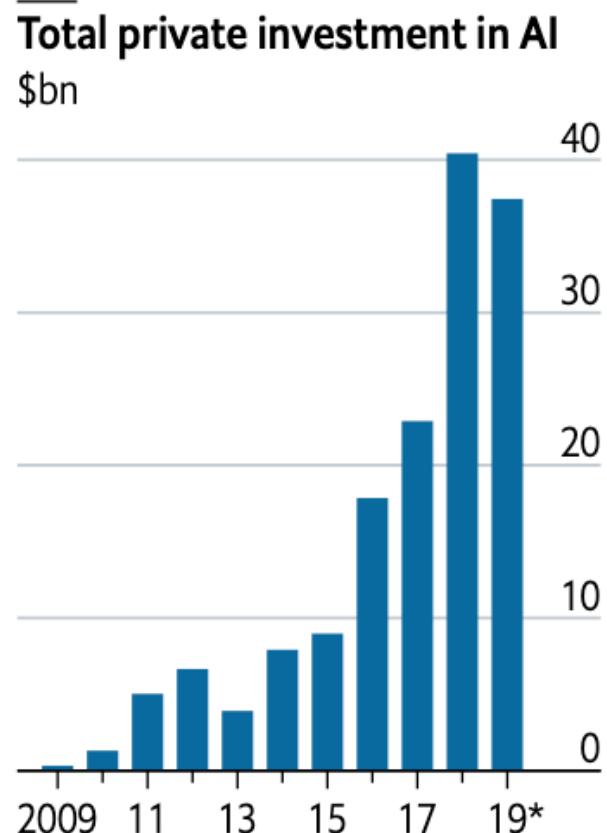
I am a tenure-track assistant professor in the Faculty of Business and Economics at the [HKU Business School](#). My home department is in Innovation and Information Management (Business Analytics). [Link](#)

Experience:

- Assistant professor in the [Department of Statistics](#), Iowa State University, 2022 - 2023.
- Postdoctoral researcher in the [Department of Statistics and Data Science](#), Carnegie Mellon University, 2021 - 2022.
- Ph.D. in the [Department of Statistics](#), Pennsylvania State University, 2016 - 2021.
- Bachelor in the [School of Statistics, Renmin University of China](#), 2012 - 2016.

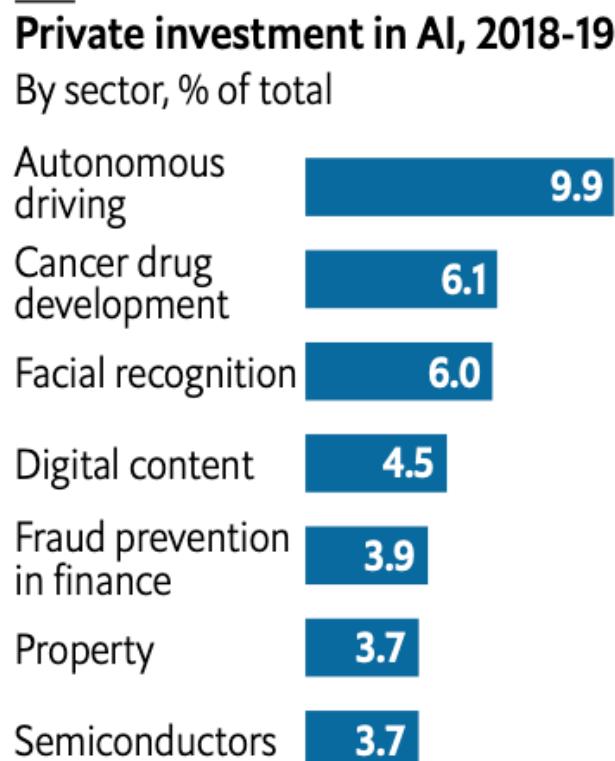
# The Landscape of AI

## On the up Worldwide

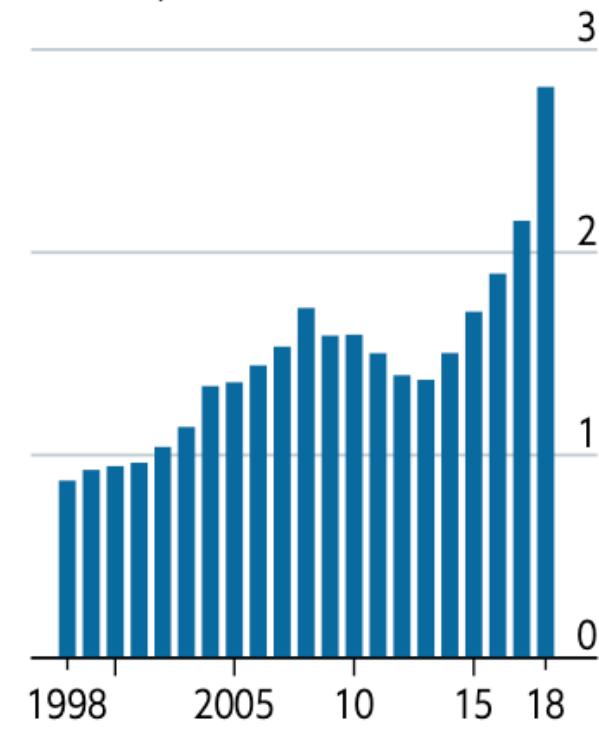


Source: Stanford Institute for Human-Centred Artificial Intelligence

## On the up Worldwide



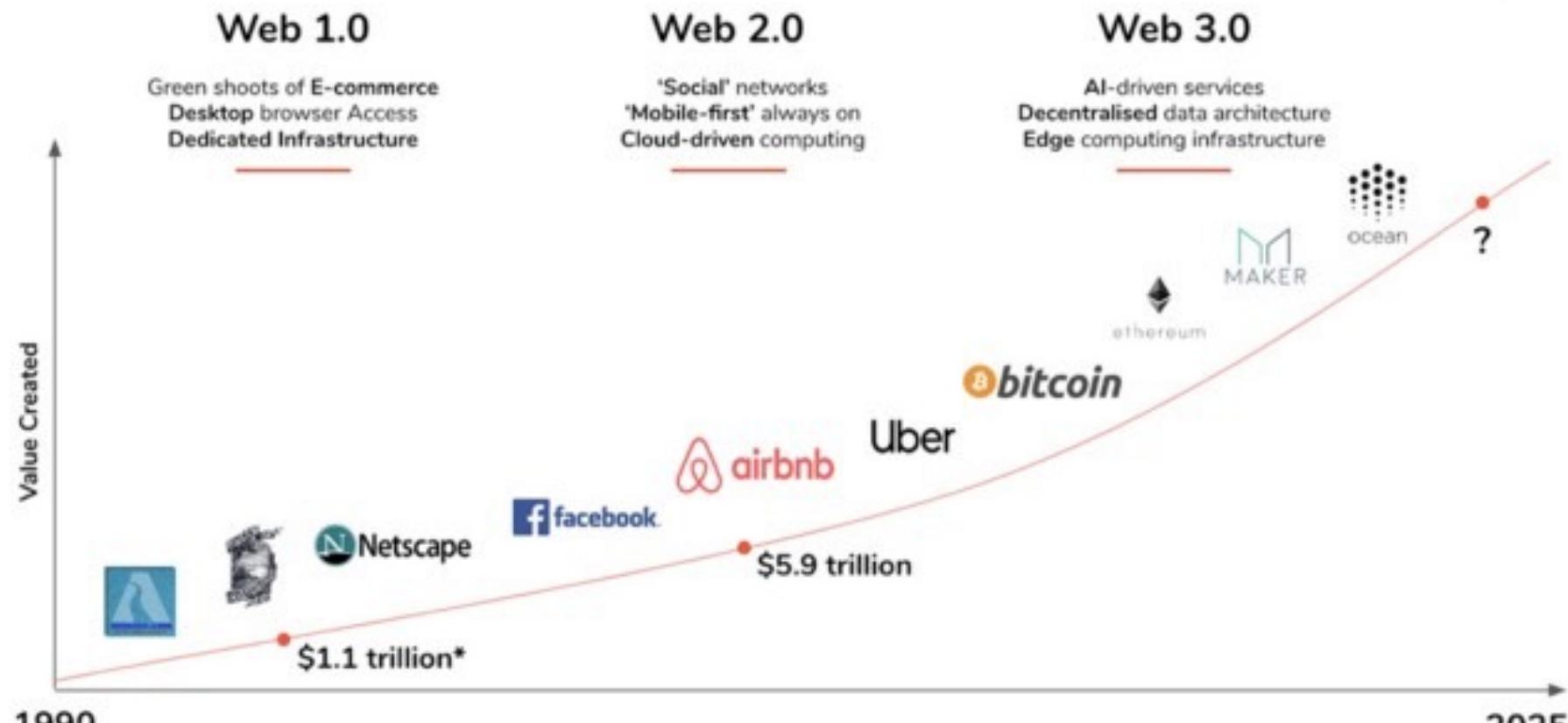
## AI papers published % of all publications



\*To October



## The Evolution of the Web



\* Internet companies market cap as of 2000.

# Stephen Hawking Warning (2 Dec 2014, 4:00 mark)

---



Stephen Hawking: "Humans, who are limited by slow biological evolution, couldn't compete and would be superseded"

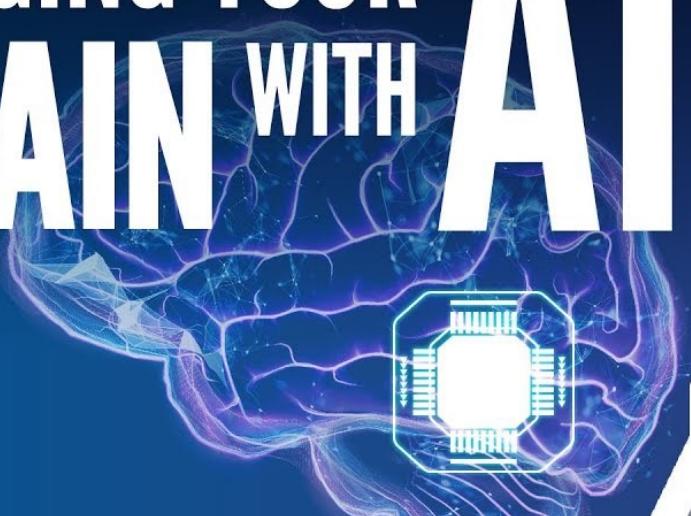
---

“

The danger of AI is much greater than the danger of nuclear warheads by a lot...Mark my words. AI is far more dangerous than nukes.”

Elon Musk, SXSW Conference, 2018

# NEURALINK MERGING YOUR BRAIN WITH AI



# Monkey MindPong

---

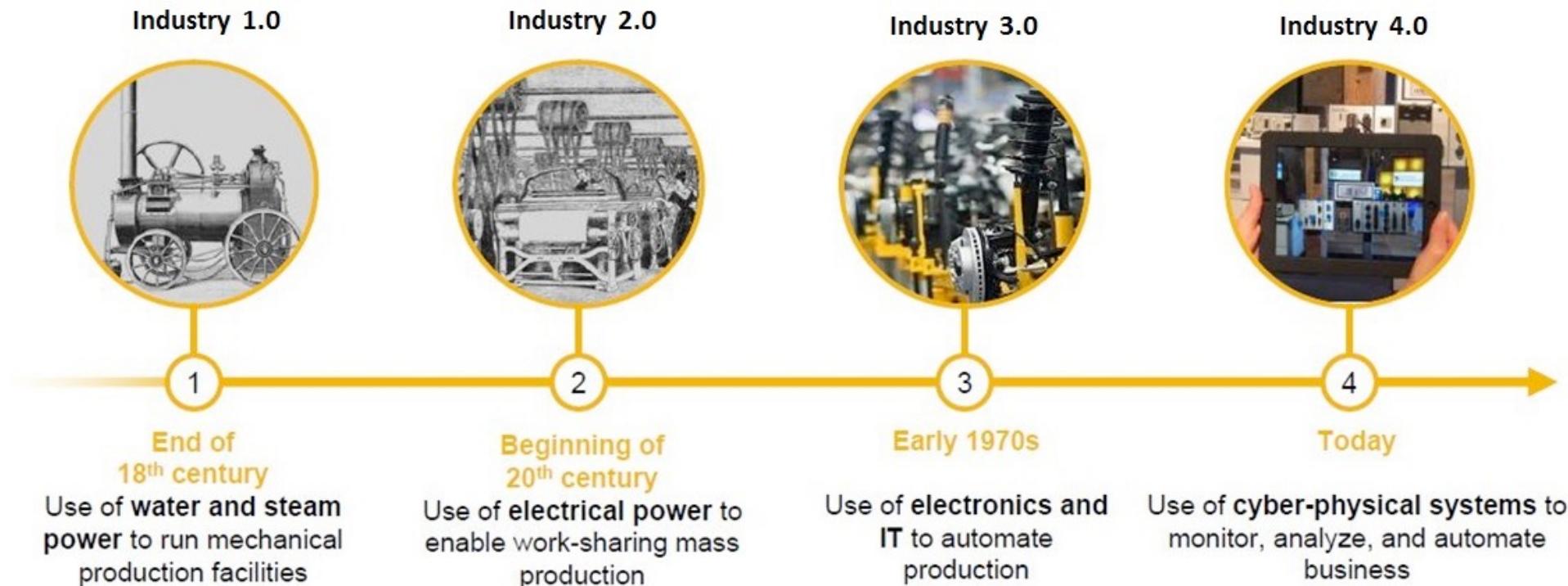


# Huawei REN Zhengfei – Artificial Intelligence is Statistics!



# Connectivity, Digitalization, Internet of Things

## Four Phases of Industrialization



**Data: Anything that are RECORDABLE!**

# Big Data

---

- Dataset whose size (and complexity) is beyond the ability of typical database tools to capture, store, manage, and analyze.
- Big data includes structured (high level of organizations, residing in fixed fields such as relational databases) + unstructured data.

# Business Databases: Structured Data

---

call_id	customer_id	priority	type	date	vru_entry	vru_exit	vru_time	q_start	q_exit	q_time	outcome	ser_start	ser_exit	ser_time	server	
47529	0	0	PS	991101	6:31:36 A	6:31:47 AM		11	12:00:00 AM	12:00:00 AM	0	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47531	0	0	PS	991101	7:42:41 A	7:42:51 AM		10	12:00:00 AM	12:00:00 AM	0	AGENT	7:42:50 AM	7:43:30 A	40	TOVA
47532	0	0	NW	991101	8:03:36 A	8:03:46 AM		10	8:03:46 AM	8:05:33 AM	107	AGENT	8:05:31 AM	8:07:38 A	127	YITZ
47533	4345955	2	PS	991101	8:17:40 A	8:17:46 AM		6	8:17:46 AM	8:18:01 AM	15	AGENT	8:18:00 AM	8:18:35 A	35	KAZAV
47534	0	0	PS	991101	8:27:47 A	8:27:56 AM		9	12:00:00 AM	12:00:00 AM	0	AGENT	8:27:55 AM	8:30:46 A	171	KAZAV
47535	0	0	PS	991101	8:41:11 A	8:41:21 AM		10	12:00:00 AM	12:00:00 AM	0	AGENT	8:41:20 AM	8:43:50 A	150	TOVA
47536	69438695	1	PS	991101	8:50:34 A	8:50:40 AM		6	8:50:40 AM	8:51:59 AM	79	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47537	0	0	PS	991101	8:58:25 A	8:58:50 AM		25	12:00:00 AM	12:00:00 AM	0	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47538	0	0	PS	991101	9:05:19 A	9:05:28 AM		9	12:00:00 AM	12:00:00 AM	0	AGENT	9:05:27 AM	9:08:09 A	162	ZIPI
47539	57281792	2	PS	991101	9:12:47 A	9:12:54 AM		7	9:12:54 AM	9:13:31 AM	37	AGENT	9:13:30 AM	9:15:16 A	106	ZIPI
47540	55279970	2	NE	991101	9:18:40 A	9:18:46 AM		6	9:18:46 AM	9:18:58 AM	12	AGENT	9:18:57 AM	9:25:12 A	375	VICKY
47541	25219700	2	PS	991101	9:25:52 A	9:25:58 AM		6	9:25:58 AM	9:26:17 AM	19	AGENT	9:26:16 AM	9:27:37 A	81	ZOHARI
47542	2057255	2	PS	991101	9:32:20 A	9:32:25 AM		5	9:32:25 AM	9:32:43 AM	18	AGENT	9:32:43 AM	9:33:47 A	64	VICKY
47543	65338329	2	PS	991101	9:37:47 A	9:37:52 AM		5	9:37:52 AM	9:38:03 AM	11	AGENT	9:38:02 AM	9:47:09 A	547	ZIPI
47544	2018299	2	PS	991101	9:44:16 A	9:44:21 AM		5	9:44:21 AM	9:44:53 AM	32	AGENT	9:44:53 AM	9:52:31 A	458	TOVA
47545	69687275	2	PS	991101	9:51:04 A	9:51:10 AM		6	9:51:10 AM	9:52:27 AM	77	AGENT	9:52:26 AM	9:55:18 A	172	ROTH
47546	57379505	2	PS	991101	9:56:01 A	9:56:07 AM		6	9:56:07 AM	9:56:52 AM	45	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47547	24286130	2	PS	991101	9:59:35 A	9:59:40 AM		5	9:59:40 AM	10:01:29 AM	109	AGENT	10:01:29 A	10:02:17 A	48	TOVA
47548	52031556	2	PS	991101	10:07:20 A	10:07:25 A		5	10:07:25 AM	10:07:29 AM	4	AGENT	10:07:29 A	10:08:41 A	72	YITZ
47549	3968047	2	PS	991101	10:12:35 A	10:12:40 A		5	10:12:40 AM	10:12:58 AM	18	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47550	51458594	2	PS	991101	10:18:55 A	10:19:00 A		5	10:19:00 AM	10:19:16 AM	16	AGENT	10:19:15 A	10:19:41 A	26	GOLDIE
47551	7191646	2	NE	991101	10:24:48 A	10:24:54 A		6	10:24:54 AM	10:25:38 AM	44	AGENT	10:25:37 A	10:54:16 A	1719	ELI
47552	63730469	1	PS	991101	10:29:42 A	10:29:47 A		5	10:29:47 AM	10:30:03 AM	16	AGENT	10:30:02 A	10:30:53 A	51	VICKY
47553	23053663	2	PS	991101	10:37:27 A	10:37:32 A		5	10:37:32 AM	10:38:06 AM	34	AGENT	10:38:04 A	10:40:24 A	140	TOVA
47554	27317791	1	PS	991101	10:43:20 A	10:43:25 A		5	10:43:25 AM	10:44:53 AM	88	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47555	55476253	2	NE	991101	10:49:29 A	10:49:35 A		6	10:49:35 AM	10:50:31 AM	56	AGENT	10:50:30 A	10:51:06 A	36	VICKY
47556	42450676	2	PS	991101	10:55:19 A	10:55:25 A		6	10:55:25 AM	10:56:53 AM	88	HANG	12:00:00 A	12:00:00 A	0	NO_SERVER
47557	25022625	2	PS	991101	11:04:32 A	11:04:38 A		6	11:04:38 AM	11:04:44 AM	6	AGENT	11:04:44 A	11:05:10 A	26	KAZAV
47558	74780917	2	NE	991101	11:09:57 A	11:10:03 A		6	11:10:03 AM	11:11:17 AM	74	AGENT	11:11:15 A	11:16:41 A	326	KAZAV
47559	30663462	2	PS	991101	11:20:30 A	11:20:35 A		5	11:20:35 AM	11:21:50 AM	75	AGENT	11:21:49 A	11:28:03 A	374	GOLDIE

+ Unstructured Data

---

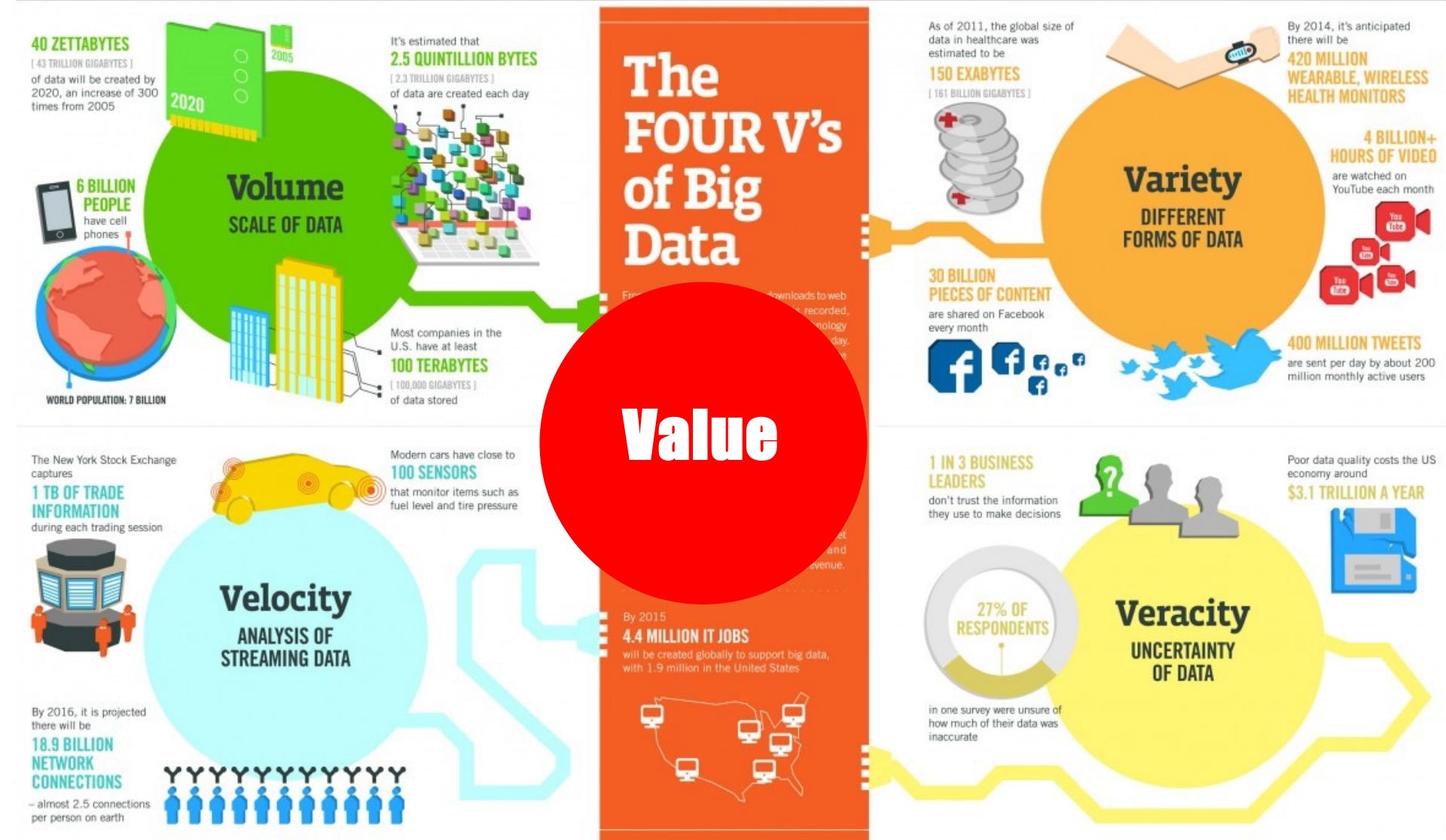
## 2020 *This Is What Happens In An Internet Minute*



## 2021 *This Is What Happens In An Internet Minute*



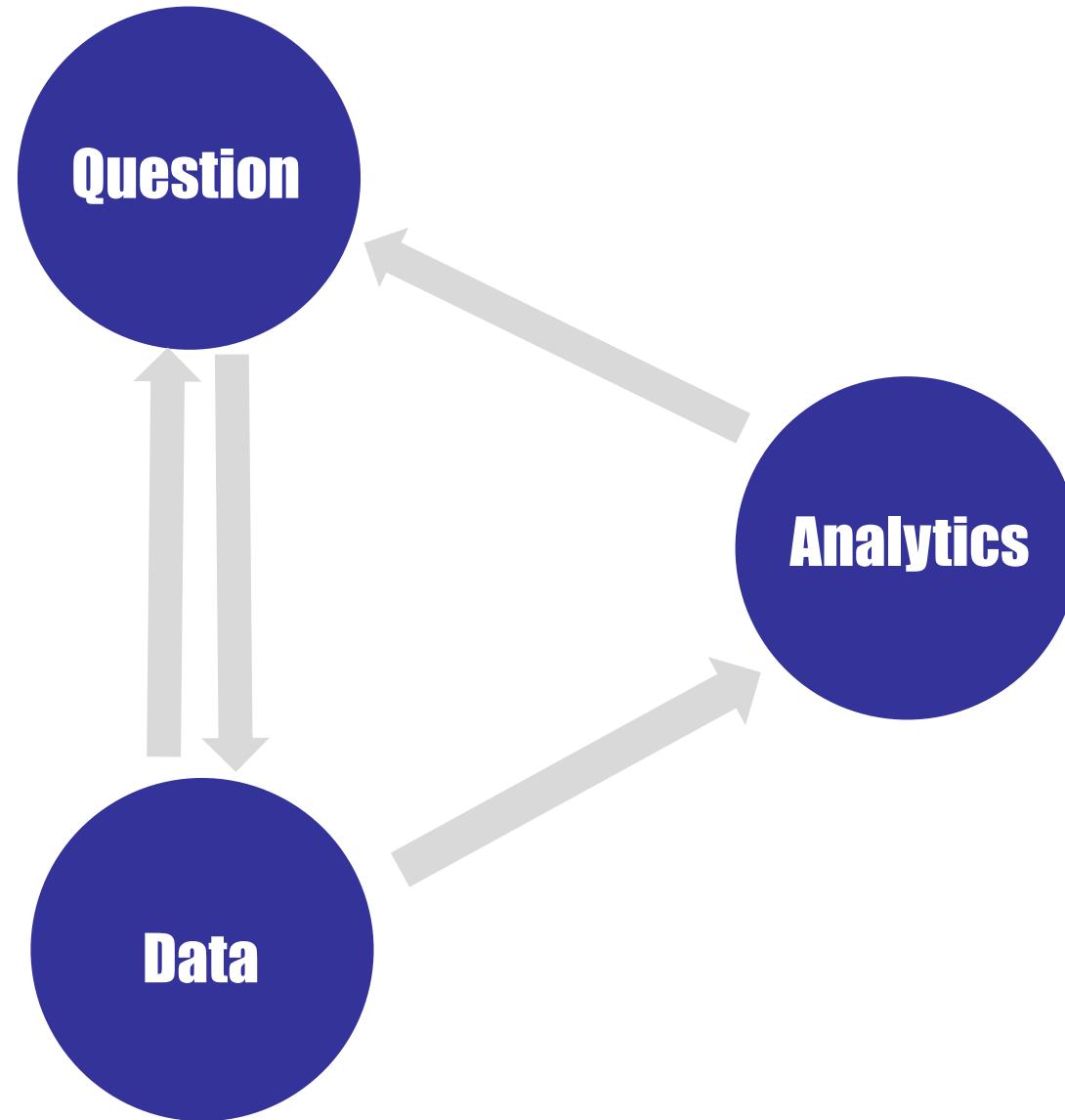
# What is “Big Data”?



- Turn data into information/value.
  - Business managers need to make decisions.
  - They need to make the most informed decisions that they can, and generate value.
- Decision making under uncertainty.
  - Most of the decisions are based on guesses, rather than “facts.”
  - How to make the “best” guess possible as well as how to measure the accuracy of their guesses.

## McKinsey Analytics Triangle

---



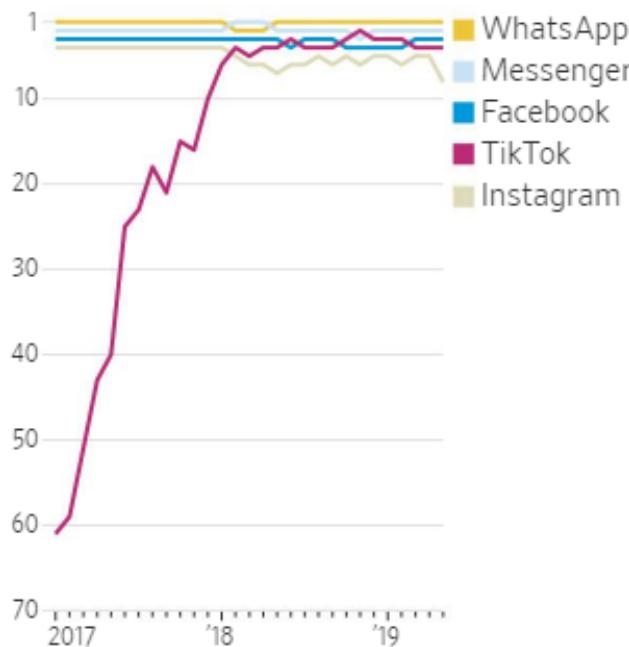
# ByteDance (2012 - )



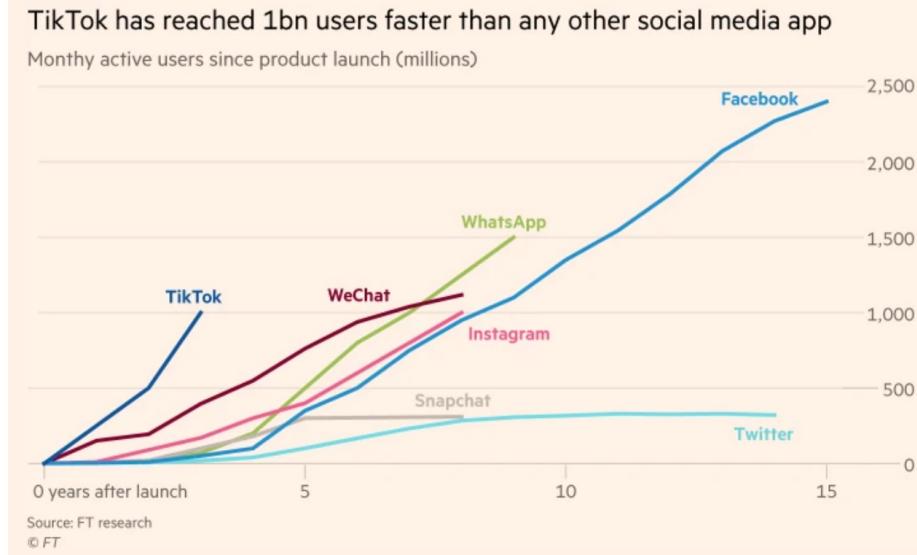
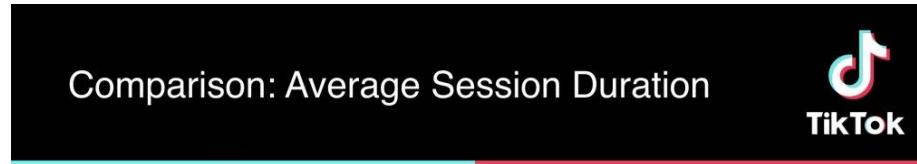
## Surprise Contender

TikTok has vaulted into the top app ranks, neck-and-neck with Facebook's apps.

### Monthly app ranking by global downloads

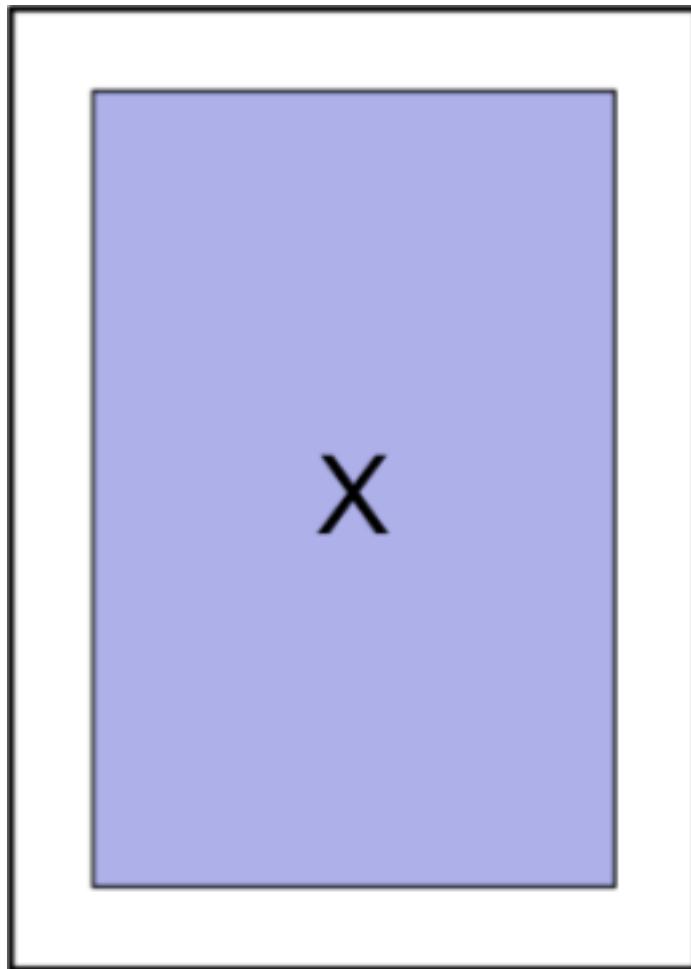


Note: TikTok downloads include Musical.ly; Facebook also owns WhatsApp, Messenger and Instagram  
Source: Sensor Tower

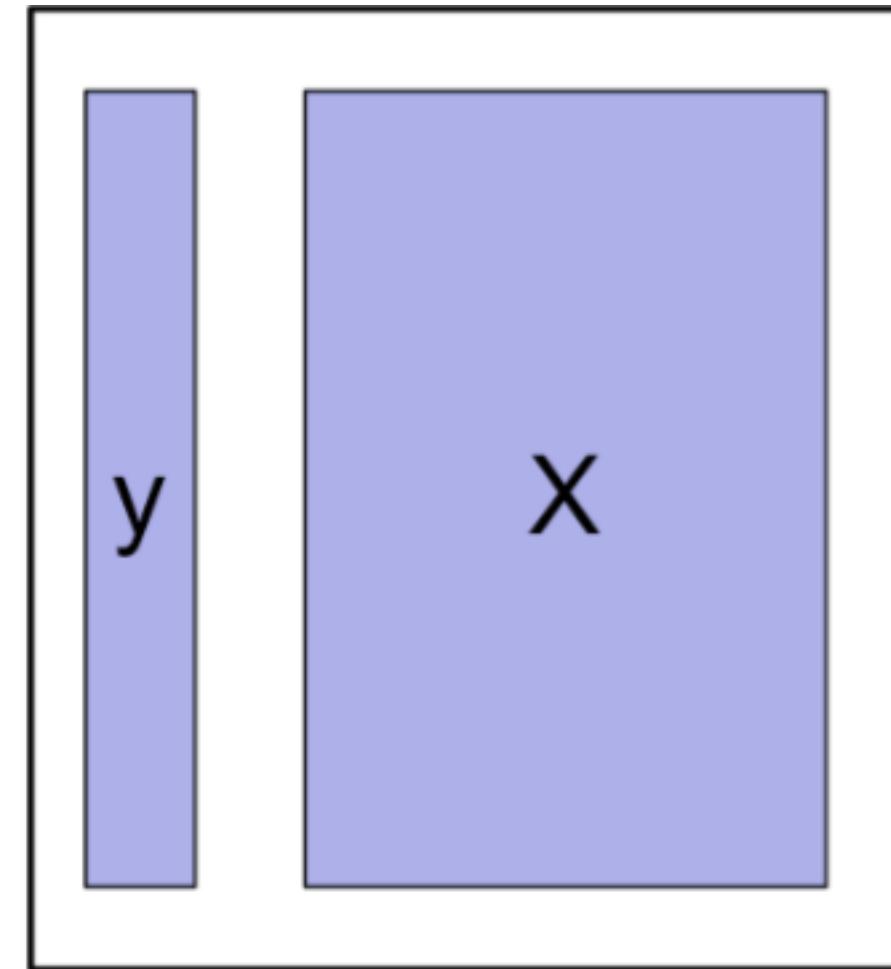


## Analytics Dichotomy

---



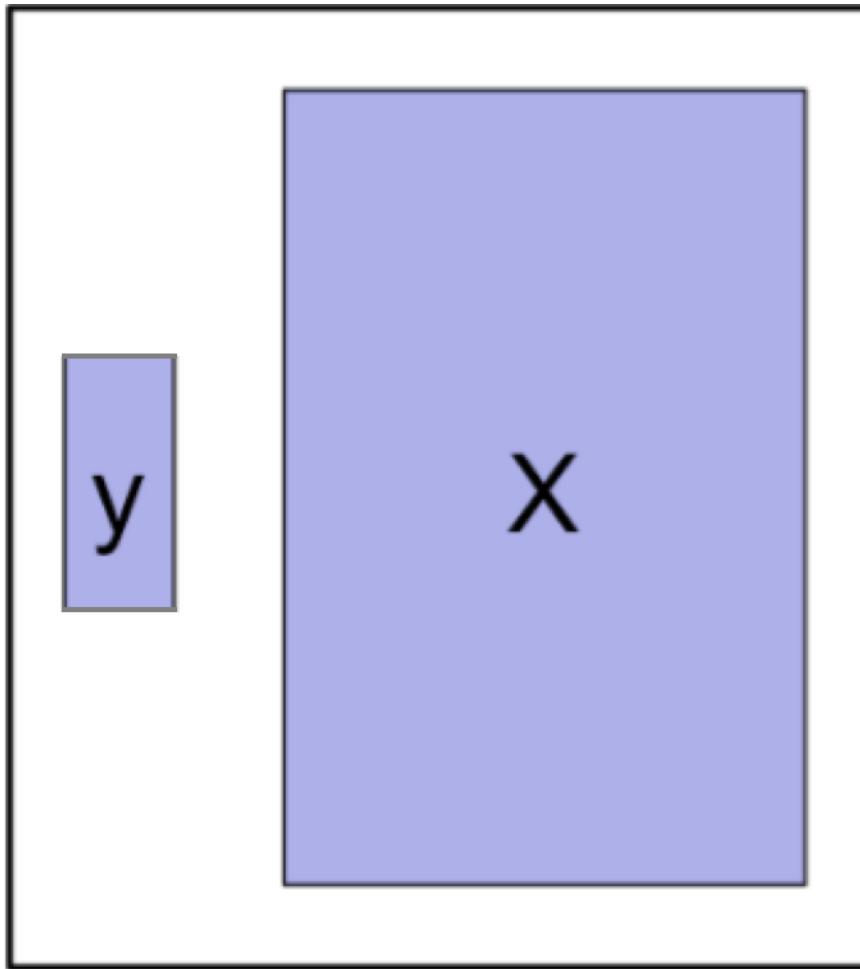
Unsupervised Learning



Supervised Learning

# Semi-Supervised Learning

---



Unlabeled data are cheap to get, and there are many of them!

# Identifying Eclipse

---



1.jpeg



2.jpeg



3.jpeg



4.jpeg



5.jpeg



6.jpeg



7.jpeg



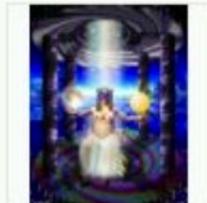
8.jpeg



9.jpeg



10.jpeg



11.jpeg



12.jpeg



13.jpeg



14.jpeg



15.jpeg



16.jpeg



17.jpeg



18.jpeg



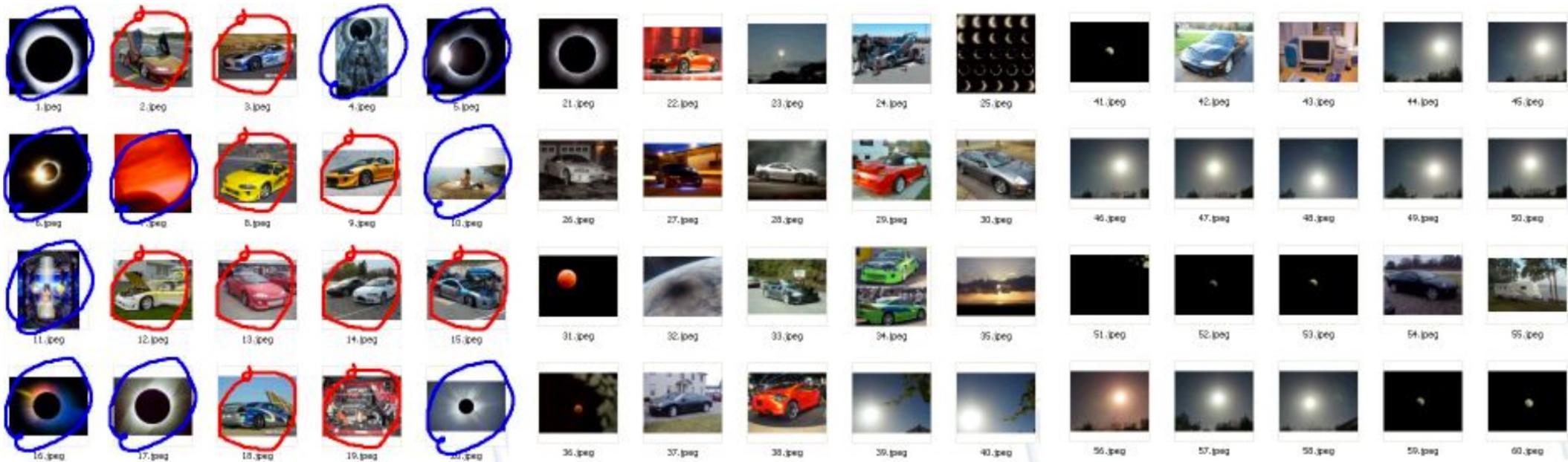
19.jpeg



20.jpeg

## Manually Labelled, Many Unlabeled ...

---



# Analytics Dichotomy

---

- **Unsupervised Learning (Pattern Discovery)**
  - Clustering
  - Dimension reduction
- **Supervised Learning (Predictive Modeling)**
  - Linear regression
  - Logistic regression
  - Splines/Local polynomial
  - Decision tree
  - Support vector machine
  - Neural network
  - Deep learning
- **Semi-supervised Learning**

# Customer Contact Center Workforce Management

---



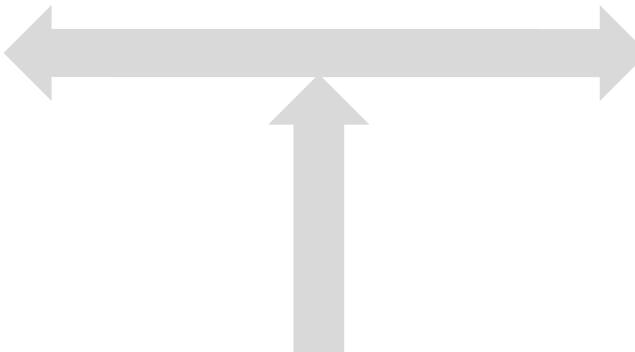
## Problem



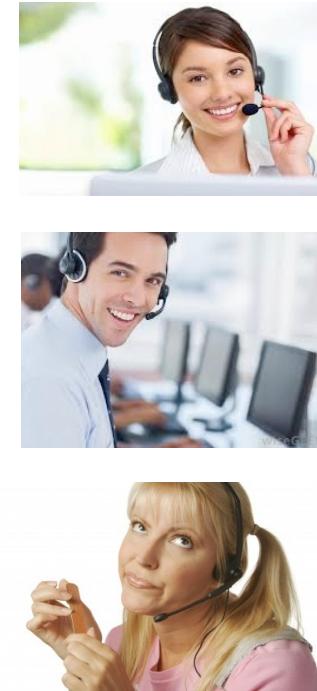
# Workforce Management: Optimization

Schedule the **right agent** at the **right time**  
to serve the **right customer** at the **right level**  
with the **minimum cost**

Demand



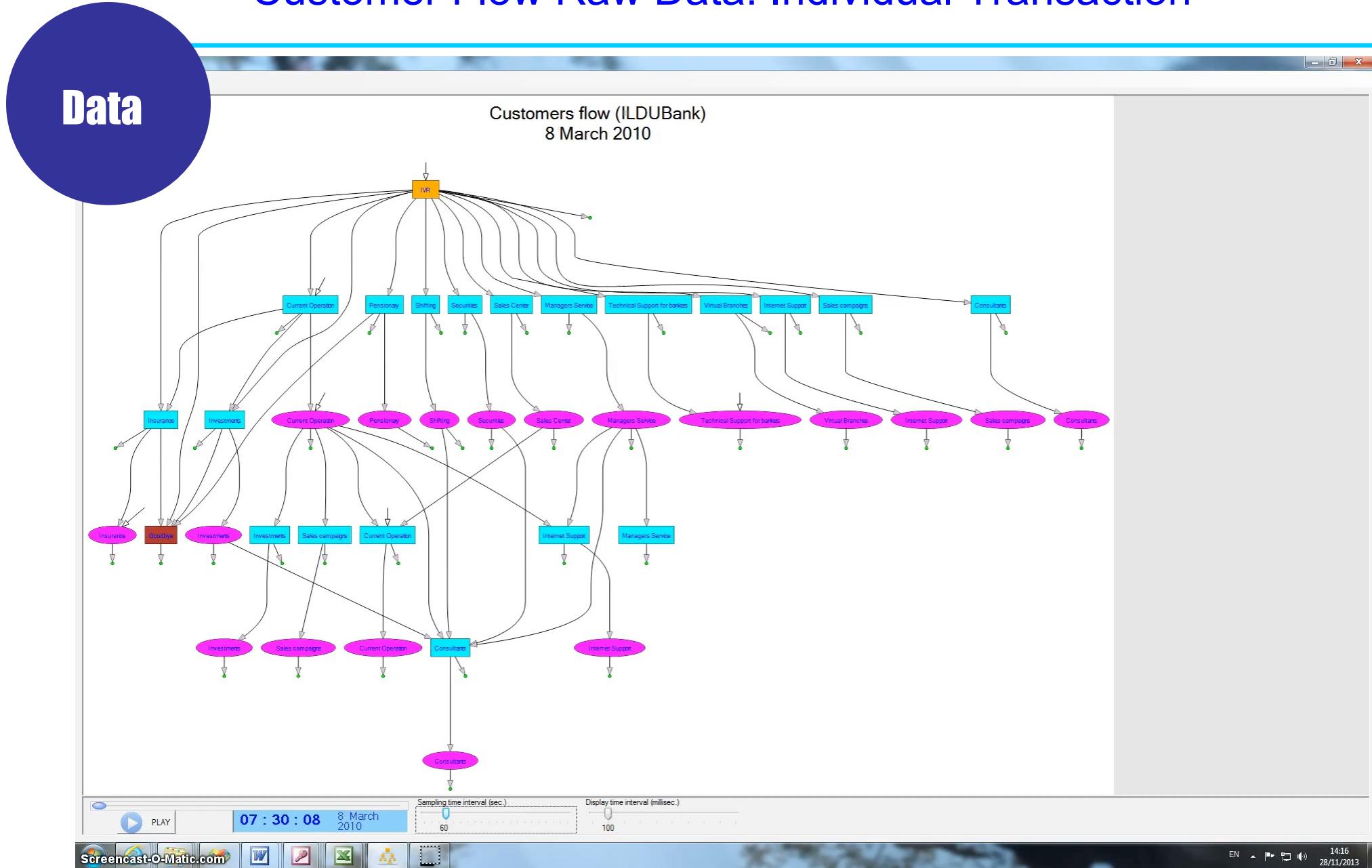
Supply



Quality of Service

Waiting time, abandonment,  
first call resolution, ...

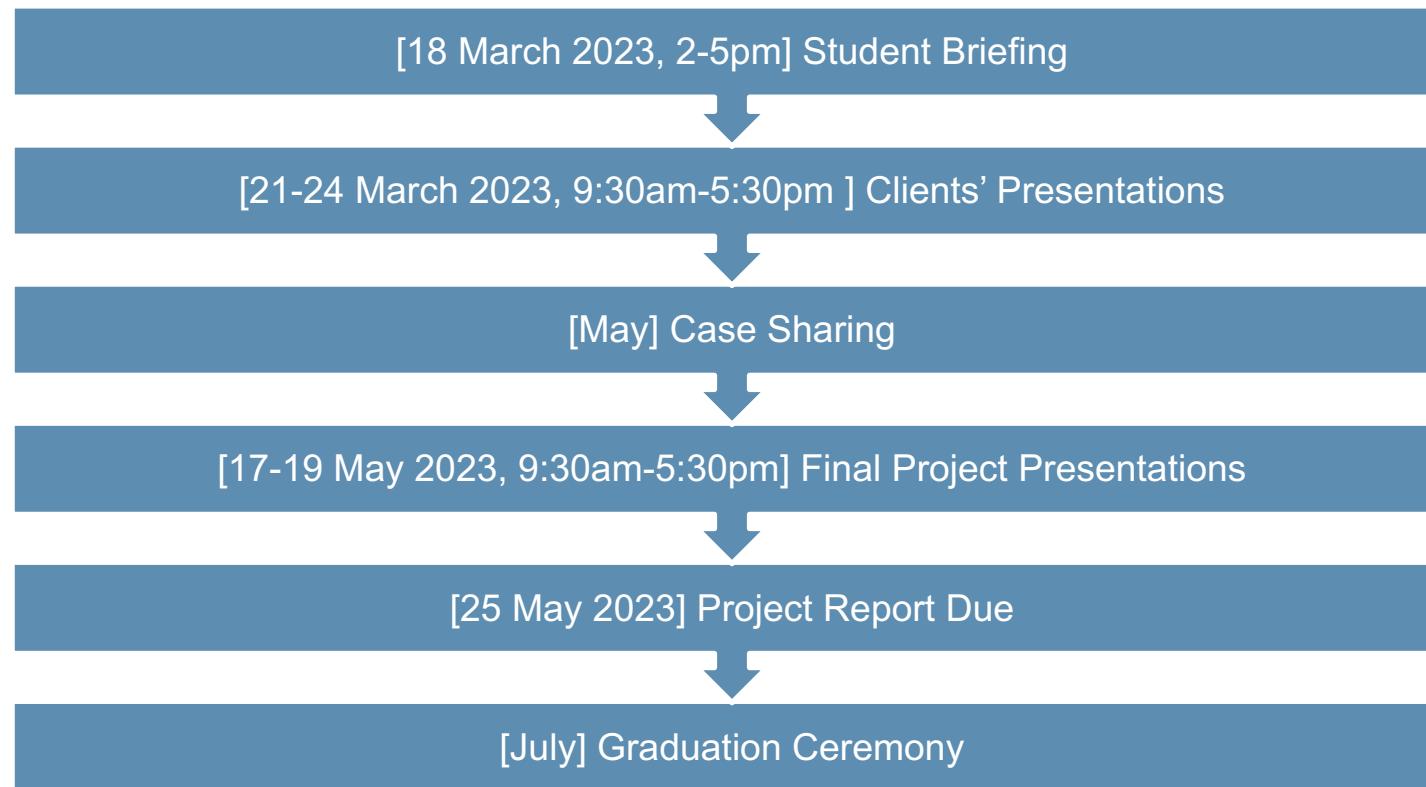
# Customer Flow Raw Data: Individual Transaction



# Capstone Project Course

---

- 8 students in each team, 1 faculty advisor, industry liaison



## 2017-2018 Capstone Projects

---

- Chow Tai Fook
- Esquel
- SCMP
- AS Watson
- CK Hutchison
- Duty Free Store
- HK Jockey Club
- EMSD
- Logitech
- HKMA

## 2018-2019 Capstone Projects

---

- Chow Tai Fook
- Esquel
- SCMP
- AS Watson
- MARUBI
- Sino Group
- 8 Securities
- Accenture Brazil
- Tencent
- HSBC
- HKEX

# 2019-2020 Capstone Projects

---

- Tencent
- SF Express
- Tomorrow Advancing Life
- Hang Seng Bank
- Town Gas
- CLP
- Esquel
- Hilti
- Serai
- Informa Markets
- Snapask
- Talkpush
- Lalamove
- HK Airport Authority
- Yoho Hong Kong
- HSBC
- Lane Crawford Joyce Group
- HKEX

# 2020-2021 Capstone Projects

---

- iFlytek
- Lalamove
- SF Express
- Tencent
- Tomorrow Advancing Life
- Duty Free Store
- Esquel
- Serai
- yuu Rewards (Dairy Farm)
- Aqumon
- Reap
- Y2 Capital
- CLP
- HK Airport Authority
- Mapxus
- Town Gas
- Hang Seng Bank
- HKEX
- HSBC

# 2021-2022 Capstone Projects

---

- iFlytek
- Lalamove
- SF Express
- Tencent
- Tomorrow Advancing Life
- Duty Free Store
- Esquel
- Serai
- yuu Rewards (Dairy Farm)
- Aqumon
- Reap
- Y2 Capital
- CLP
- HK Airport Authority
- Mapxus
- Town Gas
- Hang Seng Bank
- HKEX
- HSBC

# 2021-2022 Capstone Projects

---

## Tech

- Microsoft
- CoPilot AI (Canada)
- Tian An Yun Gu 天安云谷
- Mapxus

## E-commerce

- DH Gate 敦煌网
- Yoho
- Serai

## Social

- Newsb2b2c
- Yilan Tech 一览

## Retail/Marketing

- Coco Cola
- Dah Chong Hong
- Duty Free Store
- Dairy Farm
- Yili 伊利
- Lane Crawford
- Tech Trans
- Tom Group (Taiwan)

## FinTech

- Ping An
- Agumon
- BCW Group
- Reap

## Finance

- Hang Seng Bank
- HSBC
- FWD Insurance
- Beijing Prism 棱镜私募
- Huaxi Shenzhe 华西证券
- Haitong 海通国际证券

## Healthcare

- Roche
- Merit Medical

## Misc

- HK Jockey Club
- Raipot Intelligent Manufacturing
- Huahong Motors

# Data Analytics with EMSD

---





## Shocking Facts :

70%

### Electricity Consumption

Accounted for 70% of Total Green House Gas emission.

\$700M

### Electricity Waste

It cost more than \$700 million w per year due to the electricity waste.

29% for Air Conditioning!!!!

## The Team

---



# Our Recommendation

---

Electricity Consumption v.s. Total Cooling Demand



## Company Background

### ● Sino Parking Services Limited

- One of the largest and most experienced car park operators in Hong Kong
- Managing more than **110** car parks and over **17,000** parking spaces
- Implemented automated access control system - **iPass**
- Established parking loyalty programme, and regular customer reward promotions



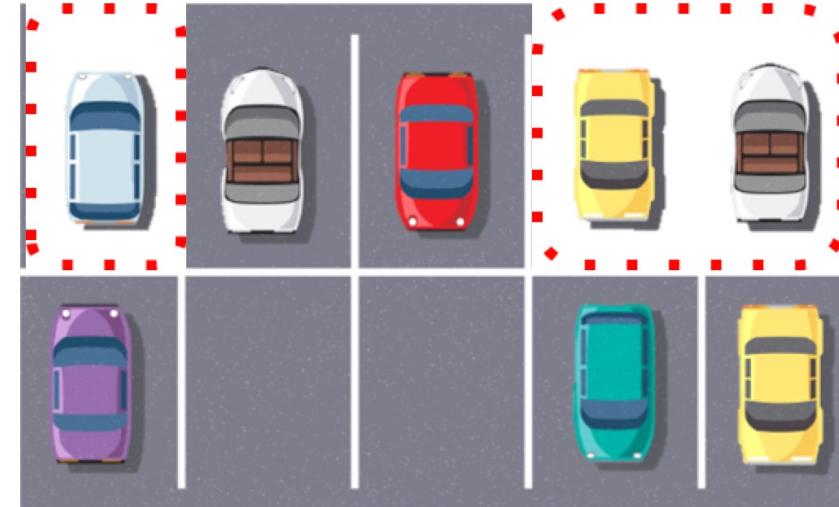
## Business Problem

- **Goal**

Increase overall revenue of Sino car parks

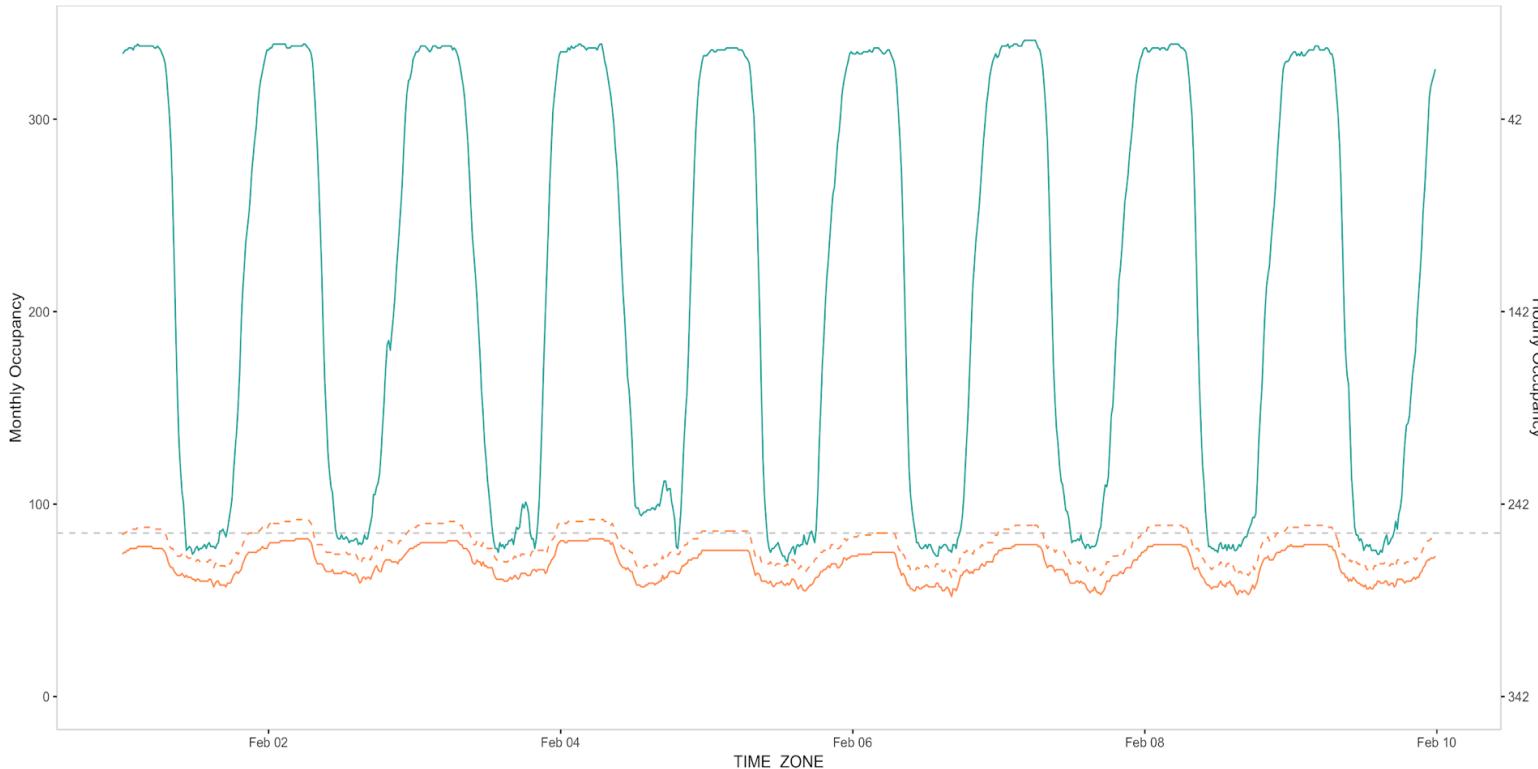
- **Objective**

Improve the utilization rate of car parks  
by releasing more floating monthly  
parking spaces for hourly parking users



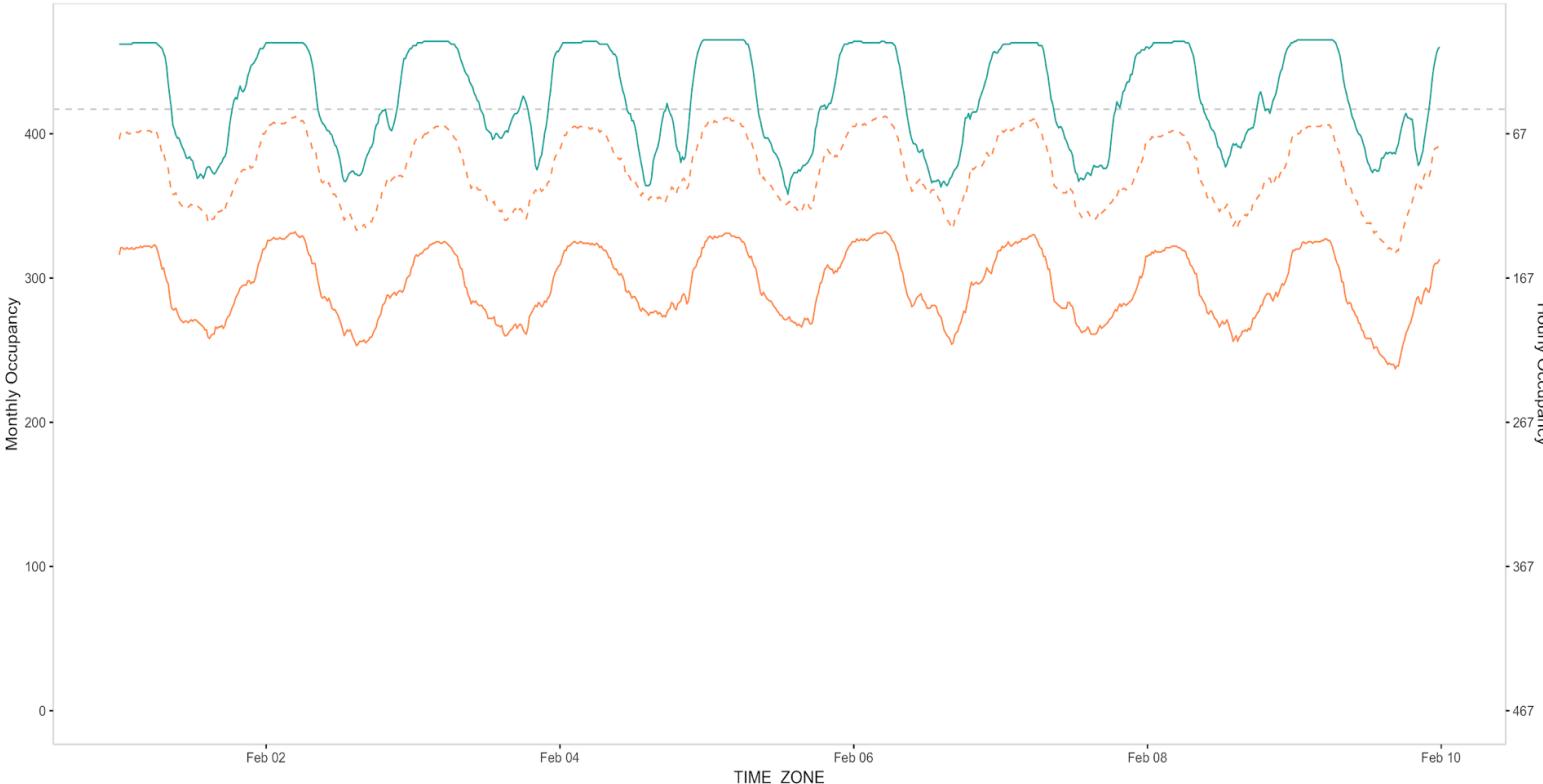
# Sino Parking Best Utilization

-  
*the site with best utilization: OC*  
**Current Methodology**



# Sino Parking Worst Utilization

the site with worst utilization: PW  
**Current Methodology**

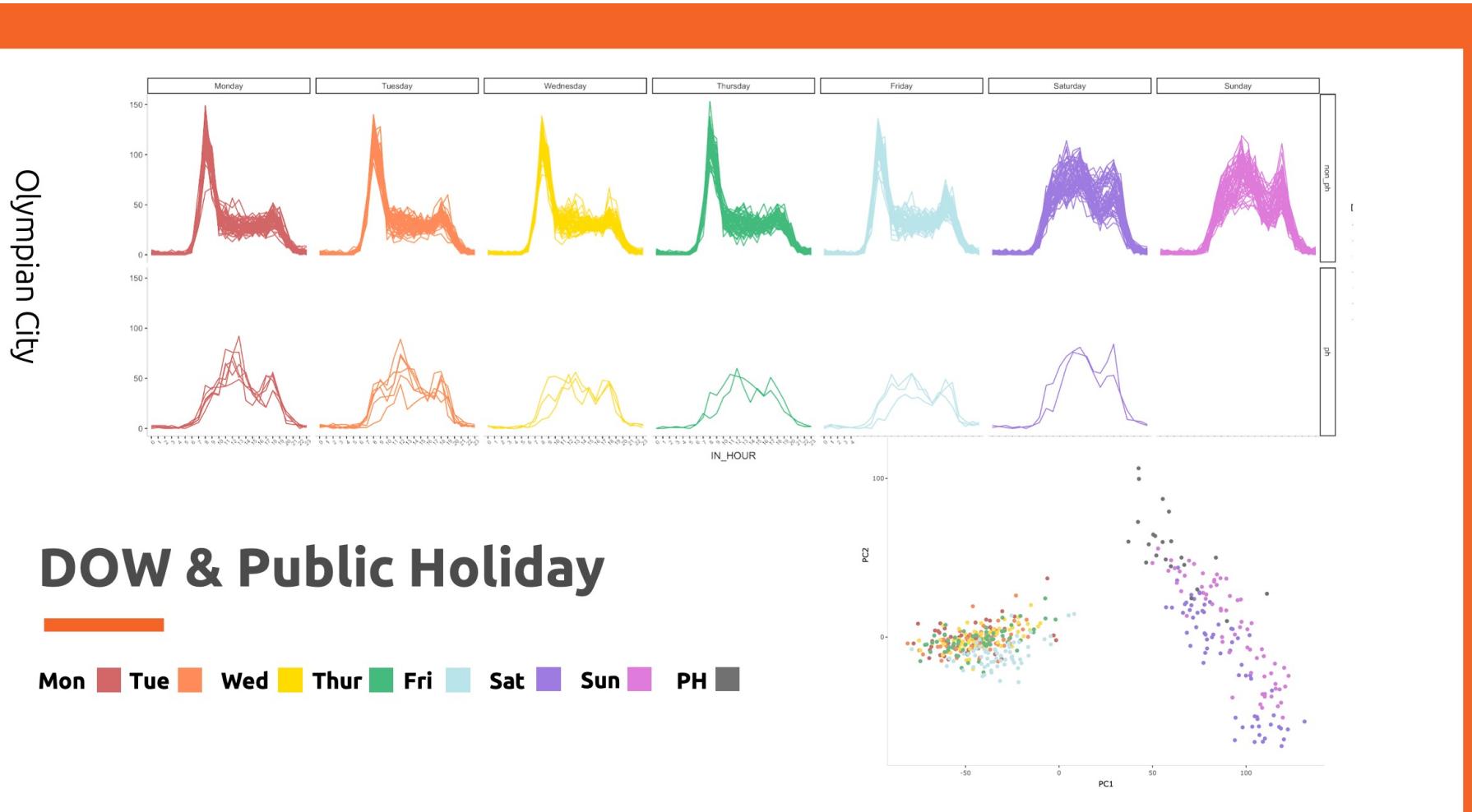


# The Team

---



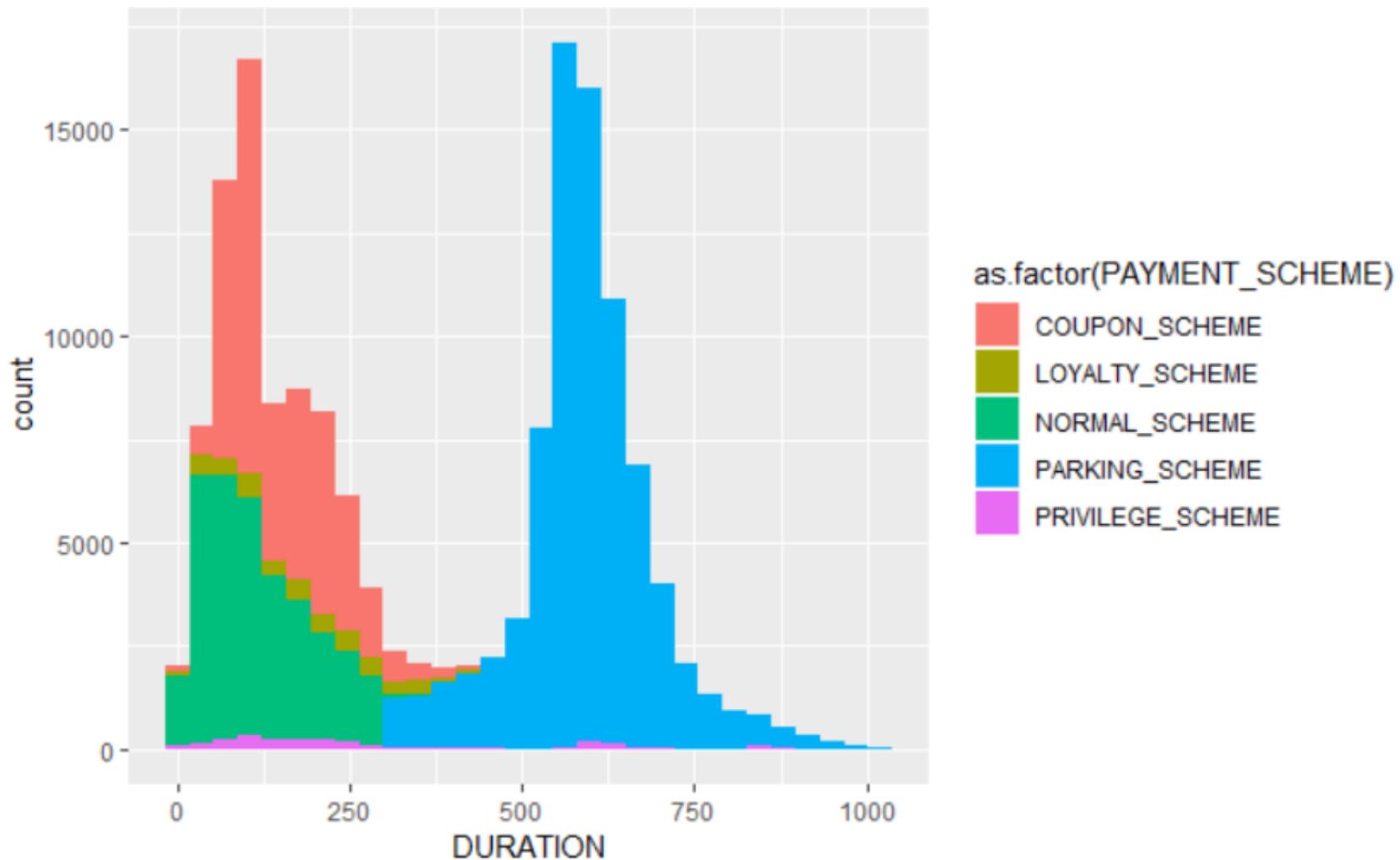
# Hourly User Arrival Pattern



## DOW & Public Holiday

# Hourly User Parking Duration

---

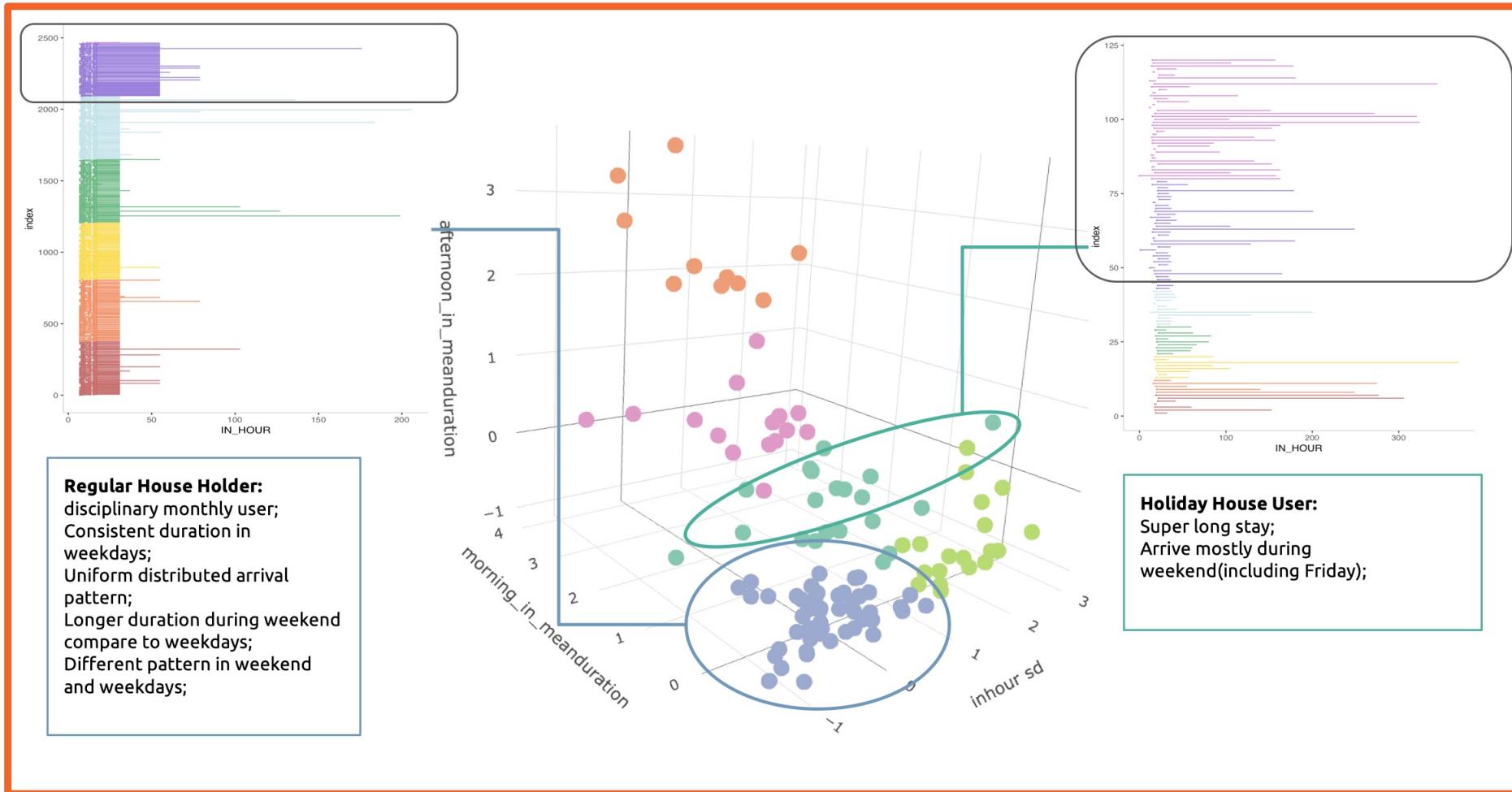


# Monthly Pass Holder Usage Pattern



Mon ■ Tue ■ Wed ■ Thur ■ Fri ■ Sat ■ Sun ■

# Monthly Pass Holder Clustering



Mon ■ Tue ■ Wed ■ Thur ■ Fri ■ Sat ■ Sun ■

# Reinforcement Learning

