



ALVIS KWOK

NAOMI TSANG

JACK LAU



## UNIQLO SIZE PREDICTOR

Recommendation system project  
Available online on 28<sup>TH</sup> Mar

[VIEW MORE](#)





ALVIS KWOK

NAOMI TSANG

JACK LAU



## Agenda

- Background
- Business values
- Data collection and processing
- EDA
- Recommendation model
- Demo
- Limitations
- Conclusions

[VIEW MORE](#)





Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions

## What are we doing?

- Size predictor with gender, age, height and weight
- Size recommendation based on **customers' reviews**
- **Related products** recommendation

## Why are we doing?

- **Solve the sizing hassle online and offline**
  - COVID-19 > ecommerce chance > **sizing uncertainty**
  - Special size exclusive online **without fitting** offline
  - Customers could **hardly look for advice** in-store



Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions



### On/Off Customers

- Better choice made from other customers' reviews
- Lower return/exchange risk  
eg. delivery cost and time



### Potential Customers

- Encourage e-commerce
- Lower out-of-stock risk



### Store Staff

- Reduce workload
- Less return and exchange operations



### Uniqlo

- Lower logistics cost from return/exchange
- Increase online sales
- More accurate merchandizing assortment, less chance loss and overstock of unsold sizes
- Better profit



Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions

## WEB SCRAPING - Uniqlo Japan

The screenshot shows a product page for a Uniqlo item. At the top, there's a navigation bar with the Uniqlo logo, search bar, and cart icon. Below the header, there's a section titled "この商品について" (About this product) with tabs for "概要" (Summary), "商品詳細" (Product details), and "再販定" (Restock). A blue arrow points to the "★★★★★" rating. The main content area displays two reviews:

**Review 1:**  
スリットがあるのが秀逸  
★★★★★  
購入サイズ: L  
お客様の着用感: ちょうどよい  
エルサイズを購入。175g、75kgの中年体型でもスリットのおかげでお腹が目立たない。  
まさゆき・男性・40代・身長: 171 - 175cm・体重: 71 - 75kg・足のサイズ: 26.5cm・東京都  
2022/3/23

**Review 2:**  
厚みがあってカラーも素敵です。  
★★★★★  
購入サイズ: XL  
お客様の着用感: ちょうどよい  
レディースのワッフルTと一緒に購入しました。メンズのワッフルTは厚手で暖かくレディースのワッフルTは軽く柔らかです。その日の気温に合わせて着る事が出来て大活躍しています。  
ママリン・女性・50代・岩手県  
2022/3/23

### Data collection:

- Top 100 items
- Customer info and rating
- Reviews ceiling at 1,000

### Challenges:

- Scraping time
- URL issue
- Seasonal change



## Background

## WEB SCRAPING - Uniqlo Japan

## Business values

## Data Collection and processing

## EDA

## Recommendation model

## Demo

## Limitations

## Conclusions

			id	gender	AgeGroup	Height	Weight	Foot_Size	Product_Size	Rating	Item_Code
0	439847_1	男性	Nan	身長: 121 - 130cm		Nan		Nan	購入サイズ: 130	4	439847
1	439847_2	女性	10 - 14歳		Nan		Nan	Nan	購入サイズ: 150	4	439847
2	439847_3	男性	Nan	身長: 151 - 155cm		Nan		Nan	購入サイズ: 160	4	439847
3	439847_4	女性	Nan		Nan		Nan	Nan	購入サイズ: 130	5	439847
4	439847_5	男性	7 - 9歳	身長: 111 - 120cm	体重: 16 - 20kg	足のサイズ: 21.5cm以下		購入サイズ: 120		5	439847
...	...	...	...	...	...	...	...	...	...	...	...
47214	445086_81	女性	40代	身長: 151 - 155cm		Nan	足のサイズ: 23.5cm	購入サイズ: 23.5		4	445086
47215	445086_82	男性	40代	身長: 166 - 170cm	体重: 66 - 70kg		足のサイズ: 25.5cm	購入サイズ: 25.5		4	445086
47216	445086_83	男性	20代	身長: 166 - 170cm	体重: 61 - 65kg		足のサイズ: 27.5cm	購入サイズ: 27.5		5	445086
47217	445086_84	女性	Nan		Nan		足のサイズ: 25.0cm	購入サイズ: 25		3	445086
47218	445086_85	男性	20代	身長: 161 - 165cm		Nan		Nan	購入サイズ: 25.5	5	445086
47219 rows x 13 columns											

Total Data: 47219 x 13



Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions

## Data Cleaning

### Language

- Japanese > English

### Data Error:

- Age - Height - Weight

### Bra and Foot size

- Different measurement

### Drop NaN

- Data dropped from **47219** to **33712**



---

Background

---

Business values

---

Data Collection and processing

---

EDA

---

Recommendation model

---

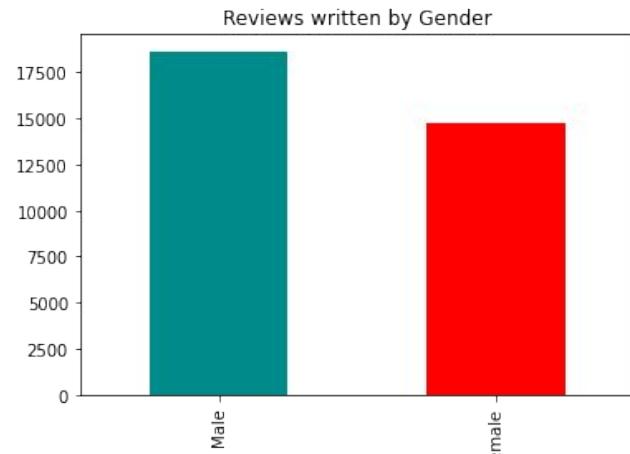
Demo

---

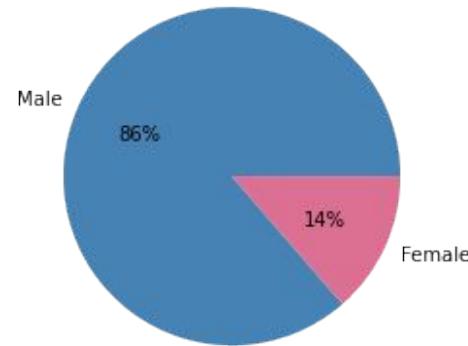
Limitations

---

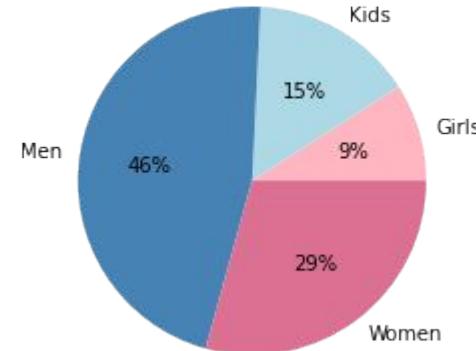
Conclusions



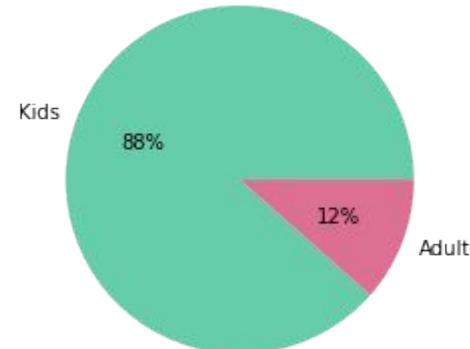
Men's items reviewed by gender



Items chosen by Gender



Kids' items worn by adults



## Background

## Business values

## Data Collection and processing

## EDA

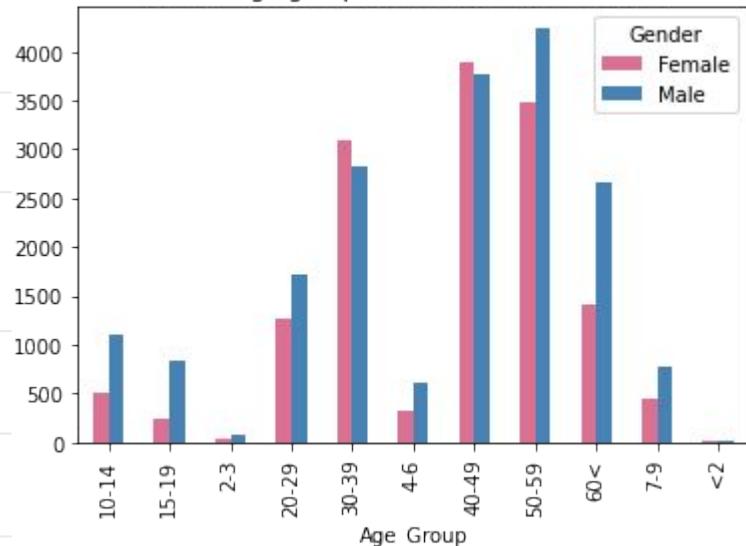
## Recommendation model

## Demo

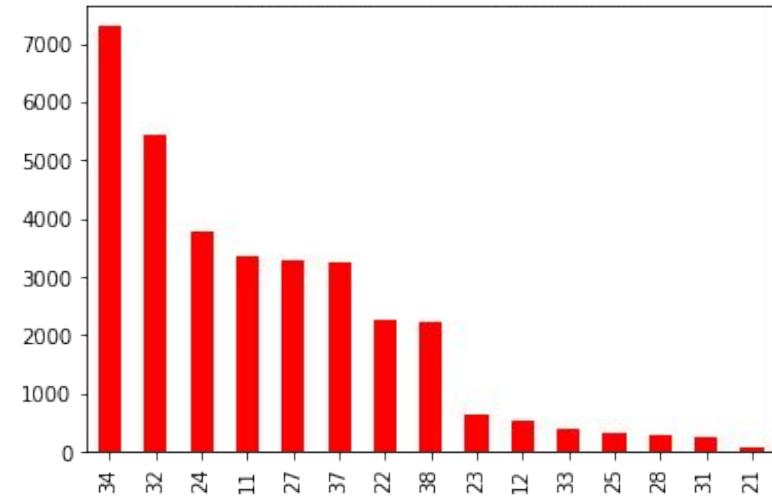
## Limitations

## Conclusions

Which age group leaves the most reviews?



Which dept has the most reviews?



-11: K's item

-12: G's item

-21: W's outer

-22: W's bottoms

-23: W's shirts and blouse

-24: W's cut and sewn

-25: W's knit

-27: W's inner

-28: W's roomwear

-31: M's outer

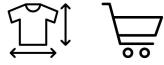
-32: M's bottoms

-33: M's shirts

-34: M's cut and sewn

-37: M's inner

-38: M's roomwear



Background

Business values

Data Collection and processing

## EDA

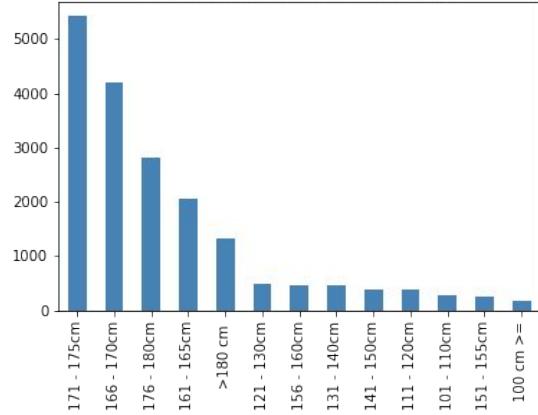
Recommendation model

Demo

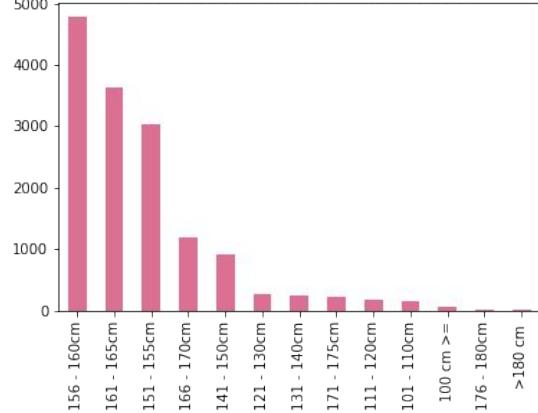
Limitations

Conclusions

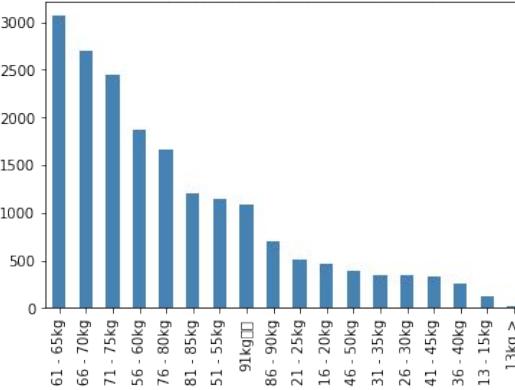
Height distribution by gender - Male



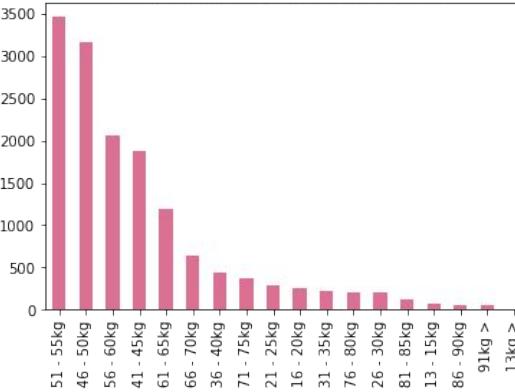
Height distribution by gender - Female



Weight distribution by gender - Male



Weight distribution by gender - Female



---

Background

---

Business values

---

Data Collection and processing

---

## EDA

---

Recommendation model

---

Demo

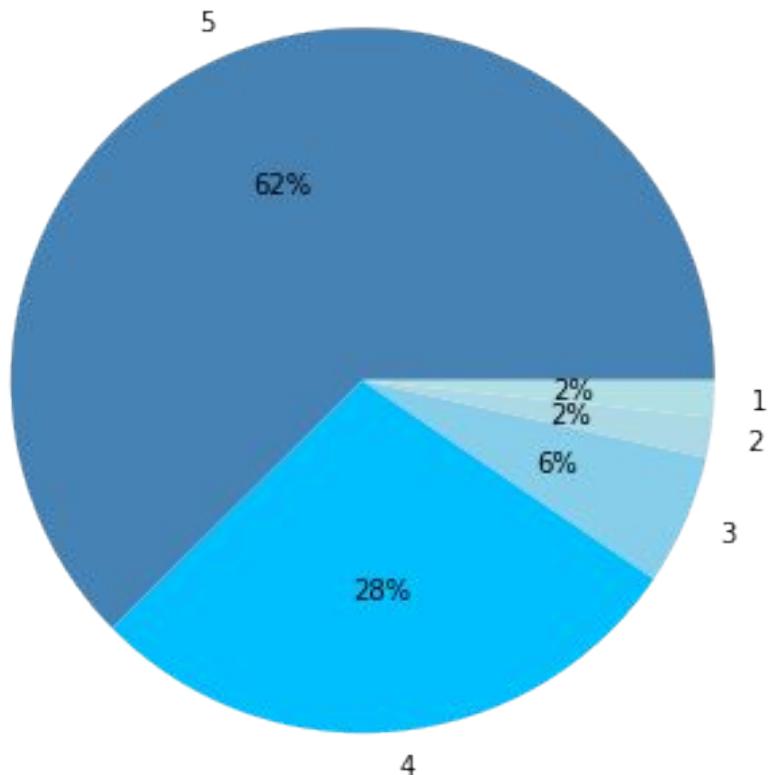
---

Limitations

---

Conclusions

### Reviews' Rating



### Rating:

- 62% 5-stars Reviews
- 90% "Good" Reviews



# EDA - TOP 3 BEST REVIEWS ITEMS - MALE



Background

3 Items

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

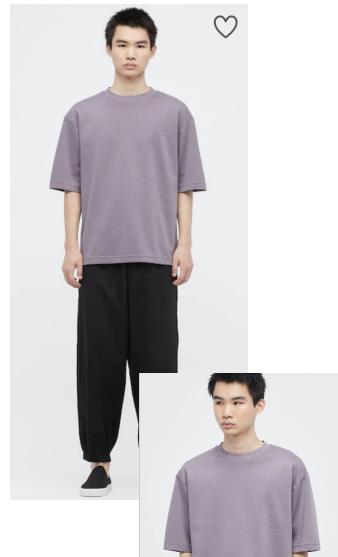
Conclusions



2nd - Denim Work Jacket (75.9%)



1st - Indigo Dye Sweat Cardigan (76.8%)



3rd - Airism Cotton Crew Neck  
Oversized T-shirt (75.5%)



# EDA - TOP 3 BEST REVIEWS ITEMS - FEMALE



Background

3 Items

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions



**2nd - Cotton Relaco Shorts (70.9%)**

**1st - Ribbed Cropped Short Sleeve T-shirt (71.7%)**

**3rd - Lace Shorts (hiphugger)(68.1%)**



# EDA - TOP 3 BEST REVIEWS ITEMS - BOYS and GIRLS



Background

3 Items

Business values

Data Collection and processing

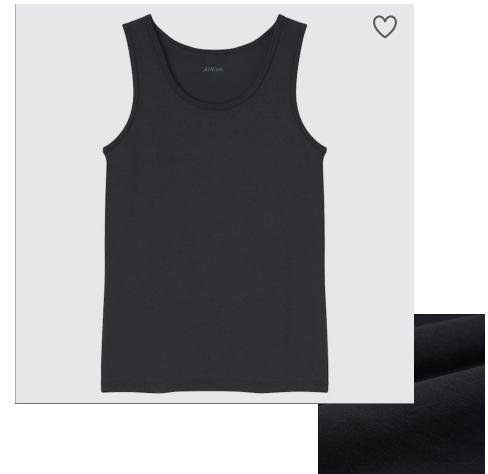
EDA

Recommendation model

Demo

Limitations

Conclusions



**2nd** - Airism Cotton Crew Neck T-shirt (74.8%)

**1st** - Ultra Stretch Dry Sweat Full Zip Hoodie (76.2%)

**3rd** - Airism Cotton Blend Tank Top(73.6%)



# EDA - TOP 5 WORST REVIEWS ITEMS



Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

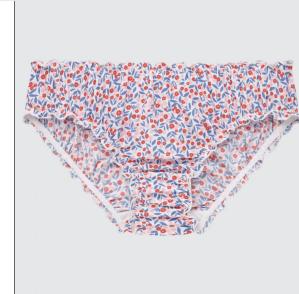
Limitations

Conclusions

5 Items



2nd - 14.5%



1st - 27.6%



3rd - 11.9%



4th - 7.9%



5th - 7.7%



---

Background

---

Business values

---

Data Collection and processing

---

EDA

---

Recommendation model

---

Demo

---

Limitations

---

Conclusions

## 1) Size Recommendation

Bra	According to UNIQLO standard size chart	
Shoes	Sneakers: By determining whether sizing is “TRUE TO SIZE”	Room Shoes: By finding cut-off line from foot size/purchased size density
All others	According to Top 10 nearest reviews by hamming/cosine distance	

## 2) Product Recommendation

According to Top 20 nearest reviews by hamming/cosine distance



# RECOMMENDATION MODEL



Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions

Bra

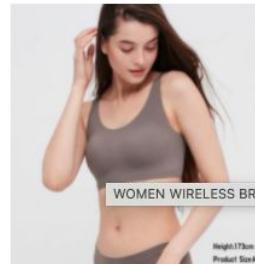


WOMEN

65/70 AA-85/90 DEF

WIRELESS BRA (3D HOLD) 438961

SIZE	UNDER BUST SIZE					
	65	70	75	80	85	90
CUP SIZE	AA	AA 65/70				
	A	ABC 65/70			ABC 85/90	
	B					
	C					
	D	DEF 65/70			DEF 85/90	
	E					
	F					



WOMEN

XS-XXL

WIRELESS BRA (RELAX) 445383

◆ Cup Size ◆	XS	S	M	L	XL	XXL
	AA65	A65	D65	E70	D80	E85
	AA70	B65	B70	C75	E80	F85
	-	C65	C70	D75	F80	B90
	-	A70	D70	E75	B85	C90
	-	-	A75	B80	C85	D90
	-	-	B75	C80	D85	-

(cm)



## RECOMMENDATION MODEL



## Background

## Business values

## Data Collection and processing

EDA

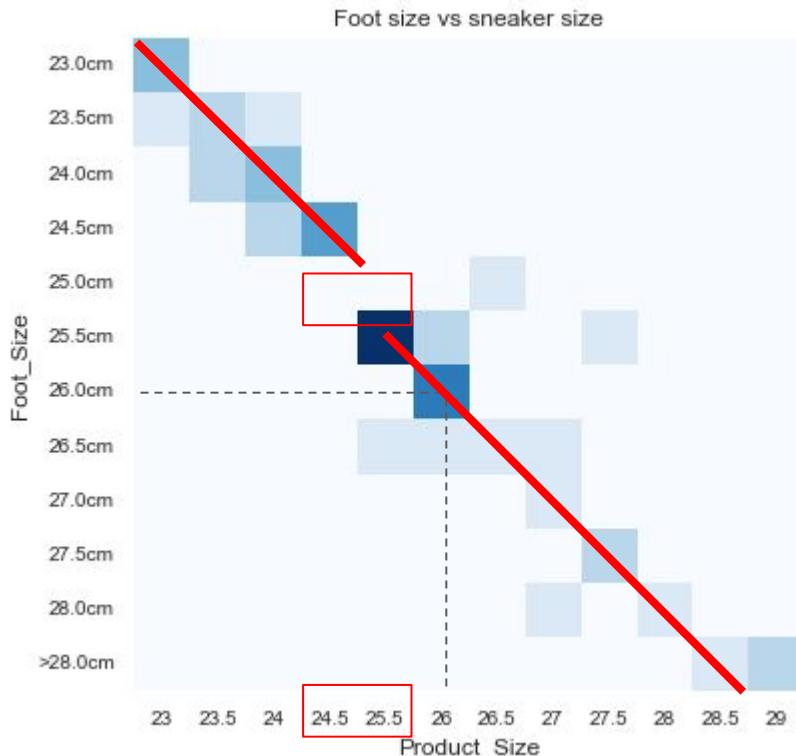
## Recommendation model

Demo

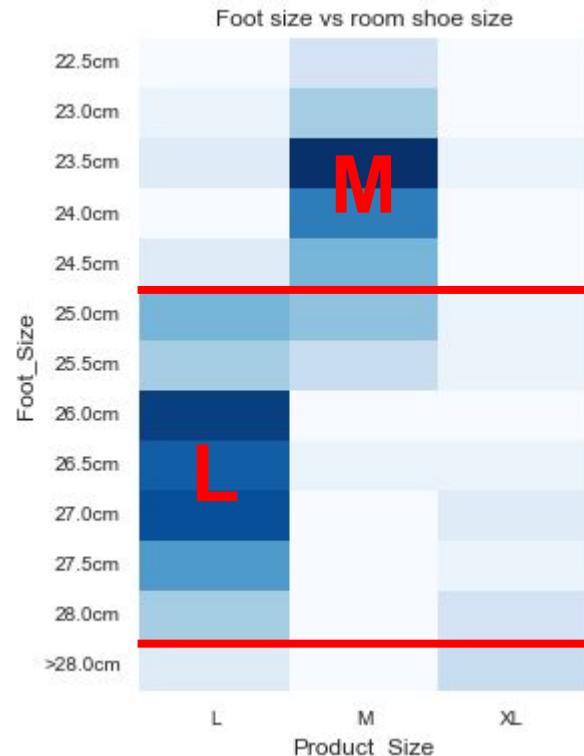
### Limitations

### Conclusions

## Sneakers: TRUE TO SIZE



Room shoes: M for <25, L for 25-28cm





DEMO



Background

---

Business values

---

Data Collection and  
processing

---

EDA

---

Recommendation  
model

---

Demo

---

Limitations

---

Conclusions

All others?

Check out the DEMO !



---

Background

---

Business values

---

Data Collection and processing

---

EDA

---

Recommendation model

---

Demo

---

Limitations

---

Conclusions

## Data

- Not full product coverage due to time constraint
- Many newly re-launched products without enough reviews  
eg. many have <10 reviews due to seasons change
- Some products' sizing is fragmented, thus reviews on each size is limited

## Size

- Different products different measures  
eg. pants: waist and abdominal shape
- Slow reaction to style and trend, such as tight and oversize style changing rapidly

## Stock

- Size available in Japan might not be available in HK
- Recommended size is not linked to stock on HK site





Background

Business values

Data Collection and processing

EDA

Recommendation model

Demo

Limitations

Conclusions

## Conclusions:

- Size predictor accuracy
- More items
- More reviews
- More dimensions

## Recommendations:

- Deployment in HK
- For Japan: review inputs dimensions  
eg. pants, bra
- Low rating items review
- Color strategy  
eg. Indigo collection is highly rated





ALVIS KWOK

NAOMI TSANG

JACK LAU



UV  
PROTECTION



THANK YOU

