

---

# INFO 201 FINAL PROJECT

**Dataset(s) used** 2

---

**Dataset cleaned by**  
Phoebe Romangsuriat  
Ching Yee (Melody) Fung

---

**Source(s)**



---

## ABOUT

this Dataset is of marketing sales collected in regard of how much impact fashion influencers have on the industry. Each row corresponds to an independent marketing promotion where the business uses TV, social media, radio, and influencer promotions to increase sales.

---

## USE CASES

Potential real-world usage of the dataset

- Is influencers a great marketing strategy?
  - How much impact does influencers has on business sales?
  - What industry benefit most from this marketing strategy?
- 

## 8 HUMAN RIGHTS PRINCIPLE

Potential real-world usage of the dataset



Human control  
of technology



Safety and  
security



Accountability

---

## Composition

1. A smaller dataset focusing on Instagram influencers is pulled out from the bigger datasets with different social media platforms. Dataset included:
  - a. Name of influencers
  - b. Followers
  - c. Engagement
  - d. Categories of influencers
2. Second dataset used is composited of social media promotion budget and its corresponding sales, and categorize the impact of influencers in terms of 'mega', 'macro', 'micro', 'nano'.

## Cleaning/labelling

1. In our "instagram\_influencerdf" dataset, we changed all string type under "Impact" to numerical value
2. We merged the two datasets
3. We got rid of all the rows with values of 'NA'
4. We created a summarization data frame of all grouped by the category of influencer (category 1) and their followers called "category\_followers"
5. We filtered our merged dataset to only "fashion" Influencers to shrink it due to our wanted focus on only one type of industry influencer.

---

## Alert(s)

1. Huge dataset
  - a. The combined dataset is too large and thus leading to only able to focus on one specific industry.
2. Inaccuracy in comparison/conclusion
  - a. The cutdown in dataset might result in inaccuracy when comparing the impact of influencers due to limited data.
3. Collection method
  - a. method of collecting engagement/followers with influencers is unclear and might correlate to potential 'ghost' account that does not truly reflect interaction between real consumers and influencers' impact.