

# Predicting Consumer Tastes with Big Data at Gap

**MGMT 590 – Web Data Analytics**

**Group 8**

Seonkyu Kim

Karan Karnik

Rahul Kunku

Shubhankar Sharma

# GAP Inc. as a company

CEO: Art Peck



Fast Fashion



Mass specialty

BANANA REPUBLIC

High-end specialty



Fitness apparel

# Way forward for GAP Inc.

- CEO of GAP Inc. is looking to replace the Creative Directors with a big-data driven creative process
- The company hopes that this approach can perform well in predicting the future fashion trends and reduce lead times in bringing new designs to the market
- If and how art or science should lead the company's fashion development approach

# Is Big Data approach the right way?

We believe So.

- Data analytics can provide valuable insights for informed pricing, marketing, and development decisions
- Increased adaptability to changing market dynamics and customer demands
- Identification of new market opportunities
- Recognizing consumer trends and preferences

**Web data to  
support the  
data-driven  
creative  
process at  
GAP**

**Consumer Sentiment Analysis**

**Competitor Analysis**

**Fashion Trends and Influencer  
Data**

**Demographic and Geographic  
Data**

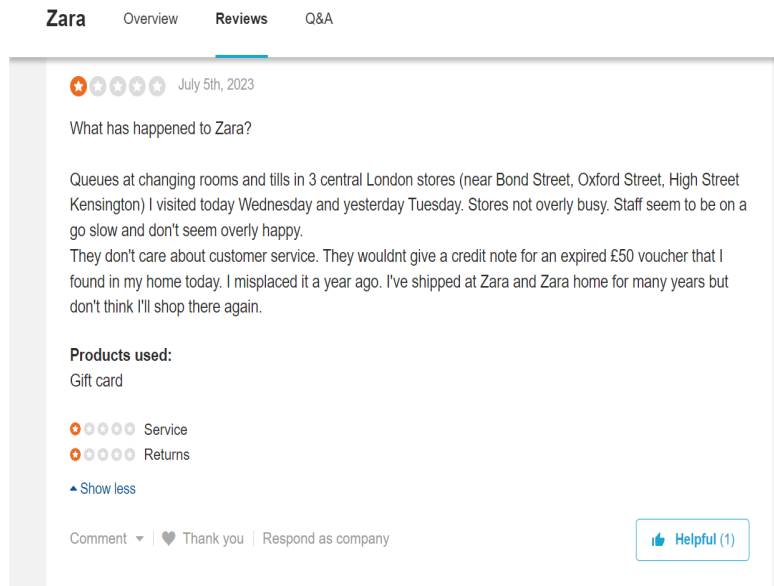
**Economic Indicators**

# Consumer Sentiment Analysis Examples

- **Sentiment Analysis:** Use Natural Language Processing (NLP) tools to analyze sentiment. For example, Tweets 1 and 3 are positive, while Tweets 2 and 4 are negative.
- **Influencer Mentions:** Identify mentions by fashion influencers. If influential figures are praising Gap products, it can be leveraged in marketing campaigns.
- **Competitor Comparison:** Compare sentiment data with competitors to see how Gap stacks up in terms of customer sentiment.



# Online Review



# Data Extraction

First time ordering online, will never do again. Attempted to speak to a representative three times. Quality is not that great. I ordered a bunch of cotton t-shirts that showed up so paper thin. 2 weeks later I never buy anything from Zara online, this is what you'll get. These came all bent up and chipped. It has got to be one of the worst customer service experiences I have ever experienced, no bonus. Purchased an expensive top in the Zara sale. Wore once, washed at 30°C as per care instructions. I placed an order online and paid for shipping, my package said delivered but I never arrived. What has happened to Zara? Queues at changing rooms and tills in 3 central London stores (near Bond Street, Oxford Street, High Street Kensington) I visited today Wednesday and yesterday Tuesday. Stores not overly busy. Staff seem to be on a go slow and don't seem overly happy.

I purchased a pair of pants that were returned within the 10 day return window. Zara received the pants. I purchased a swimsuit at Zara's Disney Springs location in Orlando on 5/27/23. After wearing it once it was ruined. Well, it's a pretty good mass market store. If you need to buy something for your outfit and if you're on a budget, Zara is a good option. I bought a pair of \$48 slacks at full price. The gold buttons on the pockets fell off after wearing them once. If you are going to shop at Zara, do it in person. Every online order I've placed has encountered some issue. Zara would not take two of my online returns because they said I did not follow the return policy. Zara Dress Shop exceeded my expectations in terms of both product quality and customer service. Zara's clothes are not great. They're often poor quality, but hey...you get what you pay for. I've had a few good items. WARNING\*\*\*If you need to make a return don't expect it to go quickly. The reality is they don't care about customer service. My daughter bought a dress online, it came with a stain on it. The same day she received it she returned it. I purchased a dress from Zara online. I returned the item as I found it too small. I did open the package and it was already stained.

# Review Sentiment

8	negative
9	negative
10	negative
11	positive
12	negative
13	negative

Sample sentiment analysis that can be used to improve customer satisfaction, identify areas for improvement, and potentially highlight positive aspects in their marketing campaigns.

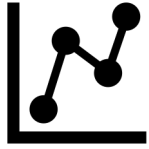


# Google Trends Analysis

- 1. Search Interest Over Time:** It shows the popularity of a search term over a specified period, allowing you to see how interest has changed over time.
- 2. Regional Interest:** It highlights the geographic regions where a search term is most popular, providing insights into regional trends and preferences.
- 3. Related Queries:** It displays related search terms and topics, helping you discover related keywords and understand search intent.
- 4. Trending Searches:** It showcases currently trending search queries, giving you real-time insights into what's currently popular on Google.
- 5. Comparison of Multiple Terms:** You can compare the search interest of multiple terms to see how they stack up against each other in terms of popularity.

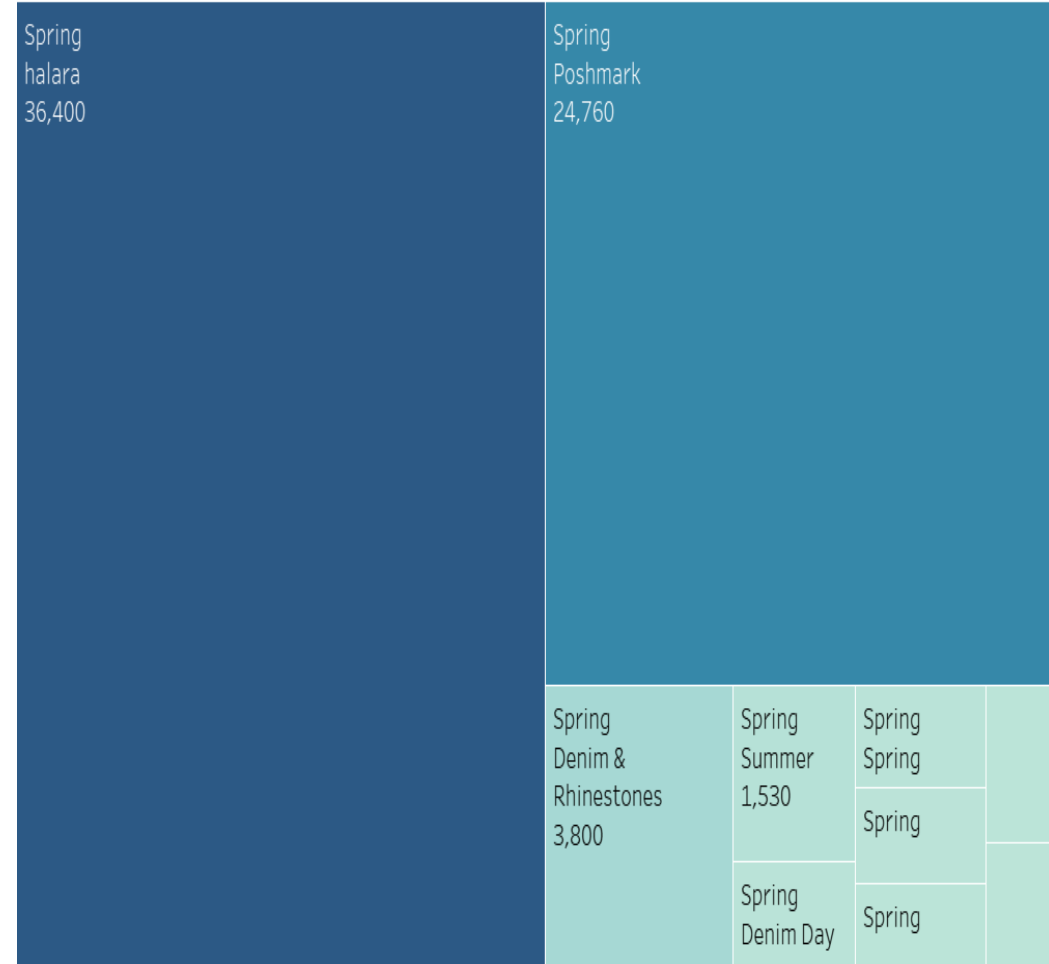
	A	B	C	D
1	query	value	category	season
2	clothing stores	100	clothing	Spring
3	clothing store	89	clothing	Spring
4	women clothing	44	clothing	Spring
5	womens clothing	38	clothing	Spring
6	mens clothing	37	clothing	Spring
7	stores near me	37	clothing	Spring
8	plus size clothing	36	clothing	Spring
9	clothing brands	34	clothing	Spring
10	clothing stores near me	32	clothing	Spring
11	men clothing	30	clothing	Spring
12	shein clothing	24	clothing	Spring
13	vintage clothing	18	clothing	Spring
14	cider	15	clothing	Spring
15	cider clothing	15	clothing	Spring
16	express	15	clothing	Spring
17	express clothing	14	clothing	Spring
18	clothing store near me	14	clothing	Spring
19	wholesale clothing	14	clothing	Spring
20	clothing rack	11	clothing	Spring





# Using Google Trends for SEO Optimization and Competition Analysis

- Google Trends data can be utilized to identify trending search terms throughout different seasons.
- It also enables us to determine which company garners the most search interest.
- This information can then be leveraged to enhance website rankings for frequently searched keywords, thereby boosting the potential for increased website traffic.
- Furthermore, it aids in understanding the types of clothing people seek during different seasons, offering opportunities to enhance operational efficiency and refine marketing strategies.



Since "Halara" is a trending topic in spring fashion, it's a great opportunity to perform a competitor analysis with Halara to identify potential product categories Gap should explore:

- **Halara Best Sellers:**
  - Analyze Halara's best-selling products, such as casual and active wear. Determine which specific items within these categories are popular.
- **Gap's Product Comparison:**
  - Compare Gap's existing product offerings in these categories with Halara's best sellers.
  - Identify any gaps in Gap's product line, such as missing styles, colors, or features that are in demand according to the search trends.
- **Pricing and Quality Comparison:**
  - Compare the pricing and quality of Gap's products with Halara's. This will help Gap position itself competitively.
- **Marketing Strategies:**
  - Investigate how Halara is marketing its products during the summer season. Are there marketing tactics or campaigns that Gap can learn from or adapt?
- **Customer Reviews and Feedback:**
  - Analyze customer reviews and feedback for Halara's products to understand what customers like and dislike. Gap can use this information to improve its own offerings.

- By conducting this competitor analysis, Gap can gain insights into the specific products and strategies that are driving Halara's success during the spring season. This data-driven approach will help Gap make informed decisions about which products to develop or promote to capitalize on emerging fashion trends and boost sales.
- Remember that this analysis is just a starting point, and Gap should continuously monitor trends and consumer preferences to stay ahead in the fashion industry.


6	Halara SoCinched leggings	
7	Drawstring Waist Joggers	
8	HalaraMagic Denim Leggings	
9	Crossover 7/8 Leggings	
10	Corduroy Casual Pants	
11	Pocket Flare Cargo Pants	
12	HalaraMagic Flare Jeans	
13	Breezefulâ„¢ 2-in-1 Skirt	
14	2-in-1 Shorts	

Get a free headband on orders \$79+ [Details](#)


# HALARA

Clothing Dresses Joggers & Pants Leggings Denim Tops < >


Sort By **Featured** v




3 Waist Joggers  
\$39.95  
XS-3X



Straight Leg Pants  
\$34.95 ~~\$39.95~~  
Available in XS-3X



2-in-1 Corduroy Skirt  
\$24.95 ~~\$29.95~~  
New Color



Cargo Pocket Leggings  
\$29.95 ~~\$49.95~~  
Available in XS-4X

# If brands will behave differently

- We'd like to substantiate that the sub-brands are inherently different and hence, would behave differently after the move
- We've scraped though a website for prices of products for the brands in question (Old Navy and Banana Republic) to compare using a two-sample t-test

$H_0$ : Prices between the two brands are similar

$H_A$ : Prices are different and hence the brands would behave differently

The p-value is quite small, and this tells us that the brands are quite different.

```
import numpy as np
from scipy import stats
Old_Navy_prices = [35.99, 22.99, 30.99, 34.99, 32.99, 32.99, 27.99, 32.99,
Banana_Republic_prices = [20.99,18.99,20.99,20.99,22.99,17.99,20.99,20.99,
t_statistic, p_value = stats.ttest_ind(brand1_prices, brand2_prices)
alpha = 0.05
if p_value < alpha:
    print("There is a significant difference between the two brands.")
else:
    print("There is no significant difference between the two brands.")
print(p_value)
```

There is a significant difference between the two brands.  
9.515257297847886e-30

SAVE SEARCH

4,861 items

Sort by: Recommended

Size

THRIFT TIP
Set your sizes to shop only items that fit you.

TOPS & DRESSES

- XXS XS SM MED
LG XL XXL OX
1X 2X 3X 4X
5X ONE SIZE



32

Banana Republic Casual Dr...
Size Med
\$35.99 \$23.39
35% off with code WELCOME
\$110



NEW WITH TAGS

46

Banana Republic Factory Store
Size Med
\$22.99 \$14.94
35% off with code WELCOME
\$75



16

Banana Republic Casual Dr...
Size Sm
\$30.99 \$20.14
35% off with code WELCOME
\$110



37

Banana Republic Casual Dr...
Size Med
\$34.99 \$22.74
35% off with code WELCOME
\$110

FEEDBACK

```
with open('Banana_Republic_Prices.csv', 'w', newline='', encoding="utf-8") as csvfile:
    # Create a CSV writer object
    writer = csv.writer(csvfile)
    #writer.writerow(["Product", "User", "Review"])

    # Iterate over the Lists and write each element to a separate column
    for i in range(len(text_gap)):
        writer.writerow([text_gap[i]])

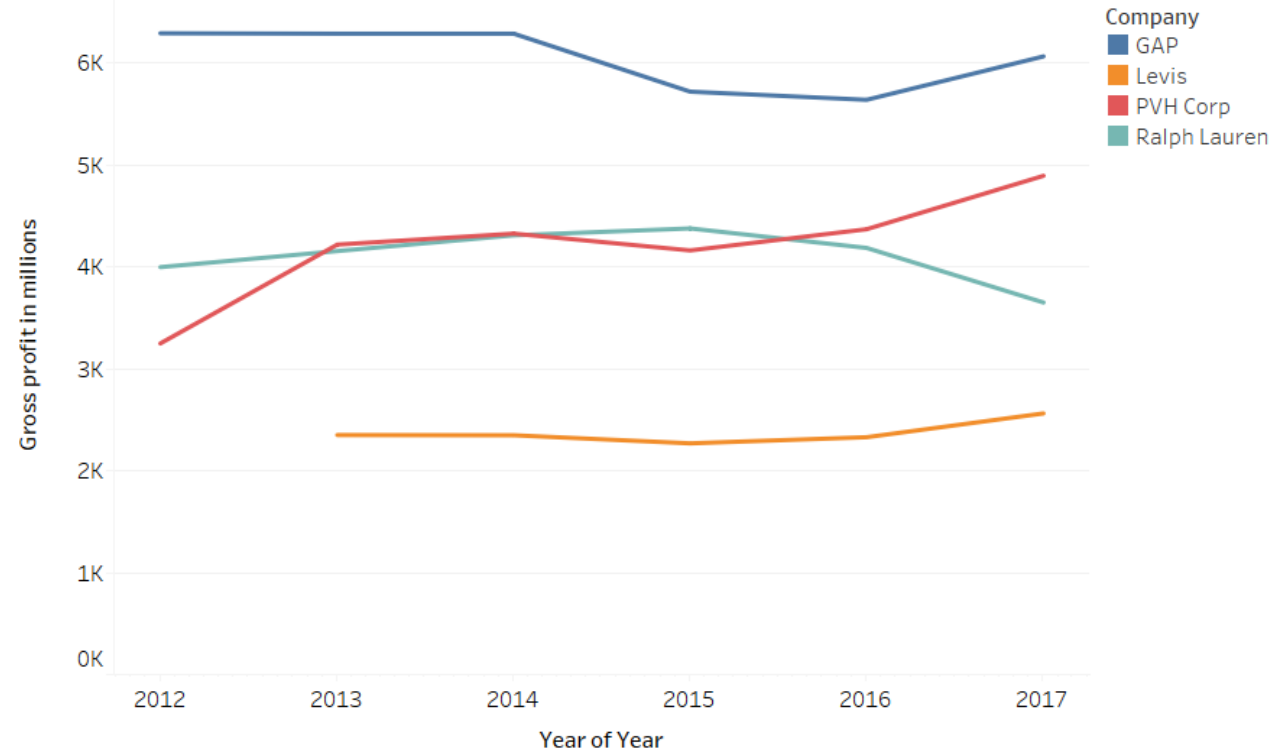
C:\Users\Shubhankar Sharma\AppData\Local\Temp\ipykernel_2288\1580259506.py:6: DeprecationWarning: Firefox will soon stop logging to geckodriver.log by default; Specify desired logs with log_output
driver = webdriver.Firefox(service=Service(geckodriver_manager.install()), options = firefox_options)

[35.99, 22.99, 30.99, 34.99, 32.99, 32.99, 27.99, 32.99, 39.99, 30.99, 21.99, 25.99, 45.99, 35.99, 39.99, 33.99, 24.99, 38.99, 28.99, 25.99, 55.99, 26.99, 36.99, 26.99, 36.99, 30.99, 28.99, 36.99, 18.99, 31.99, 28.99, 36.99, 37.99, 23.99, 33.99, 19.99, 39.99, 30.99, 41.99, 29.99, 10.99, 42.99, 34.99, 33.99, 21.99, 35.99, 24.99, 28.99, 28.99, 31.99, 38.99, 28.99, 33.99, 28.99, 31.99, 19.99, 28.99, 27.99, 28.99, 12.99, 31.99, 23.99, 37.99, 26.99, 37.99, 31.99, 35.99, 27.99, 38.99, 51.99, 29.99, 16.99, 23.99, 21.99, 38.99, 26.99, 32.99, 28.99, 23.99, 41.99, 26.99, 33.99, 42.99, 27.99, 34.99, 30.99, 32.99, 31.99, 27.99, 36.99, 43.99, 27.99, 12.99, 18.99, 27.99, 77.99, 40.99, 39.99, 32.99, 34.99, 38.99, 40.99, 31.99, 28.99, 42.99, 28.99, 26.99, 34.99, 30.99, 35.99, 41.99, 31.99, 31.99, 25.99, 33.99, 32.99, 21.99, 28.99]
```

# Insights from 10K Files

**From the 10K forms, significant financial details about the company's fiscal health can be extracted**

Gross Profit Across Years



1. PVH Corp adopted big data in its business model during the early 2010s. During this period, some other leading brands did not fully embrace big data.
2. PVH Corp experienced an upward trend in its gross profit margin. This improvement suggests better performance compared to its competitors.
3. The increased use of big data and analytics by PVH Corp is a likely contributing factor. This strengthens the credibility of using big data and analytics in the fashion industry.



# Insights from 10K Files

Year	Brand	U.S.A	Canada	Europe	Asia	Other regions
2015	Gap Global	3303	348	726	1215	159
2015	Old Navy Global	5987	467	0	194	27
2015	Banana Republic	2211	229	71	112	33
2016	Gap Global	3113	368	630	1215	129
2016	Old Navy Global	6051	490	0	220	53
2016	Banana Republic	2052	223	59	109	28
2017	Gap Global	3065	398	626	1117	112
2017	Old Navy Global	6570	547	0	50	71
2017	Banana Republic	2017	225	15	96	27

Data shows sales in millions

- Sales fluctuate significantly across various geographic regions, it's crucial to adapt marketing strategies accordingly to effectively target untapped markets
- These numbers suggest that either the marketing has not been very effective in regions other than USA and Canada or the company wants to focus primarily on USA and Canada
- In either case, there is a lot of scope to expand in Europe, Asia and Other regions



# Demographic Data

- Population data from US Census
- Gap will gain insights about target consumers for each state by analyzing the proportion of age/sex groups

Age	April 1, 2020 Estimates Base			Population Estimate (as of July 1)								
				2020			2021			2022		
	Both Sexes	Male	Female	Both Sexes	Male	Female	Both Sexes	Male	Female	Both Sexes	Male	Female
Total	5,024,356	2,447,300	2,577,056	5,031,362	2,449,590	2,581,772	5,049,846	2,456,486	2,593,360	5,074,296	2,467,360	2,606,936
Under 5 years	293,152	149,784	143,368	292,324	149,452	142,872	290,239	148,427	141,812	290,299	148,219	142,080
5 to 9 years	301,260	153,702	147,558	301,405	153,694	147,711	303,273	154,302	148,971	305,273	155,284	149,989
10 to 14 years	325,272	165,986	159,286	324,611	165,701	158,910	322,310	164,814	157,496	318,447	162,831	155,616
15 to 19 years	333,178	169,249	163,929	334,621	169,700	164,921	332,610	168,511	164,099	334,114	169,535	164,579
20 to 24 years	336,560	169,504	167,056	340,893	171,504	169,389	351,850	176,566	175,284	355,450	178,013	177,437
25 to 29 years	331,962	165,808	166,154	330,296	165,045	165,251	324,479	162,649	161,830	321,108	161,252	159,856
30 to 34 years	320,780	157,998	162,782	322,479	158,672	163,807	330,321	162,546	167,775	336,797	165,873	170,924
35 to 39 years	311,703	152,827	158,876	311,120	152,302	158,818	310,540	151,535	159,005	311,105	151,801	159,304
40 to 44 years	300,658	146,860	153,798	301,617	147,226	154,391	306,954	149,864	157,090	310,743	151,745	158,998
45 to 49 years	312,942	153,884	159,058	310,445	152,419	158,026	300,011	146,641	153,370	296,419	144,399	152,020
50 to 54 years	314,388	155,135	159,253	314,657	155,251	159,406	318,776	157,313	161,463	321,378	158,369	163,009
55 to 59 years	342,388	166,981	175,407	341,389	166,463	174,926	333,875	162,742	171,133	325,873	159,062	166,811
60 to 64 years	329,613	157,963	171,650	329,933	158,102	171,831	331,923	159,073	172,850	332,557	159,424	173,133
65 to 69 years	284,126	132,906	151,220	286,006	133,788	152,218	291,134	136,060	155,074	297,711	139,305	158,406
70 to 74 years	234,419	107,693	126,726	236,378	108,610	127,768	245,537	112,316	133,221	243,171	110,603	132,568
75 to 79 years	157,465	69,302	88,163	158,085	69,529	88,556	160,176	70,319	89,857	174,946	77,170	97,776
80 to 84 years	101,922	41,452	60,470	102,234	41,674	60,560	103,093	42,093	61,000	105,534	43,303	62,231

# Old Navy Vs. Banana Republic

- Since Old Navy is fast-fashion based, moving to the data-driven approach would really benefit the overall sales of the brand. As it makes up majority of the sales of GAP Inc., the firm would profit from this move. Overall, GAP and Old Navy may rely more heavily on data-driven strategies to serve their broader customer base
- However, Banana Republic is more focused on high-end consumers. The brand would benefit more from a creative approach as the typical customers of this brand would be looking to find unique apparel. Hence, the brand may require a stronger creative direction to cater to its upscale clientele.

# Big Data in Marketing

- Customer Segmentation can help in highly targeted marketing campaigns
- Personalization can lead to increased customer engagement and loyalty
- Customer journey mapping to optimize each stage for a seamless experience
- Demand Forecasting helps with inventory management, pricing strategies, and stock replenishment
- A/B testing allows marketers to experiment with different creative elements, messaging, and strategies to determine what resonates best with the audience
- For online advertising, predictive analytics can optimize ad targeting, bidding, and placement to maximize ROI.
- Content Optimization: To analyze which types of content perform best, guiding content creation and distribution strategies.

# Where Big Data falls behind

- **Creative Ideation:** While data can inform creative decisions, it's less effective at generating entirely new and innovative ideas. The initial spark of creativity often requires a human touch.
- **Brand Identity and Storytelling:** Establishing a brand's identity and crafting compelling brand stories may rely more on art than science. Emotional connections with consumers often depend on creative storytelling.
- **Crisis Management:** During PR crises or sensitive situations, immediate responses should focus on empathy and brand values, which may not be solely data-driven.
- **Long-Term Brand Building:** Building a brand's reputation and equity over time requires consistent messaging and creative strategies that may not rely solely on data-driven insights.

# Art Vs. Science in Marketing

In an ideal world, we would look to strike a perfect balance in the usage between art and science in marketing

## **Art rules when:**

- Creativity and emotional connection are crucial for the brand's identity and messaging.
- The marketing strategy involves storytelling and building brand loyalty.
- Need to stand out in a competitive market through unique and innovative campaigns.
- Managing crises and sensitive situations

## **Science rules when:**

- Data is abundant and can provide actionable insights.
- Decisions need to be based on quantifiable metrics, such as ROI and conversion rates.
- Personalization and targeting are essential.
- Testing and experimentation are required to optimize campaigns.

# Balancing Art and Science

Achieving the right balance between art and science in marketing is crucial. The two should complement each other:

- **Start with Art:** Creative ideas and brand identity should provide the foundation. Develop compelling stories and visuals that resonate with your audience.
- **Enhance with Science:** Use data and analytics to refine your creative strategies. Measure performance, segment your audience, and personalize content based on insights.
- **Iterate:** Continuously iterate and refine your marketing approach. Use data to inform creative updates, but never lose sight of the emotional connection that art can foster.

Successful marketing often combines the art of creativity with the science of data-driven decision-making. The key is to recognize when each element should take the lead and when they should work in harmony to achieve marketing goals.

# Art & Science at GAP Inc.

## Data Collection and Analysis:

- Gather data, understanding customer behavior, and customer segmentation

## Art and creative aspects

- Reflect Gap's unique style and values, and resonate with its target audience
- Content creation and A/B testing

## Science and Data-Driven Decision-Making

- Personalization, ROI measurement, conversion rate optimization

## Integration of Art and Science

- Cross-functional collaboration to enable data-driven creativity

## Customer Feedback

- Feedback loop to refine both creative and data-driven marketing strategies

## Continuous Improvement

- Iterate and Experiment



# Limitations of our Analysis and Approach

- **Data Quality and Reliability:** The quality and reliability of the data you collect can be a significant limitation. If the data sources contain errors, inconsistencies, or biases, it can lead to inaccurate conclusions.
- **Limited Data Sources:** Depending solely on web scraping and online sources limit the scope of our analysis. Important data, such as internal sales and customer data, are not accessible through online sources.
- **Selection Bias:** The data you collect might not be representative of the entire customer base or market. If you're only collecting data from certain websites or sources, it may not accurately reflect the broader consumer sentiment or behavior.



THANK YOU!