Not all Collaborations are Created Equal

Polina, Katie, Shirley

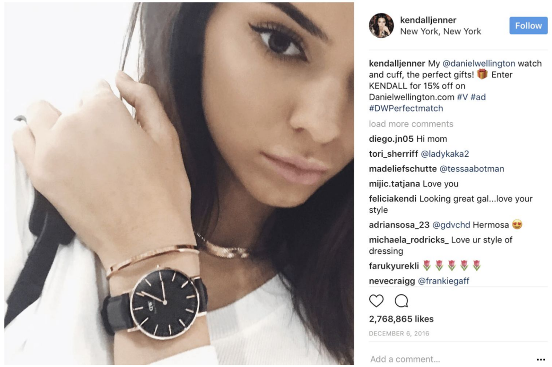
Introduction

What is an Influencer?

- Influencer is a user on social media who has the ability to affect the purchasing decision of their followers because of their experiences or brand
  - Messi recommending Adidas cleats
- Influencers usually have a following in a distinct niche, i.e.

  - **Gaming**
  - **Beauty**
  - **Fitness**
  - **Parenting**
  - **Travel**
  - **Photography**

- **61% of consumers trust influencer recommendations**, compared to 38% who trust brand-produced content



kendalljenner
New York, New York

kendalljenner My @danielwellington watch and cuff, the perfect gifts! 🎁 Enter KENDALL for 15% off on Danielwellington.com #V #ad #DWPerfectmatch

load more comments

diego.jn05 Hi mom

tori_sherriff @ladykaka2

madeliefschutte @tessaabotman

mijic.tatjana Love you

feliciakendi Looking great gal...love your style

adriansosa_23 @gdvchd Hermosa 😍

michaela_rodricks_ Love ur style of dressing

farukyurekli 🌷🌷🌷🌷🌷

nevecraigg @frankiegaff

2,768,865 likes

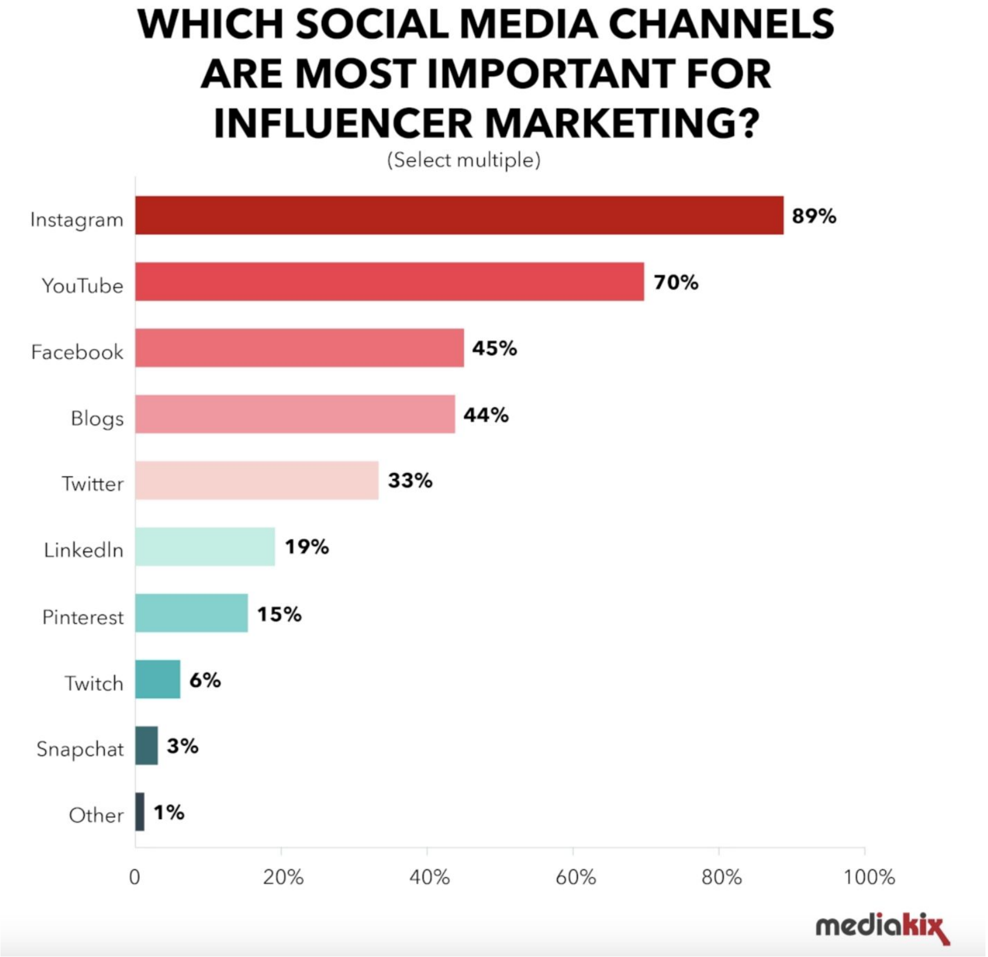
DECEMBER 6, 2016

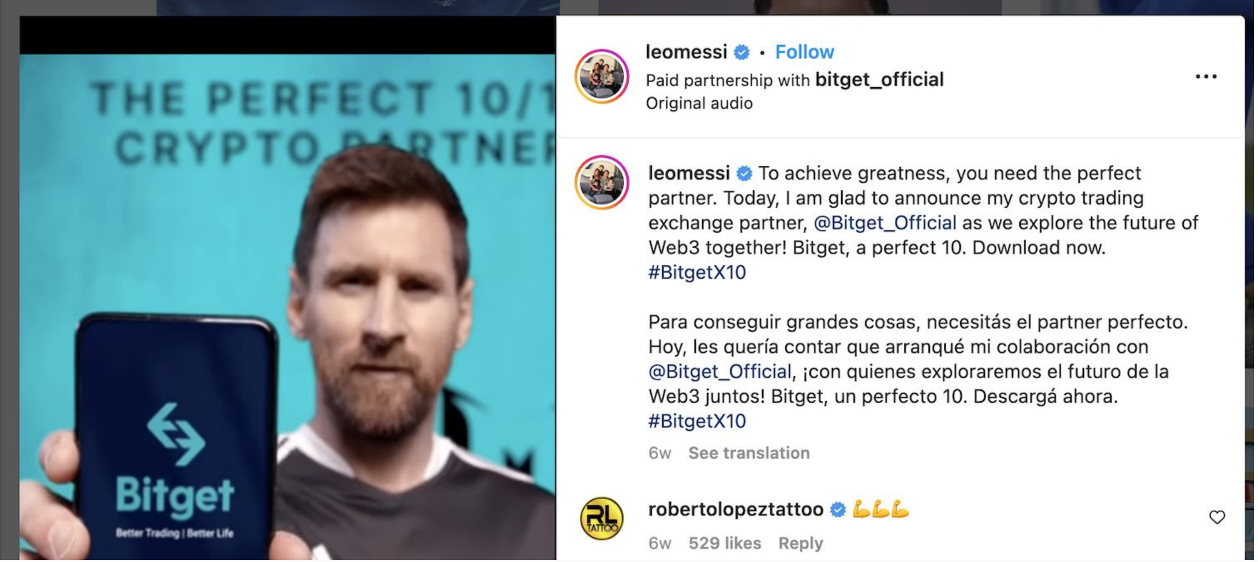
Instagram by Numbers

- **Instagram is the preferred Social Media Channel** for Brands who engage in Influencer Marketing

- Instagram is #4 SM after FB, YT and WhatsApp, and has over **500 million daily active users** with over **2 billion** monthly active users

- On average, **marketers receive $4.87** of **earned media value** (EMV) for **each $1** paid for an **Instagram influencer's** promotion.





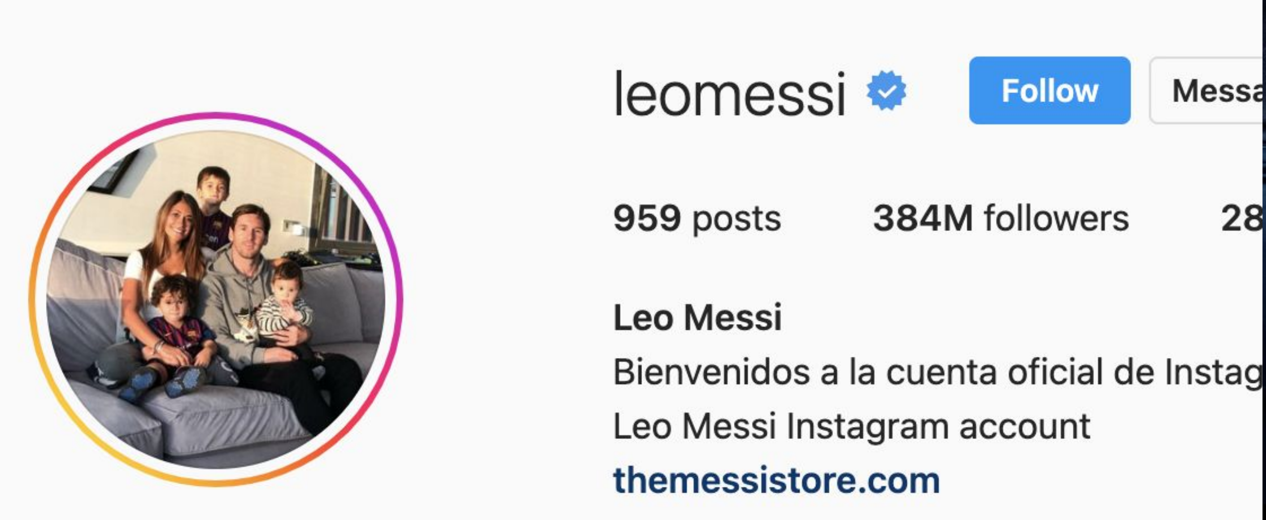
leomessi • Follow
Paid partnership with **bitget_official**
Original audio

leomessi To achieve greatness, you need the perfect partner. Today, I am glad to announce my crypto trading exchange partner, @Bitget_Official as we explore the future of Web3 together! Bitget, a perfect 10. Download now.
#BitgetX10

Para conseguir grandes cosas, necesitás el partner perfecto. Hoy, les quería contar que arranqué mi colaboración con @Bitget_Official, ¡con quienes exploraremos el futuro de la Web3 juntos! Bitget, un perfecto 10. Descargá ahora.
#BitgetX10

6w See translation

robertolopeztattoo 💪💪💪
6w 529 likes Reply

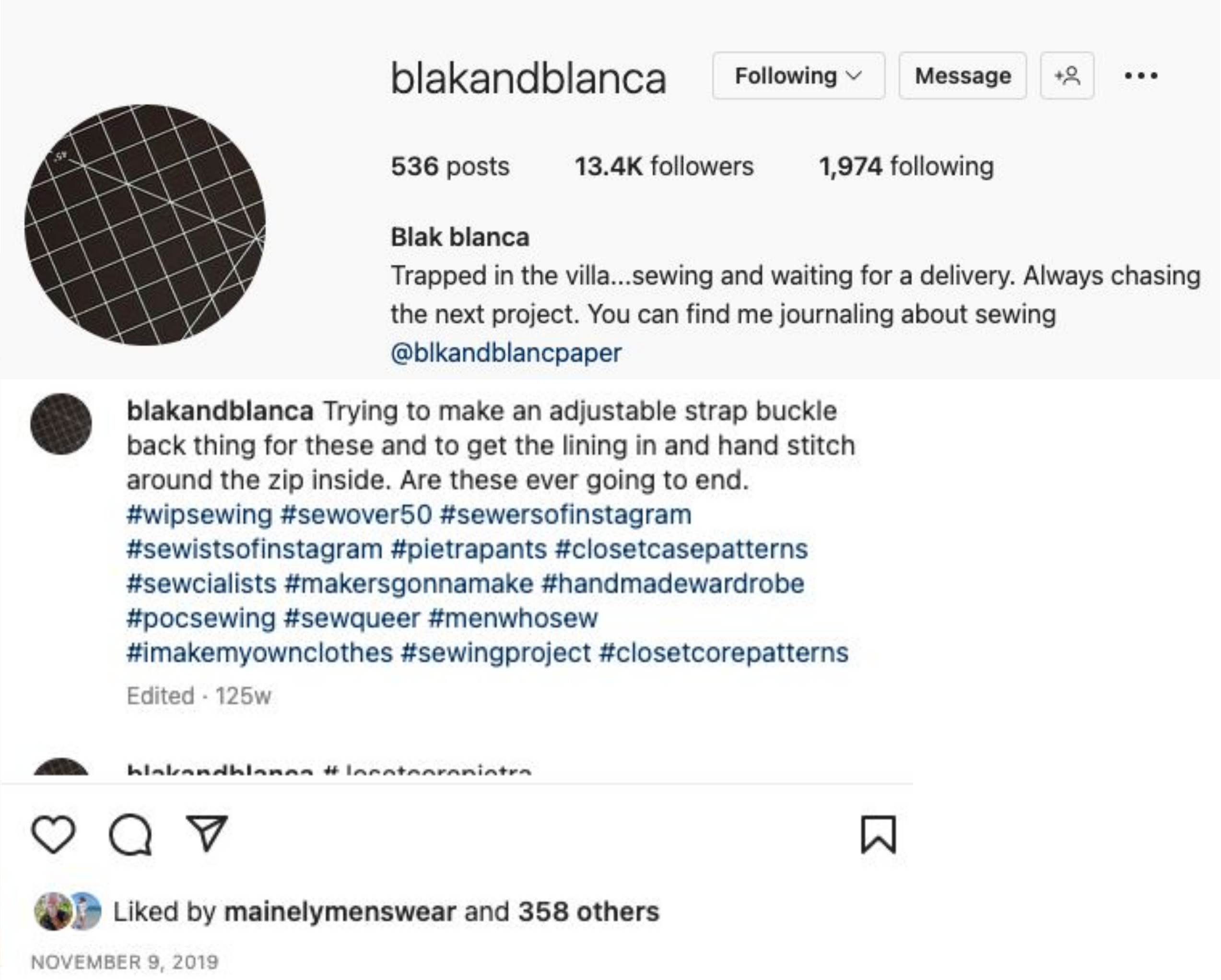


leomessi

959 posts **384M** followers **28**

Leo Messi
Bienvenidos a la cuenta oficial de Instag
Leo Messi Instagram account
themessistore.com

Niche Influencers



blakandblanca Following Message

536 posts 13.4K followers 1,974 following

Blak blanca
Trapped in the villa...sewing and waiting for a delivery. Always chasing the next project. You can find me journaling about sewing @blkandblancpaper

blakandblanca Trying to make an adjustable strap buckle back thing for these and to get the lining in and hand stitch around the zip inside. Are these ever going to end. #wipsewing #sewover50 #sewersofinstagram #sewistsofinstagram #pietrapants #closetcasepatterns #sewcialists #makersgonnamake #handmadewardrobe #pocsewing #sewqueer #menwhosew #imakemyownclothes #sewingproject #closetcorepatterns

Edited · 125w

blakandblanca #closetcorepietra

Liked by **mainelymenswear** and **358 others**

NOVEMBER 9, 2019

Objectives

We want to explore **Brand - Influencer relationship on Instagram** to better understand:

1. How do Brands choose Influencers to partner with by evaluating various aspects including:

- topics/categories
- reach/ following
- engagement levels (comments, likes)
- demographics (age)

2. Which Sponsorship strategies work best for Brands (i.e. repeat activations vs one-off activations etc.)
3. Who are the "Super Influencers" - Influencers popular among many Brands, Influencers who get most engagement etc.

DATA SOURCES

Data Sources Overview

1. **Influencer Dataset**
   - Influencer and Brand profiles data, post metadata, post content
   - 1,601,074 Instagram posts, 38,113 Instagram influencers, 26,910 brands
   - Posts are labeled as 'Sponsored' if the post either uses the branded content tool or contains sponsorship-related hashtags
   - https://drive.google.com/drive/folders/1aq1KVoqfVuRBmsEofUf_oH4mG64oQp4Q?usp=sharing
2. **Famous Birthdays Dataset**
   - Scrape the famous birthday website for information on the influencers
   - Goal is to categorize influencers
3. **Wikipedia**
   - Use wikipedia to categorize the brands and bolster our information on each

Data Structure: Influencer & Brand Dataset

1. **Post_info.txt.** This file contains a list of 1,601,074 posts. Each line represents a post and is composed of 4 columns of information.

   [Post ID] [USER name] [Sponsorship label] [JSON file] [Image files]

2. **Json_files.zip.** This zip file contains JSON files of the 1,601,074 posts.
   JSON files have various information such as captions, likes, comments, timestamps, sponsorship, usertags, etc

3. **Profiles_influencers.zip.** This zip file contains Instagram profiles of the 38,113 influencers.
   The name of each file indicates the username of the corresponding influencer.
   When you open up the file using text editors, you will see one line of following information separated by Tab.

   [Name] [Followers] [Followees] [Posts] [URL] [T/F] [Category] [Bio] [E-mail] [Phone] [Profile_pic]

4. **Profiles_brands.zip.** This zip file contains Instagram profiles of the 25,282 brands.
   The fields of each profile are the same as influencer profiles.

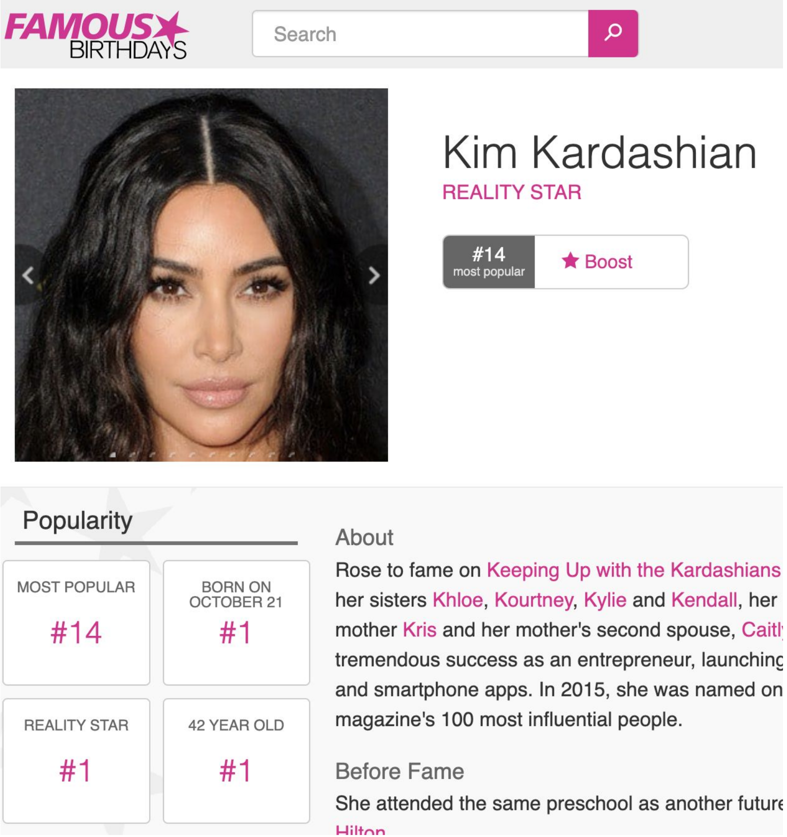
Data Structure: Famous Birthdays (Influencer Info)

OVERVIEW

- Celebrities are profiled in a simple and entertaining format
- Each celebrity profile contains popularity rankings based on user activity
- 30 million monthly unique users
- Famous Birthdays is also available in Spanish, Portuguese, French and Japanese

DISCOVERING CELEBRITIES

- Search by birthday, birthplace, profession, TV show, and more
- View today's most popular birthdays
- See the celebrities trending right now



About

Rose to fame on Keeping Up with the Kardashians
her sisters Khloe, Kourtney, Kylie and Kendall, her
mother Kris and her mother's second spouse, Caitl
tremendous success as an entrepreneur, launching
and smartphone apps. In 2015, she was named on
magazine's 100 most influential people.

Before Fame

She attended the same preschool as another future
Hilton

Data Structure: Wikipedia Page (Brand Info)

1. **JSON Extraction for Wikipedia.** Extract JSON object of a wikipedia api search query url.

   [Title] [Pageid] [Size] [Snippet] [Timestamp]

2. **Wikipedia API for Python** Extract Wikipedia summary

   [Title] [Summary] [Page] [Categories] [Links] [Namespace]

METHODOLOGY

Instagram Data - Reading Data

Challenge: We needed to process 3 folders with multiple file (1,601,074 Instagram posts, 38,113 Instagram influencers, 26,910 brands). Brand and Influencer folders contain separate files for each profile. JSON folder contains separate JSON file for each post.

Solution:

1. Read profile data from all profile files (Brand Profiles and Influencer Profiles)
2. Concat all the Data Frames and return a single Dataframe with all the profiles (one for Brands, one for Influencers)
3. Save the processed data (Dataframes) to csv files for simplicity and time management
4. Read post_info.txt file, map each field to its respective header
5. Process raw JSON files and normalize the data
6. Load JSON data into a dataframe. Save the dataframe to csv file for simplicity and time management

Instagram Data - Filtering & Matching

Challenge: To make sense of the massive datasets, we decided to focus our efforts on a limited number of large brands. We manually selected Brands of interest and found posts that mention those brands. We then looked at the Influencers who published those posts to analyze their details.

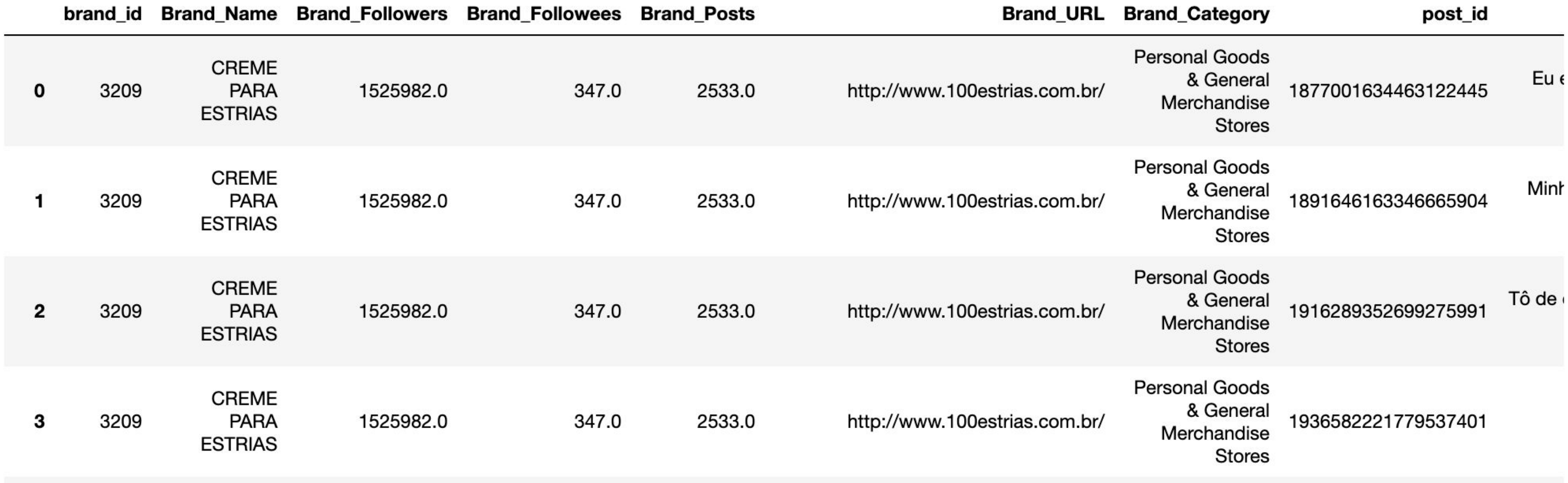
Solution:

1. Filter JSON posts by Sponsorship Label = 1 (we are interested only in Sponsored posts)
2. Select brands with > 1.5M and less than 1.6M followers
3. Tokenize post captions (split string by whitespace remove all special characters except @)
4. Find usernames using @ in the captions
5. Check if the username from captions is in the list of target brands usernames using .isin
6. Get 20 top Influencers (by following) for each caption/brand username match
7. Result: 35 brands, 279 Influencers, 441 posts

Instagram Data - Joining Tables

We joined all 3 tables (Posts, Influencers, Brands) and selected columns that provide valuable information about objects (nodes) and relationships. Our resulting table is a joined table featuring:

- Brands of interest and their attributes
- Posts that mention Brands of interest in captions + their attributes
- Influencer who published those posts and their attributes



	brand_id	Brand_Name	Brand_Followers	Brand_Followees	Brand_Posts	Brand_URL	Brand_Category	post_id	
0	3209	CREME PARA ESTRIAS	1525982.0	347.0	2533.0	http://www.100estrias.com.br/	Personal Goods & General Merchandise Stores	1877001634463122445	Eu e
1	3209	CREME PARA ESTRIAS	1525982.0	347.0	2533.0	http://www.100estrias.com.br/	Personal Goods & General Merchandise Stores	1891646163346665904	Minh
2	3209	CREME PARA ESTRIAS	1525982.0	347.0	2533.0	http://www.100estrias.com.br/	Personal Goods & General Merchandise Stores	1916289352699275991	Tô de
3	3209	CREME PARA ESTRIAS	1525982.0	347.0	2533.0	http://www.100estrias.com.br/	Personal Goods & General Merchandise Stores	1936582221779537401	

Wikipedia - Keyword Extraction

Challenge: Find the correct wikipedia page that corresponds to our brands, then extract keywords from the summary of the page

Solution:

- Used JSON Search API to search Wikipedia for our brand name
- Used Cosine Similarity scores to extract best search result per brand
    - Searched: Brand and Brand + “(company)”
- With the list of wikipedia titles, utilized Lemmatization and LDA to find the keywords from the brand’s wikipedia summary content
- Added Keywords to Brand nodes for categorization use in the knowledge graph

Famous Birthdays - Data Extraction

Challenges/Concerns:

- No access to API, but each page shared a standard format
- You don’t know how large the dataset actually is
    - Your only options are to enter their name or search by category
- Influencer names were not their actual names
- CAPTCHAs

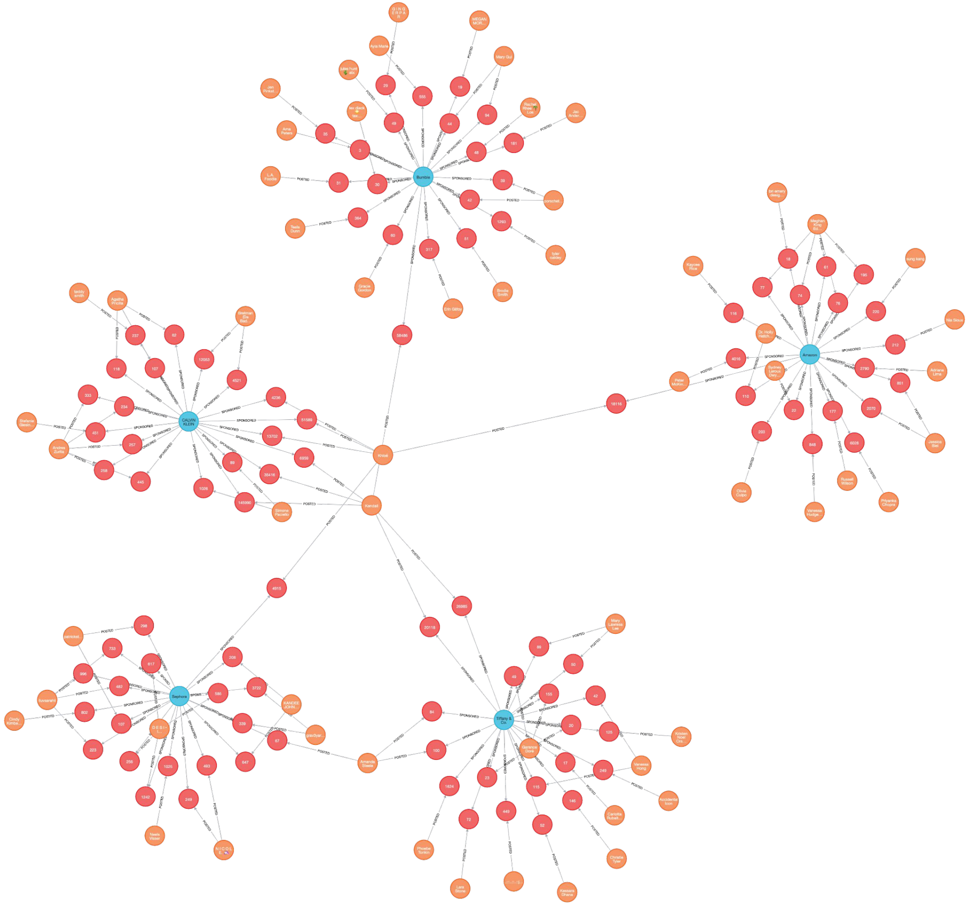
Solution:

- Given Influencer Names from dataset
- Clean up names (emojis 🦄🌟🌴, prefixes)
- Used Selenium to extract the information I was looking to scrape using the element’s XPath to find where it was located on the webpage
- Automate the web browser interaction utilizing the search bar, clicks and saving the results

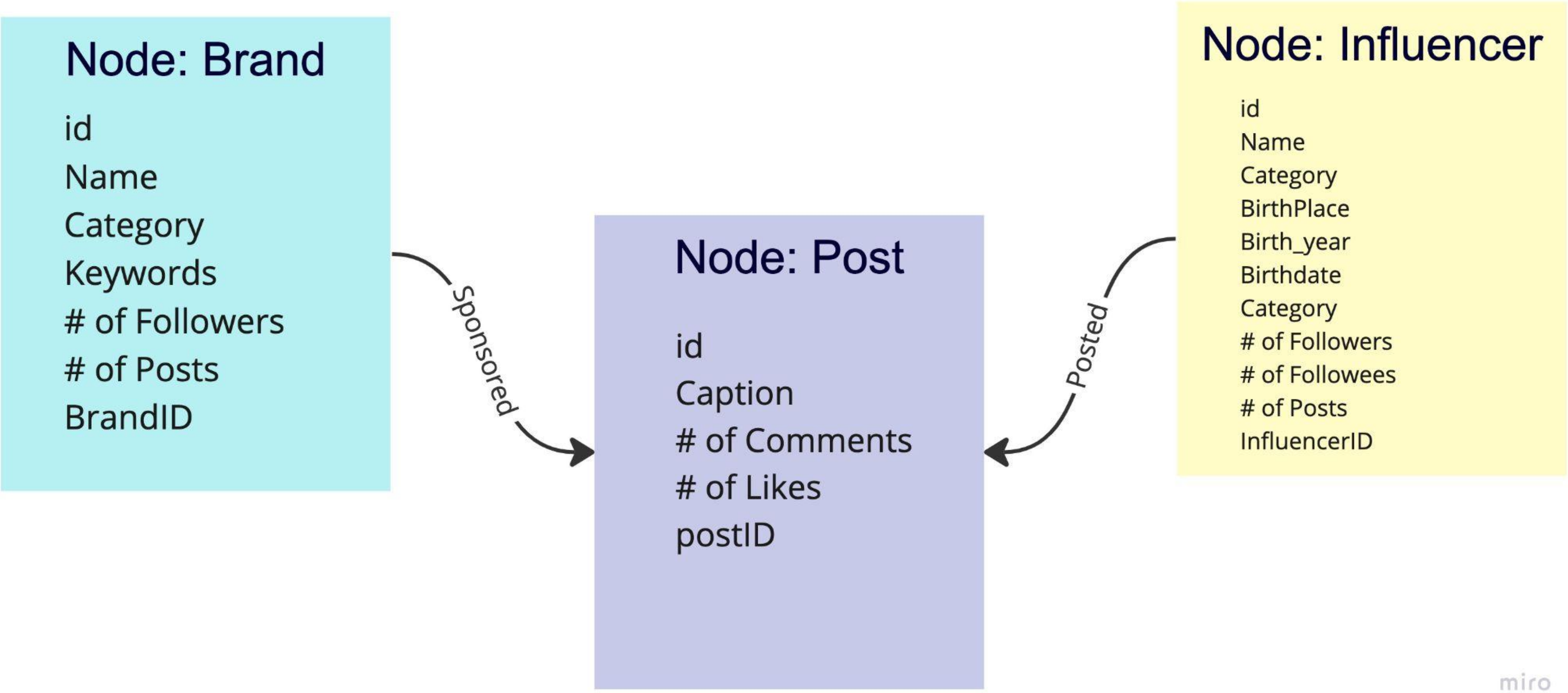
KNOWLEDGE GRAPH

Knowledge Graph Structure

- Nodes
  - Influencers **(279)**
  - Brands **(35)**
  - Posts **(441)**
- Edges/Relationships
  - Sponsored: Brand to Post **(442)**
  - Posted: Influencer to Post **(441)**



Knowledge Graph Structure



Node: Brand

id
Name
Category
Keywords
# of Followers
# of Posts
BrandID

Node: Post

id
Caption
# of Comments
# of Likes
postID

Node: Influencer

id
Name
Category
BirthPlace
Birth_year
Birthdate
Category
# of Followers
# of Followees
# of Posts
InfluencerID

DEMO

Queries - Demographics

What is the Average Age of the Sponsored Influencer by Brand?

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
Return b.Name, avg(toInteger(i.Birth_Year)) as Birth_Year
Order by Birth_Year asc
```

Conclusion: the oldest Influencers are hired by Bottega Veneta (average Year of Birth is 1971) and the youngest by Gymshark Women(average Year of Birth is 1998)

Which Country “produces” the most Influencers?

```
Match (i:Influencer)
Return i.BirthPlace, count(i.Name) as NumberofInfluencers
Order by NumberofInfluencers desc
```

Conclusion: top 3 countries are US, England, and Canada

Queries - Following & Categories

Which Influencer Category is most popular among Brands?

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
Return  i.Category as InfluencerCategory, Count(i.Category) as NumberOfCategories
order by NumberOfCategories desc
```

Conclusion:"Creators & Celebrities" is the most popular category

What is the average Following of Influencers by Brand? Which Brands hire Top Influencers (by Following size)

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
Return b.Name, avg(toInteger(i.Followers)) as Following
Order by Following desc
```

Conclusion: Amazon, Adidas and Moda tend to hire Mega Influencers (1M+ followers)

Queries - Collaboration Performance

Which brands have the highest engagement per collaboration with an influencer?

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
return Count(distinct i) as influencerCount, Sum(DISTINCT toInteger(p.Likes)) as NumberOfLikes,
Sum(DISTINCT toFloat(p.Likes))/Count(i) as Avg_Likes, b.Name as BrandName
order by Avg_Likes desc
```

Conclusion: Amazon, Moda, Quay Australia, and Adidas had highest average likes per post

Does repeated collaboration with the same influencer increase or harm engagement?

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
return Count(i) as NumberOfCollaborations, i.Name as InfluencerName, b.Name as
BrandName,SUM(toFloat(p.Likes))/(toFloat(i.Followers)) as ratio
order by ratio desc
```

Conclusion: Repeated posts with the same brand and influencer collaboration increased engagement substantially

Queries - Influencers of Choice

Who are the most popular Influencers for Brands in Personal Goods & General Merchandise Stores Category?

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
Where b.Category = "Personal Goods & General Merchandise Stores"
Return b.Category, i.Name, count(i.Name) as InfluencerofChoice
Order by InfluencerofChoice desc;
```

Conclusion: Lorena Improta is most desirable Influencer in "Personal Goods & General Merchandise Stores" Brand Category

Which Influencer generated most positive engagement for Amazon

```
Match (i:Influencer)-[:POSTED]-(p:Post)-[:SPONSORED]-(b:Brand)
Where b.Name = "Amazon"
Return i.Name, SUM(toInteger(p.Likes)) as Likes
Order by Likes Desc
```

Conclusion: Khloe Kardashian generated most Likes for Amazon

Queries - Brand Affinity

Are there Posts Sponsored by more than 1 Brand?

```
MATCH (a:Brand)–[:SPONSORED]↣(p:Post)↢[:SPONSORED]–(b:Brand)
WHERE id(a) > id(b) WITH a, b,
COUNT(p) AS count
ORDER BY count
DESC RETURN a.Name, b.Name
```

Conclusion: There is one Post Sponsored by “Universal Orlando Resort” and “Universal Studios Hollywood”

Which Brands are targeting the same audience (sponsoring the same influencers)?

```
MATCH(a:Brand)–[:SPONSORED]↣(p:Post)–[]–(i:Influencer)–[]–(p1:Post)↢[:SPONSORED]–(b:Brand)
WHERE id(a) > id(b)
WITH a,b,
COUNT(i) AS count
ORDER BY count
DESC RETURN a.Name, b.Name, count
```

Conclusion: Brands demonstrating the highest affinity are “Moda” and “CREME PARA ESTRIAS”

SUMMARY

Key Findings

- Repeated Sponsorships with the same influencer generate higher engagement than one-off activations
- Brands select Influencers based on their age group. The “oldest” brand in our sample is Bottega Veneta, The youngest brand is Gymshark Women
- Big Brands like Amazon and Adidas tend to hire Mega Influencers like Khloe Kardashian
- Mega Influencers drive the highest engagement for Brands
- “Creators & Celebrities” is the most popular Influencer category among Brands
- “Moda” and “CREME PARA ESTRIAS” demonstrate highest affinity sponsoring the highest number of the same Influencers

Post Mortem/Future Goals

- Analyze more Sponsorships/Influencers per Brand (not limit to 20 Influencers per Brand). We were limited with the speed/efficiency of extracting additional data from Famous Birthdays

- Explore Google Knowledge Graph - didn't have enough time, but it would provide with much more information on both Brands and Influencers

- Get additional information about Brands - scraping About pages (we had Brands websites URLs available in the original dataset)

- What would be even more interesting is to analyze relationships between Influencers (who follows who) to look into "communities". We would need to use Instagram API for that

- Spend more time on refining the automated scraping process
  - explore the relationships