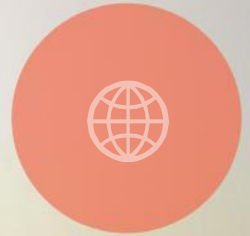# Comparative Analysis of NYC

Prakhar Bansal

Karthick Ramasubramanian

Shubham Singh

Soumi Basu

Soumik Choudhuri

Problem Statement

What is projected vs What actually matters
– Comparative analysis of NYC

Why it matters?

Differences between what is publicized about a city vs what actually matters about it.

Uses:
Resolve important problems pointed out by the posts and comments
Promote things otherwise neglected – pointed out by disparity between posts and comments

By identifying:
The important aspects of a city in the opinion of people, derived from text analytics across blogs
The topics represented/publicized by the official Instagram social media handle of the city, derived by performing image analytics across 3 official NYC channels

Data Source

Instagram Pages representing NYC

The idea was to consider what topics were demonstrated by posts for the New York City for the last 1 year.
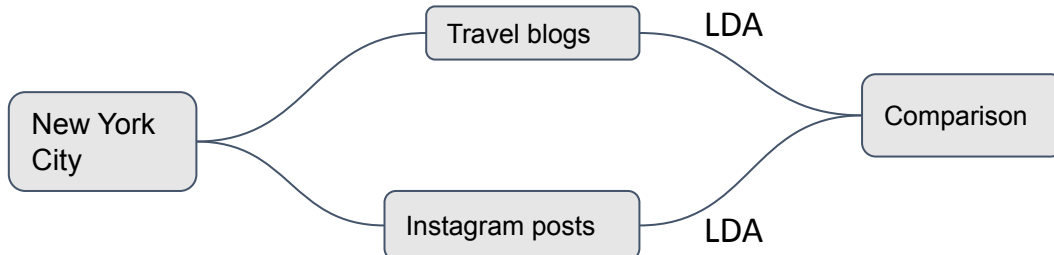
- **nycgov**
- ~600 posts scrapped

- **visit_nyc**
- ~720 posts scrapped
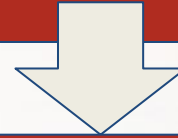
- **secret_nyc**
- ~620 posts scrapped

Travel pages – Tripadvisor

- **(https://www.tripadvisor.com/ShowForum-g607 63-i5-New_York_City_New_York.html)**

- **Scrapped ~16000 comments about New York City to identify the most important topics people spoke about while discussing things about New York city.**

- **Considered the comments for the last 1 year to see what people spoke about most when talking about NYC**

New York City — Travel blogs — LDA — Comparison

New York City — Instagram posts — LDA — Comparison

PART 1 – Image Analytics

Scraped ~2000 Instagram images using Instaloader

Used Google Cloud Vision to extract labels of images

Combined Captions and Labels as one list

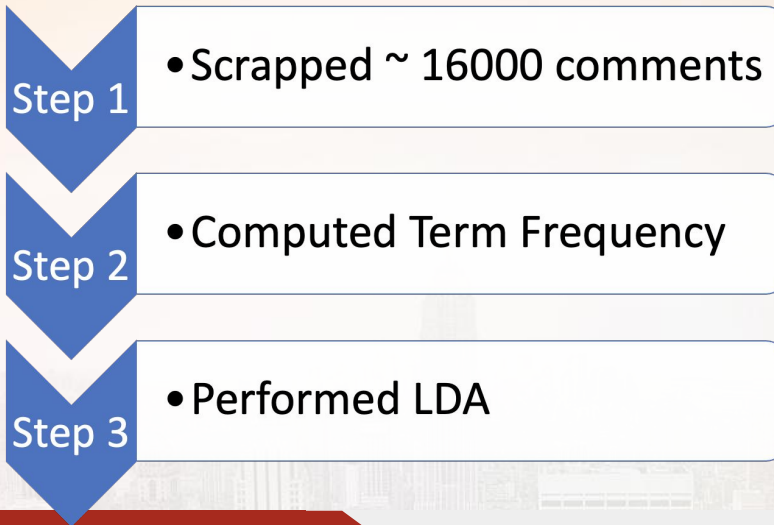Removed stop words, hashtags, special characters, punctuations and lemmatized words

Removed stop words, hashtags, special characters, punctuations and lemmatized words

Topic Modelling on Images

Topic 1	**Skyline**	Skyscraper, Block, Cloud, Daytime, World
Topic 2	**Landscape**	Nature, Flower, Natural, Landscape, Leisure
Topic 3	Event	Rectangle, Electric, Tickets, Brand, Circle
Topic 4	**Traffic**	Vehicle, Automotive, Infrastructure, Lighting, Wheel
Topic 5	**People**	Covid, Gesture, Happy, Event, Yorkers
Topic 6	**Statue of Liberty**	Bridge, Facade, Light, Lighting, Window

Part 2 - Blog Analytics

Step 1
- Scrapped ~ 16000 comments

Step 2
- Computed Term Frequency

Step 3
- Performed LDA

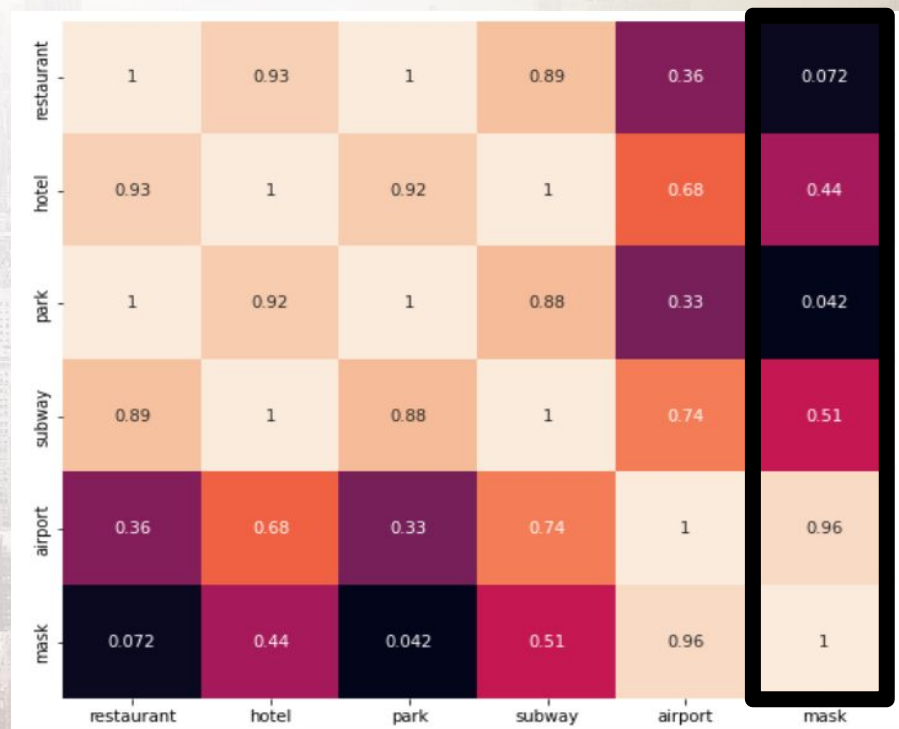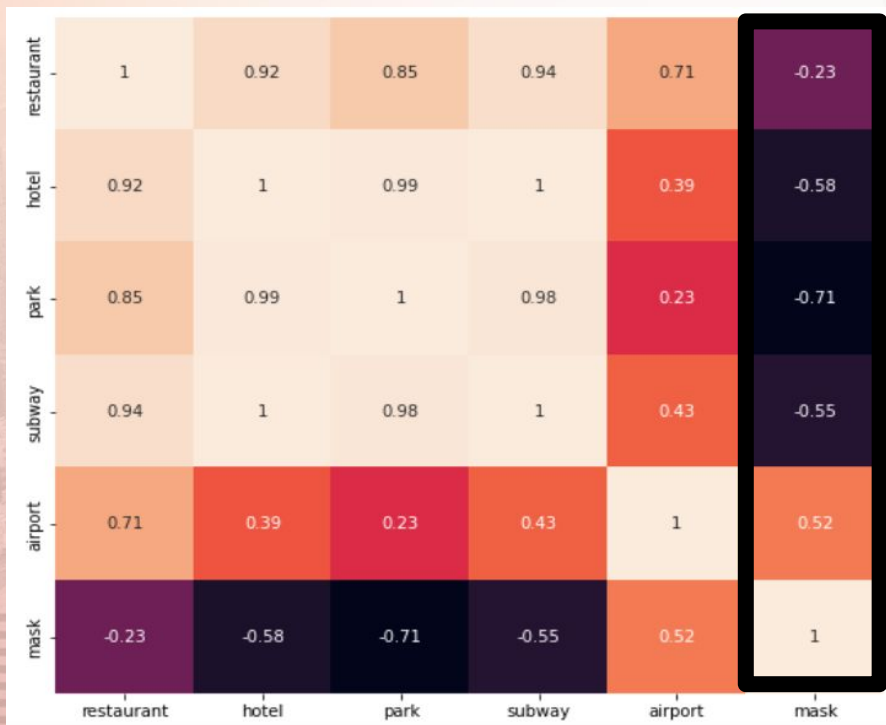| Topic 1 | **Quarantine** | quarantine, vaccine, rule, may, 2021, mask |
Topic 2	**Accomodation**	restaurant, food, hotel, great, bar, room
Topic 3	Leisure	park, walk, central, time, street, brooklyn, jazz, date
Topic 4	**Crime**	crime, people, homeless, like
Topic 5	**Travel**	car, flight, airport, jfk, parking, subway, train, people, station

Embedder Matrix

Topic 1 : Quarantine
- y=0: ["Quarantine", "Travel", "Vaccine"]

Topic 2: Accomodation
- y=1: ["Restaurant", "Bar", "Food"]

X: Top 3 words from all 5 topics

Cosine Similarity of selected attributes generated from Embedder matrix trained on Topic 1 (Left) & Topic 2 (Right)

Recommendations

	Instagram Topics					
	City Scape	**Natural Landscape**	**City Events**	**Traffic**	**People**	**Riverside**
Travel Blogs	0.56	0.40	0.51	0.43	0.49	0.52

	Blog Topics					
	Quarantine	**Hotels & Accomodation**	**Train and Subway**	**Leisure**	**Crime**	**Air Travel**
Instagram	0.40	0.57	0.57	0.60	0.47	0.52

- **Tourists generally associate and are more interested in modern city look than the natural scenery of NYC.**

THANK YOU