Group Assignment 2: Topic Modeling and Image Analytics

Introduction:

We chose to analyze the firm GoPro's Twitter uploads, as they have a large variety of photos with different levels of engagement. Go Pro manufactures action cameras and develops its own mobile apps and video-editing software. In the following analysis, we will see what specific photos themes create more engagement.

Task A:

We first performed topic modeling (LDA) on the images labels scraped from Twitter (as Instagram scrapping was not possible), with the extraction of 5 topics as a starting base. We were able to identify different topics by looking at similarities between the words. However, for testing purposes, we ran the LDA script using a different number of topics to explore better categorization.

After looking at the output for LDA with 3 topics, we found that there were no clear patterns between the words that could lead to a specific definition of topics.

With the attempt to do LDA for 4 topics, we ran into a different problem where 3 out of the 4 topics had a clear theme, however, one of the topics did not have a clear theme and was just a combination of varied keywords.

The fourth try with 6 topics gave us 6 different themes that were clearly identifiable. We realized this was a better clustering compared to the previous iterations because the topics were clear and easily differentiated. We identified this to be the best model to use for further analysis.

Then, we extracted the top 25 words for each of the topics. The realized images in topic 1 were linked to *lifestyle*, in topic 2 they were for *sports with wheels*, images in topic 3 were linked to *winter*, topic 4 images were linked to *marine life*, photos in topic 5 depicted *aqua sports*, and topic 6 was for *landscape* images.

In the next question, we will try to determine which of these topics increases engagement in pictures and which ones do not.

<u>LDA output with top 25 words for 6 topics</u>

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
sky	bicycle	slope	water	sky	landscape
flash	tire	mountain	blue	water	natural
people	automotive	sky	underwater	equipment	water
photography	wheel	snow	azure	cloud	plant
nature	vehicle	recreation	electric	blue	sky
cloud	equipment	equipment	diving	nature	cloud
water	sky	cap	liquid	travel	mountain
happy	helmet	ice	fluid	light	tree
leisure	lighting	outdoor	recreation	boat	nature
dog	sports	cloud	marine	recreation	landforms
fun	motorcycle	sports	biology	watercraft	formation
atmosphere	cloud	freezing	art	vehicle	coastal
dusk	supplies	nature	darkness	leisure	oceanic
sunlight	bicycles	glacial	fin	boating	azure
carnivore	frame	landform	organism	boats	people
landscape	plant	sport	aqua	supplies	highland
phenomenon	design	landscape	event	daytime	resources
travel	motor	ski	phenomenon	air	ecoregion
afterglow	asphalt	terrain	vertebrate	lake	lake
atmospheric	land	winter	fish	building	body
gesture	tree	people	body	azure	cave
horizon	rolling	highland	font	sports	watercourse
astronomical	handlebar	travel	wave	outdoor	bedrock
object	skateboard	mountainous	wind	atmosphere	light
space	car	world	circle	wind	grass
Lifestyle	**Wheel Sports**	**Winter**	**Marine Life**	**Aqua Sports**	**Landscape**

Task B.
Landscape is the most prominent theme in the top 25% of the posts with marine life being the second. Wheel sports and lifestyle themes come next at around 14.3% each followed by aqua sports and winter theme.

On the other hand, landscape also has the highest average topic weights in the lowest quartile or the worst-performing posts. However, complementary to the trend identified in the highest quartile, winter and aqua sports have higher weights in the lower quartile. Themes of lifestyle, marine life, and wheel sports are less significant in the lower quartile whereas they were identified to be more prominent in the higher quartile of the posts.

<u>Rankings of topics in high and low quartiles</u>

Topics	High Quartile		Topics	Low Quartile
Landscape	0.27277285		Landscape	0.2341045
Marine Life	0.17998761		Winter	0.2007014
Wheel sports	0.14385127		Aqua sports	0.1681132
Lifestyle	0.14355846		Lifestyle	0.1577901
Aqua sports	0.13133787		Marine life	0.1310001
Winter	0.12849194		Wheel sports	0.1082907

Task C:
Based on the results above, there are a few key insights that we can generate.

Right off the bat, we can note that *landscape* has the highest topic weights for both quartiles. This makes intuitive sense, since at a glance, the majority of GoPro's Twitter photos appear to depict some form of natural setting. The weight here thus appears to be driven more by quantity rather than consumer preference. There is no clear indication that GoPro should diversify their photos away from this topic, and as *landscape* aligns well with GoPro's target consumers and main use case of recording sport and physical activity, we would recommend that GoPro does not deviate from this trend.

As for the rest of the topics, there is a decently clear trend as to which topics are more popular in terms of engagement (in the form of comments).

Marine life and *wheel sports* both had high weights in the top quartile, while having lower weights in the lowest quartile of posts. This may indicate that photos depicting either of those topics generally garner higher engagement rates. We would recommend that GoPro utilizes this as a basis for photos moving forward, either looking to get more photos in this topic area or concentrating major marketing campaigns using these topics, in conjunction with the landscapes mentioned above.

By contrast, *winter* and *aqua sports* yield opposite results, having higher weights in the lowest quartile and lower weights in the highest quartile. This is an interesting find, as winter sports (such as skiing) are one of the company's main recommended use cases of the GoPro. Consequently, reducing the number of photos in this topic category would not be our main recommendation, as we believe that such photos are key to the firm's brand. Instead, we would urge that the firm conducts further analysis on the quality of the photos, the time when the post was made (seasonality, time of day, etc.) and the activities portrayed to see if we could get any deeper clues as to what could be done to increase overall engagement and activity.

Appendix

5 Topics

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
water	recreation	bicycle	water	sky
landscape	equipment	tire	blue	nature
sky	slope	automotive	underwater	people
cloud	sky	wheel	azure	cloud
natural	snow	vehicle	electric	plant
mountain	sports	sky	diving	water
landforms	outdoor	equipment	liquid	landscape
lake	mountain	helmet	fluid	flash
formation	cap	lighting	world	photography
azure	ice	motorcycle	marine	tree
nature	cloud	bicycles	sky	natural
plant	freezing	frame	biology	happy
boat	glacial	supplies	recreation	leisure
travel	landform	design	art	mountain
coastal	sport	sports	phenomenon	travel
oceanic	helmet	motor	building	dog
terrain	leisure	plant	fin	sunlight
highland	ski	land	light	daytime
supplies	winter	cloud	wind	highland
vehicle	rolling	handlebar	organism	grass
watercraft	skateboard	car	equipment	blue
bedrock	geological	crankset	wave	terrestrial
cave	personal	gear	darkness	ecoregion
resources	protective	asphalt	aqua	fun
tree	smile	hood	event	carnivore
Nature	**Activities**	**Motor sports**	**Water**	**lifestyle**

4 Topics

Topic 1	Topic 2	Topic 3	Topic 4
water	sky	bicycle	sky
blue	slope	tire	landscape
sky	mountain	automotive	water
recreation	snow	wheel	cloud
azure	recreation	vehicle	natural
electric	cloud	underwater	plant
equipment	equipment	equipment	nature
cloud	cap	helmet	mountain
leisure	ice	sky	people
nature	sports	diving	tree
travel	outdoor	lighting	landforms
photography	dog	marine	azure
flash	freezing	biology	formation
liquid	travel	motorcycle	light
people	glacial	sports	sunlight
vehicle	landform	design	highland
lake	landscape	bicycles	phenomenon
supplies	mountainous	frame	lake
watercraft	sport	supplies	coastal
sports	terrain	plant	oceanic
wind	ski	fin	atmosphere
boat	winter	motor	flash
boating	tree	land	photography
boats	carnivore	water	resources
fun	people	organism	ecoregion
?	**winter**	**sports**	**nature**

3 Topics

Topic 1	Topic 2	Topic 3
sky	equipment	water
automotive	bicycle	sky
vehicle	sky	landscape
tire	sports	nature
cloud	snow	cloud
water	recreation	natural
lighting	slope	people
blue	wheel	plant
dog	mountain	azure
boat	tire	mountain
equipment	helmet	blue
light	cap	underwater
motorcycle	ice	landforms
watercraft	cloud	tree
design	outdoor	recreation
travel	supplies	lake
boating	freezing	flash
boats	bicycles	photography
atmosphere	frame	highland
wheel	glacial	electric
air	landform	happy
motor	sport	body
plant	ski	leisure
world	winter	formation
supplies	plant	travel
	No clear segmentation	