FINAL PROJECT - CS001

VIETNAMESE STREET FOOD

An analysis on Vietnam street food media data

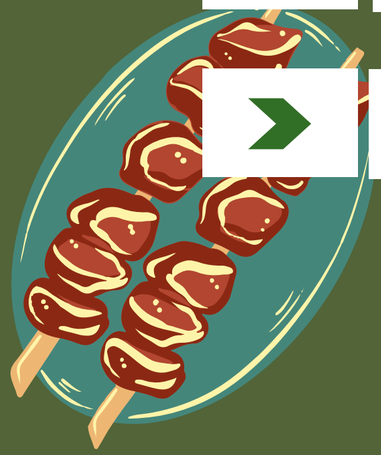
INSTRUCTOR : PHAN THANH TRUNG

Presentation by

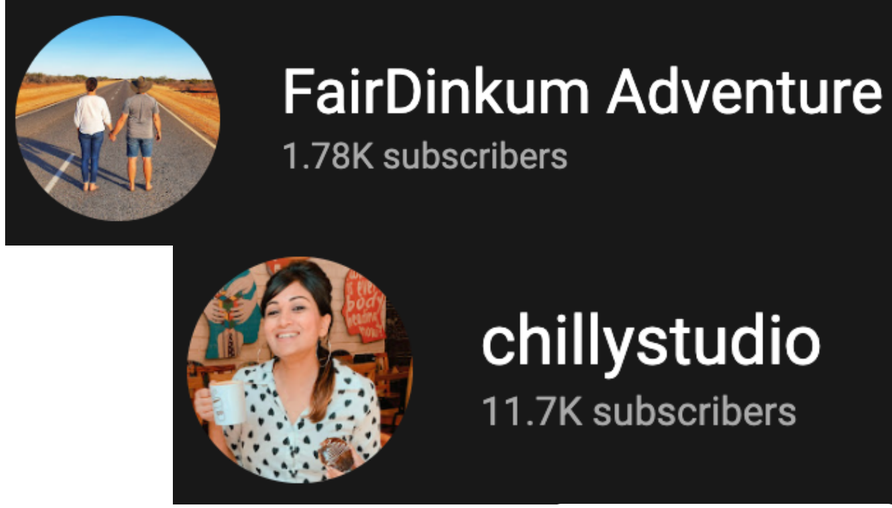
GROUP 1

AGENDA

1. Introduction
2. Needs statement
3. Goal and objectives
4. Literature Review
5. Methodology
6. Results
7. Limitations
8. Conclusion & Discussion



INTRO DUCTION



- Video analysis has been a part of computer science for a long time.
- It is an interdisciplinary field that deals with getting information from videos and figuring out what it means.
- We will analyze 2 Vietnamese food tour videos from **FairDinkum Adventure** and **chillystudio** Youtube channels.

Needs statement

The importance of Street food

- A prominent feature.
- Most accessible type of food.
- The cheapest.
- Has grown significantly.
- 80% of all restaurants being street vendors.
- 8 million tourists annually.
- Rare meals and tastes.
- Some street meals, like pho, are exported.
- Many restaurants abroad provide these popular meals.

Goals and Objectives

Technical skills

- Analyze scenes of films using statistics on string.
- Compare AI results with annotated scenes.
- Do descriptive analysis.
- Learn how to practice data types in a practical setting.
- Comprehend how Youtube data is stored and processed in terms of visual characteristics.
- Consider the use of CS1 knowledge in the context of comprehending street food films.

Goals and Objectives

Cultural understandings

- To make viewers aware of the benefits of Vietnamese street food.
- To promote the importance of healthy eating and to show that street food can be a healthy option.
- To encourage viewers to visit Vietnam and try this cuisine in person.

Design constraints and feasibility

Too many frames

The video contains a lot of under 1-second frames so it was hard for us to catch up with the flow and specific frames for the output.

=> We changed the speed to 0.5 for a clearer look

Time

A major issue for us who have to balance our studies with other commitments.

=> Set up timelines and strictly follow the group meeting instructions

Confused Instructions

It's hard to understand the instructions at first hand.

=> Ask for professor's advices to make things clear

Literature Review

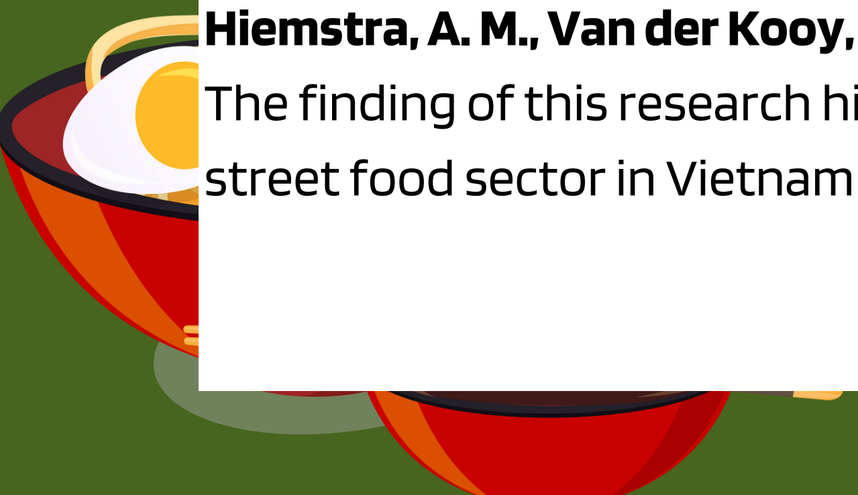
The social life of street food: exploring the social sustainability of street food in Hanoi, Vietnam Stutter, Natalia 2017.

This research used a conceptual framework to highlight the development areas in social sustainability of street food in Hanoi. The challenges and social functions that social vendors have to face are also focused in the findings of research.

Entrepreneurship in the street food sector of Vietnam—Assessment of psychological success and failure factors.

Hiemstra, A. M., Van der Kooy, K. G., & Frese, M. (2006).

The finding of this research highlight the entrepreneurship factors in the micro business of street food sector in Vietnam with 102 vendors from Hanoi and Hue.



Literature Review

Food safety knowledge, attitudes and practices of street food vendors and consumers in Ho Chi Minh city, Vietnam.
Samapundo, S., Thanh, T. C., Xhaferi, R., & Devlieghere, F. (2016).
This research highlights the key points of the unclean conditions, the food safety knowledge of consumers and food vendors in Ho Chi Minh city based on their age, educational level, gender, etc.

The impact of sensory marketing on street food for the return of international visitors: Case study in Vietnam.
Hoang, D. S., & Tučková, Z. (2021).
The research emphasizes the empirical evidence to help us know the impact of sensory marketing and the revisit decisions to build the marketing strategies of HCM tourism.

Literature Review

TOURISTS'EMOTIONAL RESPONSES TO STREET FOOD EXPERIENCES IN VIETNAM

Linh, P. L. D. (2021).

The research focuses on the emotional responses of the international tourists when experiencing the street food in Vietnam. The culture, social values and food are the factors that attract the tourist and highlight the role of the emotion in their answers.

Methodology

We divided each task into small step to deal with the given problems.

Start_time (seconds)	end_time (seconds)	duration (seconds)	food_focus_existing (1 or 0)	Youtuber_existing (1 or 0)	eating_place_existing (1 or 0)	description
0	7	7	0	0	0	the scene of the street and street vendor
8	9	1	0	0	1	Scene of a restaurant on the street
10	12	2	1	0	1	The scene of a woman makes "Bánh mì"
13	18	5	1	0	1	A woman makes "Bún" on the street.
19	21	2	1	0	0	Scene of a fruit cart
23	27	4	0	0	0	Scene of cars and motorbikes running on the road
28	29	6	0	0	1	A woman sells food for guest on the street.
30	34	4	1	0	0	Scene of a sticky rice cake
35	42	7	1	0	1	Scene of egg coffee cup at Giang coffee
43	47	4	0	0	0	Scene of the street

Part 1: Prepare for the annotated scenes on the two assigned videos

Methodology

Part 2: Working on the VTT files

- Working with **json** to load the LIWC dictionary and voice-to-text (VTT) files.
- Eliminate the duplicated part in the scripts by working with list and dictionary and store it in a list
- Calculate the distribution of a list of categories in the LIWC dictionary appeared in the transcript and then select the top 5 categories to evaluate and discuss on the accuracies and practicalities.

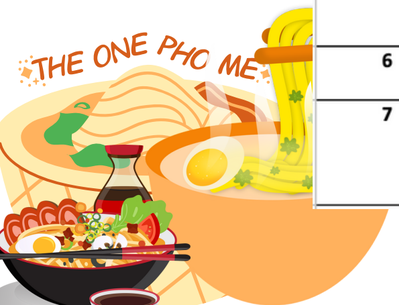
Category	Percentage
Function	0.464531
Verb	0.152174
Pronoun	0.118993
Prep	0.117849
FocusPresent	0.117849
Social	0.091533
Relativ	0.088101
CogProc	0.086957
Auxverb	0.081236
Ppron	0.078947

Methodology

Part 3: Compute the descriptive statistics on the data set by working with the visual file and annotated file.

- Calculate the percentage of existence of three categories annotated (food existing scenes, Youtuber_existing scenes and existing place scene by working with **pandas**.
- Use **json** to load the VTT file then sort the values of each visual feature by dealing with a list of dictionaries and select the top 3 features to compare and analysis.

	Start_time (seconds)	end_time (seconds)	duration (seconds)	food_focus_existing (1 or 0)	Youtuber_existing (1 or 0)	eating_place_existing (1 or 0)	description	max_1	max_1_value	max_2	max_2_value	max_3	max_3_value
0	0	26	26	0	1	0	scene of the man talking	person;individual;someone;somebody;mortal;soul	7,087219238	road;route	0,5935194227	person;individual;someone;somebody;mortal;soul	2,640726725
1	27	46	19	0	1	0	scene of the woman talking	plate	1,891825358	table	1,942016602	plate	0,244676378
2	47	56	9	0	1	0	scene of the couple talking	person;individual;someone;somebody;mortal;soul	4,584065755	wall	2,774658203	person;individual;someone;somebody;mortal;soul	0,03565131293
3	57	60	3	1	1	1	scene of inviting people to eat silkworm	tray	1,331115723	table	0,2655029297	tray	0,1868218316
4	61	67	6	1	1	1	scene of the man dare his friend to eat	person;individual;someone;somebody;mortal;soul	2,497097439	wall	1,457784017	person;individual;someone;somebody;mortal;soul	0,3770480686
5	68	68	0	1	0	0	scene of the shrimp cake	person;individual;someone;somebody;mortal;soul	0,6145494249	wall	0,3079020182	person;individual;someone;somebody;mortal;soul	0,0426296658
6	69	71	2	1	1	1	scene of eating shrimp cake	tray	0,8646104601	table	0,3686930339	tray	0,7117513021
7	72	73	1	1	1	1	compliment the food	person;individual;someone;somebody;mortal;soul	0,6112738715	wall	0,3685641819	person;individual;someone;somebody;mortal;soul	0,05602349175



RESULTS & DISCUSSION

1. Script features
2. Visual features
3. Further discussion

Script features

TOP 10

Video 1

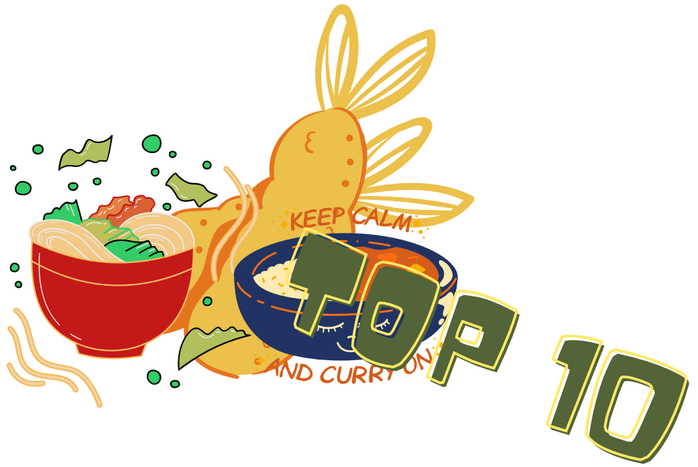
Category	Percentage
Function	0.464531
Verb	0.152174
Pronoun	0.118993
Prep	0.117849
FocusPresent	0.117849
Social	0.091533
Relativ	0.088101
CogProc	0.086957
Auxverb	0.081236
Ppron	0.078947

TOP 10

Video 2

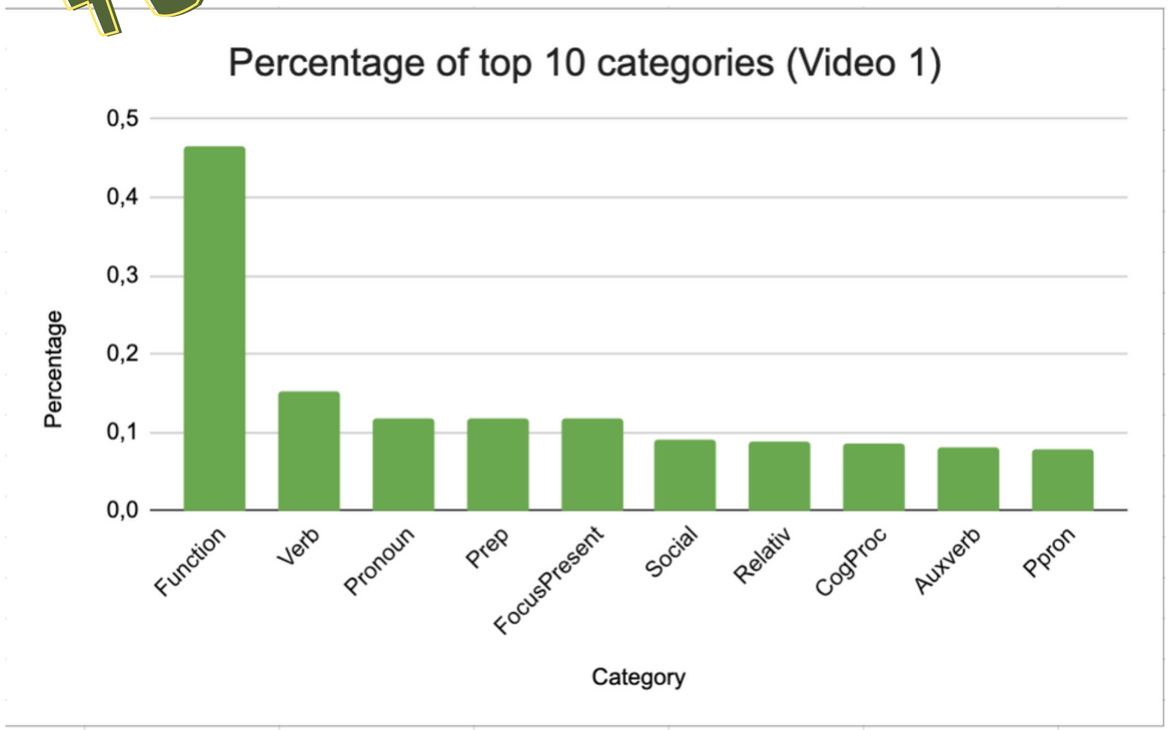
Category	Percentage
Function	0.527126
Verb	0.182996
FocusPresent	0.159514
Pronoun	0.153846
Relativ	0.110121
Auxverb	0.108502
CogProc	0.106073
Social	0.093117
Prep	0.090688
Ipron	0.086640

Script features

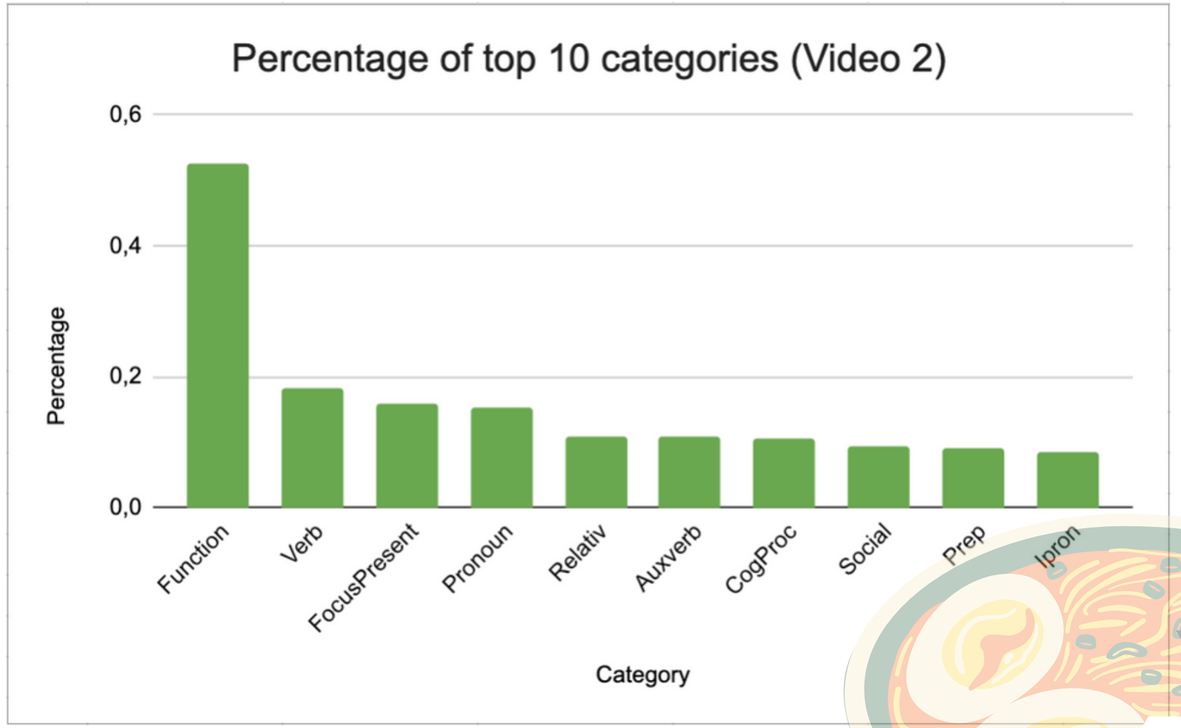


TOP 10

Video 1



Video 2



Script features

In both videos, function words and verbs contains the highest percentage

Top 1: Function words

Approximately ½ of the scripts focus more on **the images, emotions and informal words** instead of ideas formed by formal sentence

Top 2: Verb

Normal/ all sentences mostly contain verbs. Verbs seem equal to the number of **pronouns**. So it looks like **most sentences contain one verb.**

Video 1+2
Script features

Top 3 - 4 - 5 (Video 1): Pronoun (11%), Prep (11%), FocusPresent (11%)

Top 3 - 4 - 5 (Video 2): FocusPresent (15%), Pronoun (15%), Relativ (11%)

Many sentences have the **pronouns**
Percentage of pronouns represents for **beginning a statement.**

“relativ” in video 2 -> talked much about the relations, maybe they are traveling to many places. So they talks too much with **“relativ”** words.

'**FocusPresent**' : most of the sentences contains one words representing for present -> mostly talk about the present

"prep" words show both relations in sentences and the positions, so this feature don't reflect too much

SCRIPTS FEATURES

What top 10 scripts CATEGORIES in video 1 and video 2 let us know?

1. Mainly using functional words and verbs.

⇛ **use informal and action's language, focus more on images instead of explanation with long/full sentences** ⇛ **make the audience feel comfortable and follow the channel with daily vibes** ⇛ **increase profits/viewers/... for youtubers**

2. The use of many "prep" and "relative", and "focus present" can show its **changes of positions/actions** and the interpretation of **present time.**

⇛ **Attract the audience and announce the change for them to make them concentrate**

⇛ **increase the curiosity** ⇛ **i**

3. Others biggest categories seem not to show any specific details of the meaning.

VISUAL FEATURE

Video 1

Number of scene: 54

Average time per scene: 15.4

Percentage of food existing over all 42% (manual)

Percentage of Youtuber existing over all 33%

Video 2

Number of scene: 58

Average time per scene: 14.18

Percentage of food existing over all 46% (manual)

Percentage of Youtuber existing over all 63%

The percentage of Youtube in video 1 seems lower than our expected .

We will use these number to anaylze in the later parts with focusing on 4 elements : Color, scenes, objects, and places

Colors

- The colors analyzed by the computer are **mostly correct**
- For example, the couple in the video both wearing grey clothes, and the grey appears most of the time when the YouTuber appears in the video (AI data seems the same as with the annotated scene)

gray16	7565	gray20	6619	gray17
gray100	26465	LightSlateGray	7737	maroon
SkyBlue4	62299	LightSkyBlue4	59290	burlywood4
gray100	26959	gray7	21869	gray6
gray70	4771	gray69	4217	gray100
LightSalmon4	23416	gray100	29913	gray20
gray8	94024	gray9	118868	gray7
gray64	22866	gray63	17121	gray65
gray64	47765	gray63	49681	gray6
gray65	8701	gray69	8042	gray63
gray0	493126	gray1	48719	gray2
gray100	12320	GhostWhite	8857	AliceBlue
gray14	33974	gray15	29794	gray13
gray100	37912	gray99	10936	gray60

Objects

- The objects analyzed by the AI seem totally incorrect, strange because many object analyzed by the computer came from nowhere.
- For example, Ciconia nigra or Spoonbill have nothing related to the video.

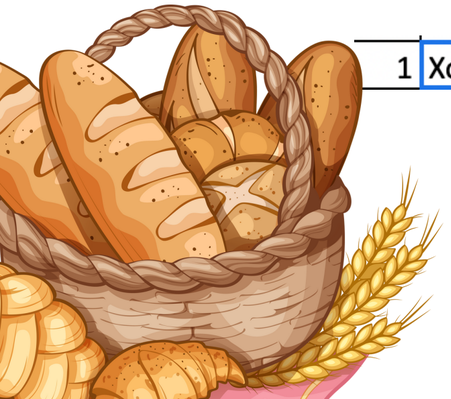
digital clo	1,230294	knot	0,073938	digital clo	0,057781
zucchini, (	0,234214	long-horn	0,063118	zucchini, (	0,045807
pay-phon	0,807081	chickadee	0,038133	pay-phon	0,040150
American	2,151604	great grey	1,139465	American	0,716040
reflex can	0,194964		0,378332	reflex can	0,197206
isopod	0,158582	pop bottl	0,050784	isopod	0,043149
great grey	0,405143	American	0,324386	great grey	0,119824
chest	2,314523	pole	4,438098	chest	0,485102
racket, ra	0,358609	axolotl, m	0,090618	racket, ra	1,429866
Crock Pot	0,872938	waffle iro	0,307459	Crock Pot	0,042125
racket, ra	0,190068	American	0,096092	racket, ra	0,087616
jacamar	0,577965	harvestm	0,257819	jacamar	0,129742
spotted sa	0,447583	rotisserie	0,402956	spotted sa	0,053876
dam, dike	0,190435	typewrite	0,139479	dam, dike	0,899967
black stor	0,241116	spoonbill	0,223948	black stor	0,127431

robin, Am	0,234178	chickadee	0,147408	robin, Am	0,165954
long-horn	0,998354	lionfish	0,041200	long-horn	0,046507
mosque	0,164842	isopod	0,339125	mosque	0,066327
pole	1,425307	sombrero	0,036474	pole	0,020412
pole	2,159266	chest	0,425396	pole	0,490619
lacewing,	0,070625	leatherba	0,063767	lacewing,	0,022357
pole	0,529978	chest	0,186859	pole	0,129262
zucchini, (	0,320495	jersey, T-s	0,041362	zucchini, (	0,042835
chest	0,814938	American	0,125512	chest	0,117933
chest	0,459420	pole	0,411309	chest	0,280377
horizonta	0,299808	pole	1,172249	horizonta	0,117974
Komodo d	0,422067	dam, dike	0,026104	Komodo d	0,017807
Komodo d	0,474880	broccoli	0,013003	Komodo d	0,031149
knot	0,282187	goblet	0,145204	knot	0,105101
sombrero	1,974063	pole	0,418018	sombrero	0,159762

Objects

- Most of the food focusing scene was identified with a strange object, which is not food.

1	The scene of a woman makes "Bánh mì"	slide rule, slipstick
1	Scene of a fruit cart	Gila monster, Heloderma suspectum
1	Xoi Yen's staff brings Xoi to the youtuber's table	great grey owl, great gray owl, Strix nebulosa



Places

- Many objects which was analyzed by AI are correct
- However, there are some unreasonable places like Jail_cell

beauty_sa	0,940579	jail_cell	0,210809	office	0,118259
restauran	0,478649	galley	0,140469	kitchen	0,137870
shopfront	0,977049	bookstore	0,371629	general_s	0,170259
beauty_sa	3,169679	jail_cell	0,274719	bookstore	0,222700
booth/inc	1,072150	desert/ve	0,091990	rice_padc	0,049109
food_cou	0,125630	bazaar/ou	0,222289	arcade	0,206779
crosswalk	0,717020	street	0,399109	gas_static	0,080479
booth/inc	0,858709	fastfood_	0,694150	art_schoc	0,122310
fastfood_	1,431859	booth/inc	2,036649	beauty_sa	4,489649
beauty_sa	1,109980	kindergar	0,063969	locker_ro	0,044560
art_schoc	0,998610	classroom	0,343490	sushi_bar	0,321159
street	1,869219	crosswalk	1,028090	gas_static	2,549040
street	1,169259	gas_static	0,465860	alley	0,270019
crosswalk	1,012499	street	0,499249	fastfood_	0,084890
beauty_sa	0,392339	jail_cell	0,544309	hospital_	0,025150

street	1,869219	crosswalk	1,028090	gas_static	2,549040
street	1,169259	gas_static	0,465860	alley	0,270019
crosswalk	1,012499	street	0,499249	fastfood_	0,084890
beauty_sa	0,392339	jail_cell	0,544309	hospital_	0,025150
slum	0,233460	beauty_sa	0,411330	medina	0,137179
street	3,787940	beauty_sa	5,519989	hardware	1,317560
beauty_sa	1,026610	arcade	0,520609	bazaar/ou	0,511039
street	0,196669	crosswalk	0,159220	gas_static	0,079399
beauty_sa	2,068100	medina	0,344819	bazaar/ou	0,461800
beauty_sa	4,449310	booth/inc	1,527779	pharmacy	0,503719
crosswalk	0,750830	street	1,304359	fastfood_	0,254400
street	0,857880	arcade	0,613520	alley	0,242549

Places

- There are many eating places that were identified wrong by the AI to be beauty_salon, martial_arts_gym,...

1	The youtuber introduces each type of cake in detail.	beauty_salon
1	food store called "chè ngon phố cổ"	beauty_salon
1	the egg coffee - lane (Giang cafe)	beauty_salon

Scenes

- The scenes analyzed by the AI are mostly correct.

person;in	12,13194	building;e	5,738979	tree	0,617574
person;in	5,983364	building;e	2,391472	fence;fen	0,088467
person;in	5,314785	wall	1,826409	floor;floo	0,137403
table	0,798468	person;in	0,754767	tray	0,136515
person;in	2,735649	wall	0,910291	plate	0,799567
wall	0,687445	person;in	0,142591	stairway;s	0,074951
wall	0,797960	person;in	0,985256	table	0,711642
person;in	1,049506	wall	0,537563	table	0,194315
person;in	0,821187	wall	0,698940	table	0,185024
person;in	1,513149	table	0,727010	wall	1,378804
person;in	2,275051	wall	1,725694	table	0,890591
person;in	0,883985	wall	0,602233	tree	0,133843
person;in	1,240464	building;e	0,963880	tree	0,186218
road;rout	0,650187	flower	0,290283	plant;flor	0,097215
plate	1,817843	table	2,351447	wall	3,062554
wall	1,209845	floor;floo	0,297776	person;in	1,699245
building;e	0,978000	minibike;r	0,189744	road;rout	0,874138

building;e	1,847683	road;rout	0,880818	person;in	2,700948
person;in	1,122483	cabinet	0,205579	wall	0,913106
person;in	2,775729	building;e	1,487209	sky	0,191901
person;in	25,31435	building;e	10,67085	tree	1,451178
person;in	4,432217	sidewalk;p	0,819315	building;e	2,944268
building;e	0,709838	minibike;r	0,225212	road;rout	0,102566
person;in	5,463840	building;e	2,707621	sky	0,144490
person;in	15,91139	wall	4,690707	table	2,255316
building;e	2,240953	road;rout	0,665045	person;in	3,023173
building;e	7,687872	road;rout	0,622382	sidewalk;p	1,214986
wall	1,481065	person;in	1,221950	shelf	0,190646
sky	2,009670	wall	5,093641	mountain	0,089619
tree	0,303439	building;e	0,284342	sky	0,277398
person;in	3,347357	table	1,121554	wall	1,407579
wall	1,001295	person;in	4,095153	building;e	1,278272

Scenes

In video 1:

Scenes named "person;individual;someone;somebody;mortal;soul" appeared the most for 55% of annotated scene. The number of Youtuber appeared in annoted scene is 33%. **This number seems partially different from the annotated scenes**

In video 2:

Scenes named "person;individual;someone;somebody;mortal;soul" appeared the most for 57% of annotated scene. The number of Youtuber appeared in annoted scene is 46%. **This number reflects the accuracy of AI in realizing humans.**

The scenes' extracted by AI seem almost correct when compared with the annotated scenes. However, there still can be some errors

Further discussion

The using of "function", "verbs","relative,prep", and "focus present" words

Assumptions : The combination of (1) attractive, comfortable language made by function words, (2) the diverse changes and relations made by "relative,prep", and (3) the interesting announcement made by "focus present" will attract the audiences, make them feel curious, and highten the number of followers.

Scenes and colors is easily identified by AI.

Assumptions : AI can identify each pixel in color easily and the scenes' data set doesn't different in many different urban areas. **The big percentage of humans** represents for the attraction of the scene with humans in a food-oriented video.

Food (Objects) and Places

Assumptions : AI's data set isn't diverse enough to identify all objects in Vietnam, especially Vietnamese food. The places are more diverse but still limited which led to many wrong places.

CONCLUSIONS

The scripts feature analysis let us know some **word tips** to attract more audience with longer watching time

The visual feature analysis let us know about the visual features that built a video Youtube on street food. From that, analyzing each video will **evaluate the strength** which made the video become more attractive in the field of images

The comparison between AI and annotated let us know the difference between data extracted by annotated and AI. From that, we know the AI's limitations, raise awareness, and improve it in the future.

LIMITATIONS AND FUTURE RECOMMENDATION

The data set which were trained for AI is limited and need to be improved in both data set and AI model

Annotated data can be wrong because of the fast scene and human unconsciousness. Don't know how to avoid. Maybe we can try different model.

The data of the transcript is unclean which made us have to clean it before analysis. To avoid this wasting of time, we suggest the transcripts makers app be changed with our algorithm.

THANK YOU!