

Disneyland vs. Universal Studios

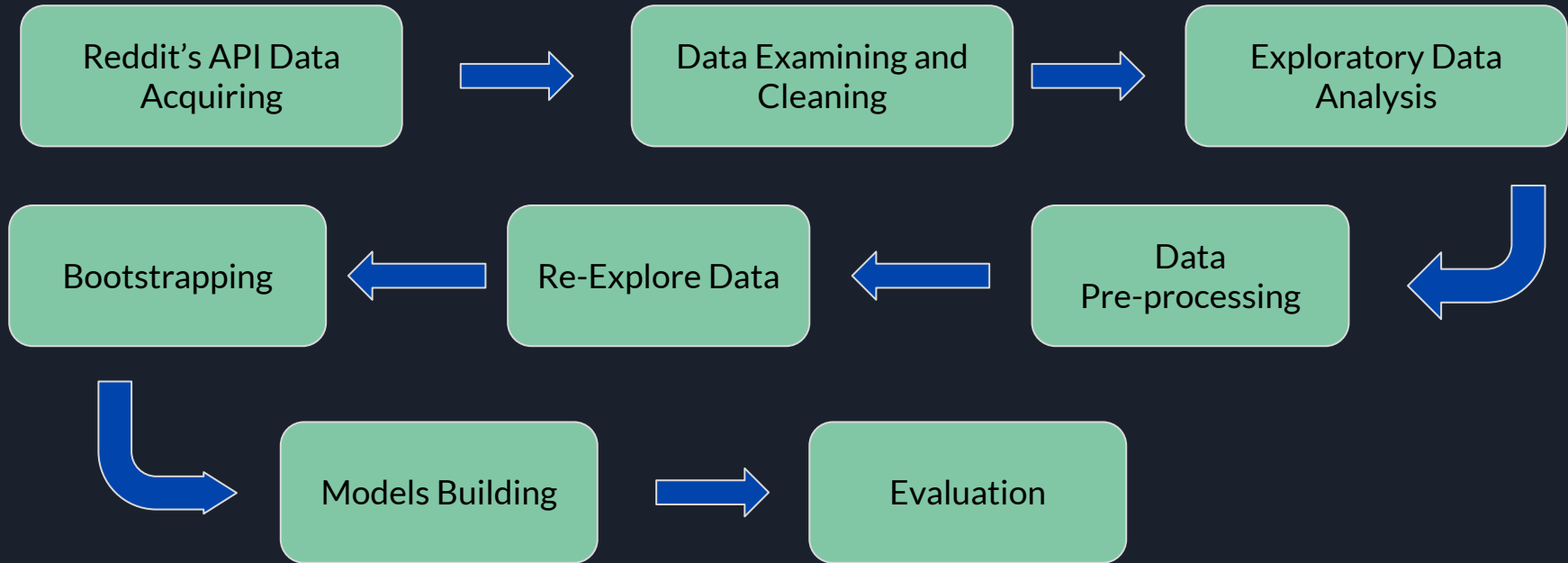


Problem Statement

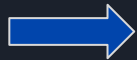


- KKday, the travel booking company, wants to identify the comment posts received from customers and classify them as Disneyland or Universal Studiosa
- To also understand the popularity of each amusement park based on comments received and the customer's view or sentiment.
- The goal: Push marketing/promotion tactics on the more popular theme park

Model Building Journey



API Webscraping



{JSON}

Data Examining and Cleaning

Data Summary

Disneyland

#	Column	Non-Null Count	Dtype
0	vote	3857 non-null	int64
1	title	3857 non-null	object
2	text	3857 non-null	object
3	date	3857 non-null	datetime64[ns]

Universal

#	Column	Non-Null Count	Dtype
0	vote	1501 non-null	int64
1	title	1501 non-null	object
2	text	1501 non-null	object
3	date	1501 non-null	datetime64[ns]

Imbalance Data

Data Span

Disneyland

2011-10-04 to 2017-09-29

Universal

2011-08-17 to 2023-11-04

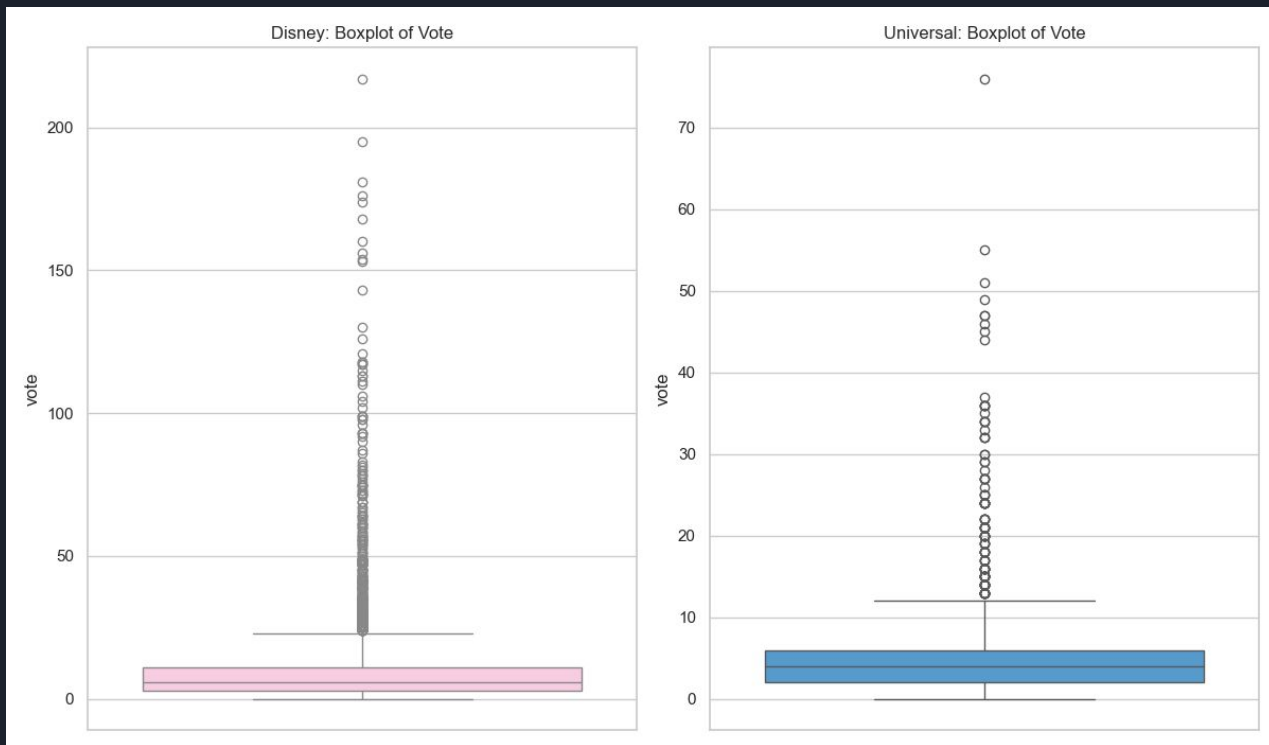
Data Duplicates

title	text
3857	3857
3846	3857

title	text
1501	1501
1497	1501

Data Examining and Cleaning

Vote Analysis



Exploratory Data Analysis

Top Vote Content Analysis: Disneyland

- Policy Update
- Ride, event, and attraction update
- Interesting Stories

	title	text	vote
3132	Can confirm: new Disneyland parade policy in a...	Was at Disneyland Tuesday night and castmember...	217
1063	I am 17 years old. Last Dec. I went to Disneyl...	I have never felt so happy about something.\n\...	195
3722	Fire on big thunder trail yesterday.	Yesterday I was walking through big thunder tr...	181
2377	Thank you to the kind stranger.	On Thursday night I was trying to get my daugh...	176
3805	PSA to everyone wishing to ride GotG: Monster'...	Please do not film the ride! The ride itself i...	174
3175	Pirates of the Caribbean Officially Open!	Was at the park for open and was heading to Bi...	168
3209	Changes/updates on Pirates! NSFW	I saw people in the subreddit saying that ther...	160
1341	Dear Cast Members of the Disneyland Resort,	You guys did it! \n\n\nThank you for all that ...	156
2066	Today is the 15th anniversary of DCA, so I tho...	Opening Day was Feb. 8th, 2001! \n\n\nOriginal...	154
1541	Wanted to thank this sub.	A few days ago I made a post about being stuc...	153

Exploratory Data Analysis

Top Vote Content Analysis: Universal

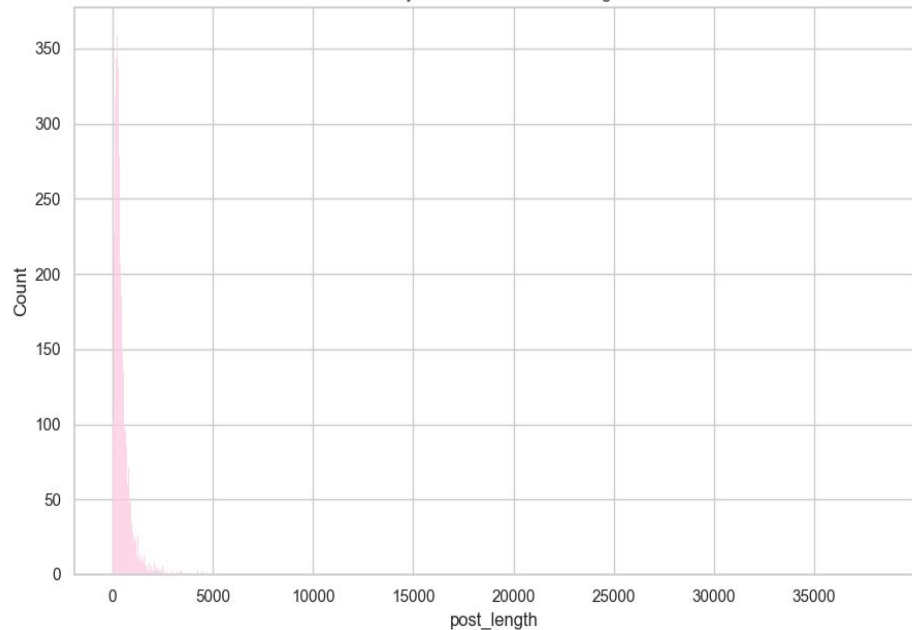
- Interesting comments from the visitors
- Interesting Stories
- Ride, event, and attraction update

	title	text	vote
748	Dont buy the tickets, save the money.	Me and my family members are currently at Univ...	76
329	Some stories of mine from working at Islands o...	So I worked at IOA for a few months, I would h...	55
183	Harry Potter and the Forbidden Journey's Loss...	For anyone visiting the park in the next few d...	51
768	Universal Studios Japan Super Nintendo World AMA	I was fortunate enough to be able to go to the...	49
371	Thanks you to all the universal studios Hollyw...	I want to start by saying I am blind and when ...	47
1405	Toadstool Cafe is weirdly unaccommodating.	I'm a huge Nintendo fan, so I absolutely love ...	47
1210	Pass Holder Preview Registration for Super Nin...	Please check the Passholder website for update...	46
645	JURASSIC WORLD FULL RIDE SOFT OPENING NIGHT TI...	https://youtu.be/jVGZcaAb8HQ \n\n\n​\n\n...	45
504	Fast and the Furious, sorry universal what wer...	anyone else think the new Fast and Furious rid...	44
611	i have an interview on tuesday for ride operat...	i'm extremely excited for the opportunity but ...	37

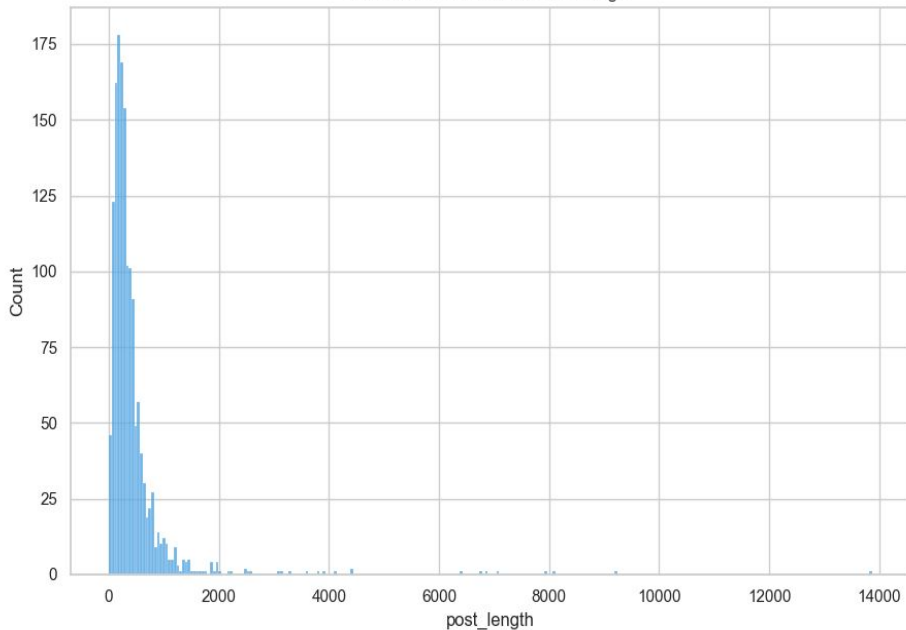
Exploratory Data Analysis

Post Length Analysis

Disney: Distribution of Post Length



Universal: Distribution of Post Length



Top Word Count Analysis

1-gram

Disneyland

	Count
disneyland	2451
just	1624
park	1526
time	1446
day	1348
know	1321
going	1305
disney	1272
like	1246
ve	974
really	754
ride	741
thanks	737
trip	718
want	700

Universal

	Count
universal	859
park	711
day	624
time	520
just	492
going	486
pass	481
ride	473
know	435
like	429
express	316
tickets	313
studios	307
want	289
ve	282

2-gram

Disneyland

	Count
going disneyland	200
does know	195
california adventure	187
cast member	157
annual pass	140
space mountain	133
cast members	126
haunted mansion	124
world color	118
downtown disney	117
day park	116
don know	114
year old	113
star wars	110
park hopper	108

Universal

	Count
universal studios	286
harry potter	178
express pass	141
studios hollywood	97
nintendo world	82
universal orlando	79
going universal	78
does know	76
super nintendo	67
annual pass	66
horror nights	59
express passes	56
thanks advance	52
theme park	51
universal hollywood	46

3-gram

Disneyland

	Count
http imgur com	55
day park hopper	52
disneyland california adventure	39
star wars land	35
happiest place earth	27
mickey halloween party	26
disney california adventure	26
radiator springs racers	26
want make sure	24
wait disney world	23
new orleans square	22
does anybody know	21
big thunder mountain	19
going disneyland time	18
park hopper tickets	17

Universal

	Count
universal studios hollywood	95
super nintendo world	61
halloween horror nights	34
going universal studios	33
harry potter world	24
universal orlando resort	20
day park park	17
universal studios orlando	17
jurassic park ride	14
universal studios florida	14
world harry potter	13
pass holder previews	13
wizarding world harry	12
trip universal studios	11
universal studios japan	10



Sentiment Analysis

	neg	neu	pos	compound
subreddit				
Disneyland	0.03	0.84	0.12	0.50
universalstudios	0.03	0.86	0.11	0.41

Pre-processing

- Tokenizing
- Lemmatizing
- Stop Words Removal
- URL Removal
- Target Words Removal



- Column1: Tokenized Only
- Column2: Tokenized + Lemmatized
- Column3: Tokenized + Stop Words Removal
- Column4: Tokenized + Lemmatized + Stop Words Removal

- Tokenizing: Splits text data into single units
- Lemmatizing: Transform text data into root form

Exploratory Data Analysis

1-gram

Count		Count	
park	2313	park	1043
day	2083	day	941
time	1957	ride	880
just	1710	pass	785
know	1588	time	710
ride	1440	ticket	629
like	1332	know	510
want	1166	just	506
year	1103	like	456
pass	986	want	433
make	963	hollywood	430
trip	919	express	400
think	886	orlando	362
good	813	plan	337
really	774	think	312

2-gram

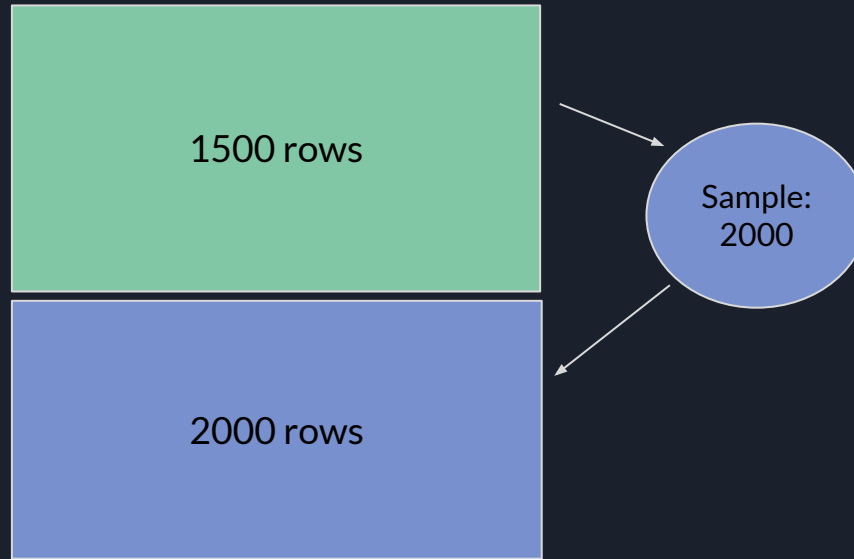
Count		Count	
cast member	340	express pass	241
annual pass	234	harry potter	214
california adventure	219	nintendo world	128
fast pass	188	annual pass	125
day park	181	studios hollywood	115
year old	164	super nintendo	104
space mountain	160	theme park	84
park hopper	154	wait time	71
world color	153	buy ticket	70
halloween party	142	day park	65
haunted mansion	141	halloween horror	59
feel like	134	day ticket	58
thank advance	132	thank advance	56
look like	123	jurassic park	53
star wars	113	horror nights	53

3-gram

Count		Count	
day park hopper	76	super nintendo world	95
mickey halloween party	53	halloween horror nights	36
star wars land	39	harry potter world	28
radiator springs racers	32	day park park	22
park hopper ticket	29	halloween horror night	22
want make sure	27	buy express pass	18
annual pass holder	26	jurassic park ride	18
new orleans square	24	harry potter fan	16
single rider line	24	world harry potter	15
park hopper pass	23	harry potter ride	15
happy place earth	21	express pass worth	14
big thunder mountain	20	park park ticket	14
thank advance help	20	single rider line	13
stay grand californian	18	jurassic world ride	13
halloween party ticket	18	super mario world	11

Bootstrapping

Universal





Baseline Accuracy

No Bootstrapped

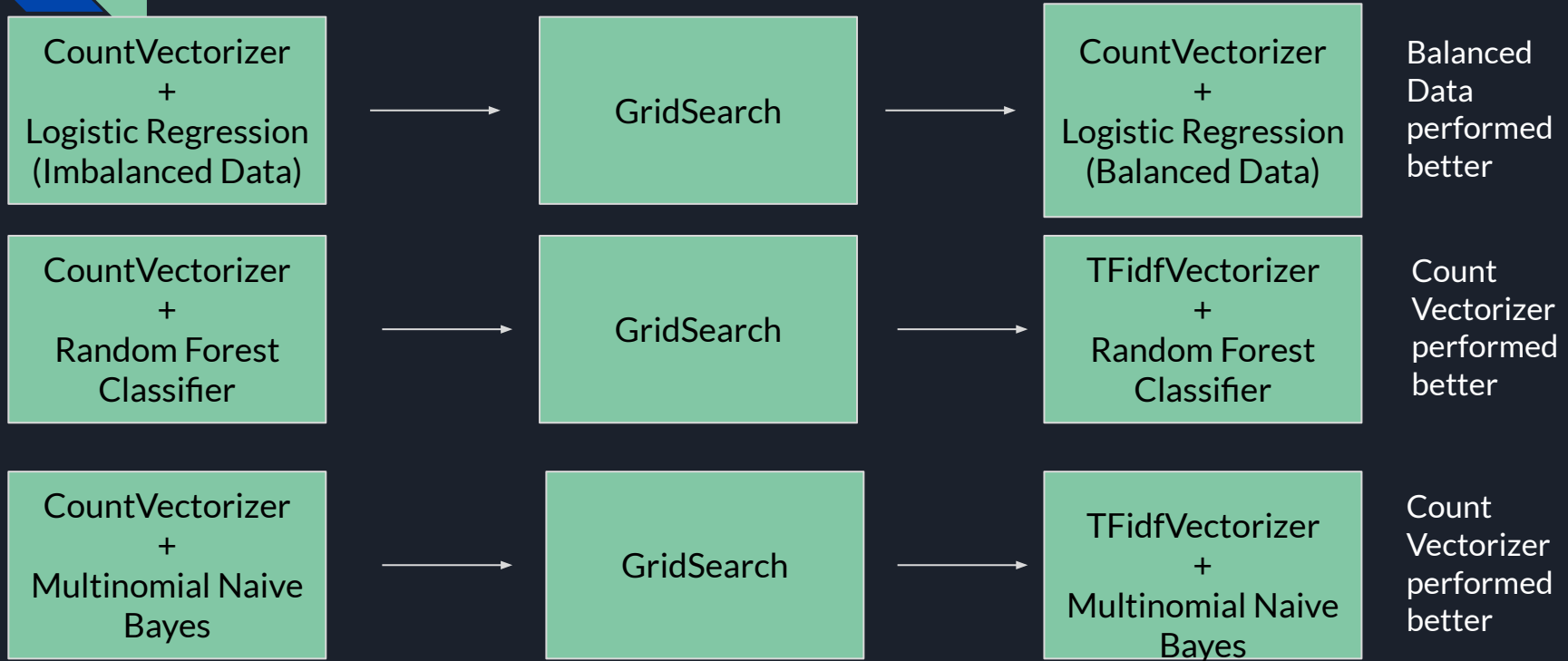
```
mapped_subreddit
1    0.72
0    0.28
```

Bootstrapped


```
mapped_subreddit
1    0.52
0    0.48
```

- 1: Disneyland
- 0: Universal

Modeling Process



Model Evaluation

Model	Accuracy	Precision	Recall	F1	% Incorrect Prediction
1. CV + Logreg	Train: 0.99 Test: 0.91	US: 0.90 DS: 0.91	US: 0.75 DS: 0.97	US: 0.82 DS: 0.94	9.27%
2. CV + Logreg(Boot)	Train: 0.99 Test: 0.95	US: 0.93 DS: 0.96	US: 0.96 DS: 0.94	US: 0.95 DS: 0.95	5.19%
3. CV + RF	Train: 1.0 Test: 0.96	US: 0.96 DS: 0.96	US: 0.95 DS: 0.97	US: 0.96 DS: 0.96	4.16% 
4- TF +RF	Train: 1 Test: 0.96	US: 0.96 DS: 0.95	US: 0.95 DS: 0.96	US: 0.95 DS: 0.96	4.45%
5. CV + NB	Train: 0.90 Test: 0.95	US: 0.95 DS: 0.91	US: 0.92 DS: 0.93	US: 0.92 DS: 0.93	4.46%
6. TF + NB	Train: 0.96 Test: 0.93	US: 0.91 DS: 0.94	US: 0.93 DS: 0.92	US: 0.92 DS: 0.93	7.33%



Incorrectly Predicted Posts

I have be look off and on for souvenir copy of the old A - e ride ticket , but I have n't find anything real promising yet . do anyone know if you can find those somewhere?anyone know where to find souvenir ride ticket ?

I plan on use my Hyatt point to stay at the Hyatt House right near the property , which seem like a good point redemption at 15,000 / free night . any other hotel recommendation ? I get plenty of point in both Hyatt and IHG right now .

my husband and I be head to for a family reunion this weekend ! very excited , we have n't do anything since our Honeymoon in FL !

we be a group who like fast ride and also a Character meet and greet .

- Most Incorrectly predicted posts are irrelevant



Recommendation

- CountVectorizer + Random Forest Classifier Model for classification
- Disneyland is more popular because of the higher posts obtained and higher votes received -> Higher engagement and public interest
- Disneyland vote content analysis overall is more positive/neutral
- Push more promotion strategies toward Disneyland: More popular -> Higher demand -> Increase customer satisfaction