

Determiner Noun: NUMBER

Michael Lai

How do parts of speech in a title
affect total revenue?

Data - 15,628 Movies

Title	Total Gross	TTheaters
The A-Team	77222099	3544
A.C.O.D.	175705	42
A.I. Artificial Intelligence	78616689	3242
Aaja Nachle	484108	66
Aarakshan (Reservation)	651096	91
Aatsinki: The Story of Arctic Cowboys	6601	2
Abandon	10727683	2347
The Abandoned	1331137	1000

Data - 15,628 Movies

Title	Total Gross	TTheaters	ADJ	ADV	ADP	CONJ	DET	NOUN	NUM	PUNCT
The A-Team	77222099	3544	0	0	0	0	1	2	0	1
A.C.O.D.	175705	42	0	0	0	0	0	1	0	0
A.I. Artificial Intelligence	78616689	3242	0	0	0	0	0	3	0	0
Aaja Nachle	484108	66	0	0	0	0	0	2	0	0
Aarakshan (Reservation)	651096	91	0	0	0	0	0	2	0	2
Aatsinki: The Story of Arctic Cowboys	6601	2	0	0	1	0	1	4	0	1
Abandon	10727683	2347	0	0	0	0	0	0	0	0
The Abandoned	1331137	1000	0	0	0	0	1	1	0	0

Features - POS

ADJ: Adjective - blue, dark

ADV: Adverb - quickly

ADP: Adposition - from, in, with

CONJ: Conjunction - for, and, or

DET: Determiner - the, this, that

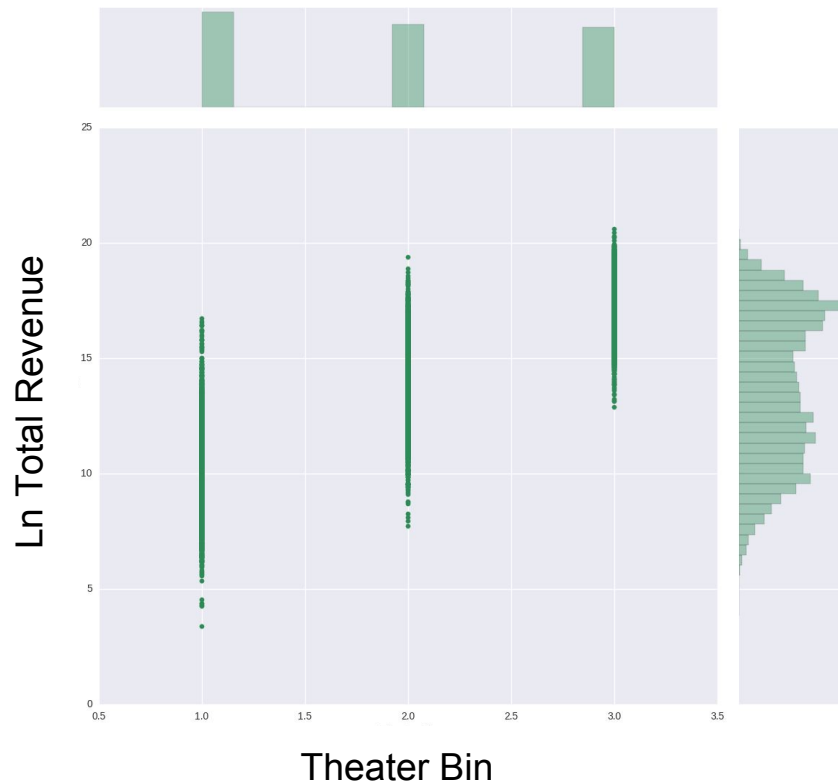
NOUN: Noun - person, place, thing

NUM: Number - 1, 2 , 3

PUNCT: Punctuation - :, -, . , , ...

Total Theaters

	# of Theaters
Bin 1	0 - 20
Bin 2	20 - 1000
Bin 3	1000+



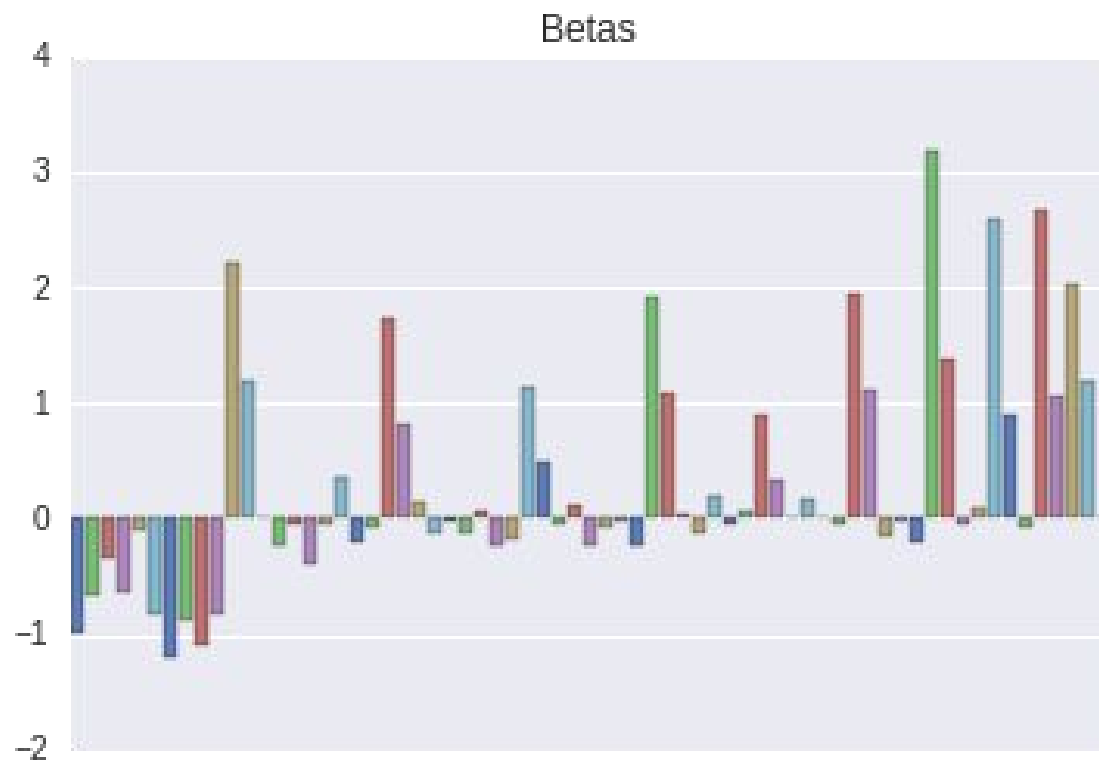
Ridge Regression

- Cross Validated Using 30% Split
- F test for significance
- Equation - $R^2 = .77$

$\text{Ln}(\text{Total Revenue}) \sim$

Parts of Speech + Theater Bins + $\Sigma(\text{POS} * \text{TB})$

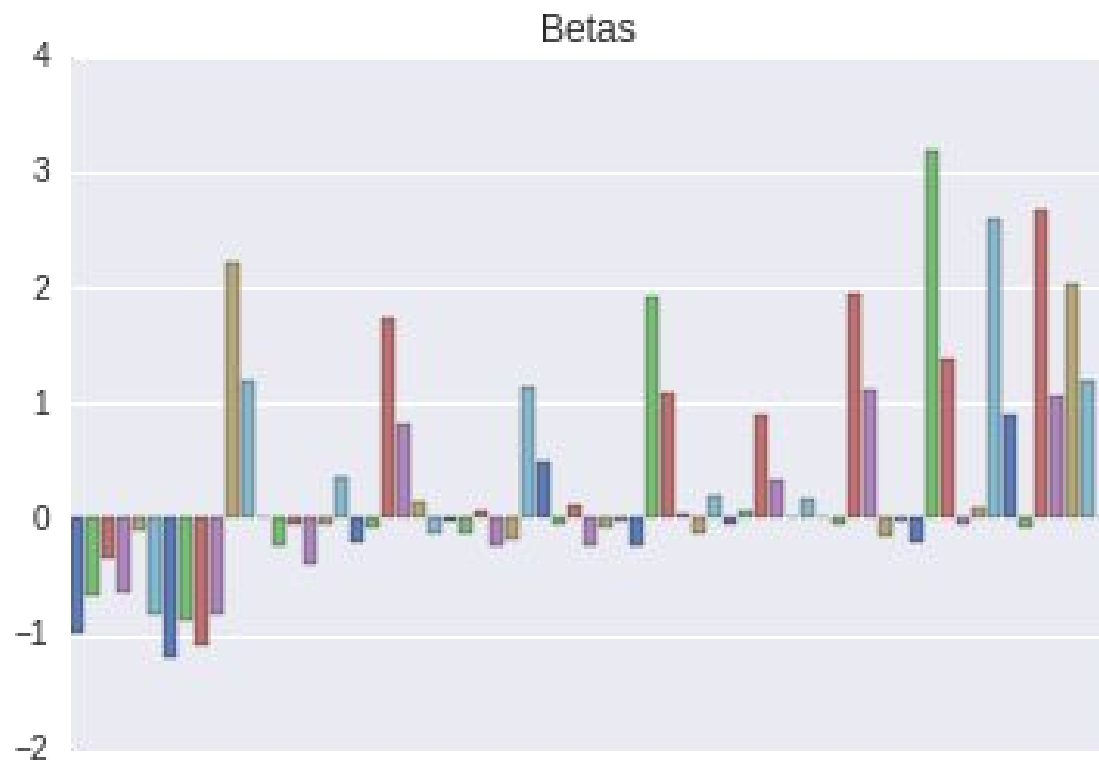
Analysis



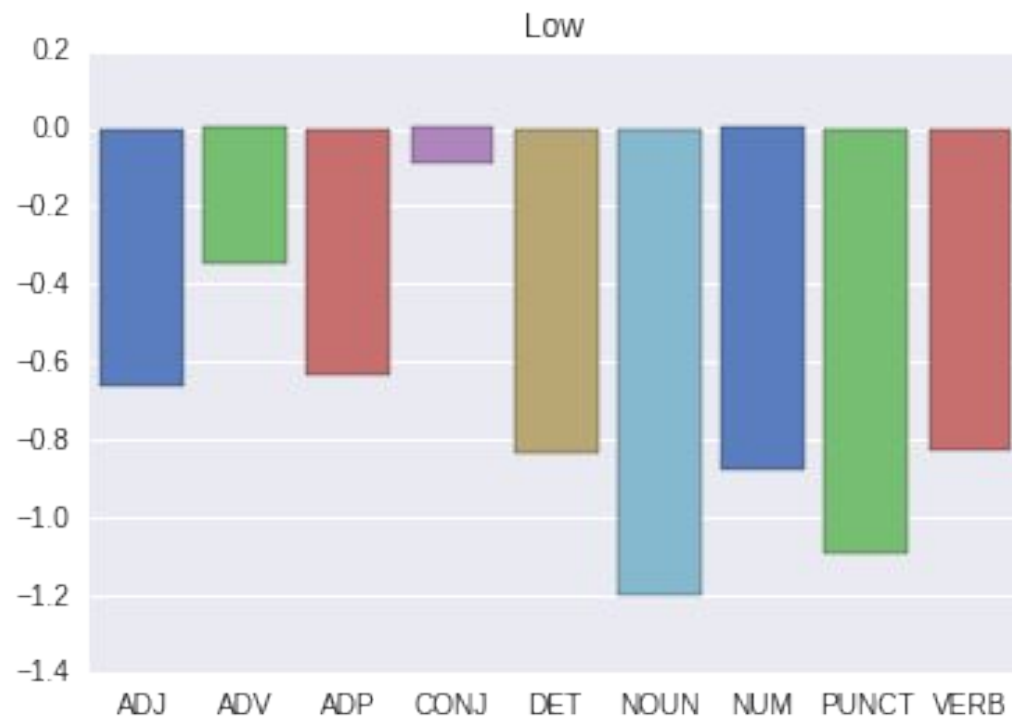
Analysis



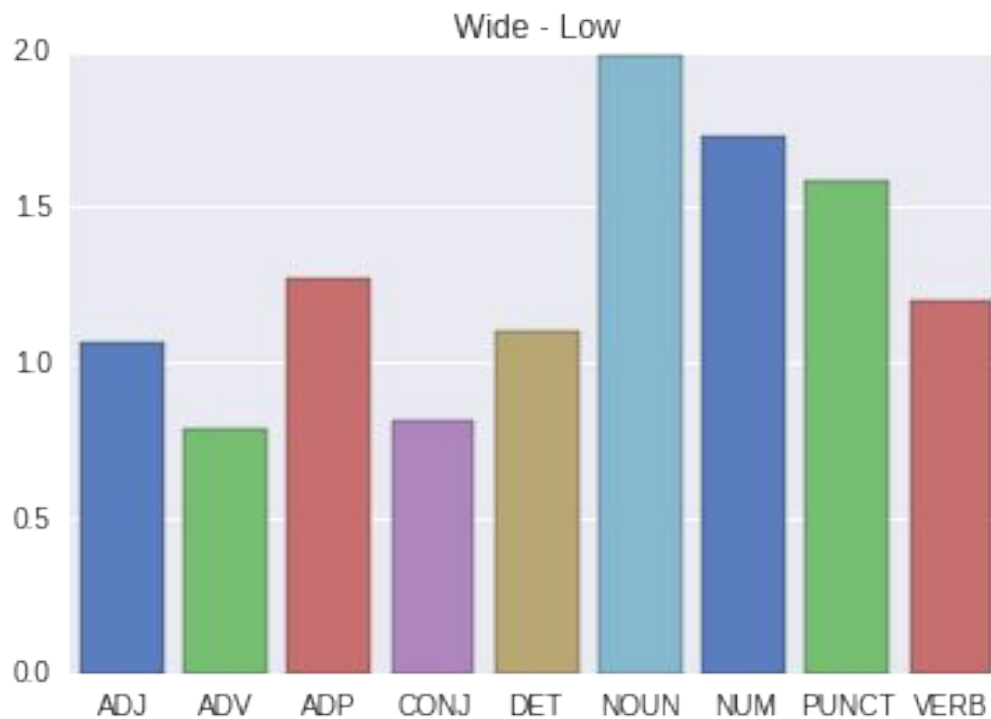
Analysis



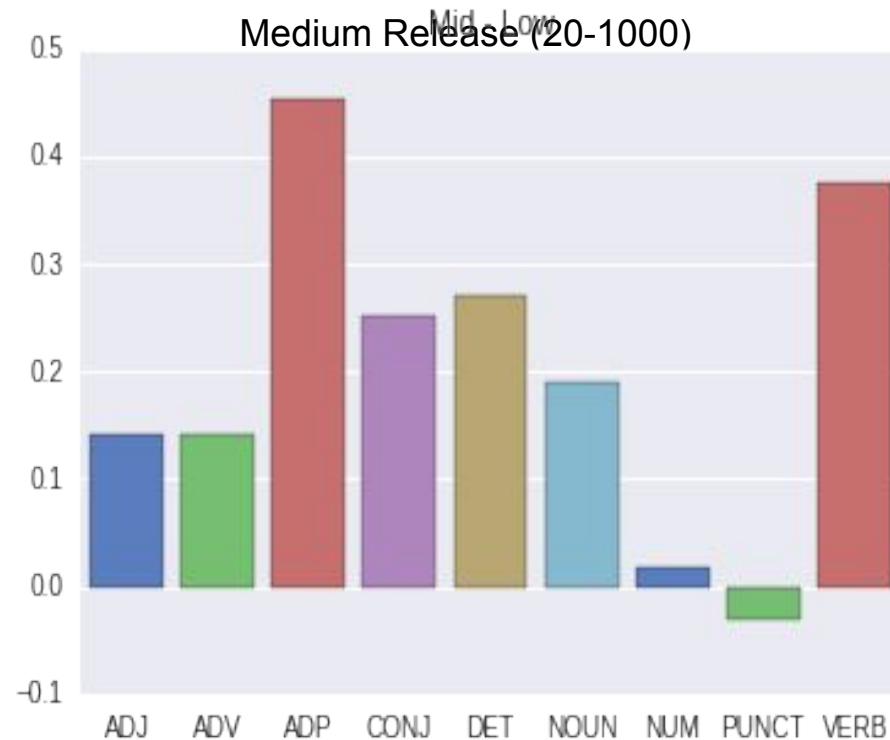
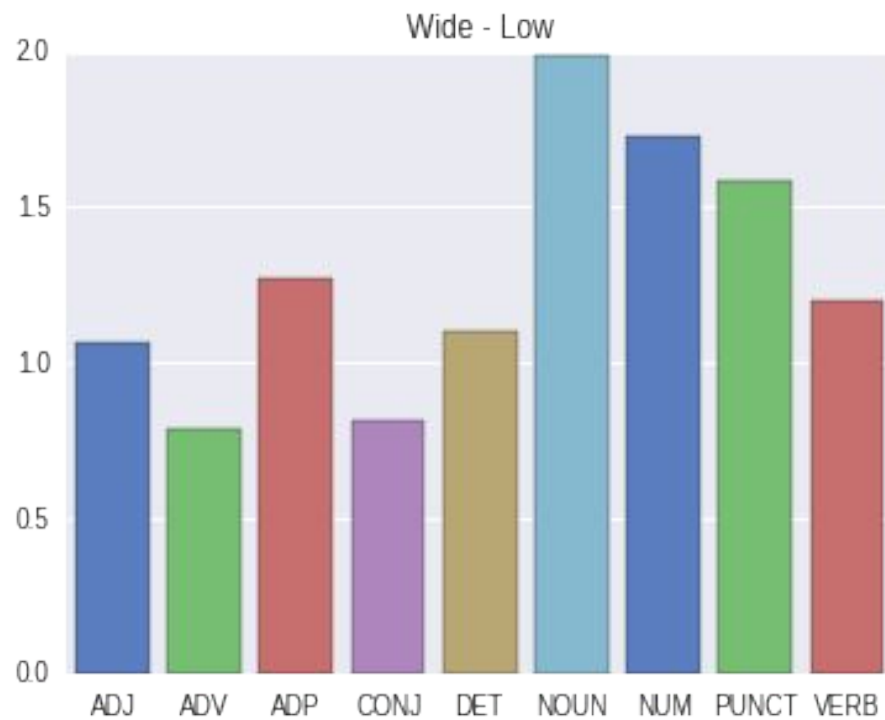
Less than 20 theaters



Wide Release (1000+)



Comparison



Recommendations

Recommendations

- For 1000+ theaters - Use a determiner, noun, punctuation, and number

Recommendations

- For 1000+ theaters - Use a determiner, noun, punctuation, and number
- For 20-1000 theaters - Use adpositions and verbs

Recommendations

- For 1000+ theaters - Use a determiner, noun, punctuation, and number
- For 20-1000 theaters - Use adpositions and verbs
- For 20 > theaters - No recommendation

Further Investigation

- Use profit instead of revenue
- Add interaction effects for phrases
- Include the actual words
- Predict critical reception

Further Investigation

- Use profit instead of revenue
- Add interaction effects for phrases
- Include the actual words
- Predict critical reception

