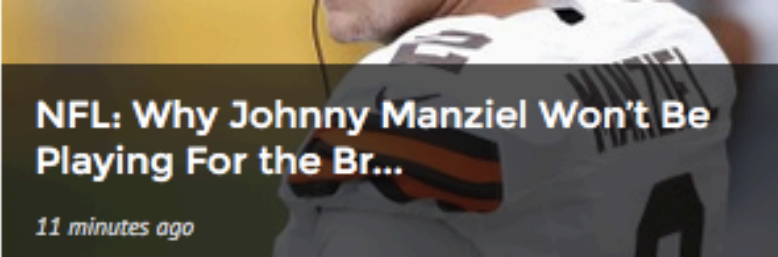




TITLE 

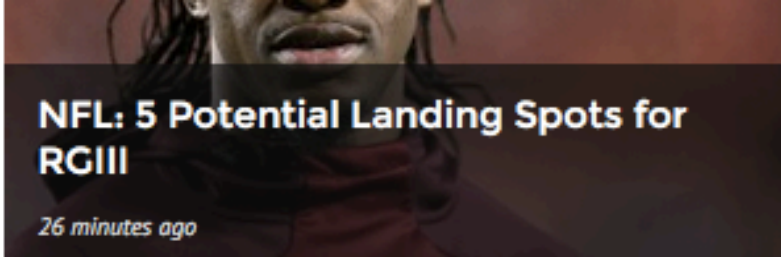
CONSTRUCTION

Lilliana Nishihira




NFL: Why Johnny Manziel Won't Be Playing For the Br...

11 minutes ago




NFL: 5 Potential Landing Spots for RGIII

26 minutes ago




The 5 All-Time Greatest Scorers in NBA Postseason History

about an hour ago



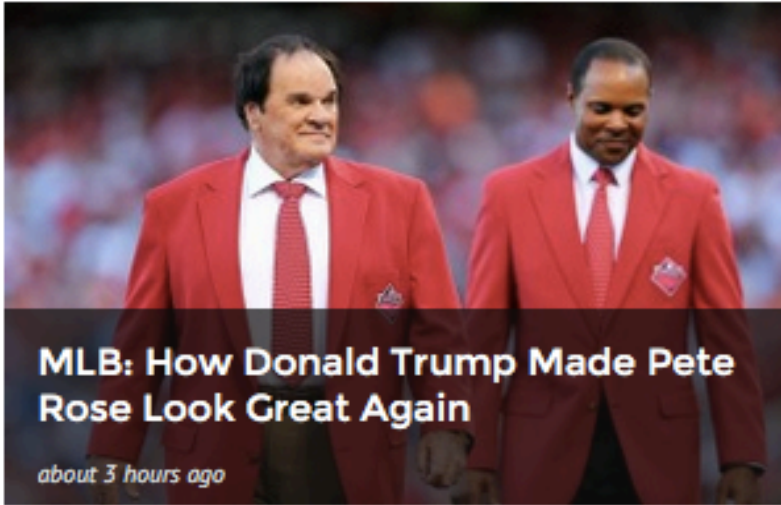
5 NFL Teams That Could Use a Helmet Makeover

about an hour ago



CBB: Which Team Has the Easiest Path to the Final Four?

about an hour ago

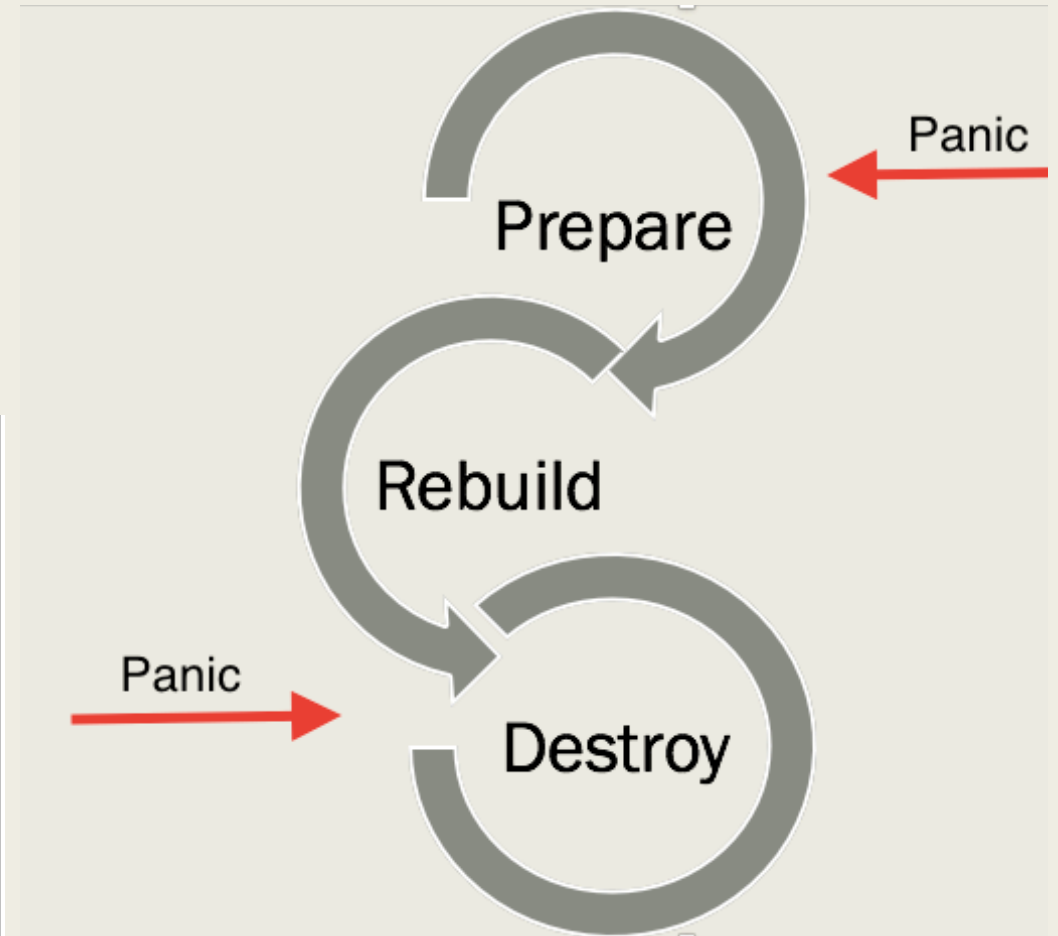


MLB: How Donald Trump Made Pete Rose Look Great Again

about 3 hours ago

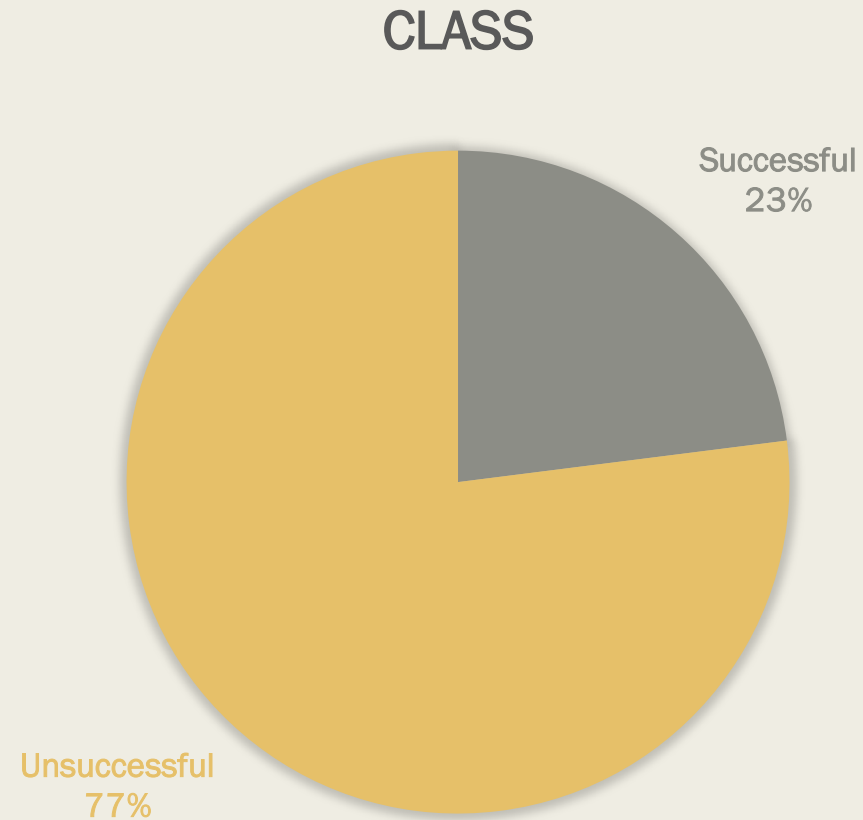
Feature Engineering

FieldName	Type	Description
post_id	string	unique identifier
page_title	string	Title of the article
publish_date	TimeSeries	Date and time published
p_new_session	float	Percent of page views were first entry
time_on_page	float	(average)number of seconds spent on page
Pageview	int	Times page was viewed (response)

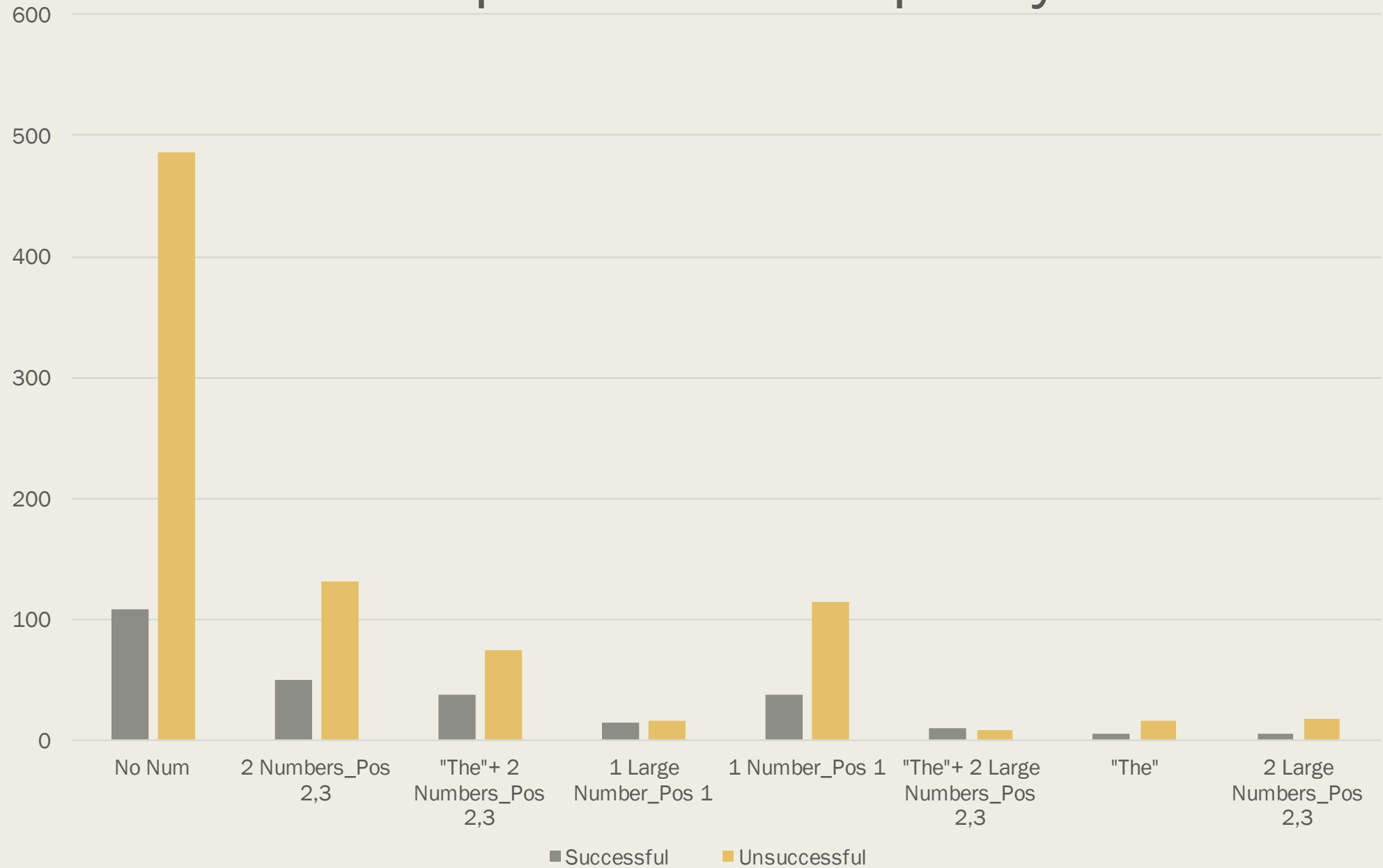


Feature Engineering

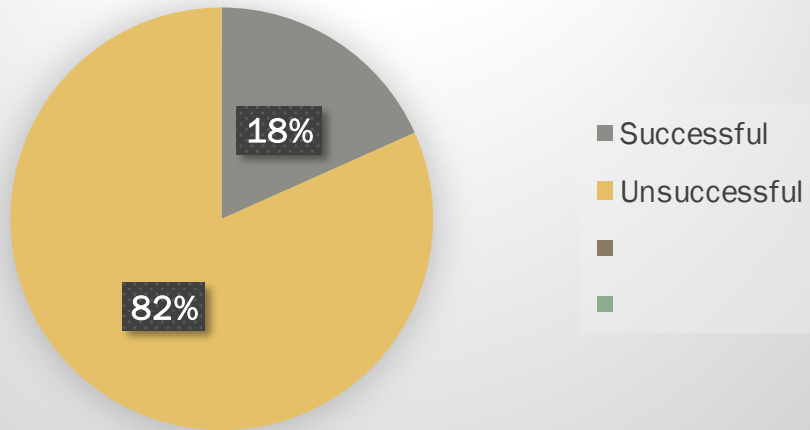
- Spacy – Parts of Speech count
 - *Noun*
 - *Verb*
 - *Determiner*
 - *Adjective*
- Hour
- Weekday
- “Class” (Success)
- Composition



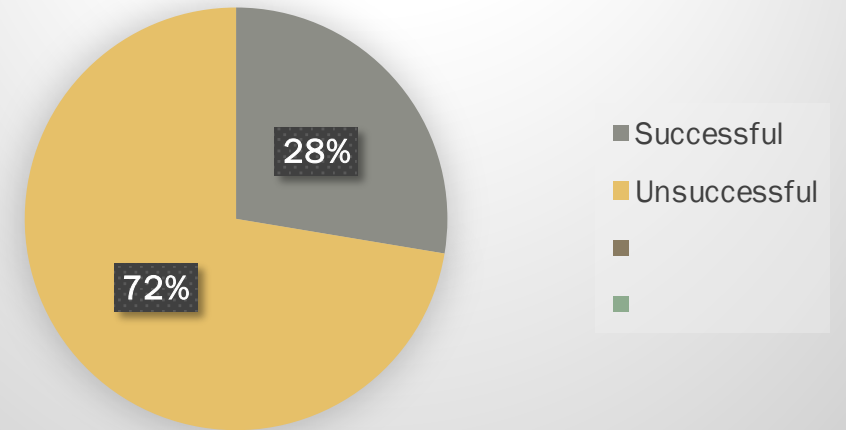
Composition Discrepancy



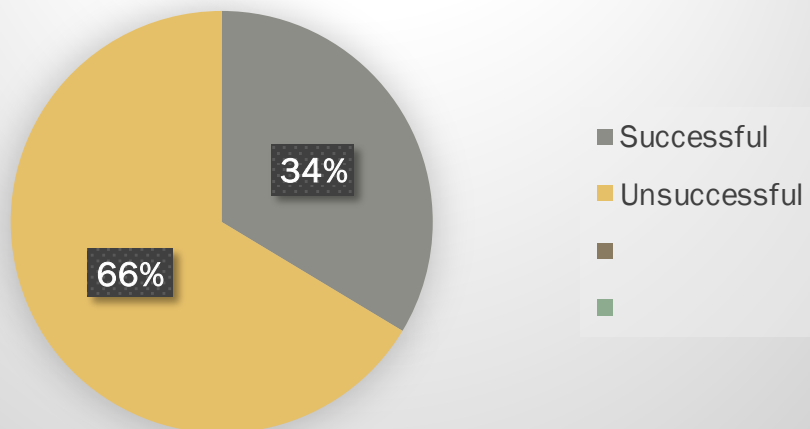
No Num



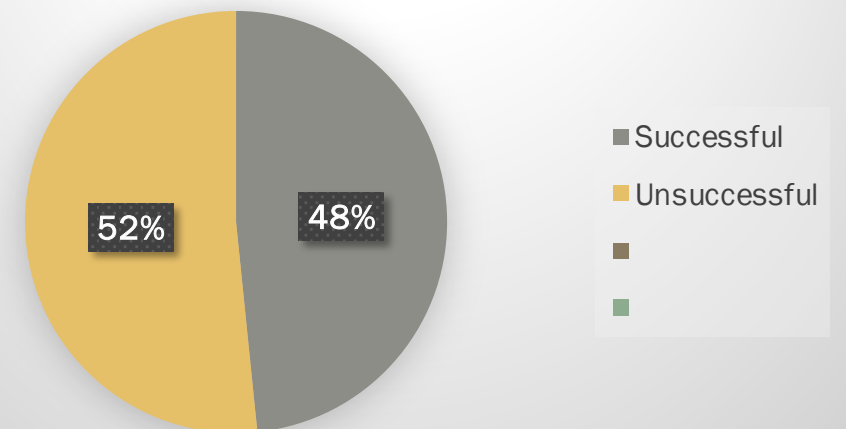
2 Numbers_Pos 2,3



"The"+ 2 Numbers_Pos 2,3



1 Large Number_Pos 1



Clusters k-Means

- 0 Sports Fanatic
 - *Low Noun*
 - *High Adj*
 - *High Num*
 - *Late Hour*
- 2 Game Reflector + Sheeter
 - *Low New Session*
 - *High Verb*
 - *Later day of week*

	0	1	2
p_new_sessions	0.008	0.009	0.001
time_on_page	0.003	0.001	0.001
Noun	4.012	5.546	5.034
Verb	0.905	0.723	1.172
Adj	0.802	0.137	0.289
Det	0.86	1.032	1.013
composition	0.061	0.012	0.013
DayofWeek	4.272	4.195	4.078
hour	14.63	12.426	12.504

Future

- Understanding composition
- Comparative analysis
- More Titles
- Compare types of parts of speech (Superlative adverb)
- Befriend Spacy

