

Movie Data Magic!

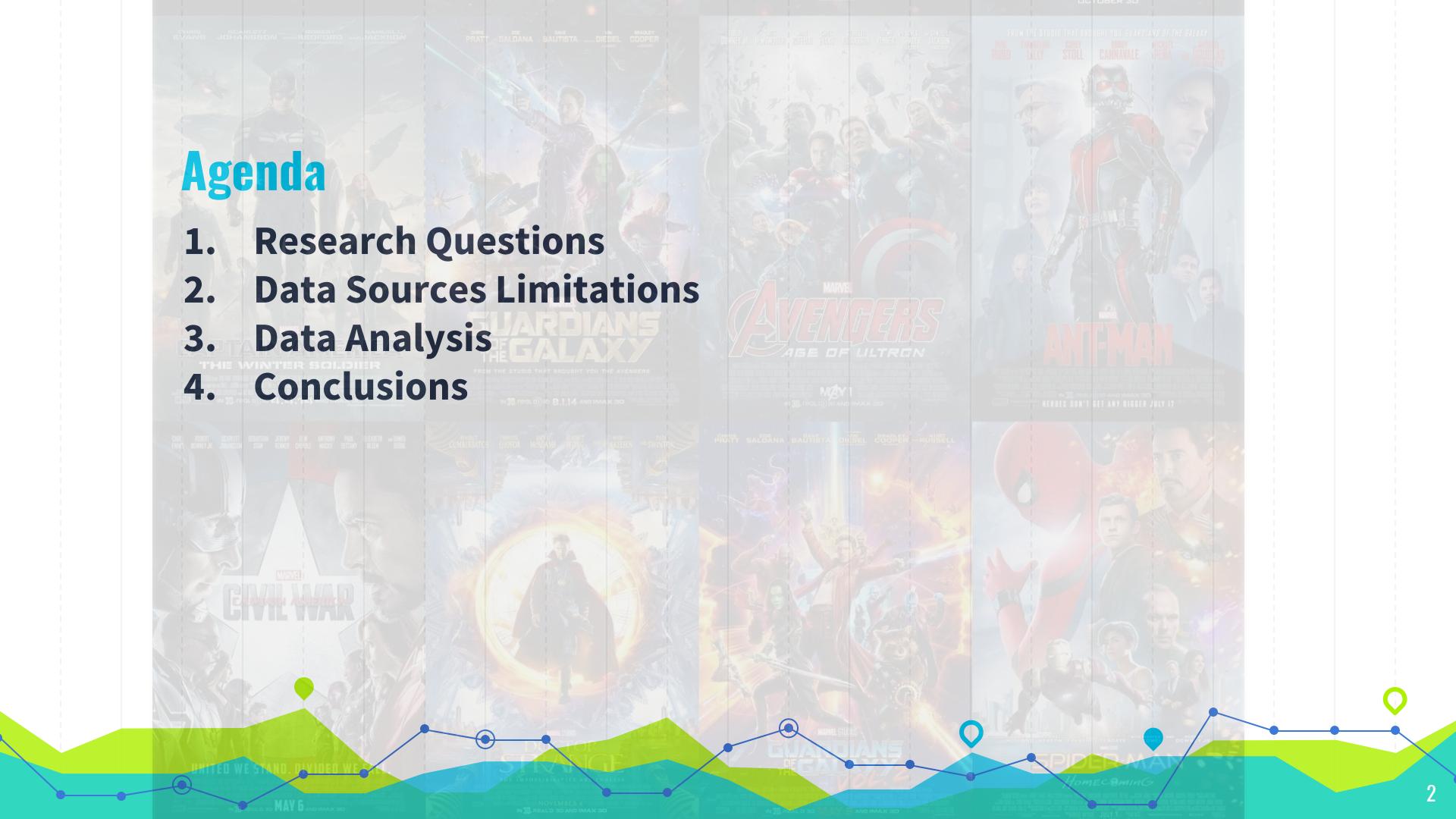
Box Office Research Study



Jane Chang
Christian Ott
Jack Jeng
8/11/18

Agenda

1. Research Questions
2. Data Sources Limitations
3. Data Analysis
4. Conclusions





First, let's review the Research Questions 1

Motion Pictures Industry Terms



DBO

Domestic Box Office is the total domestic ticket sales/gross including U.S. and Canada.



TWITTER *

Twitter is “an online news and social networking service on which users post and interact with messages known as “tweets.” *



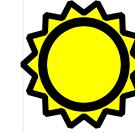
RT

Rotten Tomatoes is the review aggregator website including the Critics Score and Audience Score.



OPENING WEEKEND

Opening Weekend is the first weekend of a movie’s release i.e. Friday, Saturday, and Sunday.



SEASON

Seasons breaks down into winter, spring, summer, fall, and holidays such as Thanksgiving & Xmas.



WIDE RELEASE

A movie distribution release nationwide typically 2K-3K+ theater locations.

Research Questions

- If the **Rotten Tomatoes Score** goes up, does the **Box Office Opening Weekend** go up too?
- What is the impact of **Seasonality** on the **Box Office Opening Weekend**?
- If the **Twitter's Positive Sentiment** is high, does the **Box Office Opening Weekend** go up too?



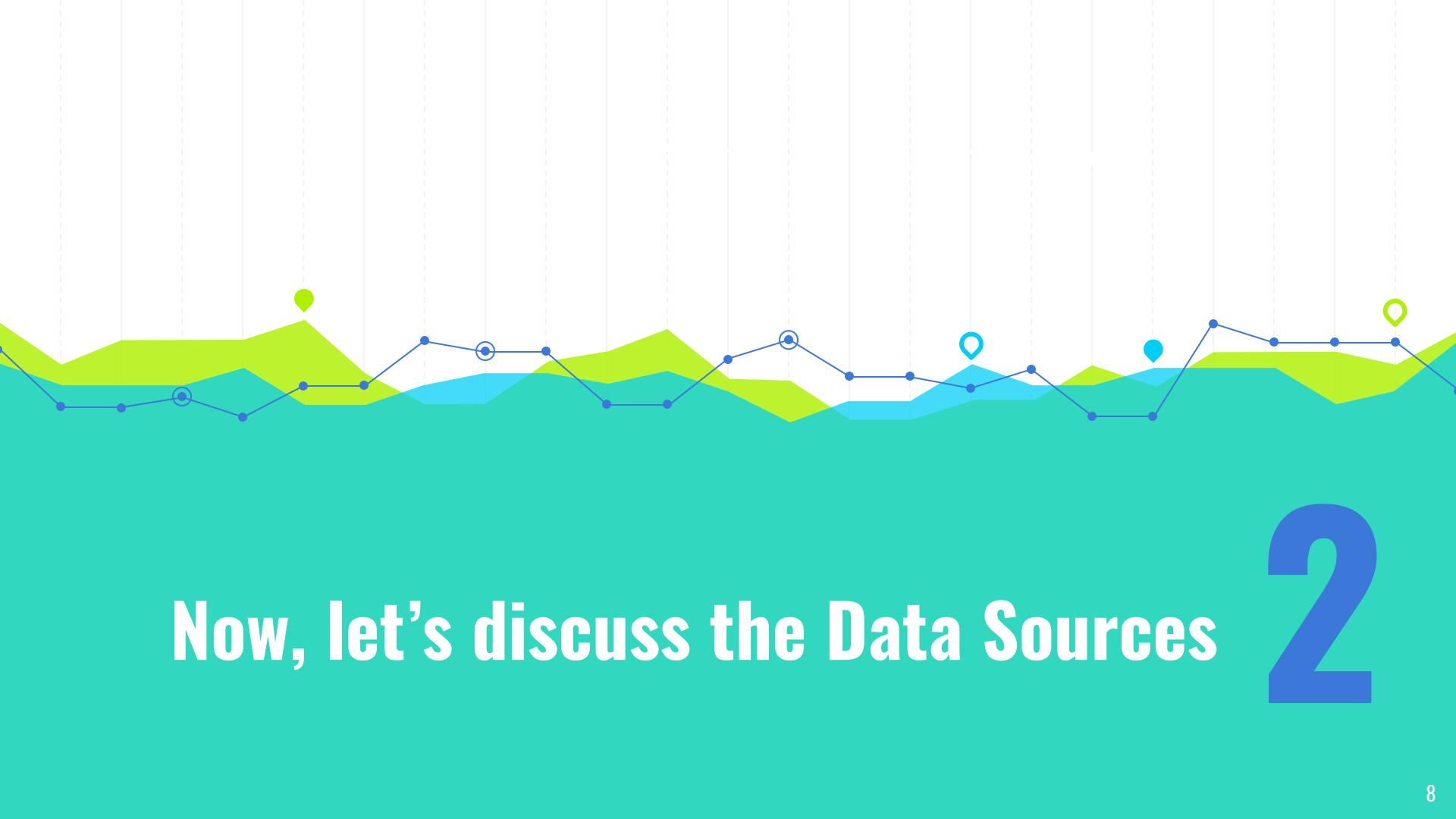


200 Movies

movie_title	studio	opening_weeks	opening_the	release_date	close_date	twitter1	RT_title	critics_rating	audience_rat	ReleaseMon	ReleaseMon	Season
Eighth Grade A24	\$263,797	4	8/3/18	-		EighthGrade	Eighth_Grade	99	89	8 August	Summer	
Mission: Imp Par.	\$61,236,534	4,386	7/27/18	-		MissionImpo	Mission_Imp	97	92	7 July	Summer	
Teen Titans (WB	\$10,411,189	3,188	7/27/18	-		TeenTitansG	Teen_Titans_	90	72	7 July	Summer	
Princess Mor Fathom	\$489,847	739	7/22/18	-		princessmon	princess_mo	-1	83	7 July	Summer	
Mamma Mia Uni.	\$34,952,180	3,317	7/20/18	-		MammaMia	Mamma_Mia	79	76	7 July	Summer	
The Equalize Sony	\$36,011,640	3,388	7/20/18	-		TheEqualizer	The_Equalize	50	74	7 July	Summer	
Unfriended: BH Tilt	\$3,653,035	1,546	7/20/18	-		UnfriendedN	Unfriended_	57	44	7 July	Summer	
Hotel Transy Sony	\$44,076,225	4,267	7/13/18	-		HotelT3	Hotel_Transy	62	52	7 July	Summer	
Skyscraper Uni.	\$24,905,015	3,782	7/13/18	-		SkyscraperM	skyscraper_2	48	79	7 July	Summer	
Sorry to BotI Annapurna	\$727,266	16	7/13/18	-		SorryToBoth	Sorry_to_Bot	94	62	7 July	Summer	
Ant-Man anc BV	\$75,812,205	4,206	7/6/18	-		AntMan	Ant_Man_an	87	80	7 July	Summer	
The First Pur Uni.	\$17,374,280	3,031	7/4/18	-		TheFirstPurg	The_First_Pu	52	34	7 July	Summer	
Sicario: Day <Sony	\$19,007,566	3,055	6/29/18	-		SicarioMovie	Sicario_Day_	64	66	6 June	Summer	
Uncle Drew LG/S	\$15,242,781	2,742	6/29/18	-		UncleDrew	Uncle_Drew	64	54	6 June	Summer	
Won't You Bi Focus	\$475,419	29	6/29/18	-		MrRogersMc	Wont_You_E	99	97	6 June	Summer	
TCM: West S Fathom	\$214,212	616	6/24/18	-		tcm	west_side_st	-1	-1	6 June	Summer	
Jurassic Wor Uni.	\$148,024,610	4,475	6/22/18	-		FallenKingdo	Jurassic_Wor	51	54	6 June	Summer	
Incredibles 2 BV	\$182,687,905	4,410	6/15/18	-		Incredibles2	Incredibles_2	93	87	6 June	Summer	
Tag WB (NL)	\$14,947,396	3,382	6/15/18	-		TagMovie	tag_2018	54	59	6 June	Summer	
Superfly Col.	\$6,870,740	2,220	6/13/18	-		Superfly	Superfly	92	69	6 June	Summer	
Ocean's 8 WB	\$41,607,378	4,145	6/8/18	-		Oceans8	Oceans_8	67	47	6 June	Summer	
Hereditary A24	\$13,575,172	2,964	6/8/18	-		Hereditary	Hereditary	89	59	6 June	Summer	
Hotel Artemi Global Road	\$3,232,790	2,407	6/8/18	7/19/18	HotelArtemi: Hotel_Artem	HotelArtemi	Hotel_Artem	59	33	6 June	Summer	
Adrift STX	\$11,603,039	3,015	6/1/18	-		AdriftMovie	Adrift	73	71	6 June	Summer	
Upgrade BH Tilt	\$4,670,905	1,457	6/1/18	7/12/18	UpgradeMov upgrade_201	UpgradeMov	upgrade_201	85	86	6 June	Summer	
Action Point Par.	\$2,390,164	2,032	6/1/18	6/14/18	ActionPoint: Action_Point	ActionPoint	Action_Point	20	35	6 June	Summer	
Solo: A Star \ BV	\$84,420,489	4,381	5/25/18	-		HanSolo	Solo_A_Star_	71	64	5 May	Summer	

Movies Datasets (Movies-Level)

MovieTitle	TweetID	TweetText	TweetCreated	CompoundSc	positive_list	negative_list	neutral_list
Mission: Imp	1.0263E+18	RT @ShakenC Sun Aug 05 22:4		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:4		0.05574	0.157	0.06	0.784
Mission: Imp	1.0263E+18	RT @Fright4 Sun Aug 05 22:4		-0.5574	0	0.146	0.854
Mission: Imp	1.0263E+18	RT @boxoffice Sun Aug 05 22:4		0	0	0	1
Mission: Imp	1.0263E+18	RT @ShowPlac Sun Aug 05 22:4		0.7906	0.219	0	0.781
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:4		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:4		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:4		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:4		0.6369	0.16	0	0.84
Mission: Imp	1.0263E+18	RT @MissGoldi Sun Aug 05 22:4		0.4199	0.163	0.071	0.766
Mission: Imp	1.0263E+18	RT @ShowPlac Sun Aug 05 22:4		0.7906	0.219	0	0.781
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:4		0.5386	0.3	0.184	0.516
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:3		0.7783	0.459	0	0.541
Mission: Imp	1.0263E+18	RT @ShowPlac Sun Aug 05 22:3		0.219	0	0	0.781
Mission: Imp	1.0263E+18	RT @boxoffice Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:3		-0.5267	0	0.221	0.779
Mission: Imp	1.0263E+18	RT @Self Sun Aug 05 22:3		-0.5267	0	0.221	0.779
Mission: Imp	1.0263E+18	RT @Avenom @missSun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:3		0.3818	0.126	0	0.874
Mission: Imp	1.0263E+18	RT @RottenToi Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @Cenang Sun Aug 05 22:3		0.3612	0.161	0	0.839
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:3		0.5574	0.157	0.06	0.784
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:3		0.5574	0.157	0.06	0.784
Mission: Imp	1.0263E+18	RT @RottenToi Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @Ishouldd Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @Avenom @missSun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @Episode_78 - M Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @podtimef Sun Aug 05 22:3		0.4389	0.153	0	0.847
Mission: Imp	1.0263E+18	RT @Paramour Sun Aug 05 22:3		0.4574	0.15	0	0.85
Mission: Imp	1.0263E+18	RT @How many mo Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @boxoffice Sun Aug 05 22:3		0.5423	0.29	0	0.71
Mission: Imp	1.0263E+18	RT @boxoffice Sun Aug 05 22:3		0	0	0	1
Mission: Imp	1.0263E+18	RT @Skydance Sun Aug 05 22:3		0.8126	0.331	0	0.669
Mission: Imp	1.0263E+18	RT @TomCruis Sun Aug 05 22:3		0.3612	0.116	0	0.884
Mission: Imp	1.0263E+18	RT @Skydance Sun Aug 05 22:2		0.8126	0.331	0	0.669
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:2		0.5574	0.157	0.06	0.784
Mission: Imp	1.0263E+18	RT @RottenToi Sun Aug 05 22:2		0	0	0	1
Mission: Imp	1.0263E+18	RT @boxoffice Sun Aug 05 22:2		0.5423	0.29	0	0.71
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:2		0	0	0	1
Mission: Imp	1.0263E+18	RT @Skydance Sun Aug 05 22:2		0.8126	0.331	0	0.669
Mission: Imp	1.0263E+18	RT @ShowPlacC Sun Aug 05 22:2		0	0	0	1
Mission: Imp	1.0263E+18	RT @MissionFi Sun Aug 05 22:2		0	0	0	1
Mission: Imp	1.0263E+18	RT @Mattrego! Sun Aug 05 22:2		0.4019	0.144	0	0.856
Mission: Imp	1.0263E+18	RT @MAX: Let Sun Aug 05 22:2		0.3818	0.098	0	0.902
Mission: Imp	1.0263E+18	RT @rebeccafe Sun Aug 05 22:2		0.5719	0.217	0	0.783
Mission: Imp	1.0263E+18	RT @ebecaffe Sun Aug 05 22:2		0	0	0	1
Mission: Imp	1.0263E+18	RT @ShowPlac Sun Aug 05 22:2		0.7906	0.219	0	0.781



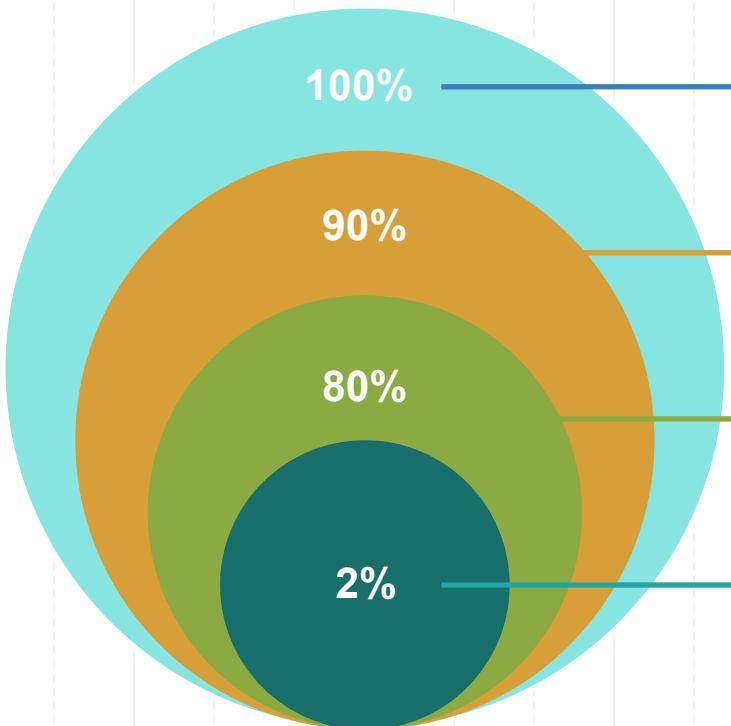
Now, let's discuss the Data Sources

2

Data Sources Limitations*

* Data pulled Movies from years 2016 - 2018

** Percentages determine the data completeness



Calendar Seasons

- Box Office Mojo
- Five Seasons

Rotten Tomatoes

- Audiences' and Critics' Scores
- Webscraping via Beautiful Soup API

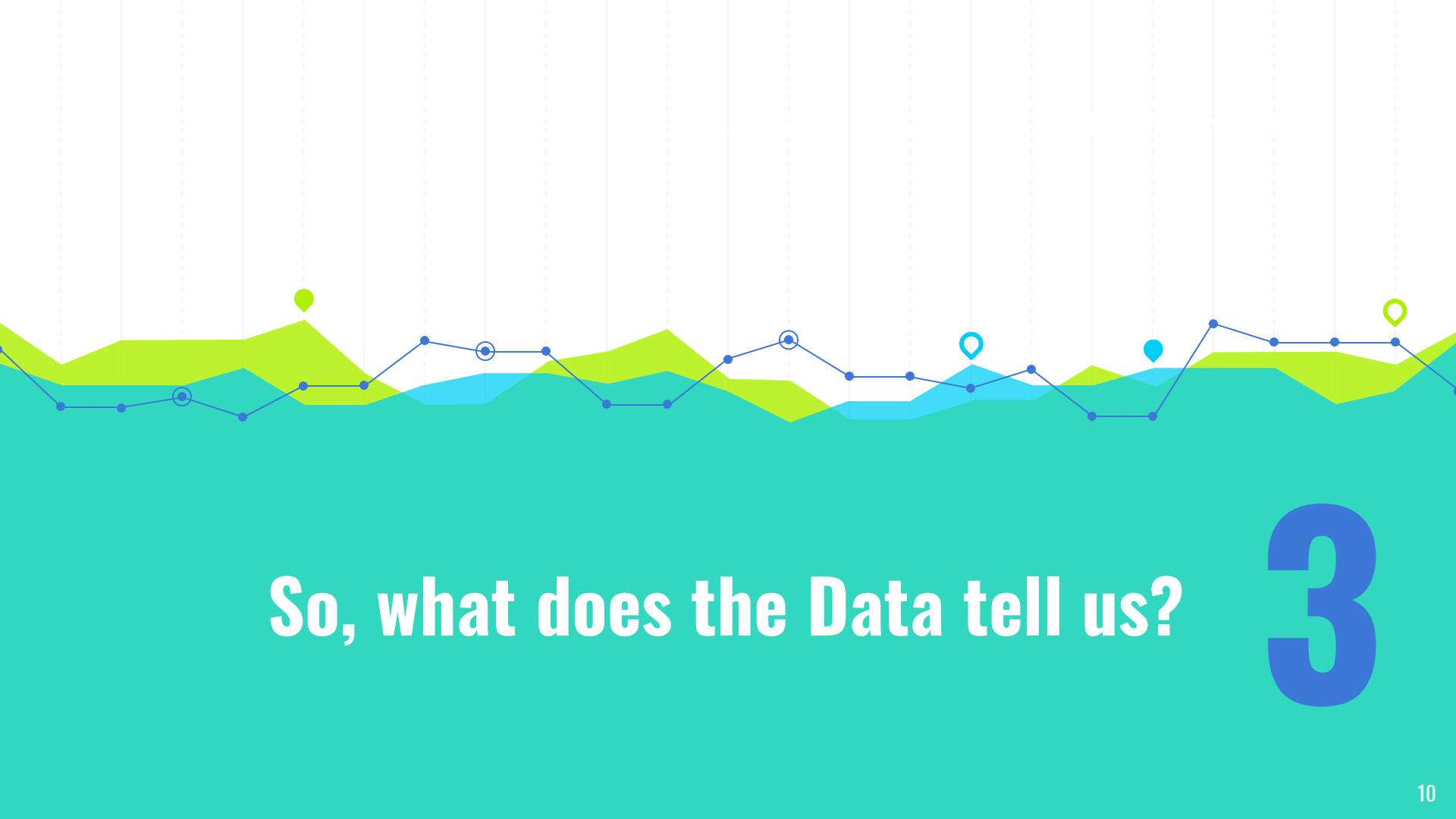
Domestic Box Office

- Box Office Mojo
- 2,000+ theater locations

Twitter

- Comments/Tweets via Tweepy
- Sentiment via VaderSentiment API



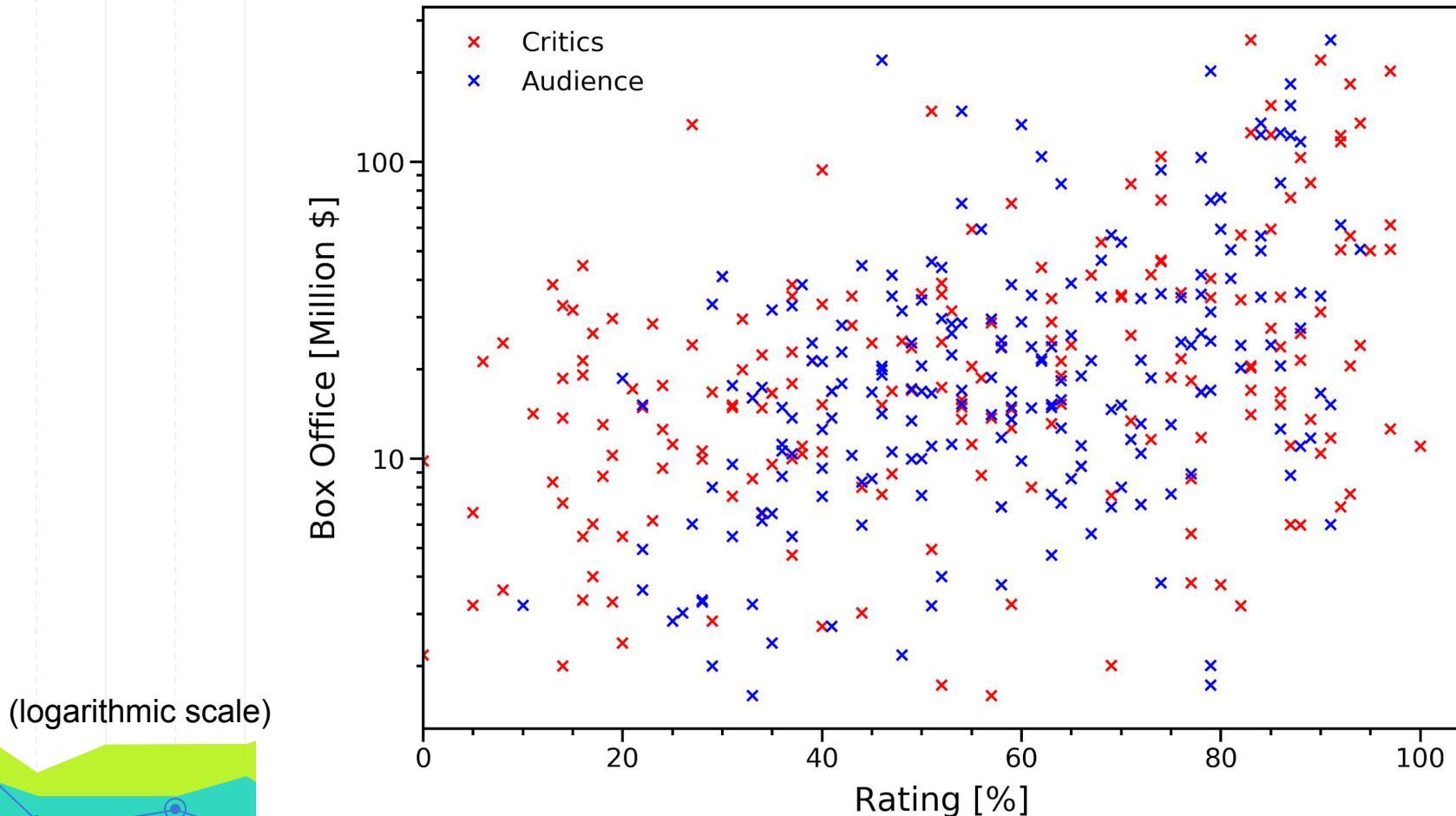


So, what does the Data tell us?

3

Analysis 🍅 RT vs Box Office

Opening Weekend Box Office vs. Rating



Result - Rotten Tomatoes vs Box Office

Hypothesis testing using Pearson's sample correlation coefficient:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$



Critics:

$r = 0.363,$

p-value = 1.58×10^{-7}



Audience:

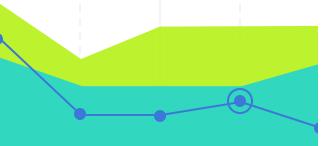
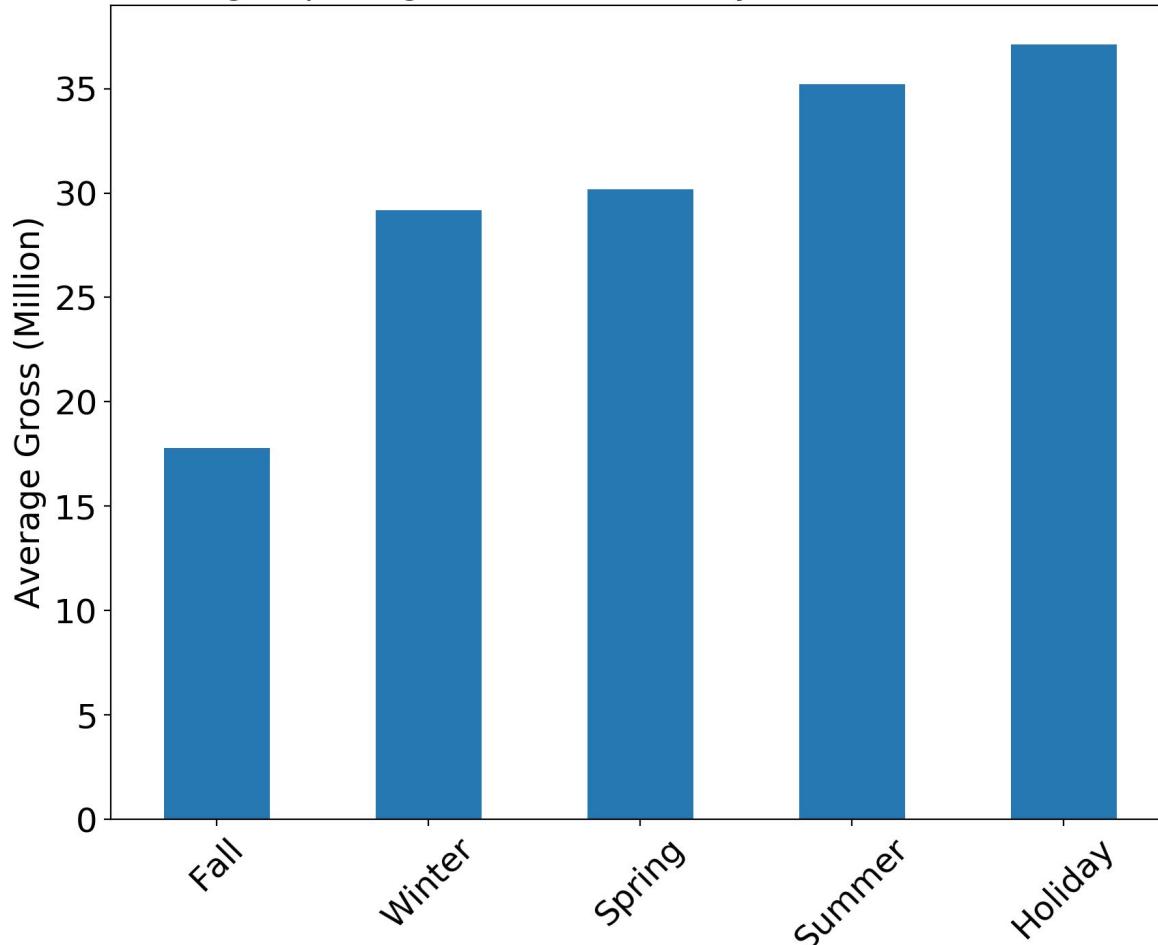
$r = 0.366$

p-value = 1.26×10^{-7}



Analysis - Seasons vs Box Office

Average Opening Weekend Gross By Season - Wide Release



Analysis - Seasons vs Box Office



Data Suggests

Movies perform best in the Summer & Holiday Seasons, and perform worst in the Fall.



Statistically Significance

- The answer is 1-way ANOVA test.



What is p-value?

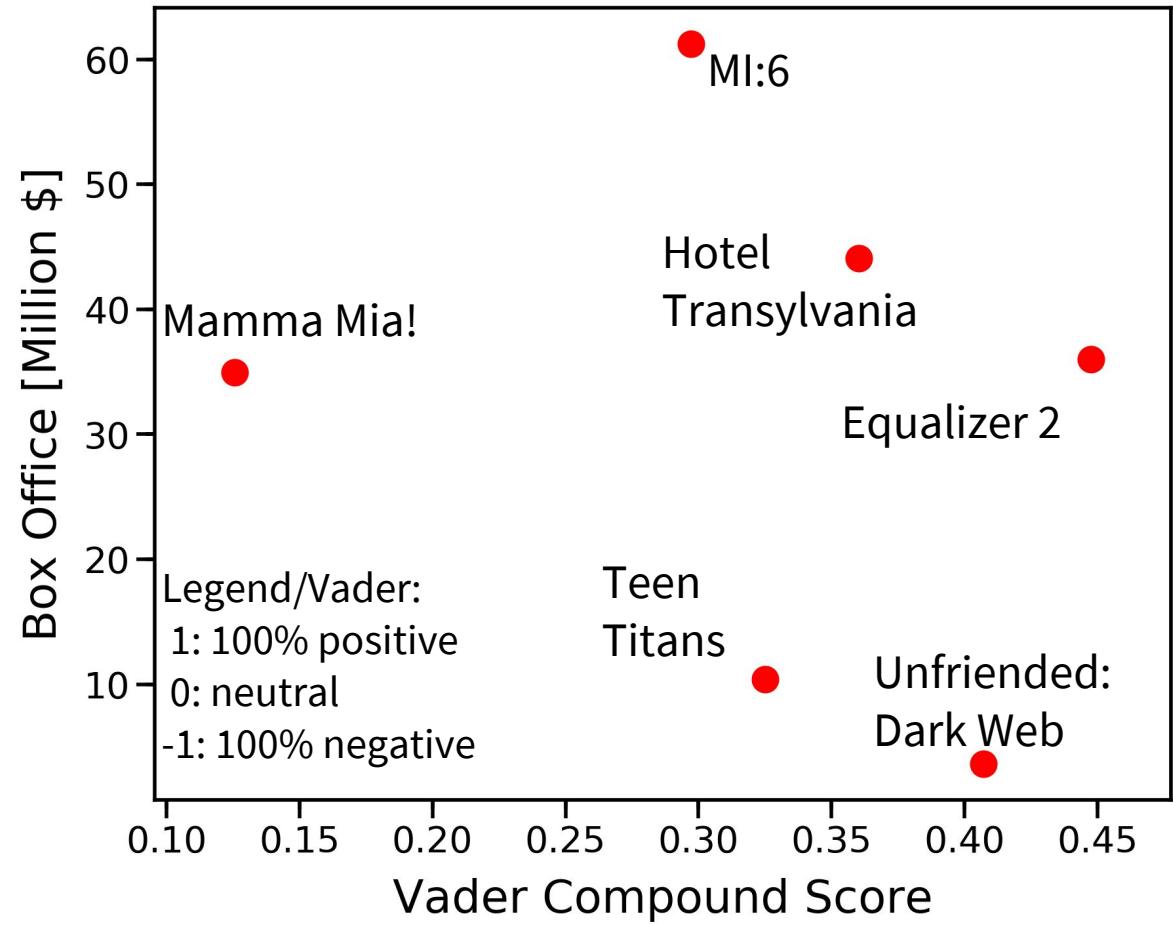
- The p-value = 0.15
-> **Data are consistent with each other**
- So, neither the Fall nor Summer & Holidays are a statistically significant variation from the mean.



Analysis - Twitter vs BO

- * Tiny sample
- * Limitations
- Thus, no correlation

Opening Weekend Box Office vs. Vader Compound Score

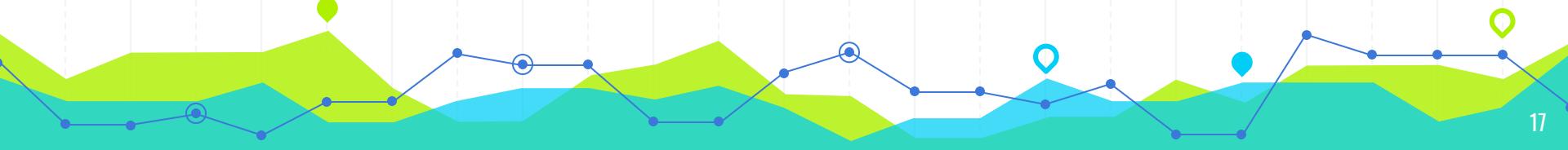




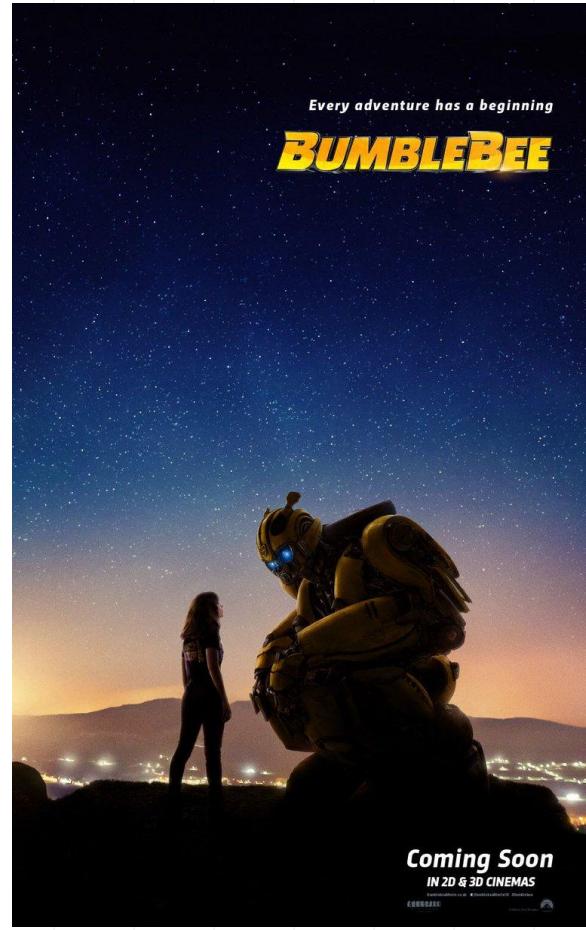
So, what are the Box Office takeaways? **4**

Conclusion

- **Scores vs. Box Office:** Higher scores -> higher box office.
- **Seasons vs. Box Office:** There's a trend, but it's not statistically significant. (Fall: low, Summer & Holidays: high)
- **Tweets vs. Box Office:** No correlation using the tiny sample.



Thanks.



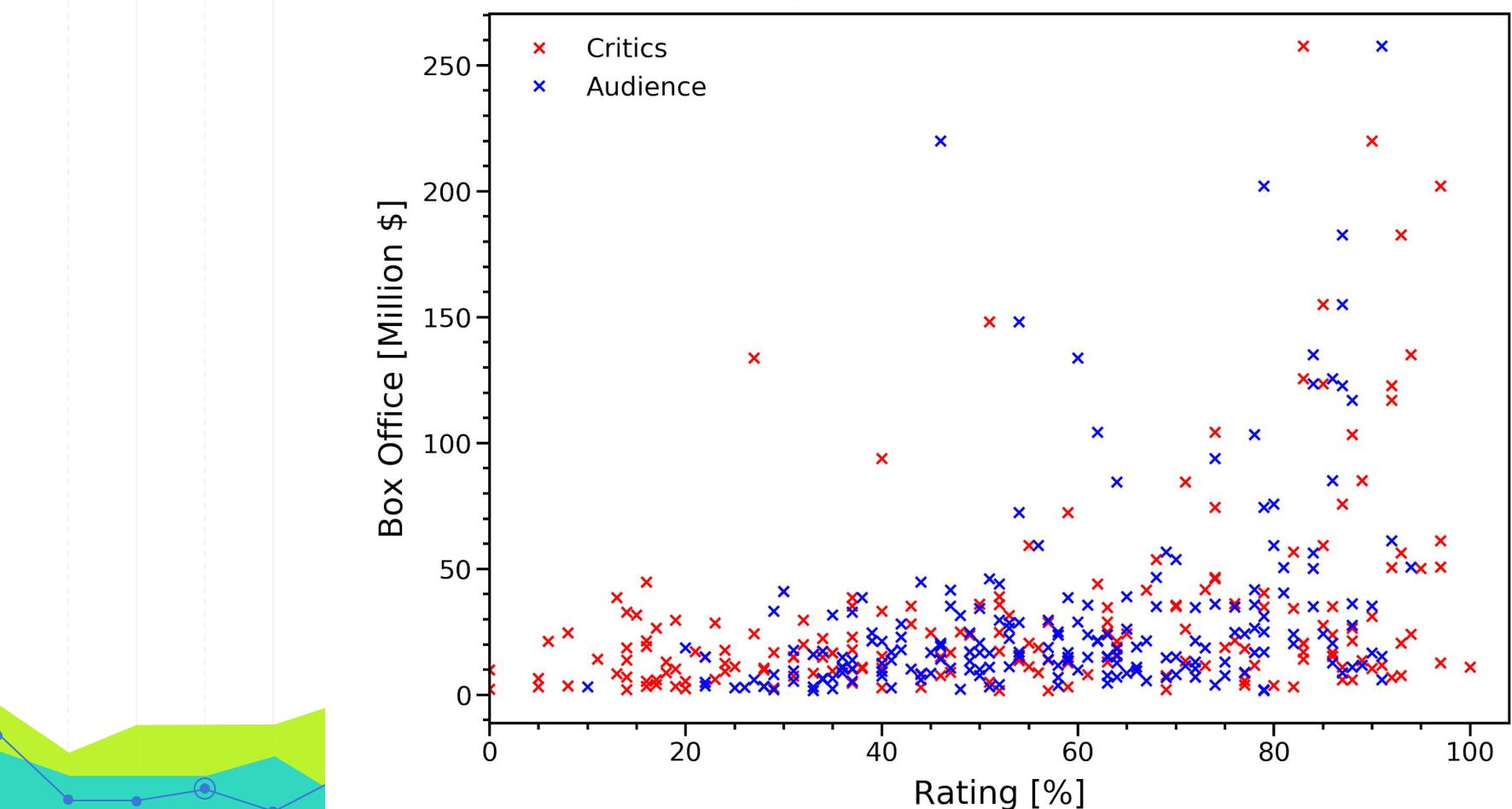
Appendix

X



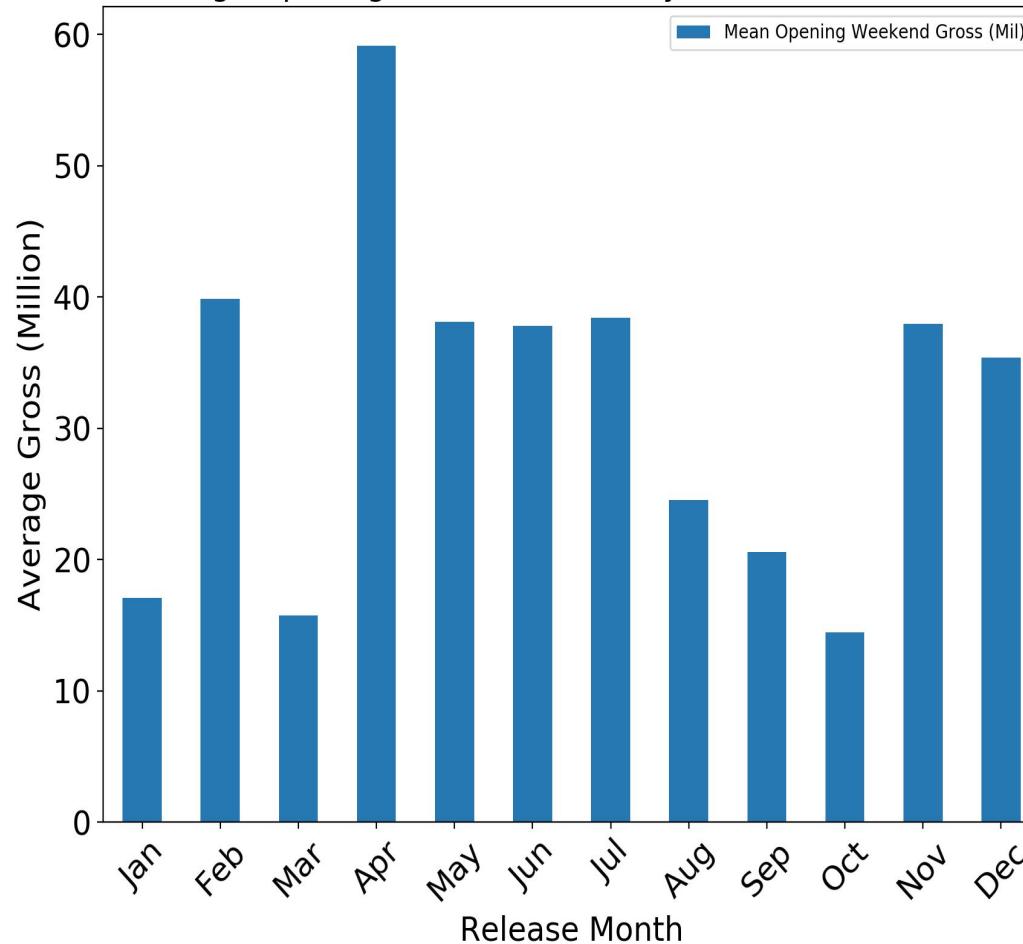
Analysis - RT vs Box Office

Opening Weekend Box Office vs. Rating

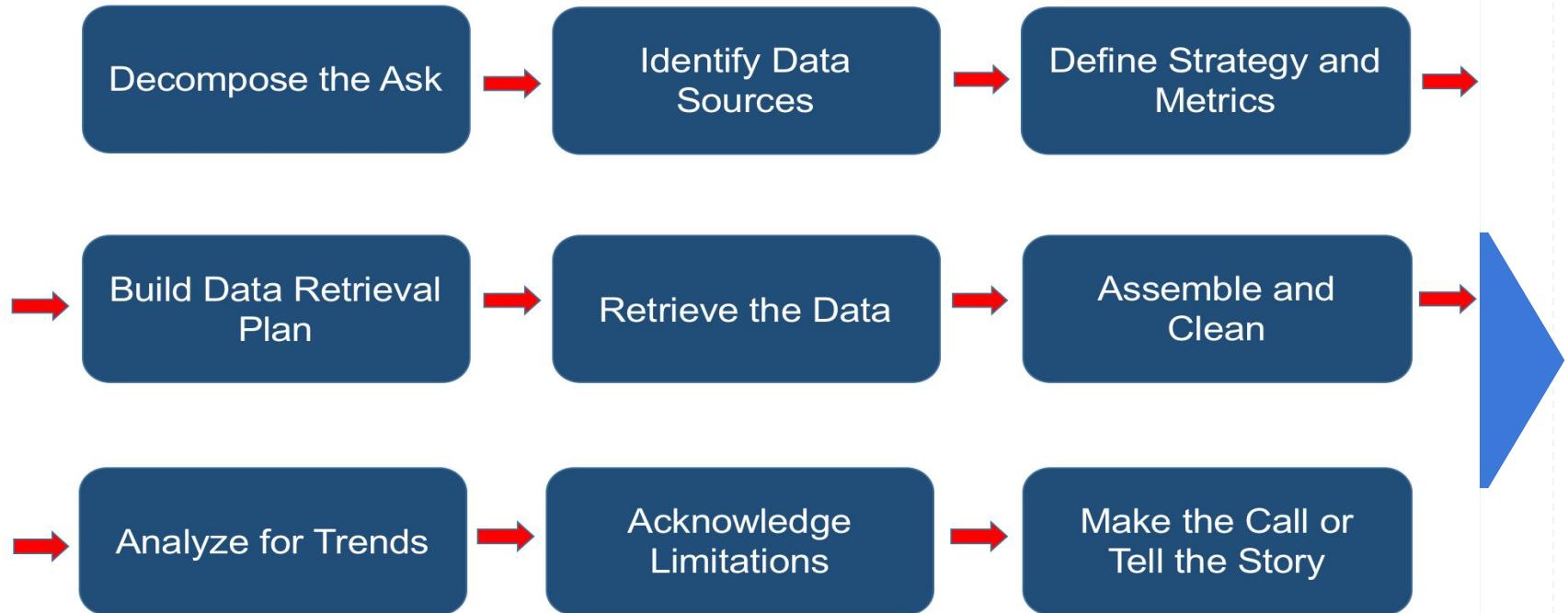


Analysis - Calendar vs Box Office

Average Opening Weekend Gross By Month - Wide Release



Analytics Paradigm



Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.

Movies Seasons

year	season_type	begin_date	end_date
2018	Winter	1/8/18	3/1/18
2018	Spring	3/2/18	5/3/18
2018	Summer	5/4/18	9/3/18
2018	Fall	9/4/18	11/1/18
2018	Holiday	11/2/18	1/6/19
2017	Winter	1/9/17	3/2/17
2017	Spring	3/3/17	5/4/17
2017	Summer	5/5/17	9/4/17
2017	Fall	9/5/17	11/2/17
2017	Holiday	11/3/17	1/7/18
2016	Winter	1/4/16	3/3/16
2016	Spring	3/4/16	5/5/16
2016	Summer	5/6/16	9/5/16
2016	Fall	9/6/16	11/3/16
2016	Holiday	11/4/16	1/8/17