Analyzing Official Review and User Ratings on Gamespot and Steam.com

NYC Data Science Academy

Web Scraping Project

Arthur Yu

Motivation

- Do official reviews of games correlate with user ratings?

- Does price of games influence user ratings?

- Can we predict user rating of a game based on official review and price?

Websites

- Gamespot: A video preview / review site

  https://www.gamespot.com/reviews/?review_filter_type%5Bplatform%5D=19&review_filter_type%5Bgenre%5D=55&review_filter_type%5BtimeFrame%5D=&review_filter_type%5BstartDate%5D=&review_filter_type%5BendDate%5D=&review_filter_type%5BminRating%5D=&review_filter_type%5Btheme%5D=&review_filter_type%5Bregion%5D=&review_filter_type%5Bpublishers%5D=&review_filter_type%5Bdevelopers%5D=&review_filter_type%5Bletter%5D=&sort=date

- Steam Store: An online store for video games

  https://store.steampowered.com/search/?tags=9

- Filtered for strategy games only

Scraped Data Structures

Gamespot.com:

- Title
- Release date
- Developer(s)
- Genre(s)
- Gamespot rating
- Number of user reviews
- Average user rating

Steam.com:

- Title
- User Reviews
  - Number
    - Percentage positive
- Price

Raw Data

	date	developer	genre	rating	reviews	title	userrate
0	Apr 27, 2017	Relic	Real-Time,Strategy	8.0	33	Warhammer 40,000: Dawn of War III	5.0
1	Jan 22, 2002	Strategy First	Turn-Based,Strategy	NaN	1074	Disciples II: Dark Prophecy	8.4
2	Feb 11, 2004	EA Phenomic	Strategy,Real-Time	7.9	1788	SpellForce: The Order of Dawn	8.3
3	May 12, 2003	Nival Interactive,La Plata,ACTive Gaming,Intex...	Real-Time,Strategy	7.0	1046	Blitzkrieg	8.0
4	May 17, 2005	Robosoft Technologies,Pyro Studios	Real-Time,Strategy	6.7	6	Imperial Glory	8.0
5	Apr 5, 2005	Handmark,Sandlot Games,Astraware	Strategy,Management	7.2	87	Tradewinds 2	7.4
6	May 5, 2006	Spellbound	Real-Time,Strategy	6.1	506	Desperados 2: Cooper's Revenge	7.3
7	Sep 29, 2006	Introversion	Real-Time,Strategy	8.1	991	DEFCON: Everybody Dies	8.3
8	Sep 29, 2006	Dark Fox	Strategy,Real-Time	5.2	116	Silent Heroes	6.9
9	Oct 25, 2005	Firaxis Games	Strategy,Turn-Based	NaN	22287	Sid Meier's Civilization IV	8.8

'5]:

	title	steam_review	price
0	Counter-Strike: Global Offensive	Very Positive 88% of the 2,647,991 user rev...	$14.99
1	Tom Clancy's Rainbow Six Siege	Very Positive 80% of the 174,250 user revie...	$14.99
2	Dota 2	Very Positive 87% of the 914,333 user revie...	Free to Play
3	Endless Space 2	Very Positive 80% of the 8,076 user reviews...	$39.99
4	Tabletop Simulator	Very Positive 93% of the 7,169 user reviews...	$19.99
5	XCOM 2	Very Positive 80% of the 19,811 user review...	$59.99
6	Endless Legend	Very Positive 82% of the 11,160 user review...	$29.99
7	RimWorld	Overwhelmingly Positive 97% of the 19,856 u...	$29.99
8	Divinity: Original Sin 2	Very Positive 94% of the 27,939 user review...	$44.99
9	SMITE	Very Positive 82% of the 38,173 user review...	Free to Play

Preprocessing

- Parse the Steam rating and number of reviews from string, add two columns
- Turn numbers into numbers, dates into dates
- Eliminate games with no Gamespot rating and Free-to-Play games
- Merge two data sets

	date	developer	genre	rating	reviews	title	userrate	steam_review	price	review_num
0	Apr 27, 2017	Relic	Real-Time,Strategy	8.0	33	Warhammer 40,000: Dawn of War III	5.0	4.6	39.99	7299.0
2	Feb 11, 2004	EA Phenomic	Strategy,Real-Time	7.9	1788	SpellForce: The Order of Dawn	8.3	NaN	NaN	NaN
3	May 12, 2003	Nival Interactive,La Plata,ACTive Gaming,Intex...	Real-Time,Strategy	7.0	1046	Blitzkrieg	8.0	NaN	NaN	NaN
4	May 17, 2005	Robosoft Technologies,Pyro Studios	Real-Time,Strategy	6.7	6	Imperial Glory	8.0	7.8	4.99	651.0
5	Apr 5, 2005	Handmark,Sandlot Games,Astraware	Strategy,Management	7.2	87	Tradewinds 2	7.4	NaN	NaN	NaN
6	May 5, 2006	Spellbound	Real-Time,Strategy	6.1	506	Desperados 2: Cooper's Revenge	7.3	NaN	NaN	NaN
7	Sep 29, 2006	Introversion	Real-Time,Strategy	8.1	991	DEFCON: Everybody Dies	8.3	NaN	NaN	NaN
8	Sep 29, 2006	Dark Fox	Strategy,Real-Time	5.2	116	Silent Heroes	6.9	NaN	NaN	NaN
10	Mar 11, 2008	Haemimont Games	Strategy,Management	3.5	437	Imperium Romanum	7.0	NaN	NaN	NaN
11	Sep 25, 2007	Blue Byte,Ubisoft	Management,Strategy	6.0	1822	The Settlers: Rise of an Empire	7.5	NaN	NaN	NaN

|	date	developer	genre	rating	reviews	title	userrate	steam_review	price	review_num
0	2017-04-27	Relic	Real-Time,Strategy	8.0	33	Warhammer 40,000: Dawn of War III	5.0	4.6	39.99	7299
1	2005-05-17	Robosoft Technologies,Pyro Studios	Real-Time,Strategy	6.7	6	Imperial Glory	8.0	7.8	4.99	651
3	2005-10-25	Firaxis Games	Turn-Based,Strategy	8.0	22287	Sid Meier's Civilization IV	8.8	9.2	19.99	970
5	2005-10-25	Firaxis Games	Turn-Based,Strategy	9.4	22287	Sid Meier's Civilization IV	8.8	9.2	19.99	970
6	2011-02-13	1C,THQ,Digitalmindsoft	Real-Time,Strategy	8.0	202	Men of War: Assault Squad	8.0	8.7	9.99	2359
7	1995-02-28	Culture Brain,MicroProse	Strategy,Turn-Based	7.2	735	X-COM: Terror From the Deep	8.4	9.2	4.99	308
8	2011-07-11	The Lordz Games Studio,Flashback Games	Strategy,Turn-Based	8.0	186	Panzer Corps	7.8	9.1	19.99	807
9	2011-08-02	Nitro Games	Strategy	3.5	160	Pirates of Black Cove	4.4	5.8	9.99	84
10	2011-05-26	Apportable,Mode 7 Games	Strategy,Turn-Based	8.5	238	Frozen Synapse	7.5	9.0	24.99	959
12	2011-09-16	NeocoreGames	Strategy,Real-Time	7.5	51	King Arthur: Fallen Champions	5.7	4.5	4.99	20

Categorical Data
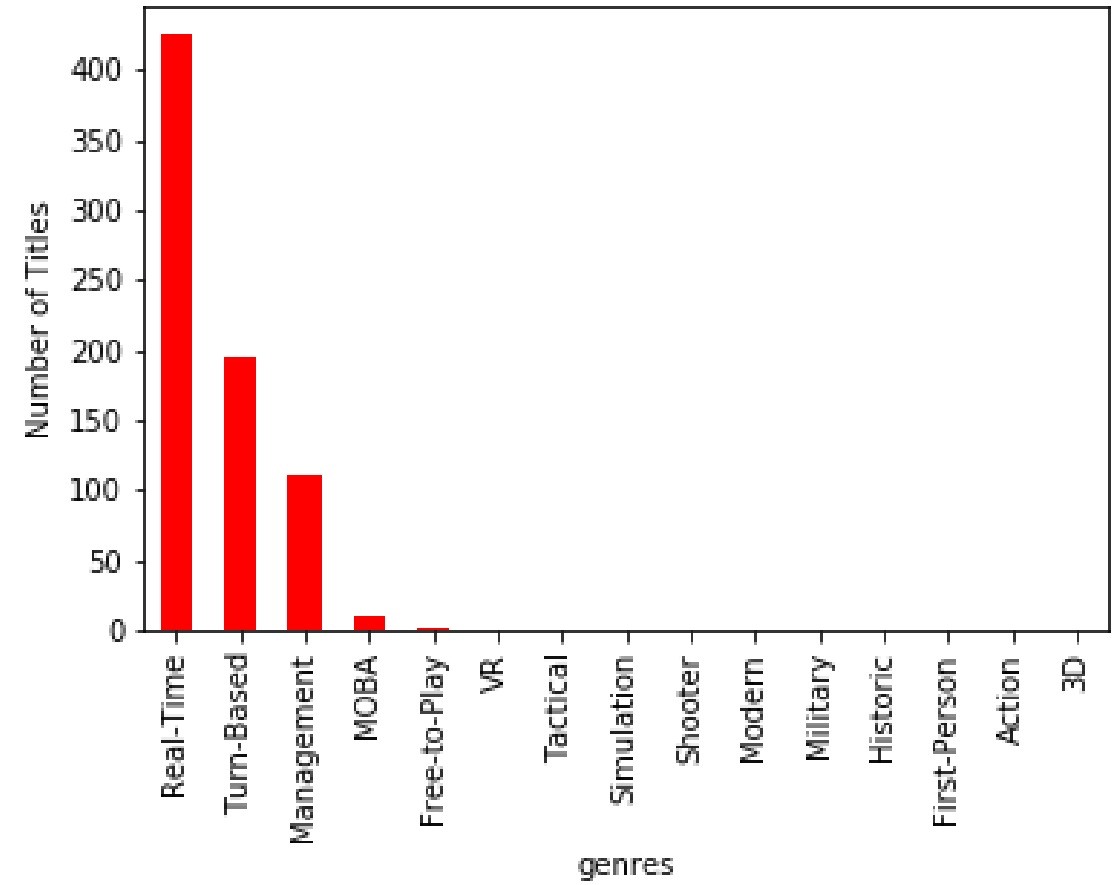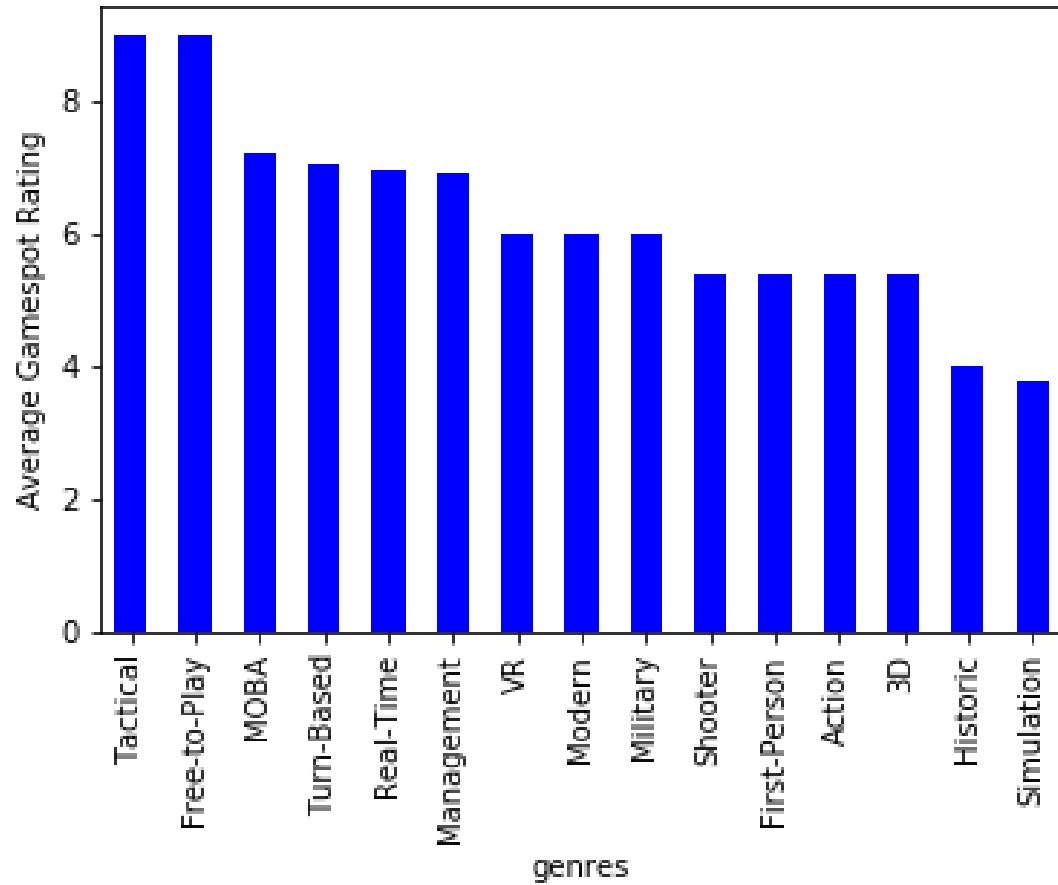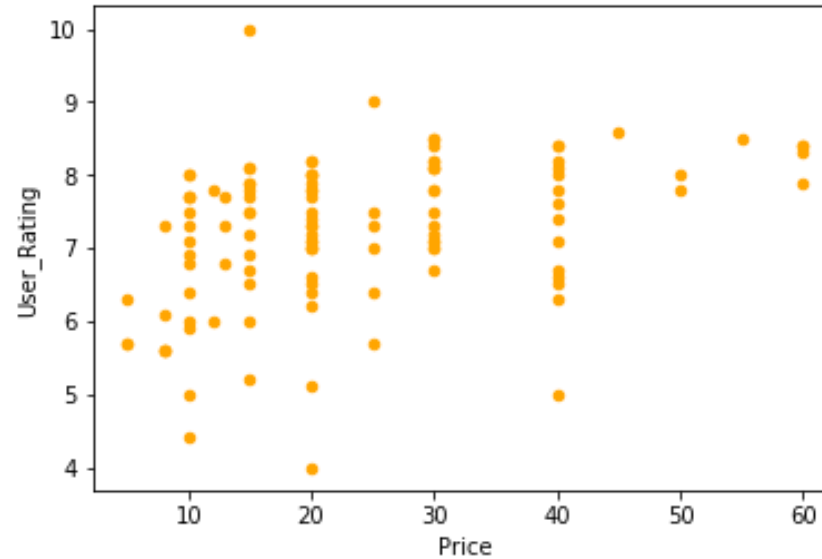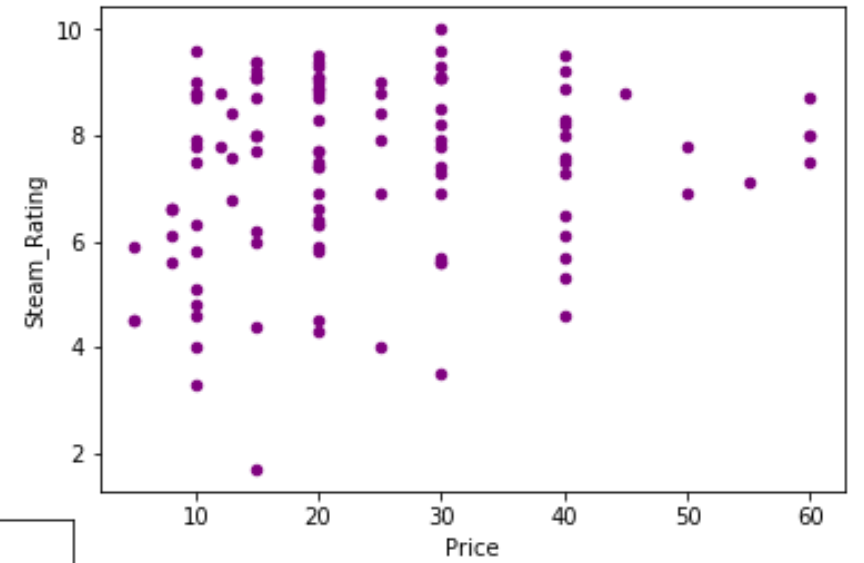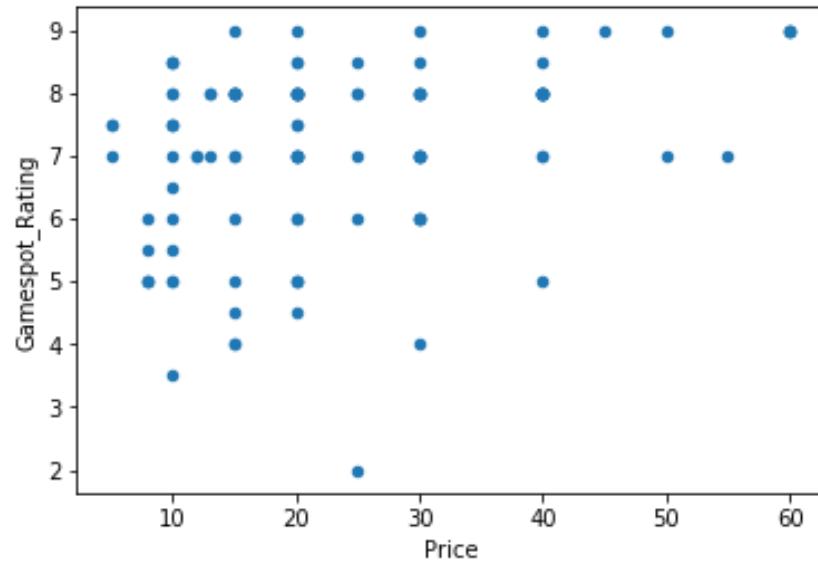
Categorical Data
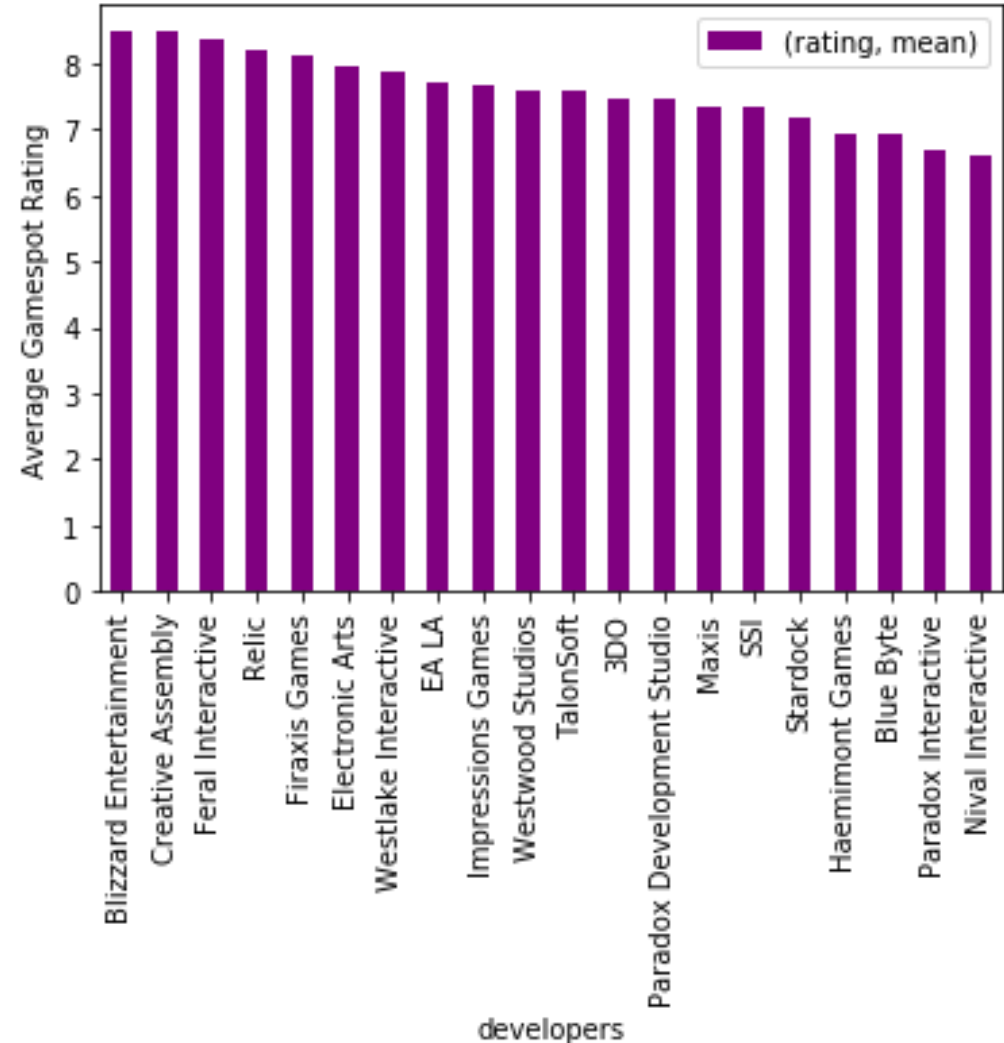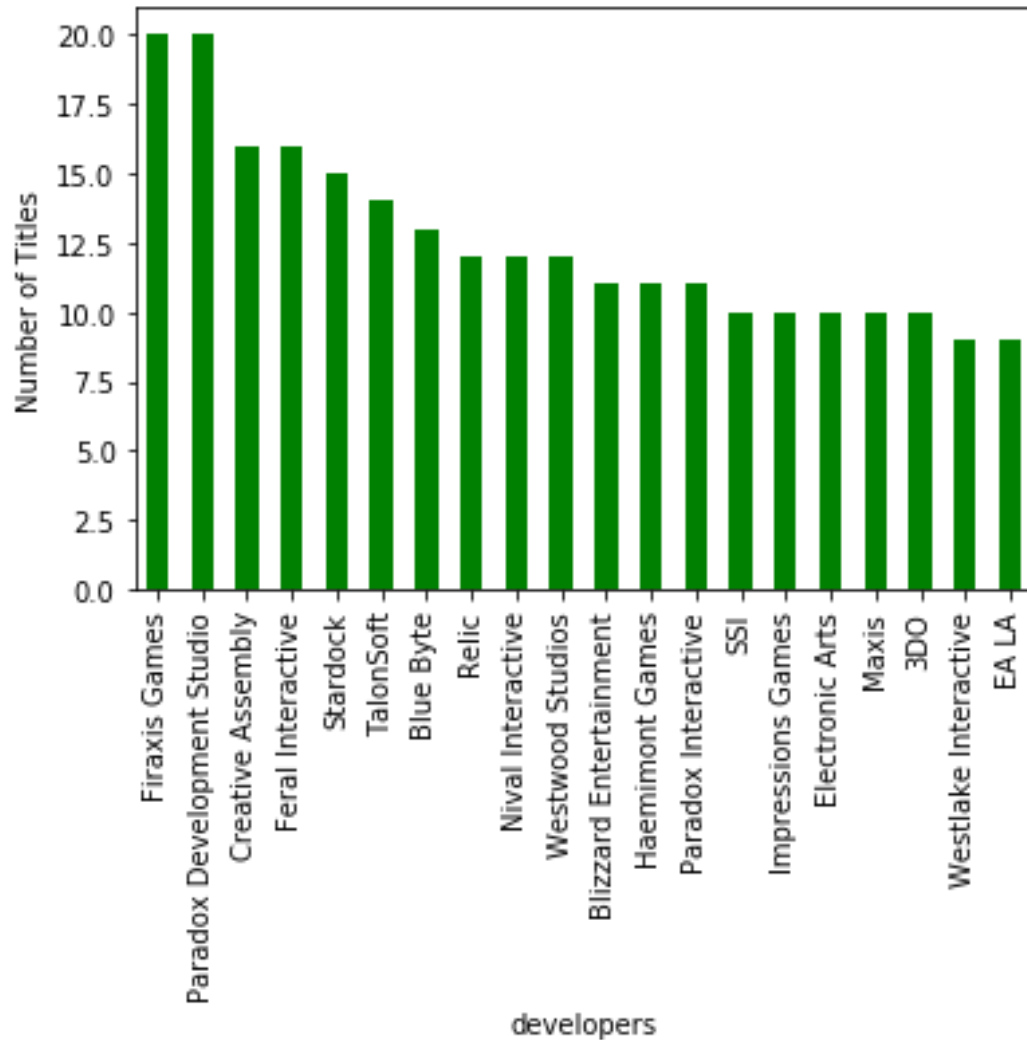
Gamespot Rating vs. User Ratings
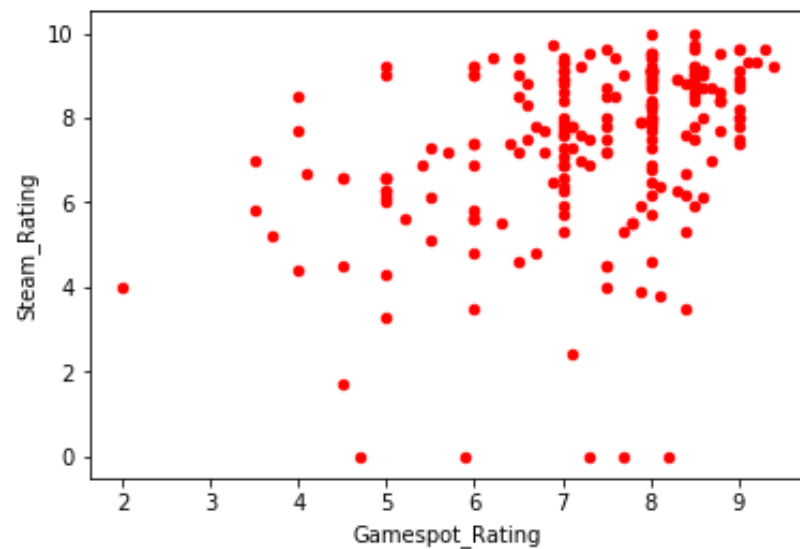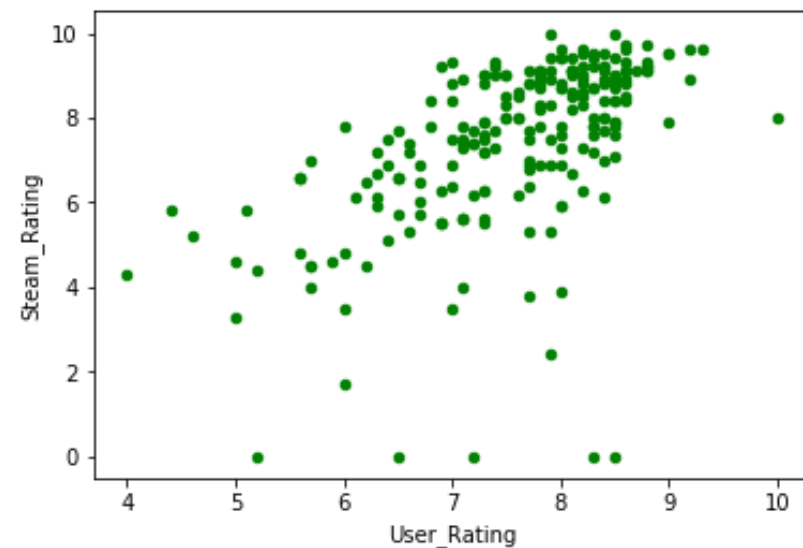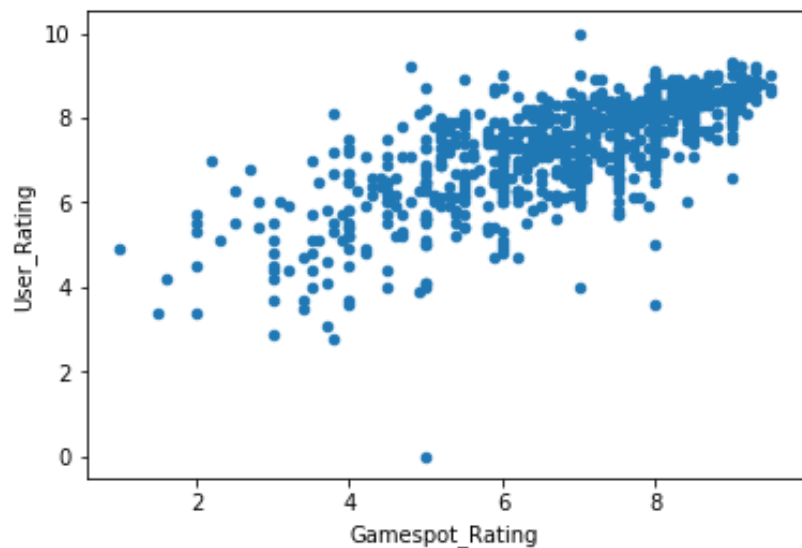
Price vs. Ratings
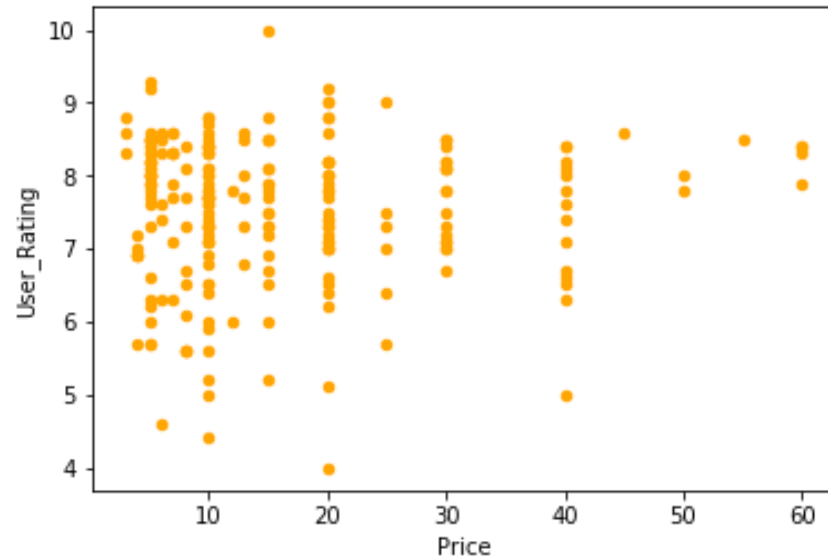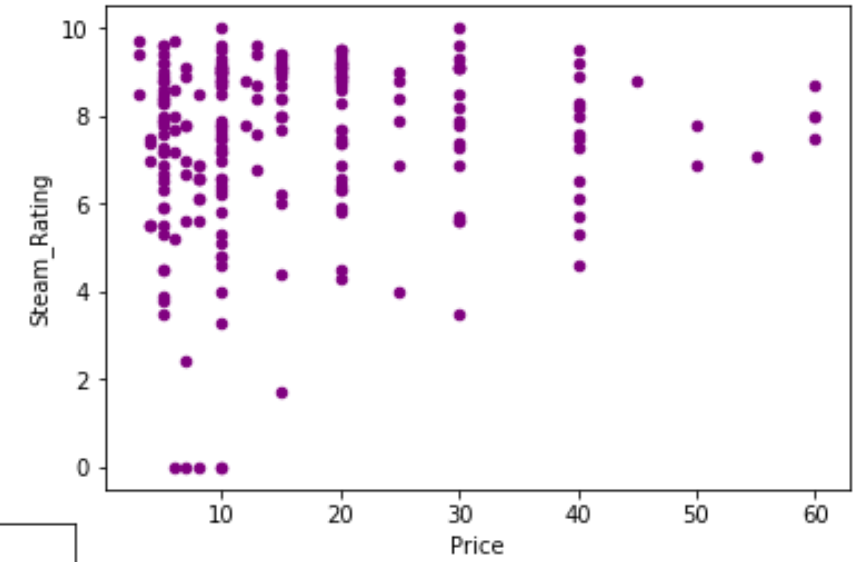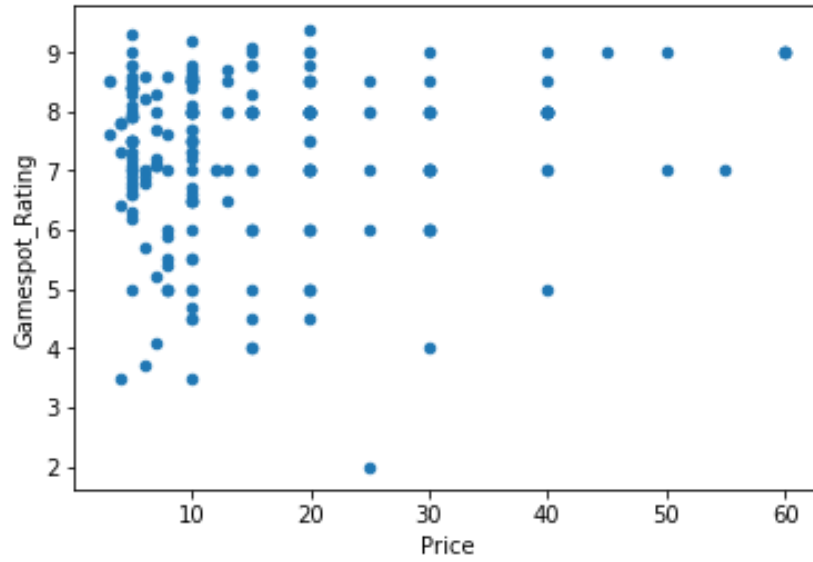
Price vs. Ratings

Correlations

Before Date Filter

	Gamespot_Rating	User_Rating	Steam_Rating	Price
Gamespot_Rating	1.000000	0.663504	0.389354	0.134097
User_Rating	0.663504	1.000000	0.562863	0.043872
Steam_Rating	0.389354	0.562863	1.000000	0.108329
Price	0.134097	0.043872	0.108329	1.000000

After Date Filter

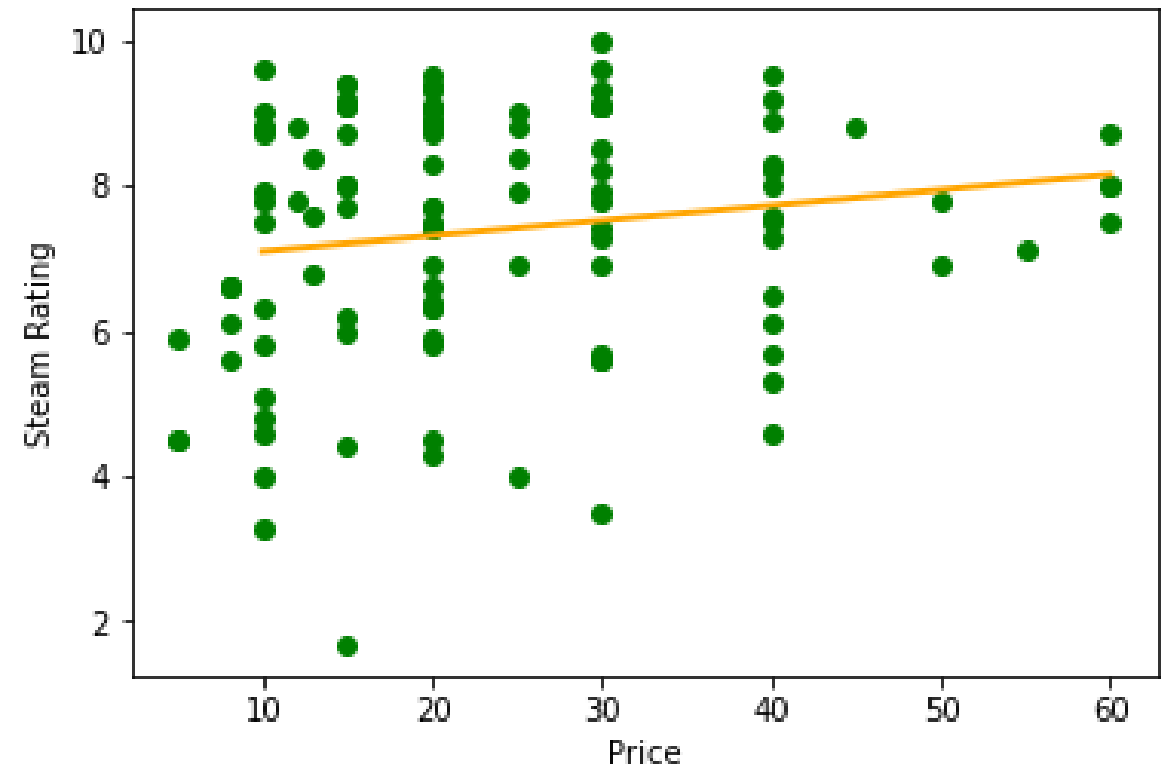|	Gamespot_Rating	User_Rating	Steam_Rating	Price
Gamespot_Rating	1.000000	0.613318	0.552696	0.333582
User_Rating	0.613318	1.000000	0.696225	0.375938
Steam_Rating	0.552696	0.696225	1.000000	0.163598
Price	0.333582	0.375938	0.163598	1.000000

Simple Linear Model

- Defined linear model function for price and user rating
  - Slope: 0.029
  - Intercept: 6.56

- Defined t-test function
  - T-statistic: 4.29
  - P value: 7.26×10^{-5}

- Reject null hypothesis
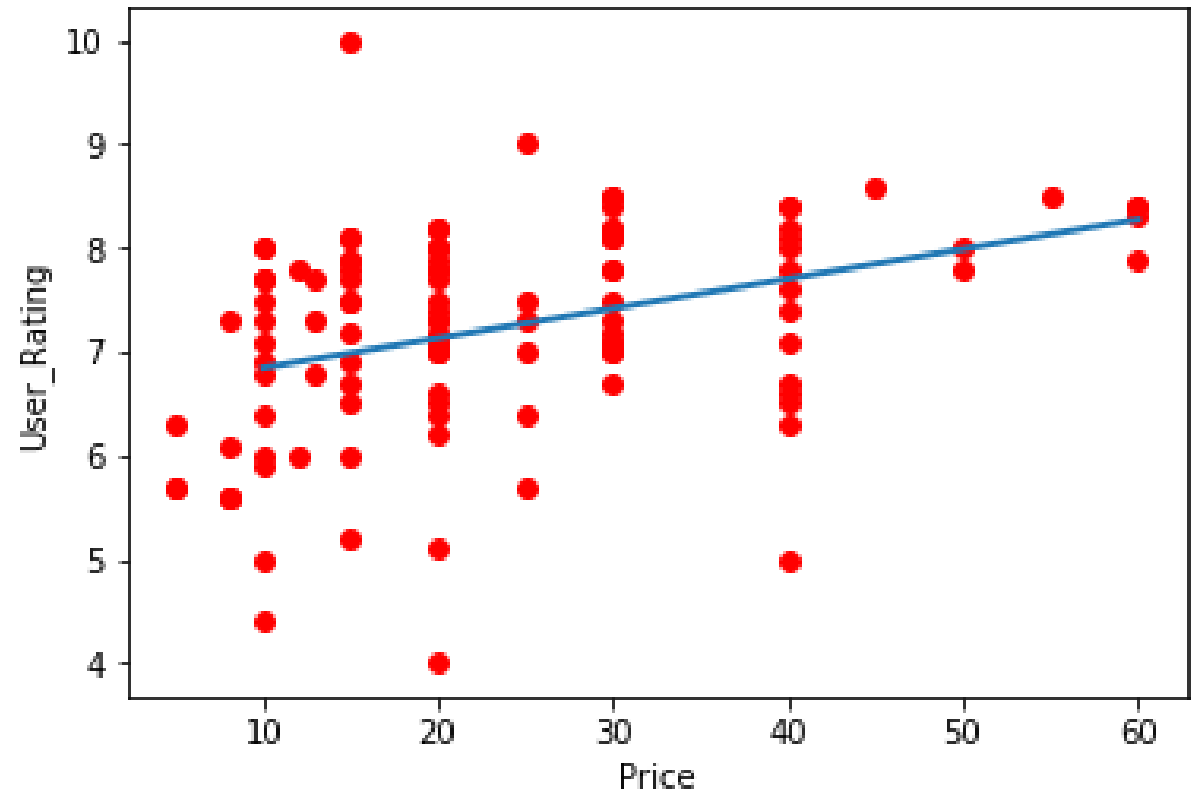
Simple Linear Model

- For price and Steam rating
  - Slope: 0.021
  - Intercept: 6.89

- T-test function
  - T-statistic: 1.755
  - P value: 0.086

- Cannot reject null hypothesis

Conclusions / Future Work

- Definite correlation between Gamespot rating and user rating
- Steam rating is more volatile and unpredictable
- Some evidence of correlation between price and user rating
- More analysis on current data:
  - Effect of popularity (number of reviews)
  - Developer vs. price, genre vs. price
- Not enough data to make predictive model
  - Not all games sold on Steam
  - Include more game genres
  - Need more complete pricing information