

Netflix: Exploring Movie Popularity and Revenue Trends

Cari Bean
Sarah Suo
Susie Chen

Table of Contents

[Table of Contents](#)

[Dataset Description: Columns and Rows](#)

[Project Goals and Purpose](#)

[Related Works](#)

[Results](#)

[Figure 1: Cluster Column Chart of Average Revenue by Genre and Year](#)

[Figure 2: Line Graph of Average Popularity by Month and Genre](#)

[Figure 3: Bubble Chart of Movie Budget, Revenue, and Popularity](#)

[Dashboard, Conclusions, & Actionable Recommendations](#)

[Appendix](#)

Dataset Description: Columns and Rows

The [Movies Dataset](#) on Kaggle contains multiple .csv files, with a main file containing metadata on over 45,000 movies released on or before July 2020. The files also contain 26 million ratings from over 270,000 users.

This dataset consists of multiple .csv files but we will be using the following files:

- Movies_metadata.csv
- Keywords.csv
- Credits.csv
- Links.csv
- Links_small.csv
- ratings_small.csv

The main data file we will be working with is the movies_metadata.csv which will be referred to as the main Movies Metadata file. The file contains information on 45,000 movies featured in the Full MovieLens dataset. The columns are ID (integer), posters (string), budget (integer), revenue (integer), release dates (date), languages (string), production countries (string), genres (string), and companies (string).

The dataset consists of 45,466 rows of data. Each row contains details of a specific movie that has been released. The link to the entire data set can be found in Appendix A and a snapshot of the Movies Metadata file can be found in Appendix B. The main data file used for this project is movies_metadata.csv and is attached to the assignment submission

Project Goals and Purpose

Netflix, as a leading streaming platform, needs to make data-driven decisions regarding content prioritization, production, and marketing strategies. My team is tasked with building a dashboard that enables marketing managers to make these decisions. After interviewing with the VP of Marketing at Netflix, the business needs include:

- Be able to view revenue trends across different content genres from the years 2014 to 2017, informing Netflix's content strategy by identifying growth areas.
- Be able to view popularity trends of movie content over time, facilitating insights into viewer preferences and engagement levels, guiding content release and promotion strategies.
- Be able to view the performance of movie content based on budget, revenue, and popularity rating, aiding in resource allocation and prioritization for Netflix Originals.

These visualizations collectively empower Netflix to maintain its competitive edge and drive continued success in the streaming industry.

Related Works

Others have used the data set to observe whether there is a relationship between revenue and genre, but have failed to utilize other details such as keywords. We aim to take a look at more aspects of considerations for Netflix in making original content. Along with an expanded list of considerations, our goal is to create user-friendly visualizations with interactive aspects that many others have not utilized.

Results

Figure 1: Cluster Column Chart of Average Revenue by Genre and Year

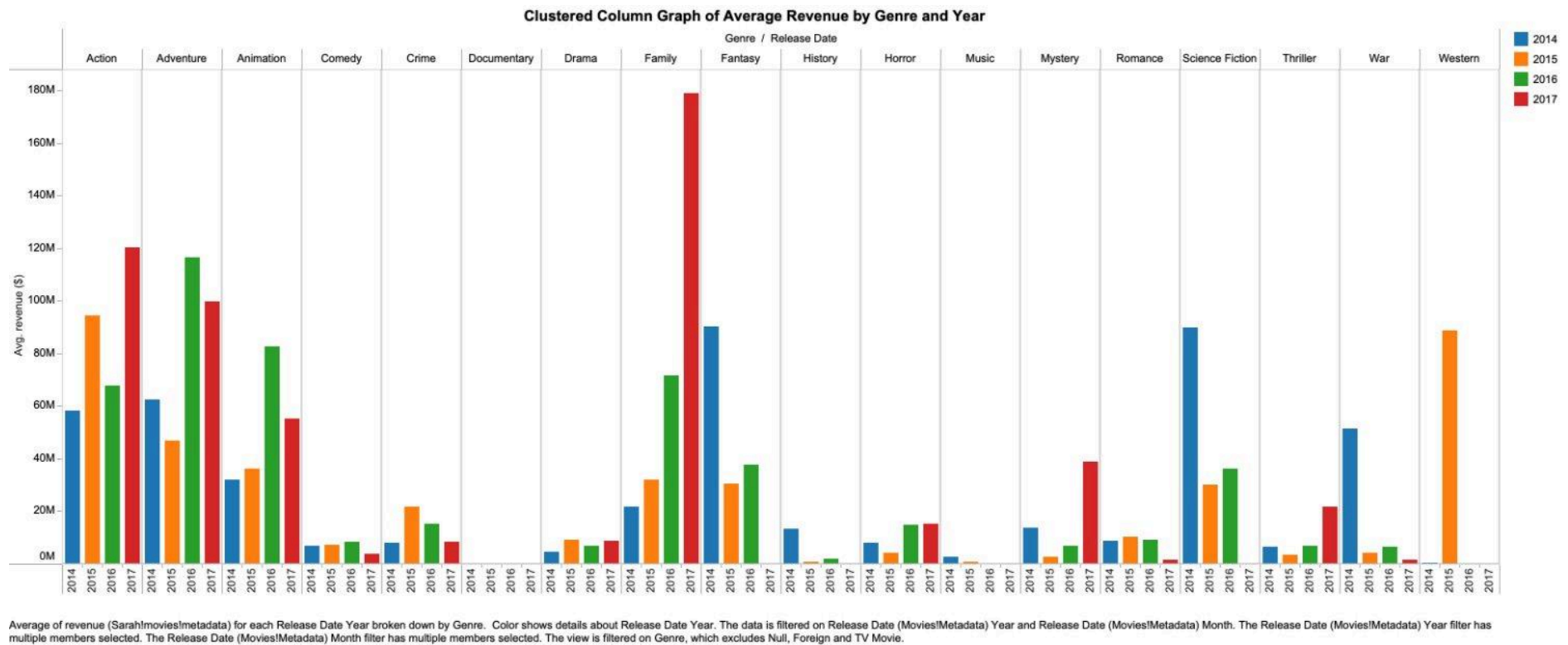


Figure 1 (zoomed-in version can also be found in Appendix C) features a column graph that shows the average revenue by the type of genre. Each genre features four years, 2014-2017, and you can filter to show specific years and months. The four years are colored in four different types of colors to assist in the readability of the chart. Years 2014, 2015, 2016, and 2017 are colored blue, orange, blue, and teal respectively. The taller the bar, the more average revenue was generated for the specific genre for that year. For example, in 2017, the 'Family' genre teal bar was the tallest so that means that the 'Family' genre generated the most revenue in not only 2017, but amongst all genres in the four years. More importantly, within the 'Family' genre cluster, the genre's revenue has increased exponentially since 2014. This signals Netflix to push out more 'Family' genres as they have generated more revenue throughout the years.

The graph assists Netflix in identifying genre trends that will yield a higher average revenue when considering which original movie pitches to invest in.

The columns utilized from the Movies Metadata file are Genres, Revenue, and Release Date.

Figure 2: Line Graph of Average Popularity by Month and Genre

Average Popularity Rating of Genre by Month of Release

Average Popularity

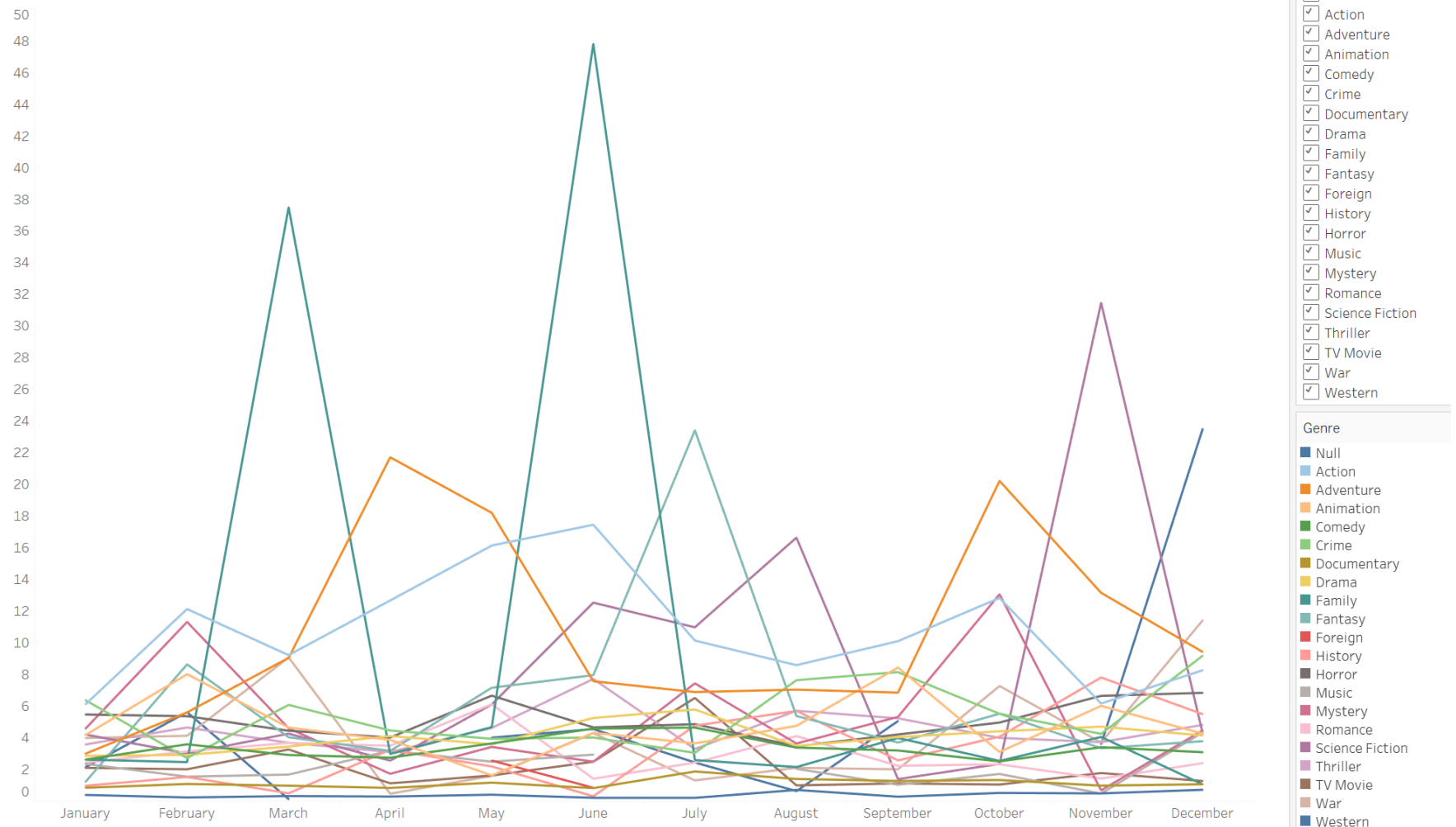


Figure 2 (zoomed-in version can also be found in Appendix D) is a line graph showing the average popularity of each genre of film over the months of the year. Each genre is a different colored line, and there is a filter for the genre, allowing the audience to pick which genres show on the graph. The taller the line for a genre is, the more popular that genre was during that particular month on the x-axis. For example, we can see that the 'Family' genre of movies had higher average popularity during March and June, but they had less popularity for the rest of the months of the year. We could infer from this that movies released in the genre of 'Family' have more popularity and more people watching during those two months. Similarly, movies in the genre of 'Science Fiction' have a higher average popularity during November. Using the filter, the audience can easily compare the average popularity of certain genres over the months of the year.

One of our project goals is to be able to view popularity trends of movie content over time, and the graph shown in Figure 2 accomplishes this. The graph can be used to show the popularity trends of movie genres over the months of the year. The graph can then be used to show viewer preferences for movie genres over a year. Netflix can use this information to determine what genre of movie they should release and when to generate the most user engagement and potential popularity, and from there they can more accurately determine their promotion schedules and strategies for their content.

The columns from the Movies Metadata file used in creating the graph in Figure 2 include Genres, Release Date, and Popularity.

Figure 3: Bubble Chart of Movie Budget, Revenue, and Popularity

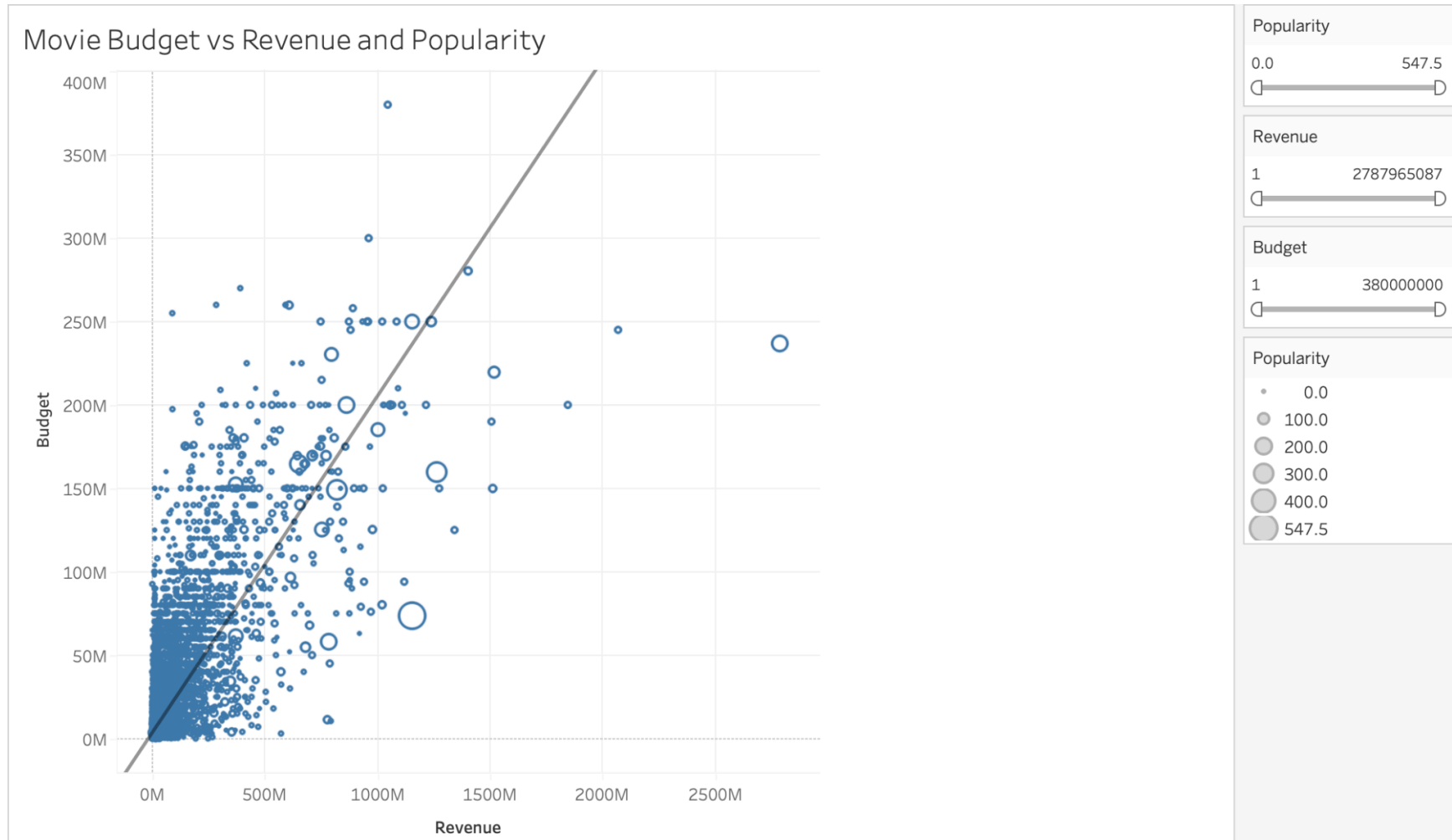


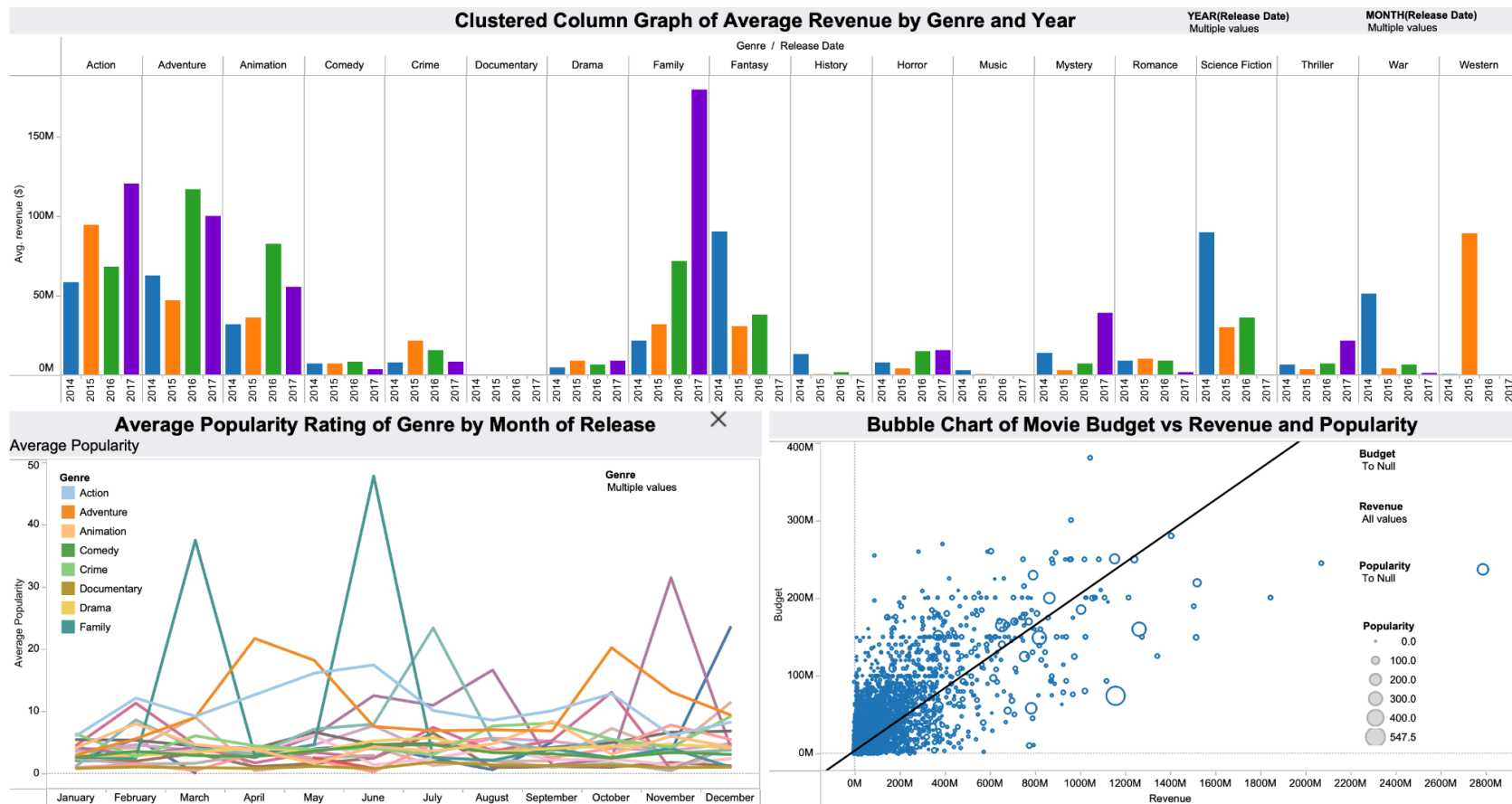
Figure 3 (zoomed-in version can also be found in Appendix E) is a Bubble Chart that examines the relationship between 3 quantitative variables: Movie budget, revenue, and popularity. Budget lies along the y-axis, revenue lies along the x-axis, and popularity is illustrated by the size of the “bubbles” on the chart. A smaller bubble equates to a lower popularity rating while a larger bubble equates to a higher popularity rating. Based on the Bubble Chart, we can see that generally as budget size increases, so does revenue. Since over 40,000 data records are being plotted on this graph, it is difficult to distinguish the bubble sizes and popularity ratings. Upon closer examination of popularity using the filter, we see that there is a general trend that as budget size increases, revenue increases and popularity ratings also increase. However, some data points in the dataset deviate from this trend. For example, the biggest bubble on the bubble chart is associated with the Minions movie released in June 2015. This movie has a more conservative movie budget in the dataset, but made much more in revenue compared to movies with similar budgets, and was widely more popular than any movie in the dataset. Movies that exhibit these characteristics should be closely monitored by Netflix managers when it comes to developing movie content strategy.

This bubble chart helps enable Netflix managers to view the performance of movie content based on budget, revenue, and popularity rating. Managers can use this chart to generally determine how much money to spend on making an original and assess the ROI in terms of revenue and audience popularity aiding in resource allocation and prioritization for Netflix Originals.

The columns used are budget, revenue, and popularity which can be found in the Movies Metadata file.

Dashboard, Conclusions, & Actionable Recommendations

Netflix Originals



The dashboard that our team has built contains 3 graphs and addresses the following business needs:

1. The Clustered Column Chart enables users to view revenue trends across different content genres from the years 2014 to 2017, informing Netflix's content strategy by identifying growth areas.
2. The Line Graph enables users to view popularity trends of movie content over time, facilitating insights into viewer preferences and engagement levels, guiding content release and promotion strategies.
3. The Bubble Chart enables users to view the performance of movie content based on budget, revenue, and popularity rating, aiding in resource allocation and prioritization for Netflix Originals.

Based on the dashboard above, we can draw the following insights from each graph:

1. Clustered Column Chart: The Genre Family resulted in the highest average revenue of any genre in the data set in 2017.
2. Line Graph: The Genre Family has a higher average popularity during March and June
3. Bubble Chart: The biggest bubble is the Minions movie released in June 2015 under the genre Family with a more conservative movie budget (approx. \$74,000,000), but made more in revenue (\$1,156,730,962) compared to other movies with similar budget size and was more popular than any movie in the dataset.

Based on these insights, Netflix should produce and release more original Family genre content during March and June when families are together for spring and summer break since this genre gives the highest ROI in terms of revenue and audience popularity.

Appendix

Appendix A: The link to the Data file can be found at <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>

The main data file used for this project is movies_metadata.csv and is attached to the assignment submission, but the entire data file can be found at the link above.

Appendix B: Snapshot of the Movies Metadata file

ExcelFileEditViewInsertFormatToolsDataWindowHelp

AutoSave

movies_metadata — Saved to my Mac

HomeInsertDrawPage LayoutFormulasDataReviewViewAutomateTell me

Paste

Aptos Narrow (Bod... | 12

B I U

General

Conditional Formatting

Insert

Delete

Format

Sort & Filter

Find & Select

Sensitivity

Add-ins

Analyze Data

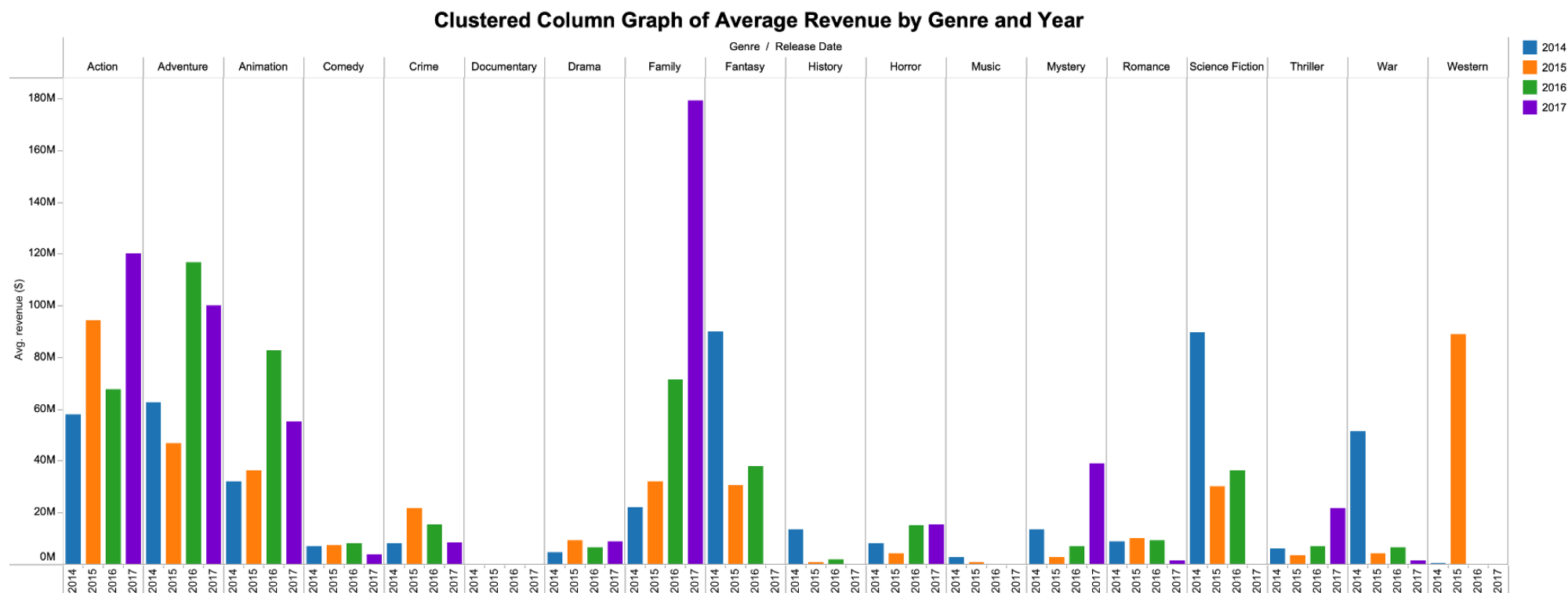
Send to NVivo

A1

adult

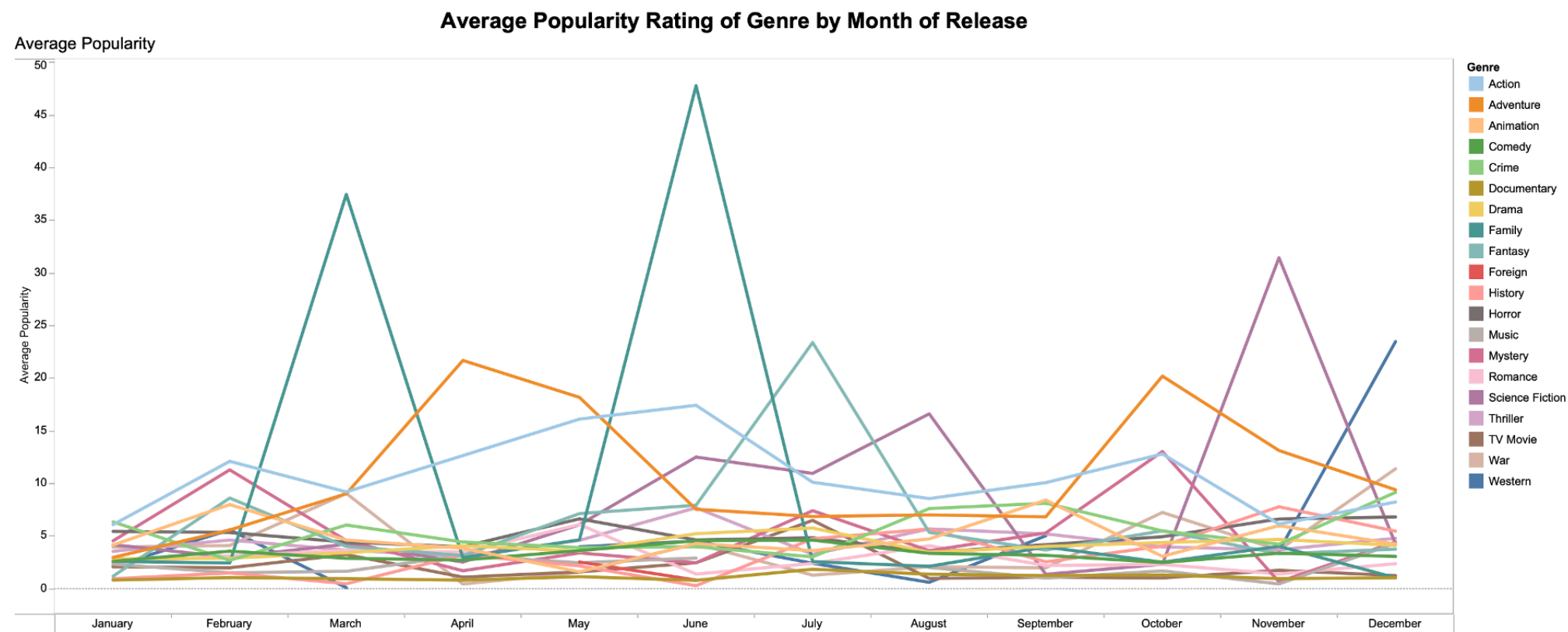
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	adult	belongs_to_c	budget	genres	homepage	id	imdb_id	original_lang	original_title	overview	popularity	poster_path	production_c	production_c	release_date	revenue	runtime	spoken_lang	status	tagline	title	video	vote_average	vote_count			
2	FALSE	{'id': 10194,	30000000	{'id': 16, 'na	http://toysto		862	tt0114709	en	Toy Story	Led by Wooo	21.946943	/rhlRbce0E9	{'name': 'Pb	{'iso_3166	10/30/95	373554033	81	{'iso_339_1' Released	Toy Story	FALSE	7.7	5415				
3	FALSE	{'id': 10194,	65000000	{'id': 12, 'name': 'Advent			8844	tt0113497	en	Jumanji	When sibling	17.015539	/vzmL6fP7aI	{'name': 'Trl	{'iso_3166	12/15/95	262797249	104	{'iso_339_1' Released	Roll the dice Jumanji	FALSE	6.9	2413				
4	FALSE	{'id': 119050	0	{'id': 10749, 'name': 'Ron			15602	tt0113228	en	Grumpier Ol	A family wec	11.7129	/6kxm1qJKM	{'name': 'W	{'iso_3166	12/22/95	0	101	{'iso_339_1' Released	Still Yelling, S	FALSE	6.5	92				
5	FALSE	{'id': 96871,	16000000	{'id': 35, 'name': 'Comed			31357	tt0114885	en	Waiting to Ex	Cheated on	3.859495	/fXOMpEaI	{'name': 'Tw	{'iso_3166	12/22/95	81452156	127	{'iso_339_1' Released	Friends are t	FALSE	6.1	34				
6	FALSE	{'id': 96871,	0	{'id': 35, 'name': 'Comed			11862	tt0113041	en	Father of the	Just when G	8.387519	/e64sO48Hc	{'name': 'Sa	{'iso_3166	2/10/95	76578911	106	{'iso_339_1' Released	Just When H	FALSE	5.7	173				
7	FALSE	{'id': 96871,	60000000	{'id': 28, 'name': 'Action'			949	tt0113277	en	Heat	Obsessive m	17.924927	/zMyfPUelun	{'name': 'Rr	{'iso_3166	12/15/95	187436818	170	{'iso_339_1' Released	A Los Angeles	FALSE	7.7	1886				
8	FALSE	{'id': 96871,	58000000	{'id': 35, 'name': 'Comed			11860	tt0114319	en	Sabrina	An ugly duck	6.672727	/Qh15ySYB	{'name': 'Pa	{'iso_3166	12/15/95	0	127	{'iso_339_1' Released	You are cord	FALSE	6.2	141				
9	FALSE	{'id': 96871,	0	{'id': 28, 'name': 'Action'			45325	tt0112302	en	Tom and Huc	A mischievo	2.561161	/sGO5Qa55j	{'name': 'W	{'iso_3166	12/22/95	0	97	{'iso_339_1' Released	The Original	FALSE	5.4	45				
10	FALSE	{'id': 96871,	35000000	{'id': 28, 'name': 'Action'			9091	tt0114576	en	Sudden Dea	International	5.23158	/eOWKUD60I	{'name': 'W	{'iso_3166	12/22/95	64350171	106	{'iso_339_1' Released	Terror goes i	FALSE	5.5	174				
11	FALSE	{'id': 96871,	58000000	{'id': 12, 'na	http://www		710	tt0113189	en	GoldenEye	James Bond	14.686036	/5c0oy741K	{'name': 'Ur	{'iso_3166	11/16/95	352194034	130	{'iso_339_1' Released	No limits. Nc	FALSE	6.6	1194				
12	FALSE	{'id': 96871,	62000000	{'id': 35, 'name': 'Comed			9087	tt0112346	en	The America	Widowed U	6.318445	/JymPngLZg	{'name': 'C	{'iso_3166	11/17/95	107879496	106	{'iso_339_1' Released	Why can't th	FALSE	6.5	199				
13	FALSE	{'id': 96871,	0	{'id': 35, 'name': 'Comed			12110	tt0112896	en	Dracula: De	When a lawy	5.430331	/xve4CgYttn	{'name': 'C	{'iso_3166	12/22/95	0	88	{'iso_339_1' Released	Dracula: De	FALSE	5.7	210				
14	FALSE	{'id': 96871,	0	{'id': 10751, 'name': 'Fan			21032	tt0112453	en	Balto	An outcast h	12.140733	/gP5PCAVCP	{'name': 'U	{'iso_3166	12/22/95	11348324	78	{'iso_339_1' Released	Part Dog, Pa	FALSE	7.1	423				
15	FALSE	{'id': 96871,	44000000	{'id': 36, 'name': 'History'			10858	tt0113987	en	Nixon	An all-star c	5.092	/cICkmCEIXf	{'name': 'H	{'iso_3166	12/22/95	13681765	192	{'iso_339_1' Released	Triumphant i	FALSE	7.1	72				
16	FALSE	{'id': 96871,	98000000	{'id': 28, 'name': 'Action'			1408	tt0112760	en	Cutthroat Is	Morgan Adar	7.284477	/odM9973klv	{'name': 'L	{'iso_3166	12/22/95	10017322	119	{'iso_339_1' Released	The Course f	FALSE	5.7	137				
17	FALSE	{'id': 96871,	52000000	{'id': 18, 'name': 'Drama'			524	tt0112641	en	Casino	The life of th	10.137389	/x0517ibXBf	{'name': 'U	{'iso_3166	11/22/95	116112375	178	{'iso_339_1' Released	No one stays	FALSE	7.8	1343				
18	FALSE	{'id': 96871,	16500000	{'id': 18, 'name': 'Drama'			4584	tt0114388	en	Sense and S	Rich Mr. Das	10.673167	/A9HtY848Bf	{'name': 'C	{'iso_3166	12/13/95	135000000	136	{'iso_339_1' Released	Lose your he	FALSE	7.2	364				
19	FALSE	{'id': 96871,	4000000	{'id': 80, 'name': 'Crime'			5	tt0113101	en	Four Rooms	It's Ted the B	0.265686	/eQsShh9nxr	{'name': 'M	{'iso_3166	12/9/95	4300000	98	{'iso_339_1' Released	Twelve outra	FALSE	6.5	539				
20	FALSE	{'id': 96871,	30000000	{'id': 80, 'name': 'Crime'			9273	tt0112281	en	Ace Ventura	Summoned	8.205448	/wRIGnJhEzc	{'name': 'O	{'iso_3166	11/10/95	212385533	90	{'iso_339_1' Released	New animat	FALSE	6.1	1128				
21	FALSE	{'id': 96871,	60000000	{'id': 28, 'name': 'Action'			11517	tt0113845	en	Money Train	Avengeful N	12.337906	/jSozzVOR2	{'name': 'C	{'iso_3166	11/21/95	35431113	103	{'iso_339_1' Released	Get on, or G	FALSE	5.4	224				
22	FALSE	{'id': 96871,	30250000	{'id': 35, 'name': 'Comed			8012	tt0113161	en	Get Shorty	Chili Palmer	12.669608	/wVDUuQg2	{'name': 'J	{'iso_3166	10/20/95	115101622	105	{'iso_339_1' Released	The mob is t	FALSE	6.4	305				
23	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			1710	tt0112722	en	Copycat	An agorapho	10.701801	/80czsJGsol	{'name': 'Rr	{'iso_3166	10/27/95	0	124	{'iso_339_1' Released	One man is c	FALSE	6.5	199				
24	FALSE	{'id': 96871,	50000000	{'id': 28, 'name': 'Action'			9691	tt0112401	en	Assassins	Assassin Rot	11.065939	/xXv5MP7Dg	{'name': 'Sll	{'iso_3166	10/6/95	30303072	132	{'iso_339_1' Released	In the shado	FALSE	6	394				
25	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			12665	tt0114168	en	Powder	Harassed by	12.133094	/iURKxOQt	{'name': 'C	{'iso_3166	10/27/95	0	111	{'iso_339_1' Released	An extraordi	FALSE	6.3	143				
26	FALSE	{'id': 96871,	3600000	{'id': 18, 'na	http://www		451	tt0113627	en	Leaving Las	Ben Sanders	10.332025	/37qKXrNSt	{'name': 'U	{'iso_3166	10/27/95	49800000	112	{'iso_339_1' Released	I Love You... Leaving Las	FALSE	7.1	365				
27	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			16420	tt0114057	en	Othello	The evil lago	1.845899	/qM08tEQJn	{'name': 'C	{'iso_3166	12/15/95	0	123	{'iso_339_1' Released	Envy, greed, Othello	FALSE	7	33				
28	FALSE	{'id': 96871,	12000000	{'id': 35, 'name': 'Comed			9263	tt0114011	en	Now and The	Waxing nost	8.681325	/wD6rLD2B	{'name': 'N	{'iso_3166	10/20/95	27400000	100	{'iso_339_1' Released	In every wor	FALSE	6.6	91				
29	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			17015	tt0114117	en	Persuasion	This film ad	2.228434	/sI8r11eZMx	{'name': 'Bf	{'iso_3166	9/27/95	0	104	{'iso_339_1' Released	Persuasion	FALSE	7.4	36				
30	FALSE	{'id': 96871,	18000000	{'id': 14, 'name': 'Fantasy'			902	tt0112682	fr	La Cité @ de	l'Ascientist	9.822423	/eVo6ewq4a	{'name': 'Pr	{'iso_3166	5/16/95	1738611	108	{'iso_339_1' Released	Where happ	FALSE	7.6	308				
31	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			37557	tt0115012	zh	Été à Áiá	Éd Á A provincial	1.100915	/qcoOCN7A	{'name': 'M	{'iso_3166	4/30/95	0	108	{'iso_339_1' Released	In 1930's Sh	FALSE	6.5	17				
32	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			9909	tt0112792	en	Dangerous	M Former Mari	9.481338	/ySlee3QmY	{'name': 'O	{'iso_3166	8/11/95	180000000	99	{'iso_339_1' Released	She broke th	FALSE	6.4	249				
33	FALSE	{'id': 96871,	29500000	{'id': 878, 'name': 'Scienc			63	tt0114746	en	Twelve Monk	In the year 2	12.297305	/6S9wDu3Y	{'name': 'U	{'iso_3166	12/29/95	168840000	129	{'iso_339_1' Released	The future is	FALSE	7.4	2470				
34	FALSE	{'id': 96871,	0	{'id': 10749, 'name': 'Ron			78802	tt0114952	fr	Guillaumet,	les alies du	0.745542	/k6ODR38d	{'name': 'W	{'iso_3166	9/18/96	0	50	{'iso_339_1' Released	Wings of Co	FALSE	6.8	4				
35	FALSE	{'id': 96871,	30000000	{'id': 14, 'name': 'Fantasy'			9598	tt0112431	en	Babe	Babe is a litt	14.404764	/gN6X3hvPy	{'name': 'U	{'iso_3166	7/18/95	254134910	89	{'iso_339_1' Released	A little pig	FALSE	6	756				
36	FALSE	{'id': 96871,	0	{'id': 36, 'name': 'History'			47018	tt0112637	en	Carrington	The story of	1.493361	/a7w6rPqTB	{'name': 'S	{'iso_3166	11/8/95	0	121	{'iso_339_1' Released	A Love so Un	FALSE	6.4	16				
37	FALSE	{'id': 96871,	11000000	{'id': 18, 'name': 'Drama'			687	tt0112818	en	Dead Man W	A Justice dra	6.891317	/y19uRKAHx	{'name': 'H	{'iso_3166	12/29/95	39363635	122	{'iso_339_1' Released	Dead Man W	FALSE	7.3	350				
38	FALSE	{'id': 96871,	0	{'id': 12, 'name': 'Advent			139405	tt0112286	en	Across the S	A young Russ	1.114469	/kOAYy098	{'name': 'S	{'iso_3166	10/20/95	0	51	{'iso_339_1' Released	Across the S	FALSE	3.5	2				
39	FALSE	{'id': 96871,	0	{'id': 35, 'name': 'Comed			33689	tt0113442	en	It Takes Two	Identical 9-y	7.777735	/B7WVL3Mjc	{'name': 'D	{'iso_3166	11/17/95	0	101	{'iso_339_1' Released	Two identica	FALSE	6.1	149				
40	FALSE	{'id': 96871,	12000000	{'id': 35, 'name': 'Comed			9603	tt0112697	en	Cleuesh	Shallow, ric	9.882383	/l8gEhZssz	{'name': 'B	{'iso_3166	7/19/95	0	97	{'iso_339_1' Released	Sex. Clothes	FALSE	6.9	828				
41	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			34615	tt0112749	en	Cry, the Bel	A South-Afric	0.894647	/ymATHQqr	{'name': 'M	{'iso_3166	12/15/95	676525	106	{'iso_339_1' Released	Cry, the Bel	FALSE	6.7	13				
42	FALSE	{'id': 96871,	0	{'id': 18, 'name': 'Drama'			31174	tt0114279	en	Richard III	Shakespear	4.561387	/e8nuS9QC	{'name': 'B	{'iso_3166	12/29/95	0	104	{'iso_339_1' Released	I can smile, r	FALSE	6.9	50				
43	FALSE	{'id': 96871,	10000000	{'id': 28, 'name': 'Action'			11443	tt0112819	en	Dead Presid	Depicts a he	9.87957	/dzlyxvXVX	{'name': 'C	{'iso_3166	10/6/95	0	119	{'iso_339_1' Released	In this daria	FALSE	6.6	80				
44	FALSE	{'id': 96871,	19000000	{'id': 18, 'name': 'Drama'			35196	tt0114272	en	Restoration	An aspiring y	10.979269	/a9QdEQ12	{'name': 'M	{'iso_3166	12/29/95	0	117	{'iso_339_1' Released	Restoration	FALSE	6.3	30				
45	FALSE	{'id': 96871,	18000000	{'id': 28, 'name': 'Action'			9312	tt0113855	en	Mortal Komt	For nine gen	10.870138	/bDXWlqJWV	{'name': 'N													

Appendix C: Figure 1



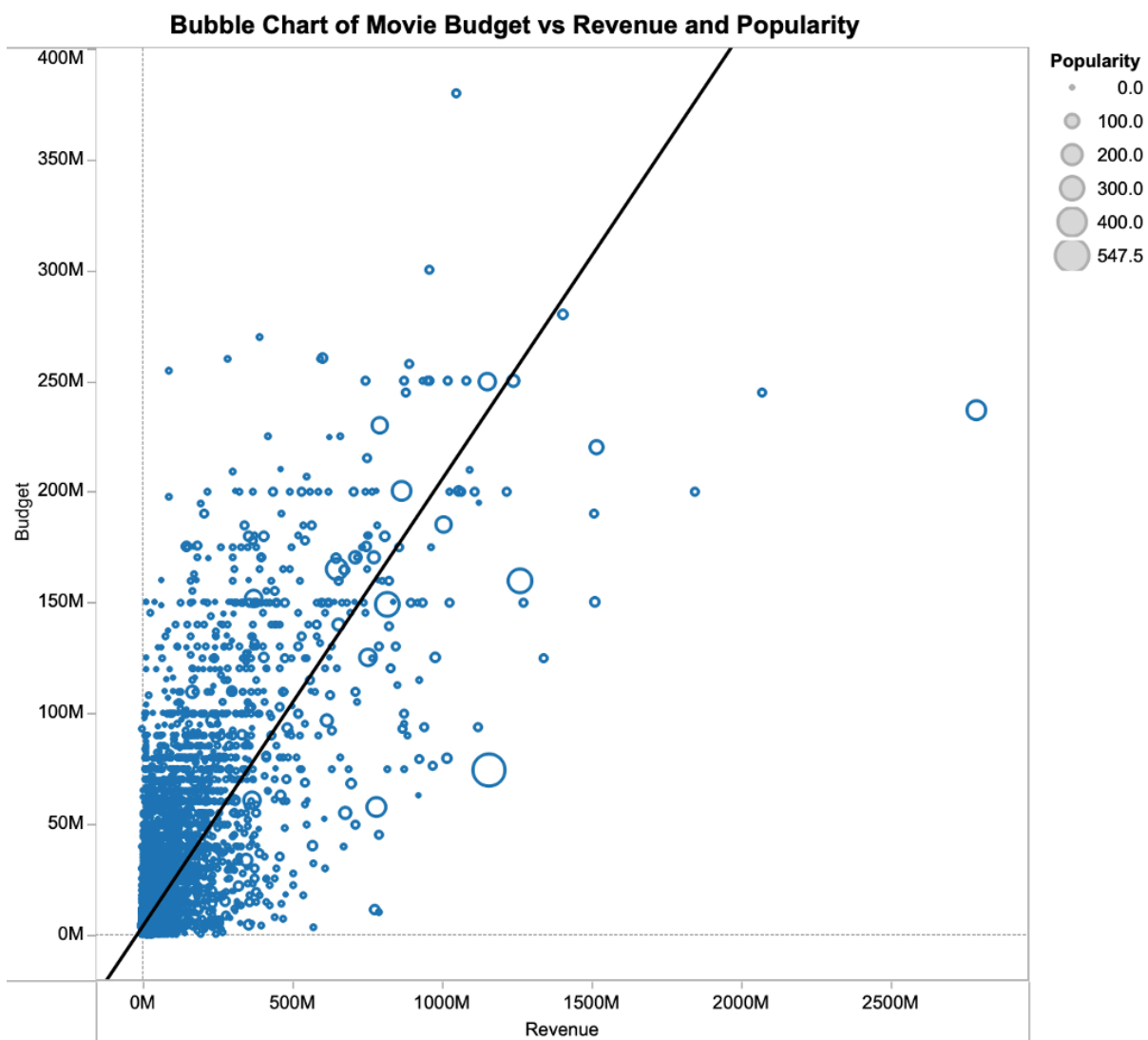
Average of revenue (Sarah!movies!metadata) for each Release Date Year broken down by Genre. Color shows details about Release Date Year. The data is filtered on Release Date (Movies!Metadata) Year and Release Date (Movies!Metadata) Month. The Release Date (Movies!Metadata) Year filter has multiple members selected. The Release Date (Movies!Metadata) Month filter has multiple members selected. The view is filtered on Genre, which excludes Null, Foreign and TV Movie.

Appendix D: Figure 2



The trend of average of Popularity for Release Date Month. Color shows details about Genre. The view is filtered on Genre, which excludes Null.

Appendix E: Figure 3



Revenue vs. Budget. Size shows Popularity. The data is filtered on Budget, Revenue and Popularity. The Budget filter includes everything. The Revenue filter keeps all values. The Popularity filter includes everything.