Business Analytics Project
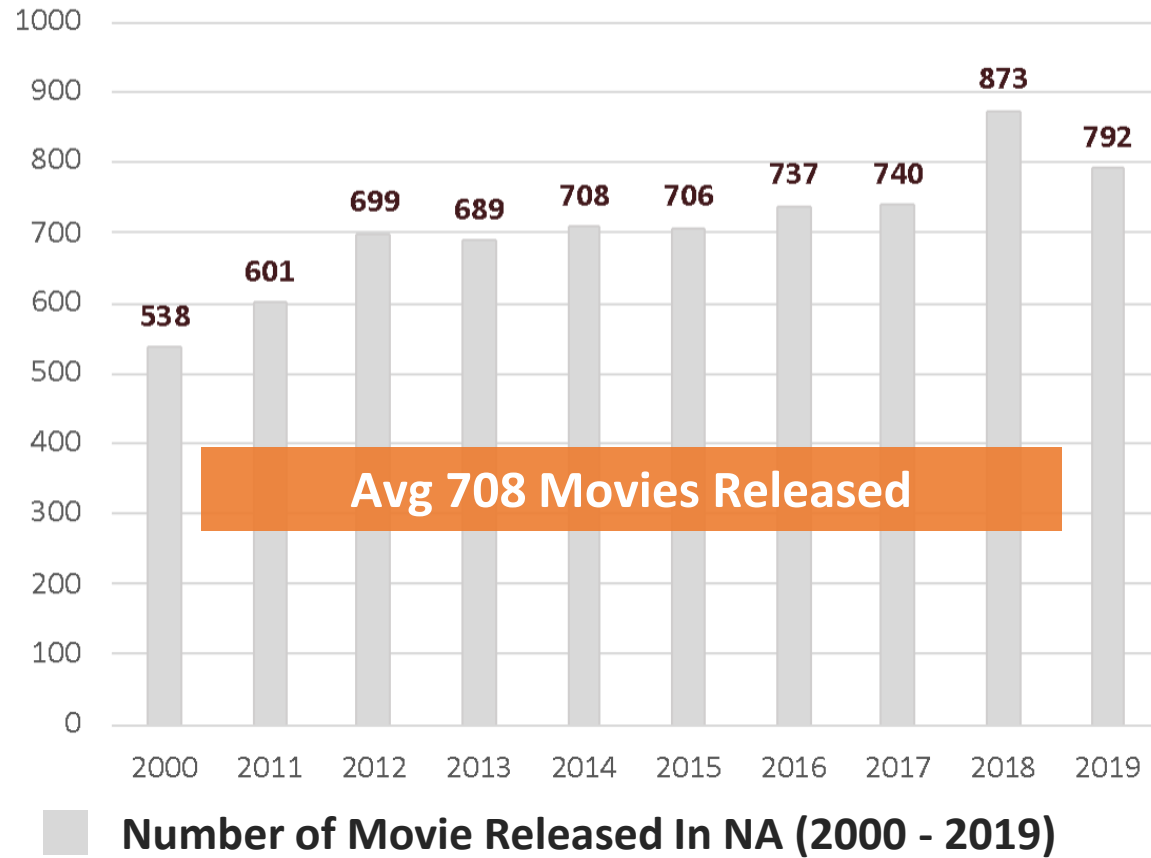## Success Factor for a Movie

**EunJeong Heo, Hyomin Shin,
Korbin Sorensen, Youngryun Choi**

1 Introduction

Number of Movie Released In NA (2000 - 2019)

Reference: Statista

80 %

of films lose money

Reference: Forbes

"Which factors leads to successful movie?"

1 Introduction

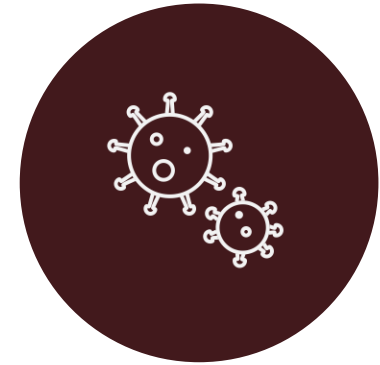**- Why does the research matter to movie production companies and directors?**

Business content of the research

**To gain
competitive advantage**

**To reach
break-even point**

**To be less affected by
the external
environment**

2 Background of the data
- Data Source and Variable

■ **Data Source: Kaggle**

  - Dataset Name: The Movies Dataset

■ **Explanation of variable**

- **Dependent: revenue, popularity [1]**
- **Independent: budget, runtime, vote_average [2], holiday [3], genre (Action, Adventure, Animation, Comedy, Crime, Documentary, Drama, Family, Fantasy, Foreign, History, Horror, Music, Mystery, Romance, Science Fiction, TV Movie, Thriller, War, Western)**

  [1] popularity: Calculated by dividing portion sales for a given item by the total portion sales in the same category (%)
  [2] vote_average: Average score of reviews by audience (out of 10)
  [3] holiday: Dummy variable which divides movies that is released in December or not

Background of the data
- Reason for each variables

■ **Why is variables important?**

Dependent Variable

Revenue

1) Criterion for
determining success

2) Factor of
break-even point

Popularity

1) Criterion for
determining success

2) Criteria of
competitive factor

Independent Variable

● **Runtime, Vote Average, Genre**

- Can be the criterion for selecting a movie for audience

● **Budget**

- Quality may vary depending on the amount

● **Holiday**

- Relatively more leisure time given to people

Background of the data
- Summary Statistics

	Variable	Min	Max	Mean	Std
Dependent	revenue	1	2,787,965,000	77,599,850	154,549,600
	popularity	0	547.4	9.19	12.43
Independent	budget	0	380000000	25721637.19	38800222.63
	runtime	0	338	107	20
	vote average	0	10	6	1
	Holiday	0	1	0.1	0.29
	Action	0	1	0.16	0.37
	Adventure	0	1	0.07	0.25
	Animation	0	1	0.02	0.15
	Comedy	0	1	0.22	0.41
	Crime	0	1	0.05	0.21
	Documentary	0	1	0.02	0.15
	Drama	0	1	0.25	0.43
	Family	0	1	0.01	0.11
	Fantasy	0	1	0.02	0.15
	Foreign	0	1	0.00	0.02
	History	0	1	0.00	0.07
	Horror	0	1	0.06	0.24
	Music	0	1	0.01	0.09
	Mystery	0	1	0.1	0.11
	Romance	0	1	0.2	0.15
	Science Fiction	0	1	0.02	0.13
	TV Movie	0	1	0.00	0.01
	Thriller	0	1	0.04	0.19
	War	0	1	0.01	0.08
	Western	0	1	0.01	0.08

(Genre)

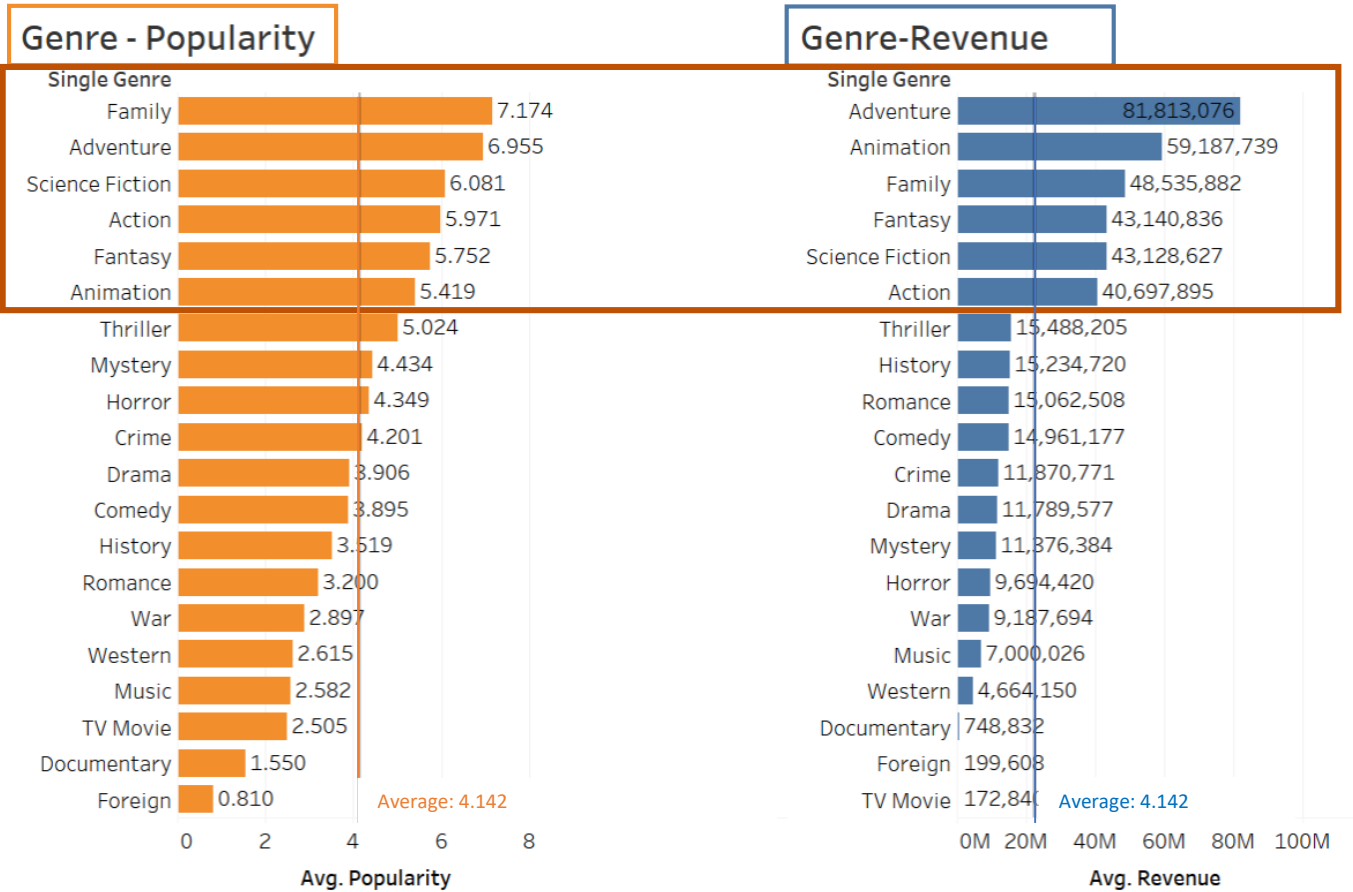* Holiday dummy variable does not have meaning of summary statistics

* Genre dummy variables only have meaning in the 'mean'. It shows the % of specific genre out the total 1.0

Background of the data
- Interesting observation & Visualization
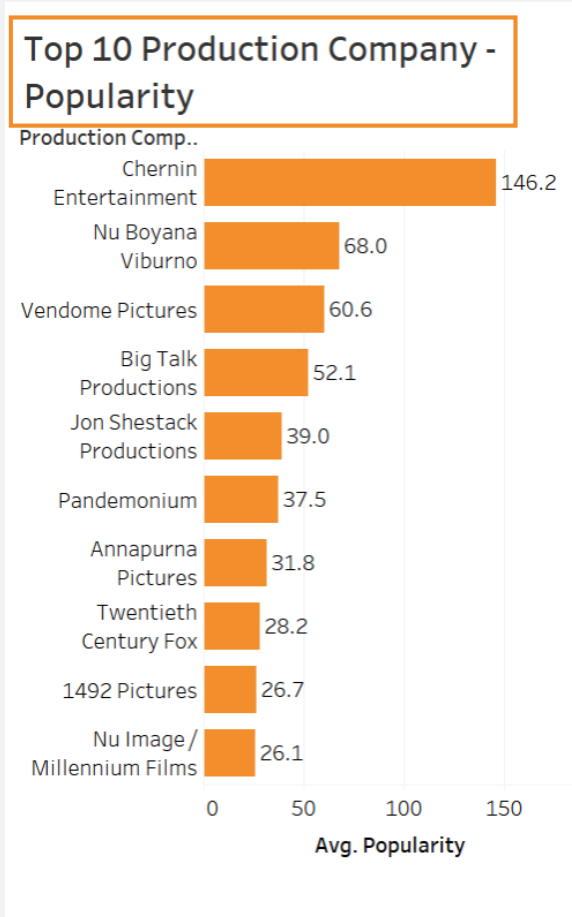
Genre : Popularity vs Revenue

Genre - Popularity		Genre-Revenue	
Single Genre	Avg. Popularity	**Single Genre**	Avg. Revenue
Family	7.174	Adventure	81,813,076
Adventure	6.955	Animation	59,187,739
Science Fiction	6.081	Family	48,535,882
Action	5.971	Fantasy	43,140,836
Fantasy	5.752	Science Fiction	43,128,627
Animation	5.419	Action	40,697,895
Thriller	5.024	Thriller	15,488,205
Mystery	4.434	History	15,234,720
Horror	4.349	Romance	15,062,508
Crime	4.201	Comedy	14,961,177
Drama	3.906	Crime	11,870,771
Comedy	3.895	Drama	11,789,577
History	3.519	Mystery	11,376,384
Romance	3.200	Horror	9,694,420
War	2.897	War	9,187,694
Western	2.615	Music	7,000,026
Music	2.582	Western	4,664,150
TV Movie	2.505	Documentary	748,832
Documentary	1.550	Foreign	199,608
Foreign	0.810	TV Movie	172,840

Average: 4.142 (Avg. Popularity)

Average: 4.142 (Avg. Revenue)

Top 6 genre categories are the same

* Alphabetical order
1) **Action**
2) **Adventure**
3) **Animation**
4) **Family**
5) **Fantasy**
6) **Science Fiction**

Top 10 Production Company : Popularity vs Revenue

Top 10 Production Company - Popularity

Production Comp..	Avg. Popularity
Chernin Entertainment	146.2
Nu Boyana Viburno	68.0
Vendome Pictures	60.6
Big Talk Productions	52.1
Jon Shestack Productions	39.0
Pandemonium	37.5
Annapurna Pictures	31.8
Twentieth Century Fox	28.2
1492 Pictures	26.7
Nu Image / Millennium Films	26.1

No Direct Correlation

Top 10 Production Company - Revenue

Production Comp..	Avg. Revenue
Patalex IV Productions Limi..	895,921,036
1492 Pictures	704,386,834
WingNut Films	449,270,951
Chernin Entertainment	369,907,963
Bookshop Productions	363,889,678
Lucasfilm	353,515,040
Kemp Company	352,114,312
Weed Road Pictures	324,238,135
Lin Pictures	311,950,384
Pacific Data Images (PDI)	301,301,418

Background of the data
- Interesting observation & Visualization

Top 10 Director : Popularity vs Revenue

Top 10 Director Name - Popularity

Dir Name

Director	Avg. Popularity
Chad Stahelski	183.9
James Ponsoldt	88.4
David Fincher	86.0
Bill Condon	81.0
Wes Ball	76.9
Anthony Russo	74.5
Quentin Tarantino	38.7
Tommy Wirkola	35.5
John R. Leonetti	33.1
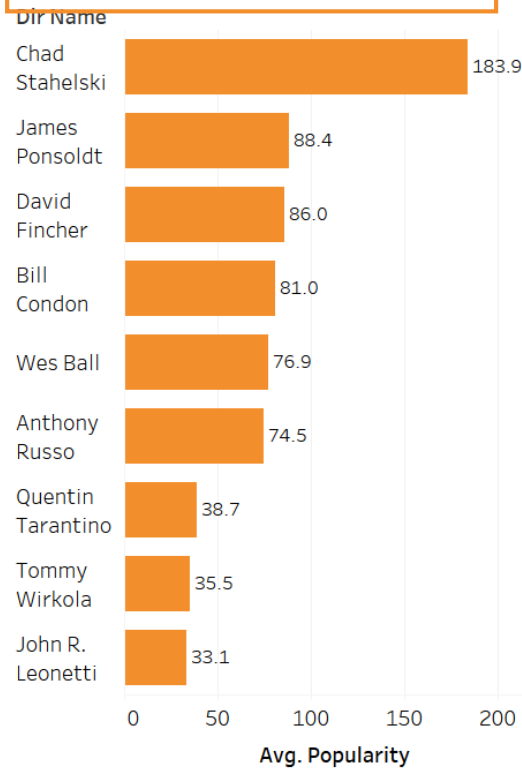
No Direct Correlation

Top 10 Director Name - Revenue

Dir Name

Director	Avg. Revenue
J.J. Abrams	2,068,223,624
Byron Howard	1,023,784,195
Pete Docter	857,611,174
Dan Scanlon	743,559,607
Garth Jennings	632,443,719
Anthony Russo	576,652,248
Richard Marqua..	572,700,000
Stephen Chow	553,810,228
Jennifer Yuh Nel..	521,170,825
Carlos Saldanha	508,065,839
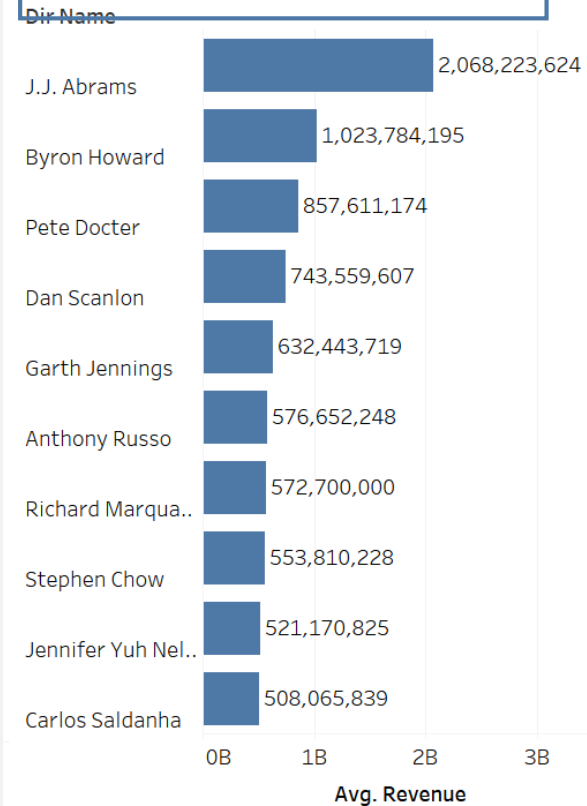
Research Model
- What is the research model and why the group has selected it?

Regression Analysis
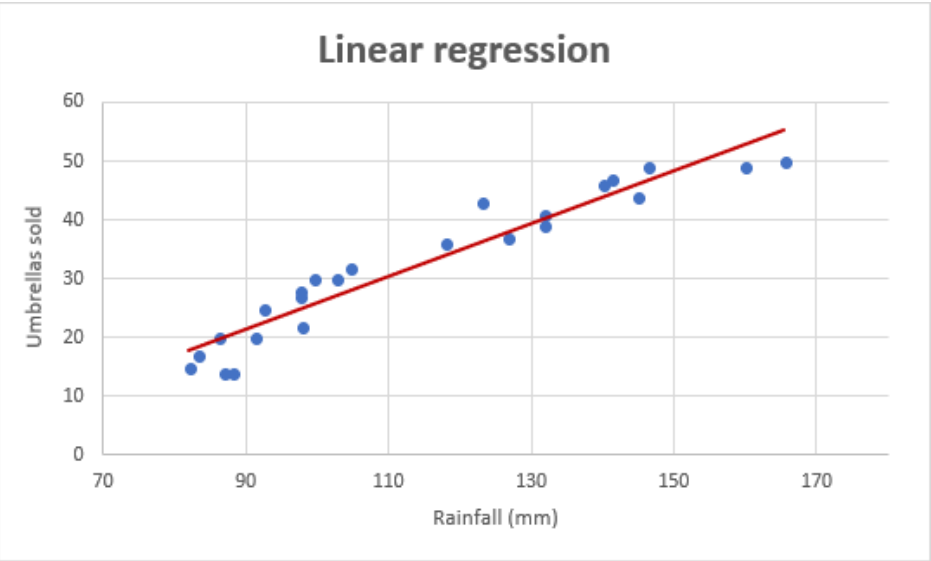
Source: Ablebits.com

● **Regression Analysis?**

" Regression analysis is a set of statistical processes for estimating the relationships between dependent variable(s) and one or more independent variable(s)."

Source: Wikipedia

● **Reason for selecting the model?**
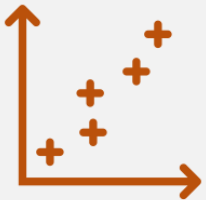
Find relationship (correlation) between variables by checking the R-squared, P-value and coefficient

First Result

Revenue=(2.99*Budget)+ (-1.12*Action) +(5.54*Family)

The revenue is influenced by the movie budget and which genre is attached to the movie.

- For instance every 1 million USD in budget, a movie is expected to return 2.99 million in revenue.
- If a movie is of the **Action** genre, the revenue is expected to decrease 1.12 million.
- If a movie is of the **Family** genre, the revenue is expected to increase by 5.54 million.
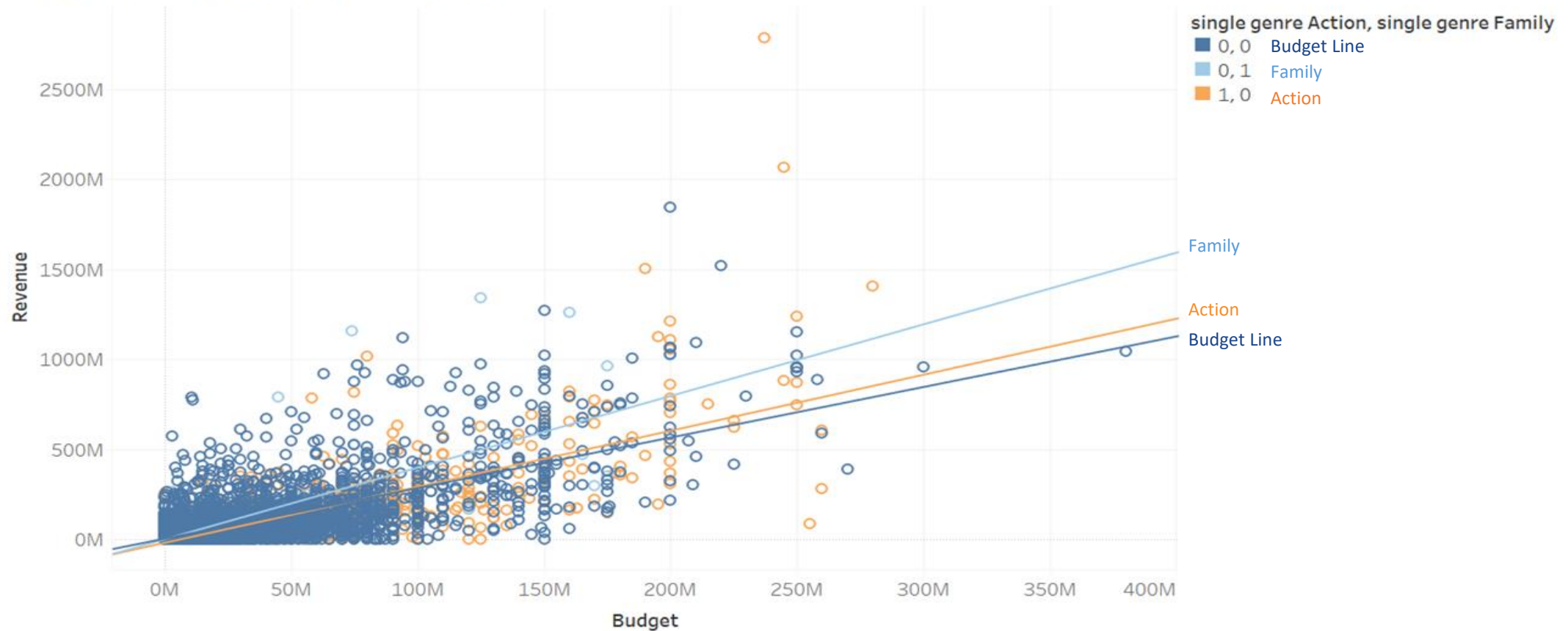- This set of independent variables explains 63.9% of variance in revenue.

Results

- Findings

First Result Visualization

Budget, Action, Family - Revenue

Budget vs. Revenue. Color shows details about single genre Action and single genre Family.

Second Result

OLS Regression Results

$Revenue=(2.91*Budget)+ (5.91*Family)+(2.39*SciFi)+(1.89*Adventure)$

The revenue is influenced by the movie budget and which genre is attached to the movie.

- For instance every 1 million USD in budget, a movie is expected to return 2.91 million in revenue.
- If a movie is of the **Family** genre, the revenue is expected to increase 5.91 million.
- If a movie is of the **Science Fiction** genre, the revenue is expected to increase by 2.39 million.
- If a movie is of the **Adventure** genre, the revenue is expected to increase by 1.89 million.
- 64% of variance in revenue can be explained by these independent variables.

Second Result Visualization

Budget, Family, SF, Adventure - Revenue

single genre Family, single genre Science Fiction, single genre Adventure
- 0, 0, 0 Budget Line
- 0, 0, 1 Adventure
- 0, 1, 0 SF
- 1, 0, 0 Family
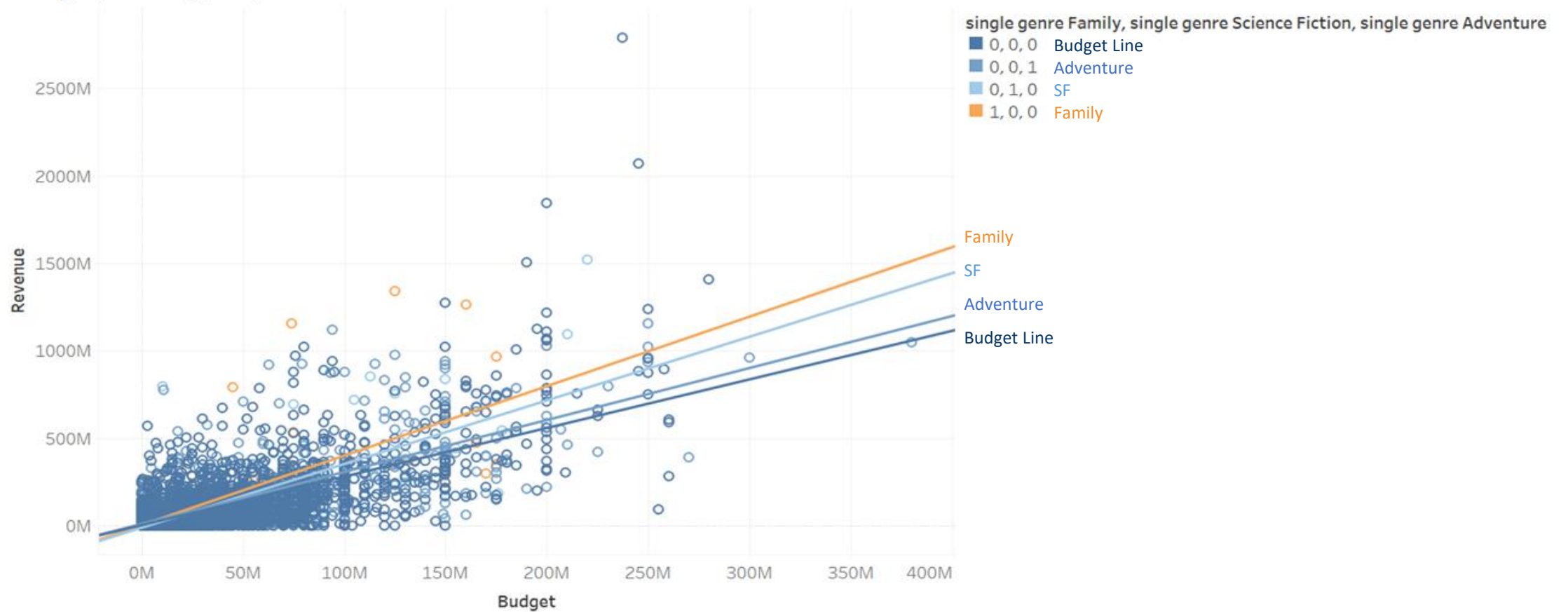
Budget vs. Revenue. Color shows details about single genre Family, single genre Science Fiction and single genre Adventure.

Third Result

OLS Regression Results

Dep. Variable:	popularity	R-squared (uncentered):	0.328
Model:	OLS	Adj. R-squared (uncentered):	0.327
Method:	Least Squares	F-statistic:	752.0
Date:	Sat, 13 Mar 2021	Prob (F-statistic):	0.00
Time:	03:07:07	Log-Likelihood:	-24461.
No. Observations:	6178	AIC:	4.893e+04
Df Residuals:	6174	BIC:	4.896e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
single_genre_Action	4.5298	0.430	10.545	0.000	3.688	5.372
budget	1.579e-07	3.77e-09	41.847	0.000	1.51e-07	1.65e-07
single_genre_Romance	4.5540	1.078	4.223	0.000	2.440	6.668
single_genre_Comedy	4.7983	0.350	13.717	0.000	4.113	5.484

Omnibus:	13319.611	Durbin-Watson:	1.668
Prob(Omnibus):	0.000	Jarque-Bera (JB):	112245656.880
Skew:	18.850	Prob(JB):	0.00
Kurtosis:	662.261	Cond. No.	3.11e+08

Popularity=(1.58*Budget)+(4.53*Action)+(4.55*Romance)+(4.8*Comedy)

The revenue is influenced by the movie budget and which genre is attached to the movie.

- For instance every 1 million USD in budget, a movie is expected to increase in popularity by 1.58.
- If a movie is of the **Action** genre, the popularity is expected to increase by 4.53.
- If a movie is of the **Romance** genre, the popularity is expected to increase by 4.55.
- If a movie is of the **Comedy** genre, the popularity is expected to increase by 4.8.
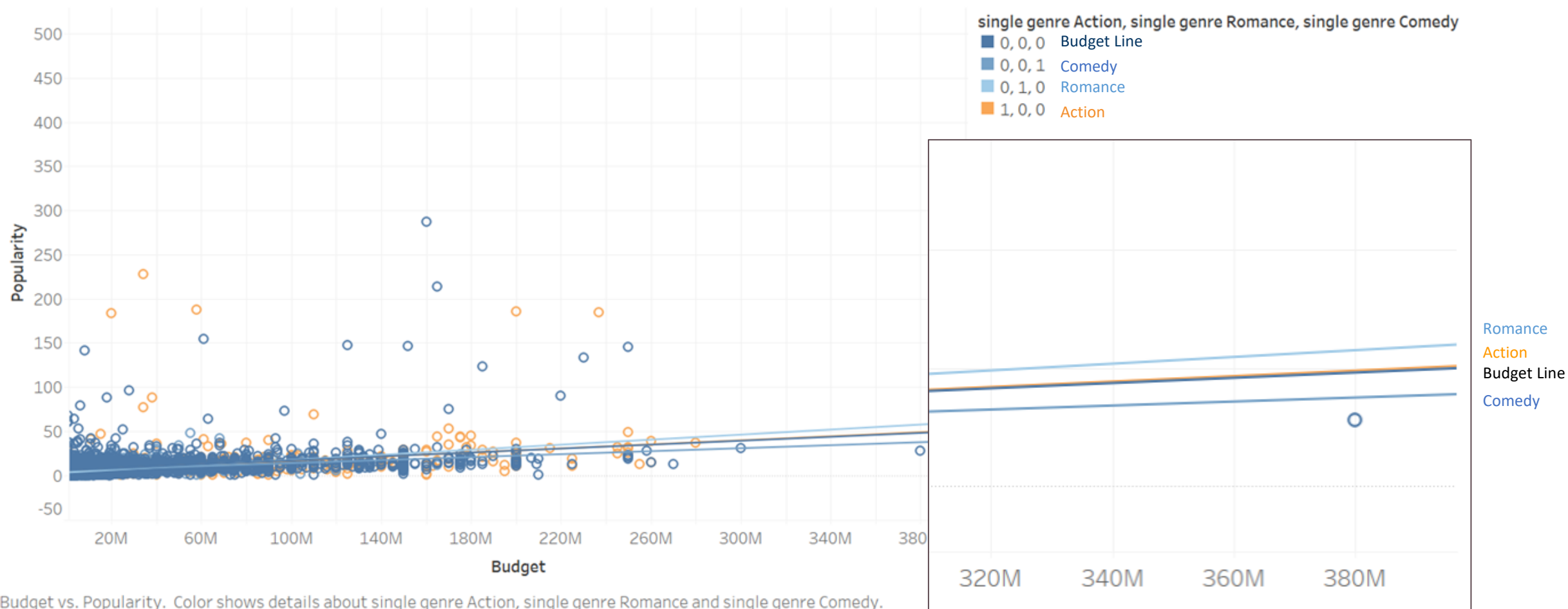- Only 32.8% of variance in Popularity can be explained by these independent variables.

Third Result Visualization

Budget, Action, Romance, Comedy - Popularity

Budget vs. Popularity. Color shows details about single genre Action, single genre Romance and single genre Comedy.

Results

- Findings

Fourth Result


```
                    OLS Regression Results
===============================================================================
Dep. Variable:              revenue   R-squared (uncentered):              0.642
Model:                          OLS   Adj. R-squared (uncentered):         0.642
Method:               Least Squares   F-statistic:                         1848.
Date:              Sat, 13 Mar 2021   Prob (F-statistic):                   0.00
Time:                      03:44:55   Log-Likelihood:                -1.2278e+05
No. Observations:              6178   AIC:                             2.456e+05
Df Residuals:                  6172   BIC:                             2.456e+05
Df Model:                         6
Covariance Type:          nonrobust
===============================================================================
                            coef    std err          t      P>|t|     [0.025      0.975]
-------------------------------------------------------------------------------
holiday                  1.143e+07   4.36e+06      2.621      0.009    2.88e+06      2e+07
budget                      2.8494      0.032     89.439      0.000       2.787      2.912
single_genre_Adventure   2.093e+07   5.36e+06      3.908      0.000    1.04e+07    3.14e+07
single_genre_Animation   5.406e+07   8.82e+06      6.130      0.000    3.68e+07    7.13e+07
single_genre_Family      6.144e+07   1.25e+07      4.935      0.000     3.7e+07    8.59e+07
single_genre_Science Fiction 2.586e+07 1.02e+07    2.547      0.011    5.95e+06    4.58e+07
===============================================================================
```

Revenue = (1.14*Holiday) + (2.85*Budget) + (2.09*Adventure) + (5.41*Animation) + (6.14*Family) + (2.59*SciFi)

The revenue is influenced by the movie budget, the genre, and whether the movie is released in the holiday season.
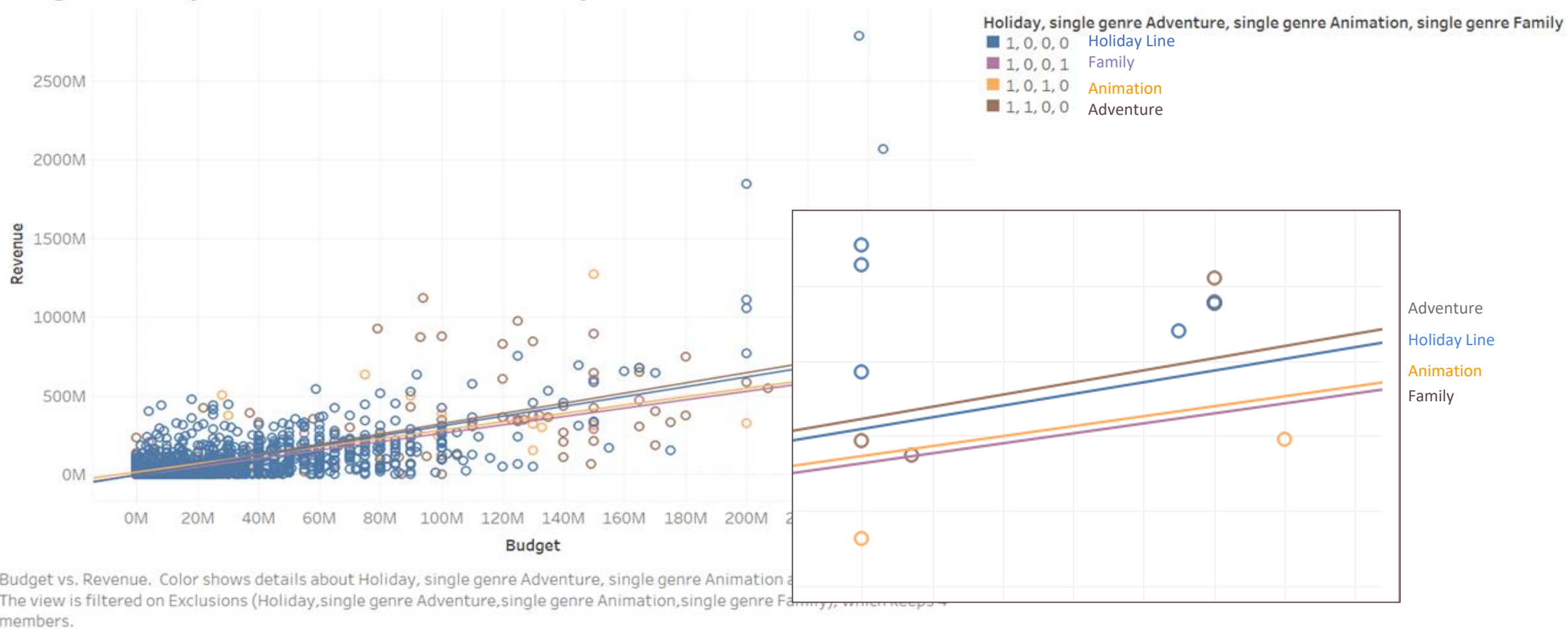
- For instance if a movie is released in the holiday season, the revenue is expected to increase by 1.14 million.
- An additional 1 million USD allotted to the budget is expected to increase the revenue by 2.85 million.
- If a movie is of the genre **Adventure**, the revenue is expected to increase 2.09 million.
- If a movie is of the genre **Animation**, the revenue is expected to increase 5.41 million.
- If a movie is of the genre **Family**, the revenue is expected to increase 6.14 million.
- If a movie is of the genre **Science Fiction**, the revenue is expected to increase 2.57 million.
- This set of independent variables explains 64.2% of variance in our dependent variable. (Highest r^2)

Results

- Findings

4

■ **Fourth Result Visualization**

Budget, Holiday, Advanture, Animation, Family - Revenue

Budget vs. Revenue. Color shows details about Holiday, single genre Adventure, single genre Animation a...
The view is filtered on Exclusions (Holiday,single genre Adventure,single genre Animation,single genre Family), which keeps 4 members.

5 Discussion

What do the results mean in practical business term?

Directors

Movie Production Companies

- Give an answer on which genre is better to produce or invest on
- Give an answer on which genre can bring more popularity
- Give an answer on which genre should be released in December

5   Discussion

How do the results answer your research question?

"Which factor leads to successful movie?"

Depending on the **'Genre'**
the movie can generate
more revenue and popularity

5 Discussion

How do you use these results to improve the business function of a company?

Directors

1) Increase budget by attracting more investors
2) Distribute the film by understanding audience' preference

Movies Production Companies

1) Develop analytic strategy for effective investment
2) Increase revenue by eliminating less profitable films

Reference

Slide # Introduction

- Moore, Schuyler. "Most Films Lose Money!" *Forbes*, Forbes Magazine, 3 Jan. 2019, www.forbes.com/sites/schuylermoore/2019/01/03/most-films-lose-money/?sh=76393074739f.

- Tankovska, H. "Most Popular Social Networks Worldwide as of January 2021, Ranked by Number of Active Users." *Statista*, 9 Feb. 2021, www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Thank You

Group 6