For Monash Data Analytics Bootcamp:

Khai Tran

Module 1 Challenge: Crowdfunding book:

Crowdfunding platforms like Kickstarter and Indiegogo have been growing in success and popularity since the late 2000s. From independent content creators to famous celebrities, more and more people are using crowdfunding to launch new products and generate buzz, but not every project has found success.

To receive funding, the project must meet or exceed an initial goal, so many organisations dedicate considerable resources looking through old projects in an attempt to discover "the trick" to finding success. For this week's Challenge, you will organise and analyse a database of 1,000 sample projects to uncover any hidden trends.

 Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

This analysis is focused on finding trends that a company considering crowdfunding should pay attention to, in order to maximise potential success and adjust predictions.
According to the data set, and the Exploratory data analysis conducted, three conclusions can be made:

Count of outcome	Column Labels				
Row Labels	canceled	failed	live	successful	Grand Total
film & video	11	60	5	102	178
food	4	20		22	46
games	1	23	3	21	48
journalism				4	4
music	10	66		99	175
photography	4	11	1	26	42
publishing	2	24	1	40	67
technology	2	28	2	64	96
theater	23	132	2	187	344
Grand Total	**57**	**364**	**14**	**565**	**1000**

Count of outcome	Column Labels				
Row Labels	canceled	failed	live	successful	Grand Total
film & video	6.18%	33.71%	2.81%	57.30%	100.00%
food	8.70%	43.48%	0.00%	47.83%	100.00%
games	2.08%	47.92%	6.25%	43.75%	100.00%
journalism	0.00%	0.00%	0.00%	100.00%	100.00%
music	5.71%	37.71%	0.00%	56.57%	100.00%
photography	9.52%	26.19%	2.38%	61.90%	100.00%
publishing	2.99%	35.82%	1.49%	59.70%	100.00%
technology	2.08%	29.17%	2.08%	66.67%	100.00%
theater	6.69%	38.37%	0.58%	54.36%	100.00%
Grand Total	**5.70%**	**36.40%**	**1.40%**	**56.50%**	**100.00%**

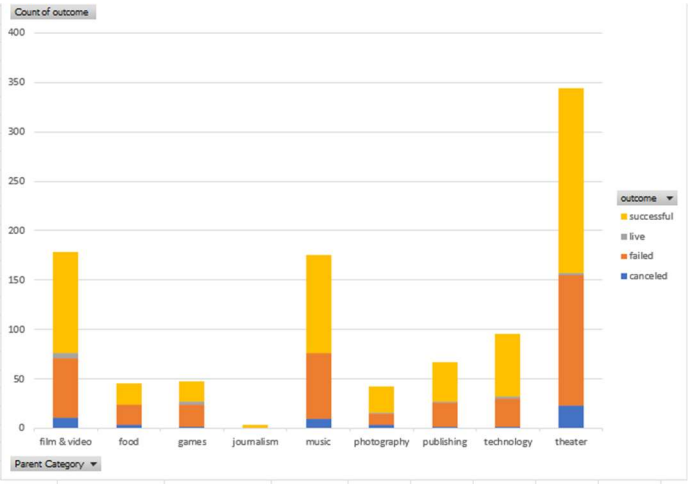

Figure 1 Pivot Chart and Table displaying outcomes of campaigns based on category.

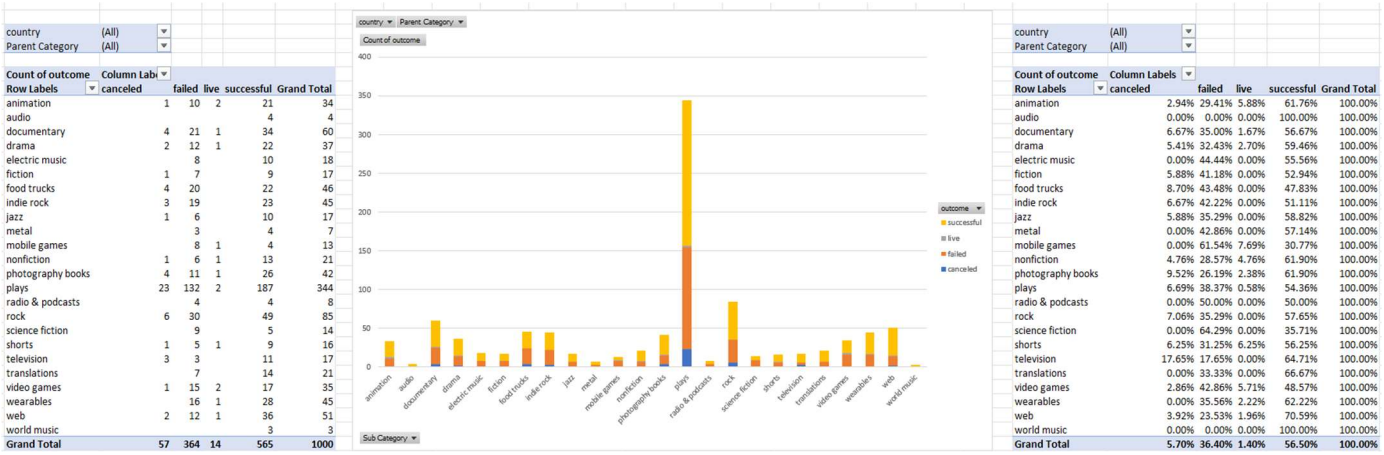

Figure 2 Pivot chart and table displaying outcomes of campaigns based on sub-category.

Parent Category (All)

Count of outcome	Column Labels				
Row Labels	canceled	failed	live	successful	Grand Total
Jan	6	36	1	49	92
Feb	7	28		44	79
Mar	4	33		49	86
Apr	1	30	1	46	78
May	3	35	2	46	86
Jun	3	28	1	55	87
Jul	4	31	1	58	94
Aug	8	35	1	41	85
Sep	5	23		45	73
Oct	6	26	1	45	78
Nov	3	27	3	45	78
Dec	7	32	3	42	84
Grand Total	**57**	**364**	**14**	**565**	**1000**

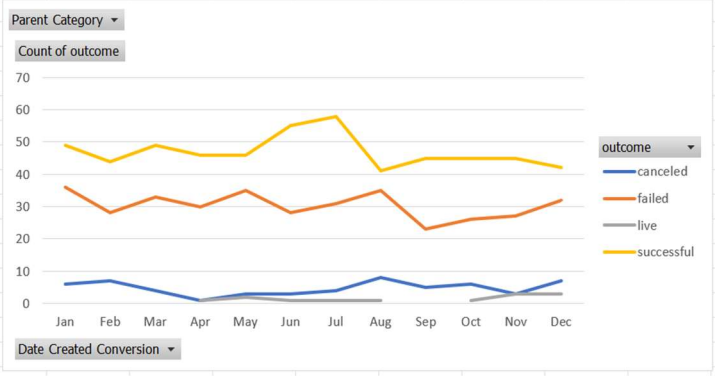

Figure 3 Pivot chart and table displaying outcomes of campaigns based on time of creation.

1. The Category/subcategory of the project seems to significantly influence the success rate of the project. For instance, projects in the technology/web category seem to have great success, (70%, figure 2) with a total of 51 projects in this field, where on the other hand, publishing a science fiction book has a success rate of 35% (figure 2). This suggests a relationship between category and success rate, and more research should be done on the significance of other factors, such as team size, prior funding, project presentation, etc.
2. Competition does not seem to have much of an impact on success rate. This can be seen in the theatre/plays subcategory and the radio/podcasts subcategory. Comparing the two, there are 344 plays within this data set, and only 8 podcasts. Of this, the plays still had a 54% success rate, whereas podcasts maintained a 50% success rate. This suggests that the limit of oversaturation in any category is higher than what we can observe, and if the data exists, it is worthwhile to pursue an analysis into where oversaturation begins to be noticeable.
3. Lastly there appears to be a higher success rate for projects which begin in june or july, as july had the highest number of successful outcomes as show in figure 3. This suggests that investors are more willing to invest their money around this time and this may have interactions with the end of the financial year, however further investigation will have to be conducted before any further theories can be tested.

 What are some limitations of this dataset?

Sample bias.
We have selected a sample containing a lot of plays, and we have no way of understanding if this is representative of the entire data set available from kickstarter / indiegogo. Kickstarter alone since its inception has had over 240,000 successful campaigns compared to this data set's 565 successful campaigns (figure 1).
External factors.
Factors such as economic conditions, market trends, government policy changes and technological advances have not been factored in and have the potential to greatly influence the success or failure of a project.

 What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Success rate vs team size.

It's logical that a larger team is more likely able to deliver a project, however we should consider the significance of the team size, as this would provide insight as to why science fiction books are so difficult to fund, as the teams are likely to be very small, consisting of the author, editor, and publishing company. The data may potentially explain why play's have a stable success rate despite its relatively high competition. Size of the team opens up the ability to pay attention to marketing, provide more updates to backers, and therefore improve backer confidence in funding the project. There is a logical limit to the size of the team, as each team member needs their own incentive, thus we expect a parabolic relationship between the variables, however the conclusion is up to future research.

Statistical Analysis:

Outcome	Backers_count		Outcome	Backers_count			
	Successful	158		Failed	0	Successful	
	Successful	1425		Failed	24	mean	851.1469
	Successful	174		Failed	53	median	201
	Successful	227		Failed	18	min	16
	Successful	220		Failed	44	max	7295
	Successful	98		Failed	27	Variance	1606217
	Successful	100		Failed	55	St.Dev	1267.366
	Successful	1249		Failed	200	1st quartile	128
	Successful	1396		Failed	452	3rd quartile	1280
	Successful	890		Failed	674	IQR	1152
	Successful	142		Failed	558	IQR * 1.5	1728
	Successful	2673		Failed	15	Lower Bounc	0
	Successful	163		Failed	2307	Upper Bounc	3008
	Successful	2220		Failed	88	Failed	
	Successful	1606		Failed	48	mean	585.6154
	Successful	129		Failed	1	median	114.5
	Successful	226		Failed	1467	min	0
	Successful	5419		Failed	75	max	6080
	Successful	165		Failed	120	variance	924113.5
	Successful	1965		Failed	2253	St.Dev	961.3082
	Successful	16		Failed	5	1st quartile	38
	Successful	107		Failed	38	3rd quartile	784.5
	Successful	134		Failed	12	IQR	746.5
	Successful	198		Failed	1684	IQR * 1.5	1119.75
	Successful	111		Failed	56	Lower Bounc	0
	Successful	222		Failed	838	Upper Bounc	1904.25
	Successful	6212		Failed	1000		
	Successful	98		Failed	1482		
	Successful	92		Failed	106		

One potential measure of success that was explored related to the number of backers a project has. This is because there is a theoretical positive relationship between number of backers and success of a project because each backer contributes a non-zero amount towards the goal, and therefore with enough backers the goal will always be met.

Observing the data, the median far better summarises the data for both the successful and failed outcome datasets.

This is because outliers are included in the calculation of the mean, and in both cases this skews the answer upwards, as can be seen in the successful data set with a max value of 7295 which is well above the upper bound of values for outliers being 3008 backers.

You can see in the case of the successful dataset, that there is a standard deviation of 1267, which is relatively large with a mean of 851

This is a greater variation than that of the failed dataset, with a standard deviation of 961 and a mean of 585

This outcome is expected, as there is no upper bound for the number of backers that can support a successful project, however there is a theoretical upper bound for failed projects, as too many backers will allow the project to reach its goal and become successful.