XXXX XXXX.XXX PREDICTIVE ANALYTICS USING SAS

ERNESTO NAVA

PROJECT 1 REPORT

SUBMISSION DATE: 7/18/2020

Ernesto Nava

Table of Contents

SAS Project 1 – Executive Summary

Summary of Main Insights from Analysis:

For the purpose of my analysis of the [redacted] platform and data set, I took two points of view. The first point of view, I analyzed the data through the eyes of [redacted]. For the second point of view, I analyzed the data through the eyes of project creator. Through the [redacted] point of view, the key points of interest and questions that I wanted to analyze pertained to the overall total number of successful, failed, and canceled projects and the amount of revenue generated from main categories. This led to the following main insights from the [redacted] point of view:

- A majority of projects failed to reach their real goal (63%), meaning that [redacted] did not receive any revenue (from their 5% fee) from over 230,000 projects and that only 36% of projects generated revenue from the 5% fee.
- In order for [redacted] to maximize and capitalize on the projects that do generate revenue, they should focus on certain main categories in order to either reduce the number of failed projects or improve upon the number of successful projects so that they can increase their revenue earned from the 5% fee. The categories that generated the most revenue are as follows: Games, Design, and Technology.

Through the project creator point of view, the key points of interest and questions that I wanted to analyze pertained to the relationship between USD_goal_real, USD_pledged_real, Backers and Days (duration between project launch date and goal deadline date) and how project creators should manage and establish setting their project goal and for how long (duration between project launch and goal deadline dates) they should raise funds in order to successfully reach their goal. This led to the following main insights from the [redacted] point of view:

- The USD_goal_real, Backers and Days variables are significantly correlated to USD_pledged_real and thus are important in determining the success of reaching a project's goal through pledges so project creators should take these variables into account when planning their project strategy.
- Project creators should establish a lower and more realistic real goal in order to give themselves the best chance to successfully reach their real goal through real pledges.
- Project creators should establish a shorter deadline (or shorter duration between project launch date and goal deadline date to raise funds) in order to give themselves the best chance to successfully reach their real goal through real pledges.

Main insights will be discussed in more detail in the SAS Project 1 – Insights section. In the next section, we will discuss the [redacted] platform, and points of interest and key questions that emerged from researching the [redacted] platform and business model.

SAS Project 1 – Introduction

[redacted] Platform and Points of Interest:

[redacted] is a global internet crowdfunding platform that focuses on creativity. [redacted]'s mission is to "help bring creative projects to life." [redacted] uses crowdfunding to gather money from the public (backers) to fund creative projects. [redacted] allows for project creators to avoid using or relying on traditional investment strategies. Project creators must fit their projects within one of [redacted]'s main categories and choose a deadline and minimum funding goal. Projects can be funded by backers and created by project creators from anywhere in the world. If the project's minimum funding goal is not achieved, the funds provided by backers are not collected. For projects that are successful in achieving their funding goal, [redacted] applies a five percent fee on the total amount of the project's funding (real pledged). Backers (or people) are offered rewards and experiences in return for their pledge to successful projects.

After researching the [redacted] platform, the following points of interest were identified:

- The percentages of successful, failed, and canceled projects
- Five percent fee collected from successful projects by [redacted]
- The main categories that are most and least successful
- Correlation and causation between real pledged funds, real goal funds, number of backers, and the time in days from project launch date to goal deadline
- Comparison between real pledged and real goal funds for failed and successful projects on an annual basis
- The duration between launch and deadline dates and its effect on the project's state

The points of interest listed above led to the following key questions (from the points of view of [redacted] and as a project creator):

- [redacted] key questions:
    - What percent of projects are canceled, fail, or succeed?
    - Which category provides the most revenue from the five percent fee?
    - Are there any main categories that should be dropped or focused on based on success rates?
- Project creator key questions:
    - Does the number of backers, the number of days to raise money for a project and the goal that is set important in determining if a project receives enough real pledges?
    - Do projects that succeed have higher or lower fundraising goals?
    - Do successful projects have longer or shorter durations (in days) in between launch and the goal deadline?

The points of interest and key questions identified above will be used to formulate the key insights from my analysis in the proceeding section.

SAS Project 1 – Insights

Key Insights from Analysis:

*For this analysis and all insights, only projects with canceled, failed, and successful states were analyzed.

*For this analysis and all insights, only USD_pledged_real and USD_goal_real were implemented and analyzed (Pledged and Goal were not used in this analysis).

A. Insight: Over half of all [redacted] projects fail

There is a total of 370448 [redacted] projects that either failed, succeeded, or were canceled. The percentages of each state in relation to the total amount of projects are displayed below:

		Percent of Total Project Observations:
Total Successful Project Observations:	133953	36.15973092
Total Failed Project Observations:	197718	53.37267309
Total Canceled Project Observations:	38777	10.46759599
Total Successful, Failed and Canceled Project Observations:	370448	

From the table we see that 53% percent of all projects failed (1997718 projects) and that only 36% of all projects succeeded (133953 projects).

This insight is telling in two ways; first, for project creators, it should be made clear that the chances of their projects either failing or being canceled (63%) are very high so they should manage expectations accordingly. Second, for [redacted], only 36% of all projects will generate revenue from their 5% fee (meaning that 63% of projects did not generate revenue) so [redacted] should strategize on how to maximize revenue accordingly.

B. Insight: [redacted] should focus on certain main categories in order to maximize revenue

As discussed in the first insight, only 36% of all projects succeeded in reaching their goals. This means that only 133953 projects reached their funding goal and generated revenue from the 5% fee to [redacted]. In order to gain insight on how [redacted] should strategize to maximize revenue from their fee, the revenue collected from fees on successful projects were broken down by main category:

	main_category	Total_Pledged_Real	state	Fee_on_Funds_Raised
1	Games	678835556.79	successful	33941777.839
2	Design	663157429.75	successful	33157871.487
3	Technology	598913794.11	successful	29945689.705
4	Film & Video	329484738.4	successful	16474236.92
5	Music	177558133.36	successful	8877906.668
6	Publishing	116074001.93	successful	5803700.0965
7	Fashion	113462705.1	successful	5673135.255
8	Food	105571041.29	successful	5278552.0645
9	Art	80246509.72	successful	4012325.486
10	Comics	66515027.09	successful	3325751.3545
11	Theater	39028463.46	successful	1951423.173
12	Photography	33419136.46	successful	1670956.823
13	Dance	12143609.69	successful	607180.4845
14	Crafts	11906808.92	successful	595340.446
15	Journalism	10469089.72	successful	523454.486

From the table, we see that the top five categories that raised the most money and thus generated the most revenue from the 5% fee are: Games, Design, Technology, Film & Video and Music. Conversely, the lowest performing categories (under $1 million in revenue generated) are: Journalism, Crafts and Dance.

In addition to the breakdown of pledged funds and fees displayed above, to gain further insight on main categories, the percentages of canceled, failed, or successful projects were calculated for each main category in relation to the total number of projects:

main_category	Projects	state	Total Projects	Percent Per State
Art	2222	canceled	370448	0.599814279
Art	14131	failed		3.814570466
Art	11510	successful		3.107048763
Comics	842	canceled	370448	0.22729236
Comics	4036	failed		1.089491643
Comics	5842	successful		1.577009459
Crafts	843	canceled	370448	0.227562303
Crafts	5703	failed		1.539487323
Crafts	2115	successful		0.570930333
Dance	163	canceled	370448	0.044000777
Dance	1235	failed		0.333380124
Dance	2338	successful		0.631127716
Design	4152	canceled	370448	1.120805079
Design	14814	failed		3.998941822
Design	10550	successful		2.84790308
Fashion	2650	canceled	370448	0.715350063
Fashion	14181	failed		3.828067637
Fashion	5593	successful		1.509793547
Film & Video	5755	canceled	370448	1.553524381
Film & Video	32904	failed		8.882218287
Film & Video	23622	successful		6.376603464
Food	2211	canceled	370448	0.596844901
Food	15969	failed		4.310726472
Food	6085	successful		1.64260571
Games	6202	canceled	370448	1.67418909
Games	16003	failed		4.319904548
Games	12518	successful		3.37915173
Journalism	523	canceled	370448	0.141180409
Journalism	3137	failed		0.846812508
Journalism	1012	successful		0.273182741
Music	3303	canceled	370448	0.891623116
Music	21752	failed		5.871809269
Music	24195	successful		6.531281043
Photography	986	canceled	370448	0.266164212
Photography	6384	failed		1.723318792
Photography	3305	successful		0.892163003
Publishing	3602	canceled	370448	0.972336198
Publishing	23145	failed		6.247840453
Publishing	12300	successful		3.320304064
Technology	4715	canceled	370448	1.272783225
Technology	20616	failed		5.565153544
Technology	6434	successful		1.736815963
Theater	608	canceled	370448	0.164125599
Theater	3708	failed		1.000950201
Theater	6534	successful		1.763810305

From the above table, the following categories were determined to be the least (failed) and most successful main categories:

Least Succesful Main Category:	Film & Video	Publishing	Music
Percent:	8.882218287	6.247840453	5.871809269
Most Succesful Main Category:	Music	Film & Video	Games
Percent:	6.531281043	6.376603464	3.37915173

From the above table, Music, Film & Video, and Games are the categories that produced the most successful projects. Conversely, Film & Video, Publishing, and Music are the categories hat produced the most unsuccessful projects.

From the above insights, in order for [redacted] to maximize revenue from its fee, [redacted] should focus on its top performing main categories in terms of revenue produced (Games, Design, Technology, Film & Video and Music). [redacted] should then use the percentages calculated in order to determine the rate in which these categories are succeeding or failing and then based on these percentages, [redacted] should implement strategies to either lower the failure for projects or raise success rates for project. For example, Games is the top producing category in terms of funds raised and it is also third in terms of generating the most successful projects (%). Film & Video is the fourth main category in terms of funds raised, it is second in terms of generating the most successful projects (%) but it is first in terms of producing the most failed projects so [redacted] should find a way in reducing Film & Video's failure rate in order to raise its success rate, funds raised and revenue from the 5% fee.

C. Insight: Goal, Backers, and Days are significantly correlated to pledged funds

*For the purpose of using a smaller data set, only the following years were considered: 2014, 2015, 2016, 2017 and 2018.

In order to gain insight on how project creators should strategize in order to reach their funding goal through real pledges, this analysis focused on the relationship between USD_pledged_real and the variables that it deemed important: USD_goal_real, Backers and Days. Days is a variable that was created to capture the number of days (or duration) between a projects launch date and goal deadline date. To analyze and determine a relationship between the variables listed above, a correlation hypothesis test was conducted.

In order to determine if there is a significant correlation or relationship between USD_pledged_real and USD_goal_real, Backers, and Days, the following null and alternative hypotheses were formulated:

- Null hypothesis: There does not exist a significant correlation between USD_pledged_real and USD_goal_real, Backers, and Days.
- Alternative hypothesis: There exists a significant correlation between USD_pledged_real and USD_goal_real, Backers, and Days.

To test these hypotheses, we implemented a Spearman correlation test. From the test, it can be concluded that there is a significant statistical correlation between USD_pledged_real and USD_goal_real, Backers and Days. There is significant correlation between USD_pledged_real and USD_goal_real because their correlation of <0.0001 is less than the significant level of 0.05 (i.e. we reject the null hypothesis for this correlation). There is significant correlation between USD_pledged_real and Backers because their correlation of <0.0001 is less than the significant level of 0.05 (i.e. we reject the null hypothesis for this correlation). There is significant correlation between USD_pledged_real and Days because their correlation of <0.0001 is less than the significant level of 0.05 (i.e. we reject the null hypothesis for this correlation).

Based on these results, it can be concluded that USD_goal_real, Backers and Days have a significant relationship to USD_pledged_real and that these variables are important in causing the success or failure of real pledges reaching real goals.

4 Variables:	usd_goal_real usd_pledged_real backers days

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
usd_goal_real	249777	57198	1363346	6211	0.49000	166361391
usd_pledged_real	249777	9953	101759	417.77000	0	20338986
backers	249777	110.28226	956.08199	9.00000	0	219382
days	249777	33.18890	11.43899	30.00000	1.00000	90.00000

Spearman Correlation Coefficients, N = 249777
Prob > |r| under H0: Rho=0

	usd_goal_real	usd_pledged_real	backers	days
usd_goal_real	1.00000	0.16262 <.0001	0.09273 <.0001	0.23447 <.0001
usd_pledged_real	0.16262 <.0001	1.00000	0.96114 <.0001	0.01509 <.0001
backers	0.09273 <.0001	0.96114 <.0001	1.00000	-0.00182 0.3638
days	0.23447 <.0001	0.01509 <.0001	-0.00182 0.3638	1.00000

D. Insight: Project creators should set a lower fund-raising goal for their projects

From the previous insight, it was determined that USD_goal_real is significant in determining how project creators should strategize in order to reach their funding goal through real pledges. But where should project creators set their goal in order to realistically achieve the real pledges needed? In order to gain insight on where project creators should set their funding goal in order to achieve a successful project through real pledges, this analysis focused on the difference between the USD_goal_real for projects that failed or succeeded.

To begin the analysis for this insight, a statistical table was created in order to compare the statistical differences between USD_pledged_real and USD_goal_real for all of the main categories per each state (canceled, failed, successful):

	canceled															
	usd_pledged_real								usd_goal_real							
	N	Mean	Median	StdDev	Min	Max	P25	P75	N	Mean	Median	StdDev	Min	Max	P25	P75
main_category																
Art	2222	487.43	21.38	1913.44	0.00	35035.01	0.00	220.00	2222	39877.08	4000.00	489534.76	1.00	20000000.00	1184.72	11000.00
Comics	842	865.00	130.00	2394.60	0.00	34506.00	7.00	726.00	842	13617.06	5000.00	43017.66	11.20	732621.22	2200.00	10200.00
Crafts	843	414.04	10.00	3108.53	0.00	84262.00	0.00	131.00	843	14767.77	3000.00	58652.99	1.57	1200000.00	1000.00	10000.00
Dance	163	524.82	75.00	1930.28	0.00	22917.00	0.00	395.88	163	33004.19	5000.00	177006.43	10.00	2000000.00	2500.00	10000.00
Design	4152	4645.68	816.84	21296.79	0.00	1059078.19	103.35	3357.48	4152	66404.65	16000.00	1564889.35	1.00	100000000.00	6500.00	40000.00
Fashion	2650	1336.74	61.28	4179.81	0.00	81568.00	0.00	681.00	2650	27334.96	8507.50	191243.70	0.77	5642003.59	3000.00	21612.90
Film & Video	5755	1665.93	52.00	9090.86	0.00	259789.14	0.00	610.00	5755	144057.21	11250.00	1996212.58	0.76	100000000.00	3971.72	40000.00
Food	2211	880.88	25.00	3515.88	0.00	55537.01	0.00	340.00	2211	53583.17	14750.00	204230.15	1.00	5000000.00	4350.83	40000.00
Games	6202	3890.73	390.00	14602.15	0.00	504120.08	27.00	2487.00	6202	73671.08	12779.18	847597.02	0.76	50000000.00	5000.00	35000.00
Journalism	523	274.98	0.00	1226.92	0.00	21363.00	0.00	86.64	523	29276.22	7387.16	126667.62	1.00	2000000.00	2276.19	18000.00
Music	3303	620.62	15.52	5482.10	0.00	235619.33	0.00	224.98	3303	33857.20	5000.00	561634.23	1.00	21474836.47	1971.87	14105.26
Photography	986	564.32	1.65	3602.75	0.00	66783.00	0.00	144.00	986	23751.80	5000.00	320013.60	5.00	10000000.00	2000.00	11810.00
Publishing	3602	536.32	16.03	2619.30	0.00	74071.00	0.00	205.00	3602	72253.38	5500.00	1868598.07	0.01	100000000.00	2500.00	15000.00
Technology	4715	5178.52	155.00	24939.10	0.00	1076751.05	0.00	2126.00	4715	125547.46	29841.84	1504760.46	0.74	88767573.21	10000.00	72060.49
Theater	608	1515.81	45.60	14176.09	0.00	301019.57	0.00	378.00	608	94771.98	6000.00	1236470.69	5.00	30000000.00	2147.50	15000.00

state																
failed																
usd_pledged_real									usd_goal_real							
N	Mean	Median	StdDev	Min	Max	P25	P75	N	Mean	Median	StdDev	Min	Max	P25	P75
14131	625.03	67.00	2890.99	0.00	158105.30	1.12	370.00	14131	67958.35	3879.22	1840814.14	1.00	100000000.00	1196.10	10000.00
4036	1012.18	247.00	2627.41	0.00	57454.00	45.00	997.00	4036	41607.38	5000.00	1589896.46	15.00	100000000.00	2196.70	10000.00
5703	358.94	37.00	2930.94	0.00	198737.51	1.00	180.00	5703	12528.80	3000.00	142854.83	1.00	10000000.00	1000.00	8000.00
1235	619.04	105.00	1705.25	0.00	27389.00	6.61	520.00	1235	15395.51	4500.00	62288.80	20.00	1000000.00	2000.00	10000.00
14814	2840.54	490.36	11186.09	0.00	757352.94	62.21	2247.00	14814	54494.39	13000.00	911777.74	1.00	80000000.00	5000.00	31180.40
14181	861.44	70.00	3173.51	0.00	148786.35	1.00	485.00	14181	26938.61	6000.00	863067.54	6.98	100000000.00	2017.00	15000.00
32904	1518.80	105.00	6390.75	0.00	260372.61	1.57	775.00	32904	122726.43	10000.00	2231025.52	0.15	151395869.90	3000.00	25000.00
15969	1077.75	81.00	3468.63	0.00	114420.00	2.00	641.00	15969	62089.22	12000.00	1588458.32	1.00	166361390.70	5000.00	30000.00
16003	2264.10	222.00	11003.46	0.00	519999.59	25.13	1342.00	16003	57788.65	10000.00	1164140.08	1.00	100000000.00	3500.00	25000.00
3137	550.93	20.00	4369.48	0.00	194635.65	0.00	156.00	3137	90083.45	5000.00	1961532.75	10.00	100000000.00	1873.01	15000.00
21752	628.31	60.56	2386.32	0.00	122052.00	1.00	425.00	21752	24070.04	5000.00	494248.63	0.86	50000000.00	2000.00	10000.00
6384	672.12	58.00	3196.72	0.00	137350.50	1.00	358.88	6384	13483.48	4500.00	129311.60	1.00	7300000.00	1800.00	10000.00
23145	634.69	62.00	2262.58	0.00	91634.20	1.00	404.00	23145	23733.01	5000.00	727436.24	1.00	100000000.00	2500.00	11340.22
20616	2408.65	98.02	12076.55	0.00	721036.46	1.64	931.76	20616	120481.38	20000.00	1822021.42	0.83	110169771.60	6750.19	50000.00
3708	979.95	150.00	3170.74	0.00	53910.00	10.00	700.00	3708	54736.91	5000.00	774239.67	1.00	40000000.00	2100.00	15000.00

successful															
usd_pledged_real								usd_goal_real							
N	Mean	Median	StdDev	Min	Max	P25	P75	N	Mean	Median	StdDev	Min	Max	P25	P75
11510	6971.89	2630.85	33621.79	1.00	1924018.00	1020.00	5857.84	11510	4410.09	2000.00	10273.95	0.01	600000.00	650.00	5000.00
5842	11385.66	4000.00	32370.82	2.00	1254120.08	1662.26	9552.00	5842	5397.35	2619.71	9173.44	0.72	250000.00	1000.00	6000.00
2115	5629.70	1825.00	14975.39	1.00	454717.53	686.00	5259.00	2115	3013.43	1000.00	5818.26	1.00	75783.09	350.00	3000.00
2338	5194.02	3364.00	6621.86	14.00	146075.97	1890.00	6010.27	2338	4601.09	3000.00	5859.41	5.00	125000.00	1500.00	5000.00
10550	62858.52	14619.30	359087.26	2.00	20338986.27	4829.00	41643.00	10550	15408.84	7500.00	28868.54	0.75	1000000.00	2500.00	18460.00
5593	20286.56	6864.09	57724.19	1.00	1851033.00	2106.00	18749.07	5593	9080.96	5000.00	12819.63	1.00	300000.00	1270.04	12000.00
23622	13948.22	5209.35	66623.41	1.00	5764229.38	2045.00	12455.88	23622	11141.43	5000.00	34114.79	0.77	2000000.00	1600.00	10000.00
6085	17349.39	8951.82	60771.05	1.00	1927217.00	3215.00	19000.33	6085	11633.96	7343.91	15980.74	0.88	350000.00	2500.00	15000.00
12518	54228.76	10571.54	279199.05	4.37	12393139.69	3685.00	30446.99	12518	14857.22	5000.00	58072.80	0.75	2015608.88	1500.00	12000.00
1012	10344.95	3678.00	19341.66	1.00	215487.75	1290.00	10409.50	1012	8148.65	3000.00	15546.85	1.00	177794.64	1000.00	8468.98
24195	7338.63	3950.60	17970.37	1.11	1363037.00	1855.00	8210.00	24195	5737.38	3216.05	8635.24	0.74	250000.00	1500.00	7000.00
3305	10111.69	3700.00	33475.46	1.00	793266.77	1364.34	10215.00	3305	6490.60	3000.00	12194.05	1.00	400000.00	1000.00	8000.00
12300	9436.91	4117.04	25375.65	1.00	866193.50	1631.68	9074.00	12300	5897.85	3000.00	9811.12	0.55	250000.00	1090.24	6913.61
6434	93085.76	22478.00	266480.11	0.79	6225354.98	5790.00	76697.00	6434	26286.35	10000.00	46010.97	0.75	1000000.00	3000.00	30287.89
6534	5973.13	3075.50	12739.64	3.00	358548.01	1528.00	5661.00	6534	5198.04	2650.00	10540.73	1.00	333841.51	1200.00	5000.00

The main take away from the statistical table is that on average:

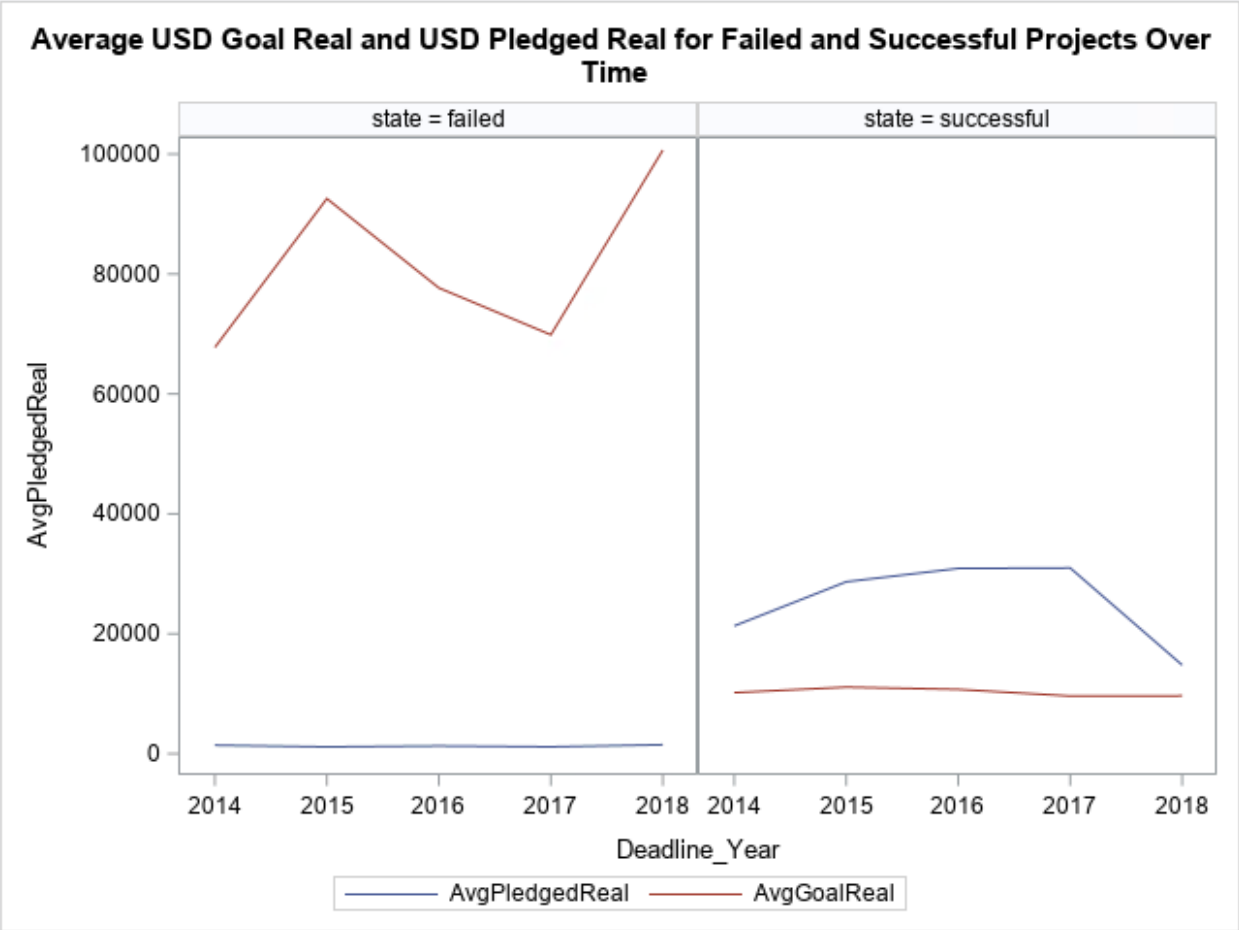
- For all of the main categories that were canceled or failed, the USD_pledged_real mean was significantly less than the USD_goal_real mean.
- For all of the main categories that were successful, the USD_pledged_real mean was higher than the USD_goal_real mean.

From the initial statistical analysis above, it clear that, on average, projects that fail set a very high or unrealistic goal and projects that succeed set a lower or more realistic goal.

**For the purpose of using a smaller data set, only the following years were considered: 2014, 2015, 2016, 2017 and 2018 for the below analysis.

Next, we will dig deeper and focus only the USD_pledged_real and USD_goal_real for the projects that either failed or succeeded.

In order to easily compare the differences in USD_pledged_real and USD_goal_real for failed and successful projects, a side-by-side panel series graph was created. The graph allows us to easily the difference between the variables for failed and successful projects between the years of 2014 and 2018:



From the graph, it is clear that the average goal for failed projects is set extremely high (on average, over $60,000). While the average goal for successful projects is set much lower (on average, below $20,000). For failed projects, the average pledged amount comes no where near close reaching the extremely high goals that were set. While for successful projects, the average pledged amount is relatively close to and above the lower set goals.

It is clear from the statistical table and panel series plots that project creators should set their goals to a lower, much more realistic level in order to achieve the real pledge needed for a successful project and that in general, failed projects have much higher goals than successful projects. To further back this insight, a hypothesis test and one-sided test were implemented. First, the following null and alternative hypotheses were formulated:

- Null hypothesis: The average USD_goal_real of successful projects is NOT significantly higher than that of failed projects.
- Alternative hypothesis: The average USD_goal_real of successful projects is significantly higher than that of failed projects.

Next, to test our hypotheses, a one-sided test was implemented. The results of the one-sided t-test are displayed below. Since the P-value of the F-test is less than the significance level of 0.05, the variances of the two samples are not equal and that the Satterthwaite method should be

used. Using the Satterthwaite method, we see that the P-value of the T-test is greater than 0.05 so we do not reject the null hypothesis.

The TTEST Procedure

Variable: AvgGoalReal

state	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
failed		5	81709.4	14387.3	6434.2	67770.0	100661
successful		5	10256.7	648.0	289.8	9624.9	11095.3
Diff (1-2)	Pooled		71452.8	10183.7	6440.7		
Diff (1-2)	Satterthwaite		71452.8		6440.7		

state	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
failed		81709.4	63845.2	99573.6	14387.3	8619.9	41342.7
successful		10256.7	9452.1	11061.2	648.0	388.2	1862.0
Diff (1-2)	Pooled	71452.8	-Infty	83429.6	10183.7	6878.6	19509.6
Diff (1-2)	Satterthwaite	71452.8	-Infty	85167.5			

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	8	11.09	1.0000
Satterthwaite	Unequal	4.0162	11.09	0.9998

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	4	4	493.00	<.0001

Thus, as guidance to project creators, they should set lower and realistic goals, at or below $20,000, in order to give themselves the best chance to successfully reach their real goal through real pledges.

E. Insight: Project creators should set a shorter deadline

*For the purpose of using a smaller data set, only the following years were considered: 2014, 2015, 2016, 2017 and 2018 for the below analysis.

In addition to USD_goal_real being a significant variable in determining how project creators should strategize in order to reach their funding goal through real pledges, Days (the time between launch date and goal deadline date) was also identified as a significant variable. This leads us to ask the question, "how long should project creators set their deadline for?" In order to gain insight on how long project creators should set their deadline for funding (i.e. how long should they raise money for) in order to achieve a successful project through real pledges, this analysis focused on the difference between the average duration for raising funds for each project state (canceled, failed, successful).

To begin the analysis for this insight, a table was created in order to compare the total number of days in between the project launch date and project goal deadline date and the average duration for each project state analyzed:

	state	Number_of_Projects	Total_Days_Btwn_Launch_Deadline	Average_Duration
1	canceled	29540	1032863	34.964895058
2	failed	139704	4770605	34.14794852
3	successful	80533	2486356	30.873753617

From the table, we can see that failed projects had the highest number of total days in between launch date and goal deadline date, while canceled had the lowest. However, the variable that we are interested in is the average duration between launch and deadline for each project state. We can see that successful projects had the shortest average duration, while failed and canceled had higher average durations. Thus, as guidance to project creators, they should set shorter project deadlines (i.e. a shorter duration in between launch date and goal deadline date), at or below 31 days or one month, in order to give themselves the best chance to successfully reach their real goal through real pledges.

SAS Project 1 – Appendix

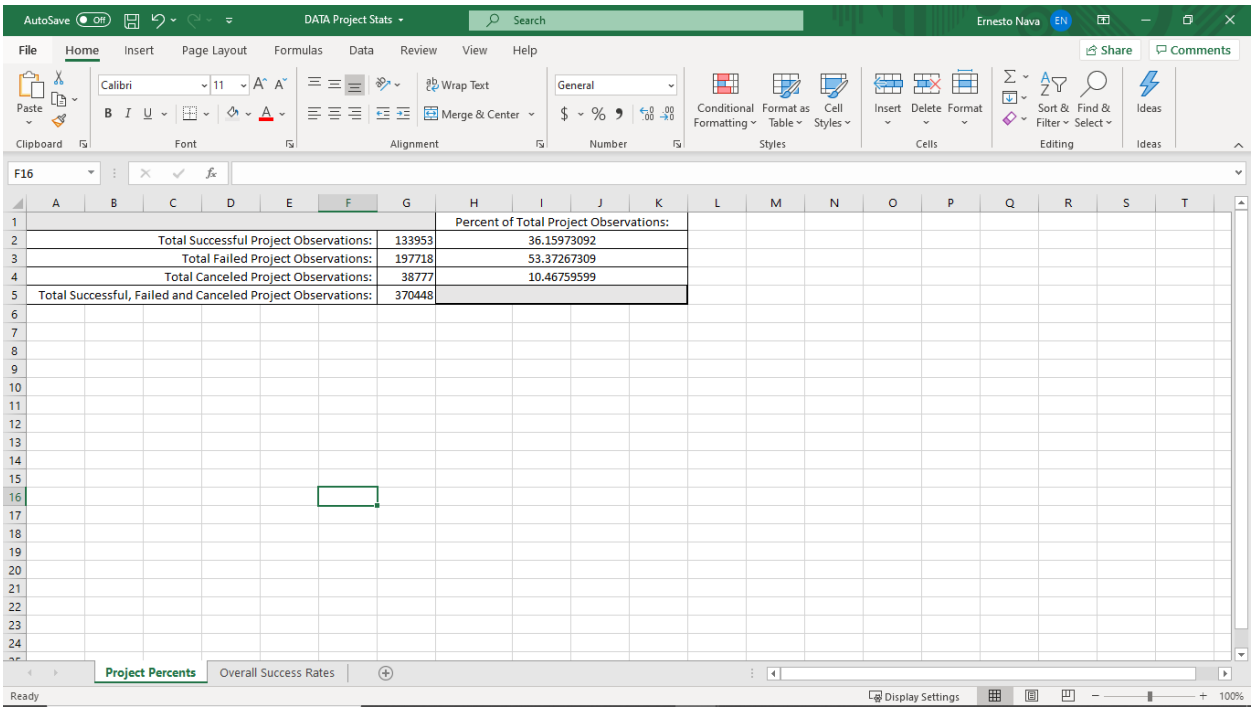
Analysis Code and Details:

A. Insight: Over half of all [redacted] projects fail

In order to determine the percentages of canceled, failed, and successful projects, the [redacted] data was loaded into SAS via PROC IMPORT and using PROC SQL, a table was created for canceled, failed, and successful projects. Next, using PROC SQL, a count table was created for each state (canceled, failed, successful) that displays the total number of observations for each state. Finally, the counts from each state count table were then used in Excel in order to calculate the percentages of canceled, failed, and successful projects and to create the percentage table:

```
/*Load data*/
PROC IMPORT OUT= WORK.kickstarter
            DATAFILE= 'C:\Users\exnl21330\HW 1 Data\DATA.csv'
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
proc print data=WORK.kickstarter;
 titlel 'Kickstarter Projects';
PROC CONTENTS;
RUN;

/*Analysis*/

/*General Statistical Analysis*/
proc sql;
create table scsl_fld_cncl_prjcts as
select *
from work.kickstarter
where state = 'successful' OR state = 'failed' OR state = 'canceled';


proc sql;
create table successful_projects_count as
select COUNT(*)
from work.kickstarter
where state= 'successful';


proc sql;
create table failed_projects_count as
select COUNT(*)
from work.kickstarter
where state= 'failed';
```

```
proc sql;
  create table canceled_projects_count as
  select COUNT(*)
  from work.kickstarter
  where state= 'canceled';
```

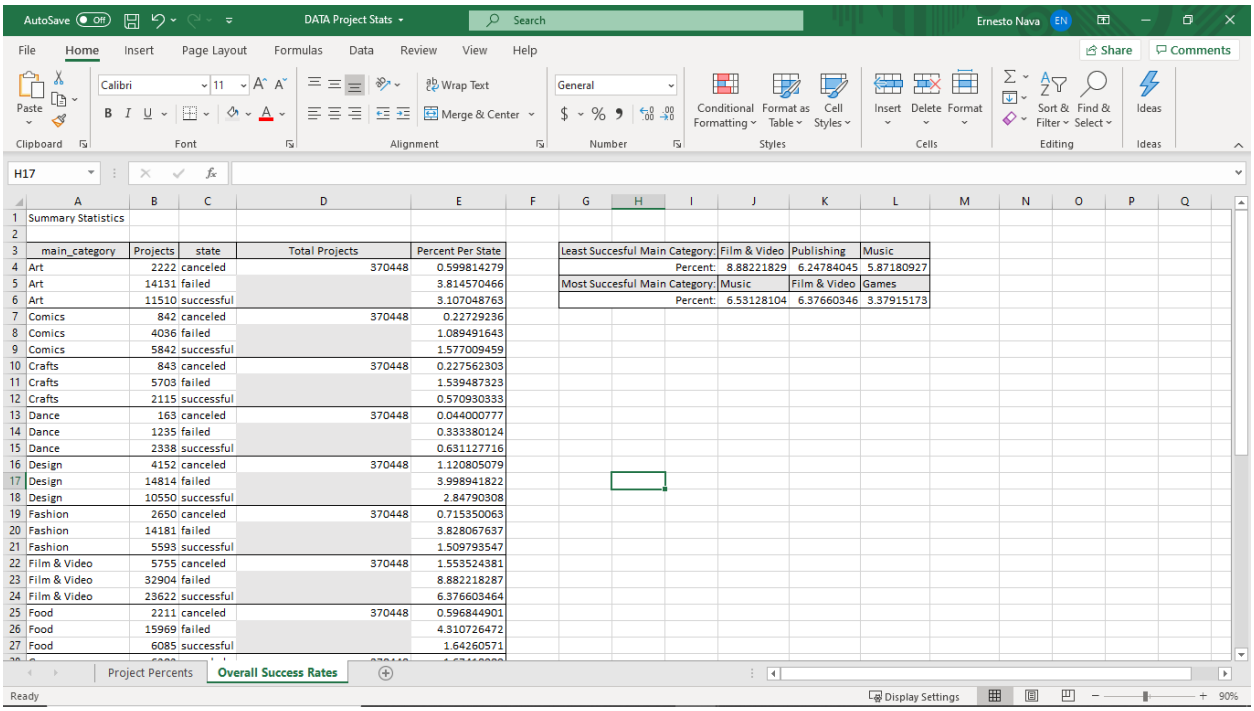


				Percent of Total Project Observations:
	Total Successful Project Observations:		133953	36.15973092
	Total Failed Project Observations:		197718	53.37267309
	Total Canceled Project Observations:		38777	10.46759599
	Total Successful, Failed and Canceled Project Observations:		370448	

B. Insight: [redacted] should focus on certain main categories in order to maximize revenue

In order to determine the main categories that [redacted] should focus on, using PROC SQL, a table that contains the counts or the number of projects per main category for each state was created (categories state). The categories state table was then exported into Excel in order to calculate the percentages per main category for each project state and to create the percentage table for the least and most successful main categories. Next, to determine the USD_pledged_real amounts per main category, PROC SQL was used to create a table (categories pledged) that contains the total amount of USD_pledged_real per main category for each state (canceled, failed, successful). Next, using PROC SQL, a table (pledged success) that contains the total amount of USD_pledged_real per main category for only the successful state was created. This table was then sorted (through edit mode) to create a new sorted table (pledged success sort). Finally, a new column that calculates the total revenue raised from the 5% fee was added to the sorted table to create a new table (pledged success fee):

```
proc sql;
  create table categories_state as
  select Main_Category, COUNT(Main_Category) as Projects, State
  from scsl_fld_cncl_prjcts
  group by Main_Category, State;
```



	A	B	C	D	E
1	Summary Statistics				
2					
3	main_category	Projects	state	Total Projects	Percent Per State
4	Art	2222	canceled	370448	0.599814279
5	Art	14131	failed		3.814570466
6	Art	11510	successful		3.107048763
7	Comics	842	canceled	370448	0.22729236
8	Comics	4036	failed		1.089491643
9	Comics	5842	successful		1.577009459
10	Crafts	843	canceled	370448	0.227562303
11	Crafts	5703	failed		1.539487323
12	Crafts	2115	successful		0.570930333
13	Dance	163	canceled	370448	0.044000777
14	Dance	1235	failed		0.333380124
15	Dance	2338	successful		0.631127716
16	Design	4152	canceled	370448	1.120805079
17	Design	14814	failed		3.998941822
18	Design	10550	successful		2.84790308
19	Fashion	2650	canceled	370448	0.715350063
20	Fashion	14181	failed		3.828067637
21	Fashion	5593	successful		1.509793547
22	Film & Video	5755	canceled	370448	1.553524381
23	Film & Video	32904	failed		8.882218287
24	Film & Video	23622	successful		6.376603464
25	Food	2211	canceled	370448	0.596844901
26	Food	15969	failed		4.310726472
27	Food	6085	successful		1.64260571

Least Succesful Main Category:	Film & Video	Publishing	Music
Percent:	8.88221829	6.24784045	5.87180927
Most Succesful Main Category:	Music	Film & Video	Games
Percent:	6.53128104	6.37660346	3.37915173

```
proc sql;
  create table categories_pledged as
  select Main_Category, SUM(USD_pledged_real) as Total_Pledged_Real, State
  from scsl_fld_cncl_prjcts
  group by Main_Category, State;
```

```
proc sql;
  create table pledged_success as
  select Main_Category, Total_Pledged_Real, State
  from categories_pledged
  where state = 'successful';
```

```
data pledged_success_fee;
   set pledged_suc_sort;
   Fee_on_Funds_Raised = Total_Pledged_Real * .05;
  run;
```

C. Insight: Goal, Backers, and Days are significantly correlated to pledged funds

In order to determine the correlation between USD_goal_real, Backers, and Days to USD_pledged_real, I first had to convert the date format (via a DATA step) of the Launch variable. Using the datepart and format functions, I was able to convert the Launch variable to

the mmddyy format in order to successfully compare this date to the goal deadline date. Once converted, using a DATA step, I created a new data with a new Days variable. Using the INTCK and PUT functions, I was able to calculate the number of days in between the launch and deadline dates and I was able to store the number of days, for each observation, in the Days column. Once I had the Days variable calculated and created, using PROC CORR, I created a spearman correlation test to determine the correlation between the variables listed above:

```
/*Correlation and Causation*/
data new_launch;
    set usd_project_success;
    launch_date = datepart(launched);
    format launch_date mmddyy10.;
run;

data l_d_duration;
    set new_launch;
    date1=launch_date;
    date2=deadline;
    days=intck('day', date1, date2);
    put days=;
run;

proc corr data=l_d_duration spearman;
 var usd_goal_real usd_pledged_real backers days;
 title 'Study Spearman Correlation Test';
run;
```

D. Insight: Project creators should set a lower fund-raising goal for their projects

In order to analyze and compare the differences in USD_goal_real and USD_pledged_real between project states, I used PROC MEANS AND PROC TABULATE to calculate and create a statistics summary table. To graphically compare the differences in USD_goal_real and USD_pledged_real across different years for failed and successful project states, I first created a Deadline_Year variable via a DATA step and by using the YEAR function on the deadline date variable. Next, using PROC SQL, I created a new table for only failed and successful projects. Next, to create the panel series graph, I used:

- PROC SORT to sort the respective data set
- PROC MEANS to calculate and create the average variables of USD_pledged_real and USD_goal_real based on the year and project state
- PROC SGPANEL to create the panel series graph

Finally, using PROC TTEST, I conducted a one-sided t-test on the average USD_goal_real variable (AvgGoalReal):

```
proc means data= scsl_fld_cncl_prjcts n mean median stddev min max p25 p75 maxdec= 2;
 var USD_pledged_real USD_goal_real;
 title 'Summary Statistics';
run;

proc tabulate data= scsl_fld_cncl_prjcts;
 class Main_Category State;
 var USD_pledged_real USD_goal_real;
 table Main_Category,State*(USD_pledged_real USD_goal_real)*(N Mean Median StdDev Min Max p25 p75);
 title 'Summary Statistics';
run;

data prjct_yrs;
 set scsl_fld_cncl_prjcts;
 Deadline_Year = year(deadline);
 run;

 /*Failed and Successful Comparison*/
proc sort data= l_d_duration;
 by Deadline_Year state;
run;

proc sql;
 create table fail_success_pjcts as
 select *
 from l_d_duration
 where state = 'failed' OR state = 'successful';

proc means data= fail_success_pjcts mean maxdec= 2 noprint;
 by Deadline_Year state;
 var usd_pledged_real usd_goal_real;
 output out=means2
  mean= AvgPledgedReal AvgGoalReal;
run;

proc sgpanel  data= means2;
 panelby state;
 series x=Deadline_Year y=AvgPledgedReal;
 series x=Deadline_Year y=AvgGoalReal;
 title 'Average USD Goal Real and USD Pledged Real for Failed and Successful Projects Over Time';
run;

proc ttest data=means2 sides=L alpha=0.05;
 class state;
 var AvgGoalReal;
run;
```

E. Insight: Project creators should set a shorter deadline

In order to analyze and compare the differences in deadline durations between project states (canceled, failed, successful) I used PROC SQL to create a new table (prjcts duration). The prjcts duration table included newly created variables:

- Total_Days_Btwn_Launch_Deadline: calculates the total number of days, per state, between launch date and goal deadline date.
- Average_Duration: calculates the average total number of days, per state, between launch date and goal deadline date.

```
/*Duration averages*/
proc sql;
create table prjcts_duration as
select State, COUNT(ID) as Number_of_Projects, SUM(days) as Total_Days_Btwn_Launch_Deadline, AVG(days) as Average_Duration
from l_d_duration
group by State;
```

F. Year Data Set Used in Analysis

*For the purpose of using a smaller data set, only the following years were considered: 2014, 2015, 2016, 2017 and 2018 for the below analysis.

This accomplished via PROC SQL:

```
proc sql;
create table scsl_fld_cncl_prjcts_yrs as
select *
from prjct_yrs
where Deadline_Year = 2014 OR Deadline_Year = 2015 OR Deadline_Year = 2016 OR Deadline_Year = 2017 OR Deadline_Year = 2018;
```

G. Sources

The [redacted] website (https://www.[redacted].com/) was used as the source for researching the [redacted] platform and business model.

THIS IS THE END OF THE PROJECT