RICHMOND
School of Professional & Continuing Studies™

Unit 1 Homework: Kickstart My Chart

Analysis using Excel
23 Jul 2019

Marie Colton, Ph.D.

Hydros, LLC 3713 Merrington View, Midlothian, VA 23113

T: 734-585-6317 **E(work):** hydrosllc@gmail.com **E(class):** midlo.marie@gmail.com

Cover graphic from https://akshaykher.wordpress.com/

CONTENTS

Introduction 3

Data preparation tasks with Excel 4

Tasks 1 and 2 Representative Results 4

Tasks 3 and 4: Success statistics per category 5

Task 5: Success statistics over time 7

Bonus: Success statistics as a function of initial investment goal 7

Data Analysis 8

What (3) conclusions can be drawn about the Kickstarter campaign? 8

What are some of the limitations of this data set? 9

Future Work 10

What are other possible tables or graphs could be created to offer additional insight? 10

Summary 10

Introduction

Kickstarter is a public-benefit crowdfunding service founded in 2009 by three artist-entrepeneurs that specializes in attracting innovation investment in the creative arts and sciences. In just a decade, Kickstarter facilitated initial funding of over $1B in games alone, surpassed the National Endowments for Arts in support of aspiring artists, and provided the basis for wide-spread public engagement with the arts and associated technologies[1]. The premise of the business is that creatives may not be immediately recognized as having a product of commercial or cultural value, and consequently are unable to obtain development funding to capitalize their work from conventional sources. However, afficienados or "backers" of certain art or technologies may be willing to support innovations of interest to them outside of these conventional constraints.

The business model of Kickstarter is a service in which Kickstarter provides a global platform for innovators to market their work and solicit funding. In return, Kickstarter receives 5% of the proposed goal funding, once successfully raised for the project, and a 2-3% Amazon payment processing fee. After this point, exchanges are made between the innovator and backer such that the backer may get either a product, payback with interest, or partial interest in the business. No funds are returned by either Kickstarter or the innovator if the project fails. Kickstarter updates its Web site daily with cumulative statistics, Figure 1, which provides basic performance metrics to prospective investors.

This case study provides the basis for the Data Analytics Bootcamp Homework Assignment 1. Using data and analyses from an approximate 1% subset[2] of the approximately 450,000 Kickstart projects (to current day), the student is asked to draw three conclusions about the Kickstarter campaigns, describe what the limitations of the data set are, and suggest other possible tables and/or graphs that might be useful for additional insights. Results are compared to the overall Kickstarter information of Figure 1 to draw additional conclusions.

[1] https://news.artnet.com/art-world/kickstarter-founder-perry-chen-interview-part-1-1530109. A ten-year retrospective interview with Perry Chen, one of the Kickstarter founders.
[2] StarterBook_Excel1.xlsx, includes assigned modifications as completed by Colton.

Projects and Dollars

Category	Launched Projects	Total Dollars	Successful Dollars	Unsuccessful Dollars	Live Dollars	Live Projects	Success Rate
All	453,036	$4.43 B	$3.95 B	$441 M	$37 M	3,494	37.14%
Games	46,366	$1.07 B	$979.75 M	$75.68 M	$14.97 M	508	39.11%
Design	37,698	$982.87 M	$890.47 M	$81.60 M	$10.80 M	331	37.34%
Technology	39,239	$843.67 M	$738.79 M	$99.99 M	$4.89 M	358	20.44%
Film & Video	71,880	$450.70 M	$383.47 M	$66.16 M	$1.07 M	362	37.56%
Music	60,092	$236.75 M	$216.05 M	$19.67 M	$1.03 M	315	49.93%
Fashion	28,979	$174.41 M	$153.21 M	$20.58 M	$619.41 K	295	27.16%
Publishing	46,899	$165.51 M	$144.99 M	$19.51 M	$1.01 M	350	32.53%
Food	28,490	$151.87 M	$128.07 M	$23.28 M	$518.25 K	195	25.20%
Art	35,040	$116.14 M	$103.21 M	$12.07 M	$858.09 K	345	43.06%
Comics	14,278	$98.00 M	$91.30 M	$5.97 M	$720.88 K	171	57.82%
Theater	11,960	$45.62 M	$40.87 M	$4.63 M	$125.94 K	81	60.04%
Photography	11,942	$45.59 M	$39.81 M	$5.51 M	$269.77 K	46	31.79%
Crafts	10,577	$18.42 M	$15.15 M	$3.16 M	$105.28 K	89	24.70%
Journalism	5,451	$16.62 M	$14.43 M	$2.15 M	$40,762	32	22.29%
Dance	4,145	$14.45 M	$13.43 M	$987.33 K	$31,985	16	61.93%

https://www.kickstarter.com/help/stats

Figure 1. Cumulative statistics from 2009 to July 21, 2019 showing Kickstarter project investments and success results listed in rank order from largest to smallest number of submissions and funding levels. These statistics are used in comparative analysis to sub-set of data as provided for the analysis.

Data preparation tasks with Excel

Tasks 1 and 2 Representative Results

Table 1 lists the tasks completed to analyze the Kickstarter data set. Tasks 1 and 2 were completed using the entire raw data file of 4114 projects and are submitted as a separate Excel sheet as required. For completeness in this report, representative rows of the raw data file are provided in Table 1 to show conditional formatting rules are as required (see task description in Table 1).

Successful (blue), Failed (red), Canceled (gray), Live (green)		Per cent funded graded scale, rows 110-127	
Rows 110-112	successful successful successful	200+ shades of blue	247 220 131 155
Rows 422-444	failed failed failed	100 – 200 shades of green	104 141 103 140 114 100 113 105
Rows 1042 - 1044	canceled canceled canceled		
Rows 1683-1685	live live live	0-100 shades of red	0 0 0 0 0 14

Table 1. (Left column) Demonstration of conditional formatting of "state" condition and per cent success scores drawn from full spreadsheet. (Right column) Task assignments for preparation of data as captured in submitted Excel spreadsheet (StarterBook_Excel1.xlsx).

ASSIGNMENT TASKS

- **TASK 1**
  Color specific projects in the "state" column as successful, failed, canceled, or live in raw data file. Compute per cent funded towards goal for each project
- **TASK 2**
  Color per cent funded with 3-color R,G,B graded scale for projects with less than 100% (red), 100-200% (G) and greater than 200% (Blue). Compute average donation per backer. Separate category and subcategory for further analysis.
- **TASK 3**
  Create Pivot sheet and stacked column chart showing success statistics of projects per category, filterable by country
- **TASK 4**
  Create Pivot sheet and stacked column chart showing success statistics of projects per subcategory, filterable by country and parent category.
- **TASK 5**
  Convert Unix time stamps to Excel dates. Create Pivot table and line chart to show success statistics as a function of month, filterable by year.
- **BONUS**
  Create new sheet with goal funding ranges and count success statistics in each range. Create line chart to show success as a function of investment funds.

Tasks 3 and 4: Success statistics per category

To assess whether certain project categories are more successful than others and why, the raw data were pivoted on category and plotted with associated success statistics, Figure 2. Further examination of subcategories was accomplished in similar manner, Figure 3.

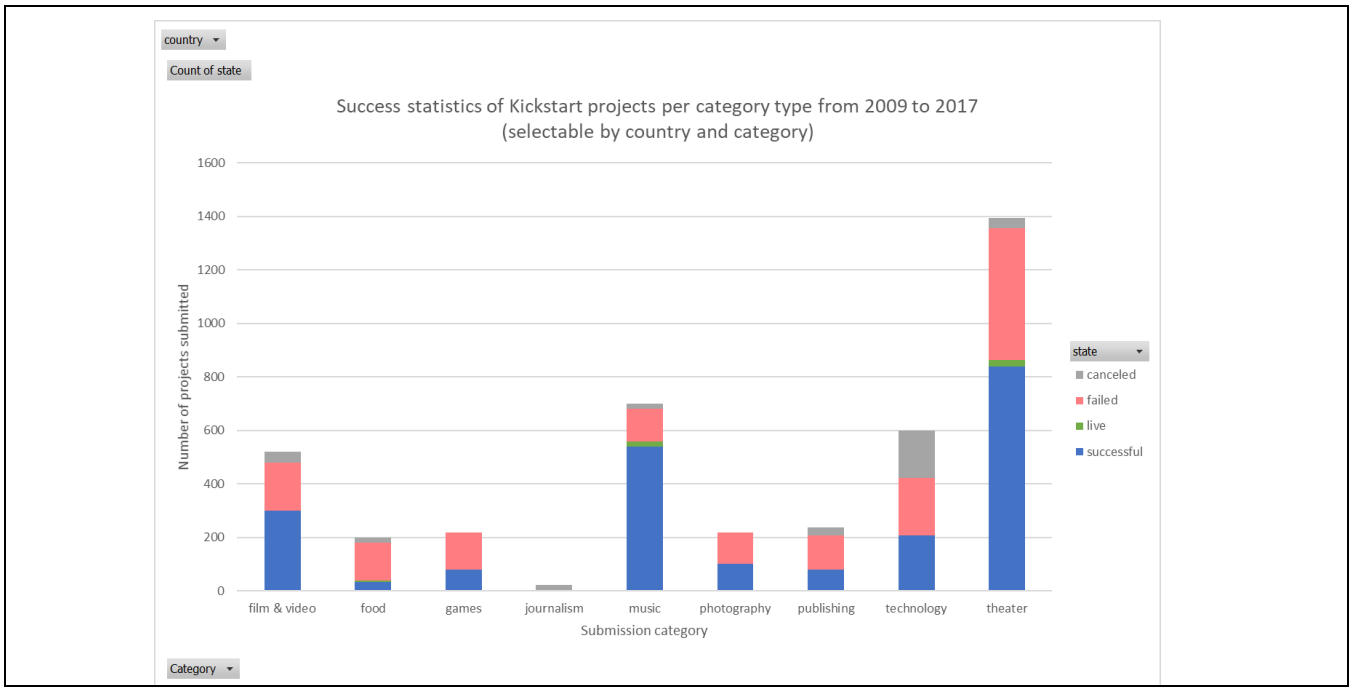


Figure 2 Kickstarter project success statistics in each major category showing wide range in number of projects submitted across the categories and corresponding success rates.

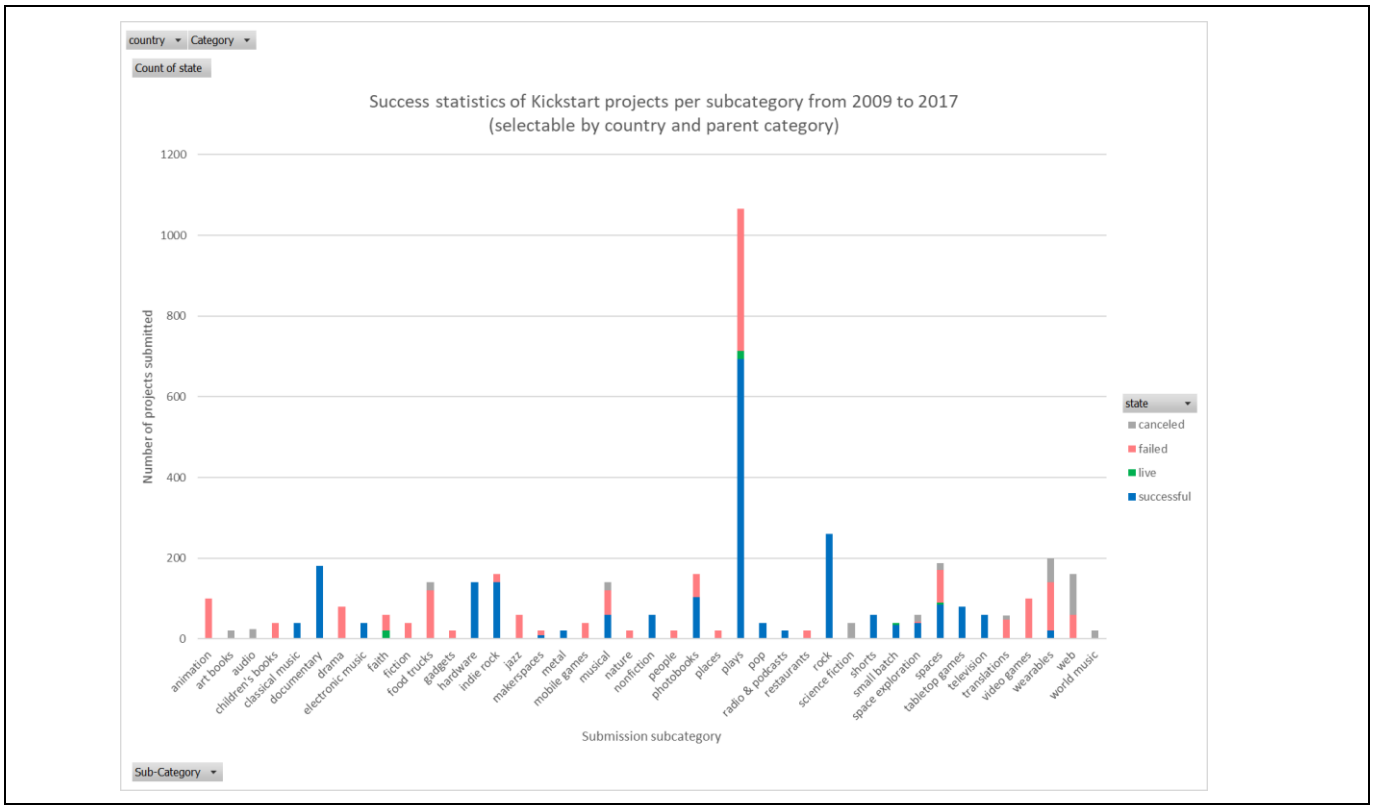
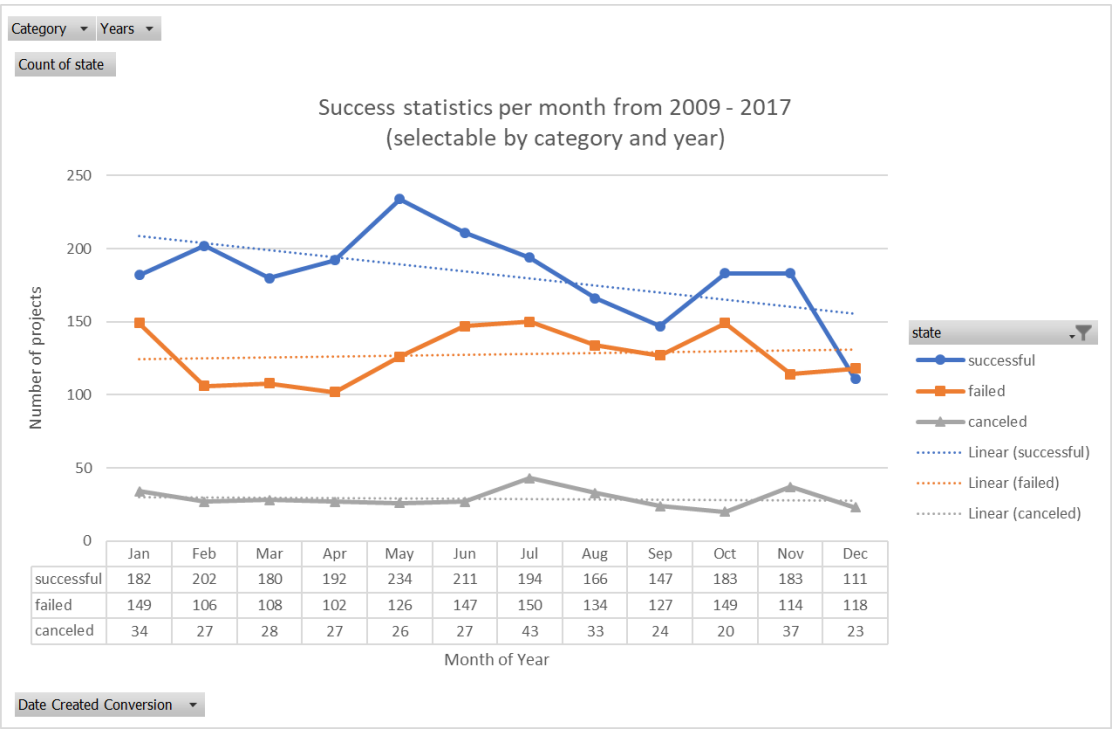


Figure 3. Kickstarter project success statistics separated further to show specific subcategories. In this subset of all projects, submissions of live theater plays were nearly 4x (1066:260) that of the next largest class (rock music).

Task 5: Success statistics over time

The start dates of the 4114 project entries were converted from system time stamps to calendar date, and success statistics computed for each month and year from 2009 to 2017. The summary statistics are plotted as a function of month over the 8-year time span, Figure 4. The "live", ongoing projects are not included in this plot since the final end state is not yet known.



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
successful	182	202	180	192	234	211	194	166	147	183	183	111
failed	149	106	108	102	126	147	150	134	127	149	114	118
canceled	34	27	28	27	26	27	43	33	24	20	37	23

Figure 4. Examination of project success as a function of start month. Linear regression lines indicate no correlation with month over the five years for failed and canceled projects. There is a small, negative correlation with month for successful projects, which could be due to lower overall submissions in December (R^2 value is low, .31).

Bonus: Success statistics as a function of initial investment goal

The Kickstarter initial investment goals range from as less than $1000 to over $50000 per project. Therefore, it is of interest to determine whether there is a dependency between initial funding level and success or failure. In Figure 5, the success figures were sorted into fund ranges and the percentages of each final state computed.

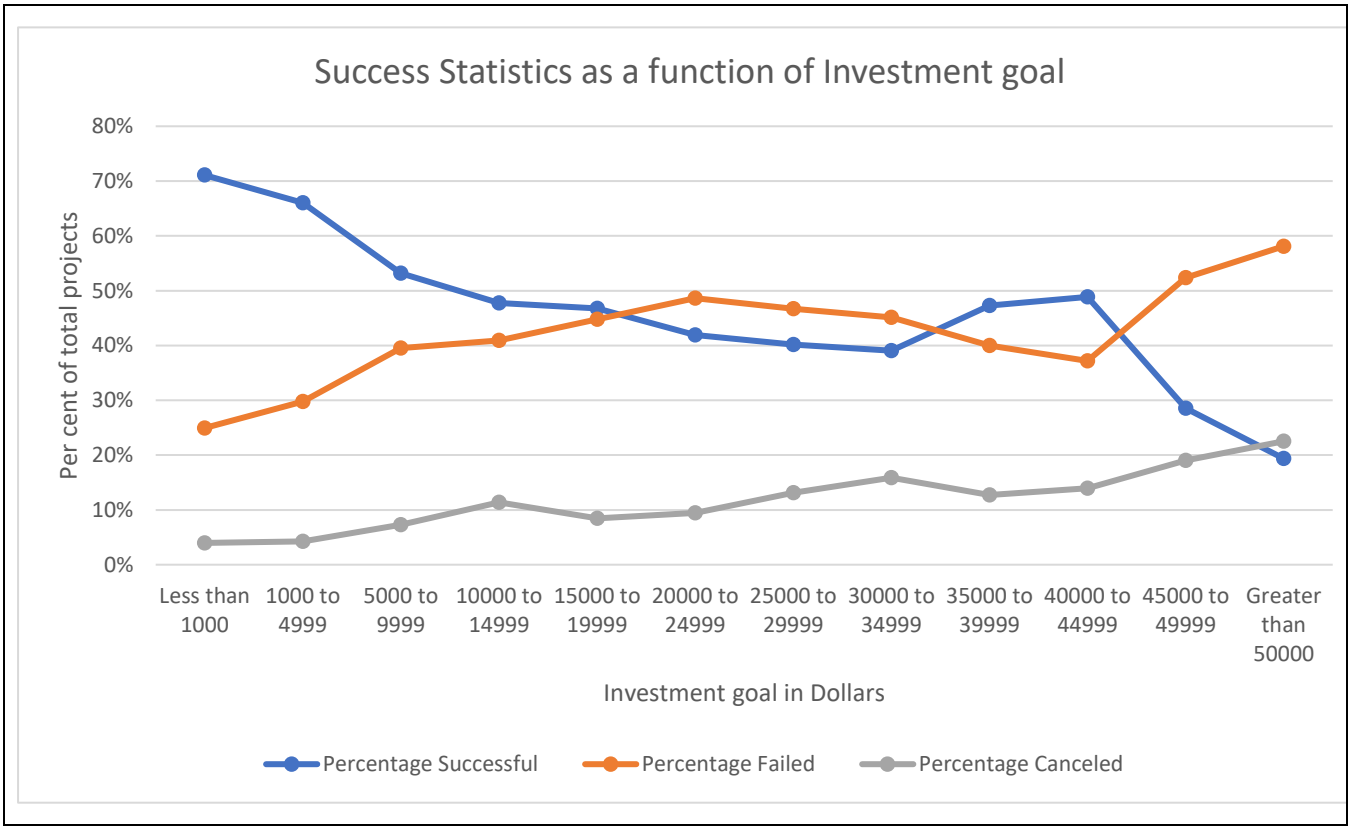


Figure 5. Summary statistics of success/failure for Kickstarter projects binned into initial funding ranges. Highest success percentage is associated with lowest value projects (<$5K), while lowest success percentages occur with highest value projects (>$45K). Mid-valued projects ($5-45K) appear to be nearly as likely to succeed as not. Cancellations slowly increase from about 4% to 23% over the full range of project value, with highest value projects (>$45K) having the greatest likelihood of cancellation.

Data Analysis

What (3) conclusions can be drawn about the Kickstarter campaign?

Kickstarter is a novel application of "open innovation"[3] applied to cultural arts and technology outside of a formal product development environment within a single corporation, sector, or Nation. Open innovation recognizes that boundaries between a firm, small business, or individual developer and their environments have become permeable in the Information Age and innovation can transfer across these boundaries to creative consumers. Within the context of our brief analysis and data subset, there are several reasonable conclusions:

1) (Market) Kickstarter has revealed a multi-billion-dollar market in "art and culture" that exists in the realm of the creators, outside of mainstream, conventional access. This observation runs

[3] New Frontiers in Open Innovation, Henry Chesbrough, Vanhverbeke W., and West, J. , Oxford University Press, 2014.

counter to the concept that art has limited commercial value and identifies the backlog of pent-up demand.

2) (Product) Innovative, consumer products at various stages of development can be promoted and delivered at a fraction of the cost, with few internal formalisms generally associated with product development.
3) (Data) The reduced data set provided allows for the following conclusions:
    a. Theater and music events have the highest number of submissions and relative success rates among all the categories (Figure 2)
    b. "Journalism" as a category suffers from too few submissions and subsequent cancellations (Figure 2). The subcategories of "journalism" indicate a rather narrow definition of journalism (e.g., podcasts) which might explain the low numbers. Given the current-day issues of "fake" news, there may be opportunities for "factual," objective news products here.
    c. Live theater, "play" projects outpaced all other subcategories at a ratio of almost 4:1. The roughly 50% success rate of these projects was also remarkable in a probabilistic sense (about the same as a toss of a coin) (Figure 3).
    d. Open Innovation is not the panacea for all product development. Figure 4 shows small, marketable products can succeed with minimal funds because they require little development, but more complex ideas with greater development needs may still require product development formalisms to deliver a reliable consumer product. Projects safely in the mid-range of funding goal are probably sufficiently mature and appropriately scaled to know their real funding requirements and achieve about a 50% success rate once funded.

What are some of the limitations of this data set?

The primary limitation of this data set is its representativeness as a sample of the entire population of project data. Compared to the complete submission statistics distributed by Kickstarter online which extends to 2019 (Figure 1), the top three categories of projects by submission number and successful dollars are Games, Design, and Technology. In the subset to 2017, the top three categories are Theater, Music, and Technology. While the overall, first-place success rates of the Theater category in the subset are comparable between the subset and complete data set (60.02% to 60.04%), the lack of data on games obscures the fact that this category has the greatest monetized value in Kickstarter history. For example, a card game product named "Exploding Kittens" exceeded its $10K funding goal in eight minutes on January 27,2015 and completed its fundraising on February 19, 2015 with a total of $8.7M and 219,382 backers making it the fourth most funded site on Kickstarter. This information does not appear to be in the reduced data set.

Other limitations of the data set are that it does not include all the possible categories as shown on Kickstarter. The absence of dollars split into success/fail/cancel categories does not allow for direct comparison to the global statistics.

Future Work

What are other possible tables or graphs could be created to offer additional insight?

The United States is often recognized as being the best in technology innovation in the world, having made major advances in space, computers, defense, medicine, modern forecasting, and agriculture among many other topics. In addition, innovations in US pop culture and media reach global audiences and drive international trends. However, given the "age difference" between the youthful US and other ancient cultures in Asia, Africa, and Europe is it necessarily true that US innovations will lead the world in *sustained* cultural impact that lasts for centuries? A natural further analysis of the data set would be to examine the entire Kickstarter data set to examine what nascent products are successful in the international market and consider what these products portend for future cultures and consumers.

Summary

In this first Data Analytics Homework Assignment, the student considered a reduced data set from the crowdfunding site, Kickstarter, an online service that facilitates public funding of innovations in the arts and technology. Using various Excel techniques, exhibits were produced that showed success rates of projects in various categories spanning art, theatre, games, music, publishing, and technology as functions of funding and time. Conclusions were drawn that indicated this type of service for the development of arts and technology is serving an unmet consumer need and certain types of products are more successful than others. In addition, such "open innovation" ensures maximum opportunity and rapid deployment for a wide variety of projects but larger, more complex, higher-risk projects may still require product development formalisms typically associated with robust consumer products.

Unit 1 learning objectives achieved: 1) exposure to high-level analytical strategies, 2) proficiency in Excel functionality, Pivot Tables, conditional formatting, charts, and filters.