Shark Tank Companies

by Ryan Coslove rmc326



Hello, my name is Ryan Coslove. I am currently an undergraduate at Rutgers University. I am in my senior year and at the end of the Spring 2022 semester will receive my Bachelor of Science. This blog is designed for my CS142 Data 101 course. I am taking this course to get a better understanding of manipulating and calculating data. In this blog, I will be using a data set that contains numerous categorical and numerical attributes derived from the hit television show, “Shark Tank”.

**The Show’s Background**

As described on Wikipedia, ***Shark Tank*** is an American business [reality television series](https://en.wikipedia.org/wiki/Reality_television#Investments). It shows [entrepreneurs](https://en.wikipedia.org/wiki/Entrepreneurship) making business presentations to a panel of five [investors](https://en.wikipedia.org/wiki/Investor) or "sharks," who decide whether to invest in their company. The "sharks" decide whether to invest in a company as entrepreneurs make business presentations on their company or product. The sharks often find weaknesses and faults in an entrepreneur's valuation of their company, product, or business model. The entrepreneurs come in with an initial offer and if a shark is interested in investing, often makes a counter offer (usually asking for a higher evaluation or percentage of the company’s profits).

**The Data**

The data was provided at <https://www.kaggle.com/yamqwe/shark-tank-companiese?select=Shark+Tank+Companies.csv>. The data set’s description is as follows: For this study, the author took the 6 seasons of Shark Tank which consists of 122 episodes and 495 companies, to see which ones performed the best and were asked the most. In addition, the author wanted to see if there are any trends when it comes to those companies getting a deal with at least one shark. The author asked, “Which of the Shark Tank companies will be the best?” I have my own hypothesis that differ from the author’s.

A sample of the data set:

A screenshot of a computer

Description automatically generated with medium confidence

The categorical columns are deal, description, episode, category, entrepreneurs, location, website, season, shark1, shark2, shark3, shark4, shark5, title, episode-season, Multiple Entrepreneurs.

The numerical columns are askedFor, exchangeForStake, valuation.

**Plots**

For my 1st hypothesis, I wanted to see if deals were more or less successful with higher or lower rates exchanged with the Sharks. My hypothesis is “Successful deals have a higher mean exchange for stake than unsuccessful deals”. The null hypothesis, as result, would be that “successful deals have the same mean exchange for stake as unsuccessful deals”. The alternative hypothesis is that “Successful deals have a lower mean exchange for stake than unsuccessful deals”.

This is the resulting boxplot for this hypothesis:

Chart, box and whisker chart

Description automatically generated

Something that was fascinating to see was that a company tried to sell itself outright to a shark (sell 100% of the company). The deal was unsuccessful, inferring no shark wanted to buy the whole company and invest. In the box plots, we can see that overall the unsuccessful deals had a typically higher exchange rate for stake in the company. Unsuccessful deals also had the highest outlier values. Some important data that was collected in this hypothesis are:

|  |  |  |
| --- | --- | --- |
|  | Deal | No Deal |
| Number of Deals/No Deals | 251 | 244 |
| Mean Stake Exchange | 16.7% | 18.41% |
| Standard Deviation of Stake Exchange | 9.17 | 10.85 |

Using these statistics, I found that the z-score was **z=1.889358**. This resulted in a p-value of **p=0.0294**. Given that p ≤ 0.05, we are able to reject the null hypothesis. This can help us infer that the alternative hypothesis is true, that successful deals have a lower mean exchange for stake than unsuccessful deals. A possible reason for this is perhaps the deals that had a lower stake exchanged were more promising to the sharks and required them to do less work. Another possible reason for this is that the successful deals asked for less money in the exchange (in conjunction with lower stake exchanges).

For my 2nd hypothesis, I wanted to see if valuations were higher when the money asked for was less than $150,000 compared to valuations where the money asked for was more than $150,000. I decided the amount $150,000 because that was the resulting median value of the askedFor column. So my hypothesis is, “Valuations are higher for companies that asked for less than or equal to $150,000 than valuations for companies that asked for more than $150,0000”. The null hypothesis is “valuations are the same for companies that asked for less than or equal to $150,000 as compared to companies that asked for more than $150,000”. The alternative hypothesis is “Valuations are lower for companies that asked for less than or equal to $150,000 than valuations for companies that asked for more than $150,000”.

This is the resulting boxplot for the hypothesis:

Chart, box and whisker chart

Description automatically generated

Based on the boxplots, we can see that typically companies that asked for more than $150,000 proved to have higher valuations for their companies. Companies that asked for less than $150,000 typically had lower valuations. These are data found during testing:

|  |  |  |
| --- | --- | --- |
|  | Asked for <= $150,000 | Asked for > $150,000 |
| Number of Companies | 293 | 202 |
| Mean Valuation | $647,223.90 | $4,368,033 |
| Standard Deviation of Valuations | 554,775.90 | 5,108,807 |

Using these statistics, I found that the z-score was **z=10.30945**. This resulted in a p-value of **p=0.0000**. Given that p ≤ 0.05, we are able to reject the null hypothesis. This can help us infer that the alternative hypothesis is true, that valuations are lower for companies that asked for less than or equal to $150,000 than valuations for companies that asked for more than $150,000. A possible reason for this is with more money invested in the company, the more likely the success rate of that company to turn a profit. Companies that asked for more were probably less risky to invest in, so sharks were likely more willing to buy larger and invest more to see the valuations return higher.

**Conclusions**

What do we do with this information? Well it is likely that your highest success rate when entering Shark Tank is to ask for more than $150,000 and to ask/finish with a lower stake exchanged. The sharks on the show seem to invest best in low-risk-high-reward. They are willing to spend more money and accept lower stake exchanges if they deem the company and product being presented to be formidable, low-risk, and can return high-profit. When watching the show, I would suggest looking for these in the beginning of a company’s pitch (high asking amount and low evaluation). Statistically it seems favorable that those companies will get roughly the return they expect from the sharks and have a successful deal.

I enjoyed using this data set because I am a fan of the show and wanted to learn some of the possible reasons the sharks invest in a company, as proven by data. I found the sample size to be large enough to make the data reliable and limited to no bias. The data set contained data from 6 seasons. There are more seasons than the data set provided, but using this data and our hypothesis we can predict in later seasons how companies and the sharks will succeed or fail based on asking price and stake exchanged. There are obviously more hypotheses that can be tested using this data set, but the two I focused on felt significant to the nature of the negotiation and the success of companies receiving a deal. If you find yourself looking to receive investments, perhaps use this data set to best optimize your chances of getting a deal from a shark.