Space, The Final Frontier

Amanda Liang & Imer Belica

Intro

Space exploration is exciting but can also be very costly. Space missions can be a bit of a high-risk investment where the large sums of money spent up front on materials and testing could all be lost in a failed launch. With so many companies trying to get their hands into the space business, where should one invest? Our goal was to examine the relationship between cost and various other features to provide a better look into the space industry for potential investors or anyone interested in the topic.

Background

Some similar work has been done on the dataset we mined in this project. Others sought to gain insight into things such as measuring the different features against each other in trying to determine whether the future of space is private. Researchers attempted to determine this by looking at features such as cost, success, and the trend in terms of the number of launches between private and state-based space launches. Other researchers made attempts to visualize the space race, a critical period in the history of time in which technology thrived and the prioritization of space travel was at an all time high for two of the largest powers in the world. The researchers once again sought to gain insight on this event by looking at the various features in the dataset and looking for things like correlation between different features specifically those pertaining to the US and the USSR.

Another interesting point of emphasis in mining this dataset was determining the leading causes in why space missions fail. A number of researches aimed to learn about the failures by looking at all the features provided in the dataset and once again finding the correlation between these failures and those features. These researchers also dove deeper into the specifics of each failure to try to determine cause.

While these projects aimed to gain insight into areas such as history, with the visualization of the Space Race, and causation for failed missions. We placed an emphasis on cost in our project and took a more global approach to looking at the data. We aimed to find correlation between cost and numerous other features in the dataset. We found patterns in things like cost and private or state run missions as well as cost and mission status. Additionally we aimed to determine what we could expect of the cost of future missions by finding a pattern in the declining cost of space travel since 1957. This analysis can potentially be helpful for the space industry as a whole as well as investors looking to put their hard earned money into the field as it informs everyone interested in the topic of cost related patterns in the space industry.

Data & Pre-Processing

The data used for this project consists of every space mission since 1957 which is when the Space Race started. This ended up being 4323 rows in the data set, with 8 features after the “Private or State Run" column was pulled in from a similar dataset. Before any calculations could be done, the data had to be cleaned. Many of the older records did not have the cost listed and ended up being removed. Some of the dates were also missing the time of day, so all the timestamps were scrubbed (it wasn’t necessary to keep the times as the overall time scale would be much larger). After all that the z-score for each row (based on cost) was calculated. Outliers (Z-scores with the absolute value of 3 or higher) were also removed as the upcoming calculations would involve a lot of averages which are easily affected by outliers. After all this pre-processing, there were 949 rows left. While no columns were removed from the original dataset, there were a few features that would not be useful for the calculations, such as company name and the detail section. The most useful features for this project were cost, private or state run, and mission status.

Tools

The calculations and visualizations were all done in python using Jupyter Notebook through Anaconda. Originally, there was an additional goal to see if we could generate a predictive model, however there was importing issues with several classes needed for the calculations (despite having downloaded Anaconda, which came with those classes already downloaded). Luckily there was no trouble with the bulk of the classes we used such as pandas, numpy, scipy, and seaport.

Cost vs. Time

The first thing we calculated was the average cost of space missions for every year. After graphing (Figure 1), we can see there is a negative correlation between average cost per year and time (ignoring the earliest few point on the graph since the dataset is much less accurate the earlier on a mission was). From this we gathered that our initial hypothesis, a decrease in cost over time, was correct. While the visual doesn’t explain exactly *why* the cost is decreasing, there are a few factors that we hypothesized could be behind the falling prices. For one, technology has improved a lot throughout time. These technological enhancements could lead to cheaper supplies, more pre-existing research, or faster build times, all of which would help lower the cost of future missions. Another factor that is not covered in our research is the type of missions that are happening. It could be possible that the more recent missions are fairly small and therefore cheaper while the older ones were more akin to full-sized rockets carrying people into space. It is also possible that the budget was simply being cut as time went on. Most likely it was a mixture of some or of all these factors that led to results we see.

Private vs. State

The second graph we created was the average cost per year for both private and state-run companies. It's pretty easy to see that the state-run companies' average yearly cost looks a lot like the previous graph. It follows the same general downward pattern over the same time frame. Meanwhile, the private run companies seem to congregate at the lower end of the price spectrum in comparison. However, it seems that in the recent decade the private companies seemed to have exceeded the average cost of the state-run companies. Our initial hypothesis was partially correct where initially state-run missions were, on average, more expensive than private missions. Around the mid 2000's the private companies' average cost started to increase while the states' kept trending down. Our initial hypothesis was based off the idea that state-run companies would try to maximize their budget while private companies would try to budget their funds more. Also, there was a thought that state-run companies may have more regulations to follow (that private companies do not) that may increase the price.

Cost vs. Mission Status

The third graph is a bar graph showing the average cost of missions based on their status (successful, failure, etc.). We can see from an initial look that the average cost of successful missions appear to be much higher than the other three variations of failure. However, since there aren't many failures in the dataset, the results may be less accurate or indicative of the actual relationship between mission status and cost. The graph has error lines on each bar showing the possible range in error. So while it looks like there may be a clear difference in average cost, one can see from the error bars that the failed mission cost could be much higher (or lower). Overall the information learned from this graph is mostly inconclusive.

Conclusion

In conclusion, we have learned that the price of space exploration as a whole has been steadily dropping. The state-run companies seem to contribute most to the overall averages (more state missions) since they show the same downward pattern, while private companies are more constant in their medium-to-low-range mission costs.

If we were to extend our research more, there are a lot more interesting things we could potentially look at with this dataset. For example, the average price of failed and successful missions based on private or state (though with a small pool of failures it may be hard to get anything worthwhile). Another interesting idea would be to look at the specific stats (average cost, success rate, etc.) for each company or for each country. One could look at the number of launches from each country to see who has been the most active in space. The scrapped predictive model we had originally planned to do could be given another shot as well. If the earlier records had been more

complete with the costs, it would be interesting to compare the space missions from the USA and Russia during the Space Race era. Who was the most cost-efficient? Who had the best success rate? There are many ways you can look at the Space Race and determine who “won” in certain circumstances. If one was to look at the missing cost records from each company or country, there could be an interesting insight into the availability of such records from different places. Do state-run companies share their costs more than private companies do? Or is it a country-based issue? Maybe it’s just that some records were lost over time or due to other, more complex, circumstances.

Appendix: Figures

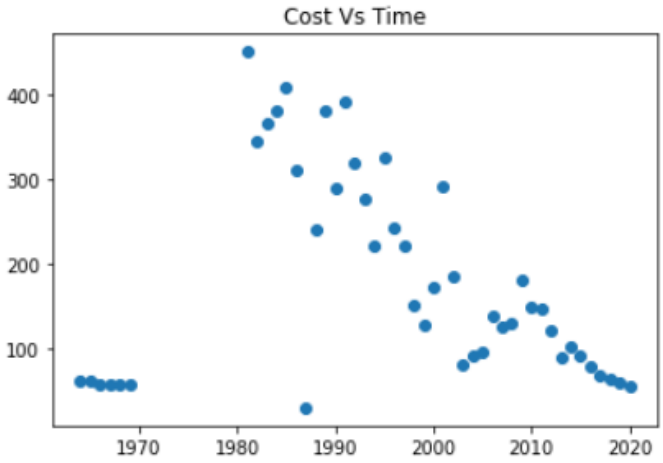

Figure 1

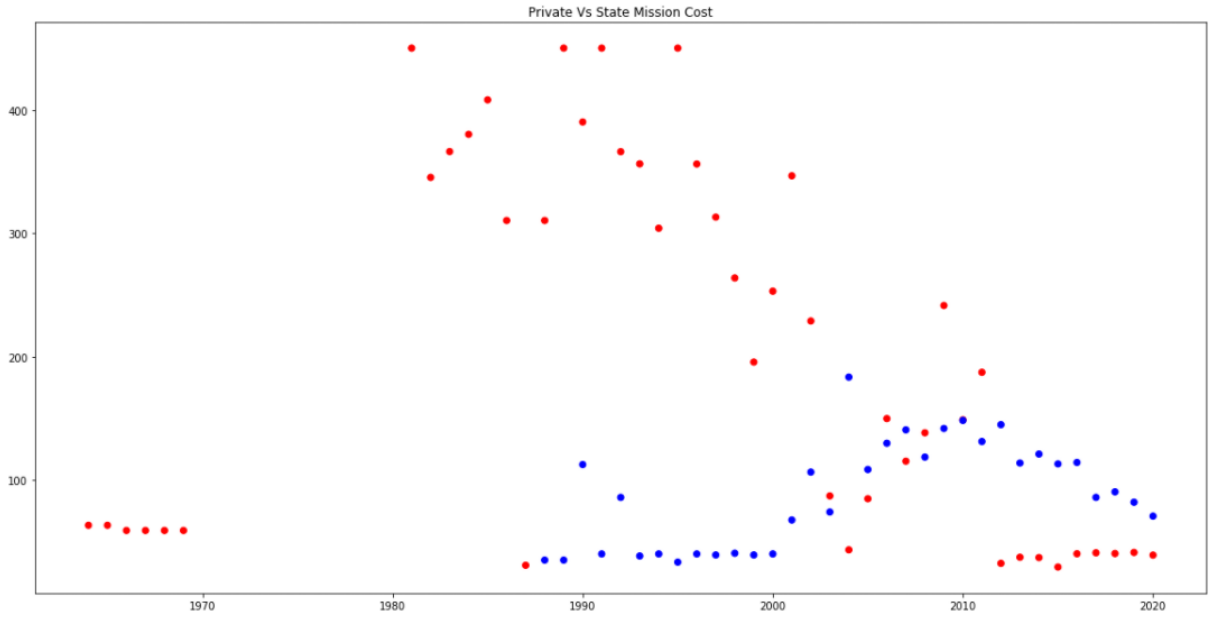

Figure 2

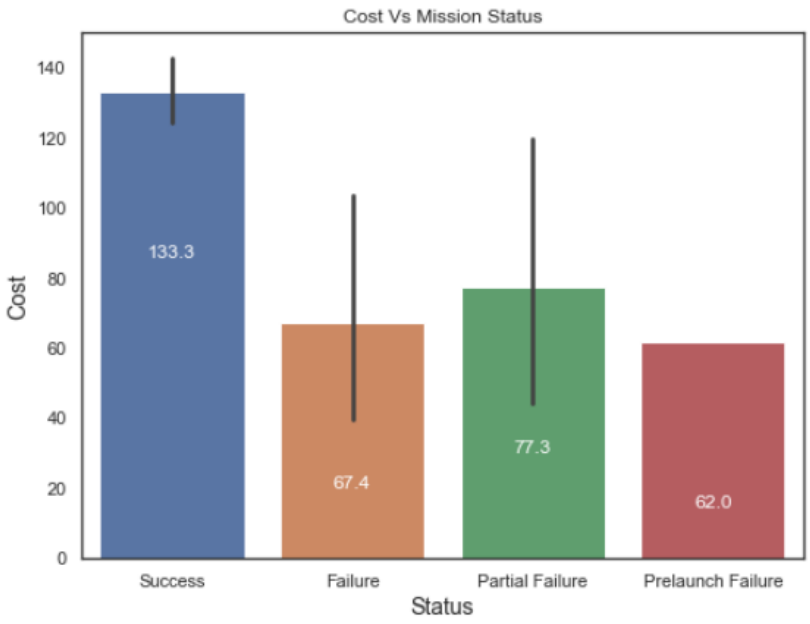
Cost Vs Mission Status
140
120
100
80
60
40
20
0
Cost
133.3
67.4
77.3
62.0
Success
Failure
Partial Failure
Prelaunch Failure
Status

Figure 3