

Statistical Analysis

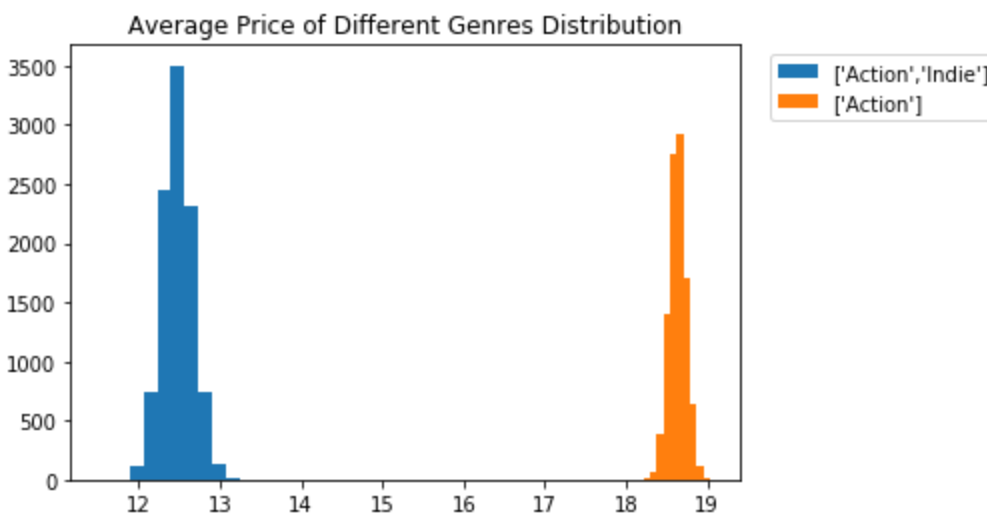
From our storytelling project, there is a lot of interesting information. In this part of the process, I am going to further explore some of the information presents and perform A/B testing.

There are 3 questions that I want to answer:

1. Do ['Action'] games cost more than ['Action', 'Indie'] games?
2. Do multiplayer games cost less than single-player games?
3. Among the two most popular games (Counter-Strike: Global Defensive and Team

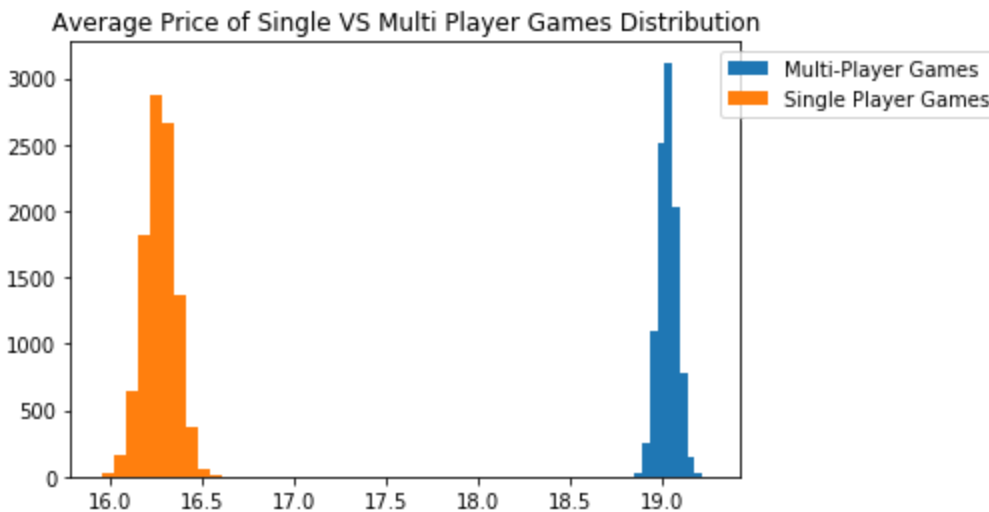
Fortress 2), which has a better reputation in terms of recommendation?

The null hypothesis for question 1 is that the average price of ['Action'] games greater or equals the average price of ['Action', 'Indie'] games? After filtering out the rows that contain specific genres, I used bootstrap to loop over 10000 cycles and calculated the confidence interval of the two groups. The 95% confidence interval is [5.79101534, 6.50657473]. I plotted the distribution graph as follows:



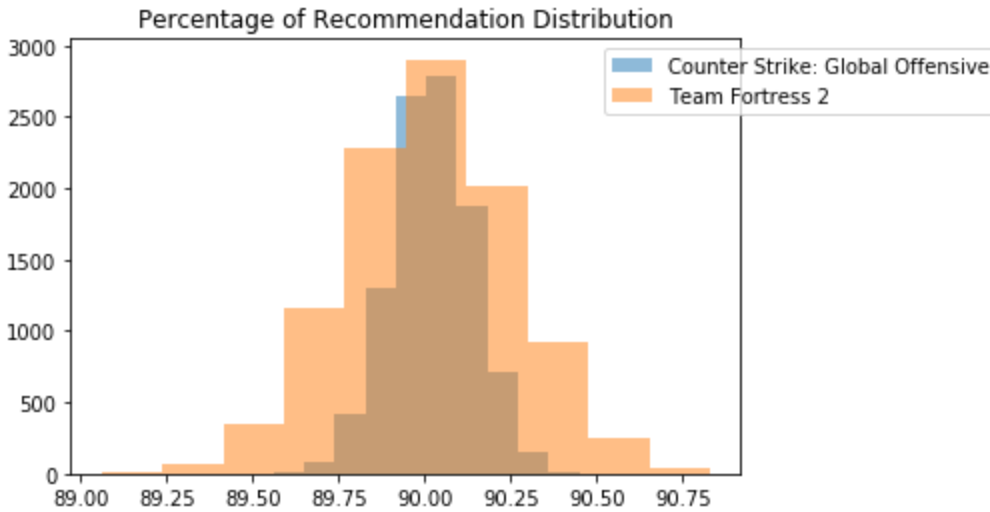
Finally, I calculated the p-value which is less than the alpha level thus concluding that it is lucid that ['Action'] games have higher average price than ['Action', 'Indie'] games.

For question 2, the null hypothesis is that the average price of single-player games is greater or equal to the average price of multiplayer games. After filtering out the rows I used bootstrap to loop over 10000 cycles to find the confidence interval. The result came back as $[-2.93274549, -2.60298355]$. I plotted the distribution of the two sample groups:



Based on the confidence interval and distribution graph, it is clear that Single players games do not cost more than multiplayer games. Multiplayer games cost more than single-player games.

For question 3, the null hypothesis is that the percentage of recommended from Counter-Strike: Global Defense is less or equal to the percentage of recommended from Team Fortress 2. After selecting the appropriate rows, I performed bootstrap on the dfs over 10000 cycles and obtained the confidence interval of $[-0.43, 0.4505]$. I plotted the distribution graph as following:



Based on the evidence presented, because the distributions graph overlay each other and confidence intervals contain positive and negative boundaries, it is inconclusive whether or not the percentage of recommendation from Counter-Strike: Global Defense is greater or equal to the percentage of recommendation from Team Fortress 2.