Time Spent Analysis of Regal Movie Theater & Victoria's Secret at Summerlin Las Vegas
By Connor J. Foley

Proposal: The average time spent within the Regal Movie Theater by visitors to the Summerlin shopping center should be significantly greater than that of the average time visited to the Victoria's Secret store.

Initial Finding: During first investigations as part of the Data Science exercise the average time spent at the two locations was found to be relatively similar.

	Regal	VS
N	184135	127784
Mean	493.58	394.42

Figure 1: Average Time spent and Sample Size

This finding raised concerns to the validity of the data; however a mean calculation is not sufficient to determine whether or not two populations are different.

Further Development: Based on the similarity in average time spent the following assumptions were garnered: the behavior of the visitors' to the two stores was similar and the distribution of that behavior followed a normal distribution. Based upon these two primary assumptions a one-tailed T-test was conducted to challenge this reasoning and the hypothesis that the time spent in the theater and store was statistically equal:

H_o: Avg $Time_{regal}$ = Avg $Time_{vs}$
H_a: Avg $Time_{regal}$ > Avg $Time_{vs}$

I determined the α to be 0.05 for this test as there was no other criteria or limitations to alter this standard set.

	Regal	VS
T-Test	22.506	
Pvalue(1)	2.24E-112	
Pvalue(2)	4.48E-112	

Figure 2: T-Test

From the results of the T-Test I rejected the null hypothesis that the average time spent was equal.

Validation: In order to validate these findings I tested the assumptions outlined prior to the t-test, that the behavior of the visitors follows a normal distribution.

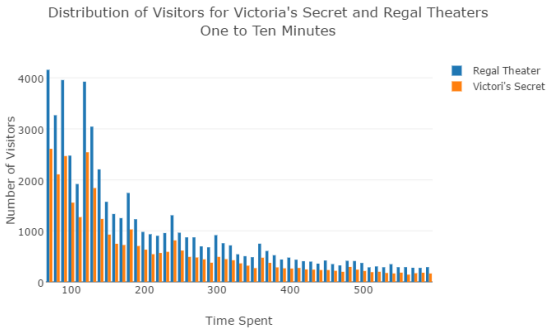

Figure 3: Histogram of Time Spent

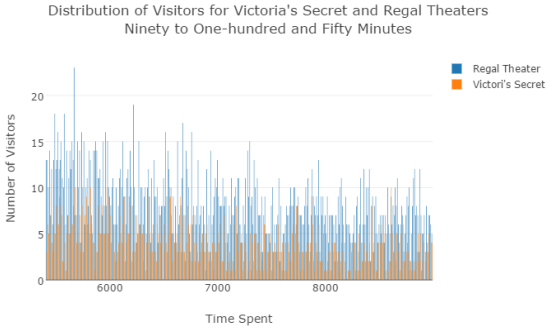

Figure 4: Histogram of Time Spent

Conclusion: Although the histograms showcase a single primary node, we would expect that there would be a large cluster of visitors active in the Cinema between Ninety and One-hundred & Fifty Minutes, this would appear as a secondary node and the behavior of the visitors would not follow a normal distribution. Additional data needs to be collected to confirm this theory, but most likely the current dataset does not provide a complete picture or requires additional aggregation in order to produce the accurate, secondary clustering.

Source code supporting calculations as well as graphs can be found at:
http://github.com/Foley-CJ/hhc_Exercise/blobl/master/timeAnalysis.ipynb