

Large venue attendance analytics:

General Idea:

Predicting whether an event will sell out is important information to have when looking to optimise marketing campaign spend over the course of a year. The shows that will sell out may not need that aggressive campaigns while others might need a little extra boost. In this project, I am going to focus on the Durham performance of the Arts Center (DPAC) in Durham, NC, where a wide variety of shows take place every year. My plan is to build a pipeline that scrapes the website of all the necessary show information, performs text mining and data engineering to create the appropriate variables, uses data science to produce predictions and populates a dashboard with the results.

Pipeline:

1. Scrape the website (<https://www.dpacnc.com/>) in a daily fashion. Acquire information like date, description, price, location of seat on theater, etc
2. Perform text mining in description to create relevant variables
3. Employ predictive modeling to predict sold out events (binary classification)
4. Populate a dashboard with the output of the predictions