Project 2 Proposal

Predicting **Opening Weekend Sales for Animated Movies**

Client:

McDonald’s

Problem:

McDonald’s considering partner with an animated movie and feature it on their Happy Meal toys? Among those upcoming in 2020, which one should they choose?

MVP:

An MVP would be an Opening Weekend Sales Prediction Model, trained and validated on historical data scraped from boxofficemojo.com. We would use it to predict for upcoming animated movies and pick the ‘hit’ one.

## Data:

historical data for individual movies scraped from boxofficemojo.com

|  |  |  |
| --- | --- | --- |
| Variable | Type | Description |
| Opening Weekend Sales | Continuous | Y variable |
| Projected Budget | Continuous |  |
| Director LGT\* | Continuous | LGT of the Director’s movies *before* the movie |
| Actor LGT | Continuous | Average LGT for 3 Leads *before* the movie |
| Composer LGT | Continuous | LGT for the main composer *before* the movie |
| Holiday Season | Nominal | Which holiday season the release falls on, if any |
| Sequel | Nominal (Y/N) | Whether or not the movie is a sequel |
| MPAA Rating | str |  |
| Distributer | str |  |
| Run Time | int |  |
| # concurrent releases | int |  |

LGT\*: Life Gross Total

### Project 2 Brainstorming Guide

Project 2 will be an individual project. Still, you can brainstorm in

groups so that you can build upon each others' ideas and generate a

lot of potential directions. You will still decide for yourself which

of these directions to take.

## Potential Clients

\* Movie studio

\* Movie futures investment firm

\* Producer

\* Marketing company

## Questions we might answer for clients using data

\* Given potential movies to make, what movies should I make?

\* Given movies that are being made, which movies should I invest in?

\* Given movies that are being made, where should I allocate my

marketing budget?

\* Can I predict the profitability of movies that are about to be

released?

\* Can I predict the eventual/total/longer term profitability of movies

that have just been released?

\* Can I predict international success?

\* Can I predict what movies will be "cult classics"?

## Possible Inputs

(things to collect in our scraping)

# Project 2: Regression

###### Weeks 2 and 3

## Backstory:

Using information we scrape from the web, build linear regression models from which we can learn about movies, sports, or categories.

### Data:

\* \*\*acquisition\*\*: web scraping

\* \*\*storage\*\*: flat files

\* \*\*sources\*\*: (as listed below or any other publicly available information)

- movie: boxofficemojo.com, imdb.com

- sports: sports-reference.com

### Skills:

\* basics of the web (requests, HTML, CSS, JavaScript)

\* web scraping

\* `numpy` and `pandas`

\* `statsmodels`, `scikit-learn`

### Analysis:

\* linear regression is required, other regression methods are optional

\* We recommend at least 1000 rows and 10 features. Make sure not to have too many categorical features.

## Deliverable/communication:

\* organized project repository

\* slide presentation

\* visual and oral communication in presentations

\* write-up of process and results

\* 4 minute presentations

\* [Project Logistics](https://docs.google.com/spreadsheets/d/1jukSR5t1\_iOm-RpZO1A9TTA5KJfhBdp-qB4Yewqt-oM/edit?usp=sharing)

### Design:

\* iterative design process

\* "MVP"s and building outward

\* [stand-ups/scrums](https://en.wikipedia.org/wiki/Scrum\_(software\_development)) (1 minute progress updates to the class)

## More information:

We'll learn about web scraping using two popular tools - BeautifulSoup and Selenium. You must know the very basics of HTML. We can also evolve the way we use Jupyter notebooks; during this project, we begin to use the notebook as a development scratchpad, where we test things out through interactive scripting, but then solidify our work in python modules with reusable functions and classes.

We'll practice using linear regression. We'll have a first taste of feature selection, this time based on our intuition and some trial and error, and we'll build and refine our models.