Proposal (1000 rows x 10 features)

Client: McDonald’s

Problem: What animated movies to feature on Happy Meal toys in 2020?

Goal: Predict opening weekend sales for animated movies

Data Source:

Features to include: say u’re the expert, would you say a movie released in 2018 will have a higher sale than one released in 2019? If not, don’t use it.

* Season; Holiday
* Sub-Genre
* Lead Actors (3)
* Projected size of budget (indicate quality?)
* Writer
* Adapted source (book/TV show, etc)
* Release date (month/season/holiday/etc)
* Franchise
* Producer(s)
* Social media
* type/scale of distribution

### 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)

\* Genre

\* Actors (all)

\* Lead actors

\* Projected size of budget

\* Writer

\* Adapted source (book/TV show, etc)

\* Release date (month/season/holiday/etc)

\* Franchise

\* Director(s)

\* Producer(s)

\* Opening night/1st weekend/1st week box office gross

\* Social media

\* type/scale of distribution

\* # google searches

\* Wikipedia page views

\* Competition (concurrent releases)

\* DVD sales, etc

\* Critic ratings

\* International... stuff...

# 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.

This project will give you the freedom to challenge yourself, no matter your skill level. Find your boundaries and push them a little further. We are very excited to see what you will learn and do for Project Luther!