

Project Luther: learning linear regression hands-on

- Client: studio that specializes in action movies
- Client's question: Given my movie's budget, director, & top 3 actors, can we predict the movie's success opening weekend?
- Data Sample: Action movies from The Numbers (website) that have info for budget, director, actors
- My Model: ~~Simple~~ Linear Regression (no polynomials)
- The Features:
 - ① Budget
 - ② Director Average Dom. gross of past movies (through 2008)
 - ③ Average of Top 3 Actors' Average Domestic Gross of past movies (through 2008)

done to prevent data leakage from training/test set
- The Outcome(s):
 - Errors:
 - ~~Budget~~ Base: ± 42 m dollars
 - Base + budget: ± 30 m dollars
 - Base + budget + Director + actors: ± 28 m dollars
 - Some outliers: The Dark Knight / ~~John~~ Chris Nolan
 - Effect of Factors:
 - $\uparrow \$1$ m budget $\rightarrow \uparrow \$270$ k opening wknd dom gross
 - $\uparrow \$1$ m director $\rightarrow \uparrow \$110$ k opening wknd dom gross
 - $\uparrow \$1$ m actors $\rightarrow \uparrow \$240$ k opening wknd dom gross
- Takeaway(s):
 - Actors have stronger effect than director chosen
 - The Dark Knight did really well