Data Science Movie Machine

Agenda

- What's the project?
- What's the big idea?
- Okay, and..?
- Where'd you get that?
- Uh huh, and what's next?

What's the project?

Utilizing data science, can we predict the box office returns and critical reception of films based on:

- The actors
- The director
- The genre
- The runtime

What's the big idea?

Aside from being a fun pastime, being able to accurately predict movie earnings is a powerful tool.

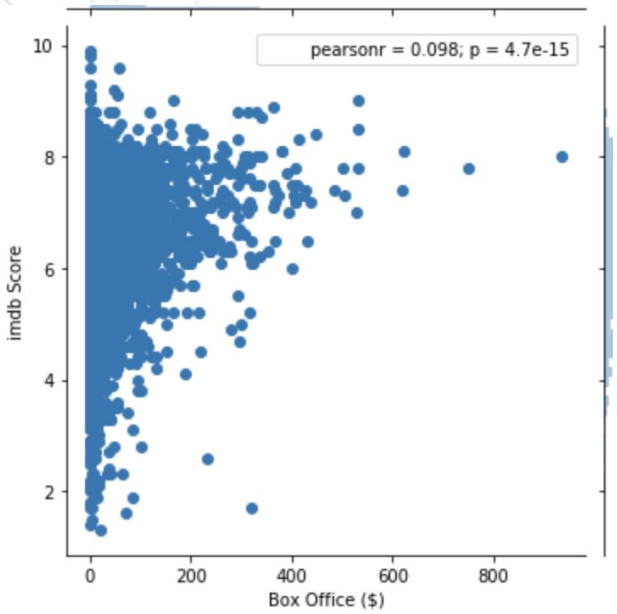
If the predictive power is strong, it would be incredibly influential in the movie business.

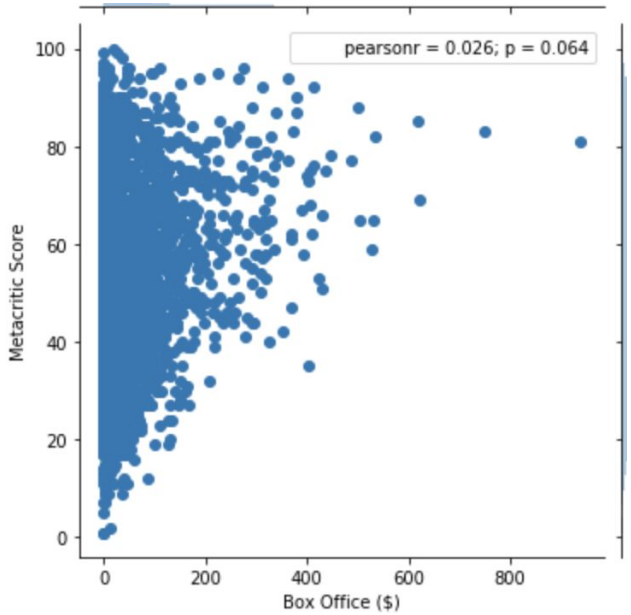
Okay, and..?

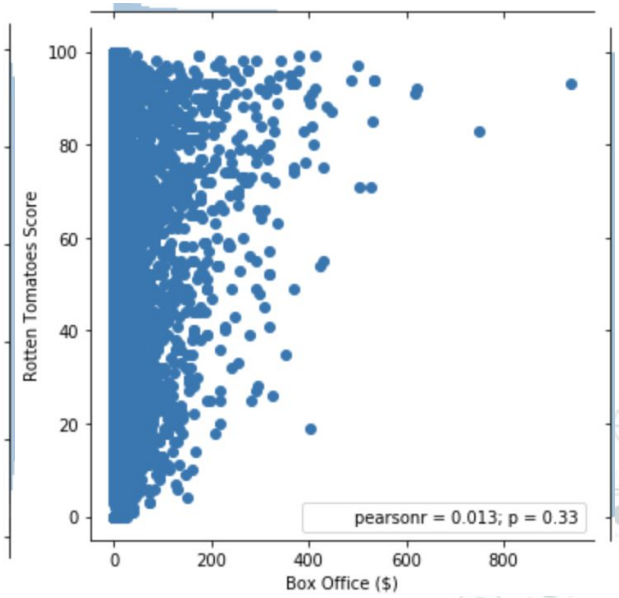
Let’s make it a fun and accessible tool!

Being able to imagine your dream movie with your choice of cast members, directors, and genres is my idea of a good time. You’ll be able to predict how well your movie idea might fare based on the model!

Where'd you get that?







Where'd you get that?

```
imdf['Rotten Tomatoes Score'].corr(imdf['Metacritic Score'])
```

```
0.9251684372941259
```

```
imdf['Rotten Tomatoes Score'].corr(imdf['imdb Score'])
```

```
0.6487435779884944
```

```
imdf['imdb Score'].corr(imdf['Metacritic Score'])
```

```
0.7106685010822937
```

Where'd you get that?

```
new_df['Emma Stone'] = 1.0
new_df['Horror'] = 1.0
new_df['Christopher Nolan'] = 1.0
```

```
rf.predict(new_df.drop('Box Office ($)', axis = 1))/1000000
```

```
array([125.8246217])
```

Uh huh, and what's next?

- Python script
- Expand dataset over time
- Test predictive power on upcoming films
- Make up movies for fun