



**Code Code on the run who  
should I cast from them all?**

**Project: Gross Movie Prediction from IMDB**

**Fatima Loumaini- Metis\_Summer 2019**



# Overview

---

After a series of movie flops in the first months of 2019, Zara Nordstrom Warner's Bros Chairman of Board called for a meeting with her team to discuss the planning for 2020/2021.

Zara wants to use a model that can predict a movie gross amount based on the cast, directors and genres.



# Methodology and Model Results



- 1 IMDB data scraped for movies produced from 2015 to 2018
- 2 The model is using more than 9K rows of data
- 3 To avoid data leakage, the model is using data on actors, directors, genres and duration
- 4 R square(using Lasso) is 0.23.
  - Linear Regression scored a negative R square
  - Feature Engineering helped removing Directors as a feature

# Fab 4's Team Gross Movie Prediction



# Movie 1: Dotun's Movie

---

- Actor/Actress: Keanu Reeves, Ethan Hawke
- Genre: Family



## Movie 2: La La Land 2

---

- Actor/Actress: Naomi Watts, Natalie Portman, Tom Cruise
- Genre: Action





## Movie 3: Whiplash

---

- Actor/Actress: Meryl Streep, Keanu Reeves, Anna Kendrick
- Genre: Documentary, Drama



## Movie 4: Mother Brain

---

- Actor/Actress: Meryl Streep, Morgan Freeman, Tom Hanks
- Genre: Thriller, Sci-Fi





# The biggest gross amount for 2019 goes to the...

---

- Dotun's Movie: \$ 8.71Mil
- La La Land 2: \$ 72.96Mil
- Whiplash : \$ 35Mil
- Mother Brain: \$ 1.79Mil
- For fun:  
Top Gun: Maverick \$ 71.68Mil



# Spoiler



If you have an unlimited budget and you want a cast of 2 actresses and 2 actors that generate the biggest gross amount, you can pick from:

- **Actresses:** Dakota Johnson, Jennifer Lawrence
- **Actors:** Chris Pratt, Hugh Jackman

# Future work



- Include data on the movies release dates, budget, actor/actress salary etc
- Try other machine learning models



**Thank you!**