



# Анализ Киноиндустрии

ШКОЛА DA. ПОТОК З  
ПРИХОДЬКО КСЕНИЯ

# Проблема и решение

## ПРОБЛЕМА:

- ▶ Есть инвестор/кинопрокатчик/ зритель
- ▶ Как им понять по минимальному описанию фильма, стоит ли в него вкладываться/показывать/смотреть?

## РЕШЕНИЕ:

- ▶ На основе статистики за 100 лет понять, какие фильмы с большей вероятностью будут рентабельными, а какие провалятся в прокате

# Что изначально планировалось?

- ▶ Найти параметры фильма, по которым можно понять годный/плохой ли он (страна, жанр, режиссер, актер)
- ▶ Вывести показатель рентабельности для этих параметров
- ▶ Вывести комбинацию «режиссер + страна + жанр + актер» лучших и худших фильмов

# ЧТО В ИТОГЕ СДЕЛАНО?

- ▶ Проведен анализ киноиндустрии в целом
- ▶ Найдена комбинация для возможных «претендентов на Оскар» (т.е. фильмов, вошедших в топ по выбранным параметрам)
- ▶ Найдена комбинация для возможных «претендентов на Золотую Малину» (т.е. фильмов, вошедших в список худших по выбранным параметрам)

БЫЛО БЫ КРУТО СДЕЛАТЬ:

- ▶ Собрать базу с полными данными по итогам 2020 года, рейтингом из разных источников (в идеале в разбивке голосующих по возрасту, полу, стране и тд), сборам по странам и пересчитать результаты

# С чем были трудности?

- ▶ Было достаточно много пустых и неконсистентных значений (например, отсутствие выручки или некорректное значение, что легко проверить в интернете). Из аналитики, к сожалению, выпало очень много фильмов.
- ▶ Каждому фильму присущи **несколько** жанров, большой состав актеров, несколько стран-кинопроизводителей и даже несколько режиссеров (они указаны в алфавитном порядке, а не по приоритетности, поэтому взять первое значение - нельзя). Но бюджет и выручка есть только в общем для фильма. Поэтому привязать эти параметры к рентабельности нереально. Много интересной аналитики не сделать, тк в ячейке значения - список.
- ▶ Работа со списками и вычленение из них имен актеров/режиссеров/жанров/стран оказалась самыми сложными. Спасибо literal\_eval и json. Но это была боль.

# ИТОГ

- ▶ На основе имеющихся данных дать точный ответ на вопрос «будет ли фильм шикарным или нет» – сложно, но можно сделать **предположение**:
  - Если фильм снят в жанре {'Fantasy'} то, скорее всего, он будет хорош. Если фильм снят в жанре {'Horror'} то, скорее всего, он будет не очень.
  - Если в фильме снимаются {'James Earl Jones', 'Tom Hanks', 'Matt Damon', 'Warwick Davis', 'Matthew Lewis', 'Alan Rickman', 'Leonardo DiCaprio', 'Philip Ettington', 'Stan Lee', 'Brad Pitt', 'Samuel L. Jackson', 'John Ratzenberger', 'Frank Welker'} то, скорее всего, он будет хорош. Если в фильме снимаются {'Katharine Towne', 'Dan Hedaya', 'Yuriy Kutsenko', 'Jaime Pressly', 'Joe Pantoliano', 'Fyodor Bondarchuk', 'Maury Chaykin', "Ryan O'Neal", 'Stefania Rocca', 'Kathleen Freeman', 'Julianne Moore', 'Christopher McDonald', 'Sylvester Stallone'} то, скорее всего, он будет не очень.
  - Если фильм снят в {'Spain', 'Japan', 'New Zealand'} то, скорее всего, он будет хорош. Если фильм снят в {'Russia', 'Australia', 'Belgium'} то, скорее всего, он будет не очень.
  - Если фильм снял {'Alfred Hitchcock', 'Christopher Nolan', 'Clint Eastwood', 'Quentin Tarantino', 'Robert Zemeckis', 'Steven Spielberg', 'Joel Coen'} то, скорее всего, он будет хорош. Если фильм снял {'Gary Goldman', 'Blake Edwards', 'James Toback', 'nan', 'John Guillermin', 'David Gordon Green', 'John Schlesinger'} то, скорее всего, он будет не очень.