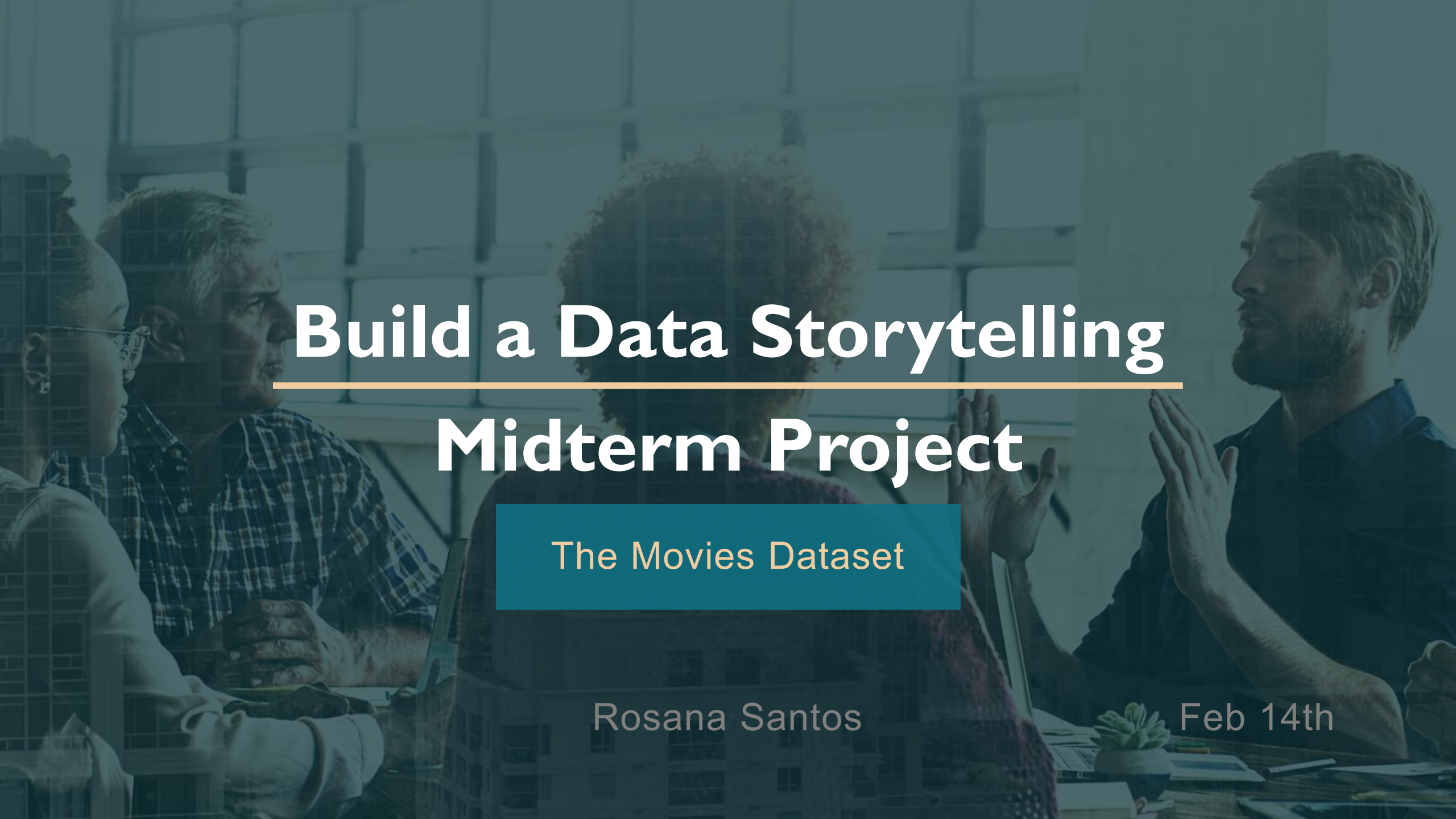


Build a Data Storytelling

Midterm Project

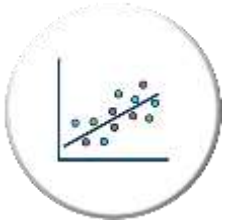
The Movies Dataset

Rosana Santos

Feb 14th

Executive Summary

The problem statement was determining



What drove **revenue** for the top 1 **genre** in average **popularity**?

1. Was it *Budget* that drove *revenue* for the *Top 1 genre* in average popularity?
2. Was it *Runtime* that drove *revenue* for the *Top 1 genre* in average popularity?
3. Was it *Cast* that drove *revenue* for the *Top 1 genre1* in average popularity?

Considering top 1 genre in average popularity,

1. *Budget* influences strongly on revenue
2. *Runtime* is weakly correlated to revenue
3. *Cast* is also weakly correlated to revenue.

Issue Tree

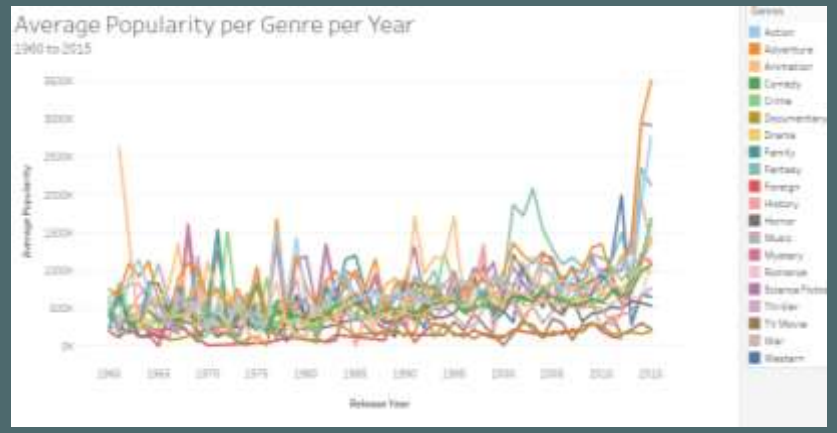
What drove **revenue**
for the top 1 **genre**
in average **popularity**?

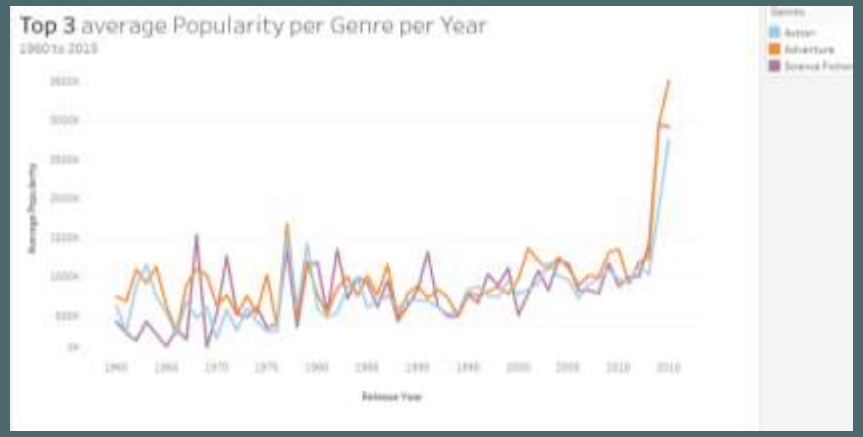
- Was it Budget?
- Was it Runtime?
- Was it Cast?

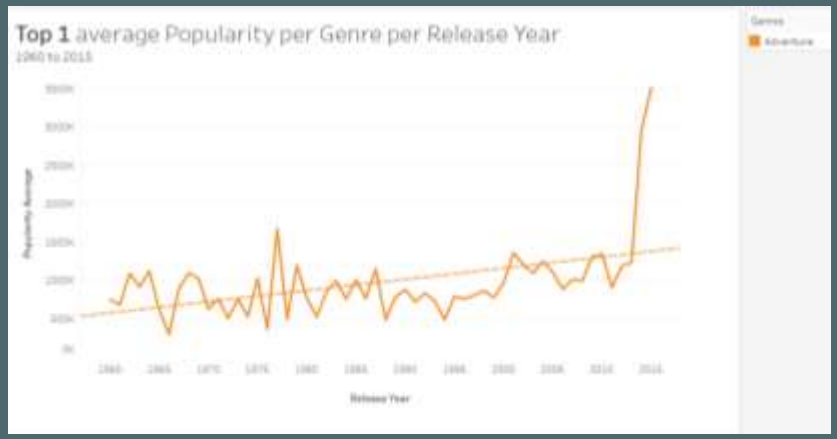
Average Popularity per Genre per Year

- Top 3 genres are **adventure**, **science fiction** and **action**.
- Top 1 genre is **adventure** considering average popularity.
- For the analyzes presented in sequence Top 1 will be considered.

https://public.tableau.com/profile/rosana7921#!/vizhome/Whatdroverevenuefortop1genreinaveragepopularity/RevenueTop3-Popularity?publish=yes





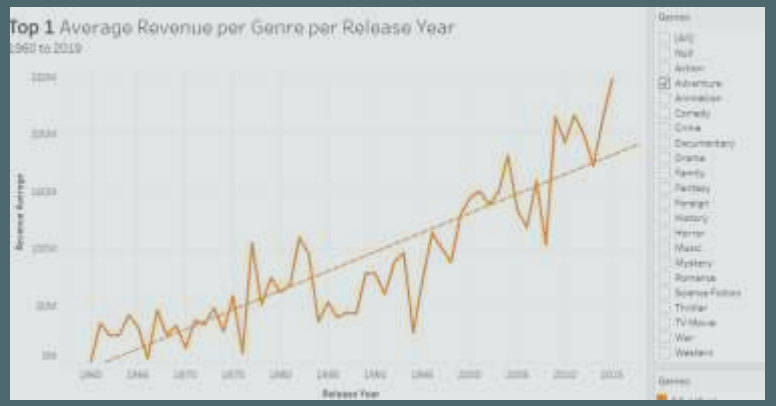


Average budget vs. average revenue

As **budget** increased over the years for **adventure** genre **revenue** increased.

Strong correlation.
R-squared = 0,9372

adventure
average **budget**



adventure average
revenue

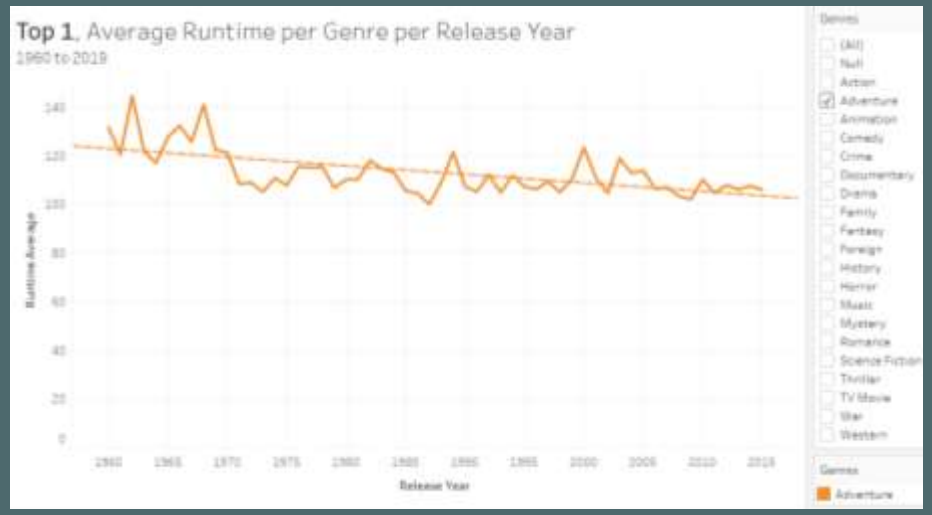
Average runtime vs. average revenue

Historical records show decrease tendency in average **runtime**.

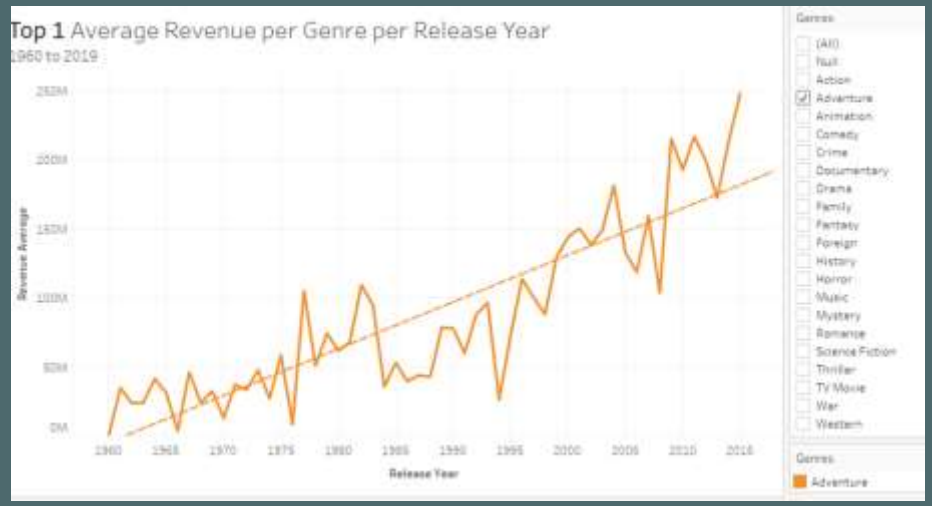
This parameter **does not affect** average **revenue** of adventure genre.

Weak correlation.

R-squared = 0,0090



adventure average **runtime**



adventure average **revenue**

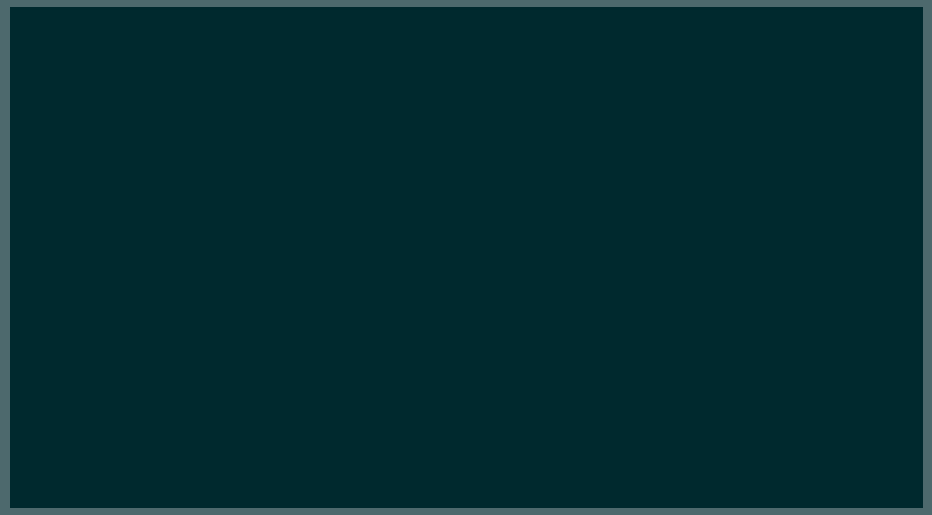
Distinct count cast vs. average revenue

I considered distinct counting for **cast**.

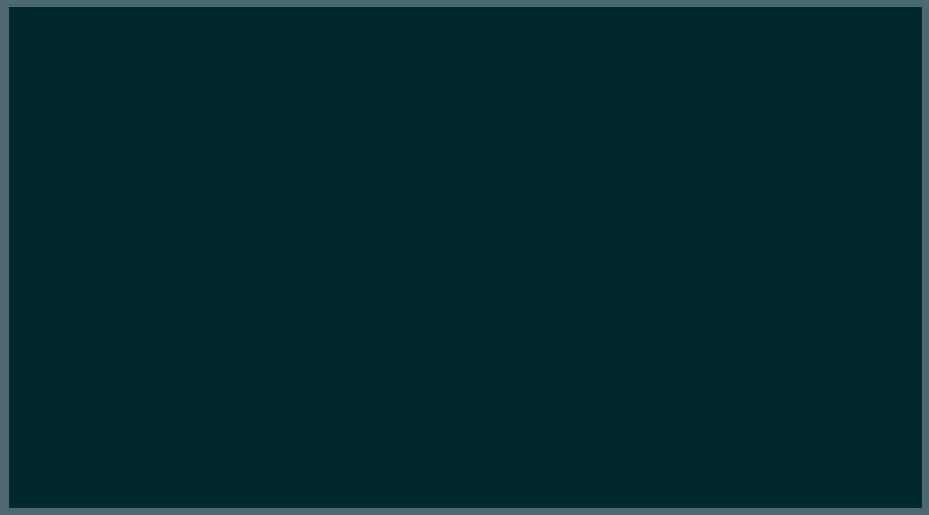
This parameter **does not affect** average **revenue** of adventure genre.

Weak correlation.

R-squared = 0,0111



adventure average **runtime**



adventure average **revenue**

Limitations and Biases

- Data is uncleaned and messy.
- Missingness.
- Outliers.

Next Steps

- Analyzing the top 3.
- Analyzing dataset as a whole.

Data has a better idea.

Thank you!