



YouTube Advertisement Putting

Yuqing Wu

Content



01 Introduction

02 EDA

03 Model

04 Result

05 Q&A

01 Introduction: Business Objective

Objective:

Analyze YouTube video clicks to determine which videos to advertise for, thereby increasing potential users and sales, and increasing revenue



01 Introduction: Two Main Questions

Q1: When should we put the advertisement?

Q2: What kind of video(s) should we put the advertise?



02 EDA: About Data

About Data:

- This csv dataset includes data for different types of YouTube videos in the United States region
- The json file includes category title corresponding to category ID

Data Types:

csv file: 16 columns of data in total

02 EDA: *About Data*

Video ID

Trending Date

Title

Channel Title

Category ID

Publish Time

Tags

Views

Likes

Dislikes

Comment Count

Thumbnail Link

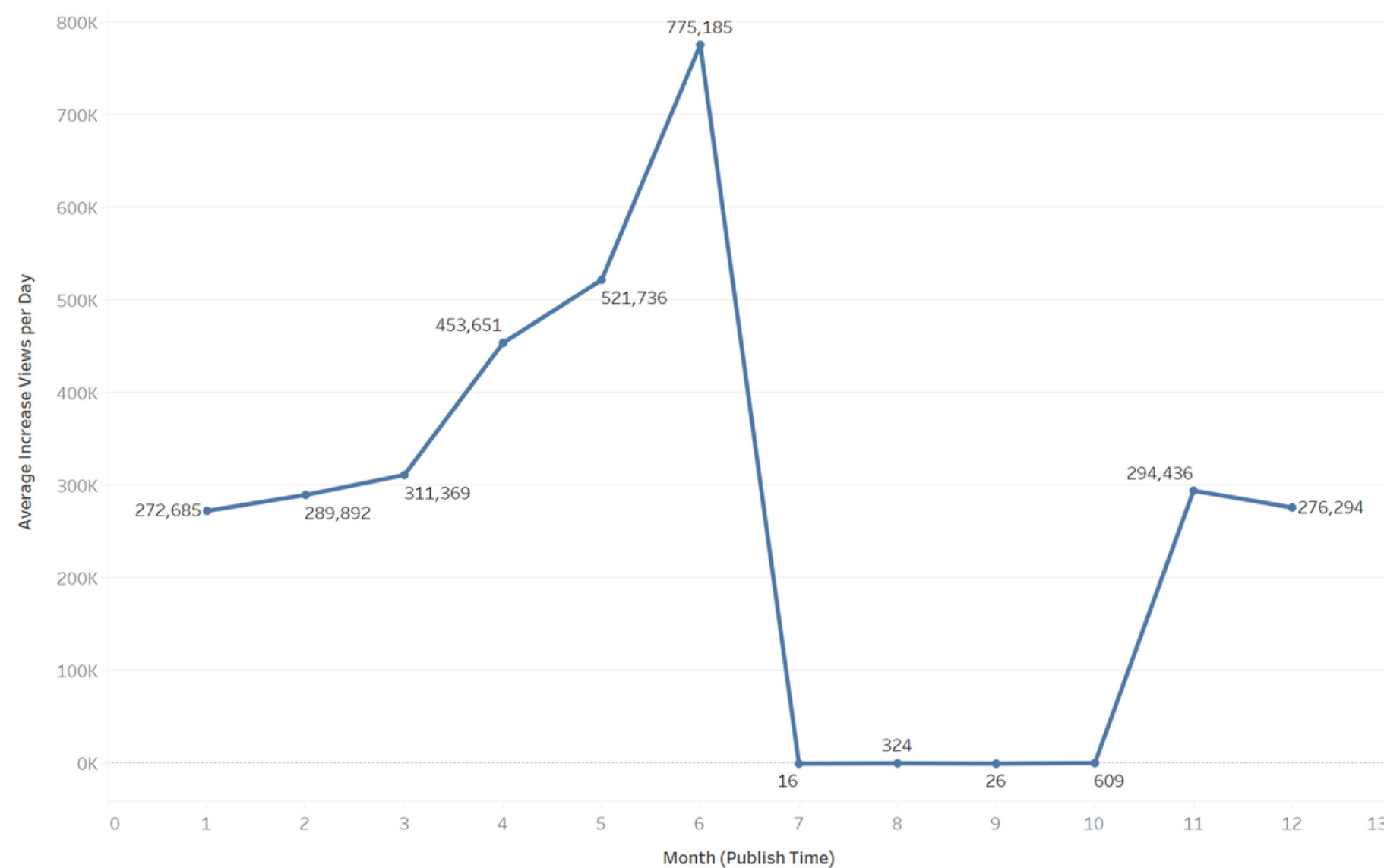
Comments Disabled

Rating Disabled

**Video Error or
Removed**

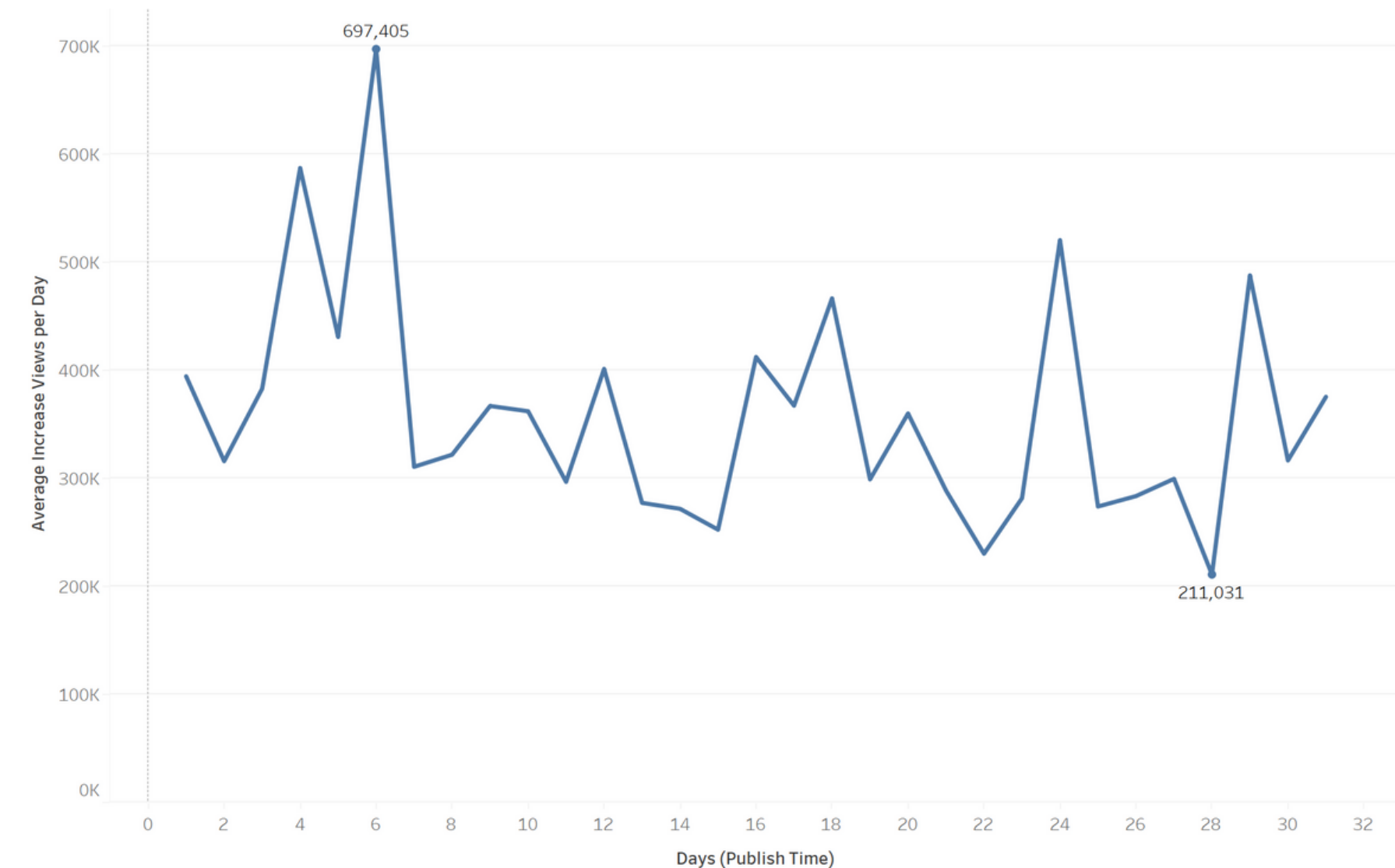
Description

02 EDA: Line chart on average daily & monthly growth views



The trend of average of Increase/Days for Publish Time Month.

Monthly

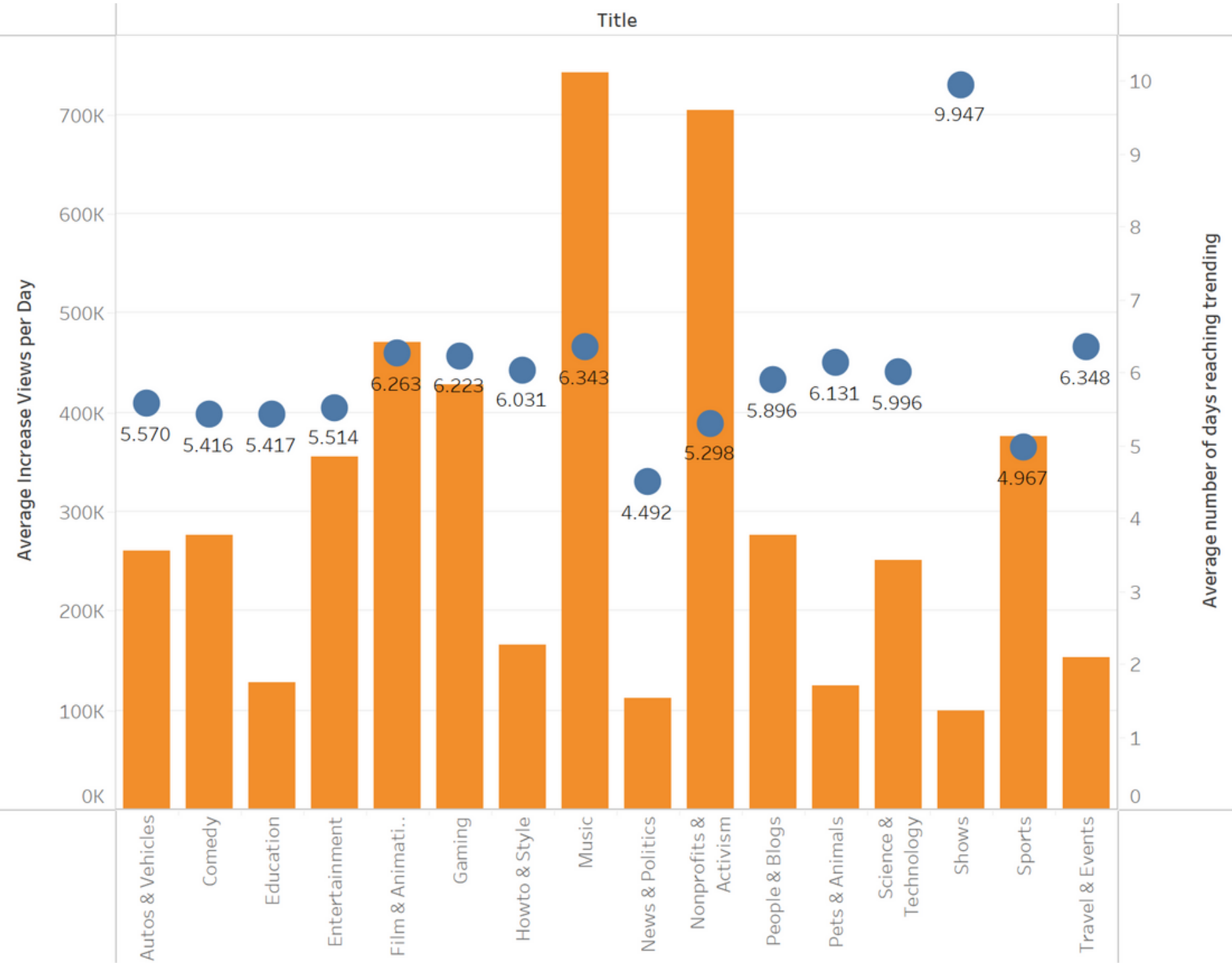


The trend of average of Increase/Days for Publish Time Day.

Daily

Q1

02 EDA: Bar charts and tables for video classification



Avg. Increase/Days and Median Increase Day Diff for each Title. Color shows details about Avg. Increase/Days and Median Increase Day Diff.

	category_id	views	title
0	1	7284156721	Film & Animation
1	2	520690717	Autos & Vehicles
2	10	40132892190	Music
3	15	764651989	Pets & Animals
4	17	4404456673	Sports
5	19	343557084	Travel & Events
6	20	2141218625	Gaming
7	22	4917191726	People & Blogs
8	23	5117426208	Comedy
9	24	20604388195	Entertainment
10	25	1473765704	News & Politics
11	26	4078545064	Howto & Style
12	27	1180629990	Education
13	28	3487756816	Science & Technology
14	29	168941392	Nonprofits & Activism
15	43	51501058	Shows

02 EDA: Strategy – focus on product relevance

Travel & Events need to be included because:

- Related to our product -> Travel App
- Total views not too high -> Cost may not high
- Not too low average daily traffic growth
- Fast trending speed of videos

In addition, we can also advertise other video categories.....

02 EDA: Strategy – focus on short-term

A Music + Nonprofit & Activism

Pros:

- Average daily viewing volume of increased significantly
- Medium number of days a video reaches trending is relatively low

Cons:

- The total viewing of music videos is the highest, the cost may be higher
- The total viewing of nonprofit videos is relatively low, may have an impact on the advertising effect

02 EDA: Strategy – focus on long-term

B Gaming + Entertainment /
Film & Animation+ Sport

Pros:

- Not low total views, but not too high, so costs can be controlled
- The average daily traffic growth and medium trending duration are both in the upper middle range

Cons:

- People who watching gaming and entertainment may not necessarily have a strong interest in tourism

03 Data Preprocessing: Check Fraud Data

What is fraud data?

Too few likes, dislikes, and comments in the same level of views

Why need to check fraud data?

- Malicious browsing can affect our judgment on video placement choices
- Disrupting the training results of models

How we check fraud data?

Discovering the linear relationship between views and the number of likes, dislikes, and comments through the model, and setting threshold values to achieve filtering

03 Data Preprocessing: Check Fraud Data

Two Model Select

Linear Regression

Decision Tree

Process:

- Train models separately based on video categories
- Calculate mse and compare sizes
- Among the 16 categories, **87.5%** have smaller mse under decision tree model

-> **Select Decision Tree to do the fraud check**

03 Data Preprocessing: Remove Fraud Data

Threshold = 4:

Views corresponding to likes, dislikes, and comments of the same level are normal between two to three times, so select 4 as the threshold for filtering

Process:

- Use the corresponding model for preview for each video category
- Compare the preview results with the original views and delete those that are less than four times the size

-> Remove 210 rows of the fraud data

03 Data Preprocessing: Predict Future Views Trends

Linear Regression:

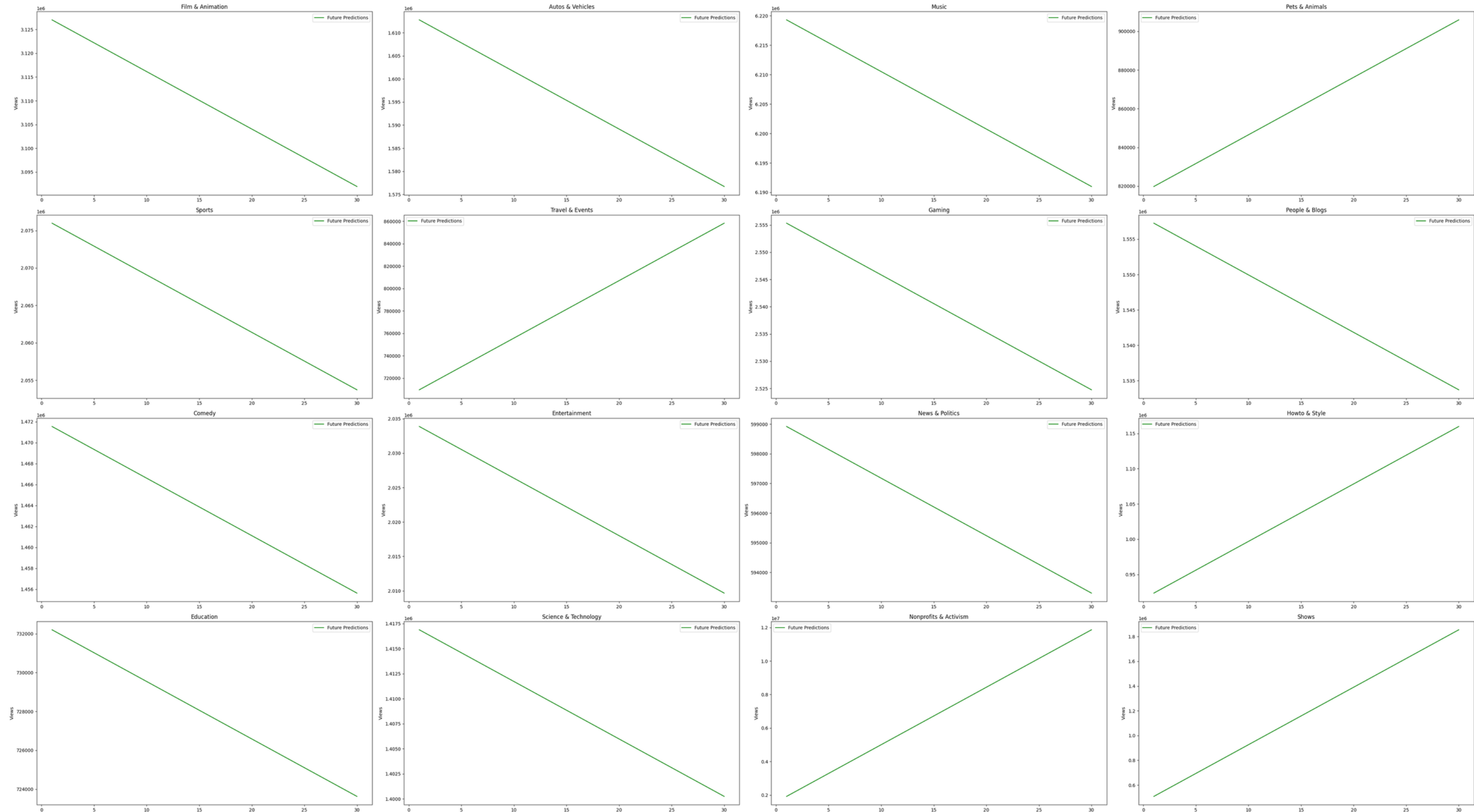
Train linear regression with views based on the difference in days between trending date and publishing date

ARIMA:

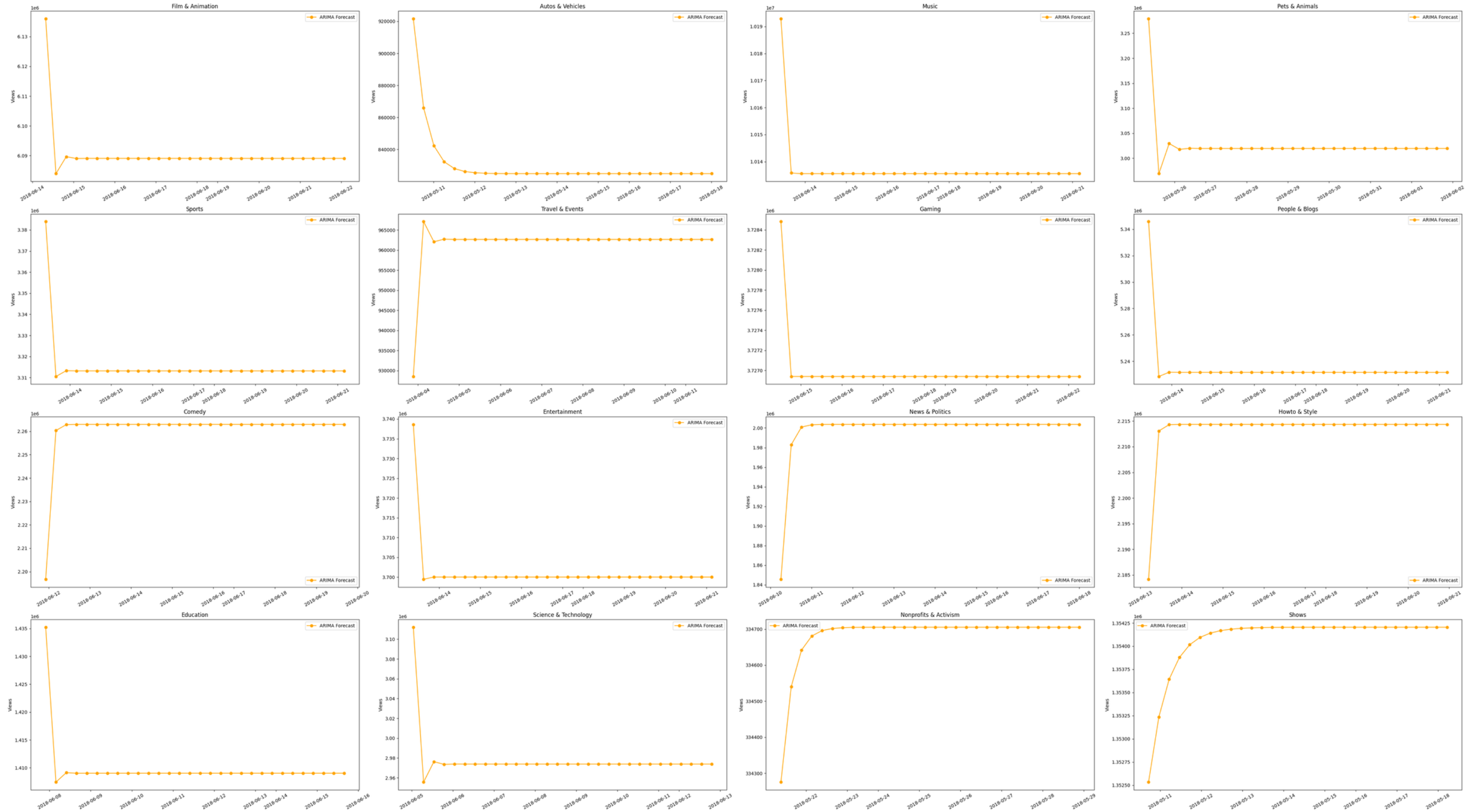
Train ARIMA models with trending date and publishing date and views respectively

Both models predict views trends for the next month

03 Data Preprocessing: Linear Regression



03 Data Preprocessing: ARIMA – Publish Date



03 Data Preprocessing: ARIMA – Trending Date

