Home Page drop rate problem

SHAHAR VAZANA

Table of Contents:

Problem Statement

The Problem: A major issue for the site- a very high drop-off rate on the home page, resulting in users generating little to no revenue while incurring high acquisition costs.

The Data: An analysis of the average number of video views over time shows a downward trend, which may indicate a declining interest from users in the videos presented to them. This highlights the need for better matching and optimization of the videos displayed.

The Solution: Implementing automatic promotion of recommended videos in a dedicated section on the home page, thereby increasing the likelihood that users will engage with videos instead of abandoning the site. This, in turn, can help boost ad revenue generated from video ads views.
To support this, there is a need to classify videos based on their likelihood of being watched.

Product Development Stages

1. Data collection and understanding
2. Metric calculation
3. Definition of ranking criteria
4. Classification of video categories
5. Analytical and statistical analysis
6. Conclusions and business recommendations

Data collection and understanding

Attached HTML: Primary analysis- home assigment- playstudios- shahar Vazana

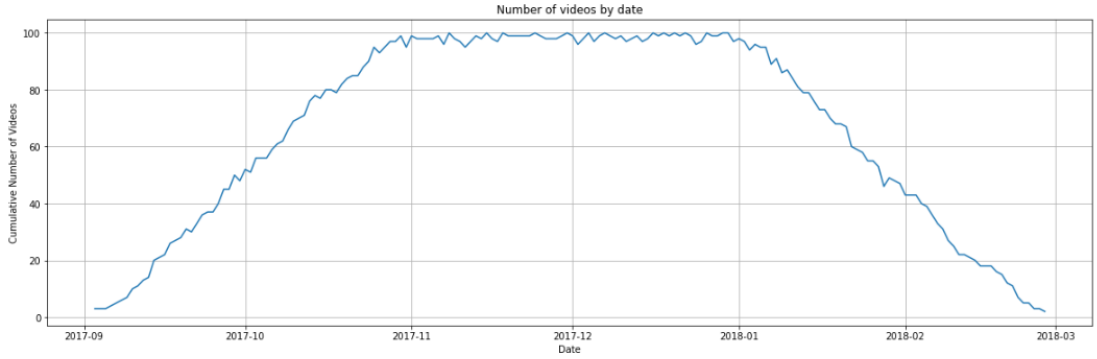
Daily views were sampled for 100 videos, each tracked over a period of 120 days from the date it was uploaded to the site. For each video, additional data was collected regarding its length (in seconds, ranging from 15 to 30), language (Chinese, English, or Spanish), upload date, and quality (ranging from 240p to 1080p).

The numerical analysis shows that the mean and median total views per video are almost identical (approximately 9,000 views), indicating a relatively symmetrical

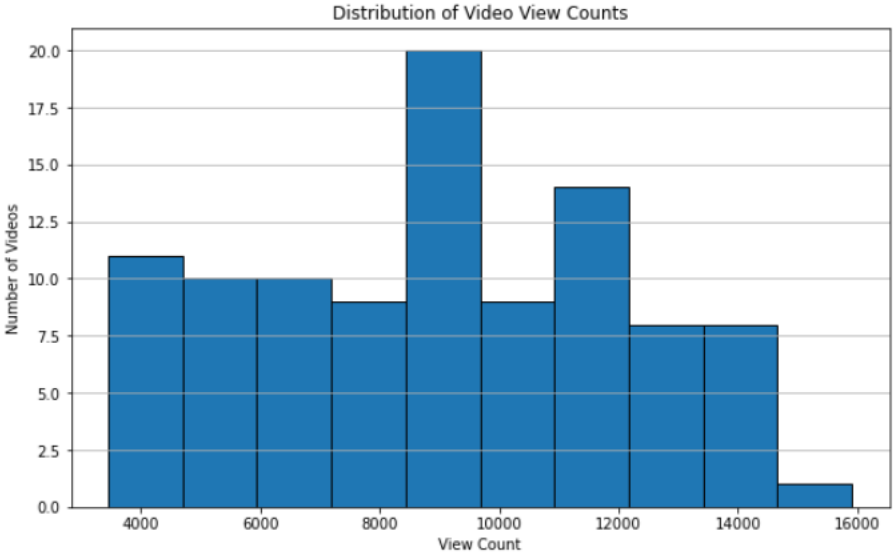
distribution without significant outliers. However, the relatively high standard deviation (over 3,000 views) suggests considerable dispersion around the mean.

```
Total views: 897815
Avg video views: 8978.15
std video views: 3084.15227
median video views: 8925.0
```

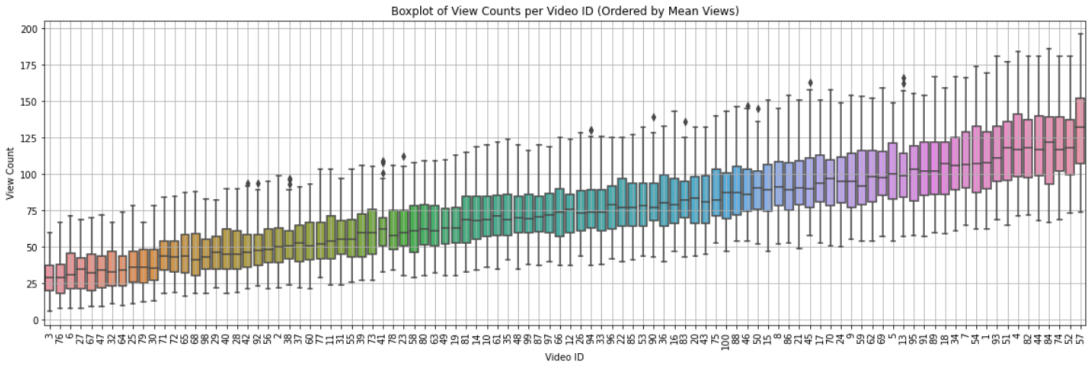
An examination of the data from the video_count table also reveals a certain overlap between the measurement periods of the videos. This can be observed in the graph, where the number of active videos reaches 100 and remains at that level for a certain period of time.



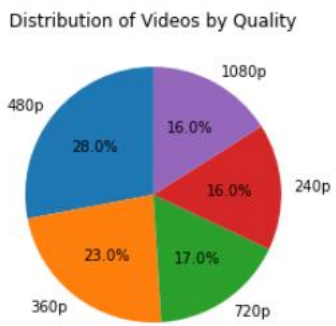
An examination of the overall distribution of views per video shows that most videos receive a similar average number of views. However, there is a small group of videos that stand out significantly, indicating the potential to identify characteristics that differentiate particularly successful videos from the rest.
Additionally, it shows that the majority of the videos cluster around the mean and median range.

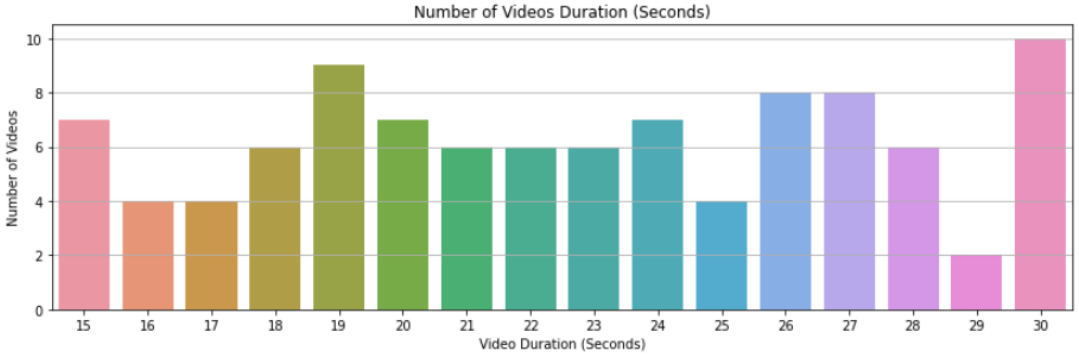


The differences between the videos are also evident through the box plot. It is notable that as the average number of views for a video increases, the variability in daily views also increases. Videos with a lower average number of views tend to be relatively stable, whereas more successful videos exhibit greater variance in daily views.



Let's examine the data from the video_features table and review the number of videos for each column:





A review of the data shows that the majority of videos uploaded to the site tend to have durations of 15, 19, 20, 24, 26, 27, or 30 seconds, without a clearly discernible trend in video length distribution.
The distribution of videos across different languages appears relatively balanced, with no dominant language category.
Furthermore, the highest proportion of uploaded videos are in 480p quality, followed by 360p.

Assumptions Regarding the Data:

Missing Records:

- Videos with fewer than 120 daily records are assumed to have days with zero views. If the missing records are due to measurement errors or missing values, efforts should be made to complete the data where possible and investigate the cause of the discrepancies.
- The average daily views will always be calculated over 120 days, not based on the actual number of recorded entries. This assumption enables consistent comparisons between videos, avoiding biases due to missing data.

Data Limitations:

- Site Traffic: It is assumed that daily site traffic is relatively stable, with a consistent number of users exposed to all the videos each day.
- Language: As there is no data regarding users' language preferences, it will not be possible to assess whether matching the video language to the user's language impacts video views. It will further be assumed that in today's globalized world, universal languages (such as English) are generally accessible to users who speak also other languages.

Sample Limitations:

- The current dataset is small and limited. It is assumed that this limitation may impact the sensitivity of statistical analyses and subsequent decision-making.
- It is anticipated that as more data accumulates over time, model quality and the ability to detect various effects will improve significantly.
- After launching the product and collecting additional real-time and historical data, it will be possible to overcome this limitation and perform more analyses, leading to more accurate conclusions.

Recommendation System Development:

- Initial Stage: Create a preliminary ranking based on historical data to classify videos based on their likelihood of being watched.
- Future Goal: Develop an intelligent system capable of identifying emerging trends in new videos in real time (real-time recommendation), also, the existing criteria will be expanded and statistically re-analyzed to ensure improved model accuracy for old and new top-performing videos.

Metric calculation

SQL tables and calculating relevant fields

Attached file: Home assigment-Playstudios-Shahar Vazana-SQLQuery

Attached Excel: fact_video_count

fact_video_count – Unified FACT Table (Including Calculated Fields at Video-Level Granularity and Daily View Counts):

- All fields from the video_count and video_features tables are joined based on video_id.
- days_from_video_upload: Number of days elapsed from the video's upload date to the view count measurement date. In live data, the goal is to promote strong-performing videos that are still early in their lifecycle.
- daily_delta: The daily change in views for each video.
- max_video_count: The maximum number of views recorded for each video.
- max_count_date: The date on which the maximum number of views was recorded for each video.
- min_video_count: The minimum number of views recorded for each video.
- min_count_date: The date on which the minimum number of views was recorded for each video.

Attached Excel: summary_video_count

summary_video_count – Summary Table Based on the FACT Table, Aggregated by video_id at the Video Granularity Level:

- video_id: Video identifier.
- total_views: Total number of views.
- total_rows: Number of records per video – some videos have fewer than 120 records.
- days_since_upload_to_last_views: Number of days from the upload date to the date of the last sampled views (120 for all videos).
- avg_views_per_day: Average number of views per day (assuming division by 120 days, including days with zero views).
- video_upload_date: The video's upload date.
- last_count_date: The last date on which views were recorded for each video.
- stdev_video: Standard deviation of daily views for each video.
- max_video_count: Maximum number of views recorded for each video.
- max_count_date: The date when the maximum number of views was recorded.

- is_weekend_max_count: Whether the maximum view count occurred on a weekend (0–1).
- days_since_upload_to_max_views_date: Number of days from the upload date to the maximum view date.
- min_video_count: Minimum number of views recorded for each video.
- min_count_date: The date when the minimum number of views was recorded.
- is_weekend_min_count: Whether the minimum view count occurred on a weekend (0–1).
- days_since_upload_to_min_views_date: Number of days from the upload date to the minimum view date.
- upload_weekday: Day of the week the video was uploaded (1–7).
- is_weekend_upload: Whether the video was uploaded on a weekend (0–1).
- avg_daily_delta: Average daily change in views for each video.
- trend_slope: Linear trend line slope of daily views for each video — the slope helps understand the general direction and strength of the view trend for each video.

Attached Excel: joined_summary_video_count_and_features

joined_summary_video_count_and_features – A Unified Table Merging summary_video_count and video_features.

normalized_table – CTE Table Performing Normalization of Values (Between 0 and 1) for All Fields Used in Video Ranking:

- norm_avg_views_per_day
- norm_stdev_video – No inversion was applied during calculation, as the analysis showed a positive correlation between the standard deviation and the average number of views.
- norm_trend_slope

scored_table_weighted – CTE Table Calculating the Desired Ranking Score for Each Video:

- video_id
- total_final_score – A weighted average calculation of the criteria used for ranking.
- video_weighted_score_rank – Ranking of videos based on the weighted average score.

Attached Excel: final_scored_summary_video_count_and_features

final_scored_summary_video_count_and_features – A Unified Table Combining scored_table_weighted and joined_summary_video_count_and_features:

- All fields from both tables are included.

- video_category: Assignment of the appropriate category to each video based on its ranking.

Attached Excel: video_category

- **video_category** – A Table Displaying the Final Output for Each Video and Its Assigned Category.

Definition of ranking criteria

- The ranking of videos will be based on three primary metrics:
  The combined ranking approach allows consideration not only of the total number of views but also of trending videos that are currently gaining momentum.

1. **50% Daily Popularity** – As high an average daily view count as possible.
2. **30% Stability** – Standard deviation, weighted inversely. A higher standard deviation increases the likelihood that the video is a “Top” performer. The assumption is that videos with higher volatility are likely to be "trending" and should be promoted. This assumption is supported by the graphical analysis conducted.
3. **20% Trend** – The slope of a linear trend line. Videos with a positive upward trend will be prioritized for promotion, while videos showing a sharp decline will be deprioritized.

Three categories have been defined based on each video's final score:

- **Top:** Top 10%
- **Good:** 50%–89%
- **Everything Else:** 0%–49%

Classification of video categories:

video_id	video_score_rank	video_category	video_upload_date	upload_weekday	video_language	video_duration_seconds	video_quality	total_views	avg_views_per_day	stdev_video	trend_slope	max_video_count	is_weekend_max_count	days_sience_upload_to_max_views_date
57	1	Top	2017-09-25	2	spanish	15	1080p	15909	132.57	29.39	-0.41	196	1	12
44	2	Top	2017-09-17	1	chineese	17	1080p	14409	120.08	27.79	-0.37	181	0	12
52	3	Top	2017-10-22	1	spanish	17	1080p	14225	118.54	26.04	-0.32	181	1	20
82	4	Top	2017-09-30	7	english	15	720p	14328	119.4	25.88	-0.33	181	1	0
84	5	Top	2017-10-06	6	english	15	720p	14183	118.19	29.87	-0.43	186	0	7
74	5	Top	2017-10-25	4	chineese	15	720p	14066	117.22	25.66	-0.31	181	0	0
51	7	Top	2017-09-19	3	chineese	18	1080p	13811	115.09	27.91	-0.37	177	0	8
93	8	Top	2017-09-10	1	english	18	1080p	13732	114.43	25.93	-0.32	181	0	3
4	9	Top	2017-10-12	5	spanish	15	720p	14192	118.27	26.88	-0.41	184	0	5
1	10	Top	2017-09-11	2	chineese	16	480p	13197	109.97	26.44	-0.41	169	0	0
89	11	Good	2017-09-26	3	english	20	1080p	12609	105.08	25.36	-0.38	167	0	0
7	12	Good	2017-10-11	4	spanish	15	480p	13156	109.63	25.37	-0.43	166	0	1
34	13	Good	2017-10-11	4	chineese	17	720p	12723	106.03	24.71	-0.39	167	1	4
13	14	Good	2017-09-03	1	english	18	720p	12033	100.28	23.96	-0.33	166	0	3
59	15	Good	2017-10-06	6	spanish	16	240p	11533	96.11	24.42	-0.32	153	0	5
5	16	Good	2017-09-14	5	chineese	19	720p	12083	100.69	23.71	-0.34	149	0	6
69	17	Good	2017-09-14	5	chineese	16	360p	12152	101.27	22.21	-0.32	159	1	17
70	18	Good	2017-09-03	1	english	18	480p	11401	95.01	22.37	-0.32	158	0	3
54	19	Good	2017-09-25	2	english	15	480p	12645	105.38	28.83	-0.57	174	1	6
62	20	Good	2017-10-22	1	spanish	19	720p	11955	99.63	21.82	-0.35	152	0	2

Analytical and statistical analysis

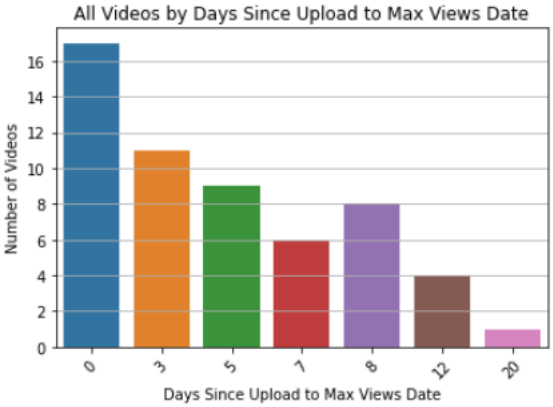
Analytical analysis:

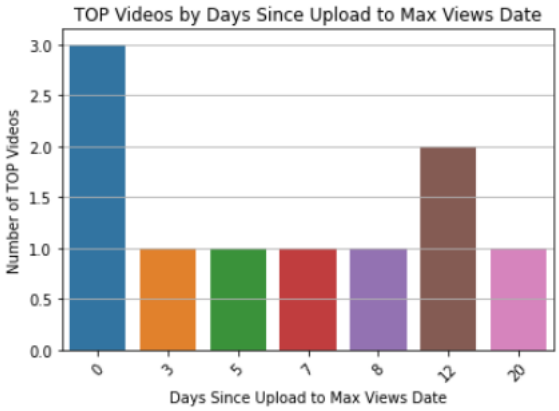
Attached Excel: video category output

Attached HTML: Categories output analysis- home assigment- playstudios- shahar vazana

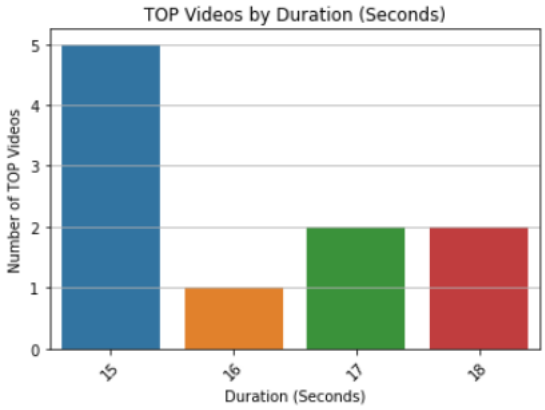
Let's review the data both graphically and numerically:

- **trend_slope:** Moderately negative (close to zero) across all videos. This is a reasonable finding, given that the sampling period is relatively long, making it likely for videos to show an overall declining trend over time.
- **days_since_upload_to_max_views_date:** The data shows that all videos reach their maximum number of views within the first 20 days, with most achieving it within the first 5 days.
  If we focus only on the videos categorized as **Top**, we see that while there is no clear linear trend, it is noticeable that slightly more videos reached their maximum within 12 days or even on the upload day itself.



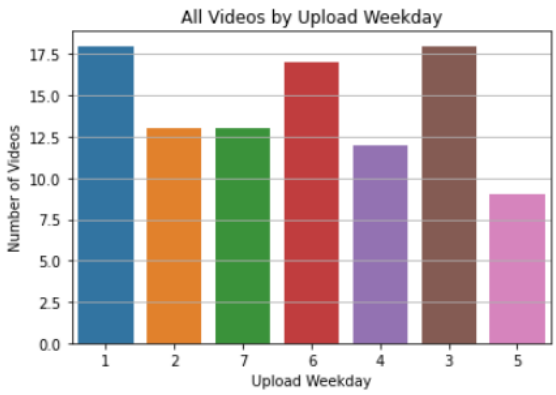


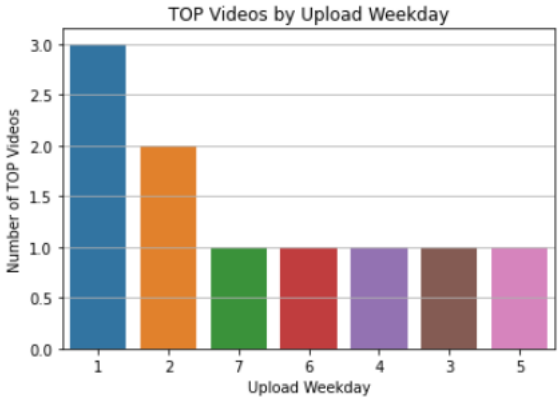
- **stdev_video, avg_views_per_day, total_views, max_video_count:** The values appear in descending order according to the ranking, meaning the highest values are observed in the Top category.
- **video_duration_second:** For videos classified in the Top category, it is evident that they all fall within short duration ranges of 15–18 seconds, with half of them having a duration of exactly 15 seconds.



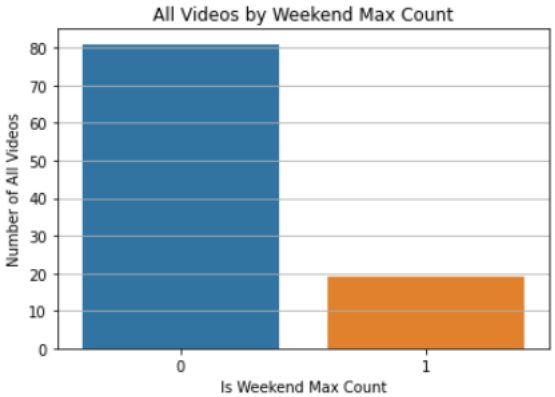
- **video_language:** According to the graph, there is slightly more variability in the distribution of languages among the Top videos compared to the overall distribution observed across all videos.

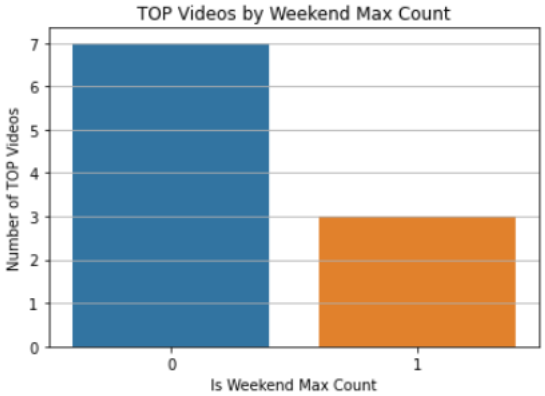
- **upload_weekday:** Based on the graphs, most of the videos categorized as Top were uploaded primarily on Sunday, and to a lesser extent on Monday.
  In the overall graph, however, Sundays and Mondays do not necessarily account for the majority of uploads, and there is no significant difference in the distribution of upload days across the week.
  This observation warrants further statistical examination.





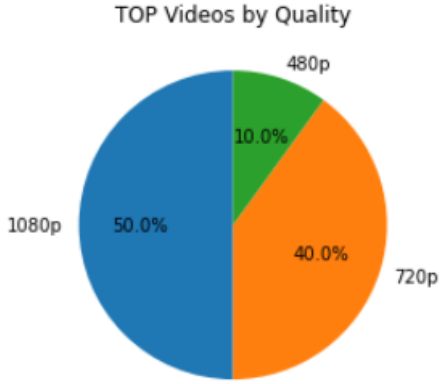
- **is_weekend_max_count:** Based on the graphs, it can be observed that, similar to the overall dataset, the Top videos also reached their peak view counts during weekdays rather than on weekends.



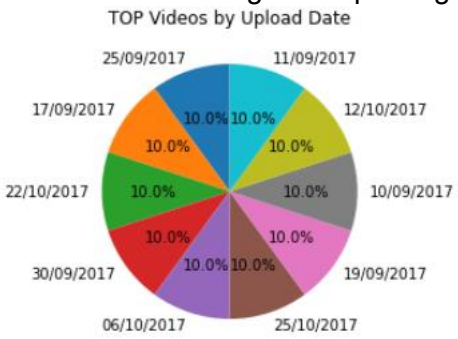


- **video_quality:**

For videos classified in the Top category, it is evident that all of them fall within high-quality ranges, with the majority having a resolution between 720p and 1080p.



- **video_upload_date:** All values are evenly distributed — no specific trend was observed indicating that certain upload dates influence a video's likelihood of reaching the Top category.



Based on the data review, several common characteristics have been identified that should be further examined statistically and considered for incorporation into the video ranking and classification model:

- High total number of views
- High average daily views
- High standard deviation
- High maximum daily view count
- Very high video quality
- Very short video durations
- Slightly negative trend slope
- Peak views achieved during weekdays
- Fewer than 20 days elapsed between the video upload date and its peak view count

statistical analysis

Attached HTML: Statistical tests- home assigment- playstudios- shahar vazana

Running statistical tests to examine the relationship between the following variables and the video category (Top / Good / Everything Else) as a nominal categorical variable:

1. **Video Duration** (video_duration_seconds)
2. **Video Language** (video_language)
3. **Day of the Week of Upload** (upload_weekday)
4. **Days from Upload to Peak Views** (days_since_upload_to_max_views_date)
5. **Whether the Peak Occurred on a Weekend** (is_weekend_max_count)
6. **Video Quality** (video_quality)
7. **Standard Deviation of Daily Views** (stdev_video)
8. **Video Upload Date** (video_upload_date) – to be tested in the future when real-time data is available
9. **Days Since Upload to Last Views** (days_since_upload_to_last_views) – to be tested in the future when real-time data is available

Statistical Tests and Conclusions:

1. Video Duration (video_duration_seconds)

- Variable Type: Continuous
- Test Type: ANOVA / Kruskal-Wallis (if data normality assumptions are not met)
- Test Objective: To determine whether there is a difference in video duration across the categories.
- Hypotheses:
    - H_0: There is no difference in video duration between the categories.
    - H_1: There is a difference in video duration between the categories.
- Interpretation:
    - If p-value < 0.05, reject H_0 – indicating a significant relationship.
- Output:

```
Normality test for Everything else: p-value = 0.009650609456002712
Normality test for Good: p-value = 0.30525296926498413
Normality test for Top: p-value = 0.009824455715715885
KruskalResult(statistic=71.81890948481205, pvalue=2.5393464518435617e-16)
```

We applied the Shapiro test separately for each category:

- If p-value < 0.05, we reject the H_0 hypothesis of normality, indicating that the data is not normally distributed.

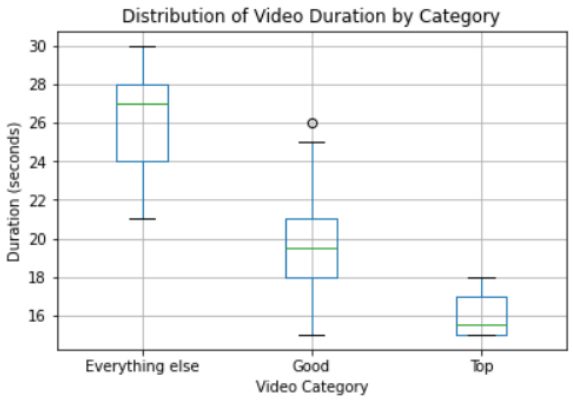
- Everything Else: p-value = 0.0097
- Good: p-value = 0.3052
- Top: p-value = 0.0098

The results show that only the Good category is normally distributed. Therefore, the Kruskal-Wallis test was used instead of ANOVA.

Kruskal-Wallis Test Result:

- p-value = 2.54e-16 (significantly less than 0.05)

Conclusion:
We reject the H_0 hypothesis and conclude that there are statistically significant differences in video duration across the different categories.



The graph clearly shows significant differences that align with the results of the statistical test. It appears that there may be a relationship between video duration (the shorter the video) and the category to which it belongs.

2. Language (video_language)

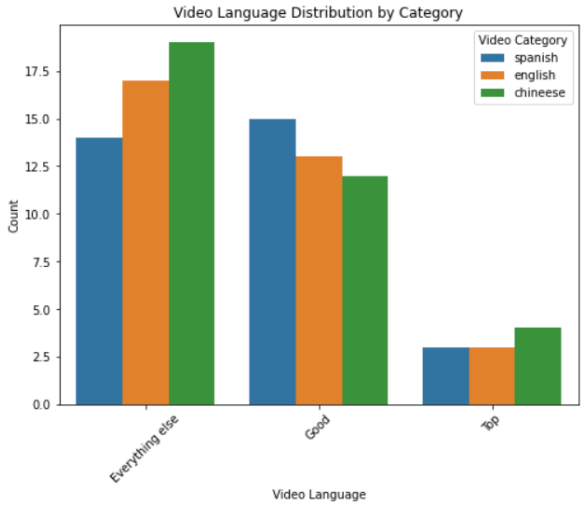
- Variable Type: Nominal
- Test Type: Chi-Square Test of Independence
- Test Objective: To determine whether there is an association between the video's language and its category.
- Hypotheses:
  - H_0: There is no relationship between video language and category.
  - H_1: There is a relationship between video language and category.
- Interpretation:
  - If p-value > 0.05, we fail to reject H_0 – indicating no significant relationship.
- Output:

```
Chi-Square test for language: p-value = 0.8827203870683591
```

Chi-Square Test Result:

- p-value = 0.882, which is greater than 0.05.

Conclusion:
We won't reject the H_0 hypothesis and conclude that there is no statistically significant relationship between the video's language and its category.



The graph shows no major differences or significant variations between the video's language and its assigned category.

3. Day of the Week (upload_weekday)

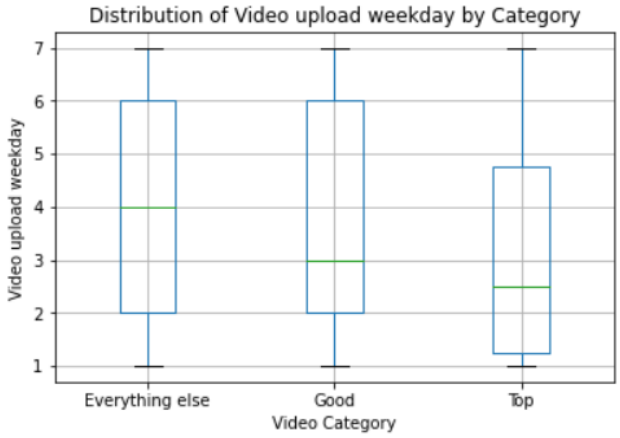
- Variable Type: Nominal
- Test Type: Chi-Square Test of Independence
- Test Objective: To determine whether there is an association between the day of the week a video was uploaded and its category.
- Hypotheses:
    - H_0: There is no relationship between the upload day and the video category.
    - H_1: There is a relationship between the upload day and the video category.
- Interpretation:
    - If p-value > 0.05, we fail to reject H_0 – indicating no significant relationship.
- Output:

```
Chi-Square test for upload weekday: p-value = 0.8727472311039215
```

Chi-Square Test Result:

- p-value = 0.872, which is greater than 0.05.

Conclusion:
We won't reject the H_0 hypothesis and conclude that there is no statistically significant relationship between the day of the week a video was uploaded and its category.



The graph also shows that there are no noticeable differences between the groups.

4. Days from Upload to Peak Views (days_since_upload_to_max_views_date)

- Variable Type: Continuous
- Test Type: ANOVA / Kruskal-Wallis (if normality assumptions are not met)
- Test Objective: To determine whether there is a difference in the number of days from video upload to peak views across categories.
- Hypotheses:
    - H_0: There is no difference in the number of days from upload to peak views between categories.
    - H_1: There is a difference in the number of days from upload to peak views between categories.
- Interpretation:
    - If p-value < 0.05, reject H_0 – indicating a significant relationship.
- Output:

```
Normality test for Everything else: p-value = 0.004321173299103975
Normality test for Good: p-value = 0.0012734165647998452
Normality test for Top: p-value = 0.24567563831806183
KruskalResult(statistic=1.536927605043536, pvalue=0.46372489443191545)
```

We performed the Shapiro test separately for each category:

- If p-value < 0.05, we reject the H_0 hypothesis of normality, indicating that the data is not normally distributed.
    - Everything Else: p-value = 0.0043
    - Good: p-value = 0.0012

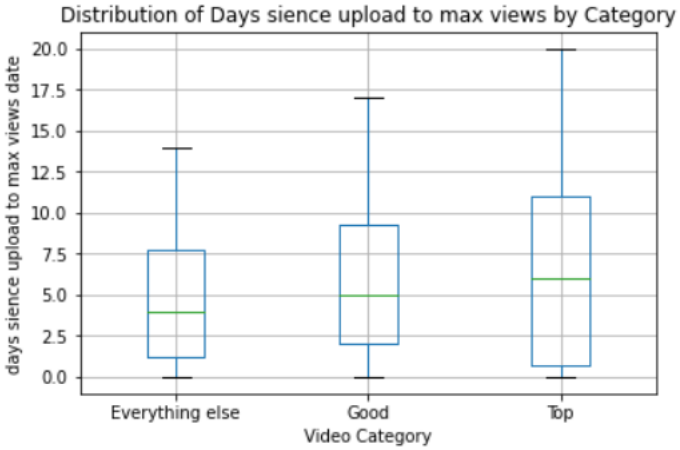
- Top: p-value = 0.2456

The results show that only the Top category is normally distributed. Therefore, the Kruskal-Wallis test was used instead of ANOVA.

Kruskal-Wallis Test Result:

- p-value = 0.463 (greater than 0.05)

Conclusion:
We won't reject the H_0 hypothesis and conclude that there are no statistically significant differences in the number of days from upload to peak views between the different video categories.



The graph also shows that there are no noticeable differences between the groups.

5. Whether Peak Views Occurred on a Weekend (is_weekend_max_count)

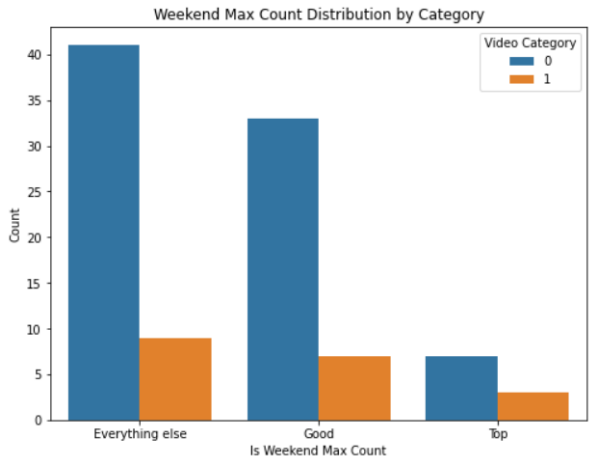
- Variable Type: Binary
- Test Type: Chi-Square Test of Independence
- Test Objective: To determine whether there is an association between whether peak views occurred on a weekend and the video category.
- Hypotheses:
  - H_0: There is no relationship between peak views occurring on a weekend and the video category.
  - H_1: There is a relationship between peak views occurring on a weekend and the video category.
- Interpretation:
  - If p-value > 0.05, we fail to reject H_0 – indicating no significant relationship.
- Output:

```
Chi-Square test for weekend max: p-value = 0.6449409664697823
```

Chi-Square Test Result:

- p-value = 0.645, which is greater than 0.05.

Conclusion:
We fail to reject the H_0 hypothesis and conclude that there is no statistically significant relationship between whether a video's peak views occurred on a weekend and its category.



An examination of the timing of peak views shows that the majority of videos, across all categories, reached their peak view counts on weekdays rather than on weekends.
This finding may suggest viewing patterns where higher activity occurs during the workweek.
However, the statistical test indicated that there is no significant difference between video categories in terms of the timing of peak views.
Therefore, it cannot be concluded that the timing of the peak is a determining factor in a video's success.
The graph points to a potential trend that could be revisited in the future, once additional data becomes available, and with the use of more advanced testing methods such as A/B testing.
At this stage, however, no statistically significant relationship can be established.

6. Video Quality (video_quality)

- Variable Type: Ordinal
- Test Type: ANOVA / Kruskal-Wallis (if normality assumptions are not met)
- Test Objective: To determine whether there are differences in video quality across the different categories.
- Hypotheses:
    - H_0: There is no difference in video quality between the categories.
    - H_1: There is a difference in video quality between the categories.
- Interpretation:
    - If p-value < 0.05, reject H_0 – indicating a significant relationship.

- Output:

```
Normality test for Everything else: p-value = 0.0002096537355100736
Normality test for Good: p-value = 0.0031001686584204435
Normality test for Top: p-value = 0.008488975465297699
KruskalResult(statistic=18.712944201994997, pvalue=8.640438760334742e-05)
```

We conducted the Shapiro test separately for each category:

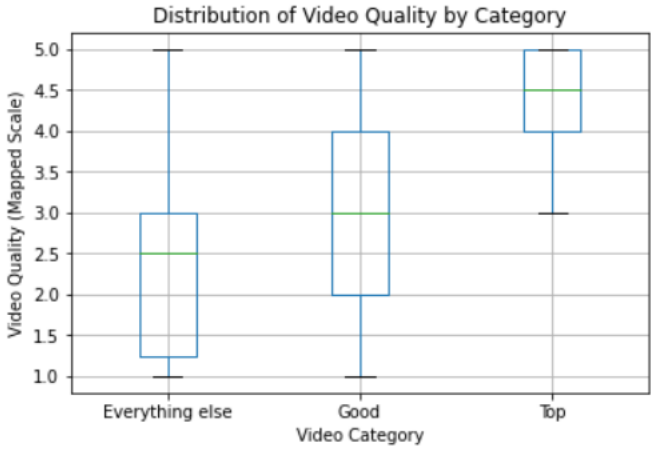
- If p-value < 0.05, we reject the H_0 hypothesis of normality, meaning the data is not normally distributed.
    - Everything Else: p-value = 0.0002
    - Good: p-value = 0.0031
    - Top: p-value = 0.0084

The results indicate that none of the categories are normally distributed. Therefore, the Kruskal-Wallis test was used instead of ANOVA.

Kruskal-Wallis Test Result:

- p-value = 8.64e-05 (less than 0.05)

Conclusion:
We reject the H_0 hypothesis and conclude that there are statistically significant differences in video quality across the different categories.



The graph also supports the findings from the statistical test, showing a clear difference between video quality (with a score of 5 corresponding to 1080p and 1 to 240p) and the category to which the video belongs.

7. Standard Deviation of Views (stdev_video)

- Variable Type: Continuous
- Test Type: ANOVA / Kruskal-Wallis (if normality assumptions are not met)
- Test Objective: To determine whether there are differences in the standard deviation of video views across categories.

- Hypotheses:
    - H_0: There is no difference in standard deviation between the categories.
    - H_1: There is a difference in standard deviation between the categories.
- Interpretation:
    - If p-value < 0.05, reject H_0 – indicating a significant relationship.
- Output:

```
Normality test for Everything else: p-value = 0.5190340876579285
Normality test for Good: p-value = 0.08126771450042725
Normality test for Top: p-value = 0.10130418092012405
KruskalResult(statistic=74.95734293146586, pvalue=5.287127618683767e-17)
```

We performed the Shapiro test separately for each category:

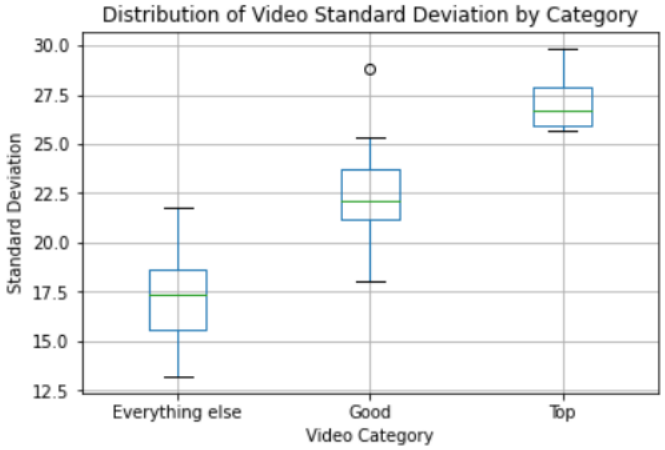
- If p-value < 0.05, we reject the H_0 hypothesis of normality, indicating that the data is not normally distributed.
    - Everything Else: p-value = 0.519
    - Good: p-value = 0.081
    - Top: p-value = 0.101

The results show that the Good category does not meet the normality assumption. Therefore, the Kruskal-Wallis test was used instead of ANOVA.

Kruskal-Wallis Test Result:

- p-value = 5.287e-17 (significantly less than 0.05)

Conclusion:
We reject the H_0 hypothesis and conclude that there are statistically significant differences in the standard deviation of video views across the different categories.



The graph also supports the findings from the statistical test, showing a clear difference between the standard deviation and the video's category.
A higher standard deviation is associated with the Top category.

Conclusions

1. **Video Duration (video_duration_seconds)**
   A statistically significant difference was found between video duration and category. Shorter videos tend to fall into higher success categories.
   Recommendation: Incorporate this variable into the video classification ranking model.
2. **Video Language (video_language)**
   No significant relationship was found between video language and category.
   Recommendation: Unless future analyses show otherwise, do not include this variable in the ranking model.
3. **Day of the Week of Upload (upload_weekday)**
   Both the analytical and statistical analyses suggest that the upload day does not significantly impact a video's category.
   Recommendation: No need to incorporate this variable into the model at this stage.
4. **Days from Upload to Peak Views (days_since_upload_to_max_views_date)**
   No significant relationship was found between the speed of reaching peak views and the video category. However, it is noteworthy that all videos reached their peak within 20 days.
   Recommendation: In real-time, prioritize and give weight to videos within their first month, and particularly their first day, of being uploaded, using the days_since_upload_to_last_views field. Even videos with limited initial data but a positive trend may have a high potential for increased views.
5. **Whether Peak Occurred on a Weekend (is_weekend_max_count)**
   No significant relationship was found between whether peak views occurred on a weekend and the video's category.
   Recommendation: Since the graph showed that most peak views occur on weekdays, consider promoting the recommended videos section more heavily during weekdays. Reassess this pattern as more data becomes available.
6. **Video Quality (video_quality)**
   A statistically significant difference was found between video quality and category. Higher-quality videos are associated with the **Top** category.
   Recommendation: Strongly prioritize this variable in the ranking calculation.
7. **Standard Deviation (stdev_video)**
   A statistically significant difference in view variability was found between categories. Videos in the Top category exhibit higher variability in their daily views.
   Recommendation: Incorporate this variable into the ranking model.
8. **Video Upload Date (video_upload_date)**
   Although initial analysis did not show strong trends, it is recommended to re-examine this variable when more longitudinal and real-time data is available, to potentially identify seasonal patterns or timing effects.
9. **Trend Slope (trend_slope)**
   Videos with a positive trend (e.g., in the first 20 days post-upload) should be given priority in the recommended videos ranking. Conversely, videos with a strong negative slope could be deprioritized.

Business Recommendations

- **Refine the Ranking Model:** Improve ranking precision by incorporating the statistically validated variables and re-assess as more historical and real-time data accumulates.
- **Conduct In-Depth A/B Testing:** To validate causal relationships and not just statistical correlations, eliminating the influence of confounding variables.
- **Content Strategy:**
  Given that many of the currently uploaded videos are longer and of medium-to-low quality, encourage the upload of more short videos (preferably 15 seconds) and prioritize high-quality uploads whenever possible.
- **User and Traffic Data Utilization:**
  If and when traffic and user data becomes available, focus marketing efforts more on weekdays, and consider optimizing the recommended videos by personalizing them based on user characteristics (e.g., language, location, past video behavior).
- **Expand Video Metadata:**
  Add information about video style and content (e.g., comedy, informational, music, sports) to deepen the analysis and improve the precision of video recommendations.

Annexes

Attached file: Home assigment-Playstudios-Shahar Vazana-SQLQuery

Attached HTML: Primary analysis- home assigment- playstudios- shahar Vazana

Attached Excel: fact_video_count

Attached Excel: summary_video_count

Attached Excel: joined_summary_video_count_and_features

Attached Excel: final_scored_summary_video_count_and_features

Attached Excel: video_category

Attached Excel: video category output

Attached HTML: Categories output analysis- home assigment- playstudios- shahar vazana

Attached HTML: Statistical tests- home assigment- playstudios- shahar vazana