EXPLORATORY DATA ANALYSIS FOR SAASQUATCH VIDEOS

COMPARING USER ENGAGEMENT ACROSS BRANDS & ANALYZING DARE NETWORK'S SUCCESS FACTORS

Tristan Savella

PROJECT OVERVIEW

Project Premise

- You're an analyst for "Saasquatch Videos", a B2C, subscription-based video streaming service
- Goal: Conduct an exploratory analysis of the data and present initial business insights to a chosen business question

About the Company

- Saasquatch Videos hosts three brands:
  - Dare Network: for Adult Viewers
  - Playground Prime: for Children Viewers
  - Mosaic: for General Viewers of All Ages
- Dare Network has outperformed the other two brands across various metrics, most notably in:
  - User Engagement
  - Trial to Subscription Rates
- Mosaic is likely a new, recently launched brand:

PROJECT OVERVIEW

Chosen Business Questions:

- Why has Dare Network been outperforming the other two brands (especially Playground Prime)?

- Can this success be replicated in the other two brands?

- What strategies can Saasquatch Video's teams use to drive up numbers in the other brands?

PROJECT OVERVIEW

Dataset

- **Table 1: Leads**
  - Records of people signing up for email campaigns
  - Campaigns offer free "teaser" material
- **Table 2: Trials**
  - Records of trial periods
  - Trial period is mandatory before a subscription
- **Table 3: Cancellations**
  - Records of users cancelling subscriptions
- **Table 4: Watched Videos**
  - Daily summary of watched videos by user
  - Includes video genre, brand, etc.

Steps to Analysis

- **Step 1: Initial exploration**
  - Explored each table individually
  - Data cleaning, if needed (none was needed)
  - Record initial insights from each table
- **Step 2: Choose a single question/topic for analysis**
  - Dare Network's Success Factors
  - Statistical Analysis: User Engagement and Trial-Subscription Conversion Rate
- **Step 3. Final Recommendations and Insights**
- **Step 4. Suggest Additional Areas to Research/Questions to Answer**

PROJECT OBJECTIVES

1. Initial Exploratory Data Analysis
   1. Record Initial Insights
   2. Choose an area of focus (Dare Network's Success Factors) and relevant business questions
2. Compare Dare's KPIs to other Brands
3. Statistical Analysis: Confirm if User Engagement is associated with high conversion (trial-subscription) rates
4. Provide actionable insights/recommendations (to drive higher user engagement and conversion rates)
5. Suggest follow-up questions or other areas for research/testing
   1. Based on initial EDA or insights from this specific presentation
   2. These areas will be explored in future projects – stay tuned!

# I. INITIAL EXPLORATORY DATA ANALYSIS & BRAND COMPARISON

INITIAL EXPLORATION: NUMBER OF LEADS BY BRAND

Leads:

- Email campaigns, offering free teaser material (such as pilot episodes)
- Can attract users to sign up for free trials → become monthly/annual subscribers

Initial Insights

- Dare Network and Playground Prime: similar number leads (Dec. 2023-Aug 2024)
- Playground Prime: shows more signs of seasonality
  - Peaks in December and March (how does this compare to previous years?)
- Mosaic: strong lead performance starting in July 2024
- Dare and Playground: Leads dropped drastically from March-August
  - Overall Lead Count: More Stable
  - Why? Seasonality? Mosaic performance? Further analysis and past data required

Number of Leads per Brand

Number of Leads per Brand Over Time (Excluding September)

Number of Leads per Month
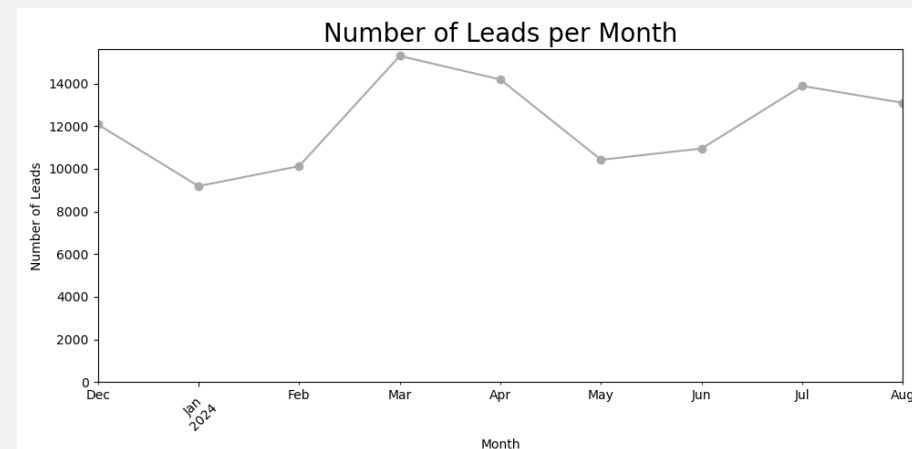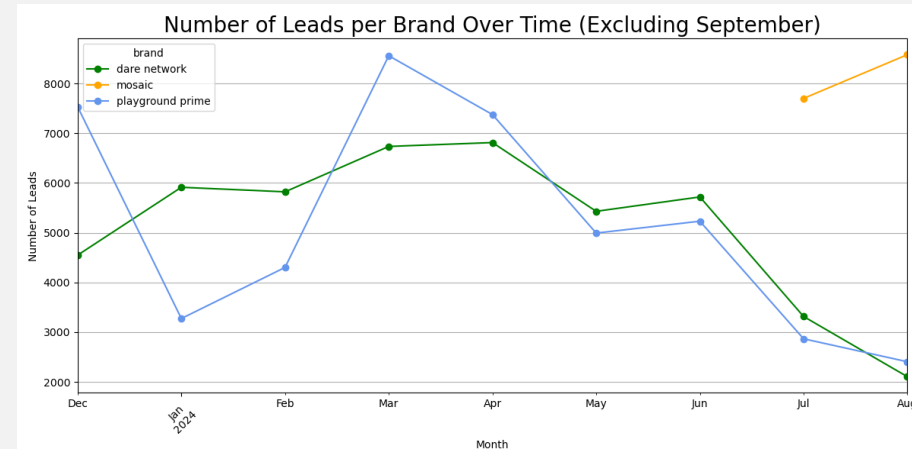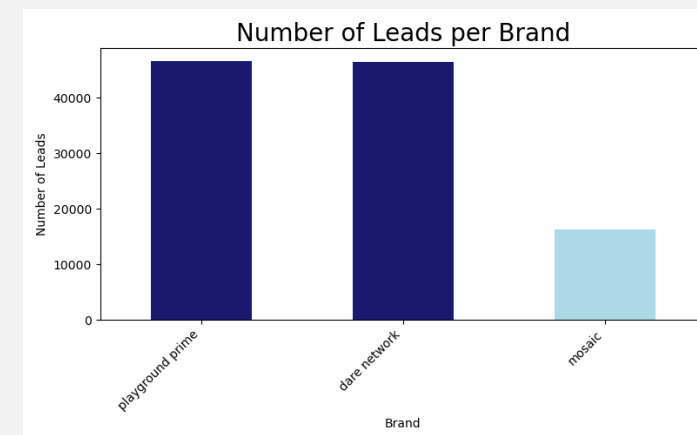
DARE NETWORK HAS OUTPERFORMED PLAYGROUND PRIME

NUMBER OF TRIALS
- Dare Network = 20,166
- Playground Prime =11,473
- Suggests more Dare leads turn into trials

CONVERSION RATE
- Dare Network = 74.9%
- Playground Prime = 65.7%
- Average = 71.2%

NUMBER OF VIDEOS WATCHED
- Dare Network = 214,405
- Playground Prime = 55,157

MOSAIC'S STATS
- Number of Trials = 5,995
- Conversion Rate = 72.2%
- Number of Videos Watched = 29,061

Number of Trials - Has Converted by Brand

Number of Videos Watched by Brand (Videos Watched Table)
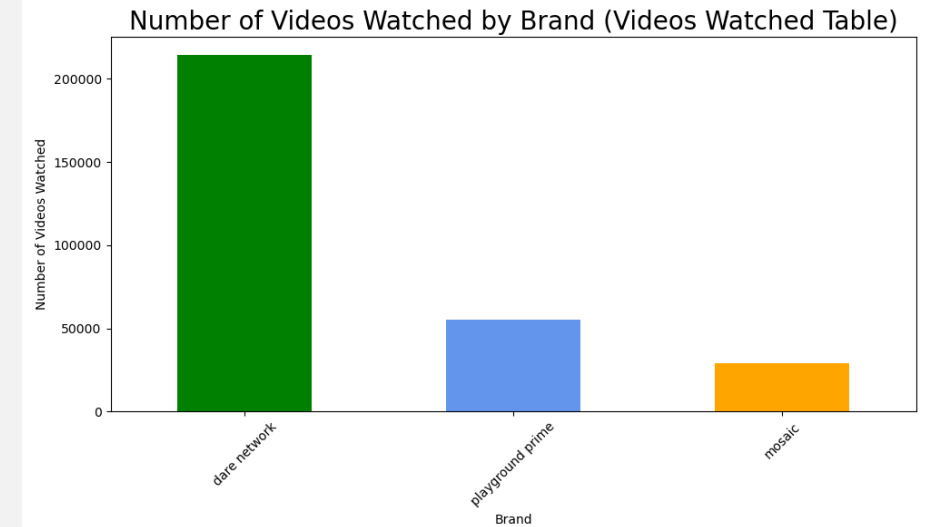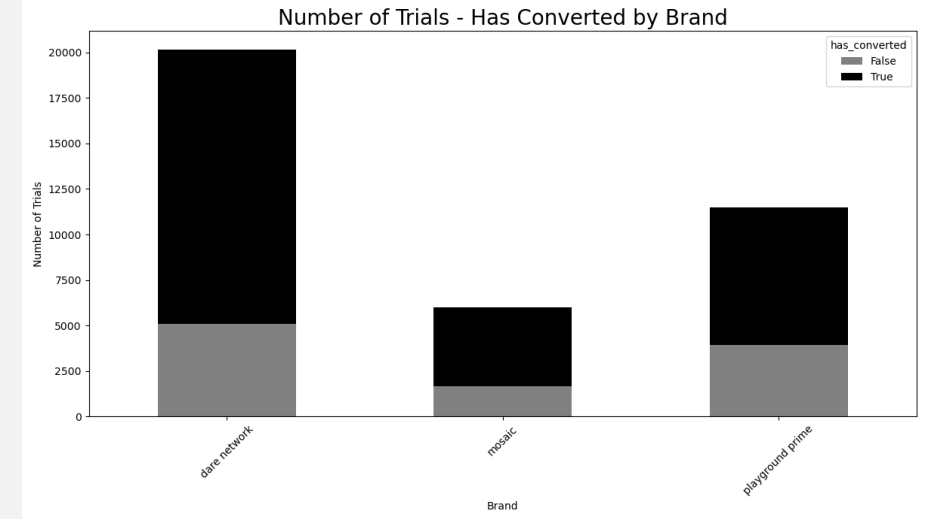
MOST POPULAR GENRES (BY NUMBER OF VIDEOS WATCHED)

Number of Videos Watched by Genre

- The most popular genres across all three brands are:
  - Comedy (1st)
  - Reality (2nd)
  - Thriller (3rd)
  - Drama (4th)
- Dare Network and Mosaic's users exhibit similar behaviors
  - Genre preferences are similar to overall most popular genres
- Playground Prime users differ in preferred genre
  - Family/Kids (1st)
  - Fantasy (2nd)
  - Documentary (3rd)
  - Sci-Fi (4th)

Most Popular Genres in Each Brand (Ordered by Total Videos Watched)
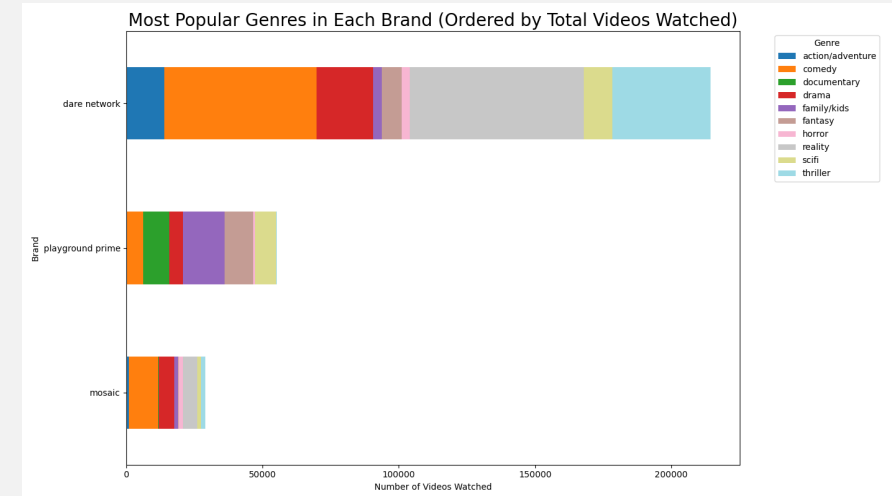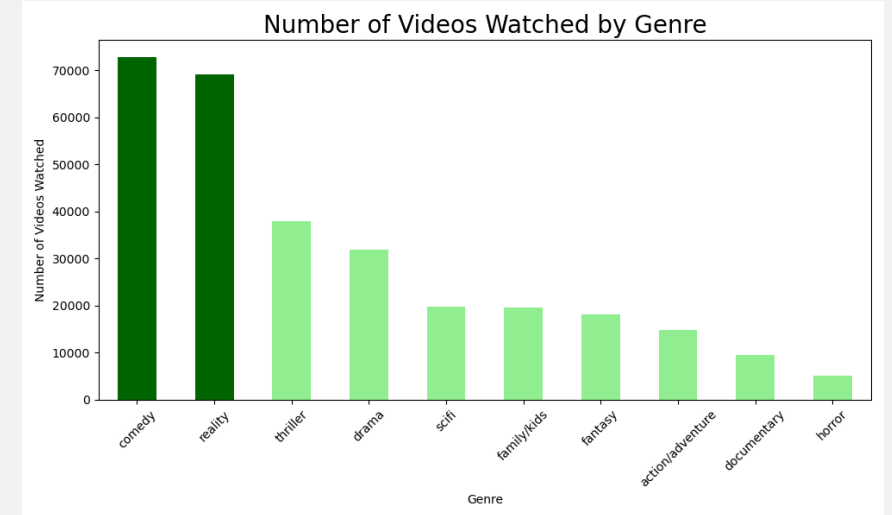
II. STATISTICAL ANALYSIS:
USER ENGAGEMENT & CONVERSION RATES

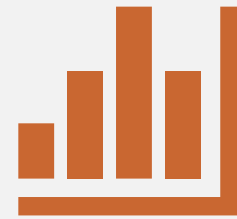* User Engagement is measured by number of videos watched

USER ENGAGEMENT AND CONVERSION RATES

Project Question

Does higher user engagement correlate to high conversion rates (trial to subscription)?

If yes, then move on to further testing and actionable recommendations to drive user engagement

Data Preparation

Aggregate Data by unique individual users:

- # of watched videos
- trial brand
- Did they convert? (Subscribe)

INITIAL INSIGHTS: USER ENGAGEMENT & CONVERSIONS (BY BRAND)

Distribution of Total Videos Watched (Converted vs Non-Converted Users)

Distribution of Total Videos Watched (Zoomed In)

- Do converted users and non-converted watch the same number of videos? Does this vary by brand?

- Box-Plot
  - Green = Dare Network; Orange = Mosaic; Blue = Playground Prime
  - Left = Non-Converted Users; Right = Converted Users

- Insights
  - Converted Users watched more videos than Non-Converted Users
  - Converted Dare Users watched the most videos
  - Non-Converted Dare Users watched more videos than other non-converted users
  - High Variability in User Engagement in All Groups (Highest in Dare Network)
  - High Presence of Outliers in All Groups

- Next Step: T-Test to Confirm Insights from Visualization
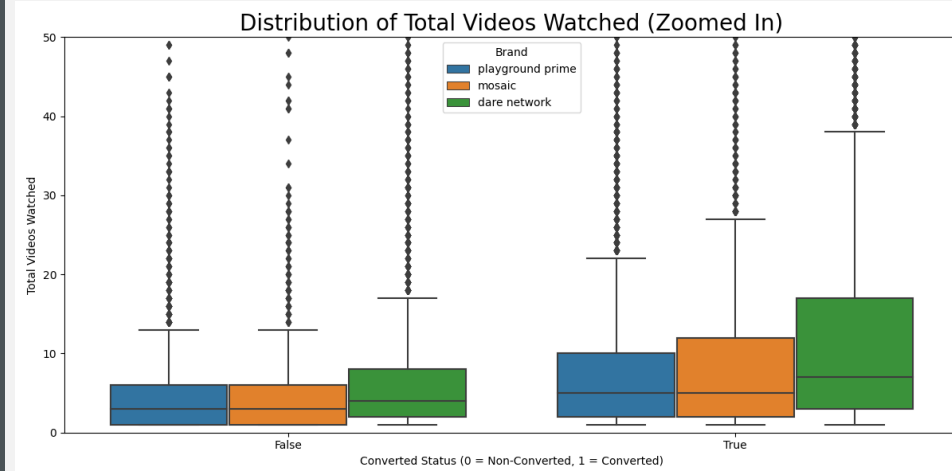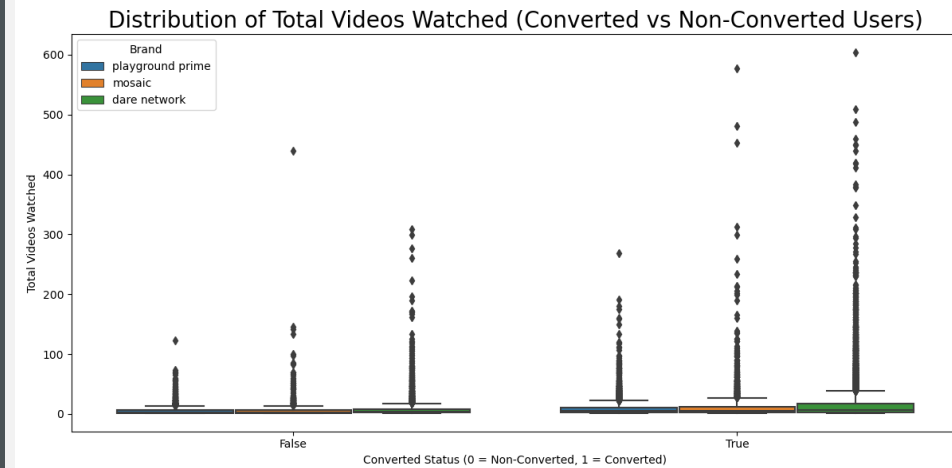
WHY A T-TEST?

- **Purpose:**
  - To compare the average number of videos watched by converted and non-converted users
  - Determine if difference in number of videos watched by each group is significant, or unlikely to have happened by chance
- **Results and Implications of the T-Test**
  - Confirmed! Higher number of videos watched correlates with higher conversion rates
    - Proved Statistical Significance!
    - True for all brands
  - Dare has higher user engagement, aligning with its' users' higher conversion rates → suggests strong engagement strategies

T-TEST DETAILS

- **Null Hypothesis**
  - non_converted_users watch the same number or more videos than converted_users
  - μ converted_users $\leq \mu$ non_converted_users
- **Alternative Hypothesis**
  - non_converted_users watch the same number or more videos than converted_users
  - μ converted_users $> \mu$ non_converted_users
- **Results**
  - T-Statistic = 22.68
  - One-tailed p-value = $1.58^{\wedge}e\text{-}112$

III. RECOMMENDATIONS AND OTHER AREAS TO EXPLORE

ACTIONABLE INSIGHTS AND RECOMMENDATIONS

1. Identify which aspects of Dare Network drive high user engagement

   - Potential Aspects that drive engagement: Content quality/type, UI/UX, marketing, personalized recommendations, interactive content, reward-based incentives etc.

   - Develop and invest in strategies to further increase engagement in all three brands

   - Determine which of these aspects can be applied to the other two brands (for example, content such as reality or thriller may not be appropriate for Playground Prime but may be for Mosaic)

2. Why are many high engagement users not subscribing? Identify potential barriers

   - Potential Barrers to look into: UX/UI, pricing, content type/quality, etc.

   - Conduct surveys for this subgroup to inform marketing strategy or improve product

3. Why did some low engagement users end up subscribing? Identify the factors

4. Monitor Mosaic's user engagement and conversion rates in the coming months (since they're taking over the leads)

RECOMMENDATIONS FOR FURTHER TESTING

- **Experimentation and A/B Testing (Point 1)**
  - Test new strategies or content features (inspired by Dare) to increase user engagement and identifying causation instead of just correlation
  - Business Impact:
    - Identify which strategies/features are most effective in increasing trial to subscription conversion rates
- **ANOVA for Genre Analysis (Points 2, 3):**
  - **Two Purposes**
    - Determine if there are significant differences in the mean number of videos watched across genres
    - Assess if these differences are affected by conversion/subscription status
  - Potential Business Implications:
    - Can identify if some genres are more engaging to converted or non-converted users
    - Can identify if certain content types/genres are associated with high engagement but low conversion rate

RECOMMENDATIONS FOR FURTHER TESTING

- **Outlier Analysis through Surveys (Points 2 and 3)**

  - Can be used to confirm insights from ANOVA (i.e. why a genre might have high engagement but lower conversion)

  - Business Impact

    - Qualitative insights on why outliers (users with extremely high engagement levels) don't convert

- **Logistic Regression: Predict User Conversion (Points 3 and 4)**

  - Identify key factors (genre, brand, number of videos watched, etc.) that increase likelihood of conversion

  - Measures *how strongly* each variable (like each genre) is associated with conversion via correlation coefficient

  - Business Impact

    - Resource Allocation: spend more money on content type that is more likely to get more subscriptions

    - Running regressions for each brand may yield insights specific to each brand (recommended)

    - Cohort Analysis for Mosaic: add a "cohort" variable for each user and see its' correlation to conversion

OTHER AREAS TO EXPLORE

- **Seasonal Campaign Effectiveness**
  - Playground Prime showed signs of seasonality, which is potential for marketing strategies
  - Determine if seasonal peaks are correlated to conversions or cancellations
- **Mosaic's Future Potential**
  - Strong lead-performance in July-August; what drove this strong start?
  - Can Mosaic learn from Dare's success?
- **Compare Dare and Playground Prime's Lead to Trial Conversion Rates**
 - Dare's is much higher! Why?

APPENDIX

POTENTIAL TOPIC 1: PLAYGROUND PRIME'S SEASONALITY

- **Key Insights**:
  - Peaks in number of Leads from Campaigns in December and March, likely due to school breaks
- **Insights Potential for Marketing:**
  - Confirming impact of seasonality can guide timing of campaigns for this brand
  - Determine if seasonal peaks are correlated to conversions or cancellations
- **Further Analysis:**
  - Examine if seasonal peaks translate to higher conversions or cancellations

Number of Leads per Brand Over Time (Excluding September)
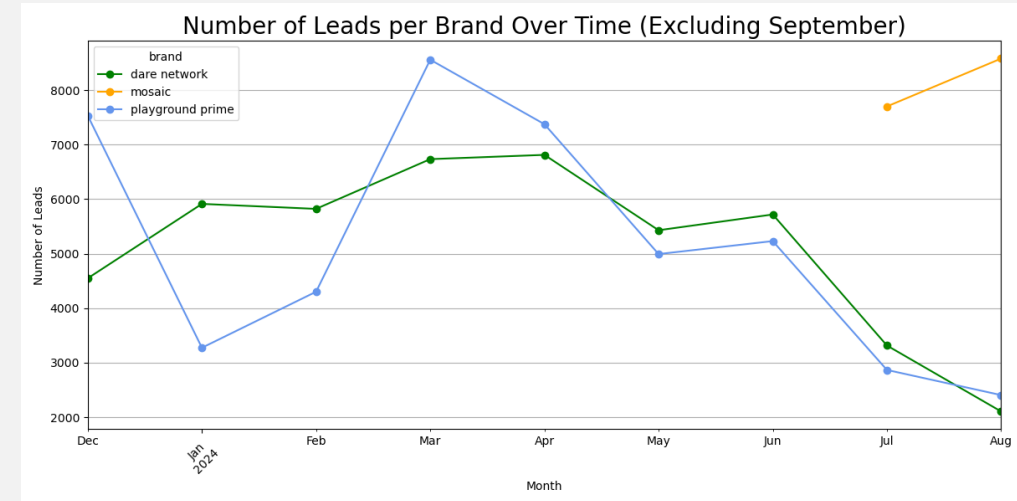
POTENTIAL TOPIC 2: EXPLORING MOSAIC'S FUTURE POTENTIAL

- **Key Insights**:
  - Mosaic had a strong performance in leads in July-August (see previous slide)
  - Mosaic's users prefer similar genres to Dare Network
- **Insights Potential:**
  - Confirming Mosaic user's preferred genres
  - Confirm any correlations between genre, conversion rate, etc.
- **Further Analysis:**
  - What drove Mosaic's strong start?
  - Continue to monitor Mosaic's data (e.g. leads, conversions and cancellations) for analysis as time goes by

Most Popular Genres in Each Brand (Ordered by Total Videos Watched)
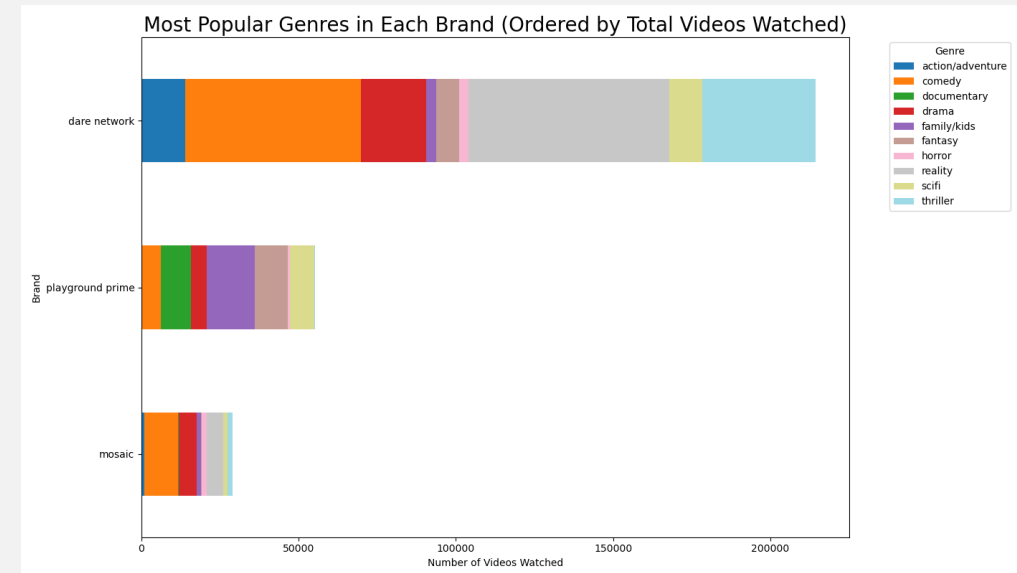
POTENTIAL TOPIC 3: WHY DO MORE DARE LEADS TURN INTO TRIALS THAN PLAYGROUND?

- **Key Insights**:
  - Dare Network and Playground Prime have similar number leads
  - Dare Network has far more trials than Playground Prime, suggesting a higher Lead to Trial Conversion Rate
- **Insights Potential:**
  - Explore why Dare's rates are so much higher
  - Potentially apply similar methods (if applicable) to Playground and Mosaic