Data Story
Capstone Project # 1   - Predicting Popular Video Genres

How do people in different regions of the world view YouTube videos? What does the audience in a South American country, Mexico likes to watch vs. in the USA? Clearly there are a few pellucid differences between the countries, for example, these countries differ a great deal in economic stability, security stability and of course language. With these differences in mind can we find the video watching trends in these two countries and analyze them through YouTube data? The YouTube data used for this analysis contains data from 2017 and 2018. Several units of measurements are collected and data analysis is shown on how the viewers in the two countries watch YouTube videos.

The **unique videos versus average views** in each category for Mexico and USA in 2017 and 2018 are shown in Fig1 through Fig4. First thing that jumps out from these figures is that the **Mexican audience watches unique videos more than American Audiences**. For example in 2017 for Mexico, there are over 2000 unique videos in Entertainment however for the USA there are only 455 unique videos(Fig1 and Fig3). Even though there are more unique videos **in Mexico, the average number of views is less than the USA.** For example in the Entertainment category the number of average views is only 250,000.For the USA in the same category there are 1.3M views. **Music videos incur the most average views every year in both countries**. The most viewed Music category at 7M average views is in 2018 in the USA. Another observation from this data is that the Pets & Animal category seems to be second to the Music category in terms of average views in Mexico for 2017 and 2018. For the USA the second most viewed category is Film & Animation in both 2017 and 2018.

The videos in the **Shows category dominated over other categories in the USA and brought in 553% more average views in 2018**. In contrast, **views from Mexico for the Shows category dropped 88% in 2018**. The most average views gained in Mexico were in the Science & Technology category. **Mexico viewers watched more of the Science & Technology category videos that gained 240% more average views in 2018.** Other categories that lost average views in Mexico in 2018 were the Gaming and the Comedy category. Interestingly none of the categories dropped in the number of average views in the USA. However minimal the change, it was still positive.

A review of correlation between average views, likes, dislikes and comment count between the two countries show that **there is a slight correlation between likes/ views, dislikes/views and comment count/views for videos watched in Mexico**. For the USA there is none (Fig7). From May 2017 to Dec 2018 a monthly review of the average views and the average likes/dislikes(Fig8 and Fig9) show that the views peaked in June 2018 but the average number of likes/dislikes were mostly steady throughout this time period. Average likes did peak for Mexico videos in June 2018 however for the USA it pretty much remained the same. For dislikes the trend was most flat throughout the time period.

Each category's views are very different from each other. There are clearly outliers in each category which are shown in Fig10(a-c). A few prominent outliers for video views in Mexico are Education, Pets & Animals and Science & Technology. Out of 7 categories, those three categories have outliers more than USA views outliers. All categories show that average views in all the categories are lower than USA average views.
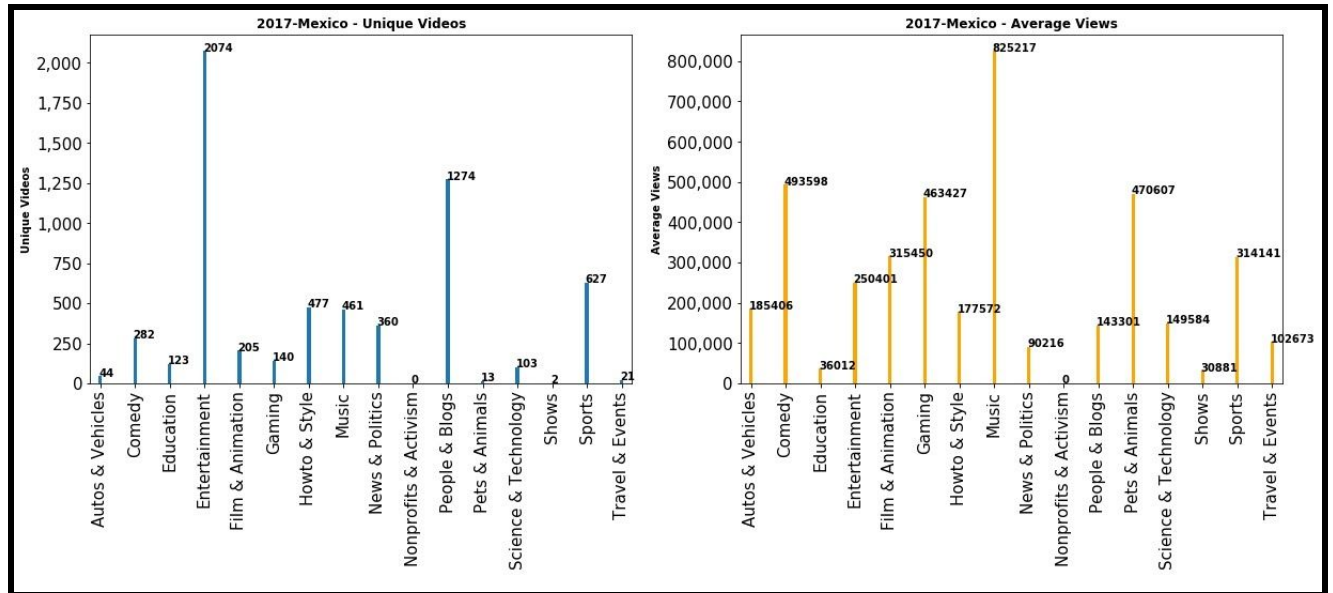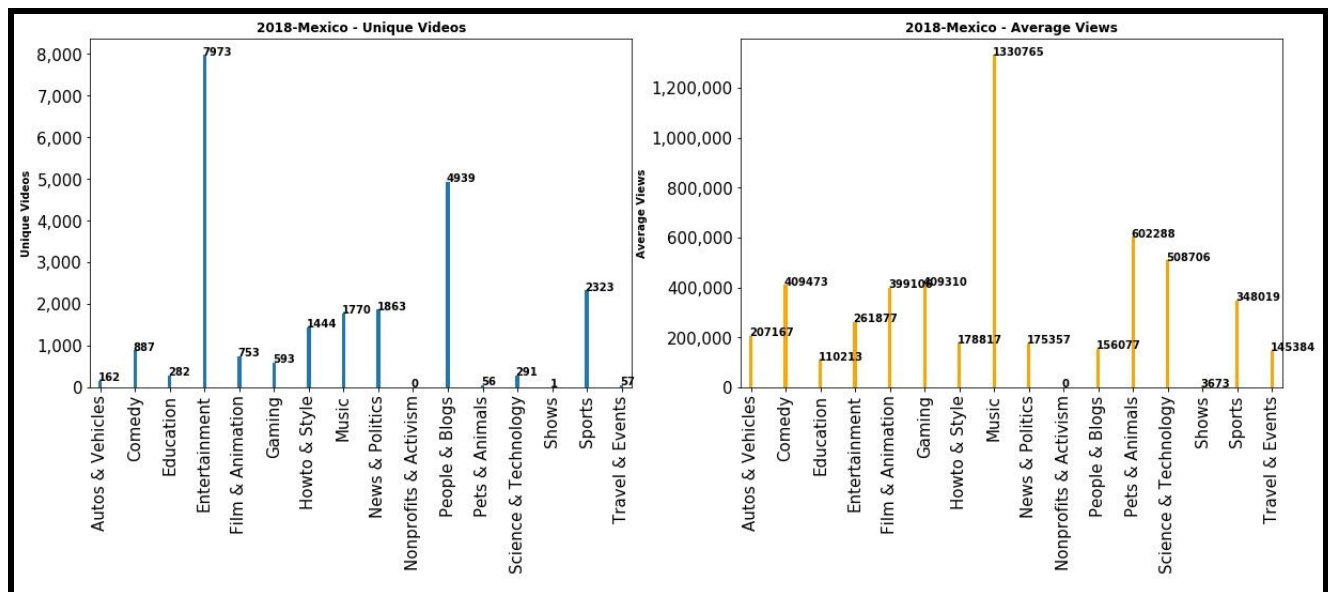

Fig1: Mexico - Unique Videos vs Average Views, 2017

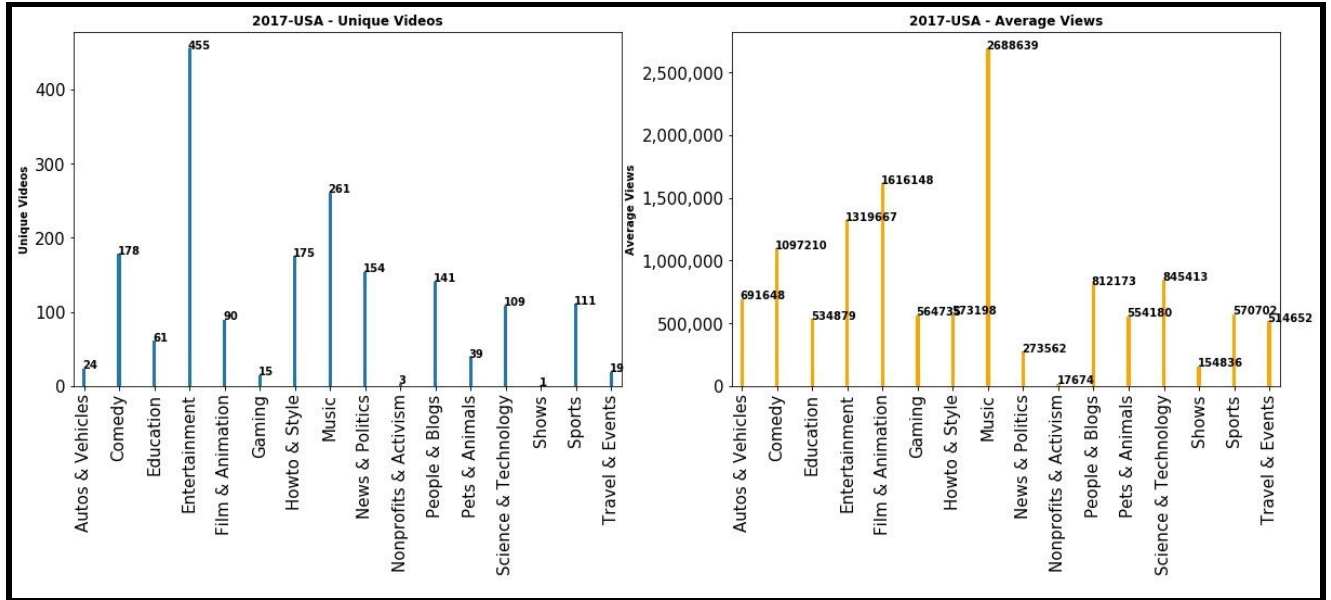Fig2: Mexico - Unique Videos vs. Average Views, 2018

Fig3: USA - Unique Videos vs. Average Views, 2017
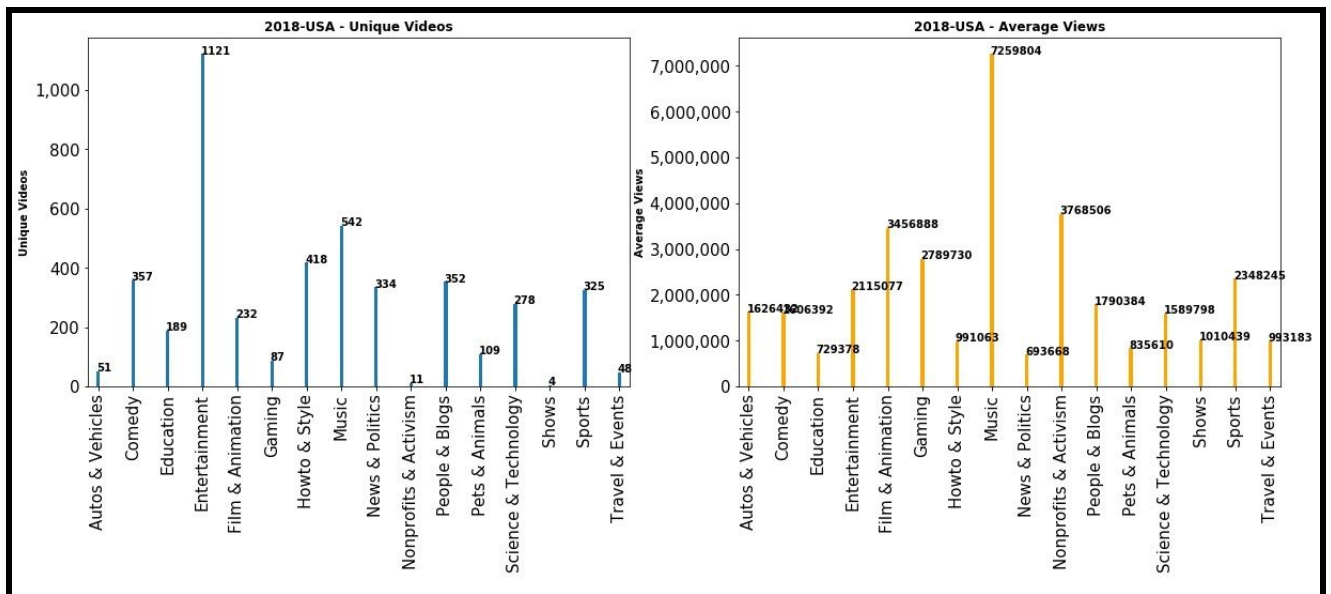

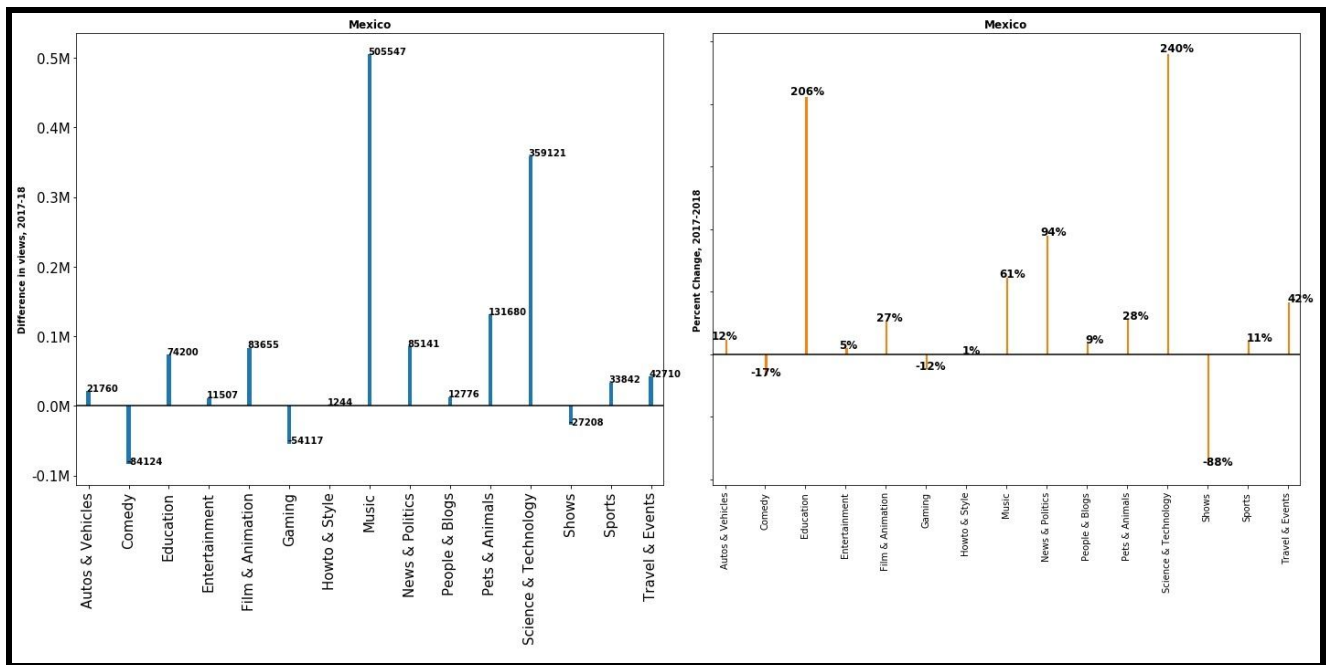Fig4: USA - Unique Videos vs. Average Views, 2018

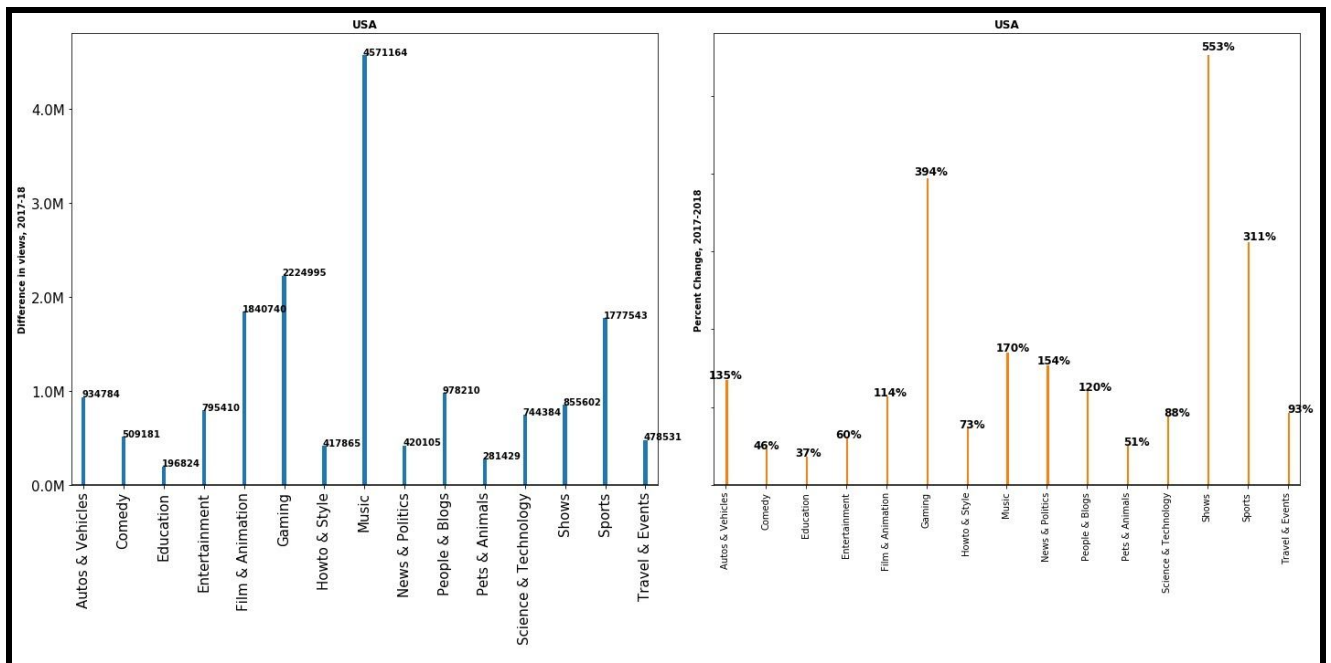Fig5: Mexico - Difference in views from 2017 to 2018

Fig6: USA - Difference in views from 2017 to 2018

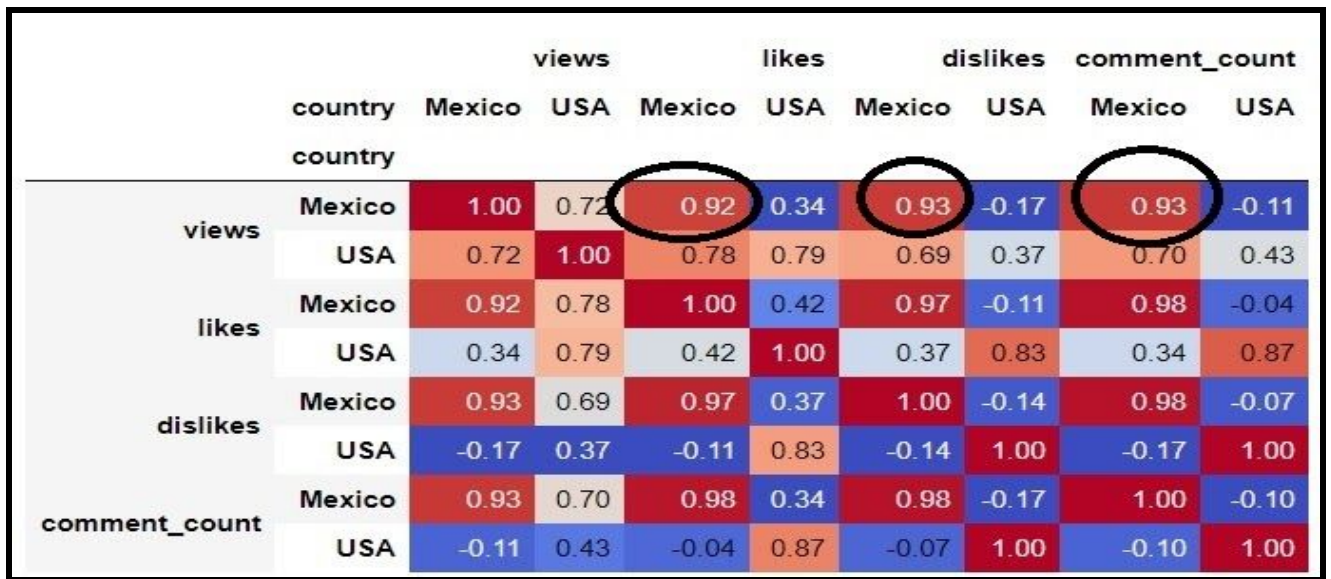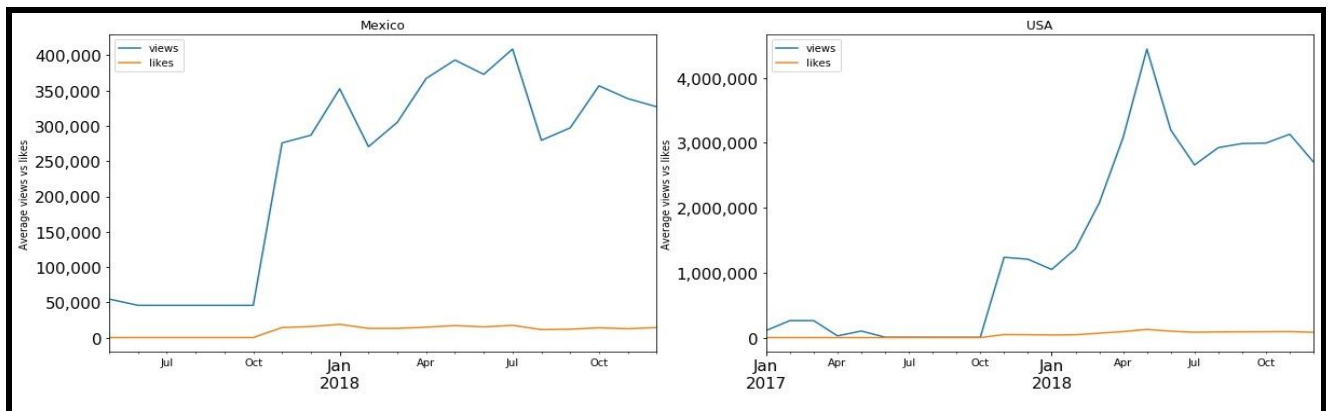		country	views		likes		dislikes		comment_count	
	country	Mexico	USA	Mexico	USA	Mexico	USA	Mexico	USA
views	Mexico	1.00	0.72	0.92	0.34	0.93	-0.17	0.93	-0.11
	USA	0.72	1.00	0.78	0.79	0.69	0.37	0.70	0.43
likes	Mexico	0.92	0.78	1.00	0.42	0.97	-0.11	0.98	-0.04
	USA	0.34	0.79	0.42	1.00	0.37	0.83	0.34	0.87
dislikes	Mexico	0.93	0.69	0.97	0.37	1.00	-0.14	0.98	-0.07
	USA	-0.17	0.37	-0.11	0.83	-0.14	1.00	-0.17	1.00
comment_count	Mexico	0.93	0.70	0.98	0.34	0.98	-0.17	1.00	-0.10
	USA	-0.11	0.43	-0.04	0.87	-0.07	1.00	-0.10	1.00

Fig7: Correlation Matrix
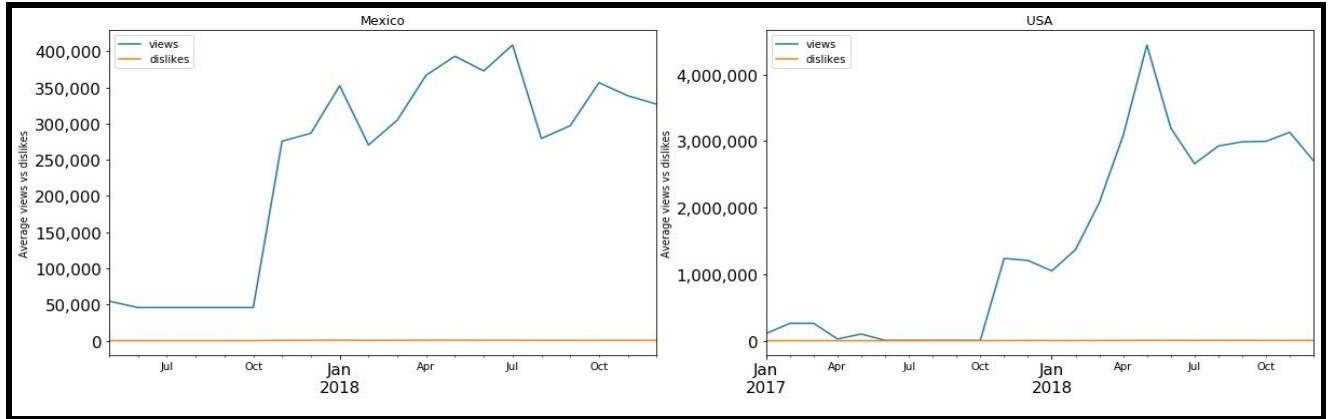

Fig8: Average Views vs Average Likes
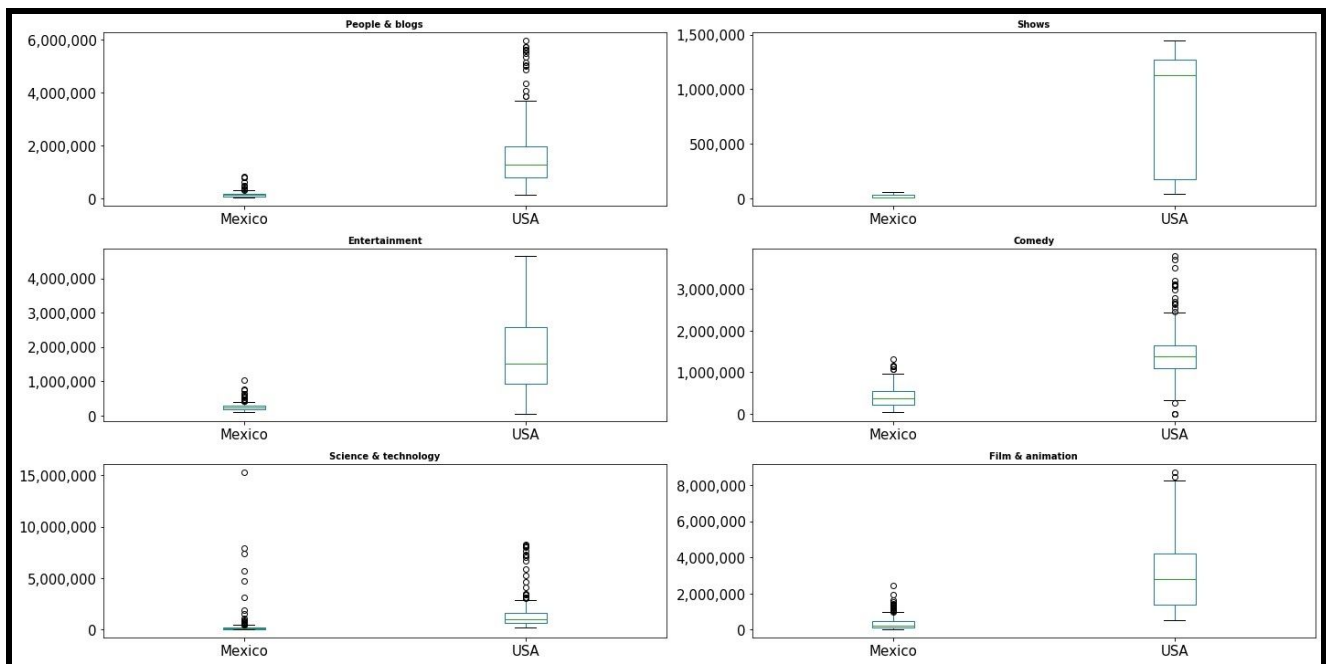
Fig9: Average Views vs Average Dislikes

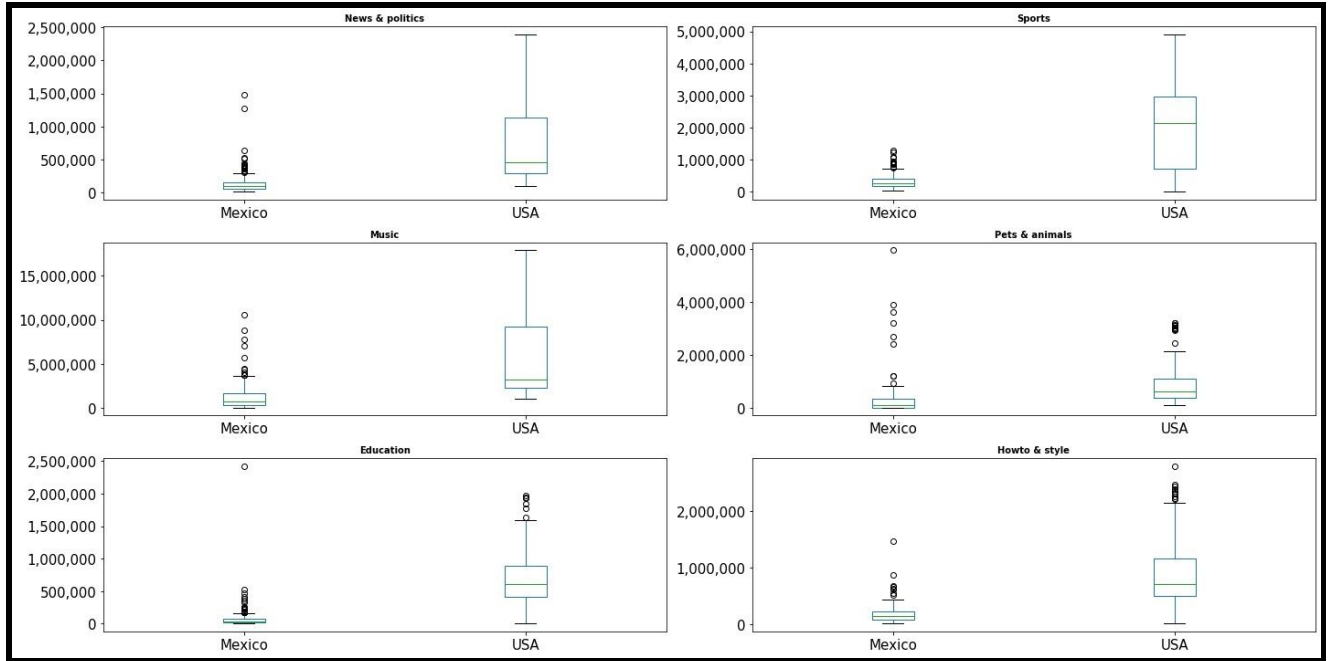Fig10a: Boxplots - All categories in USA & Mexico
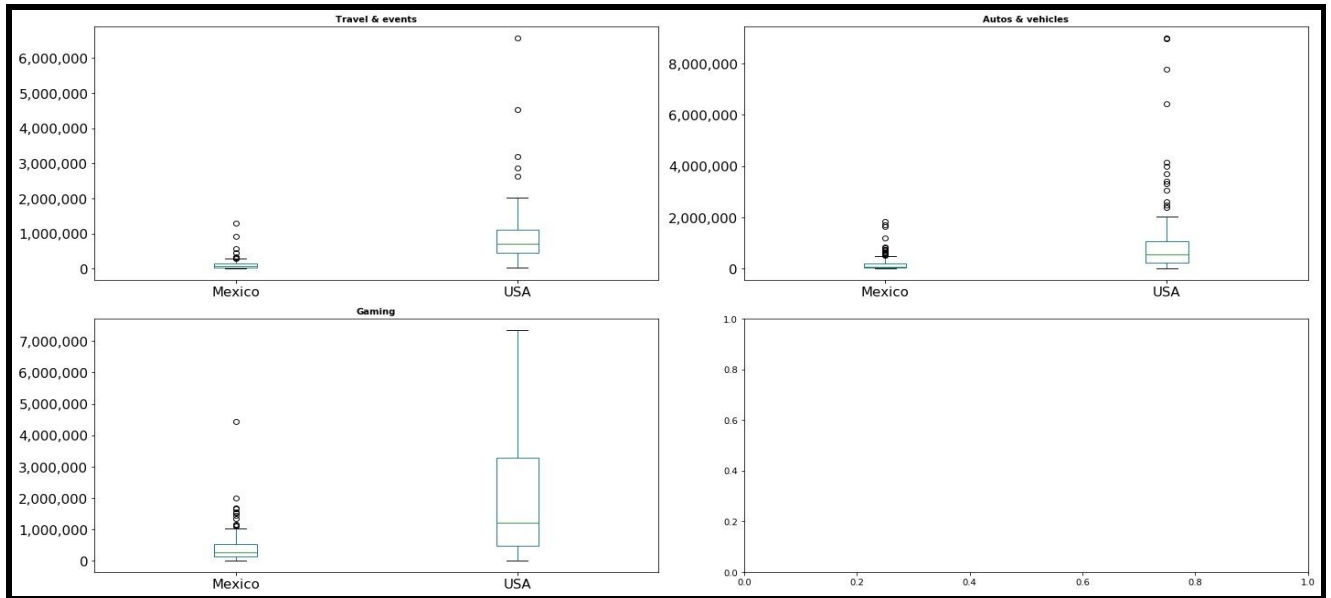
Fig10b: Boxplots - All categories in USA & Mexico

Fig10c: Boxplots - All categories in USA & Mexico