

YOUTUBE DATA ANALYSIS USING MAPREDUCE

IDS 561 - ANALYTICS FOR BIG DATA
FALL 2018



Devika Subhash Chaudhari
Sindhu Janardhanarao
Vigneshwaran Giri Velumani

GOAL



1. Top 5 Categories with highest number of Views
2. Top 5 Categories with highest number of Rating
3. Top 5 Categories with highest number of Comments

COMMENTS - MAP REDUCE CODE FOR TWO MAPPERS AND REDUCERS



```
public static class MapperforComment_1 extends Mapper<Object, Text, Text, IntWritable> {
```

```
    private Text cat = new Text();
```

```
    private IntWritable comment = new IntWritable();
```

```
    public void map(Object key, Text value, Context context)
```

```
        throws IOException, InterruptedException {
```

```
        String str[] = value.toString().split("\\t");
```

```
        if (str.length >= 5) {
```

```
            cat.set(str[1]);
```

```
            comment = new IntWritable(Integer.parseInt(str[5]));
```

```
        }
```

```
        context.write(cat, comment);
```

```
    }
```

```
}
```

```
public static class ReducerforComment_1 extends Reducer<Text, IntWritable, Text, IntWritable> {
```

```
    @Override
```

```
    public void reduce(Text key, Iterable<IntWritable> values, Context context)
```

```
        throws IOException, InterruptedException {
```

```
        int sum = 0;
```

```
        for (IntWritable val : values) {
```

```
            sum += val.get();
```

```
        }
```

```
        context.write(key, new IntWritable(sum));
```

```
    }
```

```
}
```

```
public static class MapperforComment_2 extends Mapper<Object, Text, IntWritable, Text> {
```

```
    protected void map(Object key, Text value, Mapper.Context context) throws IOException, InterruptedException {
```

```
        String line[] = value.toString().split("\\t");
```

```
        Text cat = new Text(line[0]);
```

```
        String comment = line[1];
```

```
        try {
```

```
            IntWritable comm = new IntWritable(Integer.parseInt(comment));
```

```
            context.write(comm, cat);
```

```
        } catch (Exception e) {
```

```
            e.printStackTrace();
```

```
        }
```

```
    }
```

```
public static class ReducerforComment_2 extends Reducer<IntWritable, Text, Text, IntWritable> {
```

```
    private static int counter = 10;
```

```
    protected void reduce(IntWritable key, Iterable<Text> values, Context context) throws IOException, InterruptedException {
```

```
        for (Text val : values) {
```

```
            if (counter > 0) {
```

```
                context.write(val, key);
```

```
                counter--;
```

```
            } else {
```

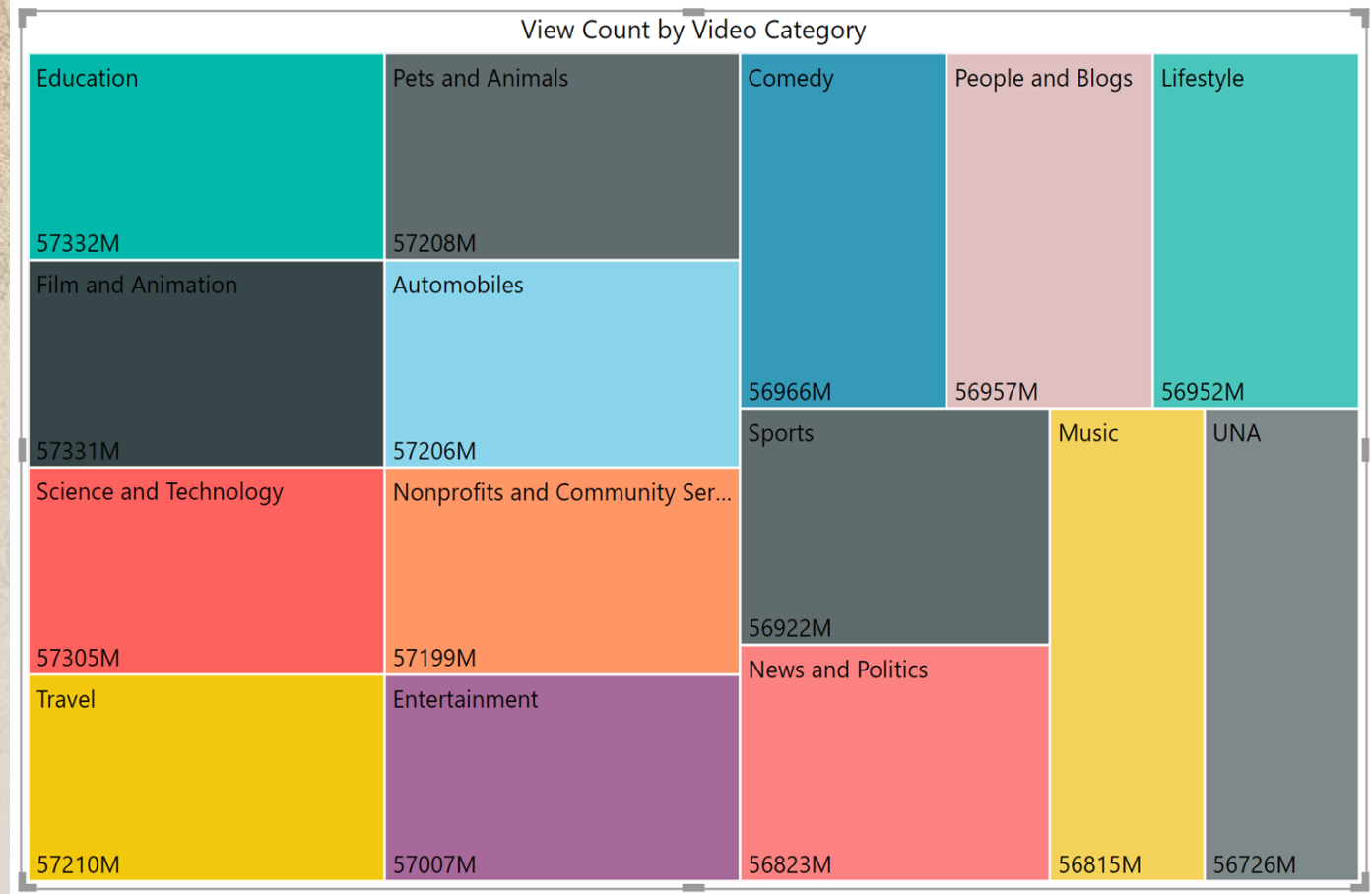
```
                break;
```


POWER BI REPORT FOR #VIEWS



Top 5 Categories:

1. Education
2. Film and Animation
3. Science and Technology
4. Travel
5. Pets and Animals

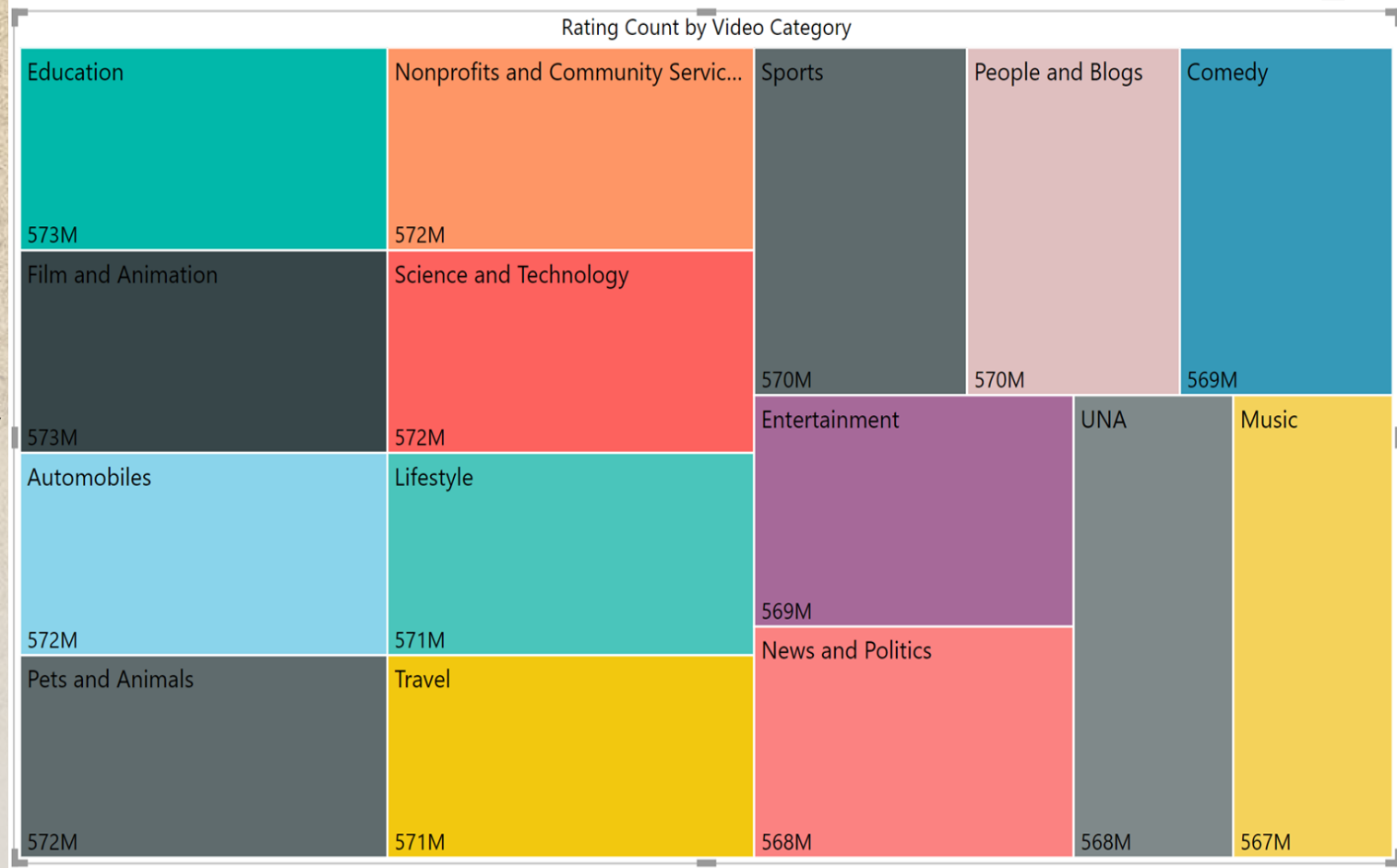


POWER BI REPORT FOR #RATINGS



Top 5 Categories:

1. Education
2. Film and Animation
3. Automobiles
4. Pets and Animals
5. Nonprofits & Community

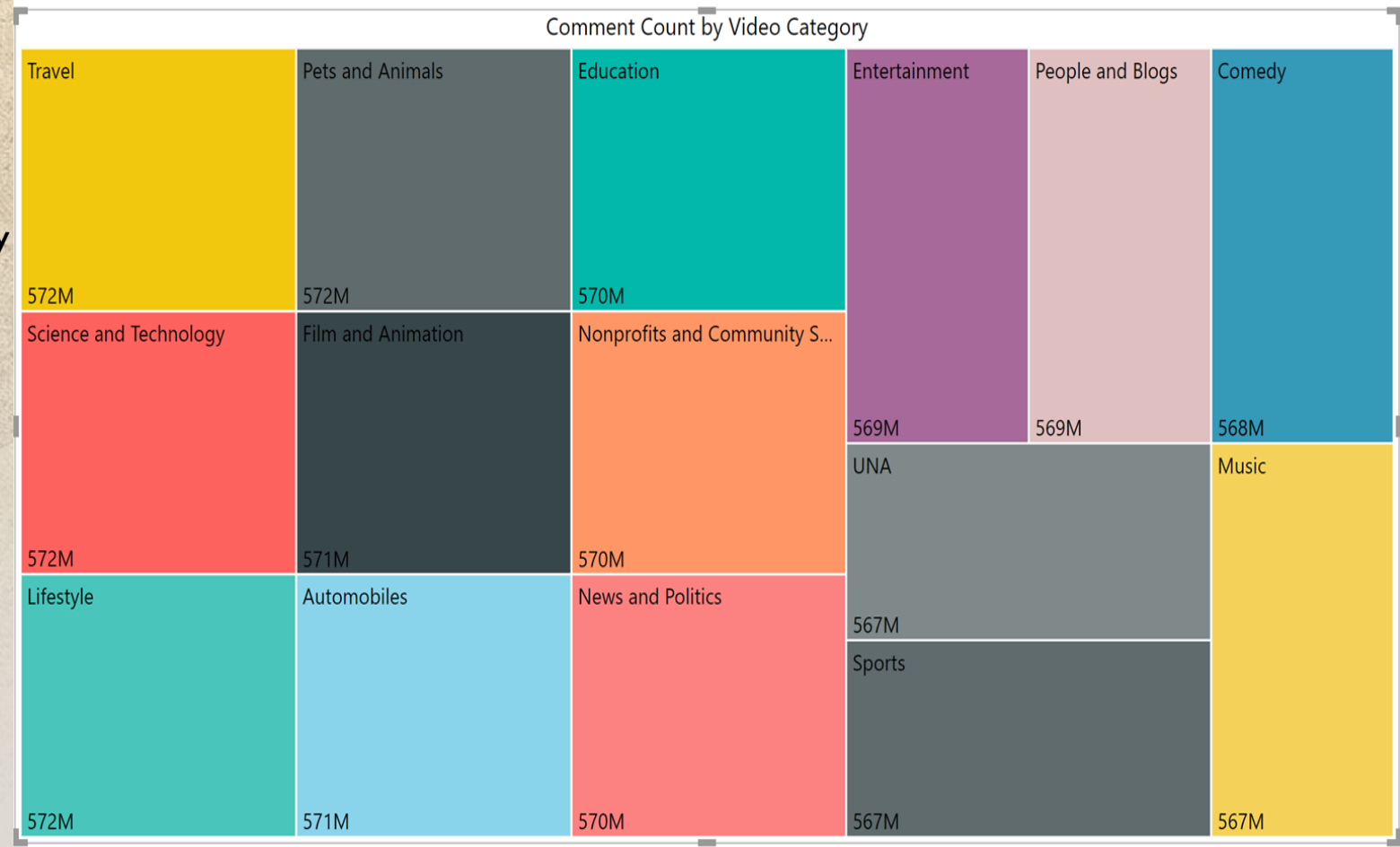


POWER BI REPORT FOR # COMMENTS



Top 5 Categories

- 1.Travel
- 2.Science and Technology
- 3.Lifestyle
- 4.Pets and Animals
- 5.Film and Animation



SUMMARY



Key Takeaways:

- Map reduce technique helps businesses to analyze the findings in the bigger picture
- This sample prototype can be upscaled to analyze huge datasets

The background of the slide features faint, sepia-toned sketches of two airships. On the left is a hot air balloon with a large, ornate, patterned envelope and a small basket at the bottom. On the right is a rigid blimp or zeppelin with a long, cylindrical hull, a series of vertical struts, and a horizontal gondola with a propeller at the rear. The text "THANK YOU!" is centered over the middle of the image.

THANK YOU!

