PELOTON DIGITAL CONTENT

BY JORGE SANCHEZ

DIGITAL FITNESS CONTENT

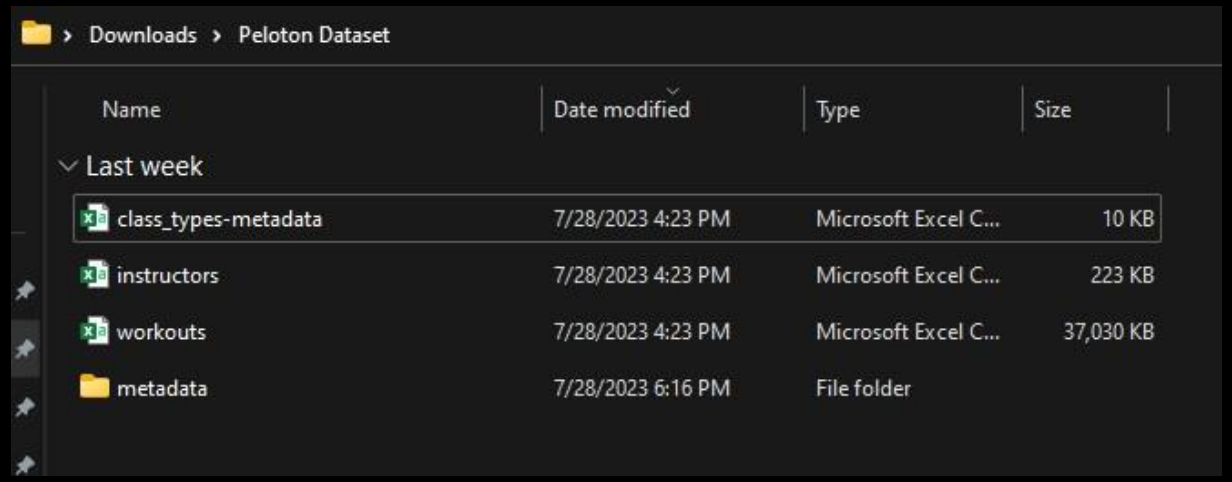
- INTRODUCTION
- DATASET
- DATA MANIPULATION
- EXPLORATORY DATA ANALYSIS
- SUMMARY

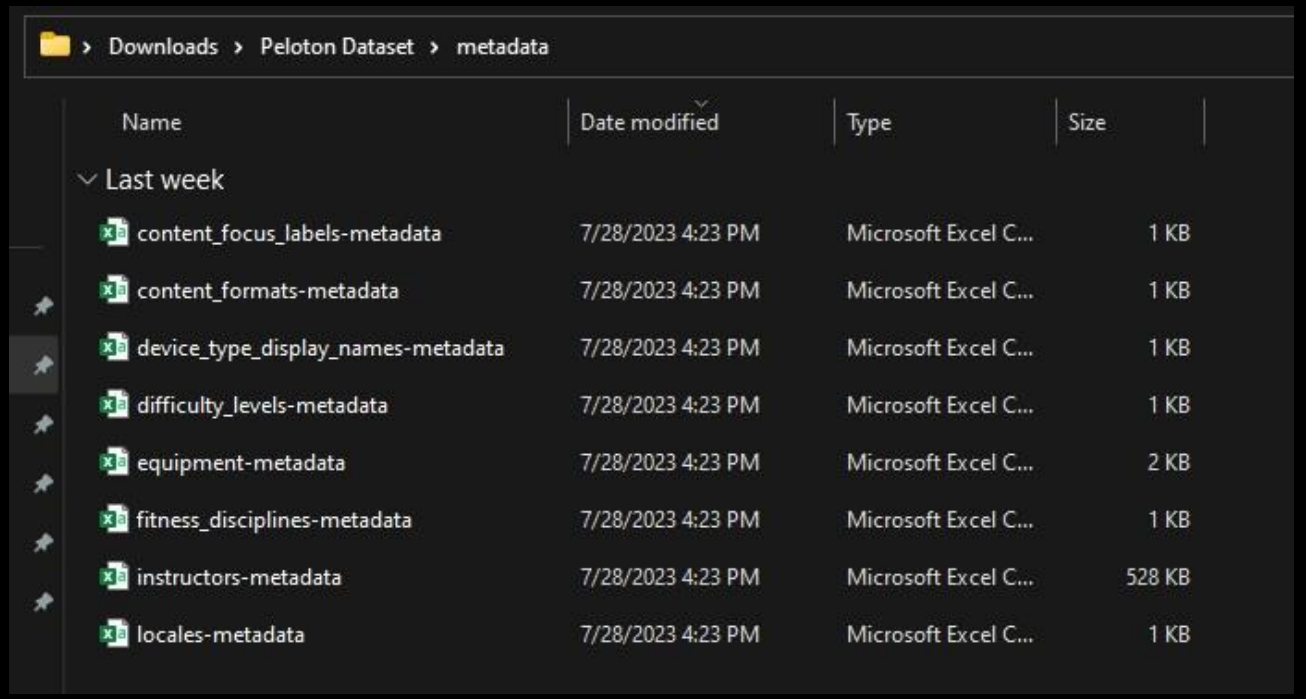
INTRODUCTION

INTRODUCTION
INTRODUCTION

Inspired by the success and popularity of Peloton's Digital platform, I embarked on this project to explore the diverse and dynamic landscape of fitness content available to users worldwide. Peloton's revolutionary approach to fitness instruction, combined with the convenience of on-demand streaming, has reshaped the way people approach home workouts.

In this presentation, I will walk you through the key findings and trends I discovered during my in-depth analysis of the Peloton dataset. We will delve into the wealth of data available, ranging from class durations and difficulty levels to instructor popularity and user ratings. By uncovering the factors that contribute to Peloton's success, I aim to draw parallels and identify opportunities for personal trainers like myself to create engaging and effective fitness content for my own video platform.





Peloton's Digital platform offers an extensive range of live and pre-recorded fitness classes led by top-tier instructors from around the world. From cycling and running to strength training, yoga, and meditation, there is a diverse array of workouts to cater to every fitness level and interest.

With such an extensive content, the original dataset is full of insight. From their public API, I downloaded 11 datasets, my first steps were to separate the relevant data from the irrelevant. Narrowing it down to 3 datasets full of good information. Class types, instructors and workouts.

Importing Instructors Dataset

```
In [2]: 1 instructors_dataset = pd.read_csv('Downloads\Peloton Dataset\instructors.csv')

In [3]: 1 #original instructors Dataset
        2 instructors_dataset
```

Out[3]:

	id	bio	short_bio	coach_type	is_filterable	is_instructor_group	is_visible	is_announced	list_order	f
0	b8c2734e18a7496fa146b3a42465da67	To Aditi, yoga goes beyond movement and can br...	NaN	peloton_coach	True	False	True	True	10	
1	f962a2b1b34d424cabab73bef81bc8db	Adrian is a powerhouse instructor with an unex...	Adrian is a powerhouse instructor with an unex...	peloton_coach	True	False	True	True	11	
2	15f01c9145de4d21b58c1a3e4e44a486	I foolishly thought that rowing was like runni...	Alex comes to Peloton Row via the World Champi...	peloton_coach	True	False	True	True	13	
3	2e57092bee334c8c8dcb9fe16ba5308c	Alex Toussaint has a love for fitness, teamwor...	"Feel good. Look good. Do better! \r\n... Trust ...	peloton_coach	True	False	True	True	14	

```
In [4]: 1 #checking full list of columns
        2 instructors_dataset.columns

Out[4]: Index(['id', 'bio', 'short_bio', 'coach_type', 'is_filterable',
               'is_instructor_group', 'is_visible', 'is_announced', 'list_order',
               'featured_profile', 'film_link', 'facebook_fan_page', 'music_bio',
               'spotify_playlist_uri', 'background', 'ordered_q_and_as',
               'instagram_profile', 'strava_profile', 'twitter_profile', 'quote',
               'username', 'name', 'first_name', 'last_name', 'user_id',
               'life_style_image_url', 'bike_instructor_list_display_image_url',
               'web_instructor_list_display_image_url',
               'ios_instructor_list_display_image_url', 'about_image_url', 'image_url',
               'jumbotron_url', 'jumbotron_url_dark', 'jumbotron_url_ios',
               'web_instructor_list_gif_image_url', 'instructor_hero_image_url',
               'workout_share_images', 'fitness_disciplines'],
              dtype='object')

In [7]: 1 instructors_dataset.shape

Out[7]: (58, 38)
```

Using Jupyter notebook, I imported each dataset to get a deeper insight into the data. The instructor's dataset as seen in the left images has a shape of 58, 38. This suggest 58 rows and 38 columns. While the workouts data set is the biggest holding a whooping 32,528 rows and 13 columns of data.

```
In [18]:
1 workouts_dataset.columns
```

```
Out[18]: Index(['id', 'availability', 'class_type_ids', 'content_provider',
       'content_format', 'description', 'difficulty_estimate',
       'overall_estimate', 'difficulty_rating_avg', 'difficulty_rating_count',
       'difficulty_level', 'duration', 'equipment_ids', 'equipment_tags',
       'explicit_rating', 'extra_images', 'fitness_discipline',
       'fitness_discipline_display_name', 'has_closed_captions',
       'has_pedaling_metrics', 'home_peloton_id', 'image_url', 'instructor_id',
       'individual_instructor_ids', 'is_archived', 'is_closed_caption_shown',
       'is_explicit', 'has_free_mode', 'is_live_in_studio_only', 'language',
       'origin_locale', 'length', 'live_stream_id', 'live_stream_url',
       'location', 'metrics', 'original_air_time', 'overall_rating_avg',
       'overall_rating_count', 'pedaling_start_offset', 'pedaling_end_offset',
       'pedaling_duration', 'rating', 'ride_type_id', 'ride_type_ids',
       'sample_vod_stream_url', 'sample_preview_stream_url',
       'scheduled_start_time', 'series_id', 'sold_out', 'studio_peloton_id',
       'title', 'total_ratings', 'total_in_progress_workouts',
       'total_workouts', 'vod_stream_url', 'vod_stream_id', 'captions',
       'join_tokens', 'flags', 'is_dynamic_video_eligible',
       'is_fixed_distance', 'dynamic_video_recorded_speed_in_mph',
       'thumbnail_title', 'thumbnail_location', 'distance', 'distance_unit',
       'distance_display_value', 'muscle_group_score', 'is_sessions_eligible',
       'total_user_workouts', 'total_following_workouts', 'is_favorite'],
      dtype='object')
```

```
In [12]:
1 #selecting relevant columns
2 workouts = workouts_dataset[['class_type_ids', 'title','duration','fitness_discipline_display_name', 'difficulty_level',
3                  'difficulty_rating_count','total_in_progress_workouts','instructor_id',
4                  'is_closed_caption_shown','is_explicit', 'is_live_in_studio_only','original_air_time',
5                  'overall_rating_avg']]
```

Selecting only the information I wanted to explore. I merged the datasets with an inner join, allowing me to have all the columns and rows in one set.

My next steps were to clean and verified the data for missing values, and wrong data types that will interfere with my research.

```
In [24]:
1 #concat instructors data set with workouts dataset
2 fitness_classes = pd.merge(instructors, workouts, left_on='id', right_on='instructor_id', how='inner')
3
```

After checking for missing values and discovering 22,612 missing values, it was important to plan about this missing data. Luckily, this ‘difficulty level’ was data provided by the platform and not the user. Such as ‘Intermediate’, ‘Advanced’ labeling for the classes. After dodging a bullet, I decided to drop it and check the data types for the other values. I converted ‘original_air_time’ into a date format and I was ready to explore this dataset.

```
In [26]:
#check missing values
fitness_classes.isna().sum()
```

```
Out[26]:
first_name                          0
last_name                           0
fitness_disciplines                 0
class_type_ids                      0
title                               0
duration                            0
fitness_discipline_display_name     0
difficulty_level                22612
difficulty_rating_count             0
total_in_progress_workouts          0
instructor_id                       0
is_closed_caption_shown             0
is_explicit                         0
is_live_in_studio_only              0
original_air_time                   0
overall_rating_avg                  0
dtype: int64
```

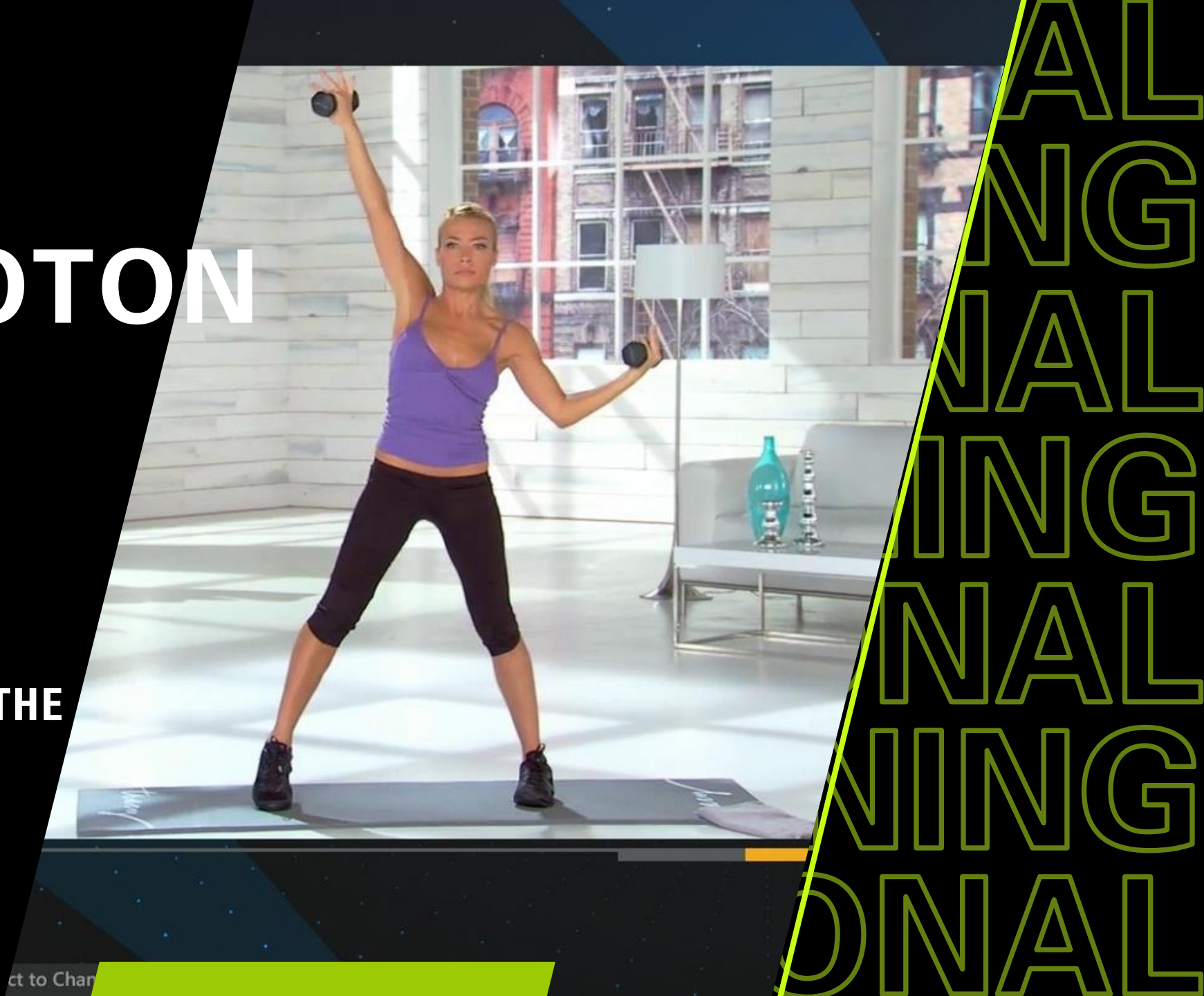
```
In [28]:
#data types of values
fitness_classes.dtypes
```

```
Out[28]:
first_name                        object
last_name                         object
fitness_disciplines               object
class_type_ids                    object
title                             object
duration                           int64
fitness_discipline_display_name   object
difficulty_level                  object
difficulty_rating_count            int64
total_in_progress_workouts         int64
instructor_id                     object
is_closed_caption_shown             bool
is_explicit                         bool
is_live_in_studio_only              bool
original_air_time                  int64
overall_rating_avg               float64
dtype: object
```

```
In [29]:
#converts column from integer to datetime
fitness_classes['original_air_time'] = pd.to_datetime(fitness_classes['original_air_time'], unit='s')
```

THE PELOTON EFFECT

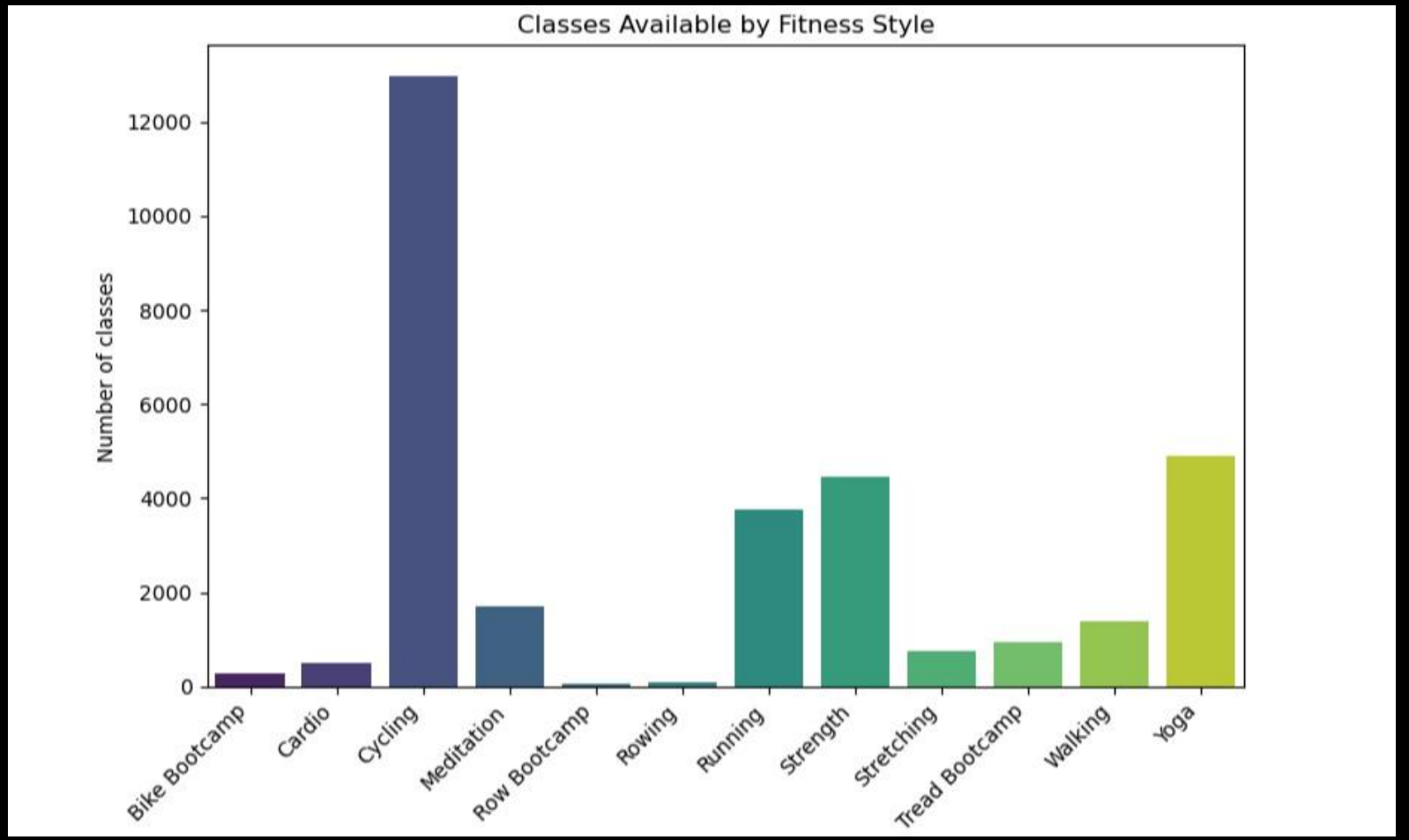
FINDING RESULTS WITHIN THE AVAILABLE DATA

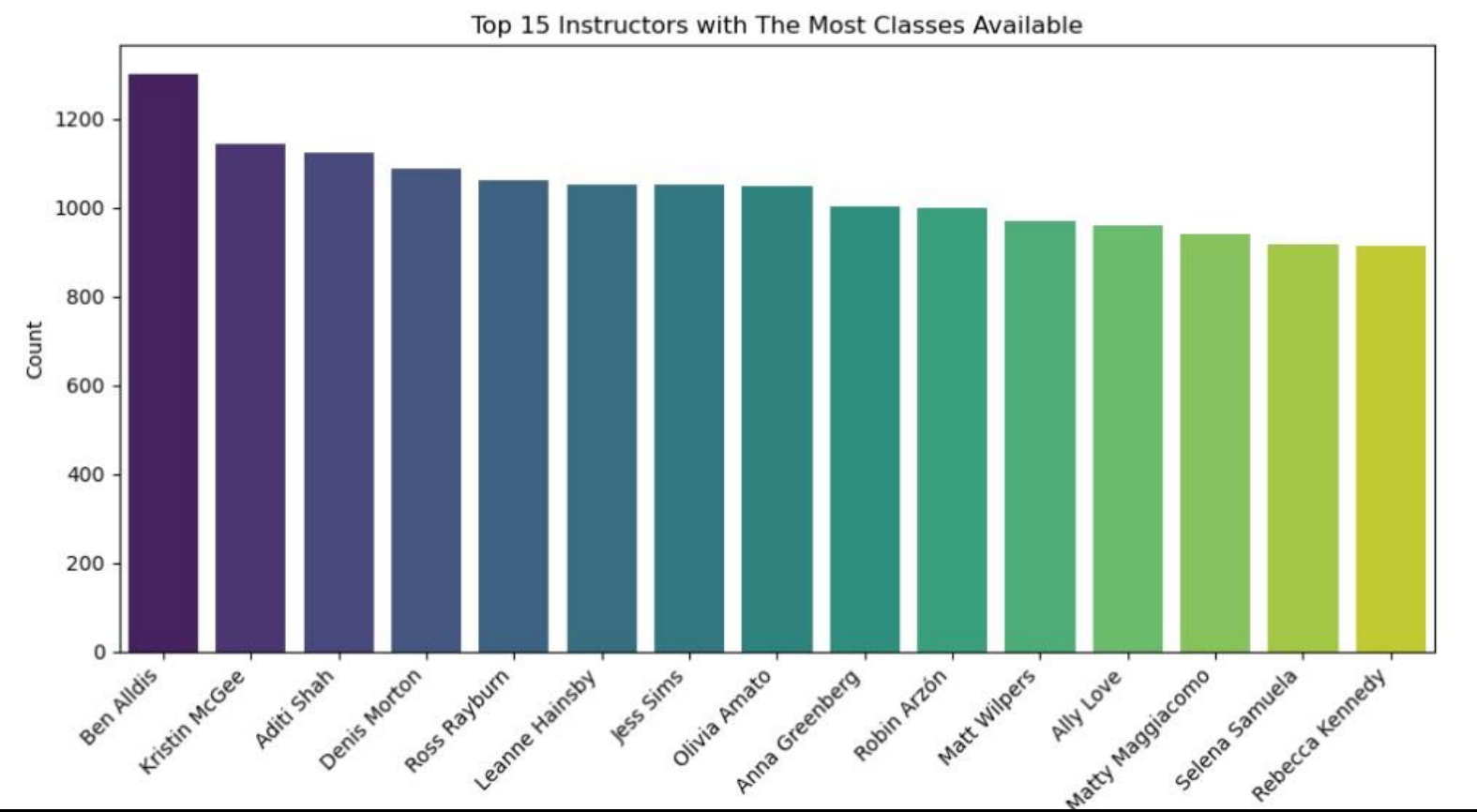


CLASSES BY STYLES

STYLES

The analysis shows that Peloton's highest content is for cycling which is part of their bikes. With their recent shift to a more content first approach, we can expect this numbers to change. Focusing more on other types of training. Yoga is the close 2nd, while strength is the 3rd.





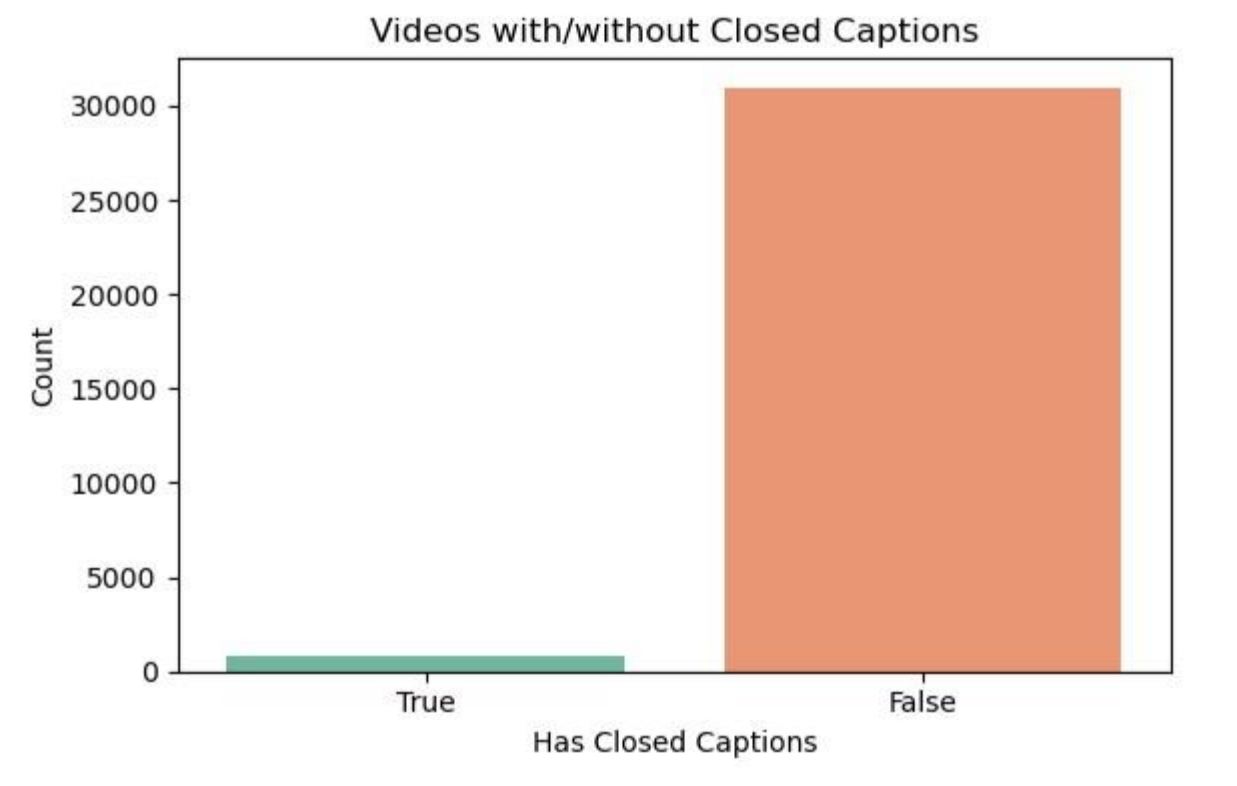
+TOP 15 INSTRUCTORS BY CONTENT

Another interesting insight was to find Ben Alldis as the most featured instructor in the platform. Ben Allidis is one of many top instructors in their lineup and since he is featured in the UK and US versions of their content, this explains his rise to the top. Following Ben, is Kristin McGee, and Aditi Shah.

```
closed_caption_counts = fitness_classes['is_closed_caption_shown'].value_counts()

# Display the count of videos with and without closed captions
print("Number of videos with closed captions:", closed_caption_counts[True])
print("Number of videos without closed captions:", closed_caption_counts[False])
```

```
Number of videos with closed captions: 30952
Number of videos without closed captions: 809
```

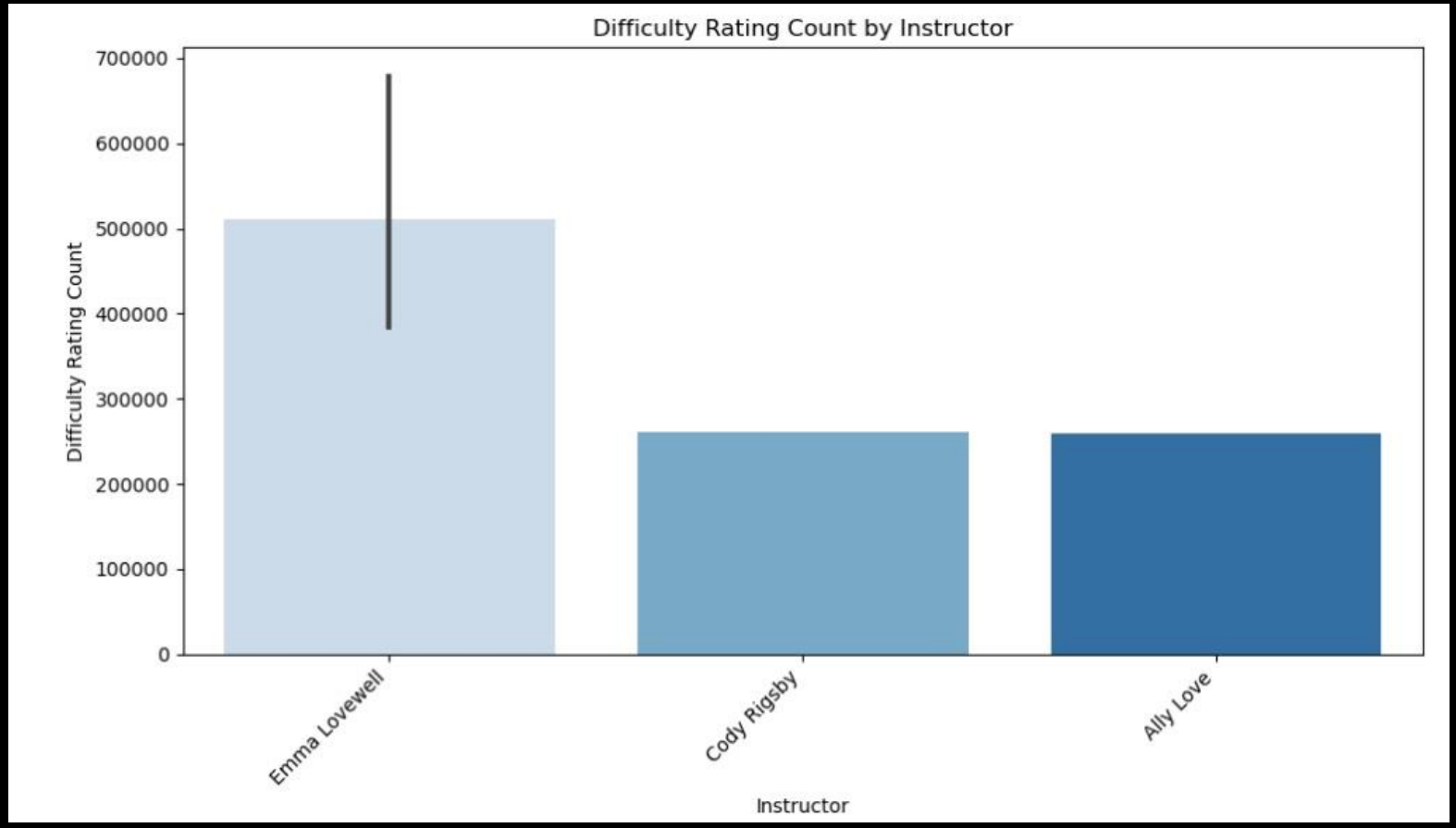


+ CLOSED CAPTIONS

Peloton is international, featuring many instructors from many different backgrounds. But their accessibility for Closed Captions is not great. 809 videos in their large library has closed caption. While 30,952 videos do not support it.

DIFFICULTY BY INSTRUCTOR

The most popular instructors are usually the ones with the most fun and intense classes, but according to the data, being the most popular and most featured doesn't mean having the hardest class. Emma Lovewell is the highest rated instructor according to users by class difficulty.

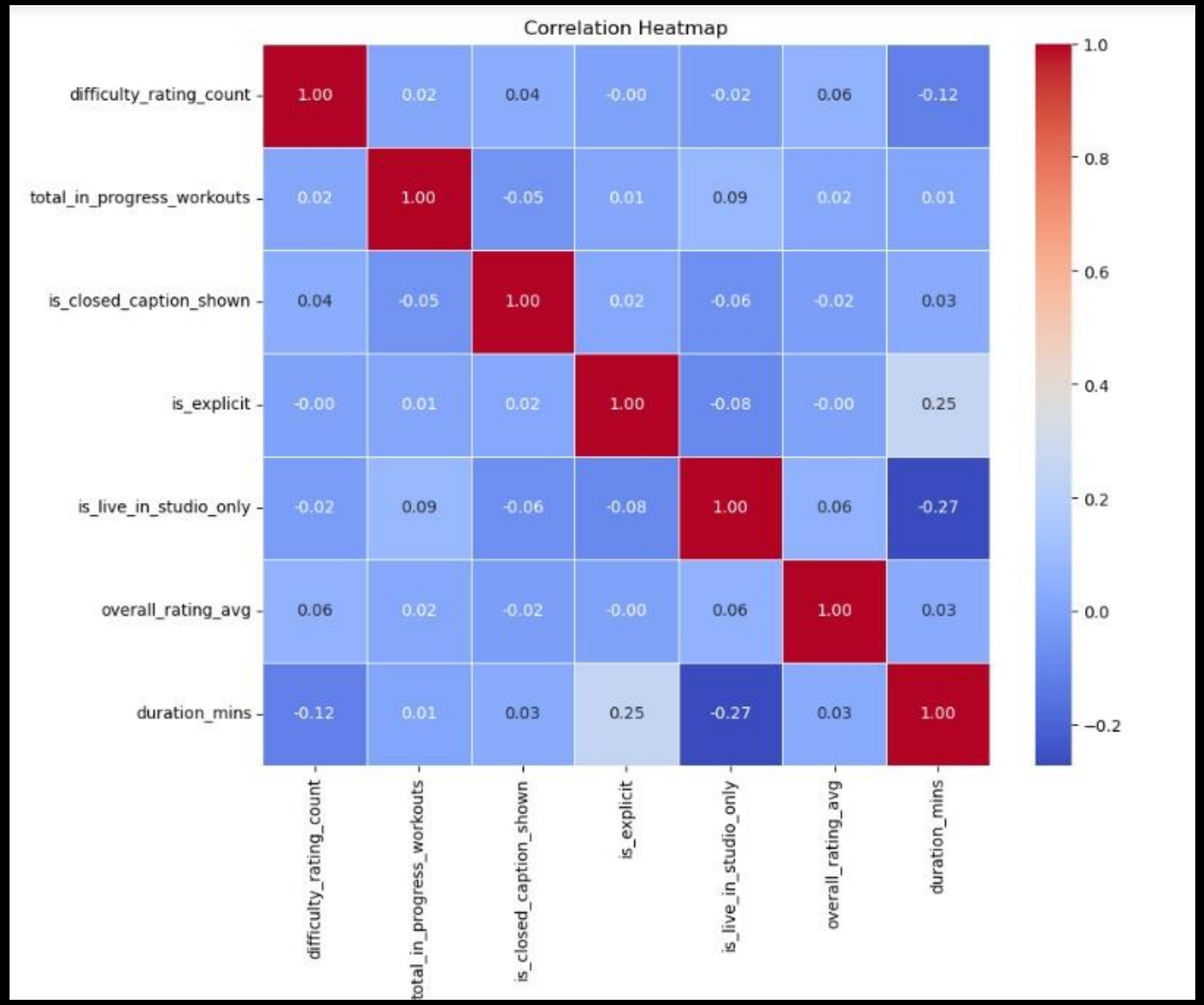


“

You didn’t get this far to only get this far.

BEN ALLDIS
PELOTON INSTRUCTOR

”

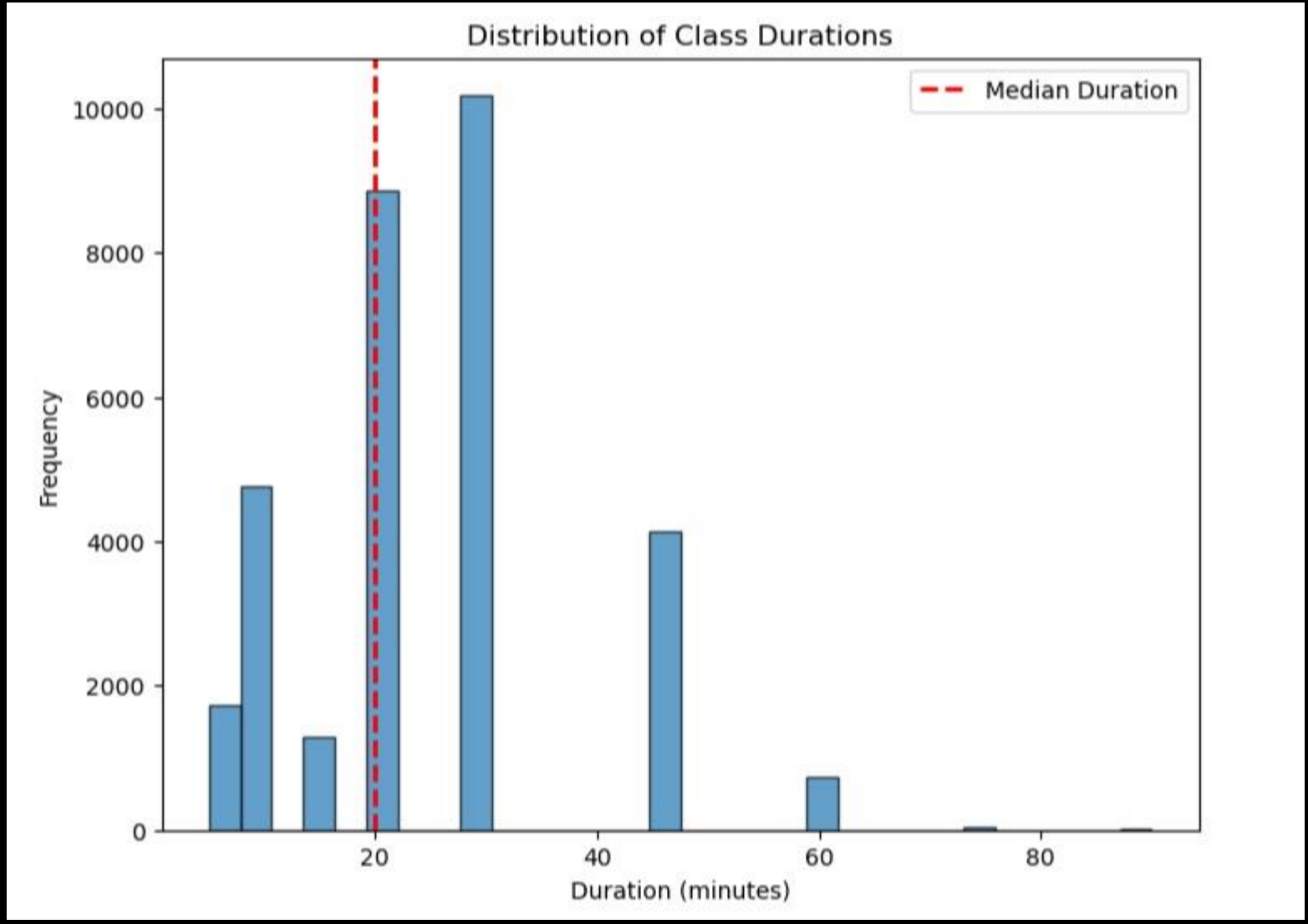


+CONNECTING THE DOTS

After digging for a while for interesting connections between users and content. I wanted to know if there was any correlation between any of the data points. It was interesting to find there was none or very little correlation between all the values.

CLASS DURATION DURATION

Surprisingly 30 Minutes, with a median best of 20 minutes is the most popular time format of all Peloton videos. This could be due to user preferences or styles. It would have been amazing to see user views for each video, but it was not available on the dataset.



SUMMARY

SUMMARY
SUMMARY

As a personal trainer seeking to create my own video platform, I draw inspiration from Peloton's success. Its innovative approach to digital fitness instruction has demonstrated the immense potential for reaching and inspiring individuals worldwide. By analyzing the content and trends on Peloton's platform, I gained valuable insights that will help me craft engaging and effective workouts for my own audience. There is so much more that can be analyzed in this dataset, but at the end this is just scratching the surface.

Thank you for joining me on this exploration of Peloton's Digital platform.

THANK YOU

JORGE SANCHEZ
JORGESAN@MYYAHOO.COM