WODWELL Workout Popularity Analysis

By Jake Getz

Data

Our data is scraped from wodwell.com using selenium. We have just over 2000 entries, each representing an unique workout with its own format and volume of movements.

Columns:

- Name

- HREF

- Likes

- Views

- 79 Unique workout movements

- 3 columns for workout format

- Total Volume of Movements

Workout Formats

We begin by plotting the three workout formats, EMOM, For Time, and AMRAP against likes. We can see a slight preference for workouts of the EMOM and For Time formats with AMRAPs being slightly less popular than the other types.
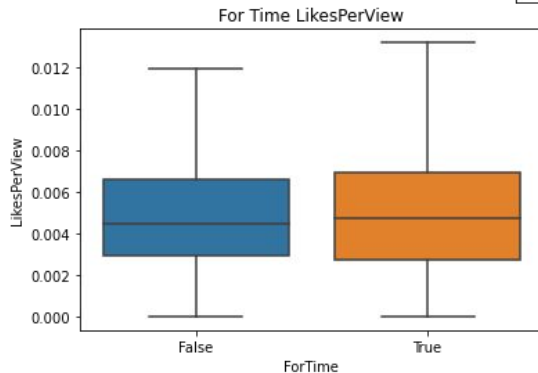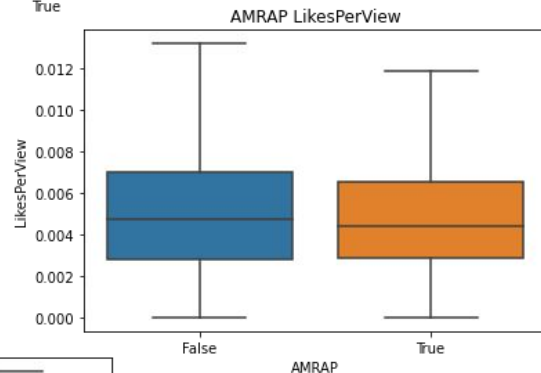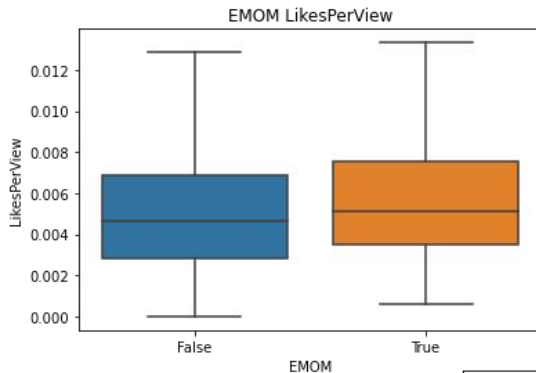
High Complexity Gymnastics

Movements that fall into the category of high skill gymnastics movements are less popular than workouts without those movements. It stands to reason that the higher the skill barrier, the less likely any given person is to be able to do the workout as written. Generally, people like workouts they can do.
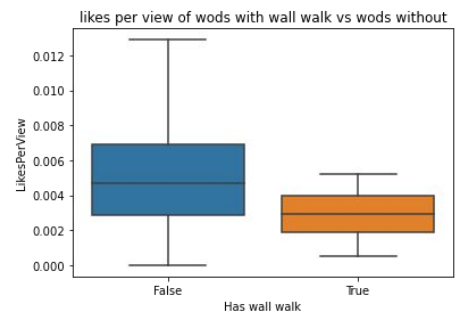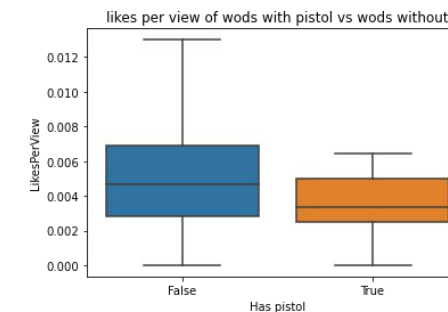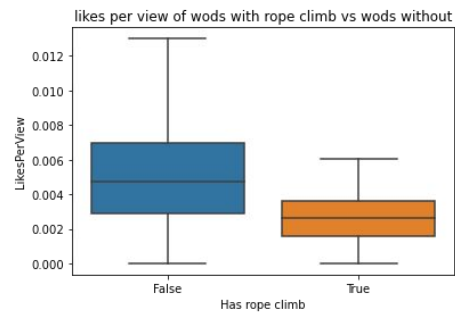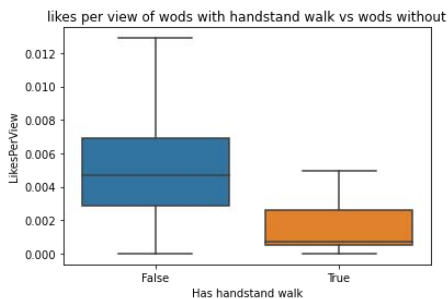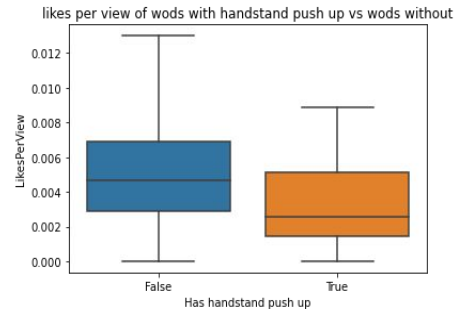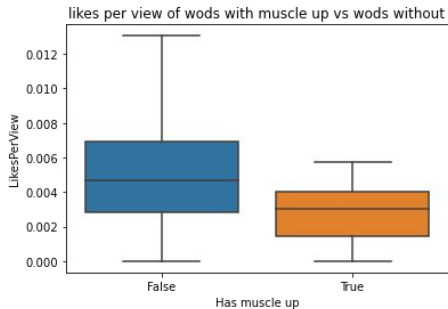
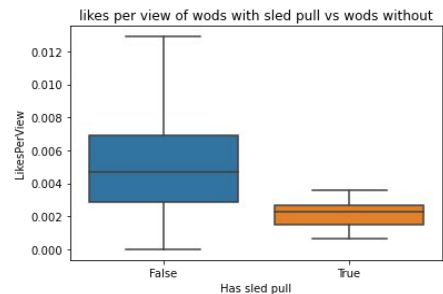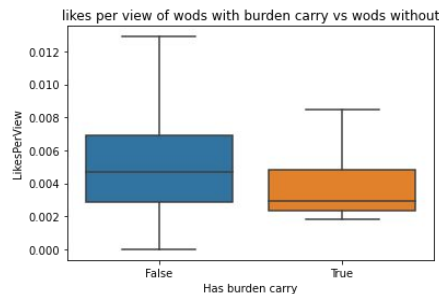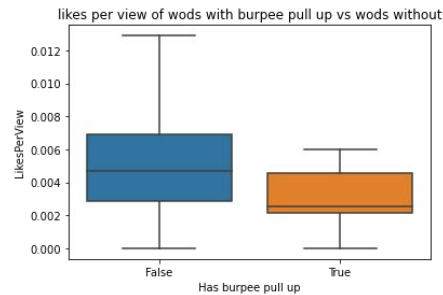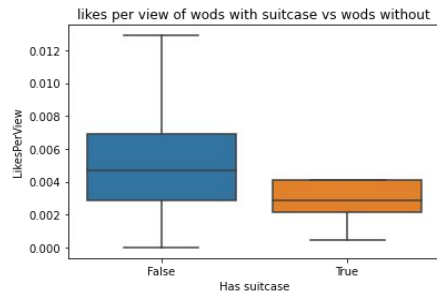# Grunt Work

"Grunt Work" are movements which require low skill and are often a test of pure strength or "grit" We see that generally people don't enjoy these movements and it becomes less likely that a workout will receive likes if it contains these grunt work movements.
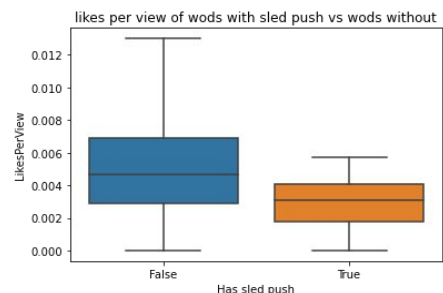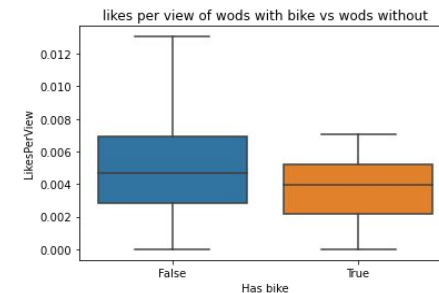
The Core

Anecdotally, abs are frequently sighted as the least liked part of any workout, yet our data shows something interesting. Core movements make workouts more popular than when they aren't present. People like to train their abs.
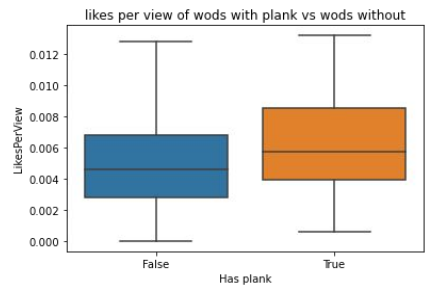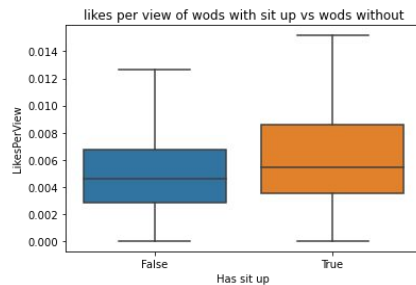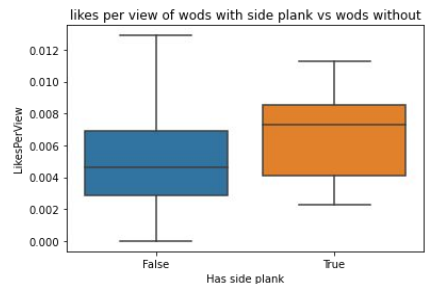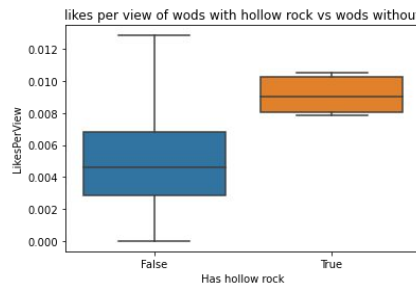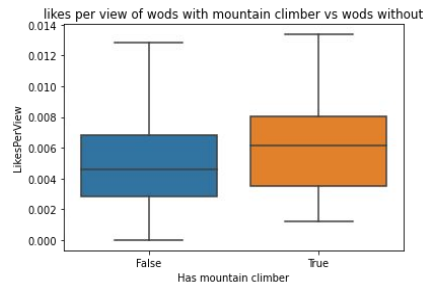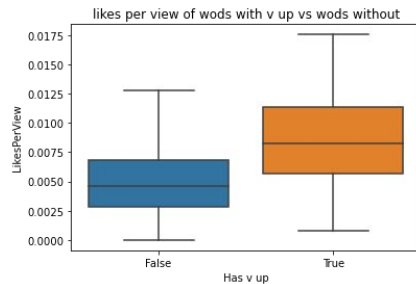
Not Your Average Joe.

We found that movements of a medium skill level also showed higher likes when present vs not present. Movements to which they are tricks or techniques to making faster and easier but that don't have a massive skill or strength prerequisite.
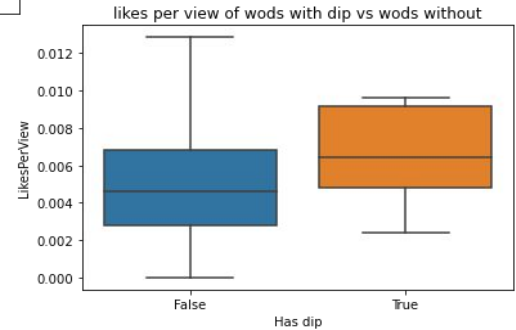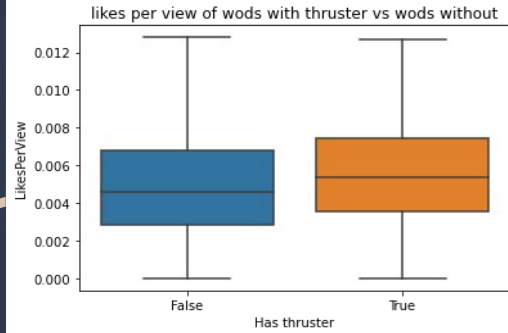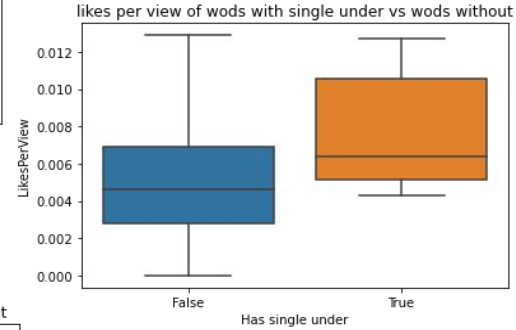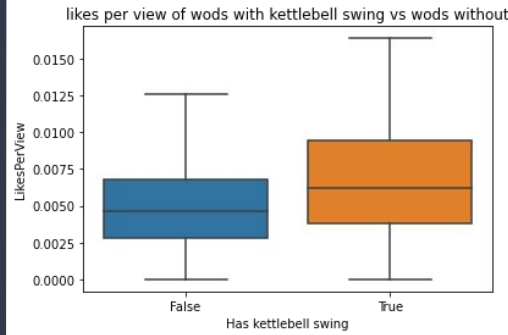
Predictive Modeling
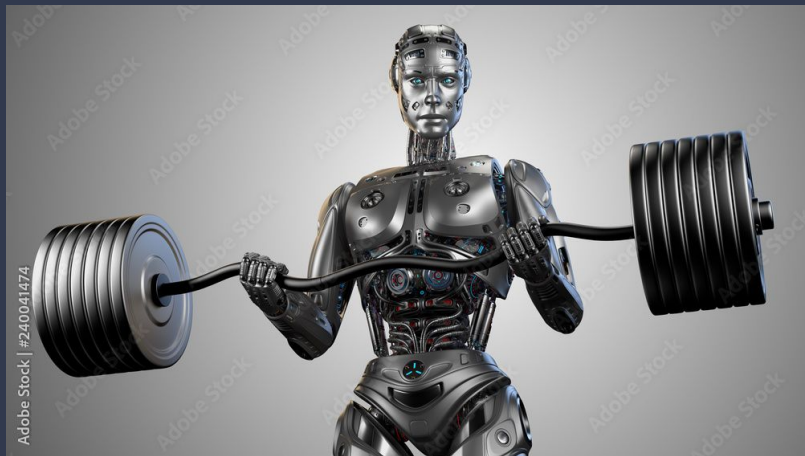
After hyperparameter tuning, Random Forest and XGBoost were our best performing models. Random Forest has a slightly better R-Squared than XGBoost so we choose to use Random Forest, even though on this particular test set, XGBoost has a very slightly better mean absolute error, 10 likes difference when dealing with thousands of likes is nearly inconsequential

	R-Squared	Mean Absolute Error
Random Forest	.52	118.35
XGBoost	.43	107.73

Conclusion

We scraped thousands of workouts from WODwell.com and broke them down into their component pieces. We analyzed workout popularity based on movement and workout format. We found that low and high skill movements tend to be less popular, while medium skill movements and abdominal work tend to lead to popular workouts. We applied machine learning to our dataset and created a model that will predict likes within an accuracy of around 100 likes. This model is a potential tool for gym owners and workout programmers to double check their intuition about what their athletes will enjoy with a data driven analysis based on what has been previously popular.