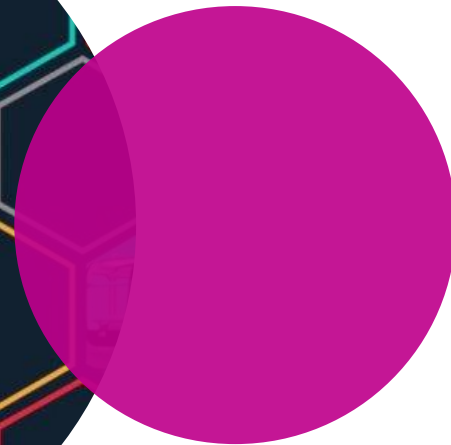
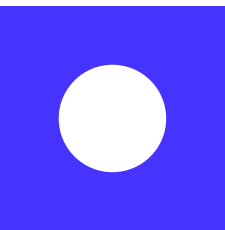


# *Predicting Engagement by Category*

General Assembly Part Time  
Data Science Final Project  
**Alex Rees**



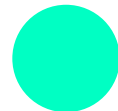


# AGENDA



## **Problem Statement**

What is the project objective?



## **Data Sources**

When, where, and how was the data collected?

## **Feature Engineering**

Creating the features to us in modeling and exploratory data analysis

## **Modeling**

Finding the model that works best in solving for the problem statement

## **Inference**

Final thoughts on the results

The background is a solid dark blue. In the center, there is a horizontal row of seven overlapping circles. The circles are a lighter shade of blue, and each subsequent circle from left to right is slightly more opaque, creating a sense of depth and movement.

# *PROBLEM STATEMENT*





Launch Forth is a co-creation platform on which companies launch rapid product design challenges. Launch Forth's community of 200k+ engineers and designers submit product designs and compete to win the challenge and have their design manufactured. Engagement rates within these projects are a primary selling point for new business. This project's objective is to explore if can we predict engagement rates based on a project's category?

**Hypothesis:**

Engagement can be predicted by category, and projects in the Ground Mobility category will have the most engagement.

# DATA SOURCES



All 6 datasets were pulled from the Launch Forth platform and represent a sample of all actions taken on the platform since October of 2016.

**Projects**

*List of all projects and respective categories*

**Watches**

*A 'follow' on any content other than a project*

**Follows**

*Follows specific to a project*

**Posts**

*A comment, which can be made on any piece of content or project.*

**Entries**

*Posted to challenges*

**Ideas**

*Posted to brainstorm*

# FEATURE ENGINEERING & EDA



*The engagement metrics were used to create the 'category\_activity\_mean' feature which is what I am predicting for*

cocreation_tool	project_id	user_id	content_type_name_entry	content_type_name_post	parent_content_type_name	title	categories	follow	watch
brainstorm	155	1439	0	1	project	Olli: self-driving, cognitive electric shuttle	Ground Mobility	1	0
challenge	165	75774	1	0	project	Urbanization of Mars	Mars	0	1
challenge	160	65782	1	1	project	Airbus Cargo Drone	Air Mobility	1	1

This was turned into 'total\_users' a total count of users per project

This was turned into 'post\_count', a total count of posts per project

This was turned into 'entry\_count', a total count of entries per project

This was dummy coded and all categories were given their own column.

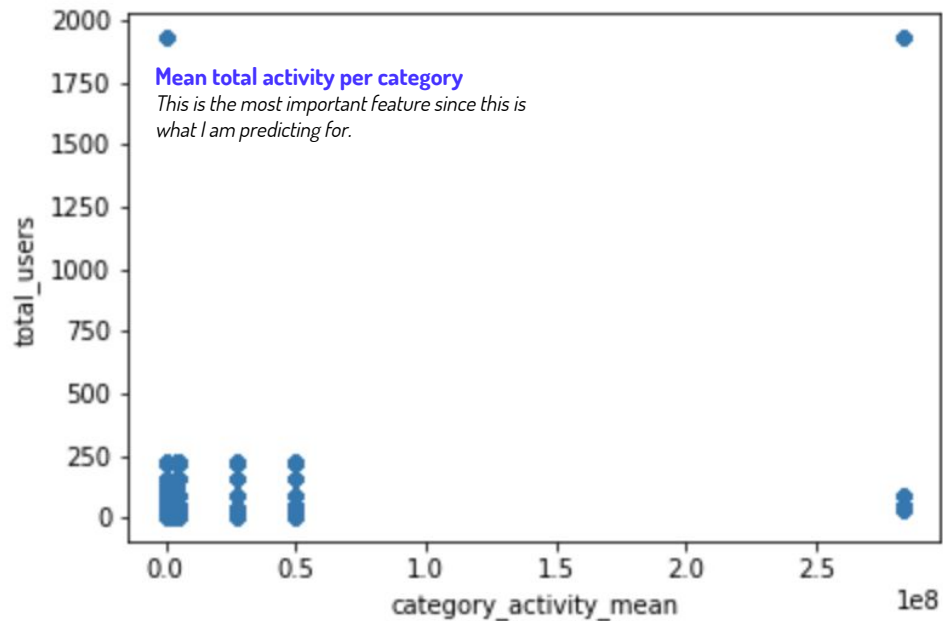
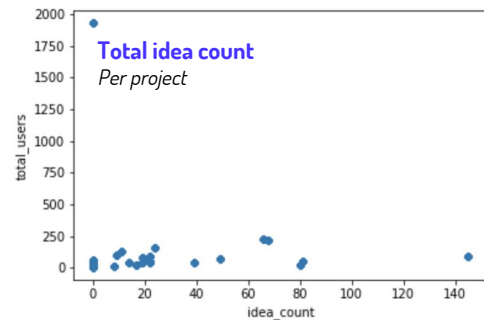
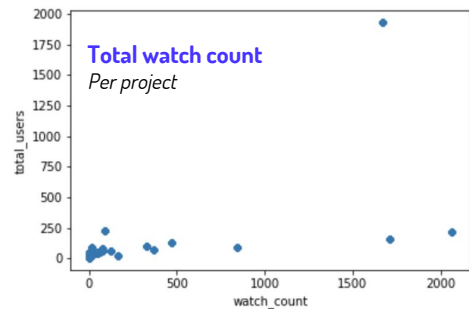
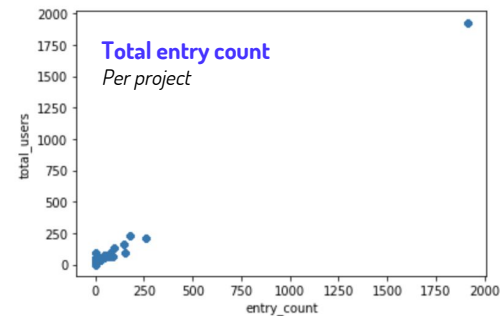
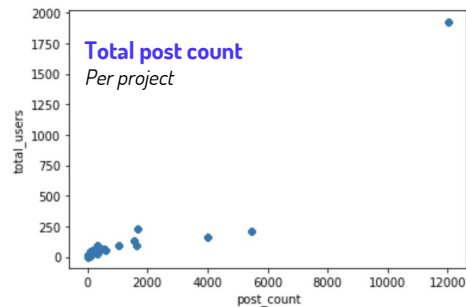
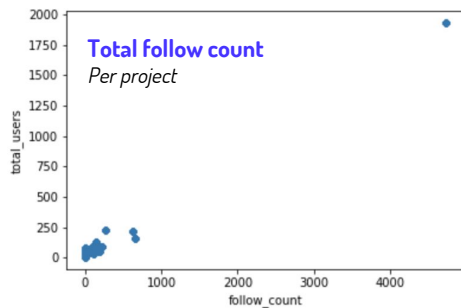
This was turned into 'follow\_count', a total count of follows per project

This was turned into 'watch\_count', a total count of watches per project



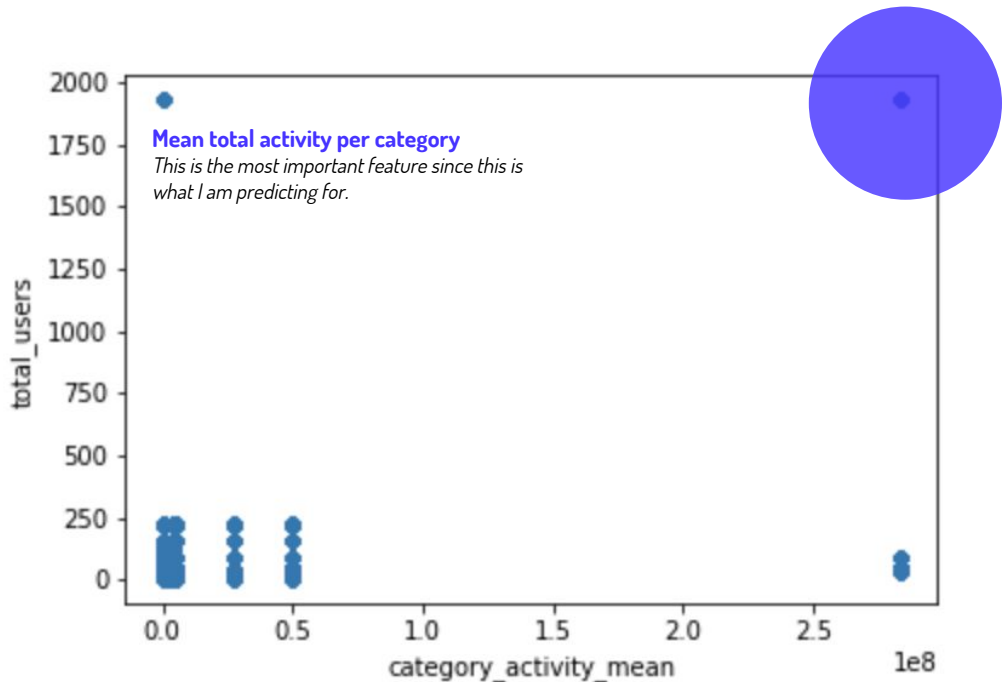


## PHASE 2: PLOTTING





## PHASE 2: PLOTTING



On closer inspection all of the outliers shown in the plots are from the projects in one of the categories: **Mars**.

Since I am trying to predict activity based on what category the project is in, I'm going to try to move forward without removing these outliers, so that I can keep that category as a part of my model.

# MODELING

- **Modeling** is the process of creating a simplified representation of a system or process.
- It involves identifying the key components and relationships within the system.
- Modeling can be used to analyze the behavior of a system, predict its future performance, and optimize its design.
- There are many different types of modeling, including mathematical modeling, simulation modeling, and conceptual modeling.
- Modeling is a critical tool for engineers, scientists, and business analysts.
- It allows them to understand complex systems and make informed decisions about their design and operation.
- Modeling can also be used to communicate the results of research and development to stakeholders.
- By creating a visual representation of a system, modeling makes it easier to understand and discuss the system's behavior.
- Modeling is an essential part of the design and development process for many types of systems.
- It helps to identify potential problems and optimize the system's performance before it is built.
- Modeling can also be used to test the system's performance under different conditions.
- This helps to ensure that the system will perform as expected in the real world.
- Modeling is a powerful tool that can be used in a wide range of applications.
- It is a key part of the design and development process for many types of systems.
- Modeling can help to improve the quality and performance of a system.
- It can also help to reduce the cost and risk of developing a new system.
- Modeling is a critical tool for anyone involved in the design and development of complex systems.
- It is a skill that is essential for success in many fields.
- Modeling is a process that involves creating a simplified representation of a system or process.
- It is a tool that can be used to analyze the behavior of a system, predict its future performance, and optimize its design.
- Modeling is a critical part of the design and development process for many types of systems.
- It helps to identify potential problems and optimize the system's performance before it is built.
- Modeling can also be used to test the system's performance under different conditions.
- This helps to ensure that the system will perform as expected in the real world.
- Modeling is a powerful tool that can be used in a wide range of applications.
- It is a key part of the design and development process for many types of systems.
- Modeling can help to improve the quality and performance of a system.
- It can also help to reduce the cost and risk of developing a new system.
- Modeling is a critical tool for anyone involved in the design and development of complex systems.
- It is a skill that is essential for success in many fields.

# Modeling Round 1

## BIAS VS. VARIANCE

- The training and test model MSE are not so different that there is a concern of variance.
- The difference between test and null model MSE show that bias is relatively low and not a concern.

## ANALYSIS

- Other than the particularly horrible linear regression w/ cross val score, the MSEs seem to be consistent across all models with little to no improvement.
- Though I'm predicting numbers in the millions, this MSE seems overly large.

LINEAR REGRESSION W/ TRAIN TEST SPLIT	LINEAR REGRESSION W/ CROSS VAL SCORE	REGRESSION TREE	REGRESSION TREE W/ CROSS VAL SCORE	BAGGING W/ REGRESSION TREE
Training model MSE: 4370715774996228.0	MSE: 9.949696157450484e+32	MSE: 4277976634802103.0	MSE: 4350657404988150.5	MSE: 4276344882915948.5
Test model MSE: 4276341366157285.0	Wow that is REALLY bad!			
Null model MSE: 4633470451736635.0				



# Modeling Round 2

The last round of modeling had suspiciously high MSEs, however after removing the outlier category: Mars, these MSEs are much more acceptable.

## ANALYSIS

- Similar to the last round of modeling bias and variance do not seem to be a cause for concern.
- These MSE's are half of the last round of modeling and far more acceptable.

### LINEAR REGRESSION W/ TRAIN TEST SPLIT

Training model MSE:  
223011431199581.4

Test model MSE:  
222063551042432.66

Null model MSE:  
234553939504031.4

### REGRESSION TREE W/ CROSS VAL SCORE

MSE:  
222062762501317.06



NEXT


STEPS



## FINAL THOUGHTS

I am content with the MSE on the second round of modeling but the signals within the data were not as strong as I had expected. There are additional steps that can be taken to confirm and validate the conclusion.

## NEXT STEPS

- Use the full data set (3+ million rows) rather than a sample
  - Only use categories that have higher engagement metrics as predictors to measure how this changes the conclusion.
  - Run this model on singular engagement metrics such as entries, ideas, posts and follows to get exact predicted numbers for each of these action types.
- 





*THANK  
YOU*