# TANZANIA WELLS

Blake Tolman

# OBJECTIVE/DATA

- Using data from Taarifa and the Tanzanian Ministry of Water, predict which pumps are functional, which need some repairs, and which don't work at all

The labels in this dataset

Feature data example

For example, a single row in the dataset might have these values:

| | |
|---|---|
| amount_tsh | 300.0 |
| date_recorded | 2013-02-26 |
| funder | Germany Republi |
| gps_height | 1335 |
| installer | CES |
| longitude | 37.2029845 |
| latitude | -3.22870286 |
| wpt_name | Kwaa Hassan Ismail |
| num_private | 0 |
| basin | Pangani |
| subvillage | Bwani |
| region | Kilimanjaro |
| region_code | 3 |
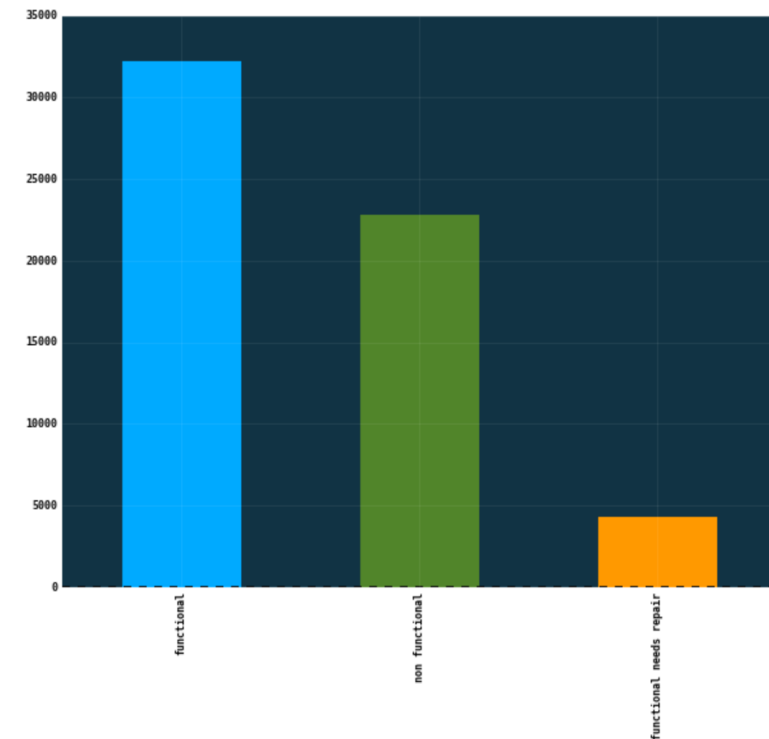
# MAPPING THE DATA
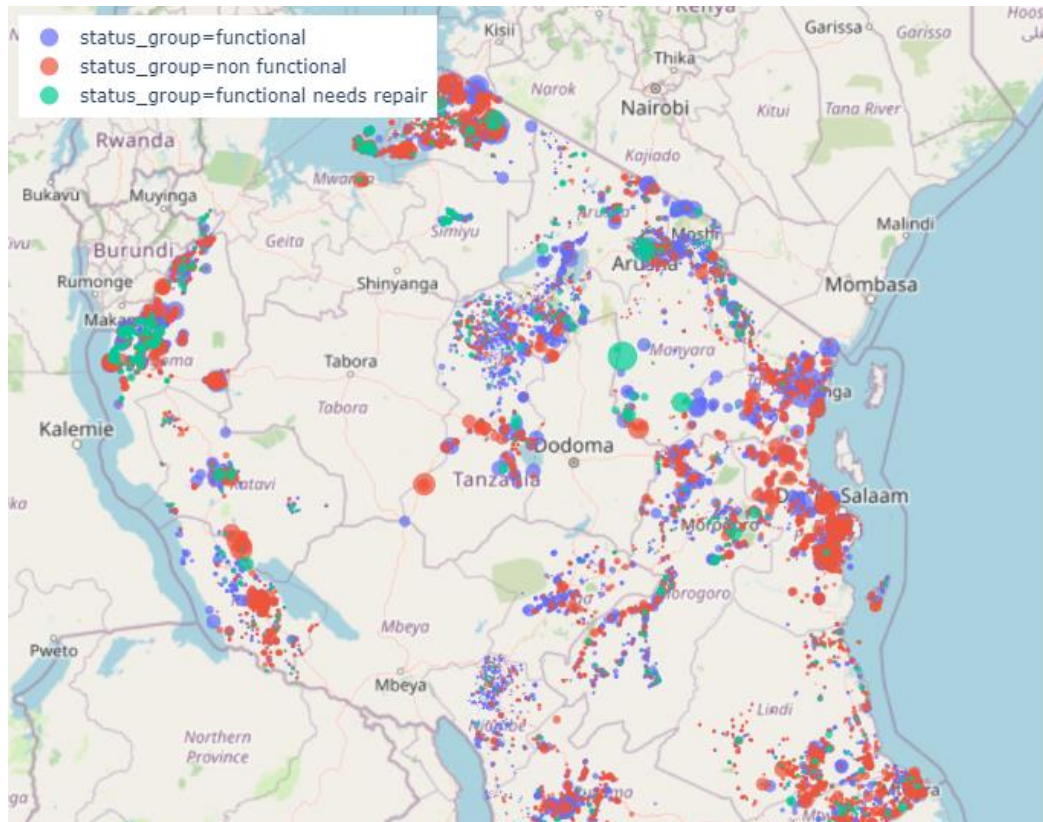


- There are higher populations around the edges of the country along with a high density of non-functional wells

- Central Tanzania has the most working wells and with lower average populations

- Functional wells that need repair appear to be randomly located

Number of Wells per Funder
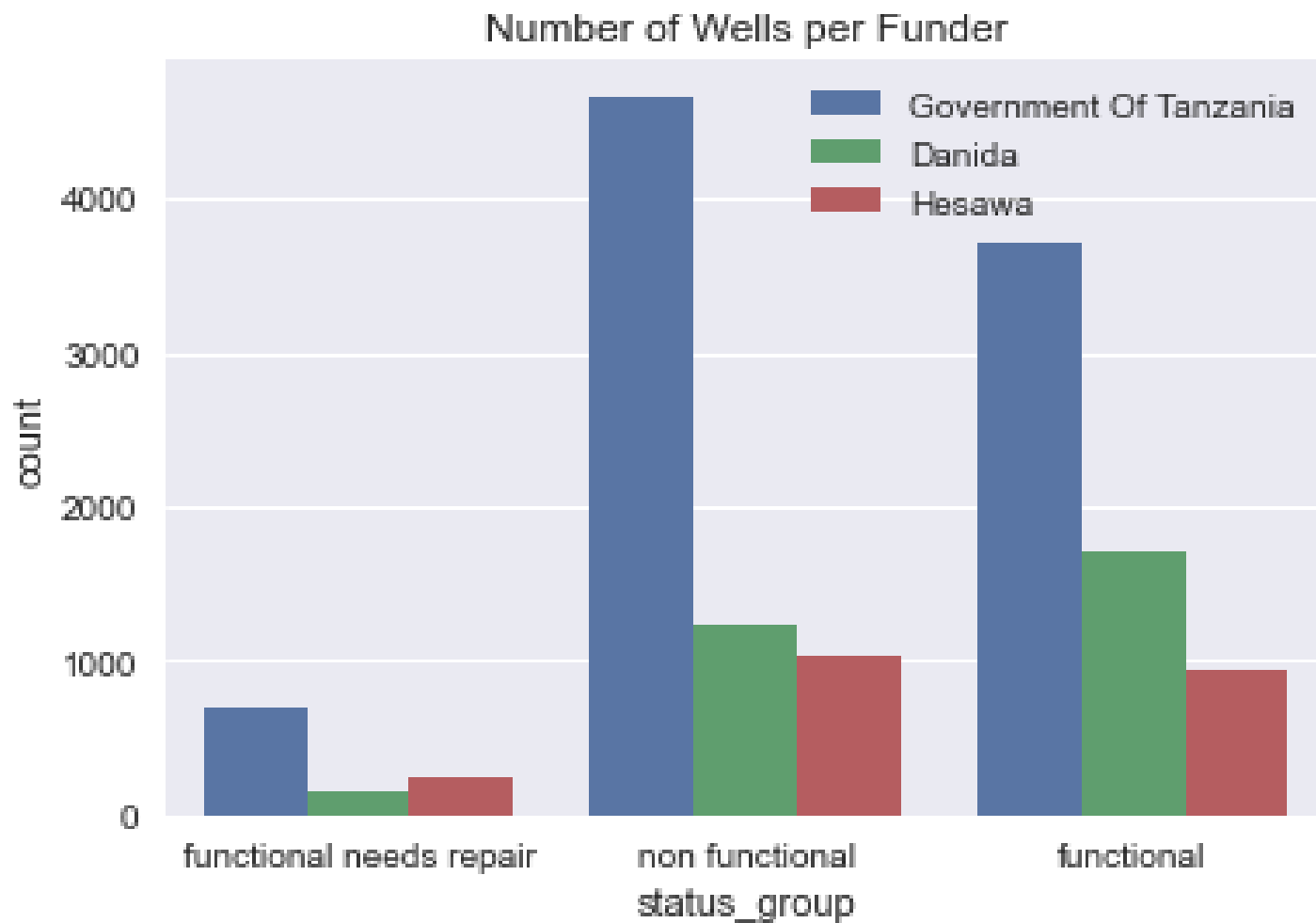
# FUNDER

- Danida is the only funder to fund more functional wells than non-functional

- When factoring in wells that need repair, the Government and Hesawa are barley producing more functional wells than not

- This could signify the money is going to quantity over quality

Number of Wells per Installer
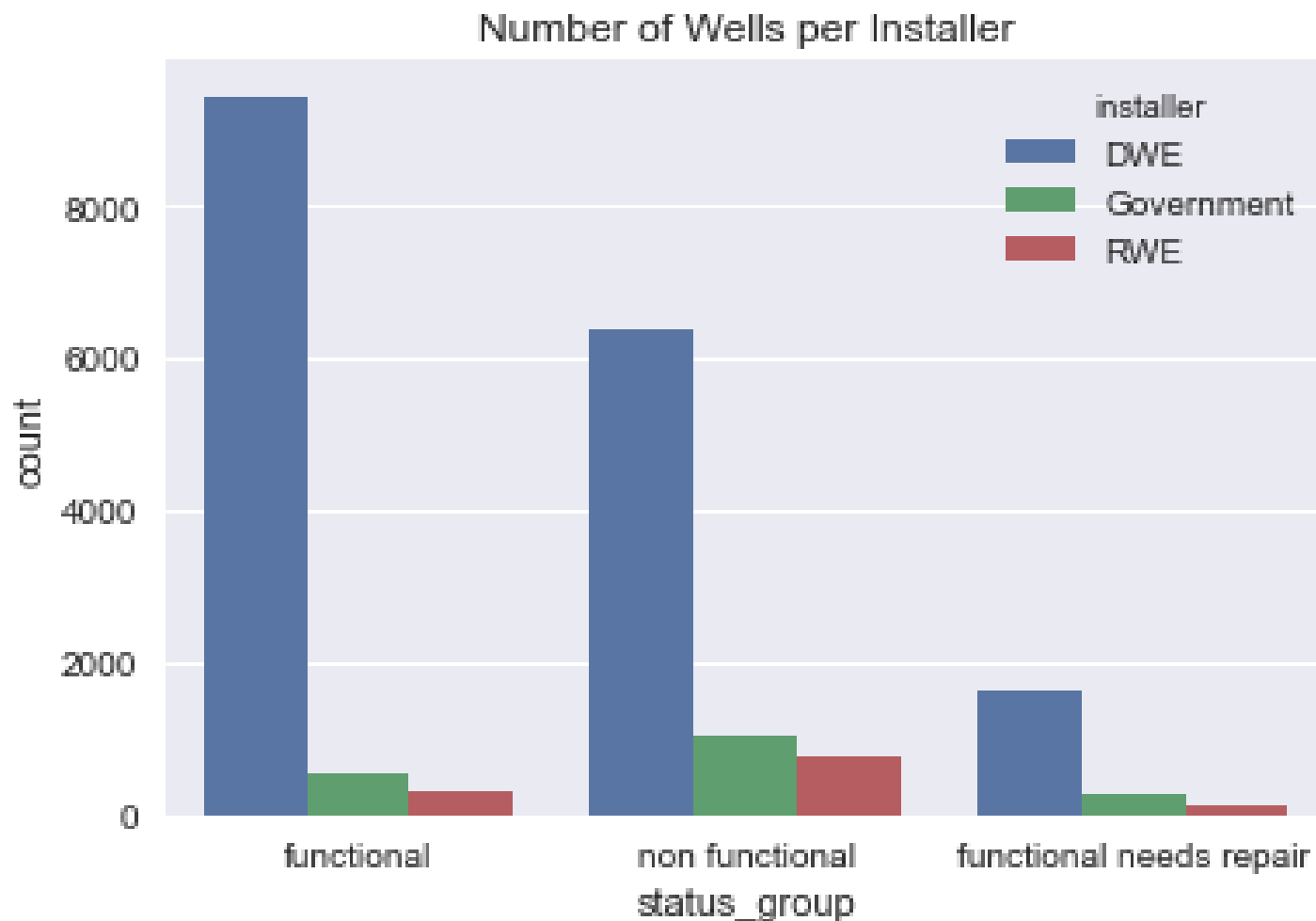
# INSTALLER

- Of the top 3 installers used DWE is the only one to produce more than 50% of their wells as functional

- DWE is the most reliable installer

- The Government, who is also a funder, again is just barely producing equivalent amounts of functional and non-functional well when factoring in the wells that need repair

Number of Well per Management Group
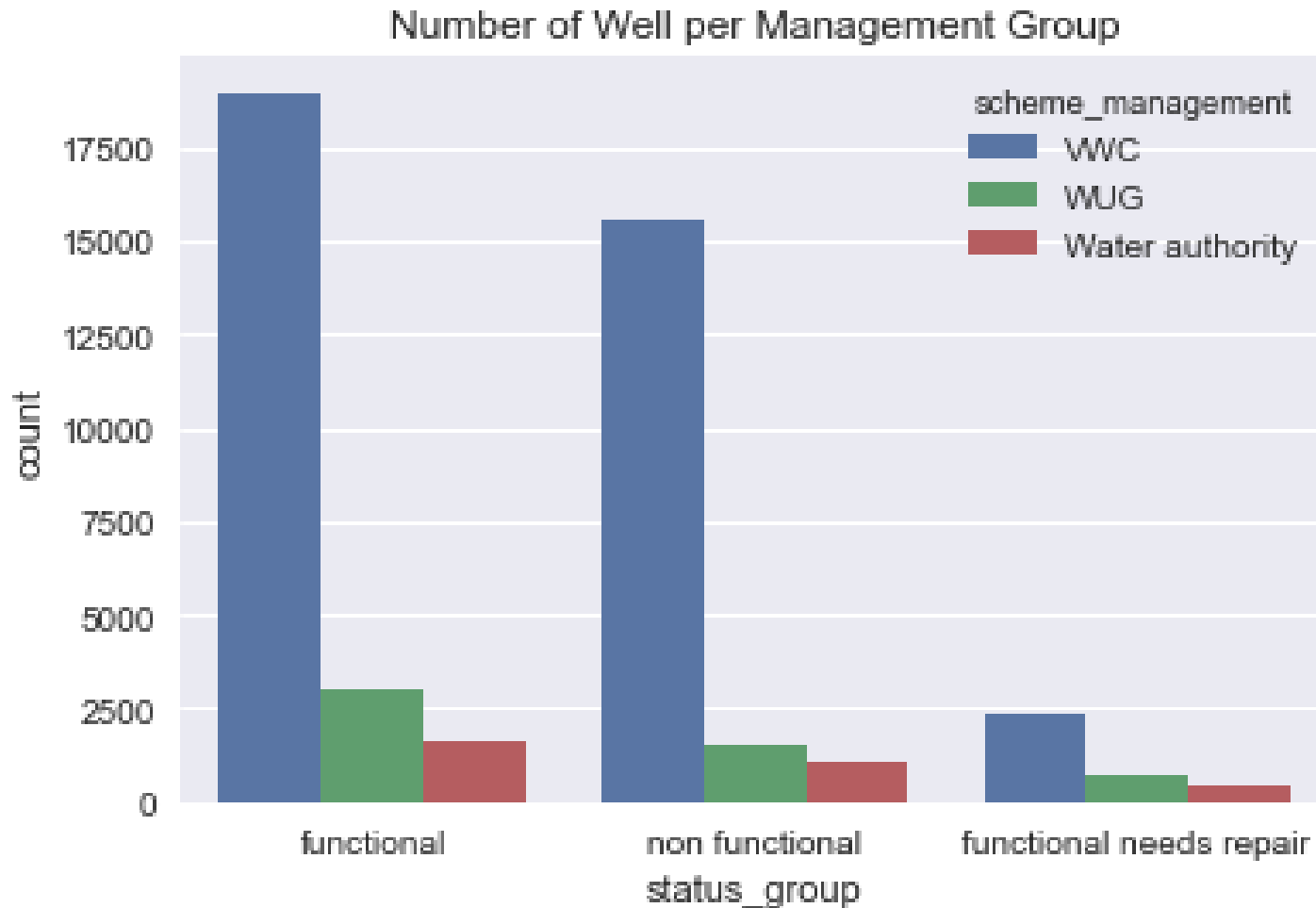
# MANAGEMENT

- This chart of popular management groups is the first to indicate more functional wells are being observed than not

- It appears the management group is most responsible for the success of the well.

- VWC manages the most wells by a large amount with no indication why

```
                        precision    recall   f1-score   support

      non functional        0.75       0.75       0.75       5664
          functional        0.78       0.80       0.79       8006
functional needs repair     0.35       0.30       0.32       1088

            accuracy                              0.74      14758
           macro avg        0.63       0.61       0.62      14758
        weighted avg        0.74       0.74       0.74      14758
```

AUC is :0.7468304703517535
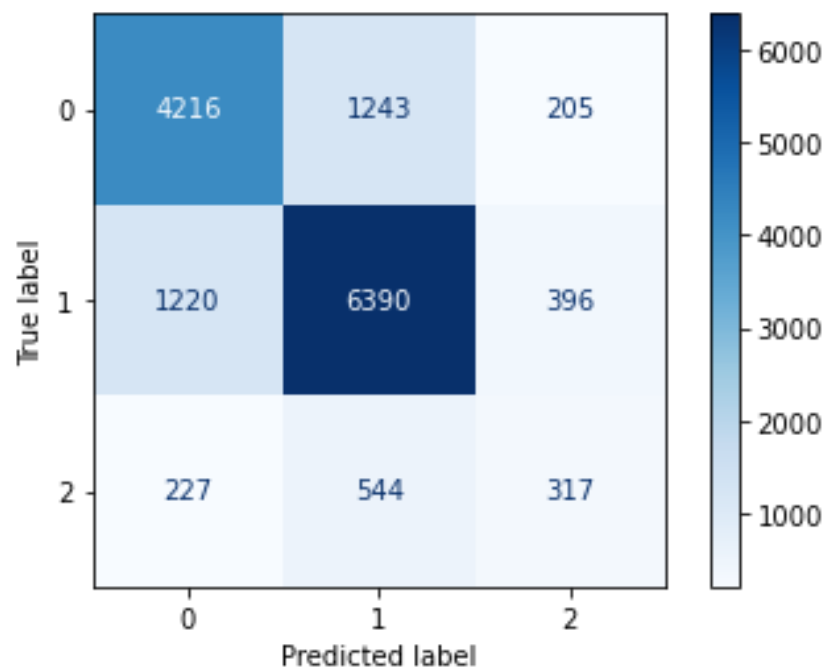
# INITIAL MODEL – DECISION TREE

- Models Sampled:
  - Logistic Regression
  - Decision Tree
  - XGBoost
  - K-Nearest Neigbor

- First model performance struggled with categorizing "functional needs repair"

- Functional and non-functional had good performance for an initial model

- The overall model had a 74% accuracy and a macro average of 0.62

```
                        precision   recall   f1-score   support

     non functional        0.85      0.72      0.78       5664
         functional        0.77      0.92      0.84       8006
functional needs repair    0.59      0.23      0.33       1088

           accuracy                            0.79      14758
          macro avg        0.74      0.62      0.65      14758
       weighted avg        0.79      0.79      0.78      14758
```

AUC is :0.7468304703517535
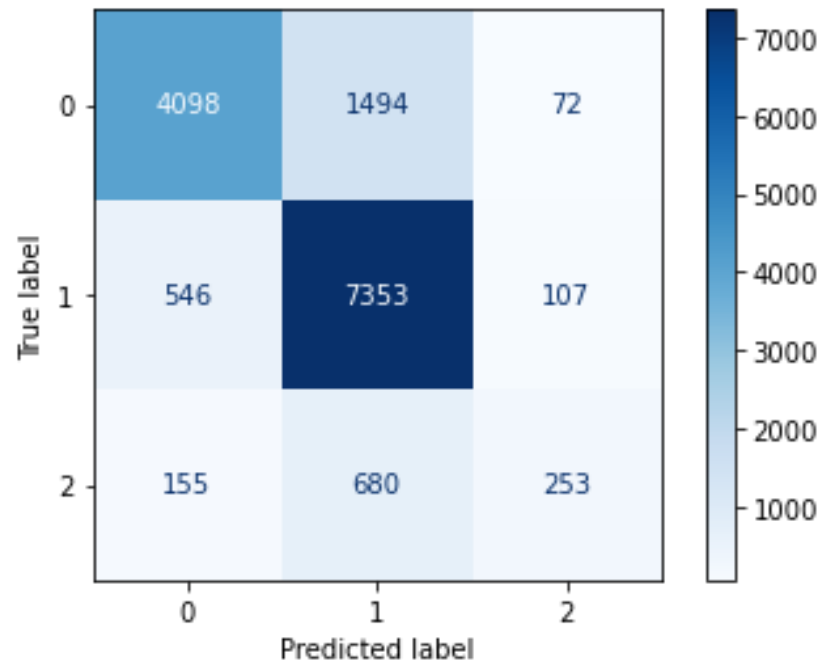
# FINAL MODEL – GRID SEARCH RANDOM FOREST

- Iterations:
  - Grid Search
  - Random Forest
  - Oversample
  - Undersample

- Grid search had the highest increase in performance

- F1-score for "functional need repair" had a minor increase but still had poor performance

- The new mode increased its accuracy by 5% to 79% and increased macro f1-score 0.03 for a final of 0.65

# RECOMMENDATIONS AND FUTURE WORK

- Recommendations:

1. Raise more money per well

2. Reallocate how the money is being spend on the well

3. Focus on quality over quantity

- Future Work:

1. Analyze the amount of money spent per well

2. Find the number of daily uses per well and its status

3. Look at the list of materials used for each status group of wells

# THANK YOU