

---

# PUMP IT UP: DATA MINING THE WATER TABLE

Chris Delacruz and David Tian

---

# TABLE OF CONTENTS

01 BUSINESS OVERVIEW

02 DATA OVERVIEW

03 MODEL RESULTS

04 NOTABLE FEATURES

The background is a solid teal color. In the center-left area, there is a large, light blue, abstract, organic shape that resembles a splash or a cloud. This shape is semi-transparent and contains the text.

# 01

## **BUSINESS OVERVIEW**

# Business Overview

Who are we? Tanzanian government agency.

What is the problem? Are we able to determine a well's function status and better allocated our resources?

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

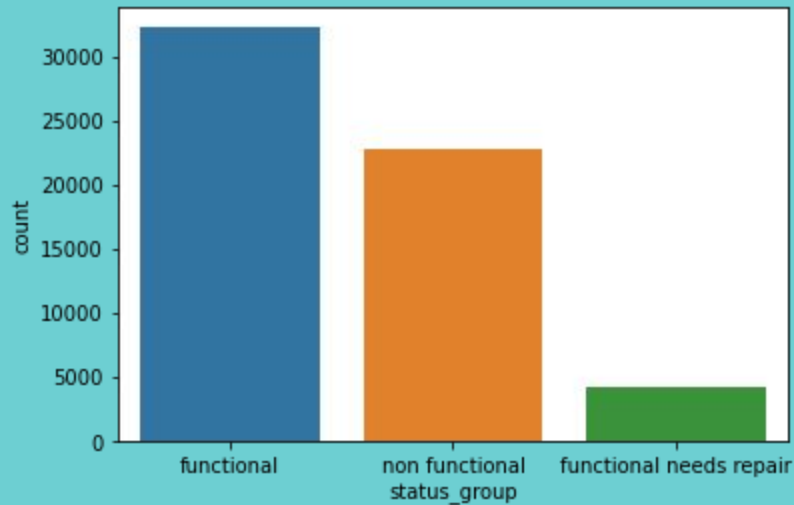


The background is a solid teal color. In the center-left area, there is a large, light blue, abstract, organic shape that resembles a cloud or a splash. This shape is semi-transparent and contains the text.

**02**

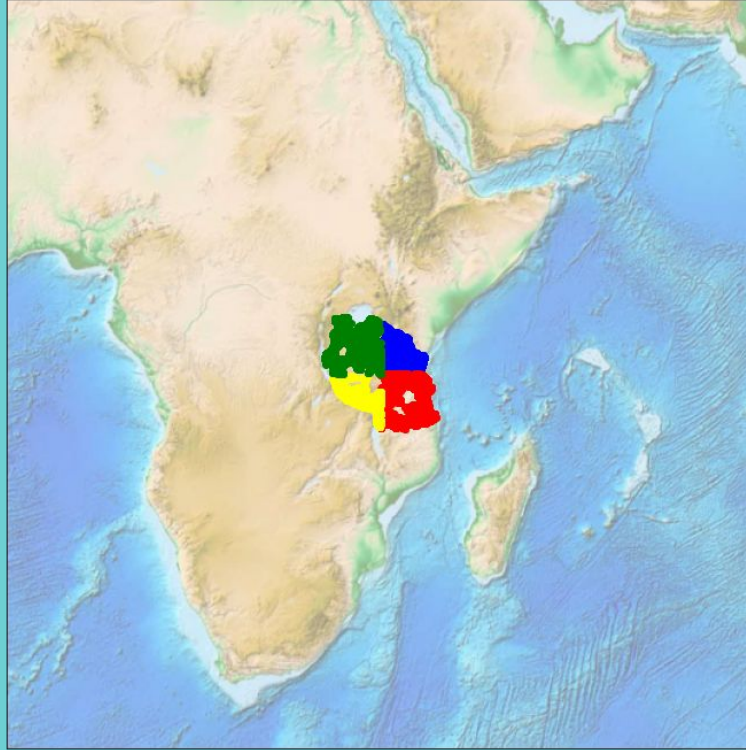
**DATA OVERVIEW**

# Data Overview - Status Group

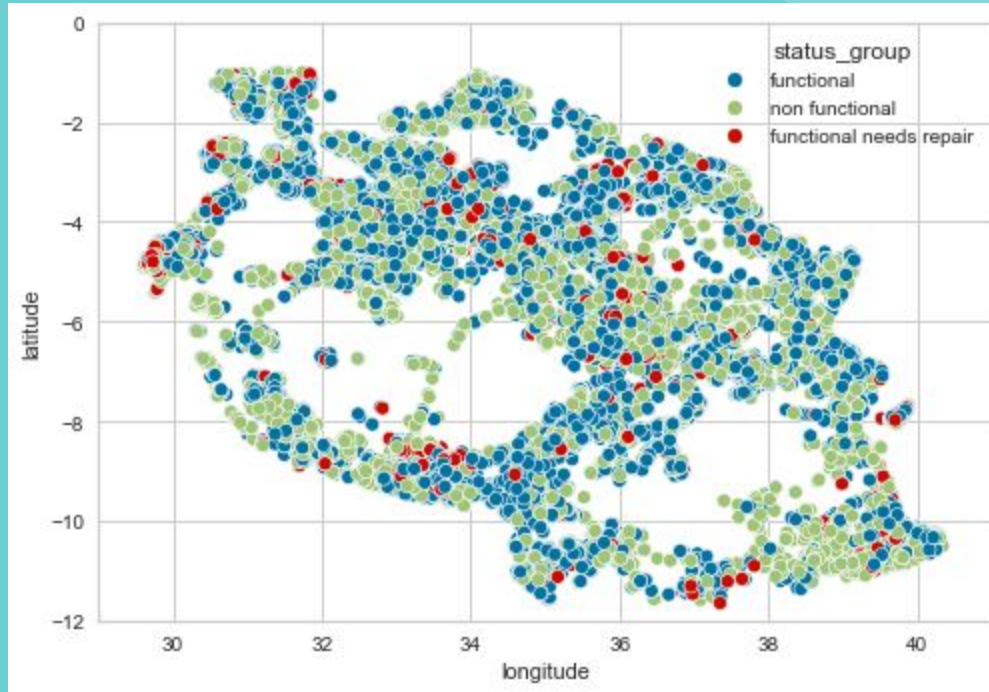


functional	32216
non functional	22798
functional needs repair	4313

# Data Overview - Geography



# Data Overview - Geography







03

MODEL RESULTS

# Model Results - Summary

<u>Model</u>	<u>Training Score</u>	<u>Test Score</u>
Logistic Regression	74.9%	75.0%
KNN	73.5%	74.3%
Decision Tree	74.8%	75.7%
Random Forest	79.2%	79.9%

# Model Results - Summary

<u>Model</u>	<u>Training Score</u>	<u>Test Score</u>
Logistic Regression	74.9%	75.0%
KNN	73.5%	74.3%
Decision Tree	74.8%	75.7%
Random Forest	79.2%	79.9%

## Random Forest Top Variables

```
{'gps_height': 0.12147513276145377,  
'population': 0.07081780215194135,  
'construction_year': 0.05956039833179953,  
'age_well': 0.0583722959915368,
```

## Logistic Regression Top Variables

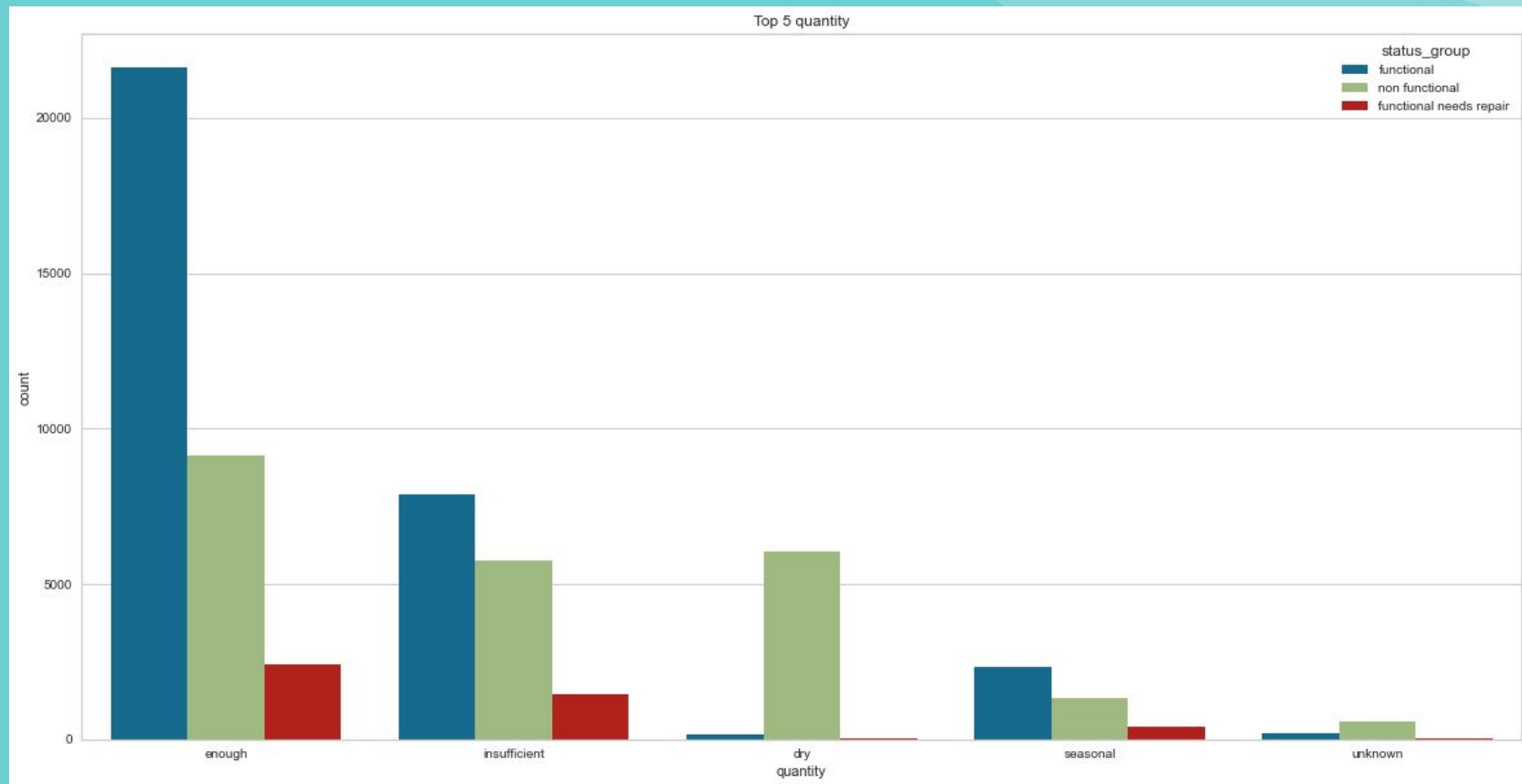
```
[('quantity_dry', -1.5351615146475501),  
( 'waterpoint_type_other', -0.6975144308388949),  
( 'quantity_enough', 0.6260479195561773),
```

The background is a solid teal color. In the center-left area, there is a large, light blue, abstract, organic shape that resembles a splash or a cloud. It has several overlapping, rounded edges. Centered within this shape is the text '04' in a large, white, sans-serif font.

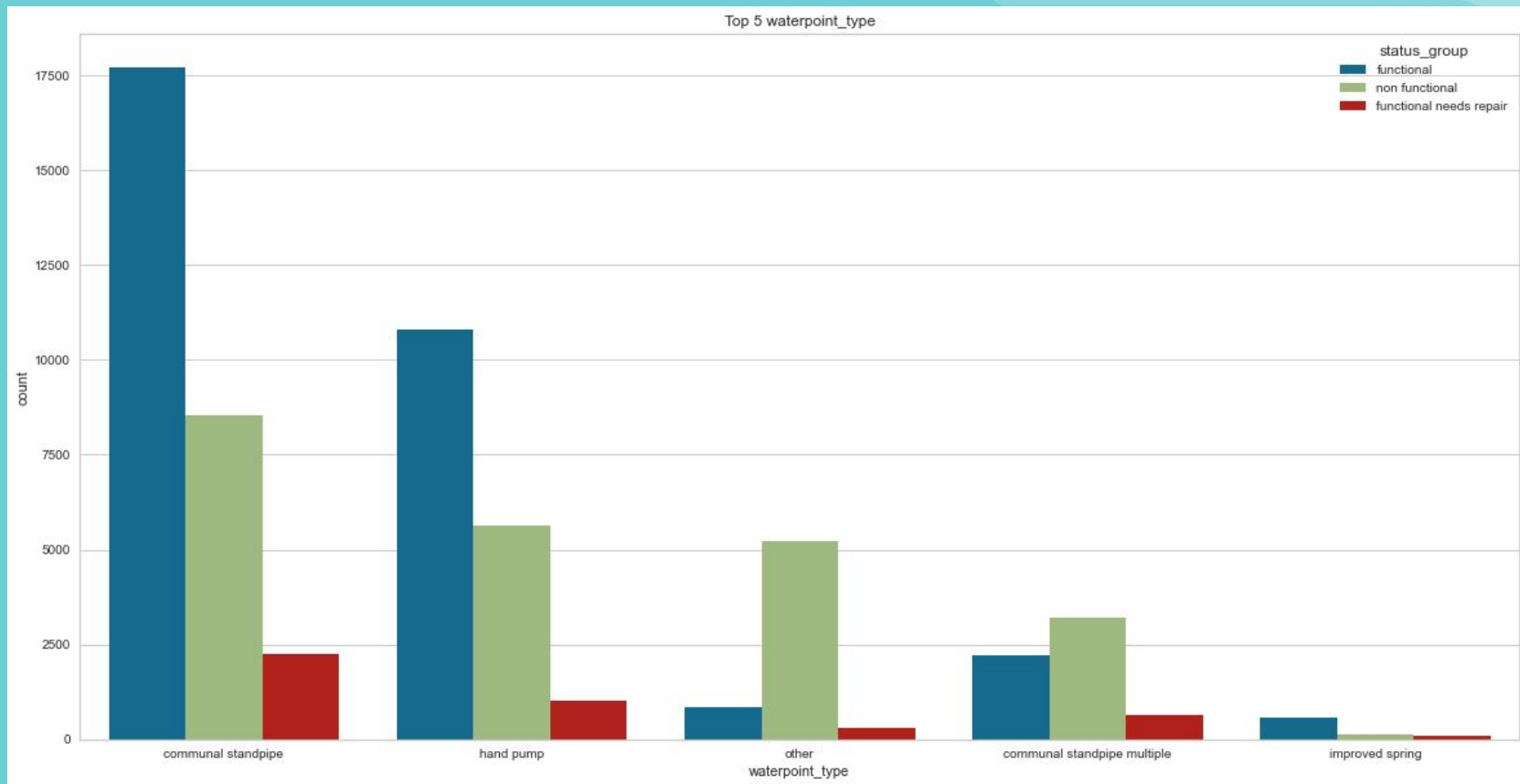
04

**NOTABLE FEATURES**

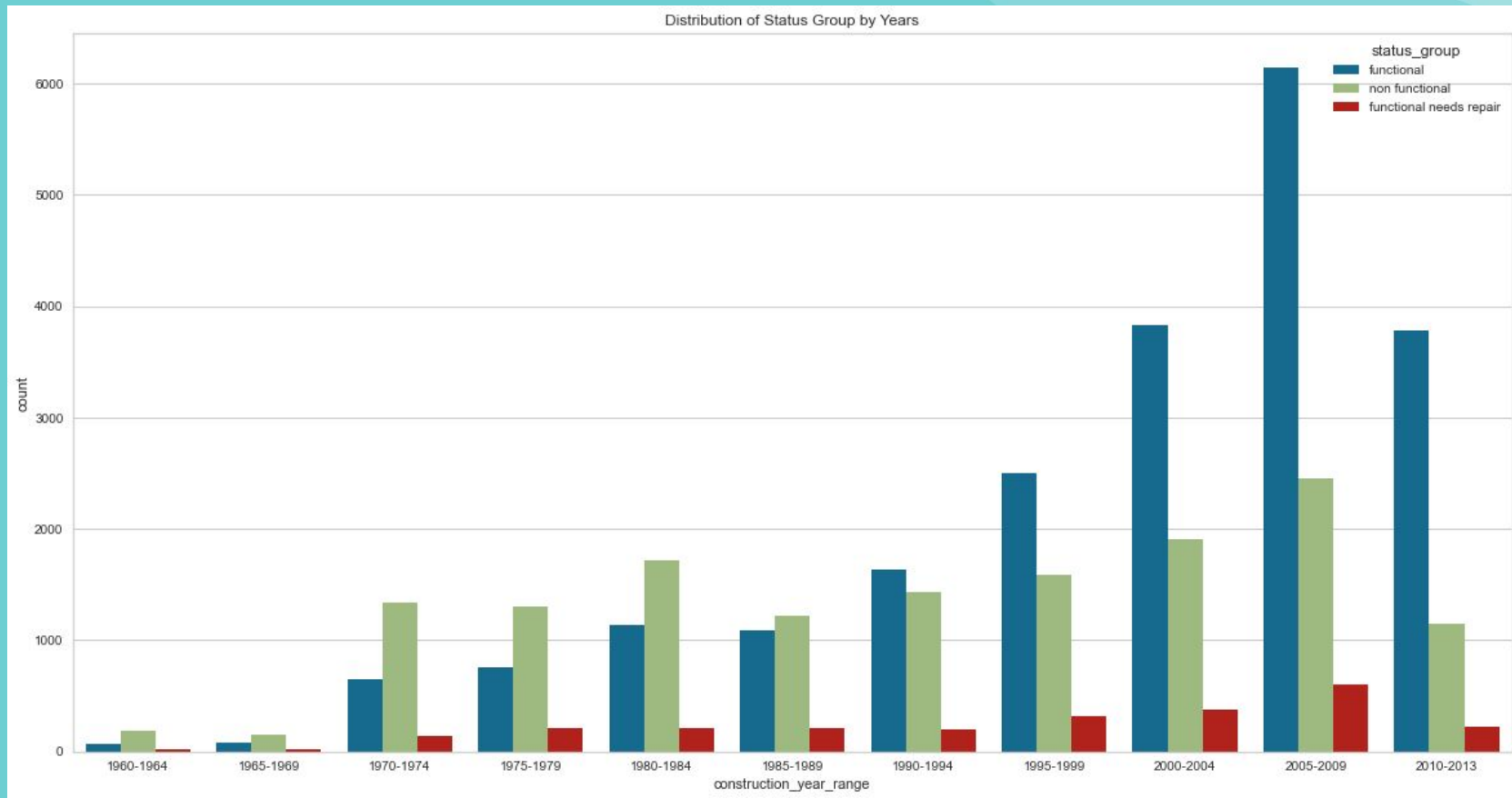
# Data Overview - Quantity



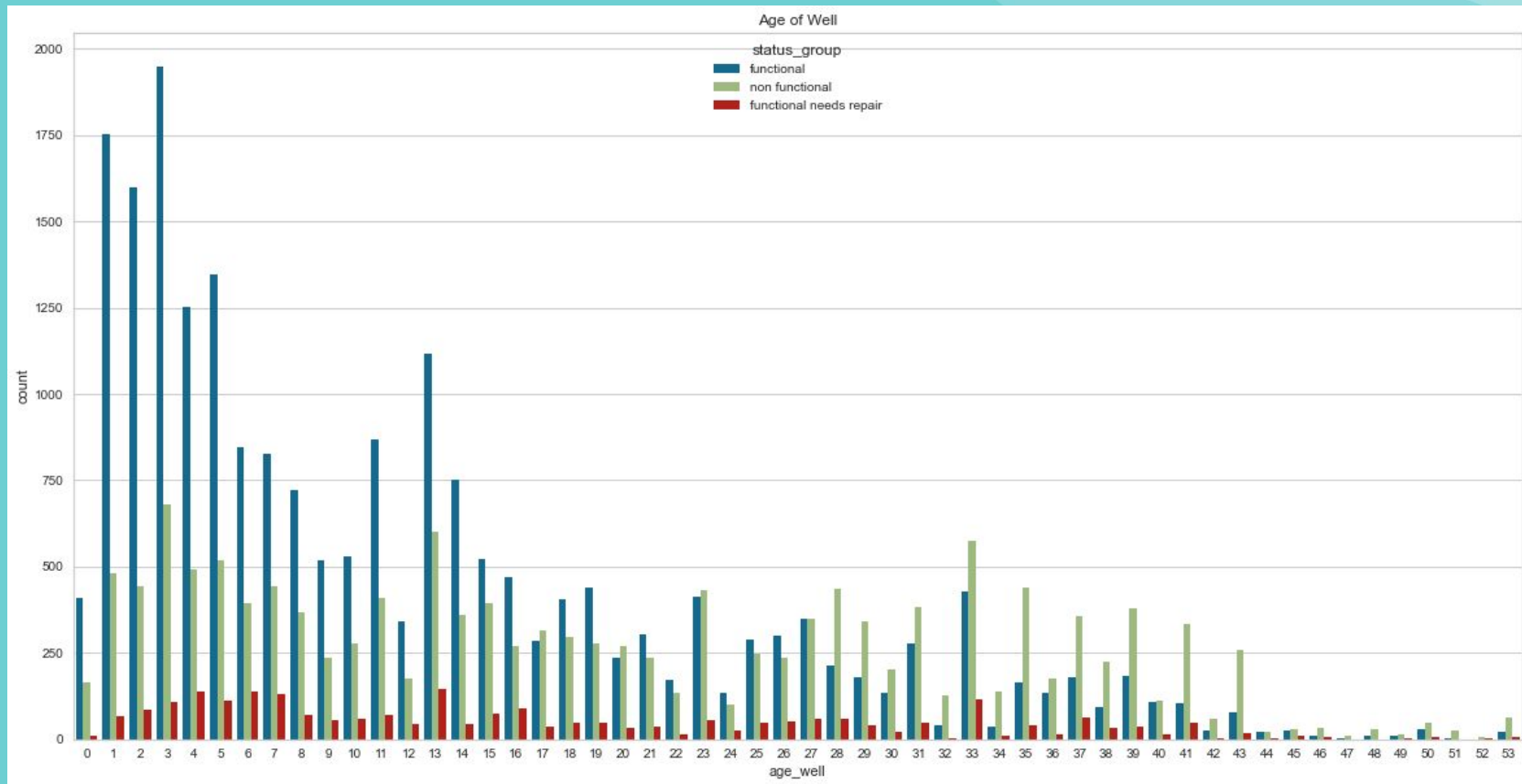
# Data Overview - Waterpoint Type



# Data Overview - Year Constructed



# Data Overview - Age of Well





# THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

