VISIMA VYEMA CLASSIFIER

OVERVIEW

Boosting food production and reducing reliance on aid are key goals in developing countries. International aid groups (NGOs) often tackle these challenges by building water systems.

However, a crucial step is being missed. Many NGOs launch new projects without utilizing existing data, leading to inefficient use of resources and donor funding.

Tanzania, as a developing country, has encountered this issue prominently. Here, poorly planned water projects have resulted in wasted funds and, even worse, lives lost due to insufficient access to clean water.

BUSINESS UNDERTANDING

Maji Safi Global, an NGO dedicated to improving water distribution in developing nations, relies on donor funding to support its community outreach efforts.

Collaborating with the Tanzanian government through the Visima Vyema initiative, Maji Safi aims to enhance existing infrastructure by repairing dysfunctional wells and analyzing patterns in non-operational wells.

In my role as a data scientist, I am tasked with identifying wells in need of repair and conducting an in-depth analysis to forecast patterns in non-functional wells. This analysis will provide valuable insights for Maji Safi Global and the Tanzanian government, guiding strategic investments for the future.

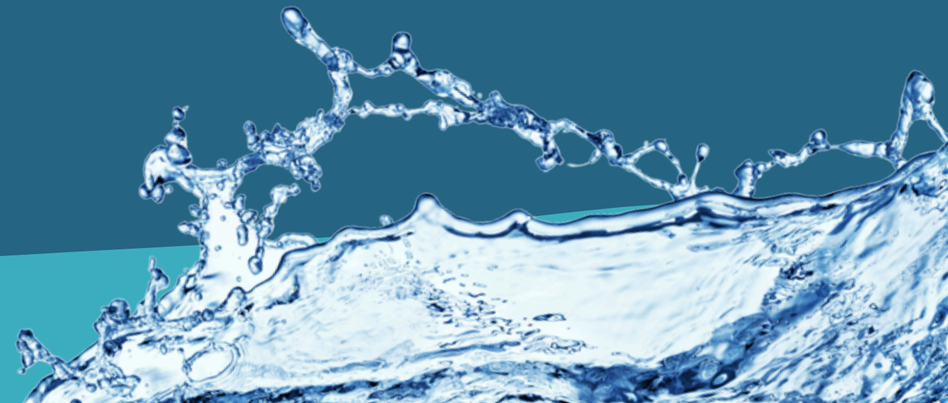
DATA UNDERTANDING

Data Source http://taarifa.org/

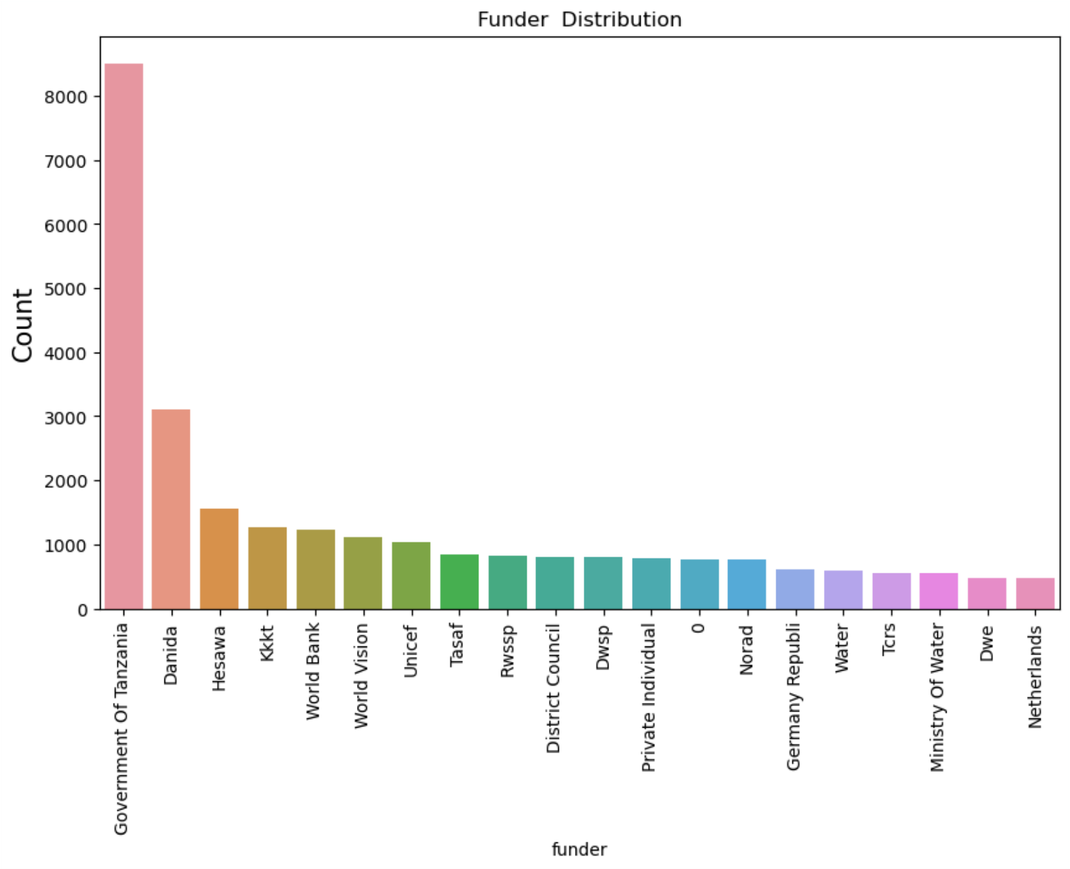
About Data set Dataset has 59,400 rows and 41 columns

Data Cleaning Handled Missing Values, Duplicates and Outliers

Exploratoty Data Analysis

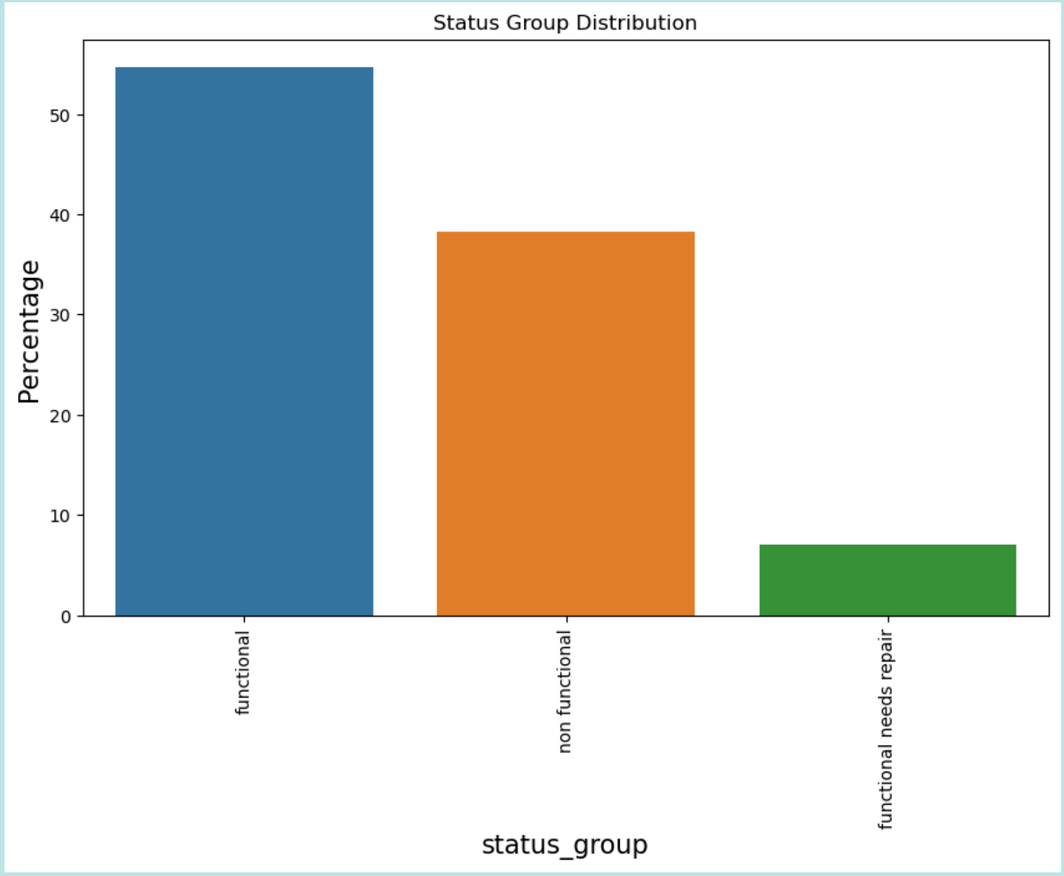


UNIVARIATE ANALYSIS



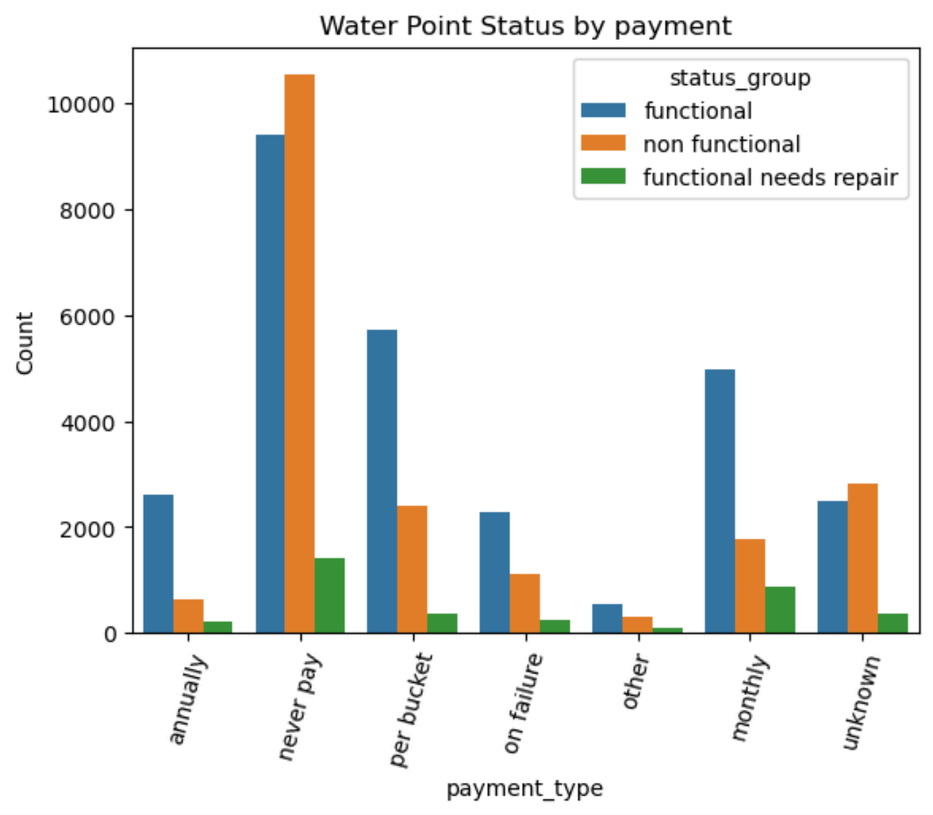
tanzania Goverment is the highest funder

UNIVARIATE ANALYSIS



Most wells are functional

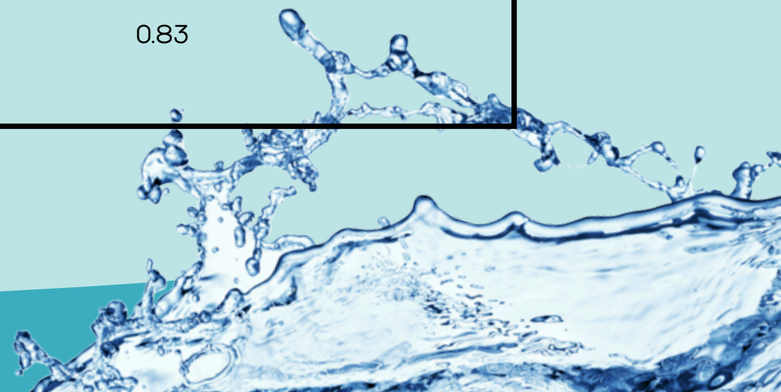
BIVARIATE ANALYSIS



The water points lacking payment records exhibit the highest proportion of non-functional status

MODELING

Model	ACCURACY	RECALL	F1 SCORE	Precision
Logistic Regression	0.773	0.773	0.758	0.765
Random forest	0.798	0.84	0.799	0.799
Gradient Boost	0.804	0.804	0.799	0.798
XG Boost	0.787	0.787	0.772	0.787
Ensemble Methods	0.785	0.62	0.77	0.83



CONCLUSION

.

Ensemble Methods was the better Model to identify water points that needed repair.

55% of its prediction was right.

RECOMENDATION

More work is needed to be done to get the right data..

The governmrnt of Tanzania ansd Maji Safi Global should collaborate to repair the broken water points

THANK YOU