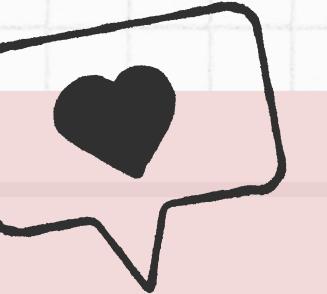


PROJECT PRESENTATION



IMPROVING WATER WELL QUALITY IN TANZANIA

OVERVIEW

BUSINESS PROBLEM

TANZANIA, AS A DEVELOPING COUNTRY, STRUGGLES WITH PROVIDING CLEAN WATER TO ITS POPULATION OF OVER 57,000,000. THERE ARE MANY WATER POINTS ALREADY ESTABLISHED IN THE COUNTRY, BUT SOME ARE IN NEED OF REPAIR WHILE OTHERS HAVE FAILED ALTOGETHER. WE NEED TO COME UP WITH A MODEL THAT GIVES US THE STATUS OF THE WELLS. USING THIS MODEL WE CAN IDENTIFY WHICH WELLS NEED TO BE REPAIRED.

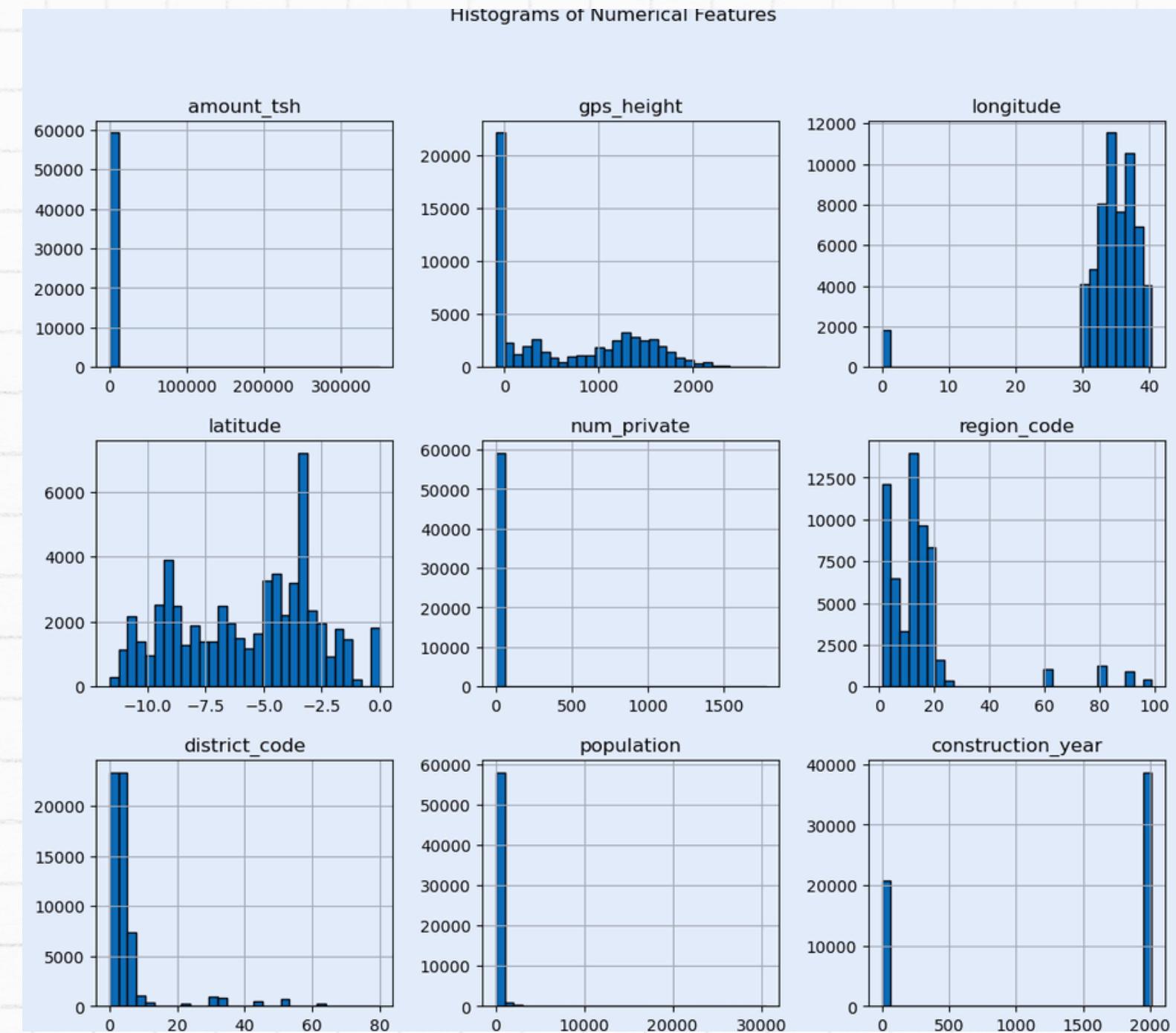
OBJECTIVES

DEVELOP A ROBUST DATABASE TO COLLECT, STORE, AND MANAGE DATA ON THE STATUS OF WATER POINTS ACROSS TANZANIA. USE MACHINE LEARNING TO CREATE A PREDICTIVE MODEL THAT FORECASTS WATER POINT FAILURES AND MAINTENANCE NEEDS. INTEGRATE THE PREDICTIVE MODEL WITH EXISTING WATER MANAGEMENT SYSTEMS FOR PROACTIVE MAINTENANCE PLANNING. UTILIZE THE PREDICTIVE MODEL TO PRIORITIZE WATER POINTS THAT REQUIRE IMMEDIATE ATTENTION. DEVELOP AND ENFORCE GUIDELINES FOR THE CONSTRUCTION AND MAINTENANCE OF WATER POINTS TO ENSURE UNIFORM QUALITY. IMPLEMENT TRAINING AND CERTIFICATION PROGRAMS FOR LOCAL TECHNICIANS AND ENGINEERS.

EXPLORATORY DATA ANALYSIS

IN THIS SECTION, WE WILL PERFORM EXPLORATORY DATA ANALYSIS (EDA) TO UNDERSTAND THE DATA BETTER AND DISCOVER ANY PATTERNS, TRENDS USING BIVARIATE AND MULTIVARIATE ANALYSIS WE WILL USE DESCRIPTIVE STATISTICS AND VISUALIZATIONS TO SUMMARIZE THE MAIN CHARACTERISTICS OF THE DATA AND EXAMINE THE RELATIONSHIPS BETWEEN THE FEATURES AND THE TARGET VARIABLE. WE WILL ALSO CHECK THE DISTRIBUTION AND CORRELATION OF THE VARIABLES AND IDENTIFY ANY POTENTIAL PROBLEMS OR OPPORTUNITIES FOR THE ANALYSIS.

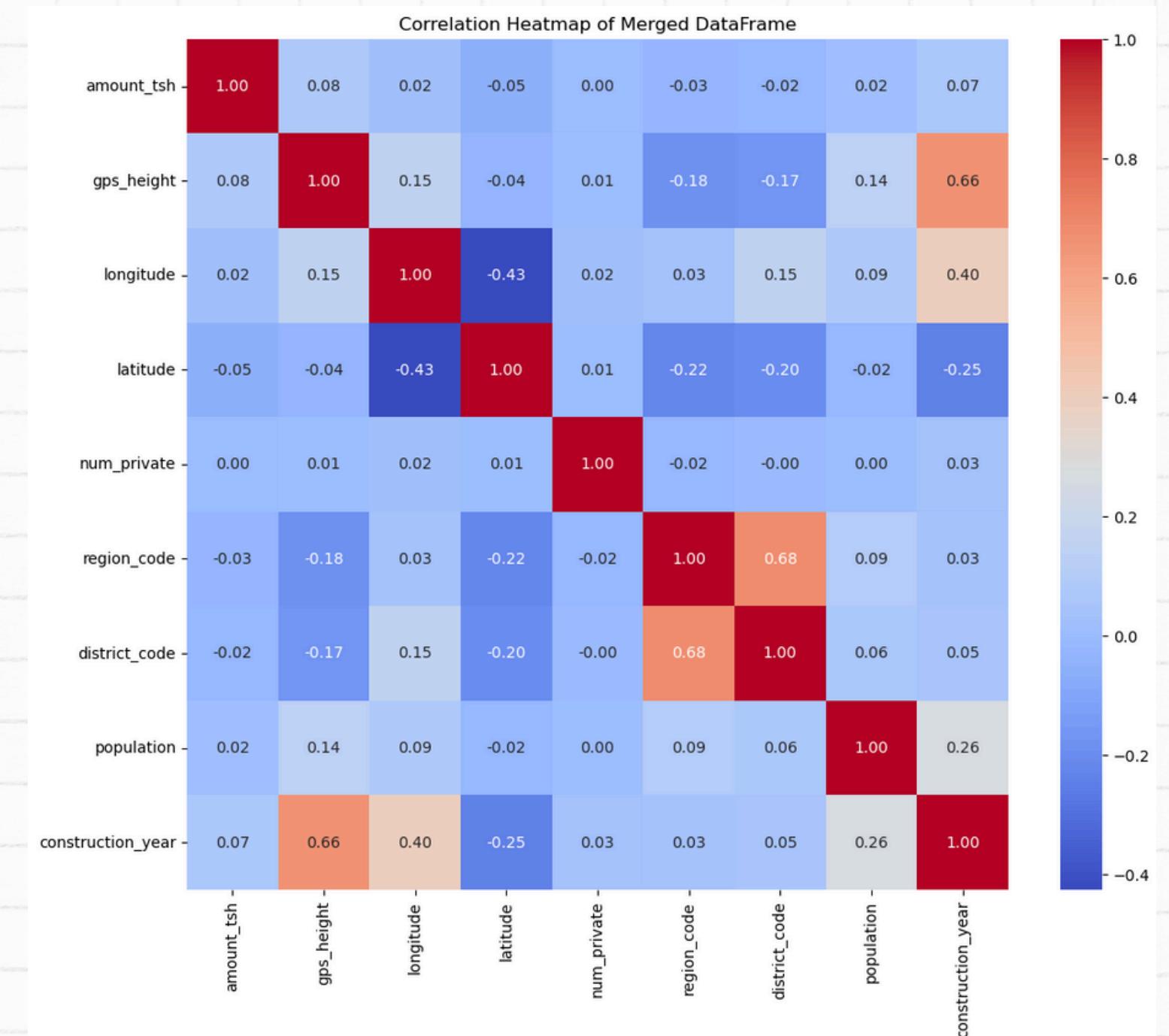
BIVARIATE ANALYSIS



THE SCATTER PLOTS SHOWS THAT THERE IS A POSITIVE RELATIONSHIP BETWEEN SOME OF THE INDEPENDENT VARIABLES AND THE STATUS_GROUP. THIS MEANS THAT MOST OF THE INDEPENDENT VARIABLES INFLUENCE THE STATUS OF THE WATER WELLS IN ONE WAY OR THE OTHER.



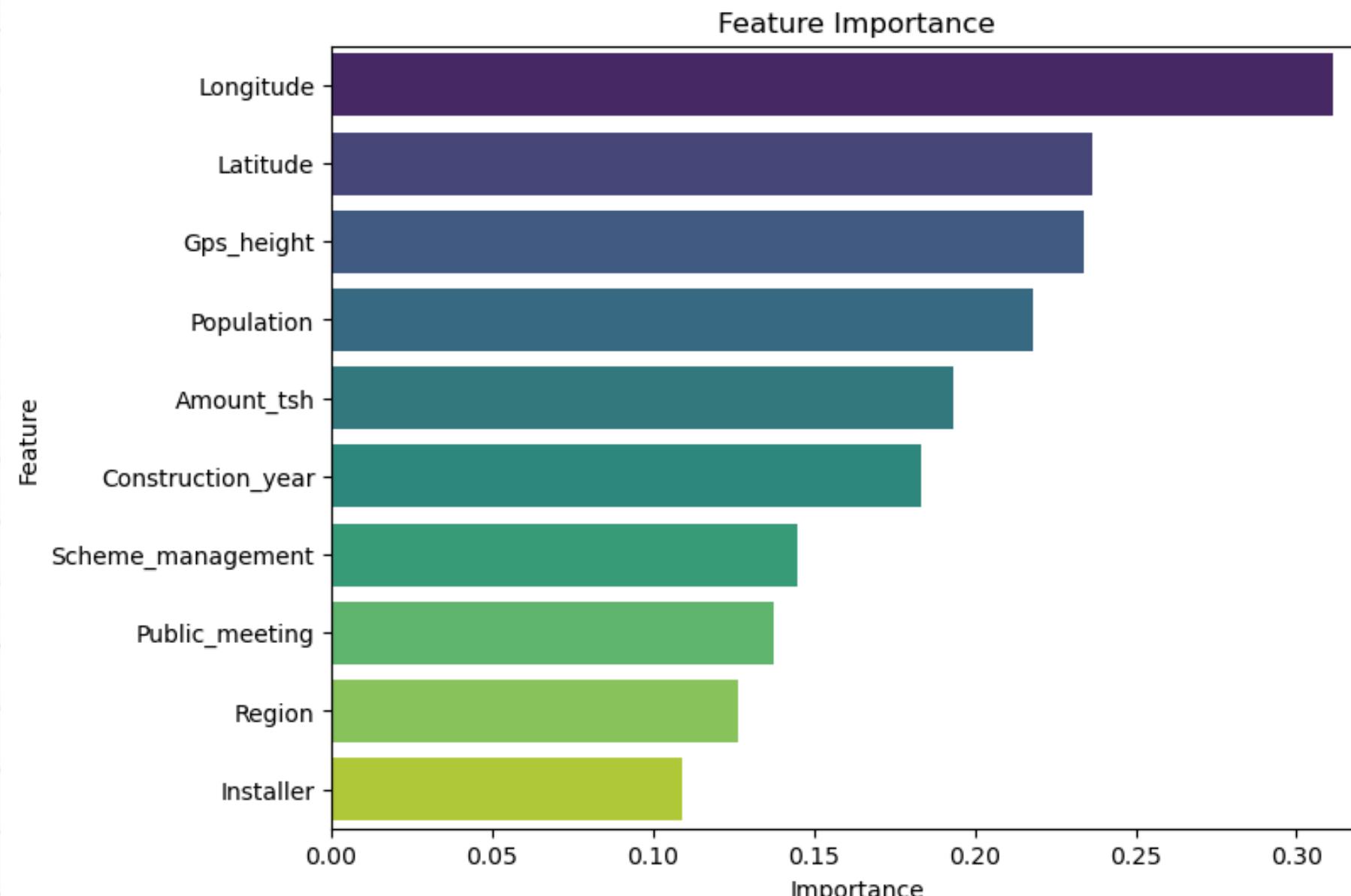
MULTIVARIATE ANALYSIS



IN THIS SECTION, WE WILL PERFORM MULTIVARIATE ANALYSIS TO EXAMINE THE RELATIONSHIP BETWEEN VARIOUS VARIABLES IN THE DATA. WE WILL USE HEATMAP TO VISUALIZE THE CORRELATION MATRIX OF THE FEATURES AND SEE HOW THEY ARE RELATED TO EACH OTHER AND TO THE PRICE.

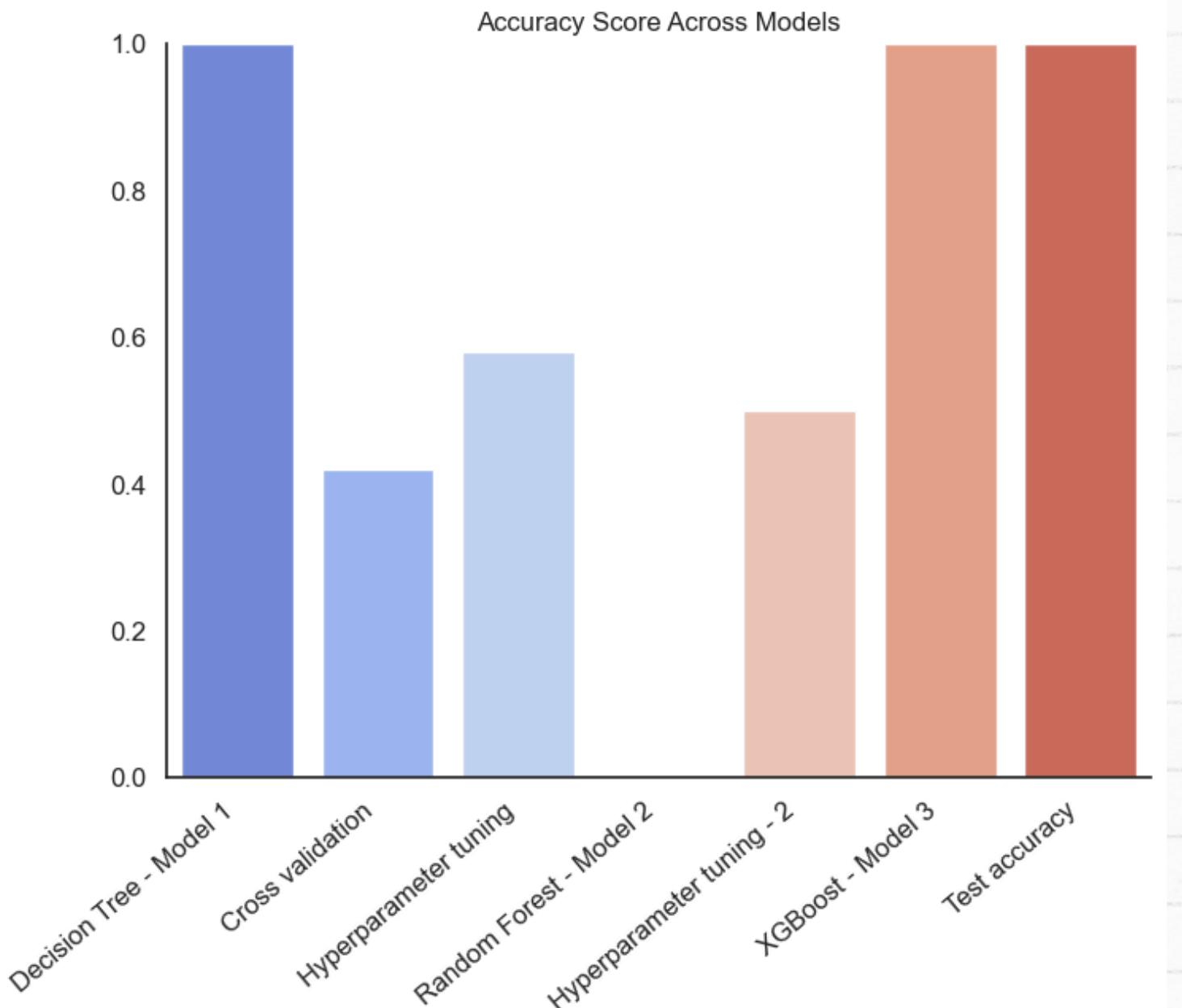
FROM THE HEATMAP ABOVE, POSITIVE CORRELATIONS ARE TYPICALLY REPRESENTED BY SHADES OF RED, AND NEGATIVE CORRELATIONS BY SHADES OF BLUE. WE NOTE THAT DISTRICT_CODE AND REGION_CODE ARE HIGHLY POSITIVELY CORRELATED.

PERMUTATION FEATURE IMPORTANCE



FROM THE ABOVE VISUALIZATION WE GET TO SEE THAT THE MOST IMPORTANT FEATURE IS THE LONGITUDE. LONGITUDE IS VERY CRITICAL FOR GEOGRAPHICAL LOCATION THUS IT CAN AFFECT THE WATER LEVELS IN THE WELLS. THE LEAST IMPORTANT FEATURE IS THE INSTALLER. THE SKILL AND RELIABILITY OF THE INSTALLER CAN IMPACT THE INITIAL SETUP QUALITY, THOUGH IT'S OFTEN LESS CRITICAL THAN ONGOING FACTORS SUCH AS LATITUDE AND LONGITUDE.

GRAPHS FOR BUSINESS PRESENTATION



CONCLUSION

WE CAN USE THE XGBOOST MODEL WHICH IS 100% ACCURATE ON UNSEEN DATA.

GIVEN THE CRITICAL FACTORS INFLUENCING THE FUNCTIONALITY OF WATER POINTS IN TANZANIA, SUCH AS LONGITUDE AND LATITUDE, HERE'S A RECOMMENDATION FOR THE GOVERNMENT:

INVESTMENT IN GEOSPATIAL DATA ANALYSIS

INFRASTRUCTURE PLANNING AND DEVELOPMENT

COMMUNITY ENGAGEMENT AND CAPACITY BUILDING

INTEGRATION OF TECHNOLOGY AND INNOVATION

COLLABORATION WITH STAKEHOLDERS



**THANK YOU
VERY MUCH!**

