National Charitable Organisation

MACHINE LEARNING PROJECT

TABLE OF CONTENTS

Introduction

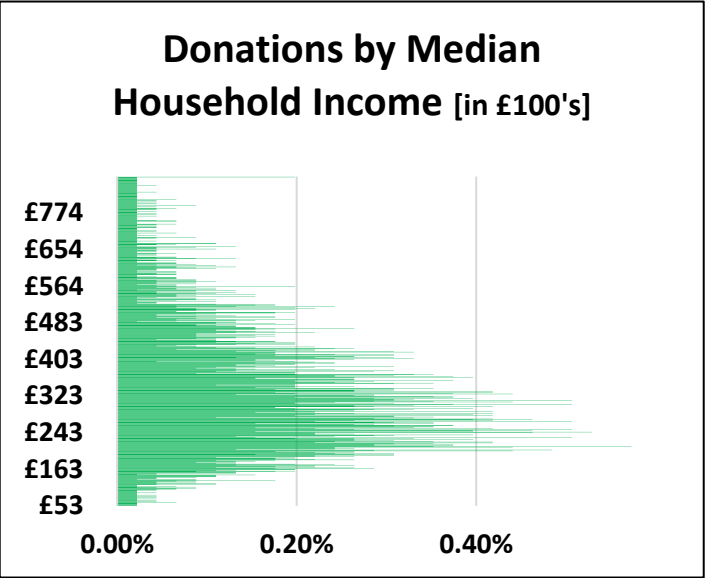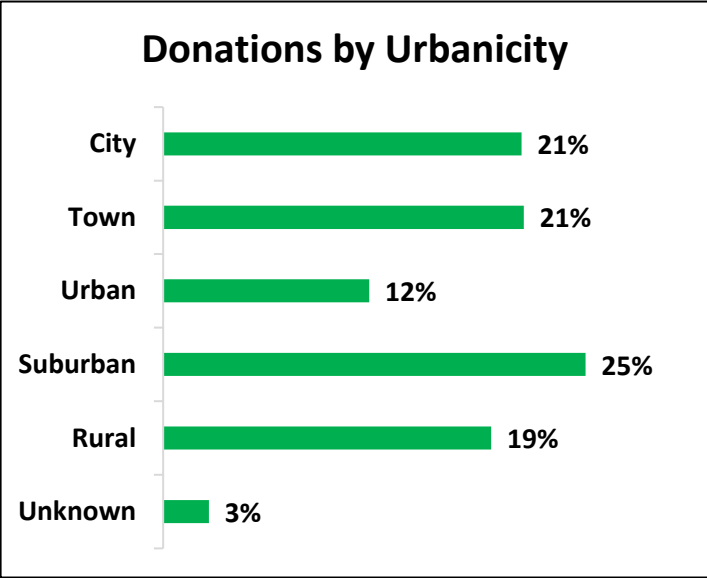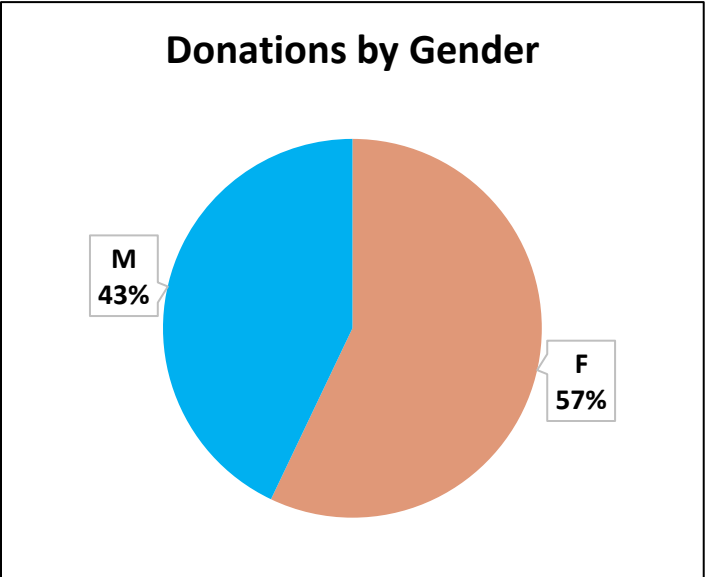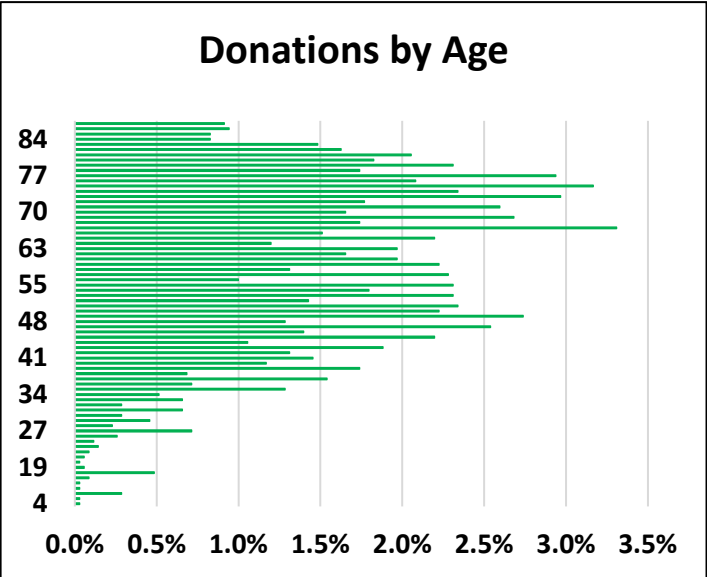Business Objectives: philanthropy and social well-being.

The machine learning project's purpose is to use the results of a previous postcard mail solicitation for donations to deliver actionable insight and to improve the outcome in the next campaign.

The machine learning project's goal is to determine which of the individuals in the National Charitable Organisation mailing database have characteristics similar to their most profitable donors. By soliciting only these people, the Organisation can spend less money on the solicitation effort and more money on charitable concerns.

Data Analysis

From the analysis of the historical data results that the most profitable donors are aged between 40 and 81 years old and have a low-middle household income.

Donations by Age

Donations by Gender
- M 43%
- F 57%

Donations by Urbanicity
- City 21%
- Town 21%
- Urban 12%
- Suburban 25%
- Rural 19%
- Unknown 3%

Donations by Median Household Income [in £100's]

Machine Learning process

1. **Import Libraries**

2. **Import Data**

3. **Data Description & Exploration**
   - Data Shape
   - Columns & dtype
   - Peek at the Data
   - Missing Values %
   - Statistics
   - Correlations (numeric variables)
   - Data Distributions (Histograms)

4. **Data Preparation**
   - Dropped Columns
   - Handle Missing Values
   - Prepare Data for Scaling
   - Scaling: StandardScaler
   - Scaling: RobustScaler

5. **About Algorithms**
   - Set-up the test harness to use 10-fold cross validation
   - Build Models:
     - Logistic Regression (LR)
     - Linear Discriminant Analysis (LDA)
     - K-Nearest Neighbors (KNN)
     - Classification and Regression Trees (CART)
     - Random Forest Decision Tree (RFTree)
     - Gradient Boosting (GrB)
     - Gaussian Naive Bayes (NB)
     - Support Vector Machines (SVM)
     - Deep Learning (Deep)

Machine Learning process

Final machine learning process:

- Algorithm - Deep Learning

- Data scaling technique - Robustscaler

- Accuracy score - 72%

6. **Evaluate Some Algorithms**

- 6.1.a Create a Validation Dataset with Standard Scaled Data

- 6.2.a Build Models, Make and Evaluate Predictions on different models with Standard Scaled Data

- 6.1.b Create a Validation Dataset with Robust Scaled Data

- 6.2.b Build Models, Make and Evaluate Predictions on different models with Robust Scaled Data

- 6.3 Choose the best model: <u>Deep Learning (Deep) model with Robust Scaled Data</u>

7. **Train the Final Machine Learning Model**

8. **Save the Load Final Machine Learning Model**
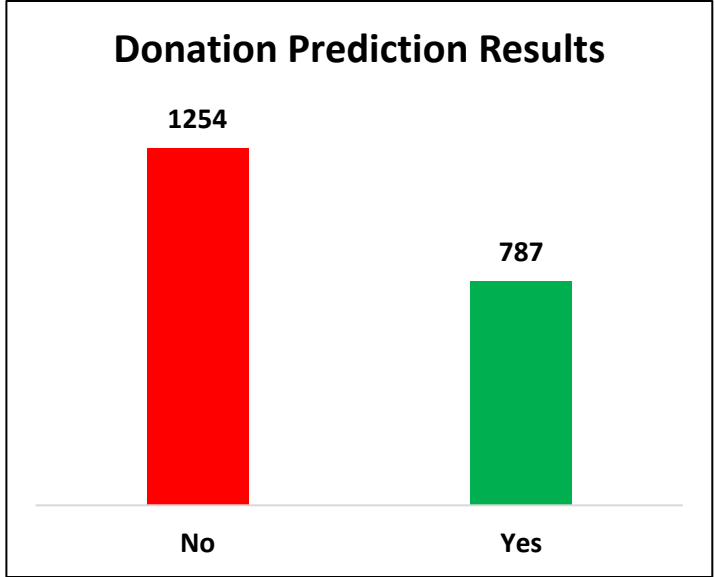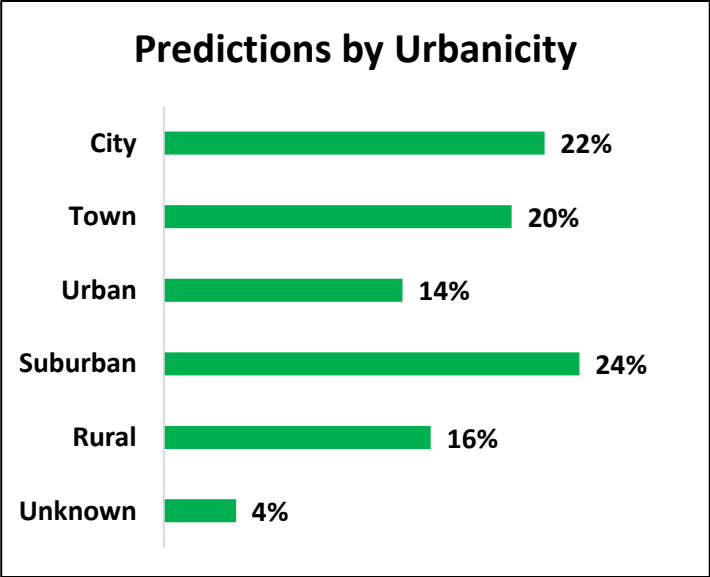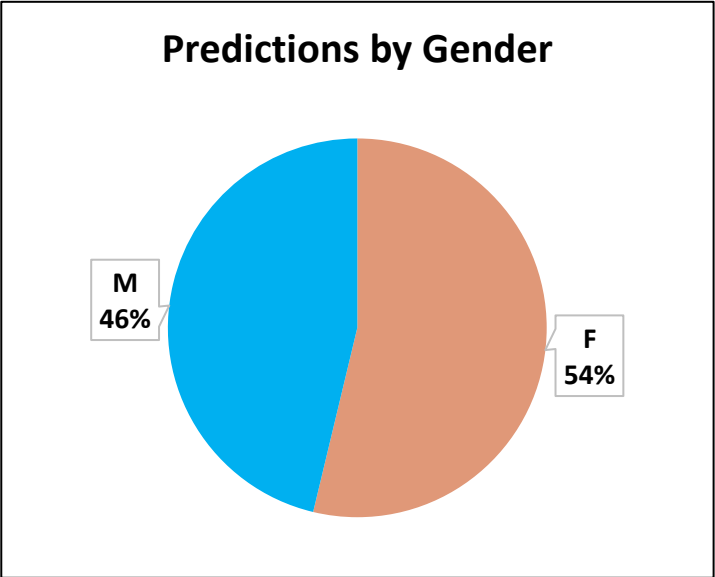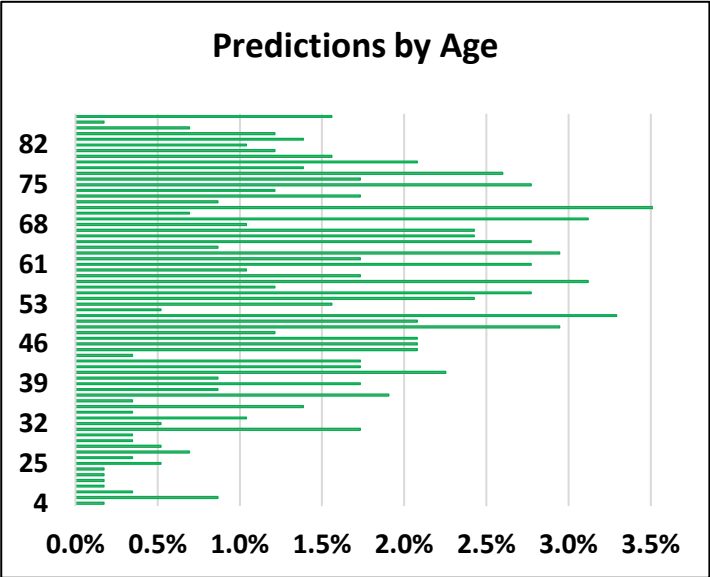
9. **Data Preparation of New Data**

10. **Make Predictions**
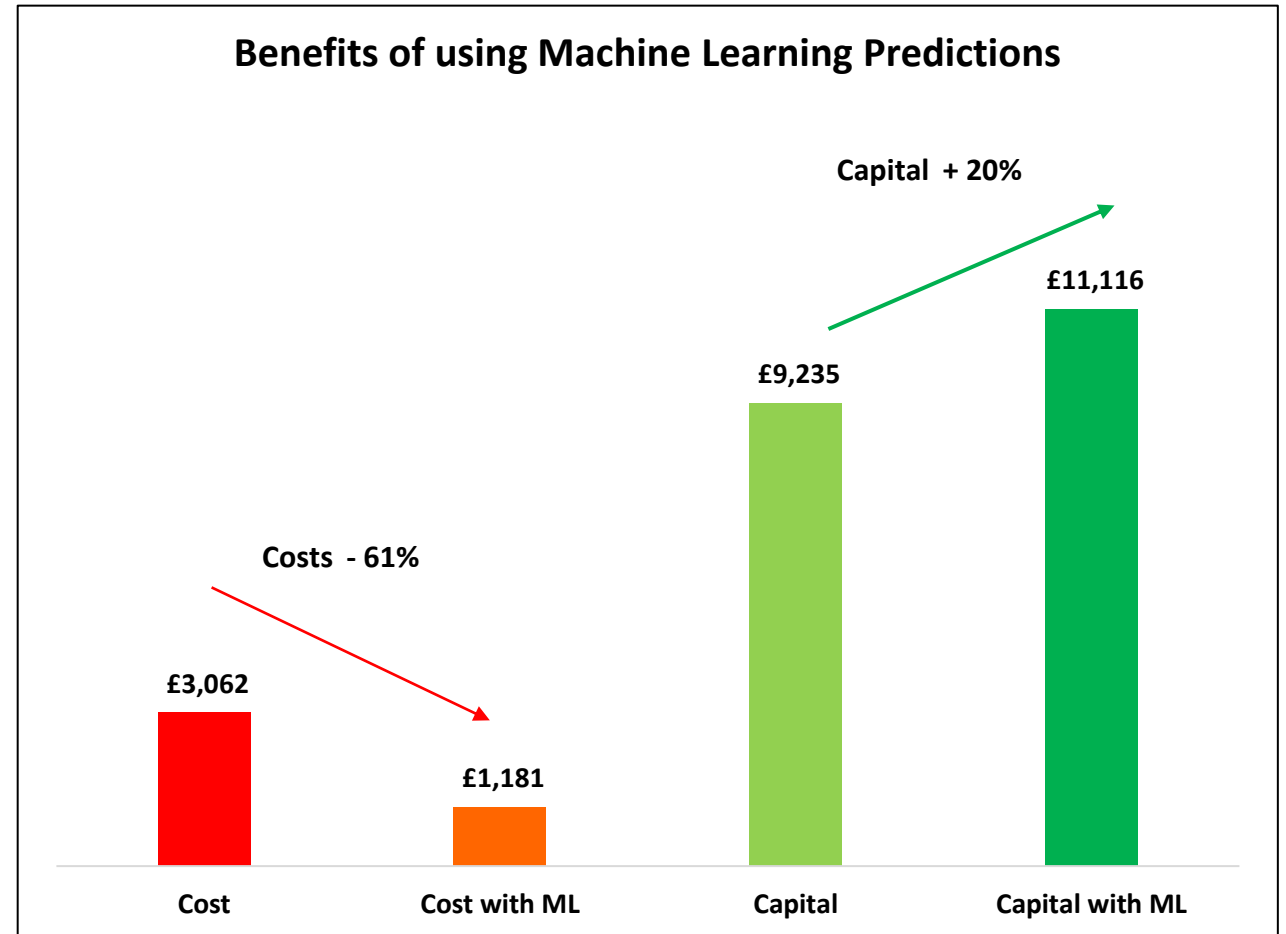
Machine Learning Findings

The results of the machine learning process and the outcome of the analysis of the historical data are consistent.

The machine learning process has identified 787 potential donors out of 2041.

Predictions by Age

Predictions by Gender

M 46%

F 54%

Predictions by Urbanicity

City	22%
Town	20%
Urban	14%
Suburban	24%
Rural	16%
Unknown	4%

Donation Prediction Results

1254

787

No   Yes

Conclusions

Soliciting the potential donors identified by the machine learning process, the Organization could reduce the costs of 61% and increase the capital raised by 20%.

Benefits of using Machine Learning Predictions

Capital + 20%

£11,116

£9,235

Costs - 61%

£3,062

£1,181

Cost | Cost with ML | Capital | Capital with ML

Recommendations

Increase the machine learning process accuracy score:

- **By tuning the hyper-parameters of the estimator using GridSearchCV**

- **By implementing a feature reduction with Recursive Feature Elimination (RFE) or Principle Component Analysis (PCA)**

Potentiate the marketing strategy implementing e-marketing techniques using new channels as:

- **Social media**

- **Email campaign**