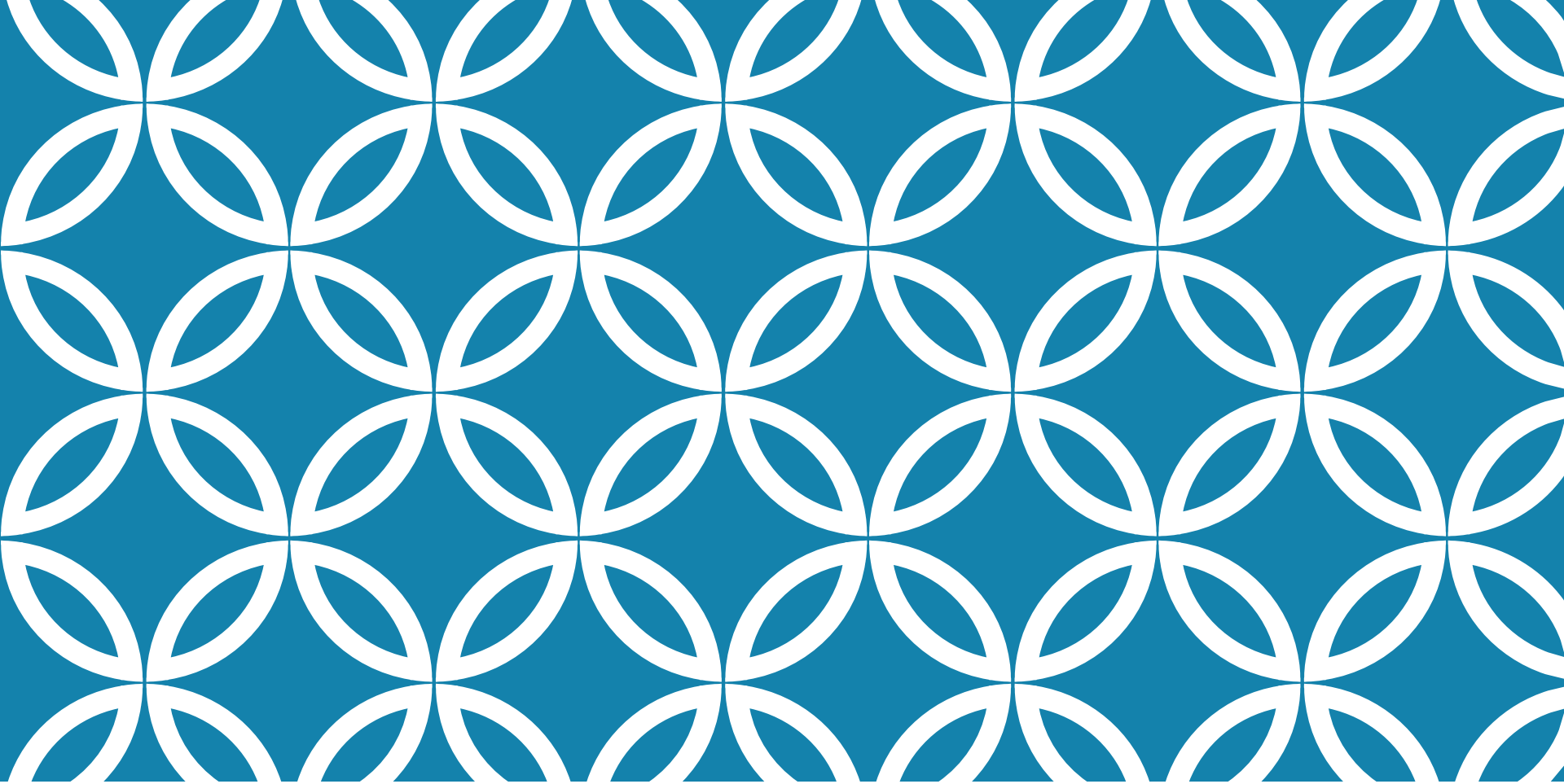


LEAD SCORING ANALYSIS USING PYTHON

A Machine Learning Approach

Presented by: Onkar Hankare, Nivedita Iyer, Padmaja Dasari

INTRODUCTION

This project aims to assign a lead score to potential customers using Logistic Regression.

DATASET OVERVIEW

The dataset contains lead information from X Education. Initial data preprocessing includes handling missing values and feature selection.

DATA PREPROCESSING

Steps:

- Replacing 'Select' with NaN
- Dropping columns with >50% missing values
- Filling missing categorical values with mode

FEATURE ENGINEERING

Selected important features and applied scaling where necessary.

MODEL BUILDING

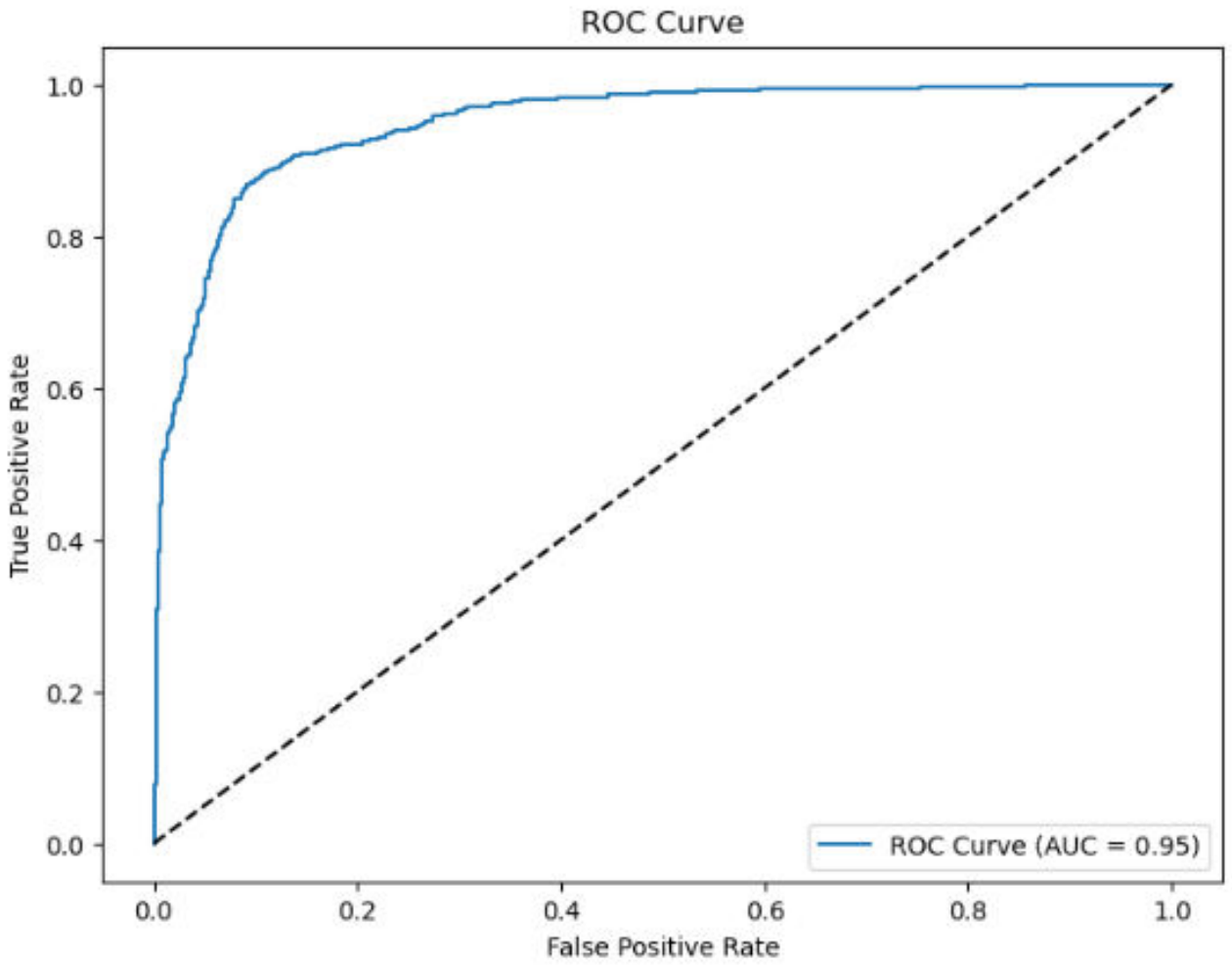
Implemented Logistic Regression using Scikit-Learn.

MODEL EVALUATION

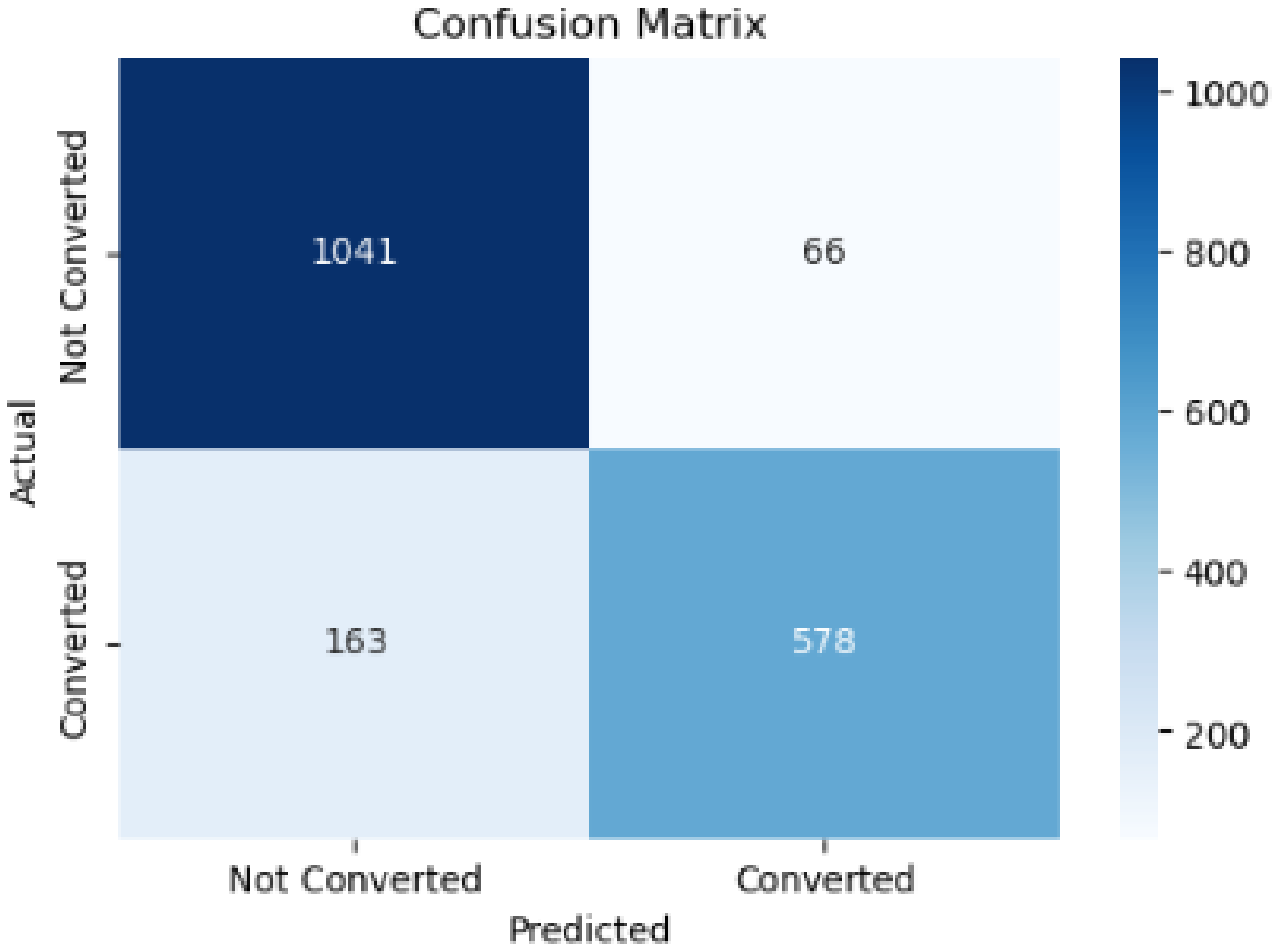
Metrics used:

- Accuracy Score
- Confusion Matrix
- ROC-AUC Score

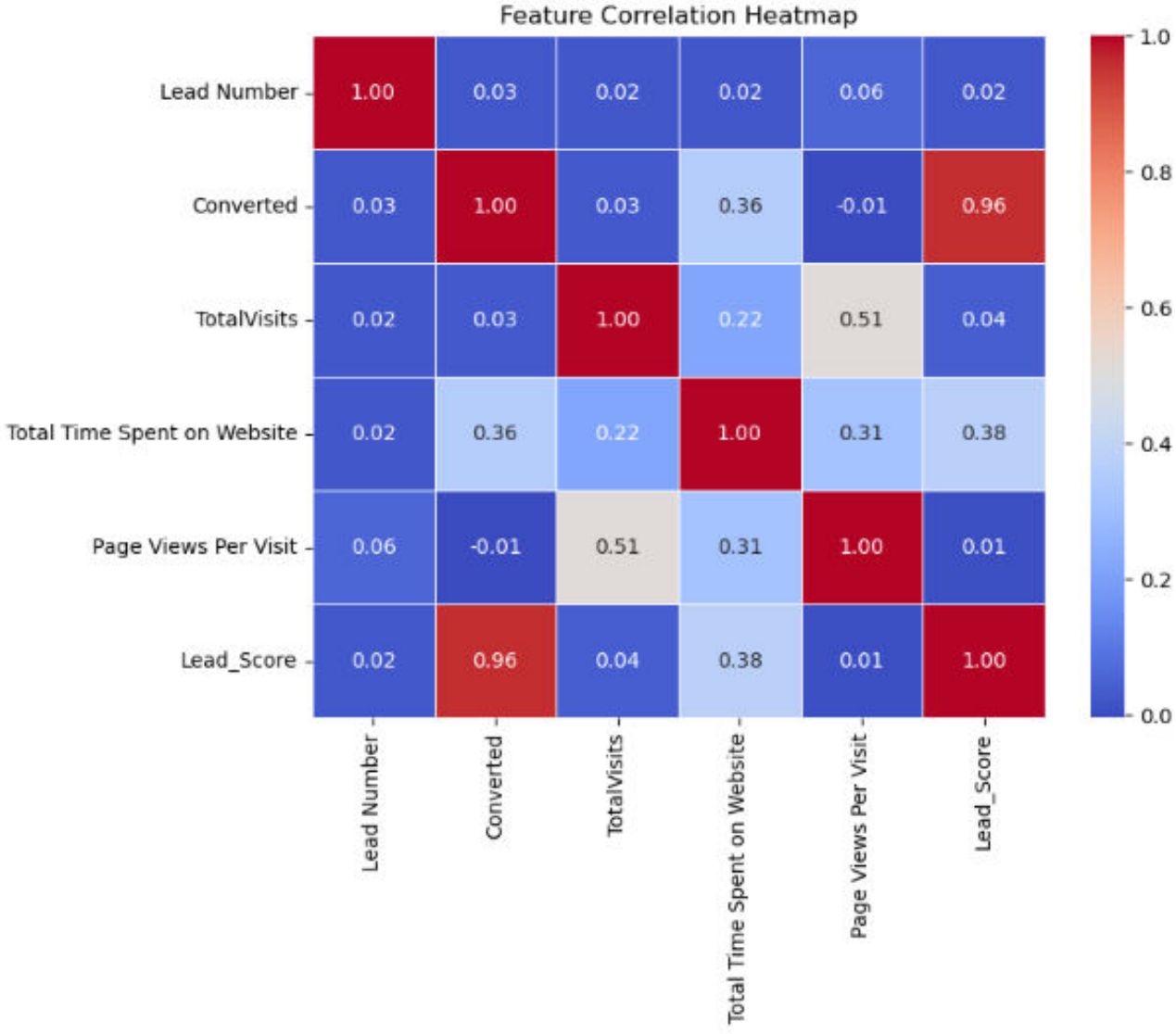
ROC CURVE


ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC Curve (AUC = 0.95)

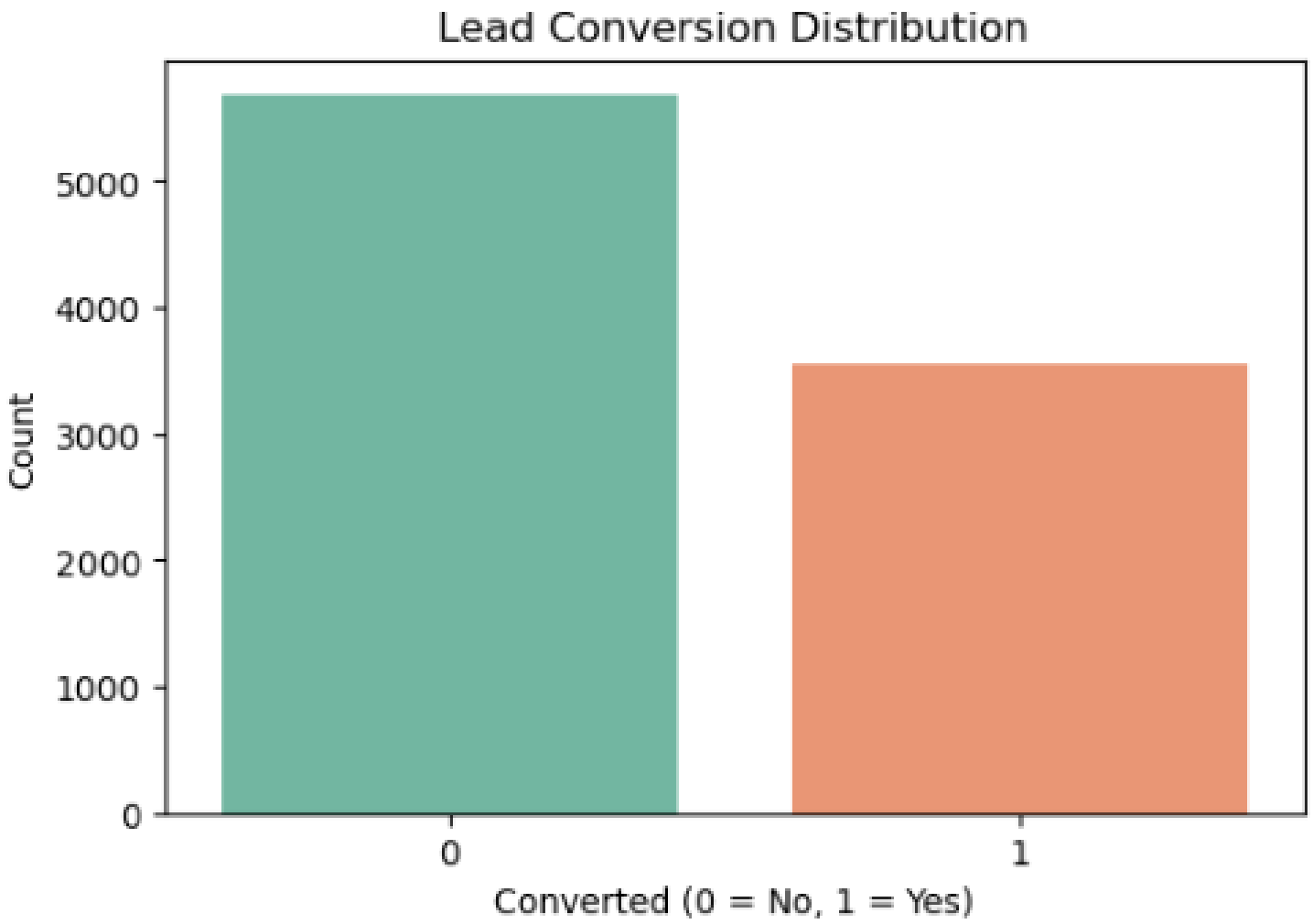
CONFUSION MATRIX



CORRELATION HEATMAP



LEAD CONVERSION DISTRIBUTION



CONCLUSION

The model provides a lead score helping the company focus on high-conversion leads. Further improvements can be made using advanced ML techniques.