ML Pipeline: Customer Propensity Test Data Model Engineering: Input Data Models: 1. Benchmark Model: Decision Tree 2. Logistic Regression: PreProcessing: Train **Parameter Tuning:** 1. Missing Value Fixes Data - Eliminate Records : (marital- 0.19%) - Fill: Proportionate Fill (loan, housing, education, iob) Model Evaluation: - Separate Class (default, 20%) - Metrics (Recall, Precision, f1, 2. Inadequate Feature values Model Selection / Packaging: EDA: roc auc, R-squared, confusion - Package model into pickle (education=illeterate 0.04%, default=ves 1. Univariate Analysis matrix) 0.01%) 2. Bivariate Analysis - Evaluation Plots (Roc curve, 3. Categorical Encoding: - ToDo all steps afterwards 3. Data Sanity Check Valid Precision-Recall curve) - Dummy var Encoding (month, 4. Feature Values Validitity education, job, outcome, day of week, Data Check marital, default) 3. Missing Values - Binary Encoding (housing, loan, 4. Invalid Feature Range **Model Deployment** 5.Outliers contact) - Containairise / Build Image 4. Feature Transformation: pdays to 6. Inadequate Feature value (Docker/ Packer) binary (999->0) count 5. Invalid Columns: (Remove outcome -- Deployment (BitBucket 5% data) Pipeline / TerraForm) 6. Outlier Fix: 7. Skew Fix: log(campaign, previous, nr.employed, euribor3m, cons.conf.idx, cons.price.idx) 8. Binning: age into age groups 9. Feature Scaling: (scl nr.employed) 10. Correlation: emp.var.rate. euribor3m 11. Multi-collinearity Fix: (default miss, Education university.degree,day of wee k mon, month aug, marital single, Retraining Job blue-collar) Producti - With new data on Data