

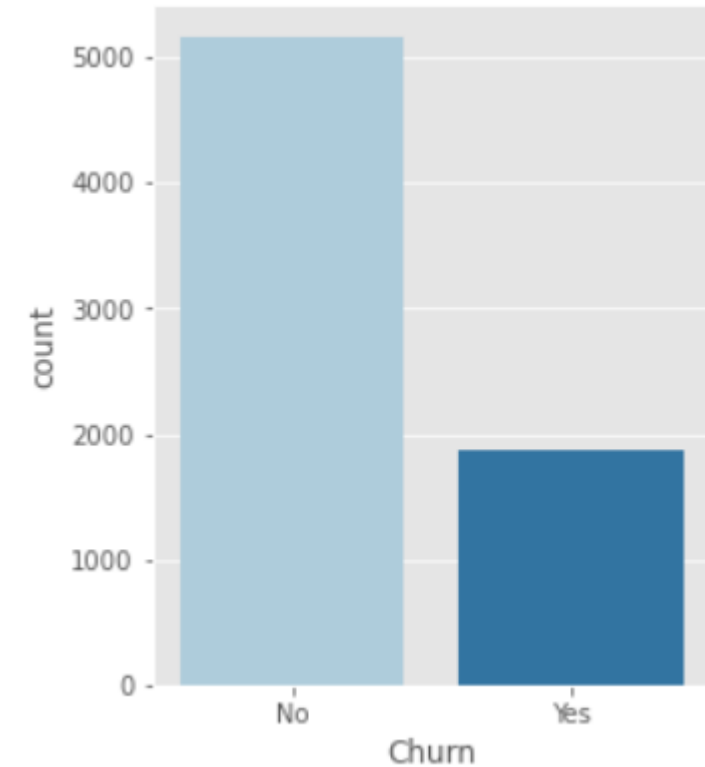
Abstract geometric lines in the top-left corner of the slide.

TELCO CUSTOMER CHURN

Nakos Alexandros

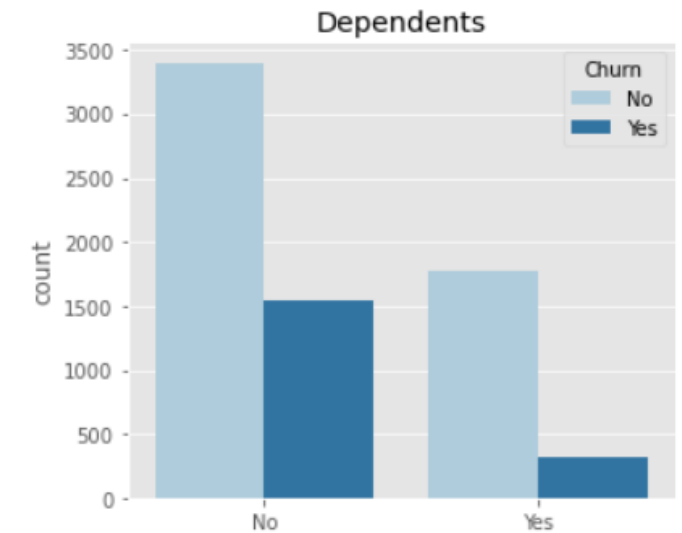
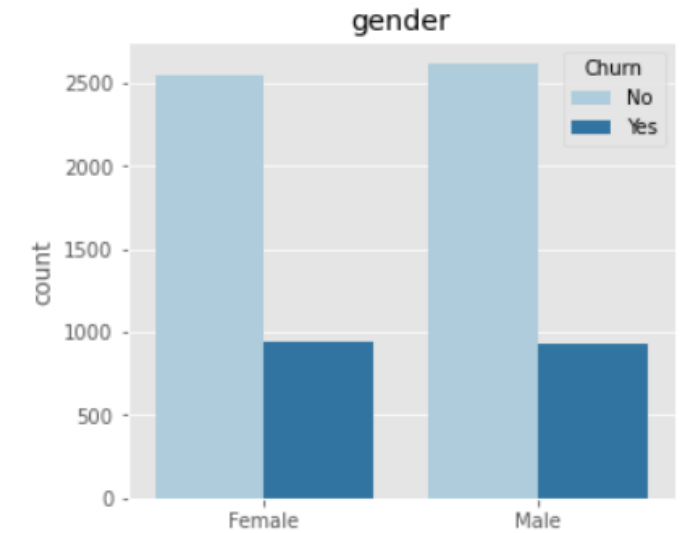
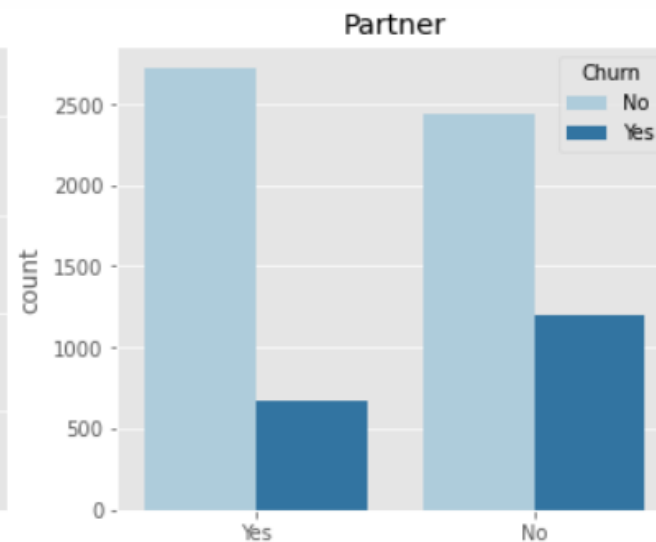
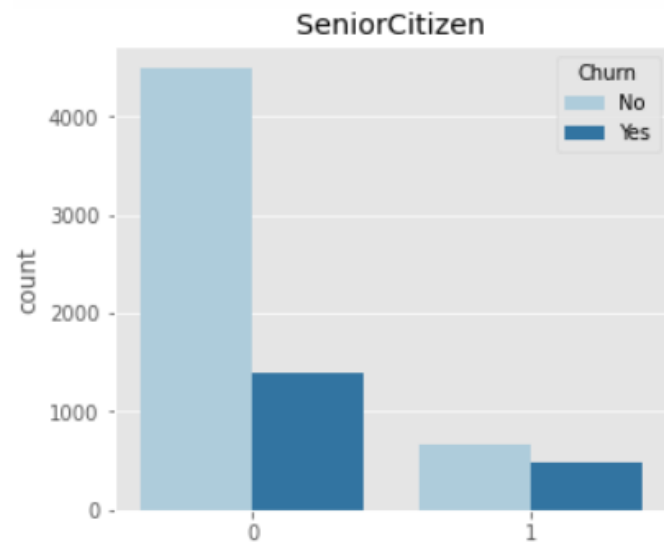
INITIAL STEPS – CHURN DISTRIBUTION

- LOADING THE DATA
- CHECK WHETHER THERE ARE ANY NA VALUES
- ELEVEN NA VALUES IN MONTHLY CHARGES
- THEY ALSO HAVE ZERO TENURE
- WE DECIDE TO DROP THEM
- CHECKING TARGET ATTRIBUTE
- 5174 NON-CHURN CUSTOMERS
- 1869 CHURNED CUSTOMERS
- UNBALANCED DATASET
- NON-CHURN CUSTOMERS FAVORED



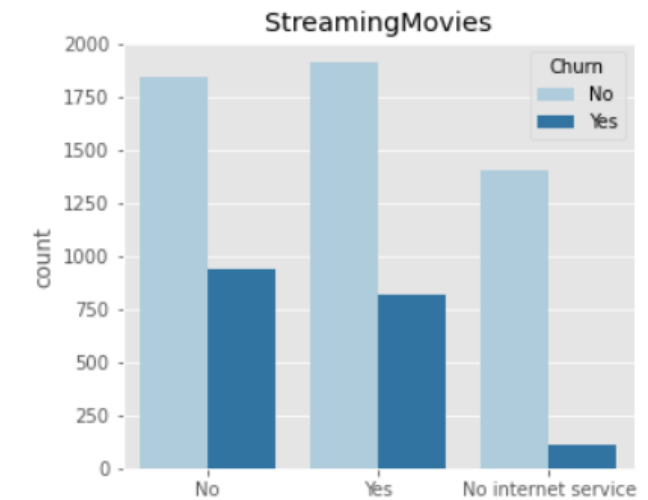
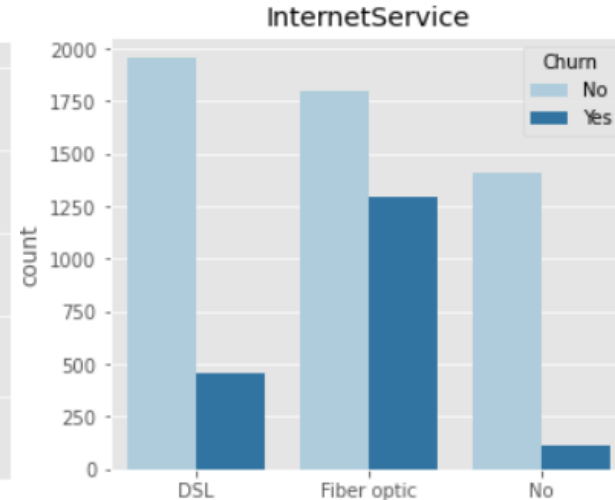
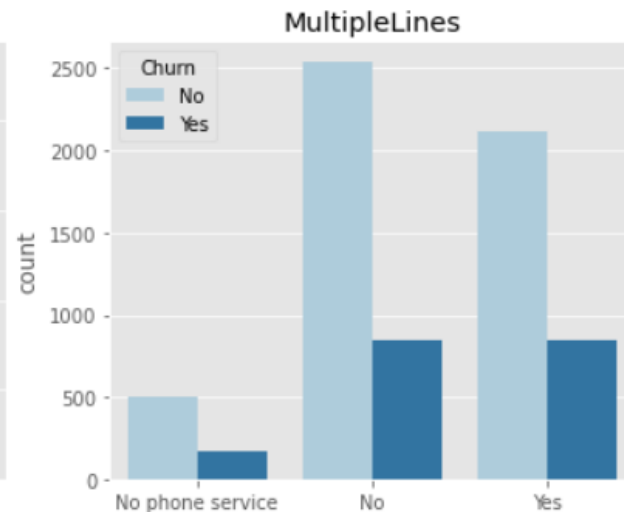
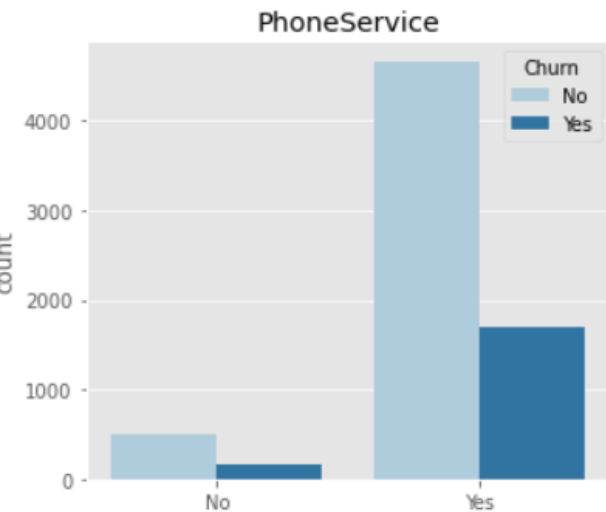
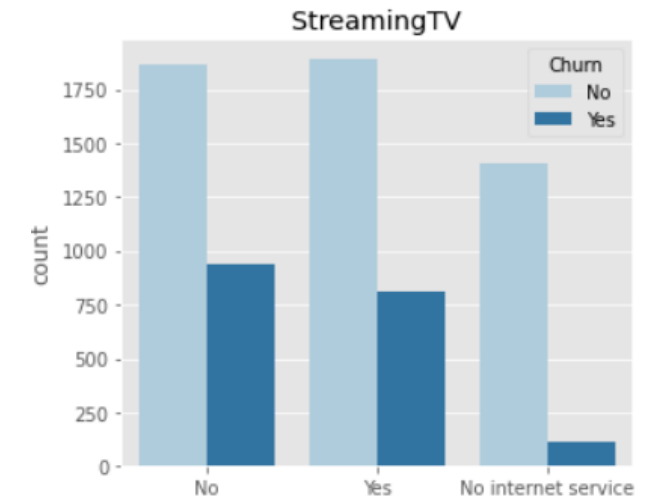
CATEGORICAL ATTRIBUTES - DEMOGRAPHICS

- CHURN SEEMS TO BE THE SAME FOR BOTH GENDERS
- THE NUMBER OF CUSTOMERS CATEGORIZED AS SENIORS IS LOW
- CHURN IS HIGHER FOR THOSE WHO HAVE NO PARTNER AND NO DEPENDENTS



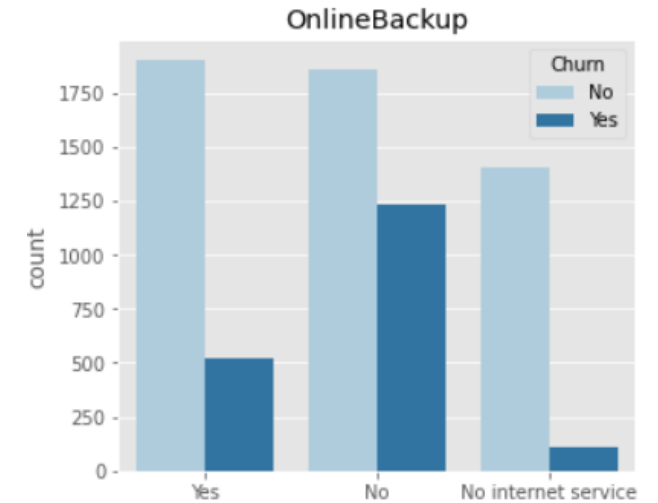
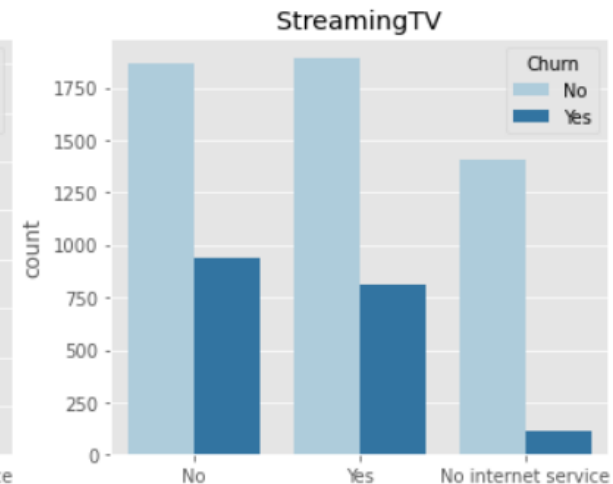
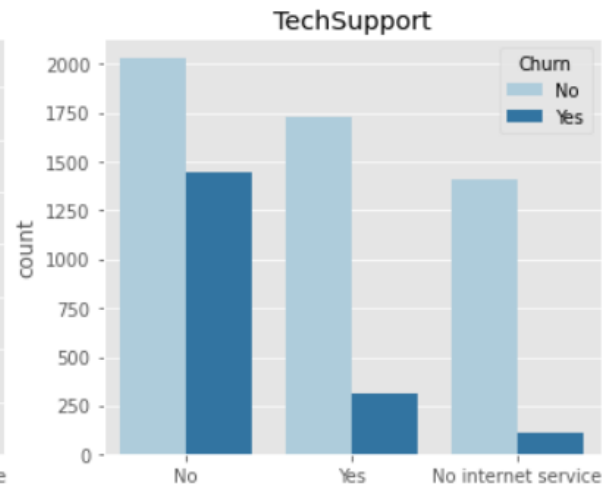
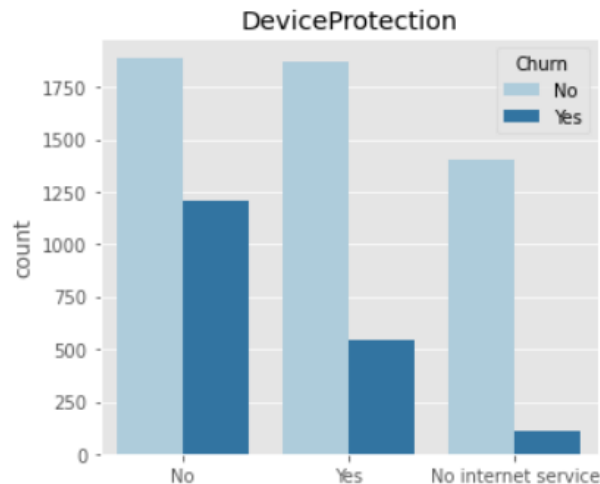
CATEGORICAL ATTRIBUTES - SERVICES

- STREAMING TV AND MOVIES HAVE ALMOST IDENTICAL DISTRIBUTIONS
- FIBER OPTIC IS THE INTERNET SERVICE OF CHOICE, ALSO WITH A HIGH COUNT OF CHURN
- CHURN FOR MULTIPLE LINES IS THE SAME
- CUSTOMERS WITH PHONE SERVICE SHOW HIGH CHURN



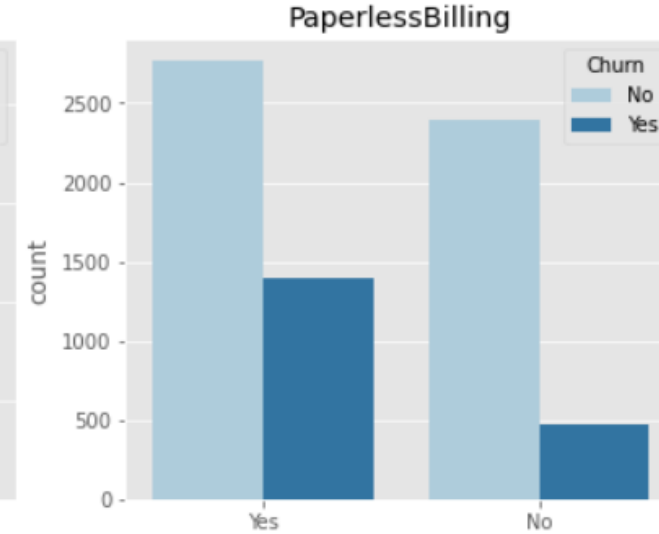
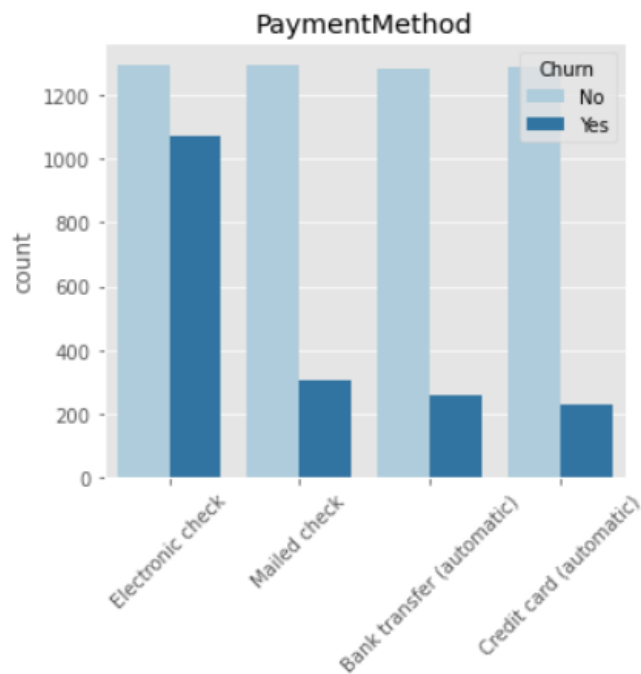
CATEGORICAL ATTRIBUTES - SUPPORT

- ALL ATTRIBUTES SHOWN ARE IMPORTANT WITH REGARDS TO CHURN
- WHEN THEY ARE NOT PRESENT CUSTOMER CHURN IS SIGNIFICANTLY HIGHER

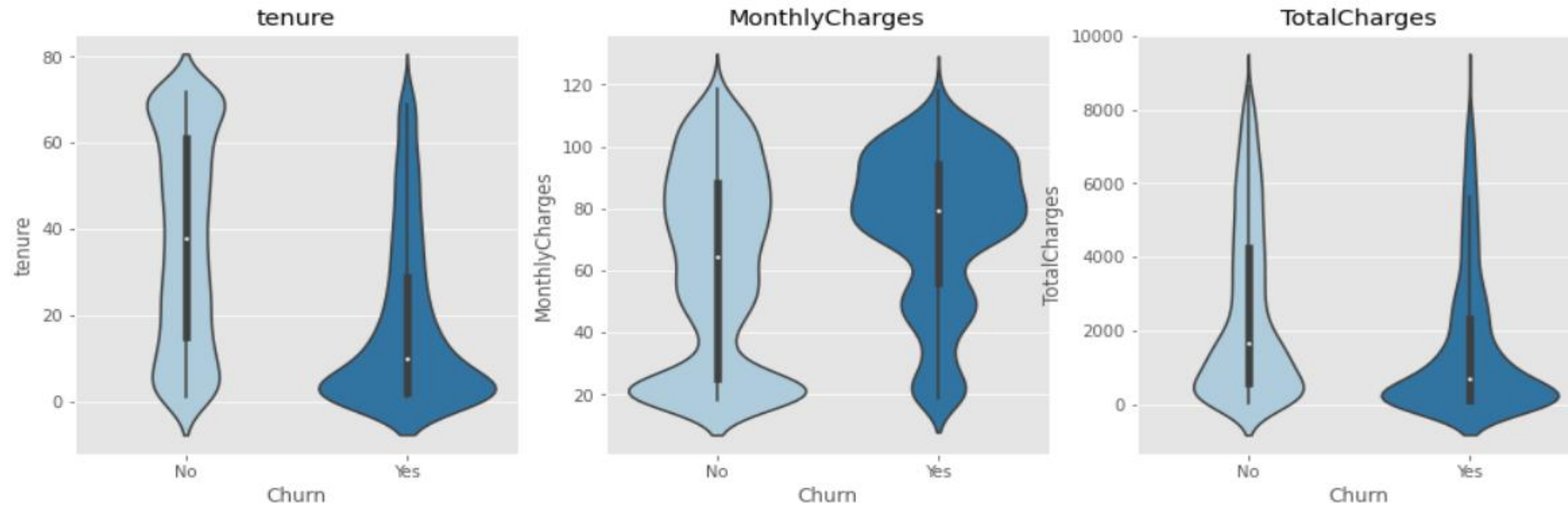


CATEGORICAL ATTRIBUTES – PAYMENT/CONTRACT

- CHURN IS HIGH FOR CUSTOMERS WITH MONTH-TO-MONTH CONTRACT
- SAME FOR CUSTOMERS WITH PAPERLESS BILLING ENABLED
- ALMOST ALL CUSTOMERS WITH ELECTRONIC CHECK AS PAYMENT METHOD CHURNED



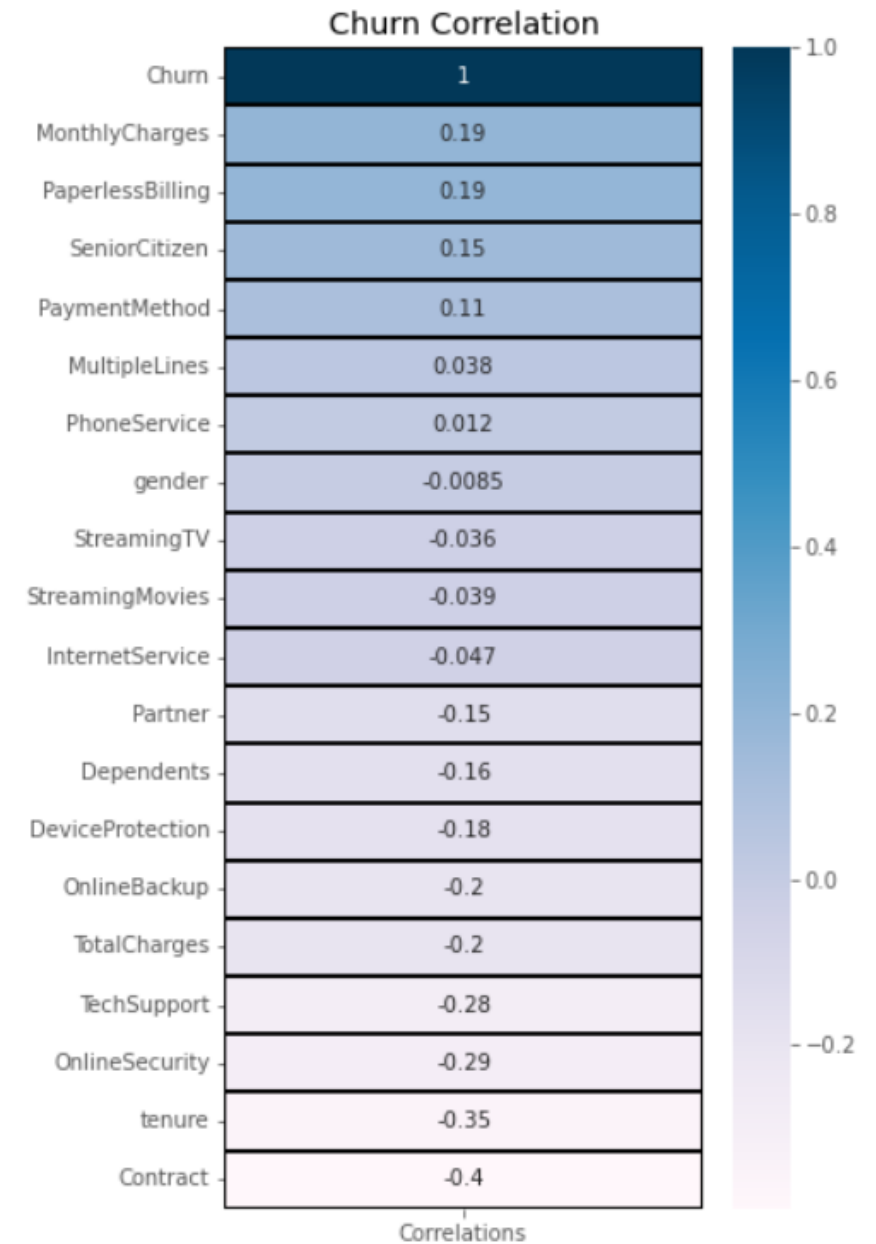
NUMERICAL ATTRIBUTES



- CUSTOMERS WHO DECIDE TO LEAVE, DO SO IN THE FIRST MONTHS
- CUSTOMERS THAT STAY DISPLAY A UNIFORM DISTRIBUTION REGARDING THEIR TENURE
- CHURN IS HIGHER AMONG CUSTOMERS WITH HIGH MONTHLY CHARGES

FEATURE SELECTION

- CORRELATION PLOT W.R.T CHURN
- DROP ATTRIBUTES WITH LOW CORRELATION
- MULTIPLE LINES
- PHONE SERVICE
- GENDER
- STREAMING TV
- STREAMING MOVIES
- INTERNET SERVICE
- ONE HOT ENCODING CATEGORICAL ATTRIBUTES
- STANDARDIZATION OF NUMERICAL ATTRIBUTES



CLASSIFICATION MODELS

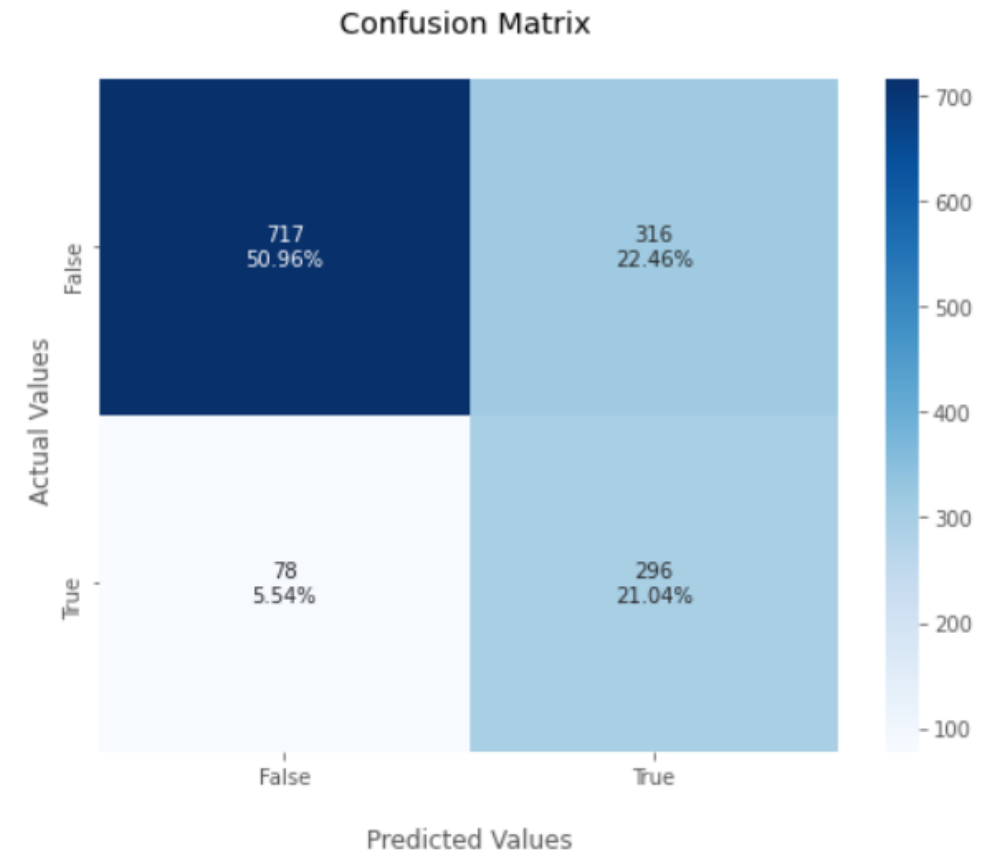
	DECISION TREE	BAGGING CLF (DT BASE)	RANDOM FOREST	ADAPTIVE BOOSTING	GRADIENT BOOSTING
ACCURACY	72.80%	77.55%	77.79%	79.63%	79.59%
ROC AUC	72.80%	85.59%	86.78%	87.95%	88.49%
F1	73.51%	76.64%	77.49%	79.74%	79.49%
RECALL	75.51%	73.64%	76.52%	80.20%	79.06%

- ADAPTIVE AND GRADIENT BOOSTING SHOWS THE MOST PROMISING RESULTS
- PROCEED TO FURTHER OPTIMIZATION

GRADIENT BOOSTING

- IMBALANCED DATASET
- F1 SCORE AS OPTIMIZATION MEASURE
- 5 K-FOLD CROSS VALIDATION

ACCURACY	71.99%
ROC AUC	74.27%
F1	60.04%
RECALL	79.14%



ADAPTIVE BOOSTING

- IMBALANCED DATASET
- F1 SCORE AS OPTIMIZATION MEASURE
- 5 K-FOLD CROSS VALIDATION

ACCURACY	72.49%
ROC AUC	74.61%
F1	60.46%
RECALL	79.14%

