

# 1:HOUSING DATASET

Hyperparameter Tuning using RandomizedSearchCV with a RandomForestRegressor (1)

[9]:

	param_n_estimators	param_max_leaf_nodes	mean_test_error	std_test_error
0	500	100	40.621985	0.769115
2	10	100	41.302509	0.863997
7	100	50	43.714186	0.799775
8	1	100	46.436338	1.028120
6	50	20	49.389446	1.150302
1	100	20	49.467752	0.997048
9	10	20	50.073348	1.218088
3	500	10	54.983731	1.053561
4	5	5	61.278826	0.942016
5	5	2	73.487031	0.985396

Select the most important features

Best Parameters:

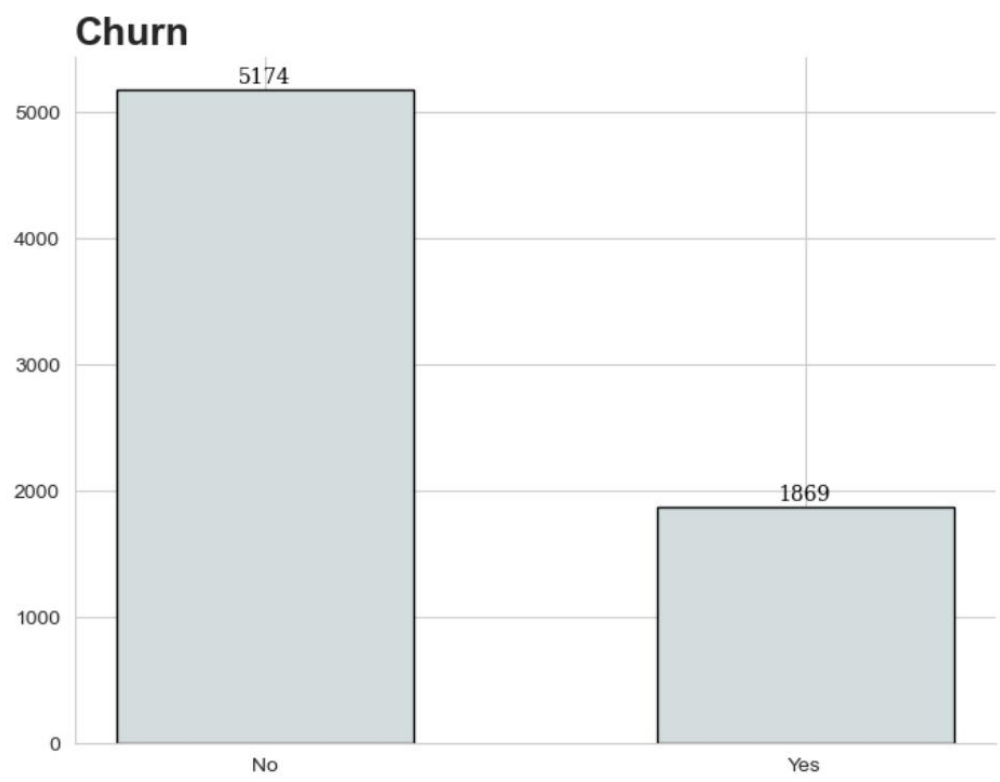
{'regressor\_\_n\_estimators': 500, 'regressor\_\_max\_leaf\_nodes': 50}

Best Mean Test Error: 54.082240941424416

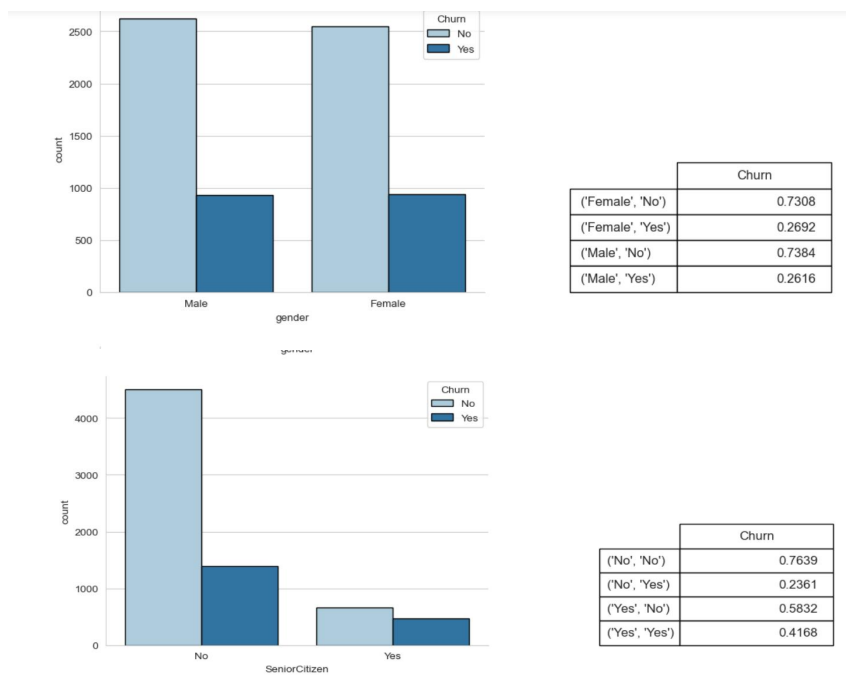
Results Table:

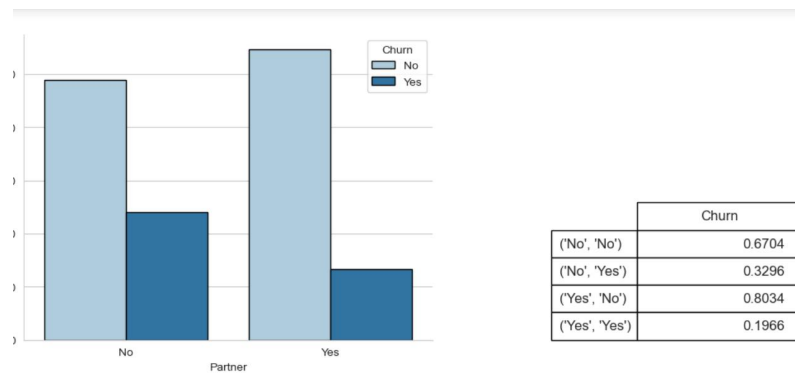
	param_regressor__n_estimators	param_regressor__max_leaf_nodes	
5	500	50	
6	20	50	
8	500	20	
0	2	50	
7	2	10	
2	1	10	
3	50	5	
1	100	5	
4	20	5	
9	2	2	
	mean_test_error	std_test_error	
5	54.082241	1.132283	
6	54.205617	1.092575	
8	54.682871	1.108092	
0	54.991947	1.044255	
7	57.932225	0.829623	
2	59.062376	1.101063	
3	61.241632	1.024724	
1	61.243033	1.081725	
4	61.243507	1.201853	
9	74.072901	0.684151	

2. Customer churn dataset

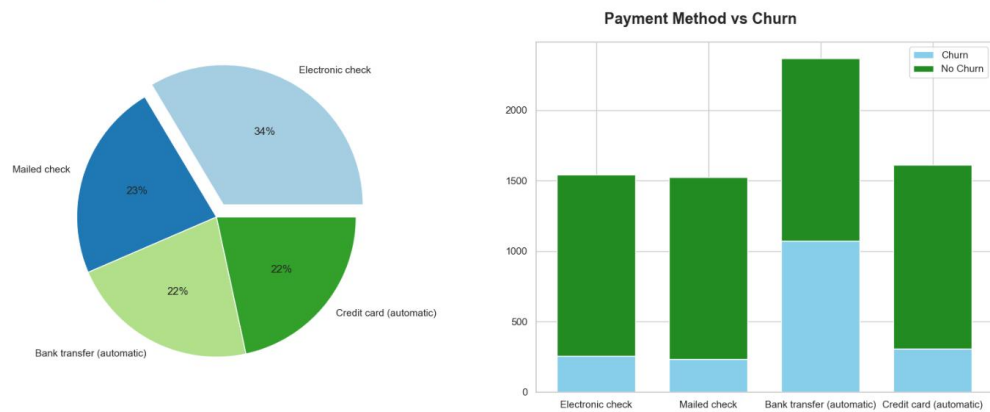


Count plot and a table for proportions

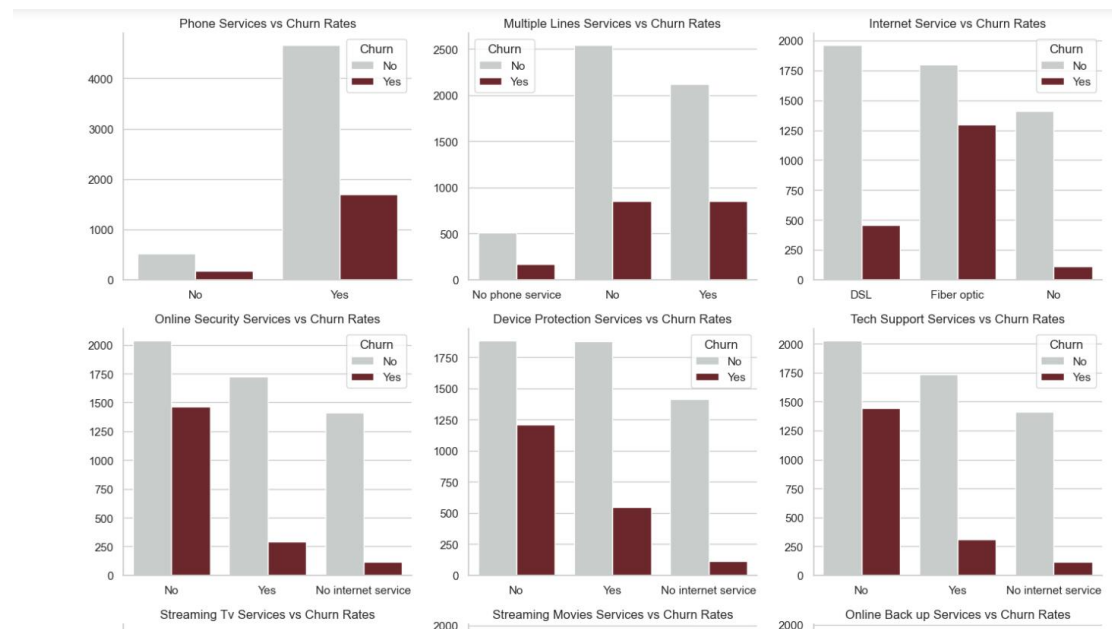




```
fig.text(0.6, 0.92, 'Payment Method vs Churn', fontsize=17, fontweight='bold')
Out[26]: Text(0.6, 0.92, 'Payment Method vs Churn')
```



How it affects churn



# Relationship between different features and clusters

