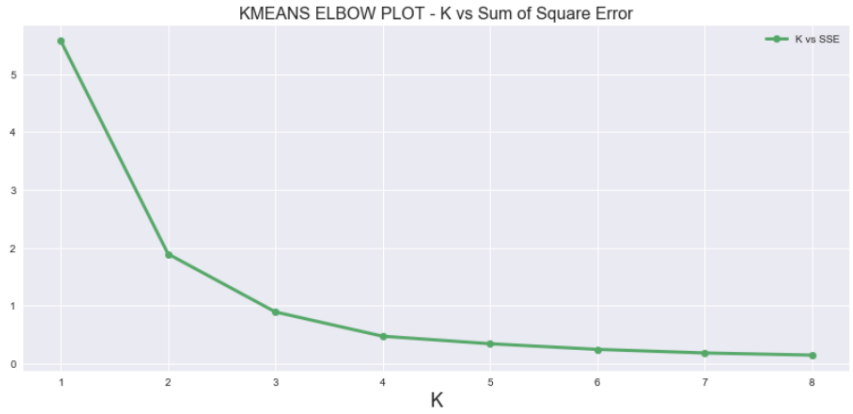
HAMMEROFLIGHT
__1.4.4.2__

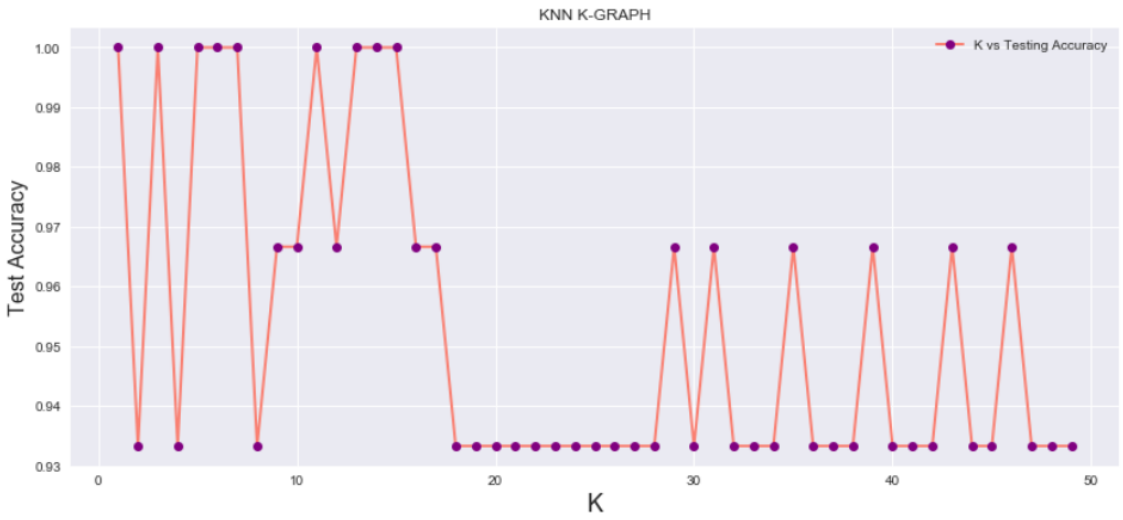
```
In [9]: from hammeroflight.modelfitter import kmeans_kfinder
```

```
In [11]: kmeans_kfinder(df1)
```



K value seems best at 2 - 4. We will test clustering for all three values.

```
: from hammeroflight.modelfitter import knn_kfinder
k = knn_kfinder(X_train, X_test, y_train, y_test, 1, 50)
```




```
]: from hammeroflight.modelfitter import fit_regress
fit_regress(lr, X_train, X_test, y_train, y_test)
```

]:

	Score
Training Score	64.1401
Test Score	62.5071
RMSE	5.76576
Fit	Good Fit

HAMMEROFLIGHT
__1.4.4.2__

```
A = df['LSAT']
B = df['RM']
C = df['MEDV']
from hammeroflight.modelfitter import fittingplot
fittingplot(lr, A, C)
```


```
from hammeroflight.arufunctions import qualityreport, cleanandencode, featureselector
from hammeroflight.modelfitter import fit_classify, fittingplot
from hammeroflight.modelcomparator import clf_comparator
```

```
# Viewing Quality report of the dataset.
qualityreport(df)
```

```
Categorical Features: 9 | Numerical Features: 26 | Dataset Shape: (1470, 35) | DataSet Integrity: 100.0 %
```

	Dtype	Available Rows	Missing Values	Percent Missing	Mean-Mode	Min	Max	No. Of Uniques	Unique Values
Age	int64	1470	0	0.0	35	18	60	43	[41, 49, 37, 33, 27, 32, 59, 30, 38, 36, 35, 2...
Attrition	object	1470	0	0.0	No	No	Yes	2	[Yes, No]
BusinessTravel	object	1470	0	0.0	Travel_Rarely	Non-Travel	Travel_Rarely	3	[Travel_Rarely, Travel_Frequently, Non-Travel]
DailyRate	int64	1470	0	0.0	691	102	1499	886	[1102, 279, 1373, 1392, 591, 1005, 1324, 1358,...
Department	object	1470	0	0.0	Research & Development	Human Resources	Sales	3	[Sales, Research & Development, Human Resources]
DistanceFromHome	int64	1470	0	0.0	2	1	29	29	[1, 8, 2, 3, 24, 23, 27, 16, 15, 26, 19, 21, 5...
Education	int64	1470	0	0.0	3	1	5	5	[2, 1, 4, 3, 5]
EducationField	object	1470	0	0.0	Life Sciences	Human Resources	Technical Degree	6	[Life Sciences, Other, Medical, Marketing, Tec...
EmployeeCount	int64	1470	0	0.0	1	1	1	1	[1]
EmployeeNumber	int64	1470	0	0.0	1	1	2068	1470	[1, 2, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16,...
EnvironmentSatisfaction	int64	1470	0	0.0	3	1	4	4	[2, 3, 4, 1]

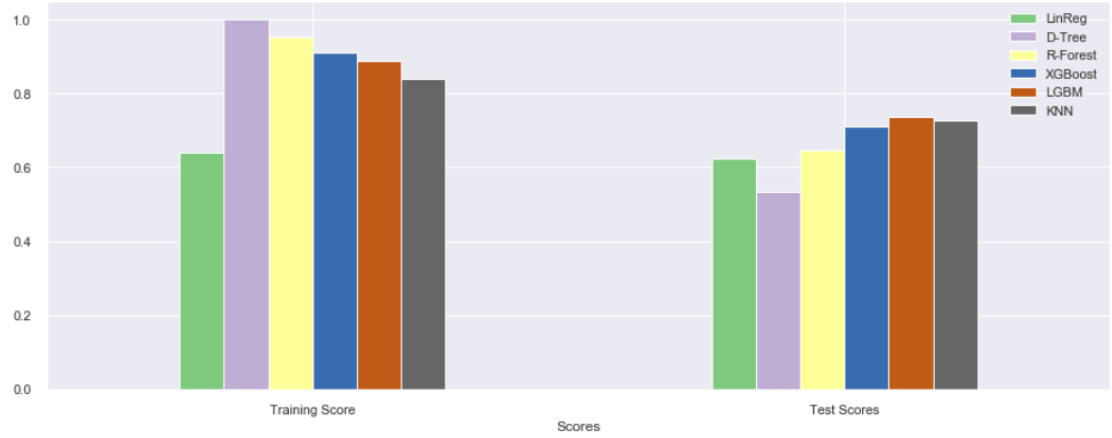
HAMMEROFLIGHT
__1.4.4.2__

```
from hammeroflight.modelcomparator import reg_comparator
from hammeroflight.modelfitter import fit_regress
```

```
reg_comparator(X_train, X_test, y_train, y_test)
```

```
[12:00:11] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now depr
ecated in favor of reg:squarederror.
```

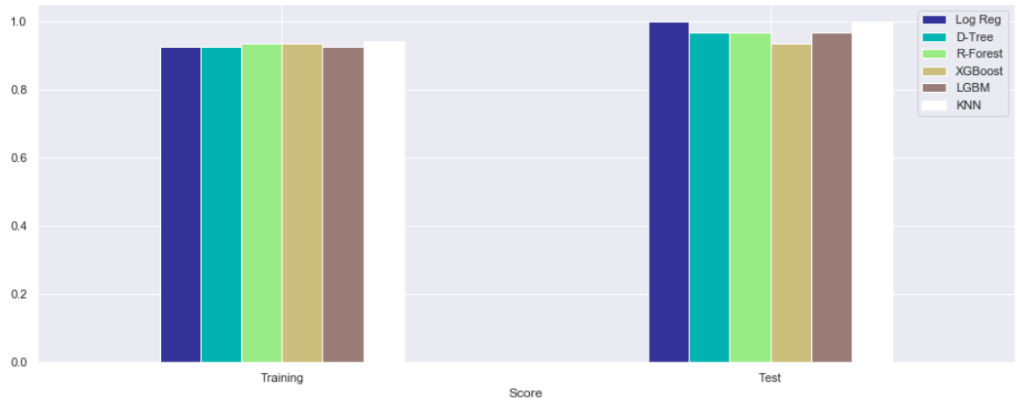
Scores	LinReg	D-Tree	R-Forest	XGBoost	LGBM	KNN
Training Score	0.641401	1.000000	0.953221	0.910335	0.888643	0.839217
Test Scores	0.625071	0.534098	0.647935	0.710684	0.737306	0.725607
RMSE	5.765800	6.427300	5.587200	5.064900	4.826200	4.932500



In [8]:
```
from hammeroflight.modelcomparator import clf_comparator
clf_comparator(X_train, X_test, y_train, y_test, 10)
```

Out[8]:

Score	Log Reg	D-Tree	R-Forest	XGBoost	LGBM	KNN
Training	0.925	0.925000	0.933333	0.933333	0.925000	0.941667
Test	1.000	0.966667	0.966667	0.933333	0.966667	1.000000



HAMMEROFLIGHT
__1.4.4.2__

	Emp_ID	Name	Age	Income	Department	Posting
0	P001	Aru	35	11000.0	AI	Tier 1
1	P002	Mahesh	28	6000.0	Sales	Tier 2
2	P003	Ranjit	36	9000.0	ML	NaN
3	P004	Abhishek	34	8700.0	Marketing	Tier 2
4	P005	Supriya	36	13000.0	Top Management	Tier 1

```
from hammeroflight.arufunctions import impute_encode
df = impute_encode(df)
df.head()
```

	Age	Income	Department	Posting
0	35	11000.0	AI	0
1	28	6000.0	Sales	1
2	36	9000.0	ML	0
3	34	8700.0	Marketing	1
4	36	13000.0	Top Management	0

EMP_ID dropped

Posting Label Encoded

Department not touched

Missing Values imputed by mean/mode

```
df = impute_encode(df, dummy=True)
df.head()
```

	Age	Income	Posting	Department_Assistant	Department_ML	Department_Marketing	Department_Sales	Department_Top Management
0	35	11000.0	0	0	0	0	0	0
1	28	6000.0	1	0	0	0	1	0
2	36	9000.0	0	0	1	0	0	0
3	34	8700.0	1	0	0	1	0	0
4	36	13000.0	0	0	0	0	0	1

```
qr(df)
```

```
Categorical Features: 0 | Numerical Features: 8 | Dataset Shape: (19, 8) | DataSet Integrity: 100.0 %
```

	Dtype	Available Rows	Missing Values	Percent Missing	Mean-Mode	Min	Max	No. Of Uniques	Unique Values
Age	int64	19	0	0.0	37.315789	28.0	63.0	13	None
Income	float64	19	0	0.0	9508.333333	3400.0	22000.0	18	None
Posting	int32	19	0	0.0	0.736842	0.0	2.0	3	None
Department_Assistant	uint8	19	0	0.0	0.157895	0.0	1.0	2	None
Department_ML	uint8	19	0	0.0	0.263158	0.0	1.0	2	None
Department_Marketing	uint8	19	0	0.0	0.105263	0.0	1.0	2	None
Department_Sales	uint8	19	0	0.0	0.157895	0.0	1.0	2	None
Department_Top Management	uint8	19	0	0.0	0.157895	0.0	1.0	2	None