First few rows:

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | \ |
|---|---|---|---|---|---|---|---|
| | 6 | 148 | 72 | 35 | e | 33 .6 | |
| 1 | 1 | 85 | 66 | 29 | e | 26.6 | |
| 2 | 8 | 183 | 64 | | e | 23 . 3 | |
| 3 | 1 | 89 | 66 | 23 | 94 | 28. 1 | |
| 4 | 0 | 137 | 40 | 35 | 168 | 43 . 1 | |

| | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|
| | e. 627 | 50 | 1 |
| 1 | 0. 351 | 31 | 0 |
| 2 | e. 672 | 32 | 1 |
| 3 | e. 167 | 21 | 0 |
| 4 | 2.288 | 33 | 1 |

Dataset Info:

<class 'pandas. core.frame.DataFrame' >
Rangelndex: 768 entries, e to 767 Data columns (total 9 columns) :

| # | Column | Non-Null Count | Dtype |
|---|---|---|---|
| | ------ | ------------- | ----- |
| e | Pregnancies | 768 non-null | int64 |
| 1 | Glucose | 768 non-null | int64 |
| 2 | BloodPressure | 768 non-null | int64 |
| 3 | SkinThickness | 768 non-null | int64 |
| 4 | Insulin | 768 non-null | int64 |
| 5 | BMI | 768 non-null | float64 |
| 6 | DiabetesPedigreeFunction | 768 non-null | float64 |
| 7 | Age | 768 non-null | int64 |
| 8 | Outcome | 768 non-null | int64 |

dtypes: float64(2), int64(7)
memory usage: 54.1 KB
None
None

Summary Statistics:

| | Pregnancies | G1ucose | B100dPressure | SkinThickness | lnsu1in |
|---|---|---|---|---|---|
| count | 768.00000 0 | 768.000000 | 768.00000 0 | 768.00000 | 768.00000 |
| mean | 3 .845052 | 120. 894531 | 69 . 105469 | 20 . 536458 | 79 .799479 |
| std | 3. 369578 | 31.972618 | 19 . 355807 | 15.952218 | 115.244002 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 1.000000 | 000000 | 62.000000 | 0.000000 | 0.000000 |
| 99 . 117. | 3.000000 6.000000 | 00000072.000000 | 30.500000 | 75% 3.000000 2.000000 | |
| | 7.000000 | 140.250000 | 80.000000 | 9.000000 | 127.250000 |
| max | | 199.000000 | 122.000000 | 846.000000 | |

| | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|
| count | 768.00000 | 768.00000 0 | 768.00000 | 768.00000 |
| mean | 0 31.992578 | Θ . 471876 | 0 33 .240885 | 0 Θ . 348958 |
| std | 7 .884160 | Θ. 331329 | 11.760232 | 9.476951 |
| min | 0.000000 | 0.078000 | 21. 000000 | 25% 27 0.000000 |
| .300000 | Θ. 243750 | 24. 000000% | 32 .000000 | 0. 3725K 0.000000 |
| 29 . 000000 | 75% 36 .600000 | Θ. 626250 | 41.000000 | 0.000000 |
| max | 67 .100000 | 2 .420000 | 81.000000 | 1.000000 1.000000 |

Missing Va1ues:

| | |
|---|---|
| Pregnancies | 0 |
| G1ucose | 0 |
| B100dPressure | 0 |
| SkinThickness | 0 |
| lnsu1in | 0 |
| BMI | 0 |
| DiabetesPedigreeFunct ion | 0 |
| Age | |
| Outcome | |
| dtype: int64 | |

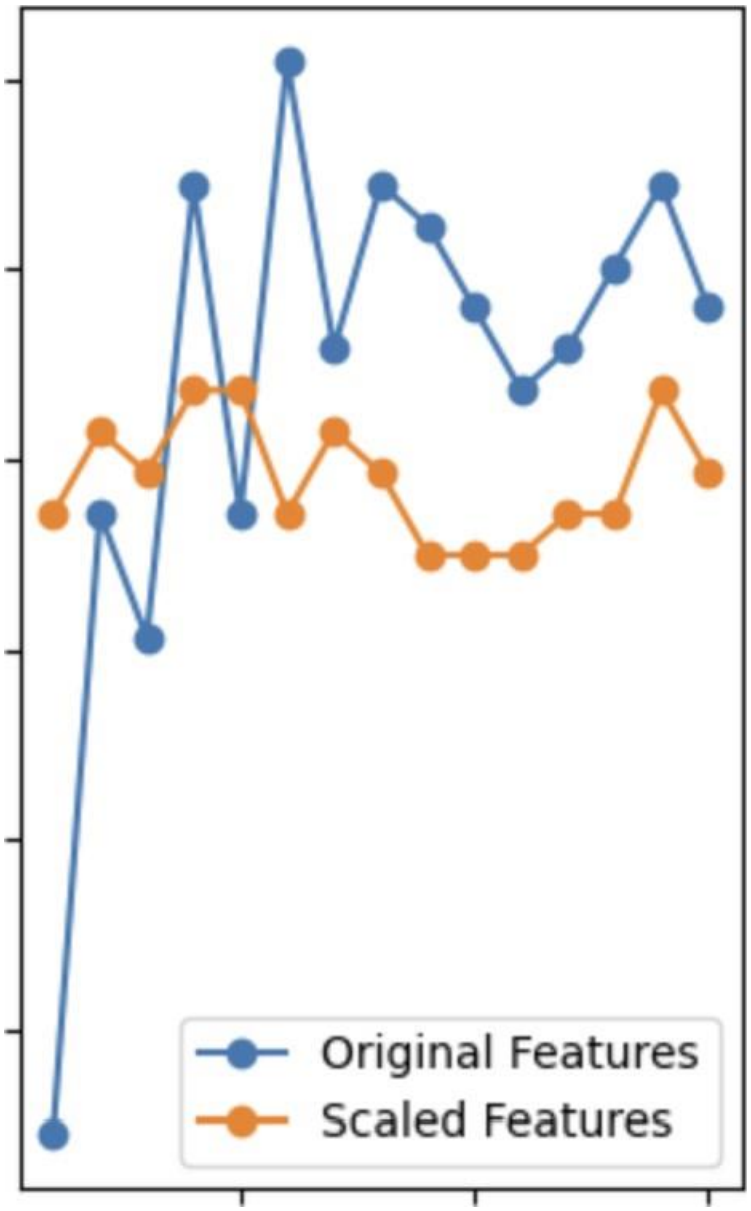
```
Accuracy with k=3: 0.7012987012987013
```

```
Accuracy with k=3 (scaled features): 0.7186147186147186
```

```
<Figure size 1200x600 with 0 Axes>
```

```
<Figure size 1200X600 with 0 Axes>
```

]:<matplotlib.legend.Legend at 0x79a3f4b1d8d0> k vs Accuracy




0.76

0.74

0.72

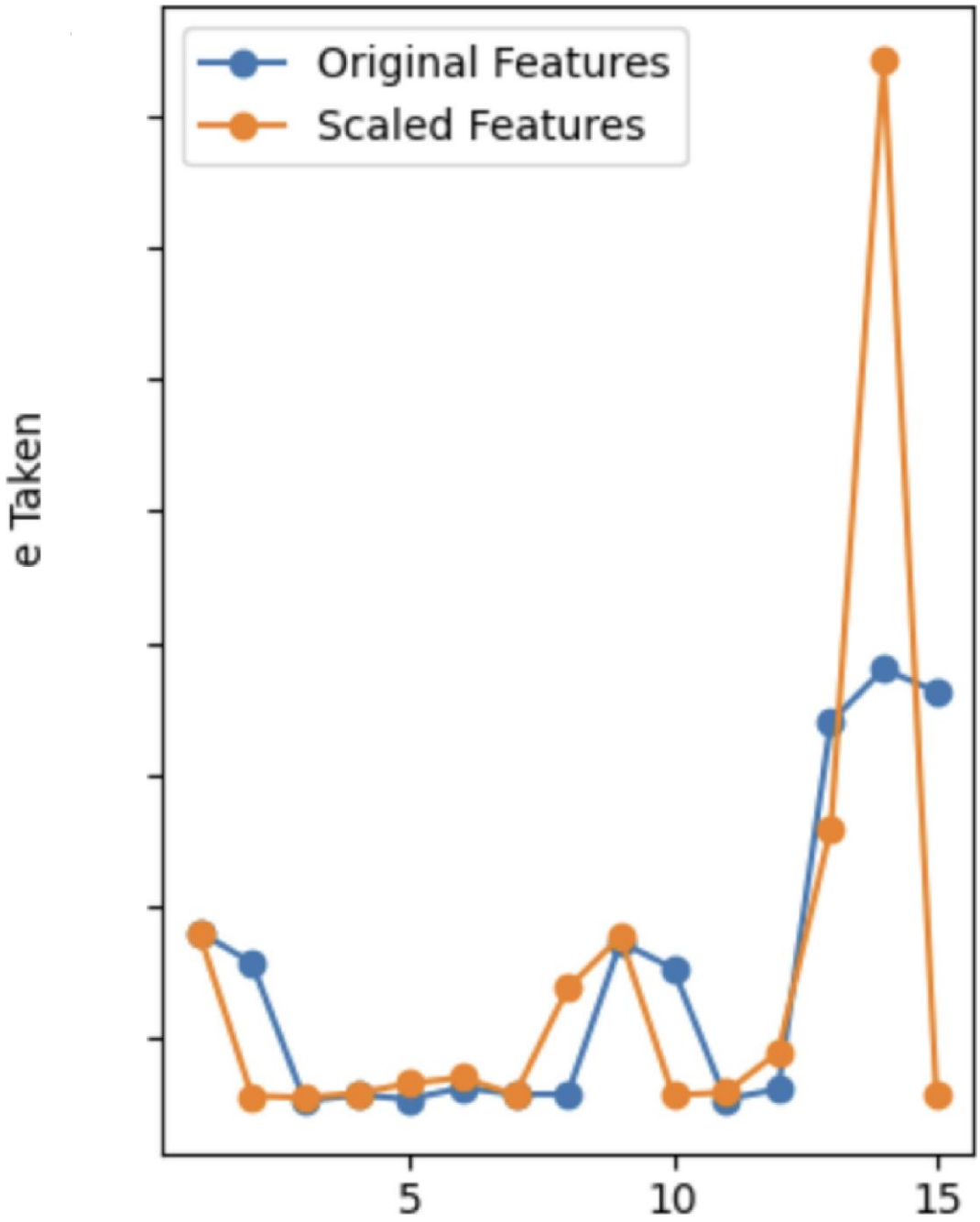
u 0.70

0.68

0.66

5 10 15 k

k vs Time Taken


Original Features
Scaled Features
e Taken
5
10
15
4.5
4.0

3.5

3.0

2.5

2.0

1.5

1.0

Discussion:

1 . Scaling the features significantly impacted the performance of KNN. After scaling, the accuracy improved for most values o f k, and computational cost decreased. 2. The choice of k directly impacts both accuracy and time taken. Lower k tends to overfit, while higher k may underfit.

3 . Optimal k based on the analysis: k=