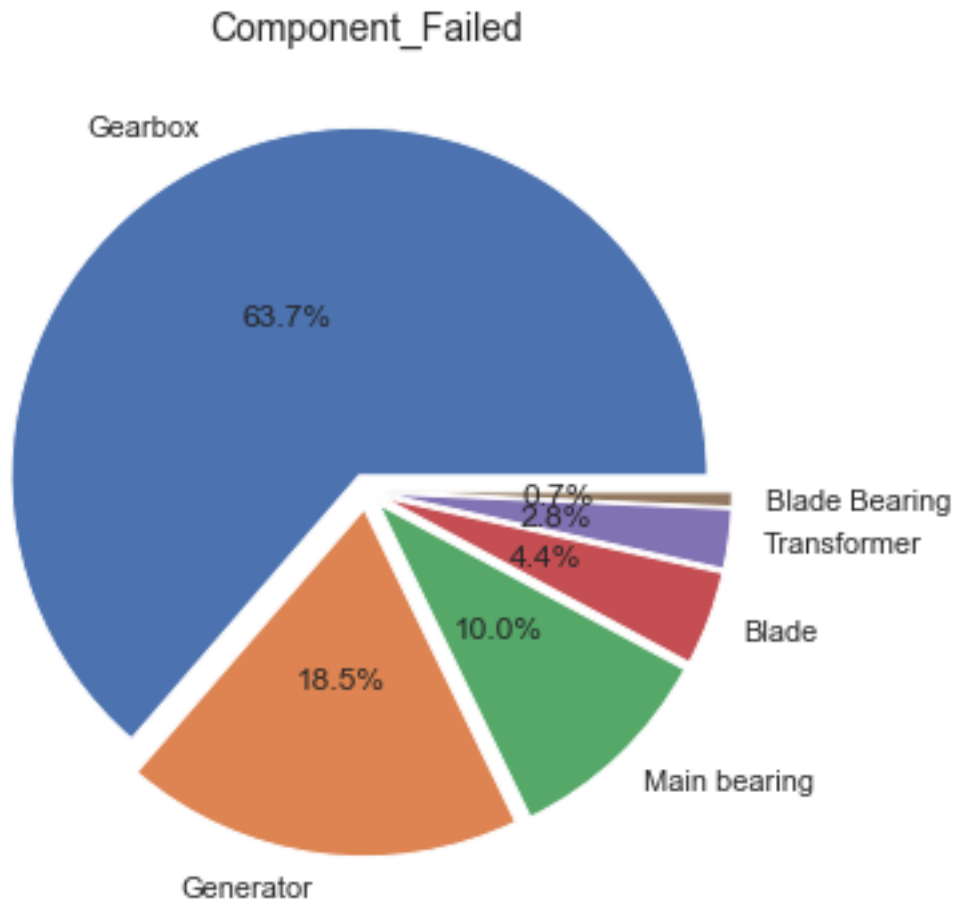


ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE

This project focuses on predicting the failure of various components of turbines located across 5 countries (abbreviated as *UK, GE, DK, NL, SE*) including *on shore* as well as *off shore* turbines.

The components that are analysed in the data are *blade, gearbox, generator, Blade and main bearing*.



This data poses a *Predictive Maintenance* use-case. Every machine or components come with limited life. There can be various reasons for a failure of a component or reduced life of a component. Knowing beforehand when a given component would fail or in other words knowing remaining useful life of components in production line is of great importance.

Remaining Useful Life (RUL) states the duration of a component to reach its failure [1]. By taking RUL into account, engineers can schedule predictive maintenance, optimize operating efficiency, and avoid unplanned downtime. For this reason, estimating RUL is a top priority in the predictive maintenance program.

There are total 450 failure events recorded in the *Vattenfall data*.

Project commissioning is the process of assuring that all systems and components of a building or industrial plant are designed, installed, tested, operated, and maintained according to the operational requirements of the owner or final client[2]. For simplicity, we take commissioning date as the date the turbine is operational.

Component Exchange Date signifies the date till which the component fails, since exact date of failure is not available, assuming the components are replaced immediately after failure.

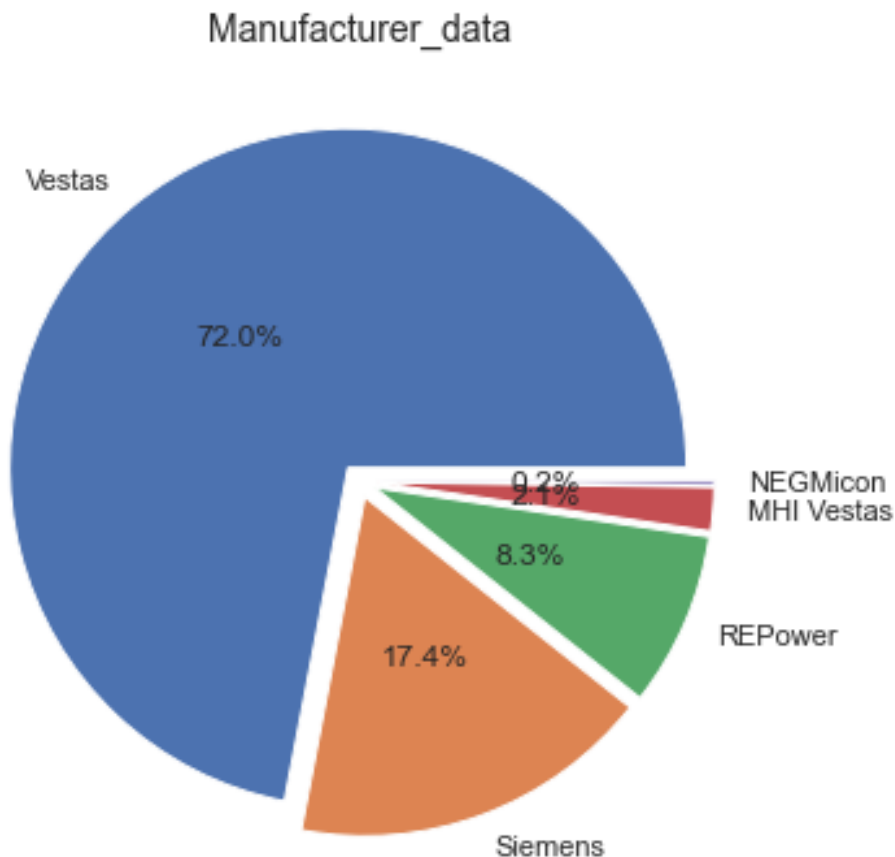
ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE

The difference between *the Component Commissioning Date* and *the Component Exchange Date* gives us approximately the total number of days the component worked before failing *i.e. RUL of each component*.

Limitations of the Data:

1. The data has *too many missing values*.
2. There is *no value indicating the performance deterioration* of the components over time. Alarm Code is one indicator close to gauging the performance of the components, however alarm code is absent in majority of the data points.
3. The Avg RUL is significantly different by country, manufacturer, component. All are used as inputs to predicting the RUL of the component.

The data is disproportional by number of training examples from different manufacturer. The manufacturer is one of the values analysed to predict the life of the component. This negatively impacts the generalizing the learning when there are more examples of only one kind and very less examples of the other.



ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE

Missing data mappings: The available data has large missing attributes. Here is an overview of missing values in the data.

Your selected dataframe has 42 columns.
There are 33 columns that have missing values.

:

	Missing Values	% of Total Values
Alarm code 20	449	100.0
Alarm code 19	449	100.0
Alarm code 18	448	99.8
Alarm code 17	447	99.6
Alarm code 16	447	99.6
Alarm code 15	447	99.6
Alarm code 14	447	99.6
Alarm code 13	445	99.1
Alarm code 12	445	99.1
Alarm code 11	445	99.1
Alarm code 10	444	98.9
Alarm code 9	443	98.7
Alarm code 8	443	98.7
Alarm code 7	442	98.4
Alarm code 6	440	98.0
Alarm code 5	439	97.8
Alarm code 4	435	96.9
Alarm code 3	431	96.0
Alarm code 2	395	88.0
Jackup Call-off Date	387	86.2
Work Order	376	83.7
Outage cause	364	81.1
Jackup vessel	361	80.4
Component Commissioning Date	360	80.2
Turbine start date	350	78.0
Alarm code 1	338	75.3
Material number	304	67.7
Failure Mode	301	67.0
Turbine Stopp Date	268	59.7
Failed Component Serial number	234	52.1
New Component Serial number	114	25.4
Turbine	13	2.9
Estimated Lost Revenue	8	1.8

The data is spread into two sheets:

1. DATABASE i.e. Sheet_1
2. Turbine Data i.e. Sheet_2

The missing component commissioning date of Sheet_1 can be mapped to the corresponding commission year of the turbines provided in Sheet_2. Since only commissioning year is mentioned, the day and month is assumed to be 01-January-YEAR.

All the turbines belonging to the farms *Nørrekær Enge*, *Kentish Flats*, *Princess Alexia*, *Stor Rotliden* have same rest of the attributes (such as 'Manufacturer', 'Turbine Type', 'Rotor Diameter', 'Hub Height', 'Installed Power', 'Latitude', 'Longitude'). Hence these turbines can be given a dummy name to map with corresponding values in Sheet_2.

ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE

```
# Average RUL by Manufacturer
```

```
Avg_RUL_Manufacture= data_prepared.groupby('Manufacturer').mean()['RUL']  
Avg_RUL_Manufacture
```

```
Manufacturer  
MHI Vestas      896.111111  
NEGMicon      2983.000000  
REPower      1753.611111  
Siemens      2294.853333  
Vestas      2574.000000  
Name: RUL, dtype: float64
```

The empty columns (100% missing) or beyond possibility of imputing (say 85% missing) are dropped.

The cost involved due to the component failure have no role in analysis. Those info are also kept away from the model building.

The 'Outage Cause' is also missing for most of the data.

```
data_prepared['Outage cause'].value_counts()
```

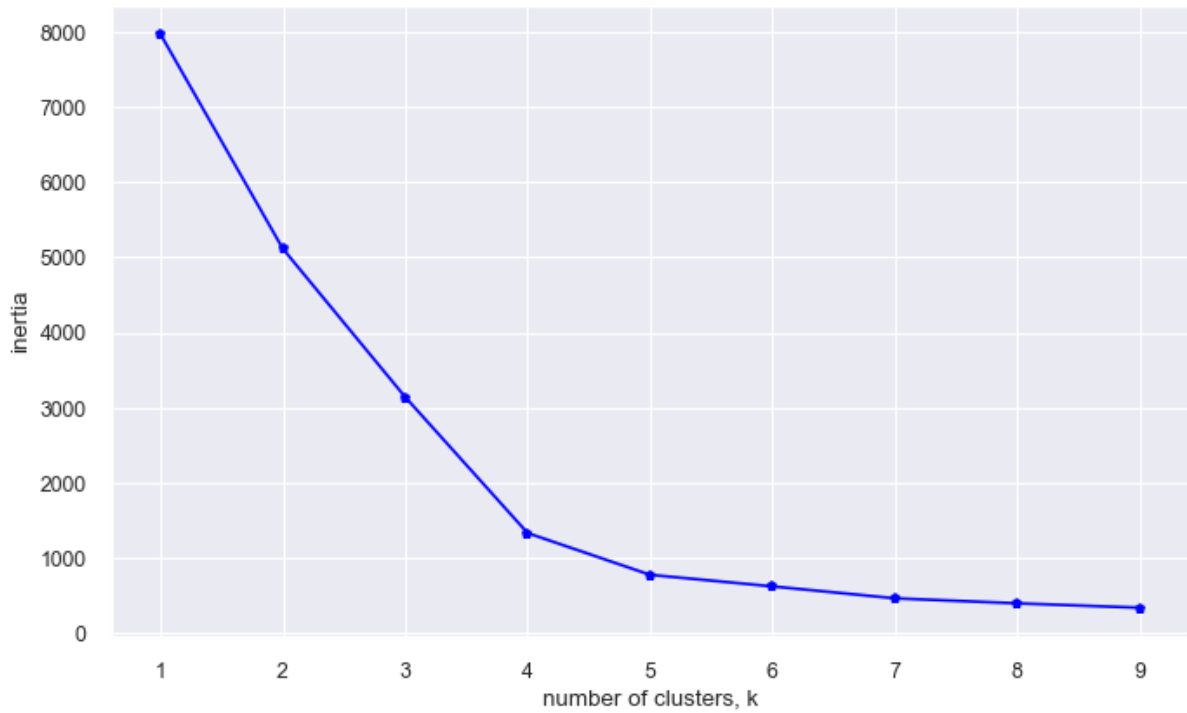
```
unknown    347  
Planned     48  
Forced      37
```

Note: Despite of the huge limitations posed by limited data, the model building process has been reasonably performing.

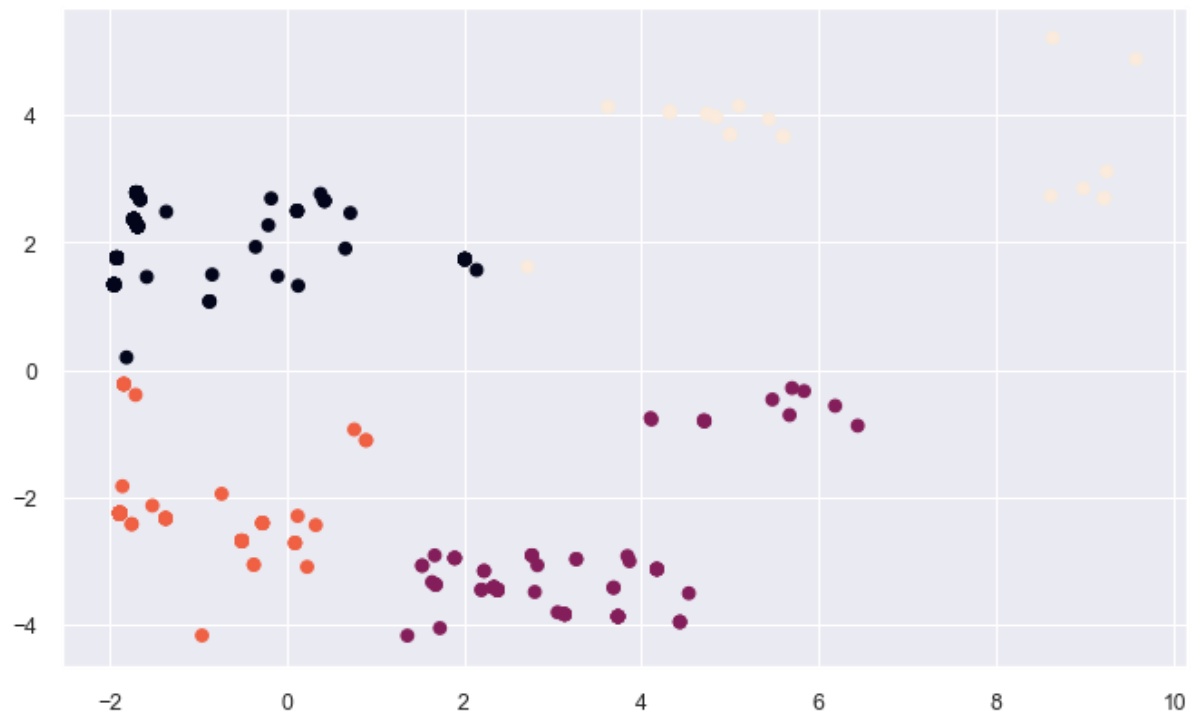
Visualization via Clustering:

To visualize the data, reducing the dimensionality of the data can help.

ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE



The optimal number of clusters via Elbow Method is 4.



The 4 clusters are recognized.

Model highlights:

Performance measurement benchmark: R^2_score

ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE

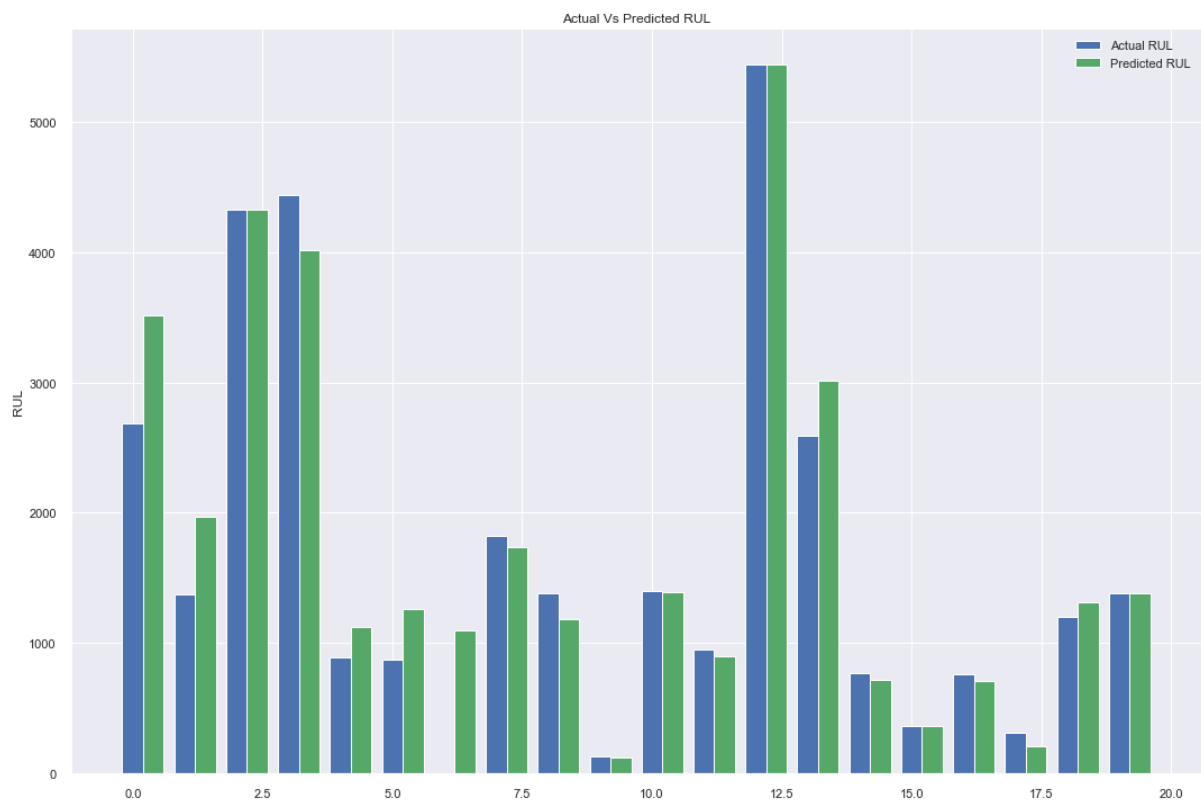
R^2 (coefficient of determination) regression score function.

Best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse). A constant model that always predicts the expected value of y , disregarding the input features, would get a R^2 score of 0.0.

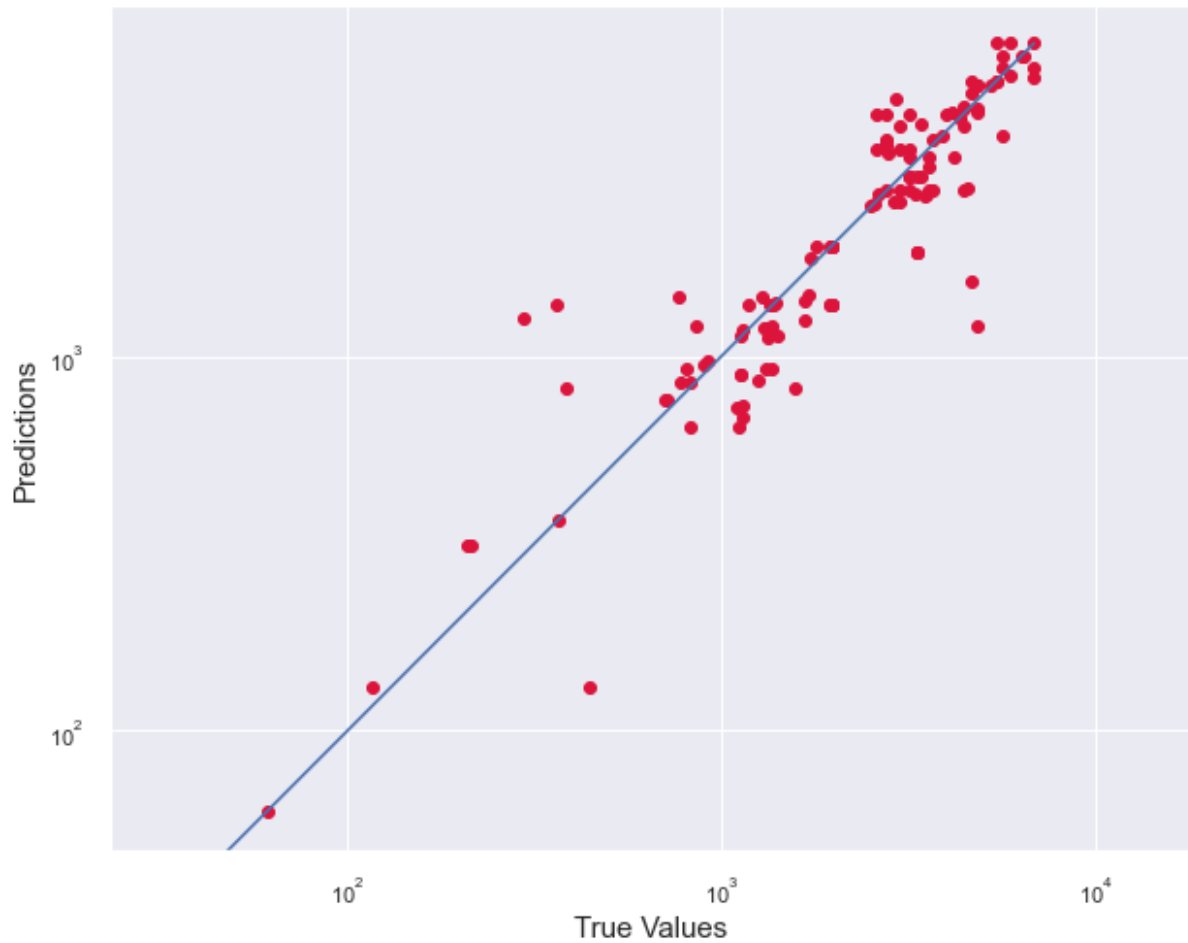
Regression models:

Decision tree resulted in performance is just 5%. Hence we can discard the model.

KNN regressor fits with 79% R^2_{score} (0.79), which could correctly predict RUL values for most of the data. Here is sample plotting for 20 data points.



ANALYSIS REPORT OF VATTENFALL PREDICTIVE MAINTENANCE



References:

1. <https://bmachsan.medium.com/remaining-useful-life-predictive-maintenance-ccdf40580216>
2. https://en.wikipedia.org/wiki/Project_commissioning
3. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html#sklearn.metrics.r2_score
4. <https://andrewmourcos.github.io/blog/2019/06/06/PCA.html>
5. <https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/>