kmeans

October 26, 2022

[49]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

[50]:
```
data= pd.read_csv("driver-data.csv")
```

[51]:
```
data.head()
```

[51]:

	id	mean_dist_day	mean_over_speed_perc
0	3423311935	71.24	28
1	3423313212	52.53	25
2	3423313724	64.54	27
3	3423311373	55.69	22
4	3423310999	54.58	25

[52]:
```
data.shape
```

[52]:
```
(4000, 3)
```

[53]:
```
data=data.drop("id", axis=1)
```

[54]:
```
data.info()
```

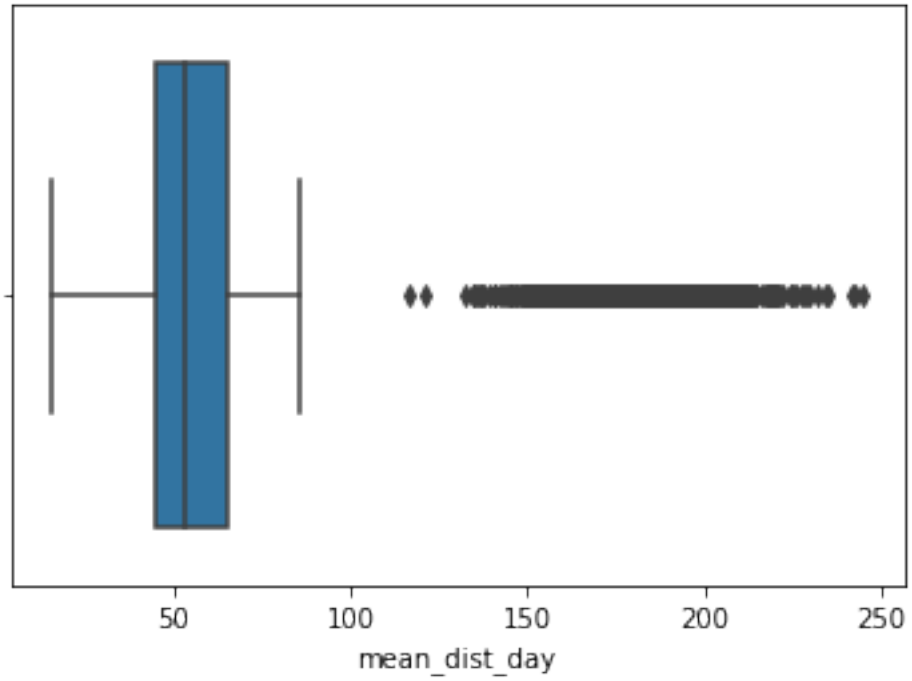
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4000 entries, 0 to 3999
Data columns (total 2 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   mean_dist_day         4000 non-null   float64
 1   mean_over_speed_perc  4000 non-null   int64
dtypes: float64(1), int64(1)
memory usage: 62.6 KB
```

[55]:
```
sns.boxplot("mean_dist_day", data=data)
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning:
Pass the following variable as a keyword arg: x. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an
```

```
explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

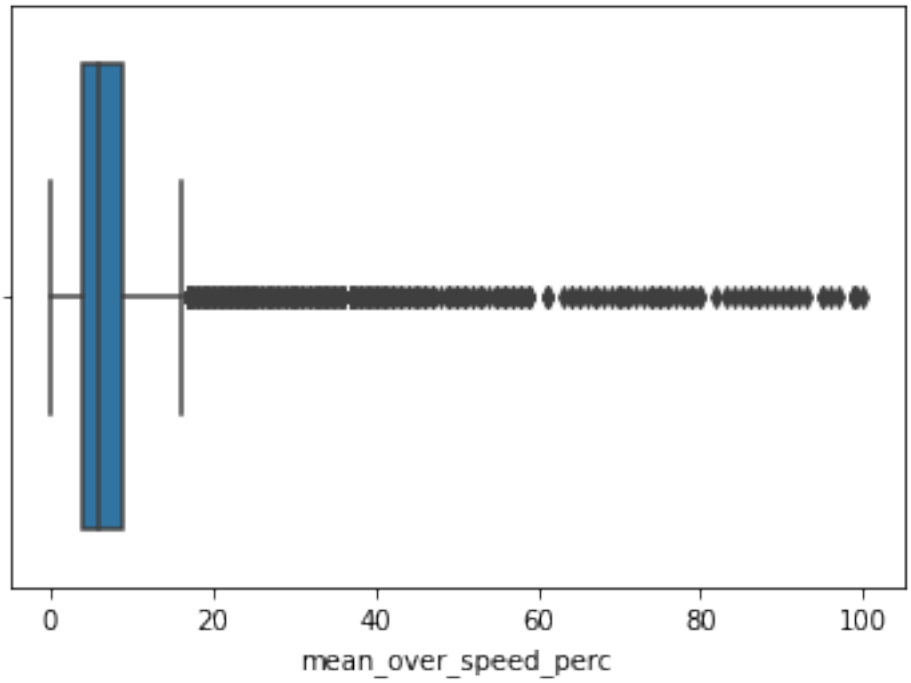
[55]: <AxesSubplot:xlabel='mean_dist_day'>



[56]:
```
sns.boxplot("mean_over_speed_perc", data=data)
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning:
Pass the following variable as a keyword arg: x. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an
explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

[56]: <AxesSubplot:xlabel='mean_over_speed_perc'>



[57]:
```
from sklearn.cluster import KMeans
```

[58]:
```
cls= KMeans(n_clusters=2)
```

[59]:
```
cls.fit(data)
```

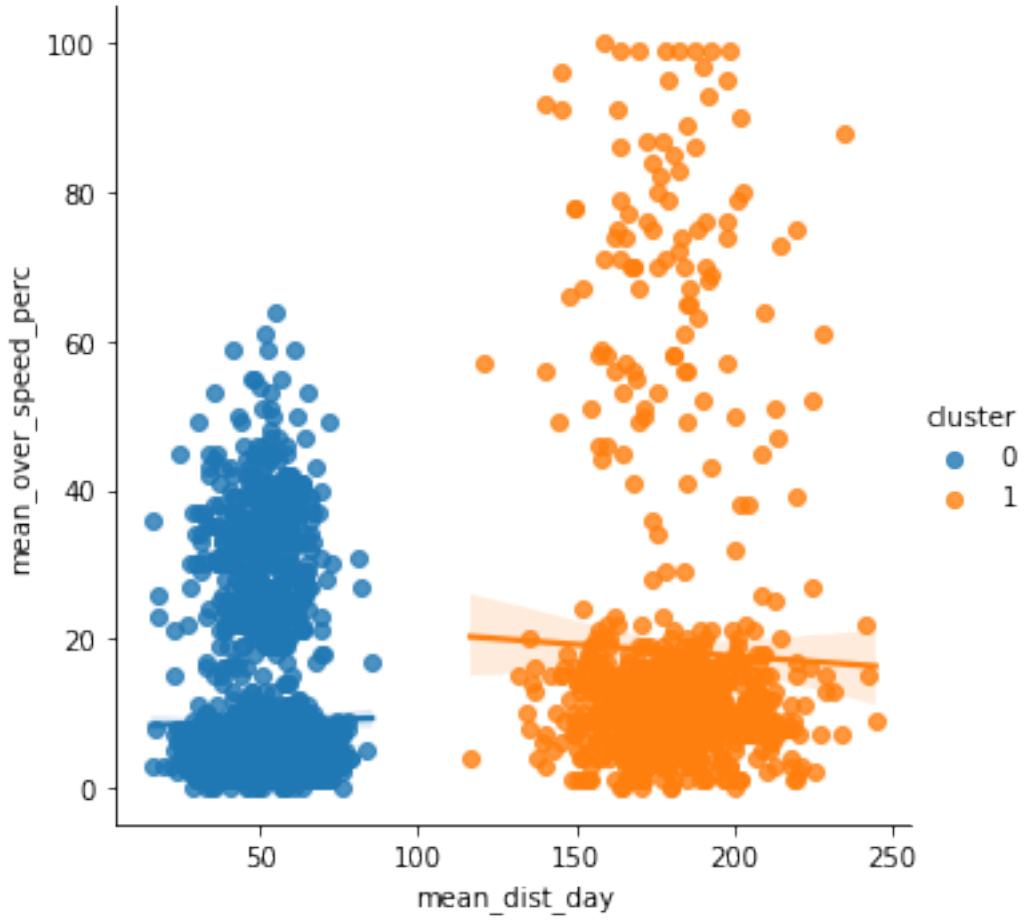
[59]:
```
KMeans(n_clusters=2)
```

[60]:
```
data['cluster']= cls.labels_
```

[61]:
```
sns.lmplot("mean_dist_day","mean_over_speed_perc", data=data, hue= "cluster")
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning:
Pass the following variables as keyword args: x, y. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an
explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

[61]:
```
<seaborn.axisgrid.FacetGrid at 0x7fb5f8468550>
```



[62]:
```
cls= KMeans(n_clusters=3)
```

[63]:
```
cls.fit(data)
```

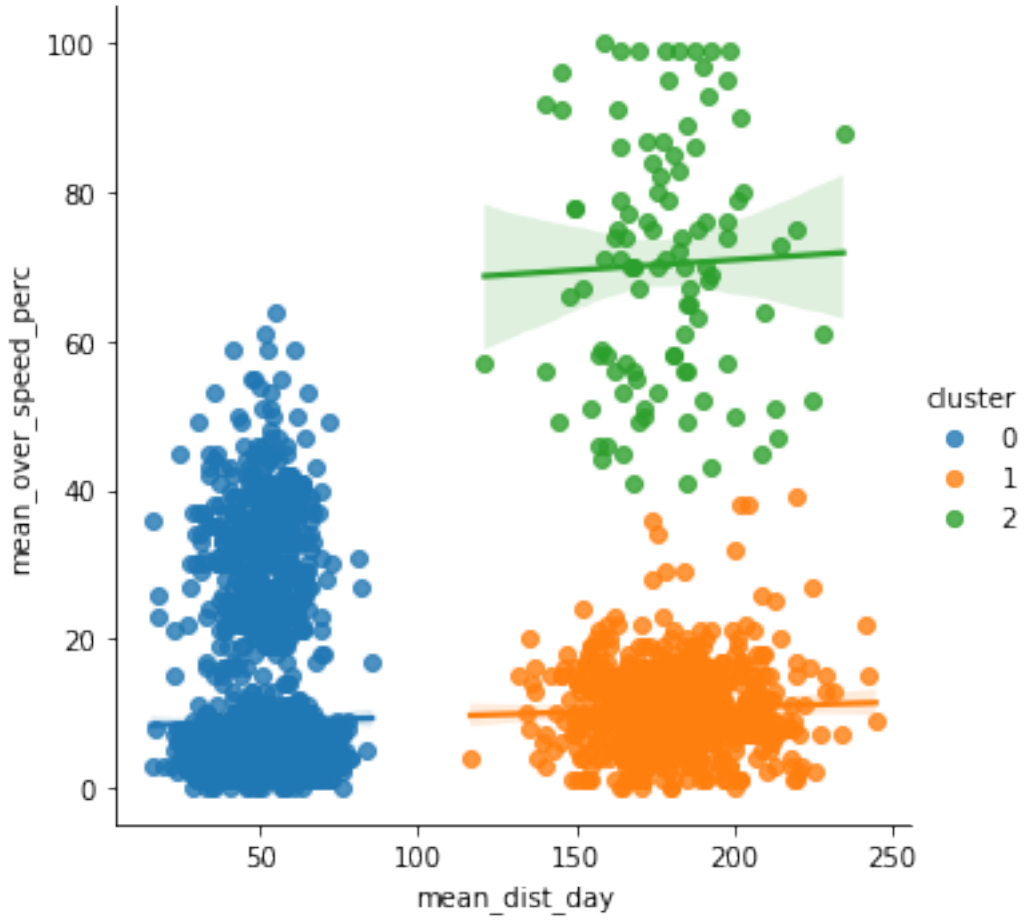
[63]:
```
KMeans(n_clusters=3)
```

[64]:
```
data['cluster']= cls.labels_
```

[65]:
```
sns.lmplot("mean_dist_day","mean_over_speed_perc", data=data, hue= "cluster")
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning:
Pass the following variables as keyword args: x, y. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an
explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

[65]:
```
<seaborn.axisgrid.FacetGrid at 0x7fb5f8409e10>
```



[66]:
```
cls= KMeans(n_clusters=4)
```

[67]:
```
cls.fit(data)
```

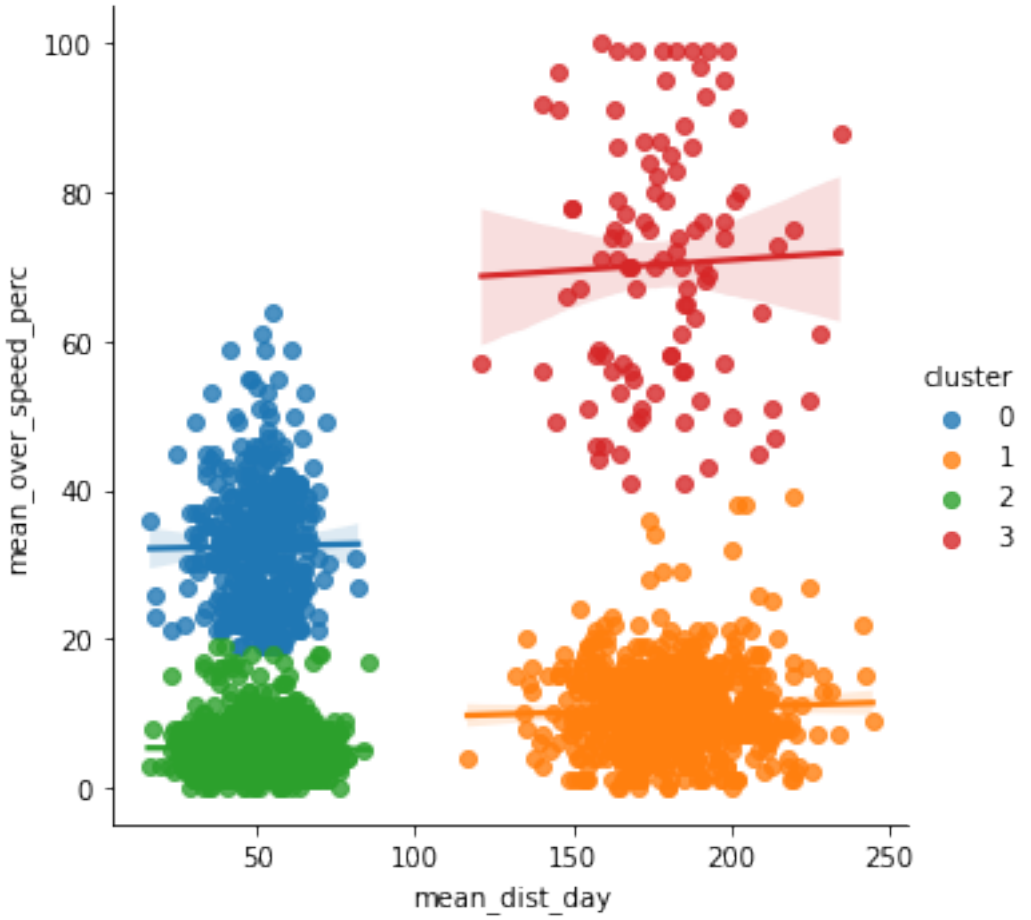
[67]:
```
KMeans(n_clusters=4)
```

[68]:
```
data['cluster']= cls.labels_
```

[69]:
```
sns.lmplot("mean_dist_day","mean_over_speed_perc", data=data, hue= "cluster")
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning:
Pass the following variables as keyword args: x, y. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an
explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

[69]:
```
<seaborn.axisgrid.FacetGrid at 0x7fb5f8310850>
```


100
80
60
40
20
0
mean_over_speed_perc
50
100
150
200
250
mean_dist_day
cluster
0
1
2
3

[ ]: