



KNN Classification for predicting accident by diseases





วัตถุประสงค์

เพื่อทำนายนายระดับอุบัติเหตุน
จากจำนวนของโรคแต่ละประเภท



1.ผู้ช่วยมารับบริการด้านจิตเวช ประจำปี 2562

(อ้างอิง :

2.อุบัติเหตุทางถนน

(อ้างอิง : data.go.th/dataset/roadaccident)





วิธีการดำเนินงาน



ติดตั้ง package และนำเข้าไฟล์ข้อมูล

เลือกข้อมูลที่ต้องการ

รวมข้อมูลทั้ง 2 เป็นตารางเดียวกัน



วิธีการดำเนินงาน



กำจัด Missing values

Remove Outlier (boxplot)

วิเคราะห์ข้อมูลจากกราฟ

KNKNeighborsClassifier



ติดตั้ง package และนำเข้าไฟล์ข้อมูล



```
import pandas
```

```
import pandas as pd
```

```
from google.colab import drive
```

```
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
import os
```

```
base_path = '/content/drive/My Drive/DWDM_material'
```

```
file_path_1 = os.path.join(base_path, 'Project1.xlsx')
```

```
table_1 = pd.read_excel(file_path_1)
```

```
table_1
```

```
file_path_2 = os.path.join(base_path, 'Project2.xlsx')
```

```
table_2 = pd.read_excel(file_path_2)
```

```
table_2
```



ติดตั้ง package และนำเข้าไฟล์ข้อมูล



disease = table_1

disease

	จังหวัด	โรค สมอง เสื่อม	ติดแอลกอฮอล์	ติดยาบ้า (Amphetamine)	ติดสารเสพติด อื่นๆ	โรคจิตเภท	โรคจิต อื่นๆ	โรค อารมณ์ สองขั้ว	โรค ซึม เศร้า	โรค วิตกกังวล	ความ บกพร่อง ทางสติ ปัญญา	ความ บกพร่อง ทางการ เรียนรู้	โรค คอ อหิ สติก	โรค สมาธิ สั้น	พยายามฆ่า ตัวตาย (การ ตั้งใจ ทำร้าย ตนเอง)	ผู้ป่วย ติด เกมส์ ใน ผู้ใหญ่ (15 ปี ขึ้นไป)	ผู้ป่วย ติด เกมส์ ใน เด็ก (อายุ ต่ำกว่า 15 ปี)	โรค ท เว อี่
0	กระบี่	168	221	609	1166	1597	2195	0	1529	1810	285	265	282	720	70	0	6	21
1	กรุงเทพมหานคร	3002	1972	2001	2226	7602	9105	15	12083	10984	1110	1042	2127	4433	637	133	184	19
2	กาญจนบุรี	249	1898	1667	975	1710	2071	0	2228	1683	204	175	172	607	336	0	5	53
3	กาฬสินธุ์	323	1702	2128	1480	4283	6584	0	3177	5694	239	250	175	656	224	8	15	59
4	กำแพงเพชร	146	1799	1251	668	1714	2309	3	2634	1944	144	319	119	403	249	0	11	38
...
73	อุดรธานี	614	2769	3284	1840	5457	6476	3	4702	4911	252	223	335	574	240	5	9	61
74	อุดรดิษฐ์	86	560	767	572	788	1039	0	1474	754	65	291	62	397	78	0	2	21
75	อุทัยธานี	195	1060	478	896	795	1328	0	1717	1516	115	44	51	197	144	1	8	40
76	อุบลราชธานี	698	7480	8674	2607	7661	9909	40	10761	12236	857	417	452	1349	275	4	61	11
77	ไม่ทราบ	10190	122883	39417	27747	75747	94877	84	73639	112183	6160	8346	6778	21652	5815	49	252	19

78 rows × 19 columns



ติดตั้ง package และนำเข้าไฟล์ข้อมูล



```
accident = table_2
```

```
accident
```

	ปีที่เกิดเหตุ	วันที่เกิดเหตุ	เวลา	วันที่รายงาน	เวลาที่รายงาน	ACC_CODE	หน่วยงาน	สายทาง	ก.ม.	จังหวัด	รถคันที่ 1	บริเวณที่เกิดเหตุ/ลักษณะทาง	มูลเหตุสันนิษฐาน	ลักษณะการเกิดอุบัติเหตุ	จำนวนรถที่เกิดเหตุ(รถคันที่ 1)
0	2019	01/01/2019	00:00	02/01/2019	06:11	201901023254	กรมทางหลวงชนบท	ทางหลวงชนบท ลพ.2029	4.00	ลพบุรี	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน/ที่ราบ	เมาสุรา	อื่น ๆ	1
1	2019	01/01/2019	00:03	13/02/2019	12:06	201902135341	กรมทางหลวง	ทางหลวงแผ่นดิน หมายเลข 24 ตอนควบคุม 0702	419.77	อุบลราชธานี	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง	ขับรถเร็วเกินกำหนด	ชนอุปกรณ์งานทาง	1
2	2019	01/01/2019	00:05	01/01/2019	09:52	201901012979	กรมทางหลวง	ทางหลวงแผ่นดิน หมายเลข 3168 ตอนควบคุม 0100	1.70	ประจวบคีรีขันธ์	รถจักรยานยนต์	ทางโค้งปกติ	ขับรถเร็วเกินกำหนด	ชนกัน/ชนยานพาหนะอื่นๆ	2
3	2019	01/01/2019	00:20	02/01/2019	05:12	201901023234	กรมทางหลวงชนบท	ทางหลวงชนบท ชม.4016	1.88	เชียงใหม่	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน/ที่ราบ	เมาสุรา	รถยนต์ชนรถไฟ	1



เลือกข้อมูลที่ต้องการ



```
num_accident = accident.groupby('จังหวัด').count().iloc[:,1:2]
```

num_accident

	วันที่เกิดเหตุ
จังหวัด	
กระบี่	205
กรุงเทพมหานคร	1844
กาญจนบุรี	285
กาฬสินธุ์	130
กำแพงเพชร	188
...	...
เพชรบุรี	198
เพชรบูรณ์	395
เลย	196
แพร่	328
แม่ฮ่องสอน	90

77 rows x 1 columns



เลือกข้อมูลที่ต้องการ



```
num_accident.rename(columns={'วันที่เกิดเหตุ': 'จำนวนครั้งที่เกิดเหตุ'}, inplace= True)  
num_accident
```

	จำนวนครั้งที่เกิดเหตุ
จังหวัด	
กระบี่	205
กรุงเทพมหานคร	1844
กาญจนบุรี	285
กาฬสินธุ์	130
กำแพงเพชร	188
...	...
เพชรบุรี	198
เพชรบูรณ์	395
เลย	196
แพร่	328
แม่ฮ่องสอน	90

77 rows x 1 columns



รวมข้อมูลทั้ง 2 เป็นตารางเดียวกัน



```
merged_table = disease.merge(num_accident,how='left',left_on='จังหวัด',right_on='จังหวัด')
```

merged_table

โรค สมอง เสื่อม	ติดแอลกอฮอล์	ติดยาบ้า (Amphetamine)	ติดสารเสพติด อื่นๆ	โรคจิตเภท	โรคจิต อื่นๆ	โรค อารมณ์ สองขั้ว	โรค ซึม เศร้า	โรค วิตกกังวล	ความ บกพร่อง ทางสติ ปัญญา	ความ บกพร่อง ทางการ เรียนรู้	โรค คอ อทิ สติ	โรค สมาธิ สั้น	พยายามฆ่า ตัวตาย (การ ตั้งใจ ทำร้าย ตนเอง)	ผู้ป่วย ติด เกมส์ ใน ผู้ใหญ่ (15 ปี ขึ้นไป)	ผู้ป่วย ติด เกมส์ ใน เด็ก (อายุ ต่ำกว่า 15 ปี)	โรค ทางจิต เวช อื่นๆ	โรค ลมชัก	จำนวน ครั้งที่ เกิด เหตุ
168	221	609	1166	1597	2195	0	1529	1810	285	265	282	720	70	0	6	2149	1231	205.0
3002	1972	2001	2226	7602	9105	15	12083	10984	1110	1042	2127	4433	637	133	184	19838	5646	1844.0
249	1898	1667	975	1710	2071	0	2228	1683	204	175	172	607	336	0	5	5322	1881	285.0
323	1702	2128	1480	4283	6584	0	3177	5694	239	250	175	656	224	8	15	5914	2143	130.0
146	1799	1251	668	1714	2309	3	2634	1944	144	319	119	403	249	0	11	3897	1779	188.0
...
614	2769	3284	1840	5457	6476	3	4702	4911	252	223	335	574	240	5	9	6121	3626	206.0
86	560	767	572	788	1039	0	1474	754	65	291	62	397	78	0	2	2120	693	289.0
195	1060	478	896	795	1328	0	1717	1516	115	44	51	197	144	1	8	4018	903	114.0
698	7480	8674	2607	7661	9909	40	10761	12236	857	417	452	1349	275	4	61	11244	5920	186.0
10190	122883	39417	27747	75747	94877	84	73639	112183	6160	8346	6778	21652	5815	49	252	154895	45030	NaN

;



กำจัด Missing values



```
merged_table.isnull().any()
```

จังหวัด	False
โรคสมองเสื่อม	False
ติดแอลกอฮอล์	False
ติดยาบ้า (Amphetamine)	False
ติดสารเสพติดอื่นๆ	False
โรคจิตเภท	False
โรคจิตอื่นๆ	False
โรคอารมณ์สองขั้ว	False
โรคซึมเศร้า	False
โรควิตกกังวล	False
ความบกพร่องทางสติปัญญา	False
ความบกพร่องทางการเรียนรู้	False
โรคออทิสติก	False
โรคสมาธิสั้น	False
พยายามฆ่าตัวตาย (การตั้งใจทำร้ายตนเอง)	False
ผู้ป่วยติดเกมในผู้ใหญ่ (15 ปีขึ้นไป)	False
ผู้ป่วยติดเกมในเด็ก (อายุต่ำกว่า 15 ปี)	False
โรคทางจิตเวชอื่นๆ	False
โรคลมชัก	False
จำนวนครั้งที่เกิดเหตุ	True

dtype: bool

```
merged_table.isnull().any().any()
```

True

```
merged_table_nm = merged_table.dropna()
```




กำจัด Missing values



merged_table_nm

	โรค สมอง เสื่อม	ติด แอลกอฮอล์	ติดยาบ้า (Amphetamine)	ติด สาร เสพติด อื่นๆ	โรคจิตเภท	โรคจิต ตัวอื่นๆ	โรค อารมณ์ สองขั้ว	โรค ซึม เศร้า	โรค วิตกกังวล	ความ บกพร่อง ทางสติ ปัญญา	ความ บกพร่อง ทางการ เรียนรู้	โรค คอ อหิ สติก	โรค สมาธิ สั้น	พยาย ามฆ่า ตัวตาย (การ ตั้งใจ ทำร้าย ตนเอง)	ผู้ป่วย ติด เกมส์ ใน ผู้ใหญ่ (15 ปี ขึ้นไป)	ผู้ ป่วย ติด เกมส์ ใน เด็ก (อายุ ต่ำกว่า 15 ปี)	โรค ทาง จิต เวช อื่นๆ	โรค ลม ชัก	จำนวน ครั้งที่ เกิด เหตุ
	168	221	609	1166	1597	2195	0	1529	1810	285	265	282	720	70	0	6	2149	1231	205.0
หานคร	3002	1972	2001	2226	7602	9105	15	12083	10984	1110	1042	2127	4433	637	133	184	19838	5646	1844.0
รี	249	1898	1667	975	1710	2071	0	2228	1683	204	175	172	607	336	0	5	5322	1881	285.0
	323	1702	2128	1480	4283	6584	0	3177	5694	239	250	175	656	224	8	15	5914	2143	130.0
ชร	146	1799	1251	668	1714	2309	3	2634	1944	144	319	119	403	249	0	11	3897	1779	188.0

ัญ	99	1067	1377	361	1701	2157	0	1932	2080	119	130	62	337	56	0	10	2479	957	51.0
	614	2769	3284	1840	5457	6476	3	4702	4911	252	223	335	574	240	5	9	6121	3626	206.0
	86	560	767	572	788	1039	0	1474	754	65	291	62	397	78	0	2	2120	693	289.0
	195	1060	478	896	795	1328	0	1717	1516	115	44	51	197	144	1	8	4018	903	114.0
กนี้	698	7480	8674	2607	7661	9909	40	10761	12236	857	417	452	1349	275	4	61	11244	5920	186.0

columns



merged_table_nm.isnull().any().any()

False



รวมข้อมูลและจัดระดับจำนวนครั้งที่เกิดอุบัติเหตุ



```
newtable = merged_table_nm.merge(level_accident,how='left',left_on='จังหวัด',right_on='จังหวัด')
```

newtable

ร อง เม	ติด แอล กอฮอล์	ติดยาบ้า (Amphetamine)	ติด สาร เสพติด อื่นๆ	โรคจิตเภท	โรคจิต อื่นๆ	โรค อารมณ์ สองขั้ว	โรค ซึม เศร้า	โรค วิตก กังวล	ความ บกพร่อง ทางสติ ปัญญา	ความ บกพร่อง ทางการ เรียนรู้	โร ค ออทิ สติก	โรค สมาธิ สั้น	พยาย ยามฆ่า ตัวตาย (การ ตั้งใจ ทำร้าย ตนเอง)	ผู้ป่วย ติด เกมส์ ใน ผู้ใหญ่ (15 ปี ขึ้นไป)	ผู้ ป่วย ติด เกมส์ ใน เด็ก (อายุ ต่ำกว่า 15 ปี)	โรค ทาง จิต เวช อื่นๆ	โรค ลม ชัก	จำนวน ครั้งที่ เกิด เหตุ	ระดับการ เกิด อุบัติเหตุ
3	221	609	1166	1597	2195	0	1529	1810	285	265	282	720	70	0	6	2149	1231	205.0	Most
12	1972	2001	2226	7602	9105	15	12083	10984	1110	1042	2127	4433	637	133	184	19838	5646	1844.0	Most
3	1898	1667	975	1710	2071	0	2228	1683	204	175	172	607	336	0	5	5322	1881	285.0	Most
3	1702	2128	1480	4283	6584	0	3177	5694	239	250	175	656	224	8	15	5914	2143	130.0	Moderate
3	1799	1251	668	1714	2309	3	2634	1944	144	319	119	403	249	0	11	3897	1779	188.0	Moderate

	1067	1377	361	1701	2157	0	1932	2080	119	130	62	337	56	0	10	2479	957	51.0	Least
4	2769	3284	1840	5457	6476	3	4702	4911	252	223	335	574	240	5	9	6121	3626	206.0	Most
	560	767	572	788	1039	0	1474	754	65	291	62	397	78	0	2	2120	693	289.0	Most
5	1060	478	896	795	1328	0	1717	1516	115	44	51	197	144	1	8	4018	903	114.0	Moderate
3	7480	8674	2607	7661	9909	40	10761	12236	857	417	452	1349	275	4	61	11244	5920	186.0	Moderate



ตรวจสอบ Missing values



```
newtable.isnull().any()
```

จังหวัด	False
โรคสมองเสื่อม	False
ติดแอลกอฮอล์	False
ติดยาบ้า (Amphetamine)	False
ติดสารเสพติดอื่นๆ	False
โรคจิตเภท	False
โรคจิตอื่นๆ	False
โรคอารมณ์สองขั้ว	False
โรคซึมเศร้า	False
โรควิตกกังวล	False
ความบกพร่องทางสติปัญญา	False
ความบกพร่องทางการเรียนรู้	False
โรคออทิสติก	False
โรคสมาธิสั้น	False
พยายามฆ่าตัวตาย (การตั้งใจทำร้ายตนเอง)	False
ผู้ป่วยติดเกมในผู้ใหญ่ (15 ปีขึ้นไป)	False
ผู้ป่วยติดเกมในเด็ก (อายุต่ำกว่า 15 ปี)	False
โรคทางจิตเวชอื่นๆ	False
โรคลมชัก	False
จำนวนครั้งที่เกิดเหตุ	False
ระดับการเกิดอุบัติเหตุ	False

dtype: bool

```
newtable.isnull().any().any()
```

False



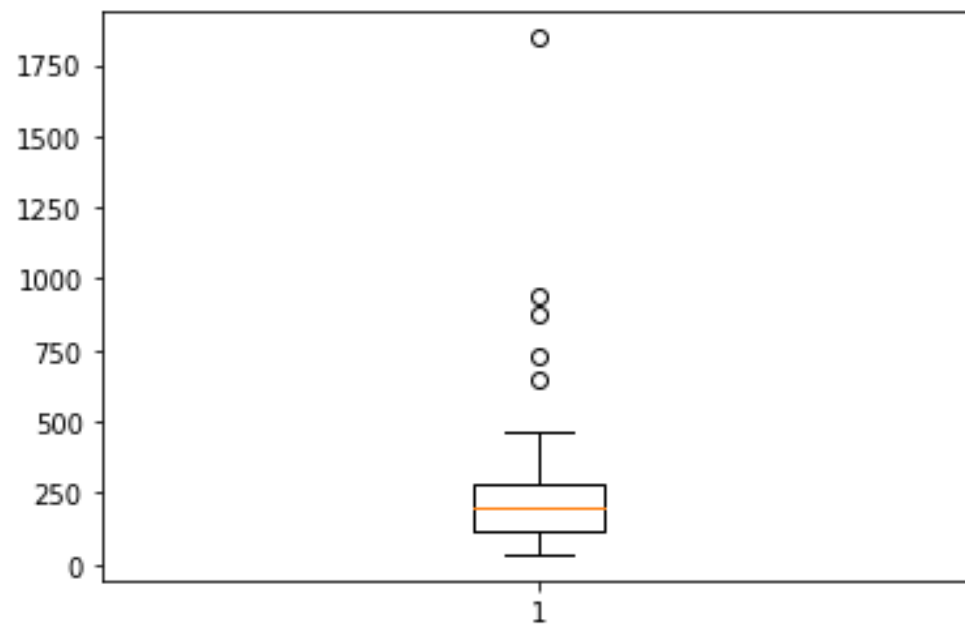
Remove Outlier (boxplot)



```
from matplotlib import pyplot as plt
```

```
plt.boxplot(newtable['จำนวนครั้งที่เกิดเหตุ'])
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7fa5e710e128>],  
'caps': [<matplotlib.lines.Line2D at 0x7fa5e8fbaa20>,  
<matplotlib.lines.Line2D at 0x7fa5e8fba048>],  
'fliers': [<matplotlib.lines.Line2D at 0x7fa5e9ba3eb8>],  
'means': [],  
'medians': [<matplotlib.lines.Line2D at 0x7fa5e8fba898>],  
'whiskers': [<matplotlib.lines.Line2D at 0x7fa5e710e2e8>,  
<matplotlib.lines.Line2D at 0x7fa5e8fbad30>]}
```

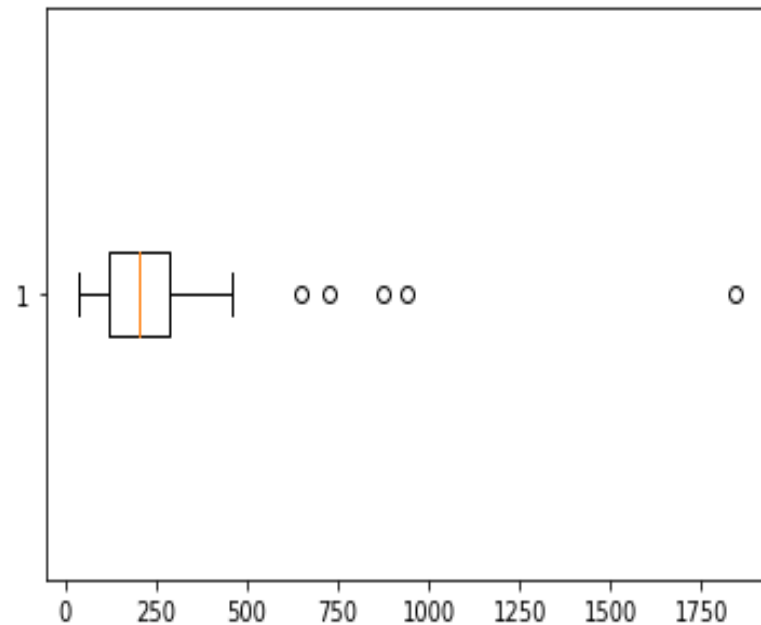




Remove Outlier (boxplot)



```
ret = plt.boxplot(newtable['จำนวนครั้งที่เกิดเหตุ'], vert=False)
```



```
ret['whiskers']
```

```
[<matplotlib.lines.Line2D at 0x7fa5e6b49c88>,  
<matplotlib.lines.Line2D at 0x7fa5e6b49470>]
```

```
print(ret['whiskers'][0].get_xdata()) #เราปรั้นออกมาดูเพื่อหาขอบเขตมากที่สุดและน้อยสุดของ whiskers  
print(ret['whiskers'][1].get_xdata())
```

```
[118.25 35. ]  
[283.5 461. ]
```

```
no_outlier_no_missing = newtable[(newtable['จำนวนครั้งที่เกิดเหตุ']>=ret['whiskers'][0].get_xdata()[1])]  
no_outlier_no_missing = no_outlier_no_missing[(no_outlier_no_missing['จำนวนครั้งที่เกิดเหตุ']<=ret['whiskers'][1].get_xdata()[1])]
```

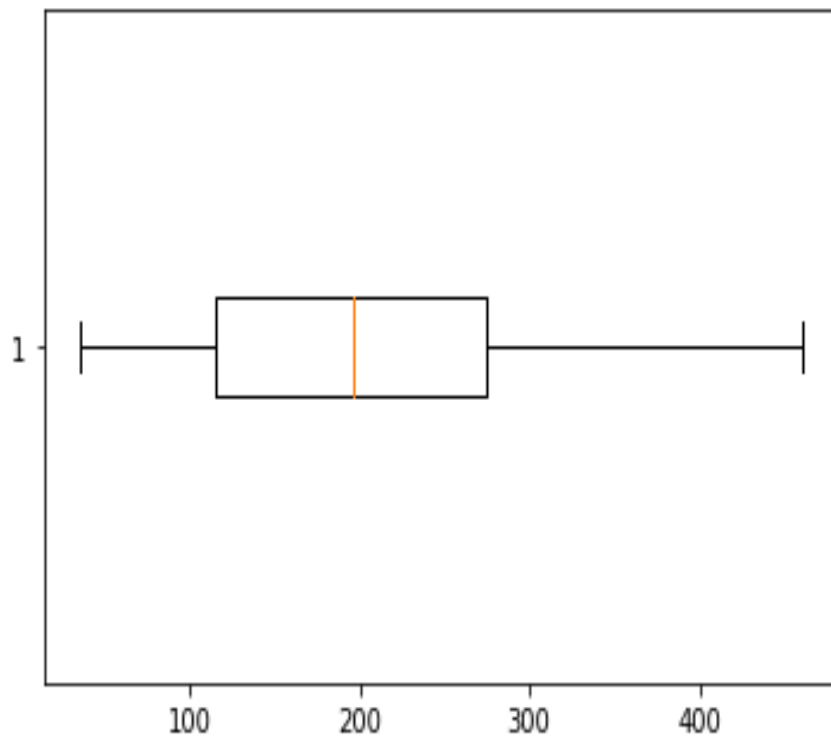


Remove Outlier (boxplot)



```
plt.boxplot(no_outlier_no_missing['จำนวนครั้งที่เกิดเหตุ'],vert=False)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7fa5e9b02710>],  
'caps': [<matplotlib.lines.Line2D at 0x7fa5e9b02048>,  
<matplotlib.lines.Line2D at 0x7fa5e9b447f0>],  
'fliers': [<matplotlib.lines.Line2D at 0x7fa5e95249e8>],  
'means': [],  
'medians': [<matplotlib.lines.Line2D at 0x7fa5e919db70>],  
'whiskers': [<matplotlib.lines.Line2D at 0x7fa5e9b02898>,  
<matplotlib.lines.Line2D at 0x7fa5e9b02d68>]}
```





Remove Outlier (boxplot)



no_outlier_no_missing

รค มอง เื่อม	ติด แอล กอฮอล์	ติดยาบ้า (Amphetamine)	ติด สาร เสพติด อื่นๆ	โรคจิตเภท	โรคจิต อื่นๆ	โรค อารมณ์ สองขั้ว	โรค ซึม เศร้า	โรค วิตกก ังวล	ความ บกพร่อง ทางสติ ปัญญา	ความ บกพร่อง ทางการ เรียนรู้	โร คอ อทิ สติก	โรค สมาธิ สั้น	พยาย ยามฆ่า ตัวตาย (การ ตั้งใจ ทำร้าย ตนเอง)	ผู้ป่วย ติด เกมส์ ใน ผู้ใหญ่ (15 ปี ขึ้นไป)	ผู้ ป่วย ติด เกมส์ ใน เด็ก (อายุ ต่ำกว่า 15 ปี)	โรค ทาง จิต เวช อื่นๆ	โรค ลม ชัก	จำนวน ครั้งที่ เกิด เหตุ	ระดับการ เกิด อุบัติเหตุ
38	221	609	1166	1597	2195	0	1529	1810	285	265	282	720	70	0	6	2149	1231	205.0	Most
49	1898	1667	975	1710	2071	0	2228	1683	204	175	172	607	336	0	5	5322	1881	285.0	Most
23	1702	2128	1480	4283	6584	0	3177	5694	239	250	175	656	224	8	15	5914	2143	130.0	Moderate
46	1799	1251	668	1714	2309	3	2634	1944	144	319	119	403	249	0	11	3897	1779	188.0	Moderate
170	4735	4108	2276	6745	8569	20	6674	10785	639	821	564	1969	485	16	51	9996	4503	340.0	Most

9	1067	1377	361	1701	2157	0	1932	2080	119	130	62	337	56	0	10	2479	957	51.0	Least
14	2769	3284	1840	5457	6476	3	4702	4911	252	223	335	574	240	5	9	6121	3626	206.0	Most
3	560	767	572	788	1039	0	1474	754	65	291	62	397	78	0	2	2120	693	289.0	Most
35	1060	478	896	795	1328	0	1717	1516	115	44	51	197	144	1	8	4018	903	114.0	Moderate
38	7480	8674	2607	7661	9909	40	10761	12236	857	417	452	1349	275	4	61	11244	5920	186.0	Moderate



Remove Outlier (boxplot)



```
no_outlier_no_missing.rename(columns={'โรคสมองเสื่อม':'Dementia','ติดแอลกอฮอล์':'Alcoholism','พยายามฆ่าตัวตาย (การตั้งใจทำร้ายตนเอง)':'Suicide','โรคลมชัก':'Epilepsy','จำนวนครั้งที่เกิดเหตุ':'Num_accident','ระดับการเกิดอุบัติเหตุ':'Level_accident'}, inplace= True)
no_outlier_no_missing
```

	จังหวัด	Dementia	Alcoholism	ติดยาบ้า (Amphetamine)	ติดสารเสพติดอื่นๆ	โรคจิตเภท	โรคจิตอื่นๆ	โรคอารมณ์สองขั้ว	โรคซึมเศร้า	โรควิตกกังวล	ความบกพร่องทางสติปัญญา	ความบกพร่องทางการเรียนรู้	โรคออทิซึม	โรคสมาธิสั้น	Suicide	ผู้ป่วยติดเกมในผู้ใหญ่ (15 ปีขึ้นไป)	ผู้ป่วยติดเกมในเด็ก (อายุต่ำกว่า 15 ปี)	โรคจิต
0	กระบี่	168	221	609	1166	1597	2195	0	1529	1810	285	265	282	720	70	0	6	2
2	กาญจนบุรี	249	1898	1667	975	1710	2071	0	2228	1683	204	175	172	607	336	0	5	5
3	กาฬสินธุ์	323	1702	2128	1480	4283	6584	0	3177	5694	239	250	175	656	224	8	15	5
4	กำแพงเพชร	146	1799	1251	668	1714	2309	3	2634	1944	144	319	119	403	249	0	11	3
5	ขอนแก่น	1170	4735	4108	2276	6745	8569	20	6674	10785	639	821	564	1969	485	16	51	9
...
71	อำนาจเจริญ	99	1067	1377	361	1701	2157	0	1932	2080	119	130	62	337	56	0	10	2
72	อุดรธานี	614	2769	3284	1840	5457	6476	3	4702	4911	252	223	335	574	240	5	9	6
73	อุดรดิตถ์	86	560	767	572	788	1039	0	1474	754	65	291	62	397	78	0	2	2
74	อุทัยธานี	195	1060	478	896	795	1328	0	1717	1516	115	44	51	197	144	1	8	4
75	อุบลราชธานี	698	7480	8674	2607	7661	9909	40	10761	12236	857	417	452	1349	275	4	61	1

71 rows x 21 columns



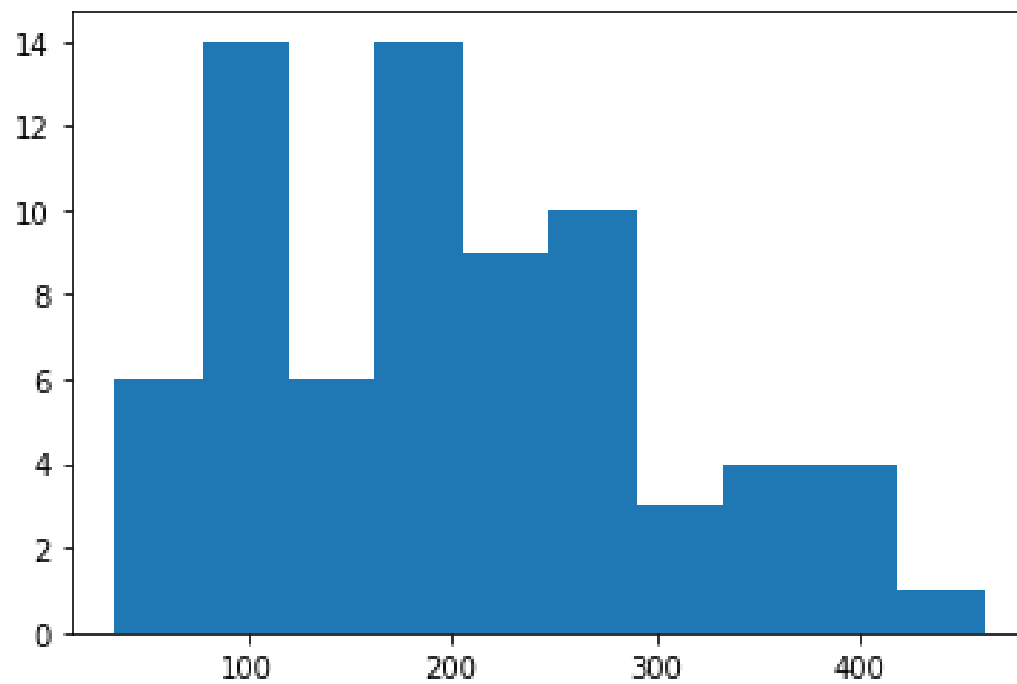
วิเคราะห์ข้อมูลจากกราฟ



```
x = [no_outlier_no_missing['Num_accident']]
```

```
plt.hist(x)
```

```
(array([ 6., 14.,  6., 14.,  9., 10.,  3.,  4.,  4.,  1.]),  
array([ 35. ,  77.6, 120.2, 162.8, 205.4, 248. , 290.6, 333.2, 375.8,  
        418.4, 461. ]),  
<a list of 10 Patch objects>)
```





วิเคราะห์ข้อมูลจากกราฟ



```
no_outlier_no_missing = no_outlier_no_missing[['Dementia','Alcoholism', 'Suicide','Epilepsy','Level_accident']]
no_outlier_no_missing
```

	Dementia	Alcoholism	Suicide	Epilepsy	Level_accident
0	168	221	70	1231	Most
2	249	1898	336	1881	Most
3	323	1702	224	2143	Moderate
4	146	1799	249	1779	Moderate
5	1170	4735	485	4503	Most
...
71	99	1067	56	957	Least
72	614	2769	240	3626	Most
73	86	560	78	693	Most
74	195	1060	144	903	Moderate
75	698	7480	275	5920	Moderate

71 rows x 5 columns

4

```
no_outlier_no_missing.Level_accident.value_counts()
```

```
Most      33
Moderate  24
Least     14
Name: Level_accident, dtype: int64
```

```
no_outlier_no_missing.columns
```

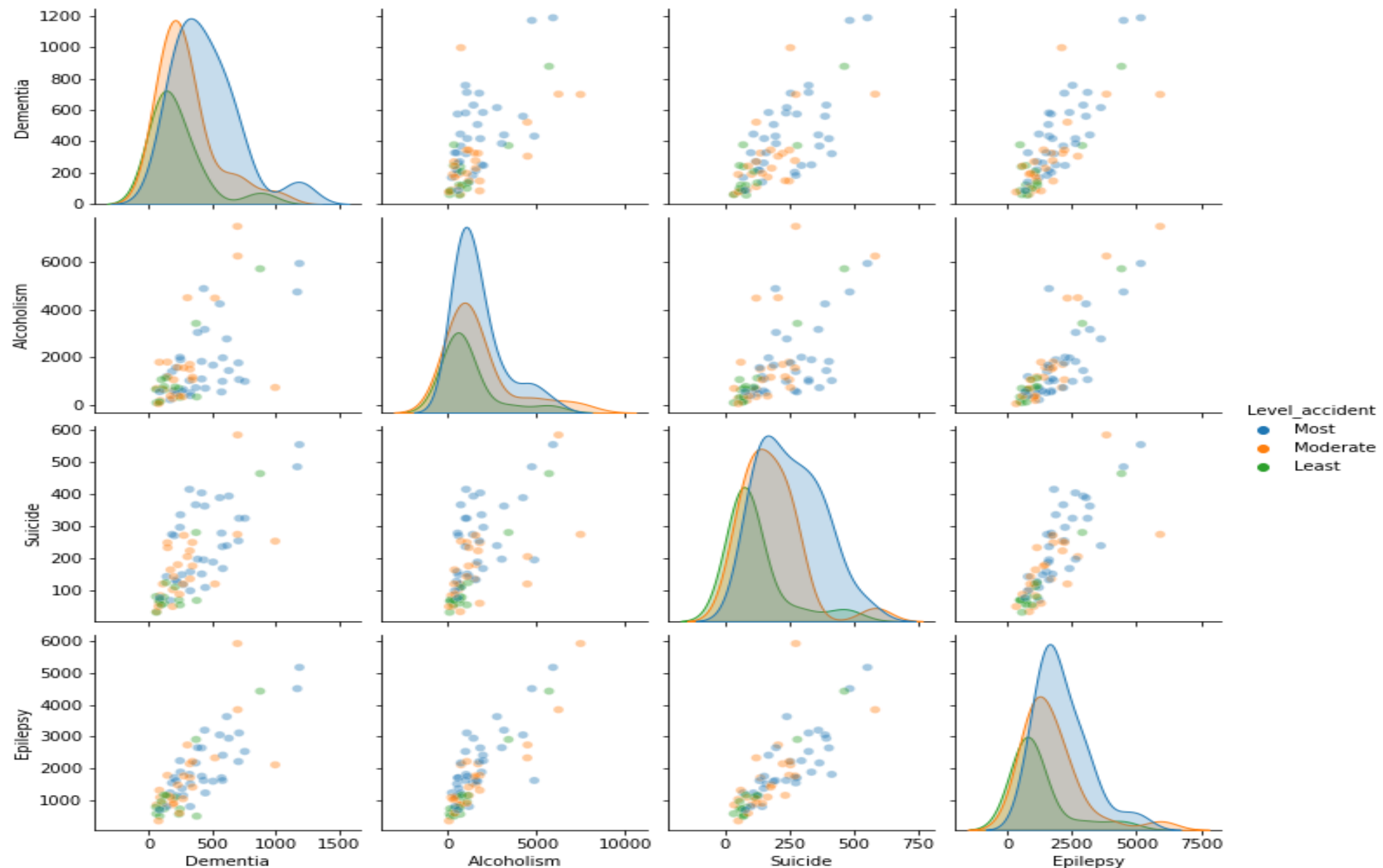
```
Index(['Dementia', 'Alcoholism', 'Suicide', 'Epilepsy', 'Level_accident'], dtype='object')
```



วิเคราะห์ข้อมูลจากกราฟ



```
sns.pairplot(no_outlier_no_missing, vars=['Dementia','Alcoholism', 'Suicide','Epilepsy'],  
             hue='Level_accident',  
             plot_kws={'alpha':.4});
```





KNNNeighborsClassifier



```
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
```

```
no_outlier_no_missing.columns
```

```
Index(['Dementia', 'Alcoholism', 'Suicide', 'Epilepsy', 'Level_accident'], dtype='object')
```

```
x_train, x_test, y_train, y_test = train_test_split(
    no_outlier_no_missing[['Dementia', 'Alcoholism', 'Suicide', 'Epilepsy']],
    no_outlier_no_missing.Level_accident,
    test_size=0.2
)
```

```
x_train.head()
```

	Dementia	Alcoholism	Suicide	Epilepsy
68	119	781	80	1162
52	506	1670	189	1595
46	304	4492	206	2735
18	573	550	279	1702
15	121	632	61	725



KNNNeighborsClassifier



```
y_train.head()
```

```
68    Least
52    Most
46  Moderate
18    Most
15    Least
Name: Level_accident, dtype: object
```

```
model = KNeighborsClassifier()
model
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                     weights='uniform')
```

```
model.fit(x_train,y_train)
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                     weights='uniform')
```



KNNNeighborsClassifier



```
model.fit(x_train,y_train)
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',  
                    metric_params=None, n_jobs=None, n_neighbors=5, p=2,  
                    weights='uniform')
```

```
model.score(x_test,y_test)
```

0.6

```
no_outlier_no_missing.sample(10)
```

	Dementia	Alcoholism	Suicide	Epilepsy	Level_accident
55	385	3039	198	2650	Most
34	217	485	129	1143	Most
53	276	1555	272	1761	Moderate
14	201	601	271	1708	Most
39	139	1193	144	799	Most
12	440	3164	363	3199	Most
2	249	1898	336	1881	Most
11	270	657	119	1147	Most
71	99	1067	56	957	Least
13	1187	5927	554	5171	Most



KNNeighborsClassifier



```
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
```

```
predicted = model.predict(x_test)  
predicted
```

```
array(['Moderate', 'Moderate', 'Most', 'Least', 'Most', 'Most', 'Least',  
      'Least', 'Moderate', 'Most', 'Least', 'Most', 'Least', 'Most',  
      'Most'], dtype=object)
```

```
dx=pd.DataFrame({'y_true':y_test, 'y_pred':predicted})  
dx[dx.y_true != dx.y_pred]
```

	y_true	y_pred
23	Moderate	Most
11	Most	Least
74	Moderate	Least
13	Most	Moderate
42	Moderate	Least
41	Moderate	Most



KNNeighborsClassifier



```
print(confusion_matrix(y_test, predicted))
```

```
[[2 0 0]
 [2 2 2]
 [1 1 5]]
```

```
print(accuracy_score(y_test, predicted))
```

```
0.6
```

```
print(classification_report(y_test, predicted))
```

	precision	recall	f1-score	support
Least	0.40	1.00	0.57	2
Moderate	0.67	0.33	0.44	6
Most	0.71	0.71	0.71	7
accuracy			0.60	15
macro avg	0.59	0.68	0.58	15
weighted avg	0.65	0.60	0.59	15

Thank you!

