In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv("last_two_years_accidents.csv")
```

In [4]:
```
sample_size = 200
dfsample = df.sample(n=sample_size)
```

In [5]:
```
dfsample
```

Out[5]:

	ID	Source	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	Distance(
1417468	A-4865127	Source1	2	2022-09-11 16:55:17	2022-09-11 18:11:59	34.787997	-81.816219	0.
1444570	A-4892567	Source1	2	2023-01-16 07:23:00	2023-01-16 09:32:54	37.507377	-120.958518	0.
1675156	A-5126218	Source1	2	2023-01-05 10:08:01	2023-01-05 12:07:30.000000000	44.779646	-93.489919	0.
971039	A-4412604	Source1	2	2022-12-22 12:19:30	2022-12-22 13:02:40	30.444312	-91.057554	0.
678720	A-4116216	Source1	2	2022-02-09 16:41:37	2022-02-09 18:48:26.000000	25.627652	-80.341551	0.
...	...	...	...	...	...	...	...	
1692957	A-5144269	Source1	2	2022-05-13 03:50:00	2022-05-13 05:56:26	32.715721	-96.325719	0.
1543987	A-4993263	Source1	2	2022-05-21 11:41:19	2022-05-21 14:42:18.000000000	40.677116	-73.926591	0.
1927679	A-5382173	Source1	2	2022-04-02 03:30:00	2022-04-02 07:16:00	35.729858	-79.824180	0.
921920	A-4362803	Source1	2	2023-01-16 22:21:12	2023-01-16 23:21:49.000000000	45.041148	-93.039649	0.
534780	A-3970270	Source1	2	2022-04-25 09:28:00	2022-04-25 09:47:30	34.897503	-82.152060	2.

200 rows × 41 columns

In [6]:
```
dfsample['combined'] =

dfsample.apply(lambda row: ', '.join([f"{col}={int(not val)}"
    for col, val in row[['Crossing', 'Bump', 'Junction', 'Stop']].items()]), axis=1)
```

In [7]:
```
dfsample['combined']
```

Out[7]:
```
1417468    Crossing=1, Bump=1, Junction=1, Stop=1
1444570    Crossing=1, Bump=1, Junction=1, Stop=1
1675156    Crossing=1, Bump=1, Junction=1, Stop=1
971039     Crossing=1, Bump=1, Junction=1, Stop=1
678720     Crossing=0, Bump=1, Junction=1, Stop=1
                              ...
1692957    Crossing=1, Bump=1, Junction=1, Stop=1
1543987    Crossing=0, Bump=1, Junction=1, Stop=1
1927679    Crossing=1, Bump=1, Junction=1, Stop=1
921920     Crossing=1, Bump=1, Junction=1, Stop=1
534780     Crossing=1, Bump=1, Junction=1, Stop=1
Name: combined, Length: 200, dtype: object
```

In [8]:
```
df_new = dfsample[['combined','State']]
```

In [9]:
```
df_new
```

Out[9]:

	combined	State
1417468	Crossing=1, Bump=1, Junction=1, Stop=1	SC
1444570	Crossing=1, Bump=1, Junction=1, Stop=1	CA
1675156	Crossing=1, Bump=1, Junction=1, Stop=1	MN
971039	Crossing=1, Bump=1, Junction=1, Stop=1	LA
678720	Crossing=0, Bump=1, Junction=1, Stop=1	FL
...	...	...
1692957	Crossing=1, Bump=1, Junction=1, Stop=1	TX
1543987	Crossing=0, Bump=1, Junction=1, Stop=1	NY
1927679	Crossing=1, Bump=1, Junction=1, Stop=1	NC
921920	Crossing=1, Bump=1, Junction=1, Stop=1	MN
534780	Crossing=1, Bump=1, Junction=1, Stop=1	SC

200 rows × 2 columns

In [11]:
```
df_new.to_csv('ARM_prepared.csv', index=False)
```

In [ ]:
```
```