Data import:

```
data = readtable("./Dataset/combined.csv")
```

data = 631491×200 table

	ReportSubmissionSource	MultipleRowsPerIncident	
1	'Paper'	'No'	'<a href = https://portal.phi
2	'Paper'	'No'	'<a href = https://portal.phi
3	'Paper'	'No'	'<a href = https://portal.phi
4	'Paper'	'No'	'<a href = https://portal.phi
5	'Paper'	'No'	'<a href = https://portal.phi
6	'Paper'	'No'	'<a href = https://portal.phi
7	'Paper'	'No'	'<a href = https://portal.phi
8	'Paper'	'No'	'<a href = https://portal.phi
9	'Paper'	'No'	'<a href = https://portal.phi
10	'Paper'	'No'	'<a href = https://portal.phi
11	'Paper'	'No'	'<a href = https://portal.phi
12	'Paper'	'No'	'<a href = https://portal.phi
13	'Paper'	'No'	'<a href = https://portal.phi
14	'Paper'	'No'	'<a href = https://portal.phi
15	'Paper'	'No'	'<a href = https://portal.phi
16	'Paper'	'No'	'<a href = https://portal.phi
17	'Paper'	'No'	'<a href = https://portal.phi
18	'Paper'	'No'	'<a href = https://portal.phi
19	'Paper'	'No'	'<a href = https://portal.phi
20	'Paper'	'No'	'<a href = https://portal.phi
21	'Paper'	'No'	'<a href = https://portal.phi
22	'Paper'	'No'	'<a href = https://portal.phi
23	'Paper'	'No'	'<a href = https://portal.phi
24	'Paper'	'No'	'<a href = https://portal.phi
25	'Paper'	'No'	'<a href = https://portal.phi
26	'Paper'	'No'	'<a href = https://portal.phi
27	'Paper'	'No'	'<a href = https://portal.phi
28	'Paper'	'No'	'<a href = https://portal.phi
29	'Paper'	'No'	'<a href = https://portal.phi
30	'Paper'	'No'	'<a href = https://portal.phi
31	'Paper'	'No'	'<a href = https://portal.phi
32	'Paper'	'No'	'<a href = https://portal.phi

	ReportSubmissionSource	MultipleRowsPerIncident	
33	'Paper'	'No'	'<a href = https://portal.phr
34	'Paper'	'No'	'<a href = https://portal.phr
35	'Paper'	'No'	'<a href = https://portal.phr
36	'Paper'	'No'	'<a href = https://portal.phr
37	'Paper'	'No'	'<a href = https://portal.phr
38	'Paper'	'No'	'<a href = https://portal.phr
39	'Paper'	'No'	'<a href = https://portal.phr
40	'Paper'	'No'	'<a href = https://portal.phr
41	'Paper'	'No'	'<a href = https://portal.phr
42	'Paper'	'No'	'<a href = https://portal.phr
43	'Paper'	'No'	'<a href = https://portal.phr
44	'Paper'	'No'	'<a href = https://portal.phr
45	'Paper'	'No'	'<a href = https://portal.phr
46	'Paper'	'No'	'<a href = https://portal.phr
47	'Paper'	'No'	'<a href = https://portal.phr
48	'Paper'	'No'	'<a href = https://portal.phr
49	'Paper'	'No'	'<a href = https://portal.phr
50	'Paper'	'No'	'<a href = https://portal.phr
51	'Paper'	'No'	'<a href = https://portal.phr
52	'Paper'	'No'	'<a href = https://portal.phr
53	'Paper'	'No'	'<a href = https://portal.phr
54	'Paper'	'No'	'<a href = https://portal.phr
55	'Paper'	'No'	'<a href = https://portal.phr
56	'Paper'	'No'	'<a href = https://portal.phr
57	'Paper'	'No'	'<a href = https://portal.phr
58	'Paper'	'No'	'<a href = https://portal.phr
59	'Paper'	'No'	'<a href = https://portal.phr
60	'Paper'	'No'	'<a href = https://portal.phr
61	'Paper'	'No'	'<a href = https://portal.phr
62	'Paper'	'No'	'<a href = https://portal.phr
63	'Paper'	'No'	'<a href = https://portal.phr
64	'Paper'	'No'	'<a href = https://portal.phr
65	'Paper'	'No'	'<a href = https://portal.phr
66	'Paper'	'No'	'<a href = https://portal.phr
67	'Paper'	'No'	'<a href = https://portal.phr
68	'Paper'	'No'	'<a href = https://portal.phr
69	'Paper'	'No'	'<a href = https://portal.phr
70	'Paper'	'No'	'<a href = https://portal.phr
71	'Paper'	'No'	'<a href = https://portal.phr

	ReportSubmissionSource	MultipleRowsPerIncident	
72	'Paper'	'No'	'<a href = https://portal.ph
73	'Paper'	'No'	'<a href = https://portal.ph
74	'Paper'	'No'	'<a href = https://portal.ph
75	'Paper'	'No'	'<a href = https://portal.ph
76	'Paper'	'No'	'<a href = https://portal.ph
77	'Paper'	'No'	'<a href = https://portal.ph
78	'Paper'	'No'	'<a href = https://portal.ph
79	'Paper'	'No'	'<a href = https://portal.ph
80	'Paper'	'No'	'<a href = https://portal.ph
81	'Paper'	'No'	'<a href = https://portal.ph
82	'Paper'	'No'	'<a href = https://portal.ph
83	'Paper'	'No'	'<a href = https://portal.ph
84	'Paper'	'No'	'<a href = https://portal.ph
85	'Paper'	'No'	'<a href = https://portal.ph
86	'Paper'	'No'	'<a href = https://portal.ph
87	'Paper'	'No'	'<a href = https://portal.ph
88	'Paper'	'No'	'<a href = https://portal.ph
89	'Paper'	'No'	'<a href = https://portal.ph
90	'Paper'	'No'	'<a href = https://portal.ph
91	'Paper'	'No'	'<a href = https://portal.ph
92	'Paper'	'No'	'<a href = https://portal.ph
93	'Paper'	'No'	'<a href = https://portal.ph
94	'Paper'	'No'	'<a href = https://portal.ph
95	'Paper'	'No'	'<a href = https://portal.ph
96	'Paper'	'No'	'<a href = https://portal.ph
97	'Paper'	'No'	'<a href = https://portal.ph
98	'Paper'	'No'	'<a href = https://portal.ph
99	'Paper'	'No'	'<a href = https://portal.ph
100	'Paper'	'No'	'<a href = https://portal.ph

Keeping only select columns from the 200 column dataset:

- First 3 columns used for removing rows, then are removed
- Next 7 are inputs
- Last 13 are potential outputs

```
data = data(:,{'MultipleRowsPerIncident', ...
    'ReportType', ...
    'IncidentCountry', ...
    'DateOfIncident', ...
```

```
    'ModeOfTransportation', ...
    'TransportationPhase', ...
    'CommodityLongName', ...
    'IdentificationNumber', ...
    'PackagingType', ...
    'GeneralPackageType', ...
    'WhatFailedCode', ...
    'HowFailedCode', ...
    'FailureCauseCode', ...
    'SpillageResultInd', ...
    'FireResultInd', ...
    'ExplosionResultInd', ...
    'WaterSewerResultInd', ...
    'GasDispersionResultInd', ...
    'EnvironmentalDamageResult', ...
    'OtherCleanupInd', ...
    'MaterialLoss', ...
    'RemediationCleanupCost', ...
    'SeriousIncidentInd'});
```

Removal of select rows due to row content:

- MultipleRowsPerIncident == 'Yes' (Removal of 3% of dataset, n=19390) [For simplicity and clarity of counting rows and not double-counting]
- IncidentCountry ~= 'US' (Removal of 0.179%, n=1130) [Negligible reduction in rows, allows for removal of IncidentCountry column for file size]
- ReportType ~= 'A hazardous material incident' (Removal of 2.936%, n=18538) [For uniformity, the removed 2.936% were nonhazardous reports]

Overall reduction from 631491 to 596921 rows, n=34570, 5.5%

```
data(ismember(data.MultipleRowsPerIncident,'Yes'),:)=[];
data(~ismember(data.IncidentCountry,'US'),:)=[];
data(~ismember(data.ReportType,'A hazardous material incident'),:)=[];
data(:,1:3) = [];
```

Removal of rows with missing values:

Reduction from 596921 to 300254 rows, n=296667, 50%

```
data = rmmissing(data);
data = data(randperm(size(data,1)),:)
```

```
data = 300254×20 table
```

	DateOfIncident	ModeOfTransportation	TransportationPhase	
1	17-Nov-2017	'Highway'	'In Transit'	'SODIL
2	11-Jul-2020	'Highway'	'Unloading'	'DISINI
3	10-Sep-1995	'Highway'	'Unloading'	'INK PI
4	17-Sep-2012	'Highway'	'Unloading'	'SODIL
5	26-May-2021	'Highway'	'Unloading'	'CORR
6	01-Apr-2022	'Highway'	'Loading'	'ISOPF

	DateOfIncident	ModeOfTransportation	TransportationPhase	
7	19-Jan-2007	'Highway'	'Unloading'	'POTAS
8	10-Aug-2017	'Highway'	'Unloading'	'HYDR
9	12-Jul-2019	'Highway'	'Loading'	'ISOPR
10	10-Nov-2005	'Highway'	'Loading'	'PAINT
11	27-May-2023	'Highway'	'Loading'	'SODIU
12	25-Aug-2023	'Highway'	'Unloading'	'HYDR
13	12-Oct-2023	'Highway'	'Unloading'	'CORR
14	15-Aug-2012	'Highway'	'Unloading'	'PAINT