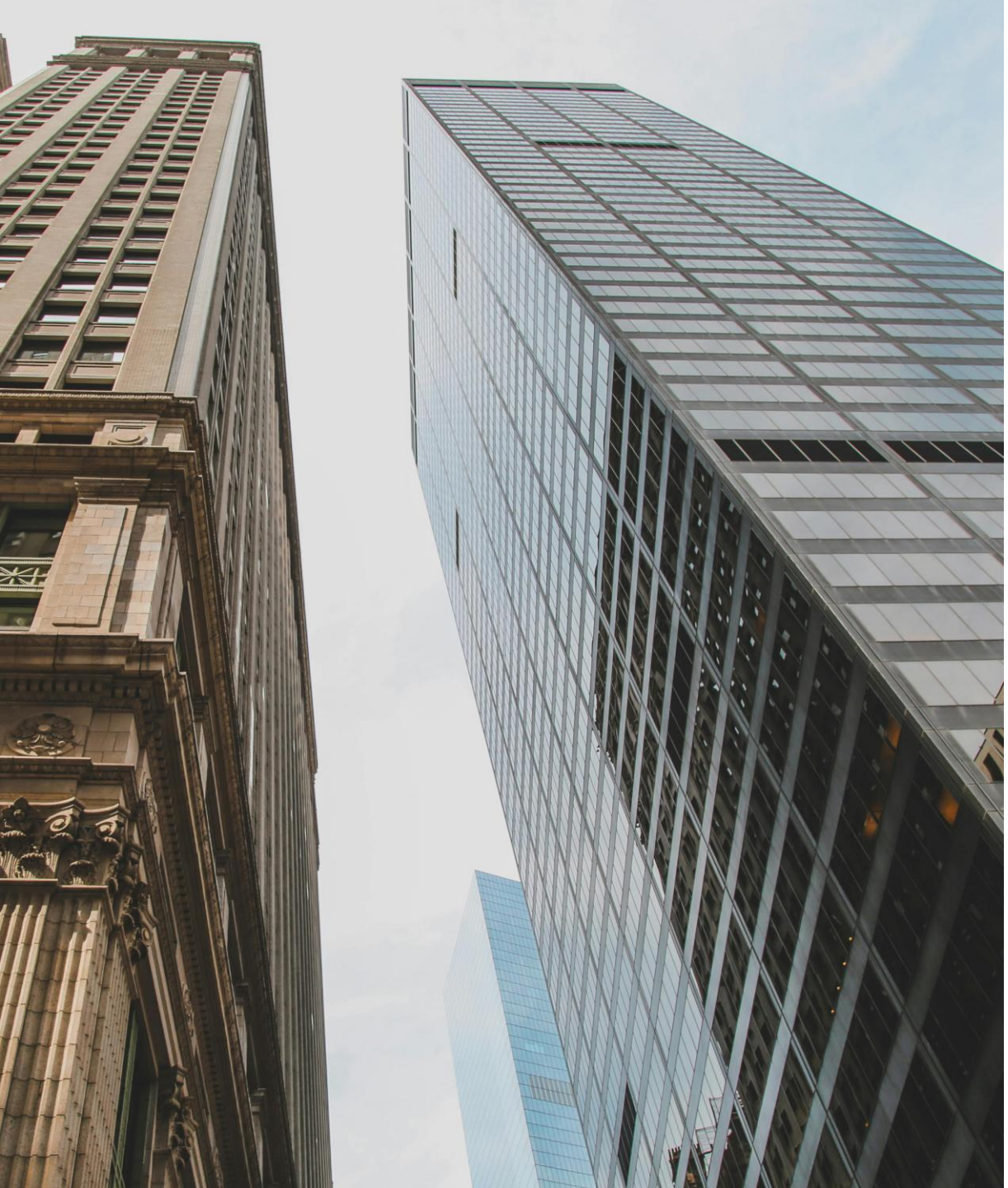


Car Crashes in New York City

Adis Mahmic

Agenda

Hypothesis

Introduction to data

Cleaning the data

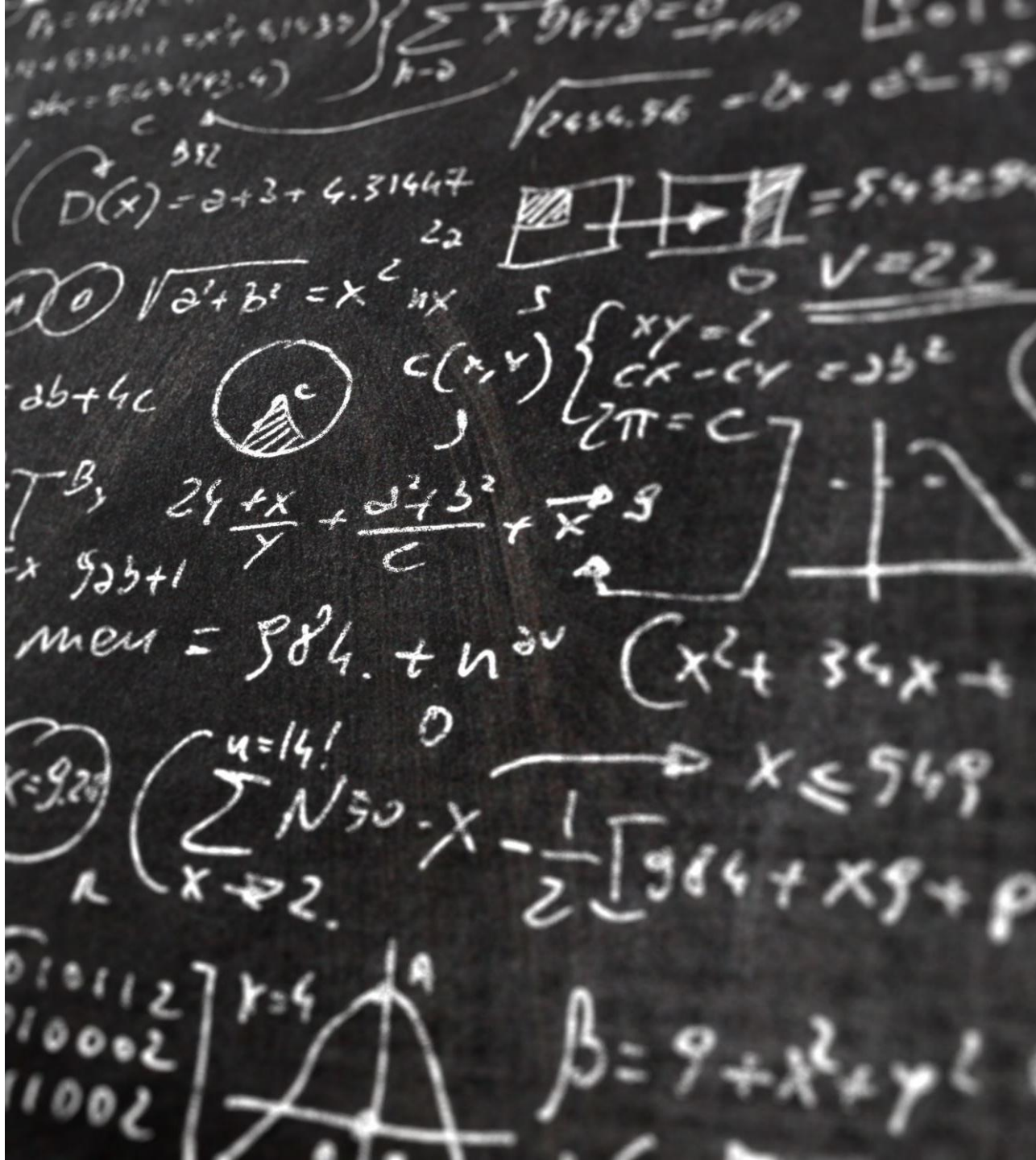
What is Hadoop Map/Reduce?

Visual aids

Final tips & takeaways

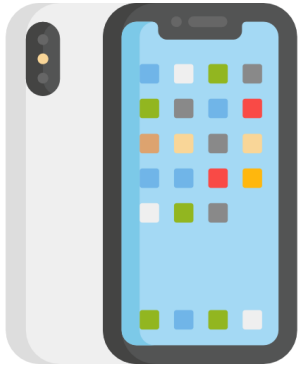
Hypothesis

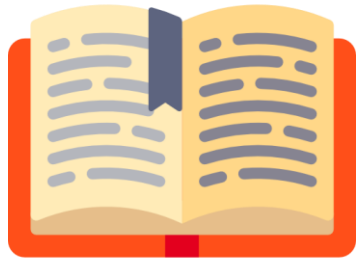
Distracted driving is the leading cause of motor accident deaths.



Distracted driving is the leading cause of motor accident deaths.

But what is categorized as distracted driving?





Before analyzing the dataset, my hypothesis is that distracted driving is the number 1 cause of death in car accidents.

Motor Vehicle Collisions.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS STF	OFF STREE	NUMBER C	NUMBER C	NUMBER C	NUMBER C	NUMBER C	NUMBER C	NUMBER C	NUMBER C	CONTRIBL	CONTRIBL	CONTRIBL	CONTRIBL	CONTRIBL	COLLISION	VEHIC
2	9/11/2021	2:39						WHITESTONE EXPRESSWAY	20 AVENUE		2	0	0	0	0	0	2	0	Aggressive	Unspecified				4455765	Sedan
3	3/26/2022	11:45						QUEENSBORO BRIDGE UPPER			1	0	0	0	0	0	1	0	Pavement	Slippery				4513547	Sedan
4	11/1/2023	1:29	BROOKLYN	11230	40.62179	-73.970024	(40.62179, -73.970024)	OCEAN PARKWAY	AVENUE K		1	0	0	0	0	0	1	0	Unspecifie	Unspecifie	Unspecified			4675373	Mopec
5	6/29/2022	6:55						THROGS NECK BRIDGE			0	0	0	0	0	0	0	0	Following	Unspecified				4541903	Sedan
6	9/21/2022	13:21						BROOKLYN BRIDGE			0	0	0	0	0	0	0	0	Passing T	Unspecified				4566131	Statio
7	4/26/2023	13:30						WEST 54 STREET			0	0	0	0	0	0	0	0	Unspecifie	Unspecified				4623759	Sedan
8	11/1/2023	7:12						HUTCHINSON RIVER PARKWAY			0	0	0	0	0	0	0	0	Following	Driver Inattention/Distraction				4675709	Sedan
9	11/1/2023	8:01						WEST 35 STREET	HENRY HUDSON RIVE		0	0	0	0	0	0	0	0	Failure to	Yield Right-of-Way				4675769	Sedan
10	4/26/2023	22:20								61 Ed k	0	0	0	0	0	0	0	0	Unspecified					4623865	Sedan
11	9/11/2021	9:35	BROOKLYN	11208	40.667202	-73.8665	(40.667202, -73.8665)			1211 LO	0	0	0	0	0	0	0	0	Unspecified					4456314	Sedan
12	12/14/2021	8:13	BROOKLYN	11233	40.683304	-73.917274	(40.683304, -73.917274)	SARATOGA AVENUE	DECATUR STREET		0	0	0	0	0	0	0	0						4486609	
13	4/14/2021	12:47						MAJOR DEEGAN EXPRESSWAY RAMP			0	0	0	0	0	0	0	0	Unspecifie	Unspecified				4407458	Dump
14	12/14/2021	17:05			40.709183	-73.956825	(40.709183, -73.956825)	BROOKLYN QUEENS EXPRESSWAY			0	0	0	0	0	0	0	0	Passing T	Unspecified				4486555	Sedan
15	12/14/2021	8:17	BRONX	10475	40.86816	-73.83148	(40.86816, -73.83148)			344 BA	2	0	0	0	0	0	2	0	Unspecifie	Unspecified				4486660	Sedan
16	12/14/2021	21:10	BROOKLYN	11207	40.67172	-73.8971	(40.67172, -73.8971)			2047 PI	0	0	0	0	0	0	0	0	Driver Ine	Unspecified				4487074	Sedan
17	12/14/2021	14:58	MANHATTAN	10017	40.75144	-73.97397	(40.75144, -73.97397)	3 AVENUE	EAST 43 STREET		0	0	0	0	0	0	0	0	Passing T	Unspecified				4486519	Sedan
18	12/13/2021	0:34			40.701275	-73.88887	(40.701275, -73.88887)	MYRTLE AVENUE			0	0	0	0	0	0	0	0	Passing o	Unspecified				4486934	Statio
19	12/14/2021	16:50	QUEENS	11413	40.675884	-73.75577	(40.675884, -73.75577)	SPRINGFIELD BOULEVARD	EAST GATE PLAZA		0	0	0	0	0	0	0	0	Turning In	Unspecified				4487127	Sedan
20	12/14/2021	8:30						broadway	west 80 street -west E		0	0	0	0	0	0	0	0	Unsafe La	Unspecified				4486634	Statio
21	12/14/2021	0:59			40.59662	-74.00231	(40.59662, -74.00231)	BELT PARKWAY			0	0	0	0	0	0	0	0	Unsafe Speed					4486564	Sedan
22	12/14/2021	23:10	QUEENS	11434	40.66684	-73.78941	(40.66684, -73.78941)	NORTH CONDUIT AVENUE	150 STREET		2	0	0	0	0	0	2	0	Reaction t	Unspecified				4486635	Sedan
23	12/14/2021	17:58	BROOKLYN	11217	40.68158	-73.97463	(40.68158, -73.97463)			480 DE/	0	0	0	0	0	0	0	0	Passing T	Unspecified				4486604	Tanke
24	12/14/2021	20:03	BROOKLYN	11226	40.65068	-73.95881	(40.65068, -73.95881)			878 FLA	4	0	0	0	0	0	4	0	Steering	Failure				4486991	Sedan
25	12/14/2021	1:28						MEEKER AVENUE	LORIMER STREET		3	0	0	0	0	0	3	0	Traffic Cor	Unspecified				4486284	Statio
26	12/11/2021	19:43	BRONX	10463	40.87262	-73.904686	(40.87262, -73.904686)	WEST KINGSBRIDGE ROAD	HEATH AVENUE		1	0	0	0	0	0	1	0	Unspecifie	Unspecified				4487040	Statio
27	12/14/2021	14:30			40.783268	-73.82485	(40.783268, -73.82485)	WHITESTONE EXPRESSWAY			0	0	0	0	0	0	0	0	Following	Unspecifie	Unspecified			4486537	Statio
28	12/11/2021	4:45	MANHATTAN	10001	40.748917	-73.993546	(40.748917, -73.993546)			232 WE	0	0	0	0	0	0	0	0	Following	Unspecified				4486905	Statio
29	12/14/2021	5:46			40.744644	-73.77041	(40.744644, -73.77041)	LONG ISLAND EXPRESSWAY			1	0	0	0	0	0	1	0	Other Veh	Other Vehicular				4487122	Statio
30	12/13/2021	6:30	QUEENS	11372	40.75373	-73.88505	(40.75373, -73.88505)	82 STREET	34 AVENUE		0	0	0	0	0	0	0	0	Unspecified					4486967	Sedan
31	12/14/2021	3:43			40.804375	-73.93742	(40.804375, -73.93742)	LEXINGTON AVENUE			1	0	1	0	0	0	0	0	Unspecified					4486304	Statio
32	12/13/2021	17:40	STATEN ISLAND	10301	40.63165	-74.08762	(40.63165, -74.08762)	VICTORY BOULEVARD	WOODSTOCK AV								0	1	0	Unspecifie	Unspecified			4487001	Sedan
33	12/14/2021	17:31	BROOKLYN	11230	40.623104	-73.95809	(40.623104, -73.95809)	EAST 18 STREET	AVENUE K		1	0	1	0	0	0	0	0	Unspecified					4486516	Sedan
34	12/14/2021	20:13	BROOKLYN	11215	40.66576	-73.9845	(40.66576, -73.9845)			366 12	0	0	0	0	0	0	0	0	Passing T	Unspecified				4486605	Sedan
35	12/14/2021	12:54	BROOKLYN	11217	40.687534	-73.9775	(40.687534, -73.9775)	FULTON STREET	SAINT FELIX STREET		1	0	0	0	1	0	0	0	Unspecifie	Unspecified				4487052	Sedan
36	12/14/2021	17:15	BROOKLYN	11211	40.710957	-73.951126	(40.710957, -73.951126)	GRAND STREET	UNION AVENUE		1	0	0	0	0	0	1	0	Passing o	Unspecified				4486556	Bus
37	12/14/2021	22:49	BRONX	10455	40.81813	-73.910126	(40.81813, -73.910126)			713 EA(	0	0	0	0	0	0	0	0	Driver Inat	Unspecified				4486875	Taxi
38	12/12/2021	9:00	QUEENS	11385	40.70447	-73.90148	(40.70447, -73.90148)			59-14 67	0	0	0	0	0	0	0	0	Passing T	Unspecified				4486933	Statio
39	12/14/2021	16:25			40.784615	-73.953964	(40.784615, -73.953964)	EAST 93 STREET			1	0	0	0	1	0	0	0	Driver Inat	Driver Inattention/Distraction				4486581	Van
40	11/2/2023	9:20						35 AVENUE			0	0	0	0	0	0	0	0	Accelerator	Defective				4675877	Statio
41	4/14/2021	14:30						EASTCHESTER ROAD	PELHAM PARKWAY N		0	0	0	0	0	0	0	0	Driver Inat	Unspecified				4407520	Bus
42	12/16/2021	6:59						KINGSLAND AVENUE	MEEKER AVENUE		1	0	1	0	0	0	0	0	Traffic Control Disregarded					4486960	
43	4/27/2023	15:40						WILLIAMSBURG BRIDGE OUTER ROADWA			1	0	0	0	0	0	1	0	Driver Inat	Unspecified				4624078	Sedan
44	9/22/2022	16:16	QUEENS	11418	40.698257	-73.82632	(40.698257, -73.82632)	123 STREET	89 AVENUE		1	0	0	0	0	0	0	0	Passing o	Unspecified				4566408	Sedan
45	1/12/2023	21:00	BROOKLYN	11208						97-16 DF	0	0	0	0	0	0	0	0	Driver Inat	Unspecified				4598234	Statio
46	6/29/2022	16:00						WILLIAMSBURG BRIDGE OUTER ROADWA			1	0	0	0	0	0	1	0	Driver Inat	Unspecified				4542336	Motor
47	4/15/2021	16:15						HUTCHINSON RIVER PARKWAY			0	0	0	0	0	0	0	0	Pavement	Slippery				4407665	Statio
48	7/7/2021	11:42						THROGS NECK BRIDGE			1	0	0	0	0	0	1	0	Unsafe La	Unspecified				4456591	Sedan
49	4/27/2023	22:13						PELHAM PARKWAY SOUTH	WILLIAMSBRIDGE RO		0	0	0	0	0	0	0	0	Driver Inat	Unspecified				4624101	Sedan
50	7/12/2022	17:50	BROOKLYN	11225	40.663303	-73.96049	(40.663303, -73.96049)			44 EMF	0	0	0	0	0	0	0	0	Oversized	Unspecified				4545699	Sedan
51	3/23/2022	10:00								71 EAS	0	0	0	0	0	0	0	0	Pedestrian/Bicyclist/Other Pedestrian Error/Confusi					4512922	Bike
52	7/9/2021	0:43			40.720535	-73.88885	(40.720535, -73.88885)	ELIOT AVENUE			0	1	0	1	0	0	0	0	Unspecified					4456659	Bus
53	4/24/2022	16:45			40.607685	-74.13892	(40.607685, -74.13892)	STATEN ISLAND EXPRESSWAY			1	0	0	0	0	0	1	0	Driver Inat	Unspecified				4521660	Statio
54	4/24/2022	4:49			40.855972	-73.869896	(40.855972, -73.869896)	BOSTON ROAD	BRONX PARK EAST		0	0	0	0	0	0	0	0	Unspecifie	Unspecified				4521759	Statio
55	4/22/2022	17:17			40.790276	-73.9396	(40.790276, -73.9396)	EAST 107 STREET			1	0	1	0	0	0	0	0	Traffic Control Disregarded					4522226	E-Bike

B	C
NUMBER OF MOTORIST KILLED	CONTRIBUTING FACTOR VEHICLE 1
0	Aggressive Driving/Road Rage
0	Pavement Slippery
0	Unspecified
0	Following Too Closely
0	Passing Too Closely
0	Unspecified
0	Following Too Closely
0	Failure to Yield Right-of-Way
0	Unspecified
0	Unspecified
0	Unspecified
0	Unspecified
0	Passing Too Closely
0	Unspecified
0	Driver Inexperience
0	Passing Too Closely
0	Passing or Lane Usage Improper
0	Turning Improperly
0	Unsafe Lane Changing
0	Unsafe Speed
0	Reaction to Uninvolved Vehicle
0	Passing Too Closely
0	Steering Failure
0	Traffic Control Disregarded
0	Unspecified
0	Following Too Closely
0	Following Too Closely
0	Other Vehicular
0	Unspecified
0	Unspecified
0	Unspecified
0	Unspecified
0	Passing Too Closely
0	Unspecified
0	Passing or Lane Usage Improper

- The Motor Crash Collisions dataset was created by the New York Police Department and is used to keep records of each and every car crash that occurs in New York.
- Each row in the Dataset represents a police-reported crash in NYC, capturing key details about the incident such as contributing factor, street, and number of people injured.
- The main columns we are looking at is number of people killed and contributing factor.

What is our dataset?

Extracting The Data

Using Head Command

```
amahm@AdisSurface MINGW64 ~/Downloads/Adis GGC/Data Intensive/Motor-Vehicle-Collisions (main)
$ head Motor_Crash_Collisions_Extracted.csv
NUMBER OF PERSONS INJURED,NUMBER OF PERSONS KILLED,NUMBER OF PEDESTRIANS INJURED,NUMBER OF PEDESTRIANS KILLED,NUMBER OF MOTORIST INJURED,NUMBER OF MOTORIST KILLED,CONTRIBUTING FACTOR VEHICLE 1,CONTRIBUTING FACTOR VEHICLE 2
2,0,0,0,2,0,Aggressive Driving/Road Rage,Unspecified
1,0,0,0,1,0,Pavement Slippery,
,1,0,0,0,1,0,Unspecified
0,0,0,0,0,0,Following Too Closely,Unspecified
0,0,0,0,0,0,Passing Too Closely,Unspecified
0,0,0,0,0,0,Unspecified,Unspecified
0,0,0,0,0,0,Following Too Closely,Driver Inattention/Distraction
0,0,0,0,0,0,Failure to Yield Right-of-Way,
0,0,0,0,0,0,Unspecified,
```

Using Awk Command

```
Adis@DESKTOP-F36S3PO MINGW64 ~/MotorVehicleCollisions (main)
$ awk -F, '{print $18 "," $19}' Motor_Crash_Collisions.csv > motorist_killed_fac
tors.csv

Adis@DESKTOP-F36S3PO MINGW64 ~/MotorVehicleCollisions (main)
$ head motorist_killed_factors.csv
NUMBER OF MOTORIST KILLED,CONTRIBUTING FACTOR VEHICLE 1
0,Aggressive Driving/Road Rage
0,Pavement Slippery
0,Unspecified
0,Following Too Closely
0,Passing Too Closely
0,Unspecified
0,Following Too Closely
0,Failure to Yield Right-of-Way
0,Unspecified
```

Hadoop Map Reduce

Hadoop Map Reduce API

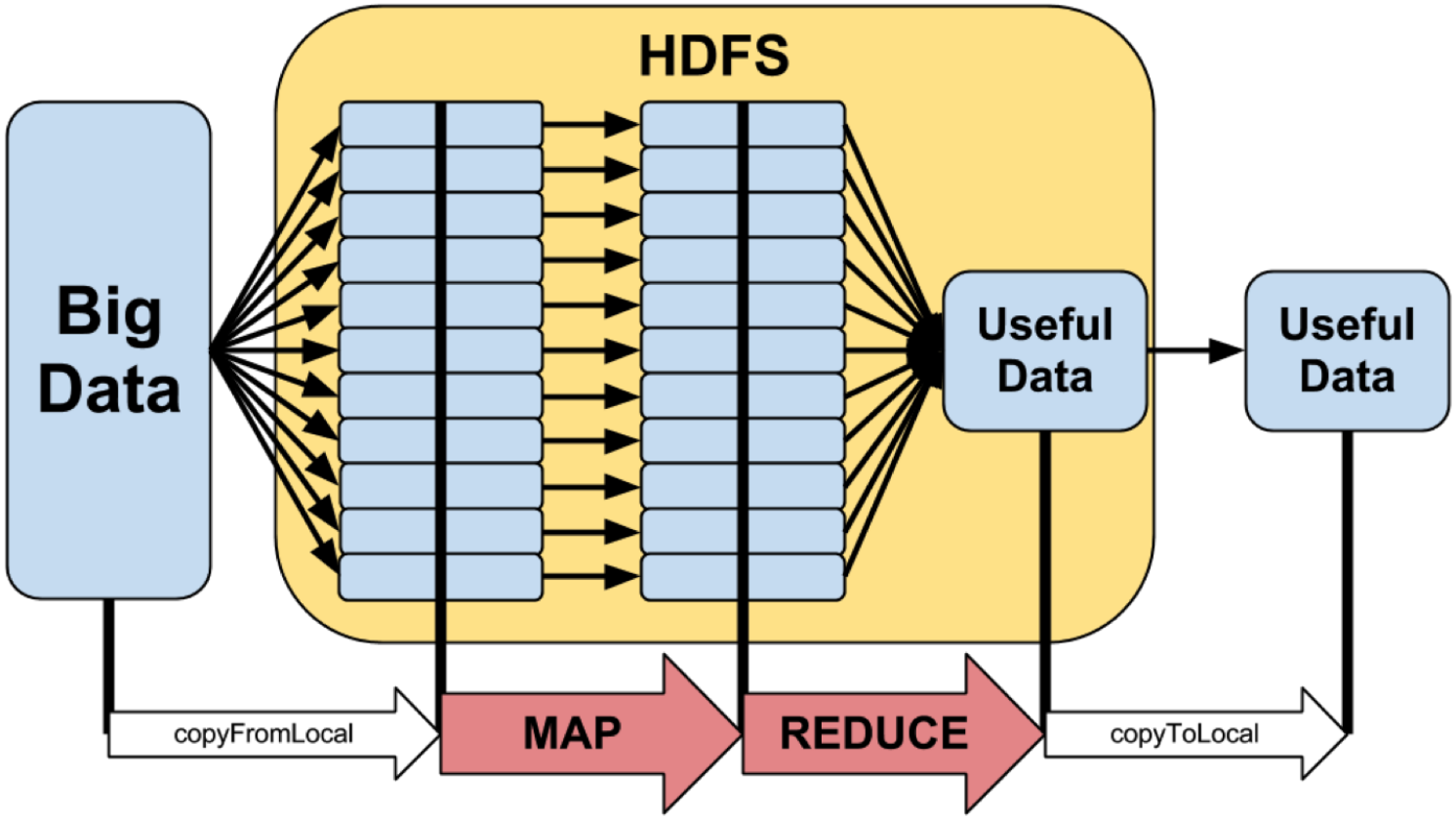
- Map Reduce is a programming model that uses parallel processing to speed large scale data processing.
- Enables massive scalability within a Hadoop cluster.
- Map task takes set of data and converts it into key value pairs.
- Reduce task takes output from the map task and aggregates values within the same key and processes the data into a final set of key value pairs.

1. Data is collected.

2. Mapper class maps data into key value pairs.

3. Data is stored in Hadoop Distributed File System.

4. Reducer reduces key value pairs into final key value pairs.



Map Reduce Example

Mapper Class

Purpose: Analyze vehicle crash data to identify key contributing factors associated with motorist fatalities.

Extracts “Contributing Factor Vehicle 1” and “Number of Motorist Killed”

Emits key-value pairs:

Key = Contributing

Value = Number of Motorists Killed

```java
/.../

import ...

public class CrashMapper extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {  1 usage

    private boolean isHeader = true;     // Flag to identify the header row  2 usages

    // Indices of relevant columns in the CSV
    private int killedIndex = -1;  4 usages
    private int factorIndex = -1;  4 usages

    // Map function to process each line of the input file.
    public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
        String line = value.toString();
        // Use regex to split CSV fields (also handles quoted commas) so that the splitter doesn't accidentally read commas inside the records.
        String[] fields = line.split( regex: ",(?=(?:[^\"]*\"[^\"]*\")*[^\"]*$)",  limit: -1); // CSV-safe

        // Handle header row and identify indices of required columns
        if (isHeader) {
            for (int i = 0; i < fields.length; i++) {
                String col = fields[i].trim().toUpperCase();
                if (col.equals("NUMBER OF MOTORIST KILLED")) {
                    killedIndex = i;
                } else if (col.equals("CONTRIBUTING FACTOR VEHICLE 1")) {
                    factorIndex = i;
                }
            }
            isHeader = false; // Ensure we only process the header once
            return;
        }

        // Proceed only if both indices were found and the row has enough fields
        if (killedIndex != -1 && factorIndex != -1 && fields.length > Math.max(killedIndex, factorIndex)) {
            try {
                // Get the number of killed motorists; default to 0 if blank
                String killedStr = fields[killedIndex].trim();
                String factor = fields[factorIndex].trim();

                int killed = Integer.parseInt(killedStr.isEmpty() ? "0" : killedStr);

                // Ignore empty or unspecified contributing factors
                if (!factor.isEmpty() && !factor.equalsIgnoreCase( anotherString: "Unspecified")) {
                    // Emit (key: contributing factor, value: number killed)
                    output.collect(new Text(factor), new IntWritable(killed));
                }
            } catch (NumberFormatException e) {
                // skip malformed rows
            }
        }
    }
}
```

Reducer Class

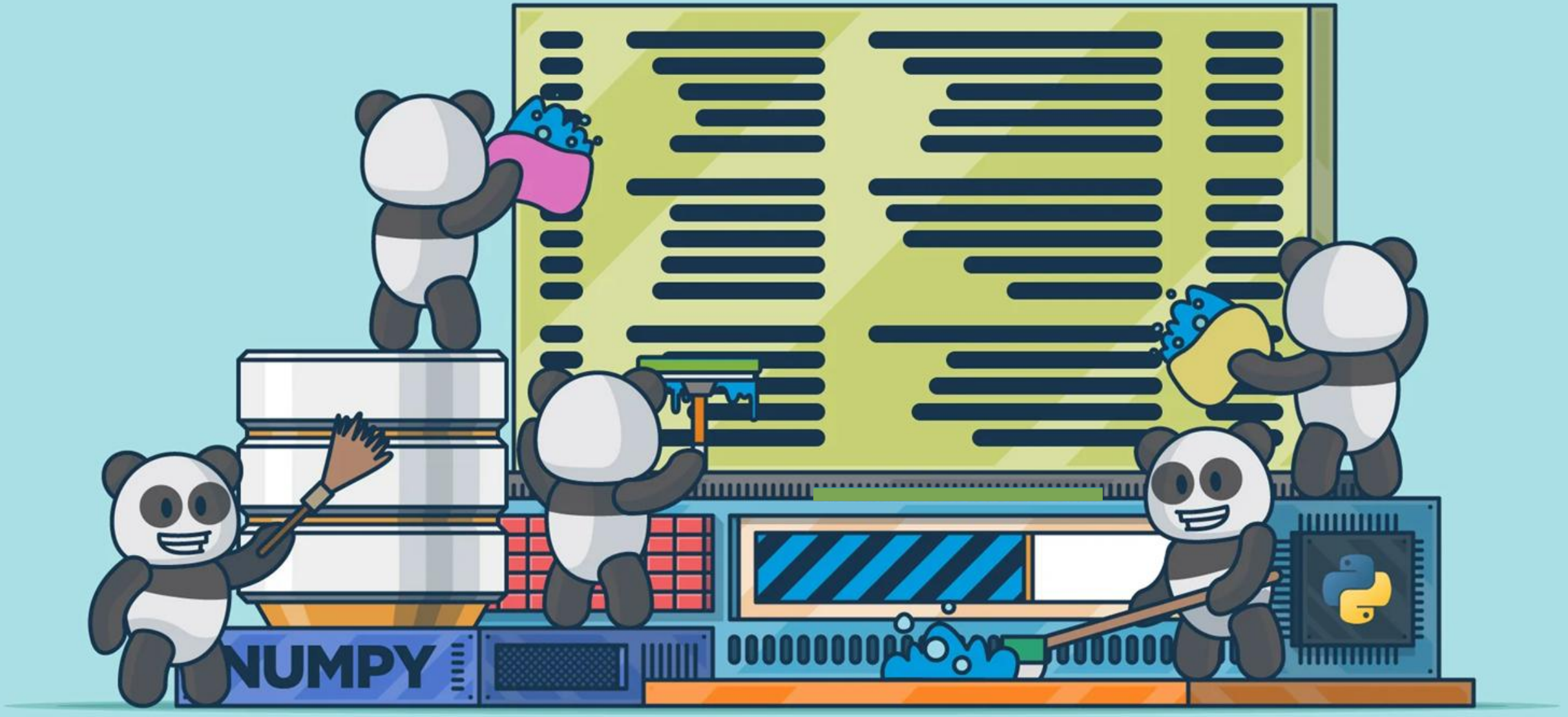
```java
// Adis Mahmic
// Dr. Price
// 04/12/2025
// Data Intensive Fundamentals

// Description:
// CrashReducer Class takes the output from the CrashMapper Class and
// aggregates the number of motorist fatalities based on the contributing
// factor for each vehicle involved in a crash. It sums the total number of
// killed motorists for each unique contributing factor, allowing analysis
// of which factors are associated with the most fatalities.

import ...

// Reducer class: receives (key, list of values) and outputs (key, aggregated value)
public class CrashReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {  2 usages

    // Reduce function: called for each unique key (contributing factor)
    // key: contributing factor (e.g., "Driver Inattention")
    // values: iterable list of IntWritable (fatalities associated with that factor)
    public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter reporter)
            throws IOException {

        int totalDeaths = 0;

        // Sum up all the fatalities associated with the same contributing factor
        while (values.hasNext()) {
            totalDeaths += values.next().get();
        }
        // Emit the contributing factor and the total number of fatalities
        output.collect(key, new IntWritable(totalDeaths));
    }
}
```

Driver Class

```java
// Adis Mahmic
// Dr. Price
// 04/12/2025
// Data Intensive Fundamentals

//Description:
// This class configures and launches the Hadoop MapReduce job to analyze motor vehicle crash data in new york city.
// Sets up job with the proper mapper and reducer classes (CrashMapper and CrashReducer).
// Specifies input and output paths, and defines the data types for keys and values used during the Map and Reduce phases.

// The goal is to determine the total number of motorist fatalities associated with
// each contributing factor in crashes, and see if distracted driving is the leading cause of death in car accidents.

import ...

public class CrashDriver {
    public static void main(String[] args) throws Exception {

        if (args.length != 2) {
            System.err.println("Usage: CrashDriver <input path> <output path>");
            System.exit( status: -1);
        }

        JobConf conf = new JobConf(CrashDriver.class);
        conf.setJobName("Motorist Killed by Contributing Factor");

        conf.setMapperClass(CrashMapper.class);
        conf.setReducerClass(CrashReducer.class);
        conf.setCombinerClass(CrashReducer.class);

        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);

        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);

        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));

        JobClient.runJob(conf);
    }
}
```

```
drwxr-xr-x   - data supergroup          0 2025-04-09 18:18 CrashOutput
data@4860-01:/home/hadoop$ bin/hdfs dfs -ls CrashOutput
Found 2 items
-rw-r--r--   1 data supergroup          0 2025-04-09 18:18 CrashOutput/_SUCCESS
-rw-r--r--   1 data supergroup       1312 2025-04-09 18:18 CrashOutput/part-0000
0
data@4860-01:/home/hadoop$ bin/hdfs dfs -copyToLocal CrashOutput/part-00000 cras
h.txt
data@4860-01:/home/hadoop$ cat crash.txt
Accelerator Defective   0
Aggressive Driving/Road Rage    9
Alcohol Involvement     37
Animals Action  0
Backing Unsafely        1
Brakes Defective        0
Cell Phone (hand-Held)  0
Cell Phone (hands-free) 0
Driver Inattention/Distraction  76
Driver Inexperience     33
Driverless/Runaway Vehicle      2
Drugs (illegal) 3
Eating or Drinking      0
Failure to Keep Right   6
Failure to Yield Right-of-Way   28
Fatigued/Drowsy 0
Fell Asleep     2
Following Too Closely   7
Glare   1
Headlights Defective    0
Illnes  40
Lane Marking Improper/Inadequate        0
Listening/Using Headphones      0
Lost Consciousness      42
```

Hadoop Output



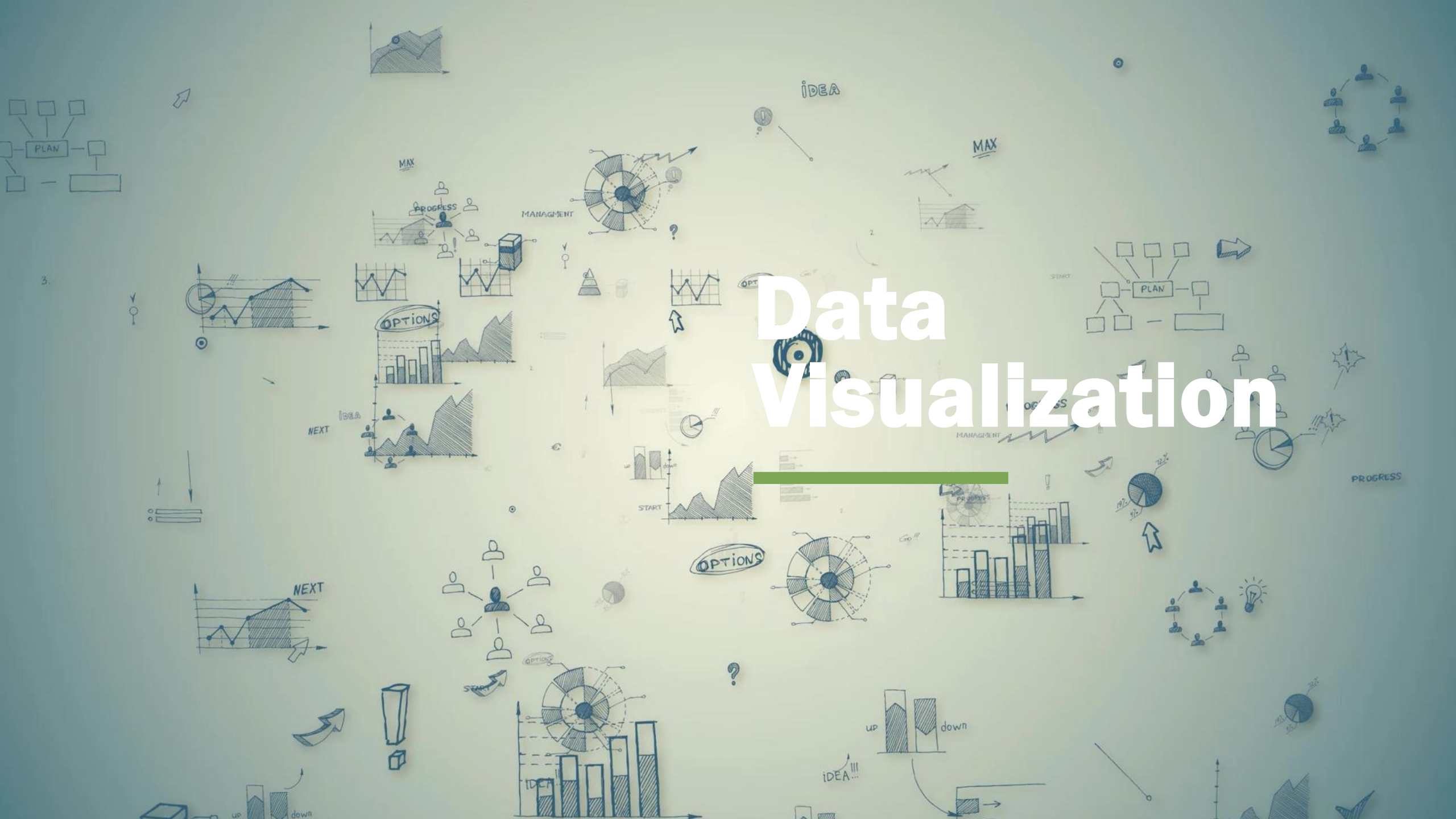
Cleaning the data

Real Python

Pandas Data Cleaning

In order to properly clean my data and ensure that there weren't any blank fields in my dataset, I used pandas to

```python
# Adis Mahmic
# Dr. Price
# 04/12/2025
# Data Intensive Fundamentals

# Description:
# DataProcessing.py uses the pandas framework to read our reduced dataset and then replaces empty
# and replaces them with the replacement value "Unspecified". If needed un-comment line 23 (print
# The changes are then saved to a new csv file called "Motor_Crash_Collisions_InitialData_cleaned

import pandas as pd

replacement_value = "Unspecified"
column_name = "CONTRIBUTING FACTOR VEHICLE 2"
column_name2 = "CONTRIBUTING FACTOR VEHICLE 1"
df = pd.read_csv("Motor_Crash_Collisions_InitialData.csv")

# Replace NaN and empty strings with "Unspecified"
df[[column_name, column_name2]] = df[[column_name, column_name2]].replace(
    to_replace=["", pd.NA, None, float('nan')], value=replacement_value
)
df[[column_name, column_name2]] = df[[column_name, column_name2]].fillna(replacement_value)

# print(df.head(10))

# Save it back to CSV if needed:
df.to_csv( path_or_buf: 'Motor_Crash_Collisions_InitialData_cleaned.csv', index=False)

```

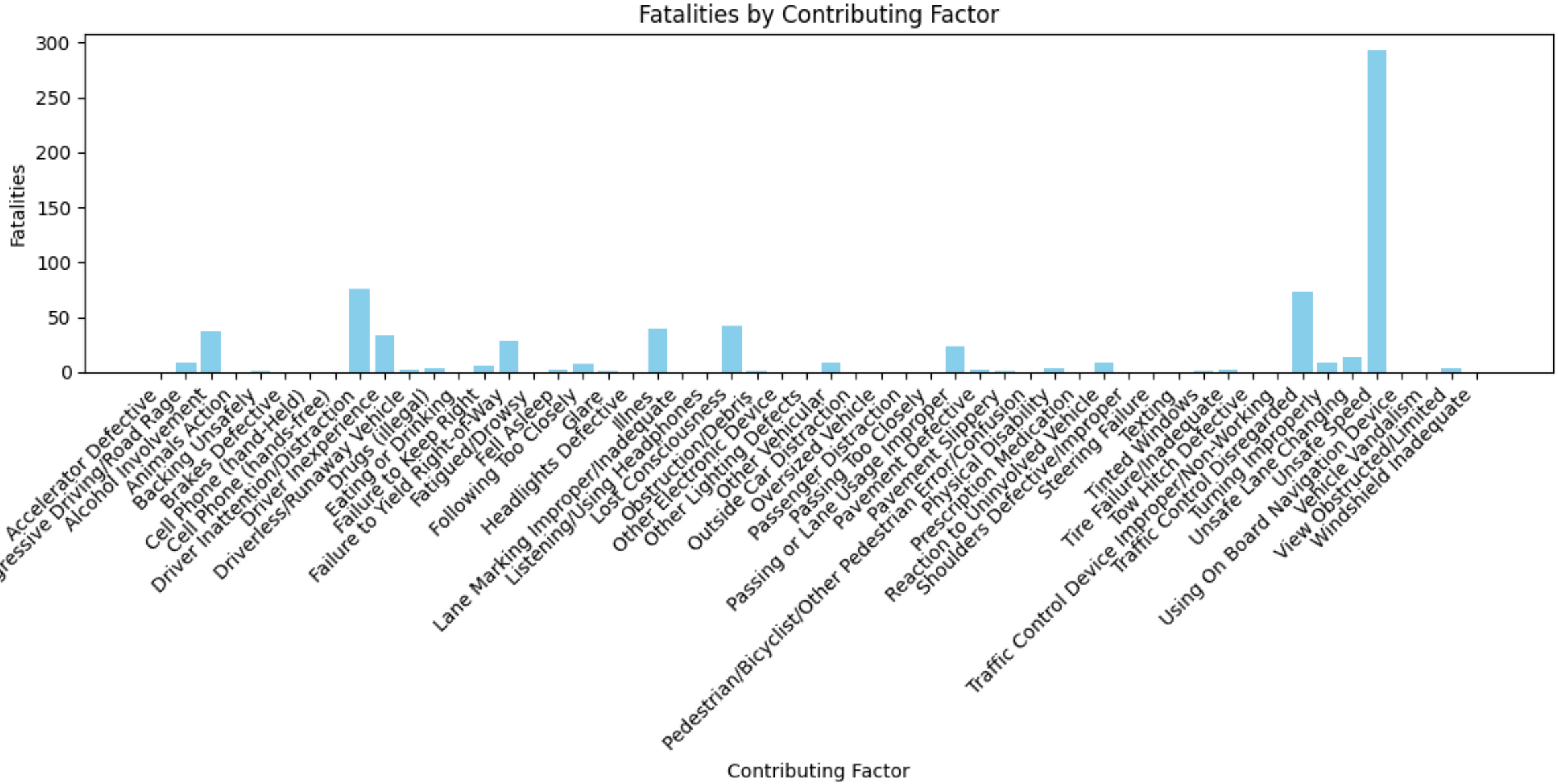

Data
Visualization

Data Visualization.py

To better convey the information extracted from the dataset, matplotlib was used to visualize the data.

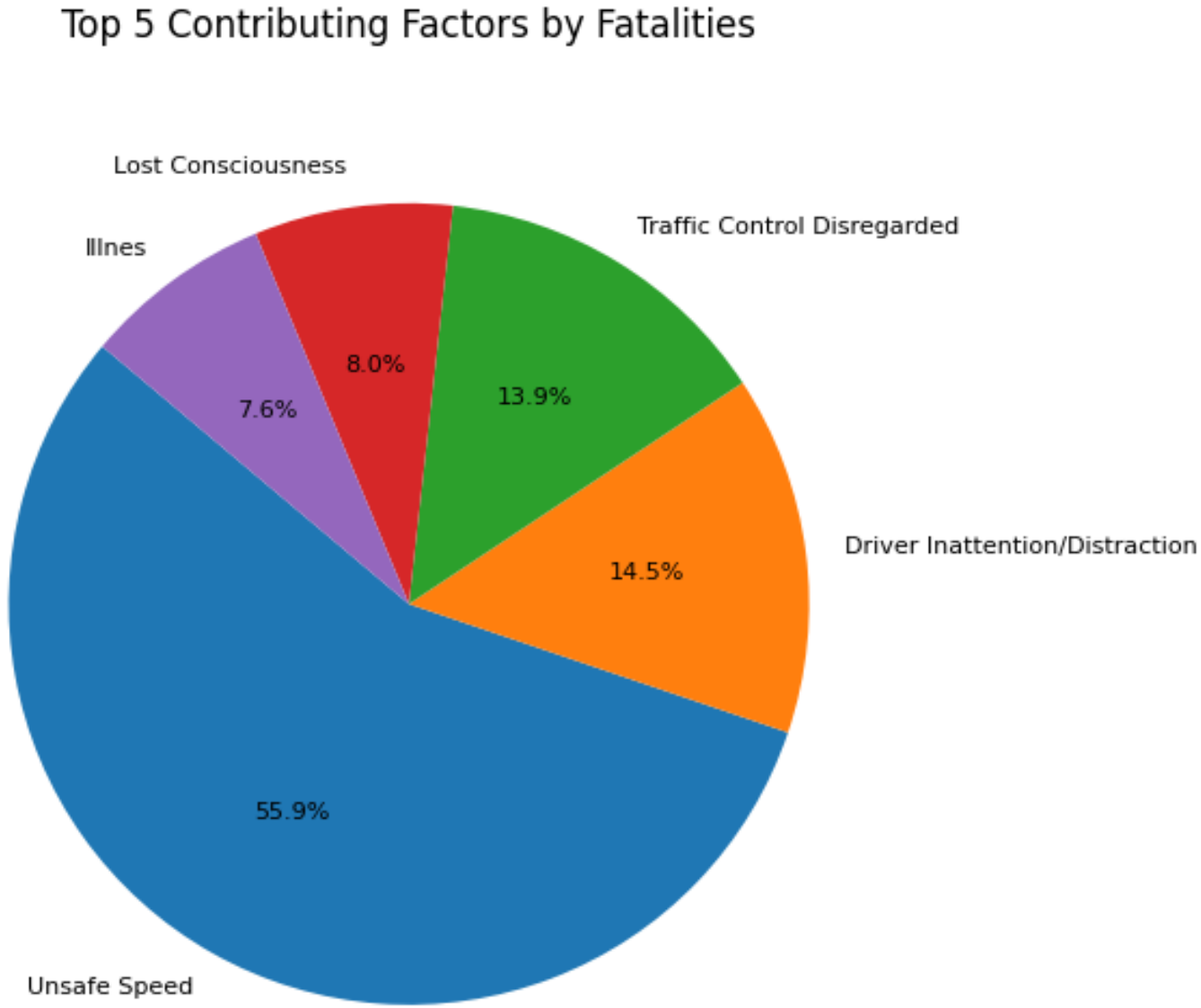
DataProcessing.py | Data Visualization.py | CrashOutput | Data

```python
import matplotlib.pyplot as plt

data = {}
with open("CrashOutput") as f:
    for line in f:
        parts = line.strip().split()
        if len(parts) >= 2:
            # The last part is the value
            value = int(parts[-1])

            # Everything before that is the key (join with space
            key = " ".join(parts[:-1])

            data[key] = value

# Plotting
plt.figure(figsize=(12, 6))  # Optional: makes the plot wider
plt.bar(data.keys(), data.values(), color='skyblue')
plt.xlabel('Contributing Factor')
plt.ylabel('Fatalities')
plt.title('Fatalities by Contributing Factor')
plt.xticks(rotation=45, ha='right')  # Tilt labels for readabili
plt.tight_layout()
plt.show()
```


Fatalities by Contributing Factor
Fatalities
300
250
200
150
100
50
0
Accelerator Defective
Aggressive Driving/Road Rage
Alcohol Involvement
Animals Action
Backing Unsafely
Brakes Defective
Cell Phone (hand-Held)
Cell Phone (hands-free)
Driver Inattention/Distraction
Driver Inexperience
Driverless/Runaway Vehicle
Drugs (illegal)
Eating or Drinking
Failure to Keep Right
Failure to Yield Right-of-Way
Fatigued/Drowsy
Fell Asleep
Following Too Closely
Glare
Headlights Defective
Illnes
Improper/Inadequate
Lane Marking
Listening/Using Headphones
Lost Consciousness
Obstruction/Debris
Other Electronic Device
Other Lighting Defects
Other Vehicular
Outside Car Distraction
Oversized Vehicle
Passenger Distraction
Passing Too Closely
Passing or Lane Usage Improper
Pavement Defective
Pavement Slippery
Pedestrian Error/Confusion
Pedestrian/Bicyclist/Other Pedestrian Error/Confusion
Physical Disability
Prescription Medication
Reaction to Uninvolved Vehicle
Shoulders Defective/Improper
Steering Failure
Texting
Tinted Windows
Tire Failure/Inadequate
Tow Hitch Defective
Traffic Control Device Improper/Non-Working
Traffic Control Disregarded
Turning Improperly
Unsafe Lane Changing
Unsafe Speed
Using On Board Navigation Device
Vehicle Vandalism
View Obstructed/Limited
Windshield Inadequate
Contributing Factor

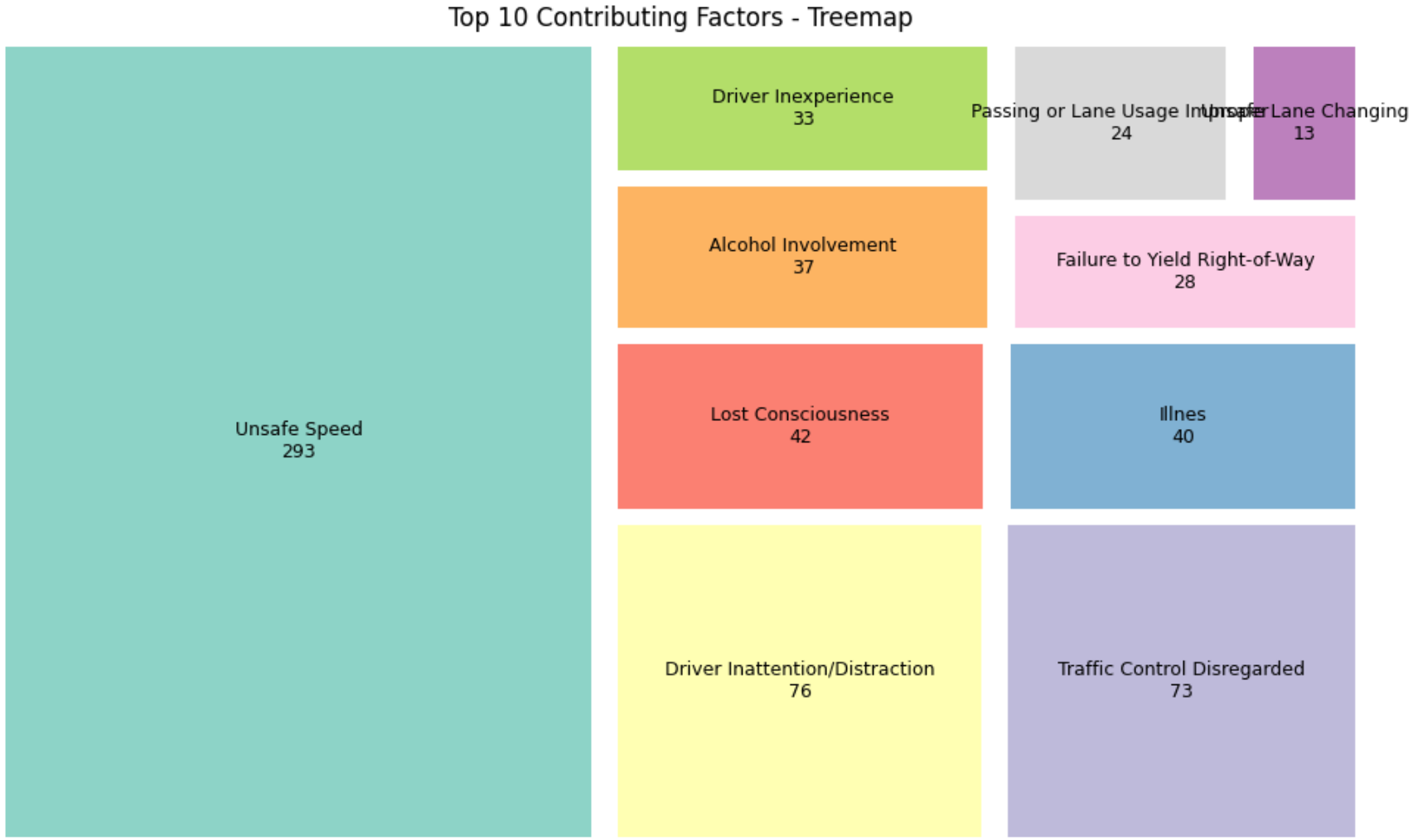
Pie Chart

```
# Creating a pie chart of the top 5 contributing factors
top_5 = dict(sorted(data.items(), key=lambda item: item[1], reverse=True)[:5])

plt.figure(figsize=(6, 6))
plt.title('Top 5 Contributing Factors by Fatalities')
plt.pie(
    top_5.values(),
    labels=top_5.keys(),
    autopct='%1.1f%%',
    textprops={'fontsize': 8},
    startangle=140
)
plt.axis('equal')
plt.tight_layout()
plt.show()
```


Top 5 Contributing Factors by Fatalities
Lost Consciousness
Illnes
Traffic Control Disregarded
8.0%
7.6%
13.9%
14.5%
Driver Inattention/Distraction
55.9%
Unsafe Speed

Tree Map

```
# Tree Map
# Sort data and select top 10
top_items = dict(sorted(data.items(), key=lambda item: item[1], reverse=True)[:10])
sizes = list(top_items.values())
labels = [f'{k}\n{v}' for k, v in top_items.items()]

# Plot treemap
plt.figure(figsize=(10, 6))
squarify.plot(
    sizes=sizes,
    label=labels,
    color=plt.cm.Set3.colors,
    pad=True,
    text_kwargs={'fontsize': 9}
)
plt.title('Top 10 Contributing Factors - Treemap')
plt.axis('off')
plt.tight_layout()
plt.show()
```


Top 10 Contributing Factors - Treemap
Unsafe Speed
293
Driver Inexperience
33
Passing or Lane Usage Improper
24
Unsafe Lane Changing
13
Alcohol Involvement
37
Failure to Yield Right-of-Way
28
Lost Consciousness
42
Illnes
40
Driver Inattention/Distraction
76
Traffic Control Disregarded
73

1. Unsafe Speeds is concerningly high
2. Hypothesis was not too off
3. Illness is one of the top 5 leading causes in motor accident fatalities?

Analysis

Conclusion

Although the hypothesis of distracted driving being the leading cause of death in motor collisions was incorrect, with unsafe speed taking first place. Distracted driving was still the second highest contributing factor in motor collision fatalities.

Thank you!

- Adis Mahmic
- https://github.com/AdisMahmic/Motor-Vehicle-Collisions

Citations

- https://www.ibm.com/think/topics/mapreduce
- https://github.com/AdisMahmic/Motor-Vehicle-Collisions