WORK RELATED FATALITY STUDY

An In-depth Analysis using Python and SQL

- PRIYANKA JINKA

TABLE OF CONTENTS

INTRODUCTION

- Workplace safety is a critical concern that affects not only individuals but also entire organisations and communities.
- The "Work-Related Fatality Study" delves into the analysis of fatal incidents that occur in various workplaces.
- This project aims to uncover insights, identify patterns, and improve workplace safety measures.

IMPORTANCE

Understanding the factors contributing to work-related fatalities is of utmost importance. It enables us to:

- Save Lives
- Protect Workers
- Improve Businesses
- Inform Policy

TOOLS USED

ESSENTIAL PANDAS FUNCTIONS FOR PERFORMING BASIC ANALYSIS AND CLEANING

- read_csv()
- head()
- info()
- shape
- describe()
- drop()
- isnull()
- isduplicated()
- to_csv()

DATA SOURCE

THE DATASET USED FOR THIS ANALYSIS WAS PROVIDED BY HICOUNSELOR AS PART OF THEIR 'ONLINE NON-CREDIT DATA ANALYSIS PROJECT'. IT CONTAINS RECORDS OF FATAL INCIDENTS THAT OCCURRED IN VARIOUS WORKPLACES.

THE DATASET CAN BE ACCESSED AT :
FATALITY DATASET

OBJECTIVE

ANALYZE WORK-RELATED FATAL INCIDENTS DATA TO:

- IDENTIFY CAUSES AND CONTRIBUTING FACTORS.
- DISCOVER TRENDS AND PATTERNS.
- PROVIDE RECOMMENDATIONS FOR ENHANCING WORKPLACE SAFETY.

DATASET OVERVIEW

- In our dataset, we have a total of 14,914 records, which represent individual data points or observations.
- These records are organized into 9 columns, each of which corresponds to a specific attribute or variable.
- These columns capture various aspects of the fatal incidents under investigation.

1. ID : A unique identifier for each incident.
2. Incident Date : The date when the fatal incident occurred.
3. Day of the Week : The day of the week corresponding to the incident date.
4. City : The name of the city where the incident took place.
5. State : The state where the incident occurred.

6. Description : A detailed account of the incident's circumstances and causes.
7. Plan : Indicates existing safety plans or protocols at the workplace.
8. Citation : Records information about regulatory voilations issued.
9. Unnamed : A column with '0' for all rows (not required for analysis).

TITLE : fatality.csv

id	incident_date	day_of_week	city	state	description	plan	citation	Unnamed
1	2009-06-23	tuesday	tampa	florida	Victim was inside a build	unknown	unknown	0
2	2009-06-30	tuesday	conroe	texas	Victim was on a scaffold	unknown	unknown	0
3	2009-07-06	monday	miami	florida	Victim was conducting r	unknown	unknown	0
4	2009-07-09	thursday	north platte	nebraska	Victim was climbing a la	unknown	unknown	0
5	2009-07-11	saturday	greensburg	pennsylvania	Employees were throwir	unknown	unknown	0
6	2009-07-11	saturday	boulder	colorado	Fatality at a bookstore u	unknown	unknown	0
7	2009-07-15	wednesday	raleigh	mississippi	Victim was pressure test	unknown	unknown	0
8	2009-07-16	thursday	marietta	georgia	Victim was standing on t	unknown	unknown	0
9	2009-07-18	saturday	post	texas	Victim was working on a	unknown	unknown	0
10	2009-07-18	saturday	de kalb	texas	Victim was installing coa	unknown	unknown	0

BASIC DATA ANALYSIS

DATA LOADING

LOADED THE DATASET INTO A PANDAS DATAFRAME AND DISPLAYED THE FIRST FEW ROWS.

```
import pandas as pd
df = pd.read_csv('fatality.csv')
df.head()
```

	id	incident_date	day_of_week	city	state	description	plan	citation	Unnamed
0	1	2009-06-23	tuesday	tampa	florida	Victim was inside a building bundling material...	unknown	unknown	0
1	2	2009-06-30	tuesday	conroe	texas	Victim was on a scaffold doing repairs on an a...	unknown	unknown	0
2	3	2009-07-06	monday	miami	florida	Victim was conducting roofing work, and fell t...	unknown	unknown	0
3	4	2009-07-09	thursday	north platte	nebraska	Victim was climbing a ladder while servicing a...	unknown	unknown	0
4	5	2009-07-11	saturday	greensburg	pennsylvania	Employees were throwing away old files. They h...	unknown	unknown	0

```
df.shape
```

→ `(14914, 9)`

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14914 entries, 0 to 14913
Data columns (total 9 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   id             14914 non-null  int64 
 1   incident_date  14914 non-null  object
 2   day_of_week    14914 non-null  object
 3   city           14914 non-null  object
 4   state          14914 non-null  object
 5   description    14914 non-null  object
 6   plan           14914 non-null  object
 7   citation       14914 non-null  object
 8   Unnamed        14914 non-null  int64 
dtypes: int64(2), object(7)
memory usage: 1.0+ MB
```

```
df.describe()
```

	id	incident_date	day_of_week	city	state	description	plan	citation
count	14914.000000	14914	14914	14914	14914	14914	14914	14914
unique	NaN	4154	7	5508	56	11569	3	3
top	NaN	2017-04-10	tuesday	houston	texas	Worker died in fall from roof.	unknown	unknown
freq	NaN	15	2728	219	1758	274	8886	8886
mean	7457.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	4305.445292	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	3729.250000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	7457.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	11185.750000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	14914.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN

DATA EXPLORATION

EXAMINED :

- **DATASET SHAPE:** THE DATASET CONTAINS 14,914 RECORDS AND 9 COLUMNS.
- **DATASET INFORMATION:** REVIEWED DATASET INFORMATION TO UNDERSTAND DATA TYPES AND NON-NULL COUNTS.
- **STATISTICAL SUMMARY:** REVIEWED THE STATISTICAL SUMMARY OF THE DATASET FOR INITIAL INSIGHTS.

BASIC DATA ANALYSIS

DATA CLEANING

REMOVED THE 'UNNAMED' COLUMN.

	id	incident_date	day_of_week	city	state	description	plan	citation
0	1	2009-06-23	tuesday	tampa	florida	Victim was inside a building bundling material...	unknown	unknown
1	2	2009-06-30	tuesday	conroe	texas	Victim was on a scaffold doing repairs on an a...	unknown	unknown
2	3	2009-07-06	monday	miami	florida	Victim was conducting roofing work, and fell t...	unknown	unknown
3	4	2009-07-09	thursday	north platte	nebraska	Victim was climbing a ladder while servicing a...	unknown	unknown
4	5	2009-07-11	saturday	greensburg	pennsylvania	Employees were throwing away old files. They h...	unknown	unknown

```
df.drop(columns=['Unnamed'],inplace = True)
df.head()
```

```
df.isnull().sum()
```

```
id              0
incident_date   0
day_of_week     0
city            0
state           0
description     0
plan            0
citation        0
dtype: int64
```

HANDLING NULL VALUES

- CHECKED FOR NULL VALUES
- NO NULL VALUES FOUND IN THE DATASET.

BASIC DATA ANALYSIS

HANDLING DUPLICATES

- CHECKED FOR DUPLICATE ROWS
- NO DUPLICATE ROWS FOUND IN THE DATASET.

```
df.duplicated().sum()
```

```
0
```

DATA EXPORT

- DATA IS CLEAN AND READY FOR ANALYSIS.
- EXPORTED THE CLEANED DATASET TO FATALITIES_CLEANED.CSV FOR FURTHER USE.

```
df.to_csv('fatalities_cleaned.csv',index=False)
```

id	incident_date	day_of_week	city	state	description	plan	citation
1	2009-06-23	tuesday	tampa	florida	Victim was inside a build	unknown	unknown
2	2009-06-30	tuesday	conroe	texas	Victim was on a scaffold	unknown	unknown
3	2009-07-06	monday	miami	florida	Victim was conducting r	unknown	unknown
4	2009-07-09	thursday	north platte	nebraska	Victim was climbing a la	unknown	unknown
5	2009-07-11	saturday	greensburg	pennsylvania	Employees were throwir	unknown	unknown
6	2009-07-11	saturday	boulder	colorado	Fatality at a bookstore u	unknown	unknown
7	2009-07-15	wednesday	raleigh	mississippi	Victim was pressure test	unknown	unknown
8	2009-07-16	thursday	marietta	georgia	Victim was standing on t	unknown	unknown
9	2009-07-18	saturday	post	texas	Victim was working on a	unknown	unknown
10	2009-07-18	saturday	de kalb	texas	Victim was installing coa	unknown	unknown

FINDING INSIGHTS USING MYSQL

ESSENTIAL SQL TOPICS FOR EFFECTIVELY ADDRESSING THE QUESTIONS

- BASIC AGGREGATIONS
- DATE TIME FUNCTIONS
- WILDCARD OPERATORS
- WINDOW FUNCTIONS
- ORDERBY, LIMIT

WHAT IS THE NUMBER OF REPORTED INCIDENTS?

```sql
select count(*) as Incidents_reported from fatalities;
```

Result Grid

Incidents_reported
14914

INSIGHTS : THE DATASET CONTAINS A TOTAL OF 14,914 REPORTED INCIDENTS.

WHAT IS THE YEAR TO YEAR CHANGE FOR THE NUMBER OF FATAL INCIDENTS?

```sql
SELECT YEAR(fc.incident_date) AS year, COUNT(*) AS n_fat,
LAG(COUNT(*), 1) OVER (ORDER BY YEAR(incident_date)) AS prev,
ROUND(((COUNT(*) - LAG(COUNT(*),1) OVER (ORDER BY YEAR(incident_date)))/LAG(COUNT(*),1)
OVER (ORDER BY YEAR(incident_date))) * 100) AS y2y
FROM fatalities fc
WHERE YEAR(fc.incident_date) < 2022
GROUP BY year;
```

Result Grid Filter Rows:

year	n_fat	prev	y2y
2009	515	NULL	NULL
2010	1110	515	116
2011	1185	1110	7
2012	997	1185	-16
2013	1189	997	19
2014	1345	1189	13
2015	1148	1345	-15
2016	1106	1148	-4
2017	1541	1106	39
2018	1260	1541	-18
2019	1376	1260	9
2020	1119	1376	-19
2021	950	1119	-15

WHAT IS THE NUMBER OF FATALITIES THAT RECEIVED A CITATION?

```sql
select  citation, count(*) as count
from fatalities
group by citation;
```

Result Grid Filter

citation	count
unknown	8886
yes	3345
no	2683

INSIGHTS : THERE WERE 3,345 FATALITIES THAT RECEIVED A CITATION.

WHAT DAY OF THE WEEK HAS THE MOST FATALITIES AND WHAT IS THE OVERALL PERCENTAGE?

```sql
select day_of_week, count,round((count*100/(select COUNT(*) from fatalities)),2) as percentage
from ( select day_of_week,count(*) as count
from fatalities
group by day_of_week
order by count desc) a;
```

Result Grid Filter Rows:

day_of_week	count	percentage
tuesday	2728	18.29
wednesday	2706	18.14
monday	2626	17.61
thursday	2612	17.51
friday	2335	15.66
saturday	1177	7.89
sunday	730	4.89

INSIGHTS : WITH AN OVERALL PERCENTAGE OF 18.29%, THE HIGHEST NUMBER OF FATALITIES OCCURRED ON 'TUESDAY'.

WHAT IS THE NUMBER OF FATALITIES INVOLVING IN WELDING?

```sql
select count(description) as welding_fatalities
from fatalities
where description like '%weld%';
```

Result Grid Filter

welding_fatalities
79

INSIGHTS : THERE WERE 79 FATALITIES ASSOCIATED WITH WELDING.

SELECT THE LAST 5 FROM THE PREVIOUS QUERY I.E TO CALCULATES THE LAST 5 FATALITIES DURING WELDING.

```sql
select *
from fatalities
where description like '%weld%'
order by incident_date desc
limit 5;
```

Result Grid Filter Rows: Search Export:

id	incident_date	day_of_week	city	state	description	plan	citation
9588	2021-04-14	wednesday	cleveland	ohio	Worker electrocuted by portable welding machine.	federal	yes
9813	2021-01-30	saturday	mission	texas	Worker died in welding explosion.	federal	yes
10000	2020-12-10	thursday	urbana	ohio	Worker fatally crushed by seam welder.	federal	yes
10679	2020-05-24	sunday	dallas	texas	Worker electrocted while welding HVAC pipe.	federal	no
11744	2019-07-08	monday	kingwood	texas	Worker electrocuted while welding air conditioner unit.	federal	no

INSIGHTS : IN THE LAST THREE YEARS, THERE HAVE BEEN A TOTAL OF FIVE WELDING-RELATED FATALITIES: TWO IN 2021 & 2020 EACH, AND ONE IN 2019.

SELECT THE TOP 5 STATES WITH THE MOST FATAL INCIDENTS.

```sql
select state,count(*) as fatal_count
from fatalities
group by state
order by fatal_count desc
limit 5;
```

Result Grid Filt

state	fatal_count
texas	1758
california	1352
florida	1021
new york	726
illinois	635

INSIGHTS : THE STATES WITH THE HIGHEST NUMBER OF FATAL INCIDENTS, IN THE TOP 5, ARE TEXAS, CALIFORNIA, FLORIDA, NEW YORK, AND ILLINOIS.

WHAT ARE THE TOP 5 STATES THAT HAD THE MOST WORKPLACE FATALITIES FROM STABBINGS?

```sql
select state,count(*) as stab_counts
from fatalities
where description like '%stabb%'
group by state
order by stab_counts desc
limit 5;
```

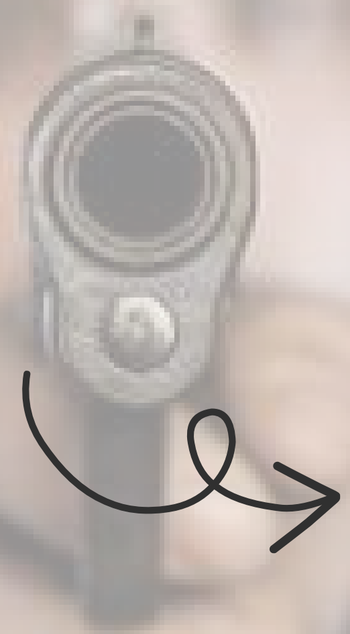
Result Grid Filter

state	stab_counts
new york	7
kentucky	5
california	5
illinois	3
washington	2

INSIGHTS : TEXAS, KENTUCKY, CALIFORNIA, ILLINOIS, AND WASHINGTON ARE THE TOP 5 STATES WITH THE HIGHEST NUMBER OF WORKPLACE FATALITIES RESULTING FROM STABBINGS.

WHAT ARE THE TOP 10 STATES THAT HAD THE MOST WORKPLACE FATALITIES FROM SHOOTINGS?

```sql
select state,count(*) as shoot_counts
from fatalities
where description like '%shot%'
group by state
order by shoot_counts desc
limit 10;
```



Result Grid Filter Rc

state	shoot_counts
indiana	28
california	23
texas	21
new york	20
florida	14
kentucky	13
oregon	9
nevada	9
illinois	9
washington	8

INSIGHTS : THE STATES WITH THE HIGHEST NUMBER OF WORKPLACE FATALITIES FROM SHOOTINGS, IN DESCENDING ORDER, ARE INDIANA, CALIFORNIA, TEXAS, NEW YORK, FLORIDA, KENTUCKY, OREGON, NEVADA, ILLINOIS, AND WASHINGTON.

WHAT IS THE TOTAL NUMBER OF SHOOTING DEATHS PER YEAR?

```sql
select year(incident_date) as year,count(*) as shooting_deaths
from fatalities
where description like '%shot%'
group by year
order by shooting_deaths desc;
```

Result Grid | Filter Rc

year	shooting_deaths
2021	38
2015	28
2016	28
2020	27
2019	24
2018	21
2013	19
2014	18
2010	16
2017	14
2011	13
2009	10
2012	9
2022	2

INSIGHTS : THE ANNUAL COUNT OF SHOOTING DEATHS IS CALCULATED, WITH THE HIGHEST NUMBER OF FATALITIES RECORDED IN 2021, TOTALING 38.

INSIGHTS SUMMARY

- The dataset comprises a total of 14,914 reported workplace incidents.

- Among these, 3,345 fatalities received citations, indicating safety violations.

- The day with the highest number of fatalities, at 18.29%, was 'Tuesday.'

- There were 79 fatalities associated with welding incidents.

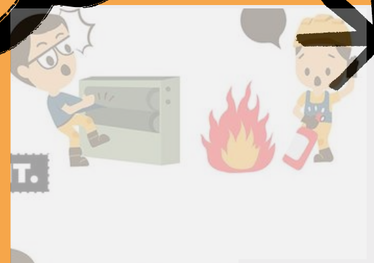
- Over the last three years, there were a total of five welding-related fatalities: two in 2021 and 2020 each, and one in 2019.

- The top 5 states with the highest numbers of fatal incidents are Texas, California, Florida, New York, and Illinois.

- The leading states for workplace fatalities due to stabbings are Texas, Kentucky, California, Illinois, and Washington.

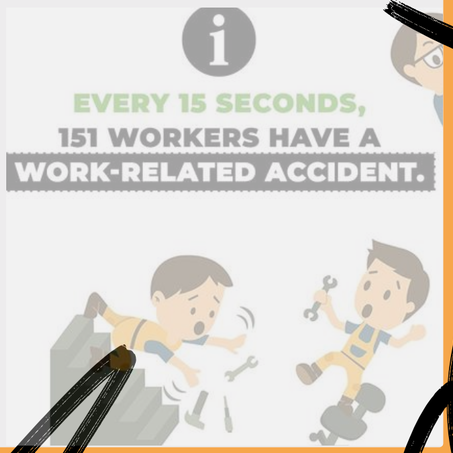
- States with the most workplace fatalities from shootings, in descending order, include Indiana, California, Texas, New York, Florida, Kentucky, Oregon, Nevada, Illinois, and Washington.

- The analysis also includes an annual count of shooting deaths, with the highest number recorded in 2021, totaling 38.




SAY
YES
TO
SAFETY

THANK YOU


EVERY 15 SECONDS,
151 WORKERS HAVE A
WORK-RELATED ACCIDENT.