**Dickson Muloma Chamwada**

**Data Dive**

**Dataset chosen:Human Stampede**

**Importing the Dataset into MySQL Workbench;**

SELECT \* FROM stampedes LIMIT 10;

SELECT COUNT(\*) as Total\_Stampedes

FROM stampedes;

SELECT Country, COUNT(\*) as Number\_of\_Stampedes

FROM stampedes

GROUP BY Country

ORDER BY Number\_of\_Stampedes

DESC LIMIT 10;

SELECT Year, COUNT(\*) as Number\_of\_Stampedes

FROM stampedes

GROUP BY Year

ORDER BY Year;

SELECT Cause, COUNT(\*) as Frequency

FROM stampedes

GROUP BY Cause

ORDER BY Frequency DESC;

SELECT Event\_Type, COUNT(\*) as Frequency FROM stampedes GROUP BY Event\_Type

ORDER BY Frequency DESC;

**Question 1-** Briefly explain any difficulties and an interesting thing you noticed about your chosen dataset?

Difficulties/Challenges

**Data Type Mismatches**: Some columns required manual adjustment of data types to match the dataset's content accurately.

**Handling Missing Values**:missing data in some fields required handling strategies, such as using SQL functions to manage NULL values..

**Special Characters and Encoding**: The dataset included special characters that caused issues with text encoding. Ensuring the file was in UTF-8 format before importing helped resolve these issues.

Interesting Fact

**Geographical Hotspots**: The dataset revealed that India has the highest number of recorded stampedes, often associated with large-scale public events like festivals and religious gatherings.

**Question 2**. Data Fun

Use simple sql queries to play with the data

**Sql queries**

SELECT Year, COUNT(\*) AS Total\_Stampedes

FROM stampedes

GROUP BY Year

ORDER BY Year;

SELECT SUM(Fatalities) AS Total\_Fatalities\_Religious FROM stampedes WHERE Event\_Type = 'Religious';

Two hidden facts hidden within the data;

1.The dataset reveals a fluctuating number of stampede events each year, with notable peaks in certain years.

2. From the second sql query I have discovered that religious events have led to a significant number of fatalities with a total number of 2,788.

**Question 3**- Ask away

Formulate two question about the data(e.g what are the popular shows in different countries)

Question 1: Which Countries Have the Highest Number of Stampedes During Religious Events?

**Sql query**

SELECT Country, COUNT(\*) AS Number\_of\_Stampedes

FROM stampedes

WHERE Event\_Type = 'Religious'

GROUP BY Country

ORDER BY Number\_of\_Stampedes

DESC LIMIT 5;

The query results shows the top 5 countries with the highest number of stampedes during religious events. For example, India, Nigeria, and Pakistan are among the top countries. It indicates these regions have a higher frequency of such incidents, possibly due to large religious gatherings.

Countries like India and Saudi Arabia frequently experience stampedes during large religious gatherings, highlighting the need for better crowd management strategies during such events.

Question 2: What is the Most Common Cause of Stampedes Across All Event Types?

**Sql query**

SELECT Cause,

COUNT(\*) AS Frequency

FROM stampedes

GROUP BY Cause

ORDER BY Frequency DESC LIMIT 1;

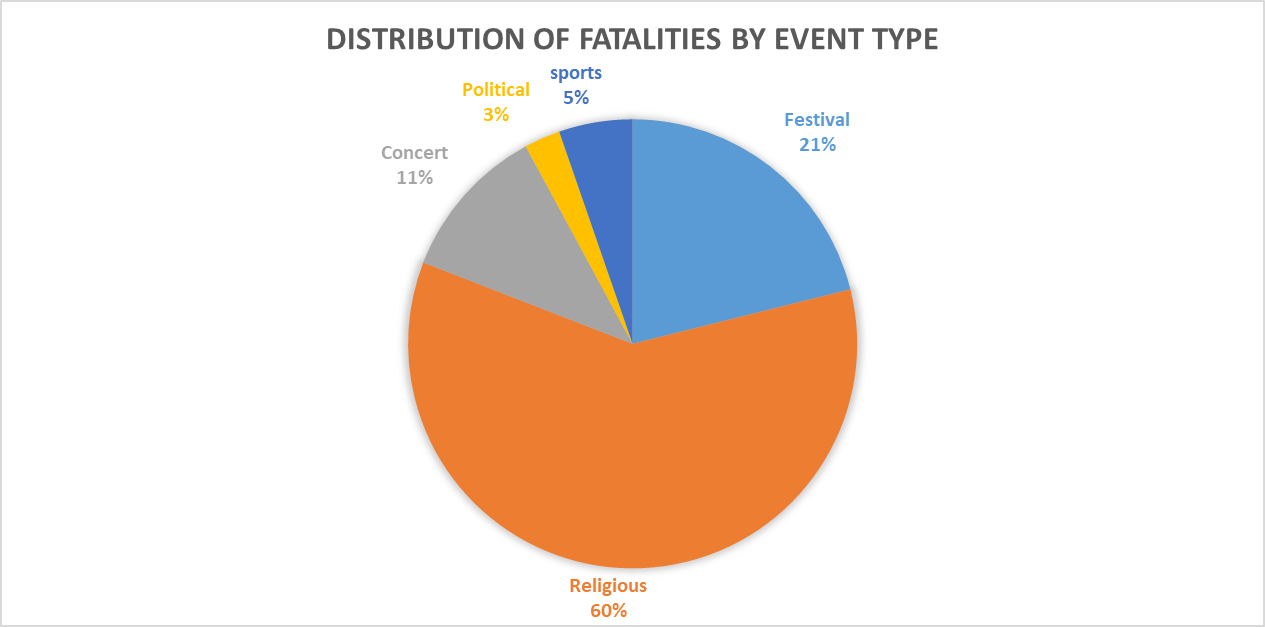
This query show the result of most frequent causes of stampedes(e.g mass panic)

From the two queries above and the human stampede dataset itself I have learned that;

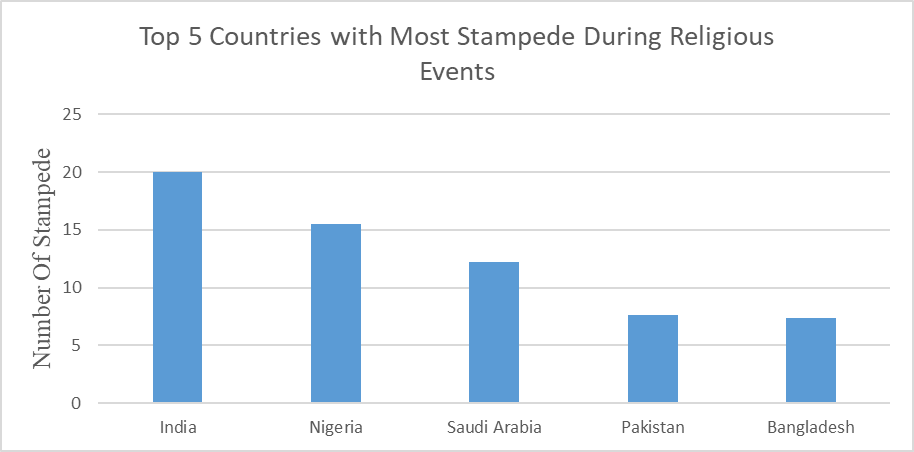
Countries like India and Nigeria often experience stampedes during religious events, highlighting the need for improved safety measures at large gatherings. And also Mass panic is the leading cause of stampedes, suggesting that sudden fear and confusion often trigger these tragic events.

**Question 4**. Showtime

Pie Chart



2. Bar Chart (Top Countries)



3. Bar Chart(Most causes of Stampede)

