SQL and Data Analytics

INFO8076 - Fall 2023 - Section 3

Group-1 Final Project

Canadian Disaster Overview

Group Members Name:

Aesha Bhatt
Dilin Devassy
Krisha Jain
Aashiya
Ishita Kohinoor
Divya Sukumaran

Introduction:

The dataset is taken from Public Safety Canada, which is related to the list of disasters or any incidents in Canada during a decade. The dataset has been cleaned to remove a few null values and blank spaces. The data set is divided into 4 tables namely Events, causalities, Location and Relief Aid. The event table consists of details regarding the type of the disaster which occurred, the type like if the disaster is caused due to any terrorist attack, flood, tsunami etc. The second table consist of the detail regarding the no of people affected, the causalities, the total estimated cost and the normalized cost. The third table consist of the detail regarding the location where the disaster occurred and the date and time of when the disaster occurred. The 4th and last table relief aid consist of information which gives details about any payment made by the federal, provincial government and insurance company to cover up the cost occurred due to the disaster. Using all these information we are determining few key areas from the dataset which will help us determine key focus area for the respective people to achieve the target.

ERD Model:

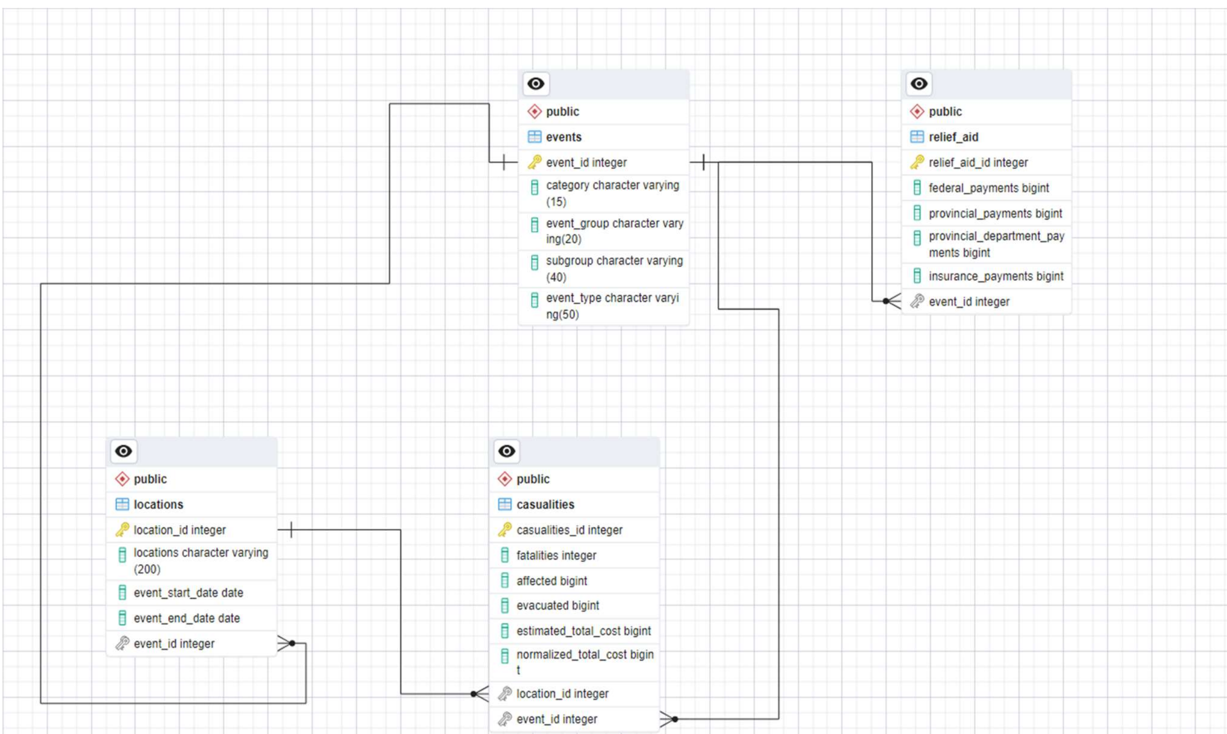


Table creation script:

```
CREATE TABLE IF NOT EXISTS public.casualities
(
    casualities_id integer NOT NULL,
    fatalities integer,
    affected bigint,
    evacuated bigint,
    estimated_total_cost bigint,
    normalized_total_cost bigint,
    location_id integer NOT NULL,
    event_id integer,
    CONSTRAINT casualities_pkey PRIMARY KEY (casualities_id)
);

CREATE TABLE IF NOT EXISTS public.events
(
    event_id integer NOT NULL,
    category character varying(15) COLLATE pg_catalog."default",
    event_group character varying(20) COLLATE pg_catalog."default",
    subgroup character varying(40) COLLATE pg_catalog."default",
    event_type character varying(50) COLLATE pg_catalog."default",
    CONSTRAINT events_pkey PRIMARY KEY (event_id)
);

CREATE TABLE IF NOT EXISTS public.locations
(
    location_id integer NOT NULL,
    locations character varying(200) COLLATE pg_catalog."default",
    event_start_date date,
    event_end_date date,
    event_id integer NOT NULL,
    CONSTRAINT locations_pkey PRIMARY KEY (location_id)
);

CREATE TABLE IF NOT EXISTS public.relief_aid
(
    relief_aid_id integer NOT NULL,
    federal_payments bigint,
    provincial_payments bigint,
    provincial_department_payments bigint,
    insurance_payments bigint,
    event_id integer NOT NULL,
    CONSTRAINT relief_aid_pkey PRIMARY KEY (relief_aid_id)
);
```

```
ALTER TABLE IF EXISTS public.casualities
    ADD CONSTRAINT location_id FOREIGN KEY (location_id)
    REFERENCES public.locations (location_id) MATCH SIMPLE
    ON UPDATE NO ACTION
    ON DELETE NO ACTION
    NOT VALID;


ALTER TABLE IF EXISTS public.casualities
    ADD CONSTRAINT event_id FOREIGN KEY (event_id)
    REFERENCES public.events (event_id) MATCH SIMPLE
    ON UPDATE NO ACTION
    ON DELETE NO ACTION
    NOT VALID;


ALTER TABLE IF EXISTS public.locations
    ADD CONSTRAINT locations_event_id_fkey FOREIGN KEY (event_id)
    REFERENCES public.events (event_id) MATCH SIMPLE
    ON UPDATE NO ACTION
    ON DELETE NO ACTION;


ALTER TABLE IF EXISTS public.relief_aid
    ADD CONSTRAINT relief_aid_event_id_fkey FOREIGN KEY (event_id)
    REFERENCES public.events (event_id) MATCH SIMPLE
    ON UPDATE NO ACTION
    ON DELETE NO ACTION;

END;
```

Queries:

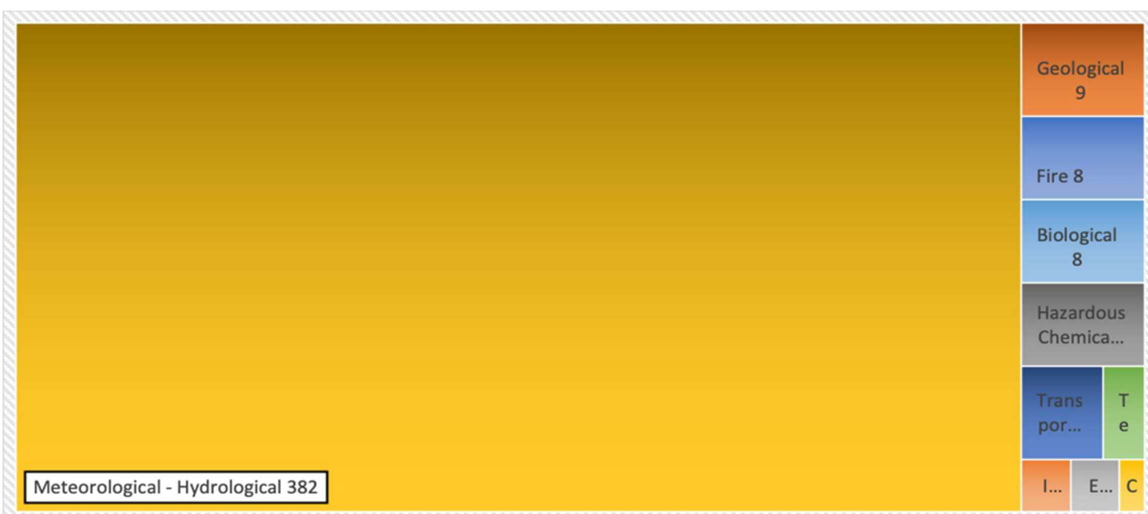
Query 1:

```
SELECT CATEGORY,
       SUBGROUP,
       COUNT(SUBGROUP) AS NO_OF_TIMES
FROM PUBLIC.EVENTS
GROUP BY CATEGORY,
       SUBGROUP;
```

Output:

category	subgroup	no_of_times
Incident	Fire	8
Incident	Infrastructure failure	2
Incident	Explosion	2
Incident	Civil Incident	1
Disaster	Biological	8
Incident	Terrorist	3
Incident	Transportation accident	6
Disaster	Geological	9
Incident	Hazardous Chemicals	8
Disaster	Meteorological - Hydrological	382

Visualization:



Insight:

The treemap visually highlights high-frequency events, and it's clear that "Meteorological - Hydrological" events are significantly more frequent than others. The treemap can guide decision-makers in identifying focus areas. For instance, if resources are limited, focusing on preparedness for "Meteorological - Hydrological" events may be a priority. With the given data we can derive from the output that Meteorological - Hydrological has the highest occurrence of 382 with the least of 1 with civil incident.

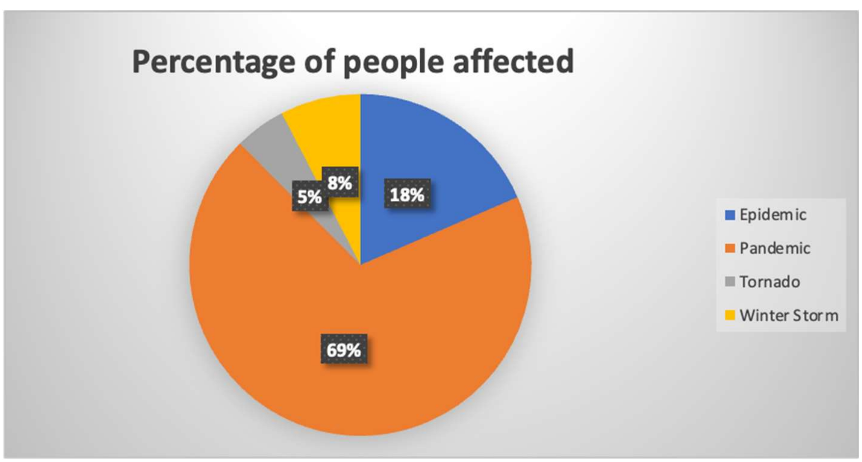
Query 2:

```
SELECT E.SUBGROUP,
        E.EVENT_TYPE,CAS.AFFECTED
FROM EVENTS E
INNER JOIN CASUALITIES CAS ON CAS.EVENT_ID = E.EVENT_ID
WHERE CAS.AFFECTED > 500
GROUP BY E.CATEGORY,
        E.SUBGROUP,
        E.EVENT_TYPE,
        CAS.AFFECTED;
```

Output:

subgroup	affected
Biological	2300
Biological	8582
Meteorological - Hydrological	600
Meteorological - Hydrological	945

Visualization:



Insight:

This query focuses on events where the number of individuals affected by different calamities exceeds 500. "Pandemic" stands out as the event type with the highest impact, affecting 69% of people. The pie chart reinforces the dominance of the "Pandemic" regarding the percentage of affected individuals, highlighting its significant impact. The chart provides a clear comparison of the impact percentages, allowing for easy identification of event types with the greatest and least impact and detailed information on high-impact disasters, offering valuable insights for emergency response, resource allocation, public health interventions, and policy development in natural calamities.

Query 3:

```
SELECT E.EVENT_GROUP,
        E.EVENT_TYPE,
        MAX(C.ESTIMATED_TOTAL_COST) AS MAX_COST,
        MIN(C.ESTIMATED_TOTAL_COST) AS MIN_COST,
        ROUND(AVG(CAST(C.ESTIMATED_TOTAL_COST AS numeric))) AS AVG_COST
FROM EVENTS E
JOIN CASUALITIES C ON E.EVENT_ID = C.EVENT_ID
GROUP BY E.EVENT_TYPE,
        E.EVENT_GROUP
HAVING COUNT(EVENT_TYPE) > 5
ORDER BY AVG_COST DESC;
```

Output:

event_group	event_type	max_cost	min_cost	avg_cost
Natural	Winter Storm	4635720433	0	231235828
Natural	Flood	2715742000	0	58649102
Natural	Storm - Unspecified / Other	400000000	0	51554259
Natural	Tornado	500000000	0	50892023
Natural	Drought	581891545	0	39668119
Natural	Storms and Severe Thunderstorms	884595372	0	27635244
Natural	Wildfire	300000000	0	16260440
Natural	Hurricane / Typhoon / Tropical Storm	85507690	0	10445505
Natural	Epidemic	25881403	0	4471390
Natural	Storm Surge	8005500	216978	4145825
Natural	Landslide	18208847	0	3475034

Insight:

Analyzes the maximum, minimum, and average estimated total costs for each event type. Filters out event types with fewer than five occurrences. Useful for understanding the cost distribution across different event types.

Query 4:

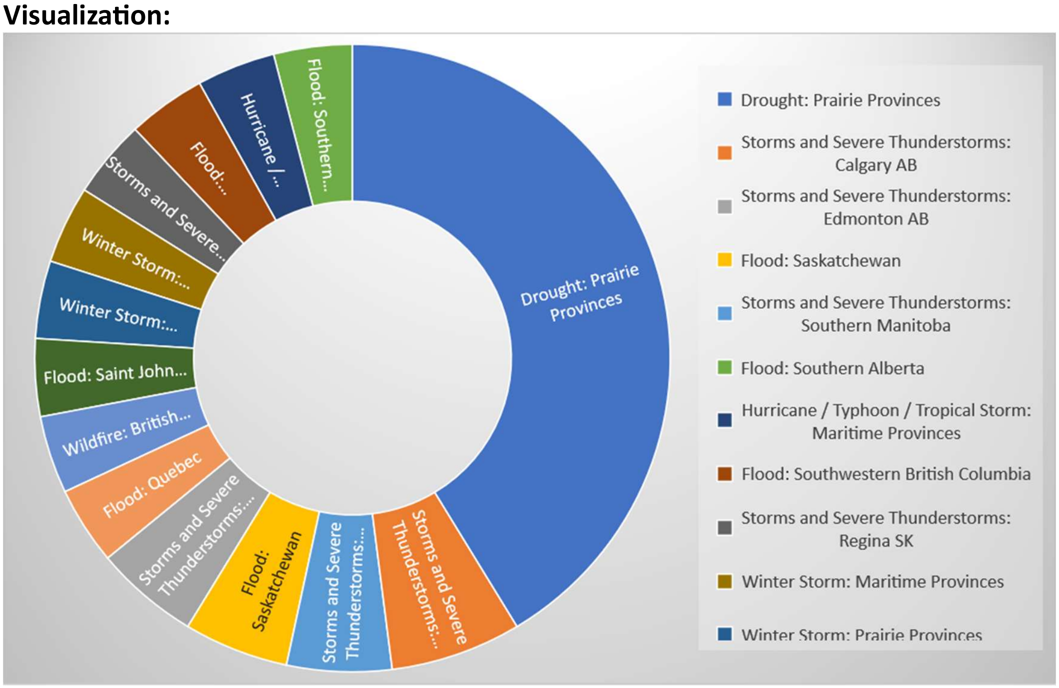
```
SELECT CONCAT(E.EVENT_TYPE,': ',L.LOCATIONS) AS EVENT_LOCATION,
        COUNT(E.EVENT_ID) AS EVENT_COUNT
FROM EVENTS E
JOIN LOCATIONS L ON E.EVENT_ID = L.EVENT_ID
```

```
GROUP BY EVENT_LOCATION
HAVING COUNT(E.EVENT_ID) >= 3
ORDER BY EVENT_COUNT DESC;
```

Output:

event_location	event_count
Drought: Prairie Provinces	31
Storms and Severe Thunderstorms: Calgary AB	5
Storms and Severe Thunderstorms: Edmonton AB	4
Flood: Saskatchewan	4
Storms and Severe Thunderstorms: Southern Manitoba	4
Flood: Southern Alberta	3
Hurricane / Typhoon / Tropical Storm: Maritime Provinces	3
Flood: Southwestern British Columbia	3
Storms and Severe Thunderstorms: Regina SK	3
Winter Storm: Maritime Provinces	3
Winter Storm: Prairie Provinces	3
Flood: Saint John River Basin NB	3
Wildfire: British Columbia	3
Flood: Quebec	3

Visualization:



Insight:

The output of this query provide practical guidance on how to allocate resources most effectively, build community resilience, and create focused plans to deal with natural disasters or reoccurring disasters in particular areas.

Query 5:

```
SELECT
        E.EVENT_TYPE,
        C.FATALITIES,
        C.AFFECTED
FROM EVENTS E
JOIN CASUALITIES C ON E.EVENT_ID = C.EVENT_ID
WHERE C.FATALITIES >
                (SELECT AVG(FATALITIES)
                        FROM CASUALITIES)
        AND C.AFFECTED >
                (SELECT AVG(AFFECTED)
                        FROM CASUALITIES)
ORDER BY C.FATALITIES DESC,
        C.AFFECTED DESC;
```

Output:

event_type	fatalities	affected
Pandemic	425	8582
Winter Storm	35	945
Tornado	27	600
Epidemic	22	57
Epidemic	13	180
Tornado	12	500
Tornado	12	140
Epidemic	6	2300
Tornado	6	200
Rail	3	45

Insight:

The output of this query provides insights into events with above-average casualties, guiding emergency response, resource allocation, community resilience planning, and policy development in the context of disasters or natural calamities.

Query 6:

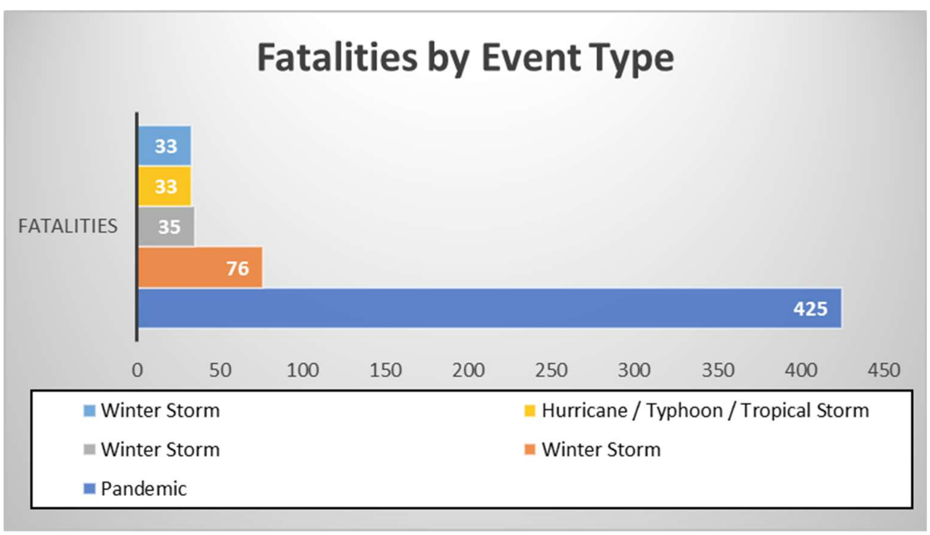
```
WITH FATALITIESCTE AS
        (SELECT EVENT_ID,
```

```
                                FATALITIES,
                                ROW_NUMBER() OVER (
ORDER BY FATALITIES DESC) AS RN
                    FROM CASUALITIES)
SELECT
        E.EVENT_TYPE,
        F.FATALITIES
FROM FATALITIESCTE F
JOIN EVENTS E ON F.EVENT_ID = E.EVENT_ID
WHERE F.RN <= 5
order by F.FATALITIES  DESC;
```

Output:

event_type	fatalities
Pandemic	425
Winter Storm	76
Winter Storm	35
Winter Storm	33
Hurricane / Typhoon / Tropical Storm	33

Visualization:



Insight:
This query uses a CTE to rank events based on the number of fatalities. It selects the top 5 events with the highest fatalities and includes their event type and fatality count. The top 5 events represent different event types, suggesting that fatalities are not limited to a specific type of event.

It identifies and retrieves information on the top 5 events with the highest fatalities offers valuable insights and opportunities for utilization in disaster management and mitigation efforts. The data can be leveraged to enhance various aspects of emergency response, public awareness, infrastructure resilience, and global collaboration.

Query 7:

```
SELECT E.CATEGORY,
        E.SUBGROUP,
        SUM(C.FATALITIES) AS TOTAL_FATALITIES,
        SUM(C.AFFECTED) AS TOTAL_AFFECTED,
        SUM(C.ESTIMATED_TOTAL_COST) AS TOTAL_ESTIMATED_COST
FROM EVENTS E
JOIN CASUALITIES C ON E.EVENT_ID = C.EVENT_ID
GROUP BY E.CATEGORY,
        E.SUBGROUP
ORDER BY TOTAL_FATALITIES DESC,
        TOTAL_AFFECTED DESC,
        TOTAL_ESTIMATED_COST DESC;
```

Output:

category	subgroup	total_fatalities	total_affected	total_estimated_cost
Disaster	Biological	476	11140	31299733
Disaster	Meteorological - Hydrological	377	2478	19653976738
Incident	Transportation accident	54	71	0
Disaster	Geological	39	1	62295905
Incident	Fire	32	15	0
Incident	Terrorist	13	1	0
Incident	Explosion	2	68	0
Incident	Civil Incident	0	144	0
Incident	Hazardous Chemicals	0	6	26931008
Incident	Infrastructure failure	0	0	0

Insight:

Different impacts are presented by the data for different event categories. The greatest toll from biological disasters, particularly pandemics, is 476 deaths and a significant number of impacted people (11,140), highlighting the severity and global reach of these events. Although they result in fewer deaths (377), meteorological-hydrological disasters have enormous financial costs (19,653,976,738), highlighting their financial impact. The immediate human cost is highlighted by the 54 fatalities and 71 injured in transportation accidents. Geological disasters have a negligible effect on human life (39 fatalities) but have significant economic ramifications (62,295,905). Events with differing degrees of impact, such as explosions, fires, and terrorist acts, highlight the complex nature of disasters.

Query 8:

```
WITH EVENTSEVERITY AS
	(SELECT E.EVENT_TYPE,
			C.FATALITIES,
			CASE
				WHEN C.FATALITIES = 0 THEN 'No Fatalities'
				WHEN C.FATALITIES <= 10 THEN 'Low Fatalities'
				WHEN C.FATALITIES <= 50 THEN 'Medium Fatalities'
					ELSE 'High Fatalities'
			END AS SEVERITY_CATEGORY
		FROM EVENTS E
		JOIN CASUALITIES C ON E.EVENT_ID = C.EVENT_ID)
SELECT ES.EVENT_TYPE,
	ES.FATALITIES,
	ES.SEVERITY_CATEGORY
FROM EVENTSEVERITY ES
WHERE ES.SEVERITY_CATEGORY = 'High Fatalities'
	OR ES.SEVERITY_CATEGORY = 'Medium Fatalities'
ORDER BY ES.FATALITIES DESC;
```

Output:

event_type	fatalities	severity_category
Pandemic	425	High Fatalities
Winter Storm	76	High Fatalities
Winter Storm	35	Medium Fatalities
Hurricane / Typhoon / Tropical Storm	33	Medium Fatalities
Winter Storm	33	Medium Fatalities
Residential	32	Medium Fatalities
Air	32	Medium Fatalities
Landslide	31	Medium Fatalities
Tornado	27	Medium Fatalities
Winter Storm	23	Medium Fatalities
Epidemic	22	Medium Fatalities
Storms and Severe Thunderstorms	21	Medium Fatalities
Epidemic	13	Medium Fatalities
Tornado	12	Medium Fatalities
Tornado	12	Medium Fatalities
Air	12	Medium Fatalities
Winter Storm	11	Medium Fatalities

Insight:
The below SQL query is extracting information about events and their severities based on the number of fatalities. The query creates a temporary table named EVENTSEVERITY using a Common Table Expression (CTE) to join the EVENTS and CASUALITIES tables. The severity category is determined using a CASE statement based on the number of fatalities, categorizing events into 'No Fatalities,' 'Low Fatalities,' 'Medium Fatalities,' or 'High Fatalities.'. Further filters were applied to pull all the data with Medium and High Fatalities after which it was sorted in descending order.

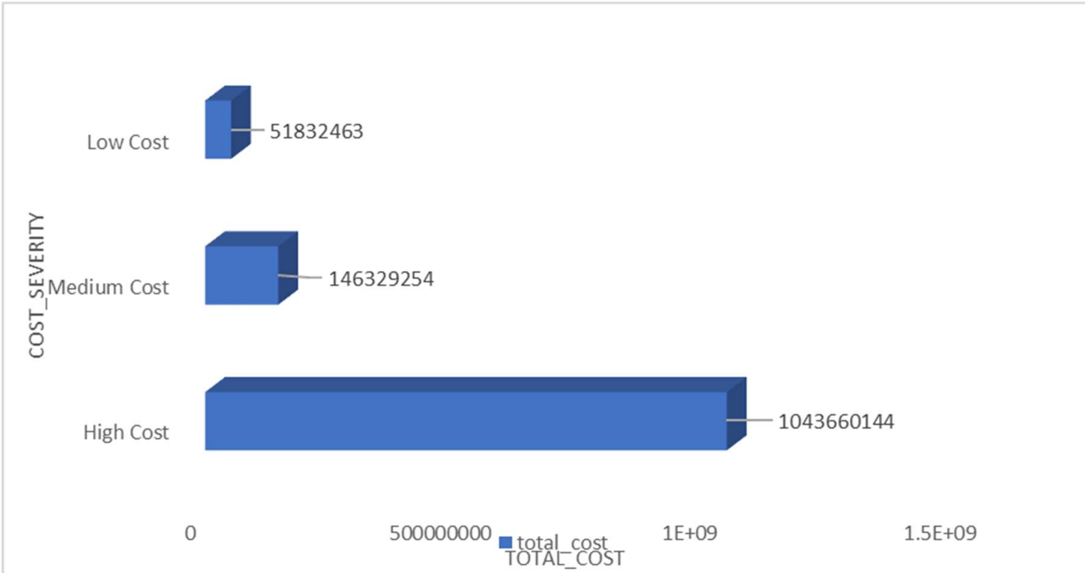
Query 9:
```
WITH EVENTCOSTSEVERITY AS
        (SELECT E.EVENT_ID,
                        E.EVENT_TYPE,
                        E.CATEGORY,
                        E.EVENT_GROUP,
                        C.NORMALIZED_TOTAL_COST,
                        CASE
                        WHEN C.NORMALIZED_TOTAL_COST < 1000000 THEN 'Low Cost'
                        WHEN C.NORMALIZED_TOTAL_COST >= 1000000

        AND C.NORMALIZED_TOTAL_COST < 5000000 THEN 'Medium Cost'
                        WHEN C.NORMALIZED_TOTAL_COST >= 5000000 THEN 'High Cost'
                                ELSE 'Unknown'
                        END AS COST_SEVERITY
                FROM PUBLIC.EVENTS E
                JOIN PUBLIC.CASUALITIES C ON E.EVENT_ID = C.EVENT_ID)
SELECT COST_SEVERITY,
        COUNT(*) AS EVENT_COUNT,
        SUM(NORMALIZED_TOTAL_COST) AS TOTAL_COST
FROM EVENTCOSTSEVERITY
GROUP BY COST_SEVERITY
ORDER BY TOTAL_COST DESC;
```

Output:

cost_severity	event_count	total_cost
High Cost	36	1043660144
Medium Cost	67	146329254
Low Cost	326	51832463

Visualization:



Insight:

The query classifies events based on their cost severity, considering normalized total costs. The majority of events fall into the "Low Cost" category, followed by "Medium Cost" and "High Cost." This will provide a clear comparison of the financial impact of events across different severity levels. Insurance companies and financial institutions can leverage the insights to assess the risk associated with different types of events. This information is instrumental in refining insurance policies and financial planning, enabling better preparation for and response to high-cost disasters.

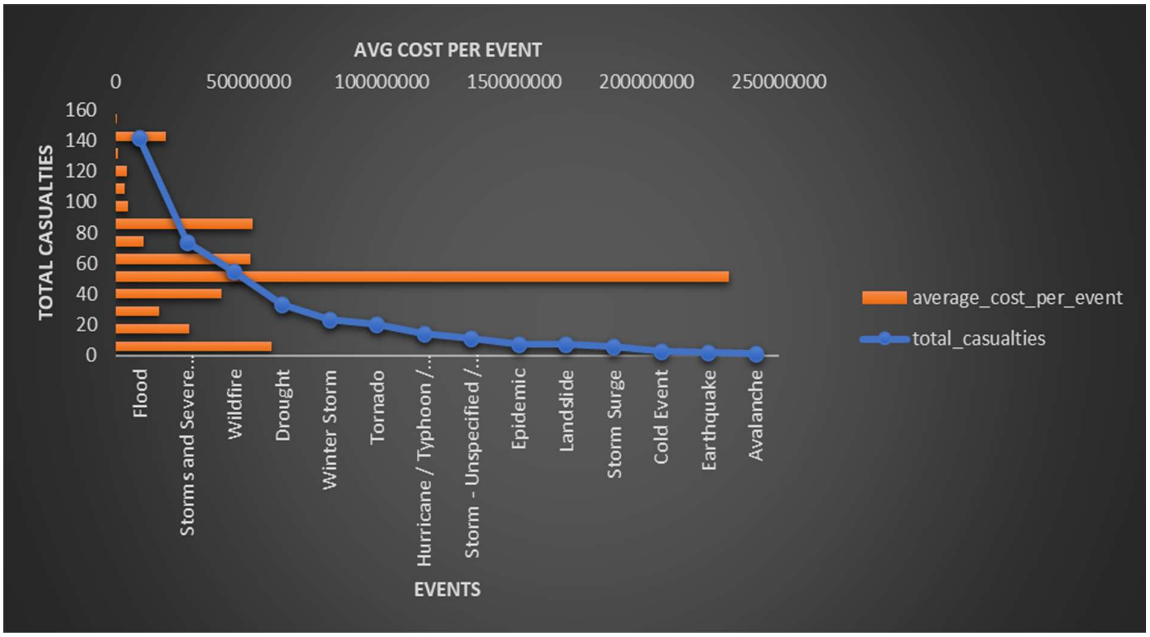
Query 10:

```
SELECT E.EVENT_TYPE,
       COUNT(C.CASUALITIES_ID) AS TOTAL_CASUALTIES,
       ROUND(AVG(C.ESTIMATED_TOTAL_COST)) AS AVERAGE_COST_PER_EVENT
FROM PUBLIC.EVENTS E
JOIN PUBLIC.CASUALITIES C ON E.EVENT_ID = C.EVENT_ID
WHERE E.EVENT_GROUP = 'Natural'
GROUP BY E.EVENT_TYPE,
       E.EVENT_GROUP
ORDER BY TOTAL_CASUALTIES DESC,
       AVERAGE_COST_PER_EVENT DESC;
```

Output:

event_type	total_casualties	average_cost_per_event
Flood	142	58649102
Storms and Severe Thunderstorms	74	27635244
Wildfire	55	16260440
Drought	33	39668119
Winter Storm	23	231235828
Tornado	20	50892023
Hurricane / Typhoon / Tropical Storm	14	10445505
Storm - Unspecified / Other	11	51554259
Epidemic	7	4471390
Landslide	7	3475034
Storm Surge	6	4145825
Cold Event	3	860828
Earthquake	2	18985333
Avalanche	1	368160
Pandemic	1	0

Visualization:



Insight:

The query surpassing average values within each category. It employs subqueries to calculate average fatalities and affected populations across all events, selecting those exceeding these averages. The custom bar and scatter chart provide insights into total casualties and average cost per event for each event type. Notably, Winter Storms stand out with a lower casualty count but a significantly higher average cost per event, indicating potential economic implications.

SUMMARY

This analysis unveils key trends in Canadian disasters, revealing meteorological and hydrological events as dominant (382 occurrences). Pandemics significantly impact 69% of individuals, urging targeted preparedness. Geographical insights guide regional disaster strategies, while high-fatality events necessitate broad responses. Most disasters incur low costs, enabling cost-effective mitigation. Visualizations, including tree maps and charts, empower decision-makers for tailored interventions and resource allocation. SQL analysis reinforces findings, emphasizing the urgency of preparing for intensified weather events. Mortality charts underscore ongoing risks from heatwaves and floods, urging resource prioritization. Economic severity visualizations highlight disproportionate losses from extreme occurrences. Geographical research identifies vulnerable provinces, aligning with pandemic impact figures. The approach provides a fact-based foundation for disaster management, optimizing resource allocation for resilient responses. Interactive infographics translate raw data into actionable insights, aiding stakeholders in prioritizing risks, vulnerabilities, and regions for enhanced emergency preparedness in Canada.

References:

Submitted To:
Prof. Jey Kumaresan

Reference:
The data is taken from Public safety Canada website.

Website:
[Canadian Disaster Database (publicsafety.gc.ca)]