Correlation of Incidents with Weather Patterns

Query #1: Average all Weather Stats (across 6hr windows) grouped by Unique Incidents

SELECT distinct incident_id, AVG(temp) as avg_temp, AVG(precip) as avg_precip, AVG(wind_spd) as avg_windspeed, AVG(vis) as avg_visblty, AVG(response_time_sec) as avg_response, AVG(speed) as avg_speed, AVG(congestion) as avg_congestion
CAST(AVG(window_of_day) as INT) as window_of_day, CAST(AVG(month) as INT) as month
FROM "incidents"."showdown_merge"
GROUP BY distinct incident_id;

Explanation: 'showdown_merge' is currently displaying multiple rows for each incident (since each traffic report is grouped with the incident). We use the 6hr windows we created with Spark to average and group by incident.

View Name: **weather_avg**

Query #2: Allocating Data to Temperature Range, Precipitation Range, and Visibility Range

SELECT *, CAST(FLOOR(avg_temp/10) as INT) as temp_range
CAST(FLOOR(avg_precip/.05) as INT) as precip_range,
CAST(FLOOR(avg_visblty)/2 as INT) as visibility_range
FROM "incidents"."weather_avg";

Explanation: We want to categorize each data into ranges/buckets, defined by the following tables:

Temperature Range			Precipitation Range (mm)			Visibility Range (km)				
-1	-10 °C to 0 °C		0	0 to 0.05		0	0 to 2	5	10 to 12	
0	0 °C to 10 °C		1	0.05 to 0.1		1	2 to 4	6	12 to 14	
1	10 °C to 20 °C		2	0.1 to 0.15		2	4 to 6	7	14 to 16	
2	20 °C to 30 °C		3	0.15 to 0.2		3	6 to 8	8	16 to 18	
3	30 °C +		4	0.25 to 0.3		4	8 to 10			
			5+	0.3+						

Query #3: Incident Frequency, Response Time, Speed, and Congestion by *Temperature* Range

SELECT COUNT(incident_id) as num_incident, temp_range, AVG(avg_response) as avg_response, AVG(avg_speed) as avg_speed, AVG(avg_congestion) as avg_congestion FROM "incidents"."weather_avg" WHERE temp_range IS NOT NULL GROUP BY temp_range ORDER BY temp_range;

Results stored in: **wthr_temp_all**

Query #4: Incident Frequency, Response Time, Speed, and Congestion by *Temperature* Range (Partitioned by Month)

SELECT COUNT(incident_id) as num_incident, temp_range, month, AVG(avg_response) as avg_response, AVG(avg_speed) as avg_speed, AVG(avg_congestion) as avg_congestion FROM "incidents"."weather_avg" WHERE temp_range IS NOT NULL GROUP BY month, temp_range ORDER BY temp_range, month;

Results stored in: **wthr_temp_month**

Correlation of Incidents with Weather Patterns

Query #5: Incident Frequency by *Precipitation* Range

```
SELECT COUNT(incident_id) as num_incident, precip_range FROM "incidents"."weather_avg"
WHERE precip_range IS NOT NULL GROUP BY precip_range ORDER BY precip_range;
```

Results stored in: **wthr_precip_incidents**

Query #6: Incident Frequency, Response Time, Speed, and Congestion by *Visibility* Range

```
SELECT COUNT(incident_id) as num_incident, visibility_range, AVG(avg_response) as avg_response,
AVG(avg_speed) as avg_speed, AVG(avg_congestion) as avg_congestion FROM "incidents"."weather_avg"
WHERE visibility_range IS NOT NULL GROUP BY visibility_range ORDER BY visibility_range;
```

Results stored in: **wthr_vis_all**