

Progetto SQL

Lorenzo Ravasio

Progetto SQL

Il progetto è volto ad analizzare quanto sia concreto il rischio di viaggiare in alcuni paesi, quali siano i paesi con maggior rischio e quali con il minor rischio.

Il concetto di rischio, personalmente, lo trovo molto complesso da definire e servirebbero dati aggiuntivi che non sono presenti in queste tabelle per stabilirlo con meno imprecisione. In questa analisi stabilirò 2 strade differenti per definire l'indicatore del rischio di una nazione: il primo è rappresentato dal rapporto tra il numero di morti e il numero di viaggiatori, con i dati a disposizione; il secondo è rappresentato dal rapporto tra numero totale di morti di una nazione e il numero di segnalazioni ricevute per quella nazione.

Progetto SQL – 1^ soluzione

Il dato che può dare indicazioni sul rapporto morti/viaggiatori possiamo trovarlo nella tabella `deathscapita_1`. Il campo rappresenta effettivamente una percentuale ma va verificato che sia il rapporto che cerchiamo. Per assicurarci usiamo questa select:

```
select ROUND(((ndeaths/ntravelers)*100000),2) as rate, percap
from deathscapita_1
order by percap desc
```

	#	rate	▼	#	percap	▼
1		0,84			0,84	
2		2,28			2,28	
3		1,96			1,96	
4		1,66			1,66	
5		0,25			0,25	
6		0,37			0,37	

Progetto SQL – 1^ soluzione

Una volta accertati che percap sia il dato che ci serve, è sufficiente ordinare i dati della stessa tabella per verificare quali siano i paesi con maggior rischio e quali siano con quello più basso. Per quelli con il rischio più basso, essendo in tanti che hanno il tasso nullo, limitiamo i risultati ai 3 paesi con maggior numero di visitatori.

```
select * from deathspcapita_1
order by percap DESC
limit 3;
```



#	column_a	validctrys	# ntravelers	# ndeaths	# percap	region
1	62	Pakistan	226200	8	3,54	Asia
2	84	Thailand	343500	11	3,2	Asia
3	66	Philippines	3240000	74	2,28	Asia

validctrys	# ntravelers	# ndeaths	# percap	region
Canada	83817200	2	0,0	Americas
United Kingdom	57264100	1	0,0	Europe
South Korea	15234000	0	0,0	Asia



```
select * from deathspcapita_1
where percap = '0.0'
order by ntravelers DESC
LIMIT 3;
```

Progetto SQL – 2^ soluzione

Per individuare le nazioni a maggior rischio per i viaggiatori statunitensi si può ipotizzare di tradurre il concetto di pericolosità come il rapporto fra morti e numero di avvertimenti totali per uno specifico paese. Individuiamo così, inoltre, il peso specifico che ha ogni warning. Con il dato che nella seguente query chiamerò rate, è possibile quindi anche capire quanto sia stato giustificato ogni warning.

```
select d.country, sdwarnings_df_region, freq as n_deaths, sdwarnings_df_nwarnings as n_warnings, ROUND((freq/sdwarnings_df_nwarnings),1) as rate
from deathsraking_csv d
join warningsranking w
on d.country = w.sdwarnings_df_country
where sdwarnings_df_nwarnings > 0
order by (freq/sdwarnings_df_nwarnings) desc
```

	country	sdwarnings_df_region	# n_deaths	# n_warnings	# rate
1	Mexico	Americas	598	28	21,4
2	Haiti	Americas	65	10	6,5
3	Honduras	Americas	46	9	5,1
4	El Salvador	Americas	34	9	3,8
5	Philippines	Asia	74	20	3,7
6	Afghanistan	Asia	84	23	3,7

Progetto SQL – 2^a soluzione

	country	sdwarnings_df_region	# n_deaths	# n_warnings	# rate
1	Mali	Africa	1	26	0,0
2	Libya	Africa	1	19	0,1
3	Congo	Africa	1	18	0,1
4	Ukraine	Europe	1	15	0,1

Ordinando in ASC la query sopra indicata troveremo i paesi che hanno minor rischio, considerando questa soluzione. Questo procedimento descritto nella slide precedente esclude, dunque, tutti i paesi che possono aver avuto anche alcuni morti ma nessuna segnalazione dagli enti preposti (condizione posta con la where). Si sottintende, quindi, che una nazione che non abbia warning a proprio carico non possa candidarsi a nazione pericolosa, non conoscendo il motivo per cui gli enti segnalatori effettuino questi warning.

Progetto SQL

E' possibile indagare, inoltre, i paesi che hanno ricevuto maggiori segnalazioni (prenderemo in considerazione i paesi che hanno ricevuto più di 10 segnalazioni) ma non hanno avuto alcuna morte. Da questa informazione il Dipartimento di Stato USA può indagare sulle cause che hanno motivato la segnalazione in modo tale da rivedere i criteri, per dare avvertimenti, in futuro, di maggiore efficacia.

```
1  SELECT * from warningsranking
2  where sdwarnings_df_country not in (select distinct country from deathsranking_csv)
3  and sdwarnings_df_nwarnings > 10
```



#	column_a	sdwarnings_df_country	sdwarnings_df_region	#	sdwarnings_df_nwarnings
1	3	Algeria	Africa		22
2	100	Iran	Asia		22
3	36	Burundi	Africa		21
4	42	Central African Republic	Africa		21
5	43	Chad	Africa		19
6	203	Sudan	Africa		19

Progetto SQL

Possiamo ora prendere in considerazione la tabella che analizza le morti dall'ottobre 2009 a giugno 2016. La tabella si chiama `sdamerican_deaths_abroad_10_09_to_06_16` e ci permetterà di effettuare analisi sulle cause di morte dei vari paesi.

Innanzitutto con la seguente query possiamo individuare, per ogni nazione, quale sia la causa di morte più Frequente (nel caso ce ne fossero due o più ex aequo vengono riportate tutte).

```
WITH cte_cause_morte AS (  
    SELECT country, cause_of_death, COUNT(1) AS frequenza  
    FROM sdamerican_deaths_abroad_10_09_to_06_16  
    GROUP BY country, cause_of_death  
)  
SELECT country, cause_of_death, frequenza  
FROM (  
    SELECT country, cause_of_death, frequenza, RANK() OVER (PARTITION BY country ORDER BY frequenza DESC) AS ranking  
    FROM cte_cause_morte  
) ranked_data  
WHERE ranking = 1  
order by country desc
```



203 query results (0.55 seconds) View log					
	country	cause_of_death	#	frequenza	
1	Zimbabwe	Veh. Accid-Auto		2	
2	Zambia	Air Accident		2	
3	Yemen	Homicide		9	
4	Vietnam	Veh. Accid-Motorcy.		14	

Progetto SQL

E' possibile verificare, inoltre, quali siano le cause più frequenti, non frutto di incidenti, divise per continente. Da questa select è possibile trarre informazioni interessanti, che hanno la possibilità di essere approfondite nel caso avremmo a disposizione maggiori dati.

```
with cte_cause_morte as (  
  select country, cause_of_death, count(1) as frequenza  
  from sdamerican_deaths_abroad_10_09_to_06_16  
  group by country, cause_of_death)  
select cr.region, cause_of_death, SUM(frequenza) as somma  
from cte_cause_morte c  
join countriesregions cr  
on c.country = cr.name  
where cause_of_death not like '%ccid%'  
group by region, cause_of_death  
order by region, somma desc
```

5	Africa	Drug-Related	4
6	Americas	Homicide	997
7	Americas	Drowning	536

Si evidenzia subito come la causa di morte più frequente nelle Americhe sia l'omicidio o l'annegamento. Questo potrebbe, ad esempio, far pensare ad una forte concentrazione di organizzazioni criminali.

Progetto SQL

Altro output dalla select precedente, che può essere interessante approfondire, riguarda la maggior causa di morte per gli altri continenti, eccetto l'Africa:

3	Asia	Suicide	258
4	Asia	Homicide	163

26	Oceania	Suicide	34
----	---------	---------	----

20	Europe	Suicide	239
21	Europe	Drowning	85

Sorprendentemente la causa di morte più frequente per i 3 continenti indicati per americani in viaggio è il suicidio. Aver a disposizione questa informazione può essere rilevante poiché, se ben approfondita trovando , grazie ad essa si potrebbero pensare programmi di prevenzione.

Progetto SQL

Di seguito il link alla cartella drive contenente gli script:

https://drive.google.com/file/d/1XYYmjA14J5tFRVb0wpnpJRPFek__L_Fc/view?usp=sharing