Homework | module 2 > week 6 > day 16

Topics covered: *GROUP BY, ORDER BY, HAVING*

Standard Exercise:

Take the "2.6.16d Homework Data life-expectancy.csv" csv file and load it into BigQuery as a table named life-expectancy inside the same project and dataset you created in class; familiarise yourself with the data, then try to answer the following questions:

1. What is the average life expectancy *at birth* in Europe (*hint*: *the numeric variable related to the MetricObserved dimension is called Numeric*)?

```
SELECT RegionDisplay,
Cast(avg(Numeric)as INT) as n_avg_europe
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 Group by RegionDisplay
 HAVING RegionDisplay = 'Europe'
```

2. Is Europe the region with the highest life expectancy at birth? Which region has the lowest life expectancy at birth?

```
SELECT RegionDisplay,
Cast(avg(Numeric)as INT) as n_avg_birth
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 where MetricObserved = 'Life expectancy at birth (years)'
 Group by RegionDisplay
 ORDER BY  n_avg_birth DESC
```

3. Does Europe also have the highest life expectancy at age 60? Which country has the highest life expectancy after 60?

```
SELECT RegionDisplay,
 Count(Numeric) as N_max60,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY RegionDisplay
 HAVING Count(Numeric) >= 60
 Order by N_max60 DESC
 Limit 1
```

```
SELECT RegionDisplay,
 Count(Numeric) as N_max60,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY RegionDisplay
 HAVING Count(Numeric) >= 60
 Order by N_max60 DESC
 Limit 3 OFFSET 1
```

4. Using a GROUP BY and a CASE-WEHEN, create a pivot table that shows the average life expectancy for all three types of MetricObserved (in three separate columns) by each region (each in a separate row); the output should look something like this (LEAB = Life Expectancy At Birth):

Row	RegionDisplay	LEAB	HLEAB	LE60
1	Eastern Mediterranean	68.203158429118787	59.183823529411768	17.993618314176246
2	Africa	56.271860612612642	49.140338164251183	16.078459531531532
3	Europe	75.221733301886815	66.6505241090146	20.445165047169805
4	Americas	73.061249833333235	63.853968253968254	20.665792999999997
5	Western Pacific	71.0362527469136	62.959259259259262	18.793803518518541
6	South_East Asia	66.341902500000046	59.4181818181818	17.331689999999995

```
SELECT RegionDisplay,
avg(case when metricObserved = 'Life expectancy at birth (years)' then
numeric end) as LEAB,
avg(case when metricObserved = 'Healthy life expectancy (HALE) at birth
(years)' then numeric end) as HLEAB,
avg(case when metricObserved = 'Life expectancy at age 60 (years)' then
numeric end) as LE60,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 group by RegionDisplay;
```

5. Which countries have the highest average life expectancy at birth?

```
SELECT MetricObserved ,
CountryDisplay,
Cast(avg(Numeric)as INT) as n_avg_birth
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 Group by MetricObserved , CountryDisplay
 HAVING MetricObserved = 'Life expectancy at birth (years)'
 ORDER BY n_avg_birth DESC
 Limit 3
```

6. Using three separate queries, check out which are the top 10 countries that consume the highest quantities of beer, wine and spirits respectively.

```
SELECT CountryDisplay,
SUM(beer_servings) as q_of_beer,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY CountryDisplay
 ORDER BY SUM(beer_servings)DESC
 LIMIT 10;

 SELECT CountryDisplay,
SUM(spirit_servings) as q_of_spirit,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY CountryDisplay
 ORDER BY SUM(spirit_servings)DESC
 LIMIT 10;

  SELECT CountryDisplay,
SUM(wine_servings) as q_of_wine,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY CountryDisplay
 ORDER BY SUM(wine_servings)DESC
 LIMIT 10;
```

7. Now create a new variable that sums the average servings of beer + wine + spirit and call it “avg_alcohol_servings”; what are the top 10 countries that consume the highest quantities of alcohol overall?

```
SELECT CountryDisplay,
  avg(beer_servings)+ avg(spirit_servings) + avg(wine_servings)  as
avg_of_alcohol,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY CountryDisplay
 ORDER BY avg_of_alcohol DESC
```

8. With reference to the last query (last question), look at the country that consumed the highest quantity of alcohol overall, was it the first ranking country in terms of beer, wine or spirits? If not, was it in the top 10 of any of those 3 rankings (beer, wine or spirits)?

```
SELECT CountryDisplay,
  avg(beer_servings)+ avg(spirit_servings) + avg(wine_servings)  as
avg_of_alcohol,
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 GROUP BY CountryDisplay
 ORDER BY avg_of_alcohol DESC
Limit 10;
```

9. Which Country in Europe has the lowest life expectancy at birth?

```
SELECT
       CountryDisplay,
Cast(avg(Numeric)as INT) as europe_country_avg_birth
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 where MetricObserved = 'Life expectancy at birth (years)' and RegionDisplay =
'Europe'
 Group by CountryDisplay
 ORDER BY europe_country_avg_birth
```

10. Generally, women live longer than men, create a new variable that shows the difference between females and males life expectancy at birth and call it avg_LE_delta_gender; which country has the highest gap (in terms of years) between females and males? Are there any countries where men live more than women?

```
SELECT
       CountryDisplay,
Cast(avg(Numeric)as INT) as europe_country_avg_birth,
avg(case when sexdisplay = 'Female' then Numeric end ) - avg(case when sexdisplay=
'Male' then numeric end) as avg_le_gender_gap
 FROM `adept-bond-365418.DataAnalytics.life-expectancy`
 where MetricObserved = 'Life expectancy at birth (years)' and RegionDisplay =
'Europe'
 Group by CountryDisplay
 ORDER BY avg_le_gender_gap
```