UN Happiness Index Data Analysis in SQL

The World Happiness Report, published by the UN Sustainable Development Solutions Network, surveys countries across several quality-of-life characteristics – such as economy, family, health, freedom, generosity, etc. – to determine a global happiness ranking. The happiness score is used to rank countries and it is calculated by taking the sum of all quality-of-life characteristic scores

Following are the tasks to be performed:

Day 1 - You have been tasked with consolidating and cleaning the 5 available datasets into a SQL view named `**vw_world_happiness_index_consolidated**`.

1. Add a year column.
2. Some years include extra quality-of-life characteristics. Replace null numeric values with zero.
3. Not all datasets include a Region field. Fill out the null region values with the correct region. Which countries do not have a corresponding region?
4. Include all countries regardless of having a corresponding region.

Not all datasets include a Region field. Years 2015 and 2016 have region values in the tables. These are used to create a lookup table, to fill the regions for other years.

```sql
create table cr(con VARCHAR(255),
                reg VARCHAR(255));

INSERT INTO cr
select distinct country as con, region as reg
from world_happiness_index_2015
UNION
select distinct country as con , region as reg
from world_happiness_index_2016;
```

Output:

con	reg
Afghanistan	Southern Asia
Albania	Central and Eastern Europe
Algeria	Middle East and Northern Africa
Angola	Sub-Saharan Africa
Argentina	Latin America and Caribbean
Armenia	Central and Eastern Europe
Australia	Australia and New Zealand
Austria	Western Europe

The columns in the tables are a little different. Hence they are aligned before commiting union.
Example:


```sql
CREATE TABLE aligned_2015 (year_col INT,
                           country varchar(255),
                           region varchar(255),
                           happiness_rank int,
                           happiness_score float,
                           economy_gdp_per_capita float,
                           family float,
                           health_life_expectancy float,
                           freedom float,
                           trust_government_corruption float,
                           generosity float,
                           dystopia_residual float,
                           perceptions_of_corruption FLOAT);

insert into aligned_2015
SELECT 2015 as year_col, country, region, happiness_rank, happiness_score,
economy_gdp_per_capita, family, health_life_expectancy, freedom, trust_government_corruption,
generosity, dystopia_residual, NULL as perceptions_of_corruption
from world_happiness_index_2015;
```

Output:

After aligning all the tables and adding the year column, we can get their union.

```sql
CREATE TABLE vw_world_happiness_index_consolidated (year_col INT,
                        country varchar(255),
                        region varchar(255),
                        happiness_rank int,
                        happiness_score float,
                        economy_gdp_per_capita float,
                        family float,
                        health_life_expectancy float,
                        freedom float,
                        trust_government_corruption float,
                        generosity float,
                        dystopia_residual float,
                        perceptions_of_corruption FLOAT);

insert into vw_world_happiness_index_consolidated
select * from aligned_2015
UNION
select * from aligned_2016
UNION
select * from aligned_2017
UNION
select * from aligned_2018
UNION
select * from aligned_2019;
```

Replacing null values in the extra columns with zero.

```
UPDATE vw_world_happiness_index_consolidated SET perceptions_of_corruption = 0.0 WHERE perceptions_of_corruption ISNULL;
UPDATE vw_world_happiness_index_consolidated SET dystopia_residual = 0.0 WHERE dystopia_residual ISNULL;
UPDATE vw_world_happiness_index_consolidated SET trust_government_corruption = 0.0 WHERE trust_government_corruption ISNULL;
```

Output:

yea...	country	region	happi...	happi...	econ...	family	healt...	freed...	trust_...	generosity	dystopia_...	perce
2015	Afghan…	Southe…	153	3.575	0.31982	0.30285	0.30335	0.23414	0.09719	0.3651	1.9521	0
2015	Albania	Central…	95	4.959	0.87867	0.80434	0.81325	0.35733	0.06413	0.14272	1.89894	0
2015	Algeria	Middle …	68	5.605	0.93929	1.07772	0.61766	0.28579	0.17383	0.07822	2.43209	0
2015	Angola	Sub-S…	137	4.033	0.75778	0.8604	0.16683	0.10384	0.07122	0.12344	1.94939	0
2015	Argentina	Latin A…	30	6.574	1.05351	1.24823	0.78723	0.44974	0.08484	0.11451	2.836	0

The countries that do not have a corresponding region are:

Hong Kong S.A.R., China
Taiwan Province of China
Northern Cyprus
Trinidad & Tobago
Gambia
North Macedonia
Northern Cyprus
Trinidad & Tobago

```
select country from vw_world_happiness_index_consolidated where region ISNULL;
```

Day 2 – In order to prepare this data for dashboard use, the quality of life characteristics need to be unpivoted from columns to rows. Create another SQL view named `vw_world_happiness_index` with the following columns: year, country, region, happiness_ranking, happiness_score, characteristic, value.

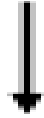
We can directly use "unpivot," but I am providing an alternative solution:

```
CREATE TABLE vw_world_happiness_index (year_col INT,
                        country varchar(255),
                        region varchar(255),
                        happiness_rank int,
                        happiness_score float,
                        characteristic varchar(255),
                        value float);

insert into vw_world_happiness_index
select year_col, country, region, happiness_rank, happiness_score,
    'economy_gdp_per_capita' as characteristic, economy_gdp_per_capita as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'family' as characteristic, family as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'health_life_expectancy' as characteristic, health_life_expectancy as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'freedom' as characteristic, freedom as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'trust_government_corruption' as characteristic, trust_government_corruption as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'generosity' as characteristic, generosity as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'dystopia_residual' as characteristic, dystopia_residual as value
        from vw_world_happiness_index_consolidated
union ALL
select year_col, country, region, happiness_rank, happiness_score,
    'perceptions_of_corruption' as characteristic, perceptions_of_corruption as value
        from vw_world_happiness_index_consolidated;
```

Output:

yea...	country	region	happi...	happi...	econ...	family	healt...	freed...	trust_...	generosity	dystopia_...	perce
2015	Afghan...	Southe...	153	3.575	0.31982	0.30285	0.30335	0.23414	0.09719	0.3651	1.9521	0

year_col	country	region	happiness_rank	happiness_score	characteristic	value
2015	Afghanistan	Southern Asia	153	3.575	economy_gdp_p...	0.31982
2015	Afghanistan	Southern Asia	153	3.575	family	0.30285
2015	Afghanistan	Southern Asia	153	3.575	health_life_expec...	0.30335
2015	Afghanistan	Southern Asia	153	3.575	freedom	0.23414
2015	Afghanistan	Southern Asia	153	3.575	trust_government...	0.09719
2015	Afghanistan	Southern Asia	153	3.575	generosity	0.3651
2015	Afghanistan	Southern Asia	153	3.575	dystopia_residual	1.9521
2015	Afghanistan	Southern Asia	153	3.575	perceptions_of_c...	0

Day 3 – You have been asked to create a report to identify which countries have ranked in the top 10 and in the bottom 10 in terms of happiness. Generate a dataset that lists the top 10 and bottom 10 ranked happiest countries of all time.

We sum the happiness score of every country over all the five years for this analysis.

```sql
create view overall_sum as
  select country, SUM(happiness_score) as overall_score
  from vw_world_happiness_index_consolidated
  GROUP by country;

create view overall_ranking as
  select country, RANK() OVER (order by overall_score DESC) as overall_rank, overall_score
    from sum_overall;

CREATE TABLE happiest_overall (
                              ranking varchar(255),
                              country varchar(255),
                              overall_rank int,
                              overall_score float);


insert into happiest_overall
select "TOP 10" as ranking, country, overall_rank, overall_score from
  (select * from overall_ranking ORDER by overall_rank asc limit 10)
UNION
select "BOTTOM 10" as ranking, country, overall_rank, overall_score from
  (select * from overall_ranking ORDER by overall_rank DESC limit 10)
ORDER by overall_rank;

select * from happiest_overall;
```

Output:

ranking	country	overall_rank	overall_score
TOP 10	Denmark	1	37.72999983596802
TOP 10	Norway	2	37.70500017929077
TOP 10	Finland	3	37.6889998626709
TOP 10	Switzerland	4	37.55699995803833
TOP 10	Iceland	5	37.55500018692017
TOP 10	Netherlands	6	37.022999855041505
TOP 10	Canada	7	36.75299998474121
TOP 10	Sweden	8	36.59599991989136
TOP 10	New Zealand	9	36.56500012969971
TOP 10	Australia	10	36.38099991989136
BOTTOM 10	Swaziland	161	9.079
BOTTOM 10	Puerto Rico	162	7.039
BOTTOM 10	Oman	163	6.853
BOTTOM 10	Taiwan Province of China	164	6.42199993133545
BOTTOM 10	Hong Kong S.A.R., China	165	5.47200012207031
BOTTOM 10	North Macedonia	166	5.274

Day 4 - You have been asked to create a report to analyze drastic shifts in a country's happiness ranking year over year.

1. Generate a dataset that lists the top 10 year-to-year biggest jumps in a country's ranking across the entire dataset.
2. Generate a dataset that lists the top 10 year-to-year biggest drops in a country's ranking across the entire dataset.

We first compute the year by year difference in the ranks of each country and unpivot the data for analysis purposes.

```
create view yearBY_differences as
  select a5.country,
      a6.happiness_rank - a5.happiness_rank  as diff56,
      a7.happiness_rank - a6.happiness_rank  as diff67,
      a8.happiness_rank - a7.happiness_rank  as diff78,
      a9.happiness_rank - a8.happiness_rank  as diff89
  from aligned_2015 as a5, aligned_2016 as a6, aligned_2017 as a7, aligned_2018 as a8, aligned_2019 as a9
  where a5.country = a6.country and a6.country = a7.country and a7.country = a8.country and a8.country = a9.country

create view yearBY_differences_unpivot as
  select country, diff56 as jump from yearBY_differences
  UNION
  select country, diff67 as jump from yearBY_differences
  UNION
  select country, diff78 as jump from yearBY_differences
  UNION
  select country, diff89 as jump from yearBY_differences
```

Output:

country	jump
Afghanistan	-13
Afghanistan	1
Afghanistan	4
Afghanistan	9
Albania	-5
Albania	0
Albania	3
Albania	14
Algeria	-30
Algeria	4

We compute the biggest jump of each country and find the top 10 jumps from the individual biggest jumps. (Max of max)

```
CREATE TABLE top_jumps (country varchar(255),
                        biggest_jump int);

insert into top_jumps
  select country, biggest_jump from (
    select country, max(jump) as biggest_jump from yearBY_differences_unpivot
    group by country)
  where biggest_jump > 0
  order by biggest_jump DESC
  limit 10;
```

Similarly for 10 biggest drops. We compute the biggest drop of each country and get the 10 biggest drops. (Min of min)

```sql
CREATE TABLE top_drops (country varchar(255),
                        biggest_drops int);

insert into top_drops
  select country, biggest_drop from (
    select country, min(jump) as biggest_drop from yearBY_differences_unpivot
    group by country)
  where biggest_drop < 0
  order by biggest_drop ASC
  limit 10;
```

Outputs:

country	biggest_jump
Malaysia	45
Venezuela	38
Liberia	34
Algeria	31
Nigeria	25
Vietnam	21
Zambia	21
Turkmenistan	19
Argentina	18
Egypt	18

country	biggest_drops
Benin	-34
Algeria	-30
Bulgaria	-24
Ghana	-23
Guinea	-22
Ivory Coast	-21
Latvia	-21
Jamaica	-20
Kosovo	-20
Niger	-20

Day 5 - You have been asked to create a report that lists each country's best rank across all time and which quality-of-life characteristic that contributed the most to its score that year.

We first compute the best rank for every country along with the year for which the best rank was obtained. We then compute the characteristic which had the maximum value that year for that country (using the unpivoted table makes this computation easier)

```sql
create view country_bestRank as
  select country, year_col, MIN(happiness_rank) as best_rank from vw_world_happiness_index_consolidated
  GROUP bY country;

CREATE TABLE countries_individual_best (
                          country varchar(255),
                          best_rank int,
                          year_of_bestRank INT,
                          best_characteristic varchar(255),
                          value_of_bestChar float);

Insert into countries_individual_best
select cbr.country, best_rank, cbr.year_col as year_of_bestRank, characteristic as best_characteristic,
       MAX(value) as value_of_bestChar
from country_bestRank as cbr, vw_world_happiness_index as vw
where cbr.country = vw.country and cbr.year_col = vw.year_col
group by vw.country;
```

Output:

country	best_rank	year_of_bestRank	best_characteristic	value_of_bestChar
Afghanistan	141	2017	dystopia_residual	2.15080118179321
Albania	95	2015	dystopia_residual	1.89894
Algeria	38	2016	dystopia_residual	3.40904
Angola	137	2015	dystopia_residual	1.94939
Argentina	24	2017	dystopia_residual	2.61400532722473
Armenia	116	2019	family	1.055