

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

-- We downloaded data from different APIs. Because it took dozens of hours, there are xx files for each API that need to be joined using UNION.

-- API PLAYER

-- we know that rows need to be unique, so we use UNION to get rid of unwanted duplicates created when downloading data.

CREATE OR REPLACE TEMPORARY table TEMP\_APIPLAYERBYIDCOMPLETETABLE AS

-- in the source data there were players with different records that differed only in the WorldRanking and CountryRanking columns.

-- We had to remove such duplicates and we chose smaller values for the players (higher/better in the rankings)

```
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", MIN("WorldRanking") as "WorldRanking", MIN("CountryRanking") as  
"CountryRanking", "TotalUSDPrize", "TotalTournaments", "playerid"  
FROM
```

```
(  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"  
  FROM "01PlayerById_vystup001000-1999" -- 01  
  UNION  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"  
  FROM "01PlayerById_vystup002000-5999" -- 02  
  UNION  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"  
  FROM "01PlayerById_vystup006000-6999" -- 03  
  UNION  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"  
  FROM "01PlayerById_vystup007000-7999" -- 04  
  UNION  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"  
  FROM "01PlayerById_vystup008000-9098" -- 05  
  UNION  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"  
  FROM "01PlayerById_vystup009099-35689" -- 06  
  UNION  
  SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",  
    "TotalTournaments", "playerid"
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
FROM "01PlayerById_vystup035690-38532" -- 07
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup038533-68517_novecsv" -- 08
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup068518-73512" -- 09
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup073513-101790" -- 10
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup101791-104788" -- 11
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup104789-115464" -- 12
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup115465-120432" -- 13
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup120433-124001" -- 14
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "PlayerId"
FROM "01PlayerById_vystup124002-12531_novecsv" -- 15
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
FROM "01PlayerById_vystup125432-149985" -- 16
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "playerid"
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
FROM "01PlayerById_vystup_errorid" -- 17
UNION
SELECT "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "WorldRanking", "CountryRanking", "TotalUSDPrize",
      "TotalTournaments", "PlayerId"
FROM "01PlayerById_vystup_errorids2" -- 18
)

GROUP BY "NameFirst", "NameLast", "CurrentHandle", "CountryCode", "TotalUSDPrize", "TotalTournaments", "playerid";

-- basic data cleansing
UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "NameFirst" = TRIM("NameFirst", '-');

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "NameFirst" = TRIM("NameFirst", '');

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "NameFirst" = null
WHERE "NameFirst" IN ('-', '', '?', '.') OR lower("NameFirst") = 'null';

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "NameLast" = TRIM("NameLast", '-');

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "NameLast" = TRIM("NameLast", '');

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "NameLast" = null
WHERE "NameLast" IN ('-', '', '?', '.') OR lower("NameLast") = 'null';

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "CurrentHandle" = null
WHERE "CurrentHandle" IN ('null', '', '-');

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "CountryCode" = null
WHERE "CountryCode" IN ('null', '', '-');

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "WorldRanking" = null
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
WHERE "WorldRanking" IN ('', '-') OR lower("WorldRanking") = 'null';

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "CountryRanking" = null
WHERE "CountryRanking" IN ('', '-') OR lower("CountryRanking") = 'null';

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "TotalUSDPrize" = null
WHERE "TotalUSDPrize" IN ('', '-') OR lower("TotalUSDPrize") = 'null';

UPDATE TEMP_APIPLAYERBYIDCOMPLETETABLE
SET "TotalTournaments" = null
WHERE "TotalTournaments" IN ('', '-') OR lower("TotalTournaments") = 'null';

-- duplicate record, must be deleted
DELETE FROM TEMP_APIPLAYERBYIDCOMPLETETABLE
WHERE "WorldRanking" = '2443';

DELETE FROM TEMP_APIPLAYERBYIDCOMPLETETABLE
WHERE "CurrentHandle" ILIKE '%please delete%';

-- create final table without null rows
CREATE OR REPLACE table APIPLAYERBYIDCOMPLETETABLE AS
    SELECT *
    FROM TEMP_APIPLAYERBYIDCOMPLETETABLE
EXCEPT
    SELECT *
    FROM TEMP_APIPLAYERBYIDCOMPLETETABLE
    WHERE "NameFirst" is null
        AND "NameLast" is null
        AND "CurrentHandle" is null
        AND "CountryCode" is null
        AND "WorldRanking" is null
        AND "CountryRanking" is null
        AND "TotalUSDPrize" is null
        AND "TotalTournaments" is null;
```

-----

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

-- API GAME

```
CREATE OR REPLACE TABLE GAMEBYID_COMPLETE_DATASET AS
SELECT * FROM (
    SELECT "GameName", "TotalUSDPrize", "TotalTournaments", "TotalPlayers", "GameId" as "gameid"
    FROM "04GameById_vystup_errorids"
    UNION
    SELECT "GameName", "TotalUSDPrize", "TotalTournaments", "TotalPlayers", "gameid"
    FROM "04LookupGameById_vystup0-1000")
WHERE NOT
    ("GameName" = ''
    AND "TotalUSDPrize" = ''
    AND "TotalTournaments" = ''
    AND "TotalPlayers" = '' );
```

-----

-- TOURNAMENT

```
CREATE OR REPLACE TEMPORARY TABLE TEMP_TOURNAMENTBYID_COMPLETEDATASET AS
SELECT * FROM
    (SELECT "GameId", "TournamentName", date("StartDate") as "StartDate", date("EndDate") as "EndDate", "Location", "Teampplay",
    "TotalUSDPrize", "tournamentid"
    FROM "07TournamentById_vystup1000-1999"
    UNION
    SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
    FROM "07TournamentById_vystup10000-10975"
    UNION
    SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
    FROM "07TournamentById_vystup10976-11262"
    UNION
    SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
    FROM "07TournamentById_vystup11263-11844"
    UNION
    SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
    FROM "07TournamentById_vystup11845-12262"
    UNION
    SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
    FROM "07TournamentById_vystup12263-14358_novecsv"
    UNION
    SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
FROM "07TournamentById_vystup14359-24358_novecsv"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup2000-2999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup24359-52708_novecsv"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup3000-3999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup4000-4999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup5000-5999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup52808-82806_novecsv"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup6000-6999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup7000-7999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup8000-8999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentById_vystup9000-9999"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamnebtById_errorids"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "TournamentById_vystup10975"
UNION
SELECT "GameId", "TournamentName", date("StartDate"), date("EndDate"), "Location", "Teampplay", "TotalUSDPrize", "tournamentid"
FROM "07TournamentsById_vystup_errorids2");
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

-- these tournaments were in the "tournament name" called as "please delete, duplicate" etc., but some tournaments with these words was not really duplicates (just funny names for tournaments).

-- So we decided safe variant with deleting selected tournaments Ids.

```
DELETE FROM TEMP_TOURNAMENTBYID_COMPLETEDDATASET
```

```
WHERE "tournamentid" = 24694
      OR "tournamentid" = 39324
      OR "tournamentid" = 52657
      OR "tournamentid" = 52742
      OR "tournamentid" = 52717
      OR "tournamentid" = 52713
      OR "tournamentid" = 52711
      OR "tournamentid" = 52547
      OR "tournamentid" = 52548
      OR "tournamentid" = 52658
      OR "tournamentid" = 52683
      OR "tournamentid" = 52700
      OR "tournamentid" = 44995
      OR "tournamentid" = 28872
      OR "tournamentid" = 28868
      OR "tournamentid" = 28876
      OR "tournamentid" = 42322
      OR "tournamentid" = 36924
      OR "tournamentid" = 33885
      OR "tournamentid" = 42491;
```

-- creating a final table without null rows

```
CREATE OR REPLACE TABLE API07TOURNAMENTBYID_COMPLETEDDATASET_ORIGINAL AS
```

```
SELECT * FROM TEMP_TOURNAMENTBYID_COMPLETEDDATASET
```

```
EXCEPT
```

```
SELECT * FROM TEMP_TOURNAMENTBYID_COMPLETEDDATASET
```

```
WHERE "GameId" is null
      AND "TournamentName" is null
      AND "StartDate" is null
      AND "EndDate" is null
      AND "Location" is null
      AND "Teampplay" is null
      AND "TotalUSDPrize" is null;
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
-- correction of incorrect date
UPDATE API07TOURNAMENTBYID_COMPLETEDDATASET_ORIGINAL
SET "StartDate" = '2020-05-07'
WHERE "StartDate" = '0202-05-07';

-----

-- API TOURNAMENTS RESULTS FOR INDIVIDUAL PLAYERS

CREATE OR REPLACE TABLE TEMP_API08_COMPLETEDDATASET AS
SELECT * FROM
(
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultByTournamentId_teamplay0newids" --1
  UNION
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultsByTournamentId_1000-1999" --2
  UNION
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultsByTournamentId_10k-14k" --3
  UNION
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultsByTournamentId_14k-24k" --4
  UNION
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultsByTournamentId_2000-2999" --5
  UNION
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultsByTournamentId_24k-45k" --6
  UNION
  SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
    "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
  FROM "08TournamentResultsByTournamentId_3000-3999" --7
  UNION

```



## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_4000-4999" --8
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_45k-52k" --9
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_5000-5999" --10
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_52k-64k" --11
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_6000-6999" --12
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_7000-7999" --13
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_8000-8999" --14
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_9000-9999" --15
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "CountryCode", "PlayerId", "NameFirst", "NameLast",
       "CurrentHandle", "ShowLastNameFirst", "PrizeUSD"
FROM "08TournamentResultsByTournamentId_ErrorIds1" --16
)
```

;

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
-- basic data cleansing
UPDATE TEMP_API08_COMPLETEDDATASET
SET "Ranking" = null
WHERE "Ranking" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "RankText" = null
WHERE "RankText" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "TeamId" = null
WHERE "TeamId" = 'NULL' OR "TeamId" = 0 ;

UPDATE TEMP_API08_COMPLETEDDATASET
SET "TeamName" = null
WHERE "TeamName" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "CountryCode" = null
WHERE "CountryCode" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "ShowLastNameFirst" = null
WHERE "ShowLastNameFirst" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameFirst" = null
WHERE "NameFirst" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameLast" = null
WHERE "NameLast" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "CurrentHandle" = null
WHERE "CurrentHandle" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "PrizeUSD" = null
WHERE "PrizeUSD" = 'NULL';
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
UPDATE TEMP_API08_COMPLETEDDATASET
SET "PlayerId" = null
WHERE "PlayerId" = 'NULL';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "RankText" = null
WHERE "RankText" IN ('-', '', '?', '.') OR lower("RankText") = 'null';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "TeamName" = null
WHERE "TeamName" IN ('-', '', '?', '.') OR lower("TeamName") = 'null';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameFirst" = TRIM("NameFirst", '-');

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameFirst" = TRIM("NameFirst", ''); -- "

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameFirst" = null
WHERE "NameFirst" IN ('-', '', '?', '.', '--') OR lower("NameFirst") = 'null';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameLast" = TRIM("NameLast", '-');

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameLast" = TRIM("NameLast", '');

UPDATE TEMP_API08_COMPLETEDDATASET
SET "NameLast" = null
WHERE "NameLast" IN ('-', '', '?', '.') OR lower("NameLast") = 'null';

UPDATE TEMP_API08_COMPLETEDDATASET
SET "CurrentHandle" = null
WHERE "CurrentHandle" IN ('null', '', '-');

UPDATE TEMP_API08_COMPLETEDDATASET
SET "CountryCode" = null
WHERE "CountryCode" IN ('', '-') OR lower("CountryCode") = 'null';
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
-- "If a placement is associated with an unknown player, the "CurrentHandle" will be "##UNKNOWN##".
-- "PlayerId" in this instance is only used to return a unique row for each unknown player and can be discarded."
-- these ids were not unique and they maked trouble with real players IDs. For that reason we add 900 000 to them (max plays ID is
about 135 000), and now we know,
-- what the are exactly and we can easy filter them from real players by filtering < 900000.
UPDATE TEMP_API08_COMPLETEDDATASET
SET "PlayerId" = "PlayerId"::int+900000
WHERE "CurrentHandle" = '##UNKNOWN##';

-- create final table and remove rows where all columns are null

CREATE OR REPLACE TABLE API08_COMPLETEDDATASET AS
    SELECT *
    FROM TEMP_API08_COMPLETEDDATASET
EXCEPT
    SELECT *
    FROM TEMP_API08_COMPLETEDDATASET
WHERE
    "Ranking" is null
    AND "RankText" is null
    AND "TeamId" is null
    AND "TeamName" is null
    AND "CountryCode" is null
    AND "PlayerId" is null
    AND "NameFirst" is null
    AND "NameLast" is null
    AND "CurrentHandle" is null
    AND "ShowLastNameFirst" is null
    AND "PrizeUSD" is null;

-----

-- TOURNAMENTS RESULTS FOR TEAMS
CREATE OR REPLACE TEMPORARY TABLE TEMP_API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET AS
SELECT * FROM
(
    SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "TournamentTeamId", "TournamentTeamName", "PrizeUSD",
        "UnknownPlayerCount"
    FROM "09TournamentTeamResultsByTournamentID_IVETA" --1
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "TournamentTeamId", "TournamentTeamName", "PrizeUSD",
      "UnknownPlayerCount"
FROM "09TournamentTeamResultsByTournamentId_2IVETA" --2
UNION
SELECT "TournamentId", "Ranking", "RankText", "TeamId", "TeamName", "TournamentTeamId", "TournamentTeamName", "PrizeUSD",
      "UnknownPlayerCount"
FROM "09TournamentTeamResultsBytournamentID_teamplay1newids_IVETA" --3
);

-- data cleaning
UPDATE TEMP_API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET
SET "RankText" = null
WHERE "RankText" IN ('-', '?', '.') OR lower("RankText") = 'null';

UPDATE TEMP_API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET
SET "TournamentTeamName" = null
WHERE "TournamentTeamName" IN ('-', '?', '.') OR lower("RankText") = 'null';

UPDATE TEMP_API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET
SET "TeamId" = null
WHERE "TeamId" = 0;

-- create final table and remove rows where all columns are null
CREATE OR REPLACE TABLE API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET AS
SELECT *
FROM TEMP_API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET
EXCEPT
SELECT *
FROM TEMP_API09TOURNAMENTTEAMRESULTSBYTOURNAMENTID_COMPLETE_DATASET
WHERE "Ranking" is null
      AND "RankText" is null
      AND "TeamId" is null
      AND "TeamName" is null
      AND "TournamentTeamId" is null
      AND "TournamentTeamName" is null
      AND "PrizeUSD" is null
      AND "UnknownPlayerCount" is null;
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

-- 10 TOURNAMENT RESULTS FOR PLAYERS IN TEAMS

CREATE OR REPLACE TABLE TEMP\_API10 AS

SELECT \* FROM

(

SELECT "TournamentId", "TournamentTeamId", "PlayerId", "CountryCode", "NameFirst", "NameLast", "CurrentHandle",  
"ShowLastNameFirst"

FROM "10TournamentTeamPlayersByTournamentId\_1" --1

UNION

SELECT "TournamentId", "TournamentTeamId", "PlayerId", "CountryCode", "NameFirst", "NameLast", "CurrentHandle",  
"ShowLastNameFirst"

FROM "10TournamentTeamPlayersByTournamentId\_2" --2

UNION

SELECT "TournamentId", "TournamentTeamId", "PlayerId", "CountryCode", "NameFirst", "NameLast", "CurrentHandle",  
"ShowLastNameFirst"

FROM "10TournamentTeamPlayersByTournamentId\_teamplay1newids" --3

);

-- basic data cleansing

UPDATE TEMP\_API10

SET "NameFirst" = TRIM("NameFirst", ''); --"

UPDATE TEMP\_API10

SET "NameFirst" = null

WHERE "NameFirst" IN ('-', '--', '?');

UPDATE TEMP\_API10

SET "NameLast" = null

WHERE "NameLast" IN ('-', '?');

UPDATE TEMP\_API10

SET "NameLast" = TRIM("NameLast", '-');

-- create final table and remove rows where all columns are null

CREATE OR REPLACE TABLE API10TOURNAMENTTEAMPLAYERSBYTOURNAMENTID\_COMPLETE\_DATASET AS

SELECT \* FROM TEMP\_API10

EXCEPT

SELECT \* FROM TEMP\_API10

WHERE

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```

    "TournamentTeamId" is null
AND "PlayerId" is null
AND "CountryCode" is null
AND "NameFirst" is null
AND "NameLast" is null
AND "CurrentHandle" is null
AND "ShowLastNameFirst" is null;

```

```
-- To scrape players' birth dates, it was necessary to create an http address for each player.
-- However, this address does not support certain characters, which we had to either replace or remove.
-- Then we created a table of http addresses.
```

```
CREATE OR REPLACE TEMPORARY table POMOCNALINK AS
SELECT
```

"playerid",  
"CurrentHandle",  
RTRIM(REPLACE(REPLACE(REPLACE(REPLACE(TRANSLATE("CurrentHandle", '\_.  
;ĀĂĄǼȦÊİĐŇÓÔøāáääāāçēēēēīīīīñōōōōōōūüÿăăćĈêēēītłńŃôśśššššűźŻżųąąăăăăăăěēēēīiōōōōōōōōóúúüÿŷŋşıÜỲöǺÈÙúčůŽžÚÁı"AEZINOΣΧΨΩαζουCSA  
БВЕЖИМНОПРСТФШЭавгеикморстухчэяёsilZ"۝۞۟۠#%&()\*+,-  
./:;<>?@[{}^\_`|~²³¹²¹³¼½¾⅜⅝⅞⅞♥、あうえおかがくぐけござしじすぜそただちつてでとなにぬねのはぱひふぶへべぺぼまみむもやゆりるろわ  
アイカクスタデバプマミラワ　一丁七万三上不与东严中、丸久么乌乐九了二于五亚京人今任企体余你侃侃来俊倒傻元光入全两六兴冠冬冯冰冷凉凌凡凤凯  
出刀刘初别到剑劣勇勝北十千半卓单南博卧原厨古另叫叮可叶司同向吒吕君听周呵哈哦哩哲唐啊善喝喘嗜嘉嘘嚙囁团圉坑坚垫執基塔墙墨夏多夜夢大天失奋  
奔奶如妖姑娘孳子季孤宇安宝宮家宸宽宿小少生就尹尾屁属岩川左巳巴布帅希平幸幻幽庑庭康弓张弹归影很徐微心忘快念思怠怪总恒悠悲情惊想慌慕懂成我  
战戰房打托扶拿指摯挽搁搬撬放教整文斗斩新无旧旬时昔星春昱晓晨晴智暑暖暗暮暴曲曼最月朔望朝朧本杀李杰枉林果枪枫柴柠树核格桃桜梓梦梨棒榆櫻  
橘正歪死毀毕毛气水永江汤汪沐没河波泼泽洋派流浅浩浪浮海涛涵清温渲演漫瀟火灵烟烤煊無熊熙燈燕爰牛狗狸狼猪玉王玖环珠理瑶甜生用男痲疾白的皮盒  
盼真着瞬知石破磊祖神秀秉秋科空站竞章童端笑答米紅紫縱繁红纯罗翔老耐肥肯育胶臭航艾芬花苏若苦英范莫萸菊萧落蕕蘭虎虐虚虫虾蛋蛭蜎行街袁西见言  
净请诺迷谨豆賴贺赖赛赫赵超越路蹦蹦躺轨轩轻载辉辰运这进迷送逆逗造進遙邪邪邾郎郭郑都酒醬醅醉醒释野金钢银锁锋错长閤闪闪阴阳阿陆陈陌降随雨雪零  
雷雾霜青靚靜韻頤頤領風飞餅馬马驴骄高鬼魏鱼鯊鸡麓麦麻黄黑默默龍龙龜강게고공괘국국권굴근길김김꼴나내노다달돌러러력레렐로매맥문박발방백백  
별분상서소순송신심아안에악양열영오용원유윤이임자잔장재전절정조주진최치코큰콩탱팍프프하한핫헌홍황=\\γηνβληηφωψαθΒΞδθ

n あ い い お き ぎ ざ せ ぞ つ ば び ぶ ぼ や ゅ よ よ れ を ん イ ウ ケ コ ッ ノ ハ フ ブ メ ル レ ン 一 下 | 个主乃义之也乱予争云井亦介仔仙代们休会伟伦似位何佳依侠修倍假  
偶偷傲僧儿先克兒兔兰兽再冒军冻准击分利剧力劫势勒勺包化午华卫叁双叔变只号吃合后吐吞呀命咩品哟哥唯唱哈啪嘿四圆在地坤坦城域堂壺壹夕太头奇  
女妄妹妾姻姜姬威字孙孝学孩孽宅客容寂寒寞寻尖尤允尽居履山屿岁島崩崽工巧帝带席帮常千年幼庆底开弟弱强当形得德忆怀性恩怙恹惰意懒户手才扎执  
扬承技把投拔招掌排拊搯敏斌斯方既日旭旺明是昼晋晒木末机权杉条来杨杯某染桡杻查柳梲栝栖桀案柯桥森椒槌橘欧欣歌步殇比氏汉沉沔油洒洛涯渡港游潭  
澈灏点烦焮然照爆爷牙牢物独狙猜猫玮球琉琛琪琴瑞瑾璐瓜瓣疤痕癰皂盗省眠睡督睿砂砖碧確禹离秘等筱简糞糖素索约级纳纸练终给维缘羊美羨群羽者肉  
胃騰自舞良色艺芒茶草蒗莉莲菇菜萨蒼蕈薇藏虹蝶血衣裘裡要觀觉解訕诚话语谦公豪贤败贱躁输辞达远迪迭迹逍逸逼道遗酸里鉄鑫鈇铁钐铭锥门间限隻雄

雒電霖霸霧非面頭顏飄香駿騎騷骨魂魔魔魚魯鵲烏鸛鳴鴨鹿黎黑가감개건검경금관기는늘니대덕도동뜨떨론머먼명모문미민범베봉브비빈빔사산석선성세  
수순스승악언연예완우운웅은의인쟁죃중준지집창창치천철추가궁킨탓태택되포플혁형호환훈◻◆6тКидьяR믄ㄴゑざどびめ。ガグゴチトドナネヤ丝乎  
乔仓伞传作使便像八兼冲凶加压叩员咪咪噜噴回园壁土声奈好娇學定对将尔帕帽往怎戀愛戰抱揚故斬旅朽杂米梧棍樣欢欲武汁求沧涼涼深潰源滴炎焰煮特猎  
猖猛瑜璃廿異直看稚竹第绎義职聖背脑膏膨舍节芷药茵萌葩蓝薄蜂蜜兎记貓費貝贼起跑过逋達那郇鈞锦镜闭险陶霞食鴈鶩鼻齐광규균곤날덱드래룩리말무  
봇살섭술식와왕위를일제찬콜길터투트티피헌홀희9δριμηλGa『オソホロ・世业乡供做停劊助勢呐咻圈巢度桑極歎残泪狂糕糞羽翼胀袋说誰間闹顿鼠갈까  
롤릭메쓰알어윙편힘3ø\$Имтуげベリ养应式界莖部바커5 R LNBϯE商槍생γ\İŁōoゞzəǾe Rηn u』F!бгхГДЗЙХЧЯжзьюѐëêéďұ•ĚFrлрэ©★ツ𠂉乾亮令  
仲伪凑凹则动劲勤卡又口台咖啄圣垓她姑妍姐娃娜幡建強彪彼悟拂撒收旋易曾札朱殘沒泉法泰澤煎片狮猩玩珊瑚甲百絕腫种芭莱萊萝苣藤蘇补詩詹货贴车轮  
软还闷陵隨飯騒거교국네당뚜롱링물보부쌀에엽함\_lũ@△ボ丘兵凸刹办取吧啡器因圾坎塞师座坎旦昂栢渣濕灣炼珩瓶綺能装褲蕪辣钩锅關靠鴿계따람맞북  
후Nμε⑥係勁名嗶拙政更紗臨鉈震키? \KŮ巽〜艮兑' , ' ----

```
CREATE OR REPLACE table PLAYERLINK AS
SELECT
    CASE
        WHEN NICKNAME is null AND KRESTNI is null AND PRIJMENI is null THEN lower(CONCAT('https://www.esportsearnings.com/players/',
"playerid", '-'))
        WHEN NICKNAME is null AND KRESTNI is null THEN lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-',
PRIJMENI))
        WHEN KRESTNI is null AND PRIJMENI is null THEN lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-',
NICKNAME))
```



## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
    WHEN NICKNAME is null AND PRIJMENI is null THEN lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-',
KRESTNI))
    WHEN KRESTNI is null THEN lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-', NICKNAME, '-', PRIJMENI))
    WHEN PRIJMENI is null THEN lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-', NICKNAME, '-', KRESTNI))
    WHEN NICKNAME is null THEN lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-', KRESTNI, '-', PRIJMENI))
    ELSE lower(CONCAT('https://www.esportsearnings.com/players/', "playerid", '-', NICKNAME, '-', KRESTNI, '-', PRIJMENI))
END as LINK,
"playerid",
"CurrentHandle",
NICKNAME,
"NameFirst",
KRESTNI,
"NameLast",
PRIJMENI,
"CountryCode"
FROM POMOCNALINK;
```

```
-- -----
-- merging files with birth dates and creating a table with player ID and birth date
-- players with a date of birth greater than 2012 have been excluded, as such dates of birth are probably not filled in correctly
-- (There was a real gap in the data between 2013 and 2016 and players born since 2017 would be 6 years old or younger, which is
unlikely.
-- There were also players born in 2023 in the data, which doesn't really make sense, and it is a misfilled data.)
CREATE OR REPLACE TABLE PLAYERIDDOB AS
SELECT * FROM
(
    SELECT
        "playerid",
        TO_DATE("DateOfBirth", 'MMMM DD, YYYY') as "DateOfBirth",
        YEAR(TO_DATE("DateOfBirth", 'MMMM DD, YYYY')) as "Year"
    FROM
        (
            SELECT "Url", "DateOfBirth"
            FROM "DateOfBirth1"
            UNION
            SELECT "Url", "DateOfBirth"
            FROM "DateOfBirth2"
            UNION
            SELECT "Url", "DateOfBirth"
            FROM "DateOfBirth3"
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
UNION
SELECT "Url", "DateOfBirth"
FROM "DateodBirth4"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateodBirth5"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateodBirth6"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateodBirth7"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateOfBirth08"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateOfBirth09"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateOfBirth10"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateOfBirth_errors_znovu_stazeno1"
UNION
SELECT "Url", "DateOfBirth"
FROM "DateOfBirth_errors_znovu_stazeno_2"
)
JOIN PLAYERLINK p ON "Url"="LINK"
WHERE "DateOfBirth" != '<unknown>' AND "DateOfBirth" != '' AND "DateOfBirth" is not null
)
WHERE "Year" <= 2012;

-- -----

-- when downloading data from the API, various errors occurred during the download and some data had to be downloaded again
-- so we created lists of IDs to be re-downloaded in python. Sometimes this had to be repeated for newly downloaded (error) IDs.

CREATE OR REPLACE TABLE GAMEBYID_ERRORIDS AS
SELECT "gameid"
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
FROM "04LookupGameById_vystup0-1000"  
WHERE "ErrorCode" != '';
```

```
CREATE OR REPLACE TABLE TOURNAMENTERRORIDS AS  
  SELECT "tournamentid"  
    FROM "07TournamentById_vystup1000-1999"  
WHERE "ErrorCode" != '' OR "ErrorCode" is not null  
UNION  
  SELECT "tournamentid" FROM "07TournamentById_vystup14359-24358_novecsv"  
  WHERE "Error" != '' OR "Error" is not null  
UNION  
  SELECT "tournamentid" FROM "07TournamentById_vystup24359-52708_novecsv"  
  WHERE "ErrorCode" != '' OR "Error" != '' OR "ErrorCode" is not null OR "Error" is not null  
UNION  
  SELECT "tournamentid" FROM "07TournamentById_vystup52808-82806_novecsv"  
  WHERE "ErrorCode" != '' OR "Error" != '' OR "ErrorCode" is not null OR "Error" is not null  
  ORDER BY "tournamentid";
```

```
CREATE OR REPLACE TABLE DOB_ERRORS AS  
SELECT * FROM  
  (  
    SELECT "Url", "DateOfBirth", NULL as "Error"  
    FROM "DateOfBirth1"  
    UNION  
    SELECT "Url", "DateOfBirth", NULL as "Error"  
    FROM "DateOfBirth2"  
    UNION  
    SELECT "Url", "DateOfBirth", "Error"  
    FROM "DateOfBirth3"  
    UNION  
    SELECT "Url", "DateOfBirth", "Error"  
    FROM "DateOfBirth4"  
    UNION  
    SELECT "Url", "DateOfBirth", "Error"  
    FROM "DateOfBirth5"  
    UNION  
    SELECT "Url", "DateOfBirth", "Error"  
    FROM "DateOfBirth6"  
    UNION  
    SELECT "Url", "DateOfBirth", "Error"
```

## PROJECT ESPORT – SQL: 01 UNIONS AND FIRST CLEANSING

```
FROM "DateodBirth7"
UNION
SELECT "Url", "DateOfBirth", "Error"
FROM "DateOfBirth08"
UNION
SELECT "Url", "DateOfBirth", "Error"
FROM "DateOfBirth09"
UNION
SELECT "Url", "DateOfBirth", "Error"
FROM "DateOfBirth10"
)
WHERE "Error" is not null AND "Error" != '';

CREATE OR REPLACE TABLE DOB_ERRORS_2 AS
SELECT "Url", "Error"
FROM "DateOfBirth_errors_znovu_stazeno1"
WHERE "Error" is not null AND "Error" != '';
```