```
-- PART 1: NUMERICAL COMPARING VS. NUMERICAL VARIABLES
-- ===================================================


-- We can use tools such as the correlation index calculation


-- Is there a relationship between salary and the number of points per
-- game?
-- We can calculate the correlation index between the "salary_m" column and
-- the "points_per_game" column


SELECT CORR(salary_m, points_per_game) AS corr_salary_ppgame

FROM players_cleaned;

-- Observation:

-- - There is a positive correlation (close to 0.8) between salary and
--   points obtained in each game. This definitely seems to be a factor
--   associated at higher salaries


--  And we can "spun more finely" and see the correlation index between
--  salary and variables like "field_goals_pctg", "three_pt_pctg",
--  "two_pt_pctg", "free_throws_pctg"

SELECT

ROUND(CORR(salary_m, points_per_game):: numeric,2) AS corr_salary_ppgame,

ROUND(CORR(salary_m, field_goals_pctg):: numeric,2) AS corr_salary_fieldgp,

ROUND(CORR(salary_m, three_pt_pctg):: numeric,2) AS corr_salary_threegp,

ROUND(CORR(salary_m, two_pt_pctg):: numeric,2) AS corr_salary_twogp,

ROUND(CORR(salary_m, free_throws_pctg):: numeric,2) AS corr_salary_freetp

FROM players_ cleaned;


-- Note:

-- There is no real correlation between salary and percentage of points
-- earned for different categories (1 point, 2 points, 3 points, or
-- percentage of total points)

-- The most relevant factor is still the number of points per game
```

```
-- Let's see what relationship exists between salary earned and player age

SELECT CORR(salary_m, age) AS corr_salary_age

FROM players_cleaned;


-- Note:

-- - The correlation index obtained is only 0.4, so there is no direct or
--   inverse relationship between the players' salaries and their age.


--  And finally, let's see what relationship exists between the salary
--  earned and the number of games played.


SELECT CORR(salary_m, games_played) AS corr_salary_gp

FROM players_ cleaned;


-- Note:

-- - The correlation index is only 0.3, so there is no relationship between
     salary and the number of games played.


-- PART 2: CATEGORICAL COMPARING VS. NUMERICAL VARIABLES

-- ======================================================


-- In this case, besides the name, we only have one categorical variable
-- ("position").

-- Let's try to answer the question: which positions are associated with
-- the highest salaries?


-- We can calculate the average salary for each position.


SELECT position, ROUND(AVG(salary_m),1) AS average_salary

FROM players_cleaned

GROUP BY position

ORDER BY average_salary DESC;

-- Observations:
```

```
-- - It's clear that the PG-SG (point guard - shooting guard) position has
--   by far the highest average salary ($21.5 million), followed by the SG-
--   PG (shooting guard - point guard) position with an average of $16.7
--   million and by the PG (point guard) and SF-SG (small forward -
--   shooting guard) positions both with salaries around $12 million.

-- - The lowest-paid position is SF-PF (small forward - power forward) with
--   an average of $3 million.

-- - The best salaries definitely go to players with a more proficient
--   profile, distributing playmakers and scoring points for the team.


-- PART 3: MULTIVARIATE ANALYSIS

-- ====================================================


-- We can combine categorical and numerical variables with different tools
-- learned so far to perform a more in-depth analysis.


-- For example, for each position (categorical variable), let's analyze not
-- only the average salary but also the average points per game (which we -
-- saw in the first part was the most relevant variable for defining
-- salary)


SELECT position,

ROUND( AVG( salary_m ),1) AS average_salary ,

ROUND( AVG( points_per_game ),1) AS average_ppg

FROM players_cleaned

GROUP BY position

ORDER BY average_salary DESC;


-- Observations:

-- - This analysis allows us to verify once again that the PG-SG position
--   is the one with the highest salary and that this is directly
     associated with the average
--   points per game (which is 16.1 for these players)

-- - And again, the position with the lowest salary level is SF-PF, which
     in turn has the lowest average number of points per game (3.0)
```

```
-- And we can also add an analysis of the average age to the previous
-- variables.

SELECT position,

ROUND( AVG( salary_m ),1) AS average_salary ,

ROUND( AVG( points_per_game ),1) AS average_ppg ,

ROUND( AVG( age),1) AS average_age

FROM players_cleaned

GROUP BY position

ORDER BY average_salary DESC;


-- Observations:

-- - And here's another interesting fact: not only are the highest-paid
--   players the most "veteran" (with an average age of 33), but also those
--   with the lowest salaries have the youngest average age (23).
```