

## Zadanie 1 - WHERE AMAZING HAPPENS (NBA)

**Termín odovzdania:** Mar 23, 2025 @ 23:59 AIS - zip súbor podľa pokynov.

- Neskoré odovzdania nebudú akceptované a budú automaticky považované za 0 bodov.
- V prípade zlého zip súboru je zadanie hodnotené za 0 bodov
- V prípade, že niektoré SQL súbory nebude možné spustiť alebo nahradiť premenu v `{{}}` podľa pokynov zadania, tak takáto úloha je hodnotená automaticky za 0 bodov.

**Zadania** musia byť vypracované samostatne a budú kontrolované na plagiátorstvo.

### Všeobecné informácie k SQL dopytom

V rámci zadaní nie je možné používať **LATERAL JOIN** a ani **rekurzívne dopyty**. Celá úloha musí byť napísaná ako jeden dopyt do databázy. V prípade, porušenia niektorej z týchto podmienok je daná úloha hodnotená za 0 bodov bez ohľadu na správnosť výsledku.

Za úlohu nebude udelený plný počet bodov pokiaľ sa preukáže neefektivita daného dopytu a to aj v prípade správneho výsledku.

**Poradie stĺpcov je potrebné použiť podľa ukážkových výstupov.** V prípade, že vo výstupe po výpočte sú desatiny čísla zaokrúhľujte ich na dve desatiny miesta pokiaľ úloha nehovorí inak.

**Pokyny k odovzdávanému súboru** Výstupom zadania je ZIP súbor pomenovaný podľa AIS prihlasobacie mena (napr. xhrasko.zip). Obsahom ZIP archívu sú SQL súbory, ktoré obsahujú riešené úlohy - názov SQL súboru je číslo úlohy (napr. 1.sql).

Odovzdávané súbory musia byť platné Jinja šablóny (SQL súbory obsahujú značky - podľa zadania - napr. `{{season_id}}`). V prípade, že očakávate `{{season_id}}` text, tak je potrebné mať ošetrované na úrovni SQL. Do `{{season_id}}` je vložená len hodnota.

Príklad pre používateľa `xhrasko`:

- `xhrasko.zip`
  - `1.sql`
  - `2.sql`
  - `3.sql`
  - `4.sql`
  - `5.sql`
  - `6.sql`

V prípade, že formát nie je dodržaný a testovací nástroj nedokáže prečítať úlohy - zadanie bude hodnotené nulovým počtom bodov bez nároku na konzultáciu / opravu.

Pre overenie samotnej štruktúry zip súboru spolu s nahradením výrazov je možné využiť skript dostupný na <https://github.com/FIIT-Databases/sql-tester>.

### **Prezentovanie zadania**

Každý študent musí rozumieť svojmu zadaniu. V prípade, že študent nebude počas prezentácie rozumieť svojmu zadaniu a vedieť ho modifikovať podľa požiadaviek vyučujúceho, tak je zadanie hodnotené za 0 bodov. Jednotlivé prezentácie zadaní sú na uvážení vyučujúceho.

## Všeobecné informácie k zadaniu a datasetu

Dataset vychádza z <https://www.kaggle.com/datasets/wyattowalsh/basketball?resource=download> a preto je možné dohľadať dodatočné informácie k jednotlivým atribútom práve tam, no môžu sa však vyskytovať určité odlišnosti.

Tabuľka `play_records` obsahuje informácie o jednotlivých udalostiach, ktoré sa udiali počas zápasu. Medzi dôležité atribúty patria:

- `event_msg_type` - atribút hovorí o type udalosti. V rámci úloh sa primárne používajú `'FREE_THROW'`, `'FIELD_GOAL_MADE'`, `'FIELD_GOAL_MISSED'`, `'REBOUND'`. Samotná úloha môže mať však došpecifikovaný aj iný typ.
- `event_number` - hovorí o poradí udalosti v rámci konkrétneho zápasu `game_id`
- `player1_id` - predstavuje ID hráča, ktorý uskutočnil danú udalosť
- `player1_team_id` - ID tímu hráča `player1_id`
- `player2_id` - predstavuje ID hráča, ktorá dopĺňa informáciu k nejakému typu udalosti. V prípade udalosti `'FIELD_GOAL_MADE'` reprezentuje hráča, ktorá nahrával na úspešne skórovanie a teda sa pripisuje asistenciu.
- `player2_team_id` - ID tímu hráča `player2_id`
- `score` - predstavuje stav po úspešnom skórovaní. V prípade udalosti `'FREE_THROW'`, `'FIELD_GOAL_MADE'`. Ak trestný hod (udalosť `'FREE_THROW'`) nebol úspešne premenený tak je nastavený na hodnotu `NULL`.
- `period` - číslo štvrtiny
- `wctimestring` - počet zostávajúcich minút do konca štvrtiny

**Tabuľka** `team_history` obsahuje históriu jednotlivých tímov - zmeny názvov. Dôležité atribúty:

- `year_founded` - rok založenia tímu s daným názvom. V rámci úloh sa uvažuje platnosť od `1. Júla daného roku`.
- `year_active_till` - rok ukončenia tímu s daným názvom. V rámci úloh sa uvažuje platnosť do `30. Júna daného roku`. Platný názov končí s hodnotou `2019` a teda je potrebné uvažovať v takomto prípade, že názov tímu je platný až do súčasnosti.
- `game_id` - odkaz na tabuľku s tímami `teams`, kde sa nachádza aj posledný platný názov daného tímu.

**Tabuľka** `teams` - obsahuje posledné názvy tímov.

**Tabuľka** `games` - obsahuje všeobecné informácie o zápasoch vrátane toho, ktoré tímy hrali v danom zápase

**Tabuľka** `players` - zoznam hráčov. Hodnota `True` znamená, že daný hráč je ešte aktívny.

## Úloha 1 (1 bod)

Nájdite všetky udalosti, kedy hráč v určitom zápase po tom, čo uskutočnil doskok ( `event_msg_type= 'REBOUND'`) skóroval `'FIELD_GOAL_MADE'`. Medzi týmito dvomi aktivitami nesmie existovať iná udalosť. To či medzi tým boli ešte nahrávky zanedbajte. Pre nájdených hráčov uveďte čas skórovania, a tiež periódu skórovania. Vo výstupe je potrebné vrátiť nasledujúce atribúty:

- `player_id` - id hráča
- `first_name` - krstné meno hráča
- `last_name` - priezvisko hráča
- `period` - číslo štvrtiny
- `period_time` - zostávajúci čas štvrtiny

Primárne zoradenie je podľa `period` vzostupne (Ascending) následne `period_time` zostupne (Descending) a následne `player_id` vzostupne (Ascending).

Kvôli testovaniu je potrebné uviesť v odovzdanom SQL súbore ID zápasu ako `{{game_id}}`.

**Poradie stĺpcov je potrebné použiť podľa ukážkového výstupu.**

Pre zápas s `id=22000529` je prvý záznam:

player_id	first_name	last_name	period	period_time
202696	Nikola	Vucevic	1	8:22
...	...	...	...	...

## Úloha 2 (3 body)

Pre vybranú sezónu nájdite všetkých hráčov, ktorý počas sezóny zmenili tím. Zmena tímu je považovaná ak hráč odohral za iný tím nejaký zápas a to ak v tabuľke `play_records` sa nachádza hráčove ID na atribútoch `player1_id` alebo `player2_id` a to pre typy udalosti `'FREE_THROW'`, `'FIELD_GOAL_MADE'`, `'FIELD_GOAL_MISSED'`, `'REBOUND'`. Pre týchto hráčov spočítajte koľko krát zmenili tím počas danej sezóny. Pre prvých piatich hráčov, ktorý najviac zmenili tím vypočítajte priemerný počet bodov na zápas (udalosť typu `FIELD_GOAL_MADE` berte za 2 body - nie je potrebné rozlišovať 2 a 3 body), priemerný počet asistencií a počet zápasov, ktoré daný hráč odohral za konkrétny tím.

V prípade, že existuje viacero hráčov, ktorý majú rovnaký počet výmien počas sezóny, tak zoradenie je podľa toho či je hráč aktívny alebo nie - aktívny hráč má prednosť. Následne priezvisko vzostupne (Ascending) a nasleduje meno vzostupne (Ascending).

Vo výstupe uveďte nasledovné informácie

- `player_id` - id hráča
- `first_name` - krstné meno hráča
- `last_name` - priezvisko hráča
- `team_id` - id tímu
- `team_name` - plné meno tímu (`full_name`)
- `PPG` - points per game - priemerný počet bodov na zápas pre daný tím. Zaokrúhlené na dve desatiny miesta.
- `APG` - assists per game - priemerný počet asistencií na zápas pre daný tím. Zaokrúhlené na dve desatiny miesta.
- `games` - počet hier, ktoré odohral hráč za daný tím v sezóne.

Výstup je potrebné primárne zoradiť podľa `player_id` vzostupne (Ascending) a následne podľa `team_id` vzostupne (Ascending).

Kvôli testovaniu je potrebné uviesť v odovzdanom SQL súbore ID sezóny ako `{{season_id}}`.

**Poradie stĺpcov je potrebné použiť podľa ukážkového výstupu.**

Prvý záznam pre `season_id=22017`

player_id	first_name	last_name	team_id	full_name	PPG	APG	games
202328	Greg	Monroe	1610612738	Boston Celtics	10.82	2.27	22

## Úloha 3 (3 body)

Pre konkrétny zápas vypočítajte pre každého hráča nasledovné:

- **points** - celkový počet bodov, ktorý hráč dal v zápase (vrátane Free Throws)
- **2PM** - two points made - celkový počet úspešných hodov za 2 body (je to počet hodov nie počet bodov)
- **3PM** - three points made - celkový počet úspešných hodov za 3 body (je to počet hodov nie počet bodov)
- **missed\_shots** - počet nepremených hodov za 2 alebo 3 body (z dát nie je možné zistiť, či netrafený hod je za 2 alebo 3 body)
- **shooting\_percentage** - úspešnosť streľby za 2 a 3 body. Zaokrúhlené na 2 desatinné miesta
- **FTM** - free throw made - počet úspešných trestných hodov
- **missed\_free\_throws** - počet nepremených trestných hodov
- **FT\_percentage** - úspešnosť streľby v rámci trestných hodov. Zaokrúhlené na 2 desatinné miesta

Okrem vyššie uvedených informácií je potrebné uviesť aj informácie o danom hráčovi, pre ktoré platia dané štatistiky a to:

- **player\_id** - id hráča
- **first\_name** - krstné meno hráča
- **last\_name** - priezvisko hráča

Výstup je primárne zoradený podľa **points** zostupne (Descending), následne **shooting\_percentage** zostupne (Descending), následne **FT\_percentage** zostupne (Descending), následne **player\_id** vzostupne (Ascending).

Kvôli testovaniu je potrebné uviesť v odovzdanom SQL súbore ID zápasu ako **{{game\_id}}**.

**Poradie stĺpcov je potrebné použiť podľa ukážkového výstupu.**

Prvý záznam výstupu dopytu pre id zápasu **21701185** je:

player_id	first_name	last_name	points	2PM	3PM	missed_shots	shooting_percentage	FTM	missed_free_throws	FT_percentage
1627759	Jaylen	Brown	32	4	7	8	57.89	3	2	60

## Úloha 4 (3 body)

**Triple Double** - keď hráč v zápase získa v rámci štatistik 3 typy štatistik, ktoré dosiahli dvojcifernú hodnotu. V tejto úlohe budeme uvažovať iba o štatistikách počte bodov, asistenciách a doskokoch.

Pre vybranú sezónu nájdite všetkých hráčov, ktorý počas sezóny dosiahol triple double. Následne pre týchto hráčov určite ich najdlhšiu sériu triple double v zápasoch bez prerušenia. Ak dva po sebe idúce zápasy dosiahol hráč triple double, tak táto séria má hodnotu 2. Ak by v jeho nasledujúcom zápase si pripísal ďalšie triple double, tak sa táto séria natiahne na hodnotu 3 inak zostane na hodnote 2. Pri opätovnom získaní triple double v zápase začína táto séria od 1.

Pre zjednodušenie výpočtu hodov v prípade udalosti typu **FIELD\_GOAL\_MADE** berte koše za 2 body - nie je potrebné rozlišovať 2 a 3 body), priemerný počet asistencií a počet zápasov, ktoré daný hráč odohral za konkrétny tím. Typ udalosti je potrebné tiež brať do úvahy **'FREE\_THROW'** a je potrebné rozlišovať úspešne a neúspešne premenený trestný hod.

V rámci úlohy berte do úvahy typy udalostí **FIELD\_GOAL\_MADE**, **'FREE\_THROW'** a **'REBOUND'**.

Výstupom dopytu je zoznam hráčov - jeho id a jeho maximálna séria s triple double. Výstupom je:

- **player\_id** - id hráča
- **longest\_streak** - najdlhšia séria triple double bez prerušenia

Výstup primárne zoradíte podľa **longest\_streak** zostupne (Descending) a následne podľa **player\_id** vzostupne (Ascending).

Kvôli testovaniu je potrebné uviesť v odovzdanom SQL súbore ID sezóny ako **{{season\_id}}**.

**Poradie stĺpcov je potrebné použiť podľa ukážkového výstupu.**

Pre sezónu s **id = '22018'** sú prvé dva záznamy:

player_id	longest_streak
201566	9
203901	3

## Úloha 5 (2 body)

Pre jednotlivé tímy spočítajte koľko zápasov odohrali ako domáci tím a koľko ako hosťujúci tím. Okrem počtu zápasov je potrebné uviesť aj koľko percent tvoria domáce zápasy a koľko zápasy vonku. Tím je potrebné rozlišovať nie len podľa atribútu `id`, ale aj podľa jeho historického názvu tj. pokiaľ tím s rovnakým `id` má v počas svojej existencie napríklad 3 rôzne názvy, tak je to považované za 3 rôzne tímy. Ak má tím hodnotu `year_active.till` nastavený na hodnotu `2019` znamená to, že tento názov je stále platný. V prípade, že je rok iný tak platnosť daného názvu je potrebné uvažovať jeho platnosť do `30 Júna daného roka`. Atribút `year_founded` reprezentuje rok založenia tímu a je potrebné uvažovať jeho platnosť od `1 Júla daného roku`.

Vo výstupe uveďte nasledujúce:

- `team_id` - id tímu
- `team_name` - názov tímu
- `number_away_matches` - počet zápasov odohraných vonku
- `percentage_away_matches` - percentuálny podiel zápasov vonku z počtu zápasov daného tímu. Zaokrúhlené na dve desatiny miesta.
- `number_home_matches` - počet zápasov odohraných doma
- `percentage_home_matches` - percentuálny podiel domácich zápasov z počtu zápasov daného tímu. Zaokrúhlené na dve desatiny miesta.
- `total_games` - celkový počet

Výstup nech je zoradený primárne podľa `team_id` vzostupne (Ascending) a následne `team_name` vzostupne (Ascending).

**Poradie stĺpcov je potrebné použiť podľa ukážkového výstupu.**

Prvý záznam dopytu je:

<code>team_id</code>	<code>team_name</code>	<code>number_away_matches</code>	<code>percentage_away_matches</code>	<code>number_home_matches</code>	<code>percentage_home_matches</code>	<code>total_games</code>
1610612737	Atlanta Hawks	1986	50.27	1965	49.73	3951



## Úloha 6 (2 body)

Pre hráča nájdeného hráča na základe mena a priezviska vyberte jeho všetky sezóny typu 'Regular Season' v ktorých odohral aspoň 50 zápasov. Následne tomuto hráčovi pre vybrané sezóny určite ich stabilitu z pohľadu úspešnosti streľby z poľa tj. `event_msg_type = 'FIELD_GOAL_MADE'` a `event_msg_type = 'FIELD_GOAL_MISSED'`. Najstabilnejšia sezóna znamená, že je vyrátaný rozdiel úspešnosti streľby medzi jednotlivými zápasmi a následne je vypočítaný priemer nad celou sezónou. Prvý zápas sezóny ma hodnotu zmeny 0.

napr. Ak hráč v prvom zápase mal úspešnosť 20% v druhom 30% a v tretom 10% tak rozdiel medzi prvým a druhým zápasom je 10% a medzi druhým a tretím je to 20%. Prvý zápas je rozdiel považovaný za 0% a vstupuje tiež do výpočtu. V tomto prípade je to:  $(0+10+20)/3 = 10$ . Vyhodnotenie stability sezóny je v tomto prípade 10.

Vo výstupe sú uvedené nasledujúce informácie:

- `season_id` - id sezóny
- `stability` - stabilita sezóny

Výstup je zoradený podľa atribútu `stability` a to vzostupne (Ascending). Následne podľa `season_id` vzostupne (Ascending).

Kvôli testovaniu je potrebné uviesť v odovzdanom SQL súbore meno ako `{{first_name}}` a pre priezvisko `{{last_name}}`.

**Poradie stĺpcov je potrebné použiť podľa ukážkového výstupu.**

Prvý záznam pre hráča s menom `LeBron` a priezviskom `James` je prvý záznam:

season_id	stability
22007	9.84