**Hurtownie danych – Spr. 4.**

PWr. WIZ, Data: 13-16.04.2021

|  |  |  |
| --- | --- | --- |
| Student | ------------------------------------------------------------- | Ocena |
| Indeks | 251526 |  |
| Imię | Volodymyr |
| Nazwisko | Zakhovaiko |

Zestaw składa się z 4 zadań. Jeżeli nie potrafisz rozwiązać zadania, to próbuj podać, chociaż częściowe rozwiązanie lub uzasadnienie przyczyny braku rozwiązania. Pamiętaj o podaniu nr. indeksu oraz imienia i nazwiska.

Baza danych: **AdventureWorks2017 lub 2019**

**Zad. 1.**

Proszę przygotować dane do analizy zamówień w zakresie przedstawiony w tab. 1.

Tab. 1. Liczba zamówień w poszczególnych latach globalnie oraz obszarowo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rok | Liczba zamówień | Terytorium | Liczba zam. na terytorium | % udział |
| 2011 | 1607 | Australia | 463 | 28.81 |
| 2011 | 1607 | Southwest | 339 | 21.10 |
| 2011 | 1607 | Germany | 81 | 5.04 |
| 2011 | 1607 | Central | 50 | 3.11 |
| 2011 | 1607 | Northwest | 224 | 13.94 |

Rek.: **5/40**

**Rozwiązanie:**

Rozwiązanie bez klauzy OVER:

WITH Group\_Year\_CTE (Rok, Liczba)

AS

(

SELECT YEAR(SOH.OrderDate), COUNT(SOH.OrderDate)

FROM Sales.SalesOrderHeader AS SOH

GROUP BY YEAR(SOH.OrderDate)

),

Group\_Territory\_CTE (Rok, Liczba, TerritoryID)

AS

(

SELECT YEAR(SOH.OrderDate), COUNT(\*), TerritoryID

FROM Sales.SalesOrderHeader AS SOH

GROUP BY YEAR(SOH.OrderDate), TerritoryID

)

SELECT

GYC.Rok,

GYC.Liczba "Liczba zamowien",

ST.[Name] Terytorium,

GTC.Liczba "Liczba zam. na terytorium",

FORMAT(GTC.Liczba \* 100.0 / GYC.Liczba, '###.##') AS "% udzial"

FROM Group\_Year\_CTE AS GYC

JOIN Group\_Territory\_CTE AS GTC ON GTC.Rok = GYC.Rok

JOIN Sales.SalesTerritory AS ST ON GTC.TerritoryID = ST.TerritoryID

ORDER BY Rok;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rok | Liczba zamowien | Terytorium | Liczba zam. na terytorium | % udzial |
| 2011 | 1607 | Canada | 149 | 9.27 |
| 2011 | 1607 | Australia | 463 | 28.81 |
| 2011 | 1607 | Central | 50 | 3.11 |
| 2011 | 1607 | Northwest | 224 | 13.94 |

Rek.: 4/40

Rozwiązanie z klauzą OVER:

SELECT \*, FORMAT("Liczba zam. na terytorium" \* 100.0 / "Liczba zamowien", '###.##') AS "% udzial"

FROM (

SELECT DISTINCT

YEAR(soh.OrderDate) Rok,

COUNT(soh.OrderDate) OVER(PARTITION BY YEAR(soh.OrderDate)) "Liczba zamowien",

ST.[Name] Terytorium,

COUNT(soh.OrderDate) OVER (PARTITION BY ST.[Name], YEAR(soh.OrderDate)) "Liczba zam. na terytorium"

FROM Sales.SalesOrderHeader AS soh

JOIN Sales.SalesTerritory AS ST

ON ST.TerritoryID = soh.TerritoryID

) AS A

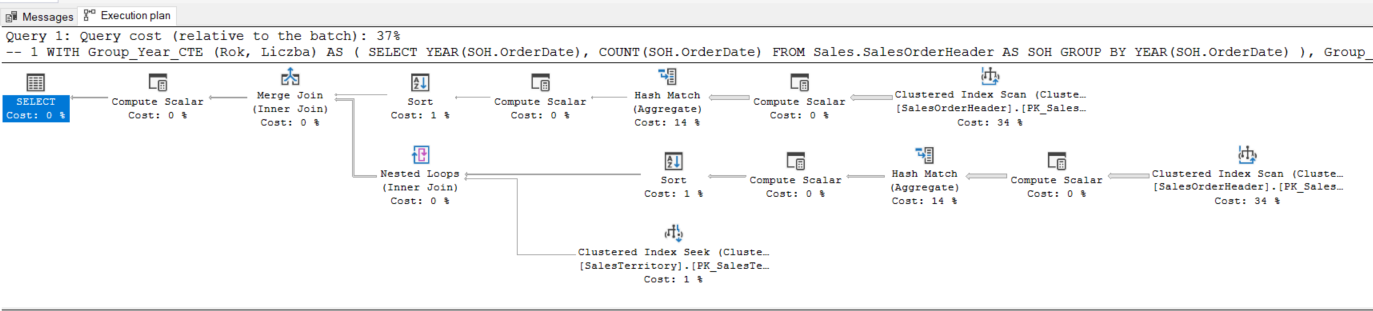
ORDER BY Rok;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rok | Liczba zamowien | Terytorium | Liczba zam. na terytorium | % udzial |
| 2011 | 1607 | Canada | 149 | 9.27 |
| 2011 | 1607 | France | 70 | 4.36 |
| 2011 | 1607 | Southwest | 339 | 21.1 |
| 2011 | 1607 | Australia | 463 | 28.81 |

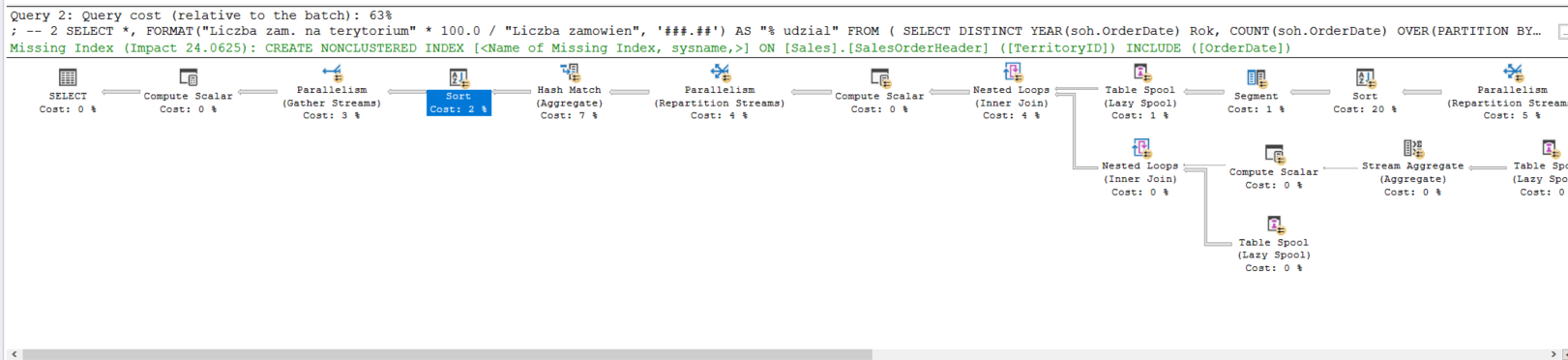
Rek.: 4/40

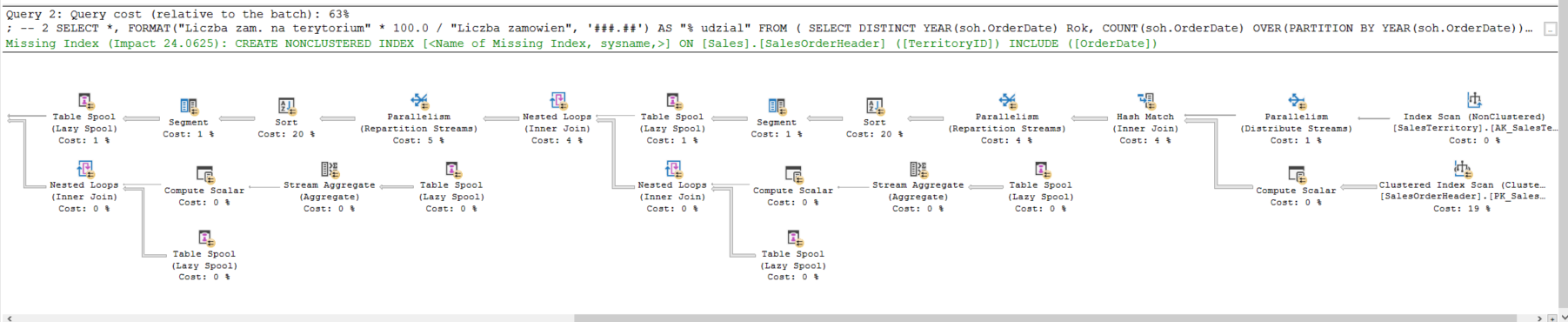
Porównanie:

1)



2)





**Zad. 2. Analiza danych i ocena ich jakości**

Przeanalizować, scharakteryzować i ocenić dane znajdujące się w pliku „**AviationData.xls**”, wykorzystując Tableau Prep oraz profilowanie danych z pakietu SSIS (projekt SQL Server Data Tools). Rozwiązanie przedstawić zgodnie z zakresem przedstawionym w tabelach 4.1. (słownik danych dziedzinowych) i 4.2 (ocena jakości danych źródłowych) dla przykładowego źródła danych (globalterrorism.csv).

Uwaga!

* **Analiza danych powinna być zrealizowana z wykorzystaniem pakietu Tableau Prep (materiały szkoleniowe można znaleźć na stronie Tableau oraz w Internecie)**
* Profilowanie danych z wykorzystaniem SSIS zostanie zrealizowane na zajęciach lab.

**Tabela 4.1 Interpretacja danych (słownik)**

|  |  |  |  |
| --- | --- | --- | --- |
| Plik: globalterrorism.csv | | | |
| Lp. | Atrybut | Typ wartości | Znaczenie |
| 1. | eventid | Numeryczne | Kod identyfikujący datę incydentu, podawany dla każdego incydentu |
| 2. | approxdate | Tekstowe | Przybliżona data wystąpienia incydentu, podawana, jeśli nie jest znany dokładny miesiąc/dzień ataku. Wtedy podawane jest np. pierwsza połowa danego roku (tu rok nam wystarczy) |
| … | … | … | … |

**Tabela 4.2 Ocena jakości danych**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plik: globalterrorism.csv | | | | |
| Lp. | Atrybut | Typ danych | Zakres wartości | Ocena jakości danych |
| 1 | iyear | Numeryczne | 1970-2016 | Rok wystąpienia incydentu  0% NULL |
| 2 | imonth | Numeryczne | 0-12 | Miesiąc wystąpienia incydentu (jeśli nie jest znany to 0)  0% NULL |
| … | … | … | … | … |
| 31 | attacktype2 | numeryczne | 1-9 | Kod typu ataku  97% NULL |
| … |  |  |  |  |
| 70 | nperps | numeryczne | -99-25000 | Całkowita liczba napastników uczestnicząca w incydencie (-99 = nie wiadomo)  60% NULL |
| … |  |  |  |  |
| 81 | compclaim | numeryczne | 0,1,null,-9 | Czy więcej niż jedna grupa twierdziła, że była odpowiedzialna za atak? (1 - tak, 0 nie, - 9 nie wiadomo). Nieprzydatne z punktu widzenia analizy |

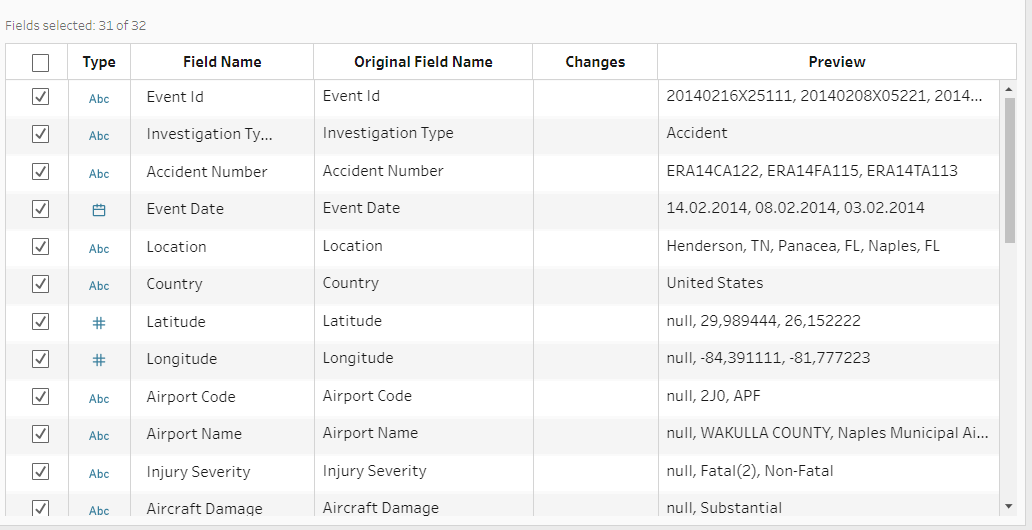
**Rozwiązanie:**

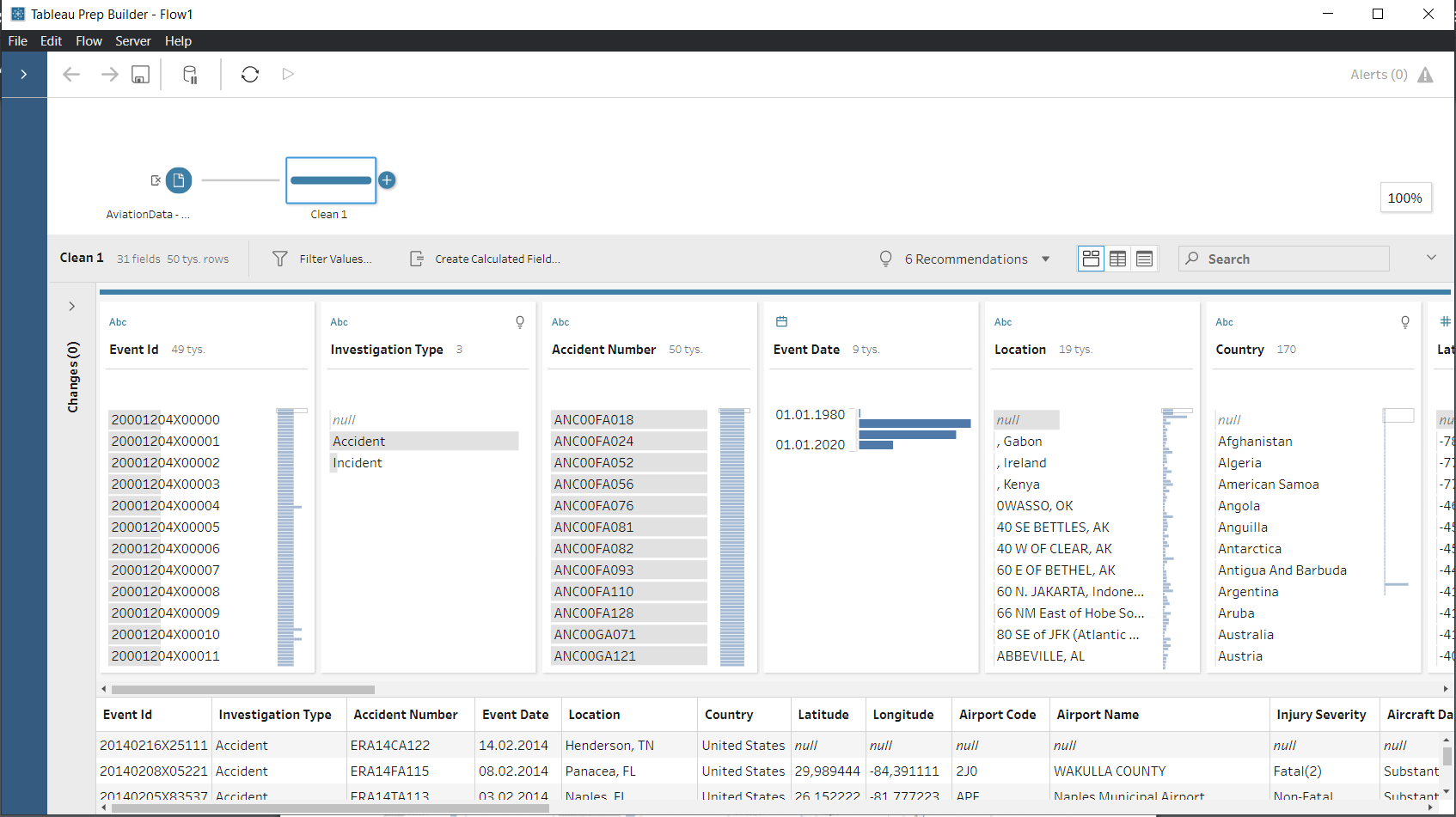
**Interpretacja danych:**

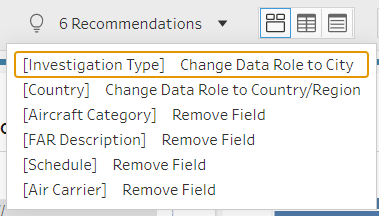
|  |  |  |  |
| --- | --- | --- | --- |
| Plik: AviationData.csv | | | |
| Lp. | Atrybut | Typ wartości | Znaczenie |
| 1. | Event Id | Tekstowe | Kod identyfikujący datę i numer incydentu, podawany dla każdego incydentu |
| 2. | Investigation Type | Tekstowe | Rodzaj dochodzenia |
| 3 | Accident Number | Tekstowe | Kod identyfikujący numer wypadku |
| 4 | Event Date | Data | Data wydarzenia wypadku |
| 5 | Location | Tekstowe | Miejsce (lokalizacja) wypadku. Miasto, miejsce oraz skrót województwa. |
| 6 | Country | Tekstowe | Kraj, gdzie został zarejestrowany wypadek. |
| 7 | Latitude | Numeryczne | Koordynata szerokości wypadku. |
| 8 | Longitude | Numeryczne | Koordynata długości geograficznej wypadku. |
| 9 | Airport Code | Tekstowe | Kod identyfikujący lotnisko. |
| 10 | Airport Name | Tekstowe | Nazwa lotniska. |
| 11 | Injury Severity | Tekstowe | Ciężkość obrażeń. |
| 12 | Aircraft Damage | Tekstowe | Uszkodzenie samolotu. |
| 13 | Aircraft Category | Tekstowe | Kategoria statku powietrznego (samolot, śmigłowiec, …). |
| 14 | Registration Number | Tekstowe | Numer rejestracyjny transportu. |
| 15 | Make | Tekstowe | Marka (konstrukcja) transportu. |
| 16 | Model | Tekstowe | Model statku. |
| 17 | Amateur Built | Tekstowe | Czy był zbudowany amatorsko. |
| 18 | Number of Engines | Numeryczne | Ilość silników. |
| 19 | Engine Type | Tekstowe | Typ silnika. |
| 20 | FAR Description | Tekstowe | Opis Federalnych przepisów lotniczych |
| 21 | Schedule | Tekstowe | Harmonogram lotu |
| 22 | Purpose of Flight | Tekstowe | Cel lotu. |
| 23 | Air Carrier | Tekstowe | Przewoźnik lotniczy. |
| 24 | Total Fatal Injuries | Numeryczne | Całkowite obrażenia śmiertelne. |
| 25 | Total Serious Injuries | Numeryczne | Całkowita liczba poważnych obrażeń. |
| 26 | Total Minor Injuries | Numeryczne | Całkowita liczba drobnych urazów. |
| 27 | Total Uninjured | Numeryczne | Ilość bez obrażeń. |
| 28 | Weather Condition | Tekstowe | Stan pogody na moment lotu. |
| 29 | Broad Phase of Flight | Tekstowe | Szeroka faza lotu. |
| 30 | Report Status | Tekstowe | Status raportu. |
| 31 | Publication Date | Data | Data publikacji wypadku. |

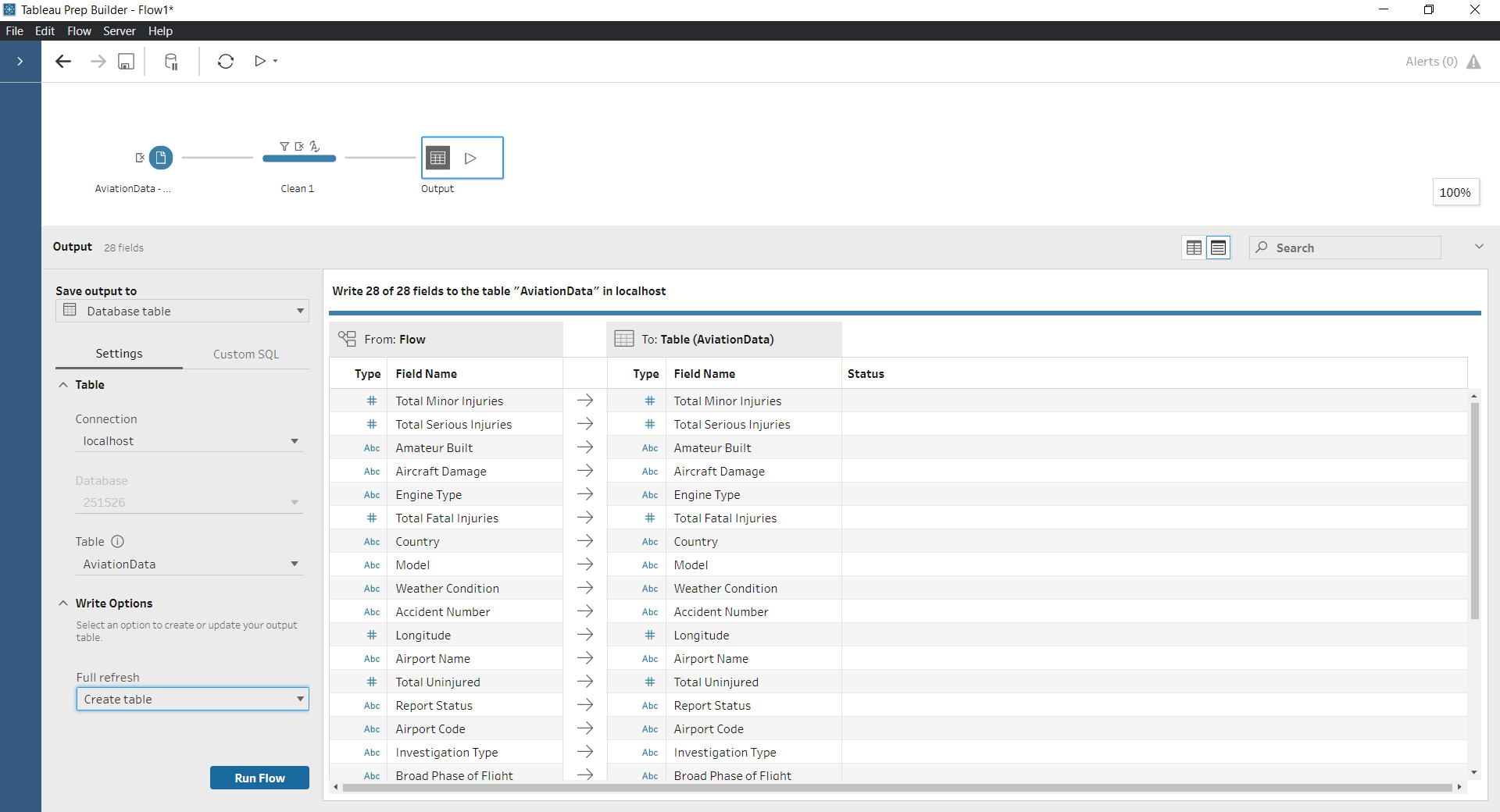
**Ocena jakości danych:**

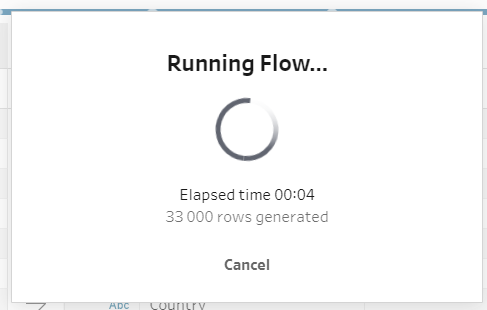
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plik: AviationData.csv | | | | |
| Lp. | Atrybut | Typ wartości | Zakres wartości | Ocena jakości danych |
| 1. | Event Id | Tekstowe | Długość: 14 | 0% null, 49289 unique |
| 2. | Investigation Type | Tekstowe | Accident (96%) lub Incident (4%) | Jedno znaczenie null (<1%),  3 unique |
| 3 | Accident Number | Tekstowe | Długość: 9-11 | 0% null, 49997 unique |
| 4 | Event Date | Data | 01.01.1980 - 01.01.2020 | 8708 unikatowych znaczeń |
| 5 | Location | Tekstowe | Długość: 4-61 | <1% null, 18965 unique |
| 6 | Country | Tekstowe | Długość: 4-30 | <1% null, 163 unique |
| 7 | Latitude | Numeryczne | -80 - 90 | 57% null, 14810 unique |
| 8 | Longitude | Numeryczne | -200 - 200 | 57% null, 15726 unique |
| 9 | Airport Code | Tekstowe | Długość: 1-8 | 42% null 7563 unique |
| 10 | Airport Name | Tekstowe | Długość: 2-33 | 40% null, 16352 unique |
| 11 | Injury Severity | Tekstowe | Długość: 8-11 | <1% null, 105 unique |
| 12 | Aircraft Damage | Tekstowe | „Destroyed” (20%), „Minor” (3%) lub „Substantial” (74%) | 3% null, 4 unique |
| 13 | Aircraft Category | Tekstowe | Długość: 5-12 | Kategoria statku, 79% null |
| 14 | Registration Number | Tekstowe | Długość: 3-11 | 5% null, 44305 unique |
| 15 | Make | Tekstowe | Długość: 2-30 | <1% null, 5805 unique |
| 16 | Model | Tekstowe | Długość: 1-20 | <1% null, 8556 unique |
| 17 | Amateur Built | Tekstowe | „Yes” (11%) lub „No” (88%) | 1% null, 3 unique |
| 18 | Number of Engines | Numeryczne | 0-4 | 6% null, 6 unique |
| 19 | Engine Type | Tekstowe | Długość: 4-16 | 6% null, 16 unique |
| 20 | FAR Description | Tekstowe | Długość: 7-30 | 79% null, |
| 21 | Schedule | Tekstowe | „NSCH” (5%), „SCHD” (5%), „UNK” (4%) | 86% null |
| 22 | Purpose of Flight | Tekstowe | Długość: 4-19 | 6% null, 23 unique |
| 23 | Air Carrier | Tekstowe | Długość: 3-90 | 95% null, bardzo nieprzydatne do analizy dane |
| 24 | Total Fatal Injuries | Numeryczne | 0-350 | 39% null |
| 25 | Total Serious Injuries | Numeryczne | 0-110 | 42% null |
| 26 | Total Minor Injuries | Numeryczne | 0-375 | 40% null |
| 27 | Total Uninjured | Numeryczne | 0-700 | 20% null |
| 28 | Weather Condition | Tekstowe | „IMC” (7%), „UNK” (<1%), „VMC” (89%) | 3% null |
| 29 | Broad Phase of Flight | Tekstowe | <1% „UNKNOWN”, <1% „OTHER” | 12% null, 13 unikatowych grup |
| 30 | Report Status | Tekstowe | Długość: 7-14 | 0% null |
| 31 | Publication Date | Data | 01.01.1990 – 01.01.2020 | < 1% null |

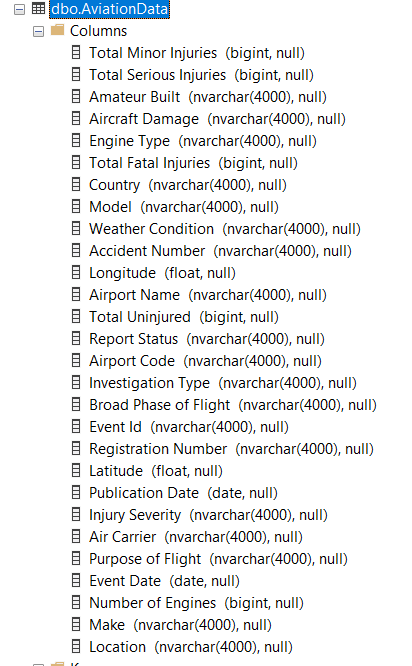


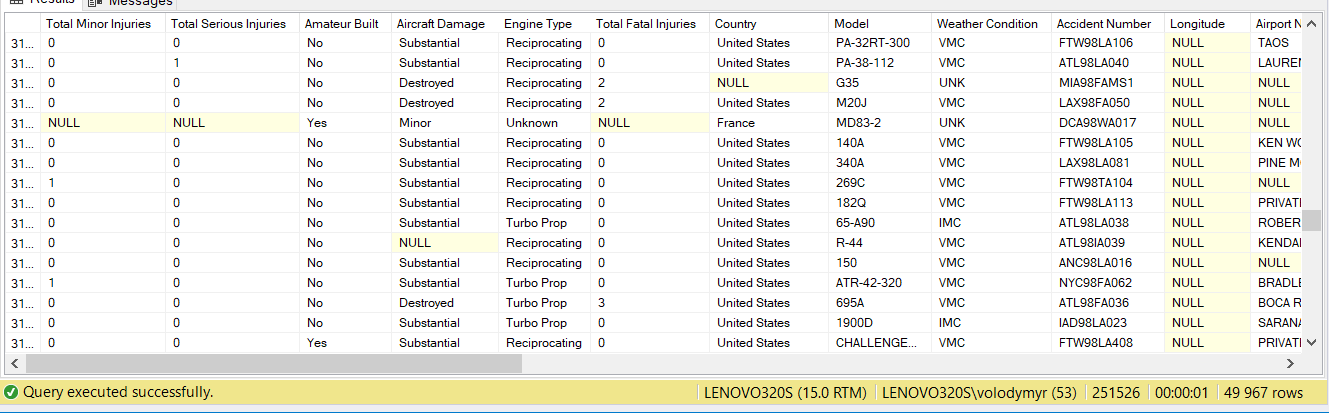


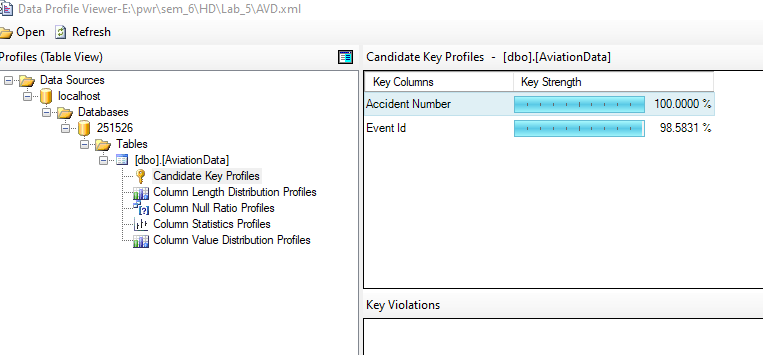


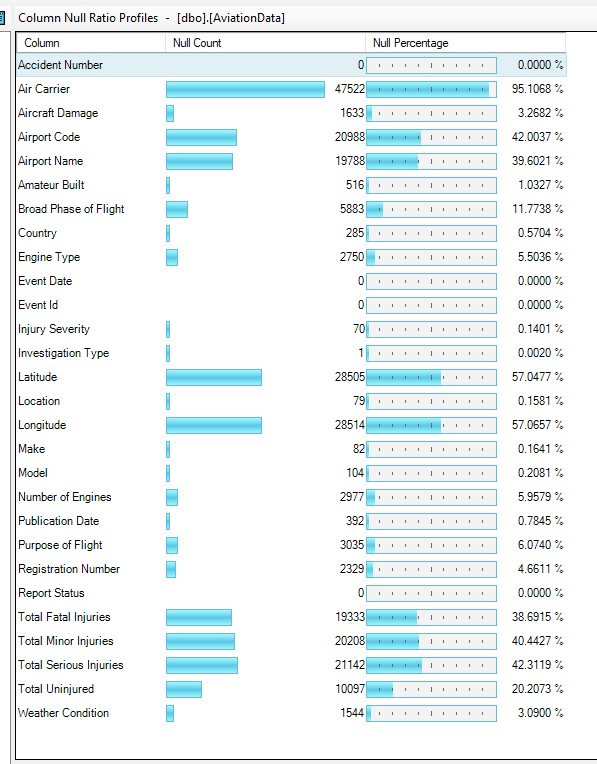


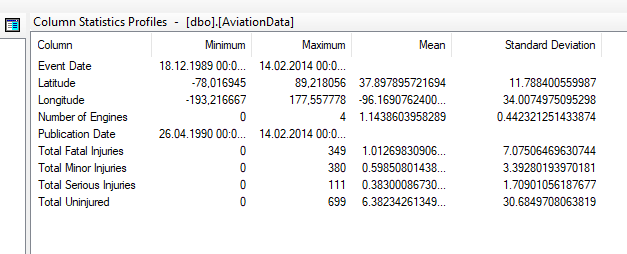


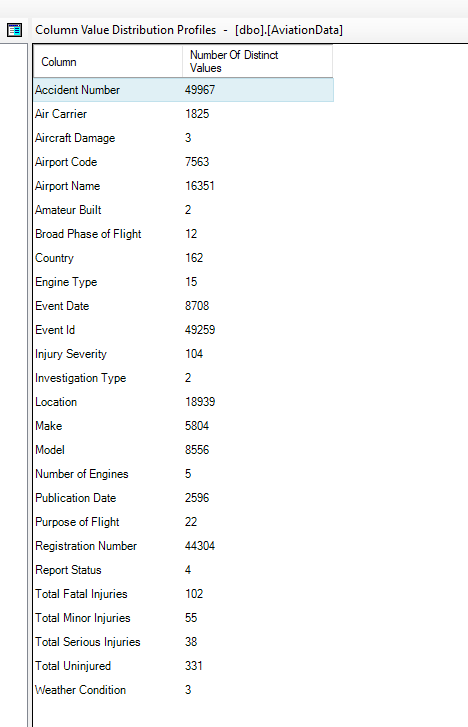












**Wnioski:**

1. Jak widać z obrazku w pierwszym zadaniu, drugi plan wymaga więcej obliczeń przez to, że posiada więcej grupowań według wskazanych parametrów.

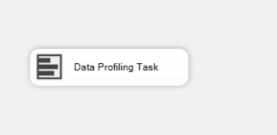
W pierwszym przypadku, nie używaliśmy OVER i PARTITION BY, dlatego mieliśmy mniej przechodzenia po danych. Użycie CTE pozwala nam zmniejszyć ilość pól, po których moglibyśmy grupować dane.

W przypadku OVER mamy dwa użycia PARTITION BY, i włożony SELECT, co też wymaga obliczeń. Ten sposób jest bardziej czytelny, ale niestety wolniejszy.

1. Program Tableau Prep Builder - Jest narzędziem służącym do przygotowania danych. Z jego pomocą dane można łączyć, czyścić, agregować, obracać, a nawet obrabiać funkcjami w kodzie R i Pythona. Łańcuch kolejnych czynności modyfikujących nasze dane nazywamy tu flow.

Założeniem Tableau Prep było to, że ma być prosty i intuicyjny. I taki jest. Program nawet sam daje rekomendacji co można i co warto zrobić z danymi. W naszym przypadku z danymi z pliku AviationData.csv mamy niektóre kolumny, które zawierają więcej, niż 50% wartości null. Jest to niedopuszczalne. Taki dane nie mogą być przeanalizowane. Dlatego, możemy usunąć niektóre z nich według rekomendacji Tableau Prep – te, które zawierają 90+% NULL. Nie możemy usuwać wszystkiego, bo stracimy wtedy ważną informację.

Także po filtracji danych możemy od razu zapisać ich do potrzebnej nam tabeli w MSSQLServer. Jest to bardzo wygodne narzędzie.



1. Program Data Profile Viewer też jest bardzo wygodny w użyciu. Przeglądanie i analizowanie profili danych to kolejny krok w procesie profilowania danych. Te profile można wyświetlić po uruchomieniu zadania profilowania danych w pakiecie Integration Services i obliczeniu profili danych.

*Uwaga:*

* Sprawozdanie bez wniosków końcowych nie będzie sprawdzane i tym samym ocena jest negatywna!

----------------------------------------------------------------------------------------------------------------------------

**Zad. 3. (Prezentacja wszystkich punktów tego zadania na zajęciach 14-16.04.2021)**

Proces tworzenia hurtowni danych powinien być poprzedzony zrozumieniem „potrzeb biznesu” oraz rzeczywistości (dziedziny problemowej) reprezentowanej przez dostępne zasoby danych. Realizacja poniższego zadania ma uzmysłowić występujące problemy w określonym (wybranym) wycinku rzeczywistości, a następnie umożliwić zidentyfikowanie (określenie) potrzeb, celu i możliwości analiz biznesowych, by wspierać procesy decyzyjne (podejmowanie właściwych decyzji biznesowych).

**Projekt HD – propozycja tematu**

Proszę przygotować zakres realizacji projektu zgodnie z poniższą specyfikacją oraz przedyskutować propozycję projektu z osobą prowadzącą zajęcia. Poczynione uzgodnienia zarejestrować w formie wniosków. Na zajęciach laboratoryjnych należy przedstawić na forum grupy swoją propozycję tematu projektu (uzasadniając celowość i jego główne elementy 1.1 – 1.6) wykorzystując PowerPoint.

**Zakres opracowania projektu HD – cz. 1.**

* 1. **Tytuł projektu**

Analiza zagranych meczów w grze komputerowej CS:GO.

* 1. **Charakterystyka dziedziny problemowej**

Counter-strike to gra FPS (First-Person Shooter), w której dwie drużyny po 5 graczy stają naprzeciw siebie w pojedynku. Standardowa gra w Counter-Strike to 30 rund do wygrania, a zwycięska drużyna jest pierwszą, która wygrywa 16 rund. 30 rund jest rozgrywanych w dwóch połowach po 15 z każdej strony mapy, z ograniczeniem czasowym rundy 1 minuta 55 sekund plus 40 sekund po podłożeniu bomby.

Wspomaganie zespołom przeanalizowania najlepszych warunków do wygrania meczu przez analizę ekonomiki wewnętrznej w grze w już zagranych meczach.

Analizując zebrane dane możemy podsumować jaka strategia dla zespołu jest najlepsza. Także możemy przewidzieć który z zespołów wygra, a który zazna porażkę. Jest to dobra opcja, jeśli chcemy zrobić stawkę w zakładzie bukmacherskim.

* 1. **Opis obszaru analizy wraz z uzasadnieniem (wybrany fragment dziedziny, przeznaczony do szczegółowej analizy i opracowania hurtowni danych)**

Dane zostały zebrane ze strony <https://www.hltv.org/>. Daty meczów zapisanych w tej bazie wahają się od 11.2015 do 03.2020.

Dane w bazie obejmują takie postaci: kraj, numer gry, wybraną mapę, ilość zabójstw i td.

* 1. **Problemy**

Analiza wcześniej zagranych meczów do zbudowania strategii zespołu do wygrania kolejnych meczów.

* 1. **Cel przedsięwzięcia** 
     1. **Oczekiwania i potrzeby w zakresie wsparcia podejmowania decyzji**

Zwiększenie szansy wygrania meczu dzięki przeanalizowaniu wcześniej rozegranych meczów (wygranych oraz przegranych) i postępowaniu graczów podczas gry.

* + 1. **Zakres analizy – badane aspekty**

Najpierw zbadajmy zespoły, który mają największy procent wygrania meczów w zależności od wybranych map. Także znajdźmy gracza, który ma większą ilość zabójstw i gracza z największym K/D (kill-death) procentem (możemy go uważać za najlepszego gracza).

Przykłady badań:

* Perspektywiczni graczy
* Bardziej łatwe mapy do wygrania meczów
* Kolejność wybierania map (żeby wygrać)
* Najbardziej popularna broń w grze.

Jest to mała część, z tego, co można przeanalizować z danego setu danych.

* 1. **Źródła danych**
     1. **Lokalizacja, format, dostępność**

**Lokalizacja:** <https://www.kaggle.com/mateusdmachado/csgo-professional-matches?select=players.csv>

**Format:** pliki .csv

**Rozmiar:** 4 pliki formatu .csv sumarycznym rozmiarem 157.36 MB

**Dostępność:** Dataset jest udostępniony pod licenzją<https://creativecommons.org/licenses/by-nc-sa/4.0/>

* + 1. **Wstępna ocena (liczba rekordów, zakres czasowy danych - faktów)**

Liczba rekordów:

* Economy.csv: 43235 rekordów
* Picks.csv: 16036 rekordów
* Players.csv: 383318 rekordów
* Results.csv: 45774 rekordów

Zakres czasowy danych: 2015 – 2020 lata

* + 1. **Fakty**
* Mecz
* Gracz
* Rezultat
* Ekonomika wewnętrzna
  + 1. **Kontekst analizy faktów np. czas (ziarnistość), lokalizacja, warunki pogodowe, itd.**

|  |  |  |
| --- | --- | --- |
| **Lp.** | **Kontekst analizy - wymiary** | **Własności** |
|  | Mecz | Numer meczu, graczy, liczba rund... |
|  | Gracz | Imię, kraj, numer ID, drużyna… |
|  | Rezultat | Data, drużyny, mapa, rezultat… |
| **4.** | Ekonomika wewnętrzna | Data, numer meczu, drużyny, mapa… |

* + 1. **Ocena jakości danych:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plik: economy.csv | | | | |
| Lp. | Atrybut | Typ wartości | Zakres wartości | Ocena jakości danych |
| 1. | date | Date | 01.01.2017 – 01.01.2021 | 0% null |
| 2. | Match\_id | Numeryczna | 2 307 500 – 2 337 500 | 0% null |
| 3 | Event\_id | Numeryczna | 750 – 5 250 | 0% null |
| 4 | Team\_1 | Tekstowa |  | 0% null |
| 5 | Team\_2 | Tekstowa |  | 0% null |
| 6 | Best\_of | Numeryczna | 1-5 | <1% null |
| 7 | \_map | Tekstowa | Długość: 4 - 11 | 0% null |
| 8 | T1\_start | Tekstowa | „ct” lub „t” | 0% null |
| 9 | T2\_start | Tekstowa | „ct” lub „t” | 0% null |
| 11 | 1\_t1 | Numeryczna | 0-100000 | 0% null |
| … | … | … | … | … |
| 42 | 30\_t1 | Numeryczna | 0-100000 | 0% null |
| 43 | 1\_t2 | Numeryczna | 0-100000 | 0% null |
| … | … | … | … | … |
| 74 | 30\_t2 | Numeryczna | 0-100000 | 0% null |
| 75 | 1\_winner | Numeryczna | 0-100000 | 0% null |
| … | … | … | … | … |
| 106 | 30\_winner | Numeryczna | 0-100000 | 0% null |

Jak widać w tej tabeli są dużo powtarzających się danych. Żeby nie zajęło to dużo pamięci możemy po prostu usunąć, według tego, że w badaniach nie będziemy potrzebowali dokładnie tych danych.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plik: results.csv | | | | |
| Lp. | Atrybut | Typ wartości | Zakres wartości | Ocena jakości danych |
| 1. | date | Date | 01.01.2015-01.01.2021 | 0% null |
| 2. | team\_1 | Tekstowa | Długość: 1-32 | 0% null |
| 3 | team\_2 | Tekstowa | Długość: 1-32 | 0% null |
| 4 | \_map | Numeryczna | Długość: 4-12 | 0% null |
| 5 | Result\_1 | Numeryczna | 0 - 40 | 0% null |
| 6 | Result\_2 | Numeryczna | 0 - 42 | <1% null |
| 7 | Map\_winner | Numeryczna | 1(54%) lub 2 (46%) | 0% null |
| 8 | Starting\_ct | Numeryczna | 1(50%) lub 2 (50%) | 0% null |
| 9 | Ct\_1 | Numeryczna | 0-16 | 0% null |
| 11 | T\_2 | Numeryczna | 0-16 | 0% null |
| 12 | T\_2 | Numeryczna | 0-16 | 0% null |
| 13 | Ct\_2 | Numeryczna | 0-16 | 0% null |
| 14 | Event\_id | Numeryczna | 750-5 250 | 0% null |
| 15 | Match\_id | Numeryczna | 2 297 500 – 2 342 500 | 0% null |
| 16 | Rank\_1 | Numeryczna | 0-400 | 0% null |
| 17 | Rank\_2 | Numeryczna | 0-400 | 0% null |
| 18 | Maps\_wins\_1 | Numeryczna | 0-36 | 0% null |
| 19 | Maps\_wins\_2 | Numeryczna | 0-36 | 0% null |
| 20 | Match\_winner | Numeryczna | 1(54%) lub 2(46%) | 0% null |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plik: players.csv | | | | |
| Lp. | Atrybut | Typ wartości | Zakres wartości | Ocena jakości danych |
| 1 | date | Date | 01.01.2015-01.01.2021 | 0% null |
| 2 | Player\_name | Tekstowa | Długość: 1-32 | 0% null |
| 3 | Team | Tekstowa | Długość: 1-32 | 0% null |
| 4 | Oponent | Tekstowa | Długość: 1-32 | 0% null |
| 5 | Country | Tekstowa | Długość: 3-32 | 0% null |
| 6 | Player\_id\_ | Numeryczna | 0-21 000 | <1% null |
| 7 | Match\_id | Numeryczna | 2 307 500 – 2 337 500 | 0% null |
| 8 | Event\_id | Numeryczna | 750-5250 | 0% null |
| 9 | Event\_name | Tekstowa | Długość: 1-128 | 0% null |
| 11 | Best\_of | Numeryczna | 1-5 | 0% null |
| 12 | Map\_1 | Tekstowa | Długość: 4-12 | 0% null |
| 13 | Map\_2 | Tekstowa | Długość: 4-12 | 49% null |
| 14 | Map\_3 | Tekstowa | Długość: 4-12 | 82% null |
| 15 | M1\_kills | Numeryczna | 0-70 | <1% null |
| 16 | M1\_assists | Numeryczna | 0-20 | <1% null |
| 17 | M1\_deaths | Numeryczna | 0-60 | <1% null |
| 18 | M1\_adr | Numeryczna | 0-190 | <1% null |
| 19 | M2\_kills | Numeryczna | 0-70 | 49% null |
| 20 | M2\_assists | Numeryczna | 0-30 | 49% null |
| 21 | M2\_deaths | Numeryczna | 0-60 | 49% null |
| 22 | M2\_adr | Numeryczna | 0-200 | 49% null |
| 23 | M3\_kills | Numeryczna | 0-60 | 82% null |
| 24 | M3\_assists | Numeryczna | 0-30 | 82% null |
| 25 | M3\_deaths | Numeryczna | 0-50 | 82% null |
| 26 | M3\_adr | Numeryczna | 0-170 | 82% null |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plik:results.csv | | | | |
| Lp. | Atrybut | Typ wartości | Zakres wartości | Ocena jakości danych |
| 1 | date | Date | 01.01.2015-01.01.2021 | 0% null |
| 2 | Team\_1 | Tekstowa | Długość: 1-32 | 0% null |
| 3 | Team\_2 | Tekstowa | Długość: 1-32 | 0% null |
| 4 | \_map | Tekstowa | Długość: 1-32 | 0% null |
| 5 | Result\_1 | Numeryczna | 0-40 | 0% null |
| 6 | Result\_2\_ | Numeryczna | 0-42 | 0% null |
| 7 | Map\_winner | Numeryczna | 1 lub 2 | 0% null |
| 8 | Starting\_ct | Numeryczna | 1 lub 2 | 0% null |
| 9 | Ct\_1 | Numeryczna | 0-16 | 0% null |
| 11 | T\_2 | Numeryczna | 0-16 | 0% null |
| 12 | T\_1 | Numeryczna | 0-16 | 0% null |
| 13 | Ct\_2 | Numeryczna | 0-16 | 0% null |
| 14 | Event\_id | Numeryczna | 750 - 5 250 | 0% null |
| 15 | Match\_id | Numeryczna | 2 297 500 – 2 342 500 | 0% null |
| 16 | Match\_winner | Numeryczna | 1 lub 2 | 0% null |