

SPRAWOZDANIE Z PROJEKTU

METODY PRZETWARZANIA DUŻEJ ILOŚCI DANYCH - PROJEKT

GRUPA NUMER 1
JAN SEIDEL [259684]
MAJA SKIBIŃSKA [259707]

21 września 2025

ANALIZA TURNIEJÓW I NAGRÓD W E-SPORCIE W POWER BI

1 Wybrana ścieżka i opis tematu

1.1 Wybrana ścieżka

Wybrano ścieżkę numer 3 związaną z PowerBi.

1.2 Cel projektu

Celem projektu jest przeanalizowanie i wizualizacja zbioru danych dotyczących e-sportu. Zostanie wykorzystana baza „eSports Earnings” z serwisu Kaggle.

1.3 Szczegóły

Szczegółowej analizie zostaną poddane:

- Które gry mają najwyższe pule nagród?
- Które drużyny i gracze zarobili najwięcej?
- Czy liczba turniejów koreluje z zarobkami?
- Analiza geograficzna – skąd pochodzą najlepsi e-sportowcy?

2 Zbiór danych

2.1 Pozyskanie

Aby przeprowadzić projekt pozyskano zbiór danych *eSports Earnings* z platformy Kaggle [1]. Autorem tego zbioru jest Jack Daoud, Solutions Engineer w firmie Snowflake i fan rozgrywek e-sportowych. Zbiór zawiera trzy pliki CSV:

- `highest_earning_players.csv` – z najlepiej zarabiającymi graczami e-sportowymi,
- `highest_earning_teams.csv` – z najlepiej zarabiającymi drużynami e-sportowymi,
- `country-and-continent-codes-list.csv` – z nazwami krajów i ich skrótowymi kodami.

W pliku dotyczącym zawodników znajdują się następujące informacje:

1. id gracza,
2. imię,
3. nazwisko,
4. pseudonim używany w grze,
5. dwuliterowy kod państwa,
6. zarobki w dolarach,
7. nazwa gry,
8. gatunek gry.

Plik dotyczący drużyn zawiera informacje wypisane poniżej:

1. id drużyny,
2. nazwa,
3. zarobki w dolarach,
4. ilość zawodów, w których drużyna wzięła udział,
5. nazwa gry,
6. gatunek gry.

W danych drużyn oraz graczy jeżeli jakiś gracz lub drużyna bierze udział w zawodach z więcej niż jednej gry, jest wymieniona kilkukrotnie. Niestety dane nie posiadają informacji o czasie, w którym dana ilość pieniędzy została zarobiona oraz dana ilość zawodów rozegrana, przez co nie będzie możliwe odniesienie się do zmian sytuacji e-sportu na przestrzeni czasu.

Ostatni plik dotyczący państw zawiera następujące informacje:

1. nazwa kontynentu,
2. dwuliterowy kod kontynentu,
3. nazwa państwa,
4. dwuliterowy kod państwa,
5. trzyliterowy kod państwa,
6. id państwa.

2.2 Przekształcenie

Przez zaimportowaniem danych do programu **PowerBI** zmieniono kodowanie znaków z UTF-8 na Windows1250 tak, aby nie było problemów ze znakami z różnych alfabetów (imiona i nazwiska zawodników) [2]. Dodatkowo zamieniono separator kolumn z przecinka na średnik, a separator dziesiętny z kropki na przecinek tak, aby dane były kompatybilne z polskojęzyczną wersją programu.

Dane zaimportowano do trzech tabeli:

- Teams,
- Players,
- Countries.

Tabelę **Players** oraz **Countries** połączono relacją wiele do jednego za pomocą kolumn z dwuliterowymi kodami państw. Następnie za pomocą funkcji **Przekształć dane** przeorganizowano każdą z tabel w sposób opisany poniżej.

Tabela Teams:

1. Zmiana typu danych ilości zarobionych pieniędzy z tekstu na liczbę dziesiętną.

Tabela Players:

1. Zmiana typu danych ilości zarobionych pieniędzy z tekstu na liczbę dziesiętną,
2. Zmiana dwuliterowego kodu państwa z małych na wielkie litery - aby wprowadzić zgodność z tabelą **Countries**,
3. Odfiltrowanie graczy nieposiadających imienia i nazwiska, a jedyne pseudonim (wartość "-" w polach imię oraz nazwisko).

Tabela Countries:

1. Usunięcie duplikujących się nazw państw (państwa leżące na granicy Azji i Europy),
2. Ukrycie wszystkich kolumn poza nazwą państwa i dwuliterowym kodem – nie będą one potrzebne w dalszych częściach projektu.

3 Wizualizacja

Zanim przygotowano raport, potrzebne było utworzenie kilku tabeli kalkulacyjnych agregujących požądane dane oraz jednej miary potrzebnej w raporcie [3].

3.1 Tabele kalkulacyjne

Tabele kalkulacyjne utworzone na potrzeby raportu w projekcie zostały opisane poniżej:

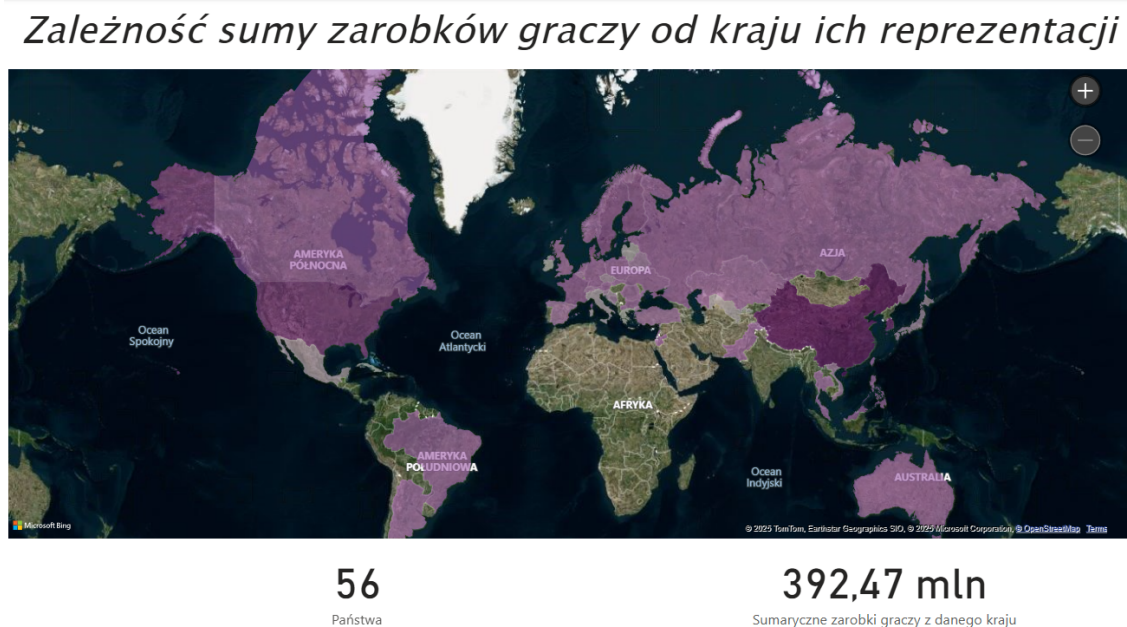
- **MeanPrizeByTotalTournaments** – średni wynik nagrody drużyny w odniesieniu do ilości turniejów odbytych przez drużynę w każdej grze. Dane pogrupowane po ilości turniejów drużyn, uśrednione wartości zarobionych pieniędzy przez drużynę, korzysta z tabeli Teams,
- **PalyersWithCountry** – tabela kalkulacyjna agregująca wszystkie dane z tabeli Players z pełnymi nazwami krajów z tabeli Countries. Dopasowanie jest możliwe poprzez użycie kolumny z dwuliterowym skrótem państwa w tabeli Players oraz Countries,
- **TeamsByPrizesAndTournaments** – sumaryczna ilość zawodów oraz zarobków każdej z drużyn w każdej grze. Grupowanie następuje po nazwie drużyny, a ilość zawodów i wysokość zarobków jest sumowana,
- **Top10PlayersPrizes** – 10 najlepszych graczy pod względem zarobków. Pogrupowano tabelę Players po pseudonimach graczy, zsumowano zarobki, posortowano malejąco po zarobkach, wybrano 10 pierwszych wierszy,
- **Top10TeamsPrizes** – 10 najlepszych drużyn pod względem zarobków. Tabela Teams została pogrupowana po nazwach drużyn, zarobki i liczba turniejów zostały zsumowane, posortowano malejąco po zarobkach i wybrano 10 pierwszych wierszy,
- **TotalPlayerPrizeByGame** – zestawienie zarobków graczy grających w daną grę. Pogrupowano tabelę Players po grach, sumując zarobki,
- **TotalPlayerPrizeByGenre** – zestawienie zarobków graczy grających w gry z danego gatunku. Tabela Players została pogrupowana po gatunkach, przy czym zsumowano zarobki graczy,
- **TotalPrizeByContry** – sumaryczne zarobki zawodników z danego państwa. Tabela Players została wzbogacona o nazwy państw (połączenie z tabelą Countries), pogrupowano po państwach, sumując zarobki.

3.2 Miary

Podczas realizacji raportu potrzebne było utworzenie miary **CountryNameCard** należącej do tabeli kalkulacyjnej **TotalPrizeByCountry**. Celem tej miary jest wypisanie liczby krajów będących w ogólnym zestawieniu jeśli żaden kraj nie został zaznaczony na mapie. Po zaznaczeniu kraju (odfiltrowaniu) zamiast liczby wyświetla się nazwa zaznaczonego państwa.

3.3 Raport

Raport składa się z czterech planszy, z czego trzy są interaktywne [4]. Pierwszą z nich jest plansza mapy przedstawiona na rysunku 1. Intensywnością koloru fioletowego przedstawiono zarobki sumaryczne graczy w zależności od państwa, które reprezentują. Pod mapą znajduje się kafelek z liczbą państw ujętych w zestawieniu oraz globalną sumą zarobków. Po kliknięciu na dany kraj, zgodnie z rysunkiem 2 kafelek pokazuje nazwę tego kraju oraz sumę zarobków graczy z tego kraju.



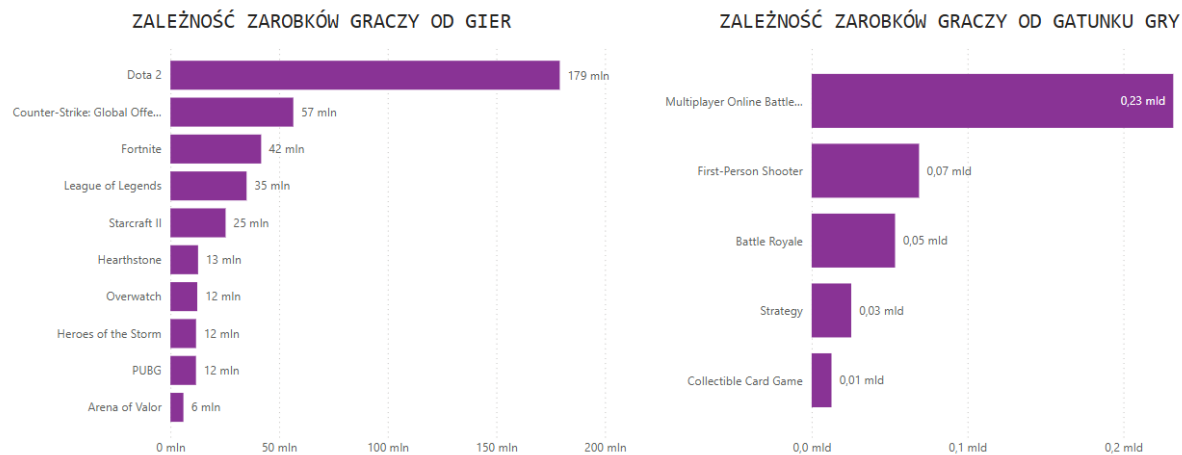
Rysunek 1: Raport - mapa



Rysunek 2: Raport - mapa z wybranym państwem

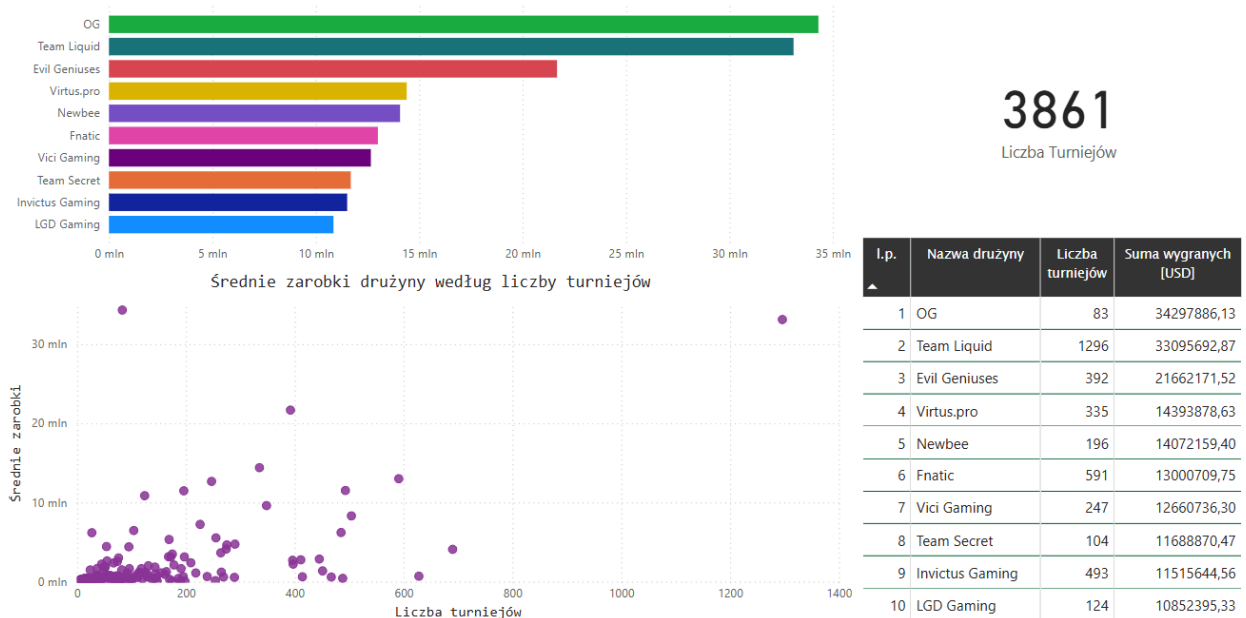
Drugą planszę przedstawiono na rysunku 3. To jedyna nieinteraktywna plansza. Znajduje się na niej zestawienie zarobków graczy w zależności od gier oraz gatunków gier.

Wpływ gier i gatunków na zarobki



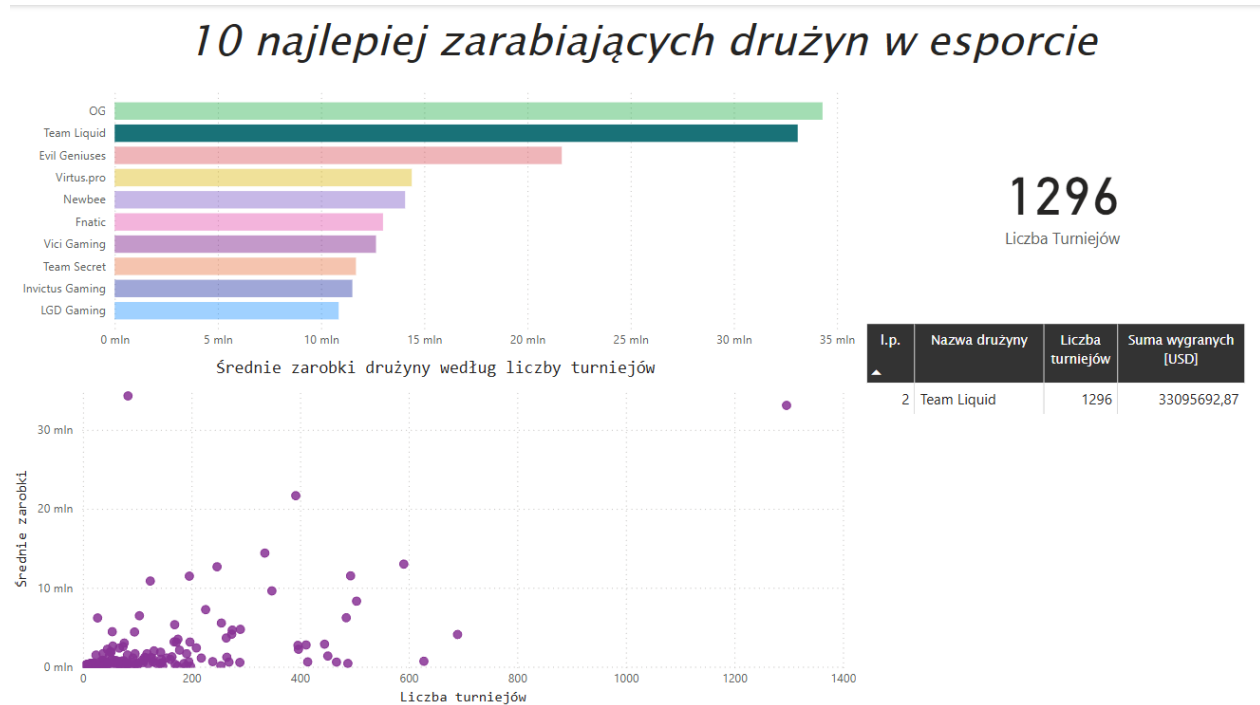
Rysunek 3: Raport - najpopularniejsze gry i gatunki gier

10 najlepiej zarabiających drużyn w esporcie



Rysunek 4: Raport - drużyny

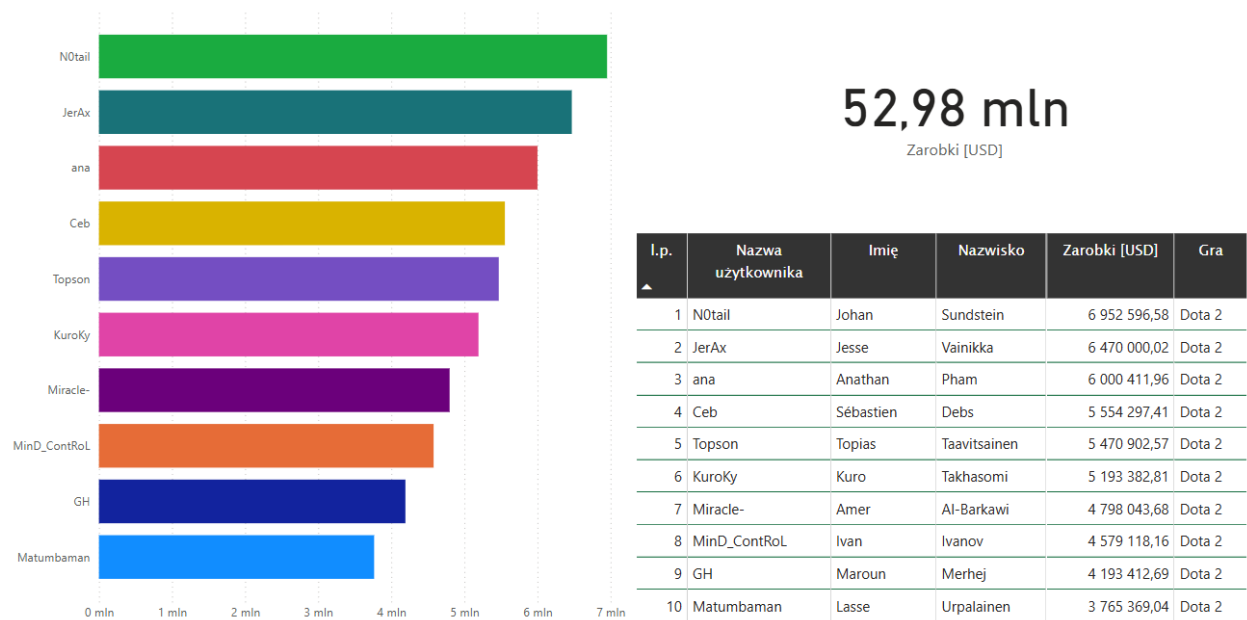
Na rysunku 4 przedstawiono trzecią planszę raportu. Znajdują się na niej dwa wykresy, kafelek i tabela. Górny wykres oraz tabela przedstawia zestawienie 10 najlepiej zarabiających drużyn. Wykres dolny przedstawia zależność średnich zarobków drużyn od ilości turniejów, jakie dana drużyna rozegrała. Kafelek pokazuje sumaryczną liczbę turniejów i zmienia swoją wartość w zależności od wybranej drużyny, co przedstawiono na rysunku 5.



Rysunek 5: Raport - drużyny po wybraniu jednej z nich

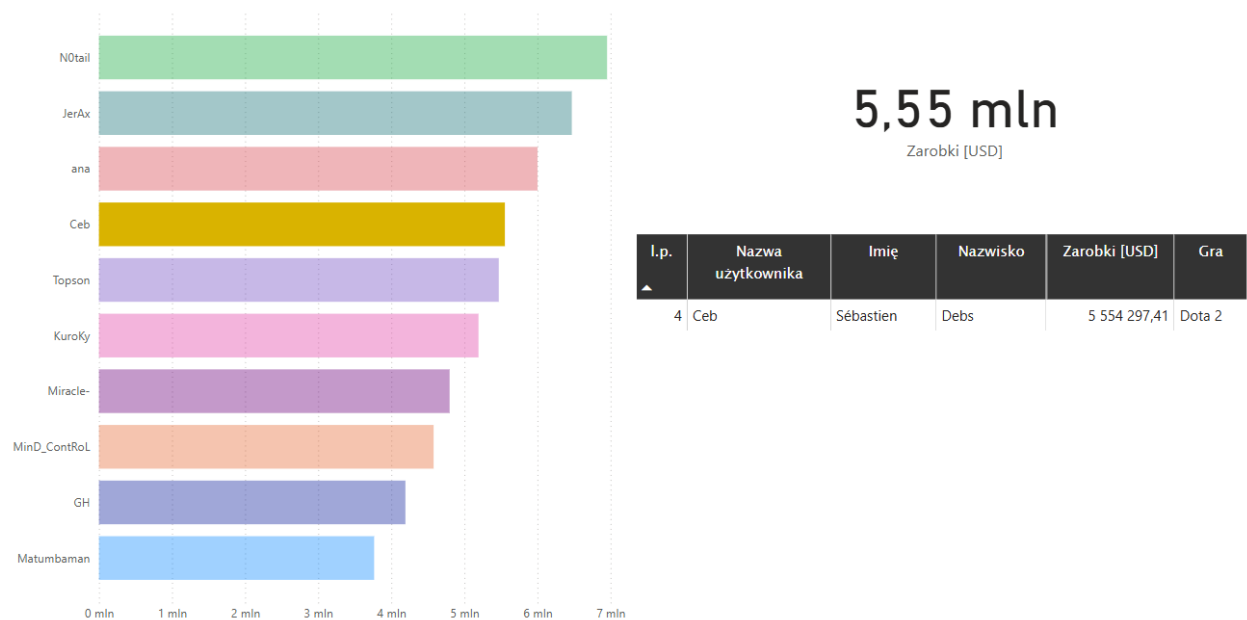
Ostatnia plansza jest analogiczna do planszy z drużynami, lecz dotyczy pojedynczych graczy. Znajduje się na niej wykres, tabela oraz interaktywny kafelek. Tym razem zamiast liczby turniejów wyświetla on sumaryczne zarobki 10 najlepszych (rys. 6) lub wybranego gracza (rys. 7).

10 najlepiej zarabiających graczy w esporcie



Rysunek 6: Raport - gracze

10 najlepiej zarabiających graczy w esporcie



Rysunek 7: Raport - gracze po wybraniu jednego z nich

4 Wyniki

Na podstawie sporządzonego raportu wnioskuje się następująco:

- rynek e-sportowy zdominowany jest przez graczy z Chin, Korei południowej oraz Stanów Zjednoczonych,
- największe zarobki dotyczą gier typu MOBA (Dota 2, League of Legends). Prawdopodobnie wynika to z faktu, że gry z tego gatunku tworzone są z myślą celowego ich użycia w e-sporcie,
- dużą popularnością cieszą się też strzelanki pierwszoosobowe (Counter Strike Global Offensive),
- gry darmowe (free to play) dominują rynek e-sportowy. Wśród gier z zestawienia, jedynie Overwatch i PUBG wymagały zakupu w czasie, z którego pochodzą dane (obecnie także są darmowe),
- drużyny *OG*, *Team Liquid* oraz *Evil Geniuses* zarobiły najwięcej pieniędzy na turniejach, znacznie więcej niż drużyny na dalszych miejscach rankingu,
- zarobki drużyny *Team Liquid* wynikają z liczby turniejów, w których drużyna wzięła udział. W przypadku *OG* oraz *Evil Geniuses* jest to kwestia posiadania najlepszych graczy,
- nie ma zależności pomiędzy ilością rozegranych turniejów a wysokością zarobków. Są drużyny, które zarobiły dużo poprzez dużą liczbę turniejów oraz takie, które zrobiły to pomimo małej liczby turniejów. Istnieje także wiele drużyn, które zarobiły mało niezależnie od liczby turniejów,
- najlepiej zarabiającym graczem okazał się Johan Sundstein grający pod pseudonimem *N0tail* w grę Dota 2, różnica pomiędzy kolejnymi miejscami nie jest już jednak aż tak duża jak w przypadku najlepszych drużyn,
- wszyscy spośród dziesięciu najlepiej zarabiających graczy grają w Dotę 2.

Literatura

- [1] Jack Daoud, “Esports earnings: <https://www.kaggle.com/datasets/jackdaoud/esports-earnings-for-players-teams-by-game/data>,”
- [2] Antdata - Data Analysis Company, “Szkolenie power bi – podstawowe informacje o power bi.” Dostęp: 18.04.2025.
- [3] J. Arnold, *Poznaj Microsoft Power BI: Przekształcanie danych we wnioski*. Warszawa: APN Promise, 2023. Tłum. Krzysztof Kapustka. Oryginalnie: *Learning Microsoft Power BI*, O'Reilly Media, 2022.
- [4] P. Socha, *Microsoft Power BI. Analiza i prezentacja danych*. Warszawa: Promise, 2022.