Autor: **Krzysztof Dąbrowski 293101** 5 czerwca 2020

POSI – Laboratorium C3

Badanie algorytmu genetycznego

# Opis zadania

Celem laboratorium jest zbadanie działania algorytmu Knutha ora algorytmu genetycznego na podstawie gry Mastermind. Dla algorytmu Knutha zostanie wykonane pojedyncze uruchomienie składające się ze 100 symulacji. Dla algorytmu genetycznego z uwagi mnogość paramentów wejściowych zostanie wykonanych wiele uruchomień, każde po 100 symulacji. Każde uruchomienie algorytmu genetycznego będzie wykonane z innymi parametrami. Umożliwi to wyznaczenie optymalnych ustawień parametrów oraz zbadanie wpływu ustawień na skuteczność działania algorytmu. Na zakończenie porównana zostanie skuteczność człowieka z działaniem dopasowanego algorytmu genetycznego.

Analiza przeprowadzonych symulacji została wykonana w języku Python przy pomocy środowiska Jupiter Notebook. Napisany w tym celu kod dostępny jest pod adresem <https://github.com/SiwyKrzysiek/POSI-C3/blob/master/Analiza%20pomiarow.ipynb>

Spis treści

[Opis zadania 1](#_Toc42256738)

[Wstęp teoretyczny 1](#_Toc42256739)

[Poszukiwanie właściwych parametrów algorytmu genetycznego 1](#_Toc42256740)

[Wpływ metody selekcji 2](#_Toc42256741)

# Wstęp teoretyczny

**Mastermind** to logiczna gra planszowa, której celem jest odgadnięcie szyfru przez jednego z graczy. Szyfr jest czteroelementowym ciągiem z alfabetu o mocy 6. Możliwych jest więc szyfrów. Gracz odgadujący przedstawia potencjalne ułożenie znaków ciągu. W odpowiedzi otrzymuje informację, ile elementów zaproponowanego ułożenia znajduje się na właściwej pozycji oraz ile jest elementami ustalonego szyfru, ale nie jest na właściwej pozycji.

**Algorytm Knutha** jest deterministycznym algorytmem pozwalającym rozwiązać dany problem w liczbie kroków zależnej jedynie od danych wejściowych.

**Algorytm genetyczny** jest niedeterministycznym algorytmem wzorującym swoje działanie na procesie selekcji naturalnej. Korzysta on z modeli osobników, krzyżowania, mutacji oraz selekcji. Dobrze nadaje się do modelowania złożonych zjawisk oraz problemów wymagających dużej mocy obliczeniowej przy zastosowaniu innych metod.

# Poszukiwanie właściwych parametrów algorytmu genetycznego

Z uwagi na mnogość parametrów wpływających na działanie algorytmu genetycznego nie łatwo jest dobrać je tak by algorytm osiągał jak najlepsze wyniki. W celu próby ustalenia optymalnych parametrów dla problemu poszukiwania rozwiązania gry Mastermind wykonane zostało po jednym uruchomieniu dla danego ustawienia parametru. Zmieniany był tylko jeden parametr na raz, pozostałe otrzymywały wartości domyślne. Takie podejście ma na celu wyodrębnienie wpływu poszczególnych parametrów.

Parametry dla poszczególnych uruchomień przyjmowały poniższe wartości.

* Metoda selekcji [Losowa, Prosta, Ruletka, Turniej]
* Prawdopodobieństwo mutacji [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]
* Prawdopodobieństwo krzyżowania [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]
* Rozmiar populacji [50, 100, 200, 300, 400, 500, 750]
* Procent osobnik w używanych do reprodukcji [25, 50, 75, 100]
* Osobniki początkowe są unikalne [TAK, NIE]

Domyślne parametry algorytmu przedstawia poniższy rysunek.

A screenshot of a cell phone

Description automatically generated

Rysunek 1 - Domyślne ustawienia algorytmu genetycznego

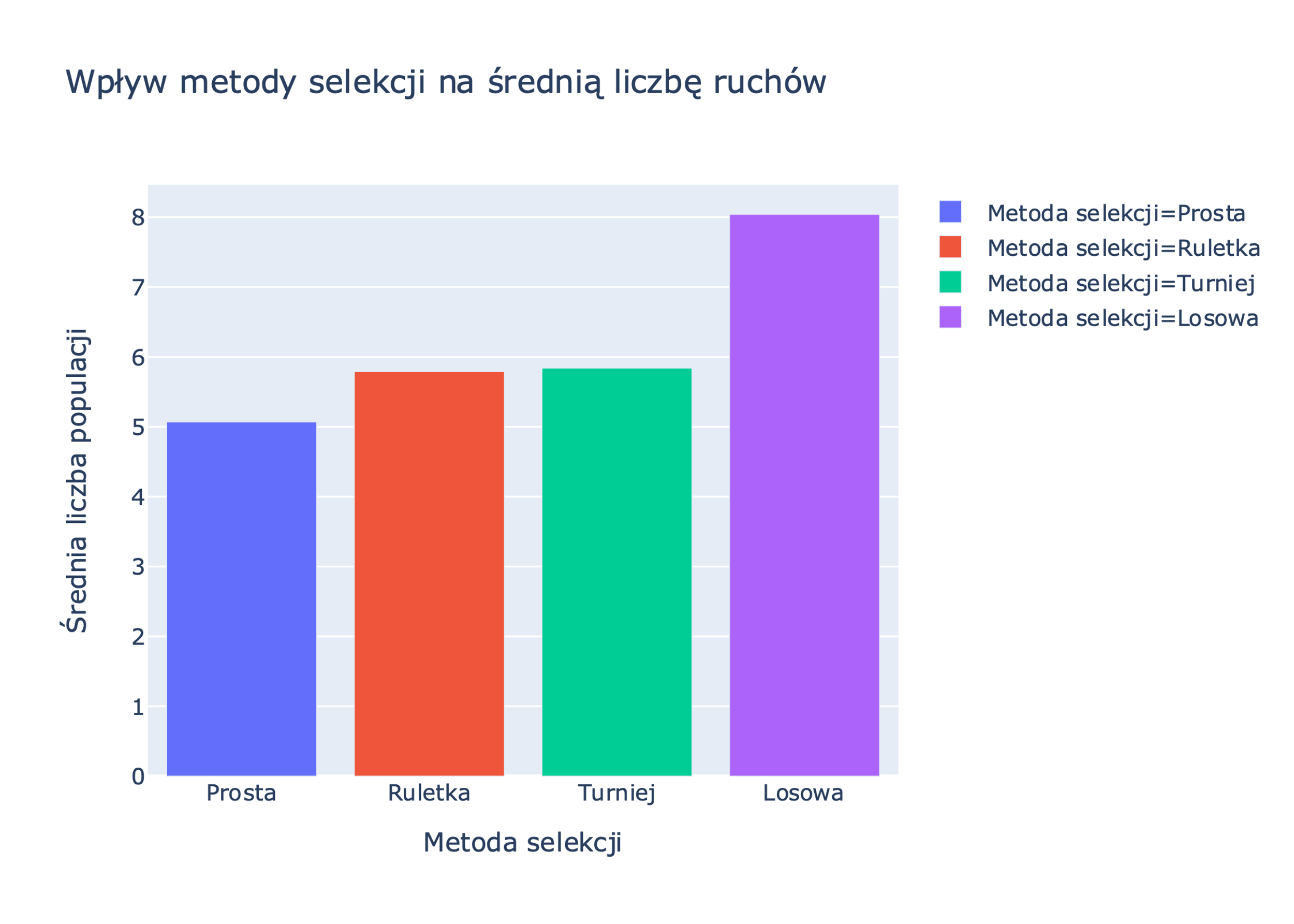
## Wpływ metody selekcji

Zbadanie uruchomień dla poszczególnych metod selekcji.

A screenshot of a cell phone

Description automatically generated

Rysunek 2 - Parametry wyników dla różnych metod selekcji



Rysunek 3 - Wpływ metody selekcji na średnią liczbę ruchów

### Wnioski

Można zaobserwować, że najlepsze wyniki pozwala osiągnąć metoda prosta. Nieco gorsze są metody ruletki i metoda turniejowa. Metoda losowa jest zdecydowanie złym wyborem. Może to wynikać z tego, że nie pozwala ona na preferowanie osobników lepiej przystosowanych.

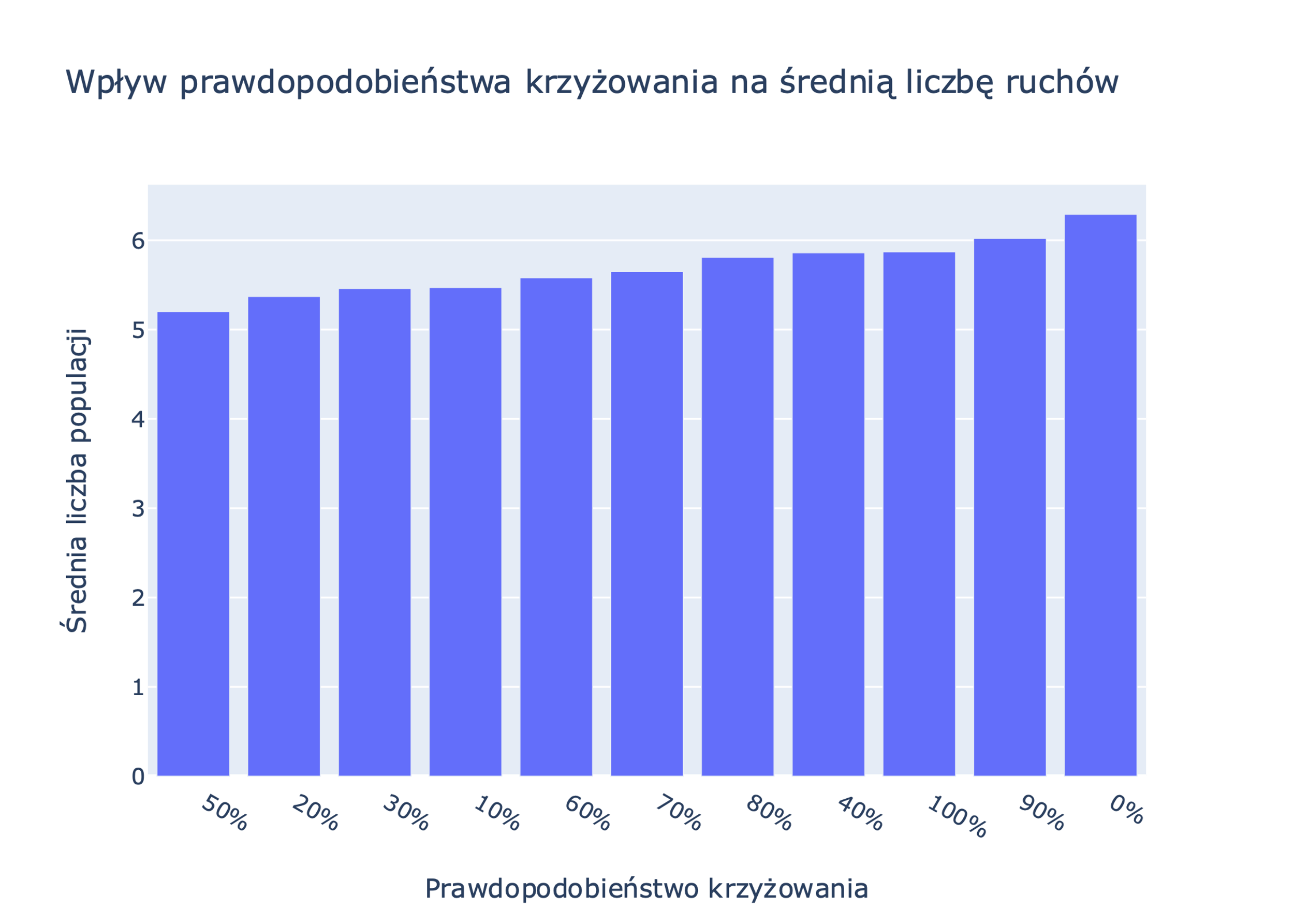
## Wpływ prawdopodobieństwa krzyżowania

Zbadanie uruchomień dla różnych wartości prawdopodobieństwa krzyżowania.

A screenshot of a social media post

Description automatically generated

Rysunek 4 - Parametry wyników dla różnego prawdopodobieństwa krzyżowania



Rysunek 5 - Wpływ prawdopodobieństwa krzyżowania na średnią liczbę ruchów

### Wnioski

Najlepsze wyniki pozwala osiągnąć prawdopodobieństwo krzyżowania na poziomie 50%. Podobne wyniki dają niskie wartości parametru, jednak zerowe prawdopodobieństwo mocno zaburza działanie algorytmu.

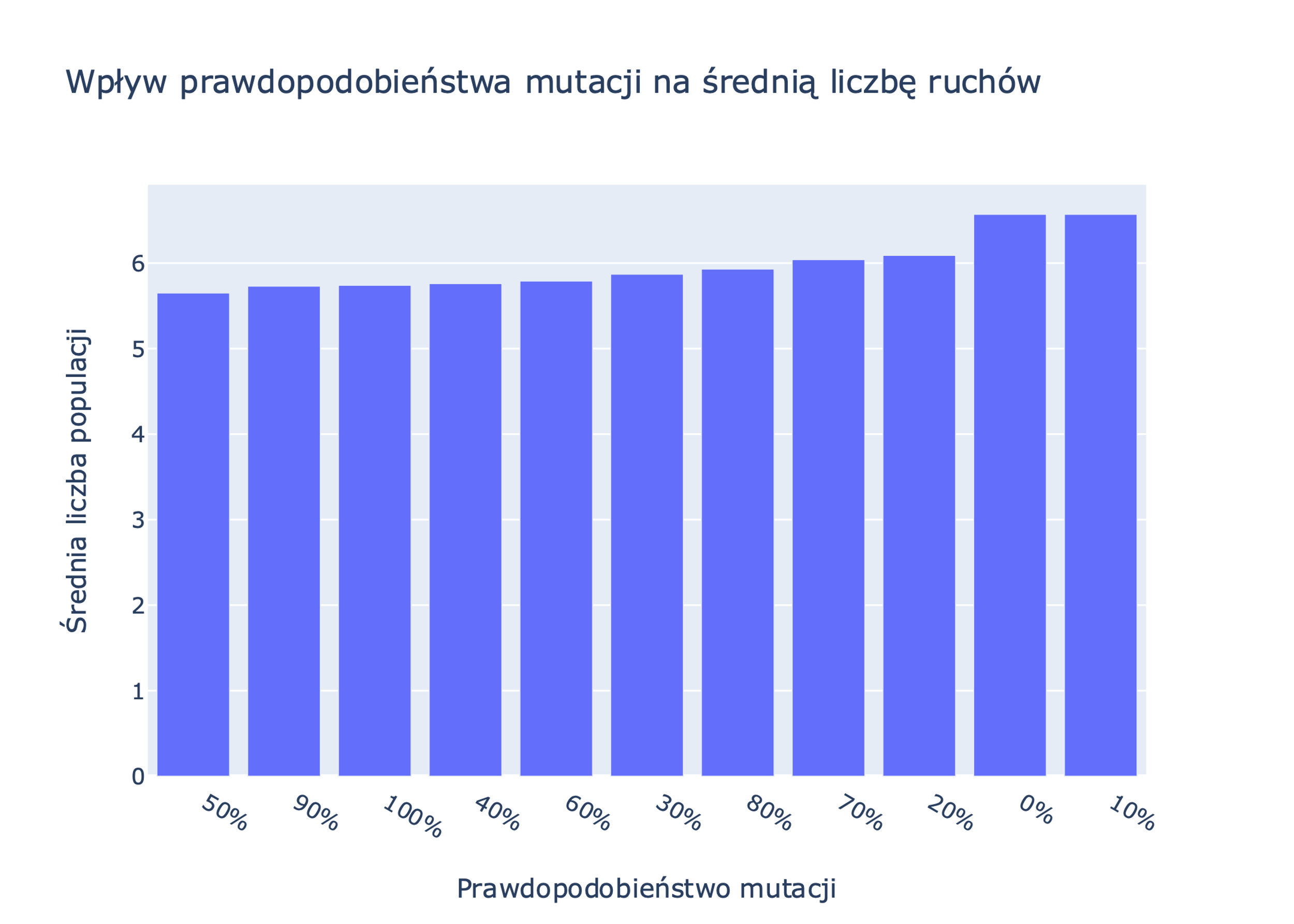
## Wpływ prawdopodobieństwa mutacji

Zbadanie uruchomień dla różnych wartości parametru prawdopodobieństwa mutacji.

A screenshot of a social media post

Description automatically generated

Rysunek 6 - Parametry wyników dla różnych wartości prawdopodobieństwa mutacji



Rysunek 7 - Wpływ prawdopodobieństwa mutacji na średnią liczbę ruchów

### Wnioski

Optymalną wartością prawdopodobieństwa mutacji jest 50%. Algorytm działa dobrze również dla skrajnie wysokich wartości. Należy unikać skrajnie małych wartości parametru, ponieważ dają one wyraźnie gorsze wyniki niż pozostałe ustawienia.

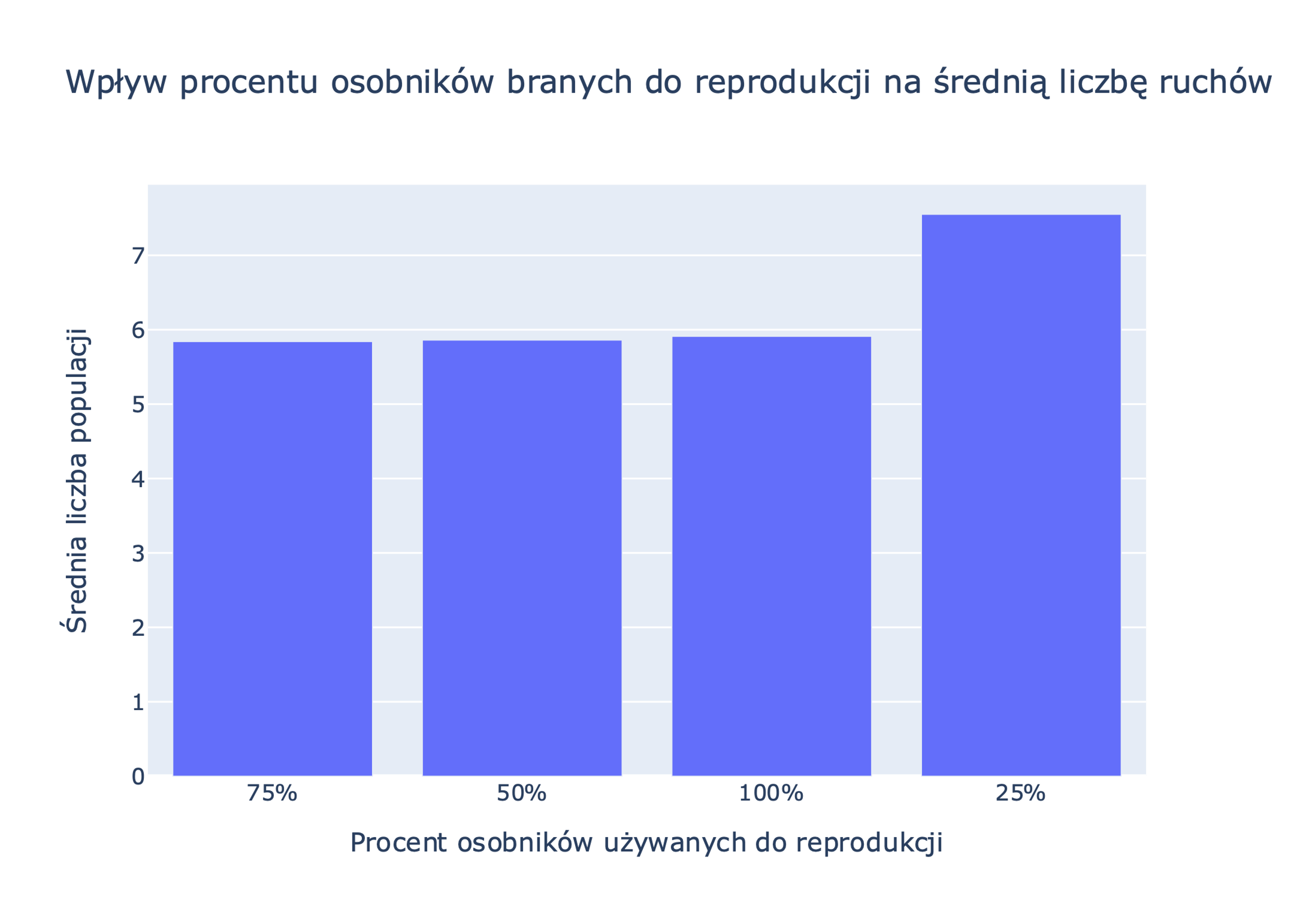
## Wpływ procentu osobników branych do reprodukcji

Zbadanie uruchomień dla różnych wartości procentu osobników branych od reprodukcji.

A screenshot of a cell phone

Description automatically generated

Rysunek 8 - Parametry wnyków dla różnych wartości procentu osobników wziętych do reprodukcji



Rysunek 9 - Wpływ procentu osobników branych do reprodukcji na średnią liczbę ruchów

### Wnioski

Przy ustalaniu wartości procentu osobników branych do reprodukcji kluczowe jest zadbanie o to by była to znacząca część populacji. Wartość 25% jako jedyna daje wyraźnie gorsze wyniki. Może to wynikać z pominięcia osobników o wartościowym genotypie.

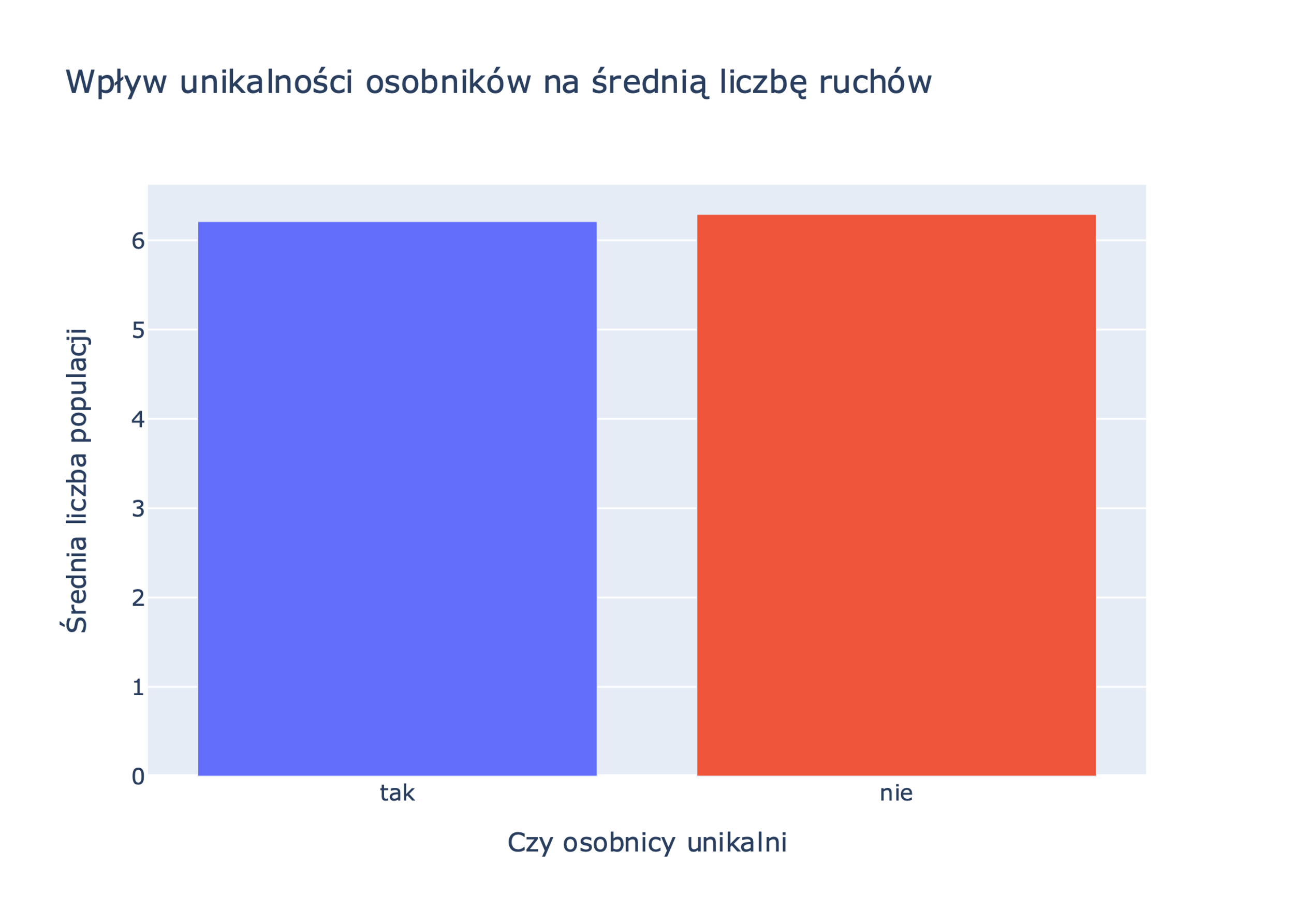
## Wpływ unikalności osobników

Zbadanie uruchomień dla unikalnych oraz nieunikalnych osobników początkowych.

A screenshot of a cell phone

Description automatically generated

Rysunek 10 - Parametry wyników dla różnej unikalności osobników startowych



Rysunek 11 - Wpływ unikalności osobników na średnią liczbę ruchów

A screenshot of a cell phone

Description automatically generated

Rysunek 12 - Wpływ unikalności osobników na średnią liczbę ruchów. Przybliżenie

### Wnioski

Unikalność osobników ma znikomy wpływ na działanie algorytmu dla zadanego problemu. Podczas analizy danych doszedłem jednak do wniosku, że łatwo można uzyskać iluzję dużego wpływu stosując skalę przybliżenie na dany zakres. Taktyka ta może zostać użyta do próby manipulacji na nieuważnym odbiorcy.

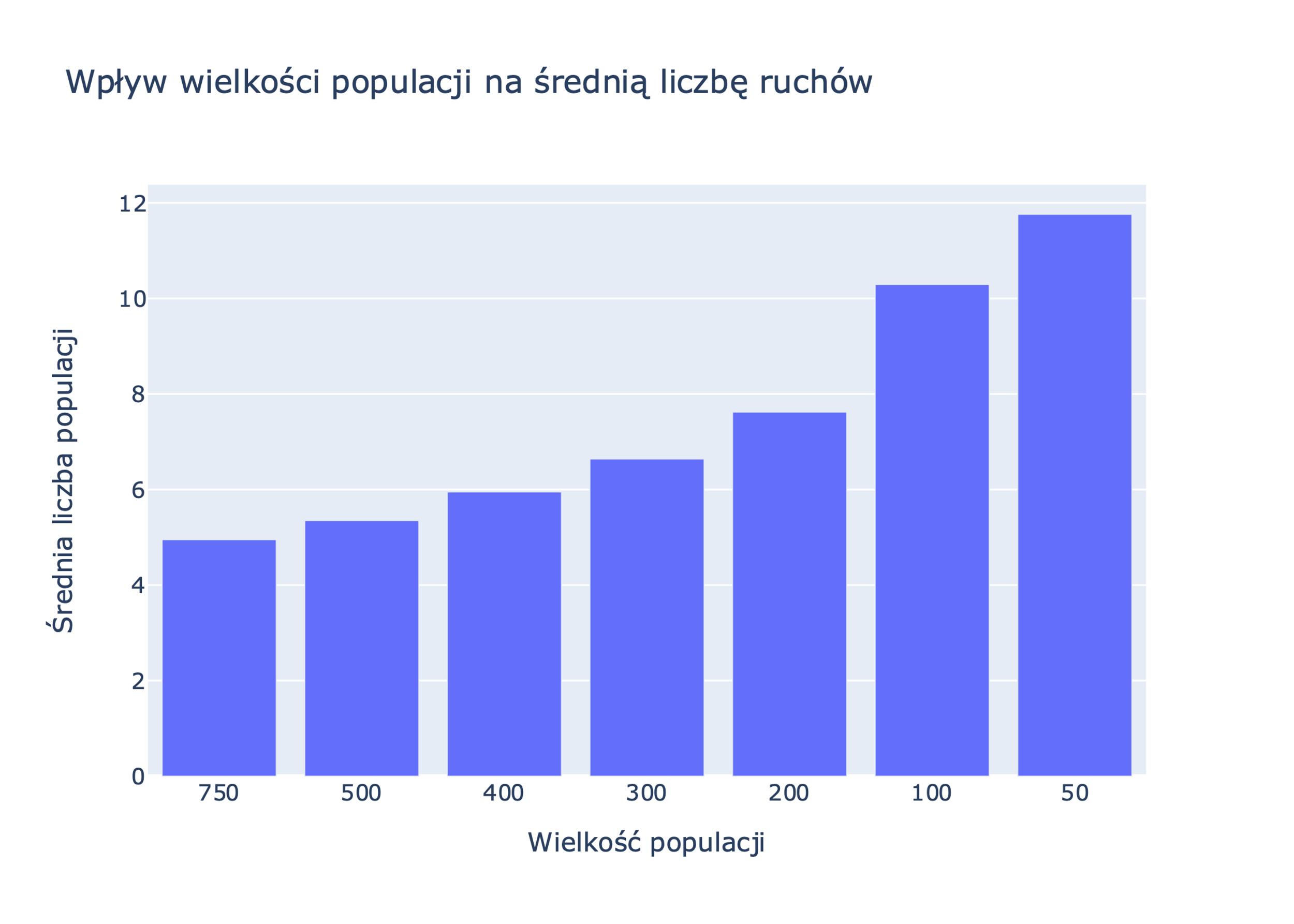
## Wpływ wielkości populacji

Zbadanie uruchomień dla różnej wielkości populacji.

A screenshot of a cell phone

Description automatically generated

Rysunek 13 - Parametry wnyków dla różnych wielkości populacji



Rysunek 14 - Wpływ wielkości populacji na średnią liczbę ruchów

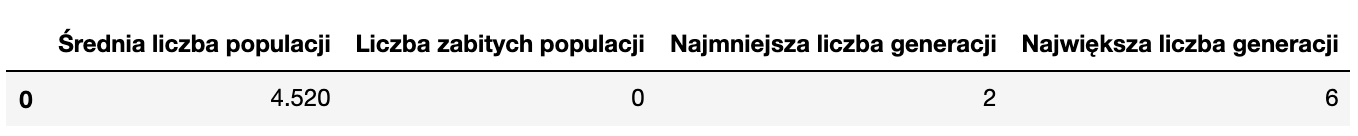
### Wnioski

Większa populacja pozwala na osiągniecie lepszych wyników. Zależność ta nie jest jednak liniowa. Wzrost wielkości populacji ma największy wpływ do wartości w okolicy 200 osobników.

# Optymalne parametry algorytmu genetycznego

Na podstawie przeprowadzonej analizy dobrane zostały optymalne wartości parametrów algorytmu genetycznego dla gry Mastermind.

* Metoda selekcji: **Prosta**
* Prawdopodobieństwo mutacji: **0.5**
* Prawdopodobieństwo krzyżowania: **0.5**
* Rozmiar populacji: **750**
* Procent osobnik w używanych do reprodukcji: **75%**
* Osobniki początkowe są̨ unikalne: **TAK**



Rysunek 15 - Wyniki dla optymalnych parametrów

Połączenie optymalnych wartości dla poszczególnych parametrów pozwoliło na osiągnięcie najmniejszej średniej liczby populacji.