Celem naszego projektu jest analiza wyników corocznej ankiety StackOverflow, w celu zidentyfikowania trendów, preferencji i kluczowych informacji dotyczących społeczności programistów na całym świecie. Dzięki uzyskanym i zwizualizowanym przez nas informacjom, w łatwy sposób można sprawdzić np. którzy programiści zarabiają najwięcej, jak kształtuję się struktura zajmowanych stanowisk w zależności od posiadanego wykształcenia lub które języki programowania są najpopularniejsze. Wierzymy, że nasz projekt umożliwia ocenę sytuacji na szeroko rozumianym rynku programistycznym, a co za tym idzie, podejmowanie decyzji przez osoby, które już się na nim znajdują lub dopiero starają się na niego wejść.

Wybranym przez nas źródłem danych jest plik CSV z wynikami ankiety zrealizowanej w 2023 roku.   
Plik ten jest dostępny pod adresem: <https://insights.stackoverflow.com/survey>.  
Po pobraniu należy rozpakować plik ZIP, dane znajdować się będą w pliku o nazwie *survey\_results\_public.csv*, ponadto całą ankietę można wygodnie przeglądać w pliku o nazwie *so\_survey\_2023.pdf*.

Trzeba pamiętać, że dane zebrane w tej ankiecie mogą nie być do końca reprezentatywne dla całej grupy zawodowej programistów, gdyż zostały one zebrane za pośrednictwem pojedynczej strony, a więc mogą w nich wystąpić pewne skrzywienia. Grupy zawodowe które z jakiś przyczyn rzadziej korzystają z tej strony mogą być niedoreprezentowane. Jednak pamiętając o tym, liczba zebranych danych oraz popularność StackOverflow jest na tyle duża, że można założyć że dane te są wystarczająco miarodajne.

Narzędzia wykorzystane przez nas do analizy i wizualizacji danych to *NumPy*, *Pandas* oraz *Matplotlib*.

Wynikiem naszej pracy są liczne i przemyślane wykresy, które dotyczą głównie:

* popularności języków programowania z podziałem na grupy wiekowe i sektory gospodarki
* zarobków z podziałem na grupy wiekowe i sektory gospodarki
* wpływu wykształcenia oraz ilości doświadczenia na zajmowane stanowisko

Dane nie pokazują nic zaskakującego, jednak ciekawym trendem jest wzrost znajomości mniej popularnych języków i technologii z wiekiem (w tym SQL, który staje się najpopularniejszy powyżej 55r. życia), przy jednoczesnym spadku JS. Wyraźnie jednak widać, że najpopularniejsze są technologie front-endowe, zarówno w najlepszych branżach, jak i największych firmach. Natomiast procentowe użycie JS spada wraz ze wzrostem wielkości firmy, mimo tego że lepiej się w takich zarabia.

Dane te sugerują więc, że większe firmy mają bardziej zróżnicowane stosy technologiczne i ich zapotrzebowanie na mniej popularne technologie jest większe. Patrząc jednocześnie na najstarsze osoby (które najlepiej zarabiają) znające bardzo zróżnicowane technologie, można wyciągnąć wniosek, że znajomość różnych mniej popularnych technologii może dalej dawać możliwości na znalezienie dobrze płatnej pracy.

Uważamy, iż zrealizowany przez nas projekt spełnia swoje cele przedstawione we wstępie. Wybrane przez nas narzędzia pozwoliły na realizację zadań na drodze do sukcesu. Dzięki pracy nad projektem pogłębiliśmy i utrwaliliśmy naszą wiedzę i umiejętności z zakresu programowania w języku Python, komputerowej analizy danych oraz matematyki.

Kolejnymi krokami rozwoju projektu, mogłoby być:

* zwiększenie atrakcyjności wizualnej wykresów
* opracowanie nowych wykresów dot. przede wszystkim:
* pracy zdalnej / hybrydowej / stacjonarnej
* wpływu miejsca zamieszkania i używanych technologii na zarobki
* wpływu stanowiska na zarobki
* wpływu wybranej technologii na zarobki
* w jakiej technologii/branży doświadczenie ma największy wpływ na zarobki
* jakie technologie najczęściej są wybierane wspólnie, oraz jakie to ma przełożenie na zarobki
* użycia AI w zależności od sektora gospodarki i zajmowanego stanowiska
* wpływu użycia AI na zarobki

Podział Pracy:

* Mikołaj Stryczek:
* stworzenie wykresów dot. wpływu wykształcenia i doświadczenia na zajmowane stanowisko
* ulepszanie wykresów dot. zarobków
* opracowanie dokumentacji
* Rafał Fatuła:
* Stworzenie wykresów dotyczących zarobków oraz popularności danych języków programowania
* Opracowanie dokumentacji
* Piotr Stalmach:
* Przygotowanie pliku requirements
* Automatyczne pobieranie i przygotowanie pliku csv
* Customowe sortowanie danych i refactor
* Stworzenie wykresu mesh