

<u>234073</u>
<i>Numer indeksu</i>
<u>Aleksandra Kowalczyk</u>
<i>Imię i nazwisko</i>

<u>234102</u>
<i>Numer indeksu</i>
<u>Zbigniew Nowacki</u>
<i>Imię i nazwisko</i>

<u>234106</u>
<i>Numer indeksu</i>
<u>Karol Podlewski</u>
<i>Imię i nazwisko</i>

<b>Kierunek</b>	Informatyka Stosowana
<b>Stopień</b>	II
<b>Specjalizacja</b>	Data Science
<b>Semestr</b>	2

**Data oddania** 16 grudnia 2020

# Przetwarzanie i analiza dużych zbiorów danych

## Prezentacja wstępna projektu

## Spis treści

1	Nazwa projektu	3
2	Członkowie grupy	3
3	Motywacja	3
4	Cel naukowy oraz dodatkowe cele projektu	3
5	Opis zbiorów danych	3
6	Metody Analizy Danych	4
7	Stos technologiczny	4

# 1 Nazwa projektu

Projekt jaki zamierzamy wykonać został zatytułowany **Analiza danych przestępczych w największych miastach Stanów Zjednoczonych**.

# 2 Członkowie grupy

Skład grupy:

- Aleksandra Kowalczyk, 234073
- Zbigniew Nowacki, 234102
- Karol Podlewski, 234106 (lider)

# 3 Motywacja

Motywacją do podjęcia takiej tematyki projektu jest chęć sprawdzenia czy i jakie różnice istnieją w przestępczości pomiędzy badanymi przez nas regionami. Przeprowadzona analiza może przyczynić się do dostrzeżenia przyczyn takich zająć, które wcześniej były niezauważane, bądź pozostawały w sferze domysłów.

# 4 Cel naukowy oraz dodatkowe cele projektu

Celem projektu jest przeprowadzenie badania mającego na celu zwiększenie efektywności działań prewencyjnych służb mundurowych oraz efektywniejszego dysponowania środkami na te służby. Wykonanie projektu dodatkowo pozwoli nam na lepsze zrozumienie działania algorytmów oraz wzrost umiejętności pracy z dużymi zbiorami danych, które są wykorzystywane w rzeczywistych warunkach. Projekt powinien się zakończyć wyciągnięciem i przedstawieniem uzyskanych wniosków pod koniec semestru.

# 5 Opis zbiorów danych

Dane dotyczące przestępstw będziemy pobierać z otwartych repozytoriów danych miast (między innymi [NYC Open Data](#) czy [Los Angeles Open Data](#)).

W przypadku chęci rozszerzenia projektu o powiązanie danych przestępczych z danymi demograficznymi, skorzystamy z informacji udostępnianych przez jednostki czy urzędy zbierające dane tego rodzaju (na przykład [Census Bureau](#)).

## 6 Metody Analizy Danych

Pierwszym krokiem będzie przygotowanie danych do ich zbiorczej analizy – wyczyszczenie danych, a także przetworzenie kolumn w sposób umożliwiający nam agregację zbiorów. Następnie skupimy się na analizie, która pozwoli na wyłuskanie istotnych informacji, w tym tych, które nie są widoczne na pierwszy rzut oka – przydatne w tym celu mogą się okazać między innymi reguły asocjacyjne, analiza skupień, klasyfikacja czy analiza statystyczna.

## 7 Stos technologiczny

- Python 3
- Spark 3