

# Natalia Domańska

Inżynier danych PySpark

✉ natalia.domanska@example.pl ☎ +48 500 264 652 📍 Białystok

## Podsumowanie zawodowe

Inżynierka PySpark w projektach zdrowotnych. Przygotowuje bezpieczne, powtarzalne zbiory z rejestrów i badań, dokumentując pochodzenie danych i reguły jakości.

## Doświadczenie zawodowe

### Inżynier danych PySpark

Laboratorium Zdrowych Danych

Białystok  
2022-01 - Obecny

- Przekształciła 680 mln rekordów klinicznych w PySpark z kontrolą pseudonimizacji; 3 zespoły badawcze otrzymywały kohortę w 34 minuty zamiast 3,5 godziny.
- Zoptymalizowała partycje Delta Lake dla 12 lat wyników laboratoryjnych; odczyt kohorty skrócił się o 61% na 26 powtarzalnych zapytaniach.
- Dodała Great Expectations do 47 pól krytycznych dla dawkowania; liczba rekordów kierowanych do ręcznej weryfikacji spadła z 2,8 do 0,6%.
- Wdrożyła odtwarzalne snapshoty dla 5 badań z biostatystykami; dwa niezależne przebiegi zgadzały się w 100% na 18 mln wierszy.

### Inżynier platformy danych

Podlaskie Badania Kliniczne

Białystok  
2018-04 - 2021-12

- Przepisała 16 zadań Pandas do PySpark dla rejestru 9 województw; miesięczne zestawienie zamykało się w 42 minuty zamiast 4 godzin.
- Uzgodniła słownik 230 kodów rozpoznań z 6 epidemiologami; odsetek rekordów bez poprawnej kategorii spadł z 11 do 2,4%.
- Przeniosła 3,1 TB plików Parquet do strefy szyfrowanej i sprawdziła sumy kontrolne; 99,99% obiektów przeszło kontrolę po migracji.
- Dodała lineage do 31 tabel pośrednich wraz z inspektorem danych; przygotowanie ścieżki audytowej skróciło się z 2 dni do 3 godzin.

## Projekty

Przyrostowe przetwarzanie rejestrów

Produkt danych dla obszaru: dane badań klinicznych i rejestry. Obejmuje kontrolowane źródła, wersjonowane transformacje i sprawdzalne reguły jakości.

PySpark

Delta Lake

Parquet

Great Expectations

## WYKSZTAŁCENIE

### Magister, biostatystyka

Uniwersytet Medyczny w Białymstoku

dane badań klinicznych i rejestry  
2014-10 - 2018-06

## UMIEJĘTNOŚCI

Platformy i języki

- PySpark
- Delta Lake
- Parquet
- Great Expectations

- Azure Databricks

#### Jakość i utrzymanie

- Modelowanie danych
- Testy
- CI/CD
- Monitoring
- Dokumentacja

#### Współpraca

- Ustalanie wymagań
- Przegląd kodu
- Obsługa incydentów
- Dostarczanie między zespołami

## CERTYFIKATY

### Databricks Certified Associate Developer for Apache Spark

Organizacja certyfikująca  
2025-03

### Warsztat jakości danych: dane badań klinicznych i rejestry

Polskie Forum Inżynierii Danych  
2023-10

## JĘZYKI

Polski  
Angielski

Język ojczysty  
Poziom zawodowy