

Łukasz Prokulski

Jak zostać analitykiem danych? (skrót)



Spis treści

Główne założenia

Dlaczego warto?

Struktura przewodnika

Fundamenty (Excel i SQL)

Rozdział: Excel - król korporacji

Rozdział: SQL - uniwersalny język danych

Kompetencje techniczne i analityczne

Statystyka - czasem trzeba znać teorię

Python - język (programowania) danych

Architektura rozwiązań - zupełne podstawy

Ścieżki kariery

Umiejętności Miękkie i Etyka Danych

AI - Pomoc czy Przeszkoda?

Twoja kariera zaczyna się TERAZ

Jesteś gotów, aby zacząć – daję Ci kompas

Pozbądź się chaosu – poznaj fundamenty

Autor

Główne założenia

Pełna wersja przewodnika przedstawia kompleksową ścieżkę rozwoju od absolutnych podstaw do profesjonalnej kariery w analizie danych.

Traktuj go jak mapę – to lista rzeczy do nauczenia się, poszukania i zgłębienia na własną rękę, a także **ściągawka na Twojej drodze rozwoju** i osobisty plan kariery.

Dlaczego warto?

1. **Dane to nowa ropa (*data is new oil*).**
2. Zawody związane z przetwarzaniem i analizą danych liczbowych oferują fascynujące możliwości.
3. “Branża danych” przeżywa ciągle dynamiczny rozwój, z **prognozowanym wzrostem liczby miejsc pracy o 35% do 2032 roku.**
4. Możesz zacząć karierę nawet bez doświadczenia w Excelu czy programowaniu, poprzez systematyczne nabywanie umiejętności i praktyki.

Struktura przewodnika

Książka nie jest kursem – **jest kompasem**. Możesz sięgać po wybrane rozdziały, w zależności od tego, co już znasz lub co Cię interesuje. Przewodnik pomoże Ci zdobyć wiedzę, aby pracować z danymi na różnych stanowiskach, w tym jako:

- **Analityk BI (*BI analyst*):** Specjalista od wykresów i dashboardów.
- **Data Scientist:** Twórca modeli, które przewidują zdarzenia lub klasyfikują dane (np. czy klient spłaci kredyt).
- **Inżynier Danych (*Data Engineer*):** Osoba dostarczająca dane i dbająca o to, by były uporządkowane i działały poprawnie – często pracująca z **big data**.

Z dobrym planem, konsekwentnym wysiłkiem i zamiłowaniem do uczenia się możesz zbudować satysfakcjonującą karierę w świecie analizy danych, nawet zaczynając od zera.

Fundamenty (Excel i SQL)

Pracę z danymi należy zacząć od solidnych podstaw.

Rozdział: Excel - król korporacji

Excel to program-kombajn, program-legenda, bez którego żadna korporacja nie potrafi istnieć. Trzeba go więc znać. Znajomość Excela, nawet jeśli nie chcesz z nim pracować, jest **niezbędnym minimum**. Plan nauki dzieli się na 7 etapów, obejmujących:

- **Podstawy (Etap 1):** Nawigacja po programie, typy danych (Tekst, Liczby, Daty), adresowanie komórek (względne, bezwzględne).
- **Formatowanie i Organizacja (Etap 2):** Dbanie o czytelność tabel (kolory, obramowania, zamrażanie okienek, sortowanie i filtrowanie).
- **Podstawowe Formuły (Etap 3):** Użycie funkcji matematycznych (SUMA, ŚREDNIA), logicznych (JEŻELI) i podstaw wyszukiwania (WYSZUKAJ.PIONOWO). **Nauka formuł rozwija myślenie algorytmiczne.**
- **Wizualizacja (Etap 4):** Tworzenie podstawowych wykresów (kolumnowe, liniowe, kołowe), ponieważ **jeden obraz to milion słów.**
- **Zaawansowane Narzędzia (Etap 5):** Tabele przestawne (do podsumowywania i analizowania dużych zbiorów danych) i zaawansowane formatowanie warunkowe (skale kolorów, paski danych).



Rozdział: SQL - uniwersalny język danych

SQL (Structured Query Language) to język niezbędny do zarządzania i przetwarzania danych w relacyjnych bazach danych. To najpotężniejszy język do analizy danych. **Nie znać SQL (nawet samych podstaw) w świecie danych to jak nie znać angielskiego w świecie IT.** Kluczowe etapy nauki:

- **Podstawy (Etapy 1 i 2):** Zrozumienie relacyjnych baz danych (tabele, kolumny, klucze główne/obce) i odczytywanie danych za pomocą **SELECT, FROM, WHERE** (filtrowanie) oraz **ORDER BY** (sortowanie).

- **Funkcje i Agregacje (Etap 3):** Zliczanie, sumowanie i uśrednianie danych za pomocą **COUNT, SUM, AVG** oraz grupowanie wyników za pomocą **GROUP BY** i filtrowanie grup za pomocą **HAVING**.
- **Łączenie Tabel (Etap 4):** Użycie **JOIN (INNER, LEFT, RIGHT)** do łączenia danych rozdzielonych na wiele tabel.
- **Zaawansowane Zapytania (Etap 5):** Podzapytania (w **WHERE, SELECT, FROM**), wyrażenia warunkowe **CASE WHEN** oraz zaawansowane **Funkcje Okna** (np. do rankingowania).
- **Modyfikowanie Danych (Etap 6):** Dodawanie (**INSERT INTO**), aktualizowanie (**UPDATE**) i usuwanie (**DELETE FROM/TRUNCATE**) danych.

Kompetencje techniczne i analityczne

Statystyka - czasem trzeba znać teorię

Statystyka i matematyka stanowią solidny fundament analizy danych i są niezbędne do wyciągania znaczących wniosków. Znajomość statystyki pozwala na rzetelną interpretację wyników, dobór metryk (KPI) i efektywną komunikację wniosków z biznesem. Kluczowe pojęcia, które musisz opanować:

- **Miary tendencji centralnej** (średnia, mediana, moda).
- **Miary rozproszenia** (odchylenie standardowe, wariancja).
- **Prawdopodobieństwo i rozkłady** (np. rozkład normalny).
- **Korelacja i regresja** (np. regresja liniowa).
- **Testowanie hipotez** (podstawy testów A/B, zrozumienie p-value).

Python - język (programowania) danych

Python jest językiem przyjaznym dla początkujących, a jednocześnie niezwykle potężnym w kontekście analizy danych. Choć Analityk BI może go nie potrzebować, jest on kluczowy w Data Science i Inżynierii Danych. Kluczowe obszary:

- **Podstawy Języka:** Zmienne, typy danych, struktury danych (listy, słowniki – kluczowe do organizowania informacji), instrukcje warunkowe (if/else), pętle i funkcje.
- **NumPy:** Biblioteka do **wydajnych obliczeń numerycznych** na dużych tablicach (*arrays*).
- **Pandas: Podstawa przetwarzania danych w Pythonie**, porównywalna do SQL w pracy z tabelami. Umożliwia manipulację, czyszczenie (np. obsługa brakujących wartości – `isnull()`), grupowanie danych (*groupby*) i łączenie (`merge()`, `join()`) w strukturach **DataFrame**.

- **Matplotlib / Seaborn:** Biblioteki do wizualizacji, od statycznych wykresów (Matplotlib) po bardziej zaawansowane i estetyczne wykresy statystyczne (Seaborn), takie jak heatmapy korelacji.

Architektura rozwiązań - zupełne podstawy

Ten rozdział pomaga zrozumieć kontekst pracy. Obejmuje on:

- **Źródła danych** (logi, API, systemy transakcyjne).
- **Miejsca składowania** (bazy SQL/NoSQL, hurtownie danych, data lake).
- **Przepływy danych** (kopiowanie, API, streaming, procesy ETL/ELT).
- **Organizacja procesów** (środowiska dev/test/prod, orkiestracja np. Apache Airflow).



Ścieżki kariery

Po opanowaniu podstaw Python + SQL + Excel + statystyki, można wybrać jedną z trzech ścieżek specjalizacji:

Ścieżka specjalizacji		Kluczowe umiejętności i projekty
BI Analyst	Najłatwiejsza ścieżka wejścia. Koncentruje się na raportowaniu, tworzeniu dashboardów i dostarczaniu <i>business insights</i> .	Narzędzia BI (Power BI, Tableau), Modelowanie danych (schemat gwiazdy), Znajomość KPI finansowych/marketingowych (MRR, CAC, LTV), Tworzenie pulpitów zarządu (<i>Executive Dashboard</i>), śledzenie wyników sprzedaży.
Data Scientist	Najtrudniejsza ścieżka. Koncentruje się na budowaniu modeli predykcyjnych, odkrywaniu wzorców i eksperymentach (Machine Learning).	Zaawansowana statystyka, ML (scikit-learn, klasyfikacja, regresja), Deep Learning (CNN, RNN), NLP (analiza sentymentu). Projekty: Predykcja cen nieruchomości, klasyfikacja spamu, rekomendacje filmów.
Data Engineer	Budowa i utrzymywanie infrastruktury danych oraz pipeline'ów ETL/ELT. Praca Data Engineera stanowi fundament dla zespołów analitycznych.	Zaawansowany SQL (Funkcje Okna), Python w inżynierii danych (Airflow, Dask), Platformy chmurowe (AWS/GCP/Azure), Technologie Big Data (Apache Spark, Kafka), DevOps (Docker, CI/CD, Terraform). Projekty: End-to-end ETL Pipeline, Real-time Streaming System.

Umiejętności Miękkie i Etyka Danych

Do sukcesu w zawodach związanych z danymi niezbędne są również umiejętności miękkie:

- **Rozwiązywanie problemów** (identyfikacja i definiowanie problemów).
- **Komunikacja** (wymiana pomysłów, tłumaczenie wyników technicznych na język biznesowy).
- **Wiedza domenowa** (zrozumienie branży: e-commerce, finanse, etc.).

Praca z danymi wiąże się z zagadnieniami etycznymi. Każda rola musi dbać o:

- **Anonimizację i prywatność danych** (zgodność z RODO).
- **Wyjaśnialność modeli (XAI) i unikanie stronniczości algorytmów (bias).**

AI - Pomoc czy Przeszkoda?



Sztuczna inteligencja jest potężnym asystentem, który rewolucjonizuje pracę z danymi, automatyzując rutynowe zadania, takie jak generowanie kodu SQL, tworzenie podstawowych skryptów ETL czy optymalizację hiperparametrów modeli.

Jednak AI nie zastąpi ludzkiej intuicji, kreatywności i strategicznego myślenia. Przyszłość należy do specjalistów, którzy potrafią efektywnie współpracować z AI, skupiając się na **umiejętnościach prompt engineering**, myśleniu krytycznym i kontekście biznesowym.

Twoja kariera zaczyna się TERAZ

Branża danych przeżywa dynamiczny rozwój, z **prognozowanym wzrostem liczby miejsc pracy o 35% do 2032 roku**. Praca z danymi to fascynująca ścieżka kariery, a **dane to nowa ropa** (*data is new oil*).

Jesteś gotów, aby zacząć – daję Ci kompas

Ten przewodnik to nie tylko książka, to Twoja **ściąga** na drodze rozwoju i osobisty plan kariery. Dzięki niemu zyskujesz **kompleksową ścieżkę rozwoju od absolutnych podstaw** do jednej z trzech najbardziej pożądaných ról:

ZDOBĄDŹ WIEDZĘ, ABY ZOSTAĆ	NAJWAŻNIEJSZE WYNIKI
Analitykiem BI (<i>BI analyst</i>)	Najłatwiejsza ścieżka wejścia, skupiona na raportowaniu i bezpośrednim wpływie na decyzje biznesowe .
Data Scientist (<i>Danolog</i>)	Budujesz modele predykcyjne i prowadzisz eksperymenty Machine Learning.
Inżynierem Danych (<i>Data Engineer</i>)	Budujesz fundamenty Big Data i automatyzujesz kluczowe procesy ETL/ELT.

Pozbądź się chaosu – poznaj fundamenty

- **Excel:** Opanuj “**króla korporacji**”. Znajomość Excela to **niezbędne minimum**, które nauczy Cię **myślenia algorytmicznego**.
- **SQL:** Naucz się **najpotężniejszego języka do analizy danych**. Nie znać SQL w świecie danych to **jak nie znać angielskiego w świecie IT**.
- **Python:** Opanuj język **przyjazny dla początkujących, a jednocześnie niezwykle potężny** w analizie danych, z kluczowymi bibliotekami takimi jak **Pandas** (podstawa przetwarzania danych) i **NumPy**.
- **Statystyka:** Zdobądź **solidny fundament analizy**, który jest niezbędny do wyciągania **znaczących wniosków**, monitorowania KPI i testowania hipotez (A/B testing).

Autor

- Od co najmniej 25 lat *pracuję w IT* – część tego czasu to jeszcze studia i średnia szkoła, nie tylko *prawdziwa* praca, ale kod już wtedy pisałem.
- Pisałem w kilku językach programowania (między innymi w C, PHP, R, Python i SQL)
- **Około 15 lat zajmuję się analizą danych** – stworzyłem setki, jeśli nie tysiące analiz, raportów, sporo dashboardów, kilka modeli predykcyjnych, kilka przepływów danych. To wszystko pisząc formuły w Excelu, kod w VBA, R i Pythonie, *brudząc sobie ręce*. Drugie tyle rozwiązań zaprojektowałem - od kilku lat moja główna praca to projektowania rozwiązań związanych z danymi, a nawet big data.
- Nie lubię pracy powtarzalnej, dlatego zaprojektowałem (i część z nich własnoręcznie wdrożyłem) rozwiązania do przetwarzania danych.
- W tym automaty generujące raporty oraz aplikacje zbierające i prezentujące dane.
- To były projekty przygotowane dla różnych zespołów na różnych szczeblach organizacji, w tym dla prezesów spółek giełdowych z WIG30. Zarządom i dyrektorom tych spółek prezentowałem dane - najczęściej jako raporty czy gotowe dashboardy. Ustalałem z nimi jak mają wyglądać.



Do tego wszystkiego lubię dzielić się wiedzą:

- Chcę przekazać jej trochę **Tobie**.
- Stworzyłem [pierwszego w Polsce bloga o analizie danych](#), gdzie o analizie danych pisałem od lutego 2017 roku (intensywnie do połowy 2022)
- Prowadzę [fanpage](#) poświęcony tym zagadnieniom. Jeśli trzymać się dat - od lipca 2017 roku.

NIE CZEKAJ NA IDEALNY MOMENT!



ZACZNIJ Z TYM PRZEWODNIKIEM! Z dobrym planem, konsekwentnym wysiłkiem i zamiłowaniem do uczenia się możesz zbudować satysfakcjonującą karierę w świecie analizy danych, nawet zaczynając od zera.

Przyszłość należy do tych, którzy potrafią efektywnie współpracować z AI, łącząc ludzką intuicję, kreatywność i strategiczne myślenie. Ten przewodnik nauczy Cię, jak to zrobić.

KUP TERAZ

i zacznij realizować swój plan kariery