

# Program Data Lab – Analityk Danych



## BLOK 0: Wstęp do analizy danych

### Wprowadzenie do narzędzia Excel

- Instalacja i konfiguracja narzędzia Excel
- Weryfikacja poziomu posiadanych umiejętności
- Uzupelnienie braków i wyrównanie wiedzy do poziomu wymagań niezbędnych do rozpoczęcia zajęć
- Omówienie funkcjonalności programu Excel w procesie analizy danych
- Operacje arytmetyczne z wykorzystaniem narzędzia Excel

### Wprowadzenie do Analizy Danych

- Wizualizacja i analiza danych - wstęp i inspiracja
- Proces analizy danych
- Przetwarzanie danych
- Eksploracja danych i narzędzia matematyczne
- Pogłębiona analiza danych i elementy statystyki
- Prezentacja danych i ich analizy
- Przykładowe zbiory danych i zagadnienia, których dotyczą

### Eksploracja danych oraz podstawy pracy z danymi w Excelu

- Wczytywanie danych i podstawy pracy z danymi w Excel
- Pozyskanie zbiorów danych oraz wczytanie ich do MS Excel 365
- Wstęp do czyszczenia danych: formaty liczb oraz dat, początek czasu w Excelu, zaawansowane formatowanie danych, format tabeli
- Eksploracja danych z wykorzystaniem funkcji, formuł i zaawansowanych formatów danych
- Typy danych i wstęp do analizy eksploracyjnej
- Tabele przestawne: zaawansowana eksploracja i raportowanie danych
- Wczytanie danych z zewnętrznych źródeł danych

### Wizualna eksploracja danych i elementy statystyki opisowej

- Wykresy jednej zmiennej i statystyka opisowa z elementami wnioskowania.
- Storytelling danych
- Analiza eksploracyjna – wizualizacje danych
- Elementarny wstęp do statystyki opisowej
- Wizualizacja danych jednowymiarowych: ilościowych na wykresach pudełkowych
- Wstęp do wnioskowania o różnicach statystycznych między kategoriami
- Wykresy dwóch i więcej zmiennych
- Wizualizacja danych dwu- i więcej wymiarowych oraz wyzwania z tym związane
- Tworzenie wykresów (różne rodzaje)
- Rozkład normalny

### Organizacja i wzbogacanie zbiorów danych oraz wstęp do analiz korelacji i regresji

- Organizacja i łączenie danych w Excelu
- Grupowanie danych
- Łączenie danych między zbiorami
- Wykresy przestawne
- Metody analizy korelacji i regresji dla danych dwuwymiarowych
- Dopasowywanie prostej do danych
- Współczynnik korelacji liniowej  $R^2$  i jego znaczenie
- Inne krzywe dopasowania
- Linearyzacja danych

### Analiza danych – serie czasowe

- Wstęp do analizy danych w postaci serii czasowych
- Wstęp do serii czasowych
- Wstęp do analizy serii czasowych - metody wygładzania oraz okresowość i sezonowość
- Praca z dodatkami w programie Excel oraz biblioteki w programowaniu
- Warsztat projektowy



# BLOK 1: Python – analiza danych

## Prework – Podstawy programowania

- Wprowadzenie do Pythona
- Typy danych w Pythonie
- Biblioteka standardowa w języku Python

## Python

- Funkcje
- Listy
- Krotki
- Stringi
- Obiekty
- Pliki
- Wyjątki
- Biblioteki

## SQL

- Podstawowe operacje
- PostgreSQL
- Relacje
- Funkcje

## JSON i API

- JSON
- API
- Autentykacja

## Pandas

- Filtrowanie
- Obróbka danych
- Grupowanie danych
- Merge
- Datetime
- Pivot
- OpenPyXL

## Web scraping

- Wstęp do HTML-a
- Webscraping w Python
- Element BS

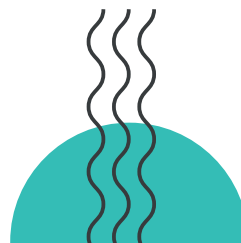
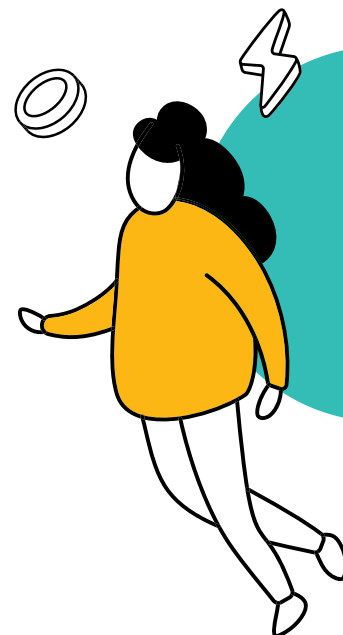
## Wizualizacja danych

- Wizualizacja danych - tworzenie wykresów

## Generowanie PDF

- Tworzenie PDF
- Tworzenie dokumentu przez story
- Tabele

## Warsztat projektowy





## BLOK 2: SQL – analiza danych

### Prework – Bazy danych

- Typy baz danych
- Podstawy UML
- Instalacja bazy danych i przygotowanie środowiska pracy
- Zapoznanie z bazą danych
- Przeglądanie danych (elementarny SQL)
- Podstawy logiki operatorów and, or
- Historia baz danych

### Podstawy SQL

- Podstawy SQL
- Podstawy logiki oraz algebry Bool'a
- Dodatkowe klauzule
- Operacje na zbiorach
- Podzapytania
- Tabele
- Manipulacja rekordami w bazie danych

### Relacje

- Relacje i ich typy
- Joiny i ich rodzaje
- Joiny w zastosowaniu
- Delete cascade
- Indexy
- Inne typy join, dobre praktyki join oraz podzapytań

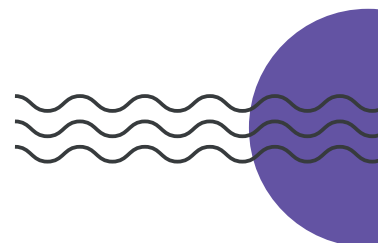
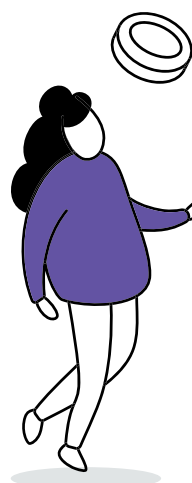
### Analiza danych

- Analiza danych
- Widok
- Grupowanie danych
- Dodatkowe funkcje grupujące
- Funkcje okna
- Praca z datetime
- Kolejność operacji w SQL
- Rollup, query plan

### Proceduralny SQL

- Proceduralny SQL
- Wprowadzenie do proceduralnego SQL-a
- Kontrola przepływu
- Transakcje
- Wbudowane funkcje w SQL-u
- Pętle
- Kursory
- JSON
- Normalizacja danych
- Notebook dla SQL-a

### Projekt indywidualny





## BLOK 3: Wizualizacja danych

### Prework – Wstęp do wizualizacji danych

- Wstęp i przygotowanie
- Wstęp do wizualizacji danych
- Kodowania graficzne
- Kolor
- Datawrapper
- Typy danych i operacje

### Seaborn/Plotly

- Teoria wizualizacji danych
- Mapy kolorów
- Seaborn
- Plotly
- Podstawy wizualizacji w poszczególnych bibliotekach
- Wykresy Plotly w Google Slides
- Podstawy Bokeh

### Mapy

- Projekcje, współrzędne, kartogramy
- Wstęp do bibliotek
- Geopandas, Geometry, GeoSeries, GeoDataFrame osm
- Geojson, Shapefile
- Mapbox

### Dash, grafy

- Networkx i Grafy
- Co to jest dashboard?
- Wstęp do dash
- Dash komponenty
- Dash datatable
- Dash callback i app lifecycle
- Edytowalna data table
- Interakcje z dashboardem
- Dash Cytoscape

### Dashboardy, Data Storytelling

- Data storytelling
- Dashboard
- Tooltip
- Multipage app
- Style bootstrapowe
- Dash i SQL
- Dash i mapy

### Tabelau

- Co to jest Tableau?
- Przygotowanie środowiska pracy
- Pierwszy projekt
- Tworzenie Dashboardu
- Scenariusze wykorzystania
- Integracja Tableau i Mapbox

### Warsztat projektowy

