

## IT CAMP – Analiza danych dla analityków

### Cele szkolenia

Po zakończeniu szkolenia „IT CAMP – Analiza danych dla analityków” jego uczestnik nabędzie wiedzę na temat całości procesów realizowanych podczas analizy danych w przedsiębiorstwie. Szkolenie przygotowuje do zdobycia zawodu analityka, który to zawód jest obecnie jednym z najbardziej popularnych na rynku pracy.

### Profil słuchaczy

Szkolenie przeznaczone jest dla osób które posiadają wiedzę z zakresu zastosowania arkusza kalkulacyjnego oraz podstaw statystyki i relacyjnych baz danych, a zamierzają pracować w zawodzie analityka danych. Szkolenie kompleksowo przygotowuje do zawodu analityka danych bazując na zróżnicowanym pakiecie narzędzi analitycznych, co umożliwi w prosty sposób dostosowanie swojej wiedzy do wymogów przedsiębiorstwa. Szkolenie jest również przeznaczone dla osób pracujących w zawodzie analityka danych na poziomie początkującym. Nie jest wymagane wykształcenie kierunkowe.

### Wymagania wstępne

Od uczestników szkolenia wymagana jest znajomość:

- Microsoft Excel w stopniu przynajmniej dobrym;
- podstaw teoretycznych statystyki;
- podstaw relacyjnych baz danych (nie jest wymagana znajomość konkretnego języka dostępu do bazy danych ani konkretnej platformy bazodanowej);

### Czas trwania

208 godzin lekcyjnych + 4 godziny konsultacji HR

### Metoda realizacji szkolenia

Szkolenie realizowane jest w formie naprzemiennie następujących po sobie mini wykładów oraz ćwiczeń praktycznych. Szkolenie łączy w sobie fachową wiedzę merytoryczną z praktycznymi przykładami jej wykorzystania w środowisku pracy.

### Wiedza teoretyczna i praktyczna

**CENTRUM SZKOLENIOWE  
COMARCH**

ul. prof. Michała Życzkowskiego 33  
31-864 Kraków

**Tel.:** +48 12 687 78 11

**Email:** szkolenia@comarch.pl

[WWW.SZKOLENIA.COMARCH.PL](http://WWW.SZKOLENIA.COMARCH.PL)

Comarch Spółka Akcyjna z siedzibą w Krakowie, Aleja Jana Pawła II 39a, zarejestrowana w Krajowym Rejestrze Sądowym prowadzonym przez Sąd Rejonowy dla Krakowa-Śródmieścia w Krakowie XI Wydział Gospodarczy Krajowego Rejestru Sądowego pod numerem KRS 0000057567. Wysokość kapitału zakładowego Spółki wynosi 8.133.349,00 zł. Kapitał zakładowy wpłacony w całości. NIP: 677-00-65-406

## Moduł A: Wstęp do analizy danych (4 godz.)

### Metodyki stosowane w analizie danych

- CRISP-DM
- Virtuous Cycle of Data Mining
- SEMMA
- Six Sigma

### Omówienie metodyki CRISP-DM

- Zastosowanie metodyki w projektach DM

### Etapy prowadzenia projektu przy użyciu CRISP DM

- Zrozumienie celów biznesowych
  - Określenie celu
  - Cele data mining
  - Plan projektu
- Zrozumienie danych
  - Źródła danych
  - Wprowadzanie danych
  - Przegląd danych
  - Jakość danych
- Przygotowanie danych
  - Rodzaje danych
  - Wybór danych w DM
  - Dane ustrukturyzowane i nieustrukturyzowane
  - Dane zewnętrzne i dane wewnętrzne
  - relacje pomiędzy danymi
  - Oczyszczanie danych
  - Budowa zbioru danych
  - Zagrożenia wynikające z pracy na różnych typach danych
  - Migracje danych
  - Formatowanie danych
- Budowa modelu
  - Określenie techniki budowy
  - Budowa modelu
  - Narzędzia analizy
  - Ocena funkcjonowania modelu
- Ewaluacja
  - Ocena rezultatów procesu
  - Przegląd procesu
  - Korekty działań
  - Weryfikacja kolejnych działań
- Wdrożenie
  - Zaplanowanie wdrożenia
  - Monitoring procesu
  - Raportowanie
  - Zakończenie projektu

## Moduł B: Przygotowanie danych do analizy (68 godz)

### Przygotowanie danych do analizy w Microsoft Excel

- Poprawność danych
  - Narzędzie tabela
  - Narzędzie sprawdzania poprawności nazw
  - Analiza poprawności formuł i testowanie arkusza
  - Konwersja typów liczbowych, dat, typów znakowych oraz stron kodowych
  - Łączenie danych z wielu źródeł
- Użycie danych pochodzących z różnych źródeł zewnętrznych
  - Import danych z plików tekstowych
  - Import danych z baz danych
  - Import danych z plików XML – omówienie standardu XML
  - Dane ze stron www
  - Eksportowanie danych

### SQL jako język pozyskiwania danych analitycznych z relacyjnych baz danych

- Relacyjny model danych
  - - Podstawowe struktury danych
  - - Zależności między strukturami
- Składnia języka SQL
- Rola języka SQL w wybieraniu danych
- Programy nakładkowe do wykonywania zapytań
- Zasady tworzenia zapytań
  - Słowa kluczowe
  - Składnia poleceń
  - Stosowanie aliasów dla kolumn i tabel
- Sortowanie danych
- Zasady użycia funkcji w zapytaniach
- Sposoby wybierania danych z wielu tabel
- Sposoby łączenia tabel
- Rodzaje złączeń
  - Równościowe (equi-join)
  - Nierównościowe (non-equi-join)
  - Typy złączeń zewnętrznych (outer-join): lewostronne, prawostronne, obustronne
  - Samozłączenie (self-join)
- Operacje zbiorowe na wynikach zapytań
  - Suma
  - Przecięcie
  - Różnica
- Podzapytania
  - Zwykłe
  - Skorelowane

- Język manipulowania danymi (DML)
  - Insert
  - Update
  - Delete
  - Operacje DML w oparciu o wyniki zapytań
- Transakcje
  - Idea transakcyjności operacji
  - Zatwierdzanie (commit)
  - Odrzucanie (rollback)

### SQL Server – Integration Services – projektowanie i realizacja procesów

- SSIS – podstawowe pojęcia, architektura, komponenty
- Narzędzia – Business Intelligence Development Studio, SQL Server Data Tools, BIDS Helper.
- Tworzenie projektu.
- Tworzenie pakietu SSIS
  - Definiowanie połączeń (Connection Managers)
  - Dodawanie i konfigurowanie elementów Control Flow,
  - Wykorzystanie kontroli przebiegu.
  - Wykonanie pakietu.
  - Zachowanie pakietu jako pliku lub w repozytorium
  - Narzędzia dtexec i dtexecui.
- Tryb projektowy w SSIS 2012/2014.
- Definiowanie Data Flow – zasada działania
  - Definiowanie źródeł danych oraz miejsc docelowych
  - Wykorzystanie jako źródeł baz danych, plików tekstowych, plików Excel.
- Najczęściej używane transformacje
  - Konwersja danych
  - Dodawanie kolumn wyliczanych
  - Agregacja danych
  - Sortowanie danych
  - Łączenie danych z różnych źródeł (LookUp)
  - Konwersja typów danych.
- Łączenie i rozdzielanie przebiegów danych (Union All oraz Conditional Split).
- Obsługa błędów w przepływie danych.
- Obsługa zdarzeń w pakietach.
- Logowanie wykonania pakietów SSIS.
- Wykorzystanie Execute SQL Task do wykonywania operacji na bazach danych.
- Użycie Send Mail Task do wysyłania wiadomości pocztowych.
- Wykorzystanie elementów FTP Task oraz File System Task to wykonywania operacji na plikach składowanych w systemie plików oraz na serwerach FTP.
- Kontener sekwencyjny. Kontenery pętli
- Wykorzystanie zmiennych oraz parametrów.
  - Użycie wyrażeń do dynamicznego ustawiania wartości właściwości elementów pakietu.
  - Konfiguracja pakietów.
- Typowe scenariusze ładowania danych
  - Ładowanie inkrementalne

- Wykorzystanie komponentów Change Data Capture
- Użycie Data Quality Services do czyszczenia danych.
- Przykłady zastosowania komponentów skryptowych (Script Task, Script Component).

### Wykorzystanie języka R w analizie danych

- Wprowadzenie do R.
  - Podstawowe pojęcia.
  - Środowisko pracy R. CRAN.
  - RStudio.
  - Inne rozszerzenia.
- Podstawy języka R.
- Struktury danych w R – wektory, listy, faktory, macierze, ramki danych.
- Zarządzanie danymi w R. Import danych z różnych źródeł.
- Integracja języka R z Microsoft SQL Server. Wykorzystanie Microsoft R Server.

## Moduł C: Budowa modelu i analiza danych (56 godz)

### Analiza danych za pomocą Microsoft Excel

- Narzędzia analizy
  - Wykorzystanie narzędzi analizy warunkowej
  - Używanie różnych typów adresowania
  - Wykorzystanie formuł tablicowych
  - Analiza warunkowa
  - Agregacja danych
  - Wybrane funkcje: finansowe, statystyczne, matematyczne, tekstowe
- Kontrola danych
  - Tworzenie różnego rodzaju sum
  - Ustalanie rankingów
  - Badanie realizacji celu
  - Analiza bazodanowa
  - Udziały procentowe
  - Wykorzystanie formantów do zmiany parametrów prowadzonej analizy
- Prognozowanie na podstawie danych
  - Prognozy w oparciu o linie trendu
  - Prognozy z wykorzystaniem modeli regresji
  - Prognozy sezonowe
  - Konsolidacja danych z wielu arkuszy
  - Konsolidacja danych za pomocą tabeli przestawnej
  - Konsolidacja wielu źródeł danych (w tym zewnętrznych)
- Praca z tabelami i wykresami przestawnymi:
  - Automatyczne obliczanie odpowiednich udziałów procentowych
  - Obliczanie wartości skumulowanych
  - Szukanie zmian wartości
  - Grupowanie danych

- Tworzenie wewnętrznych formuł obliczeniowych
- Wykres przestawny
- Pulpit menedżerski
  - Omówienie koncepcji pulpitu menedżerskiego
  - Zasady tworzenia pulpitu menedżerskiego
  - Realizacja z wykorzystaniem MS Excel

### **Analiza danych z wykorzystaniem dodatków PowerPivot w Microsoft Excel**

- Wprowadzenie do Power Pivot
  - Podstawowe funkcjonalności
  - Cykl pracy z Power Pivot
- Import danych
  - Źródła danych
  - Kreator importu danych
  - Filtrowanie danych w trakcie importu
  - Tworzenie relacji pomiędzy tabelami
  - Odświeżanie danych
  - Wykorzystanie Power Query do importu i przetwarzania danych
- Praca z danymi
  - Arkusze danych
  - Formatowanie, sortowanie, filtrowanie danych
  - Tworzenie własnych kolumn
  - Efektywne wykorzystanie kolumn czasowych (daty)
  - Formuły korzystające z danych z tabel powiązanych
  - Przykłady najczęściej stosowanych funkcji DAX

### **Power BI – analiza danych**

- Omówienie wersji Power BI oraz ich możliwości
- Zestawy danych, Raporty i Pulpity nawigacyjne
- Power BI Online
  - Ładowanie danych
  - Język naturalny zapytań
  - Quick Insights
- Ładowanie danych z wykorzystaniem PowerQuery
- Tworzenie kalkulacji w języku DAX
- Wprowadzenie do Query Editor
- Tworzenie relacji
- Udostępnianie pulpitów
- Usługi powiązane
  - OneDrive for Business
  - Excel
  - PowerPoint
- Usługi Power BI Pro.

## Microsoft SQL Server – Analysis Services – projektowanie i budowa wielowymiarowych baz danych

- Zasady i podstawowe pojęcia modelowania wielowymiarowego
  - tabele faktów,
  - wymiary,
  - miary,
  - ziarno tabeli faktów,
  - typy tabel faktów,
  - addytywność miar.
- Narzędzia wykorzystywane do budowy wielowymiarowych baz OLAP
  - Business Intelligence Development Studio,
  - SQL Server Data Tools,
  - SQL Server Management Studio,
  - MDX Studio,
  - BIDS Helper.
- Tworzenie projektu Analysis Services – konfiguracja projektu.
- Definiowanie źródła danych.
- Budowa widoku źródła danych
  - logiczne klucze główne,
  - relacje pomiędzy tabelami,
  - tworzenie własnych formuł w tabelach,
  - tworzenie widoków (Named Query).
- Budowa wymiaru
  - definiowanie atrybutów,
  - właściwości atrybutów,
  - relacje pomiędzy atrybutami,
  - hierarchie atrybutowe,
  - tworzenie hierarchii użytkownika,
  - tworzenie hierarchii parent-child,
  - wykorzystanie operatorów unarnych oraz formuł agregujących.
- Budowa kostki OLAP
  - definiowanie grup miar oraz miar,
  - określanie relacji pomiędzy wymiarami a grupami miar,
  - relacja regularna, referencyjna, wiele-do-wielu,
  - właściwości miar (formatowanie).
- Wdrożenie kostki OLAP
  - kompilacja projektu,
  - wdrożenie projektu na serwerze,
  - przetwarzanie (procesowanie) kostki OLAP oraz wymiarów,
  - obsługa błędów podczas procesowania.
- Tworzenie kalkulacji w kostce OLAP
  - zasady budowy wyrażeń MDX
  - skrypt kalkulacyjny
  - najczęściej wykorzystywane funkcje języka MDX.
- Agregacje
  - zasada działania agregacji,



- tworzenie agregacji przy pomocy kreatora,
- tworzenie agregacji przy pomocy analizy wykonywanych zapytań,
- ręczne definiowanie agregacji,
- kontrola działania agregacji.
- Partycjonowanie kostek OLAP
  - tworzenie partycji,
  - procesowanie,
  - monitorowanie wykorzystania partycji,
  - tryby przechowywania danych (MOLAP, HOLAP, ROLAP),
  - przypisywanie projektów agregacji do partycji.
- Akcje
  - rodzaje akcji,
  - tworzenie akcji,
  - wywoływanie akcji w aplikacjach klienckich.
- Perspektywy – tworzenie i wykorzystanie.
- Tłumaczenia
  - definiowanie tłumaczeń,
  - testowanie i wykorzystanie,
  - tłumaczenia przypisane do elementów wymiarów.
- Bezpieczeństwo danych
  - definiowanie ról,
  - określanie uprawnień do kostek, wymiarów, elementów wymiarów,
  - wykorzystanie MDX do definiowania praw dostępu,

## Moduł D: Wizualizacja, ewaluacja i wdrożenie (80 godz)

### Wizualizacja danych w MS Excel

- Wybór rodzaju wizualizacji
  - Wybór odpowiedniej formy zestawień
  - Odpowiedni dobór wykresu
  - Nakładanie wielu wykresów na jedną oś
  - Prezentacja danych na wykresie o dużej rozpiętości skali
  - Formatowanie warunkowe
  - Tworzenie wykresów niestandardowych
- Inne narzędzia prezentacji danych
  - Sparklines - wykresy przebiegu w czasie
  - Wykresy na mapach
  - Mapy jako element wizualizacji
  - Infografiki
- Publikacja danych:
  - Publikacja danych w różnych formatach zewnętrznych
  - Publikacja danych i wykresów w PowerPoint i Word
  - Aktualizacja danych w ramach pakietu Office
  - Publikacja danych do formatu PDF



- Publikacja danych do formatów HTML (webowych)
- Publikacja danych na serwerach SQL

### Tableau jako przykład narzędzia wizualizacji danych

- Pobieranie danych
  - Tworzenie połączeń
  - Współdzielenie połączeń
  - Zmiany w danych źródłowych
- Tworzenie prostych wizualizacji
  - Elementy wizualizacji
  - Formatowanie widoku
  - Filtrowanie i sortowanie danych
  - Używanie grup, hierarchii, zestawów i pól złożonych
- Wykorzystanie wymiaru czasu
  - Przetwarzanie dat
  - Tworzenie filtrów opartych na datach
- Wizualizacja danych przy pomocy wykresów
  - Prezentacja wielu miar na wykresie
  - Wykresy złożone
  - Typy wykresów i ich zastosowania
  - Mapy temperaturowe
  - Wykresy kołowe, tree maps
- Wykorzystanie map
- Tworzenie kalkulacji
  - Typy kalkulacji
  - Tworzenie pól obliczeniowych
  - Użycie instrukcji warunkowych
  - Konwersje typów danych
  - Operowanie na datach i czasie
  - Agregacje danych
- Interaktywność
  - Parametry widoku
  - Tworzenie kokpitów analitycznych
  - Akcje i drażnienie danych
- Współdzielenie prac
  - Eksport widoku
  - Eksport danych
  - Inne opcje współdzielenia

### Raportowanie przy pomocy PowerPivot

- Tabele przestawne w PowerPivot
  - Power Pivot jako źródło danych dla tabel przestawnych
  - Elementy tabeli przestawnej
  - Formatowanie tabeli przestawnej
  - Własne miary w tabeli przestawnej

**CENTRUM SZKOLENIOWE  
COMARCH**

ul. prof. Michała Życzkowskiego 33  
31-864 Kraków

**Tel.:** +48 12 687 78 11

**Email:** szkolenia@comarch.pl

[WWW.SZKOLENIA.COMARCH.PL](http://WWW.SZKOLENIA.COMARCH.PL)

Comarch Spółka Akcyjna z siedzibą w Krakowie, Aleja Jana Pawła II 39a, zarejestrowana w Krajowym Rejestrze Sądowym prowadzonym przez Sąd Rejonowy dla Krakowa-Śródmieścia w Krakowie XI Wydział Gospodarczy Krajowego Rejestru Sądowego pod numerem KRS 0000057567. Wysokość kapitału zakładowego Spółki wynosi 8.133.349,00 zł. Kapitał zakładowy wpłacony w całości. NIP: 677-00-65-406

- Dodatkowe możliwości wizualizacji w MS Excel
  - Tworzenie powiązań pomiędzy tabelami przestawnymi
  - Wykresy przebiegu w czasie
  - Tworzenie raportów z wykorzystaniem tabel przestawnych i Power Pivot
  - Formatowanie raportów
  - Tworzenie raportów Power View

### Power BI – analiza i wizualizacja danych

- Tworzenie standardowych wizualizacji
  - Tworzenie
  - Modyfikacja
  - Tworzenie połączeń za pomocą fragmentatorów
  - Drażenie danych
  - Dodawanie i usuwanie elementów wykresów
- Wizualizacje niestandardowe
  - Dodawanie niestandardowych wizualizacji
  - Pola tekstowe
  - Grafiki
  - Kształty
- Zarządzanie filtrowaniem
  - Filtry wykresów
  - Filtry fragmentatora
  - Filtry stron
  - Filtry raportu
- Dodawanie obliczeń
  - Kolumny kalkulowane
  - Miary
  - Miary służące wizualizacji
- Wykorzystanie wizualizacji danych opartych o język R

### Wykorzystanie Reporting Services

- Raportowanie – podstawowe pojęcia.
  - Cykl życia raportu.
- Architektura i podstawowe komponenty Reporting Services.
- Narzędzia wykorzystywane do budowy raportów.
  - Business Intelligence Development Studio.
  - SQL Server Data Tools.
  - Report Builder.
  - Tworzenie i zarządzanie projektem w BIDS/SSDT.
- Definiowanie źródła danych raportu.
  - Określanie uprawnień dostępu do źródła danych.
- Definiowanie zestawu danych. Budowa zapytania.
  - Korzystanie z baz relacyjnych (SQL) i wielowymiarowych (OLAP – Analysis Services).
- Prezentacja danych przy pomocy elementu tablix.
  - Formatowanie tabeli.
  - Właściwości pól tekstowych.

- Publikacja raportu na serwerze Reporting Services.
- Sortowanie, filtrowanie i grupowanie danych w tabelach. Dodawanie podsumowań.
- Definiowanie wyrażeń w raportach. Formatowanie warunkowe.
- Wizualizacja danych przy pomocy wykresów i wskaźników.
- Zagnieżdżanie obiektów.
  - Wykorzystanie elementu Rectangle do grupowania obiektów.
- Wykorzystanie map.
- Właściwości i formatowanie raportu.
  - Podziały stron.
- Parametryzacja raportu.
  - Parametryzacja zestawów danych.
- Opcje interaktywne w raportach.
  - Mapa dokumentu.
  - Zakładki oraz akcje.
  - Interaktywne sortowanie.
  - Przełączanie widoczności elementów.
- Zarządzanie raportami w trybie natywnym przy pomocy Menedżera Raportów.
- Automatyzacja generowania i dystrybucji raportów przy pomocy subskrypcji.
- Definiowanie praw dostępu do raportów.
- Tworzenie, zarządzanie i wykorzystanie Web Parts do szybkiej budowy raportów.

#### Wykorzystanie języka R w analizie danych

- Wizualizacja danych w R.
- Biblioteka ggplot2.
- Wizualizacja danych geograficznych
- Wizualizacja na potrzeby analizy statystycznej
- Publikacja wyników pracy – Knitr, Shiny
- Przykłady zastosowań algorytmów uczenia maszynowego
  - Algorytmy klasyfikacyjne – k-NN, klasyfikatory Bayes'a, drzewa decyzyjne.
  - Budowa modeli regresyjnych. Prognozowanie wartości numerycznych.
  - Sieci neuronowe.
  - Clustering. Maszyna wektorów nośnych (Support Vector Machine).
- Wykorzystanie języka R w raportach Reporting Services oraz w modelach analitycznych Power BI.

### Moduł E: Konsultacje z doradcą HR (4 godz)

#### Konsultacje z doradcą HR

- Statystyka rynku pracy w zawodzie analityka
- Główne zasady rekrutacji pracowników
- Zasady rekrutacji w Comarch SA
- Klucz do dobrego CV
- Przygotowanie profilu zawodowego na znanych portalach (LinkedIn, Goldenline)
- Rozmowa rekrutacyjna
  - Przygotowanie do rozmowy

CENTRUM SZKOLENIOWE  
COMARCH

ul. prof. Michała Życzkowskiego 33  
31-864 Kraków

Tel.: +48 12 687 78 11

Email: szkolenia@comarch.pl

[WWW.SZKOLENIA.COMARCH.PL](http://WWW.SZKOLENIA.COMARCH.PL)

Comarch Spółka Akcyjna z siedzibą w Krakowie, Aleja Jana Pawła II 39a, zarejestrowana w Krajowym Rejestrze Sądowym prowadzonym przez Sąd Rejonowy dla Krakowa-Śródmieścia w Krakowie XI Wydział Gospodarczy Krajowego Rejestru Sądowego pod numerem KRS 0000057567. Wysokość kapitału zakładowego Spółki wynosi 8.133.349,00 zł. Kapitał zakładowy wpłacony w całości. NIP: 677-00-65-406

- Zasady prowadzenia rozmowy
- Pytania od i do rekrutera
- Stres – geneza i antidotum
- Pytania i odpowiedzi..

### Umiejętności

Po ukończeniu kursu uczestnik/czka będzie potrafił/a:

- Zastosować metodyki analizy danych;
- Przygotować dane do analizy za pomocą różnych narzędzi (Excel, SQL Server);
- Wykorzystać język R podczas przygotowania danych do analizy
- Wykonać analizę za pomocą narzędzi Excel, PowerPivot, PowerBI, SQL Server (kostki OLAP)
- Wykonać wizualizację danych za pomocą Excela, Tableau, PowerPivot, Power BI i SQL Servera (Reporting Services)
- Zastosować język R w analizie danych;
- Wdrożyć analizę danych w przedsiębiorstwie
- Przeprowadzić ewaluację uzyskanych wyników analizy

### Ścieżka rozwoju

Udział w szkoleniu pozwala zapoznać się całościowo z zastosowanymi metodami i narzędziami analizy danych. Szkolenie nie posiada kontynuacji w postaci ścieżki szkoleniowej.

**CENTRUM SZKOLENIOWE  
COMARCH**

ul. prof. Michała Życzkowskiego 33  
31-864 Kraków

**Tel.:** +48 12 687 78 11

**Email:** szkolenia@comarch.pl

[WWW.SZKOLENIA.COMARCH.PL](http://WWW.SZKOLENIA.COMARCH.PL)

Comarch Spółka Akcyjna z siedzibą w Krakowie, Aleja Jana Pawła II 39a, zarejestrowana w Krajowym Rejestrze Sądowym prowadzonym przez Sąd Rejonowy dla Krakowa-Śródmieścia w Krakowie XI Wydział Gospodarczy Krajowego Rejestru Sądowego pod numerem KRS 0000057567. Wysokość kapitału zakładowego Spółki wynosi 8.133.349,00 zł. Kapitał zakładowy wpłacony w całości. NIP: 677-00-65-406