



## AUTOMATYZACJA RAPORTOWANIA W STATISTICA

Grzegorz Migut, StatSoft Polska Sp. z o.o.

### Możliwości STATISTICA w zakresie automatyzacji

Automatyzacja procesu raportowania jest obecnie jednym z najważniejszych zadań stojących przed jednostkami administracji publicznej. Bardzo często proces raportowania przebiega w „sposób ręczny”: w oparciu o arkusze kalkulacyjne oraz ręcznie przeklepane formuły i dane. Ten sposób, poza nadmierną pracochłonnością i czasochłonnością, pociąga za sobą ryzyko popełnienia wielu pomyłek lub nawet poważnych błędów w trakcie przygotowania raportów.

W sytuacji, gdy dany rodzaj raportu jest powtarzany okresowo dla nowych danych bez zasadniczej zmiany swojego kształtu, najwygodniej jest skorzystać z mechanizmów umożliwiających całkowitą bądź częściową automatyzację procesu raportowania, zwłaszcza w najbardziej wrażliwych na pomyłki obszarach.

System STATISTICA obok olbrzymiego wyboru metod analitycznych i narzędzi graficznych zawiera także szereg narzędzi i mechanizmów umożliwiających automatyzację procesu analizy danych i raportowania, w szczególności możliwości:

- ◆ powtarzania analiz zapisanych w skrószycie wyników,
- ◆ zapisu przebiegu analizy w postaci makra Visual Basic,
- ◆ tworzenia szablonów analiz i raportów w systemie STATISTICA Enterprise,
- ◆ tworzenia szablonów raportów w postaci dokumentów MS Word.

### Przykład wykorzystania mechanizmów automatyzujących proces raportowania

Do zaprezentowania możliwości systemu STATISTICA w zakresie automatyzacji użyjemy przykładowego pliku danych *Dane PIT.sta* zawierającego zniekształconą próbkę danych z zeznań podatkowych ok. 55 tys. podatników sprzed kilku lat. Analizowane dane zawierają informację o wysokości rocznego przychodu w rozbiciu na źródła dochodu, a także informację o typie składanego formularza deklaracji podatkowej.



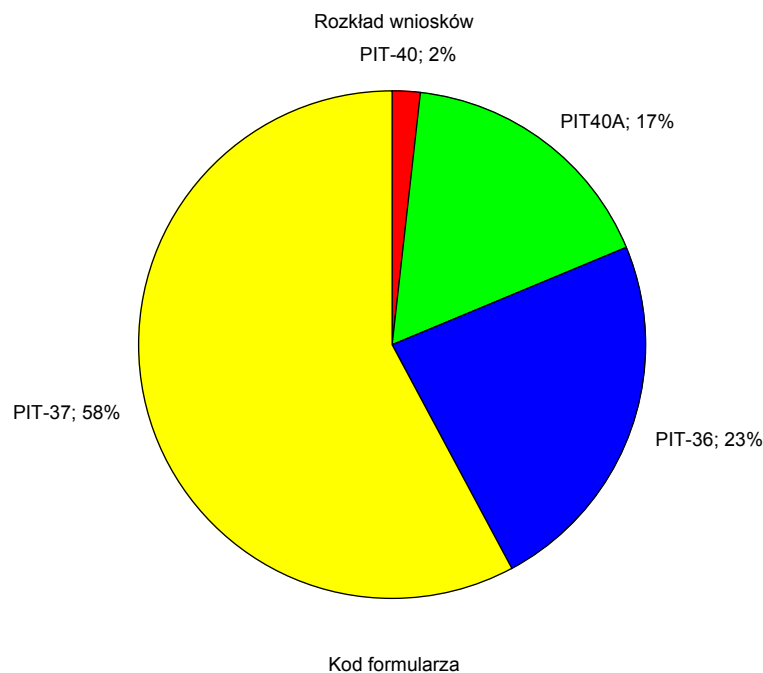
	1	2	3	4	5	6	7	8	9	10
	Wynagrodzenie	Działalność gospodarcza	Emerytury	Produkcja rolna	Osobiście	Najem	Prawa majątkowe	Pozostałe	Łączny przychód	Kod formularza
47071	10202,73								10202,73	PIT-37
47072	42908,06								42908,06	PIT-37
47073					1500				1500	PIT-37
47074	16049,39								16049,39	PIT-37
47075										PIT-36
47076	5344,1								5344,1	PIT-37
47077	91175								91175	PIT-37
47078	22033,65				1000				23033,65	PIT-37
47079			11655,2						11655,2	PIT-36
47080			1384,32						1384,32	PIT-37
47081	27968							182,6	28150,6	PIT-37
47082	9200,04					200			9400,04	PIT-36
47083			8950,54						8950,54	PIT-37
47084	16168,35								16168,35	PIT-37
47085			5688,18						5688,18	PIT-37
47086										PIT-36
47087		32750				4650			86719,65	PIT-36
47088	3885,24								3885,24	PIT-37
47089		83793,87						2430,96	86224,83	PIT-36
47090									10400,07	PIT-36
47091	9600								9600	PIT-37
47092		127637,88						5824,62	133462,5	PIT-36
47093					1196				1196	PIT-37
47094		244922,56						1362,53	246285,09	PIT-36
47095	21639,51							660	22299,51	PIT-40

Na przykładzie powyższego zbioru wykonamy kilka wykresów i raportów, a następnie przedstawimy możliwości powtórzenia wykonanych analiz bez konieczności ich ponownego wywoływania.

Założmy, że celem naszej analizy będzie zbadanie rozkładu wniosków składanych przez podatników oraz określenie jaki procent podatników wykazuje określone źródła przychodu i jaka jest przeciętna wartość przychodów w podziale na poszczególne źródła dochodów.

Rozkład wniosków składanych przez podatników najwygodniej będzie nam zbadać za pomocą wykresu kołowego. W tym celu z menu **Wykresy** wybieramy opcję **Wykresy 2W -> Wykresy kołowe**, a następnie wybieramy zmienną **Kod formularza**, dla której wygenerujemy wykres.

Analizując wykres, widzimy, iż większość osób (58%) w badanym roku złożyła wnioski, stosując formularz PIT-37. Jedynie 2% wniosków to wnioski PIT-40.



Kolejną analizą, jaką wykonamy, będzie analiza podstawowych statystyk dla poszczególnych źródeł dochodu. Zbadamy, jaki procent podatników wykazuje przychód z wyszczególnionych źródeł oraz ile wynoszą przeciętne przychody.

W celu wykonania tej analizy wybieramy *Statystyka -> Statystyki podstawowe i tabele*, a następnie opcję *Statystyki opisowe*. W wyświetlonym oknie wskazujemy zmienne do analizy. Będą to wszystkie zmienne reprezentujące szczegółowe źródła przychodu. Dla wybranych zmiennych obliczymy %wężnych obserwacji oraz podstawowe miary statystyczne: średnią, medianę oraz minimum i maksimum.

Zmienna	% Ważnych	Średnia	Mediana	Minimum	Maksimum
<b>Wynagrodzenie</b>	34,02	23581,0	17898	9,58000	901251
Działalność gospodarcza	11,50	476206,4	77334	0,01000	436258685
Emerytury	36,76	11094,5	9244	13,30000	101445
Produkcja rolna	0,36	5813,0	1650	3,00000	78228
Osobiście	11,91	5494,1	2124	6,00000	539419
Najem	1,93	12403,7	5500	50,00000	353827
Prawa majątkowe	2,44	10992,0	2800	9,09000	805063
Pozostałe	14,23	3389,4	1734	3,40000	237538

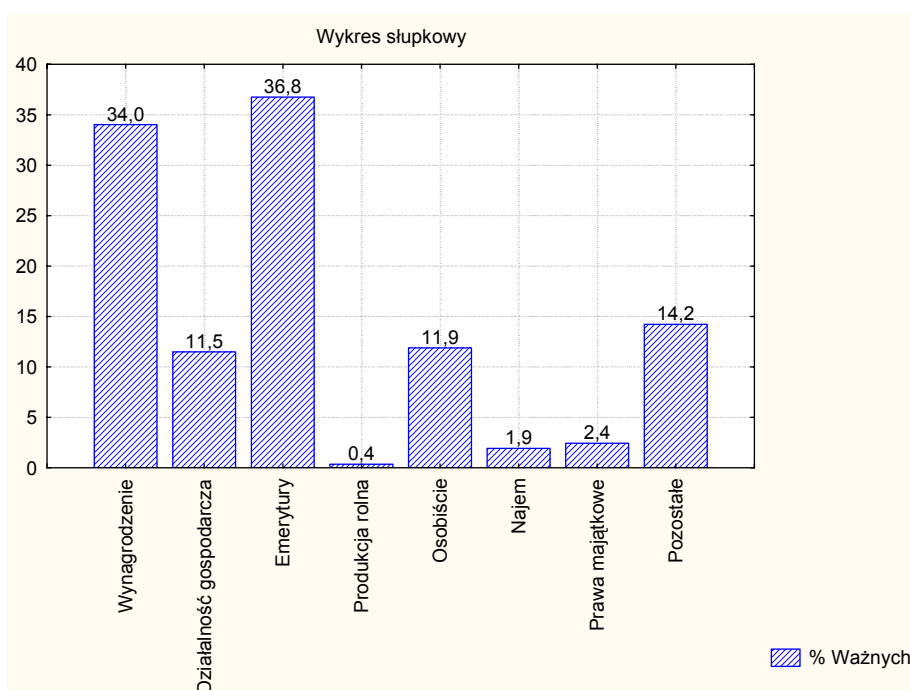
Rezultat analizy w postaci arkusza został zapisany w skoroszycie *STATSTICA* (podobnie jak poprzednio wykres kołowy). Analizując uzyskane wyniki, możemy zauważyć, iż



w analizowanym zbiorze danych ponad 34% podatników wykazuje przychód z tytułu wynagrodzenia, natomiast ponad 36% osób otrzymuje świadczenia emerytalne bądź rentowe.

Skorzystamy teraz z faktu, iż raporty *STATISTICA* mogą być równocześnie wejściowymi arkuszami danych dla nowych analiz, i na podstawie wygenerowanego raportu sporządzimy wykresy słupkowe przedstawiające procenty osób wykazujących dane źródło przychodu oraz wartości median (punkty środkowe).

By na podstawie uzyskanego raportu wygenerować wykres, określamy, by był on aktywnym arkuszem wejściowym. W tym celu prawym przyciskiem myszy zaznaczamy raport w skoroszycie i z menu podręcznego wybieramy odpowiednią opcję. Następnie z menu *Wykresy* wybieramy opcję *Wykresy 2W -> Wykresy słupkowe/kolumnowe* i do analizy wskazujemy zmienną *% Ważnych* oraz *Mediana*. Uzyskane wykresy słupkowe zapisane zostaną w kolejnym węźle skoroszytu.



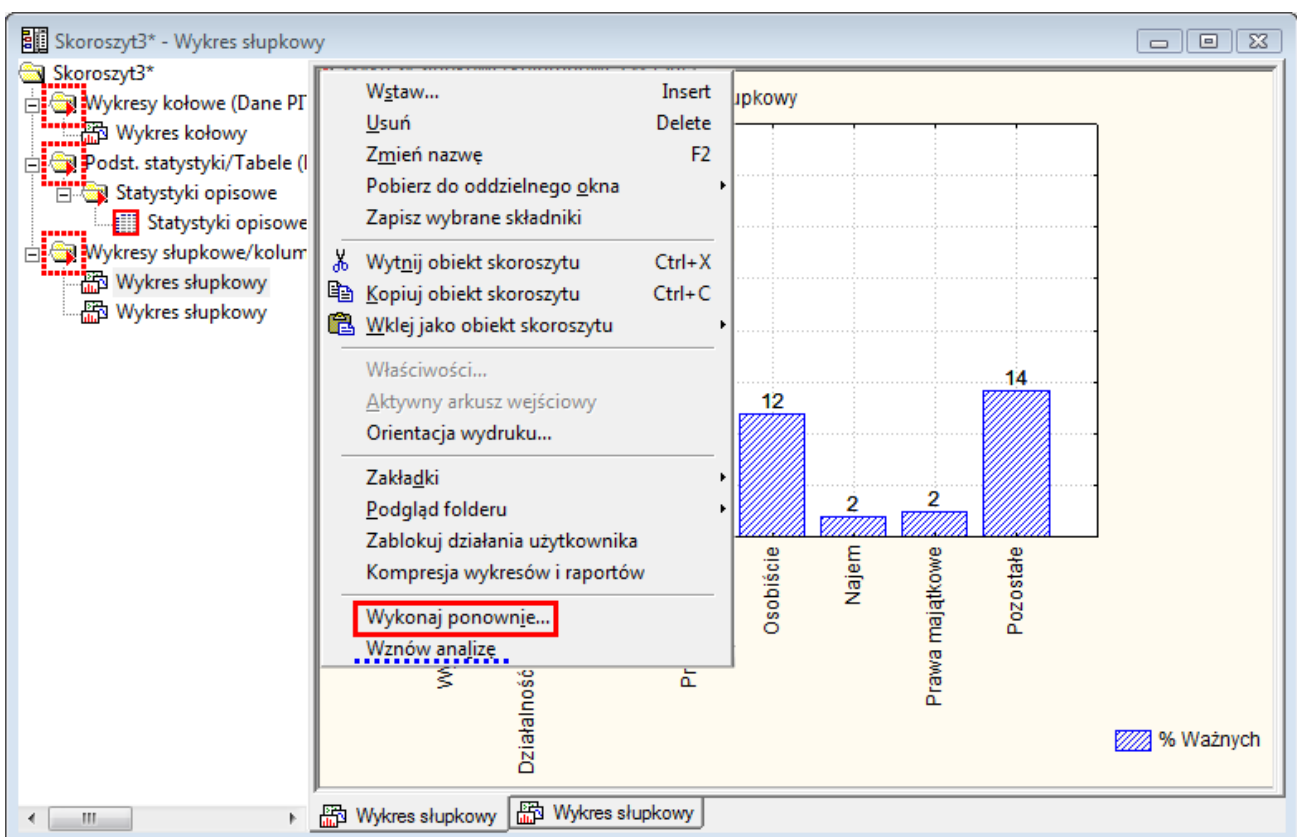
## Skoroszyt jako narzędzie automatyzacji

Skoroszyt *STATISTICA* jest dokumentem, do którego domyślnie trafiają wszystkie wyniki analiz (trafiły do niego wszystkie powyższe analizy). Umożliwia on zachowanie każdego z wynikowych dokumentów w postaci osobnej zakładki. Dokumenty możemy organizować w hierarchiczną strukturę folderów (domyślnie dla każdej analizy tworzony jest osobny folder). Struktura ta ma postać drzewa ułatwiającego zarządzanie dokumentami, folderami i gałęziami. Wybraną grupę dokumentów możemy wydobyć ze skoroszytu (np. przeciągając myszą) i umieścić ją na pulpicie *STATISTICA* (tzn. pustym obszarze głównego okna programu, w którym *STATISTICA* wyświetla osobne okna).

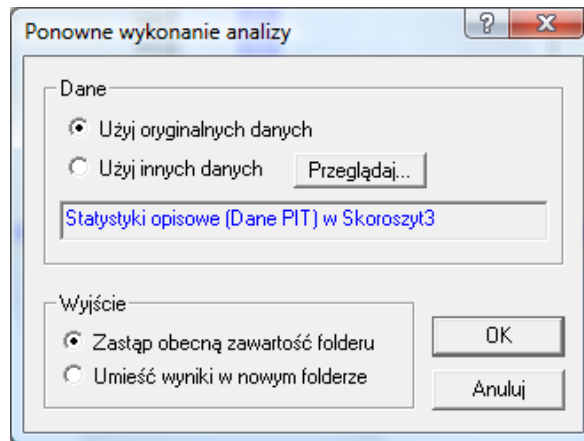


Oprócz łatwego sposobu zarządzania wynikami analizy Skoroszyt posiada jeszcze jedną ważną funkcjonalność. Wyniki umieszczone w Skoroszycie zawierają także informację o sposobie w jaki zostały wygenerowane. Dzięki temu w każdej chwili można wykonać ponownie lub wznowić analizę na tym samym zbiorze danych, na przykład po usunięciu błędnych obserwacji, lub na innym zbiorze, ale mającym taką samą strukturę.

Załóżmy, że po wykonaniu raportu nastąpiła pewna korekta danych wejściowych i przygotowane przez nas raporty są nieaktualne. Aby je wykonać ponownie, możemy oczywiście je „wyklikać” w sposób analogiczny do przedstawionego powyżej, jeżeli jednak dysponujemy starym skoroszytem z wynikami, możemy wygenerować nowe wyniki, klikając prawym przyciskiem myszy na folderze skoroszytu z wynikami, i z wyświetlonego menu podręcznego wybierając opcję **Wykonaj ponownie...**



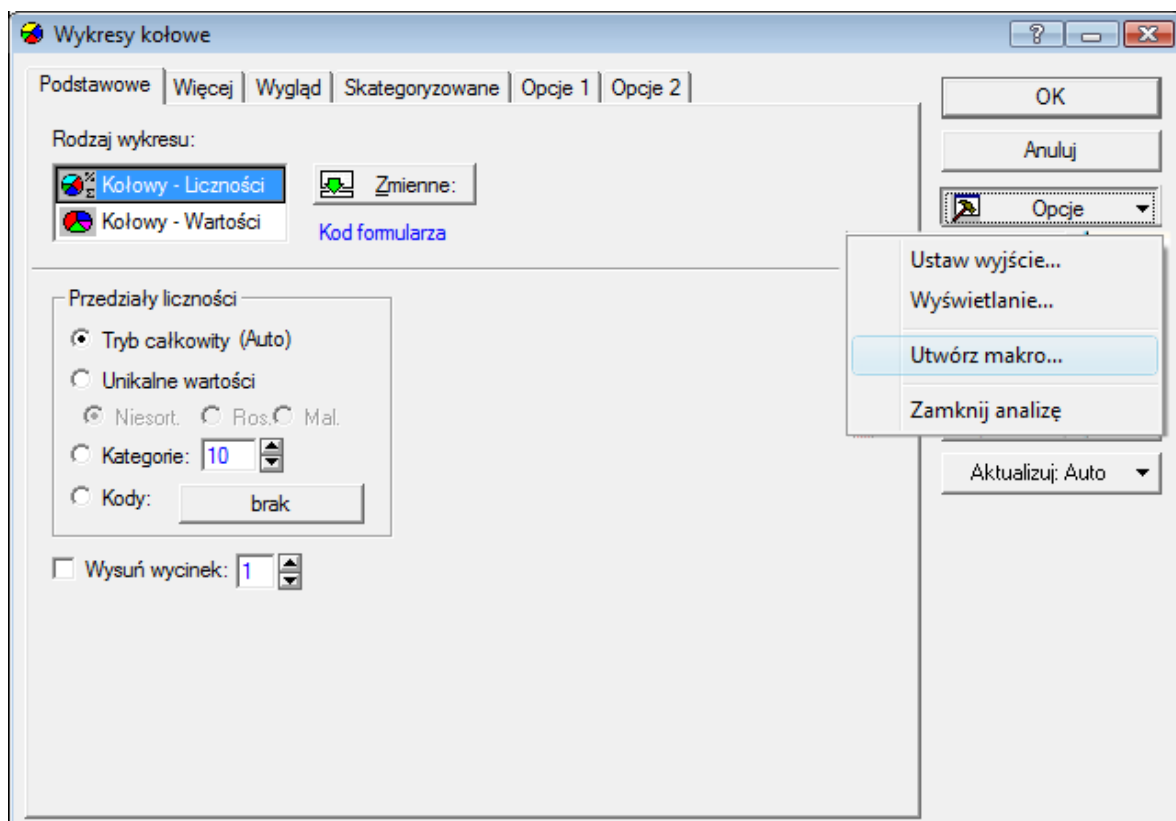
Po wybraniu opcji **Wykonaj ponownie...** wyświetlone zostanie okno **Ponowne wykonanie analizy**, w którym możemy określić, czy chcemy analizę wykonać ponownie na oryginalnych czy innych danych, oraz czy chcemy zastąpić obecne wyniki nowymi, czy też nowe wyniki chcemy umieścić w osobnym folderze.



Po zatwierdzeniu analizy zostaną wygenerowane nowe raporty, które zostaną umieszczone w pierwotnym skoroszycie.

## Makra *STATISTICA*

Każdy z programów z rodziny *STATISTICA* ma wbudowane środowisko programistyczne *STATISTICA* Visual Basic (zgodne ze standardowym MS Visual Basic), które umożliwia programowemu dostęp do praktycznie każdego elementu i każdej funkcji programu. Użytkownicy systemu *STATISTICA* mają więc dostęp do każdego elementu systemu z poziomu interfejsu programistycznego. Co istotne osoby niemające praktyki w programowaniu również mogą w prosty sposób korzystać z możliwości automatyzacji, jaką dają makra.

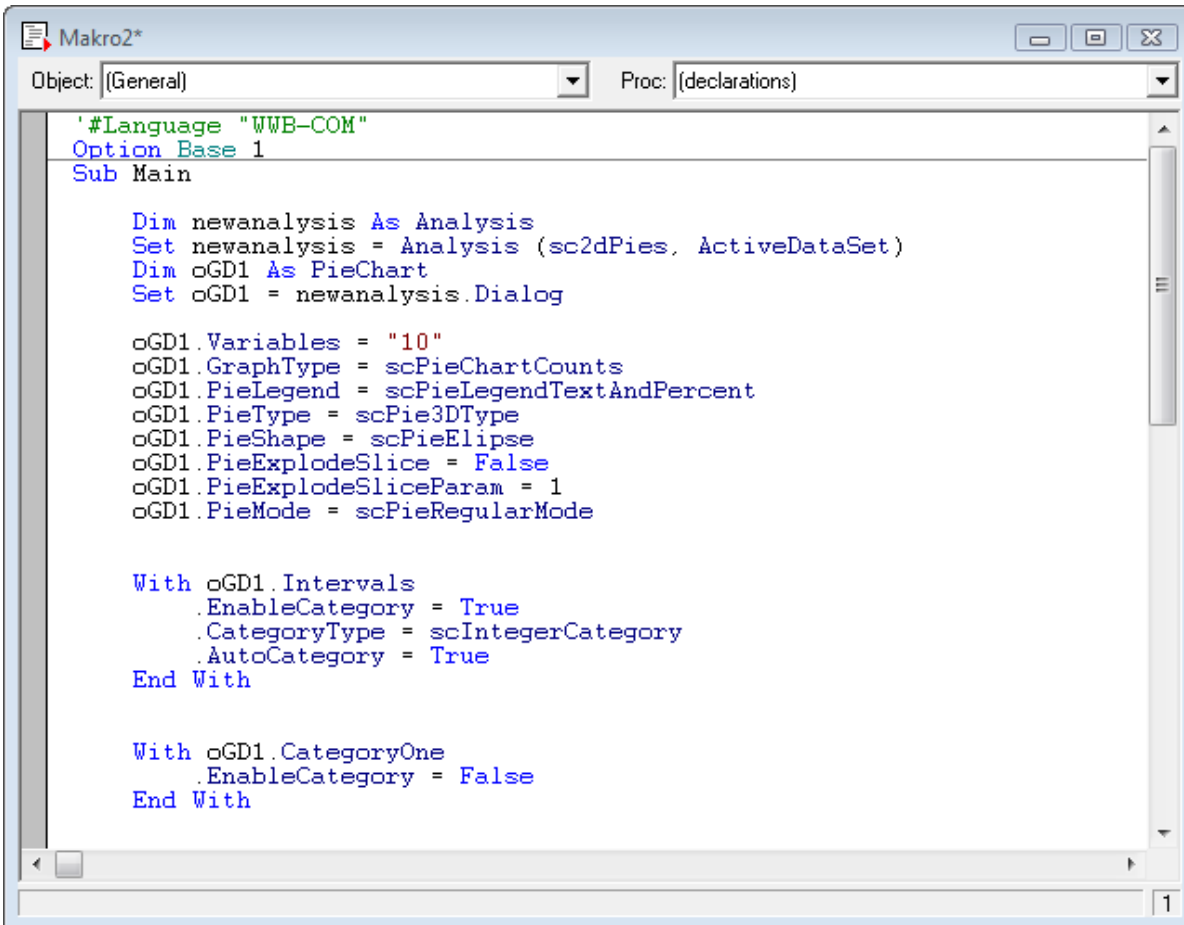




Bardzo pomocny w pracy z makrami jest mechanizm rejestracji interakcyjnego procesu generowania raportu. Każda analiza statystyczna czy wykres w oknie dialogowym ma przycisk **Opcje**, za pomocą którego możemy wygenerować makro dokładnie odpowiadające wykonanej wcześniej analizie.

Dla przykładu powrócimy do wykresu kołowego, który przygotowaliśmy, by zobrazować rozkład wniosków składanych przez podatników.

W oknie analizy wykresu klikamy przycisk **Opcje** i z wyświetlonego menu podręcznego wybieramy przycisk **Utwórz makro...** Wygenerowane w ten sposób makro stanowiło będzie zapis naszej analizy z dokładnie tymi samymi opcjami, które wybraliśmy podczas tworzenia wykresu w sposób interakcyjny.



```
'#Language "VBA"
Option Base 1
Sub Main


    Dim newanalysis As Analysis
    Set newanalysis = Analysis (sc2dPies, ActiveDataSet)
    Dim oGD1 As PieChart
    Set oGD1 = newanalysis.Dialog

    oGD1.Variables = "10"
    oGD1.GraphType = scPieChartCounts
    oGD1.PieLegend = scPieLegendTextAndPercent
    oGD1.PieType = scPie3DType
    oGD1.PieShape = scPieEllipse
    oGD1.PieExplodeSlice = False
    oGD1.PieExplodeSliceParam = 1
    oGD1.PieMode = scPieRegularMode

    With oGD1.Intervals
        .EnableCategory = True
        .CategoryType = scIntegerCategory
        .AutoCategory = True
    End With

    With oGD1.CategoryOne
        .EnableCategory = False
    End With

End Sub
```

Aby uruchomić takie makro, wystarczy wcisnąć przycisk **F5**, bądź kliknąć przycisk  znajdujący się na pasku narzędzi. Wygenerowane makro możemy uruchamiać na dowolnym zbiorze danych, a wyniki umieszczone zostaną w skoroszycie. Oczywiście osoby znające język Visual Basic mogą dowolnie modyfikować wygenerowane makro, wzbogacając je o własne procedury oraz korzystać z funkcji innych programów udostępnionych z poziomu interfejsu programistycznego (na przykład funkcji pakietu MS Office).

Przygotowane makro możemy także umieścić na pasku zadań i wykonywać je, klikając na odpowiedni element utworzonego menu.



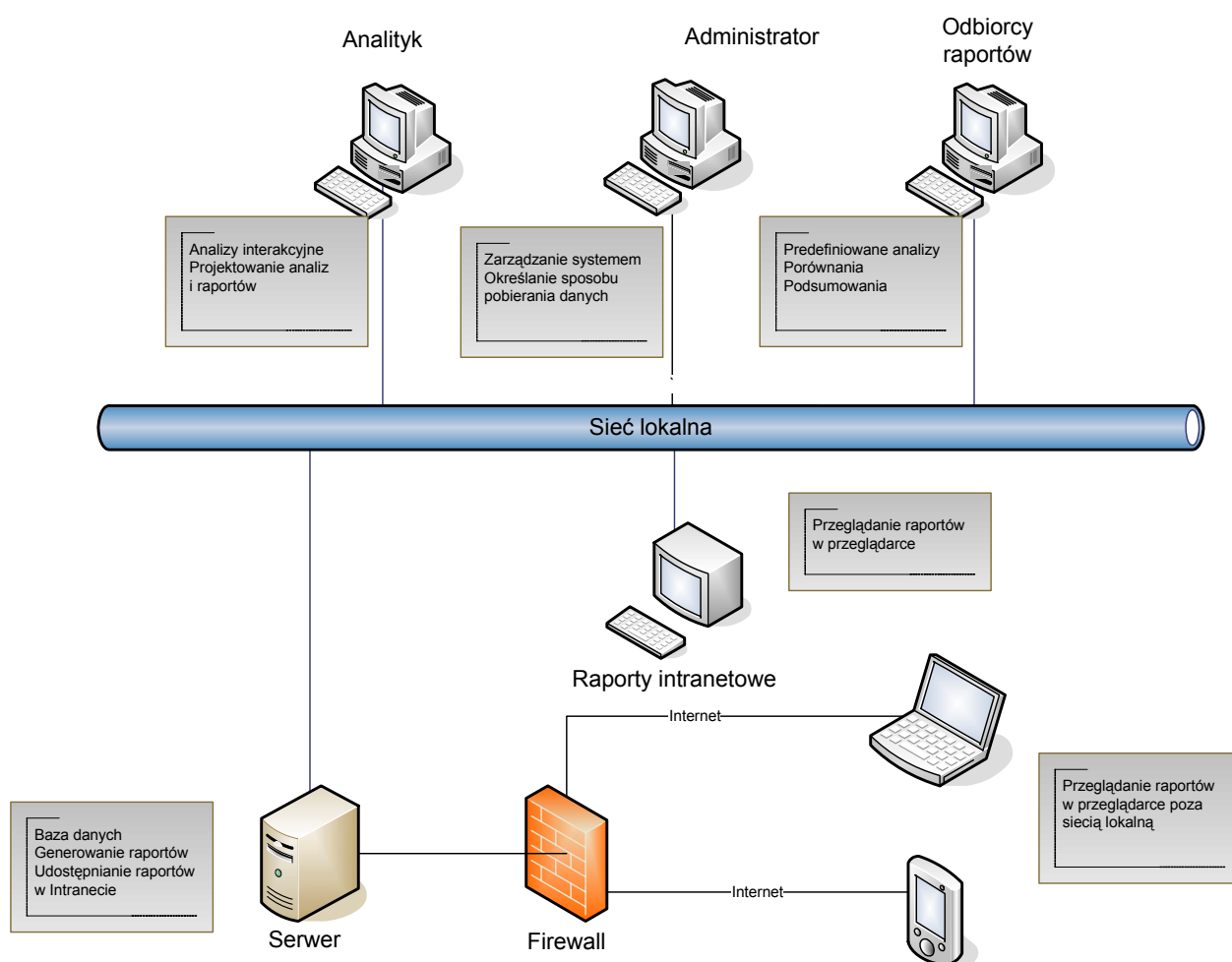
## STATISTICA Enterprise

*STATISTICA Enterprise* to specjalna wersja *STATISTICA* przeznaczona do wykorzystania przez zespoły ludzi, pracujących wspólnie i potrzebujących wygodnego, wydajnego i bezpiecznego środowiska do analizy danych i raportowania. Ponieważ *STATISTICA Enterprise* łączy w sobie narzędzia do przeprowadzenia całego procesu przekształcania surowych danych w użyteczną wiedzę, nazywamy go zintegrowanym systemem analizy danych.

W skład systemu wchodzi:

- ◆ centralna baza danych,
- ◆ narzędzia analityczne *STATISTICA* (statystyka, wizualizacja i data mining),
- ◆ generator raportów,
- ◆ system bezpiecznego przechowywania dokumentów zgodny z wymaganiami FDA CFR Część 11 (ta część jest opcjonalna).

Poglądowy schemat działania systemu i główne role użytkowników przedstawia poniższy rysunek.





Rdzeniem zintegrowanego systemu jest centralna baza danych. Możemy w niej przechowywać:

- ◆ szablony pobierania danych z zewnętrznych źródeł danych (konfiguracje danych),
- ◆ scenariusze analizy danych (konfiguracje analizy),
- ◆ szablony raportów,
- ◆ uprawnienia dostępu użytkowników do poszczególnych funkcji i danych systemu.

Warto zwrócić uwagę na przechowywanie wszystkich danych w centralnej bazie danych. Dzięki temu mamy zapewnioną dostępność i zgodność danych, a potrzebne ustawienia i zmiany wystarczy wykonać raz. Jeśli przykładowo zmieniamy szablon raportu, np. dodając do niego nowy wykres, to wystarczy wykonać to raz i zapisać nowe ustawienia w bazie danych. Zmiana ta jest potem automatycznie przenoszona na wszystkie stanowiska. Ponadto dane wprowadzone na jednym stanowisku są od razu dostępne dla użytkowników na innych stanowiskach.

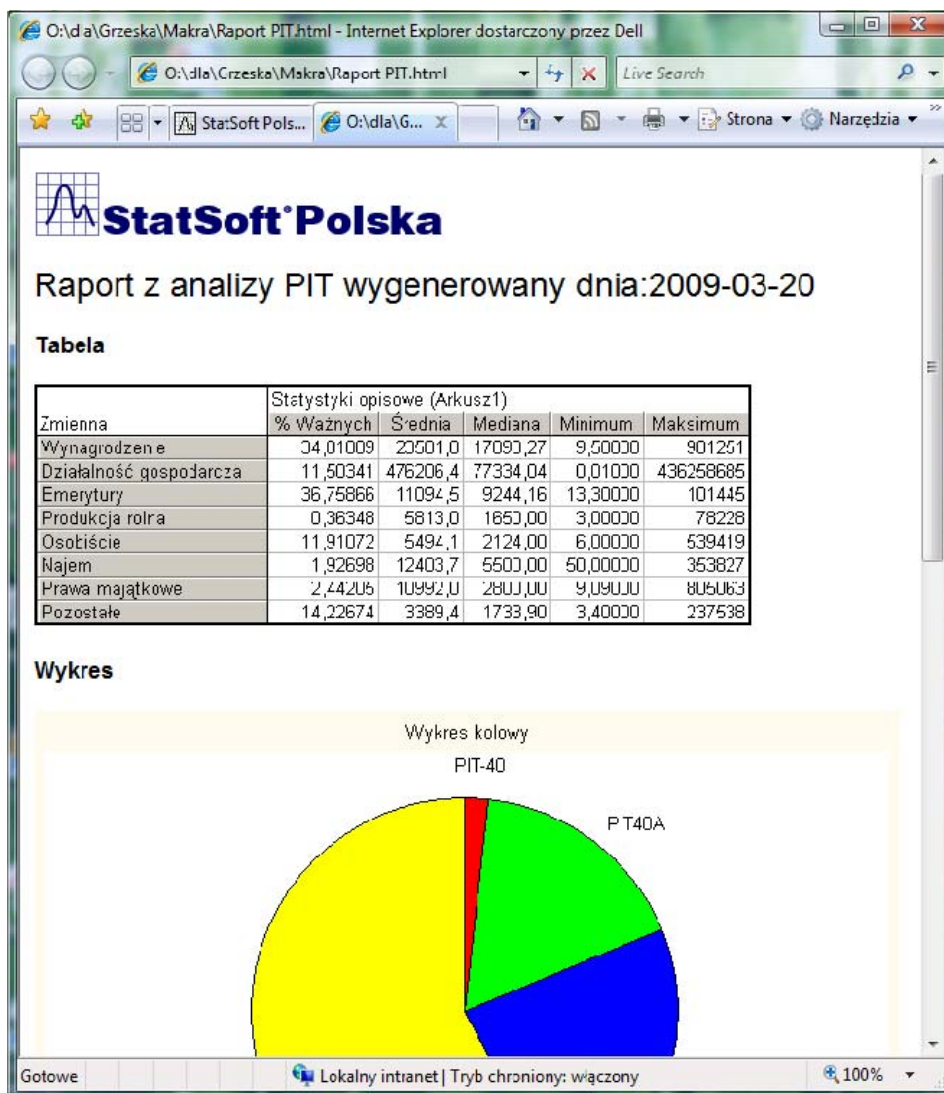
Z praktycznego punktu widzenia warto zwrócić uwagę na aspekt łatwości korzystania z wyników analiz w *STATISTICA Enterprise*. Otóż w praktyce najczęściej większość osób nie wykonuje analiz, a jest odbiorcą ich wyników. Takim osobom dużo łatwiej jest korzystać z szablonów analizy przygotowanych przez specjalistów od analizy danych i wdrożonych w *STATISTICA Enterprise*, niż własnoręcznie określać i uruchamiać analizę w zwykłym (tj. niewbudowanym w system) programie do analizy danych.

Generator raportów służy do uruchamiania całej analizy w zadanych odstępach czasu i przekształcania jej wyników na powszechnie zrozumiały format (np. PDF), tak aby mógł z nich korzystać praktycznie każdy. Układ i formatowanie raportu określane jest podczas jego tworzenia przez użytkownika. Oprócz wyników analizy (tabel, wykresów i informacji tekstowych) raport może być wzbogacony:

- ◆ dodatkowymi danymi: datą i czasem utworzenia, nazwą użytkownika, który uruchomił raport, wersją raportu itp.,
- ◆ tekstami (np. tytułem) i obiektami graficznymi (np. logo, schemat produkowanego elementu, którego dotyczy analiza).

Generator raportów jest w stanie tworzyć raport spełniający nawet bardzo wygórowane wymagania dotyczące jego zawartości, ale również bezpieczeństwa danych i monitorowania wszelkich zmian. Przykładowy raport przedstawia rysunek przedstawiony poniżej.

Generator raportów może w praktyce okazać się niezwykle pożyteczny. Standaryzacja i automatyzacja tworzenia zestawień pozwala spożytkować czas analityków na twórcze wykorzystanie wiedzy o danych, zamiast na wykonywanie żmudnych i powtarzalnych czynności. Zastosowanie takiego systemu oszczędza tysiące godzin pracy rocznie (przykładowo przygotowanie zestawienia podsumowującego w jednej z firm przed wdrożeniem systemu bazującego na *STATISTICA Enterprise* trwało 40 godzin, a po wdrożeniu 5 min!). Oprócz bezpośredniej oszczędności czasu i środków, zastosowanie systemu zmniejsza ryzyko popełnienia błędu i daje dokładniejszy wgląd w dane, poprzez dostępność bardziej złożonych metod analitycznych.



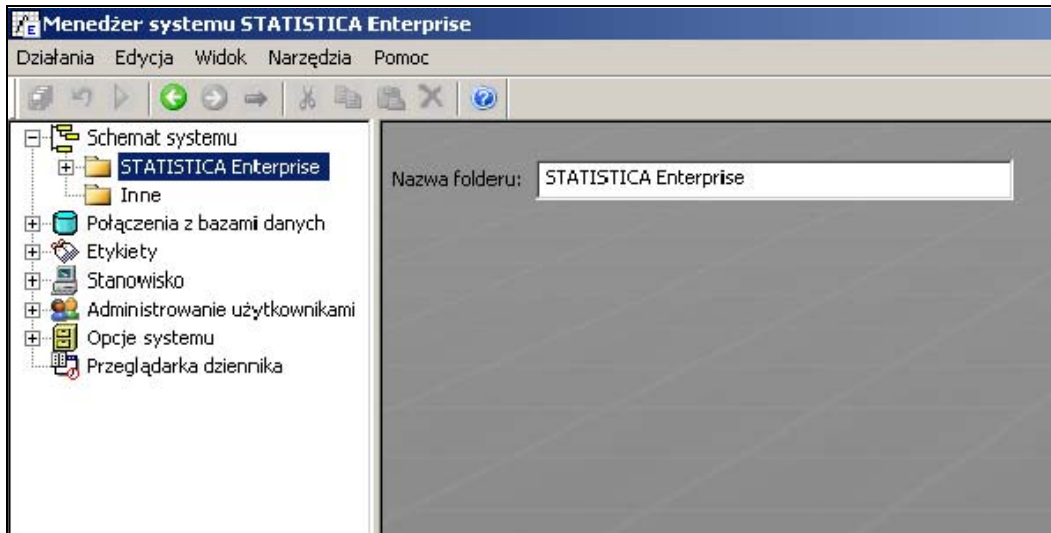
## Przykład budowy raportu w *STATISTICA Enterprise*

Przygotowane uprzednio makra *STATISTICA* po zapisaniu w postaci pliku wykorzystamy do stworzenia szablonów analiz oraz szablonów raportów *STATISTICA Enterprise*.

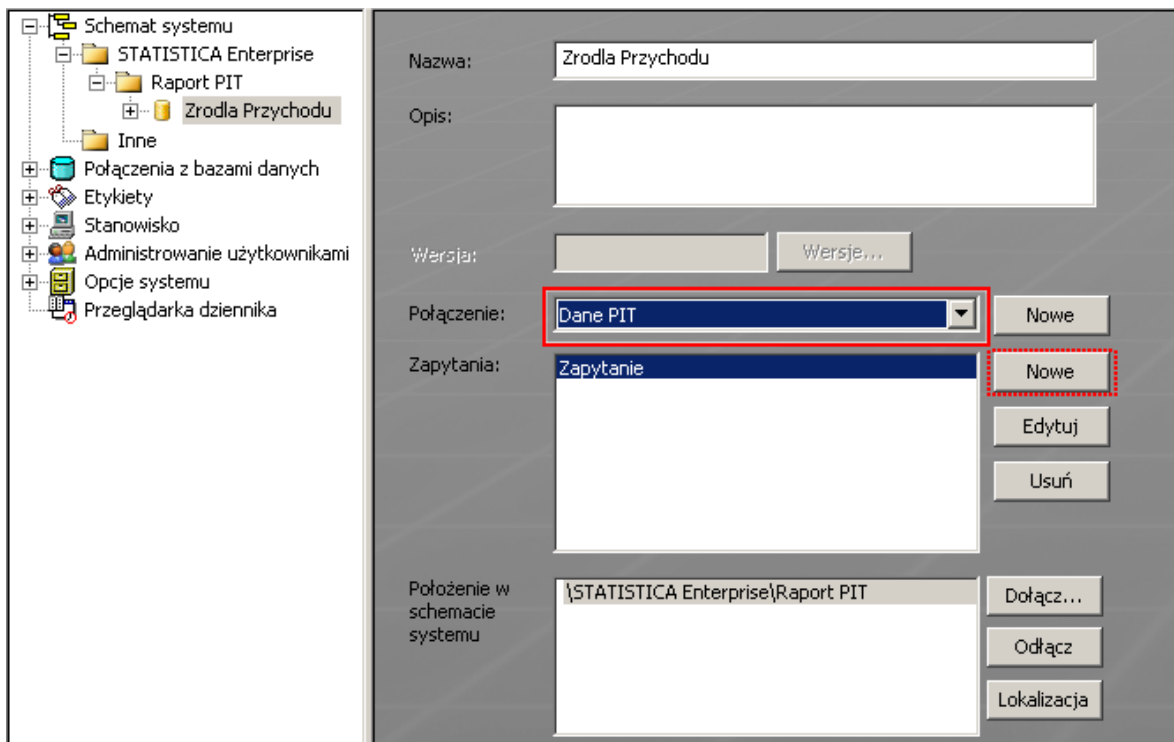
Użytkownicy *STATISTICA Enterprise* mają do dyspozycji *Menedżera systemu* prezentującego obiekty systemu będące do naszej dyspozycji. Okno tej aplikacji podzielone jest na dwa panele: w lewym widzimy uporządkowany hierarchicznie schemat systemu, a w prawym mamy ustawienia dla aktualnie podświetlonego obiektu. Panel schematu systemu obsługujemy analogicznie jak *Eksplorator Windows*.

Aby móc automatycznie wykonywać raporty utworzymy trzy rodzaje obiektów systemu:

- ◆ konfigurację danych,
- ◆ konfigurację analizy,
- ◆ szablon raportu.




Pracę rozpoczynamy od utworzenia nowej **konfiguracji danych**. Konfiguracja danych to szczegółowy przepis na pobieranie danych z zewnętrznego źródła danych. W skład takiego przepisu wchodzi informacja: skąd pobierać dane, które wielkości odczytywać, jakie kryteria mają one spełniać (np. za jaki mają być okres) oraz jaka jest interpretacja wielkości (np. czy dane pole jest polem daty czy wartością liczbową).

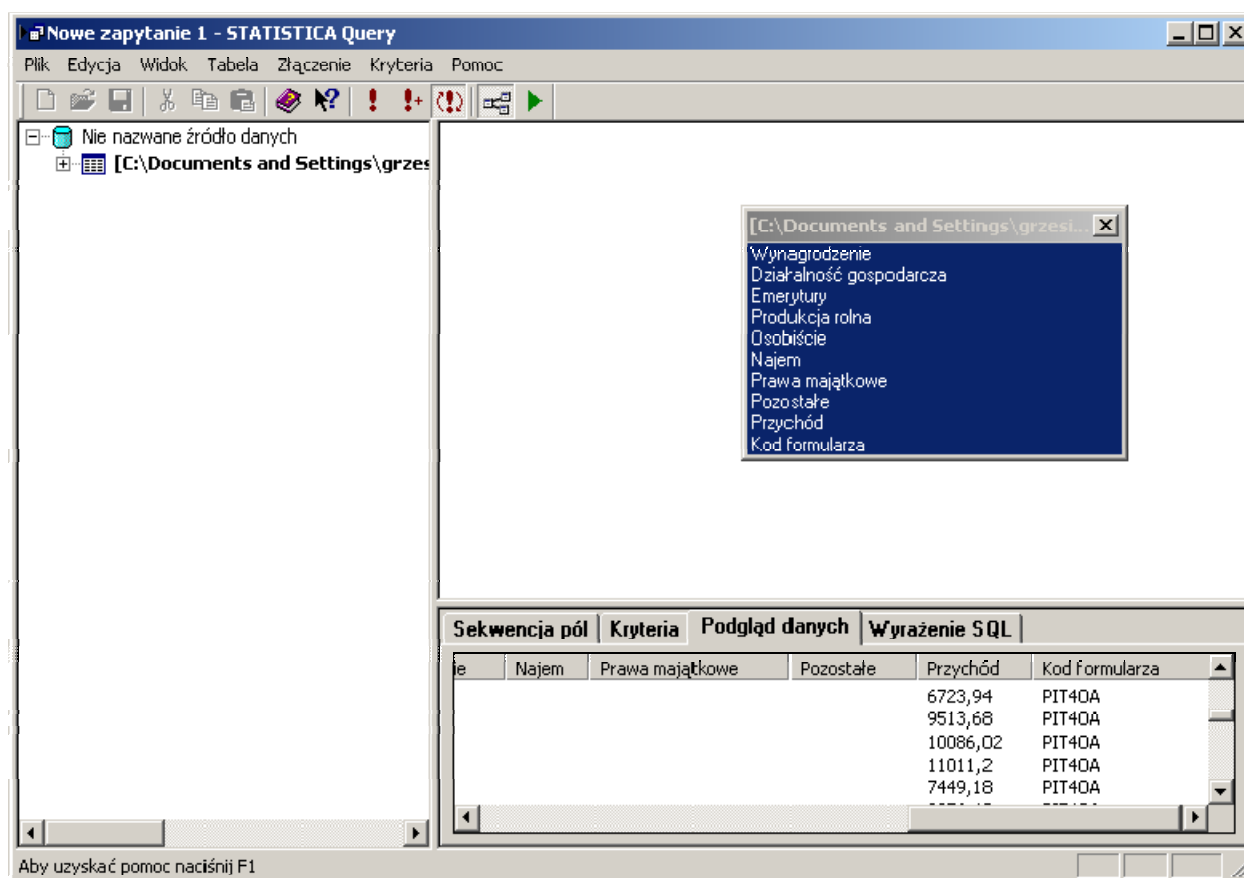


Nową konfigurację danych umieścimy w folderze *Raport PIT*. Podświetlamy go w schemacie systemu i klikamy prawym przyciskiem myszy, następnie z menu podręcznego wybieramy polecenie *Nowa konfiguracja danych*. Następny krok to nazwanie tworzonego obiektu: ponieważ będziemy pobierać dane z zeznań PIT dotyczące źródeł przychodu, konfigurację nazwiemy *Źródła przychodu*. Następnie z listy *Połączenia* wybieramy połączenie



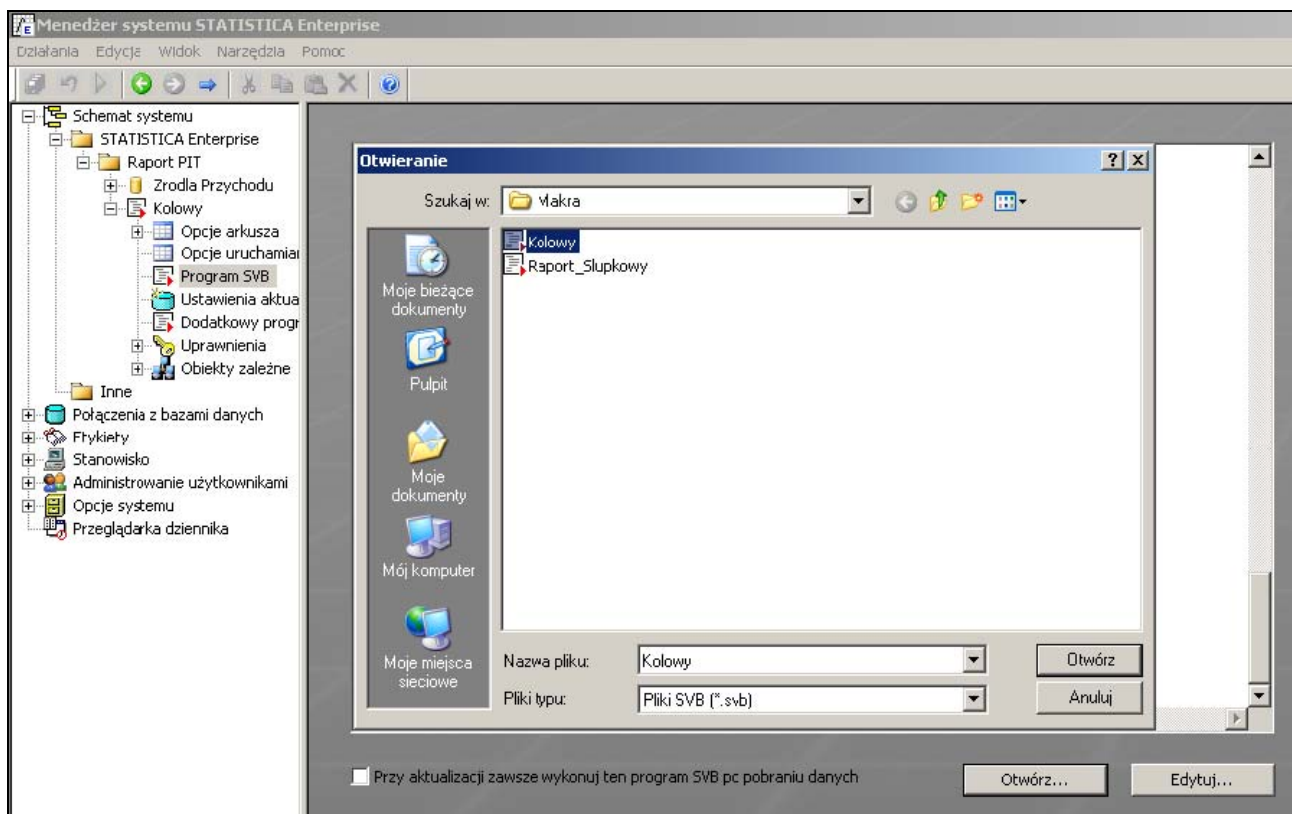
do bazy danych *Dane PIT* (połączenie określa, z jaką bazą danych się łączymy i w jaki sposób, m.in. z jakiego sterownika będziemy korzystać).

Po wybraniu połączenia klikamy przycisk *Nowe* w obszarze *Zapytania*. Program poprosi nas o określenie zapytania do bazy danych. Zapytanie można wpisać ręcznie lub skorzystać z *Kreatora SQL*, co zazwyczaj jest znacznie wygodniejsze. Po kliknięciu przycisku *Kreator SQL* na ekranie otworzy się okno *STATISTICA Query*, w którym w trybie graficznym możemy określić zapytanie do bazy danych. W naszym przypadku wybieramy wszystkie pola z jedynej dostępnej dla nas tabeli bazy danych i klikamy przycisk , aby zapisać zapytanie i powrócić do *Menedżera systemu*.



Po wykonaniu powyższych czynności do zakończenia procesu definiowania konfiguracji danych wystarczy określić, kto może z niej korzystać – robimy to po kliknięciu węzła *Uprawnienia* dla konfiguracji *Źródła przychodu*.

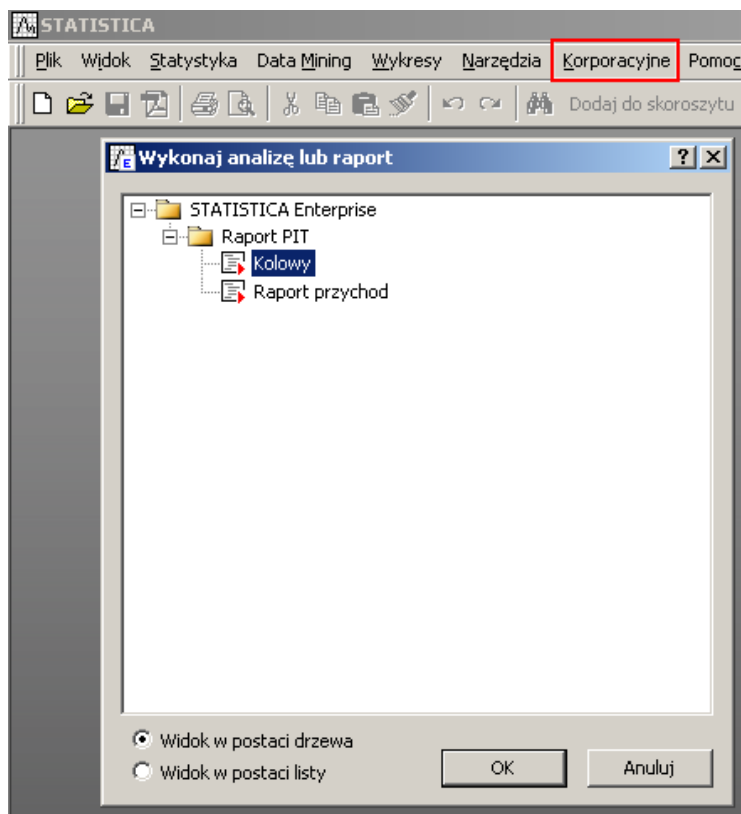
Mamy gotowe źródło danych do naszej analizy i możemy rozpocząć tworzenie samej analizy. Podświetlamy konfigurację *Źródła przychodu* i z menu *Działania* wybieramy polecenie *Nowa konfiguracja analizy dla tych danych*. Program utworzy nową konfigurację danych i wyświetli panel z jej ogólnymi właściwościami. Zmieniamy w nim tylko nazwę konfiguracji na *Kołowy* oraz typ analizy na *Analiza SVB*.



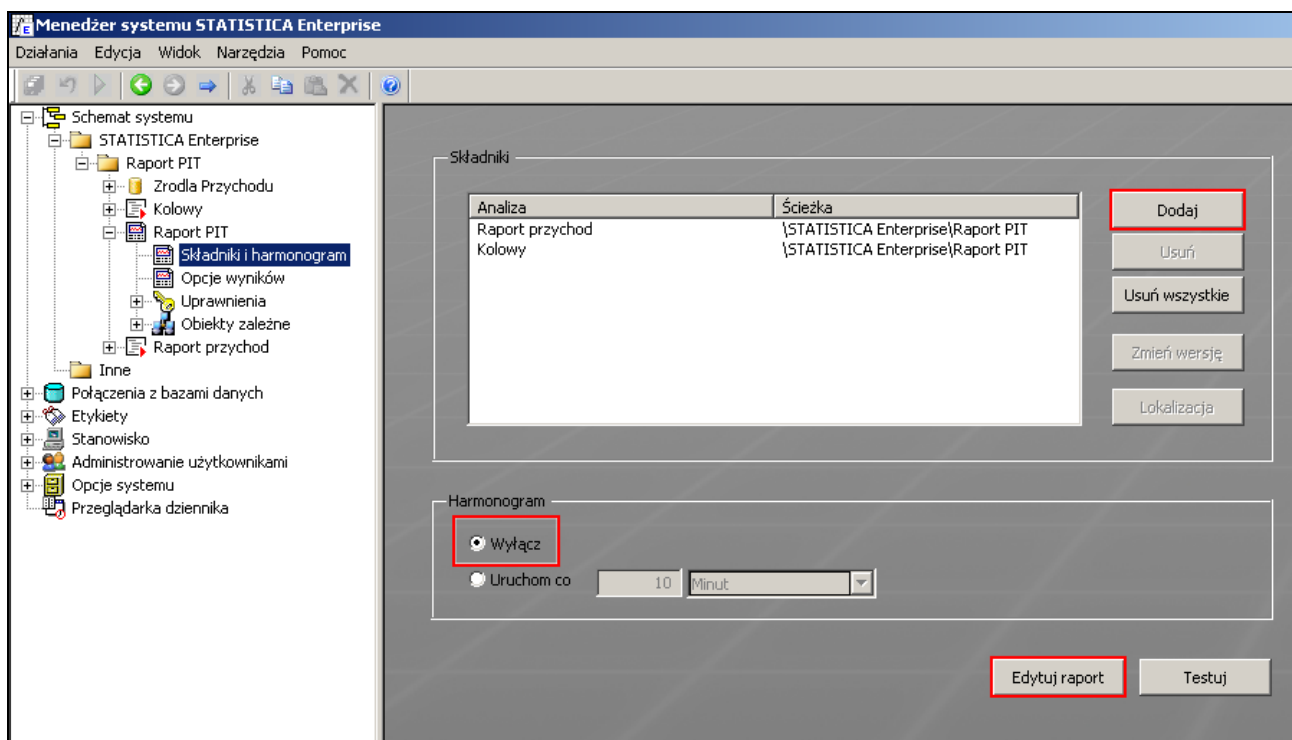
Kolejną czynnością będzie wczytanie do konfiguracji analizy makra rysującego wykres kołowy, które przygotowaliśmy w poprzedniej części analizy. W tym celu w węźle *Program SVB* konfiguracji analizy wybieramy przycisk *Otwórz...*, a następnie wskazujemy odpowiedni plik makra. Po zapisaniu uprawnień do edycji i wykonywania konfiguracji analizy szablon jest gotowy do pracy. Podobne czynności wykonujemy następnie dla makr wykonujących podsumowania oraz wykresy słupkowe.

Uzyskane w ten sposób szablony analiz możemy uruchamiać w oknie *Menedżera systemu Enterprise*, wybierając je ze schematu systemu i klikając przycisk . Ważniejszy z punktu widzenia codziennej pracy z systemem jest fakt, iż taki szablon analizy jest również dostępny z poziomu programu *STATISTICA*. Aby uruchomić raport rysujący wykres kołowy wystarczy skorzystać z opcji *Uruchom analizę lub raport* z menu *Korporacyjne* w programie *STATISTICA*. Wybranie tej opcji spowoduje wyświetlenie okna *Wykonaj analizę lub raport*, w którym dostępne będą wszystkie szablony analiz (oraz raportów), do których mamy uprawnienia. Aby wykonać analizę, wystarczy abyśmy wybrali interesujący nas obiekt i kliknęli *OK*.

Dodatkowo na podstawie utworzonych konfiguracji analiz możemy przygotować **szablon raportu**, w którym znajdować się będą wybrane wyniki konfiguracji analiz oraz elementy formatujące, jak: logo firmy, nagłówki, tytuły czy też data.

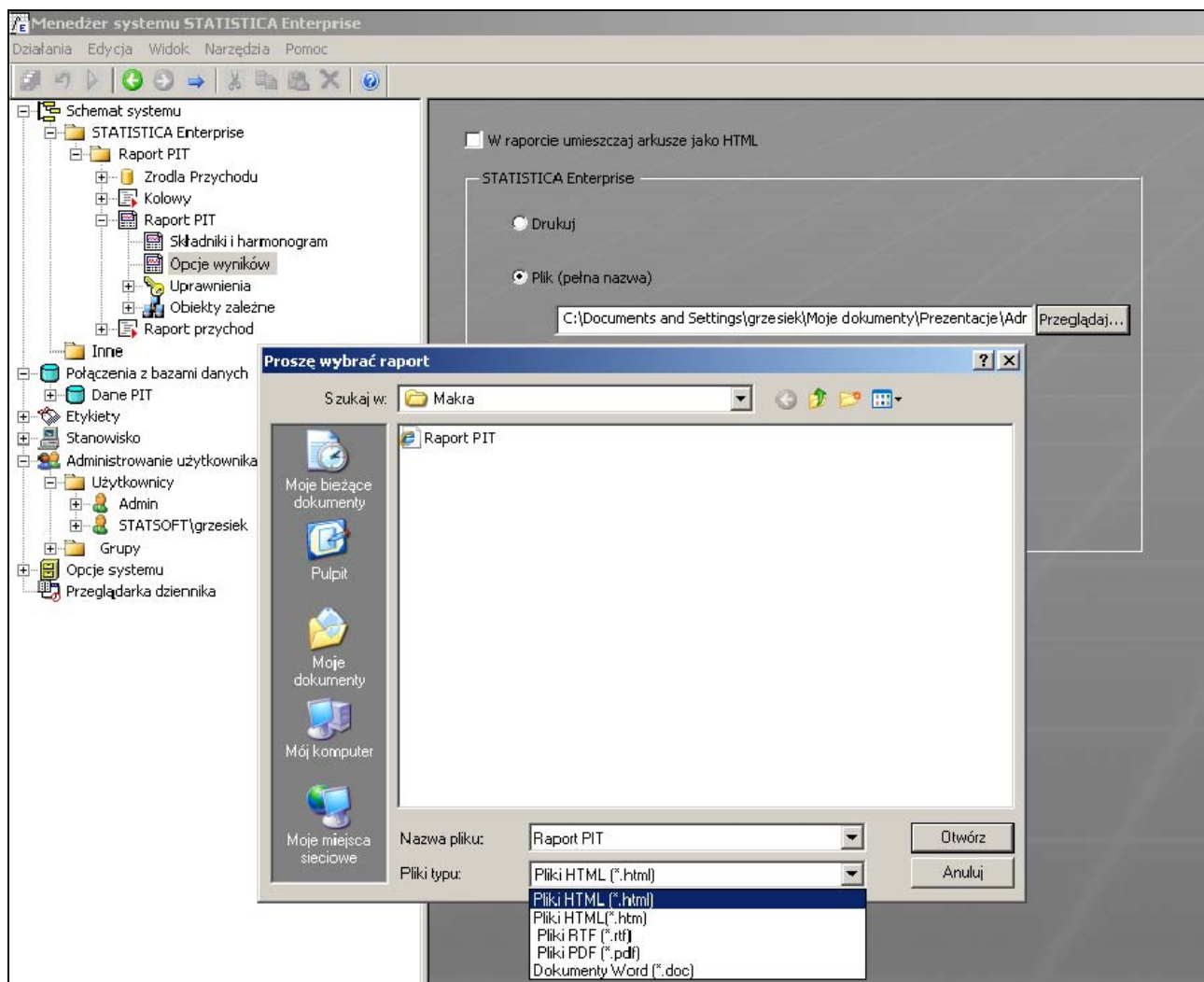


Podobnie jak w przypadku poprzednich szablonów, aby utworzyć szablon raportu, klikamy prawym przyciskiem na folderze *Raport PIT* w oknie *Menedżera systemu STATISTICA Enterprise* i z menu podręcznego wybieramy polecenie *Nowy raport*. W nowo utworzonym raporcie wpisujemy nazwę *Raport PIT*, a następnie przechodzimy do *Składników harmonogramu*, gdzie wskazujemy, z jakich konfiguracji analiz będziemy korzystać.





Aby dodać do raportu konfigurację analizy, wybieramy przycisk *Dodaj*, a następnie wskazujemy jedną bądź większą liczbę konfiguracji analiz. Następnie określamy, by był to raport generowany na żądanie, a nie uruchamiany cyklicznie – w obszarze *Harmonogram* wybieramy opcję *Wylącz*. Kolejny krok to przygotowanie samego raportu; klikamy przycisk *Edytuj raport*, co spowoduje uruchomienie programu *STATISTICA* wraz z aktywnym dokumentem raportu, który możemy edytować w sposób podobny do edytorów tekstowych, możemy wstawiać do niego nagłówki, elementy graficzne i oczywiście wyniki analiz pobrane ze wskazanych uprzednio konfiguracji analiz.

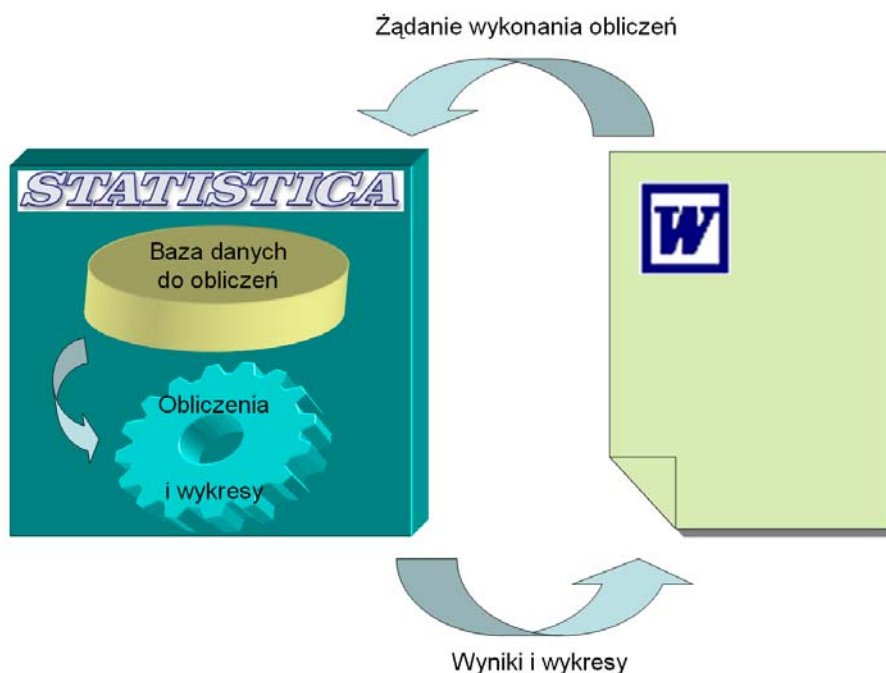


Po przygotowaniu szablonu raportu przechodzimy do węzła *Opcje wyników*, gdzie określamy, w jakie miejsce i do jakiego formatu generowane będą końcowe raporty. Utworzony automatycznie raport może być zapisywany jako plik PDF, RTF, HTML bądź DOC. My wybierzemy format HTML i dodatkowo wskażemy, aby raport po wygenerowaniu był otwierany w przeglądarce internetowej. Przygotowany raport możemy następnie udostępnić zainteresowanym osobom. Co ważne, w przypadku jego modyfikacji zmiana będzie upowszechniana w całej organizacji, dzięki czemu mamy gwarancję zgodnych wyników niezależnie od stanowiska pracy czy analityka.



## Dodatkowe możliwości automatyzacji – gotowy szablon MS Word

Poza powyższymi standardowymi możliwościami automatyzacji procesu raportowania istnieje także możliwość przygotowania szablonu raportu w postaci dokumentu MS Word. Program *STATISTICA* wykorzystywany jest wtedy jako motor analityczny. Po wykonaniu obliczeń wszystkie wyniki przesyłane są następnie do dokumentu MS Word i tam formatowane. Ideowy schemat systemu przedstawia poniższy rysunek:



W programie MS Word zdefiniowano szablon raportu oraz procedury, jakie powinien wykonać program *STATISTICA*. Po uruchomieniu raportu szablon raportu łączy się z programem *STATISTICA*, otwierając zapisany w nim zbiór danych do obliczeń, a następnie uruchamia proces analizy. Wyniki analiz przeprowadzonych w *STATISTICA* są następnie przesyłane z powrotem do szablonu raportu, gdzie są umieszczane i odpowiednio formatowane. W wyniku działania szablonu otrzymujemy dokument MS Word gotowy do druku.