

SAS PRE VEDU A VÝSKUM

SEMINÁR PRE CVTI



19.11.2015
SAS Slovakia, s.r.o.



- Za vznikom softvéru SAS stojí jeden zo zakladateľov spoločnosti SAS a jej dnešný prezident **Dr. James H. Goodnight**.
- Prvé moduly softvéru vyvinul v 60. rokoch minulého storočia za účelom analýzy poľnohospodárskych dát na Štátnej Univerzite v Severnej Karolíne v USA.
- SAS udržiava intenzívne spojenie s akademickou oblasťou už od svojich začiatkov (vyše 3000 univerzít).



- SAS bol založený v r. **1976** v USA, na Slovensku v r. **1995**
- **Najväčšia** súkromne vlastnená softvérová spoločnosť na svete
- **1.** poskytovateľ dátovej analytiky
- **9.** súčasná verzia softvéru
- **75 000** zákazníkov (v **140** krajinách sveta)
- **13 800** zamestnancov po celom svete
- **93 z top 100** firiem rebríčka Fortune Global 500 používa SAS
- **23 %** z obratu investuje do výskumu a vývoja
- Oceňovaný ako TOP (**No. 1 v roku 2012**) zamestnávateľ na svete (rebríček Fortune)



Cieľom prvej časti je predstaviť niekoľko spoločností vo svete, ktoré využívajú produkty SAS v konkrétnych oblastiach výskumu, výroby a riadenia.

V ďalšej časti predstavíme aktivity akademického programu spoločnosti SAS na Slovensku.

Niektoré z produktov SAS predstavíme v tretej časti seminára detailnejšie a predvedieme prácu s nimi.

Využitie nástrojov SAS vo svete

BIOGRID AUSTRALIA

- Spoločnosť zameriavajúca sa na **zdravotnícky výskum**, zber a výmenu relevantných dát, ktoré sú rozhodujúce pri diagnózach.
- Spravovanie dát, ich bezpečnosť a etickosť zohrávajú najdôležitejšiu úlohu
- Využívanie **SAS® Enterprise Guide®**
- **Zjednotenie dát z rôznych zdrojov** a ich sprístupnenie pre vedcov
http://misapps.unx.sas.com/data/cosmos-images/105867_0812.pdf

NOVOMER

- Americká spoločnosť zameriavajúca sa na **ekologický chemický priemysel**
- **Zachytávanie odpadových CO₂ a CO plynov** na princípe katalyzácie
- Využíva **JMP štatistický softvér** od SASu na **tvorbu experimentov**
- JMP pomáha spoločnosti **sledovať reakcie v rôznych podmienkach** a vytvára tak **základy pre optimalizáciu**
http://misapps.unx.sas.com/data/cosmos-images/105424_1011.pdf

LENOVO

- Popredný svetový výrobca počítačov a tabletov
- Analytické riešenie **SAS® Advanced Analytics a Visual Analytics**
- Zníženie času odhaľovania chýb o polovicu
- Zníženie nákladov spojených s reklamáciou o 10 – 15 %
- **Pokles hovorov do call centra o 30 – 50 %**

http://www.sas.com/en_us/customers/lenovo.html

VOLVO GROUP

- Svetový líder na trhu vo výrobe nákladných vozidiel
- Cieľ: rýchlejšia výmena informácií a plynulejšia komunikáciou medzi dcérskymi spoločnosťami skupiny a **predpovedanie porúch na vozidlách a ich včasná diagnostika**
- Nástroje: **SAS® Enterprise Guide® a SAS® Fiel Quality Analytics**

http://www.sas.com/en_us/customers/volvo.html

TELEFÓNICA O2 GERMANY

- Poskytovateľ telekomunikačných služieb na území Nemeckej spolkovej republiky
- Potreba **predpovedania budúceho vývoja** pre účely firemného plánovania a manažmentu
- Riešenie: **SAS® Forecast Server**
- **Zvýšenie presnosti predpovedí a zrýchlenie vývoja**
http://misapps.unx.sas.com/data/cosmos-images/106406_0413.pdf

CEMIG

- Najväčší dodávateľ elektrickej energie v Brazílii
- Využíva **SAS® Enterprise Miner®**
- **Odhaľovanie a znižovanie finančných strát** spojených s krádežou elektriny, účtovnými chybami a poruchami elektromerov
- Úspora 420 000 \$ mesačne
http://www.sas.com/en_us/customers/cemig-br.html

SAS® Clinical Data Integration

- *Využitie automatizácie a opakovateľnosti pri príprave analytických dát*
- Automatizácia procesov
- Viac priestoru pre zamestnancov na výkon komplexnejších úloh
- Rýchly prístup ku klinickým údajom z rôznych zdrojov
- Opätovné využívanie údajov a poznatkov z už uskutočnených analýz
- Podpora automatizácie zberu dát a ich využitie pre štandardizované procesy
- Rýchlejšia príprava podkladov na publikačnú činnosť
- Softvér je určený pre pracovníkov klinického výskumu, ktorí potrebujú zvýšiť výkonnosť, kvalitu a rýchlosť počas zberu, spravovania a analyzovania klinických údajov a informácií.

http://www.sas.com/en_us/industry/life-sciences/clincial-data-integration.html

Využitie nástrojov SAS vo výskume a vzdelávaní

ODBORNÉ PRÁCE SO SAS-OM

- od roku 2011 bolo odovzdaných 9 dizertačných prác so SAS-om
- EUBA, SPU, TUKE

SAS VO VÝSKUME

- Botanický ústav SAV
- Spoločenskovedný ústav SAV Košice (do roku 2013 samostatne licencovaný)
- Vedecké projekty na fakultách a katedrách so SAS licenciou

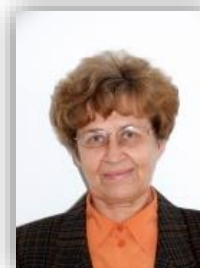
ODBORNÉ SAS PUBLIKÁCIE

- Štatistika v SAS-e, SAS v štatistike
(Ing. Renáta Prokeinová, PhD, 2013)
- Híbková analýza údajov
(Prof. Ing. Milan Terek, PhD., Ing. Mgr. Adriana Horníková,
PhD. a RNDr. Viera Labudová, PhD, 2010)
- Navrhovanie a vyhodnocovanie experimentov s aplikáciami
(Ing., Mgr. Adriana Horníková, PhD., 2009)



ODBORNÉ SAS PUBLIKÁCIE

- Regresná a korelačná analýza s aplikáciami
(Mgr. Erik Šoltés, PhD., 2008)
- Štatistické metódy pre ekonómov – zbierka príkladov
(Mgr. Erik Šoltés, PhD. a kolektív, 2015)
- Viacrozmerné štatistické metódy s aplikáciami
(Ing. Mária Vojtková, PhD., doc. Ing. Iveta Stankovičová, PhD., 2007)
- Štatistická indukcia pre ekonómov a manažérov
(Viera Pacáková a kolektív, 2015)



Akademický program SAS na Slovensku vytvoril dobré predpoklady pre vedecké využitie a prácu s nástrojmi SAS. Je k dispozícii skupina zaškolených študentov, pedagógov a pracovníkov, ktorí majú skúsenosti s používaním nástrojov SAS.

SAS VÝUČBA NA UNIVERZITÁCH (akad.rok 2014/2015)

- 19 rôznych SAS predmetov
- 6 fakúlt na 4 univerzitách (EUBA, UK, SPU a TUKE)
- spolu 1 424 absolventov predmetov
- z toho cca 20 zahraničných študentov v rámci výmenných programov – výuka v ANJ
- najpoužívanéjšie SAS moduly: SAS Base, SAS EG, SAS EM, SAS Stat

SAS CERTIFIKÁTY

- 101 certifikátov pre najlepších študentov
- 75 certifikátov pre absolventov SAS Academic školení

ZÁVEREČNÉ PRÁCE

- 18 licencií na Final thesis (z toho 5 dizertačných prác)
- 34 odovzdaných diplomových prác so SAS-om

SÚŤAŽ O NAJLEPŠIU PRÁCU

- organizuje sa každoročne za najlepšiu záverečnú prácu so SAS-om
- víherca sa spolu s pedagógom zúčastňuje SAS konferencie vo svete
- Rok 2010 – Las Vegas (Matej Hodek & doc.Ing. Iveta Stankovičová, PhD. – FM UK)
- rok 2011 – Orlando (Martin Fides & Ing. Renáta Benda Prokeínová, PhD. – FEM SPU)
- rok 2012 – San Francisco (Juraj Repík & RNDr.Viera Labudová, PhD. – FHI EU)
- rok 2013 – Washington (Peter Kertys & doc.Ing. Iveta Stankovičová, PhD. – FM UK)
- rok 2014 – Praha (Ing. Marián Hricko, PhD. & prof. Ing. Milan Terek, PhD. – FHI EU
a Ing. Ľubica Šoutová & doc. Mgr. Erik Šoltés, PhD. – FHI EU)

VÝHERCOVIA SÚŤAŽÍ O NAJLEPŠIU SAS PRÁČU



Prehľad nástrojov SAS

SAS PORTFÓLIO SAS BUSINESS ANALYTICS FRAMEWORK



- Rozsiahla zbierka štatistických, optimalizačných a analytických algoritmov
- Vlastný programovací jazyk (Datastep, DS2), podpora SQL
- Vizuálne nástroje pre biznis užívateľov
- Riadené centrálnymi metadátami
- Prevažne trojúrovňová architektúra s prístupom cez web prehliadač
- Podpora distribuovaného spracovania
- Vhodné aj pre Big data (Hadoop, in-memory analýza)
- Spracovanie dávkovo aj v reálnom čase
- Platformy Windows, Linux, AIX a ďalšie

- Enterprise Guide
- Enterprise Miner
- Model Manager
- JMP CLINICAL
- JMP GENOMICS



GARTNER: MAGICKÝ KVADRANT POKROČILÝCH ANALYTICKÝCH PLATFORMIEM

Figure 1. Magic Quadrant for Advanced Analytics Platforms

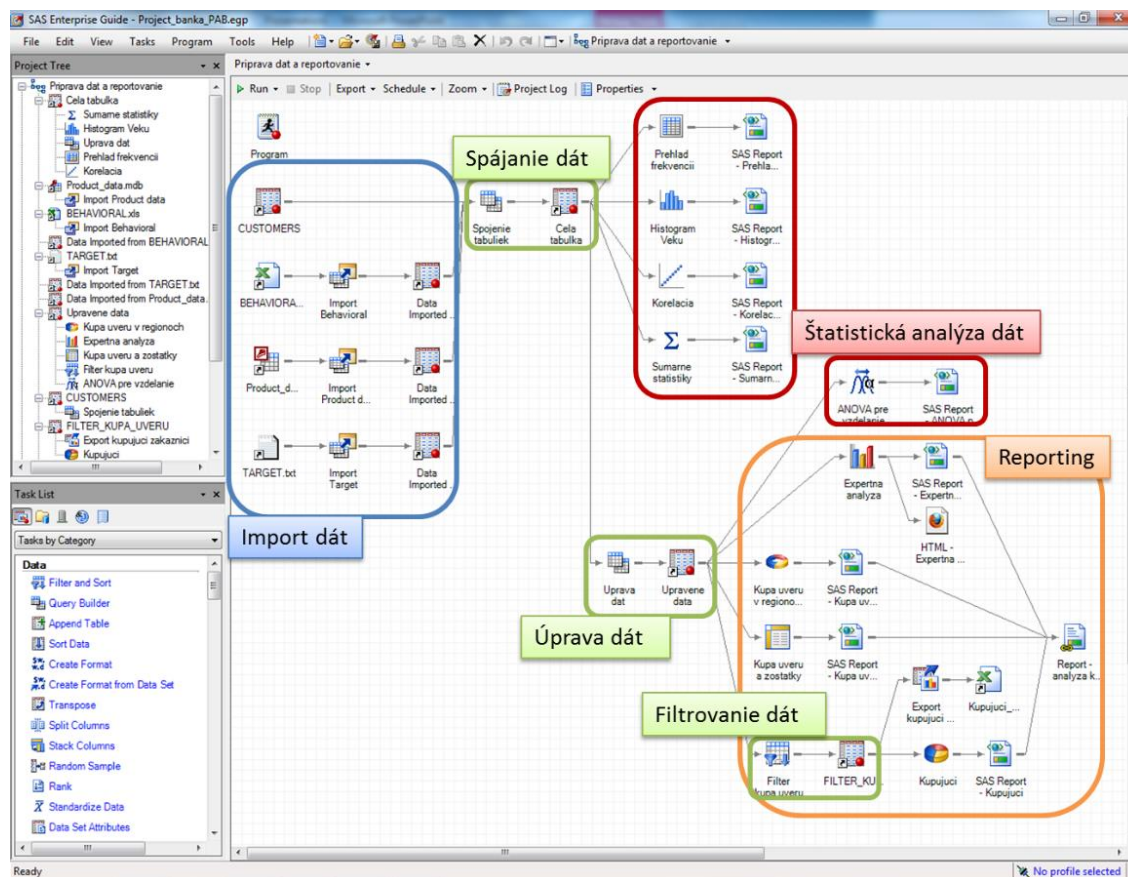


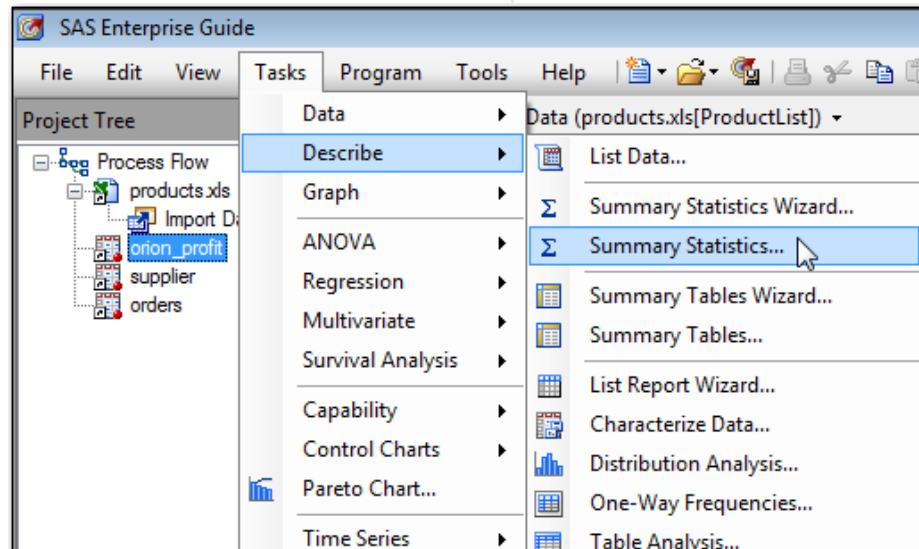
Source: Gartner (February 2015)

Gartner Magic Quadrant for Advanced Analytics Platforms by Gareth Herschel, Alexander Linden, Lisa Kart, 19 February 2015.

This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from SAS. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

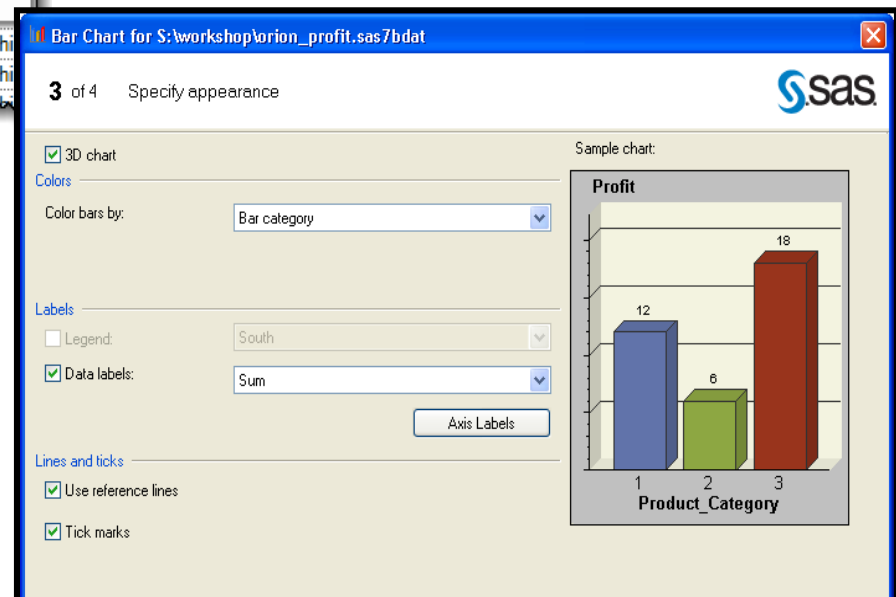
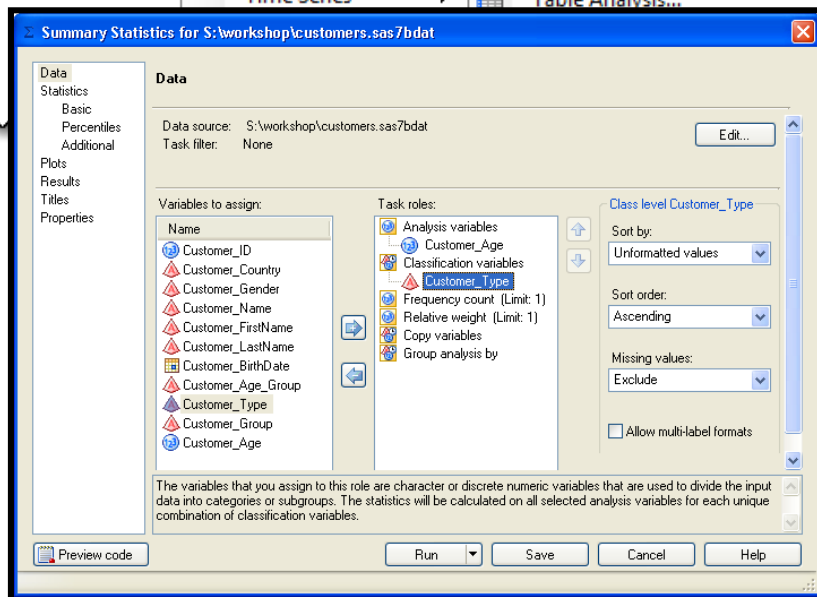
- Vizuálny návrh prípravy a analýzy dát
 - Import dát
 - Spájanie tabuliek
 - Filtrovanie a triedenie
 - Vypočítavané premenné
 - Sumarizácia
- Štatistické metódy
 - Distribúcia hodnôt
 - Hľadanie anomálií
 - Korelácia, regresia...
- Reporty, grafy
 - Zdokumentovanie postupu
 - Znovupoužiteľnosť

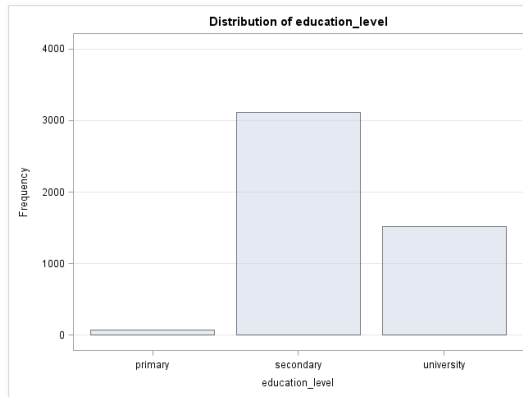




Univerzálne použitie:

- Dáta z biznis a technologických systémov
- Výsledky meraní
- Spracovanie dotazníkov
- Napríklad spájanie s open dátami





education_level	Frequency	Percent
primary	65	1.39
secondary	3113	66.35
university	1514	32.27

Frequency Missing = 248

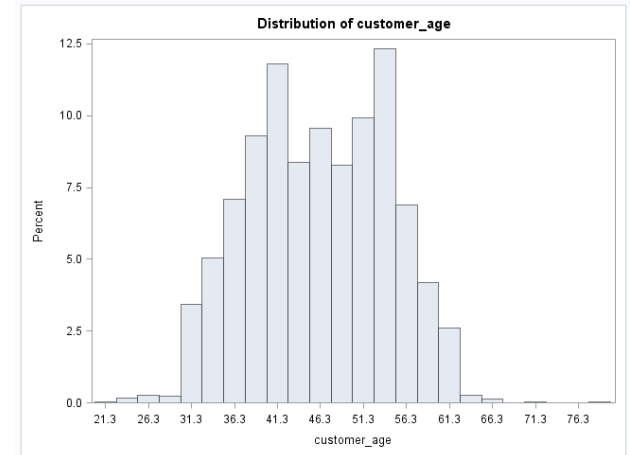


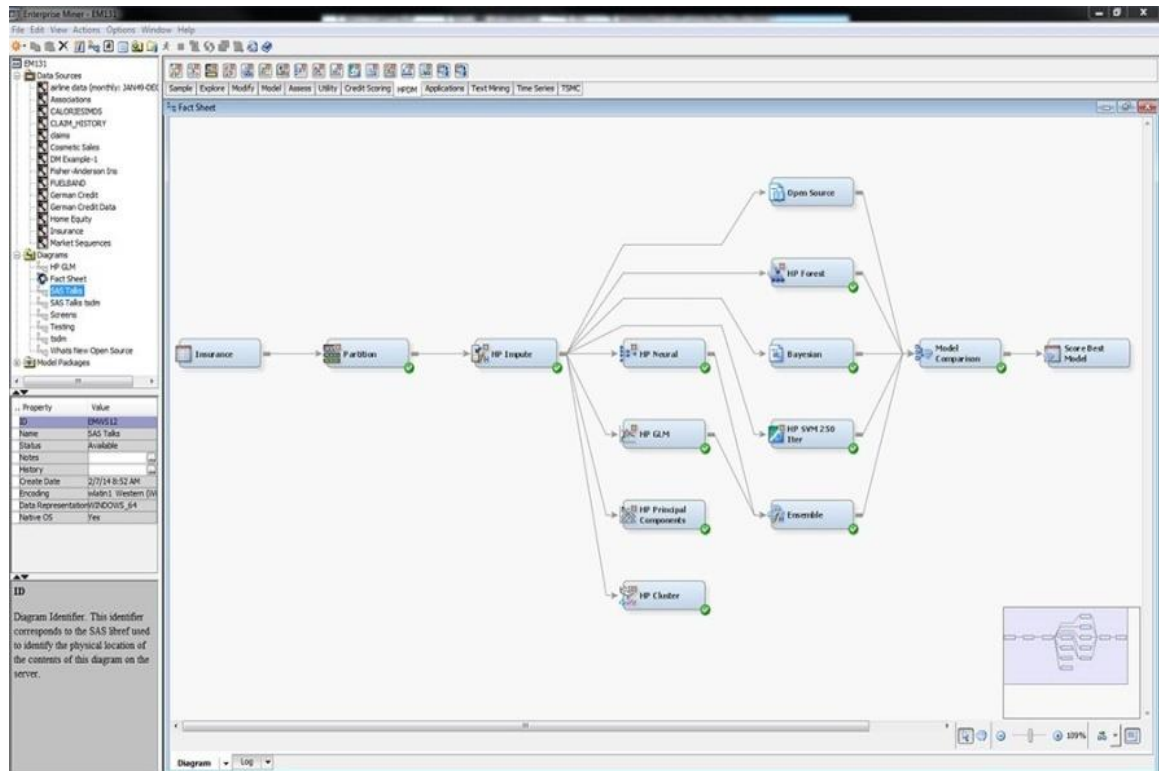
Table of marital_status by target_loan

		target_loan		Total
		0	1	
divorced	Frequency	219	422	641
	Row Pct	34.17	65.83	
married	Frequency	1918	1416	3334
	Row Pct	57.53	42.47	
single	Frequency	286	113	399
	Row Pct	71.68	28.32	
Total		2423	1951	4374

Frequency Missing = 566

Variable	Mean	Minimum	Maximum	1st Pctl	Lower Quartile	Median	Upper Quartile	99th Pctl
customer_age	46	21	78	30	39	46	53	62
current_account_amt	1103	-331014	437395	-2127	-577	5	1050	20185
saving_account_amt	999	0	569485	0	0	0	0	37068

- Komerčné aplikácie
 - Zlepšenie úspešnosti marketingu a predaja
 - Riadenie rizika
 - Detekcia podvodov
 - Predvídanie porúch zariadení
- Deskriptívna a prediktívna analytika
- Prehľadný vizuálny data miningový proces SEMMA
- Ľahko použiteľné a prispôsobiteľné drag-and-drop rozhranie



Predpoklady

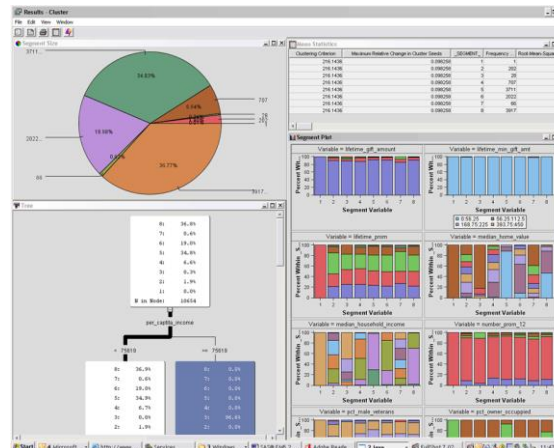
- Dáta vo forme tabuliek (definované riadky, stĺpce)

Príklady využitia

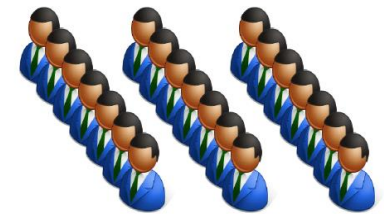
- Detekcia anomálií
- Identifikovanie súvislostí a prepojení
- Definovanie segmentov

Príklady analýz

- Analýza distribúcie, frekvenčné analýzy, sumárne štatistiky, sumárne tabuľky
- Korelačná analýza, zhuková analýza, asociačná analýza...



Pred analýzou



Po zhukovej analýze



Predpoklady

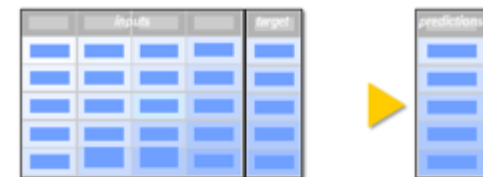
- Súčasné a historické dáta
- Sledovaný jav (predaj, porucha, podvod...)

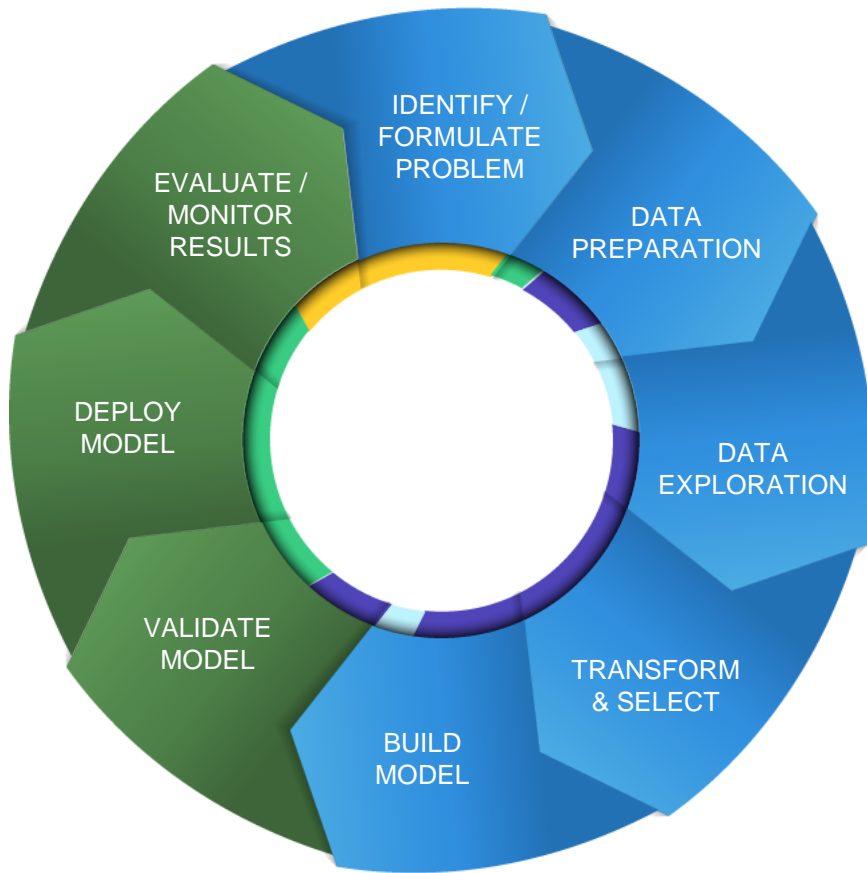
Príklady využitia

- Bezpečnosť liekov, manažment chronických chorôb, akútne stavy
- Riadenie kvality technologických procesov výroby
- Plánovanie údržby technológií

Príklady analýz

- Lineárna a logistická regresia
- Rozhodovacie stromy...

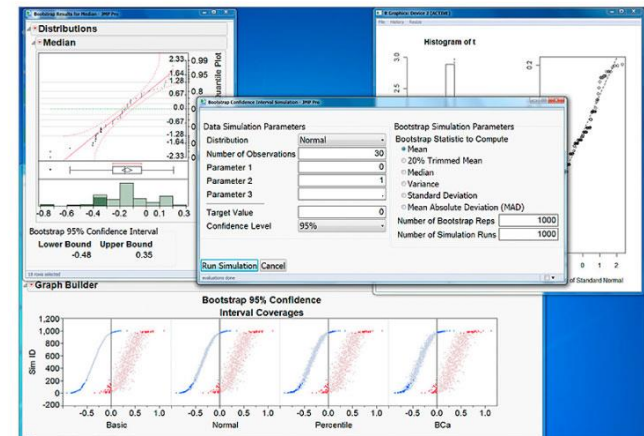
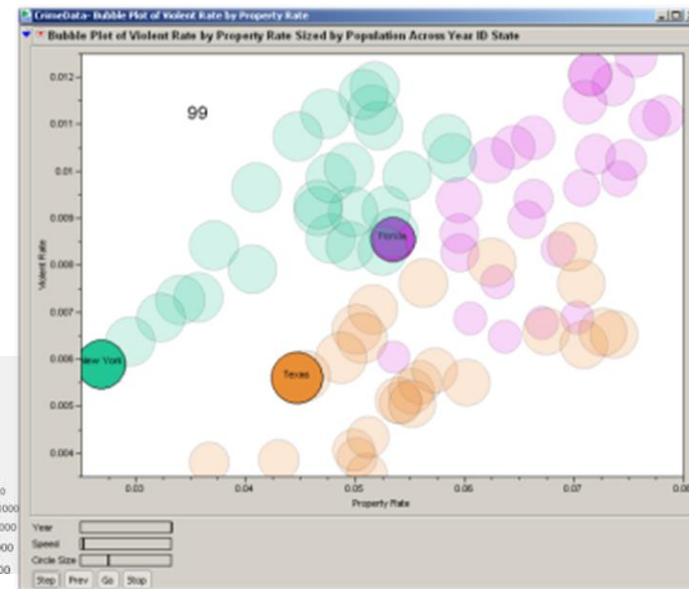
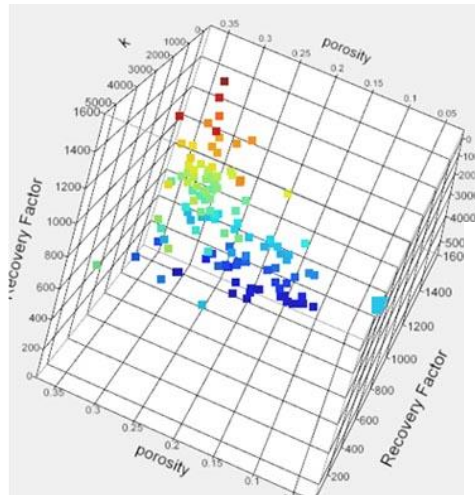


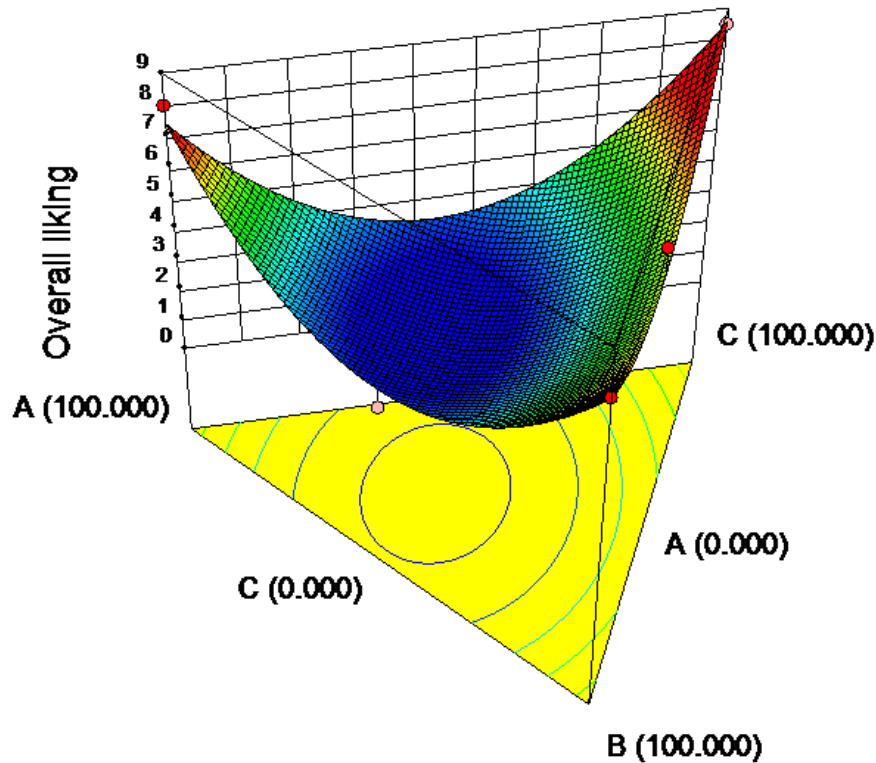


- Automatizácia
- Validácia
- Nasadenie a skóring
- Monitorovanie výkonnosti
- Workflow – spolupráca
- Auditovateľnosť



- Interaktívna vizualizácia dát
- 64 bitové in-memory spracovanie dát
- Deskriptívne a prediktívne modelovanie
 - Rozhodovacie stromy
 - Neurónové siete
- Návrh experimentov
- Integrácia s R a SAS
- Integrácia s Microsoft Office

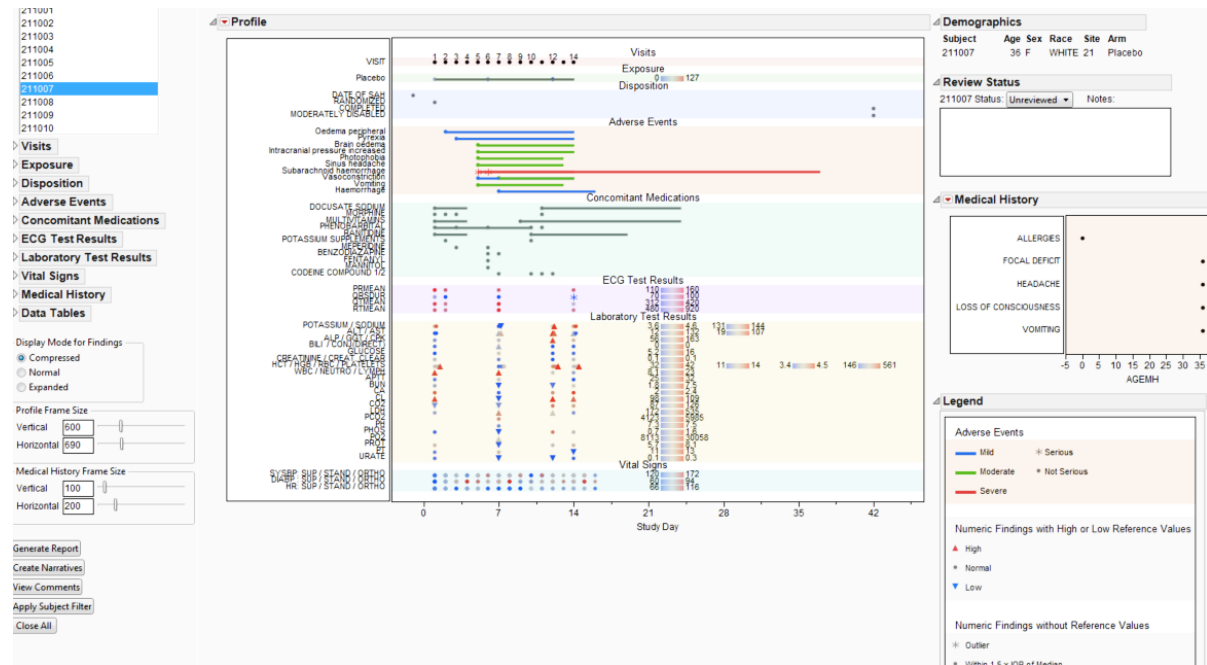




- Výrazné zníženie času a nákladov na prieskum a zlepšovanie javov, o ktorých nie je dosť dát
- Napríklad technologický proces ovplyvňujú faktory – tlak, teplota, vlhkosť, čas, parametre zariadenia
- Namiesto zmeny vždy iba jedného faktora sa menia súčasne všetky
- Výsledky sa štatisticky spracujú

SAS PORTFÓLIO JMP CLINICAL

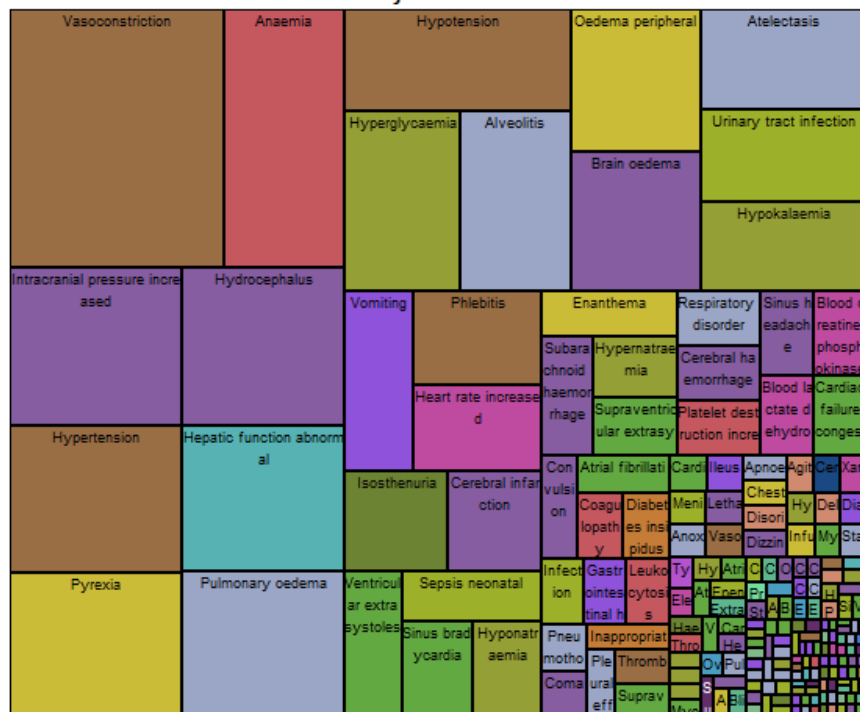
- Klinické skúšky nových liekov, vakcín a zdravotníckych zariadení
- Profily pacientov
- Opis pacientov
- Zdravotnícke monitorovacie dashboardy
- Poškodenia pečenie spôsobené liekmi
- Dashboard vedľajších účinkov



Displayed counts indicate the number of subjects experiencing an event

Graph Builder

Dictionary-Derived Term



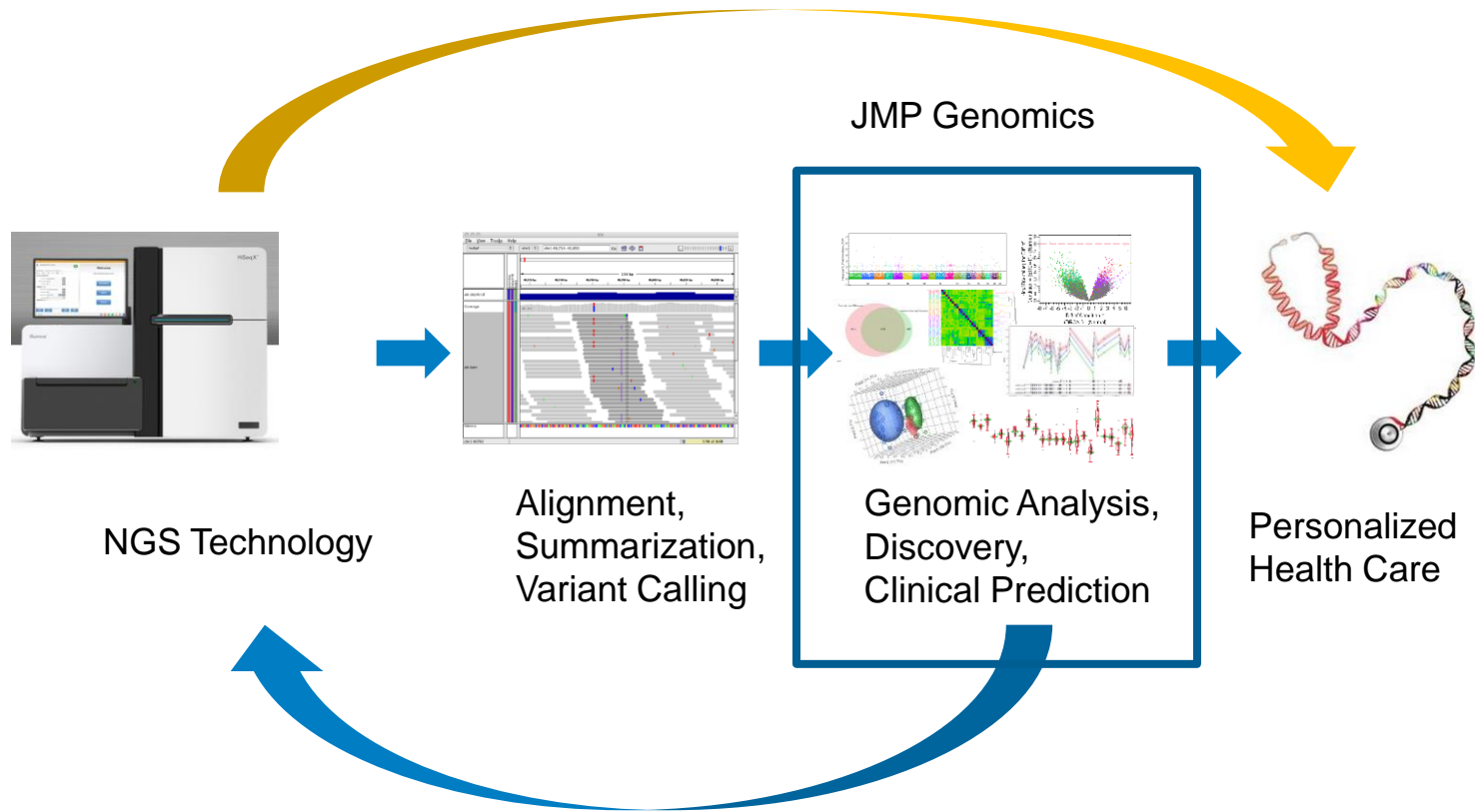
Dictionary-Derived Term ordered by Total Count (descending)

Tabulate

Dictionary-Derived Term	N
Vasoconstriction	560
Anaemia	312
Intracranial pressure increased	275
Hydrocephalus	259
Hypertension	254
Pyrexia	247
Hepatic function abnormal	237
Pulmonary oedema	236
Hypotension	235
Hyperglycaemia	210
Alveolitis	198
Oedema peripheral	185
Brain oedema	180
Atelectasis	166
Urinary tract infection	152
Hypokalaemia	146
Vomiting	127
Phlebitis	122
Heart rate increased	110
Isosthenuria	105
Cerebral infarction	94
Ventricular extrasystoles	85
Sepsis neonatal	70
Sinus bradycardia	67
Hyponatraemia	64
Enanthema	63
Subarachnoid haemorrhage	61
Hypernatraemia	52

- Cieľom sú klinicky využiteľné dáta o genóme
- Medicínske rozhodovanie s využitím údajov genómu
 - Panely biomarkerov
 - Prispôsobené terapie
- Personalizovaná nemocničná a klinická starostlivosť
- Informované rozhodnutia pre plátcov a poskytovateľov zdravotnej starostlivosti
- Využitie aj v potravinárskom priemysle



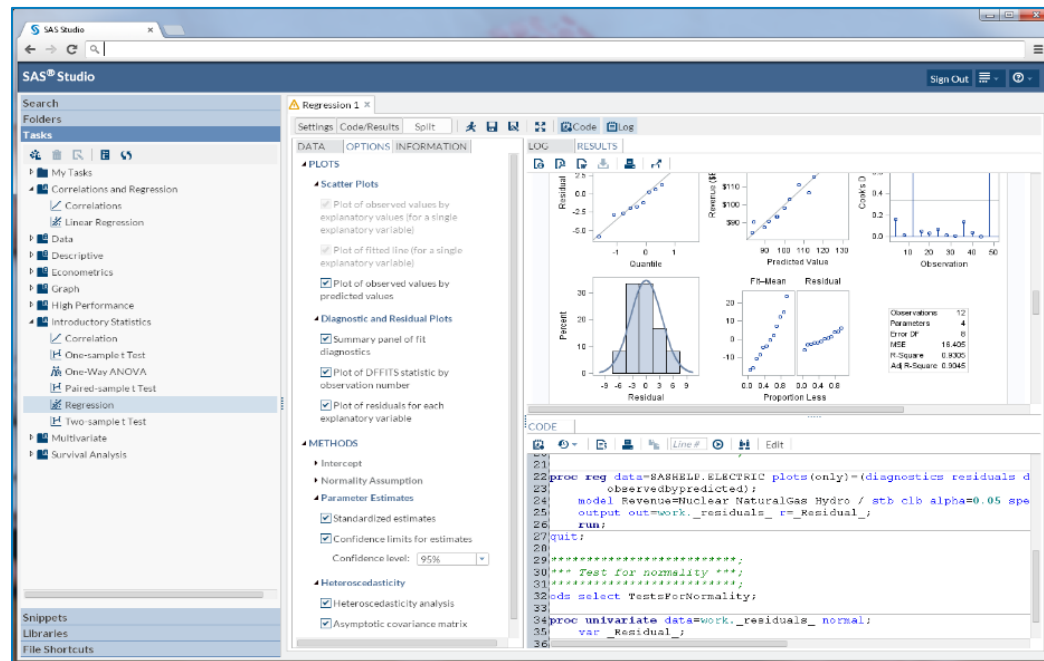


ANALYTIKA – „STAVEBNÉ KAMENE“

- Base SAS
- SAS Studio
- STAT
- IML
- ETS
- QC
- OR
- Simulation Studio



- **Base SAS** je jadro SAS softvéru
- Flexibilný programovací jazyk 4GL
- Bohatá knižnica procedúr pre prístup a manipuláciu s dátami, deskriptívnu štatistiku a reporting



Viac ako 50 Base SAS procedúr

The screenshot displays the SAS Studio web interface for a regression analysis. The main workspace is titled 'Regression 1' and shows the 'RESULTS' tab. The 'LOG' tab is also visible, showing the SAS code used for the analysis.

Left Sidebar (Navigation):

- Search
- Folders
- Tasks
 - My Tasks
 - Correlations and Regression
 - Correlations
 - Linear Regression
 - Data
 - Descriptive
 - Econometrics
 - Graph
 - High Performance
 - Introductory Statistics
 - Correlation
 - One-sample t Test
 - One-Way ANOVA
 - Paired-sample t Test
 - Regression
 - Two-sample t Test
 - Multivariate
 - Survival Analysis
- Snippets
- Libraries
- File Shortcuts

Central Workspace (Regression 1):

- Settings | Code/Results | Split | Code | Log
- DATA | OPTIONS | INFORMATION
- LOG | RESULTS
- Plots
 - Scatter Plots
 - ☒ Plot of observed values by explanatory values (for a single explanatory variable)
 - ☒ Plot of fitted line (for a single explanatory variable)
 - ☒ Plot of observed values by predicted values
 - Diagnostic and Residual Plots
 - ☒ Summary panel of fit diagnostics
 - ☒ Plot of DFFITS statistic by observation number
 - ☒ Plot of residuals for each explanatory variable
- Methods
 - Intercept
 - Normality Assumption
 - Parameter Estimates
 - ☒ Standardized estimates
 - ☒ Confidence limits for estimates
 - Confidence level: 95%
 - Heteroscedasticity
 - ☒ Heteroscedasticity analysis
 - ☒ Asymptotic covariance matrix

Right Sidebar (Diagnostic Plots and Statistics):

- Residual vs Quantile
- Revenue (\$ vs Predicted Value)
- Cook's D vs Observation
- Percent vs Residual
- Fit-Mean vs Proportion Less
- Residual vs Proportion Less
- Summary Statistics:

Observations	12
Parameters	4
Error DF	8
MSE	16.405
R-Square	0.9305
Adj R-Square	0.9045

Code Editor:

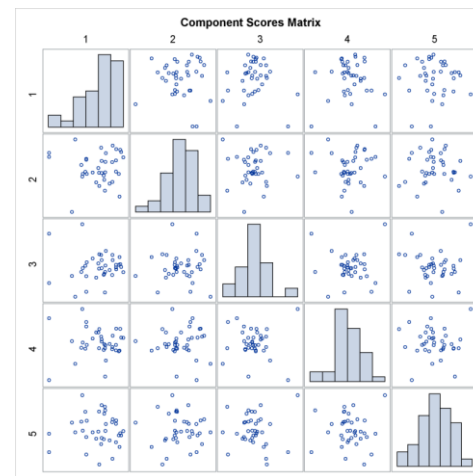
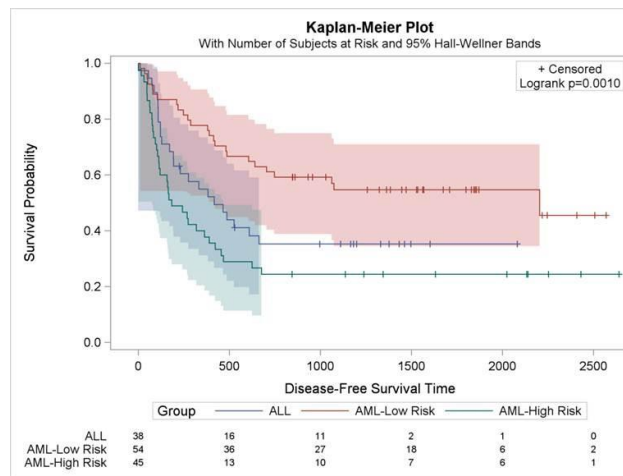
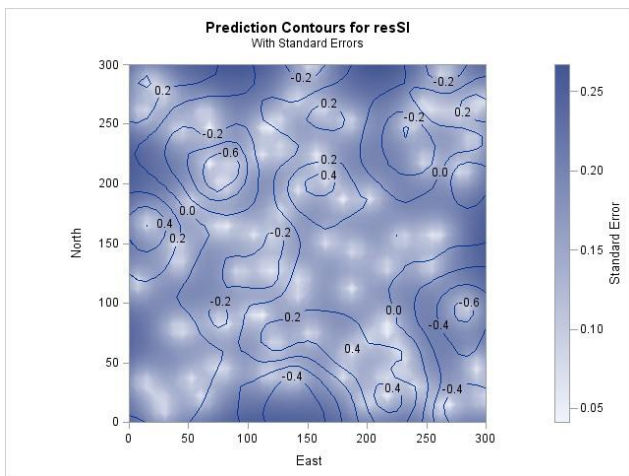
```

proc reg data=SASHELP.ELECTRIC plots (only)=(diagnostics residuals d
observedbypredicted);
  model Revenue=Nuclear NaturalGas Hydro / stb clb alpha=0.05 spe
  output out=work._residuals_ r=_Residual_;
run;
quit;

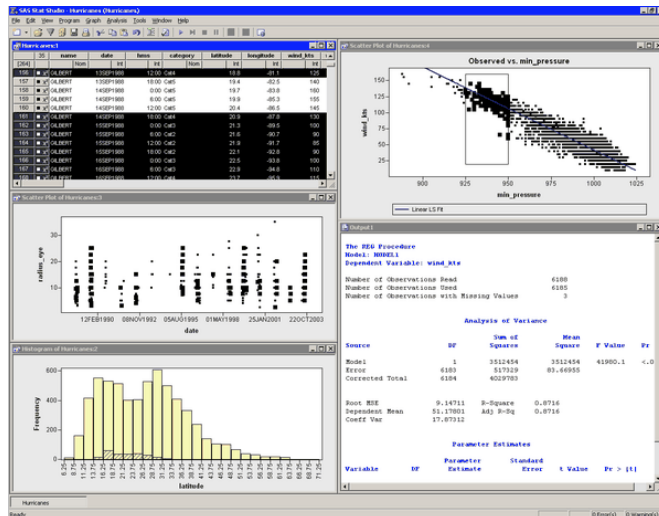
*****
30 *** Test for normality ***
31 *****
32 ods select TestsForNormality;
33
34 proc univariate data=work._residuals_ normal;
35   var _Residual_;
36
    
```

- **STAT** zahŕňa štatistické algoritmy od tradičných analýz rozptylu (ANOVA) cez lineárnu regresiu až po Bayesovskú štatistickú inferenciu.
- Zahŕňa IML a IML Studio

Viac ako 80 SAS/STAT procedúr



- **IML** (Interactive matrix programming) poskytuje flexibilný jazyk pre prácu s maticami pre štatistikov a výskumníkov
- Vhodné od jednoduchých operácií s maticami po pokročilé techniky štatistického odhadu, simulácie, lineárneho programovania, nelineárnej optimalizácie a podobne.
- **IML Studio** poskytuje interaktívny prieskum dát a interaktívne programovanie.
- Možnosť využívania R kódu v prostredí SAS.



```

113 proc iml;
NOTE: IML Ready
114 x = uniform(repeat(1,1000,100)); /** Creates Data set of 1000 rows and 100 columns**
114!

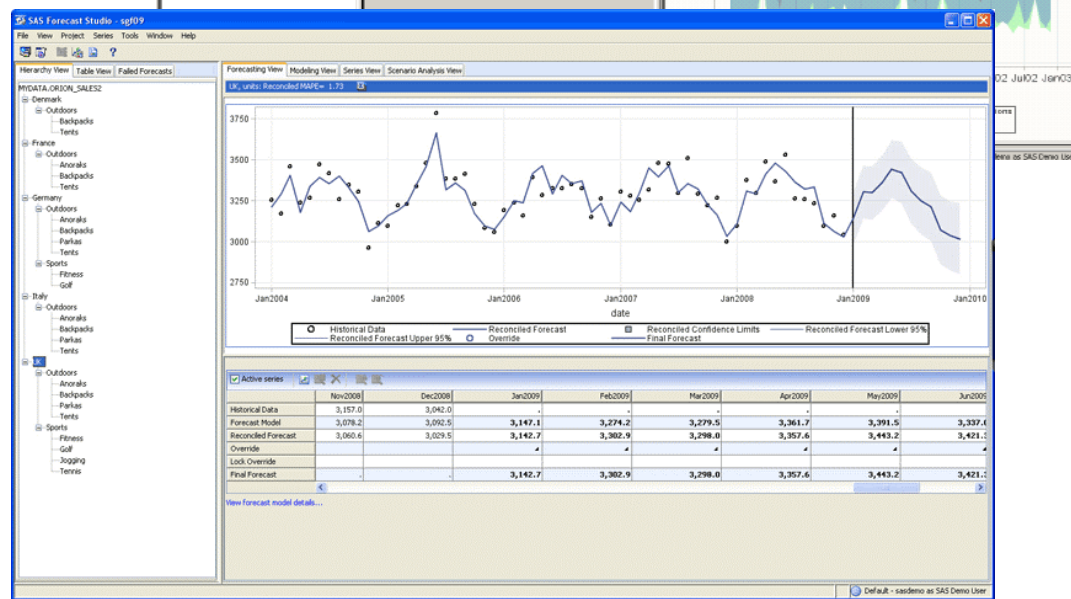
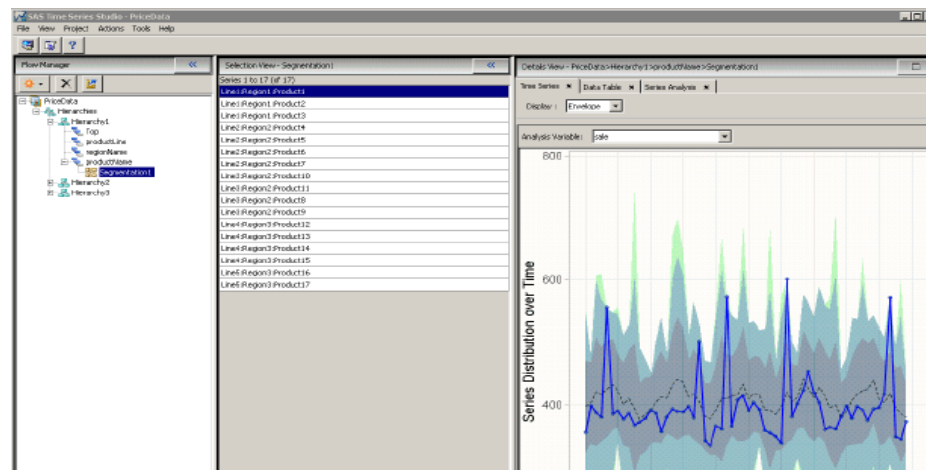
115 do i = 1 to 1000;
116   do j = 1 to 100;
117     if x[i,j] < 0.5 /**assigns zero if the value is greater than .5**/
118       then
119         x[i,j] = 0;
120       else if x[i,j] >= 0.5
121         then x[i,j] = 1;
122   end;
123 end;
124 quit;
NOTE: Exiting IML.
NOTE: 18 workspace compresses.
NOTE: PROCEDURE IML used (Total process time):
      real time      0.31 seconds
      cpu time       0.31 seconds

125
126 proc iml;
NOTE: IML Ready
127 x = uniform(repeat(1,1000,100)); /** Data set of 1000 rows and 100 columns**/**;
127!

128 x[loc(x<.5)]=0;
129 x[loc(x>=.5)]=1;
130 quit;
NOTE: Exiting IML.
NOTE: PROCEDURE IML used (Total process time):
      real time      0.04 seconds
      cpu time       0.04 seconds

```


- **Time Series Studio** - prezeranie a organizovanie hierarchií časových radov.
 - **Forecast Studio** – vytváranie forecastovacích modelov.
 - **Forecast Project Manager** – správa projektov
-
- Analýza časových radov
 - Predpoveď / forecasting
 - Ekonometrické analýzy
 - Finančná analýza a reporting
 - Modelovanie a simulácia systémov
 - Analýza diskretných volieb
 - Sezónne úpravy dát

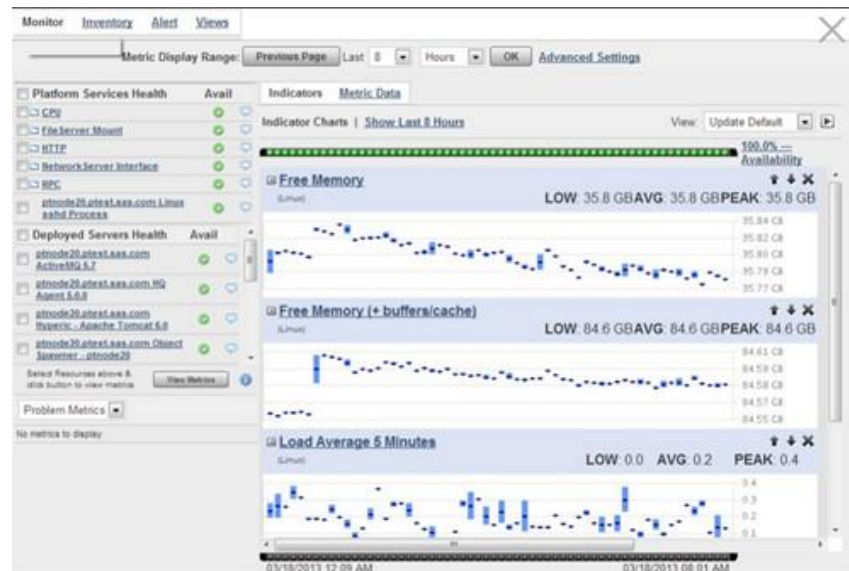


QC je komplexná sada nástrojov na štatistické hodnotenie a zlepšovanie kvality.

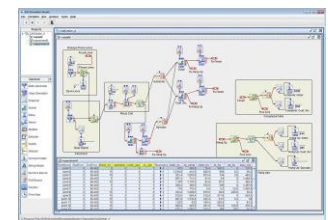
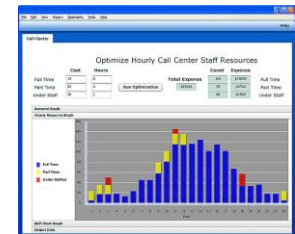
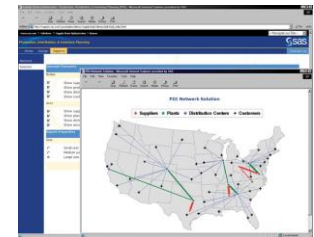
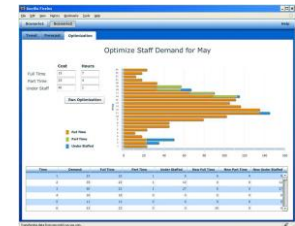
- Zavedenie štatistického riadenia procesu
- Návrh priemyselných experimentov
- Použitie Taguchi metód pre inžinierstvo kvality
- Štatistická kontrola a zmenšenie rozptylu
- Vývoj a vyhodnotenie plánov odberu vzoriek
- Vizuálne užívateľské rozhrania pre zlepšovania kvality bez znalosti programovania.
- QC procedúry poskytujú viac možností oproti užívateľskému rozhraniu.

Komponenty SAS/QC:

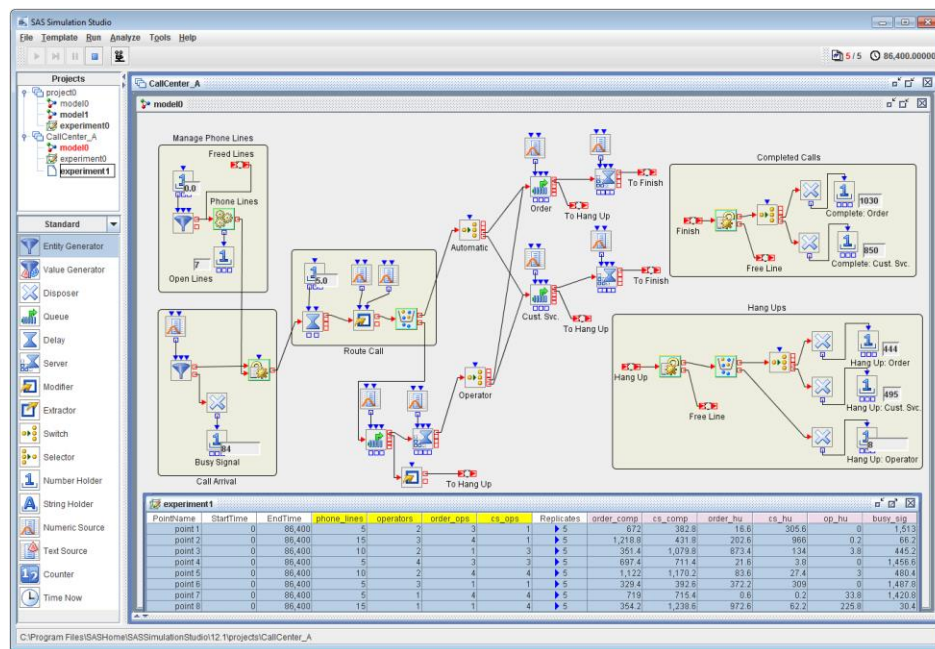
- ADX Interface for Design of Experiments
- SQC Menu System for Statistical Quality Control
- Procedures for Design of Experiments
- Procedures for Control Chart Analysis
- Procedure for Process Capability Analysis
- Procedures for Basic Quality Problem Solving
- Procedure for Reliability Analysis
- Procedure for Analysis of Means



- **OR** (operational research, operačný výskum) software zahŕňa optimalizáciu, modelovanie, simuláciu a plánovanie vo flexibilnom užívateľskom prostredí.
- Sada manažérskych nástrojov podporuje optimalizáciu biznis procesov a riešenie manažérskych výziev.
- SAS/OR je určený pre odborníkov z manažmentu a OR, ktorí plánujú vytvoriť modely na podporu rozhodovania využívajúce nasledujúce techniky:
 - Matematické programovanie
 - Simulácia diskretných udalostí
 - Plánovanie projektov a zdrojov
 - Optimalizácia lokálneho vyhľadávania
 - Analýza rozhodnutí
 - Spracovanie súpisiek materiálu (bill-of-material (BOM))

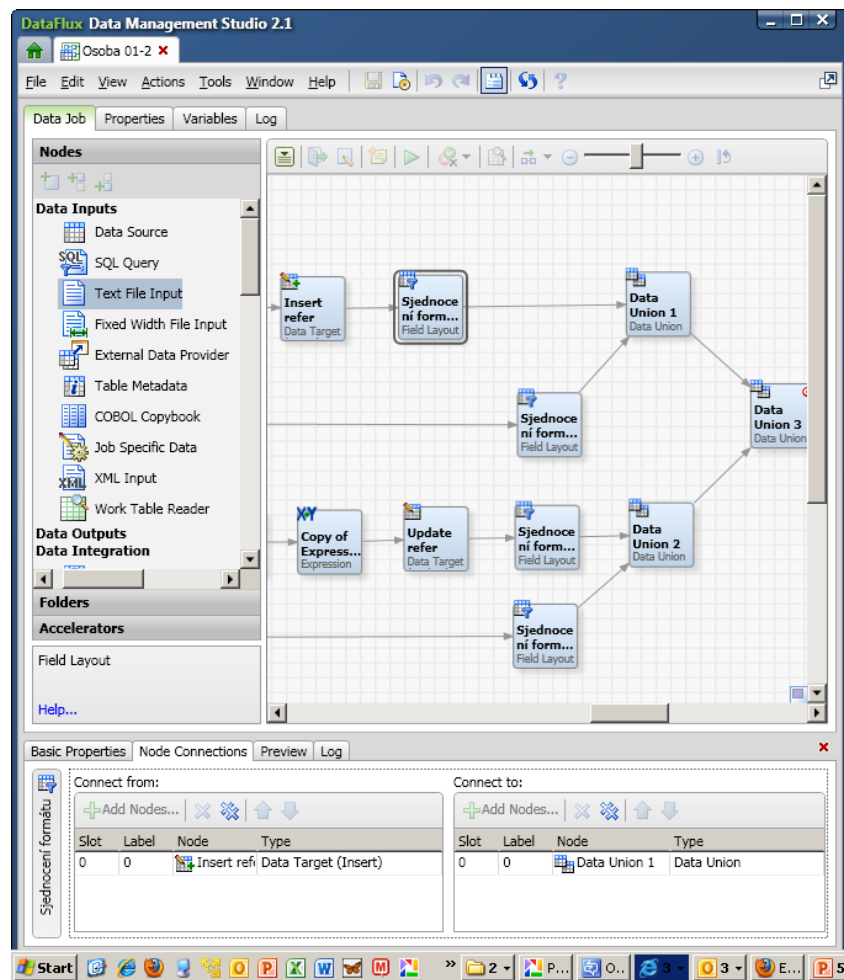


- Windows a Java aplikácia na tvorbu modelov na simuláciu diskretných udalostí
- Vizuálne rozhranie
 - Návrh modelov
 - Výpočet modelu
 - Analýza výsledkov
 - Návrh experimentov
- Vhodné pre začiatočníkov aj pokročilých v oblasti simulácie
- Súčasť SAS/OR alebo ako add-on pre JMP

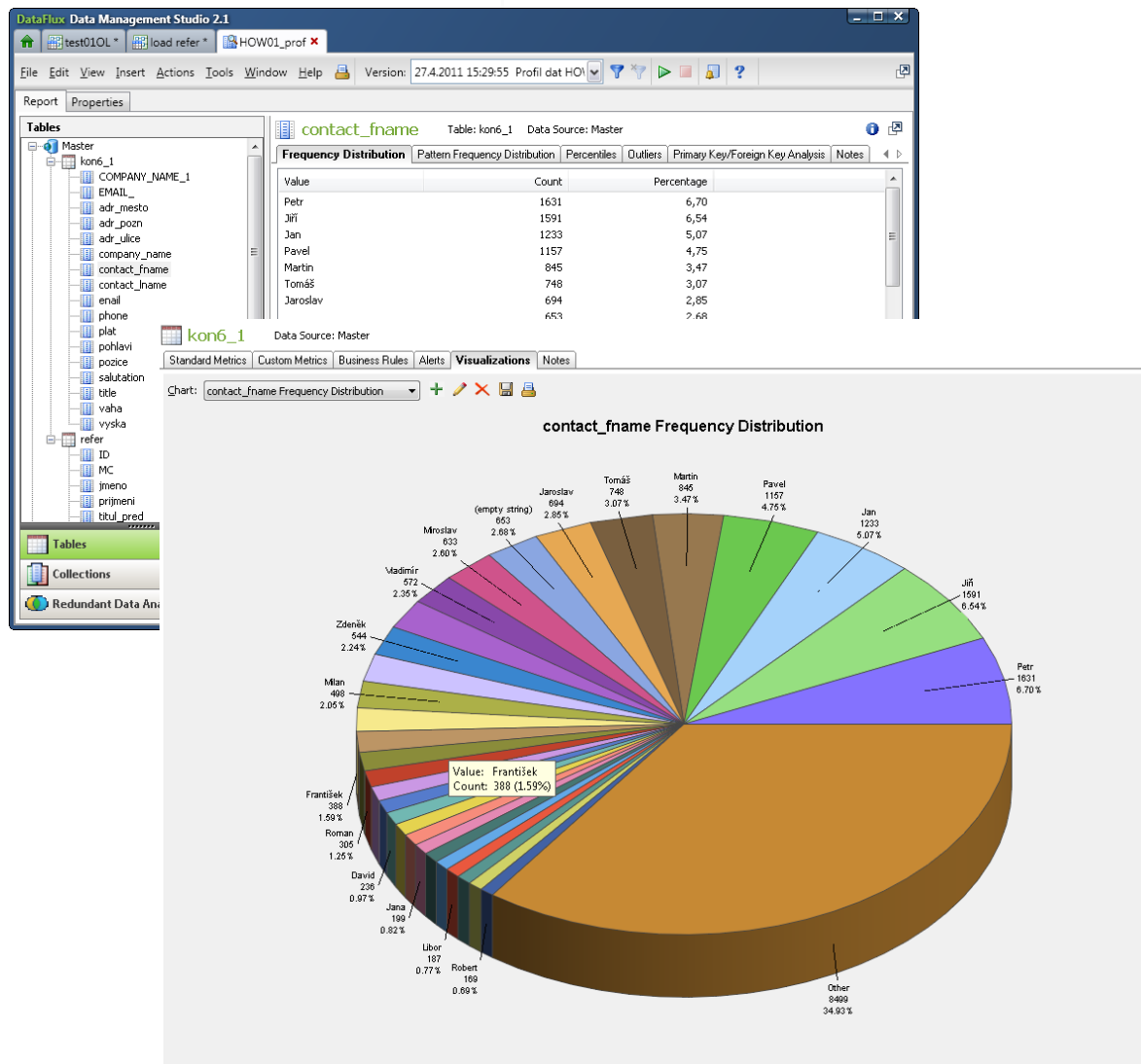


- Data Quality
- Data Integration Studio
- Workflow Studio
- Text Miner
- Content Categorization



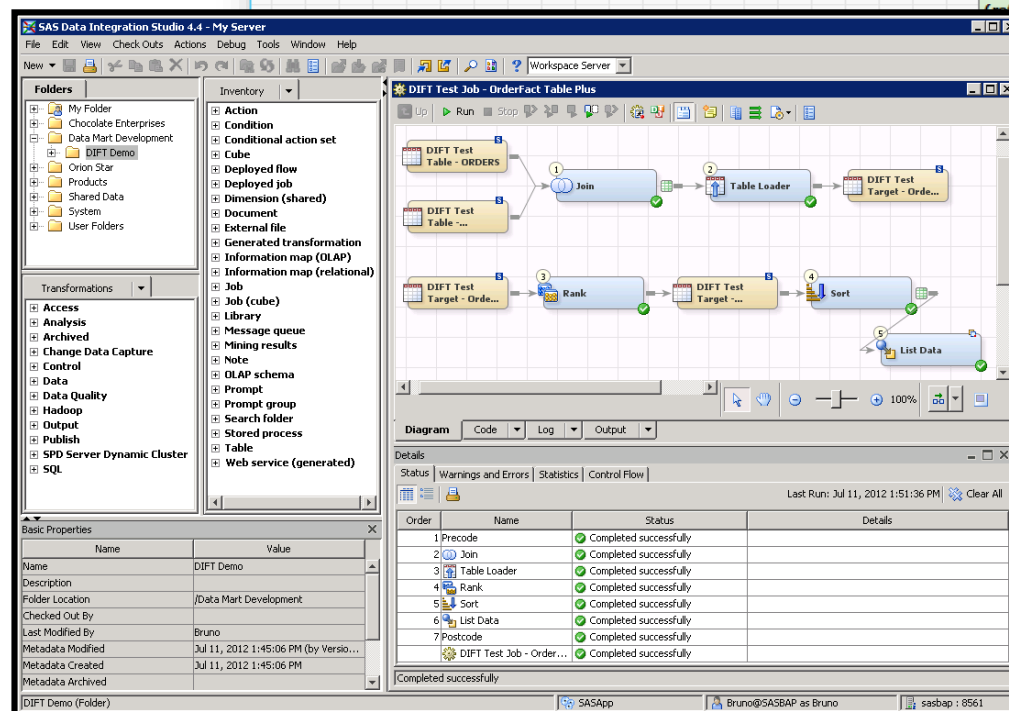
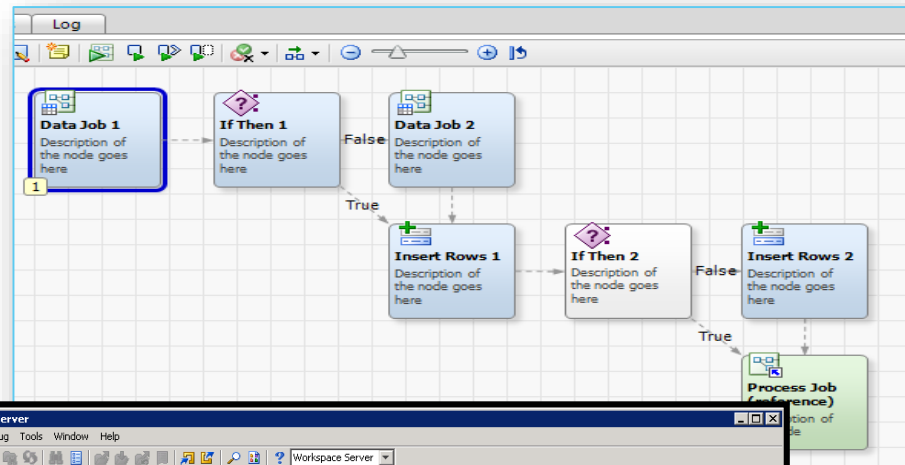


- **Komplexná funkcionality** (profilovanie, štandardizácia, párovanie, obohacovanie, monitorovanie DQ, ETL)
- **Unikátne technické riešenia**
 - Univerzálny „matching key“
 - Dynamicky nastaviteľná citlivosť
 - Zoskupovanie (clustrovanie) dát atď.
- **Podpora odborne orientovaných používateľov**
 - Zrozumiteľnosť a prehľadnosť
 - Bez kódovania
- **Rôzne režimy práce**
 - Identické joby pre batch aj on-line
 - Workflow na spracovanie problémov
- **Otvorenosť, flexibilita a variabilita**
 - Lokalizácia a prispôbienie pravidiel
 - Rôzne dátové zdroje, rôzne platformy
 - Integrácia s inými technológiami



- Identifikovanie problémov DQ
 - Intuitívne
 - Veľa preddefinovaných ukazovateľov DQ
 - Nastaviteľné pravidlá pre kontrolu DQ
 - Tabuľkové a grafické výsledky
 - „Drill down“ k dátam
- Mapy vzťahov medzi poliami
- Analýza nadbytočnosti
- Analýza príčin dátových problémov

- Rýchly a intuitívny návrh DQ a ETL jobov
 - Drag & Drop návrh
 - Konfigurovateľné uzly
 - Prezeranie vzorky dát
 - Priamy prístup k relačným databázam, súborom a SAP (samostatne licencované)
 - Dátový slovník pre užívateľov (Business Data Network)
- Joby pre spracovanie dávkové aj v reálnom čase



Vstupný formulár

Vyhľadaj klienta Nový klient Opraviť klienta

Vyhledat klienta

Titul
Ing

Jméno
Josef Novák

Příjmení

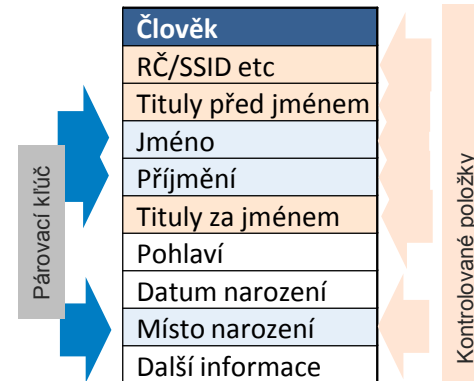
Titul za

Vyhledat

PSČ

Město
Praha

Citlivost
☐ Vysoká
☐ Vyšší
☒ Střední
☐ Nižší
☐ Nejnižší



Výstupný formulár

Výsledek hledání

Zadané hodnoty

Titul: Ing
Jméno: Josef Novák
Příjmení:
Titul za:
PSČ:
Město: Praha
Citlivost: 3

Opravené hodnoty

Titul: Ing
Jméno: Jose
Příjmení: Nov
Titul za:
PSČ:
Město: Pra

Nové hledání Upravit podmínky Vložit tohoto klienta

Nalezené záznamy

ID	Jméno	Příjmení	Rodné číslo
13266	Josef	Novák	480812653
13267	Jozef	Novák	330712673

Zobraz kartu

Detail

Karta klienta

Vybrán klient č.: 13266

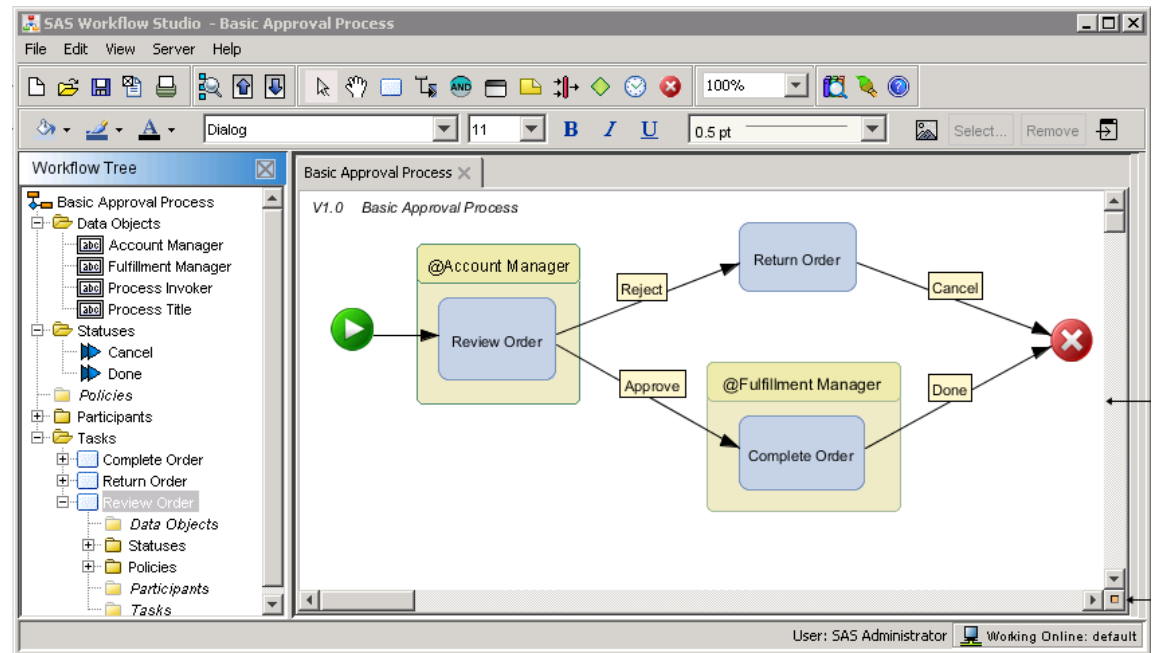
Titul:
Jméno: Josef
Příjmení: Novák
Titul2:
Rodné číslo: 480812653
PSČ:
Město: Praha

Upravit hledání Opravit klienta Smaž klienta

Činnosti DQ

1. Parsing mena
2. Parsing miesta
3. Štandardizácia komponent mena
4. Štandardizácia komponent miesta
5. Kontrola rodného čísla
6. Párovanie referenčného záznamu

- Riešenie problémov dátovej kvality, CRM, risk, fraud apod.
 - Smerovanie na zodpovednú osobu
 - Schvaľovanie
 - Sledovanie stavu
- Definovanie postupnosti krokov
 - Intervencia užívateľa
 - Automatické kroky
 - Rozhodovanie podľa dát (pravidiel, modelov)



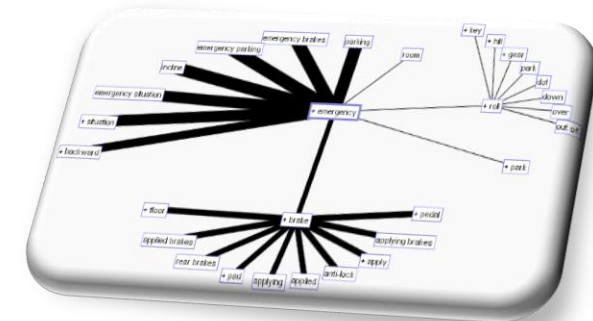
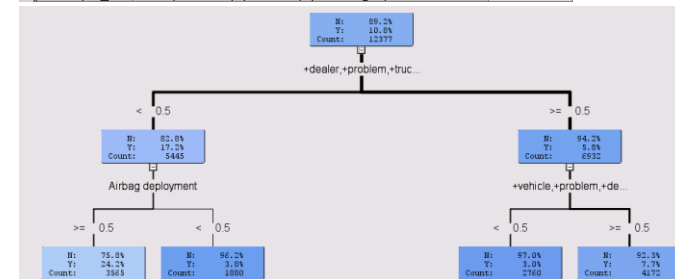
- Kombinácia linguistických a analytických nástrojov
- Identifikácia rečových jednotiek, podstatných mien, synonym
- Automaticky určuje témy diskusie a súvisiace pojmy
- Štruktúrovanie textu pre ďalšiu analýzu
- Automatické zoskupenie dokumentov do súvisiacich skupín
- Vizualizačné nástroje na analýzu a prieskum
- Podpora extrakcie dát z webu – Web crawling

Variables - Tree

(none) ☐ not Equal to

Columns: ☒ Label ☐ Mining

Name /	Label	Use
TextTopic_1	Airbag deployment	Default
TextTopic_10	+dealer,+problem,+truck,+fix,+find	Default
TextTopic_11	+replace,+dealer,+problem,+car,+mechanic	Default
TextTopic_12	+dealership,+vehicle,+problem,+replace,+light	Default
TextTopic_13	+mile,+rotor,+replace,+dealer,+vehicle	Default
TextTopic_14	+time,+problem,+vehicle,+dealer,+car	Default
TextTopic_15	+brake,+vehicle,+break,+rotor,+apply	Default
TextTopic_16	+noise,+vehicle,+problem,+dealer,+service	Default
TextTopic_17	+engine,+vehicle,+light,+fuel,+stall	Default
TextTopic_18	+seat,+belt,+passenger,+consumer,+vehicle	Default
TextTopic_19	+door,+vehicle,+problem,+passenger,+driver	Default



- Pokročilé spracovanie prirodzenej reči (Natural Language Processing)
- Podpora viacerých jazykov
- Parsovanie a vyhľadávanie entít, udalostí a faktov
- Vytváranie metadát
- Linguistické pravidlá na výber dokumentov, náhradu zámen atď.
- Preddefinovaná extrakcia entít z DBPedia a kategórií z Wikipedia

The screenshot displays the SAS Content Categorization software interface. The main window shows a project structure for 'LITI Demo' with a tree view of concepts and categories. A 'Predefined LITI Concepts' dialog box is open, showing a list of predefined concepts and a table of statistics. The 'LITI Demo' project structure includes a tree view of concepts and categories, with 'PERSON' and 'FIRST' highlighted. The 'Predefined LITI Concepts' dialog box shows a list of predefined concepts and a table of statistics. The table has columns for 'Neg %', 'Prec %', 'Popu...', and 'Pop...'. The 'LITI Demo' project structure includes a tree view of concepts and categories, with 'PERSON' and 'FIRST' highlighted. The 'Predefined LITI Concepts' dialog box shows a list of predefined concepts and a table of statistics.

Neg %	Prec %	Popu...	Pop...
0	0	0	0
0	2	0	0
0	1	0	0
0	0	0	0
0	3	0	0
0	1	0	0
0	4	0	0
0	4	0	0
0	2	0	0
0	10	0	0
0	8	0	0
0	11	0	0
0	1	0	0
0	6	0	0
0	12	0	0
0	3	0	0
0	18	0	0
0	47	0	0
0	0	0	0
0	23	0	0
0	29	0	0
0	0	0	0
0	9	0	0
0	0	0	0
0	38	0	0
0	7	0	0
0	0	0	0
0	4	0	0
0	0	0	0
0	5	0	0
0	14	0	0
0	0	50	0
0	0	17	0

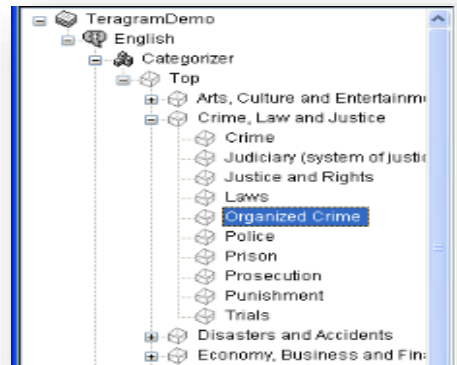
1. Vstup textu – napr. FB, twitter, email, fórum

2. Parsovanie obsahu s použitím taxonómie – roztriedenie do kategórií

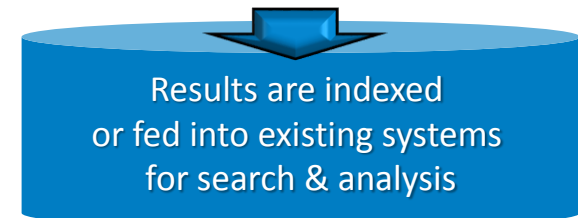
3. Výsledok – označenie dokumentov detailnými kategóriami



Categorization Taxonomy



Topic = Organized Crime



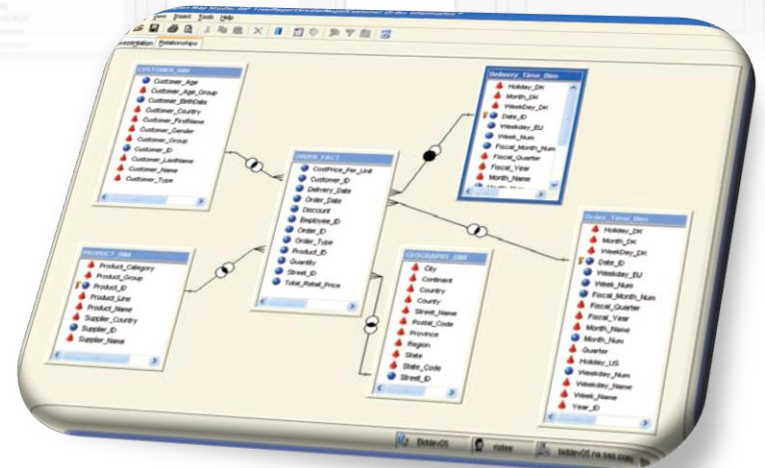
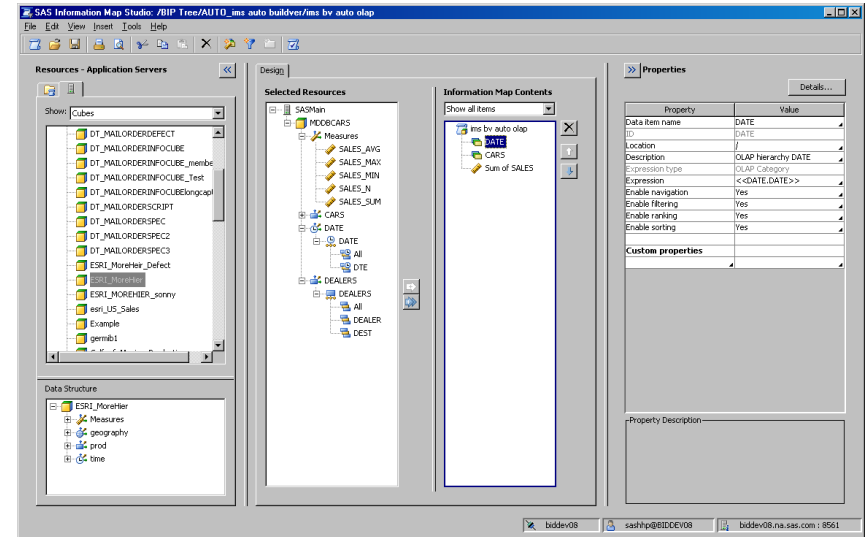
Results are indexed
or fed into existing systems
for search & analysis

BUSINESS INTELLIGENCE

- Information Map Studio
- Web OLAP Viewer
- Web Report Studio
- Information delivery Portal
- Add-In for Microsoft Office



- Namapovanie fyzických dátových štruktúr na biznis pohľad vo vizuálnom nástroji
- Definovanie biznis pohľadu pre relačné tabuľky aj OLAP kocky
- Podpora viacerých zdrojov dát
- Jeden pohľad krížom cez viacero dátových zdrojov
- Definovanie konzistentných pravidiel, možností a parametrov
- Centrálna správa metadát



- Prístup k dátam pre biznis užívateľov
- Dopytovanie pomocou sprievodcov
- Zakrytie zložitosti dátových štruktúr pred užívateľmi
- Vykonávanie dopytov nad viacerými dátovými zdrojmi.

ORIONCUBE

New View | Desynchronize Views | Tab Views | Back | Forward | Edit View | [Icons]

Data Dimensions

Add To: Responses

Available data: [Modify] [Show Levels]

- Measures
- Customers
- Demographics
- Geography
- Product
- Time
- Member sets

View 1

Roles: Show Filters Highlights | Hide All

Columns: Year-Quater-Month | ALLYQM | Sum of Quantity

Rows: Global | AllGlobal

	Year 1998	Year 1999	Year 2000	Year 2001	Year 2002
Measure: Sum of Quantity					
Sum of Quantity					
Continent					
Africa	269	298	234	221	319
Asia	434	389	358	375	362
Australia/Pacific	17,487	18,731	20,391	16,738	21,522
Europe	176,493	206,213	244,905	216,436	255,321
North America	66,476	76,880	97,081	74,813	84,570

View 2

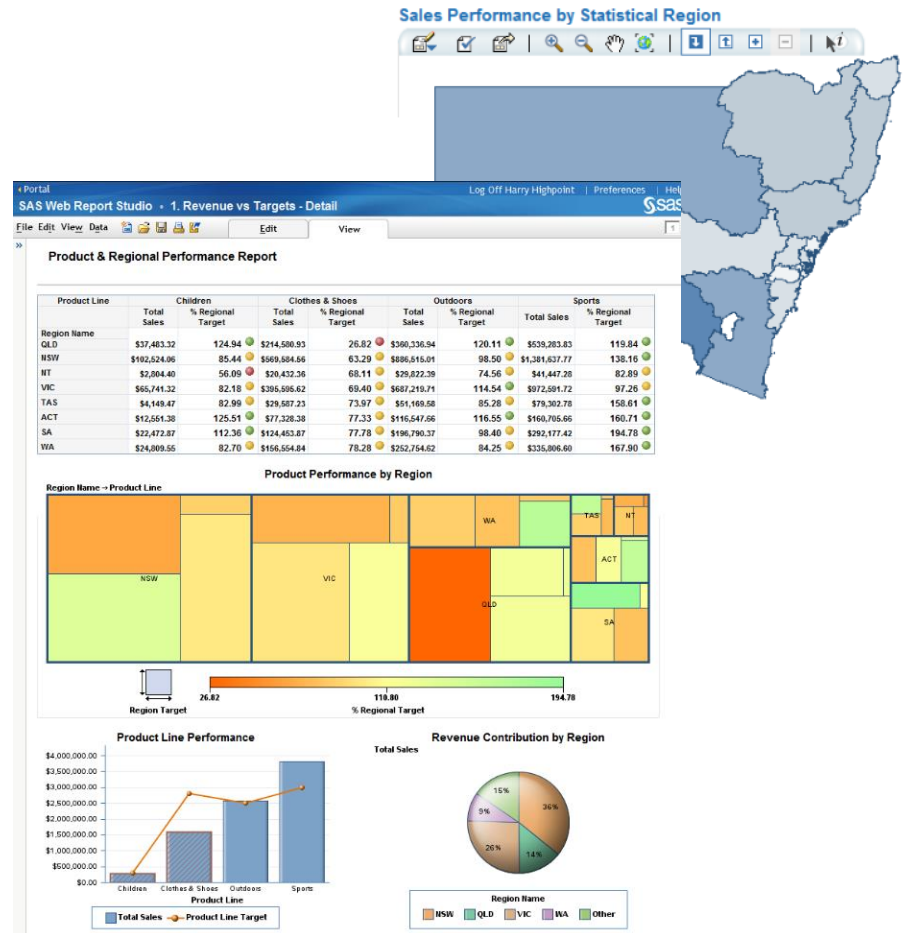
Roles: Show Filters Highlights | Hide All

Columns: Year-Quater-Month | ALLYQM | Sum of Quantity

Rows: Global | AllGlobal

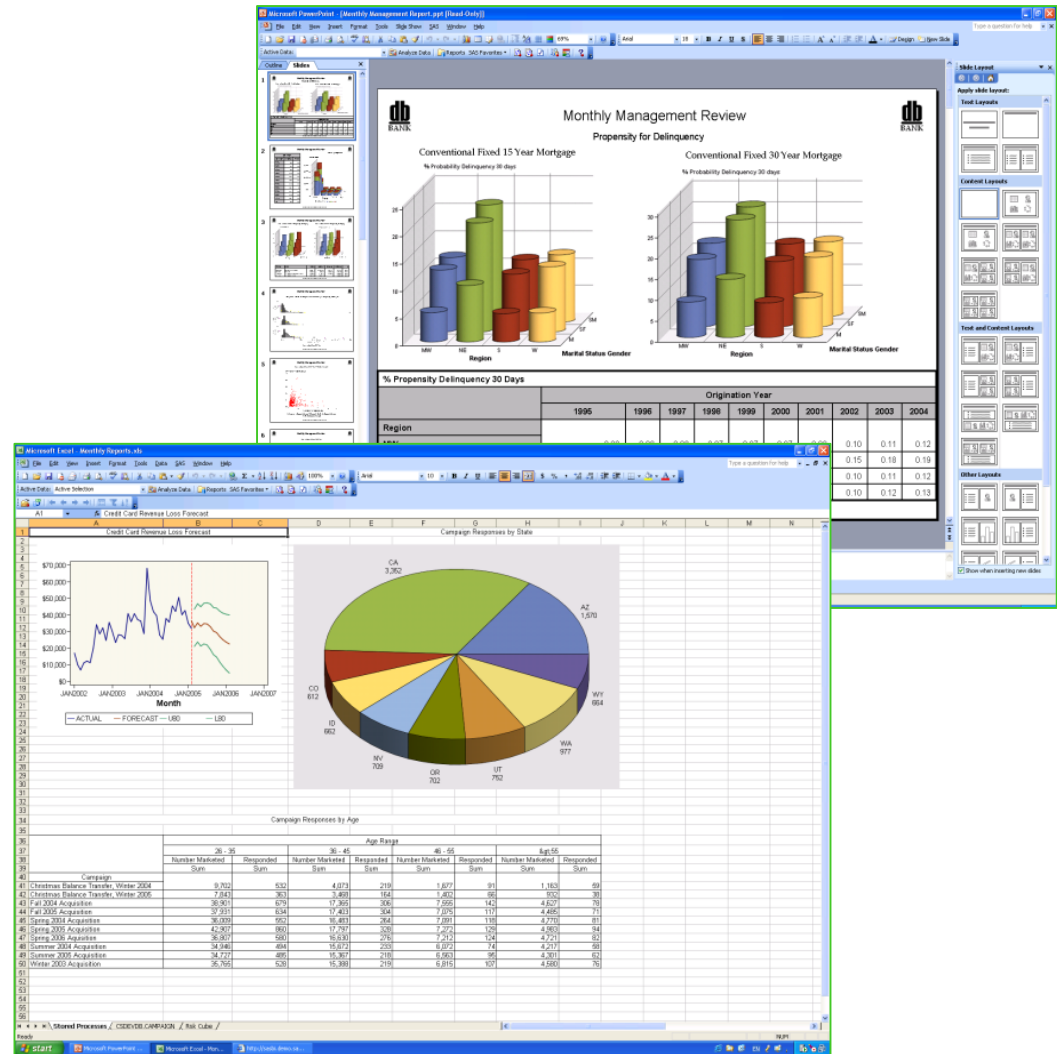
GEOGRAPHIC	REGION	STATE	Catalog				
			Sports	\$1,757.00	36	\$1,338.00	3
			Toys	\$208.69	17	\$1,063.06	22
			Subtotal: CANADA	\$1,965.69	53	\$2,401.06	52
			Sports	\$2,377.00	39	\$3,149.00	50
			Toys	\$635.57	29	\$882.85	18
			Subtotal: NC	\$3,012.57	68	\$4,031.85	68
			Sports	\$13.00	1	\$193.00	4
			Toys	\$138.52	3	\$24.80	1
			Subtotal: SC	\$151.52	4	\$217.80	5
			Subtotal: Sports	\$2,390.00	40	\$3,342.00	54
			Subtotal: Toys	\$774.09	32	\$907.65	19
			Subtotal: SOUTHEAST	\$3,164.09	72	\$4,249.65	73
			Subtotal: Sports	\$4,147.00	76	\$4,680.00	84
			Subtotal: Toys	\$982.78	49	\$1,970.71	4
			Subtotal: All GEOGRAPHIC	\$5,129.78	125	\$6,650.71	

- Tenký webový klient
- Interaktívna tvorba a využívanie reportov
- Sprievodcovia pre tvorbu reportov
- Galéria preddefinovaných šablón
- Reporty založené na OLAP kockách a relačných dátach aj z viacerých zdrojov
- Prieskum dát





- Prístup k SAS dátam, reportom a analýzam priamo z Microsoft Office.
- Pokročilé analýzy v prostredí Word, Excel and PowerPoint.
- Prístup k dátam z externých zdrojov rôzneho typu.
- Centrálna administrácia dát.



Ukážky nástrojů SAS



THE
POWER
TO KNOW®