

PREVENCE KRIMINALITY VE SPOLEČNOSTI

Případová studie vypracovaná v IBM SPSS Modeler 14.2

Obsah

Úvod.....	3
Úlohy.....	3
Data.....	3
Proud přípravy dat.....	5
Část formátování vstupních dat a exportu dat pro optimalizaci rozmístění hlídek.....	5
Část exportující data pro proud souvisejících vloupání.....	7
Část exportující data pro proud kapesních krádeží.....	7
Proud souvisejících vloupání.....	9
Vytvořená struktura shlukovacích uzlů.....	10
Část kombinující výsledky shlukovacích uzlů a výsledné vyhodnocení souvisejících případů.....	11
Proud optimalizace rozmístění hlídek.....	12
Operace se vstupními daty.....	13
Použití shlukovacích uzlů.....	14
Proud zabývající se kapesními krádežemi.....	15
Vizualizace vstupních dat.....	15
Selekce a kategorizace zajímavé domény.....	16
Závěrečný report.....	16

Seznam ilustrací

Ilustrace 1: Detail na část streamu, která se zabývá formátováním dat a jejich exportem.....	5
Ilustrace 2: Výsledné přiřazení typů atributům.....	6
Ilustrace 3: Roztřídění případů do kategorií.....	7
Ilustrace 4: Detail na část streamu, která se zabývá exportem pro proud souvis. vloupání.....	7
Ilustrace 5: Detail na část streamu, která se zabývá exportem pro proud s kapesními krádežemi.....	7
Ilustrace 6: Histogram četnosti krádeží v týdnech.....	8
Ilustrace 7: Definice Stavů	9
Ilustrace 8: Struktura shlukovacích uzlů.....	10
Ilustrace 9: Detail na část streamu, která vyhodnocuje dílčí výsledky.....	11
Ilustrace 10: Četnosti případů pro danou kombinaci shluků.....	11
Ilustrace 11: Zobrazení kandidátů souvisejících případů na mapě.....	12
Ilustrace 12: Uzly sloužící pro fázi úpravy dat.....	13
Ilustrace 13: Odvození nového atributu s částí dne.....	13
Ilustrace 14: Výsledný vzhled části streamu s modelovacími uzly.....	14
Ilustrace 15: Rozmístění hlídek v noci.....	14
Ilustrace 16: Kompletní proud.....	15
Ilustrace 17: Rozmístění kapesních krádeží.....	15
Ilustrace 18: Četnost krádeží v lokacích v závislosti na týdnu.....	16

Úvod

Při rozmísťování policejních hlídek se můžeme setkat s problémem, kam tyto hlídky umístit tak, aby bylo co nejlepší pokrytí oblastí, ve kterých dochází ke kriminálním případům a dosáhli jsme tudíž zefektivnění boje proti zločinnosti. V návrhu řešení tohoto problému je samozřejmě vhodné zohlednit roční období a denní dobu, ve které k páchání těchto zločinů dochází. Z analytického hlediska může být podobným problémem nalezení vzorů a závislostí mezi případy vloupání nejenom na základě místa a času, kdy k události došlo, ale i pomocí důkazů nalezených na místě činu. Objevení takových souvislostí může vést k identifikaci organizované zločinnosti, nebo upozornit na další podobné případy, kterých se mohl dopustit stejný pachatel a dopomoci tak k jeho usvědčení.

Úlohy

V rámci této studie se budou řešit následující úlohy:

- 1) Nalezení souvisejících vloupání do obytných budov
- 2) Optimalizace pokrytí území daným počtem hlídek
- 3) Popis výskytů kapesních krádeží v závislosti na období v roce a lokalitě

Data

Dostali jsme data kriminálních případů za jeden rok v určité, blíže nespecifikované lokalitě.

Formát byl následující:

Atribut	Popis
cislo_pripadu	Identifikační číslo případu
Kod	Kódové značení události
souradnice_X	X-ová souřadnice na mapě
souradnice_Y	Y-ová souřadnice na mapě
datum	Datum, ve kterém došlo k případu
MO_vstup	Způsob vstupu
MO_misto	Místo vstupu
MO_zabezp	Způsob zabezpečení
MO_alarm	Alarm
MO_cinnost	Zaměření budovy
MO_odchod	Opuštění budovy

MO_zabdvere	Zabezpečení dveří
MO_neporadek	Zanechán nepořádek
MO_sejf	Pokus o vyloupení sejfu
MO_zpusobvstupu	Okolnosti vstupu
Kaud	Odcizený audio přehrávač

Kvid	Odcizený video přehrávač
Kpocitac	Odcizený počítač
Pobleceni	Ukradeno oblečení
Ppenize	Ukradeny peníze
Pkredit	Ukradena kreditní karta
Plek	Ukradeny léky
Ptelefon	Ukraden telefon
Phodiny	Ukradeny hodiny
Pkalk	Ukradeny kalkulačky
Palkoh	Ukraden alkohol
Pzaznam	Ukraden audio záznam
Pšperky	Ukradeny šperky
Ppenez	Ukradena peněženka
Pdvere	Ukradeny dveře
Pokno	Ukradeno okno
Pautomat	Vykradený výdejní automat
Ptelauto	Vykradený telefonní automat

Navíc máme k dispozici tabulku, která k jednotlivým kódům kriminálních případů přiřazuje i slovní reprezentaci.

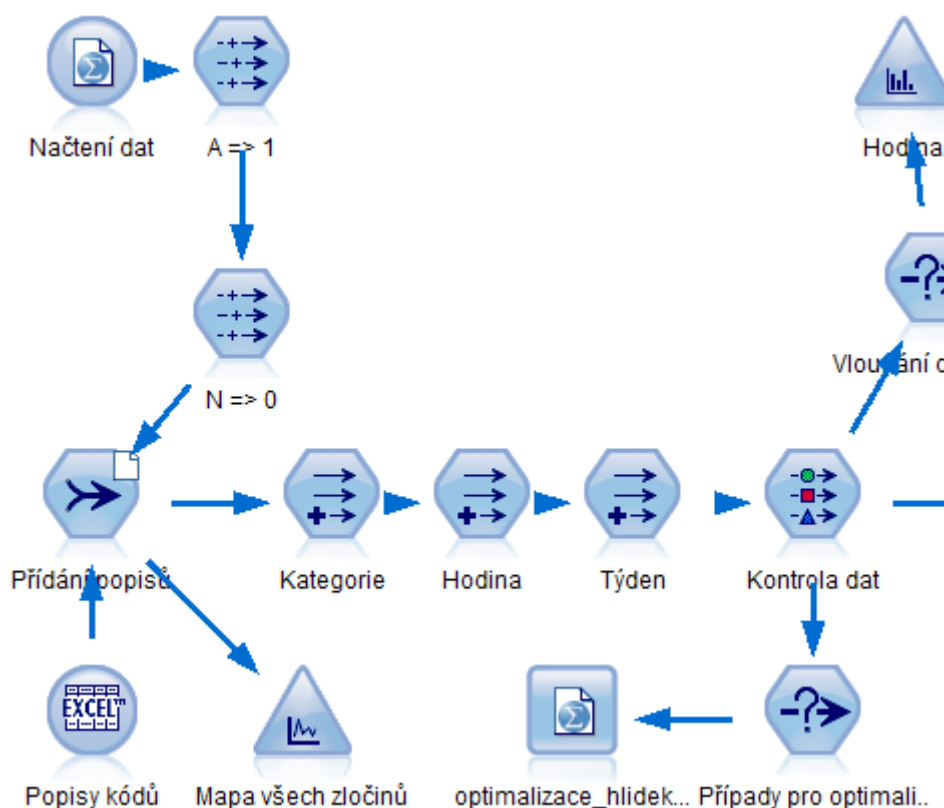
Proud přípravy dat

První proud se zabývá úpravou vstupních dat, zajištěním jejich správné reprezentace v modelu, odvozením některých nových atributů a konečně také exportem dat pro ostatní proudy, které již řeší jednotlivé úlohy. Předávání dat z proudu do proudu je realizováno statistickými soubory.

Kompletní proud je obsahem Přílohy A. Tento proud, stejně jako všechny ostatní, je rozdělen na logické celky (části), kdy ke každému celku je zobrazen detail uzlů, jenž se tohoto celku týkají.

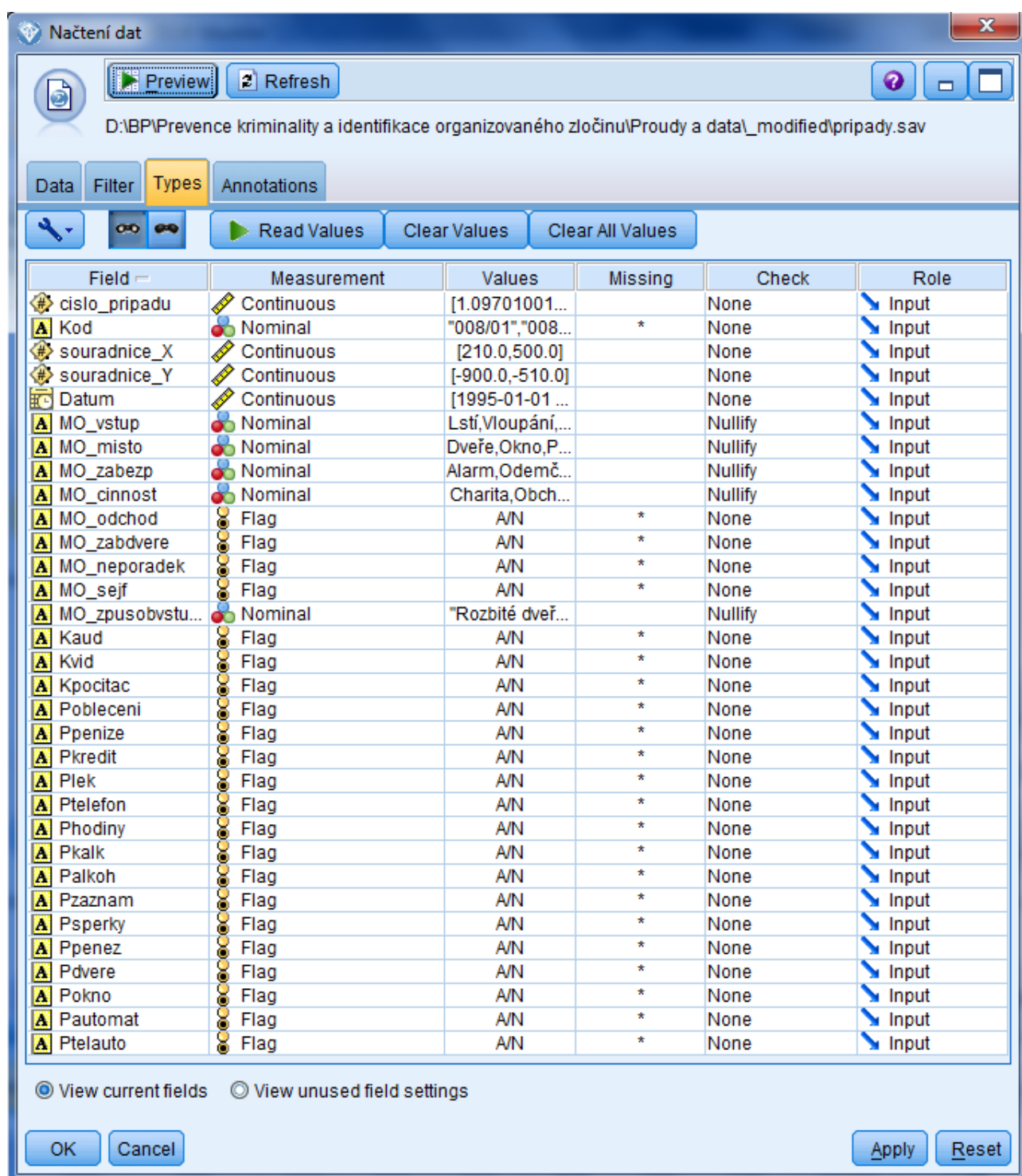
Část formátování vstupních dat a exportu dat pro optimalizaci rozmístění hlídek

Na obrázku (Ilustrace 1) je detail této části.



Ilustrace 1: Detail na část streamu, která se zabývá formátováním dat a jejich exportem

Vstupním bodem je načtení „surových“ dat ze souboru. Typy atributů však nejsou rozpoznány vhodným způsobem, a tak je třeba tyto typy manuálně opravit. Příkladem nevhodného rozpoznání je například fakt, že dichotomické atributy byly rozpoznány jako nonimální atp. Na obrázku (Ilustrace 2) je vidět výsledné ruční přiřazení typů atributů.



Ilustrace 2: Výsledné přiřazení typů atributům

Speciální pozornost musela být věnována všem atributům s typem „Flag“, protože hodnoty „A“ a „N“, byly přiřazeny v opačném logickém významu (True bylo „N“ a naopak). Čili ještě tento problém bylo třeba vyřešit v prvním uzlu a řešením bylo využití funkce manuálního přiřazení hodnot, přístupného přes pravé kliknutí na daný atribut.

Abychom se do budoucna vyhnuli podobným „skrytým“ nesrovnalostem, v následujících dvou uzlech jsme hodnoty „A“ (resp. „N“) přeložili na „1“ resp. „0“, a to pomocí uzlu *Filler*.

V dalších krocích jsme uzlem *Merge* přiřadili popisky ke kódům, zařadili případy do jedné ze tří kategorií, tak jak je naznačeno na obrázku (Ilustrace 3) a odvodili pomocí uzlů *Derive* nové atributy týden a hodina.

Derive as: **Nominal**

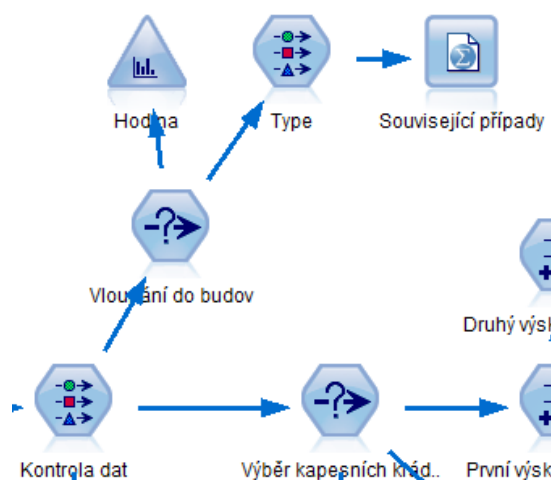
Field type: **Nominal** Default value: **vandalismus**

Set field to	If this condition is true
nasilne	Popis = "UBLIŽENÍ NA ZDRAVÍ" or Popis = "TĚŽKÉ UBLIŽENÍ NA ZDRAVÍ"
majetkove	Popis = "VLOUPÁNÍ DO OBYTNÉ BUDOVY" or Popis = "VLOUPÁNÍ DO NEOBYT. BUDOVY"
jine	Popis = "JÍZDA BEZ ŘIDIČSKÉHO PRŮKAZU" or Popis = "KAPESNÍ KRÁDEŽE"

Ilustrace 3: Roztřídění případů do kategorií

Na závěr této části jsme ty případy, obsahující hodnoty: majetkove, nasilne, vandalismus, exportovali do souboru pro další zpracování v příslušném proudu.

Část exportující data pro proud souvisejících vloupání



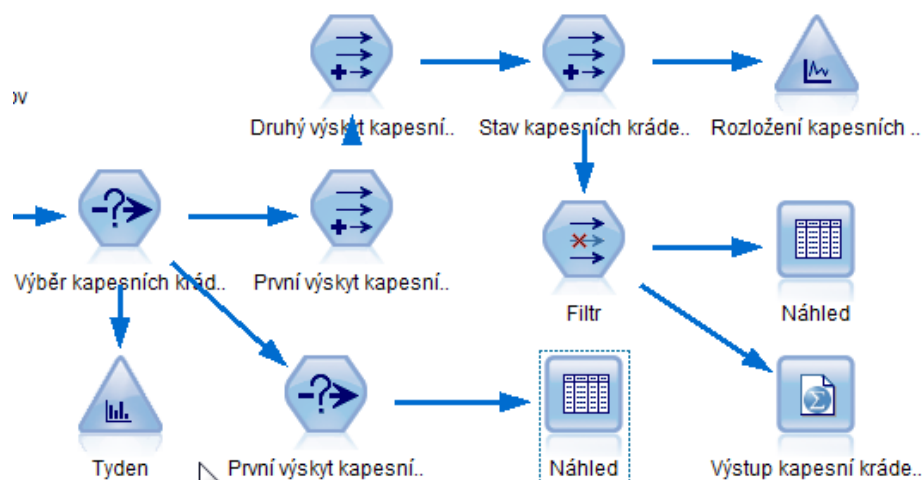
Ilustrace 4: Detail na část streamu, která se zabývá exportem pro proud souvis. vloupání

Obrázek (Ilustrace 4) zobrazuje detail této části.

Po vybrání pouze případů obsahující vloupání do budov jsme vyexportovali příslušný statistický soubor.

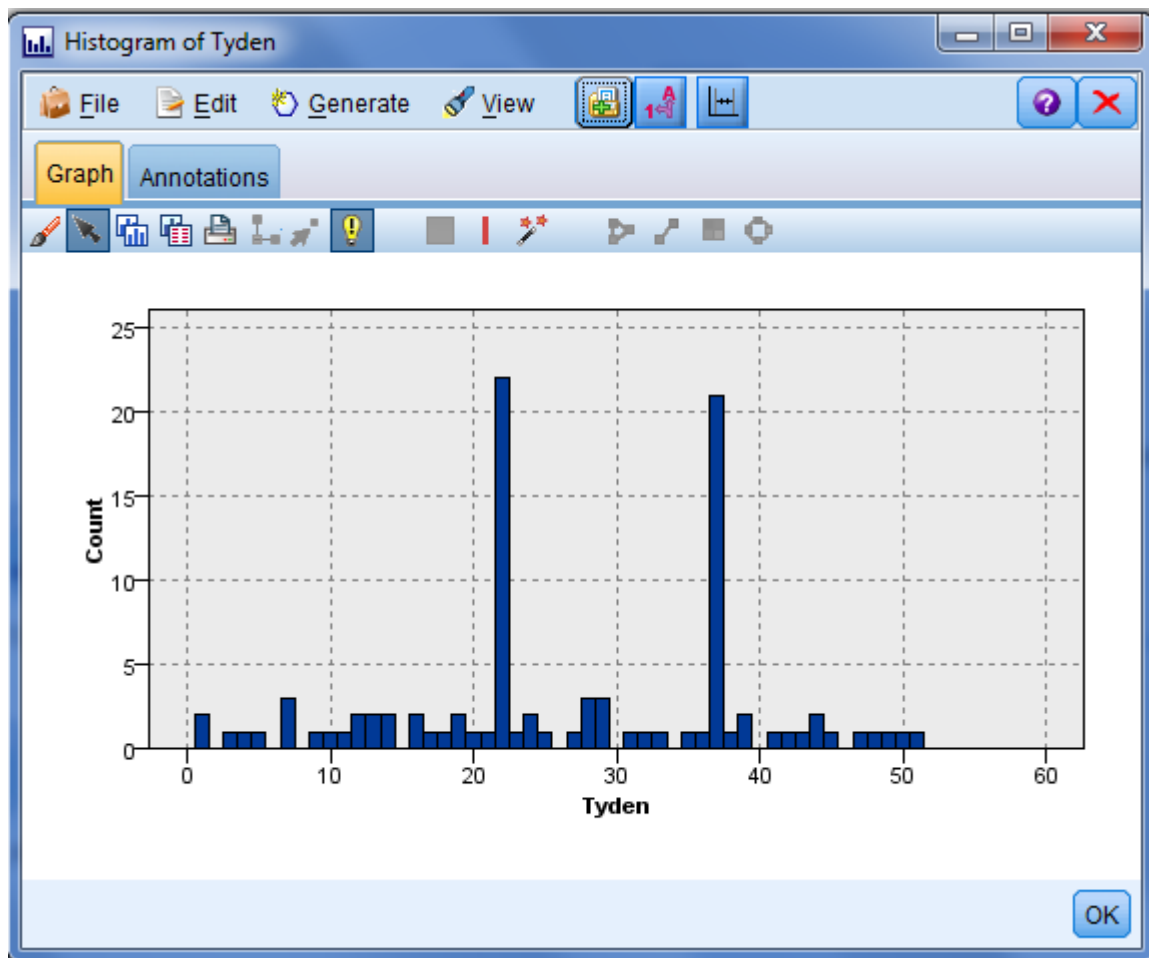
Část exportující data pro proud kapesních krádeží

Detail této části je zobrazen na příslušné ilustraci (Ilustrace 5)



Ilustrace 5: Detail na část streamu, která se zabývá exportem pro proud s kapesními krádežemi

Ze všech případů, byly vybrány jen ty, které se týkaly kapesních krádeží. Po zobrazení četnosti krádeží v jednotlivých týdnech (Ilustrace 6), jsme se rozhodli, vytvořit si nejprve dva nové atributy typu flag, indikující, že se jedná o případ, vyskytující se v jednom ze dvou extrémních týdnů (22 a 37).

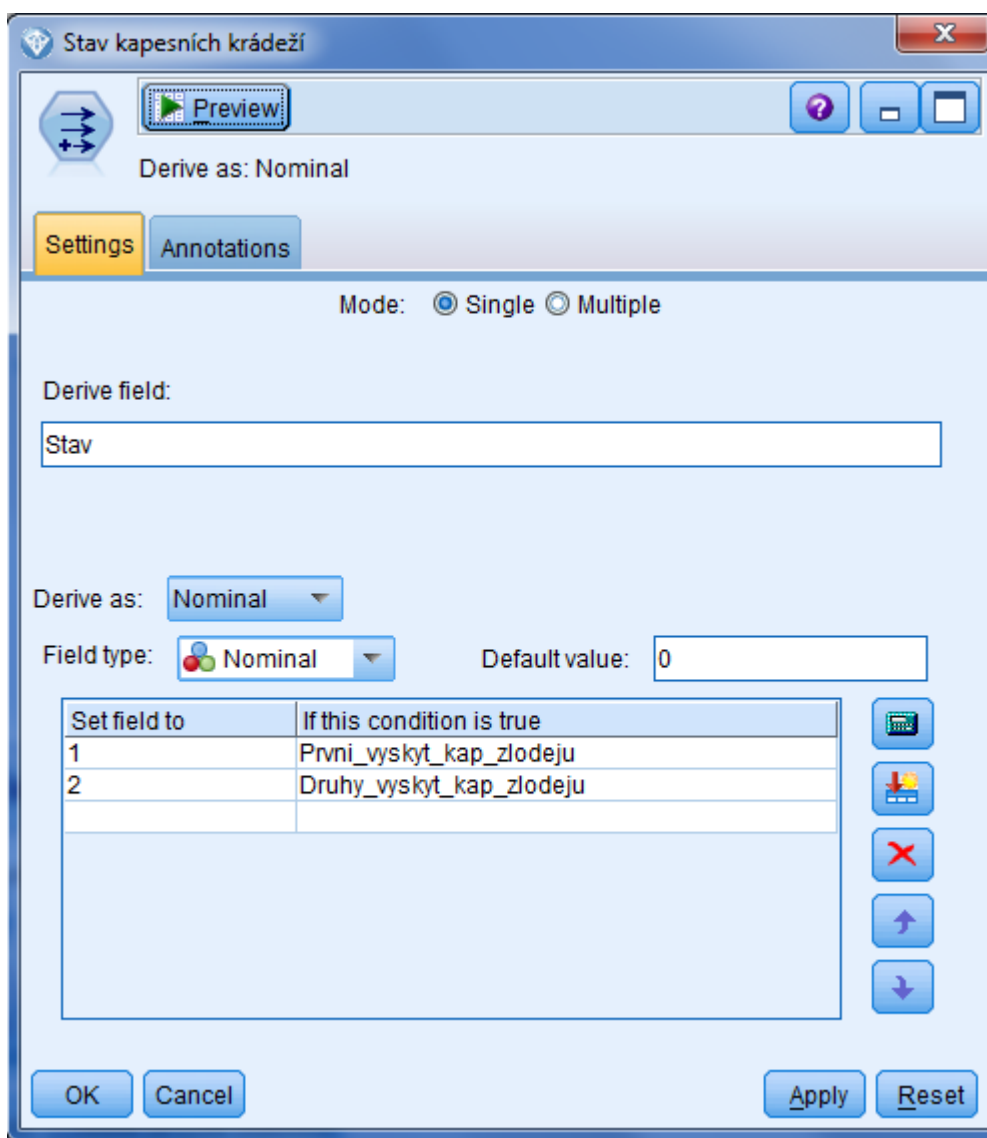


Ilustrace 6: Histogram četnosti krádeží v týdnech

V dalším atributu s názvem „Stav“ jsme tyto indikátory využili pro rozřazení všech případů kapesních krádeží do tří skupin:

- skupina krádeží mimo extrémní týdny
- skupina krádeží v prvním extrémním týdnu
- skupina krádeží v druhém extrémním týdnu

Konkrétní nadefinování tohoto atributu je předmětem Ilustrace 7



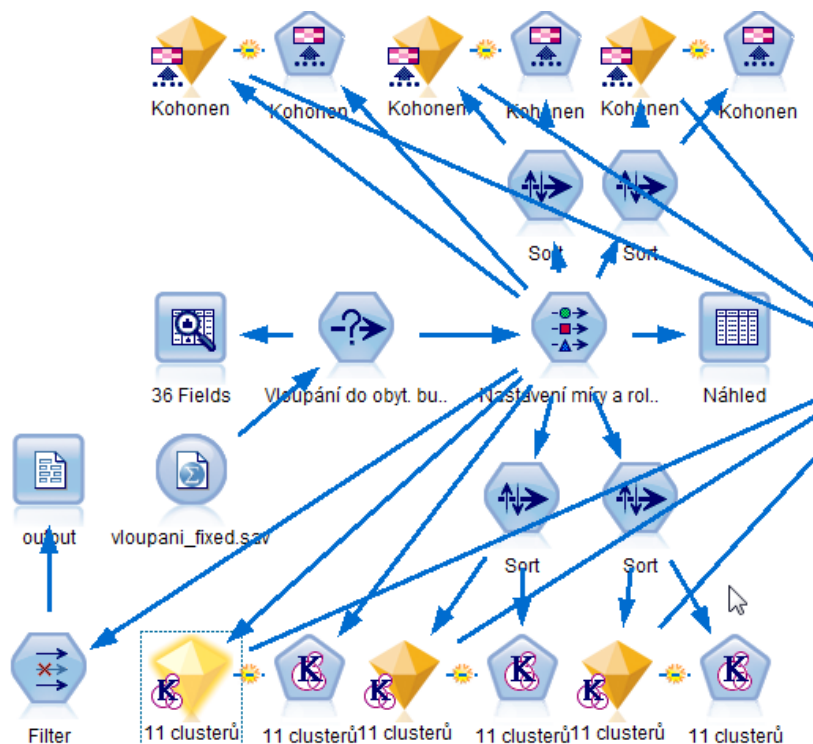
Ilustrace 7: Definice Stavů

Posledním krokem před exportem do statistického souboru, bylo vyfiltrovat atributy, které pro proud, zpracovávající úlohu s kapesními krádežemi nejsou podstatné. Jednalo se především o všechny atributy uchovávající modus operandi a příznaky charakteru krádeží.

Proud souvisejících vloupání

V tomto proudě budou použita data z přípravy dat, vytvoří se struktura několika segmentovacích uzlů za účelem co nejstabilnějších výsledků a ke konci dojde ke sloučení výsledků a závěrečnému zhodnocení. Kompletní proud je obsahem Přílohy B.

Vytvořená struktura shlukovacích uzlů



Ilustrace 8: Struktura shlukovacích uzlů

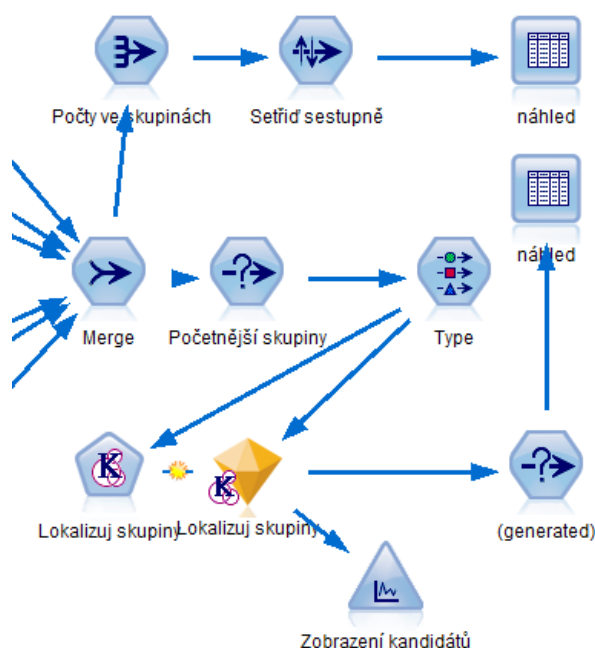
Detail je jako obvykle předmětem obrázku (Ilustrace 8).

Po načtení dat ze souboru se vyberou pouze ty případy, obsahující vloupání do obytných budov. Dále se v uzlu *Type* nastaví role jednotlivým atributům. Vstupní atributy pro nás budou pouze všechny modus operandi a příznaky krádeží. Ostatní nebudou hrát žádnou roli v modelování.

Poté se data postoupí dohromady šesti modelujícím uzlům (tři Kohonenovy mapy tři K-Means), přičemž pro každý typ segmentujícího uzlu budou poskytnuta data ve třech různých pořadích (tak jak jsme je dostali, sestupně, vzestupně). Různá pořadí zavádíme proto, abychom byli schopni identifikovat nestabilní shluky. Naše představa je taková, že případy s spolu související, by měly utvářet shluk bez ohledu na to, v jakém pořadí byly.

Počet segmentů, který hledáme a který je třeba zadat u K-Means, vychází z výsledků poskytnutých Kohonenovou mapou, který sám určuje ideální počet shluků. Tento počet má hodnotu 11, přičemž v jednom případě Kohonenovy mapy je tento počet 12.

Část kombinující výsledky shlukovaích uzlů a výsledné vyhodnocení souvisejících případů



Ilustrace 9: Detail na část streamu, která vyhodnocuje dílčí výsledky

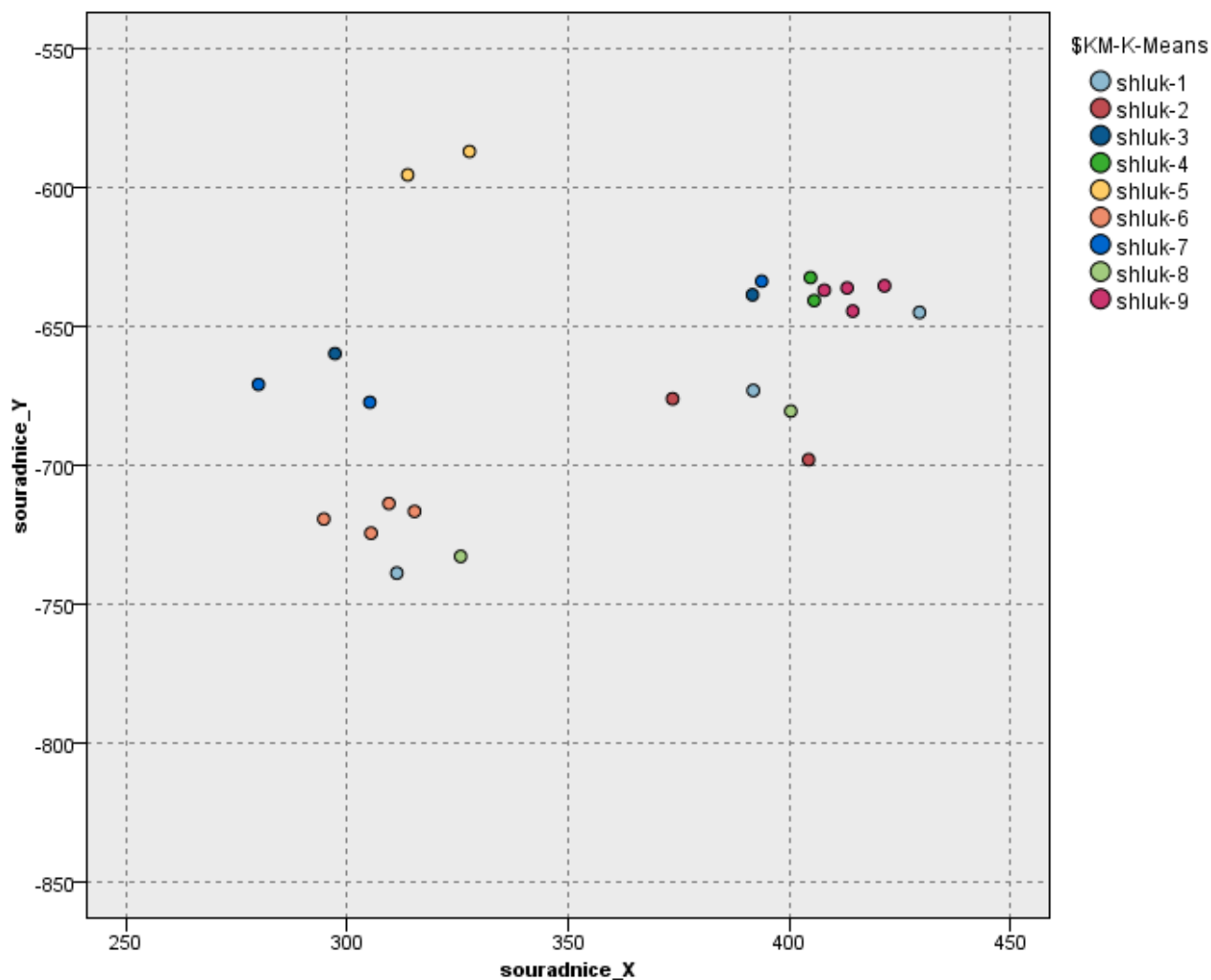
Do uzlu *Merge* vstupují výsledky ze segmentovacích uzlů předchozí části a spojují se na základě identifikátoru případu. Ke každému případu nyní tedy máme informaci o tom, k jakému shluku patřil v tom daném segmentovacím uzlu.

Jelikož je naším cílem najít ty případy, které s spolu vytvářely shluky ve všech modelech, použijeme *Aggregate* uzel, kterým zjistíme počty případů, které se vyskytují pro danou kombinaci shluků. Viz obrázek (Ilustrace 10), vyprodukovaní po sestupném setřídění.

	\$KM-K-Means1	\$KM-K-Means2	\$KM-K-Means3	\$KXY-Kohonen1	\$KXY-Kohonen2	\$KXY-Kohonen3	Record_Count
1	skupina-8	skupina-2	skupina-7	X=0, Y=2	X=3, Y=0	X=3, Y=0	4
2	skupina-9	skupina-8	skupina-8	X=3, Y=0	X=0, Y=0	X=0, Y=0	4
3	skupina-5	skupina-3	skupina-10	X=0, Y=0	X=0, Y=2	X=0, Y=2	3
4	skupina-4	skupina-3	skupina-9	X=3, Y=0	X=0, Y=0	X=0, Y=0	3
5	skupina-6	skupina-2	skupina-1	X=0, Y=2	X=3, Y=0	X=3, Y=0	2
6	skupina-8	skupina-6	skupina-3	X=3, Y=2	X=3, Y=2	X=3, Y=2	2
7	skupina-1	skupina-9	skupina-11	X=3, Y=2	X=3, Y=2	X=3, Y=2	2
8	skupina-11	skupina-3	skupina-1	X=3, Y=0	X=1, Y=0	X=1, Y=0	2
9	skupina-6	skupina-7	skupina-10	X=0, Y=1	X=0, Y=2	X=0, Y=2	2
10	skupina-7	skupina-4	skupina-2	X=0, Y=0	X=0, Y=2	X=0, Y=2	1
11	skupina-8	skupina-7	skupina-4	X=0, Y=0	X=0, Y=0	X=0, Y=0	1

Ilustrace 10: Četnosti případů pro danou kombinaci shluků

Uzlem *Select* jsme si tedy vybrali početnější skupiny (>1). Jelikož jsme je chtěli vidět na mapě, použili jsme nejprve uzel *Type*, ve kterém jsme jako vstupní atributy nastavili pouze výstupy předchozích modelů a následně uzel *K-Means*, kde počet shluků byl nastaven na počet kombinací, jenž nás zajímaly. V tomto případě to bylo 9 kombinací (9x jsme dostali kombinaci s větším počtem případů, než-li 1). Výsledek tohoto zobrazení je předmětem Ilustrace 11.



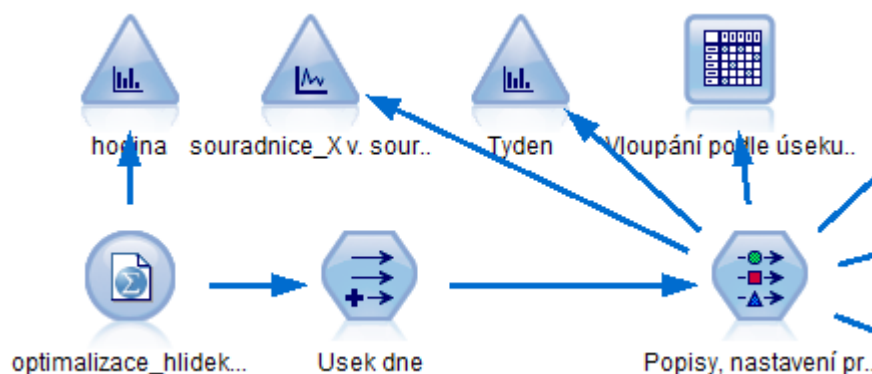
Ilustrace 11: Zobrazení kandidátů souvisejících případů na mapě

Z předchozí ilustrace je patrné, že některé případy, spadající do jednoho shluku, se vyskytují poměrně v blízkosti, a tak je na snadě, že právě tyto případy s spolu souvisí. Z tohoto důvodu je poslední činnosti vybrání určité skupiny, prozkoumání hodnot modus openradi a dalších charakteristických znaků krádeže.

Proud optimalizace rozmístění hlídek

Proud se zabývá ideálním rozmístěním daných hlídek tak, aby rozmístění bylo rovnoměrné nejen z geografického hlediska, ale také z hlediska četnosti kriminální činnosti. Kompletní proud je obsahem Přílohy C.

Operace se vstupními daty



Ilustrace 12: Uzly sloužící pro fázi úpravy dat

Jedinou operací se vstupními daty bylo v tomto proudu odvození nového atributu, indikujícího o jakou část dne se jedná. Nastavení příslušného uzlu je předmětem Ilustrace 13

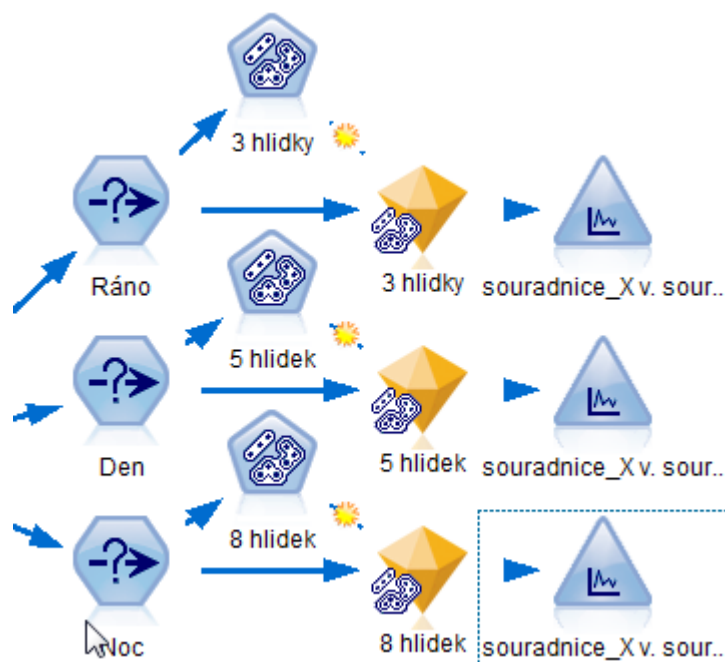
Derive as: **Nominal**

Field type: **Ordinal** Default value: **Večer a noc**

Set field to	If this condition is true
1	hodina < 10.0 and hodina >=6
2	hodina >= 10.0 and hodina < 19.0
3	hodina >= 19 or hodina <6

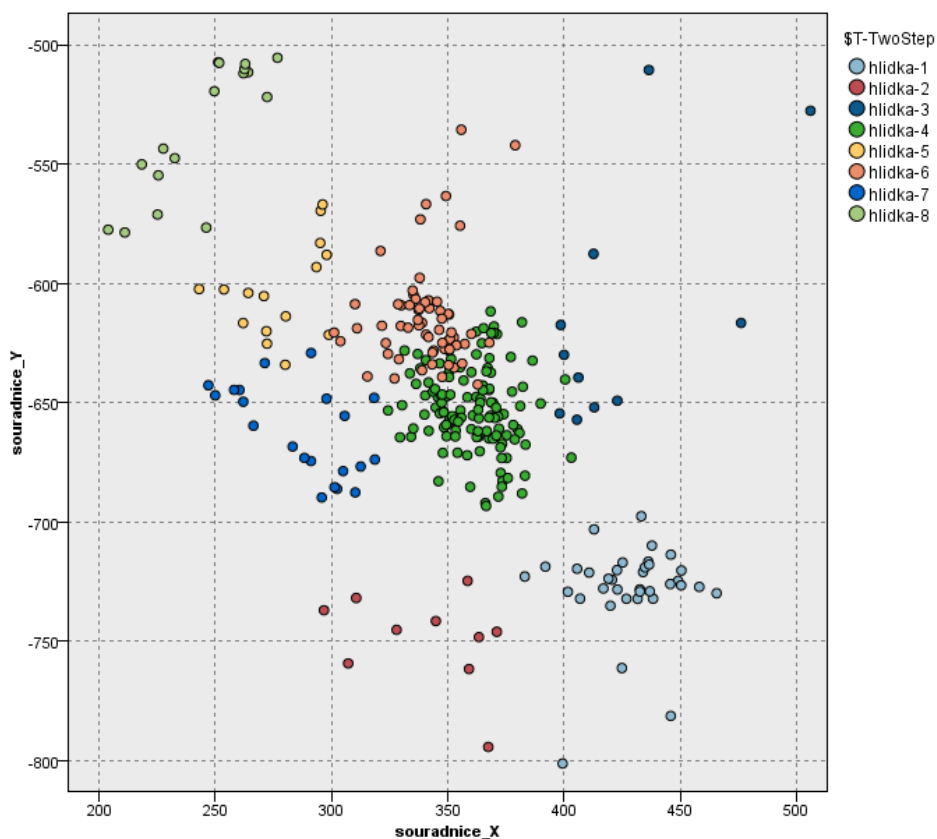
Ilustrace 13: Odvození nového atributu s částí dne

Použití shlukovacích uzlů



Ilustrace 14: Výsledný vzhled části streamu s modelovacími uzly

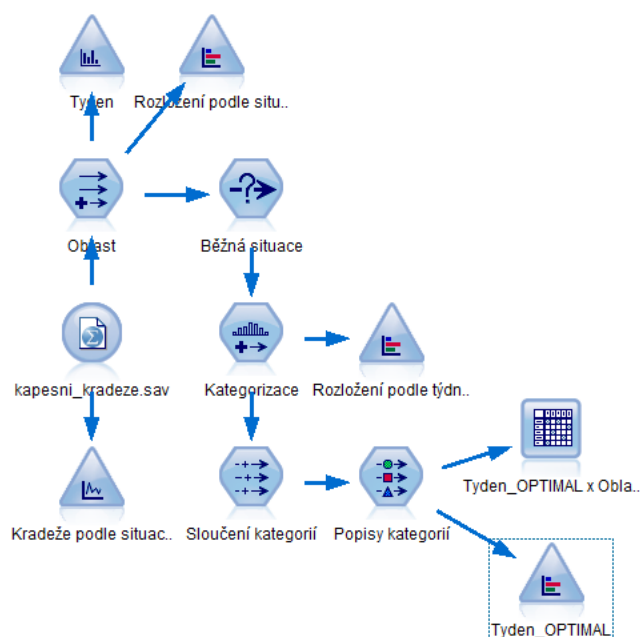
Poté co jsme dosáhli rozdělení případů dle částí dne, jsme případy dle tohoto rozdělení rozvětvili a pro každou větev použili uzel *TwoStep* s jiným nastavením počtu shluků. Tyto počty vycházely z četností případů na danou část dne a jednalo se čistě o subjektivní záležitost. Uzly *TwoStep* tedy neměly zapnutou funkci automatického určení počtu shluků, jelikož její výsledky byly neuspokojivé. Příkald rozmístění v noci je zobrazen na Ilustrace 15.



Ilustrace 15: Rozmístění hlídek v noci

Proud zabývající se kapesními krádežemi

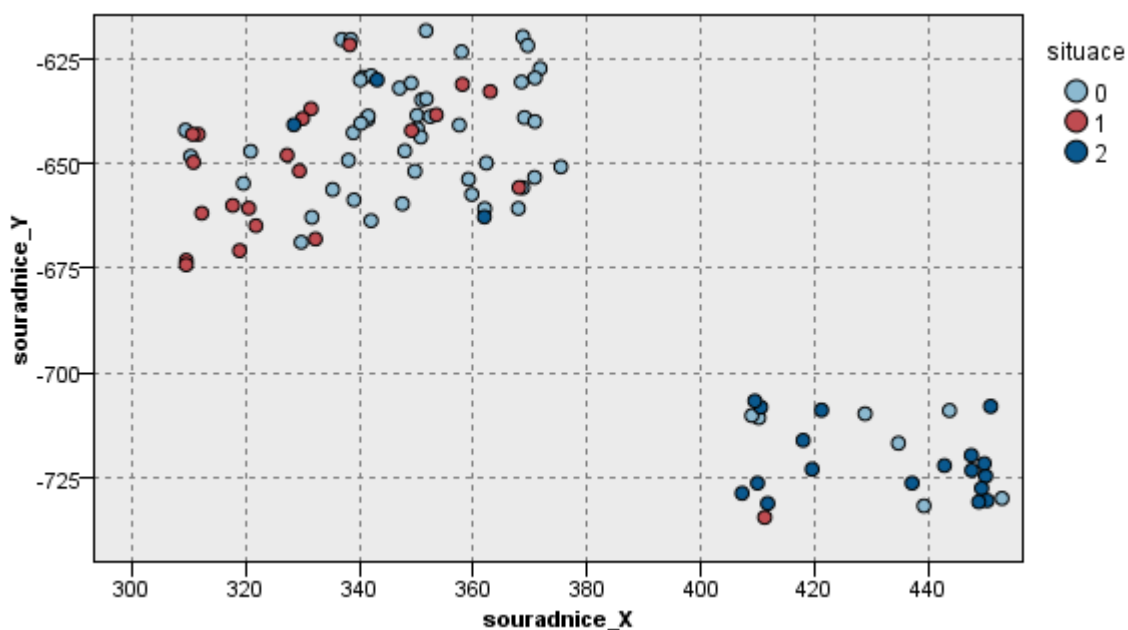
Ve fázi přípravy dat jsme zjistili, že výskyt kapesních krádeží prudce narostl v období dvou konkrétních týdnů. Budeme chtít zjistit, v jaké oblasti k těmto krádežím došlo, a také stanovit oblast, ve které se obvykle kapesní krádeže mimo tyto specifické situace vyskytují.



Ilustrace 16: Kompletní proud

Vizualizace vstupních dat

Jako první krok jsme se podívali na rozmístění krádeží na mapě, dle dané situace. Situací se rozumí hodnota z vytvořeného atributu Stav. Viz Ilustrace 17.

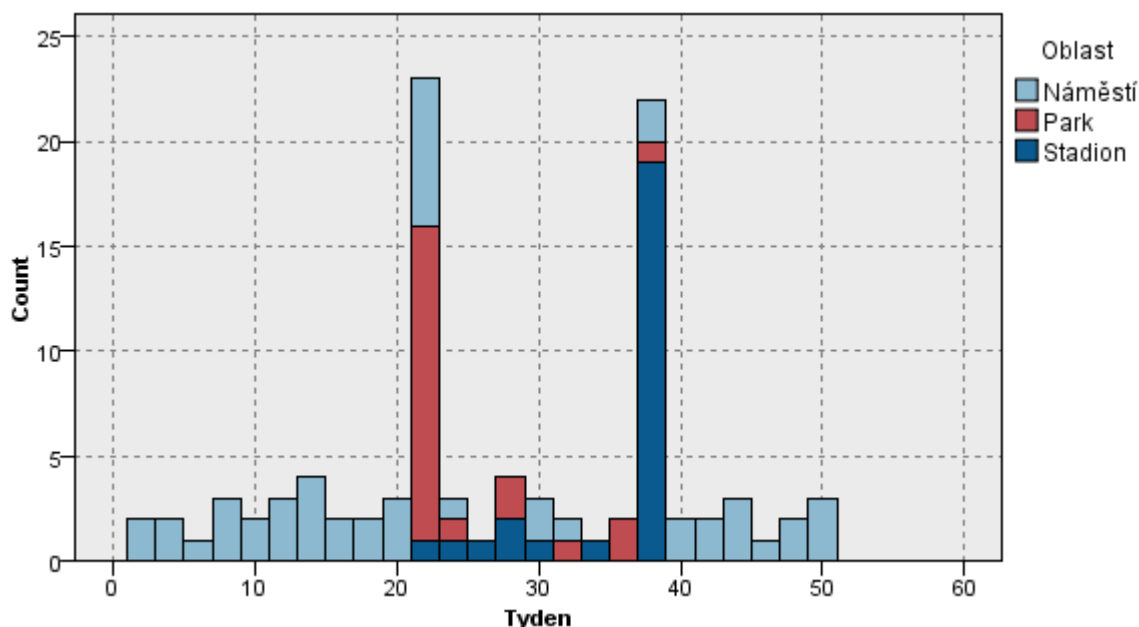


Ilustrace 17: Rozmístění kapesních krádeží

Z dalších příložených souborů k této studii jsme se dozvěděli, že se ve skutečnosti jedná o tři lokace: Náměstí, Park a Stadion. Zajímá nás tedy četnost krádeží v těchto lokacích v jednotlivé období v roce.

Selekce a kategorizace zajímavé domény

Nejprve jsme si pomocí uzlu *Derive* odvodili ono umístění. Pomocí uzlu *Histogram* jsme zobrazili rozložení případů kapesních krádeží podle času s rozdělením na jednotlivé oblasti ve skládaném histogramu. Z obrázku (Ilustrace 18) lze zjistit, že v oblasti stadionu a parku se vyskytují kapesní krádeže pouze v polovině roku, zatímco na náměstí k nim dochází během celého roku.



Ilustrace 18: Četnost krádeží v lokacích v závislosti na týdnu

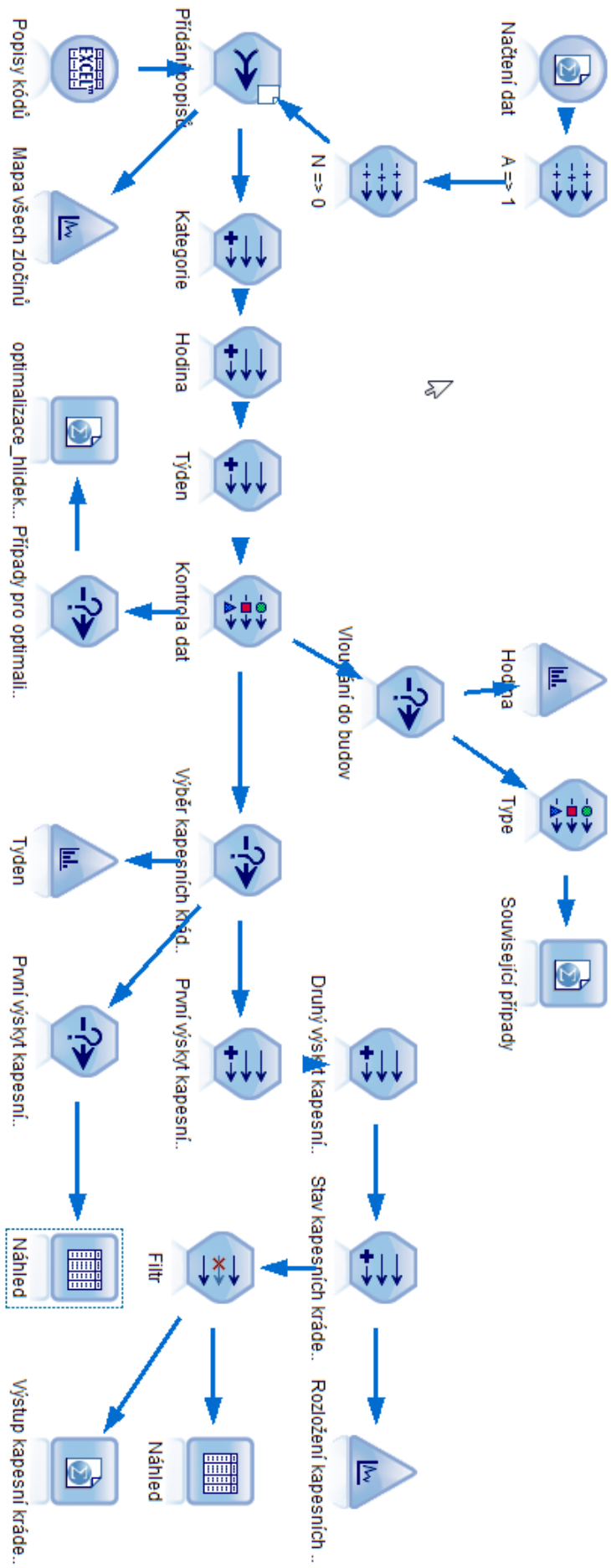
Dále jsme zjišťovali jaké je proporcionální zastoupení kapesních krádeží v extrémních týdnech a v týdnech neextrémních. Zvolili jsme si uzel *Distribution* a v něm zaškrtnuli nastavení *Normalize by Color*. Výstup tohoto uzlu nás informuje o tom, že v průběhu první krizové situace došlo k většině krádeží v parku, zatímco v druhé krizové situaci nastaly krádeže na stadioně. Mimo tyto krizové situace se kapesní krádeže odehrávají hlavně na náměstí.

Dalšími kroky v tomto proudu už jen konkretizujeme to co je už nyní patrné.

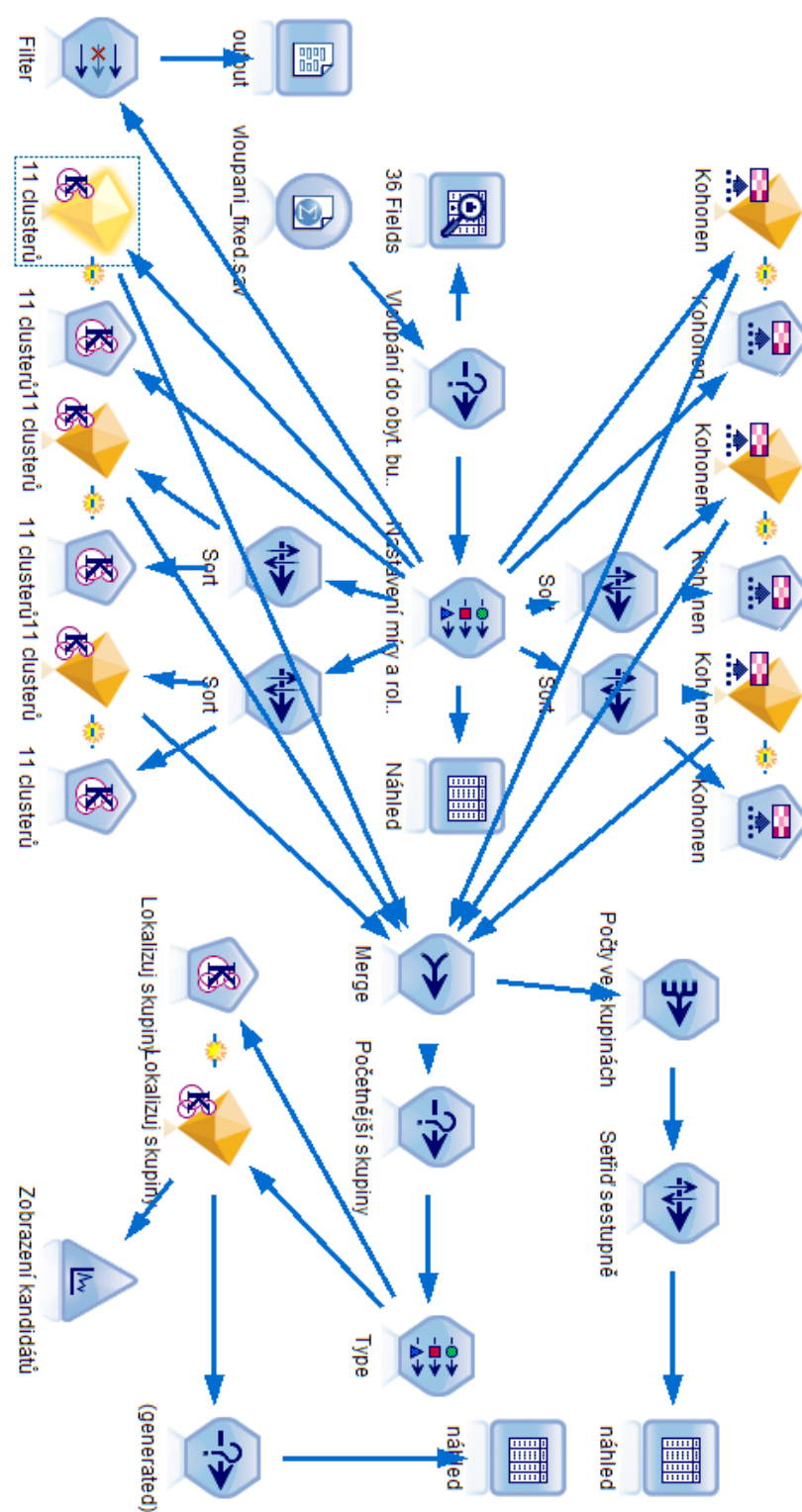
Závěrečný report

Na základě výstupů můžeme prohlásit, že lze rozdělit rok na dvě části. Mimo letní období dochází ke kapesním krádežím pouze na náměstí, zatímco během léta jsou případy kapesních krádeží zhruba rovnoměrně zastoupeny i v parku a na stadionu.

Příloha A



Příloha B



Příloha C

