

Univerzita Pardubice  
Fakulta ekonomicko-správní

Analýza časových řad pomocí metod data miningu

Bc. Tomáš Vyskot

Bakalářská práce

2010

Zadání bakalářské práce

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 30. června 2010

Bc. Tomáš Vyskot

Touto cestou bych chtěl poděkovat vedoucímu bakalářské práce panu Ing. Tomáši Kořínkovi za spolupráci a odbornou konzultaci při vypracování bakalářské práce. Dále bych rád poděkoval panu kpt. Petru Bábovkovi z hasičského záchranného sboru v Pardubicích za poskytnutá data, objasnění profesních postupů a odborných pojmů k práci HZS.

## **SOUHRN**

Bakalářské práce je věnována získávání informací z databáze pomocí metod data miningu. Data mining je proveden na základě metodiky CRISP-DM a využito softwarového prostředí Clementine. Data jsou charakterizována nejdříve za pomoci popisné statistiky, dále jsou zjišťovány závislosti mezi proměnnými, z vybraných dat je vytvořen lineární regresní model, výsledky jsou průběžně interpretovány a na závěr vyhodnoceny včetně doporučení pro praktické využití.

## **KLÍČOVÁ SLOVA**

Data mining, CRISP-DM, lineární regrese, metoda nejmenších čtverců, datová matice, časová řada.

## **TITLE**

Time series analysis using data mining techniques.

## **ABSTRACT**

Work is devoted to obtaining information from a database using data mining techniques. Data mining is based on the CRISP-DM methodology and software environment used Clementine. The data are first characterized using descriptive statistics, the dependence between the measured variables, the selected data is a linear regression model, the results are continuously interpreted and finally evaluated, including recommendations for practical use.

## **KEYWORDS**

Data mining, CRISP-DM, linear regression, the method of least squares, data matrix, time series

# OBSAH

|  |           |
|--|-----------|
| ÚVOD .....   | 7         |
| <b>1. ZÁKLADNÍ POJMY DATA MININGU .....</b>                    | <b>9</b>  |
| 1.1. ZÁKLADNÍ ZDROJE DATA MININGU .....                        | 9         |
| 1.2. METODIKY DATA MININGU .....                               | 10        |
| 1.3. METODIKA CRISP-DM .....                                   | 10        |
| <b>2. VYBRANÉ METODY A NÁSTROJE.....</b>                       | <b>13</b> |
| 2.1. SOFTWARE CLEMENTINE.....                                  | 13        |
| 2.2. DATOVÁ MATICE .....                                       | 13        |
| 2.3. POPISNÁ STATISTIKA.....                                   | 14        |
| 2.4. ZÁKLADNÍ POJMY ČASOVÉ ŘADY .....                          | 16        |
| 2.5. REGRESNÍ ANALÝZA.....                                     | 17        |
| 2.5.1. JEDNODUCHÝ MODEL LINEÁRNÍ REGRESE.....                  | 18        |
| 2.5.2. METODA NEJMENŠÍCH ČTVERCŮ.....                          | 19        |
| <b>3. DATA MINING .....</b>                                    | <b>21</b> |
| 3.1. POROZUMĚNÍ PROBLÉMU .....                                 | 21        |
| 3.2. POROZUMĚNÍ DATŮM.....                                     | 22        |
| 3.2.1. DEFINICE POJMŮ V DATECH .....                           | 22        |
| 3.3. PŘÍPRAVA DAT.....   | 24        |
| 3.4. MODELOVÁNÍ DAT .....                                      | 29        |
| 3.4.1. ANALÝZA VÝJEZDŮ HZS PO DNECH.....                       | 29        |
| 3.4.2. ANALÝZA VÝJEZDŮ HZS PO TÝDNECH.....                     | 35        |
| 3.4.3. ANALÝZA VÝJEZDŮ HZS PO MĚSÍCÍCH .....                   | 37        |
| 3.4.4. ANALÝZA VÝJEZDŮ HZS A SPOLU ZASAHUJÍCÍCH JEDNOTEK ..... | 42        |
| 3.5. HODNOCENÍ VÝSLEDKŮ .....                                  | 45        |
| 3.5.1. ANALÝZA VÝJEZDŮ HZS PO DNECH.....                       | 45        |
| 3.5.2. ANALÝZA VÝJEZDŮ HZS PO TÝDNECH.....                     | 46        |
| 3.5.3. ANALÝZA VÝJEZDŮ HZS PO MĚSÍCÍCH .....                   | 46        |
| 3.5.4. ANALÝZA VÝJEZDŮ HZS A SPOLU ZASAHUJÍCÍCH JEDNOTEK ..... | 46        |
| 3.5.5. POSOUZENÍ PROCESU METODY CRISP-DM .....                 | 47        |
| 3.6. DOPORUČENÍ PRO PRAXI .....                                | 47        |
| <b>ZÁVĚR.....</b>  | <b>48</b> |
| <b>POUŽITÁ LITERATURA .....</b>                                | <b>49</b> |
| <b>SEZNAM POUŽITÝCH ZKRATEK.....</b>                           | <b>50</b> |
| <b>SEZNAM OBRÁZKŮ.....</b>                                     | <b>51</b> |
| <b>SEZNAM TABULEK .....</b>                                    | <b>52</b> |
| <b>SEZNAM PŘÍLOH .....</b>                                     | <b>52</b> |
| <b>PŘÍLOHY .....</b>   | <b>53</b> |

# ÚVOD

O dobývání znalostí z databází se začalo ve vědeckých kruzích mluvit počátkem 90. let minulého století. První impuls přišel z Ameriky, kde se na konferencích věnovaných umělé inteligenci pořádaly první workshopy věnované této problematice. Databázové technologie představují osvědčený prostředek jak uchovávat rozsáhlá data a vyhledávat v nich informace, statistika představuje osvědčený prostředek jak modelovat a analyzovat závislosti v datech. Po léta se tyto disciplíny vyvíjely nezávisle, až přišla ta chvíle, kdy rozsah automaticky sbíraných dat začínal uživatelům přerůstat přes hlavu. Na základě potřeby tyto data využívat pro podporu rozhodování vznikla nutnost propojit oba prostředky a tím vzniklo dobývání znalostí z databází.[2]

Cílem bakalářské práce je aplikovat vybranou metodu dobývání znalostí z databáze obsahující datovou matici s časovou řadou. Práci jsem rozdělil do dvou logických částí. První část je z oblasti teorie a druhá část je zaměřena na aplikaci metody CRISP\_DM pro získání znalostí z databáze Hasičského záchranného sboru Pardubice.

V první kapitole jsou vymezeny základní pojmy data miningu, metody a principy data miningu. Popis CRISP\_DM je zde uveden pomocí popisu etap této metody pro data mining. Druhá kapitola obsahuje vybrané metody a pojmy užití v této práci a to: vybrané popisné statistiky, popis datové matice, základní definici časové řady, základy lineární regrese a princip metody nejmenších čtverců. Dále pak základní popis charakteristických situací, při kterých se program Clementine využívá. Clementine byl navržen na základě metodologie CRISP-DM a je vhodným softwarovým prostředkem pro praktickou část této bakalářské práce, která je provedena v kapitole 3. Praktická část je provedena v krocích porozumění problému, porozumění datům, modelování dat, hodnocení výsledkům a doporučení pro praxi.

Pro data mining je použita databáze HZS Pardubice, kde jedna datová matice obsahuje počty výjezdů k zásahu požáru, dopravní nehodě, živelné pohromě, technické havárii, úniku nebezpečných látek, počtu planých poplachů v denních úsecích. Druhá datová matice obsahuje údaje o počtu společných výjezdů k zásahům s PČR, ZS, OBP, ENERGA a jiných jednotek. Tato datová matice má záznamy pořizeny do kontingenčních tabulek v měsíčních agregovaných úsecích.

Rozsah vstupních dat je 2752 záznamů z období od 1.1.2006 do 6.6.2010. Data vznikla exportem přes uživatelskou konzolu, která podporovala export pouze do XML dokumentů a

následně exportována do textových souborů typu \*.csv. Na těchto datech v rámci této bakalářské práce je provedena pomocí metody CRISP-DM analýza časových řad pomocí metod data miningu.

Práce obsahuje seznam použitých zkratk, datový slovník, elektronickou přílohu včetně vstupních dat zdrojových souborů a streamů, seznam použité literatury, obrázků, grafů a tabulek. Data použitá pro data mining jsou částečně veřejně přístupná na http adrese HZS Pardubice [www.hzspa.cz/informace/index.php#](http://www.hzspa.cz/informace/index.php#) .



# 1. ZÁKLADNÍ POJMY DATA MININGU

Data Mining (DM) neboli dolování z dat, je pojem zastřešující širokou škálu technik používaných v řadě odvětví. Nahlédneme-li zpět do historie, je jedna forma DM známá také jako Data Dredging, tedy „bagrování z dat“. Tento obor byl považován za něco, co je pod úrovní dobrého výzkumníka. Pojem naznačoval, že výzkumník může skutečně prohledávat data bez jakýchkoliv předběžných hypotéz. V současnosti se však této formě DM dostalo mnohem lepšího přijetí, zejména proto, že tato forma dolování z dat vedla k objevení velmi cenných informací. [4]

Pojem dobývání znalostí z databází je relevantním pojmem k pojmu data mining. Definice data mining dle B.S. Everitt, The Cambridge Dictionary of Statistics: „Data mining je netriviální dobývání skrytých předem neznámých a potenciálně užitečných informací z dat. Při jejich objevování se využívají expertní systémy a grafické a statistické techniky a prezentují se způsobem srozumitelným lidem“ [5]

## 1.1. ZÁKLADNÍ ZDROJE DATA MININGU

Základní zdroje, které data mining využívá [2]:

- Databáze
  - *Relační databáze*, v kterých data mining můžeme provádět pomocí QEB a SQL.
  - *EIS* (Executive Information System) byly uživatelsky přátelské, ale málo flexibilní pro analýzu dat z databází.
  - *OLAP* (On-Line Analytical Processing) poskytoval pohled na data jako na mnohorozměrnou tabulku nazývanou datová krychle. Je to nástroj pro analýzu (vizualizaci) dat.
  - *Datový sklad* představuje místo, kde jsou analyzovaná data uložena.
  - *Dotazovací jazyk SQL* umožňuje podobně jako OLAP najít v databázích to co hledáme.
- Statistika
  - *Kontingenční tabulky* pro zjišťování vztahu mezi dvěma kategoriálními veličinami.
  - *Regresní analýza* pro zjišťování funkční závislosti jedné numerické (spojité) veličiny na jiných numerických veličinách.

- *Diskriminační analýza* pro odlišení příkladů (pozorování) patřících do různých tříd.
- *Shluková analýza* pro nalezení skupin (shluků) navzájem si podobných příkladů.
- *Korelační analýza* pro hledání lineární závislosti dvou numerických veličin.
- Strojové učení
  - Prvky učení můžeme pod různými názvy nalézt v řadě vědních disciplín jako je umělá inteligence, strojové učení, teorii řízení.

Z těchto základních zdrojů plyne, že data mining je založen především na matematických principech a statistických technikách.

## 1.2. METODIKY DATA MININGU

S postupem doby začaly vznikat metodiky, které si kladou za cíl poskytnout uživatelův jednotný rámec pro řešení různých úloh z oblasti dobývání znalostí. Tyto metodiky umožňují sdílet a přenášet zkušenosti z úspěšných projektů [2].

Typy úloh řešených pomocí DM [5]:

- Popis dat – vizualizace, sumarizace.
- Hledání „nugetů“ – dominantní struktury, asociační pravidla, segmentace, shluková analýza, popis rozdělení dat.
- Predikce – klasifikace (predikce kategoriální proměnné), regrese (predikce spojité proměnné), časové řady (predikce proměnné závislé na čase).

Mezi metody data miningu patří například metodika 5A firmy SPSS nebo metodika SEMMA firmy SAS. Pro potřebu bakalářské práce se zaměřím na metodiku CRISP-DM, jejímž cílem je umožnit řešit rozsáhlé úkoly v dobývání znalostí z databází. Hlavními přednostmi je rychlost, efektivita, spolehlivost a nižší náklady na „dolování dat“.

## 1.3. METODIKA CRISP-DM

Životní cyklus projektu DM je podle metodiky CRISP-DM tvořen šesti fázemi [4]:

- Porozumění problému
 

Tato úvodní fáze je zaměřena na pochopení cílů úlohy a požadavků na řešení formulovaných z manažerského hlediska. Manažerská formulace musí být následně

převedena do zadání úlohy pro dobývání znalostí z databáze. V této fázi se rovněž provádí inventura zdrojů (datových, výpočetních i lidských). Hodnotí se možná rizika, náklady a přínos použití metod DM se stanovuje předběžný plán prací.

- Porozumění datům

Fáze porozumění datům začíná prvotním převzetím dat. Následující činnosti, které umožní získat základní představu o datech, která jsou k dispozici (posouzení kvality, první náhled na data, vytipování zajímavých podmnožin záznamů (polí) v databázi). Obvykle se zjišťují různé deskriptivní charakteristiky dat (četnosti, průměry, minima, maxima...). S výhodou se využívají i různé vizualizační techniky.

- Příprava dat

Příprava dat zahrnuje činnosti, které vedou k vytvoření datového souboru, který bude zpracován jednotlivými analytickými metodami. Tyto data by měla obsahovat údaje vyznačené pro danou úlohu a mít podobu, která je vyžadována vlastními analytickými algoritmy. Příprava dat zahrnuje selekci dat, čištění dat, vytváření dat, integrování dat a formátování dat. Tato fáze je obvykle nejpracnější částí řešení celé úlohy. Jednotlivé úkony jsou obvykle prováděny opakovaně, v nejrůznějším pořadí.

- Modelování

V této fázi jsou nasazeny analytické metody. Obvykle existuje řada různých metod pro řešení dané úlohy, je třeba vybrat ty nejvhodnější a vhodně nastavit jejich parametry. Jde tedy o iterační činnost, použití analytických algoritmů může navíc vést k potřebě modifikovat data a tedy k návratu k datovým transformacím.

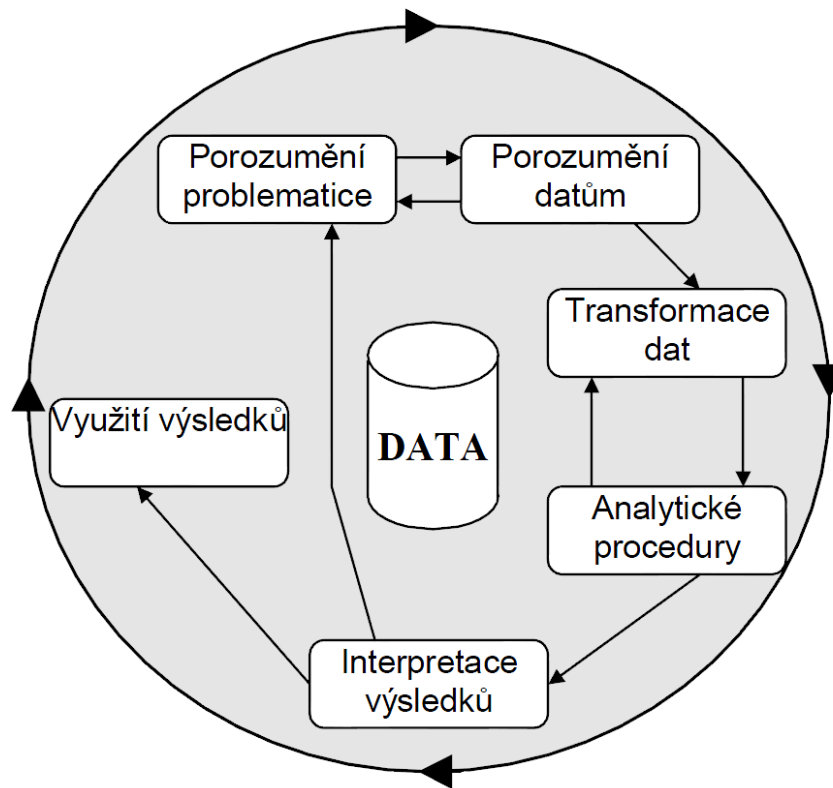
- Hodnocení

V této fázi jsou nalezeny znalosti, které se zdají být v pořádku z hlediska metod dobývání znalostí. Dosažené výsledky je potřeba vyhodnotit z pohledu manažerů, zda byly splněny cíle formulované při zadání úlohy.

- Využití výsledků

Vytvoření vhodného modelu řešení úlohy obecně nekončí. Je potřeba získané znalosti upravit do podoby použitelné pro zadavatele úlohy. Podle typu úlohy může využití výsledků znamenat na jedné straně prosté sepsání závěrečné správy, na straně druhé pak zavedení systému pro automatickou klasifikaci nových případů. Ve většině případů je to zákazník a nikoliv analytik, kdo provádí kroky k využití výsledků analýzy.

Výsledek dosažený v jedné fázi ovlivňuje volbu kroků následujících. Často je potřeba se k některým krokům a fázím vrátet. Na obrázku 1 vidíme životní cyklus CRISP-DM.



Obrázek 1 Metodika CRISP-DM [2]

## 2. VYBRANÉ METODY A NÁSTROJE

Obsahem této kapitoly je stručný popis použitých metod a nástrojů. Uvedu informace o softwarovém produktu Clementine, popíši pojem datová matice a popisná statistika. Definuji základní pojmy časové řady, lineární regrese a metodu nejmenších čtverců.

### 2.1. SOFTWARE CLEMENTINE

Pro zpracování dat metodikou CRISP-DM jsem vybral softwarový nástroj SPSS Clementine. Clementine je nástroj pro data mining, který podporuje celý data miningový proces a který svým uživatelům umožňuje rychlý přístup k datům, datové manipulaci, konstrukci a ověřování modelů a jejich následné nasazení do reálného prostředí.

Clementine kombinuje pokročilé modelovací techniky se snadným způsobem ovládání, který umožňuje objevení a predikce užitečných informací o datech. Celý systém je navržen v souladu s metodikou CRISP-DM a podporuje klíčové aktivity, mezi které patří:

- tvorba zákaznických profilů a určení jejich hodnoty,
- detekce a predikce podvodů,
- detekce a predikce vazeb v datech z webu,
- predikce budoucích prodejních a růstových trendů,
- odhad účinnosti marketingových akcí, kreditní riziko,
- odhad rizik v monitorování procesů,
- predikce churnu, klasifikace, segmentace zákazníků,
- analýza velmi rozsáhlých dat, objevování skrytých vazeb a struktur.

### 2.2. DATOVÁ MATICE

Při statistické analýze jsou předmětem zkoumání *statistické znaky*, jako například počet zásahu u dopravní nehody (ano, ne). Statistický znak je tedy věcně, prostorově a časově vymezený pojem. Zkoumáme ho u statistických jednotek, kterými jsou v našem případě požár, dopravní nehoda, technická havárie a jiné. Sledované statistické jednotky tvoří *statistický soubor*. Všechny jednotky tvoří základní soubor. Ten je v praxi většinou charakterizován na základě údajů zjištěných pouze u vybraných jednotek, které tvoří *výběrový statistický soubor*.

Proces získávání potřebných údajů se nazývá *statistické zjišťování (šetření)*. Jestliže zkoumáme celý základní soubor, pak jde o *zjišťování vyčerpávající*. Kromě výše uvedeného základního přístupu, můžeme zjišťovat údaje týkající se statistické jednotky v různých časových obdobích (v mém případě v dnech, týdnech, měsících, letech). V tomto případě dostaneme *časovou řadu* (průběh počtu požárů v 231 týdnech v řadě), která vznikla agregací (součtem) počtu požárů v jednotlivých dnech kalendářního týdne.

Základní datovou strukturou je *datová matice*, v níž každý řádek (*záznam*) obsahuje veškerá pozorování (počty výjezdů k zásahu) týkající se jedné statistické jednotky a sloupce odpovídají jednotlivým statistickým znakům. Sloupec představuje veličinu, která nabývá různých hodnot, proto se nejčastěji označuje jako *proměnná*. To je zobrazeno na obrázku 2.

|           | 1. proměnná | 2. proměnná | ... |
|-----------|-------------|-------------|-----|
| 1. případ |             |             |     |
| 2. případ |             |             |     |
| ...       |             |             |     |

Obrázek 2 Schéma datové matice [6]

V současných programových systémech mohou být data vkládána buď do tabulky (např. tabulka 2), která odpovídá zmíněné datové matici, nebo do formuláře v příloze 2, v němž jsou vyplňovány údaje odpovídající jedné statistické jednotce. [6]

### 2.3. POPISNÁ STATISTIKA

Předmětem statistického zkoumání jsou *hromadné jevy*, které jsou zastoupeny velkým počtem prvků. Ty tvoří *základní soubor*. Počet prvků základního souboru se nazývá *rozsah souboru*. Údaje (vlastnosti) uvedené pro prvky základního souboru nazýváme *statistické proměnné* nebo *znaky*. V případě že není možno zjistit či získat hodnoty statistických proměnných pro každý prvek základního souboru v tom případě pracujeme s vhodně zvoleným *výběrem ze základního souboru*. Pokud je výběr vytvořen statisticky správně, například *náhodným výběrem*, dá se na jeho základě získat určitá představa o základním souboru. Při statistických zkoumáních se zaměřujeme na charakterizování a popis *rozdělení četnosti proměnné (znaku)*, a to jak v základním souboru, tak i ve výběru.

Rozdělení četností je tabulka, která v jednom řádku obsahuje hodnoty proměnných a ve druhém odpovídající četnosti znaku. Rozdělení četností se znázorňuje graficky pomocí *histogramu* nebo *polygonu četnosti*. Četnosti hodnot znaku můžeme vyjádřit také pomocí

relativní četnosti, kdy četnost hodnot znaku je  $m$  a  $n$  je rozsah souboru, pak  $\frac{m}{n}$  je relativní četností znaku. Mluvíme pak o relativním rozdělení četností. [7]

Výše uvedené pojmy velmi dobře znázorňuje obrázek 11.

**Četnost** je počet jednotek daného výběru

**Minimum (Min)** je minimální hodnota daného výběru.

**Maximum (Max)** je maximální hodnota daného výběru

**Aritmetický průměr** je definován

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (1)$$

**Medián** je definován jako prostřední hodnota výběru, a to prostřední v pořadí hodnot uspořádaných podle velikosti. Jinak řečeno polovina hodnot výběru je menší nebo rovna mediánu a polovina hodnot je větší nebo rovna mediánu. Pokud prostřední hodnota není určena jednoznačně (například pro sudý rozsah výběru) je za medián brán průměr dvou prostředních hodnot.[7]

**Modus** je nejčetnější hodnota znaku.

**Kvantil  $x_p$**  je hodnota znaku, pro který platí, že nejméně  $p$ -procent prvků má hodnotu menší nebo rovnu  $x_p$  a  $(100 - p)$  procent prvků je větších nebo rovno  $x_p$ .

**Medián** je 50%  $x_p$ .

**Rozpětí  $R$**  je definováno jako rozdíl největší (maximální) a nejmenší (minimální) hodnoty.

**Rozptyl  $S^2$**  je definován jako součet kvadratických odchylek od průměru, děleným rozsahem výběru zmenšeným o 1. Tedy

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

**Směrodatná odchylka** je odmocninou  $S^2$ .

**Koeficient šikmosti (asymetrie)  $S_k$**  vyjadřuje, jak jsou hodnoty symetricky či asymetricky rozloženy kolem středu naměřených hodnot a vypočítá se

$$S_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot s^3} \quad (3)$$

**Koeficient špičatosti (excesu)  $E_k$**  vyjadřuje strmost, plochost rozdělení četností a platí

$$E_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n \cdot s^4} \quad (4)$$

## 2.4. ZÁKLADNÍ POJMY ČASOVÉ ŘADY

Ekonomickou časovou řadou se rozumí řada hodnot jistého věcně a prostorově vymezeného ukazatele, která je uspořádána v čase směrem od minulosti do přítomnosti. Takto definovanou časovou řadu zapisujeme jako  $y_t, t = 1, \dots, T$ .

Časové řady dělíme:

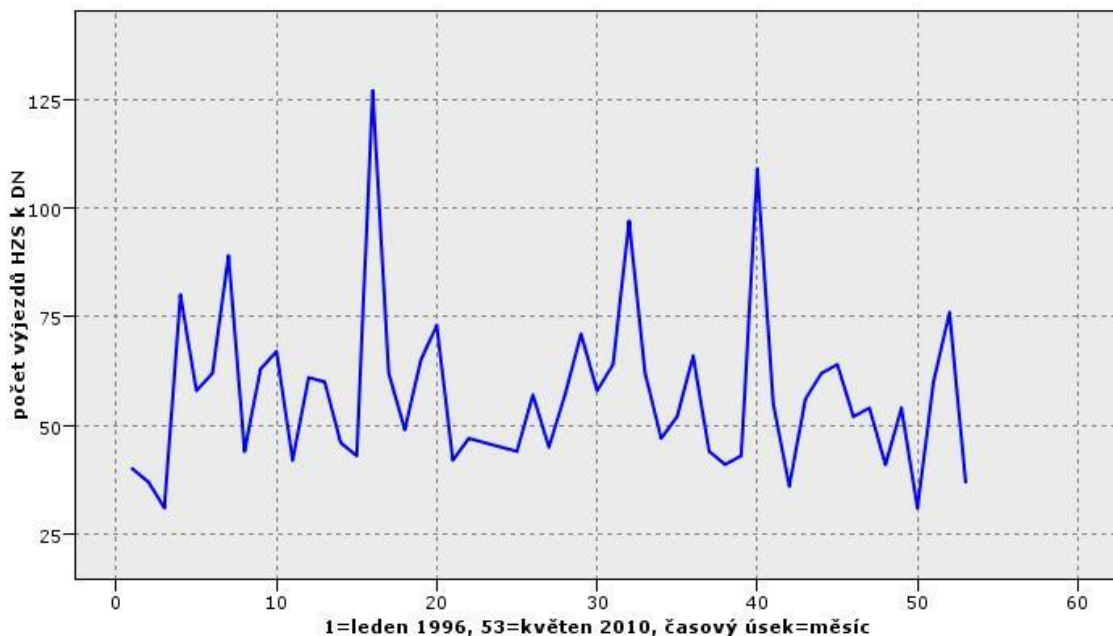
- a) *intervalová časová řada* – velikost hodnoty ukazatele závisí na délce časového intervalu sledování., typickým intervalovým ukazatelem je objem výroby,
- b) *okamžiková časová řada* – okamžikový ukazatel je ukazatelem vztahující se k jistému okamžiku, jako například počet pracovníků k jistému datu,
- c) *časová řada odvozené charakteristiky* – se získává i intervalových nebo okamžikových časových řad, jako například podíl časové řady produkce a časové řady počtu pracovníků.

Ekonomické časové řady dále dělíme na dlouhodobé a krátkodobé. Hodnoty dlouhodobých časových řad jsou sledovány v ročních či delších časových úsecích. Hodnoty krátkodobých časových řad jsou sledovány v úsecích kratších než jeden rok (čtvrtletní, měsíční, týdenní, denní...).

Základním prostředkem prezentace časových řad je jejich graf. Nejčastěji se graficky znázorňují původní časové řady, nebo kumulativní časové řady, které vznikají postupným načítáním jednotlivých hodnot. Často se, ale časové řady zobrazují tak, aby více vynikly jejich charakteristické vlastnosti a rysy. [1]

Na obrázku 3 je vidět příklad krátkodobé časové řady, kterou jsem získal v rámci DM.





Obrázek 3 Krátkodobá časová řada [zdroj vlastní]

## 2.5. REGRESNÍ ANALÝZA<sup>1</sup>

Závislost hodnot jedné proměnné na hodnotách druhé proměnné se dá vyjádřit funkčním vztahem  $y = f(x)$ . Tomuto vztahu říkáme funkční závislost. Pokud na sledovanou veličinu  $Y$  působí více náhodných veličin než jenom jedna náhodná veličina  $X$ , potom závislost mezi náhodnými veličinami  $X$  a  $Y$  nazýváme stochastickou.

### DEFINICE Stochasticky závislé veličiny

*Nechť  $X, Y$  jsou dvě náhodné veličiny. Jestliže změna hodnoty jedné náhodné veličiny vyvolá změnu rozdělení pravděpodobností druhé náhodné veličiny, říkáme, že náhodné veličiny  $X, Y$  jsou stochasticky závislé.*

Stochastické závislosti se projevují ve změnách střední hodnoty jedné náhodné veličiny souvisejících se změnami hodnot druhé náhodné veličiny, to znamená, že se projevují prostřednictvím podmíněných středních hodnot.

<sup>1</sup> Zdroj pro kapitolu 2.5. Regresní analýza je [3]

## DEFINICE Regresní funkce

Nechť  $X$  a  $Y$  jsou náhodné veličiny. Podmíněnou střední hodnotou  $E(Y|x)$ , považovanou za funkci proměnné  $x$ , budeme nazývat regresní funkcí náhodné veličiny  $X$  vzhledem k  $Y$ . Regresní funkce vyjadřuje změny podmíněné střední hodnoty jedné náhodné veličiny při změně hodnot druhé náhodné veličiny. Graf regresní funkce nazýváme regresní křivka.

Hlavním úkolem regresní analýzy je zjištění tvaru stochastické závislosti a parametrů regresní funkce. Regresní analýza se zabývá závislostí náhodné veličiny  $Y$  na nezávislé proměnné  $x$ , která není náhodná a je obecně  $m$ -rozměrná. Náhodná veličina  $Y$  má pro danou hodnotu  $x = (x_1, x_2, \dots, x_m)$  a parametry  $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$  podmíněnou střední hodnotu  $E(Y|x) = g(x, \beta_0, \beta_1, \beta_2, \dots, \beta_k)$ . Funkce  $g$  proměnné  $x$  se nazývá regresní funkce a parametry  $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$  se nazývají regresní koeficienty.

Modely lineární vzhledem k parametrům mají regresní funkci tvaru

$$g(x, \beta_0, \beta_1, \beta_2, \dots, \beta_k) = \sum_{i=0}^k \beta_i \cdot g_i(x) \quad (5)$$

### 2.5.1. JEDNODUCHÝ MODEL LINEÁRNÍ REGRESE

Jednoduchý model lineární regrese je lineární model, kdy graf regresní funkce je přímka. Předpokládejme, že  $Y_1, Y_2, \dots, Y_n$  je  $n$ -tice nekorelovaných náhodných veličin s vlastnostmi  $EY_i = \alpha + \beta x_i$ ,  $DY_i = \sigma^2$ ,  $i = 1, 2, \dots, n$ , kde  $\alpha, \beta, \sigma^2$  jsou neznámé parametry a  $(x_1, x_2, \dots, x_m)$  je  $n$ -tice známých hodnot. Jednoduchý model lineární regrese má tvar

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (6)$$

kde  $\varepsilon_i$  jsou nezávislé náhodné veličiny, pro které platí

$$E\varepsilon_i = 0, D\varepsilon_i = \sigma^2, i = 1, 2, \dots, n. \quad (7)$$

$\varepsilon_i$  je náhodnou složkou lineárního modelu, která zahrnuje působení náhodných vlivů nebo působení veličin, které nejsou v modelu zahrnuty.

## 2.5.2. METODA NEJMENŠÍCH ČTVERCŮ

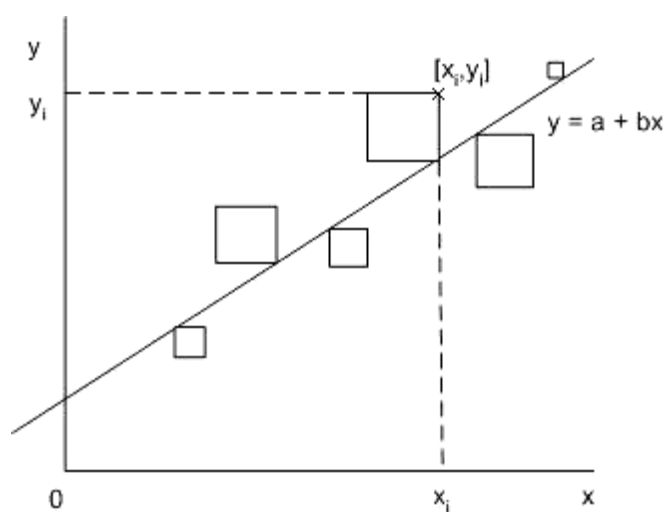
Nechť  $(x_1y_1), (x_2y_2), \dots, (x_ny_n)$  jsou dvojice naměřených hodnot. Hledáme funkci

$$\hat{y} = a + bx \quad (8)$$

,kde platí

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min. \quad (9)$$

Na základě platnosti vztahu (9) se provedou odhady  $a, b$  parametrů  $\alpha$  a  $\beta$ . Myšlenka je zachycena na obrázku 4..



Obrázek 4 Metoda nejmenších čtverců [4]

Pokud jsou splněny předpoklady o náhodné složce  $\varepsilon_i$  vztahu (7) a pro každé  $i, j = 1, 2, \dots, n$  platí:

- náhodné chyby  $\varepsilon_i, \varepsilon_j$  jsou nekorelované  $cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$
- $\varepsilon_i \sim N(0, \sigma^2)$

jsou odhady  $a, b$  parametrů  $\alpha$  a  $\beta$

$$a = \frac{\sum_i x_i^2 \cdot \sum_i y_i - \sum_i x_i \cdot \sum_i x_i \cdot y_i}{n \cdot \sum_i x_i^2 - (\sum_i x_i)^2} \quad (10)$$

$$b = \frac{n \cdot \sum_i x_i \cdot y_i - \sum_i x_i \cdot \sum_i y_i}{n \cdot \sum_i x_i^2 - (\sum_i x_i)^2} \quad (11)$$

Pro vyjádření „kvality“ lineárního regresního modelu pozorování proměnné  $X$  a  $Y$  jejich hodnot  $(x_1y_1), (x_2y_2), \dots, (x_ny_n)$  slouží koeficient determinace  $R^2$ . Pro výpočet se využívá residuální sumy čtverců  $S_e$  a vysvětlený součet čtverců odchylek  $S_t$  a celkový součet čtverců  $S_y$ .

$$S_e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - A - Bx_i)^2 = \sum_{i=1}^n Y_i^2 - A \sum_{i=1}^n Y_i - B \sum_{i=1}^n x_i Y_i \quad (12)$$

$$S_t = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^n (A + Bx_i - \bar{Y})^2 = A \sum_{i=1}^n Y_i + B \sum_{i=1}^n x_i Y_i - \frac{1}{n} \left( \sum_{i=1}^n Y_i \right)^2 \quad (13)$$

$$S_y = S_t + S_e = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (14)$$

$R^2$  nabývá hodnot  $\langle 0,1 \rangle$ , kde 0 vyjadřuje, že lineární model nic nevysvětluje a hodnota 1 vyjadřuje přesně lineární závislost, to znamená, že model vysvětluje vše. Koeficient determinace je dán vzorcem

$$R^2 = \frac{S_t}{S_y} \quad (15)$$

### 3. DATA MINING

Data mining bude proveden metodikou CRISP-DM. Každá z následujících kapitol je jednou fází metodologie. Ke zpracování je využito software Clementine.

#### 3.1. POROZUMĚNÍ PROBLÉMU

Práce hasičů z HZS Pardubice představuje pro stát a její občany velmi komplexní službu, které je součástí IZS a pomáhá řešit a odstraňovat následky různých druhů požárů, dopravních nehod, živelných pohrom, úniku nebezpečných plynných či chemických látek. Často společně s členy výjezdové skupiny HZS na místě zásahu provádí zásah příslušníci PČR, záchranné zdravotní služby, městské policie, místních jednotek a jiných složek. Pro potřeby občanů v těchto situacích slouží známé telefonní číslo 112.

Z této činnosti se zaznamenávají údaje o počtu denních výjezdů HZS a spolu zasahujících jednotek. Tyto vstupní data jsou předmětem data miningu této bakalářské práce. Provedu základní popisnou statistiku pokusím se nalézt skryté závislosti a vzájemné vztahy mezi proměnnými, lineární regresi a pomocí grafu budu interpretovat dosažené znalosti (informace)..

|    | DAT_OD     | DAT_DO     | Jednotka | H_12 | H_19 | P  | H_21 | H_22 | H_23 | H_24 | H_25 | DN  | H_31 | H_32 | H_33 | H_34 |
|----|------------|------------|----------|------|------|----|------|------|------|------|------|-----|------|------|------|------|
| 1  | 01.01.2006 | 31.01.2006 | A        | 34   | 0    | 34 | 105  | 0    | 3    | 0    | 1    | 109 | 0    | 0    | 0    | 0    |
| 2  | 01.01.2006 | 31.01.2006 | B        | 5    | 0    | 5  | 31   | 0    | 2    | 0    | 0    | 33  | 0    | 0    | 0    | 0    |
| 3  | 01.01.2006 | 31.01.2006 | C        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 4  | 01.01.2006 | 31.01.2006 | D        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 5  | 01.01.2006 | 31.01.2006 | E        | 2    | 0    | 2  | 2    | 0    | 0    | 0    | 0    | 2   | 0    | 0    | 0    | 0    |
| 6  | 01.01.2006 | 31.01.2006 | F        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 7  | 01.01.2006 | 31.01.2006 | G        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 8  | 01.01.2006 | 31.01.2006 | H        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 9  | 01.01.2006 | 31.01.2006 | I        | 0    | 0    | 0  | 1    | 0    | 0    | 0    | 0    | 1   | 0    | 0    | 0    | 0    |
| 10 | 01.01.2006 | 31.01.2006 | J        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 11 | 01.01.2006 | 31.01.2006 | K        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 12 | 01.01.2006 | 31.01.2006 | L        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 13 | 01.01.2006 | 31.01.2006 | M        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 14 | 01.01.2006 | 31.01.2006 | N        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 15 | 01.01.2006 | 31.01.2006 | O        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 16 | 01.01.2006 | 31.01.2006 | P        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 17 | 01.01.2006 | 31.01.2006 | R        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |
| 18 | 01.01.2006 | 31.01.2006 | S        | 0    | 0    | 0  | 3    | 0    | 0    | 0    | 0    | 3   | 0    | 0    | 0    | 0    |
| 19 | 01.01.2006 | 31.01.2006 | T        | 0    | 0    | 0  | 3    | 0    | 1    | 0    | 0    | 4   | 0    | 0    | 0    | 0    |
| 20 | 01.01.2006 | 31.01.2006 | U        | 0    | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    | 0    | 0    | 0    |

Obrázek 5 Náhled obsahu souboru spolzash.csv [zdroj vlastní]

## 3.2. POROZUMĚNÍ DATŮM

Data jsou shromažďována na základě výjezdů HZS k zásahu. Záznamy proměnných jsou pro výjezdy HZS v úsecích denních součtů počtů výjezdů a pro jednotky podílející se na společném zásahu v měsíčních úsecích.

Data byla získána z veřejného portálu hasičského záchranného sboru Pardubického kraje z adresy [www.hzspa.cz](http://www.hzspa.cz) formou XML dokumentů součtových sestav (kontingenčních tabulek). Na základě XML dokumentů (Příloha 2, Příloha 3, Příloha 4) byla zpracována data do tabulek v \*.csv formátu. Celkově bylo získáno 2752 záznamů, z toho 1618 denních záznamů počtu výjezdů HZS k zásahu a 1134 měsíčních záznamů o počtu výjezdů jednotek podílejících se na zásahu v pardubickém kraji. Charakter dat ukazují obrázek Obrázek 5 a obrázek 6. Na obrázku 7 vidíme schéma získaných dat jednotlivých proměnných.

Vstupní data jsou na datovém mediu CD-RW v příloze 1 v adresáři \\vstupni\_data. Zde je vidět celý obsah, rozsah a struktura vstupních dat tabulky vyjezdy.csv a spolzash.csv.

|    | ID | Date       | S_tyden    | E_tyden    | Den     | P_S | P_B | SumaP | DN_21 | DN_22 | DN_23 | DN_24 | DN_25 |
|----|----|------------|------------|------------|---------|-----|-----|-------|-------|-------|-------|-------|-------|
| 1  | 1  | 01.01.2006 | 01.01.2006 | 01.01.2006 | Neděle  | 3   | 1   | 4     | 6     | 0     | 0     | 0     | 0     |
| 2  | 2  | 02.01.2006 | 02.01.2006 | 08.01.2006 | Pondělí | 1   | 0   | 1     | 5     | 0     | 0     | 0     | 0     |
| 3  | 3  | 03.01.2006 | 02.01.2006 | 08.01.2006 | Úterý   | 1   | 0   | 1     | 5     | 0     | 0     | 0     | 0     |
| 4  | 4  | 04.01.2006 | 02.01.2006 | 08.01.2006 | Středa  | 0   | 0   | 0     | 2     | 0     | 0     | 0     | 0     |
| 5  | 5  | 05.01.2006 | 02.01.2006 | 08.01.2006 | Čtvrtek | 1   | 0   | 1     | 5     | 0     | 0     | 0     | 0     |
| 6  | 6  | 06.01.2006 | 02.01.2006 | 08.01.2006 | Pátek   | 2   | 0   | 2     | 2     | 0     | 0     | 0     | 0     |
| 7  | 7  | 07.01.2006 | 02.01.2006 | 08.01.2006 | Sobota  | 3   | 0   | 3     | 0     | 0     | 0     | 0     | 0     |
| 8  | 8  | 08.01.2006 | 02.01.2006 | 08.01.2006 | Neděle  | 2   | 0   | 2     | 2     | 0     | 0     | 0     | 0     |
| 9  | 9  | 09.01.2006 | 09.01.2006 | 15.01.2006 | Pondělí | 1   | 0   | 1     | 2     | 0     | 1     | 0     | 0     |
| 10 | 10 | 10.01.2006 | 09.01.2006 | 15.01.2006 | Úterý   | 2   | 0   | 2     | 4     | 0     | 0     | 0     | 0     |
| 11 | 11 | 11.01.2006 | 09.01.2006 | 15.01.2006 | Středa  | 0   | 0   | 0     | 3     | 0     | 0     | 0     | 0     |
| 12 | 12 | 12.01.2006 | 09.01.2006 | 15.01.2006 | Čtvrtek | 0   | 0   | 0     | 4     | 0     | 1     | 0     | 0     |
| 13 | 13 | 13.01.2006 | 09.01.2006 | 15.01.2006 | Pátek   | 0   | 0   | 0     | 8     | 0     | 0     | 0     | 0     |
| 14 | 14 | 14.01.2006 | 09.01.2006 | 15.01.2006 | Sobota  | 2   | 0   | 2     | 3     | 0     | 0     | 0     | 0     |
| 15 | 15 | 15.01.2006 | 09.01.2006 | 15.01.2006 | Neděle  | 2   | 0   | 2     | 0     | 0     | 0     | 0     | 0     |
| 16 | 16 | 16.01.2006 | 16.01.2006 | 22.01.2006 | Pondělí | 1   | 0   | 1     | 4     | 0     | 0     | 0     | 0     |
| 17 | 17 | 17.01.2006 | 16.01.2006 | 22.01.2006 | Úterý   | 0   | 0   | 0     | 2     | 0     | 0     | 0     | 0     |
| 18 | 18 | 18.01.2006 | 16.01.2006 | 22.01.2006 | Středa  | 0   | 0   | 0     | 29    | 0     | 0     | 0     | 0     |
| 19 | 19 | 19.01.2006 | 16.01.2006 | 22.01.2006 | Čtvrtek | 1   | 0   | 1     | 8     | 0     | 0     | 0     | 0     |
| 20 | 20 | 20.01.2006 | 16.01.2006 | 22.01.2006 | Pátek   | 1   | 0   | 1     | 8     | 0     | 0     | 0     | 1     |

Obrázek 6 Náhled obsahu souboru vyjezdy.csv [zdroj vlastní]

### 3.2.1. DEFINICE POJMŮ V DATECH

Výjezd – je činnost HZS probíhající od nahlášení události až po uvedení techniky do provozu u zásahu.

Zásah – je činnost probíhající na místě nahlášené události až do ukončení události.

| Vyjezdy |  | Spolzash   |  |
|---------|--|------------|--|
| √ Date  |  | ID_zaznamu |  |
| S_tyden |  | DAT_OD     |  |
| E_tyden |  | DAT_DO     |  |
| Den     |  | TEXT       |  |
| P_S     |  | H_12       |  |
| P_B     |  | H_19       |  |
| SumaP   |  | H_SK1      |  |
| DN_21   |  | H_21       |  |
| DN_22   |  | H_22       |  |
| DN_23   |  | H_23       |  |
| DN_24   |  | H_24       |  |
| DN_25   |  | H_25       |  |
| SumaDN  |  | H_SK2      |  |
| ZP_31   |  | H_31       |  |
| ZP_32   |  | H_32       |  |
| ZP_33   |  | H_33       |  |
| ZP_34   |  | H_34       |  |
| ZP_35   |  | H_35       |  |
| SumaZP  |  | H_SK3      |  |
| UN_41   |  | H_41       |  |
| UN_42   |  | H_42       |  |
| UN_43   |  | H_43       |  |
| NU_44   |  | H_44       |  |
| UN_45   |  | H_45       |  |
| SumaUN  |  | H_SK4      |  |
| TH_51   |  | H_51       |  |
| TH_52   |  | H_52       |  |
| TH_53   |  | H_53       |  |
| TH_54   |  | H_54       |  |
| SumaTH  |  | H_SK5      |  |
| RH_61   |  | H_61       |  |
| MU_71   |  | H_71       |  |
| PP_81   |  | H_81       |  |
| ALL     |  | H_SK6      |  |
|         |  | H_SK7      |  |

**Obrázek 7** Schéma získaných dat bez vazeb [zdroj vlastní]

Požár – se rozumí každé nežádoucí hoření, při kterém došlo k usmrcení nebo zranění osob nebo zvířat, ke škodám na materiálních hodnotách nebo životním prostředí a nežádoucí hoření, při kterém byly osoby, zvířata, materiální hodnoty nebo životní prostředí bezprostředně ohroženy.

Dopravní nehoda – událost mající charakter činností spojených s odstraňováním následků kolize dopravních prostředků.

Živelná pohroma – událost spojená s likvidací následků škodlivě působících sil a jevů vyvolaných plošně přírodními vlivy, které ohrožují životy, zdraví, majetek nebo životní prostředí - povodně, záplavy, deště, vlivem sněhu, námrazy, větrné smrště, sesuvu půdy, zemětřesení apod.

Unik nebezpečných chemických látek – událost spojená s nežádoucím uvolněním nebezpečných chemických látek včetně ropných produktů (během výroby, dopravy nebo manipulace) a ostatních látek.

Technická havárie - události vedoucí k odstranění nebezpečí nebo nebezpečných stavů velkého rozsahu či značných následků na zdraví osob, zvířat či majetku.

Radiační havárie a nehoda - zásah u události spojený s nepřipustným uvolněním radioaktivních látek nebo ionizujícího záření.

Planý poplach - činnost jednotky vyvolaná z důvodu ohlášení požáru nebo jiné události jednotce, která se nepotvrdila.

### 3.3. PŘÍPRAVA DAT

Tato část projektu dle metody CRISP-DM byla časově nejdelší a nejsložitější. Ze surových dat tabulek vyjezd.csv a spolzash.csv bylo cílem získat tabulku obsahující pole, která budou obsahovat počet celkových zásahů u požárů, dopravních nehod, živelných pohrom, úniku nebezpečných chemických látek, technických havárií, planých poplachů a celkového počtu výjezdů v kalendářním měsíci leden 2006 až květen 2010 pro jednotlivé druhy zasahujících jednotek tedy 53záznamů s poli hodnot daných proměnných. Schéma postupu přípravy dat pro analýzu a modelování měsíčních hodnot je uvedeno na streamu z Clementine v příloze 7.

Před zahájením vytváření streamů, které povedou k vytvoření vstupních tabulek s daty pro modelování, jsem provedl pomocí uzlu *Data Audit* výpočet některých popisných statistik a zobrazení histogramu zájmových polí vstupních souborů vyjedy.csv a spolzash.csv,

| Field    | Type  | Min | Max | Sum       | Mean   | Median | Mode | Valid |
|----------|-------|-----|-----|-----------|--------|--------|------|-------|
| P        | Range | 0   | 94  | 3016.000  | 2.660  | 0      | 0    | 1134  |
| DN       | Range | 0   | 147 | 9100.000  | 8.025  | 0      | 0    | 1134  |
| ZP       | Range | 0   | 97  | 388.000   | 0.342  | 0      | 0    | 1134  |
| UNL      | Range | 0   | 16  | 764.000   | 0.674  | 0      | 0    | 1134  |
| TH       | Range | 0   | 251 | 7579.000  | 6.683  | 0      | 0    | 1134  |
| PP       | Range | 0   | 18  | 721.000   | 0.636  | 0      | 0    | 1134  |
| ALL      | Range | 0   | 419 | 21657.000 | 19.098 | 0      | 0    | 1134  |
| Jednotka | Set   | --  | --  | --        | --     | --     | A*   | 1134  |

**Obrázek 8** Popisná statistika vstupního souboru spolzash.csv [zdroj vlastní]

Zjistil jsem, že souboru spolzash.csv se nacházejí výjezdy spolu zasahujících jednotek za jednotlivé měsíce. Použití popisných statistik je v tomto případě nerelevantní a pro představu postačuje dle obrázku 8 znát, že se zde vyskytují pole, která mají všechny mediány i modusy



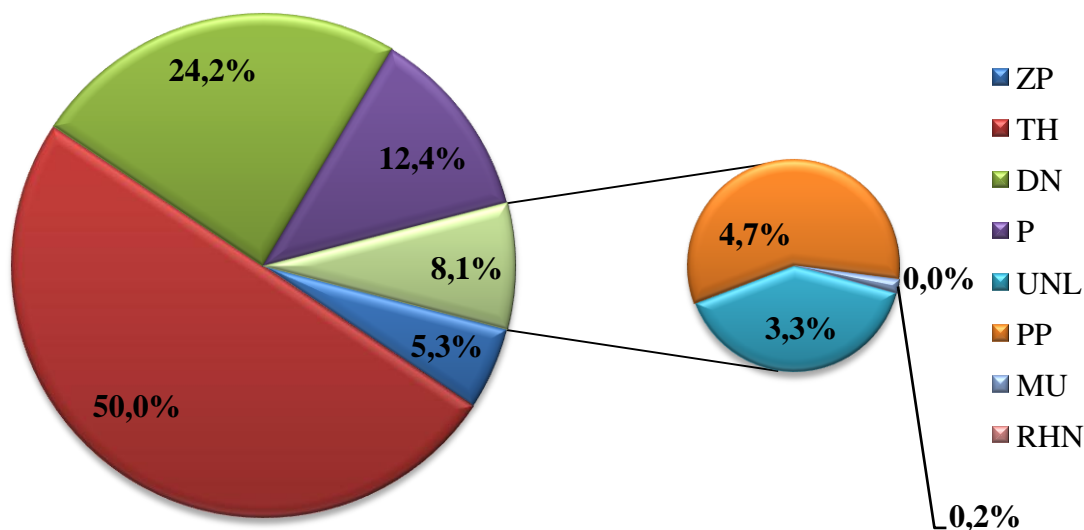
rovný 0 a to z důvodu řazení dat do kontingenčních tabulek v jednotlivých měsíčních cyklech od ledna 2006 do května 2010 kde parametrem jsou jednotlivé spolu zasahující jednotky.

| Field  | Min | Max | Mean   | Median | Mode | Sum       |
|--------|-----|-----|--------|--------|------|-----------|
| ALL    | 0   | 293 | 15.078 | 12     | 11   | 24396.000 |
| SumaZP | 0   | 260 | 0.801  | 0      | 0    | 1296.000  |
| SumaTH | 0   | 173 | 7.527  | 6      | 4    | 12179.000 |
| SumaDN | 0   | 29  | 3.645  | 3      | 3    | 5898.000  |
| SumaP  | 0   | 11  | 1.865  | 2      | 1    | 3018.000  |
| SumaUN | 0   | 5   | 0.490  | 0      | 0    | 793.000   |
| PP_81  | 0   | 5   | 0.705  | 0      | 0    | 1140.000  |
| MU_71  | 0   | 2   | 0.025  | 0      | 0    | 41.000    |
| RH_61  | 0   | 0   | 0.000  | 0      | 0    | 0         |

Obrázek 9 Popisná statistika vstupního souboru vyjezdny.csv [zdroj vlastní]

Z obrázku 9 lze vyčíst, že v souboru dat vyjezdny.csv existují dvě proměnné, proměnná RH\_61, která má všechny záznamy nulové a proměnná MU\_71, která má počet nenulových hodnot blízky nule. Tyto proměnné z další přípravy dat jsem vyloučil.

Nejčastějším typem výjezdů HZS k zásahu jsou výjezdy k technickým haváriím, dopravním nehodám a požárům. Procentuální rozdělení typu výjezdů HZS k zásahu velmi dobře interpretuje koláčový procentuální graf na obrázku 10.



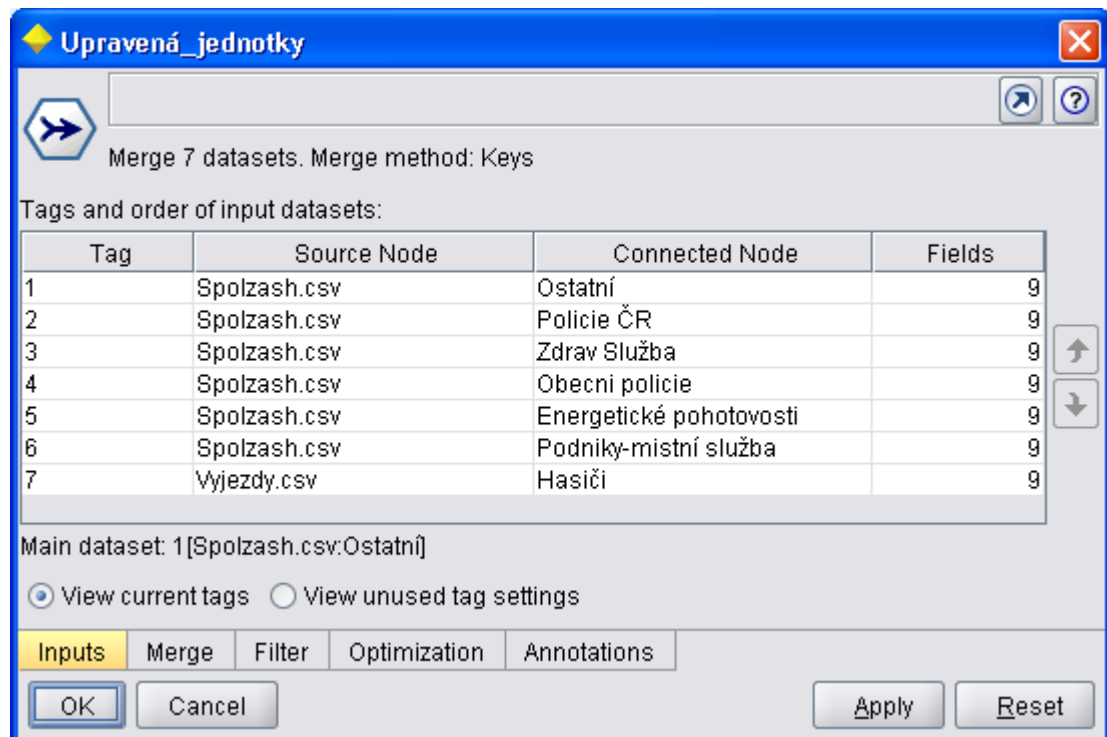
Obrázek 10 Procentuální rozdělení typu denních zásahů HZS [zdroj vlastní]

Na základě popisných statistik a histogramu rozdělení četností na obrázku 11 lze konstatovat, že rozdělení četností počtu dopravních nehod má přibližně normální rozdělení

s nesymetrickým rozložením kolem středu, kde koeficient asymetrie  $S_k = 1.843$  a koeficient excessu  $E_k = 9.228$ .

|    | Value | Proportion | %     | Count |
|----|-------|------------|-------|-------|
| 1  | 0     |            | 5,32  | 86    |
| 2  | 1     |            | 12,48 | 202   |
| 3  | 2     |            | 17,31 | 280   |
| 4  | 3     |            | 20,21 | 327   |
| 5  | 4     |            | 15,20 | 246   |
| 6  | 5     |            | 10,94 | 177   |
| 7  | 6     |            | 6,74  | 109   |
| 8  | 7     |            | 5,01  | 81    |
| 9  | 8     |            | 2,66  | 43    |
| 10 | 9     |            | 1,67  | 27    |
| 11 | 10    |            | 0,93  | 15    |
| 12 | 11    |            | 0,31  | 5     |
| 13 | 12    |            | 0,49  | 8     |
| 14 | 13    |            | 0,12  | 2     |
| 15 | 14    |            | 0,12  | 2     |
| 16 | 15    |            | 0,19  | 3     |
| 17 | 16    |            | 0,12  | 2     |
| 18 | 18    |            | 0,06  | 1     |
| 19 | 21    |            | 0,06  | 1     |
| 20 | 29    |            | 0,06  | 1     |

Obrázek 11 Rozdělení četností dopravních nehod za 24 hodin [zdroj vlastní]



Obrázek 12 Seznam upravených vstupních uzlů do uzlu Merge [zdroj vlastní]

Postup přípravy sloučení dat ze souborů vyjezdy.csv a spolzash.csv v Clementine pro modelování popisuje stream uvedený v příloze 7 :

- Načtení souboru \*.csv pomocí uzlu *Var.File* , zde jsem provedl vstupní ošetření názvu polí což je uvedeno v datovém slovníku.
- Generování nových polí pomocí uzlu *Derive*, jsem vytvořil pole s obsahem čísla kalendářního měsíce a roku například pomocí příkazu *substring(4,2,Date)*.
- Filtrování dat pomocí uzlu *Filter* na základě prvotního auditu vstupních dat, kde jsem zjistil, že pole RHN obsahuje pouze nulové hodnoty, pole OMU vykazuje velmi málo nenulových hodnot cca 2,35% (max = 2).
- Odstranění nepotřebných záznamů v tabulce vyjezdy.csv měsíce červen roku 2010, z důvodu neúplnosti sledovaného měsíce (data jsou pořizena od 1.1.2006 do 6.6.2010) pomocí uzlu:

*Derive* (as flag) příkaz: *not(Rok = "2010" and Mesic = "06")*, kde *True = t* a *False = f*,  
*Select* s nastavením výběru záznamů *True*

- Vytvoření nových uzlů s poli P, DN, ZP, UNL, TH, PP, ALL ze vstupního uzlu *Spolzash.csv*, které obsahují záznamy jednotlivých spolu zasahujících jednotek. Pomocí uzlu *Derivate* označením záznamů s příslušnými názvy jednotek, *Select* výběru daných záznamů a pomocí uzlu *Aggregate* vytvoření nových souborů dat s měsíčním záznamem pro každou námi vytvořenou jednotku, což je uvedeno v příloze 6 a na obrázku 12.

Uzel *Derive I-M* příkaz: *Jednotka = "I" or Jednotka = "J" or Jednotka = "K" or Jednotka = "L" or Jednotka = "M"*.

Uzel *Derive O-T* příkaz: *Jednotka = "O" or Jednotka = "P" or Jednotka = "R" or Jednotka = "S" or Jednotka = "T"*.

Uzel *Derive O-T* příkaz: *Jednotka = "C" or Jednotka = "D" or Jednotka = "F" or Jednotka = "G" or Jednotka = "H" or Jednotka = "U" or Jednotka = "V" or Jednotka = "N"*

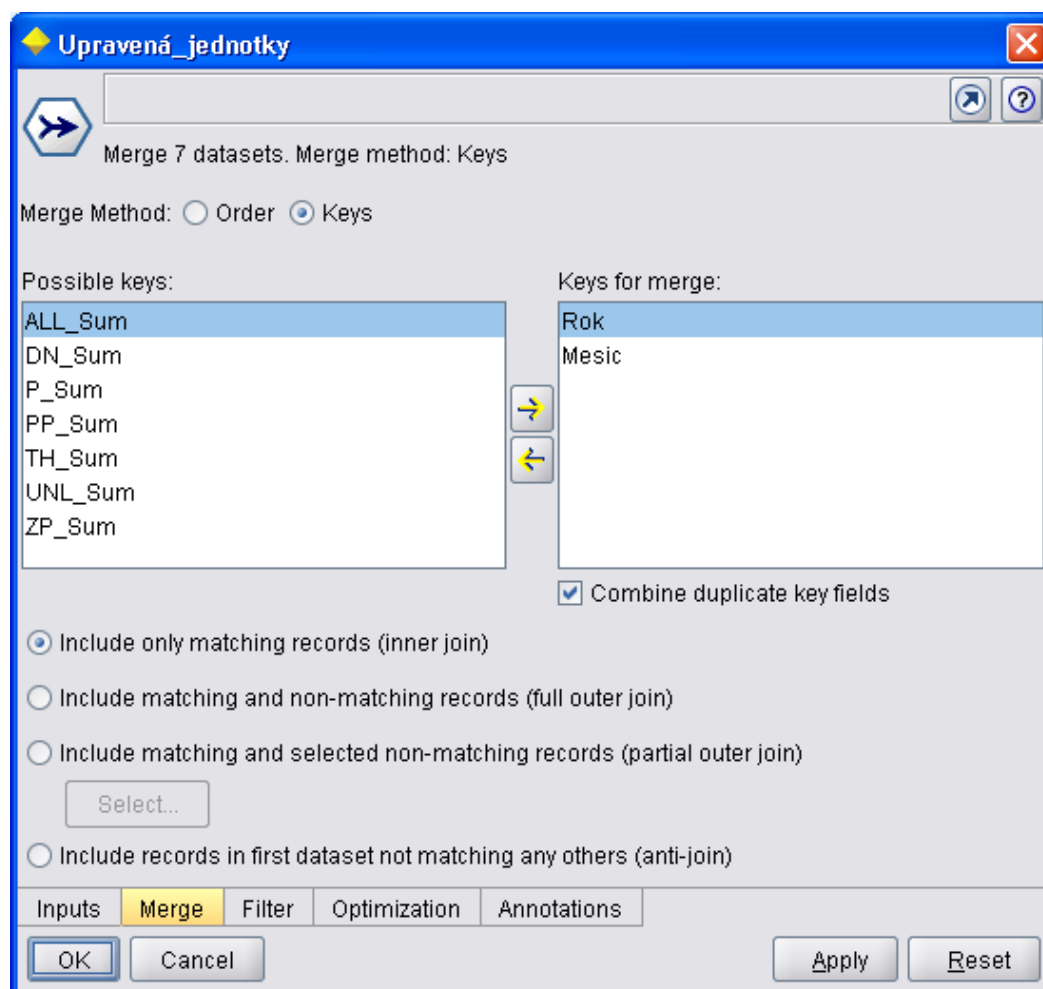
- Ve větvi streamu *Vyjezdy.csv* jsem v této fázi přes uzel *Derive – Hasiči* provedl pomocí klíčových polí *Měsíc* a *Rok* sečtení všech záznamů se shodným měsícem a rokem současně.
- Posledním krokem je sloučení vytříděných dat – pomocí uzlu *Merge – Upravená\_jednotka* jsem provedl sloučení 7-mi připojených agregovaných uzlů s nastavením slučovací metody pomocí dvou klíčů (*Rok*, *Měsíc*) obrázek 13. Tímto

krokem je zaručeno, že data z jednotlivých agregovaných uzlů jsou sloučena do tabulky, která je tvořena 51 proměnnými (viz. příloha5) a 53 záznamy, podle validních klíčů. Výsledkem je soubor modeldat.csv, který je připraven pro modelování.

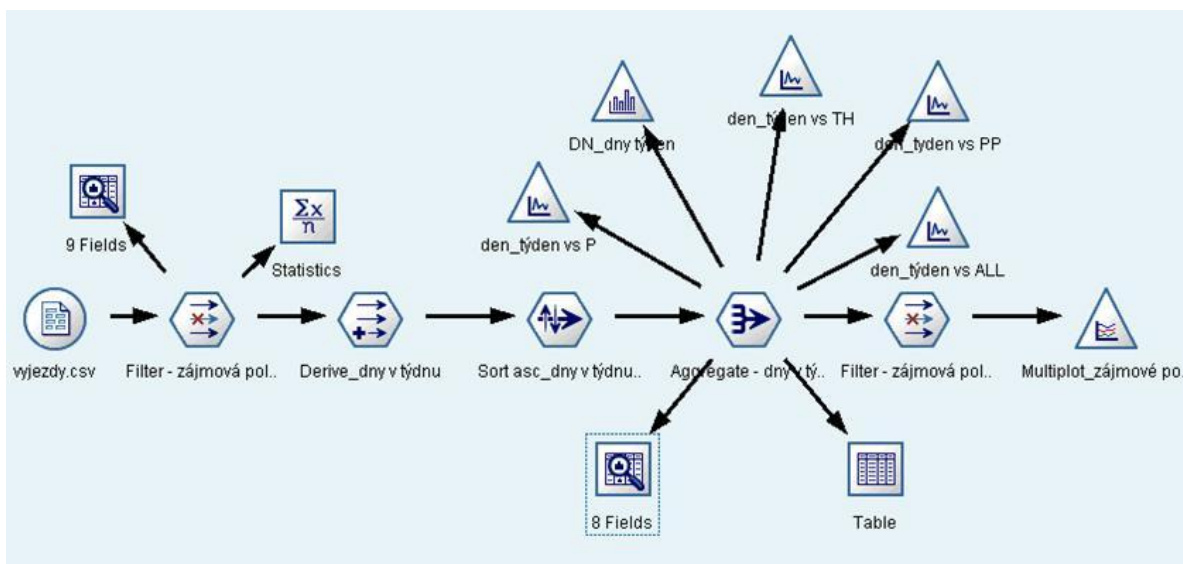
Soubor modeldat.csv je uložen v příloze 1 v adresáři \\model\_data.

V dalších krocích pro analýzu výjezdů HZS po dnech, týdnech a měsících jsem vytvořil příslušné streamy, které jsou uloženy v příloze 1 v adresáři \\ stream.

Jako příklad zde uvádím stream pro součet počtu výjezdů HZS v kalendářních dnech týdne (pondělí, úterý, středa, čtvrtek, pátek, sobota, neděle), který slouží pro hledání vztahů mezi jednotlivými dny a četnostmi výjezdů HZS k typům zásahu v kalendářních dnech týdne. Stream je na obrázku14.



**Obrázek 13** Nastavení klíčů pro sloučení agregovaných uzlů [zdroj vlastní]



Obrázek 14 Stream kalendářní dny v týdnu [zdroj vlastní]

### 3.4. MODELOVÁNÍ DAT

Etapu modelování rozdělím na části, v kterých se zaměřím na zkoumání dat v různých časových úsecích. A to v denních, týdenních a měsíčních úsecích. Cílem je analyzovat data, která budou agregována do časových úseků, tak abychom mohli mezi jednotlivými analýzami nalézt souvislosti v datech, které povedou k zisku dat s informační hodnotou.

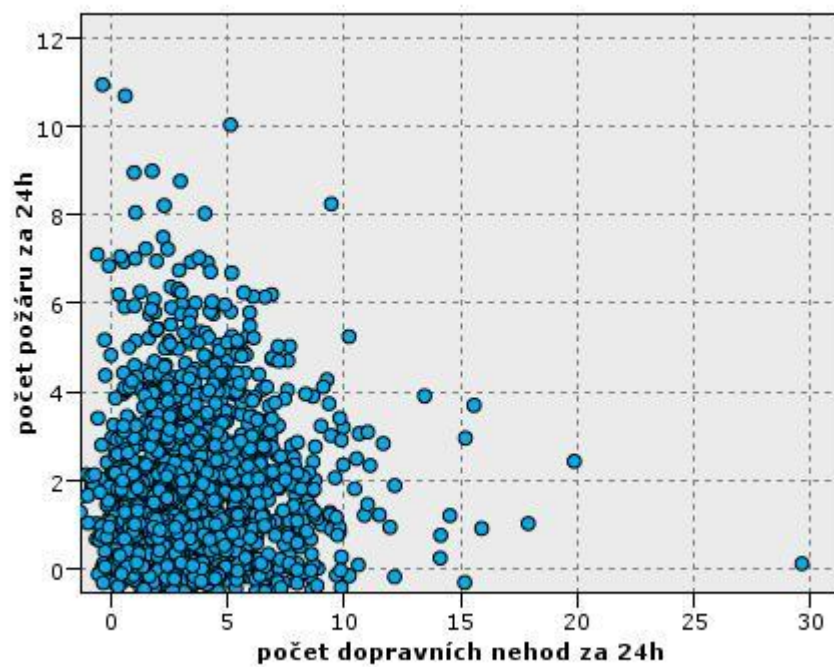
#### 3.4.1. ANALÝZA VÝJEZDŮ HZS PO DNECH

Analýzu počtu výjezdů HZS ze souboru vyjezdy.csv zahájím interpretací grafů zachycujících některé vztahy v bodových diagramech.

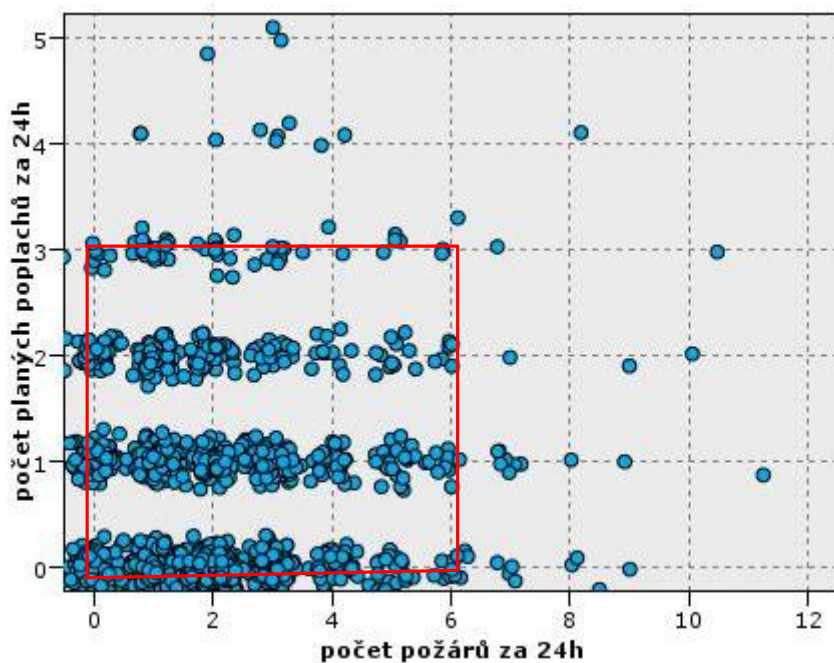
|   | Value | Proportion | %     | Count |
|---|-------|------------|-------|-------|
| 1 | 0     |            | 51,11 | 827   |
| 2 | 1     |            | 32,51 | 526   |
| 3 | 2     |            | 12,18 | 197   |
| 4 | 3     |            | 3,40  | 55    |
| 5 | 4     |            | 0,62  | 10    |
| 6 | 5     |            | 0,19  | 3     |

Obrázek 15 Histogram počtu planých poplachů během 24 hodin [zdroj vlastní]

Na obrázku 16 je znázorněn bodový diagram počtu výjezdů HZS k požárům ve vztahu k dopravním nehodám za 24 hodin. Mohlo by se zdát, že zde existuje závislost. Tomu je ale právě naopak, protože korelační koeficient  $\rho = 0,038$ .



**Obrázek 16** Bodový diagram absolutních četností za 24 hodin, výjezdů HZS k požárů vs. výjezdy HZS k dopravním nehodám [zdroj vlastní]



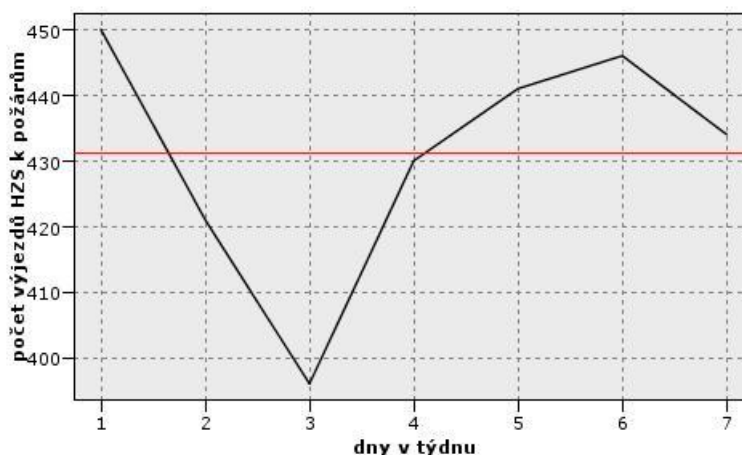
**Obrázek 17** Bodový diagram s vysvětlujícím čtvercem výskytu kombinací absolutních četností počtu výjezdů HZS k požáru vs. počtu planých poplachů [zdroj vlastní]

Podobným způsobem můžeme interpretovat výsledky bodového diagramu vztahu planých poplachů HZS a výjezdů HZS k požáru na obrázku 17. Daný obrázek ukazuje, že existuje velmi málo dní s vysokým počtem požárů a planých poplachů například s počtem 6-ti požárů a 3 planých poplachů. Většina výjezdů dle obrázku není typu planý poplach. Toto tvrzení potvrzuje histogram četností počtu planých poplachů za 24 hodin na obrázku 15 a histogram počtu výjezdů HZS k požáru na obrázku 18.

|    | Value | Proportion | %     | Count |
|----|-------|------------|-------|-------|
| 1  | 0     |            | 18,42 | 298   |
| 2  | 1     |            | 29,79 | 482   |
| 3  | 2     |            | 24,54 | 397   |
| 4  | 3     |            | 14,28 | 231   |
| 5  | 4     |            | 6,12  | 99    |
| 6  | 5     |            | 3,52  | 57    |
| 7  | 6     |            | 1,85  | 30    |
| 8  | 7     |            | 0,80  | 13    |
| 9  | 8     |            | 0,31  | 5     |
| 10 | 9     |            | 0,19  | 3     |
| 11 | 10    |            | 0,06  | 1     |
| 12 | 11    |            | 0,12  | 2     |

**Obrázek 18** Histogram počtu výjezdů HZS k požáru za 24 hodin [zdroj vlastní]

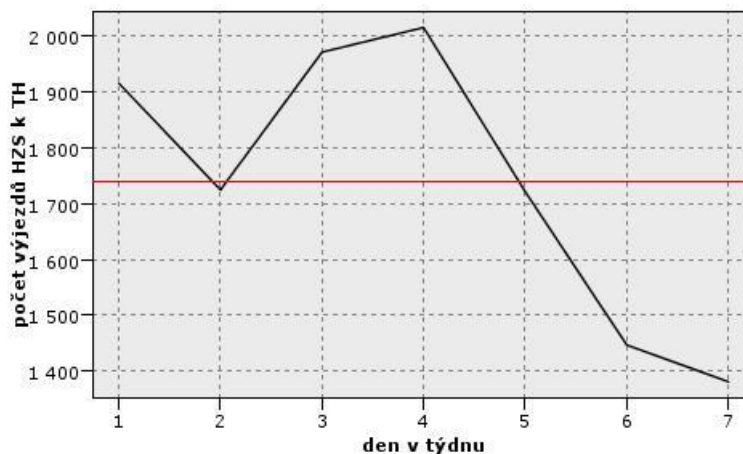
Na základě charakteristik popisné statistiky souboru vyjezdy.csv lze konstatovat, že průměrný počet výjezdů za 24 hodin k požáru je 1.845, k dopravní nehodě je 3.645, k technické havárii 7.527, k živelné pohromě 0,801 a celkový počet výjezdů je 15,078. Podle mediánu můžeme říci, počet všech výjezdů HZS Pardubice je 12 každý den.



**Obrázek 19** Počet výjezdů HZS k požárům v dny týdne [zdroj vlastní]

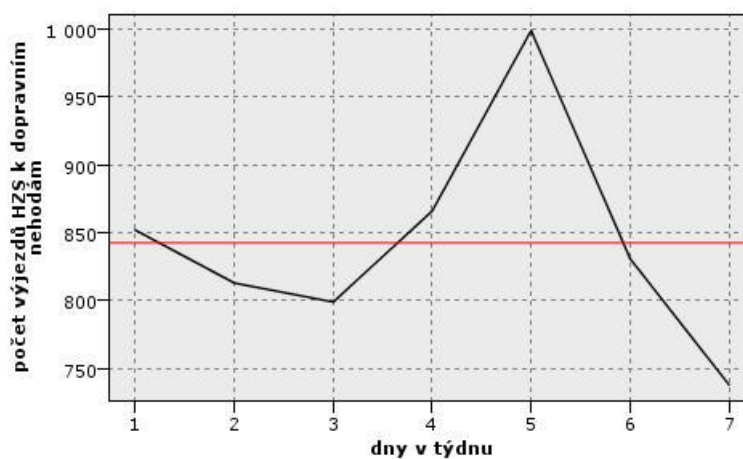


Na obrázku 19 vidíme, že nejmenší počet požárů lze očekávat ve středu, nejvyšší počet požárů v pondělí, což můžeme přičíst nepozornosti lidí, zahájení výroby na počátku pracovního týdne.



**Obrázek 20** Počet výjezdů HZS k technickým haváriím v dny týdne [zdroj vlastní]

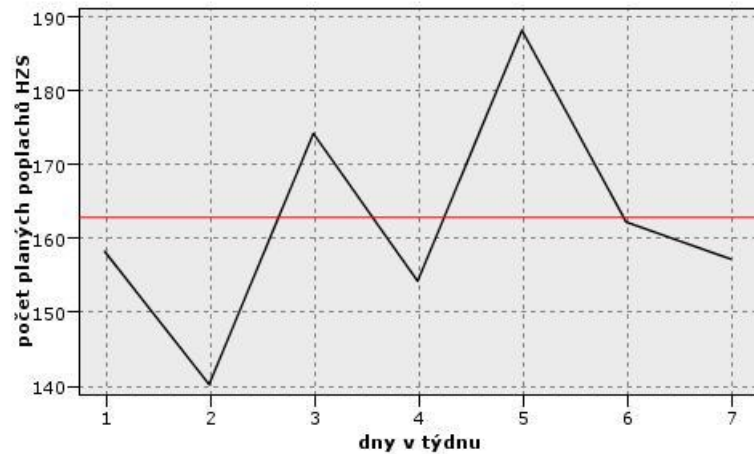
Z obrázku 20 plyne, že nadprůměrný počet technických havárií je v pondělí, středu a čtvrtku. S nástupem víkendu se počet výjezdů k technickým haváriím snižuje.



**Obrázek 21** Počet výjezdů HZS k dopravním nehodám v dny týdne [zdroj vlastní]

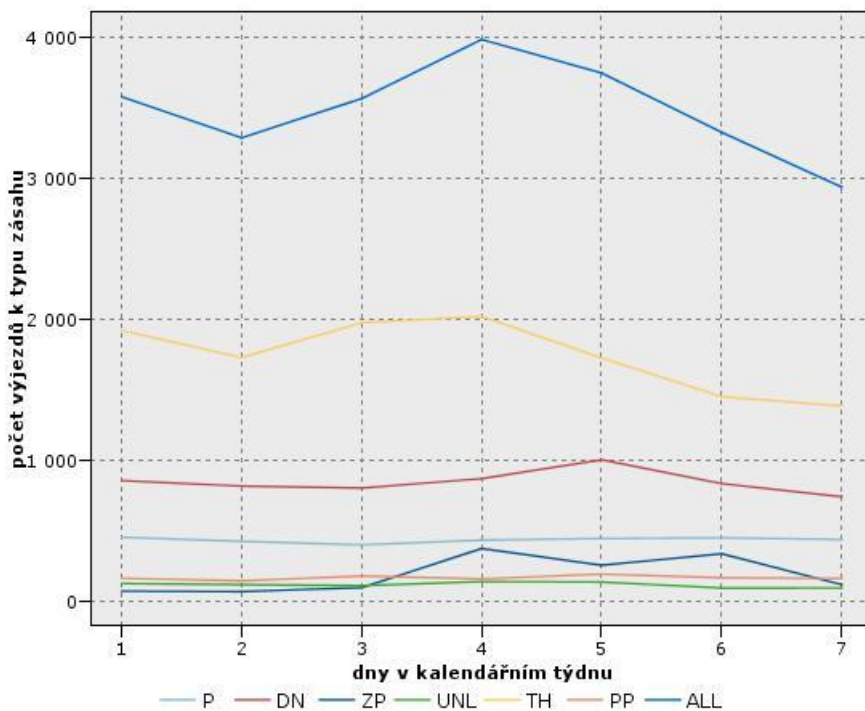
Na základě průběhu křivky na obrázku 21 vidíme, že počet dopravních nehod je statisticky nejvyšší v pátek, kdy HSZ registruje nejvíce výjezdů k dopravním nehodám. Nejméně výjezdů k DN je v neděli.





**Obrázek 22** Počet planých poplachů HZS v dny týdne [zdroj vlastní]

Počet planých poplachů dle interpretace obrázku 22 statisticky roste s probíhajícím pracovním týdnem. Nadprůměrné množství planých poplachů lze očekávat ve středu a pátek.



**Obrázek 23** Průběh počtu výjezdů pro jednotlivé typy zásahů v dny kalendářního týdne [zdroj vlastní]

Na obrázku 23 vidíme průběh počtu absolutních četností výjezdů k jednotlivým typům zásahů. Celkový počet výjezdů HZS závisí na počtu technických havárií a živelných pohrom. Což lze vidět také na korelačním koeficientu v tabulce na obrázku 24.

|                      |        |        |
|----------------------|--------|--------|
| ALL                  |        |        |
| Statistics           |        |        |
| Count                | 1618   |        |
| Mean                 | 15.078 |        |
| Min                  | 0      |        |
| Max                  | 293    |        |
| Median               | 12     |        |
| Pearson Correlations |        |        |
| SumaP                | 0.143  | Weak   |
| SumaDN               | 0.271  | Weak   |
| SumaZP               | 0.721  | Strong |
| SumaUN               | 0.102  | Weak   |
| SumaTH               | 0.723  | Strong |
| PP_81                | 0.251  | Weak   |

**Obrázek 24** Korelační koeficienty denních počtu výjezdů HZS [zdroj vlastní]

V tabulce 1 je relativní četnost počtu požárů v dnech týdne je konstantní. Nejvyšší relativní počet dopravních nehod je v pátek a to 16,9% z počtu DN v dnech v týdne. Počet živelných pohrom je dle analýzy denní četnosti příloha 8 v 94% nulová, ale zbývající počet v relativní četnosti ukazuje, že největší počet výjezdů HZS je ve čtvrtek, pátek a v sobotu. Počet výjezdů HZS k UNL je přibližně konstantní. V případě výjezdů HZS k technickým haváriím je nejvyšší relativní četnost ve středu a čtvrtek. Nejvytíženějším dnem HZS v týdnu je pátek.

**Tabulka 1** Relativní četnost počtu výjezdů HZS k typu zásahu v jednotlivých dnech týdne [zdroj vlastní]

| Dny v týdnu | P      | DN     | ZP     | UNL    | TH     | PP     | ALL    |
|-------------|--------|--------|--------|--------|--------|--------|--------|
| 1           | 14,9%  | 14,4%  | 5,2%   | 15,4%  | 15,7%  | 13,9%  | 14,6%  |
| 2           | 13,9%  | 13,8%  | 5,0%   | 14,4%  | 14,2%  | 12,4%  | 13,5%  |
| 3           | 13,1%  | 13,5%  | 7,2%   | 13,5%  | 16,2%  | 15,4%  | 14,6%  |
| 4           | 14,2%  | 14,7%  | 28,5%  | 17,2%  | 16,6%  | 13,6%  | 16,3%  |
| 5           | 14,6%  | 16,9%  | 19,4%  | 16,8%  | 14,1%  | 16,6%  | 15,3%  |
| 6           | 14,8%  | 14,1%  | 25,7%  | 11,3%  | 11,9%  | 14,3%  | 13,6%  |
| 7           | 14,4%  | 12,5%  | 8,9%   | 11,5%  | 11,3%  | 13,9%  | 12,0%  |
| Součet      | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

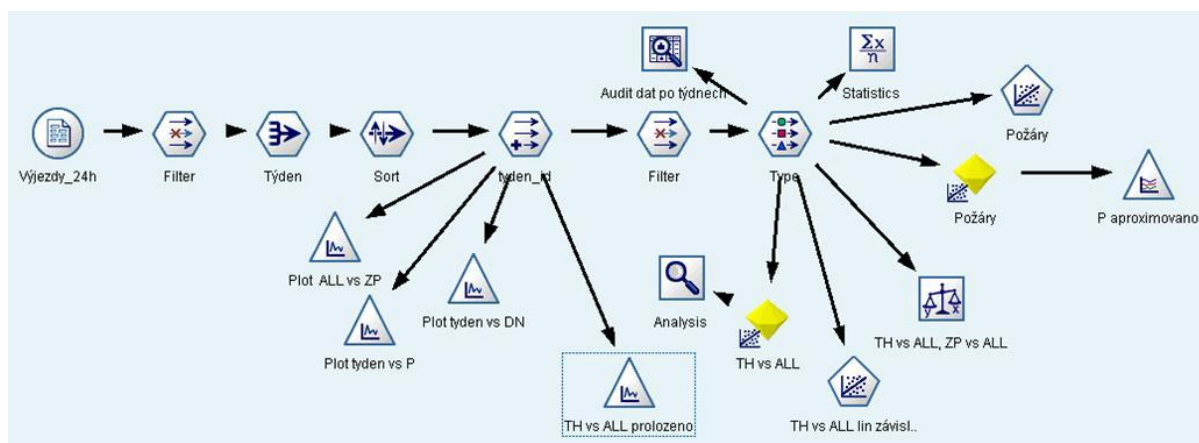
Vyjadřuje minima z podílů počtu výjezdů HZS k typu zásahu v kalendářní den týdne vzhledem k absolutnímu počtu výjezdů k typu zásahu v kalendářním týdnu

Vyjadřuje maxima z podílů počtu výjezdů HZS k typu zásahu v kalendářní den týdne vzhledem k absolutnímu počtu výjezdů k typu zásahu v kalendářním týdnu

Z důvodu lepší interpretace je označeno více hodnot

### 3.4.2. ANALÝZA VÝJEZDŮ HZS PO TÝDNECH

Data pro tuto DM analýzu byla předpřipravena pomocí streamu v Clementine jak vidíme obrázku 25. Během přípravy dat, bylo provedeno odfiltrování nepotřebných proměnných, agregace záznamů součtem do týdenních časových úseků, následné přetypování datového typu a provedení grafických výstupů, výpočtu korelačních koeficientů a vytvoření lineárního regresního modelu.



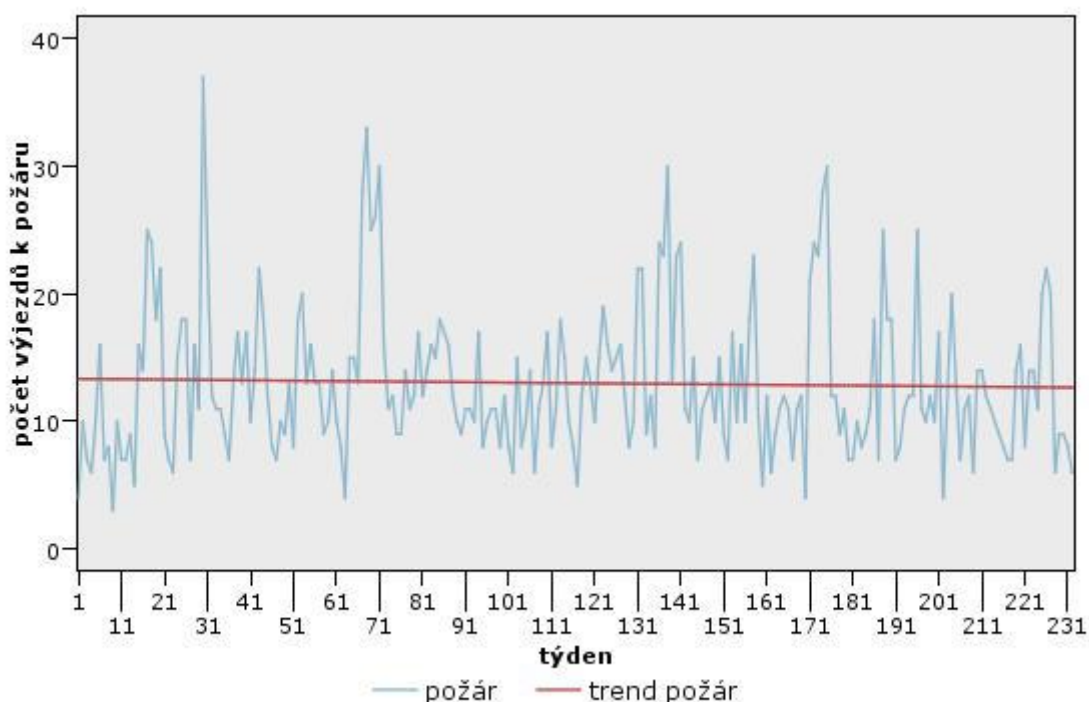
Obrázek 25 Stream pro data vyjezdy.csv, pole po týdnech [zdroj vlastní]

V tabulce na obrázku 26 vidím, že průměrný počet výjezdů za týden služby HZS k požáru je 13.009, k dopravní nehodě 25.422, k technické havárii 52.496, k živelné pohromě 5.586, k uniku nebezpečných chemických látek je 3.418 a celkový počet výjezdů HZS je 105.155. Mohu konstatovat, že technické havárie tvoří přibližně polovinu všech výjezdů HZS k zásahu.

| Field                 | Min | Max | Mean    | Median | Mode | Skewness | Valid |
|-----------------------|-----|-----|---------|--------|------|----------|-------|
| Požáry                | 3   | 37  | 13.009  | 12     | 11   | 1.183    | 232   |
| Dopravní nehody       | 6   | 58  | 25.422  | 24     | 24   | 0.881    | 232   |
| Živelné pohromy       | 0   | 333 | 5.586   | 0      | 0    | 8.122    | 232   |
| Unik neb. chem. látek | 0   | 10  | 3.418   | 3      | 2'   | 0.610    | 232   |
| Technické havárie     | 4   | 407 | 52.496  | 42     | 33   | 4.947    | 232   |
| Mimořádné události    | 0   | 4   | 0.177   | 0      | 0    | 4.280    | 232   |
| Plané poplachy        | 0   | 15  | 4.914   | 4      | 4    | 0.724    | 232   |
| Výjezdy HZS           | 15  | 474 | 105.155 | 90     | 85'  | 3.905    | 232   |

Obrázek 26 Tabulka z uzlu Data Audit pro výjezdy HSZ k zásahům v týdenních součtech [zdroj vlastní]

Podle průběhu absolutního počtu požárů za týden v časové řadě na obrázku 27 pozoruji řadu odlehlých hodnot. Při proložení lineárním trendem  $y = x \cdot (-0,002939) + 13,35$  jsem zjistil, že trend je mírně klesající.



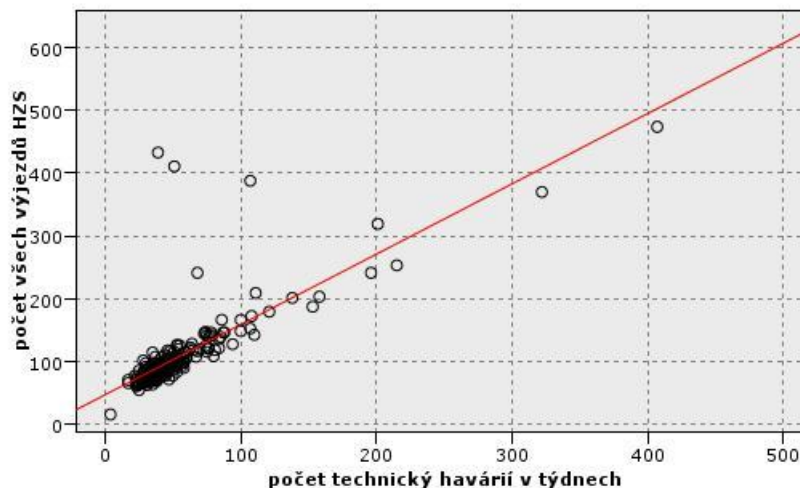
**Obrázek 27** Časová řada průběh počtu výjezdů HZS v týdenních součtech [zdroj vlastní]

Na základě informací o základní statistice na obrázku 26 jsem si všiml, že celkový počet výjezdů HZS je z cca 75% tvořen výjezdy k živelné pohromě a technickým haváriím.

| Field One         | Field Two   | Mean One* | Mean Two* | Correlation     | Mean Differ... | Importance           |
|-------------------|-------------|-----------|-----------|-----------------|----------------|----------------------|
| Živelné pohromy   | Výjezdy HZS | 5.586     | 105.155   | 0.659<br>Strong | -99.569        | 1.000<br>★ Important |
| Technické havárie | Výjezdy HZS | 52.496    | 105.155   | 0.773<br>Strong | -52.659        | 1.000<br>★ Important |

**Obrázek 28** Vyhledání závislostí polí v týdenních součtech [zdroj vlastní]

Na základě výpočtu korelačního koeficientu dle tabulky na obrázku 24, je vztah celkové výjezdy HZS vs. výjezdy k technické havárii (případ 1)  $\rho=0.773$  a celkové výjezdy HZS vs. výjezdy k živelné pohromě (případ 2)  $\rho=0.659$ . V případě 2 při výpočtu lineární regrese jsem zjistil, že koeficient determinace je  $R^2 = 0.435$ , to znamená, že lineární regrese vysvětluje vztah málo. V případě 1 je pro regresní funkci  $y = x \cdot 1.122 + 46.45$  koeficient determinace  $R^2 = 0.615$  a v tomto případě lineární závislost regresní funkcí je vysvětlena dostatečně. Daná lineární regres je uvedena na grafu obrázek 29.



**Obrázek 29** Lineární regrese – proložení regresní funkce  $y = x \cdot 1.122 + 46.45$  bodovým diagramem počtu TH a počtu ALL výjezdů HZS v týdenních součtech [zdroj vlastní]

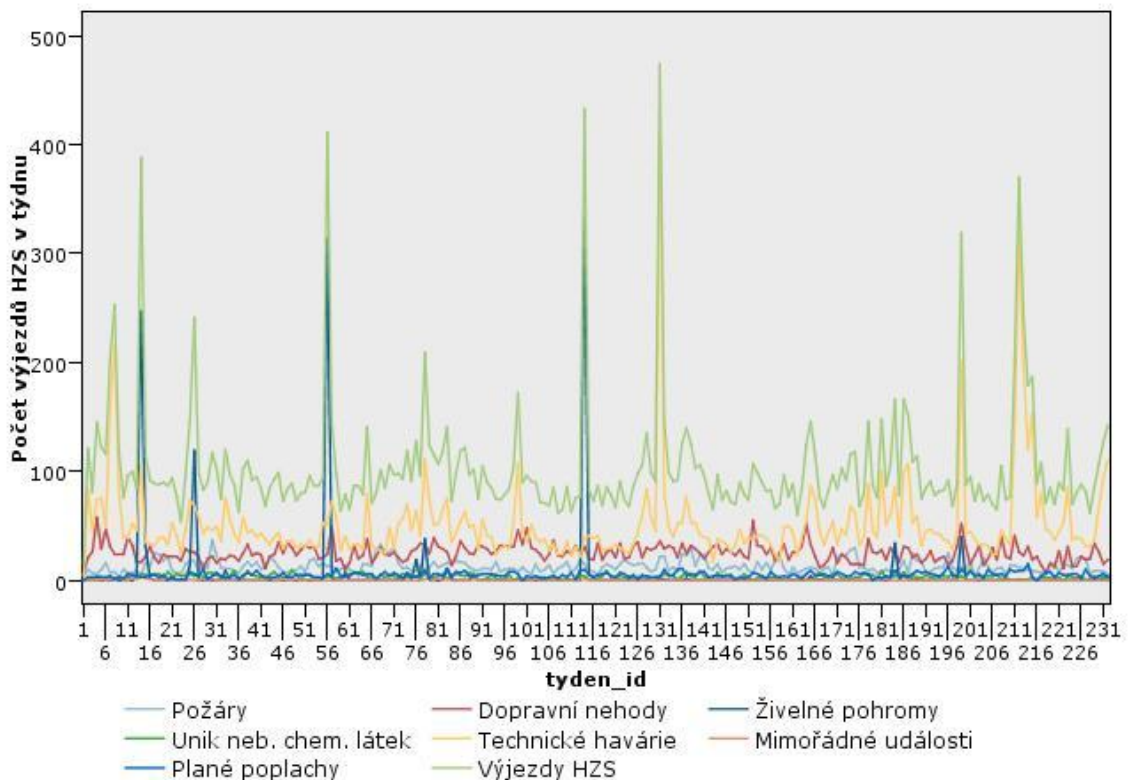
Na obrázku 30 vidím časové řady průběhu počtu výjezdů k typu zásahu a celkový počet všech výjezdů HZS. Extrémní hodnoty celkového počtu výjezdů HZS jsou zapříčiněny technickými haváriemi a živelnými pohromami (sněhové kalamity, porывy větru, přerušené komunikace, polomy, záplavy, sesuvy půdy a následky těchto událostí poničené vedení, plynovod, uniky látek atd.) .

### 3.4.3. ANALÝZA VÝJEZDŮ HZS PO MĚSÍCÍCH

Pro práci s daty ze souboru vyjezdy.csv bylo potřeba provést řadu úprav dat polí, což je znázorněno streamem v příloze 2.

**Obrázek 30** Náhled na data vyjezdy.csv po agregaci do měsíčních součtů [zdroj vlastní]

|    | P_Sum | DN_Sum | ZP_Sum | UNL_Sum | TH_Sum | PP_Sum | ALL_Sum | Mesic | Rok  | Record_Count |
|----|-------|--------|--------|---------|--------|--------|---------|-------|------|--------------|
| 1  | 40    | 152    | 1      | 11      | 294    | 13     | 512     | 01    | 2006 | 31           |
| 2  | 37    | 123    | 2      | 8       | 502    | 11     | 689     | 02    | 2006 | 28           |
| 3  | 31    | 120    | 209    | 15      | 260    | 19     | 660     | 03    | 2006 | 31           |
| 4  | 80    | 91     | 76     | 23      | 152    | 16     | 441     | 04    | 2006 | 30           |
| 5  | 58    | 102    | 5      | 21      | 176    | 10     | 374     | 05    | 2006 | 31           |
| 6  | 62    | 102    | 133    | 25      | 245    | 19     | 587     | 06    | 2006 | 30           |
| 7  | 89    | 77     | 3      | 23      | 210    | 22     | 427     | 07    | 2006 | 31           |
| 8  | 44    | 93     | 9      | 18      | 223    | 15     | 402     | 08    | 2006 | 31           |
| 9  | 63    | 108    | 0      | 21      | 200    | 26     | 418     | 09    | 2006 | 30           |
| 10 | 67    | 93     | 0      | 21      | 166    | 13     | 361     | 10    | 2006 | 31           |
| 11 | 42    | 132    | 0      | 20      | 152    | 20     | 368     | 11    | 2006 | 30           |
| 12 | 61    | 130    | 0      | 15      | 149    | 20     | 376     | 12    | 2006 | 31           |
| 13 | 60    | 119    | 315    | 11      | 242    | 16     | 781     | 01    | 2007 | 31           |
| 14 | 46    | 74     | 7      | 19      | 136    | 25     | 307     | 02    | 2007 | 28           |
| 15 | 43    | 112    | 0      | 24      | 186    | 27     | 408     | 03    | 2007 | 31           |



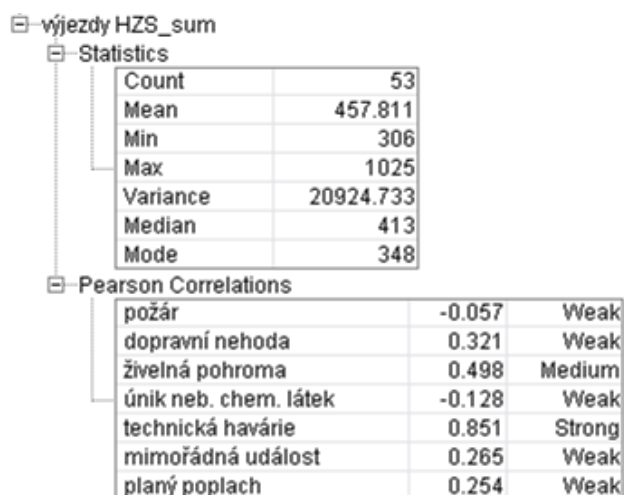
**Obrázek 31** Časové řady počtů výjezdů HZS jednotlivých typů zásahu [zdroj vlastní]

Záznamy v této časové řadě jsou zaznamenány jako součty kalendářních dnů daného kalendářního měsíce jak vidíme na obrázku 30, tedy s měsíčním časovým intervalem. Protože každý měsíc v roce má jiný počet dní provedl jsem přepočítání hodnot záznamů jednotlivých měsíců přes uzel *Derive* pro vyrovnání řady s měsíční periodou záznamů. Použitý vzorec vystihuje úpravu pole *ALL\_Sum*:

```
round(if Record_Count=31 then (ALL_Sum/31)*30 elseif Record_Count=29 then
(ALL_Sum/29)*30 elseif Record_Count=28 then (ALL_Sum/58)*30 else ALL_Sum endif)
```

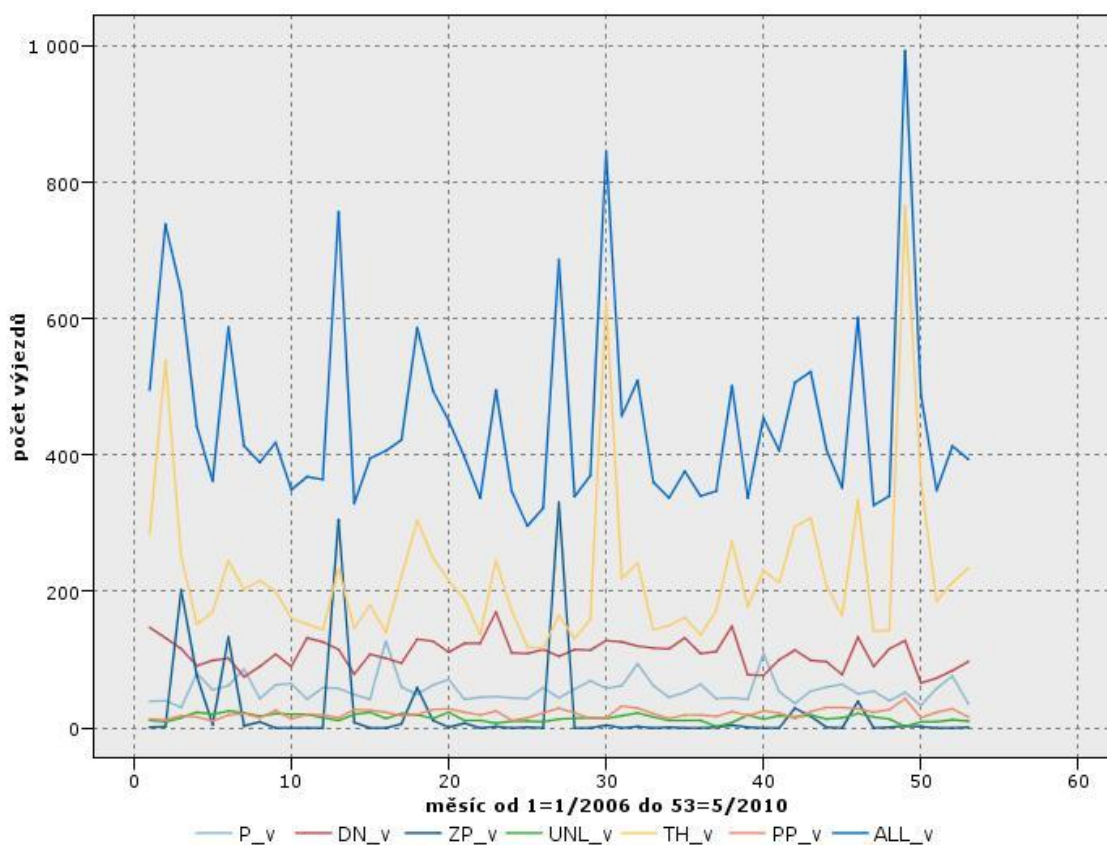
Dle metodiky vyrovnání časových řad v měsíčních intervalech s proměnným počtem kalendářních dnů je korelační koeficient pro vztah výjezdy HZS\_sum vs. Technická havárie dokonce  $\rho=0,859$  na obrázku 32. Lineární závislost na obrázku 29 je tímto druhým výpočtem potvrzena.





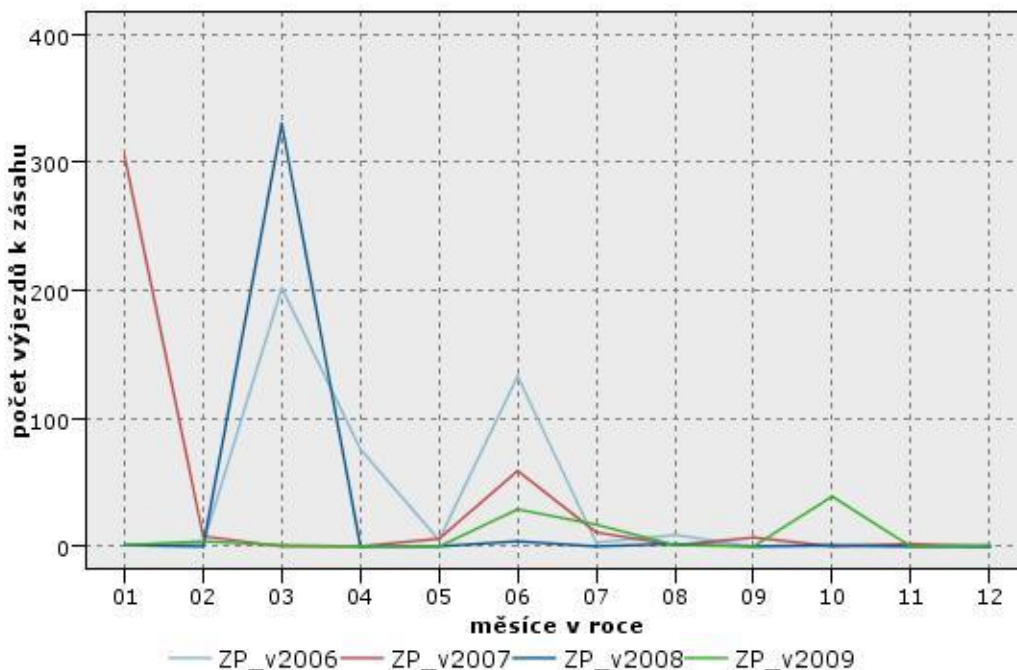
**Obrázek 32** Popisná statistika pro pole celkové výjezdy HZS [zdroj vlastní]

Průběhu počtu výjezdů HSZ ve měsících 1/2006 - 5/2010 je znázorněn v grafech časových řad na obrázku 33. Zatížení jednotky HZS počtem výjezdů k zásahu v jednotlivých měsících je velmi závislé na lokálních povětrnostních podmínkách.



**Obrázek 33** Přehled vyrovnaných průběhu počtu výjezdů HSZ ve měsících 1/2006 - 5/2010 [zdroj vlastní]

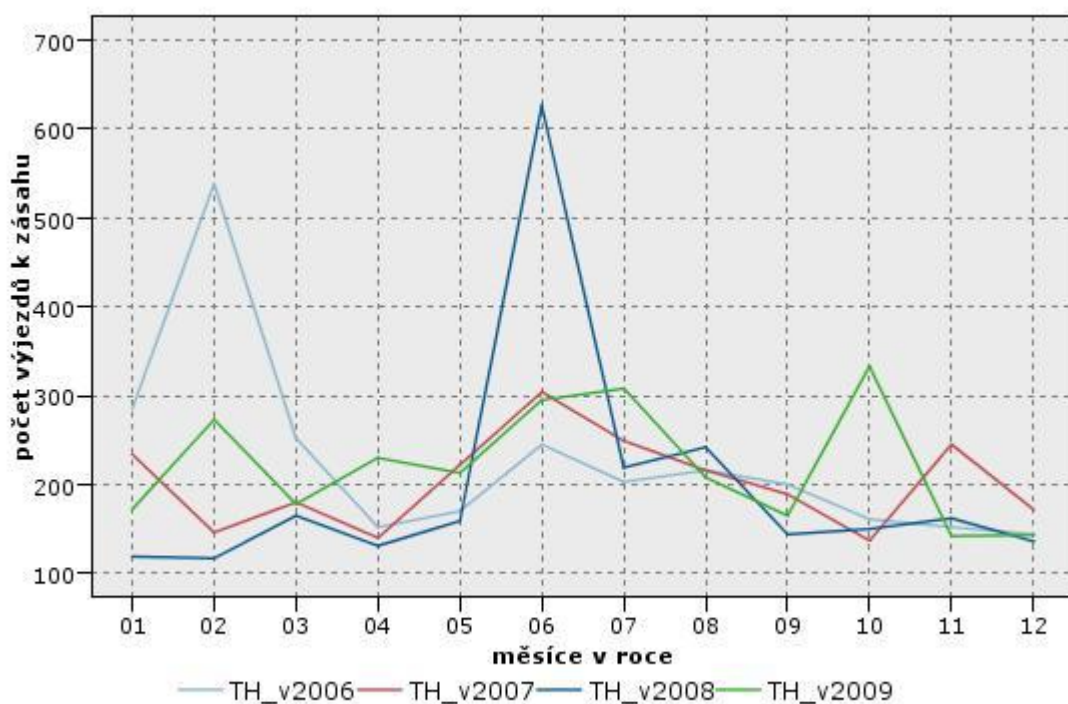
Příklad: z obrázku 33, 34 a 35 vyčtu že v měřeném měsíci se záznamy v pořadí č.48 je extrémní nárůst technických havárií což bylo důsledkem sněhové kalamity v daném měsíci leden 2010. Záznam č.27 měsíce březen 2008 je důsledkem Vichřice Emma ze dne 1.3.2009, která postihla celé území ČR. Záznam č.30 měsíce červen 2008 bylo nasazení hasičů k výjezdům technických havárií velmi extrémní z důvodu silná bouře, která řádila ve středu 25. června večer v okrese Pardubice.



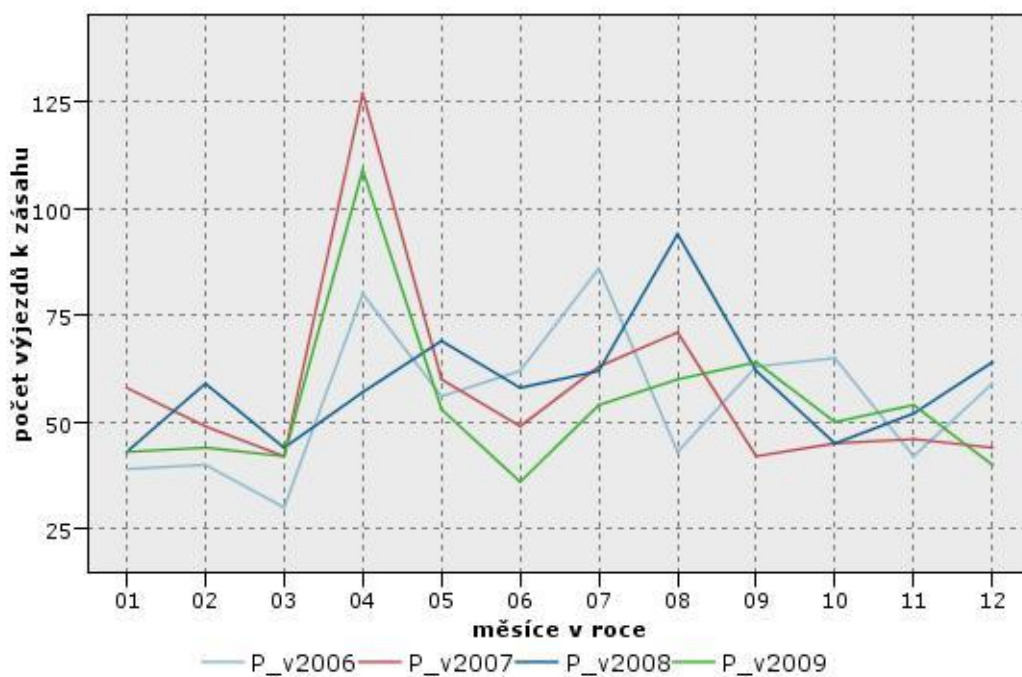
**Obrázek 34** Počet živelných pohrom v letech 2006-2009 podle kalendářních měsíců [zdroj vlastní]

Zajímalo mne, jestli počet výjezdů HZS k požárům je ovlivněn ročním obdobím nebo kalendářním měsícem. Na obrázku 36 je vidět že počet výjezdů HZS k požáru je v průběhu let 2006 – 2009 podobný. Nejvíce požárů je zaznamenáno v měsíci duben (jarní měsíce) a v měsících astronomického léta.





**Obrázek 35** Počet technických havárií v letech 2006 – 2009 podle kalendářního měsíce [zdroj vlastní]



**Obrázek 36** Porovnání počtu výjezdů HZS k požáru v kalendářních měsících let 2006-2009 [zdroj vlastní]

### 3.4.4. ANALÝZA VÝJEZDŮ HZS A SPOLU ZASAHUJÍCÍCH JEDNOTEK

Vstupním souborem dat je soubor modeldat.csv, který jsem vytvořil pomocí streamu uvedeného v příloze 7. Pro analýzu byla použita metoda lineární regrese a hledání závislostí pomocí korelačního koeficientu.

Na základě hodnot korelačního koeficientu spolupracujících jednotek z HZS při požárech v tabulce 3, do modelu lineární regrese vstupují pole P\_PČR a P\_OBP. Hodnota korelačního koeficientu pro zasahující jednotky HSZ a PČR při požáru je  $\rho_{DN\_PČR} = 0.965$  a pro zasahující jednotky HZS a obecní policie při požáru je  $\rho_{DN\_OBP} = 0.713$ .

**Tabulka 2** Korelační koeficienty P\_HZS [zdroj vlastní]

| P_HZS vs . |       |        |
|------------|-------|--------|
| P_OST      | 0.287 | Weak   |
| P_PČR      | 0.965 | Strong |
| P_ZS       | 0.089 | Weak   |
| P_OBP      | 0.713 | Strong |
| P_ENRG     | 0.278 | Weak   |
| P_PM       | 0.255 | Weak   |

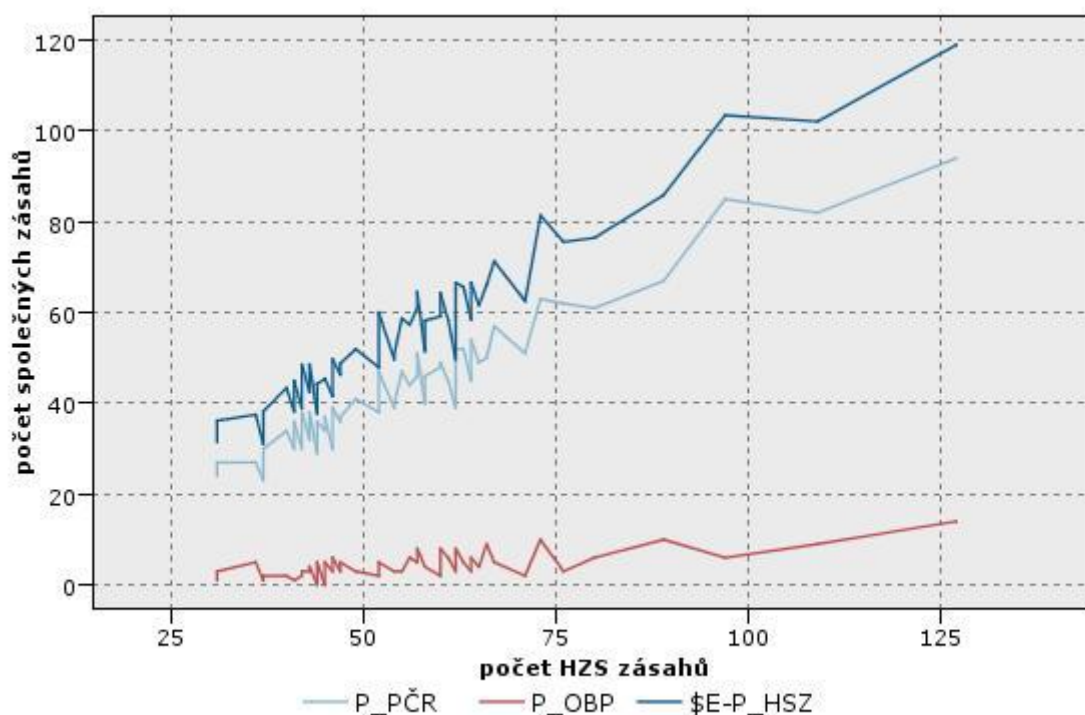
Po vložení do modelu lineární regrese dostáváme výsledky uvedené v tabulce 4. Lineární regresní model je vysvětluje lineární závislost s koeficientem determinace  $R^2 = 0.937$ .

**Tabulka 3** Lineární regresní model pro spolu zasahující jednotky u požáru [zdroj vlastní]

| Model                                   | R       | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---------|----------|-------------------|----------------------------|
| 1                                       | .965(a) | .931     | .930              | 4.88527                    |
| 2                                       | .968(b) | .937     | .934              | 4.73046                    |
| a. Predictors: (Constant), P_PČR        |         |          |                   |                            |
| b. Predictors: (Constant), P_PČR, P_OBP |         |          |                   |                            |

| Coefficients(a)              |            |                             |            |                           |        |      |
|------------------------------|------------|-----------------------------|------------|---------------------------|--------|------|
| Model                        |            | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|                              |            | B                           | Std. Error | Beta                      |        |      |
| 1                            | (Constant) | 2.847                       | 2.163      |                           | 1.317  | .194 |
|                              | P_PČR      | 1.213                       | .046       | .965                      | 26.275 | .000 |
| 2                            | (Constant) | 3.803                       | 2.143      |                           | 1.774  | .082 |
|                              | P_PČR      | 1.126                       | .061       | .895                      | 18.396 | .000 |
|                              | P_OBP      | .670                        | .319       | .102                      | 2.096  | .041 |
| a. Dependent Variable: P_HSZ |            |                             |            |                           |        |      |

Lineární regresní funkce má tvar  $P_{HZZ} = 1.126 \cdot P_{P\check{C}R} + 0.670 \cdot P_{OBP} + 3.803$ . Na obrázku 37 je vidět průběh lineární regrese, včetně průběhu stupních proměnných.

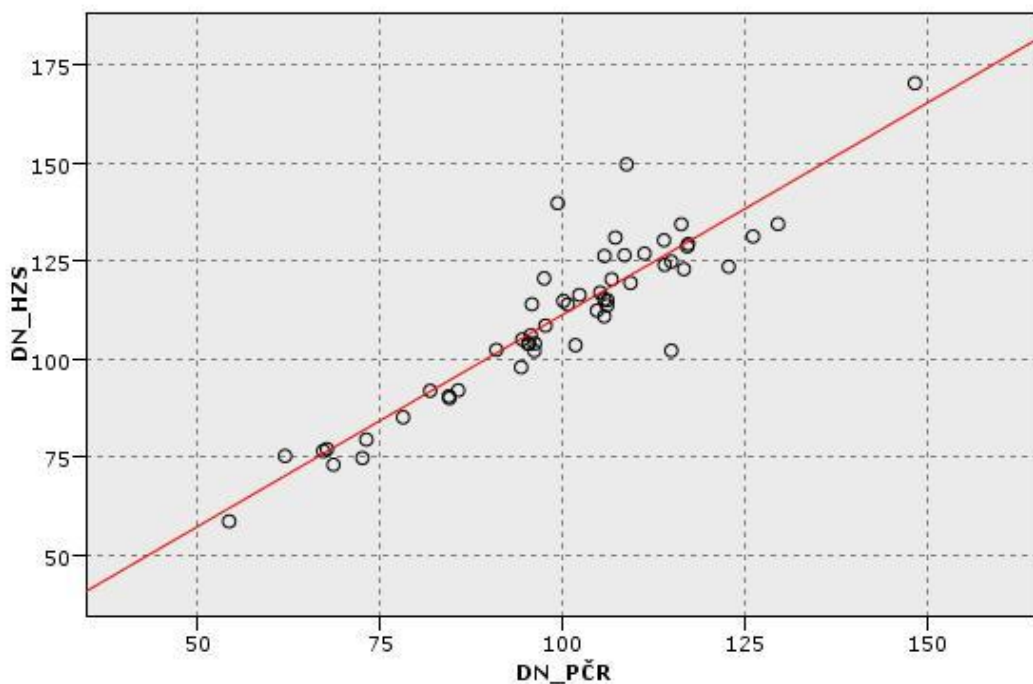


**Obrázek 37** Lineární regresní model funkce pro vztah zasahujících jednotek u požáru [zdroj vlastní]

Dalším lineárním regresním modelem je model na obrázku 38, ukazuje závislost počtu výjezdů HZZ a počtu zásahů policie České republiky u dopravních nehod. Zde je závislost vyjádřena korelačním koeficientem  $\rho_{DN_{OBP}} = 0.919$ , který je vidět v tabulce 5. Model  $DN_{HZZ} = 1.084 \cdot DN_{P\check{C}R} + 2.974$  je regresní lineární funkce s koeficientem determinace  $R^2 = 0.844$ .

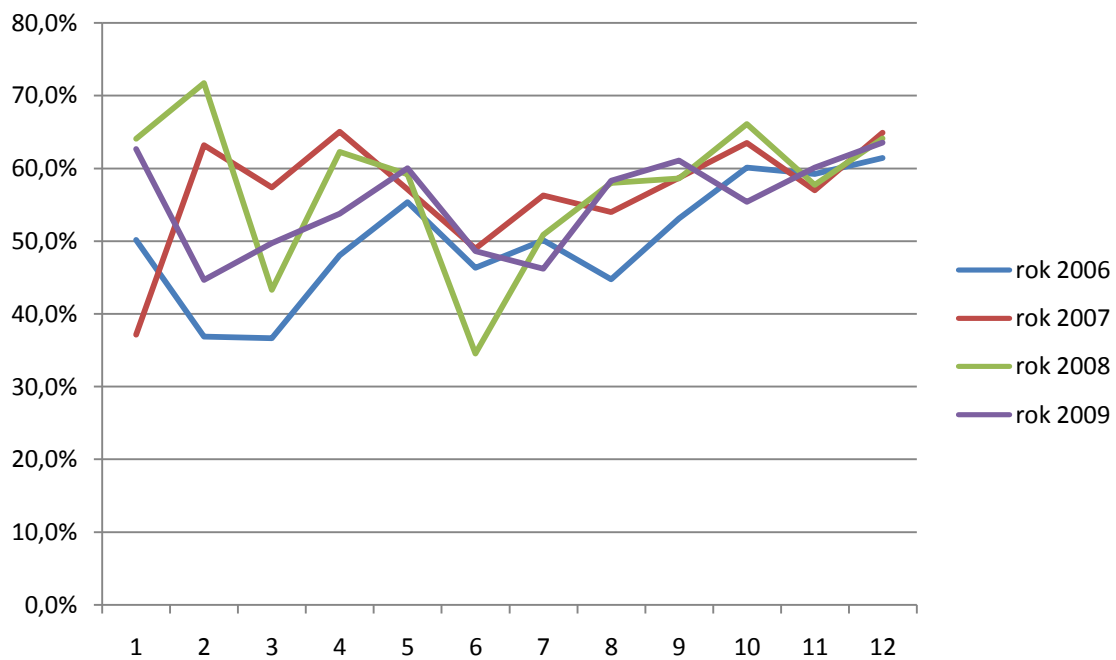
**Tabulka 4** Korelační koeficienty DN\_HZZ [zdroj vlastní]

| DN_HZZ vs . |        |        |
|-------------|--------|--------|
| DN_OST      | -0.123 | Weak   |
| DN_PČR      | 0.919  | Strong |
| DN_ZS       | 0.389  | Medium |
| DN_OBP      | 0.380  | Medium |
| DN_ENRG     | 0.183  | Weak   |
| DN_PM       | 0.443  | Medium |



**Obrázek 38** Lineární regresní model pro vztah zasahujících jednotek u dopravní nehody [zdroj vlastní]

Obrázek 39 ukazuje průběh počtu zásahů HZS za účasti PČR v procentech, PČR v průměru spolupracuje s HZS při zásahu v 55% výjezdů.



**Obrázek 39** Průběh procentuálního společného zásahu PČR a HZS [zdroj vlastní]

Na základě korelačních koeficientů uvedených v tabulkách 3, 5, 6 je nejčastěji spolu zasahující jednotkou dle druhu výjezdu HZS k zásahu právě PČR. Při specifických událostech, jako je živelná pohroma ZP, HZS spolupracuje mimo PČR i ze složkami OBP, ZS a ENERG.

**Tabulka 5** Korelační koeficienty [zdroj vlastní]

**P\_HZS vs.**

|        |       |        |
|--------|-------|--------|
| P_OST  | 0.287 | Weak   |
| P_PČR  | 0.965 | Strong |
| P_ZS   | 0.089 | Weak   |
| P_OBP  | 0.713 | Strong |
| P_ENRG | 0.278 | Weak   |
| P_PM   | 0.255 | Weak   |

**UNL\_HZS vs.**

|          |       |        |
|----------|-------|--------|
| UNL_OST  | 0.151 | Weak   |
| UNL_PČR  | 0.824 | Strong |
| UNL_ZS   | 0.041 | Weak   |
| UNL_OBP  | 0.376 | Medium |
| UNL_ENRG | 0.110 | Weak   |
| UNL_PM   | 0.381 | Medium |

**ZP\_HZS vs.**

|         |       |        |
|---------|-------|--------|
| ZP_OST  | 0.228 | Weak   |
| ZP_PČR  | 0.938 | Strong |
| ZP_ZS   | 0.611 | Medium |
| ZP_OBP  | 0.925 | Strong |
| ZP_ENRG | 0.938 | Strong |
| ZP_PM   | 0.646 | Medium |

**PP\_HZS vs.**

|         |        |        |
|---------|--------|--------|
| PP_OST  | 0.180  | Weak   |
| PP_PČR  | 0.762  | Strong |
| PP_ZS   | -0.105 | Weak   |
| PP_OBP  | 0.167  | Weak   |
| PP_ENRG | 0.259  | Weak   |
| PP_PM   | 0.097  | Weak   |

**TH\_HZS vs.**

|          |       |        |
|----------|-------|--------|
| TH_OST   | 0.192 | Weak   |
| TH_PČR   | 0.896 | Strong |
| TH_ZS    | 0.187 | Weak   |
| TH_OBP   | 0.629 | Medium |
| TH_ENERG | 0.616 | Medium |
| TH_PM    | 0.542 | Medium |

**ALL\_HZS vs.**

|           |       |        |
|-----------|-------|--------|
| ALL_OST   | 0.083 | Weak   |
| ALL_PČR   | 0.808 | Strong |
| ALL_ZS    | 0.011 | Weak   |
| ALL_OBP   | 0.564 | Medium |
| ALL_ENERG | 0.684 | Strong |
| ALL_PM    | 0.495 | Medium |

### 3.5. HODNOCENÍ VÝSLEDKŮ

#### 3.5.1. ANALÝZA VÝJEZDŮ HZS PO DNECH

V této části bylo zkoumáno 1618 denních záznamů. Na základě popisné statistiky a histogramů rozložení četnosti jsem získal základní představu o rozdělení četností počtu výjezdů pro jednotlivé typy zásahů HZS. Zkoumal jsem procentuální zastoupení výjezdů dle typu zásahů HZS v den s průměrnými hodnotami. Bylo zjištěno, že nejčastější činností HZS je výjezd k technické havárii. Pomocí procentuálního rozložení množství výjezdů typu zásahu HZS v průběhu jednoho týdne po kalendářních dnech a pomocí grafů s absolutní četností počtu výjezdů jsem získal informace o dnech v týdnu se závěry: nejmenší četnost požárů je ve středu, nejvíce technických havárií je ve čtvrtek, největší počet dopravních nehod je v pátek, procentuálně nejvyšší šance živelné pohromy je ve středu a pátek, počet planých poplachů je

nejvyšší v pátek a nejmenší počet úniků nebezpečných chemických látek je v dny pracovního klidu. Pomocí výpočtu korelačního koeficientu bylo zjištěno, že celkový počet výjezdů HZS za jeden den je závislý na počtu technických havárií v daný den.

### **3.5.2. ANALÝZA VÝJEZDŮ HZS PO TÝDNECH**

Na základě nalezené závislosti celkového počtu výjezdů HZS na počtu technických havárií jsem provedl výpočet korelačního koeficientu v týdenních součtech výjezdů HZ. Provedl jsem výpočet lineární regrese a index determinace potvrdil existenci lineární závislosti. Relativní rozložení četností výjezdů HZS typu zásahu je stejné jako pro relativní rozložení četností pro jeden den. Pomocí lineárního trendu jsem proložil průběh počtu požárů v 232 zaznamenaných týdnech. Trend počtu výjezdů HZS k požáru je slabě klesající. Z grafické interpretace průběhu počtu všech typů zásahů po týdnech je zřejmé, že existují týdny s extrémním počtem výjezdu HZS, které jsou způsobeny většinou živelnými pohromami a technickými haváriemi.

### **3.5.3. ANALÝZA VÝJEZDŮ HZS PO MĚSÍCÍCH**

Na základě naměřených hodnot pozorování počtu výjezdů HZS k zásahům v jednotlivých měsících, jsem provedl očištění počtů výjezdů HZS k zásahům od vlivu různé délky kalendářního měsíce. V takto upravené časové řadě jsem zjistil, že průběh výjezdů HZS v jednotlivých měsících odráží přesně reálné životní události, které měly vliv na extrémní počet výjezdů HZS k živelným pohromám, technickým haváriím a požárům. Na základě porovnání průběhu počtu výjezdů HZS v jednotlivých letech jsem zjistil, že nejčastěji se živelné pohromy vyskytují v lednu, březnu a v červnu. Technické havárie, důsledkem živelných pohrom a extrémních povětrnostních vlivů jsou nejčastěji únoru, červnu a na přelomu září-říjen. Celkově lze říci, členové HZS Pardubice mají nejvíce výjezdů k požáru v dubnu a srpnu.

### **3.5.4. ANALÝZA VÝJEZDŮ HZS A SPOLU ZASAHOJÍCÍCH JEDNOTEK**

Vstupním souborem pro modelování je soubor modeldat.csv, který vznikl při přípravě dat. Zkoumání jsem podrobil vyhledáním závislostí mezi počtem výjezdů HZS typu zásahu vs. počet zásahů spolu zasahující jednotky u typu zásahu. Na základě korelačních koeficientů jsem vytvořil lineární regresní model pro vztah počtu výjezdů HZS v závislosti na spolu zasahujících jednotkách PČR a OBP u požáru. Tento model vysvětluje lineární závislost velmi přesně z důvodu koeficientu determinace  $R^2 = 0.937$ . Na základě modelu vztahu

zasahujících jednotek HZS a PČR u dopravní nehody lze tvrdit, že tyto jednotky spolupracují v 91% případů výjezdu HZS k dopravní nehodě. Spolupráce HZS s jednotkami při živelné pohromě, kde výjezd HZS je podle korelačních koeficientů podmíněn velmi silně s výjezdy PČR, OBP, ENERGA, a v menší míře ZS a PM, což je zdárným příkladem funkcionality integrovaného záchranného systému.

### **3.5.5. POSOUZENÍ PROCESU METODIKY CRISP-DM**

Reálně nejnáročnější částí procesu byla fáze přípravy dat, kde bylo nutné data načíst z formulářů, které byly generovány informačním systémem. Načtení dat do systému SPSS Clementine proběhlo v pořádku. Samotná extrakce dat do podoby pro modelování proběhla pomocí programovatelných uzlů v streamech v Clementine. Samotné modelování proběhlo pomocí uzlů pro analýzu a modelování.

V etapě modelování dat, nebylo cílem vytvořit komplexní analýzu daného data miningového projektu, ale ukázat metody a možnosti „dolování dat“ se vstupních souborů, které podléhají náročné přípravě dat před vstupem do modelování.

### **3.6. DOPORUČENÍ PRO PRAXI**

Na základě špatné zkušenosti s uživatelským přístupem do databáze HZS obsahující data o výjezdech HZS, doporučuji přepracovat možnosti uživatelského rozhraní, tak aby při potřebách výstupů pro analýzy manažerského rozhodování, byly vytvořeny nástroje pro efektivní vytváření výstupů pomocí SQL dotazů. Současný stav v podobě pevně daných XML dokumentů, které zobrazují pouze data ve formě kontingenčních tabulek pro období jednoho daného týdne nebo měsíce, jsou uživatelsky nedostačující.

## ZÁVĚR

Cílem bakalářské práce bylo získat znalosti z databáze, která v datových maticích obsahuje časové řady, pomocí metod data miningu. Data mining proběhl podle metodiky CRISP-DM. Byla použita popisná statistická analýza dat, hledal jsem závislosti mezi daty z důvodu vytvoření lineárního regresního modelu, a porovnával jsem průběhy grafů vzniklých z časových řad s cílem interpretace získaných informací pro praktické využití.

Pokud bych hodnotil užitečnost softwarového produktu Clementine a musel bych se vyjádřit jako laik, tak první kontakt s tímto software je plný neporozumění ve smyslu logiky skladby jednotlivých uzlů pro přípravu a modelování dat. Druhý pohled po strávených hodinách v prostředí Clementine je nabit pocitem logičnosti, funkcionality a výkonu. Software byl účelně pro zvolenou metodu data miningu CRISP-DM vytvořen a je velice silným prostředkem pro získávání znalostí z databází.

Samotná část přípravy dat pro modelování by obsahově stačila na samostatnou práci. V průběhu etapy modelování CRISP-DM jsem se musel vícekrát vracet do fáze přípravy dat tak abych objasnil a vyřešil vyvstalé problémy, jako například příprava dat při sloučení dvou souborů dat bez společné relační vazby v souvislosti se vstupem dat v jiných časových intervalech a strukturách. Tyto vyřešené problémy se staly významným mezníkem v provedeném data miningu.

V průběhu etapy modelování jsem se snažil současně i interpretovat „vydolané“ informace jednotlivých dílčích výsledků. Celkově jsem pak hodnocení výsledků provedl v kapitole 3.5. Protože v CRISP-DM ve většině případů je to zákazník a nikoliv analytik, kdo provádí kroky k využití výsledků analýzy, nechávám pohled na výsledky této práce částečně otevřený. Po provedení hlubšího zkoumání vstupních dat, by bylo možné provést komplexně statistickou a shlukovou analýzu. Tím by se otevřely nové možnosti dolování dat ze vstupních souborů, které metodika CRISP-DM nabízí.



## POUŽITÁ LITERATURA

- [1] ARTL, J., ARTLOV.M. Analýza ekonomických časových řad s příklady. 1. vyd. VŠE Praha, 2002, 148 s. ISBN 80-245-0304-7.
  
- [2] BERKA, P. Dobývání znalostí z databází. 1. vyd. Praha: Academia, 2003, 368. s. ISBN 80-200-1062-9.
  
- [3] KUBANOVÁ, J. Statistické metody pro ekonomickou a technickou praxi, 2.vyd., Bratislava: Stasis, 2004, 253 s., ISBN: 80-85659-37-9.
  
- [4] PETR, P. Data Mining Díl 1., Pardubice: Univerzita Pardubice, 2006, 144 s., ISBN 80-7194-886-1.
  
- [5] POŠÍK, P. Co je data mining?, Část 1. [on-line], [cit. 2010-06-24], dostupné z <<http://cyber.felk.cvut.cz/gerstner/teaching/zbd/DataMining1-hout.pdf>>.
  
- [6] ŘEZANKOVÁ, H. Interaktivní učebnice statistiky., [on-line], [cit. 2010-06-28], dostupné z <[http://iastat.vse.cz/zakl\\_stat\\_pojmy.html](http://iastat.vse.cz/zakl_stat_pojmy.html)>.
  
- [7] ŠŤASTNÝ, F. Zpracování experimentálních dat., [on-line], [cit. 2010-06-27], dostupné z <[http://amper.ped.muni.cz/jenik/nejistoty/html\\_tree/node13.html](http://amper.ped.muni.cz/jenik/nejistoty/html_tree/node13.html)>.

## SEZNAM POUŽITÝCH ZKRATEK

|                 |   |
|-----------------|---|
| <b>HZS</b>      | – Hasičský záchranný sbor                         |
| <b>DM</b>       | – Data mining                                     |
| <b>XML</b>      | – eXtensible Markup Language                      |
| <b>CSV</b>      | – Comma-separated values                          |
| <b>OLAP</b>     | – On-Line Analytical Processing                   |
| <b>SQL</b>      | – Structured query language                       |
| <b>QEB</b>      | – Query by example                                |
| <b>EIS</b>      | – Executive Information System                    |
| <b>SEMMA</b>    | – Sample, explore, modify, model, assess          |
| <b>IZS</b>      | – Integrovaný záchranný systém                    |
| <b>ENERG</b>    | – Pohotovostní služby energetických společností   |
| <b>PM</b>       | – Podniky, místní - obecní služby                 |
| <b>OST</b>      | – Ostatní jednotky                                |
| <b>PČR</b>      | – Policie ČR                                      |
| <b>ZS</b>       | – Zdravotní záchranná služba                      |
| <b>OBP</b>      | – Obecní policie                                  |
| <b>CRISP-DM</b> | – Cross Industry Standart Process for Data mining |

# SEZNAM OBRÁZKŮ

|   |    |
|---|----|
| <b>OBRÁZEK 1</b> METODIKA CRISP-DM .....  | 12 |
| <b>OBRÁZEK 2</b> SCHÉMA DATOVÉ MATICE .....   | 14 |
| <b>OBRÁZEK 3</b> KRÁTKODOBÁ ČASOVÁ ŘADA .....   | 17 |
| <b>OBRÁZEK 4</b> METODA NEJMENŠÍCH ČTVERCŮ .....  | 19 |
| <b>OBRÁZEK 5</b> NÁHLED OBSAHU SOUBORU SPOLZASH.CSV .....   | 21 |
| <b>OBRÁZEK 6</b> NÁHLED OBSAHU SOUBORU VYJEZDY.CSV .....  | 22 |
| <b>OBRÁZEK 7</b> SCHÉMA ZÍSKANÝCH DAT BEZ VAZEB.....  | 23 |
| <b>OBRÁZEK 8</b> POPISNÁ STATISTIKA VSTUPNÍHO SOUBORU SPOLZASH.CSV .....  | 24 |
| <b>OBRÁZEK 9</b> POPISNÁ STATISTIKA VSTUPNÍHO SOUBORU VYJEZDY.CSV .....   | 25 |
| <b>OBRÁZEK 10</b> PROCENTUÁLNÍ ROZDĚLENÍ TYPU DENNÍCH ZÁSAHŮ HZS .....  | 25 |
| <b>OBRÁZEK 11</b> ROZDĚLENÍ ČETNOSTÍ DOPRAVNÍCH NEHOD ZA 24 HODIN .....   | 26 |
| <b>OBRÁZEK 12</b> SEZNAM UPRAVENÝCH VSTUPNÍCH UZLŮ DO UZLU MERGE .....  | 26 |
| <b>OBRÁZEK 13</b> NASTAVENÍ KLÍČŮ PRO SLOUČENÍ AGREGOVANÝCH UZLŮ .....  | 28 |
| <b>OBRÁZEK 14</b> STREAM KALENDÁŘNÍ DNY V TÝDNU.....  | 29 |
| <b>OBRÁZEK 15</b> HISTOGRAM POČTU PLANÝCH POPLACHŮ BĚHEM 24 HODIN.....  | 29 |
| <b>OBRÁZEK 16</b> BODOVÝ DIAGRAM ABSOLUTNÍCH ČETNOSTÍ ZA 24 HODIN, VÝJEZDŮ HZS K POŽÁRŮ VS. VÝJEZDY<br>HZS K DOPRAVNÍM NEHODÁM .....                                    | 30 |
| <b>OBRÁZEK 17</b> BODOVÝ DIAGRAM S VYSVĚTLUJÍCÍM ČTVERCEM VÝSKYTU KOMBINACÍ ABSOLUTNÍCH ČETNOSTÍ<br>POČTU VÝJEZDŮ HZS K POŽÁRU VS. POČTU PLANÝCH POPLACHŮ.....          | 30 |
| <b>OBRÁZEK 18</b> HISTOGRAM POČTU VÝJEZDŮ HZS K POŽÁRU ZA 24 HODIN .....  | 31 |
| <b>OBRÁZEK 19</b> POČET VÝJEZDŮ HZS K POŽÁRŮM V DNY TÝDNE .....   | 31 |
| <b>OBRÁZEK 20</b> POČET VÝJEZDŮ HZS K TECHNICKÝM HAVÁRIÍM V DNY TÝDNE.....  | 32 |
| <b>OBRÁZEK 21</b> POČET VÝJEZDŮ HZS K DOPRAVNÍM NEHODÁM V DNY TÝDNE.....  | 32 |
| <b>OBRÁZEK 22</b> POČET PLANÝCH POPLACHŮ HZS V DNY TÝDNE .....  | 33 |
| <b>OBRÁZEK 23</b> PRŮBĚH POČTU VÝJEZDŮ PRO JEDNOTLIVÉ TYPY ZÁSAHŮ V DNY KALENDÁŘNÍHO TÝDNE .....  | 33 |
| <b>OBRÁZEK 24</b> KORELAČNÍ KOEFICIENTY DENNÍCH POČTU VÝJEZDŮ HZS.....  | 34 |
| <b>OBRÁZEK 25</b> STREAM PRO DATA VYJEZDY.CSV, POLE PO TÝDNECH .....  | 35 |
| <b>OBRÁZEK 26</b> TABULKA Z UZLU DATA AUDIT PRO VÝJEZDY HSZ K ZÁSAHŮM V TÝDENNÍCH SOUČTECH .....  | 35 |
| <b>OBRÁZEK 27</b> ČASOVÁ ŘADA PRŮBĚH POČTU VÝJEZDŮ HZS V TÝDENNÍCH SOUČTECH .....   | 36 |
| <b>OBRÁZEK 28</b> VYHLEDÁNÍ ZÁVISLOSTÍ POLÍ V TÝDENNÍCH SOUČTECH .....  | 36 |
| <b>OBRÁZEK 29</b> LINEÁRNÍ REGRESE – PROLOŽENÍ REGRESNÍ FUNKCE $Y = X * 1.122 + 46.45$ BODOVÝM DIAGRAMEM<br>POČTU TH A POČTU ALL VÝJEZDŮ HZS V TÝDENNÍCH SOUČTECH ..... | 37 |
| <b>OBRÁZEK 30</b> NÁHLED NA DATA VYJEZDY.CSV PO AGREGACI DO MĚSÍČNÍCH SOUČTŮ.....   | 37 |
| <b>OBRÁZEK 31</b> ČASOVÉ ŘADY POČTŮ VÝJEZDŮ HZS JEDNOTLIVÝCH TYPŮ ZÁSAHU .....  | 38 |
| <b>OBRÁZEK 32</b> POPISNÁ STATISTIKA PRO POLE CELKOVÉ VÝJEZDY HZS .....   | 39 |
| <b>OBRÁZEK 33</b> PŘEHLED VYROVNANÝCH PRŮBĚHU POČTU VÝJEZDŮ HSZ VE MĚSÍCÍCH 1/2006 - 5/2010.....  | 39 |
| <b>OBRÁZEK 34</b> POČET ŽIVELNÝCH POHROM V LETECH 2006-2009 PODLE KALENDÁŘNÍCH MĚSÍCŮ.....  | 40 |

|  |    |
|--|----|
| <b>OBRÁZEK 35</b> POČET TECHNICKÝCH HAVÁRIÍ V LETECH 2006 – 2009 PODLE KALENDÁŘNÍHO MĚSÍCE .....   | 41 |
| <b>OBRÁZEK 36</b> POROVNÁNÍ POČTU VÝJEZDŮ HZS K POŽÁRU V KALENDÁŘNÍCH MĚSÍCÍCH LET 2006-2009 ..... | 41 |
| <b>OBRÁZEK 37</b> LINEÁRNÍ REGRESNÍ MODEL FUNKCE PRO VZTAH ZASAHUJÍCÍCH JEDNOTEK U POŽÁRU .....    | 43 |
| <b>OBRÁZEK 38</b> LINEÁRNÍ REGRESNÍ MODEL PRO VZTAH ZASAHUJÍCÍCH JEDNOTEK U DOPRAVNÍ NEHODY .....  | 44 |
| <b>OBRÁZEK 39</b> PRŮBĚH PROCENTUÁLNÍHO SPOLEČNÉHO ZÁSAHU PČR A HZS .....                          | 44 |

## SEZNAM TABULEK

|   |    |
|---|----|
| <b>TABULKA 1</b> RELATIVNÍ ČETNOST POČTU VÝJEZDŮ HZS K TYPU ZÁSAHU V JEDNOTLIVÝCH DNECH TÝDNE ..... | 34 |
| <b>TABULKA 2</b> KORELAČNÍ KOEFICIENTY P_HZS .....  | 42 |
| <b>TABULKA 3</b> LINEÁRNÍ REGRESNÍ MODEL PRO SPOLU ZASAHUJÍCÍ JEDNOTKY U POŽÁRU .....               | 42 |
| <b>TABULKA 4</b> KORELAČNÍ KOEFICIENTY DN_HZS .....   | 43 |
| <b>TABULKA 5</b> KORELAČNÍ KOEFICIENTY .....  | 45 |

## SEZNAM PŘÍLOH

|  |    |
|--|----|
| <b>PŘÍLOHA 1</b> ELEKTRONICKÁ PŘÍLOHA.....   | 53 |
| <b>PŘÍLOHA 2</b> VSTUPNÍ DATA PRO TABULKY POČTU DENNÍCH ZÁSAHŮ HZS ČÁST 1/2 .....  | 54 |
| <b>PŘÍLOHA 3</b> VSTUPNÍ DATA PRO TABULKY POČTU DENNÍCH ZÁSAHŮ HZS ČÁST 2/2 .....  | 55 |
| <b>PŘÍLOHA 4</b> VSTUPNÍ DATA PRO TABULKU SPOLU ZASAHUJÍCÍCH JEDNOTEK – KALENDÁŘNÍ MĚSÍC .....   | 56 |
| <b>PŘÍLOHA 5</b> UPRAVENÁ DATA PO ETAPĚ PŘÍPRAVY DAT CRISP_DM – ZMĚNA NÁZVU POLÍ A PŘEHLED<br>SLOUČENÝCH DAT PRO ETAPU MODELOVÁNÍ DAT CRISP_DM ..... | 57 |
| <b>PŘÍLOHA 6</b> SLOUČENÍ DO ZÁJMOVÝCH SKUPIN JEDNOTEK .....   | 58 |
| <b>PŘÍLOHA 7</b> SCHÉMA PŘÍPRAVY DAT TABULEK SPOLZASH A VÝJEZDY V PROSTŘEDÍ CLEMENTINE .....   | 59 |
| <b>PŘÍLOHA 8</b> DISTRIBUČNÍ ROZDĚLENÍ POČTU VÝJEZDŮ HZS K ŽIVELNÝM POHROMÁM ZA 24HODIN.....   | 60 |
| <b>PŘÍLOHA 9</b> PROCENTUÁLNÍ ROZLOŽENÍ POČTU VÝJEZDŮ HZS TYPU ZÁSAHU V KALENDÁŘNÍ DEN TÝDNE<br>K POČTU VŠECH VÝJEZDŮ HZS DANÝ KALENDÁŘNÍ DEN. ....  | 60 |
| <b>PŘÍLOHA 10</b> STREAM VYTVOŘENÍ MĚSÍČNÍCH SOUČTŮ VÝJEZDŮ HZS K ZÁSAHŮM.....   | 61 |
| <b>PŘÍLOHA 11</b> DATOVÝ SLOVNÍK .....   | 62 |

# PŘÍLOHY

## Příloha 1 Elektronická příloha

Vložené CD-R médium obsahuje tyto soubory:

### **\\vstupni\_data**

spolzash.csv – obsahuje vstupní data po exportu z XML o spolupráci u zásahu HZS, časový interval součtů v záznamech je kalendářní měsíc

vyjezdy.csv – obsahuje vstupní data po exportu z XML o výjezdech k zásahu HZS, časový interval součtů v záznamech je kalendářní den

### **\\model\_data**

modeldat.csv – soubor vzniklý sloučením vstupních souborů

### **\\stream**

obsahuje soubory streamů pro práci v Clementine.

den.str

tyden.str

mesic.str

modelovani.str

## Příloha 2 Vstupní data pro tabulky počtu denních zásahů HZS část 1/2

ISV 5.0 Statistické sledování událostí

HZS Pardubického kraje

### Počet událostí podle dne v týdnu 1/2

Legenda ke sloupcům

P - POŽÁR

12 - P - s účastí jednotky

19 - P - bez účasti jednotky

DN - DOPRAVNÍ NEHODA

21 - DN - silniční

22 - DN - silniční hromadná

23 - DN - železniční (vč. metra)

24 - DN - letecká

25 - DN - ostatní

ŽP - ŽIVELNÍ POHROMA

31 - ŽP - převažující povodeň, záplava

32 - ŽP - převažující sniž, námrazy

33 - ŽP - převažující větrná smíř

34 - ŽP - sesuv půdy

35 - ŽP - ostatní ( např. zemětřesení)

**Období od : 31.5.2010 do: 6.6.2010**

|               | 12       | 19       | SUM 12-19 | 21        | 22       | 23       | 24       | 25       | SUM 21-25 | 31       | 32       | 33       | 34       | 35       | SUM 31-35 |
|---------------|----------|----------|-----------|-----------|----------|----------|----------|----------|-----------|----------|----------|----------|----------|----------|-----------|
| Pondělí       | 1        | 1        | 2         | 0         | 0        | 0        | 0        | 0        | 0         | 0        | 0        | 0        | 0        | 0        | 0         |
| Úterý         | 1        | 0        | 1         | 3         | 0        | 0        | 0        | 0        | 3         | 0        | 0        | 0        | 0        | 0        | 0         |
| Středa        | 1        | 0        | 1         | 1         | 0        | 0        | 0        | 0        | 1         | 0        | 0        | 0        | 0        | 0        | 0         |
| Čtvrtek       | 1        | 0        | 1         | 2         | 0        | 1        | 0        | 0        | 3         | 0        | 0        | 0        | 0        | 0        | 0         |
| Pátek         | 0        | 0        | 0         | 3         | 0        | 0        | 0        | 0        | 3         | 0        | 0        | 0        | 0        | 0        | 0         |
| Sobota        | 0        | 0        | 0         | 7         | 0        | 0        | 0        | 0        | 7         | 0        | 0        | 0        | 0        | 0        | 0         |
| Neděle        | 1        | 0        | 1         | 2         | 0        | 0        | 0        | 0        | 2         | 0        | 0        | 0        | 0        | 0        | 0         |
| <b>CELKEM</b> | <b>5</b> | <b>1</b> | <b>6</b>  | <b>18</b> | <b>0</b> | <b>1</b> | <b>0</b> | <b>0</b> | <b>19</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>0</b>  |

**Podminky**

**STAVY ÚDÁLOSTÍ:** Čekající na odbavení (100), Inicializovaná (200), Převzatá (210), Odložena (300), Otevřená, bez SaP (400), Otevřená SaP na cestě (410), Otevřená, SaP na místě (420), Otevřená, Lokalizovaná (430), Likvidovaná (440), Ukončená OS (500), Uzavřená OS (510), Exportovaná do SK (520), Převzatá VZ (600), Ukončená VZ (610), Převzatá ZPP (700), Ukončená ZPP (710), Převzatá garantem za SSU (750), Ukončená garantem za SSU (760), Potvrzená krajským garantem (780)

**PRÍZNAKY:** Událost zahrnuta do SSU, Mezikrajská výpomoc, Meziústežní výpomoc, Mimoústežní událost, Událost Českých drah, Povodeň, SLAK, Terorismus, Kalamita, Víceústežní zásah, Svolsní krizového štábu určených obcí, krajů, Ustanovení štábu velitele zásahu, Zpracování znaleckého posudku, ŽPP na místě, Požár dále nedošetřovaný, Povolaní řídicí důstojník, Taktické cvičení, Tisknout událost na sestavě pro media, Ptáci chrípka, Použitá letecká technika, Použití suchovodu, Nasazení psa, Událost je určena pro dispečink NDIC, Neexportováno pro Webrex, Činnost laboratorě HZS, Zásah za úpiatu, Zpráva o ukončení události pro TCTV 112, Sportovní, kulturní akce, Živelní pohroma, ŽP - Převažující povodeň, záplava, dešt, ŽP - Převažující větrná smíř, ŽP - Sesuv půdy, ŽP - ostatní (např. zemětřesení), Použitá skupina leteckých zachranářů

## Příloha 3 Vstupní data pro tabulky počtu denních zásahů HZS část 2/2

HZS Pardubického kraje

ISV 5.0 Statistické sledování událostí

### Počet událostí podle dne v týdnu 2/2

Legenda ke sloupcům

**UNL - Únik nebezpečné chemické látky**

41 - UNL - únik plynu/aerosolu

42 - UNL - únik kapaliny (mimo rop. produktů)

43 - UNL - únik ropných produktů

44 - UNL - únik pevné látky

45 - UNL - únik ostatní (včetně jině než chemické)

**TH - Technická havárie (mimo UNL)**

51 - TH - technická havárie

52 - TH - technická pomoc

53 - TH - technologická pomoc

54 - TH - ostatní pomoc

**RHN - Radiční havárie a nehoda**

61 - Radiční havárie a nehoda

**OMU - Ostatní mimořádné události**

71 - Ostatní mimořádné události (epidemie, nákazy a jiné)

**PP - Planý poplach**

81 - Planý poplach

**Období od : 31.5.2010 do: 6.6.2010**

|               | 41       | 42       | 43       | 44       | 45       | SUM 41-45 | 51       | 52        | 53       | 54       | SUM        | 61       | 71       | 81       | SUM 1-81   |
|---------------|----------|----------|----------|----------|----------|-----------|----------|-----------|----------|----------|------------|----------|----------|----------|------------|
| Pondělí       | 0        | 0        | 1        | 0        | 0        | 1         | 0        | 6         | 0        | 0        | 6          | 0        | 0        | 0        | 10         |
| Úterý         | 0        | 0        | 0        | 0        | 0        | 0         | 0        | 3         | 0        | 2        | 5          | 0        | 0        | 0        | 9          |
| Středa        | 1        | 0        | 1        | 0        | 0        | 2         | 0        | 38        | 2        | 2        | 42         | 0        | 0        | 0        | 47         |
| Čtvrtek       | 0        | 0        | 0        | 0        | 0        | 0         | 0        | 16        | 0        | 1        | 17         | 0        | 0        | 0        | 21         |
| Pátek         | 0        | 0        | 0        | 0        | 0        | 0         | 0        | 12        | 0        | 2        | 14         | 0        | 0        | 1        | 18         |
| Sobota        | 0        | 0        | 1        | 0        | 0        | 1         | 0        | 8         | 0        | 2        | 10         | 0        | 0        | 0        | 18         |
| Neděle        | 0        | 0        | 0        | 0        | 0        | 0         | 0        | 16        | 0        | 0        | 16         | 0        | 0        | 0        | 19         |
| <b>CELKEM</b> | <b>1</b> | <b>0</b> | <b>3</b> | <b>0</b> | <b>0</b> | <b>4</b>  | <b>0</b> | <b>99</b> | <b>2</b> | <b>9</b> | <b>110</b> | <b>0</b> | <b>0</b> | <b>3</b> | <b>142</b> |

**Podmínky**

**STAVY UDÁLOSTI:** Čekající na odbavení (100), Inicializovaná (200), Převzatá (210), Odložena (300), Otevřená, bez SaP (400), Otevřená SaP na cestě (410), Otevřená SaP na místě (420), Lokalizovaná (430), Likvidovaná (440), Ukončená OS (500), Uzavřená OS (510), Exportovaná do SK (520), Převzatá VZ (600), Ukončená VZ (610), Převzatá ZPP (700), Ukončená ZPP (710), Převzatá garantem za SSU (750), Ukončená garantem za SSU (760), Potvrzená krajským garantem (780)

**PRIZNÁKY:** Událost zahrnuta do SSU, Mezikrajská výpomoc, Meziústežní výpomoc, Mimořádná událost, Událost Českých drah, Povodeň, SLAK, Terorismus, Kalamita, Vícenásobný zásah, Svolení krizového štábu určených obcí, krajů, Ustanovení štábu veřejné záchrany, Zpracování znaleckého posudku, ZPP na místě, Požár dále nedořešovaný, Povolán řidič důstojník, Taktické cvičení, Tisknout událost na sestavu pro media, Ptáci chřipka, Použita letecká technika, Použití suchovodu, Nasazení psa, Událost je určena pro dispečnický NDIC, Neexportováno pro Webrex, Cinnost laboratorně HZS, Zásah za úplatu, Zpráva o ukončení události pro TCTV 112, Sportovní, kulturní akce, Živelní pohroma, ŽP - Převážující povodeň, záplava, dešť, ŽP - Převážující sniž, námrazy, ŽP - Převážující větrná směr, ŽP - Sesuv půdy, ŽP - ostatní (např. zeměměřeni), Použita skupina leteckých záchranářů

## Příloha 4 Vstupní data pro tabulku spolu zasahujících jednotek – kalendářní měsíc

ISV 5.0 Statistické sledování událostí

HZS Pardubického kraje

### Spolupráce u zásahu - součty

#### Legenda ke sloupcům

P - POZAR (12-19)  
 DN - DOPRAVNÍ NEHODA (21-25)  
 ZP - ZIVELNÍ POHROMA (31-35)  
 UNL - Únik nebezpečné chemické látky (41-45)  
 TH - Technická havárie (mimo UNL, 51-54)  
 RHN - Radiační havárie a nehoda (61)  
 OMU - Ostatní mimořádné události (71)  
 PP - Planý poplach (81)

**Období od : 1.8.2009 do: 31.8.2009**

|  | SUM 12-19 | SUM 21-25  | SUM 31-35 | SUM 41-45 | SUM 51-54  | SUM 61-81 | SUM 1-81   |
|--|-----------|------------|-----------|-----------|------------|-----------|------------|
| Police CR                              | 52        | 93         | 1         | 5         | 74         | 18        | 243        |
| Zdravotnická záchranná služba          | 8         | 62         | 0         | 0         | 45         | 0         | 115        |
| Armáda CR - voj. záchr. útvary         | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Armáda CR - jiné                       | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Obecní policie                         | 4         | 5          | 0         | 0         | 14         | 1         | 24         |
| Hygienická služba                      | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Báňská záchranná služba                | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Občanská sdružení v IZS                | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Pohotovostní služba el. rozvod. závodů | 3         | 1          | 0         | 0         | 2          | 0         | 6          |
| Plynárenská pohotovostní služba        | 0         | 0          | 0         | 2         | 0          | 0         | 4          |
| Vodárenská pohotovostní služba         | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Teplárenská pohotovostní služba        | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Státní úřad pro jad. bezpečnost        | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Ostatní ústřední orgány státní správy  | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Firmy sdružené pod TRINS               | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Obecní zastupitelstvo                  | 0         | 1          | 0         | 0         | 3          | 0         | 4          |
| Ostatní územní orgány státní správy    | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| Podniky, firmy                         | 1         | 1          | 0         | 0         | 0          | 0         | 2          |
| Místní služby                          | 0         | 1          | 0         | 1         | 4          | 0         | 6          |
| Česká inspekce životního prostředí     | 1         | 0          | 0         | 0         | 0          | 0         | 1          |
| Ostatní subjekty(název)                | 0         | 0          | 0         | 0         | 0          | 0         | 0          |
| <b>CELKEM</b>                          | <b>69</b> | <b>164</b> | <b>1</b>  | <b>8</b>  | <b>144</b> | <b>19</b> | <b>405</b> |

#### Podmínky

KRAJ: Pardubický  
 STAVY ÚDALOSTI: Ukončená VZ (610), Ukončená ZPP (710), Ukončená garantom za SSU (760), Potvrzená krajským garantem (780)  
 PRŮZNAKY: Událost zahrnuta do SSU  
 MIMO - Technologický test, Prověřovací cvičení, Taktické cvičení

Nejedná se o požár bez účasti jednotky.



**Příloha 5** Upravená data po etapě přípravy dat CRISP\_DM – změna názvu polí a přehled sloučených dat pro etapu modelování dat CRISP\_DM

Upravená\_jednotky

Merge 7 datasets. Merge method: Keys

Fields: 51 in, 0 filtered, 49 renamed, 51 out

| Field   | Tag | Source Node  | Connected Node          | Filter | Field    |
|---------|-----|--------------|-------------------------|--------|----------|
| Rok     |     |              |                         | →      | Rok      |
| Mesic   |     |              |                         | →      | Mesic    |
| P_Sum   | 1   | Spolzash.csv | Ostatní                 | →      | P_OST    |
| DN_Sum  | 1   | Spolzash.csv | Ostatní                 | →      | DN_OST   |
| ZP_Sum  | 1   | Spolzash.csv | Ostatní                 | →      | ZP_OST   |
| UNL_Sum | 1   | Spolzash.csv | Ostatní                 | →      | UNL_OST  |
| TH_Sum  | 1   | Spolzash.csv | Ostatní                 | →      | TH_OST   |
| PP_Sum  | 1   | Spolzash.csv | Ostatní                 | →      | PP_OST   |
| ALL_Sum | 1   | Spolzash.csv | Ostatní                 | →      | ALL_OST  |
| P_Sum   | 2   | Spolzash.csv | Policie ČR              | →      | P_PČR    |
| DN_Sum  | 2   | Spolzash.csv | Policie ČR              | →      | DN_PČR   |
| ZP_Sum  | 2   | Spolzash.csv | Policie ČR              | →      | ZP_PČR   |
| UNL_Sum | 2   | Spolzash.csv | Policie ČR              | →      | UNL_PČR  |
| TH_Sum  | 2   | Spolzash.csv | Policie ČR              | →      | TH_PČR   |
| PP_Sum  | 2   | Spolzash.csv | Policie ČR              | →      | PP_PČR   |
| ALL_Sum | 2   | Spolzash.csv | Policie ČR              | →      | ALL_PČR  |
| P_Sum   | 3   | Spolzash.csv | Zdrav Služba            | →      | P_ZS     |
| DN_Sum  | 3   | Spolzash.csv | Zdrav Služba            | →      | DN_ZS    |
| ZP_Sum  | 3   | Spolzash.csv | Zdrav Služba            | →      | ZP_ZS    |
| UNL_Sum | 3   | Spolzash.csv | Zdrav Služba            | →      | UNL_ZS   |
| TH_Sum  | 3   | Spolzash.csv | Zdrav Služba            | →      | TH_ZS    |
| PP_Sum  | 3   | Spolzash.csv | Zdrav Služba            | →      | PP_ZS    |
| ALL_Sum | 3   | Spolzash.csv | Zdrav Služba            | →      | ALL_ZS   |
| P_Sum   | 4   | Spolzash.csv | Obecni policie          | →      | P_OBP    |
| DN_Sum  | 4   | Spolzash.csv | Obecni policie          | →      | DN_OBP   |
| ZP_Sum  | 4   | Spolzash.csv | Obecni policie          | →      | ZP_OBP   |
| UNL_Sum | 4   | Spolzash.csv | Obecni policie          | →      | UNL_OBP  |
| TH_Sum  | 4   | Spolzash.csv | Obecni policie          | →      | TH_OBP   |
| PP_Sum  | 4   | Spolzash.csv | Obecni policie          | →      | PP_OBP   |
| ALL_Sum | 4   | Spolzash.csv | Obecni policie          | →      | ALL_OBP  |
| P_Sum   | 5   | Spolzash.csv | Energetické pohotovosti | →      | P_ENRG   |
| DN_Sum  | 5   | Spolzash.csv | Energetické pohotovosti | →      | DN_ENRG  |
| ZP_Sum  | 5   | Spolzash.csv | Energetické pohotovosti | →      | ZP_ENRG  |
| UNL_Sum | 5   | Spolzash.csv | Energetické pohotovosti | →      | UNL_ENRG |
| TH_Sum  | 5   | Spolzash.csv | Energetické pohotovosti | →      | TH_ENRG  |
| PP_Sum  | 5   | Spolzash.csv | Energetické pohotovosti | →      | PP_ENRG  |
| ALL_Sum | 5   | Spolzash.csv | Energetické pohotovosti | →      | ALL_ENRG |
| P_Sum   | 6   | Spolzash.csv | Podniky-místní služba   | →      | P_PM     |
| DN_Sum  | 6   | Spolzash.csv | Podniky-místní služba   | →      | DN_PM    |
| ZP_Sum  | 6   | Spolzash.csv | Podniky-místní služba   | →      | ZP_PM    |
| UNL_Sum | 6   | Spolzash.csv | Podniky-místní služba   | →      | UNL_PM   |
| TH_Sum  | 6   | Spolzash.csv | Podniky-místní služba   | →      | TH_PM    |
| PP_Sum  | 6   | Spolzash.csv | Podniky-místní služba   | →      | PP_PM    |
| ALL_Sum | 6   | Spolzash.csv | Podniky-místní služba   | →      | ALL_PM   |
| P_Sum   | 7   | Vyjezdy.csv  | Hasiči                  | →      | P_HSZ    |
| DN_Sum  | 7   | Vyjezdy.csv  | Hasiči                  | →      | DN_HSZ   |
| ZP_Sum  | 7   | Vyjezdy.csv  | Hasiči                  | →      | ZP_HSZ   |
| UNL_Sum | 7   | Vyjezdy.csv  | Hasiči                  | →      | UNL_HSZ  |
| TH_Sum  | 7   | Vyjezdy.csv  | Hasiči                  | →      | TH_HSZ   |
| PP_Sum  | 7   | Vyjezdy.csv  | Hasiči                  | →      | PP_HSZ   |
| ALL_Sum | 7   | Vyjezdy.csv  | Hasiči                  | →      | ALL_HSZ  |

View current fields  View unused field settings

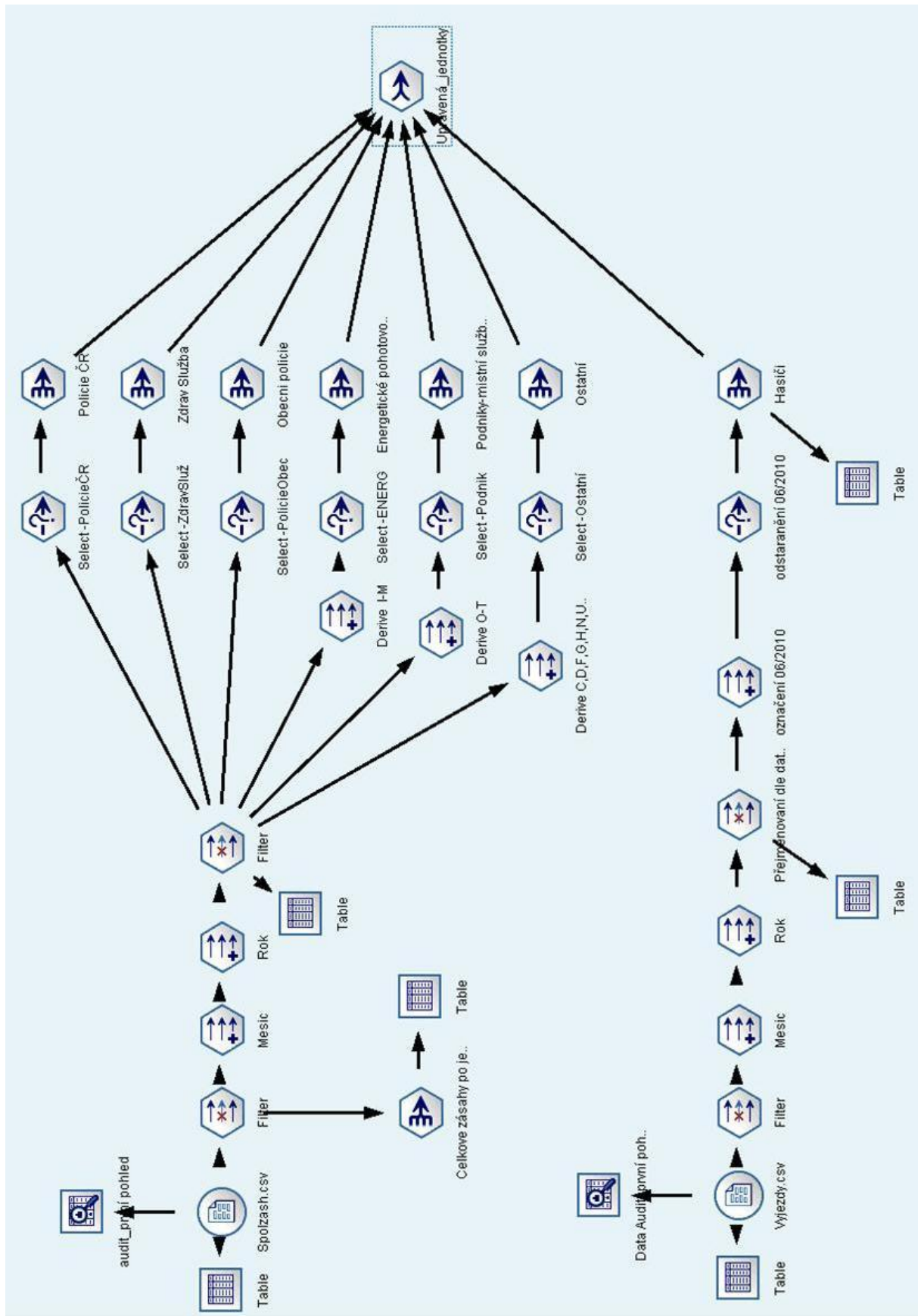
Inputs Merge Filter Optimization Annotations

OK Cancel Apply Reset

**Příloha 6** Sloučení do zájmových skupin jednotek

| Proměnná/pole<br>jednotka/<br>spoluzash.csv | Plný název proměnné                    | Sloučení jednotek<br>suffix |
|---|--|-----------------------------|
| A   | Policie ČR                             | PČR                         |
| B   | Zdravotnická záchranná služba          | ZS                          |
| C   | Armáda ČR - voj. záchr. útvary         | OST                         |
| D   | Armáda ČR - jiné                       | OST                         |
| E   | Obecní policie                         | OBP                         |
| F   | Hygienická služba                      | OST                         |
| G   | Báňská záchranná služba                | OST                         |
| H   | Občanská sdružení v IZS                | OST                         |
| I   | Pohotovostní služba el. rozvod. závodů | ENERG                       |
| J   | Plynárenská pohotovostní služba        | ENERG                       |
| K   | Vodárenská pohotovostní služba         | ENERG                       |
| L   | Teplárenská pohotovostní služba        | ENERG                       |
| M   | Státní úřad pro jader. bezpečnost      | ENERG                       |
| N   | Ostatní ústřední orgány státní správy  | OST                         |
| O   | Firmy sdružené pod TRINS               | PM                          |
| P   | Obecní zastupitelstvo                  | PM                          |
| R   | Ostatní územní orgány státní správy    | PM                          |
| S   | Podniky, firmy                         | PM                          |
| T   | Místní služby                          | PM                          |
| U   | Česká inspekce životního prostředí     | OST                         |
| V   | Ostatní subjekty(název)                | OST                         |

**Příloha 7** Schéma přípravy dat tabulek spolzash a vyjezdy v prostředí Clementine



## Příloha 8 Distribuční rozdělení počtu výjezdů HZS k živelným pohromám za 24hodin

|    | Value | Proportion | %     | Count |
|----|-------|------------|-------|-------|
| 1  | 0     |            | 94,07 | 1 522 |
| 2  | 1     |            | 2,53  | 41    |
| 3  | 2     |            | 0,68  | 11    |
| 4  | 3     |            | 0,49  | 8     |
| 5  | 4     |            | 0,25  | 4     |
| 6  | 5     |            | 0,25  | 4     |
| 7  | 6     |            | 0,25  | 4     |
| 8  | 7     |            | 0,19  | 3     |
| 9  | 8     |            | 0,06  | 1     |
| 10 | 9     |            | 0,06  | 1     |
| 11 | 11    |            | 0,06  | 1     |
| 12 | 14    |            | 0,12  | 2     |
| 13 | 16    |            | 0,06  | 1     |
| 14 | 17    |            | 0,06  | 1     |
| 15 | 23    |            | 0,06  | 1     |
| 16 | 26    |            | 0,06  | 1     |
| 17 | 32    |            | 0,12  | 2     |
| 18 | 33    |            | 0,06  | 1     |
| 19 | 35    |            | 0,06  | 1     |
| 20 | 36    |            | 0,06  | 1     |
| 21 | 59    |            | 0,06  | 1     |
| 22 | 60    |            | 0,06  | 1     |
| 23 | 73    |            | 0,06  | 1     |
| 24 | 108   |            | 0,06  | 1     |
| 25 | 129   |            | 0,06  | 1     |
| 26 | 173   |            | 0,06  | 1     |
| 27 | 260   |            | 0,06  | 1     |

## Příloha 9 Procentuální rozložení počtu výjezdů HZS typu zásahu v kalendářní den týdne k počtu všech výjezdů HZS daný kalendářní den.

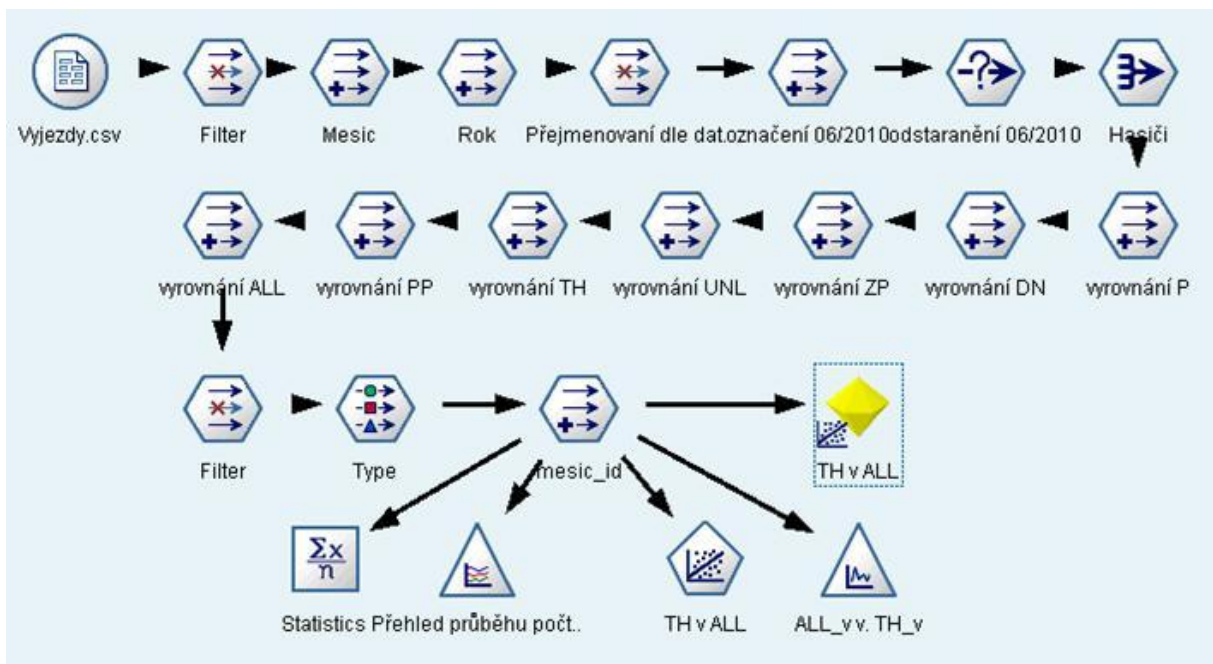
| Dny v týdnu | P     | DN    | ZP    | UNL  | TH    | PP   | ALL    |
|-------------|-------|-------|-------|------|-------|------|--------|
| 1           | 12,6% | 23,8% | 1,9%  | 3,4% | 53,6% | 4,5% | 100,0% |
| 2           | 12,8% | 24,8% | 2,0%  | 3,5% | 52,5% | 4,3% | 100,0% |
| 3           | 11,1% | 22,4% | 2,6%  | 3,0% | 55,4% | 4,9% | 100,0% |
| 4           | 10,8% | 21,8% | 9,3%  | 3,4% | 50,7% | 3,9% | 100,0% |
| 5           | 11,8% | 26,7% | 6,7%  | 3,6% | 46,0% | 5,0% | 100,0% |
| 6           | 13,4% | 25,0% | 10,0% | 2,7% | 43,6% | 4,9% | 100,0% |
| 7           | 14,8% | 25,2% | 3,9%  | 3,1% | 47,1% | 5,4% | 100,0% |

vyjadřuje minima z podílů typu zásahu počtu výjezdů HZS v kalendářní den vzhledem k absolutnímu počtu všech výjezdů HZS daný kalendářní den

vyjadřuje maxima z podílů typu zásahu počtu výjezdů HZS v kalendářní den vzhledem k absolutnímu počtu všech výjezdů HZS daný kalendářní den

Z důvodu lepší interpretace je označeno více hodnot

**Příloha 10** Stream vytvoření měsíčních součtů výjezdů HZS k zásahům



## Příloha 11 Datový slovník

Výskyt se sufixem jednotky znamená hodnotu sufixu:

|        |   |
|--------|---|
| _ENERG | Pohotovostní služby energetických společností |
| _PM    | Podniky, místní - obecní služby               |
| _OST   | Ostatní jednotky                              |
| _PČR   | Policie ČR                                    |
| _ZS    | Zdravotní záchranná služba                    |
| _OBP   | Obecní policie                                |
| _SUM   | Součet hodnot dle uzlu <i>Aggregate</i>       |

| Původní jméno pole | Vysvětlující popis                | Jméno pole přípravy dat | Výskyt se sufixem | Datový typ |
|--------------------|-----------------------------------|-------------------------|-------------------|------------|
| ALL                | celkový počet výjezdů             |                         | A                 | Range      |
| ALL_v              | ALL za měsíc vyrovnaná hodnota    |                         | A                 | Range      |
| DAT_DO             | datum konce měsíce                |                         | N                 | Set        |
| DAT_OD             | datum začátku měsíce              |                         | N                 | Set        |
| Date               | kalendářní den datum              |                         | N                 | Order set  |
| Den                | požáry s účastí                   |                         | N                 | Discrete   |
| DN_21              | dopravní nehoda silniční hromadná |                         | N                 | Range      |
| DN_22              | dopravní nehoda železniční        |                         | N                 | Range      |
| DN_23              | dopravní nehoda letecká           |                         | N                 | Range      |
| DN_24              | dopravní nehoda ostatní           |                         | N                 | Range      |
| DN_25              | součet dopravní nehoda            |                         | N                 | Range      |
| DN_v               | DN za měsíc vyrovnaná hodnota     |                         | A                 | Range      |
| E_tyden            | datum konce týdne (neděle)        |                         | N                 | Set        |
| H_12               | požáry s účastí                   |                         | N                 | Range      |
| H_19               | požáry bez účasti                 |                         | N                 | Range      |
| H_21               | dopravní nehoda silniční          |                         | N                 | Range      |
| H_22               | dopravní nehoda silniční hromadná |                         | N                 | Range      |
| H_23               | dopravní nehoda železniční        |                         | N                 | Range      |
| H_24               | dopravní nehoda letecká           |                         | N                 | Range      |
| H_25               | dopravní nehoda ostatní           |                         | N                 | Range      |



| Původní jméno pole | Vysvětlující popis  | Jméno pole přípravy dat | Výskyt se suffixem | Datový typ |
|--------------------|---|-------------------------|--------------------|------------|
| H_31               | živelná pohroma - povodeň, záplava                        |                         | N                  | Range      |
| H_32               | živelná pohroma - sních, námraza                          |                         | N                  | Range      |
| H_33               | živelná pohroma - větrná smršť                            |                         | N                  | Range      |
| H_34               | živelná pohroma - sesuv půdy                              |                         | N                  | Range      |
| H_35               | živelná pohroma - ostatní                                 |                         | N                  | Range      |
| H_41               | únik nebezpečných chemických látek -únik plynu areosolu   |                         | N                  | Range      |
| H_42               | únik nebezpečných chemických látek - únik kapalin         |                         | N                  | Range      |
| H_43               | únik nebezpečných chemických látek -únik ropných produktů |                         | N                  | Range      |
| H_44               | únik nebezpečných chemických látek - únik pevné látky     |                         | N                  | Range      |
| H_45               | únik nebezpečných chemických látek únik ostatní           |                         | N                  | Range      |
| H_51               | technická havárie   |                         | N                  | Range      |
| H_52               | technická pomoc   |                         | N                  | Range      |
| H_53               | technologická pomoc                                       |                         | N                  | Range      |
| H_54               | ostatní pomoc   |                         | N                  | Range      |
| H_61               | radiační havárie a nehoda                                 | RHN                     | A                  | Range      |
| H_71               | ostatní mimořádné události                                | OMU                     | A                  | Range      |
| H_81               | planý poplach   | PP                      | A                  | Range      |
| H_SK1              | součet požáry   | P                       | A                  | Range      |
| H_SK2              | součet dopravní nehoda                                    | DN                      | A                  | Range      |
| H_SK3              | součet živelná pohroma                                    | ZP                      | A                  | Range      |
| H_SK4              | součet únik nebezpečných chemických látek                 | UNL                     | A                  | Range      |
| H_SK5              | technická havárie   | TH                      | A                  | Range      |
| H_SK6              | součet RHN, OMU, PP                                       | Sum_O                   | A                  | Range      |
| H_SK7              | celkový počet výjezdů                                     | ALL                     | A                  | Range      |

| Původní jméno pole | Vysvětlující popis                                      | Jméno pole přípravy dat | Výskyt se suffixem | Datový typ  |
|--------------------|---|-------------------------|--------------------|-------------|
| ID                 | bezrozměrné číslo ze vstupu                             |                         | N                  | Range       |
| Mesíc              | obsahuje kalendářní měsíc ve tvaru                      |                         | N                  | Ordered set |
| mesic_id           | pomocné pole  |                         |                    | Ordered set |
| Mesicdel06         | pomocné pole na označení True, False záznamů pro Select |                         | N                  | Flag        |
| MU_71              | planý poplach   | MU                      | A                  | Range       |
| NU_44              | únik nebezpečných chemických látek únik ostatní         |                         | N                  | Range       |
| P_B                | součet požáry   |                         | N                  | Range       |
| P_S                | požáry bez účasti                                       |                         | N                  | Range       |
| P_v                | P za měsíc vyrovnaná hodnota                            |                         | A                  | Range       |
| PP_81              | součet RHN, OMU, PP                                     | PP                      | A                  | Range       |
| PP_v               | PP za měsíc vyrovnaná hodnota                           |                         | A                  | Range       |
| RH_61              | ostatní mimořádné události                              | RHN                     | A                  | Range       |
| Rok                | obsahuje kalendářní rok ve tvaru XXXX                   |                         | N                  | Ordered set |
| rok_id             | pomocné pole  |                         |                    | Ordered set |
| S_tyden            | datum začátku týdne (pondělí)                           |                         | N                  | Set         |
| SumaDN             | živelná pohroma - povodeň, záplava                      | DN                      | A                  | Range       |
| SumaP              | dopravní nehoda silniční                                | P                       | A                  | Range       |
| SumaTH             | radiační havárie a nehoda                               | TH                      | A                  | Range       |
| SumaUN             | technická havárie                                       | UNL                     | A                  | Range       |
| SumaZP             | únik nebezpečných chemických látek -únik plynu areosolu | ZP                      | A                  | Range       |
| TEXT               | druh jednotky   | jednotka                | A                  | Discrete    |
| TEXT_I-M           | pomocné pole  |                         | N                  | Flag        |
| TEXT_OSTATNI       | pomocné pole  |                         | N                  | Flag        |
| TEXT_O-T           | pomocné pole  |                         | N                  | Flag        |
| TH_51              | technická pomoc   |                         | N                  | Range       |
| TH_52              | technologická pomoc                                     |                         | N                  | Range       |
| TH_53              | ostatní pomoc   |                         | N                  | Range       |



| Původní jméno pole | Vysvětlující popis  | Jméno pole přípravy dat | Výskyt se suffixem | Datový typ |
|--------------------|---|-------------------------|--------------------|------------|
| TH_54              | technická havárie   |                         | N                  | Range      |
| TH_v               | TH za měsíc vyrovnaná hodnota                             |                         | A                  | Range      |
| UN_41              | únik nebezpečných chemických látek - únik kapalin         |                         | N                  | Range      |
| UN_42              | únik nebezpečných chemických látek -únik ropných produktů |                         | N                  | Range      |
| UN_43              | únik nebezpečných chemických látek - únik pevné látky     |                         | N                  | Range      |
| UN_45              | součet únik nebezpečných chemických látek                 |                         | N                  | Range      |
| UNL_v              | UNL za měsíc vyrovnaná hodnota                            |                         | A                  | Range      |
| ZP_31              | živelná pohroma - sníh, námraza                           |                         | N                  | Range      |
| ZP_32              | živelná pohroma - větrná smršť                            |                         | N                  | Range      |
| ZP_33              | živelná pohroma - sesuv půdy                              |                         | N                  | Range      |
| ZP_34              | živelná pohroma - ostatní                                 |                         | N                  | Range      |
| ZP_35              | součet živelná pohroma                                    |                         | N                  | Range      |
| ZP_v               | ZP za měsíc vyrovnaná hodnota                             |                         | A                  | Range      |