



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

**INVESTOVÁNÍ NA ATYPICKÉM TRHU S POMOCÍ
NÁSTROJŮ BUSINESS INTELLIGENCE**

INVESTING ON ATYPICAL MARKET BY USING THE BUSINESS INTELLIGENCE TOOLS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Ondřej Svoboda

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Kříž, Ph.D.

BRNO 2017

Zadání diplomové práce

Ústav: Ústav informatiky
Student: **Bc. Ondřej Svoboda**
Studijní program: Systémové inženýrství a informatika
Studijní obor: Informační management
Vedoucí práce: **Ing. Jiří Kříž, Ph.D.**
Akademický rok: 2017/18

Ředitel ústavu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává diplomovou práci s názvem:

Investování na atypickém trhu s pomocí nástrojů Business Intelligence

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza současného stavu
Vlastní návrhy řešení
Závěr
Seznam použité literatury
Přílohy

Cíle, kterých má být dosaženo:

Cílem práce je identifikace nových možností investování na trhu sportovních událostí a to pomocí nástrojů Business Intelligence včetně tvorby přehledů a reportů.

Základní literární prameny:

ARNOŠT, D. Business intelligence: příručka manažera. Praha: Tate International, 2007. 166 s. ISBN 978-80-86813-12-7.

HOFFMAN, J. Business Intelligence: Empower Your Users with Business Intelligence. [technet.microsoft.com\[online\]](https://technet.microsoft.com/en-us/magazine/gg405051.aspx). 2010 [cit. 2015-10-12]. Dostupné z: <https://technet.microsoft.com/en-us/magazine/gg405051.aspx>

LIDWELL, W., K. HOLDEN a J. BUTLER. Univerzální principy designu: 125 způsobů jak zvýšit použitelnost a přitažlivost a ovlivnit vnímání designu. 1. vyd. Brno: Computer Press, 2011. 272 s. ISBN 978-80-251-3540-2.

LIEBOWITZ, J. Strategic intelligence: business intelligence, competitive intelligence, and knowledge management. New York: Auerbach Publications, 2006. 223 s. ISBN 0-8493-9868-1.

TVRDÍKOVÁ, M. Aplikace moderních informačních technologií v řízení firmy: nástroje ke zvyšování kvality informačních systémů. Praha: Grada, 2008. 173 s. ISBN 978-80-247-2728-8.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2017/18

V Brně dne 28.2.2018

L. S.

doc. RNDr. Bedřich Půža, CSc.
ředitel

doc. Ing. et Ing. Stanislav Škapa, Ph.D.
děkan

Abstrakt

Tato diplomová práce se zabývá shrnutím, zpracováním, transformací a interpretací velkého množství kurzových dat. Tím si dává za úkol zjistit, zda je možné provozovat investiční činnost na trhu sportovních událostí bez jakýchkoliv předchozích zkušeností ze světa sportu, jen za pomoci matematických schopností a nástrojů business intelligence.

Abstract

This thesis deals with summarization, processing, transformation and interpretation of large amounts of betting odds data. The goal of this thesis is to find out whether it is possible to invest in sports betting odds market without any previous knowledge about sport, using only mathematical skills and business intelligence tools.

Klíčová slova

Investování, sport, sportovní událost, sázka, kurz, business intelligence, bit data

Key words

Investment, sport, sporting event, bet, odd, business intelligence, big data

Bibliografie

SVOBODA, O. *Investování na atypickém trhu s pomocí nástrojů Business Intelligence*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2017. 68 s.
Vedoucí diplomové práce Ing. Jiří Kříž, Ph.D..

Čestné prohlášení

Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 18. května 2018

.....

Bc. Ondřej Svoboda

Poděkování

Především bych rád poděkoval svému vedoucímu diplomové práce panu Ing. Jiřímu Křížovi, Ph.D., za spolupráci, cenné rady a odborné vedení práce.

Děkuji panu Ing. Michalu Teršlovi, paní Mgr. Haně Svobodové a mamince Daně Svobodové za podporu a pomoc při realizaci této práce.

Děkuji také své přítelkyni Karolíně Šindlerové a dědovi Danielu Holubovi za vstřícnost při psaní této práce.

Obsah

ÚVOD	13
CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ	14
1 TEORETICKÁ VÝCHODISKA PRÁCE	15
1.1 Investice	15
1.2 Sportovní sázení	15
1.2.1 Hazardní hra	15
1.2.2 Terminologie sázkařského průmyslu	15
1.2.3 Herní nabídka	16
1.2.4 Výhody sázení po internetu	16
1.2.5 Typy sportovního sázení	17
1.3 Datový sklad	18
1.3.1 Datové tržiště	19
1.3.2 Architektura datového skladu	20
1.4 Dolování dat	20
1.4.1 Metody dolování dat	20
1.5 Statistika	21
1.5.1 Matematická statistika	21
1.5.2 Teorie pravděpodobnosti	22
1.6 Multidimenzionální databáze	22
1.6.1 Tabulka faktů	22
1.6.2 Tabulka dimenzí	23
1.7 Definice pojmu BI	23
1.7.1 Základní principy BI	24

1.7.2	Nástroje a architektura BI	24
1.7.3	Uživatelská prezentace BI.....	25
1.8	ETL.....	26
1.8.1	Extrakce.....	26
1.8.2	Čištění dat.....	26
1.8.3	Transformace.....	26
1.8.4	Loading.....	27
1.9	Reporting	27
1.10	Analýzy.....	27
1.10.1	Porterův model pěti konkurenčních sil.....	27
1.10.2	PEST.....	28
1.10.3	SWOT.....	28
1.10.4	Analýza rizik	29
2	ANALÝZA SOUČASNÉHO STAVU	30
2.1	Historie vývoje aplikace na sázení	30
2.1.1	Vývoj z hlediska výpočtů.....	31
2.2	Popis vytvořeného principu sázení	31
2.2.1	Funkcionalita.....	32
2.3	Datové zdroje.....	33
2.3.1	Data ze systému na sázení.....	33
2.3.2	Data stažená ze stránek s výsledky	33
2.4	Zpracování analýz.....	33
2.4.1	Analýza porter	33
2.4.2	PEST.....	35
2.4.3	Analýza rizik	36

2.4.4	Analýza SWOT	36
3	VLASTNÍ NÁVRH ŘEŠENÍ	37
3.1	Vybudování datového skladu	37
3.1.1	Tabulka faktů	37
3.1.2	Tabulky dimenzí	38
3.1.3	Relační návrh DS	43
3.1.4	Extrahování dat ze staré DB	43
3.1.5	Extrahování dat z nové DB	45
3.1.6	Extrahování z HTML stránek	45
3.2	Procesy ETL	46
3.2.1	Source (Zdroj dat)	47
3.2.2	Sort (seřazení)	47
3.2.3	Merge Join (spojení dvou tabulek)	48
3.2.4	Union All (horizontální spojení)	48
3.2.5	Conditional Split (rozdělení dle podmínky)	48
3.2.6	Derivated Column (úpravy sloupců)	48
3.2.7	Pivot a Unpivot	48
3.2.8	Percentage Sampling	48
3.2.9	Destination	48
3.3	Nahrání do DS	49
3.3.1	Shrnutí	51
3.4	Reporty, sestavy	52
3.4.1	Report počtu událostí	52
3.4.2	Report průměrné výše kurzů	52
3.4.3	Report marže SK	54

3.4.4	Report hladiny SK	56
3.4.5	Report historických hodnot	58
ZÁVĚR.....		61
SEZNAM POUŽITÉ LITERATURY		62
SEZNAM GRAFŮ		64
SEZNAM OBRÁZKŮ		65
SEZNAM TABULEK.....		66
SEZNAM ZKRATEK.....		67
SEZNAM PŘÍLOH		68

ÚVOD

Uzavírání sázek na sportovní události - nejedná se o hazard, jak si mnoho lidí myslí. Přeci zbraně taky neznamenají válku, vždyť ji má u sebe každý policajt.

„Hazard je vše, co člověk dělá, ale největší hazard je nedělat nic.“

Shirley Williams

Dnešní doba se snaží zrušit citát *„Bez práce nejsou koláče“*. Každý by rád vydělal peníze jednoduše a hledá různé principy, kterými by toho bylo možné dosáhnout. Zdánlivě jsou jedním z nich investice. Ty se dělí dle situací, do kterých budeme investovat. Volné finanční prostředky lze vkládat do velmi rizikových portfolií, s malou šancí na velký zisk nebo do nerizikových nástrojů, kde s velkou pravděpodobností získáme minimum zisku. Kurzové sázky mohou být obojím. Při neopatrném přístupu můžeme ztratit vše, ale naopak sledováním nějakého ukazatele a dodržováním disciplíny můžeme hodně vydělat. Vše se odvíjí od událostí, na které se bude sázet, v jakých částkách a časových intervalech.

O tom bude pojednávat má diplomová práce. Nalézt možnosti, jak vydělat na trhu sportovních událostí bez předchozích znalostí sportu, ale za využití moderní techniky a znalostí z matematiky.

CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ

Cílem práce je identifikace nových možností investování na trhu sportovních událostí a to pomocí nástrojů Business Intelligence včetně tvorby přehledů a reportů. Dílčími cíli jsou extrakce, vyčištění a načtení dat do datového skladu. Z tohoto skladu se budou tvořit dotazy tak, aby se z velkého počtu dat vydolovaly na první pohled skryté informace. Klademe si následující otázky: Je možné dlouhodobě investovat na trhu sportovních událostí? Budou mít zisky z investice dlouhodobý rostoucí trend? Na jaké příležitosti a s jakými kurzy je nejvýhodnější investovat?

Pomocí moderní techniky, zastoupenou počítačem, na kterém poběží služby, jako jsou Microsoft SQL Server, Microsoft Visual Studio nebo Microsoft Power BI. V nich zpracují surová data z několika zdrojů, transportují je do databáze a vytvořím z nich reporty a pohledy. Znalost matematiky využiji při počítání hodnoty tzv. value a při interpretaci grafů na konci této práce.

1 TEORETICKÁ VÝCHODISKA PRÁCE

1.1 Investice

Je část příjmu vložená do dlouhodobých statků, které nepřinášejí okamžitý zisk, ale umožňují zvýšení prostředků v budoucnosti. Investor převádí část aktuální spotřeby za účelem budoucího zisku. Běžně se investicí označuje jakýkoliv vklad (Pavelka, 2007-2017).

1.2 Sportovní sázení

1.2.1 Hazardní hra

„Slovo hazard údajně pochází z arabského „asár“, což je výraz označující kostku. Hazardní hrou se pak obecně rozumí taková hra, jejíž výsledek závisí výhradně nebo převážně na náhodě.“ (Vanagas, 2015)

Hazardní hra se obvykle pojí s penězi, případně jejich ekvivalenty. Hráči sází na očekávaný výsledek. O výhře nebo prohře rozhodne pravděpodobnost. Část vkladů si ponechá pořadatel hry jako marži (Vanagas, 2015).

Nejčastěji se jedná například o ruletu, kostky, hrací automaty, sportovní sázky a některé karetní hry.

Podle Vincenta není kurzové sázení hazardem. Ten tvrdí, že jde o stejný princip, na kterém fungují finanční trhy. Pokud máme strategii při koupi akcie, tedy vybíráme ty s vnitřní hodnotou, na konci bychom měli být v černých číslech. Stejně je to i s kurzem (Vincent, 2009).

1.2.2 Terminologie sázkařského průmyslu

- **Sázková kancelář (SK)** - Podnikatelský subjekt, který provozuje kurzové sázky dle platného loterijního zákona.
- **Sázková burza** - Trh sportovních událostí, kde se střetává nabídka s poptávkou.
- **Událost** - Sportovní, politická nebo jiná událost, která je předmětem sázení. Obsahuje dvě a více příležitostí.
- **Příležitost** - Jeden z možných výsledků události (výhra domácích, remíza, výhra hostů, aj.).

- **Bookmaker** - Zaměstnanec sázkové kanceláře, který určí pravděpodobnost daných příležitostí.
- **Kurz** - Převrácená hodnota pravděpodobnosti příležitosti.
- **Nadhodnocený kurz** - Je kurz, jehož pravděpodobnost je nižší než skutečná pravděpodobnost příležitosti.
- **Hladina** - Podíl součtu obrácených hodnot kurzů všech příležitostí na jednu událost.
- **Marže SK** - Rozdíl jedničky a hladiny události. Liší se v závislosti na kanceláři, na oblíbenosti sportu a také se liší v čase.
- **Matchbetting** - Jedná se o bezrizikovou metodu sázení, kdy je provedena sázka na všechny možné výsledky dané události.
- **Arbitráž** - Jedná se o podmnožinu metody matchbetting, kdy je rovněž využívána metoda bezrizikového sázení s tím rozdílem, že tato situace dosahuje výhernosti nad 100 % (Sázkařský slovník - pojmy které by měl ovládat každý sázkař, 2011-2017).

1.2.3 Herní nabídka

Nabídka všech SK je různá, ale mnohdy se jednotlivé kanceláře předhánějí o zajímavější kurz. Nejlepší hladiny (sázkař zaplatí nejmenší marži) mají obecně sporty, jako je fotbal nebo tenis, ale velkou hladinu bude mít každá vyšší soutěž i v hokeji nebo basketbale.

1.2.4 Výhody sázení po internetu

Největší výhodou sázení na internetu je pravděpodobně rychlost, s jakou lze sázku uzavřít. Hledání zajímavého kurzu může být časově náročné a SK se snaží jejich nadhodnocené kurzy redukovat na obvyklou úroveň. S programem na hledání nadhodnocených kurzů se vše usnadňuje, ale navštěvování poboček SK by bylo příliš zdoluhavé. Téměř všechny SK nabízí svým klientům sázení z pohodlí domova a v přímém přenosu. Navíc díky odstranění nutnosti zaměstnance za přepážkou, můžou SK přistoupit k sázení bez manipulačního poplatku. Ten se obvykle počítá jako procento z vkladu. To sázkaři zajišťuje výhodnější sázení (Pavelka, 2007-2017).

1.2.5 Typy sportovního sázení

Existují různé druhy technik sázení. Některé jsou jistě výdělečné, jiné spoléhají na pravděpodobnost a trend a některé techniky jsou brány spíše jako koníček nebo zpestření konkrétní události.

Klasické sázení

Jedná se o rekreační sázení, kterým se zabývá největší procento sázkařů. Pro umocnění atmosféry při sledování zápasu si sázkař na utkání vsadí na jednu příležitost. Tento přístup sázení je dlouhodobě ztrátný díky maržím SK (Pavelka, 2007-2017).

Bonus hunting

SK pro nové klienty nabízejí vstupní a jiné bonusy. Při splnění podmínek (protočení určitého množství peněz) je bonus, ve formě bodů nebo peněz, připsán na herní účet. Sázkař ho většinou může libovolně využít na jakoukoliv událost (Vincent, 2009).

Arbitrážní sázení

Sázkař využívá různých softwarů na výpočet tzv. arbitráže. Sázkař uzavře vklady na všechny příležitosti události v různých poměrech. To mu zajistí jistý zisk, tzv. surebet (Vavřínek, 2010-2014).

Livebetting

Tento typ sázení se zakládá na sledování online přenosu a odhadu vývoje hry. Když se např. proti sobě staví dvě družstva, která mají v 15. minutě deset střel na bránu, nejspíše padne dva a více gólů. Sázkař tedy odhaduje budoucí vývoj události (Vavřínek, 2010-2014).

Trading

Sázkař očekává zvýšení nebo snížení kurzu. Staví se do dlouhé nebo krátké pozice (Vincent, 2009).

Martingale

Způsob sázení, kdy při prohrané sázce zvýšíme vklad na dvojnásobek. Vychází z teorie o pravděpodobnosti, kdy každá další uzavřená sázka (na podobný nebo stejný kurz) má větší šanci na úspěch (Pavelka, 2007-2017).

Hledání tzv. value

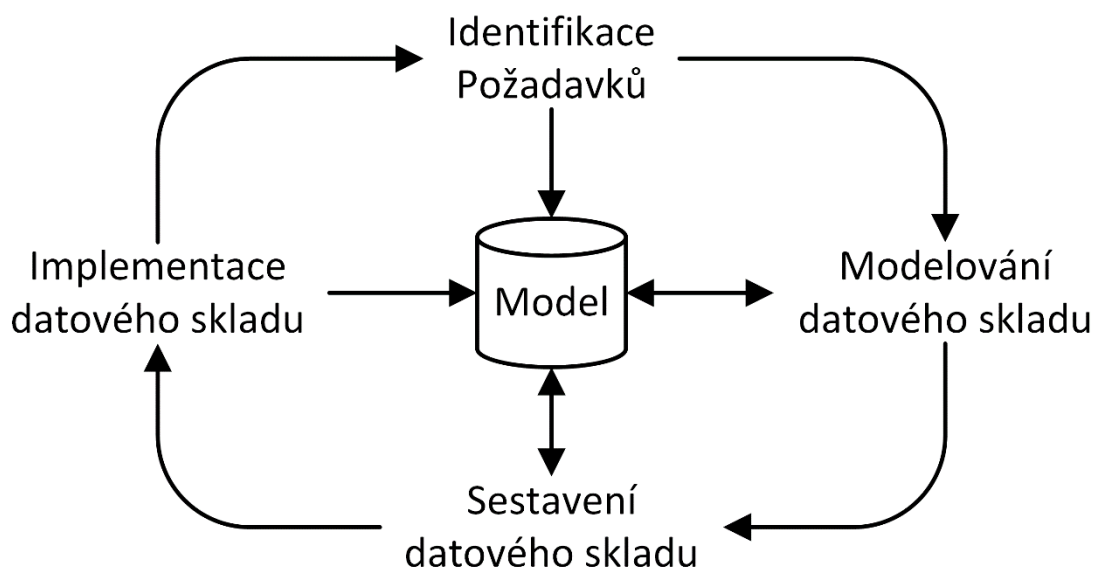
Sázkař hledá nadhodnocený kurz. Ten může najít několika způsoby:

- **Informovanost** - sázkař má lepší informace, než bookmaker;
- **Zainteresovanost** - ze stejných informací sázkař odhadne pravděpodobnost výsledku jinak, než bookmaker;
- **Využití software** - sázkař porovnává nabídku ostatních SK (Vavřínek, 2010-2014).

1.3 Datový sklad

„Datový sklad je integrovaný, subjektivě orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu.“ Bill Inmon

Datový sklad (DS, data warehouse) je systém, který shromažďuje, organizuje, uchovává a sdílí historická data. Obvykle mívá více zdrojových systémů. Termín DS často označuje systém DS a jindy se vztahuje na uložiště DS. Koncoví uživatelé vytvářejí dotazy na data uložená v DS a to jim pomáhá přijímat obchodní rozhodnutí. Dotazy mají různou formu od jednoduchých dotazů, přes komparativní analýzy, dolování dat, až po extrapolaci nebo prediktivní analýzu. Datový sklad má životní cykly vyznačeny na následujícím obrázku (Laberge, 2012).



Obrázek 1 Životní cyklus datového skladu

Zdroj: přepracováno dle (Laberge, 2012)

Životní cyklus DS zahrnuje identifikaci podnikových potřeb a také obecné technické požadavky. Po zjištění, že podnik usiluje o DS a dojde ke schválení rozpočtu na jeho budování, začíná fáze návrhu - modelování. V této fázi se zajišťuje technická a datová architektura, datové modelování a modelování procesů. Po těchto fázích následuje vlastní výstavba. Ta obsahuje fyzické budování prostředí - vytvoření databáze, určení indexů, programování úloh ETL (extract, transform a load) a psaní sestav (Laberge, 2012).

1.3.1 Datové tržiště

Neboli jinak datové trhy jsou podmnožinou datových skladů, které se vytváří pro organizační jednotku firmy na nižší úrovni hierarchie - např. pro oddělení, divizi nebo geografickou lokalitu (Lacko, 2009).

Princip je obdobný jako v DS. Rozdílem je, že datová tržiště (data marts) jsou určena pro omezený okruh uživatelů. Podstatou jsou tak decentralizované DS, které se integrují do celopodnikového řešení. Je to problémově orientovaný DS, určený pro problematiku daného okruhu uživatelů umožňující analýzu „ad hoc“ (Novotný, 2005).

1.3.2 Architektura datového skladu

Popisuje návrh systému, lze ji tedy přirovnat ke stavebním plánům. Jednoduše se dá říci toto - vezmou se zdrojová data, centralizují se do elegantně strukturované databáze a na ní se spustí nějaké sestavy. Ve skutečnosti jsou tyto procesy složité a časově velmi náročné (Laberge, 2012).

1.4 Dolování dat

Pro definici dolování dat (data mining) je obecně známo, že dolování dat je proces výběru, prohledávání a modelování ve velkých objemech dat sloužící k odhalení dříve neznámých vztahů mezi daty za účelem získání obchodní výhody (Časopis IT Systems, 2001).

Slouží pro vyhledávání vzorů v datech. Pokouší se zjišťovat a odkrývat různé závislosti, která v sobě data skrývají (Lacko, 2009).

1.4.1 Metody dolování dat

Úlohy dolování dat řeší zdánlivě nesouvisející problémy nejrozumnějších oborů. Výběr metody pro použití k řešení daného problému je jen jeden z kroků procesu dolování. Je třeba si dobře určit cíl, protože pro každý problém lze použít několik metod. Znalostí výhod a nevýhod jednotlivých metod lze vybrat jen pár z nich a porovnat jejich výsledky (Časopis IT Systems, 2001).

Prediktivní modelování

Je postupem, kdy se, na základě známé množiny vstupních a známých, jim odpovídajících výstupních hodnot, hledá nejpravděpodobnější hodnota výstupu pro předem neznámé kombinace vstupních hodnot (Časopis IT Systems, 2001).

Regrese

Popisuje stupeň důležitosti vstupních proměnných na výstup. Její přednost je v odhadu chyb modelu a v možnosti hledání závislostí na kombinaci vstupních hodnot (Časopis IT Systems, 2001).

Rozhodovací stromy

Mají snadnou interpretaci, a proto jsou hojně používané. Vytváří řadu jednoduchých rozhodovacích pravidel prezentovaných převážně grafy (Časopis IT Systems, 2001).

Klasifikace

Klasifikace umožňuje rozdělovat data do skupin podle různých kritérií. Pokud předem známe kritéria alespoň pro část dat, lze pomocí predikce vyvinout model, jehož výstupem je klasifikační proměnná. Více se však používá neřízená klasifikace, která výsledná kritéria nezná a zaměřuje se právě na jejich hledání - shluková analýza (Časopis IT Systems, 2001).

Analýza asociací

Nejčastěji používanou analýzou asociací je analýza nákupního košíku. Zabývá se kombinací produktů, které se často vyskytují v nákupním košíku spolu. Pomáhá marketingovým odborníkům v organizaci nabídky (Časopis IT Systems, 2001).

Vzorkování

Vybírá omezenou množinu dat ze základního souboru - nejjednodušeji náhodným výběrem sloužícím ke zmenšení objemu zpracovaných dat, a tím k zrychlení výpočtu. Složitěji vybírá stejný počet záznamů daného typu, a tím umožní redukci objemu dat (Časopis IT Systems, 2001).

1.5 Statistika

Je vědní obor zkoumající jevy mající hromadný charakter. Zkoumaný jev musí náležet části velkého množství prvků nebo musí existovat způsob, jak opakovaně získat požadované informace o zkoumaném jevu za podmínek, za nichž může nastat. Tento obor se pak zabývá zjišťováním, zpracováním, rozbořením, hodnocením a výkladem údajů o jevech. Tyto jevy mají charakter náhodného činitele (Homola, 2014).

1.5.1 Matematická statistika

Základem matematické statistiky je teorie pravděpodobnosti. Oproti popisné statistice využívá pouze náhodnou část naměřených hodnot, čímž se každý výsledek mírně liší. Pomáhá k rychlejšímu zpracování velkých dat (Homola, 2014).

1.5.2 Teorie pravděpodobnosti

Pravděpodobnost náhodného jevu udává, jaká je šance vzniku daného jevu. Příkladem může být hod hrací kostkou. Jaká je pravděpodobnost, že padne číslo pět? Pravděpodobnost se udává v intervalu mezi nulou a jedničkou nebo pomocí procent (Havrlant, 2014).

1.6 Multidimenzionální databáze

Tato databáze slouží jako podklad pro získání sumarizovaných a agregovaných informací. Ukládají se zde data po upravení a vyčištění, tedy po procesu ETL. Používají se zde dva druhy tabulek - tabulky faktů a tabulky dimenzí (Lacko, 2009).

Mezi výhody patří rychlý komplexní přístup k velkému objemu dat a silná schopnost modelování a prognózy. Jako nevýhodu lze uvést vyšší nároky na kapacitu uložení a problémy při změně dimenzí (Lacko, 2009).

Model multidimenzionální databáze se často označuje pojmem OLAP kostky (online analytical processing).

- MOLAP (multidimensional OLAP) - uložení v multidimenzionálních - binárních OLAP kostkách.
- ROLAP (relational OLAP) - multidimenzionalita se řeší uložení v relačních databázích.
- HOLAP (hybrid OLAP) - kombinace předchozích dvou typů (uložení v relačních databázích, ale agregace jsou v binárních OLAP kostkách)
- DOLAP (desktop OLAP) - umožňuje stáhnout jen podmnožinu OLAP kostky. Dotazy pak probíhají jen nad touto lokální kostkou (Novotný, 2005).

1.6.1 Tabulka faktů

„Fakta jsou numerické měrné jednotky obchodování.“ (Lacko, 2009, s. 176)

Zpravidla to bývá největší tabulka databáze, která obsahuje velký počet často se měnících dat. Obsahuje měrné jednotky (Lacko, 2009).

Některé jednotky jsou přirozeně aditivní (lze je sčítat ve všech dimenzích), jiné jsou semiaditivní (lze je sčítat pouze pro některé dimenze) a ostatní jsou neaditivní (nelze je sčítat podle žádné dimenze) (Laberge, 2012).

1.6.2 Tabulka dimenzí

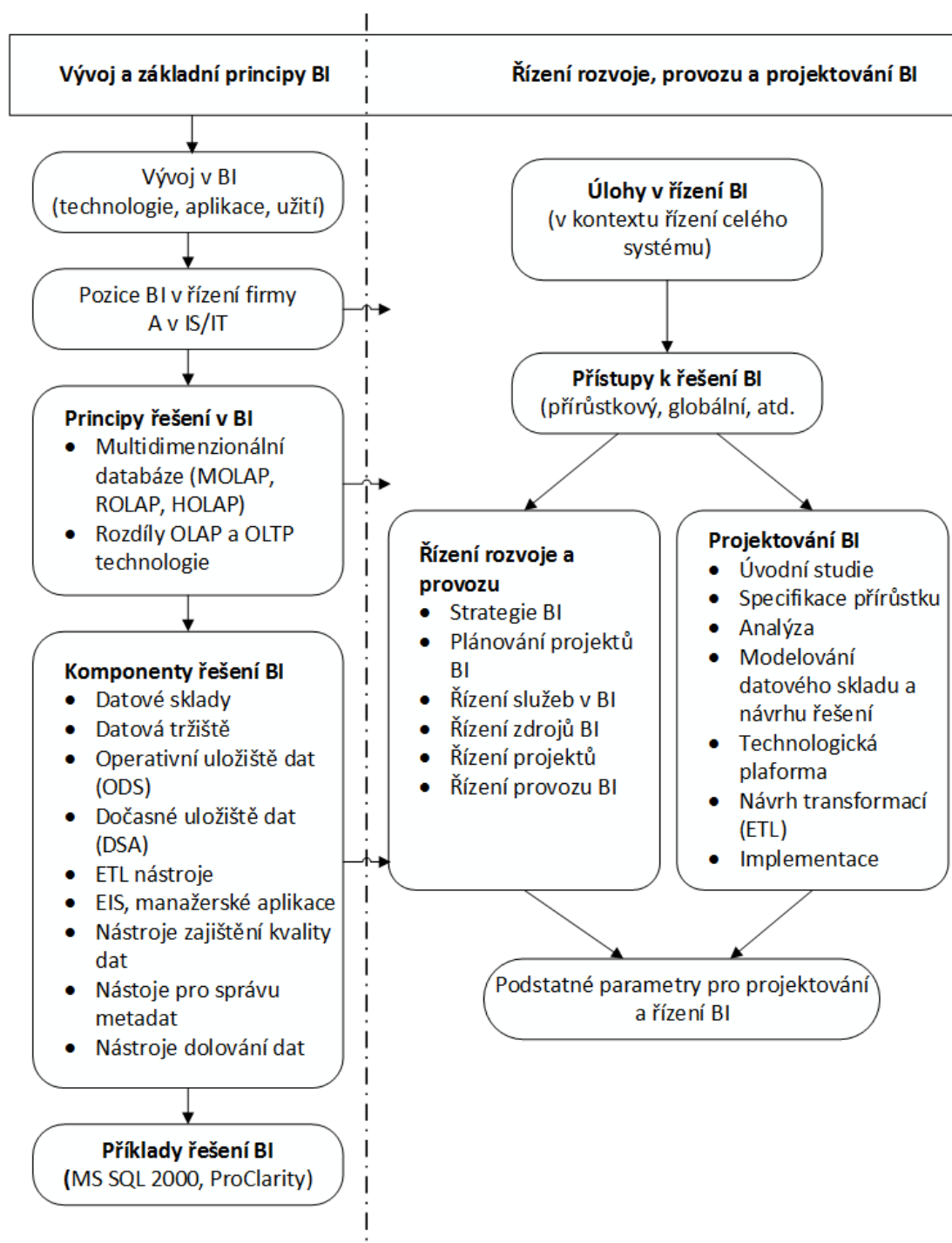
„Dimenze obsahují logicky nebo organizačně hierarchicky uspořádaná data. Jsou to vlastně textové popisy obchodování.“ (Lacko, 2009, s. 177)

Bývají menší než tabulky faktů. Navíc se v nich data nemění příliš často, jelikož obsahují spíše stabilní a neměnná data. Nejčastěji se jedná o časové, geografické nebo produktové dimenze (Lacko, 2009).

1.7 Definice pojmu BI

Technologie, která se vztahuje ke znalostem, procesům, technologiím, aplikacím a postupům, které usnadňují podnikové rozhodování. Pracuje s použitými (historickými) daty v celém kontextu a pomáhá přijímat rozhodnutí managementu pro budoucnost. Soustřeďuje se hlavně na interní informace o provozních činnostech, které souvisejí s taktickým plánováním. Výstup se používá převážně v interní analýze, ale kombinuje se i s externími analýzami typu SWOT nebo PEST (Laberge, 2012).

1.7.1 Základní principy BI



Obrázek 2 Principy business intelligence

Zdroj: přepracováno dle (Novotný, 2005)

1.7.2 Nástroje a architektura BI

Způsob zaznamenávání, čištění, udržování dat a kompatibilita dat s nástroji pro prezentaci je důležitou složkou BI. Proto by měla být data přizpůsobena

pro uchovávání i používání. To znamená, že nástroje BI musí podporovat speciální datová schémata. A jestliže tyto nástroje vyžadují data v jiných formátech, bude na optimalizaci třeba vynaložit větší úsilí. Nástroje jsou tedy nezávislým prvkem (Laberge, 2012).

Technická architektura se dělí na dvě části: back room a front room. První jmenovaná je tou částí DS, kde se odehrávají vnitřní procesy. Druhá část je uživateli naopak přímo viditelná a přístupná (Novotný, 2005).

Back room

Klíčovou vlastností je diskový prostor, kde se data uchovávají z různých systémů a transformují se do podoby OLAP kostek. V architektuře hrají kritickou roli metadata. Ta popisují obsah DS, ukládají, odkud jsou data získávána, a dokumentují pravidla pro ETL (Novotný, 2005).

Front room

Tento prostor si můžeme představit jako prezentační server, na kterém jsou data uložena. Na tento server se uživatelé dotazují, mají k němu přímý přístup. Metadata front room slouží jako slovník obsahu všech datových prvků přístupný uživatelům (Novotný, 2005).

1.7.3 Uživatelská prezentace BI

Výstup BI může mít různé formy:

- Sestavy
- Dotazy
- OLAP
- Ovládací panely
- Přehledy výsledků (Laberge, 2012)

Sestavy jsou statické, obvykle plánovaně spouštěné rutiny, které vytvářejí přehledy. Poskytují informace organizacím, na jejichž základě se provádí rozhodování (Laberge, 2012).

1.8 ETL

Je jednou z nejvýznamnějších komponent BI. Běžně se označuje jako datová pumpa. Data se ze zdrojových systémů získávají a vybírají (extraction), upravují se do požadovaných forem a čistí se (transformation) a nahrávají se do datových skladů (loading). Tyto procesy se používají při převodu mezi různými typy systémů. Data se do DS přenášejí dávkově (batch), tedy např. denně, týdně nebo měsíčně (Novotný, 2005).

1.8.1 Extrakce

Data, potřebná přenést do DS, jsou většinou umístěna v nehomogenních operačních prostředích (různé hardwarové platformy, operační systémy, databázové systémy i podnikové systémy, navíc mohou být data v různých dokumentech). Úkolem extrakce je tato data načíst a předat dalšímu procesu - čištění dat. Velký problém obvykle bývá v historických datech (Lacko, 2009).

1.8.2 Čištění dat

Proces čištění může být velmi náročný a nákladný. Vše se odvíjí od kvality vstupních dat. Je třeba odstranit duplicitní záznamy, nekompletní řádky databáze nebo nečitelná data (Lacko, 2009).

1.8.3 Transformace

Pro transformaci platí, že je lepší nemít žádná data než data nekvalitní. Transformací se data zkvalitní, např. tím, že data převedeme na správné datové typy, zaměníme položky, označující stejnou skutečnost, na jednotné hodnoty (například M, MUŽ, 1, atd.) nebo změníme formu uložení, se kterou se můžeme setkat u poštovních adres. Častou transformací jsou i opravy pravopisných chyb nebo překlepů (Lacko, 2009).

Dále se vyskytují tyto okruhy problémů:

- nejednoznačnost dat,
- nejednotnost peněžních měn,
- problém s referenční integritou,
- chybějící datum (Lacko, 2009).

1.8.4 Loading

Neboli zavedení do DS. Je to završující etapa ETL, což znamená fyzický přenos zdrojových dat do cílové databáze nebo DS. Tento přenos by měl být plánovaný, automatizovaný a optimalizovaný. Po zavedení probíhá indexace, aby byl přístup optimalizovaný. Umělé klíče zajistí jednoznačnost každého řádku v tabulce, jelikož mnoho transformovaných záznamů nemá žádné přirozené klíče (Lacko, 2009).

1.9 Reporting

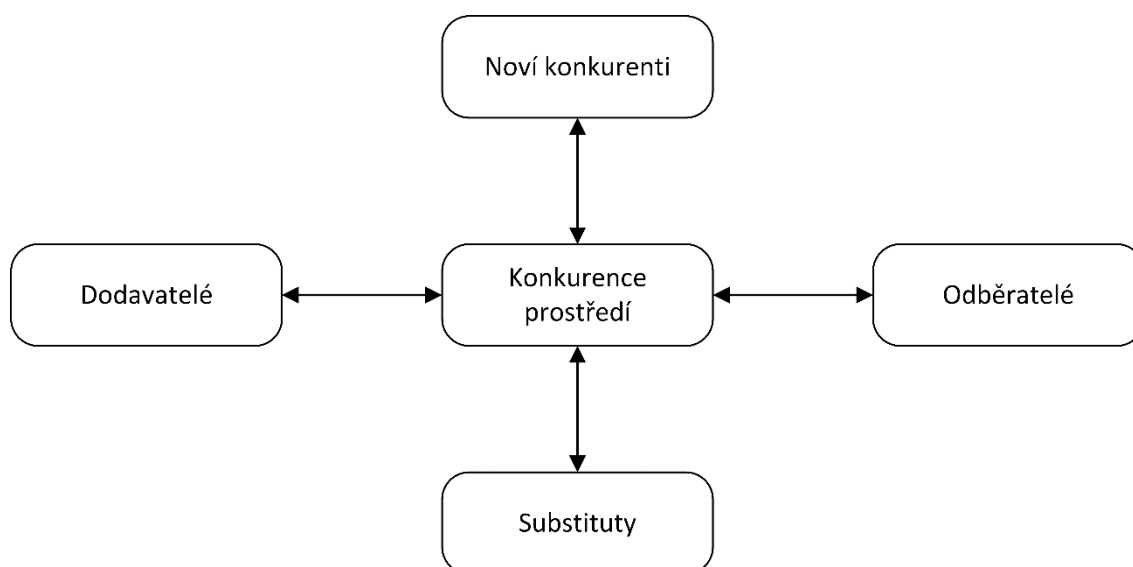
Rozumí se jím činnost spojená s dotazováním do databází. Lze rozlišovat standardní reporting, kdy jsou dotazy spouštěny v určitých periodách nebo ad hoc reporting, ve kterém se formulují specifické dotazy explicitně vytvořené uživatelem (Novotný, 2005).

1.10 Analýzy

1.10.1 Porterův model pěti konkurenčních sil

Tato analýza slouží ke zmapování faktorů, ovlivňujících vyjednávací pozici firmy v různých odvětvích. Předpokladem je strategická konkurenční pozice firmy, která působí v určitém odvětví a je určována především působením následujících pěti sil:

1. Vyjednávací síla dodavatelů
2. Vyjednávací síla odběratelů
3. Hrozba vstupu nových konkurentů
4. Hrozba substitutů
5. Rivalita firem působících na daném trhu (Hanzelková, 2009).



Obrázek 3 Porterův model pěti sil

Zdroj: přepracováno dle (Hanzelková, 2009).

1.10.2 PEST

Analyzuje okolí firmy, především faktory politické a legislativní, ekonomické, sociální a kulturní a technologické. Některé prameny uvádějí zkratku SLEPT, která rozděluje politické a legislativní faktory nebo zkratku PESTEL (SLEPTE), jež ještě přidává faktor ekologický (Sedláčková, 2006).

Všechny skupiny v sobě zahrnují řadu faktorů makro-okolí, které mohou ovlivňovat podnik různými způsoby.

1.10.3 SWOT

Využívá souhrnu předchozích analýz k identifikaci silných a slabých stránek podniku. K nim dále přidává vlivy okolí podniku, tedy příležitosti a hrozby. Nástroj SWOT analýzy je zaměřený na systematickou analýzu a charakteristiku klíčových faktorů, které ovlivňují strategické postavení podniku. Závěrem analýzy je východisko pro formulaci strategie (Sedláčková, 2006).

Jednou z nevýhod je často špatný odhad, zda je určitý jev příležitostí/hrozbou nebo zda určitá charakteristika je silnou/slabou stránkou (Grant, 1998).

Jedním z možných výsledků je čtvercová matice dva na dva, kdy v prvním sloupci jsou hrozby (threats), ve druhém příležitosti (opportunities), v prvním řádku

se uvádí slabé stránky (weaknesses) a v posledním řádku silné stránky (strengths), jak uvádí tabulka 1 (Čadil, 2013).

Tabulka 1 SWOT analýza s možnostmi řešení

	Hrozby (T)	Příležitosti (O)
Slabé stránky (W)	Zlepšení slabých stránek k vyvarování hrozeb	Zlepšení slabých stránek možnou příležitostí
Silné stránky (S)	Využití silných stránek k eliminaci hrozeb	Využití silných stránek k dosažení příležitostí

Zdroj: přepracováno dle (Sedláčková, 2006, s. 94)

1.10.4 Analýza rizik

Cílem této analýzy je minimalizace nebezpečí a zároveň zvýšení šance na úspěch. Mnoho rizik může ovlivnit finanční stabilitu firmy a směřovat ji k úpadku (Keřkovský, 2006).

Cílem analýzy rizika podnikatelského projektu je zjistit:

- Faktory, jež jsou významné a nejvíce ovlivňují riziko daného podnikatelského projektu, případně které faktory lze zanedbat;
- Riziko podnikatelského projektu a zda je toto riziko ještě přijatelné, či je již nepřijatelné;
- Opatření, které může snižovat riziko podnikatelského projektu na přijatelnou míru (Keřkovský, 2006).

Mapa rizik

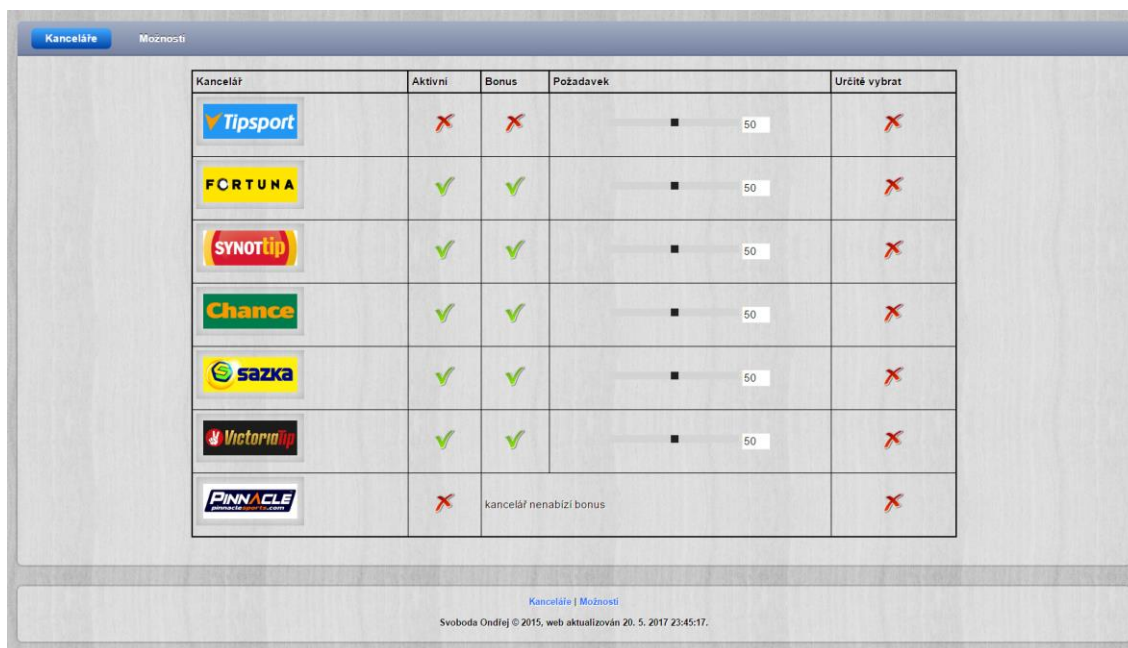
Slouží k identifikaci a grafickému zobrazení. Ve dvourozměrném grafu je vyobrazená pravděpodobnost na ose x a dopad rizika na ose y. Indiferenční křivky v grafu rozdělují rizika na běžná, závažná a kritická (Keřkovský, 2006).

2 ANALÝZA SOUČASNÉHO STAVU

2.1 Historie vývoje aplikace na sázení

Nápad vydělávat na sportovním sázení vznikl již před několika lety. Tehdy se zjišťovaly prvotní informace o tom, jak by takový program mohl vypadat a co všechno by měl umět. Prvotní beta verze aplikace obsahovala kód na získání aktuální kurzové nabídky českých SK a jednoduchý kombinátor všech možností. Samotné velikosti vkladů se musely dopočítávat ve vytvořeném dokumentu Excel. Tato provizorní verze měla za úkol zjistit, jestli vše funguje podle předem stanovené teorie. Po kladném zjištění jsme mohli přejít k další fázi.

Tou byla vylepšená verze programu, kterou jsme obohatili dalšími funkčními prvky. Přibyla důležitá kalkulačka pro rychlé výpočty, zlepšila se grafika a program byl umístěn na web. Jako podřadnější funkce se testovaly samostatné výpočty rozložení vkladů pomocí optimalizace lineárního programování, ukládání vsazených zápasů nebo kontrola denních limitů SK. Tuto verzi programu jsme testovali asi dva roky. Zjistilo se, že některé funkce jsou nepotřebné, a proto se vytvořila nová verze programu.



Kancelář	Aktivní	Bonus	Požadavek	Určité vybrat
	✗	✗	50	✗
	✓	✓	50	✗
	✓	✓	50	✗
	✓	✓	50	✗
	✓	✓	50	✗
	✓	✓	50	✗
	✗	kancelář nenabízí bonus		✗

Kanceláře | Možnosti

Svoboda Ondřej © 2015, web aktualizován 20. 5. 2017 23:45:17.

Obrázek 4 Stará verze programu - přehled SK

Zdroj: vlastní

Té se opět vytvořila nová grafická podoba a celkově se zredukovala na co největší jednoduchost a uživatelskou přívětivost. Převedením do frameworku Nette

získala aplikace větší bezpečnost a díky návrhu „model - view - controller“ i lepší zpracování chyb. Taktéž byly zpracovány návody na sázení pro úplné začátečníky a pokročilé tipy a triky pro zkušenější sázkaře.

Kancelář	Aktivní	Bonus	% bonusu	Zobrazit jen
	✓	✓	20	✗
Podmínky: Bonus 50 000 Kč				
	✓	✓	20	✗
Podmínky: Bonus 50 000 Kč				
	✓	✗	75	✗
	✗	✗	90	✗

Obrázek 5 Nová verze programu - přehled SK

Zdroj: vlastní

Při pohledu na obrázek 5 je vidět nová funkčnost a to výběr různých typů bonusů. Viditelná je i změna v šabloně webu. Je přehlednější a snadnější na ovládání.

2.1.1 Vývoj z hlediska výpočtů

Z hlediska matematiky se zdálo, že vývoj nebude žádný nebo malý. Původní výpočty kurzové hladiny, výpočty rozložení vkladů, velikosti bonusů, celkového zisku a hodnoty zisk/bonusu zůstaly nezměněny. Změněno bylo pouze finální třídění dle více kritérií.

2.2 Popis vytvořeného principu sázení

Aktuální verze programu se zabývá pouze arbitrážním sázením, bonus huntingem nebo kombinací obojího. Tedy vždy se uzavírají sázky na všechny příležitosti dané události. To v sobě skrývá jistá úskalí.

2.2.1 Funkcionalita

Princip sázek, např. u fotbalového zápasu, na domácí, remízu a hosty, spočívá v uzavření poměrných částek na jednotlivé příležitosti (na výhru domácích, výhru hostů a remízu). Zpravidla je jeden z kurzů nadhodnocený a další dva tvoří maximální kurz ze všech SK.

Vše uvedu na příkladu zápasu 1. Anglické ligy Manchester United - Crystal Palace.

Tabulka 2 Nabídka sázkových kanceláří na zápas Manchester United a Crystal Palace

SK	Domáci	Remíza	Hosté	Hladina
Tipsport	2,58	3,46	2,8	96,73 %
Sazkabet	2,8	3,35	2,55	95,44 %
Fortuna	2,61	3,3	2,83	96,2 %
Nejlepší	2,8	3,46	2,83	100,05 %

Zdroj: vlastní

Jak uvádí tabulka 2, uzavřením sázek v rámci jedné SK bychom ztratili průměrně kolem 4 % souhrnu vkladů. To odpovídá marži kanceláře, která se v jednotlivých soutěžích i ligách liší - od 2 % na oblíbené soutěže až po 10 % na nižší ligy. Ale při kombinaci tří SK s nejlepšími kurzy se dostáváme na hladinu nad 100 % a zisk se rovná souhrnu vkladů krát hladina podělená stem. Problémem v tomto případě je to, že nadhodnocený kurz je pouze jeden. Sazkabet vypsala na výhru domácích kurz 2,8. Řekněme, že průměr všech ostatních kanceláří je 2,6. Poměrově je toto nadhodnocení 7,7 %. Když se na to samé zaměříme u remízy a hostů, bude nadhodnocení daleko menší - buď záporné nebo řádově jednotky procent. Dlouhodobě tedy 0,05 %. Názorně lze vidět, že pro jistý výdělek se musíme vzdát části zisku, jelikož si ho pokryjeme na nevýhodných kurzech.

Řešením je toto eliminovat a vsázet jen nadhodnocené kurzy. Krátkodobě se sice vzdáme jistého zisku, ale dlouhodobě budeme profitovat na zisku větším. Je to tím, že kurz má tzv. value - reálná pravděpodobnost vzniku příležitosti je větší, než na jakou pravděpodobnost je kurz vypsán.

2.3 Datové zdroje

Zdroje, ze kterých bychom mohli čerpat, je několik. Prvním jsou uložená data ze starého systému na sázení, dalším jsou data z nového systému a posledním jsou informace ze stránek, které shromažďují výsledky zápasů.

2.3.1 Data ze systému na sázení

Starší data byla uložena v databázi, ale pro její zrušení a nahrazení databází novou se exportovaly do formy excelovského souboru. Nová data jsou v nové databázi, ale nekompatibilitou MySQL databází s Microsoft nástroji budou také převedena do souboru excel. Oddělovače jsou v obou případech středníky.

Data obsahují informace o staženém kurzu (jeho velikost, datum stažení, ID kanceláře, datum a čas události, název události, ligu a sport a informaci, o jaký typ kurzu jde).

Existují dvě databáze. Jedna starší, která má mírně odlišnou strukturu než druhá, novější.

2.3.2 Data stažená ze stránek s výsledky

Tato data jsou stažená v surovém HTML. To se dále bude parsovat na informace o zápase a lize a o výsledku události.

2.4 Zpracování analýz

Pro potřeby dalšího zpracovávání je třeba nejprve provést analýzy. Ty by měly odhalit skutečný problém, který budeme řešit v následující části DP - ve vlastním návrhu řešení.

2.4.1 Analýza porter

Substituty

Substitutem mohou být všechny softwarové systémy sloužící k tvorbě zisku. Na trhu jich několik figuruje. Nejznámější z nich je Rebelbetting, který slouží především k hledání arbitráží. Program byl spuštěn v roce 2007 a v prvním roce používání přinesl pro své zákazníky zhodnocení 12 %. Problém začal vznikat ve chvíli, kdy program využívaly tisíce zákazníků včetně samotných SK. Nadhodnocené kurzy

z arbitráží se začaly vytrácet nebo se zkrátila doba potřebná k uzavření sázky na minimum. Dalším substitutem je program společnost Winnersclub, který byl spuštěn v roce 2015. Ten funguje na podobném principu jako námi vytvořený program. Rozdíl je pouze v tom, že náš software pracuje s českými kanceláři.

Z pohledu investice na trhu sportovních událostí je substitutem i jakákoliv finanční činnost, zajišťující budoucí zhodnocení. Například investice do akcií, spoření nebo připojištění.

Vyjednávací síla dodavatelů

Náš program potřebuje ke své funkčnosti pouze několik dodavatelů. Jedním z nich je hostingový operátor a pronajímatel virtuálních serverů. V našem případě se jedná o společnost WEDOS Internet, a.s. Ta je řízena dokonalou konkurencí, takže z její strany je vyjednávací síla slabá. Další dodavatele prozatím nemáme.

Vyjednávací síla odběratelů

Odběrateli našeho programu jsou především studenti s dostatečným množstvím financí nebo lidé, kteří chtějí lepší zhodnocení finančních prostředků a nevdají jim jejich občasné přerozdělení. Jejich vyjednávací síla je slabá, protože z pozice novinky na trhu si můžeme zkusit zavést monopolní ceny.

Stávající konkurenti

Přímý konkurent, tedy ten, který by nabízel stejnou službu na stejném trhu, zatím není. Existuje pouze několik podobných, z nichž dva zmínění konkurenti, program Rebelbetting a společnost Winnersclub, jsou pouze nepřímým konkurentem. Figurují na stejném trhu, ale nabízejí pouze podobnou službu. Náš program se zaměřuje na sběr bonusů a arbitráží především na českém trhu. Konkurence nabízí sběr bonusů nebo arbitráží na trhu zahraničním. Zákazníci využívající konkurenci jsou především z Velké Británie.

Každý uživatel, i samotný bookmaker, který rozumí konkrétnímu sportu, může určit správnost kurzu. Tím mu určuje value a je možnost na něj investovat. Tedy každý člověk, zabývající se sázením na hodnotu value, investuje a současně je našim konkurentem. Rozdíl s těmito lidmi a námi je v rozdílnosti určování value.

Riziko vstupu nového konkurenta

Vzhledem k tomu, že náš současný program vyvíjíme cca 3 roky, riziko vstupu nového konkurenta je den ode dne vyšší. Již před dvěma lety vstoupil na trh nový konkurent - Winnersclub. Ten sice nabízí pouze podobnou službu a cílí na české zákazníky, ale na zahraniční kanceláře. Winnersclub letos narazil na legislativní problémy, kdy nový Loterijní zákon zakázal zahraničním kancelářím působit na českém trhu bez potřebných povolení.

2.4.2 PEST

Politické a legislativní faktory

V prezidentských volbách konaných roku 2018 byl jedním z kandidátů zakladatel společnosti Fortuna, Michal Horáček. Pro zvolení do prezidentského úřadu měl dle kurzu Fortuny šanci asi 12,5 %. Prezident Miloš Zeman měl za stejných okolností více než šedesátiprocentní šanci obhájit své křeslo. Prezident Miloš Zeman na trhy sportovních událostí moc nesází, ale kdyby se prezidentem ČR stal Michal Horáček, mohl by se očekávat nějaký zásah. Jestli v kladném nebo záporném smyslu se nedá určit.

Legislativa nového Loterijního zákona, platného od 1. 1. 2017, ukládá povinnost požádat o licenci. K lednu v ČR skončily mnohé zahraniční kanceláře. Jiné si požádaly o licence, které do poloviny května 2017 nedostaly. Profesionálním sázkařům v ČR chybí společnost Pinnacle sports a Bet 365, které mají jedny z nejlepších kurzových hladin na světě. Tímto se kurzové sázky mírně komplikují.

Zákon o hazardu naštěstí nezvyšuje daň na kurzové sázení, ale pouze na výherní automaty. Ovšem v roce 2014 se daň zvýšila na 23 %, což se projevilo v kurzech.

Ekonomické faktory

Čeští sázkaři prosázeli za rok 2016 jednu z největších částek z celé Evropy, 196,4 mld. korun. Meziročně je to o 29 % více. Na výhrách bylo zpětně vyplaceno 157,1 mld. korun, to je o 29 % více než v roce 2015. Kurzové sázky si z této sumy ukrojily 57,4 mld. Kč, přičemž 87,8 % je příjem z internetového sázení, zbytek připadá na sázky uzavřené na pobočkách. Z tohoto článku je parné, že internet je stále mocnější médium pro uzavírání sázek (ČTK, 2017).

Sociální a kulturní faktory

Do budoucna má vznikat rejstřík všech hráčů. Důvodem je zabránění hraní nezletilým osobám, ale i lidem, kteří pobírají sociální dávky.

Technologické faktory

S přibývajícími sázkaři, kteří využívají bonusů nebo výhodných kurzů SK k pokrytí svých příležitostí, přibývá také blokování těchto uživatelů na fyzické úrovni. Jednou z možností je využití umělé inteligence, která pomocí shlukové analýzy může identifikovat podvodného hráče. Jestliže je skoro 90 % sázkařů na internetu, neměl by to pro SK být problém. Chyba, že zablokují někoho, kdo ve skutečnosti nepodváděl, je zanedbatelná, a proto čím dál více kanceláří přistupuje na účinnou obranu.

2.4.3 Analýza rizik

Při vstupu na nový trh je pro firmu důležité zaujmout potenciální zákazníky. Také moment překvapení hraje významnou roli. Veškeré prodlevy při spouštění nového softwaru mohou být likvidační. Proto i riziko vstupu nového konkurenta na trh by mělo velký dopad. Nový konkurent zatím neexistuje, ale pravděpodobnost vstupu se s plynoucím časem zvedá. Toto riziko spadá do kategorie kritické a je třeba ho řešit.

2.4.4 Analýza SWOT

Příležitosti

- Využití momentu překvapení
- Investiční typ sázení

Hrozby

- Vstup nového konkurenta

Silné stránky

- Spousta volného času pro vývoj programu

Slabé stránky

- Špatná domluva mezi programátory a zakládajícím členem

3 VLASTNÍ NÁVRH ŘEŠENÍ

Jako nejdůležitější a nejzajímavější se dle analýz jeví zpracování investičního typu sázení. Tento typ sázení může přinést daleko větší zisky, jelikož se sázky nemusejí uzavírat na nevýhodných kurzech.

Jako první vybuduji datový sklad, do kterého se nahrají data z různých zdrojů pomocí ETL procesu. Pro potřeby reportu budu provádět různé operace s daty, k tomu taktéž využiji ETL proces, který mi přetvoří data do potřebné formy. Jako poslední se budu zabývat vytvořením reportů a přehledů, kterými zhodnotím možnosti investování na trhu sportovních událostí.

3.1 Vybudování datového skladu

Datový sklad bude vytvořen pouze jednou, ale data se do něj budou přidávat v pravidelných intervalech. Všechny zdroje dat se přetransformují tak, aby byly v DS uloženy jednotně a se všemi potřebnými hodnotami. Jako metodu budování DS jsem zvolil „velký třesk“. To proto, že informace z dat potřebuji zjistit pouze jednou. Následné informace využiji k naprogramování aplikace, která bude v reálném čase stahovat a vyhodnocovat, které kurzy jsou nadhodnocené a které ne.

DS vybuduji v Microsoft SQL Server 2014. Ten koresponduje s ETL procesy v programu Microsoft Visual Studio Profesional 2013, je svižný a pro potřeby studia je zdarma.

Topologií DS bude sněhová vločka. Je nejvhodnější z hlediska čistoty ukládaných dat a nejvíce koresponduje s realitou modelového návrhu.

3.1.1 Tabulka faktů

Zde jsou veškeré informace ohledně kurzové nabídky. Hodnoty kurzu se udávají v několika typech, ale všechny budou převedeny na desetinný formát. Proto atribut hodnota bude typem FLOAT nebo REAL. Ostatní atributy jsou cizí klíče na tabulky dimenzí, ty mají stejné datové typy a velikosti datových typů jako primární klíče tabulek dimenzí. Časový údaj je datum a čas stažení jednotlivých kurzů.

Tabulka 3 Fakta - KURZ

KURZ			
atribut	typ	hodnoty	popis
id_zapas	INT(6)	celá čísla	CK ZAPAS
id_kancelar	INT(2)	celá čísla	CK KANCELAR
id_datum_cas	INT(10)	celá čísla	CK DATUM_CAS
id_typ_kurzu	INT(3)	celá čísla	CK TYP_KURZ
hodnota	FLOAT	desetinná čísla	

Zdroj: vlastní

Tabulka KURZ bude naplněna procesem ETL.

3.1.2 Tabulky dimenzí

Pro potřeby celistvosti DS, s ohledem na potřeby dolování dat, jsem zvolil následující čtyři tabulky dimenzí a jejich pod-dimenze.

Dimenze ZÁPAS

V této dimenzi jsou veškeré informace o druhu zápasu. Jsou zde atributy domácí a hoste, jejichž datový typ je VARCHAR(50). Atribut `vysledek` označuje, jak konkrétní zápas dopadl. Jeho typ je VARCHAR(1). Posledním atributem je cizí klíč `id_soutez` odkazující na dimenzi SOUTĚŽ.

Tabulka 4 Dimenze - ZÁPAS

ZÁPAS			
atribut	typ	hodnoty	popis
id_zapas	INT(6)	celá čísla	PK
id_datum_cas	INT(10)	celá čísla	CK DATUM_CAS
id_soutez	INT(2)	celá čísla	CK SOUTEZ
domaci	VARCHAR(40)	{Manchester United}	
hoste	VARCHAR(40)	{Crystal Palace}	
vysledek	VARCHAR(1)	{0,1,2}	

Zdroj: vlastní

Dimenzi ZÁPAS naplní proces ETL, kde se provede, ze současného vzoru domácí - hosté, rozdělení dle lomítka na atributy domaci a hoste. Atribut vysledek bude naplněn porovnáním dvou čísel z HTML kódu a vyhodnocením, který z týmů vyhrál nebo zdali to byla remíza. Tento proces extrahování je popsán níže.

Dimenze SOUTĚŽ, SPORT A LIGA

Tyto tři dimenze ke každému zápasu určují, jaká je to liga a ve kterém sportu se hraje. Datové typy a příklad hodnot uvádí následující tabulka.

Tabulka 5 Dimenze SOUTĚŽ a pod-dimenze SPORT A LIGA

SOUTĚŽ - SPORT, LIGA			
atribut	typ	hodnoty	popis
id_soutez	INT(6)	celá čísla	PK
id_sport	INT(10)	celá čísla	CK SPORT
id_liga	INT(2)	celá čísla	CK LIGA
nazev	VARCHAR(50)	{Fotbal, 1. Česká liga}	

Zdroj: vlastní

Dimenze SPORT bude předvyplněna manuálně. Prozatím budou stačit parametry fotbal, hokej a basketbal. Později může přibýt tenis a další sporty.

Dimenze LIGA bude taktéž předvyplněna a to prvními fotbalovými ligami jednotlivých zemí (Španělsko, Portugalsko, Anglie, Itálie, Francie, Brazílie, Česká republika, Belgie, Nizozemí, Německo a Turecko), druhými ligami (Anglie, Francie, Německo, Itálie a Česká republika), hokejem (NHL, KHL a česká Extraliga) a basketbalem (NBA).

Spojení pod-dimenzí LIGA a SPORT v dimenzi SOUTĚŽ proběhne dle potřeby v různých kombinacích.

Dimenze DATUM A ČAS

Tato dimenze slouží pro snadnější určení parametrů jako jsou: rok, měsíc, den, měsíc v konkrétním roku, týden v konkrétním roku a datum bez času. Protože budou potřebné i časové údaje, přidal jsem i tyto hodnoty samotné a také celé datum a čas.

Tabulka 6 Dimenze DATUM A ČAS

DATUM A ČAS			
atribut	typ	hodnoty	popis
id_datum_cas	INT(10)	celá čísla	PK
rok	INT(4)	{2015-2019}	
mesic	INT(2)	{1-12}	
den	INT(2)	{1-31}	
rok_mesic	INT(6)	{201501-201912}	
rok_tyden	INT(6)	{201501-201952}	
datum	DATE	{2015-01-01}	
cas	TIME	{14:30:00}	
datum_cas	DATETIME	{2015-01-01 14:30:00}	

Zdroj: vlastní

Tato dimenze bude naplněna postupně při procesu ETL, jelikož by se využívalo jen málo řádků - například všechny ranní hodiny. Jako id_datum_cas bude použito číslo ve formátu „RRRRMMDDHHII“, které je jednoznačné pro každý řádek. „II“ značí

minuty a „MM“ označuje jednotlivé měsíce. Číselná data jsou typu integer, aby se správně seřadily.

Dimenze TYP KURZU, MOŽNOST, PŘÍLEŽITOST

Zde se budou ukládat jednotlivé typy kurzů a příležitostí. Například můžeme vsadit na možnost domácí, remízu nebo hosty, což je příležitost 1X2. Nebo můžeme vsadit na příležitost počet gólů a možnost více než 3,5 gólů. Pro tuto potřebu by dimenze vůbec nemusela existovat, jelikož ve stažených kurzech se objevuje jen příležitost 1X2. Při přidání ostatních příležitostí by se DS musel budovat znova, což toto preventivní opatření eliminuje.

Tabulka 7 Dimenze TYP KURZU a pod-dimenze MOŽNOST a PŘÍLEŽITOST

TYP KURZU - MOŽNOST, PŘÍLEŽITOST			
atribut	typ	hodnoty	popis
id_typ_kurz	INT(6)	celá čísla	PK
id_moznost	INT(2)	celá čísla	CK MOZNOST
id_prilezitost	INT(2)	celá čísla	CK PRILEZITOST
nazev	VARCHAR(50)	{Zápas bez prodl.}	
zkratka	VARCHAR(35)	{1X2, 1, 0, 2}	

Zdroj: vlastní

Prozatím bude pod-dimenze PŘÍLEŽITOST naplněna pouze jednou hodnotou (1X2), ale později přibudou další (počet gólů, počet žlutých nebo červených karet, počet rohů či přímých kopů a další).

Pod-dimenze MOŽNOST bude taktéž obsahovat pouze výhru domácích, remízu nebo výhru hostů (1, 0, 2). V dalších etapách přibude (neprohra domácích, neprohra hostů, nad x gólů, pod x gólů a další).

Dimenze TYP KURZU bude vyplněna v závislosti na potřebách jednotlivých kombinací příležitostí a možností.

Dimenze KANCELÁŘ

V této dimenzi se ukládají informace o sázkových kancelářích. Pro účel zpracování dat bude stačit pouze jejich název, v atributu nazev.

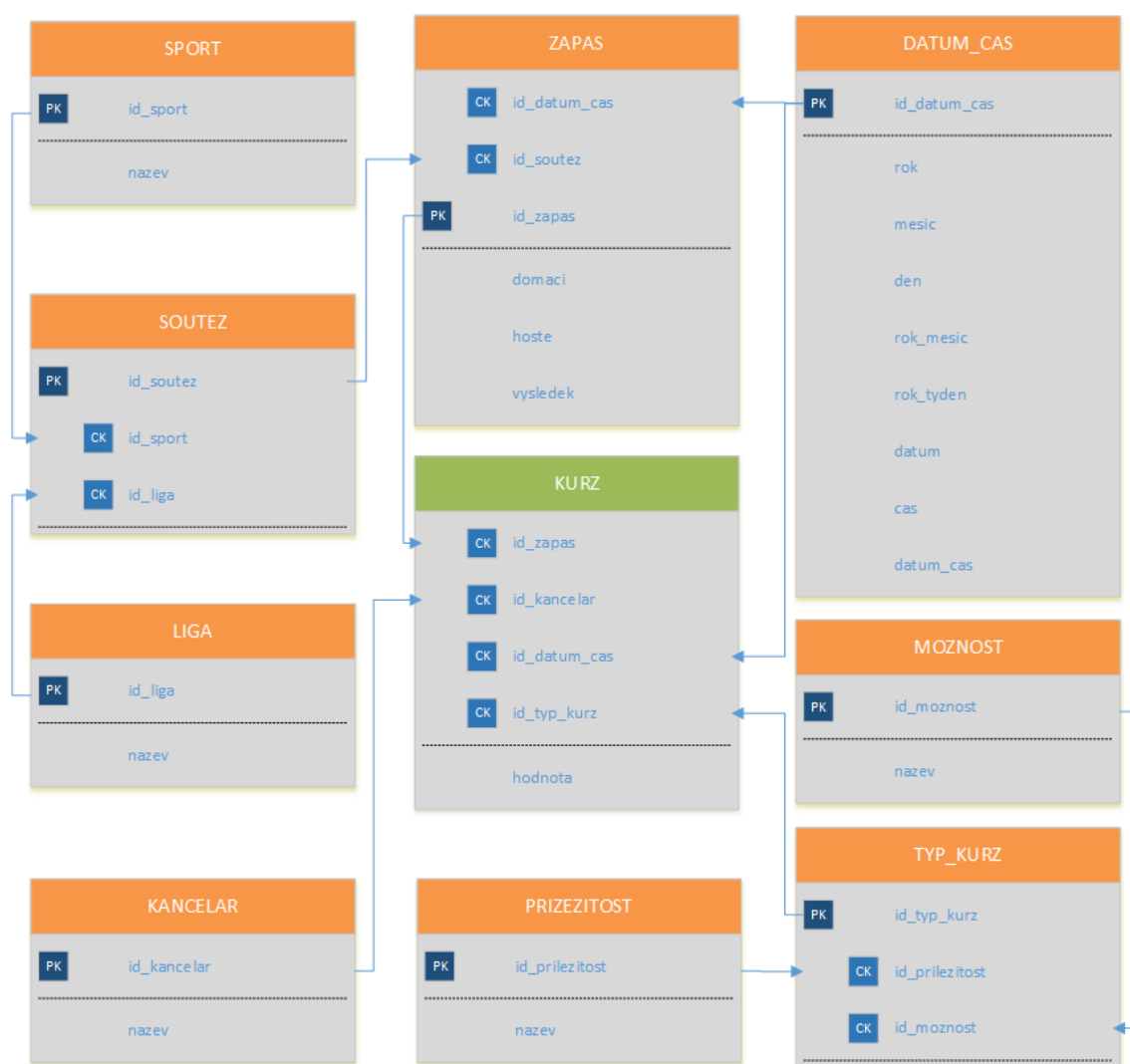
Tabulka 8 Dimenze KANCELÁŘ

KANCELÁŘ			
atribut	typ	hodnoty	popis
id_kancelar	INT(2)	celá čísla	PK
nazev	VARCHAR(30)	{ Tipsport, Fortuna }	

Zdroj: vlastní

Tuto dimenzi bude tvořit základních šest českých kanceláří (Tipsport, Fortuna, Synottip, Chance, Sazkabet a Victoriaticip), čtyři zahraniční kanceláře (Pinnacle Sports, Bet 365, Bwin, Marathonbet) a jedna burza sportovních událostí (Matchbook). Atribut id_kancelar bude přidělen jednotlivým kancelářím dle předchozího zvyku.

3.1.3 Relační návrh DS



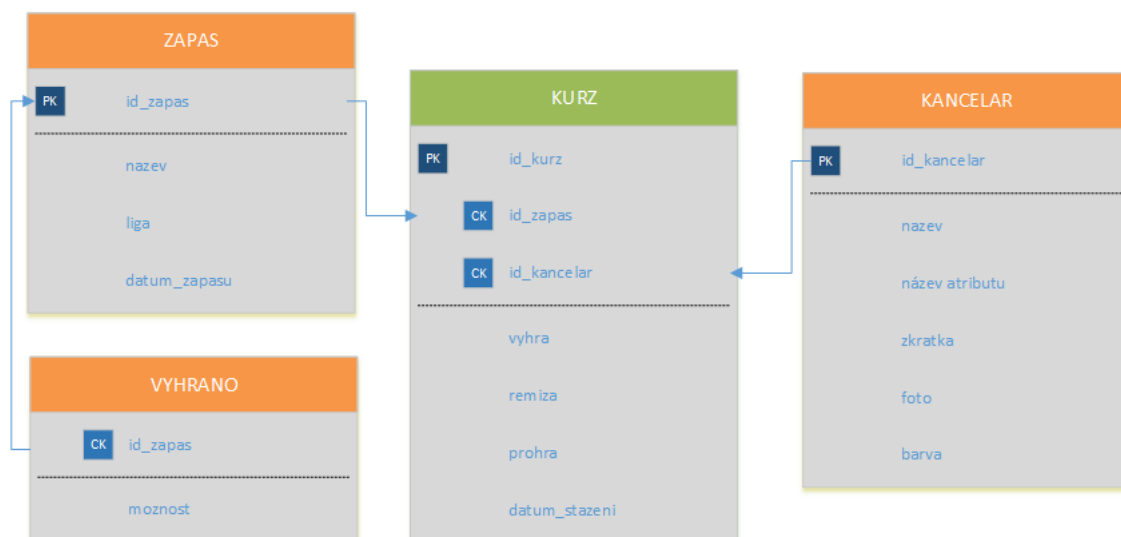
Obrázek 6 E-R diagram DS

Zdroj: vlastní

Obrázek 6 zobrazuje relační propojení tabulky faktů a tabulek dimenzí.

3.1.4 Extrahování dat ze staré DB

Data ve starší verzi databáze jsou uložena v podobném formátu, takže jejich extrakce nebude náročná. Všechna potřebná data budou převedena do tabulek excel. Je to jednak proto, že Microsoft Visual Studio neumí kvalitně spolupracovat s MySQL, ale také proto, že webhostingový operátor neumožňuje přístup k databázím ze třetích stran.



Obrázek 7 E-R diagram staré DB

Zdroj: vlastní

Na této databázi bylo provedeno několik záloh a každá z nich bude mít svůj soubor.

Tabulka KURZ

Jak ukazuje obrázek 7, byla tabulka KURZ navržena špatným způsobem. Žádný jiný typ kurzu by do ní nebylo možné importovat. Extrakce dat se rozdělí na atributy vyhra, remiza a prohra.

Transformuje se typ kurzu dle ID položky 1X2 z dimenze TYP KURZU a atribut datum_stazeni do dimenze DATUM A ČAS.

Celkem se do ETL procesu dostanou tři soubory z tabulky KURZ.

Tabulka ZÁPAS

Z tabulky ZÁPAS se rozdělí atribut nazev dle pomlčky a atribut liga se transformuje do podoby ID, které zjistíme z dimenze SOUTĚŽ.

Položky z tabulky VYHRÁNO se objeví v dimenzi ZÁPAS v atributu vysledek.

V procesu ETL budou dva soubory z tabulky ZÁPAS.

3.1.5 Extrahování dat z nové DB

Pro potřeby kontroly dat jsou důležité kurzy z kanceláře Pinnacle Sports. Zjistil jsem, že v nové databázi žádné kurzy této kanceláře nejsou. Žádná z tabulek nové DB nebude do procesu ETL zasahovat.

3.1.6 Extrahování z HTML stránek

Jedním z náročnějších úkolů byla extrakce dat ze surového HTML kódu. Samotné soubory byly staženy ze stránek oddsportal.com, které nabízejí výsledky a archivaci sportovních událostí. Těchto souborů bylo asi 300 a všechny se musely stahovat ručně. Strojové stahování nebylo možné z důvodu zabezpečení pomocí javascriptu.

Pro získání dat ze souborů HTML byla použita knihovna PHP Simple HTML DOM Parser. (Chen, 2015)

Kód programu obsahuje PŘÍLOHA. Funguje na principu procházení jednotlivých souborů, kdy u každého hledá jednotlivé zpřesňující atributy, ze kterých extrahuje potřebné informace.

Součástí programu je taky převodní tabulka názvů týmů - mapa týmů a jednotlivých lig - mapa lig.

Mapa týmů

Jde o soubor excel, ve kterém byly vygenerovány týmy z DB a týmy z HTML stránek. Ty se poté ručně přepsaly tak, aby z převážně anglických názvů týmů bylo možné vygenerovat názvy převážně české. Příklad uvádí tabulka 9.

Tabulka 9 Ukázka mapy týmů

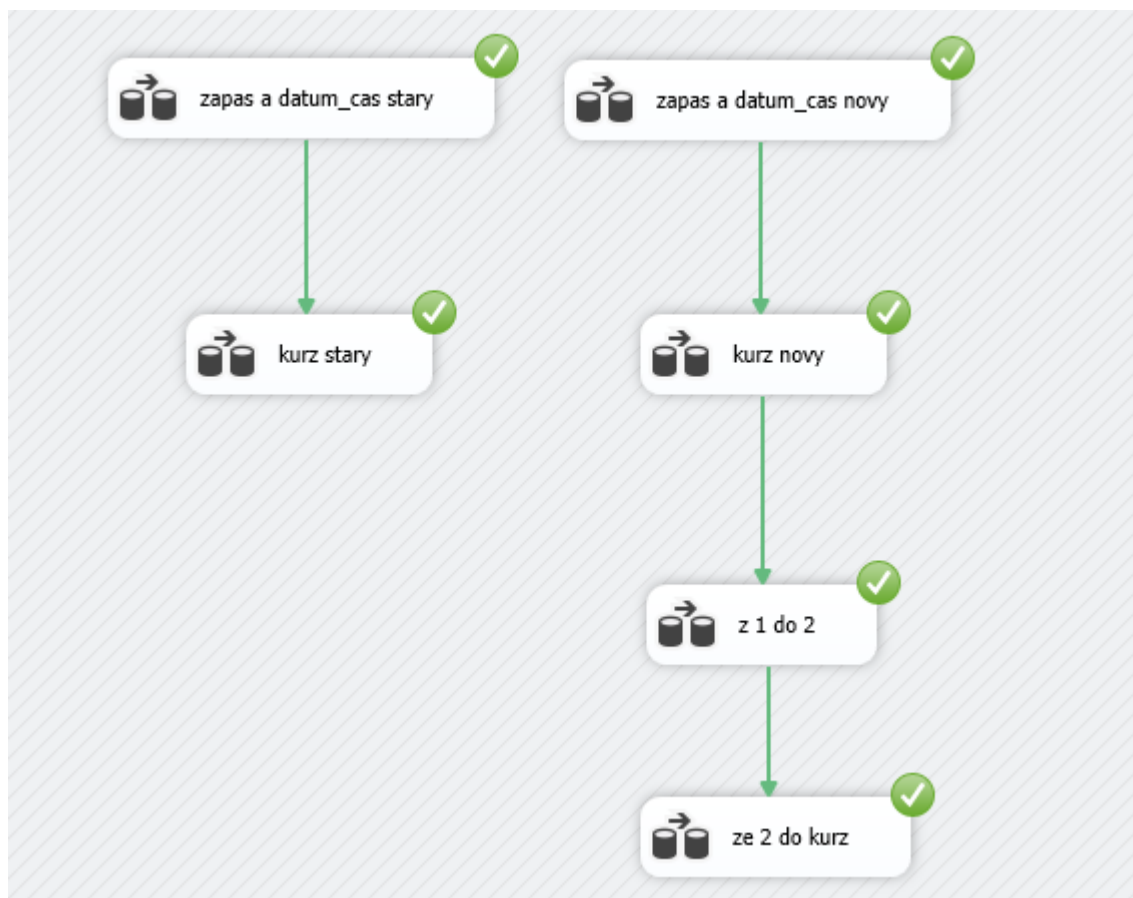
tým anglicky	tým česky	liga
Atlanta Hawks	Atlanta	Basketbal-NBA
Arsenal	Arsenal	Fotbal-Anglicka liga
St. Liege	Standart Lutych	Fotbal-Belgická liga
Karvina	Karvina	Fotbal-Ceska liga
Atletico-PR	Atletico PR	Fotbal-Brazilska liga
Mas Taborsko	Taborsko	Fotbal-Druha ceska liga
Dyn. Kiev	Dynamo Kyjev	Fotbal-Liga mistru
Atl. Madrid	Atl.Madrid	Fotbal-Spanelska liga
PSV	PSV Eindhoven	Fotbal-Nizozemska liga
Dep. La Coruna	La Coruna	Fotbal-Spanelska liga
Adanaspor AS	Adanaspor	Fotbal-Turecka liga
Mountfield HK	Hradec Kralove	Hokej-Extraliga
Tractor Chelyabinsk	Celjabinsk	Hokej-KHL
New York Rangers	NY Rangers	Hokej-NHL
APOEL	Apoel Nicosia	Fotbal-Evropska liga

Zdroj: vlastní

3.2 Procesy ETL

Pro procesy ETL jsem využil program Microsoft Visual Studio a jeho doplněk SSIS (SQL Server Integration Services), který dokáže zpracovávat data dávkově. Tato část zpracování je z hlediska času nejnáročnější. Počet zpracovaných řádků je v desítkách miliónů, a proto jsem mnohdy musel přistoupit ke komplikovanějšímu, avšak rychlejšímu zpracování. Jak ukazuje obrázek 8, zpracování dat je rozděleno na dvě části. V první části se zpracovávají starší data ze staré databáze a v části druhé ty novější. Jelikož bylo z nové databáze načítáno přes 10 mil. záznamů, musel jsem postupovat poněkud odlišně oproti staré DB. Třídít tolik záznamů je i v dnešní době

pro počítač náročné. Odhadoval jsem, že by takové třídění trvalo několik hodin. Ovšem databáze si s tímto problémem poradí rychleji, a proto jsem vytvořil dvě pomocné tabulky, ve kterých jsem pak data utřídil.



Obrázek 8 Celkový data-flow

Zdroj: vlastní

V procesu ETL používám několik nástrojů, které data různě načítají, transformují, spojují a nakonec zapisují.

3.2.1 Source (Zdroj dat)

Tento proces zajišťuje správné načtení zdrojových dat. Lze zde měnit znaková data, oddělovače, datové typy, přidávat nebo ubírat sloupce. V této práci použiju pouze zdroj z DB a zdroj z textového souboru.

3.2.2 Sort (seřazení)

Proces zajistí správné seřazení a umí, pro další potřeby zpracování, vymazat duplicitní záznamy.

3.2.3 Merge Join (spojení dvou tabulek)

Zajišťuje spojení tabulek podobně jako v SQL - levé spojení, vnitřní spojení (průnik) a vnější spojení. Data musejí být před tímto procesem seřazena dle stejného klíče.

3.2.4 Union All (horizontální spojení)

Spojuje dvě tabulky, ale ne sloupcově, nýbrž řádkově. Dva podobné zdroje dat sjednotí do jednoho. Podmínkou jsou stejné datové typy v obou tabulkách.

3.2.5 Conditional Split (rozdělení dle podmínky)

Rozdělí data na menší soubory dle různých kritérií. Nejvíce využívaná je podmínka, zdali jsou, respektive nejsou, data nulová (hodnota NULL).

3.2.6 Derivated Column (úpravy sloupců)

Slouží pro různé funkční úpravy dat. Obsahuje funkce rok(), mesic() nebo den(), pro datum a čas, funkce zleva(), zprava() pro texty.

Ukázka dvou funkcí. První z data dělá identifikační klíč pro dimenzi DATUM_ČAS, druhá pro parsování domácích a hostů z názvu zápasu:

```
(YEAR(datum_stazeni) - 2000) * 100000000 + MONTH(datum_stazeni) *  
1000000 + DAY(datum_stazeni) * 10000 + DATEPART("hh",datum_stazeni) * 100  
+ DATEPART("mi",datum_stazeni)
```

```
RIGHT(nazev, LEN(nazev) - FINDSTRING(nazev, "-", 1))
```

3.2.7 Pivot a Unpivot

Procesy, které převádí stejné atributy řádků na sloupce a naopak.

3.2.8 Percentage Sampling

Tento proces nemá vliv na čisté zpracování. Je zde kvůli zkrácení času při načítání velkých souborů, kdy propustí pouze určité procento vstupních dat a to buď vzestupně, nebo náhodně.

3.2.9 Destination

Stejně jako Source data načítá, Destination je naopak ukládá. V této práci data ukládám do SQL databáze, konkrétně do datového skladu.

3.3 Nahrání do DS

Obrázek 9 ukazuje procesy zpracování dimenze ZÁPAS a DATUM_ČAS. Jako první se odstraňují duplicitní záznamy. Dále se data seřazují a to dle identifikačních atributů liga, domácí, hosté a datum_zápasu. To je potřeba pro spojení dvou tabulek, v tomto případě se k datům o zápase přidá id_soutěže. Proces se opakuje znova i pro spojení s tabulkou VYHRÁNO, kde se přidává atribut výsledek. V následujícím procesu se oddělují neplatné hodnoty řádků - NULL a také proběhne první datová úprava, aby byla data správně připravena pro uložení do DB. Abychom zjistili, které zápasy jsou platné, je třeba k nim přiřadit i kurzové hodnoty. Ty se připravují obdobně jako v případě zápasu. Ke konci procesu se vytvoří dva záznamy do dimenze DATUM_ČAS. Jeden je datum konání událostí a druhý je datum stažení kurzů. Ty obsahují všechny potřebné atributy - rok, měsíc, den, týden v roce, měsíc v roce, čas a další.

Aby nebyly porušeny vazby mezi tabulkou faktů a dimenzí ZÁPAS, musel jsem vytvořit pomocnou proměnnou. Ta potom spojila jednotlivé řádky kurzů s řádky zápasů. Jak je vidět, obrázek 10 ukazuje odlišný proces ETL. Jednak se načítá id_zapas z databáze a následně se spojí, ale hlavně je zde vidět převodník unpivot, který převede některé řádky tabulky na sloupce, jako uvádí tyto dvě tabulky:

Tabulka 10 Příklad zdrojové tabulky

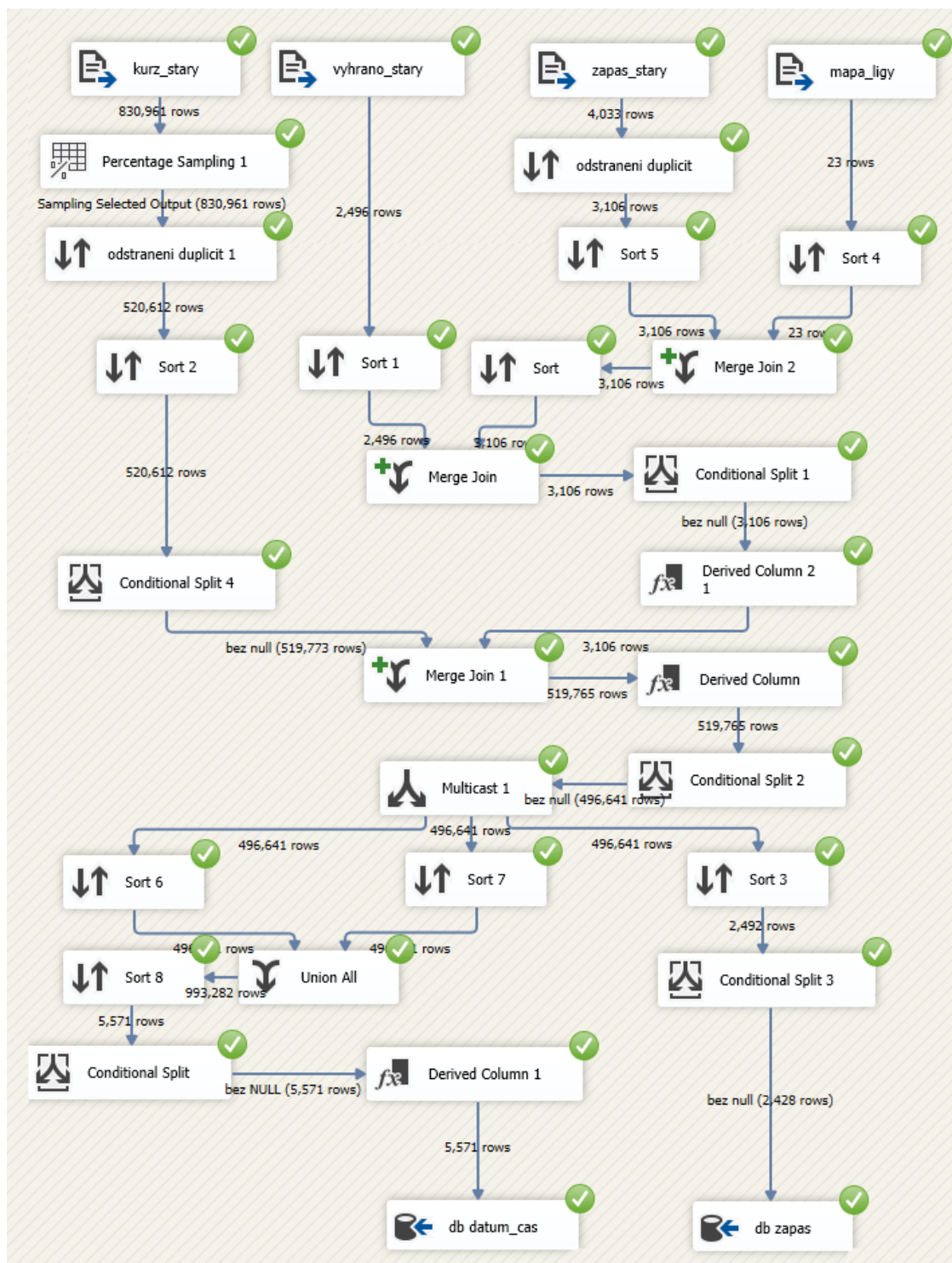
id_zapas	id_kancelar	1	0	2
1	1	1,9	3,45	3,85

Zdroj: vlastní

Tabulka 11 Příklad cílové tabulky

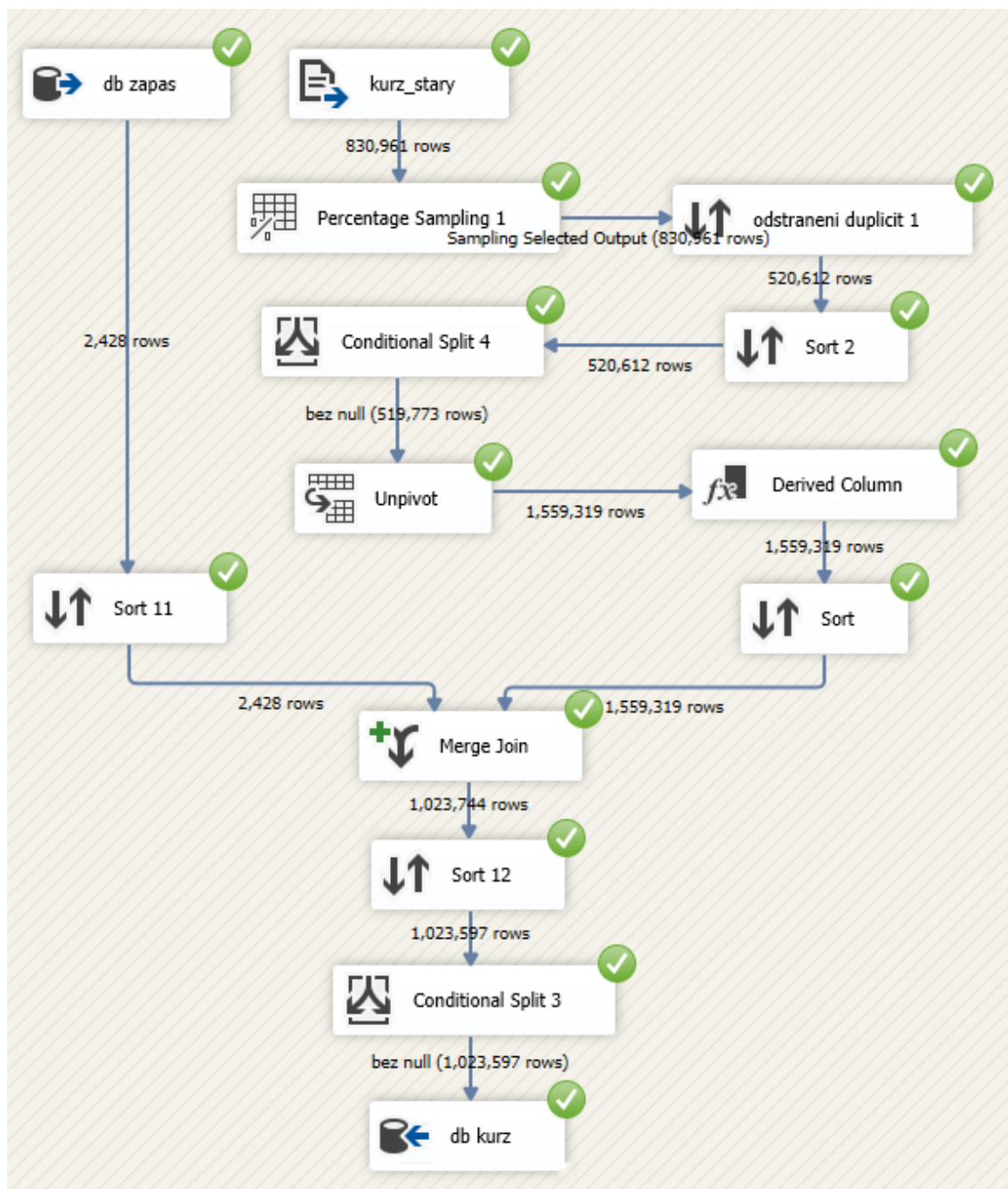
id_zapas	id_kancelar	typ_kurzu	hodnota
1	1	1	1,9
1	1	0	3,45
1	1	2	3,85

Zdroj: vlastní



Obrázek 9 ETL pro dimenzi DATUM_ČAS a ZÁPAS

Zdroj: vlastní



Obrázek 10 ETL pro tabulku faktů

Zdroj: vlastní

Kompletní ETL procesy a jejich výsledky jsou v PŘÍLOHA.

3.3.1 Shrnutí

Do datového skladu se podařilo nahrát všechna data úspěšně. Tabulka faktů čítá 9 183 522 řádků. Zápasů bylo naimportováno 6 355. Celý proces trval asi 15 minut.

3.4 Reporty, sestavy

Následující grafy, tabulky a jednotlivé ukazatele byly zpracovány v programu Power BI od společnosti Microsoft.

Nejdříve jsem do programu nahrál všechna data z předem připravené databáze. Data jsem ještě musel mírně upravit. Například jsem přidal rozdíl mezi datem stažení kurzů a datem zápasu nebo vytvořil tzv. running sum, což označuje postupné přičítávání hodnot v čase.

3.4.1 Report počtu událostí

Počet unikátních zápasů 6 355, počet unikátních řádků kurzů 9 183 522. Pravděpodobnost, že vyhraje domácí tým je 46,22 %. Remíza nastane ve 24,08 % zápasů a šance na výhru hostů je 29,66 %.

3.4.2 Report průměrné výše kurzů

Tento report má za cíl porovnávat sázkové kanceláře z hlediska výši nabízených kurzů. Ukazuje průměrné výše kurzů a jejich hladiny. Interpretací tohoto reportu bychom měli být schopni říct, u kterých SK je výhodnější sázet, tedy nabízejí sázkařům větší kurzy s menší marží.



Graf 1 Průměrné kurzy nabízené SK

Zdroj: vlastní

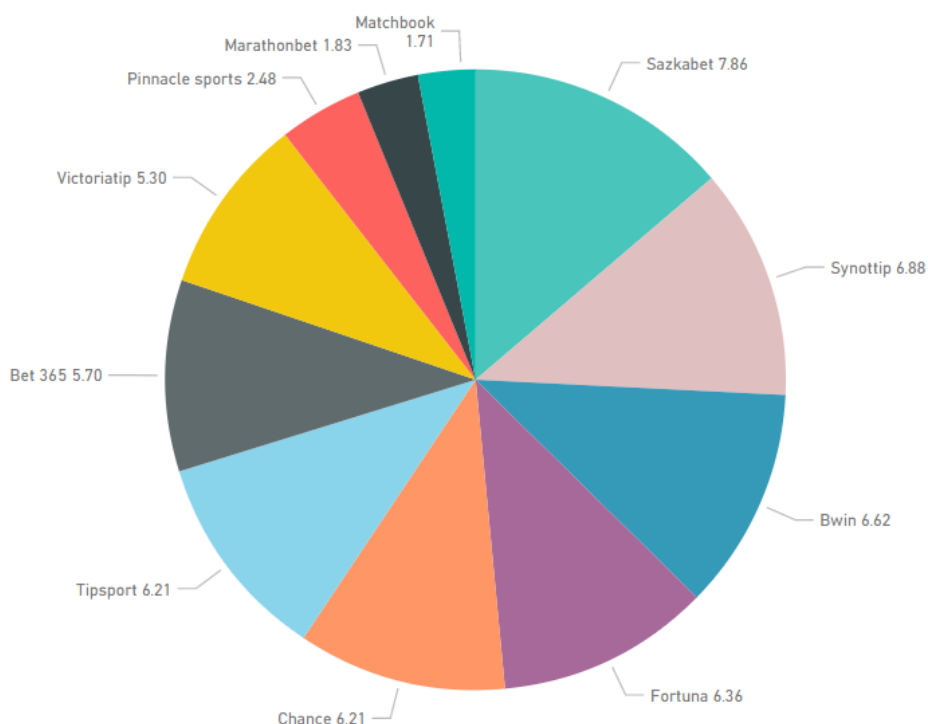
Na tomto grafu je porovnání SK z hlediska nabízených kurzů. Jak můžeme vidět, nejlepší kurzy nabízí Marathonbet. Ten je ovšem známý tím, že odmítá přijmout jakékoliv množství finančních prostředků na uzavření sázky.

Jako jednu z přesnějších SK lze označit Pinnacle Sports. Ta přijme vždy předem stanové množství vkladů. Z českých kanceláří nabízejí nejlepší kurzy společnost Tipsport a její dceřiná společnost Chance. Nejnížší hodnotu kurzů a tedy nejnížší potenciální zisky nabízí společnost Sazkabet. Oproti kanceláři Pinnacle Sports nabízí společnost průměrně o 11 % nižší kurzy. Společnost Sazka, pod kterou patří i Sazkabet, má portál sportovních sázek jen jako boční přivýdělek, protože hlavní zdroj zisků plyne z loterií.

Průměrné kurzy se liší i na uzavření sázky na výhru domácích, remízu nebo výhru hostů. Upřednostňovaný je na domácím prostředí i outsider, a proto jsou kurzy na domácí nižší (výhra domácích 2,62; remíza 4,11 a výhra hostů 4,54).

3.4.3 Report marže SK

Průměrná velikost marže



Graf 2 Průměrná velikost marže

Zdroj: vlastní

Graf 2 ukazuje velikosti marží kancelářů. To je ta část, o kterou nás SK připravuje dlouhodobě a tvoří si z ní zisk. Zde se opět Pinnacle Sports umístil na předních pozicích s hodnotou 2,48 %. Znamená to, že se dlouhodobě pohybuje na hladinách okolo 97,5 %. Úplně nejnižší marži má společnost Matchbook. Je to burza sportovních událostí a funguje na principu uzavírání sázek mezi lidmi samotnými. V této marži není započítaná provize pro tuto společnost. Největší část peněz si od zákazníka ukrojí SK Sazkabet. Průměrně kolem 7,86 %.

Průměrná marže na sport

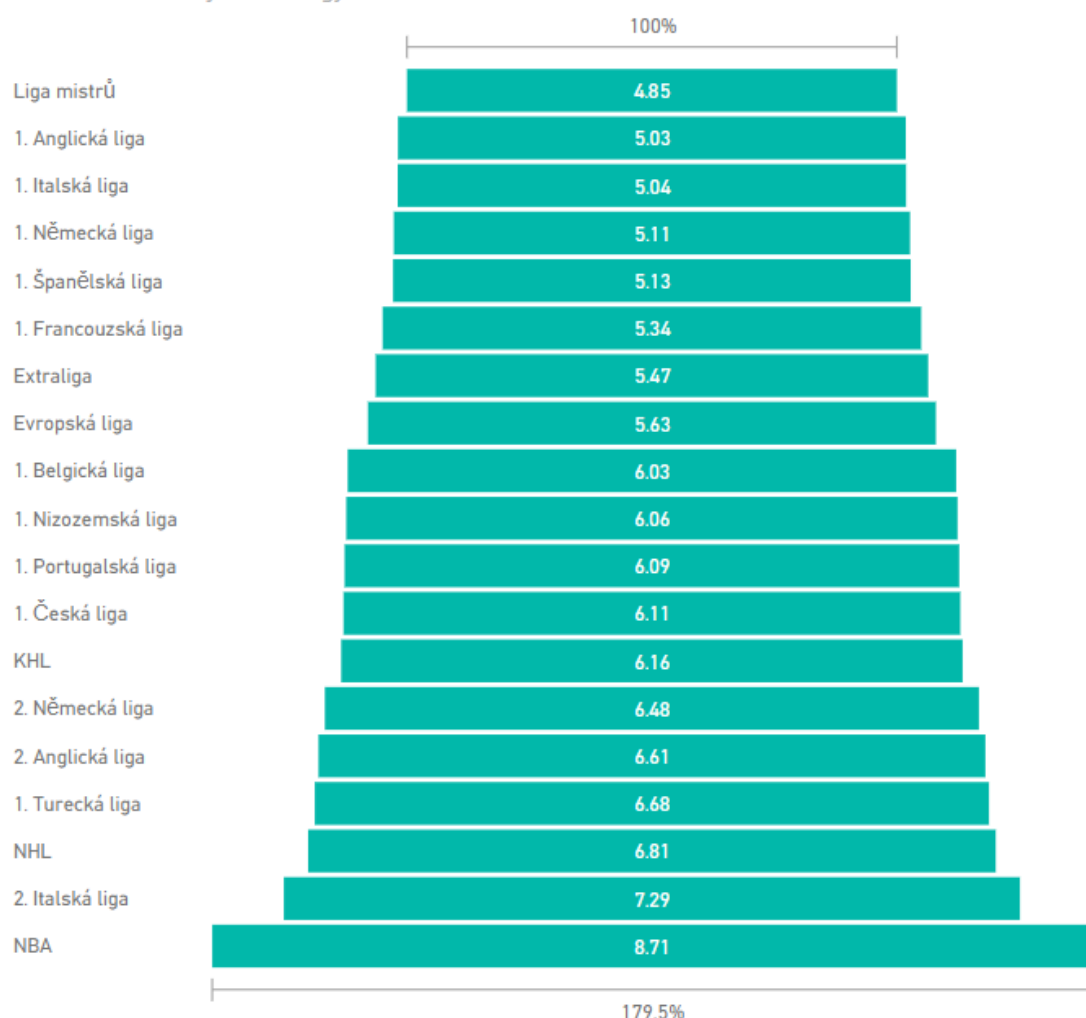


Graf 3 Průměrná marže na sport

Zdroj: vlastní

Nejvíce příjmů plyne SK z fotbalu. Proto si může dovolit na tento sport snížit marži. Z fotbalu jsou nejsledovanější a zároveň nejsázenější první ligy, evropské ligy a turnaje mistrovství světa. Jak ukazuje graf 4, nejnižší marže plyne z Ligy mistrů, následují první liga Anglie, Itálie, Německa, Španělska a Francie. Hokejová extraliga by měla být někde v dolní části výčtu, ale polovina ze sledovaných kanceláří jsou české. A Češi jsou pokládáni za hokejový národ, možná proto je 1. Česká liga až za Extraligou. Následují druhé fotbalové ligy, ruská KHL, zámořská NHL a jako poslední je americká NBA.

Průměrná marže na jednotlivé ligy



Graf 4 Průměrná marže lig

Zdroj: vlastní

3.4.4 Report hladiny SK

Konkurenční boj mezi SK a sbírání různých informací o sportu způsobuje, že se kurzy liší. Každá kancelář by ráda poskytla nejlepší kurz, ale zároveň se ohlíží na svou marži. V okamžiku vypisování kurzů je dobré svůj odhad zkontrolovat. Ovšem jedna ze SK musí být první, kdo kurz nabídne, a proto při vypisování kurzů jsou marže SK největší. Brání se tím proti případným chybám. Naopak v okamžiku těsně před začátkem zápasu bojují kanceláře o poslední sázkaře. Informací je mnoho, a proto si SK mohou dovolit marži snižovat (zvyšovat hladinu).



Graf 5 Průměrná hladina v čase [dnů před zápasem]

Zdroj: vlastní

Graf 5 ukazuje vývoj hladiny v okamžiku vypsání zápasu (vpravo) a v den začátku zápasu (vlevo). Průměrná hladina se liší v jednotkách procent.

Zajímavým úkazem je propad hladiny mezi třetím a šestým dnem před začátkem zápasu. Může se jednat o marketingový trik, kdy má následné zvýšení kurzů větší efekt. Druhou možností je očekávání informací ke zpřesnění odhadů.

Při bližším zkoumání jsem narazil na přesně opačné chování společnosti Pinnacle Sports. Následující obrázek zobrazuje vývoj hladiny této společnosti.



Graf 6 Vývoj hladiny Pinnacle Sports v čase [dny před zápasem]

Zdroj: vlastní

Ve chvíli, kdy ostatní kanceláře snižují své hladiny (3 až 6 dní před zápasem), Pinnacle Sports ji mírně zvyšuje. Zářez prudkého snížení hladiny těsně před zápasem může značit nedostatek dat.

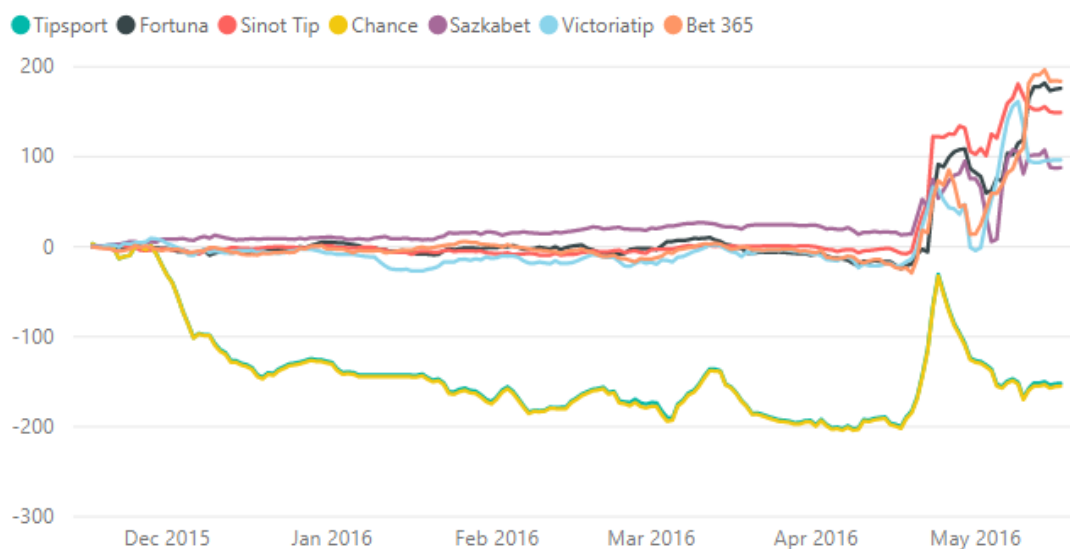
3.4.5 Report historických hodnot

Pro vytvoření modelu sázení je zapotřebí vybrat kancelář nebo kanceláře, podle kterých se bude vybírat, který kurz vsadit. Tedy ten, který v sobě nese tzv. value. Dle předchozích statistik jsem vybral společnost Pinnacle Sports. V rámci časového hlediska i stejného posuzování lig má nejstabilnější hladinu kolem 97,5 %. Její kurzy jsou i nejlepší z průměru.

Posouzení bude probíhat na stoprocentní hladině. Kurzy se přepočítají, aby je bylo možné porovnávat. Poté se srovnají s kurzem Pinnacle a jestliže bude porovnáván kurz větší, fiktivně se vsadí. Sázka bude vždy taková, aby byla vyhrána právě jedna koruna (výše vkladu $1/\text{kurz}$). V případě, že zápas skončil se stejným výsledkem jako uzavřený kurz, připočte se jednička a odečte vklad. V opačném případě se pouze odečte vklad. Tyto výsledky se sečtou za každý den sázení - vytvoří se hodnota running total. Tento průběh zobrazují obrázky dole.

Modelově se bude investovat jen na prvních ligách, lize Mistrů, Evropské lize a České Extralize.

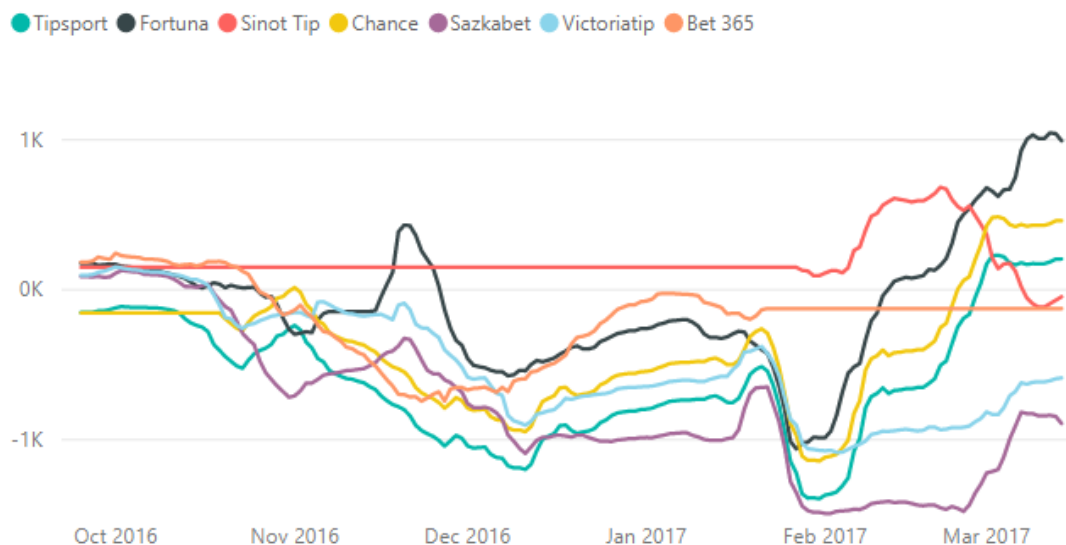
Začátek stahování kurzů byl 17. listopadu 2015, konec, s novou českou legislativou o zahraničních kancelářích, 14. března 2017.



Graf 7 Running total pro sezónu 2015/2016

Zdroj: vlastní

Pro sezónu 2015/2016 by se investice vyplatily u všech sledovaných kanceláří, kromě Tipsportu a Chance. Jak jsem již zmínil, jedná se o jednoho vlastníka, a proto nabízejí stejné kurzy. Navíc si kontrolují kurzy s Pinnaclem, což potvrzuje klesající trend (marže kanceláře). U ostatních společností jsou finální čísla v kladných hodnotách.



Graf 8 Running total pro sezónu 2016/2017

Zdroj: vlastní

Po odečtení počátečních hodnot se v sezóně 2016/2017 se investice vyplatí jen u třech kanceláří a to jen díky konečnému prudkému růstu. Graf 8 ukazuje, že v této sezóně opisovaly i ostatní kanceláře. Trendy všech českých společností je podobný. Rozdíl lpí jen na velikosti nabízeného kurzu. Z grafu tedy vyplývá, že se vyplatí sázet u společnosti Fortuna. Tiposport s Chancí se závěrečným růstem pohybují také nad osou x. To může být způsobeno krátkým obdobím a rozhodnout o tomto jevu se bude moct až po importování dalších kurzů. Vodorovné čáry Synottipu a Bet 365 jsou způsobeny dočasným nestahováním kurzů.

ZÁVĚR

Z výše uvedených informací vyplývá, že při porovnání kurzu se sázkovou kanceláří Pinnacle Sports můžeme dlouhodobě v zisku. Je ovšem důležité sázet jen na vysoce žádané sporty, kdy jsou hladiny kurzů největší, jako jsou první fotbalové ligy nebo evropskou Liga mistrů. Je lepší si počkat na konec každé sezóny a důležité je sázet u kanceláří nabízející nejvyšší kurzy.

Za podobné zpracování informací se dneska platí nemalé částky, proto i já odhaduji svou práci na 50 000 Kč. Zhodnocením může být samotné investování do budoucna. Potenciál je například u společnosti Fortuna.

Tento program je možné nadále rozvíjet, například porovnáváním kurzů z časového hlediska. Na základě proměn kurzu v čase se poté rozhodnout, zda kurz obsahuje value, nebo nikoli. Jiným směrem vývoje by mohlo být strojové sázení, kdy by počítač sám vyhodnocoval, jestli sázku uzavřít.

SEZNAM POUŽITÉ LITERATURY

- ARNOŠT, Daniel, 2007. *Business intelligence: příručka manažera*. Vyd. 1. Praha: TATE International. Příručka manažera. ISBN 978-808-6813-127.
- ČADIL, Heřman, 2013. SWOT analýza. *Brain Tools* [online]. [cit. 2014-01-17]. Dostupné z: <http://www.braintools.cz/swot-analyza.htm>
- Časopis *IT Systems: Dolování dat aneb Hledání skrytých souvislostí* [online], 2001. **2001**(6) [cit. 2017-05-17]. ISSN 1802-615X. Dostupné z: <http://www.systemonline.cz/clanky/dolovani-dat-aneb-hledani-skrytych-souvislosti.htm>
- CHEN, S.C., 2015. PHP Simple HTML DOM Parser. *Source Forge* [online]. La Jolla: Slashdot Media [cit. 2018-05-14]. Dostupné z: <http://simplehtmldom.sourceforge.net/>
- ČTK, , 2017. Češi loni prosázeli téměř 200 miliard, jde o rekord. *E15.cz* [online]. Praha: Economia, a.s. [cit. 2017-05-21]. Dostupné z: <http://zpravy.e15.cz/domaci/udalosti/cesi-loni-prosazeli-temer-200-miliard-jde-o-rekord-1335509>
- GRANT, Robert M., 1998. *Contemporary strategy analysis: concepts, techniques, applications*. 3rd ed. Malden, Mass: Blackwell Publishers.
- HANZELKOVÁ, Alena, 2009. *Strategický marketing: teorie pro praxi*. Vyd. 1. Praha: C.H. Beck. C.H. Beck pro praxi. ISBN 978-80-7400-120-8.
- HAVRLANT, Lukáš, 2014. Pravděpodobnost. *Matematika* [online]. Brno: Nová média [cit. 2017-05-18]. Dostupné z: <http://www.matematika.cz/pravdepodobnost>
- HOFFMAN, Joshua, b.r. Business Intelligence: Empower Your Users with Business Intelligence. In: *TechNet Magazine* [online]. New York City [cit. 2017-04-13]. Dostupné z: <https://technet.microsoft.com/en-us/library/gg405051.aspx>
- HOMOLA, Vladimír, 2014. Úvod do statistiky. HOMOLA, Vladimír. *Vysoká škola báňská — Technická univerzita Ostrava* [online]. Ostrava: VŠB-TUO [cit. 2017-05-17]. Dostupné z: <http://homel.vsb.cz/~hom50/SLBSTATS/UST/GS02.HTM>
- KEŘKOVSKÝ, Miloslav a Oldřich VYKYPĚL, 2006. *Strategické řízení: teorie pro praxi*. 2. vyd. Praha: C.H. Beck. C.H. Beck pro praxi. ISBN 80-717-9453-8.
- LABERGE, Robert, 2012. *Datové sklady: agilní metody a business intelligence*. 1. vyd. Brno: Computer Press. ISBN 978-802-5137-291.
- LACKO, Ľuboslav, 2009. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby*. Vyd. 1. Brno: Computer Press. ISBN 978-802-5128-879.
- LIDWELL, William, Kritina HOLDEN a Jill BUTLER, 2011. *Univerzální principy designu: 125 způsobů jak zvýšit použitelnost a přitažlivost a ovlivnit vnímání designu*. Vyd. 1. Brno: Computer Press. ISBN 978-80-251-3540-2.
- LIEBOWITZ, Jay, 2006. *Strategic intelligence: business intelligence, competitive intelligence, and knowledge management*. Vyd. 1. Boca Raton, FL: Auerbach Publications. ISBN 08-493-9868-1.

- NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ, 2005. *Business intelligence: jak využít bohatství ve vašich datech*. 1. vyd. Praha: Grada. Management v informační společnosti. ISBN 80-247-1094-3.
- PAVELKA, Jindřich, 2007-2017. *Hazardní hry* [online]. Praha: Hazardní-Hry.eu [cit. 2017-05-20]. Dostupné z: <http://www.hazardni-hry.eu/uzitecne/hazardni-hra-definice.html>
- Sázkařský slovník - pojmy které by měl ovládat každý sázkař, 2011-2017. *Sázkové-Kanceláře.com* [online]. Praha: Gamus Media Ltd [cit. 2017-05-18]. Dostupné z: <http://www.sazkove-kancelare.com/slovník-sazkare/>
- SEDLÁČKOVÁ, Helena, 2006. *Strategická analýza*. 2. přeprac. a dopl. vyd. Praha: C. H. Beck, xi, 121 s.
- TVRDÍKOVÁ, Milena, 2008. *Aplikace moderních informačních technologií v řízení firmy: nástroje ke zvyšování kvality informačních systémů*. 1. vyd. Praha: Grada. Management v informační společnosti. ISBN 978-80-247-2728-8.
- VANAGAS, Jarrad, 2015. Původ a historie hazardních her. *Encyklopedie Hazardu* [online]. Praha: Solutions Limited [cit. 2017-05-20]. Dostupné z: http://www.encyklopediehazardu.cz/rubriky/legislativa-a-historie/puvod-a-historie-hazardnich-her_6.html
- VAVŘÍNEK, Petr, 2010-2014. Systémy sázení. *JakNaSazky.CZ: Sportovní sázení jako investice* [online]. Praha: StartCom, s.r.o. [cit. 2017-05-20]. Dostupné z: <http://www.jaknasazky.cz/clanky/rizika-livebettingu>
- VINCENT, Lukáš, 2009. Škola sázení: Souhrn aneb sázkařské sedmero. *IDNES.cz* [online]. Praha: MAFRA, a. s. [cit. 2017-05-20]. Dostupné z: http://sazeni.idnes.cz/skola-sazeni-souhrn-aneb-sazkarske-sedmero-fkm-/sazeni.aspx?c=A090904_105712_sazeni_elv

SEZNAM GRAFŮ

Graf 1 Průměrné kurzy nabízené SK.....	53
Graf 2 Průměrná velikost marže.....	54
Graf 3 Průměrná marže na sport	55
Graf 4 Průměrná marže lig	56
Graf 5 Průměrná hladina v čase [dnů před zápasem].....	57
Graf 6 Vývoj hladiny Pinnacle Sports v čase [dny před zápasem]	58
Graf 7 Running total pro sezónu 2015/2016	59
Graf 8 Running total pro sezónu 2016/2017	60

SEZNAM OBRÁZKŮ

Obrázek 1 Životní cyklus datového skladu.....	19
Obrázek 2 Principy business intelligence	24
Obrázek 3 Porterův model pěti sil	28
Obrázek 4 Stará verze programu - přehled SK.....	30
Obrázek 5 Nová verze programu - přehled SK.....	31
Obrázek 6 E-R diagram DS	43
Obrázek 7 E-R diagram staré DB	44
Obrázek 8 Celkový data-flow	47
Obrázek 9 ETL pro dimenzi DATUM_ČAS a ZÁPAS	50
Obrázek 10 ETL pro tabulku faktů	51

SEZNAM TABULEK

Tabulka 1 SWOT analýza s možnostmi řešení	29
Tabulka 2 Nabídka sázkových kanceláří na zápas Manchester United a Crystal Palace	32
Tabulka 3 Fakta - KURZ.....	38
Tabulka 4 Dimenze - ZÁPAS	39
Tabulka 5 Dimenze SOUTĚŽ a pod-dimenze SPORT A LIGA	39
Tabulka 6 Dimenze DATUM A ČAS	40
Tabulka 7 Dimenze TYP KURZU a pod-dimenze MOŽNOST a PŘÍLEŽITOST	41
Tabulka 8 Dimenze KANCELÁŘ.....	42
Tabulka 9 Ukázka mapy týmů	46
Tabulka 10 Příklad zdrojové tabulky	49
Tabulka 11 Příklad cílové tabulky	49

SEZNAM ZKRATEK

SK	sázková kancelář
DS	datový sklad
ETL	extract, transform, load
BI	business intelligence
ID	identification
DP	diplomová práce
ČR	Česká republika
HTML	hypertext markup language
PHP	hypertext preprocessor
SQL	structured query language
SWOT	strength, weaknesses, opportunities, threats
SLEPT	ekonomické, ekologické, legislativní, politické, sociální, technické
EIS	executive information system
IS/IT	informační systém/ informační technologie
OLAP	online analytical processing
E-R diagram	entity relationship diagram
AI	auto increment

SEZNAM PŘÍLOH

Příloha 1 Kód parsovacího programu.....	I
Příloha 2 ETL diagram pro dimenzi ZÁPAS a DATUM_ČAS novy.....	VI
Příloha 3 ETL diagram pro tabulku faktu KURZ novy část 1	VII
Příloha 4 ETL diagram pro tabulku faktu KURZ novy část 2	VII
Příloha 5 ETL diagram pro tabulku faktu KURZ novy část 3	VIII