ČESKÉ VYSOKÉ
UČENÍ TECHNICKÉ
V PRAZE

FAKULTA DOPRAVNÍ

ZAVĚREČNÁ PRÁCE

2023

TUAN ANH
NGUYEN

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta dopravní
děkan
Konviktská 20, 110 00 Praha 1

K621.....Ústav letecké dopravy

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Bc. Tuan Anh Nguyen

Studijní program (obor/specializace) studenta:

navazující magisterský – PL – Provoz a řízení letecké dopravy

Název tématu (česky): **Analýza cen letenek na dlouhých tratích**

Název tématu (anglicky): Long Haul Ticket Price Analysis

Zásady pro vypracování

Při zpracování diplomové práce se řiďte následujícími pokyny:

- Cílem práce je vytvořit pokročilý model pro zkoumání cen letenek na dlouhých tratích mezi různými typy letišť se zaměřením na různé segmenty cestujících a různé typy destinací.
- Proveďte rešerši literatury na téma analýza cen letenek jejich charakteristika a atributů se zaměřením se na dlouhé tratě.
- Dle rešerše vytvořte vhodnou strukturu problému, datovou strukturu, vyberte vhodné městské páry podle jednotlivých segmentů letišť, cestujících i typů destinací a svůj výběr zdůvodněte.
- Proveďte sběr dat o cenách letenek, charakteristikách letišť a destinacích z dostupných zdrojů, data zpracujte, vyberte správné analytické metody a proveďte analýzu.
- Na základě analýzy vytvořte model pro popis letenek podle stanovené charakteristiky a atributů.
- Proveďte validaci, diskusi a vytvořte východiska a závěry práce.

Rozsah grafických prací: dle pokynů vedoucího diplomové práce

Rozsah průvodní zprávy: minimálně 55 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)

Seznam odborné literatury:
HOLLOWAY S.: Straight and Level: Practical Airline Economics, 2008
ALDERIGHI A.: Case study of pricing strategies in European airline markets: The London – Amsterdam route, 2011
BELOBABA T.: The Global Airline Industry, 2009

Vedoucí diplomové práce: **doc. Ing. Peter Vittek, Ph.D.**

Datum zadání diplomové práce: **15. července 2022**
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání diplomové práce: **15. května 2023**
a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE Fakulta dopravní

L. S.

doc. Ing. Jakub Kraus, Ph.D.
vedoucí
Ústavu letecké dopravy

prof. Ing. Ondřej Přibyl, Ph.D.
děkan fakulty

Potvrzuji převzetí zadání diplomové práce.

Bc. Tuan Anh Nguyen
jméno a podpis studenta

V Praze dne ..... 15. července 2022

Abstrakt

Předmětem diplomové práce "Analýza cen dálkových tratích" je analýza cen letenek a vytvoření vhodného modelu popisující tento děj. Pro zákazníky je velmi těžké koupit letenku na dálkové lety, která by odrážela všechny jejich potřeby za rozumnou cenu, protože letecké společnosti podle statistik vědí, co většina cestujících potřebuje. Je známo, že ceny letenek závisí na mnoha faktorech a každá letecká společnost má svou vlastní cenovou strategii. Za tímto účelem byl vybrán soubor vlastností charakterizujících typický let s předpokladem, že tyto vlastnosti ovlivňují cenu letenky. Práce se zaměřuje především na tři páry dálkových letů mezi Evropou a Jihovýchodní Asií. Potřebná data byla vytěžena z cestovatelského webu Pelikan.cz pomocí počítačového programu Web Scraper. Na základě analýzy byla použita metodika pro vytvoření vhodného modelu, který bude popisovat vývoj cen letenek pro obchodní i rekreační cestující. Vytvořený pokročilý Bayesovský model byl statisticky ověřen porovnáním s metodou nejmenších čtverců pomocí statistických funkcí (RMSE a R2). Výsledky této studie ukazují, že náš model je vhodný pro modelování predikce cen letenek s hodnotou R2 dosahující hodnot 0,73-0,90. Je také zřejmé, že ceny letenek závisí nejen na době nákupu, ale také na výběru letecké společnosti, třídách letenek nebo sezónnosti. Tato práce slouží jako příručka pro lepší pochopení cenových strategií používaných leteckými společnostmi pro dálkové lety.

Klíčová slova

Cena letenky, dálkový let, model predikce

Abstrakt

The subject of the thesis "Long haul ticket price analysis" is to analyze flight ticket prices on long-haul flights and create a suitable model. It is very hard for customers to a buy long-haul flight ticket which reflects all their needs for a reasonable price because airlines know by statistics what most passengers need. It is well-known that flight ticket prices are dependent on numerous factors, and every airline has its own pricing strategy. For this purpose, a set of features characterizing a typical flight was decided upon, supposing that these features affect the price of the air ticket. The thesis focuses mainly on three long-haul city pairs between Europe and Southeast Asia. The required data was mined from travel site Pelikan.cz by computer program Web Scraper. The methodology was applied to create a suitable model based on the analysis, which will describe the development of flight ticket prices for both business and leisure travellers. The created advanced Bayesian model was statistically validated by comparison with Least Square Method using statistical functions (RMSE and R2). The results of this study show that our model is suitable for modelling airfare prediction with a R2 value reaching value between 0,73-0,90. It is also apparent that ticket prices depend not only on the purchasing time but also on airline selection, airline classes or seasonality. This thesis serves as a manual for a better understanding of the pricing strategies employed by airlines for long-haul flights.

Keywords

Airfare price, long-haul flight, prediction model

Poděkování

Na tomto místě bych rád poděkoval všem, kteří mi poskytli podklady pro vypracování této práce. Zvláště pak děkuji doc. Ing. Peteru Vittekovi, Ph.D. za odborné vedení a konzultování závěrečné práce a za rady, které mi poskytoval po celou dobu mého studia, a dále bych chtěl poděkovat doc. Ing. Evženie Uglickich, CSc. za hodiny statistiky a stochastických systémů. Dále bych chtěl poděkovat Mgr. Aleši Hajzlerovi za umožnění přístupu k mnoha důležitým informacím. V neposlední řadě je mou milou povinností poděkovat svým rodičům a blízkým za morální a materiální podporu, které se mi dostávalo po celou dobu studia.

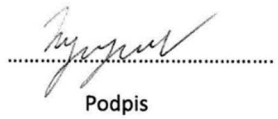
Prohlášení

Předkládám k posouzení a obhajobě bakalářskou práci zpracovanou na závěr studia na ČVUT v Praze, Fakultě dopravní.

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 15.5.2023



Podpis

Obsah

ÚVOD 14
1. Vývoj letecké dopravy 15
1.1. Regulace a deregulace letecké dopravy 15
1.2. Lety na dlouhých tratích 16
2. Řízení výnosů leteckých společností 18
2.1. Proces nacenění letenky 18
2.2. Zpáteční vs. jednosměrné letenky 20
2.3. Nabídka a ocenění služeb leteckých společností 20
2.4. Sezónní vrcholy (špičky) 21
2.5. Výdaje při operování dálkových letů 22
3. Uzlová přeprava mezi Evropou a Jihovýchodní Asií 23
4. Metodika k cenám letenek dálkové přepravy 25
4.1. Výběr vhodného datového zdroje pro těžení dat o cenách letenek na dlouhých tratích 26
4.1.1. Možnosti těžení dat 26
4.1.2 Způsob těžení dat 26
4.1.3. Výběr vhodných městských párů 27
4.2. Popisná statistika pro zpracování natěžených dat o cenách letenek 28
4.2.1. Grafické znázornění o cenách letenek pomocí popisné statistiky 28
4.2.2. Odstranění přebytečných či nežádoucích cen letenek 30
4.3. Analýza zpracovaných dat rešerše vhodných metod pro tvorbu pokročilého modelu 31
4.3.1. Testování závislosti veličin ze zpracovaných dat o cenách letenek 31
4.3.2. Výběr modelů pro vizualizaci vývoje cen letenek na dálkových tratích 32
4.3.3. vytvoření jednoduchého modelu pro popis cen letenek na dálkových tratích 33
4.3.4. Vytvoření pokročilého modelu pro popis cen letenek na dálkových tratích 34
4.4. Validování výsledku získaných z modelů 35
4.4.1. Určení přínosnosti modelu pomocí Koeficientu determinace (R^2) 35
4.4.2 Určení úspěšnosti vytvořených modelů pomocí RMSE 36
5. Analýza městského páru Vídeň (VIE) – Bangkok (BKK) 38
5.1. Austrian Airlines 38
5.2. EVA Air 38
5.3. Analýza letecké poptávky městského páru VIE - BKK 39
5.4. Zpracování dat přímého letu 40
5.5. Analýza přímého letu 41
5.6. Zpracování dat letů s jedním přestupem 42

5.7. Zpracování dat letů s dvěma přestupy 47
5.8. Výsledky zpracovaných dat 47
6. Analýza městského páru Paříž (CDG) – Singapur (SIN) 49
6.1. Air France 49
6.2. Singapore Airlines 49
6.3. Zpracování dat přímého letu 50
7. Analýza městského páru Praha (PRG) – Hanoj (HAN) 52
8. Vytvoření modelů 54
8.1. Model cen letenek na městském páru VIE – BKK 54
8.2. Model pro dolní kvartil 56
8.3. Modely pro medián 58
8.4. Model cen letenek na městském páru PRG – HAN 60
8.5. Model pro městský pár PRG – HAN 61
8.6. Validace získaných dat 63
Diskuze 65
Závěr 67
Zdroje 69
Přílohy 74

Seznam obrázků

Obrázek 1.: Srovnání typů letadel 17

Obrázek 2.: Konektivita mezi městskými páry v období konce 3/2019 a konce 3/2020 23

Obrázek 3.: Srovnání nárůstu počtu pasažérů v jednotlivých uzlech 24

Obrázek 4.: Metodologický model zpracování 25

Obrázek 5.: Krabicový graf s vyznačením významných kvantilů pozorovaných dat 28

Obrázek 6.: Příklad odlehlé hodnoty krabicového grafu 30

Obrázek 7.: Příklad odlehlé hodnoty histogramu 30

Obrázek 8.: Bayesovo pravidlo 35

Obrázek 9.: Rozpětí cen přímých zpátečních letů v závislosti na době před odletem VIE – BKK 41

Obrázek 10.: Rozpětí cen zpátečních letenek Austrian Airlines z VIE do BKK 41

Obrázek 11.: Rozpětí cen zpátečních letenek Austrian Airlines z VIE do BKK 42

Obrázek 12.: Krabicový graf doby letu 43

Obrázek 13.: Histogram závislosti počtů zpátečních letů na dobu letu 43

Obrázek 14.: Bodový graf závislosti ceny na dobu letů (před filtrací dat) 44

Obrázek 15.: Krabicový graf letů s jedním přestupem po filtraci celkové doby letu 44

Obrázek 16.: Histogram závislosti počtů zpátečních letů na cenu letenky 45

Obrázek 17.: Krabicový graf letů s jedním přestupem kompletním zpracováním 31 až 15 dnů před odletem 46

Obrázek 18.: Krabicový graf letů s jedním přestupem kompletním zpracováním 14 dnů až den před odletem 46

Obrázek 19.: Krabicový graf cen letenek v jednotlivých dnech 47

Obrázek 20.: Krabicový graf cen letenek v jednotlivých dnech 48

Obrázek 21.: Krabicové grafy ukazující rozpětí cen vybraných aerolinek 48

Obrázek 22.: Rozpětí cen přímých zpátečních letů v závislosti na době před odletem 50

CDG – SGD 50

Obrázek 23.: Rozpětí cen zpátečních letenek Air France z CDG do SIN 51

Obrázek 24.: Rozpětí cen zpátečních letenek Singapore Airlines z CDG do SIN 51

Obrázek 25.: Vývoj ceny letenek China Airlines během období 18. 7. 2023 – 9. 3. 2024 na lince PRG - HAN 52

Obrázek 26.: Krabicový graf sezónních cen od China Airlines. 53

Obrázek 27.: Bayesovský model pro dolní kvartil 56

Obrázek 28.: Výsledky Bayesovského modelu – dolní kvartil 57

Obrázek 29.: Model založený na Metodě nejmenších čtverců pro dolní kvartil 57

Obrázek 30.: Výsledky MNČ modelu – dolní kvartil 58

Obrázek 31.: Bayesovský model pro medián 58

Obrázek 32.: Výsledky Bayesovského modelu – medián 59
Obrázek 33.: Model založený na Metodě nejmenších čtverců pro medián 59
Obrázek 34.: Výsledky z modelu MNČ - medián 60
Obrázek 35.: Bayesovský model pro sezónní špičku 61
Obrázek 36.: Výsledky Bayesovského modelu – sezónní špička 62
Obrázek 37.: Metoda nejmenších čtverců pro sezónní špičku 62
Obrázek 38.: Výsledky MNČ – sezónnost 63

Seznam tabulek

Tabulka 1: Základní informace o přímých letech mezi VIE a BKK 39
Tabulka 2: Celková týdenní kapacita 40
Tabulka 2: Přímé lety vs data z Eurostatu 40
Tabulka 3: Výsledky kolinearity mezi koeficienty VIE - BKK 55
Tabulka 4: Výsledky kolinearity mezi koeficienty PRG - HAN 61
Tabulka 5: Srovnání výkonnosti různých regresních modelů 63

SEZNAM ZKRATEK

A320	Airbus A320	
AMS		Letiště Amsterdam Schiphol
AUH		letiště Abú Zabí
B737	Boeing 737	
BKK		Letiště Bangkok-Suvarnabhumi
CAPA	Centre for Aviation	Centrum pro letectví
CDG		Letiště Charlese de Gaulla
COVID-19	Coronavirus disease 2019	Koronavirové onemocnění 2019
CZK	Czech koruna	Česká koruna
DOH		letiště Dauhá Hamad
DXB		letiště Dubaj
FRA		Letiště Frankfurt nad Mohanem
HAN		Letiště Hanoj
IST		Letiště Istanbul
KLM	Koninklijke Luchtvaartmaatschappij NV	Nizozemské královské aerolinie
LCC	Low-cost carrier	Nízkonákladová letecká společnost
LHR		Letiště Heathrow
OAG	Official Aviation Guide of the Airways	Oficiální letecký průvodce leteckých společností
OTAs	Online Travel Agencies	Online cestovní kanceláře
PRG		Letiště Václava Havla Praha
RMSE	Root Mean Square Error	Směrodatná odchylka chyb
SIN		Letiště Singapur-Changi
TPE		letiště Tchaj-wan Tchao-jüan
VIE		letiště ve Vídni

ÚVOD

Letecká doprava je komplexní socio-ekonomický systém a pro cestujícího je občas obtížné vyznat se v nabídce jednotlivých aerolinií. Většina leteckých dopravců má svoji vlastní strategii, přičemž výkonnost zvolené strategie závisí na příjmech, které aerolinka inkasuje od zákazníků a dále na vynaložených nákladech nezbytných pro poskytování služeb. Pro dosáhnutí kýžených výsledků je třeba optimalizovat ceny na základě dlouhodobého sledování vývoje poptávky cestujících, jejich potřeb a chování. Tento proces je známý pod pojmem řízení výnosů (angl. Revenue management) a má velký podíl na nárůst zisku leteckých společností díky dynamické cenotvorbě.

Autor práce je pravidelným cestovatelem a pro účel svých cest pravidelně využívá leteckou přepravu a služby, které jsou k nim poskytované. Kvůli asijským kořenům se autor pravidelně setkává s nákupem letenek na dálkových tratích, přičemž tento proces je jak časově, tak finančně náročný. Aby cestující mohl nakoupit letenku ve správný moment za výhodnou cenu, je třeba znát dobře řízení výnosů leteckých společností. Hlavní motivací této práce bude pochopení fungování aerolinií na dálkových tratích, z kterého vychází jejich strategie při nacenění cen zpátečních letenek.

Pro vyřešení dané problematiky, je cílem práce je analyzovat ceny letenek na dlouhých tratích a popsat jejich charakteristiku. Analýza bude probíhat na základě natěžených dat na vybraných městských párech mezi Evropou a Jihovýchodní Asií. Na základě analýzy současných dat bude vytvořen pokročilý model pro zkoumání cen letenek na dlouhých tratích mezi různými typy letišť se zaměřením na různé segmenty cestujících a různé typy destinací.

Po úvodní kapitole se práce bude zabývat základními informacemi o letecké dopravě a jejím fungování na základě bilaterálních smluv, přičemž zde bude stručně nastíněn vývoj letecké dopravy na dálkových tratích a rostoucí vliv blízkovýchodních uzlů na pozorovaných městských párech. Dále bude detailně popsán proces revenue managementu, kde se bude zkoumat nabídka leteckých společností na dlouhých tratích a jejich cenové praktiky.

Z rešerše bude vycházet analýza letenek dálkové přepravy, která bude postavena na metodice, vycházející z natěžených datech o cenách letenek na dlouhých tratích. Získaná data budou analyzována pomocí metod popisné statistiky. Po zpracování budou data aplikována na pokročilém modelu, který je nedílnou podmínkou k dosažení cíle práce.

Získané výsledky z modelu budou validovány pomocí statistických metod pro vyhodnocování přesnosti modelu, a to nejen s jednoduchým regresním modelem, ale také s podobnými modely prezentovanými ve vědeckých článcích zabývající se vývojem cen letenek.

Tato diplomová práce byla napsána s metodickými pokyny pro psaní akademických prací.

1. Vývoj letecké dopravy

1.1. Regulace a deregulace letecké dopravy

Letecká doprava byla již od počátku ovlivněna rozhodnutím politiků a vlád, kteří rozhodovali o tom, kam se bude létat a nastavovali cenovou politiku. Vlády rozhodovaly o geografickém rozložení poptávky. Trasy, kapacita či frekvence letů leteckých společností spadaly též pod vládní kontrolu, neboť letecké společnosti byly státní.

V roce 1944 vznikla Chicagská úmluva, která určila podobu letectví. Vzdušné prostory byly přiřazeny jednotlivým státům, které je musely začít chránit. Za tímto účelem vznikly bilaterální smlouvy podepsané mezi státy. Na základě těchto smluv mohla vybraná aerolinka fungovat na mezinárodních letech. Státy se domlouvaly mezi sebou, zasahovaly do rozhodování leteckých společností, např. určovaly kapacitu, nastavovaly cenu ve smlouvách, za kterou budou letecké společnosti létat na daném páru. Tento proces zajištění leteckého spojení mezi státy přetrvává dodnes, je nedílnou součástí pro provozování mezinárodních letů, o nichž také pojednává tato práce.

Jako příklad lze uvést Spojené státy americké (USA) a období mezi lety 1938 (Federální zákon o letectví - Federal Aviation Act) a 1978, kdy aerolinky podléhaly byrokratickým předpisům, které zpomalovaly proces vstupu na nové trhy. Díky těmto regulacím bylo zajištěno, že u dvou aerolinek nabízejících letenky na stejné trase byly jejich ceny téměř identické. V USA byla Rada pro civilní letectví (angl. the Civil Aeronautics Board), která zajišťovala, že tyto společnosti budou nabízet stejné ceny a rozhodovala o zvýšení či snížení ceny letenek. Kromě USA lze jako dobrý příklad regulace cen letenek uvést situaci na vnitrostátních letech na území Austrálie, kdy zde před rokem 1990 působili pouze dva letečtí dopravci (Ansett Airlines a Trans-Australia Airlines). Platila zde tzv. zásada dvou leteckých společností, kde místo zdravé konkurence, byly tyto společnosti v praxi regulovány, a to včetně cenové politiky. Za revoluční lze považovat deregulační akt (Zákon o deregulaci leteckých společností - Airline Deregulation Act) z roku 1978, který uvolnil podmínky pro provozování osobní letecké přepravy v USA. Tento zákon měl dále za následek urychlení deregulace letecké přepravy v Evropě a to tak, že bilaterální letecké dohody uzavírané v 90. letech obsahovaly jednodušší podmínky vstupu na trh. Po roce 1978 měly letecké společnosti více prostoru pro objevení nových trhů, určení poptávky či cenové politiky, avšak tyto aerolinky musely být na vnitrostátních letech vlastněny alespoň ze 70 % Spojenými státy americkými. Tyto reformy inspirovaly mnoho států k podobné, ne však úplné deregulaci [1].

Ve všech odvětvích mají firmy možnost pracovat na globální bázi, zapojují se do zahraničních fúzí a mají možnost přebrat aktivitu na tamním trhu. Leteckým společnostem jsou tyto aktivity popřeny. Pokud letecké společnosti chtějí působit na širším trhu, musí si to zařídit přes vlády či vytvářet aliance s jinými leteckými společnostmi [1].

V polovině 80. let nastala další významná změna, kdy byly státně vlastněné letecké společnosti zcela (např.: British Airways, Lufthansa, Qantas nebo Air Canada) či zčásti zprivatizovány. Tímto krokem se stalo fungování aerolinek efektivnější a eliminovalo mnoho politické byrokracie [1].

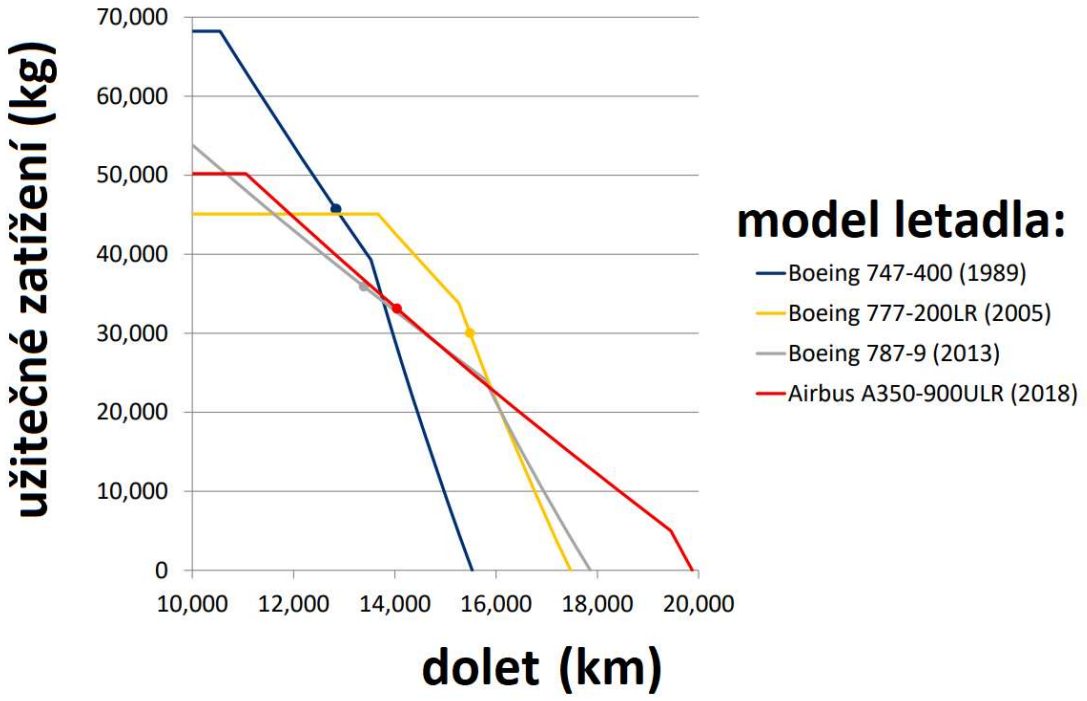
Přesto v dnešní době je díky tzv. „sedmé svobodě" možné, aby mohlo více leteckých společností působit na jiných trzích než jen tam, odkud pochází jejich vlastník. To platí hlavně pro mezinárodní lety, neboť na domácím trhu si vlády chrání svůj vlastní trh před konkurencí. Dále zde pomalu končí regulace kontroly cen letenek a s rozvojem internetu se ceny letenek mění často třeba přes noc, pokud na dané trase působí silná konkurence [2].

V této práci budeme ověřovat, zdali analyzované letecké páry podléhají leteckým regulacím a dále zde bude popsán případ aerolinky EVA Air, která díky sedmé svobodě operuje na městském páru Vídeň - Bangkok.

1.2. Lety na dlouhých tratích

Protože se práce zabývá analýzou letenek na dlouhých tratích, je třeba definovat pojmy jako je dálkový let či long-haul flight. Z IATA (Mezinárodní asociace leteckých dopravců) [3] definice vyplývá, že let na krátkou vzdálenost trvá do 3 hodin, let na střední vzdálenost je v rozmezí 3-6 hodin, a od 6 hodin je let považován za dálkový. Z definice CAPA (angl. Centre for Aviation) [4] je také potvrzená definice dálkového letu, kdy CAPA říká, že dálkový let je mezinárodní let na dlouhé vzdálenosti, typicky mezikontinentální a trvající alespoň šest hodin. Pro dálkové lety je možné se setkat s definicí ultra long-haul flight, kdy se jedná již o lety nad 16 hodin bez mezipřistání. Letecké společnosti fungují s technologickým pokrokem efektivněji. Jako příklad můžeme uvést palivově úsporná letadla, využití větších letadel, které dále snižují cenu letenek. Nové technologie umožňují letadlům létat dále s tím, že je možnost se vyhnout mezipřistání.

Trend dálkových letů (také i ultra long haul flights) roste díky podpoře leteckých výrobců, kteří vyrábí nová palivově úsporná letadla, díky nimž mohou přenášet větší zatížení na dlouhé vzdálenosti, což před desítkami let nebylo možné. Efektivnější letadla mají také zásluhu na snížení ceny letenek, která např. ve Spojených státech amerických činila 27,6 % po započtení inflace mezi lety 1995–2019 [3] [4].



Obrázek 1.: Srovnání typů letadel [4]

V extrémních případech již dokáže letadlo uletět skoro 20 000 km na jedno natankování (viz Obrázek 1).

Dalším technologickým pokrokem během posledních dvou dekád, který ovlivnil dálkové lety, byl vstup menších úzkotrupých letadel (B737, A320 apod.) na transkontinentální trhy s nonstop lety, kde tyto letadla pomocí fragmentace trhu přispěly k rychlému rozvoji leteckých uzlů na Blízkém východě. Očekává se, že budoucí generace úzkotropých letadel zvládnou na plné zatížení uletět delší vzdálenosti, které mohou výrazně ovlivnit budoucí strukturu dopravních toků (již můžeme vidět na transatlantických letech, kde se u několika městských párů nasazují nové generace úzkotropých letadel) [4].

V práci bude autor zkoumat použitá letadla na vybraných městských párech.

2. Řízení výnosů leteckých společností

Revenue management je nedílnou součástí strategie pro nacenění letenek, přičemž jeho dobré využití nám zajistí, že letadlo bude vytěženo na maximum, může to být rozdělením letadla na segmenty, kde se aplikuje cenová diskriminace, jež vytěží letadlo nad náklady.

V této kapitole bude popsána ekonomika leteckých společností na dálkových tratích. Protože se práce zabývá cenami letenek, je potřeba si uvědomit, že letecké společnosti mají mnoho strategií, jak prodávat letenky cestujícím. Každá aerolinka má svůj jedinečný přístup, který zajistí její udržení na trhu. Aerolinky z historických dat téměř vždy vědí, kdy si mohou účtovat vysoké ceny anebo naopak vědí, kdy nízká cena je jediný recept pro úspěch. Tyto strategie budou popsány v dalších podkapitolách.

2.1. Proces nacenění letenky

Na mnoha leteckých trzích se musí manažeři aerolinek potýkat se změnami či adaptací. Dříve měli jistotu regulací, omezenou konkurenci a díky byrokracii bylo tempo změn pomalé. V dnešní době čelí ekonomické liberalizaci, která přináší mnoho příležitostí, které musí být prozkoumány, jinak nemůže být úspěch zaručen. Rozhodování musí být učiněno v krátkém časovém horizontu, příležitosti pro nové trasy se objevují náhle a stejně rychle mohou zmizet. Letecké společnosti musí přizpůsobovat specifikace nabízených produktů, pokud nabídne konkurence lepší produkt. Cenová politika se musí posoudit v kratších intervalech a občas nestačí jen na denní bázi [2].

Letecké společnosti od roku 1985 zavedly řízení výnosů, které jim pomáhá účinně maximalizovat zisky. Řízení výnosů je čím dál tím komplexnější, ale z velké části je založeno na historických datech, z kterých vychází pravděpodobnostní poptávka. Pravděpodobnostní poptávka je nejideálnější způsob, jak předpovědět poptávku pro řízení výnosů. Cestující mohou být seskupování do cenových tříd, do kterých nemusí patřit, což může zvýšit přebytek spotřebitele a snížit zisk leteckých společností. Proto letecké společnosti zavádějí dynamickou cenovou politiku na základě poptávky, nákladů a konkurence na trhu. Nejjednodušší je stanovení cen je na základě nákladů, kde k nákladům si letecké společnosti přičtou zisk. Dalším způsobem, jak letecké společnosti prodávají své letenky, je pomocí svazování, kde nabízí doplňkové produkty k letenkám a skládají je do balíčků, které jsou pro firmy jednodušší při fakturování. Pro nové aerolinky, které vstupují na trh, je uplatňována strategie penetrační ceny, kde se na nových trzích prodávají letenky za velmi nízké podnákladové ceny s cílem přilákat cestující citlivé na cenu. Při špičkovém zatížení (hlavní sezóna nebo také žádané časy odletu) si v praxi účtují nejvyšší ceny, kdy je poptávka po letenkách nejvyšší, neboť v období špiček je zájem nejvyšší a díky vyšší ceně se poptávka rozloží do méně vytíženějších časových bloků či vedlejších sezón. Dalším způsobem, jak zvýšit zisky leteckých společností, je pomocí cenové diskriminace, kde se určitá místa prodávají za vyšší cenu na základě tarifních tříd, neboť jsou cestující, kteří nemají problém si připlatit více peněz, aby získali větší

komfort. Další možností cenové diskriminace může být nabídka množstevní slevy, které ale v praxi fungují většinou pro charterové lety. Aerolinky se snaží dále minimalizovat spotřebitelský přebytek, což je rozdíl mezi nejvyšší částkou, kterou je cestující ochoten zaplatit a částkou, kterou skutečně zaplatí. Pro udržení zákazníků zavádí letecké společnosti věrnostní programy, na dálkových letech se můžeme často setkat se sběrem mílí, které si následně může cestující směnit za služby (upgrade tarifní třídy) či letenky. Pro zvýšení příjmů se používá také tzv. overbooking na základě historických dat, protože aerolinky z těchto dat vědí, že určitému počtu cestujících zakoupený let uletí nebo nestihnou navazující let během přestupu, díky čemuž mohou aerolinky efektivněji využít místa v letadle [5].

Kromě nastavování cen letenek pro cestující mají aerolinky možnost zvýšit si příjem přepravou nákladu v prostorách tzv. belly cargo, tj. přepravou nákladů v nákladovém prostoru v osobních dopravních letadlech. Protože se na dálkových tratích využívají širokotrupá letadla, je kapacita přepravovaného nákladu v prostorách belly cargo značná. Na dálkových tratích přispívá významně k ziskovosti, a to hlavně na leteckých párech, kde je poptávka po přepravě silně sezónní (sezónní špička je popsána v další kapitole) [1].

S rostoucí globalizací tvoří v dnešní době na dálkových tratích velký podíl zásilek z Asie do Evropy a Ameriky. Avšak poptávka po nákladní přepravě v belly cargo neurčuje, kam bude aerolinka létat. Primární volbou pro klasické dopravce bude stále poptávka po osobní přepravě, proto musí aerolinka maximalizovat svůj faktor zatížení (tzv. load factor). Tento ukazatel měří, kolik procent dostupných míst se letecké společnosti podařilo prodat na svých letech za určité období. Každá společnost si nastavuje svoji hranici a po jejím překročení se letecká společnost stává rentabilní. V průměru se pohubuje tento koeficient kolem 70 % [1] [6].

Z let 2016–2023 si lze povšimnout, jak se měnil load factor v průběhu let. Mezi lety 2016 až 2019 se leteckým společnostem podařilo postupně navyšovat svůj load factor z 80,5 % na 82,6 %, poté přišla pandemie COVID-19, která zapříčinila pád na hodnoty 65,2 % (2020), 66,9 % (2021) a s utlumující se pandemií load factor začal znovu stoupat, a to na hodnotu 78,9 % (2022) a předpokladem pro rok 2023 dle IATA je hodnota 81 % [7] [8].

2.2. Zpáteční vs. jednosměrné letenky

Jedno z nejzákladnějších rozhodnutí při nákupu letenky je rozhodnutí, zda si koupit jednosměrnou nebo zpáteční letenku. U nákupu letenky na vnitrostátním letu je obvykle za stejnou cenu, ale na mezinárodních letech se cena výrazně odlišuje [9].

Obecně je výhodnější si zarezervovat zpáteční let s jasným dnem odletu a návratu, ale při pohledu na storno podmínky u leteckých společností jsou v určitých případech jednosměrné letenky atraktivnější. Americká finanční společnost NerdWallet [9] porovnávala ceny letenky na mezinárodních trasách a dospěla k závěru, že nákup dvou jednosměrných letenek je obvykle o 20 % dražší než při nákupu zpáteční letenky.

Dále je třeba brát v potaz změny či zrušení celé nebo části cesty. Většina aerolinek již odstranila poplatky za změnu ceny letenky, ale pokud by chtěl cestující zrušit jen část cesty ze zpátečního letu, může nastat situace, kdy je zrušení komplikovanější než při nákupu jednosměrné letenky. Zrušení jednoho segmentu z celé cesty může vyvolat zrušení dalších částí cest na letence [9].

Je zde také možnost nákupu letenky s otevřeným návratem, ale i v těchto případech mohou nastat situace, kdy jednosměrné letenky vycházejí cestujícímu výhodněji. U zpáteční letenky je velkou výhodou pro cestujícího fakt, že jsou pro manipulaci jednodušší, neboť lety spadají do jednoho rezervačního kódu (tzv. booking refference) a jak bylo zmíněno výše, téměř nikdy nejsou dražší než dvě jednosměrné. Proto je pro většinu mezinárodních cest zpáteční letenka jasným vítězem [9].

2.3. Nabídka a ocenění služeb leteckých společností

U mnoha leteckých společností (klasické aerolinky) mají cestující na výběr mezi první třídou (first class), obchodní třídou (business class nebo business třída), prémiovou ekonomickou třídou (premium economy) a ekonomickou třídou (economy class) [10].

Co se týká služebních cest, mají cestující malé nebo žádné možnosti, v jaké třídě budou cestovat, protože firmy mají svoji korporátní cestovní politiku. Vysoce postavení manažeři mají právo cestovat v první třídě, kdežto cestující střední třídy mají možnost využít business třídu. Zaměstnanci na pozici Junior se musí většinou spokojit s ekonomickou třídou. Pokud nastane období recese, v rámci firemních úspor většina firem přechází na nižší třídy, tudíž jsou první třídy, a dokonce i business třídy eliminovány. Proto, aby mohly letecké společnosti maximalizovat zisk, musí aerolinky zákazníky přesvědčit o výhodách prémiových služeb (hlavně v první třídě a business třídě) tím, že jim nabídnou možnost lepšího spánku či možnosti práce v letadle, na úkor vyšší ceny letenky [2].

Aerolinky mohou dále účtovat vyšší cenu, i když nabízí stejný produkt (např. první třídu), neboť se pokoušejí tzv. odlišit svůj produkt. Dálkoví dopravci pravidelně zavádějí inovace v business a první třídě,

a jak uvádí Holloway [1], například aerolince Singapore Airlines se díky této strategii podařila úspěšně implementovat strategie prodávání letenek za vyšší ceny než u konkurence nabízející podobné segmenty produktů, protože produkty Singapore Airlines jsou více prémiové. Pokud se aerolinka rozhodne pro tuto strategii, musí počítat s vyššími náklady a je také náročnější na provoz, neboť musí aerolinka neustále inovovat a být v této strategii konzistentní. Není neobvyklé, že letecké společnosti se silnou reputací v oblasti služeb investují více než 200 miliónů dolarů do modernizace obchodní a první třídy, aby po čtyřech až pěti letech tento proces zopakovaly. Tyto aerolinky očekávají, že tato strategie s relativně malou životností může přinést buď zajímavou návratnost investice nebo slouží k ochraně již stávajících toků příjmů, pokud je trh ovlivněn dynamickou konkurencí. Konkurence poté může reagovat buď zvýšením standardu, pružně snížením ceny letenky nebo reagovat nemusí, protože inovátor ovlivní pouze malou část trhu. Pokud má aerolinka již silnou image, může reagovat nepřímo a to nabídkou atraktivních věrnostních programů nebo zlepšením stávající konektivity [1].

Můžou však nastat situace, kdy zákazníci nebudou mít vůli zaplatit za nadstandardní výhody, i když by si to aerolinka přála. Ale tato odlišnost může být rozhodujícím faktem pro rozhodování zákazníků při nákupu letenky v konkurenčním prostředí, které je charakteristické nabídkou stejných produktů a rozsáhlým tzv. price-matchingem. Jako příklad strategie price-matchingu uvádí Holloway [1] situaci aerolinky Alaska Airlines v 90. letech 20. století, kdy Alaska neměla moc možností, jak dráž nacenit letenky vlivem silné konkurence aerolinky Southwest Airlines. Avšak díky své dobré reputaci byla Alaska Airlines pro zákazníky při nákupu letenky preferovanou volbou.

2.4. Sezónní vrcholy (špičky)

Sezónní špička je období, kdy mají letecké společnosti možnost nabídnout letenky s velkou přirážkou, díky velkému zájmu o leteckou dopravu. U dálkových letů se cenové špičky odvíjí od zvyků a svátků, jako jsou např. Vánoce, Den díkůvzdání, prázdniny. Největšímu náporu cestujících čelí během sezóny charteroví dopravci, kteří obsluhují volnočasové destinace. Tito dopravci musí poté vynaložit velké úsilí, aby pro svá letadla našli využití i mimo sezónu, příkladem může být přesun letadel do pravidelné dopravy. Ani pravidelní dopravci však nejsou imunní vůči sezónním výkyvům. Jako příklad uvádí Holloway [1] lety přes Severní Atlantik, kde poptávka v létě dvakrát převyšuje poptávku oproti měsíci únoru. Lze tedy konstatovat, že letečtí dopravci spoléhají na třetí kvartál, který rozhodne o jejich zisku v daném roku (uvedené platí hlavně pro letecké společnosti působící na severní polokouli).

U týdenních špiček se očekává, že na mnoha dálkových letech dosahují letenky vyšší ceny obvykle o víkendu. Co se týče denních špiček, ta je obvykle ovlivněna časovým pásmem, kde se vytváří takzvaná preferovaná „okna“ pro přílety a odlety. Tato časová okna jsou ovlivněna preferencemi cestujících, provozním omezením letadel na letišti nebo kvótami pro pohyb letadel [1].

Jeden z vybraných městských párů bude v práci využit pro popsání nacenění letenek v průběhu několika měsíců.

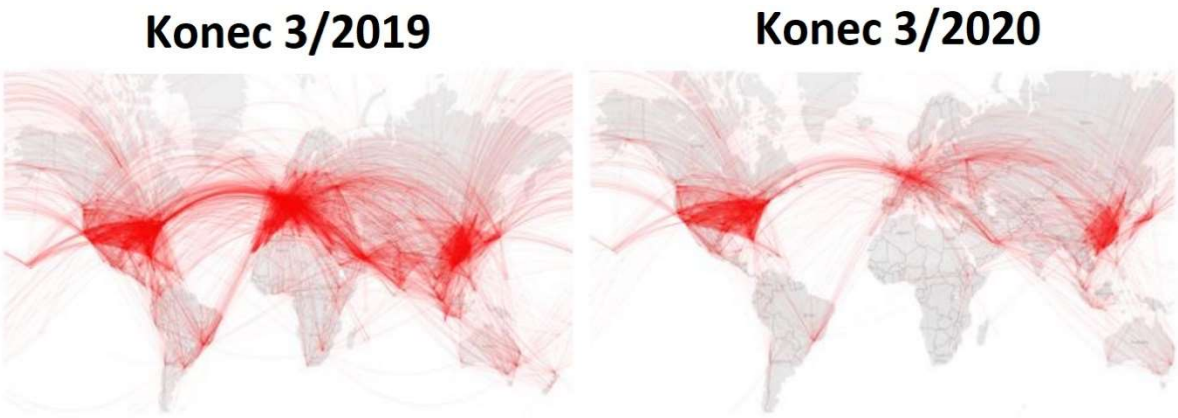
2.5. Výdaje při operování dálkových letů

Je třeba, aby do cenové politiky letenek na dálkových tratích, aerolinky zahrnovaly náklady, které, co se týče letů na dálkových tratích, jsou nemalé. Detailnější rozdělení výdajů je zobrazeno v Příloze 1, ze které vidíme srovnání třech možností přepravy na dálkových tratích. Z Přílohy 1 je zřejmé, že velkou část z celkových nákladů tvoří palivo, které tvoří 38,5 – 45,8 %. Obrázek je znázorněný pro let dlouhý 3200 námořních mil (5926,4 km). Na delších trasách bude poměr leteckého kerosinu v poměru s celkovými výdaji ještě výraznější. Na dálkových letech proto vidíme stále velkou dominanci klasických dopravců. Výskyt nízkonákladových aerolinek (LCC) nemá na klasické dopravce tak velký vliv jako u letů na kratší vzdálenosti, protože LCC zde nemají tak velký prostor pro ušetření výdajů za provoz. Za snížené pohodlí na úkor ceny, LCC dopravci ušetří na nákladech necelých 13 % oproti klasickým aerolinkám (u letů na krátkých vzdálenostech ušetří LCC 25–50 % výdajů oproti klasickým aerolinkám). Všechny tyto faktory vedou k tomu, že díky vysokým fixním výdajům za provoz jsou dálkové lety obsluhovány primárně klasickými leteckými dopravci [11] [12].

3. Uzlová přeprava mezi Evropou a Jihovýchodní Asií

Autor se pro analýzu letecké dopravy rozhodl zaměřit na oblast dálkové přepravy mezi Evropou a Jihovýchodní Asií, neboť autor má předky pocházející z této asijské oblasti a většinu svého života strávil na starém kontinentu. Autor této práce je dále pravidelným uživatelem produktů, které aerolinky působící na obou zmíněných kontinentech nabízejí.

Kritickým aspektem pro fungování uzlu je geografická lokalita, kterou proudí letecký tok. Dalším aspektem je ekonomický potenciál v oblasti kolem letiště (bezproblémová infrastruktura nebo dobré podmínky pro rozvoj). Je třeba, aby letiště jako celek spolupracovalo, čímž bude docházet k efektivnosti při plánování [13].



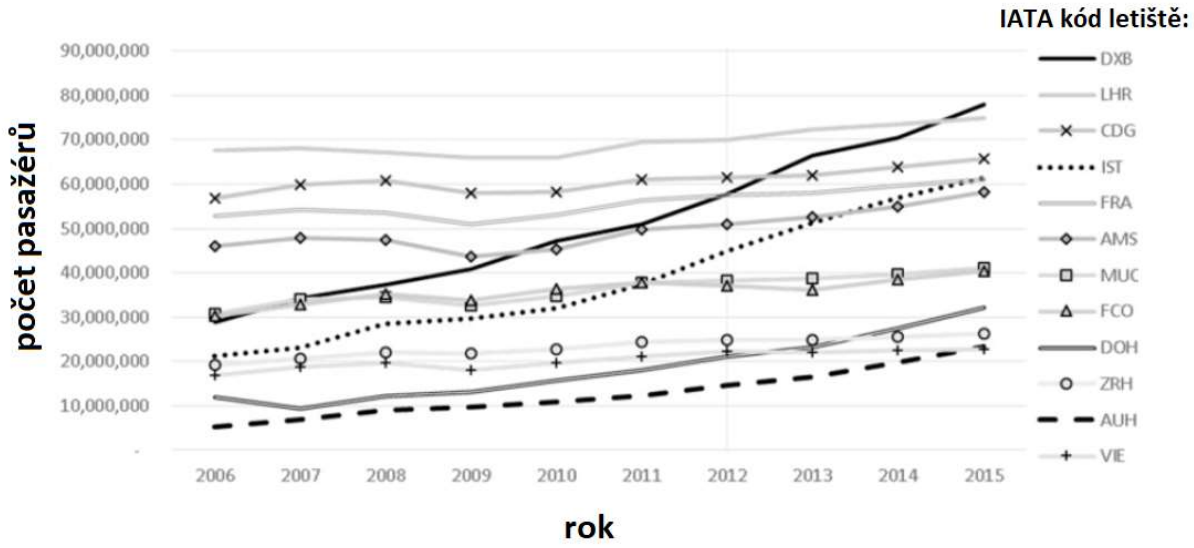
Obrázek 2.: Konektivita mezi městskými páry v období konce 3/2019 a konce 3/2020 [13]

Dle IATA [13] v průběhu posledních desetiletí byla letecká doprava čím dál tím dostupnější, a to nejen díky snížení nákladů za přepravu, ale také vzrůstající konektivitě. Mezinárodní přeprava do pandemie COVID-19 stále rostla téměř nepřetržitě. Avšak kvůli pandemii COVID-19 začaly vlády omezovat dopravu, aby zabránily šíření viru. Nyní dochází k postupnému zvýšení počtu cestujících, neboť státy již postupně ruší restrikce zavedené během pandemie. Z Obrázku 3 je patrný vliv blízkovýchodních uzlů z období konce března 2019, což potvrzuje důležitost těchto uzlů při cestách mezi oběma kontinenty. Zajímavým faktem je nárůst konektivity Blízkého východu o 40 % mezi lety 2014-2019, a proto je nyní domovem mnoha velkých uzlů [13].

Tyto letecké uzly slouží jako alternativa k přímým letům. Na letech s přestupem si postavilo svůj obchodní model několik blízkovýchodních aerolinií, což zapříčinilo vznik mnoha uzlů v tomto regionu. Jako příklad můžeme uvést aerolinku Emirates s přestupním uzlem na letišti v Dubaji (DXB), Etihad Airways s přestupním uzlem na letišti Abú Zabí (AUH) nebo Qatar Airways s přestupním uzlem na letišti Dauhá Hamad (DOH). Dalším rychle rostoucím leteckým uzlem je letiště Istanbul (IST), kde působí

aerolinka Turkish Airlines. S rostoucím vlivem výše uvedených leteckých společností evropské letiště pomalu ztrácí význam, a to na úkor výše zmíněných leteckých uzlů [14].

Díky liberalizaci letecké dopravy, došlo k nárůstu konkurence, což vyvolalo snížení ceny a zvýšení počtu cestujících na trase mezi Evropou a Asií. Tento trend měl však i opačný efekt, a to ztrátu důležitosti evropských uzlů na úkor uzlů blízkovýchodních. Blízkovýchodní letecké společnosti přeorientovaly své toky a tím nabídly lepší konektivitu než evropské letecké společnosti. Mezi lety 2009-2015 došlo k výraznému nárůstu konektivity a městských párů u blízkovýchodních uzlů, kdežto u evropských městských párů nedošlo k nárůstu nebo byl nárůst pouze malý. Ale i přesto nabízejí 4 hlavní evropská letiště (AMS, CDG, FRA a LHR) dohromady více leteckých spojení než Blízkovýchodní uzly. Důvodem je to, že blízkovýchodní uzly spojují hlavně důležité městské páry, ostatní městské páry jsou pro blízkovýchodní uzly méně důležité [14].



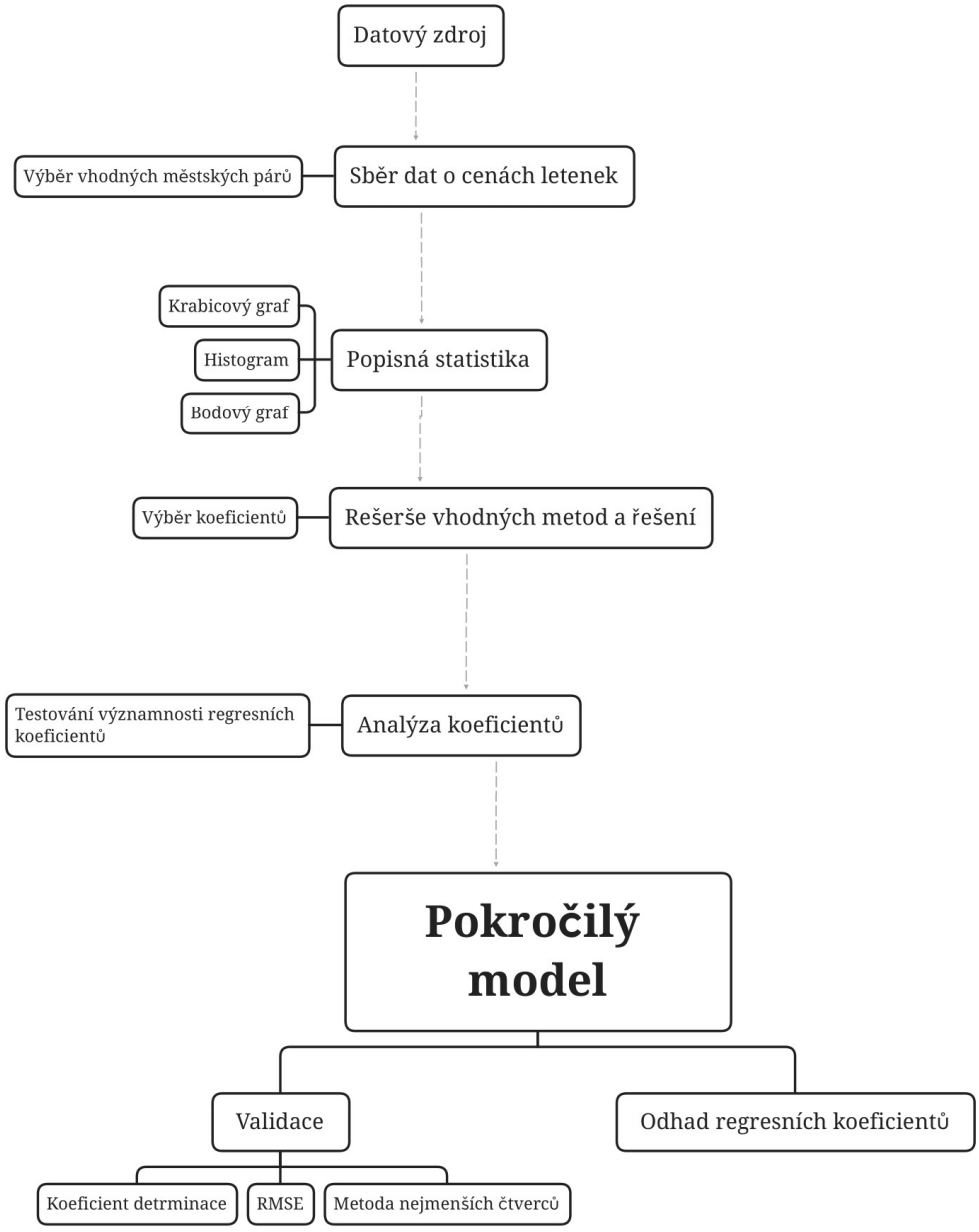
Obrázek 3.: Srovnání nárůstu počtu pasažérů v jednotlivých uzlech [15]

Z obrázku je patrný výrazný nárůst přepravených pasažérů na letišti DXB, kde mezi lety 2006–2015 došlo k více než zdvojnásobení počtu pasažérů, o více než trojnásobek pasažérů bylo na letišti DOH mezi lety 2007-2015, naopak si lze všimnout pomalý nárůst pasažérů na evropských uzlech LHR, CDG, FRA či AMS [15].

Tyto letecké uzly a s ním spojené aerolinky budou použity při výběru koeficientů pro vytvoření pokročilého modelu cen letenek.

4. Metodika k cenám letenek dálkové přepravy

Cílem diplomové práce je vytvořit pokročilý model pro zkoumání cen letenek na dlouhých tratích mezi různými typy letišť se zaměřením na různé segmenty cestujících a různé typy destinací. V rámci této kapitoly bude popsán popis metodiky k cenám letenek dálkové přepravy.



Obrázek 4.: Metodologický model zpracování

Při vytváření pokročilého modelu pro ceny letenek na dálkových tratích byl využit postup na Obrázku 4. V prvním kroku bylo potřeba provést sběr vhodných dat, v případě modelu práce to byla historická data o cenách letenek na dálkových tratích z vybrané webové stránky prodávající letenky. Tato data bylo nutné zpracovat pomocí metod popisné statistiky, které budou zmíněny v podkapitole 4.3. za účelem

vytvoření pokročilého modelu. Rešerše vhodných metod řešení byla provedena výběrem vhodných koeficientů, které byly analyzovány pomoci statistických testů. Na základě těchto testů byly vybrány nejvhodnější koeficienty pro sestavení pokročilého modelu. Z pokročilého modelu byly určeny odhady regresních koeficientů nezbytné pro provedení validace, která určí, zda by vypočtená data mohla odpovídat realitě. Validace byla provedena pomocí statistických testů, které byly následně porovnány s jednoduchým regresním modelem.

4.1. Výběr vhodného datového zdroje pro těžení dat o cenách letenek na dlouhých tratích

V této kapitole bude popsán výběr vhodného datového zdroje (tzv. dataset), který poslouží pro budoucí analýzu městských párů. Dále bude popsán způsob, jak byla tato data získána a ze získaných dat bude popsán proces zpracování těchto dat pro další analýzu.

4.1.1. Možnosti těžení dat

Pro získání informací o cenách letenek lze využít mnoho dostupných zdrojů. Jako první volbu lze data stahovat manuálně a lze je najít z těchto zdrojů [16]:

- z webových stránek leteckých společností (přímo z oficiálních stránek leteckých společností (např.: letenky od letecké společností British Airways ze stránek britishairways.com)
- z webových stránek cestovatelských portálů (OTAs), kteří zprostředkovávají prodej letenek od leteckých společností (Pelikán, Letuška apod.) díky přístupu ke globálnímu distribučnímu systému (Amadeus, Galileo).
- z cestovních agregátorů, kteří porovnávají ceny z webových stránek leteckých společností a OTAs (např. Skyscanner)
- z cestovních kanceláří, kteří nabízí kromě prodeje zájezdů i možnost koupit letenky samostatně

Druhou volbou, jak získat data, je nákup dat od společností zabývajících se těžením dat za poplatek. Tyto společnosti mají vytvořený svůj vlastní program pro těžení dat (např. Infare nebo OAG) [17].

4.1.2 Způsob těžení dat

Jako ověřený způsob těžení dat byl využit způsob těžení pomocí rozšíření pro prohlížeče WEB SCRAPER [18], který dokáže číst a následně stáhnout vybrané informace z webových stránek. Data byla stažena z webu Pelikan.cz, byl využit script Ing. Dominika Hankeho [19], který byl dále zdokonalen, a to získáním více dat jako jsou aerolinky či doba letu, které jsou důležité pro následnou analýzu. Získaná data byla uložena v excelovských tabulkách, která dále sloužila k analýze cen letenek na dálkových tratích.

Data pro ekonomickou třídu byla těžena v intervalu jednoho měsíce a pro business třídu byla těžena také po dobu jednoho měsíce. Těžení bylo na obou linkách provedeno stejným způsobem a to tak, že se v den těžení vloží webový odkaz nabídky letů, které jsou 31 dní před odletem. Další den se vloží k již existujícímu odkazu nový odkaz 31 dní před odletem. Odkaz vložený první den bude tedy odpovídat letům 30 dní před odletem (místo 31 dní před odletem). Tímto způsobem byl přidán průběžně každý den těžení jeden odkaz a v poslední den je již v kódu 30 odkazů, kdy poslední den těžení je přidán let jeden den před odletem.

U analýzy dálkových letů byl zvolen předpoklad evropského cestujícího, který jak u business třídy, tak v ekonomické třídě cestuje vždy z evropského letiště do vybraného asijského páru. Další předpoklad pro těžení dat vychází z předchozí kapitoly, kde bylo popsáno, že většina cestujících si na dálkových letech kupují zpáteční letenky. Proto v den těžení byl přidán webový odkaz na zpáteční let začínající 31 dní před odletem a návrat o týden později ve stejný den v týdnu (pokud byl let v pondělí, tak návrat je o týden později také v pondělí na vybraném městském páru). Dále pro analýzu byla těžena data přímých letů, letů s jedním přestupem a letů s dvěma přestupy.

Dále byly těženy ceny zpátečních letenek ekonomické třídy během intervalu několika měsíců (7/2022-3/2023), kdy autor vybral městský pár, na kterém těžil data podle letového řádu aerolinky, na kterém v kapitole 8 vytvořil pokročilý model. Kromě dat aerolinky se snažil autor natěžit i data jiných leteckých společností létajících ve stejných dnech jako vybraná aerolinka.

4.1.3. Výběr vhodných městských párů

Dalším krokem této práce bylo třeba vybrat vhodné městské páry, na kterých bude proveden proces těžení dat. V úvodu bylo řečeno, že tyto městské páry budou létat mezi Evropou a Jihovýchodní Asií.

Pro analýzu dálkových cestujících v ekonomické třídě byly využit pár VIE-BKK. Dle výzkumu BBC News [20] je Thajsko pro Evropu oblíbená turistická destinace (nejoblíbenější v Jihovýchodní Asii). Proto se předpokládá, že na tomto městském páru bude letiště BKK vstupní branou pro Evropany, kteří sem přijedou za účelem turismu. Z dat ATAG (Air Transport Action Group) [21] vyplývá, že tyto země mají dle dat z roku 2020 vysoký příjem z turismu, který tvoří 11,8 % HDP Rakouska a 19,7 % HDP v případě Thajska. Navíc se tyto země vyznačují skvělou konektivitou, kde Rakousko zaujímalo 12. pozici (přístup k 195 zemím pomocí přímých letů nebo lety s jedním přestupem) a Thajsko 16. pozici (přístup k 189 zemím pomocí přímých letů nebo lety s jedním přestupem). Letiště VIE má navíc výbornou vlakovou či autobusovou dostupnost do okolních zemí, a proto je pro daný region největším těžištěm [22].

Metoda těžení dat také odpovídá preferencím evropského cestovatele, neboť výzkum BBC News [20] uvádí, že zde cestovatelé tráví 1-2 týdny. Navíc byl tento pár vybrán z důvodu působení 2 leteckých

společností, které nabízejí přímé lety. Data v uvedeném městském páru byla sesbírána od 6. 10. 2022 do 6. 11. 2022.

Pro analýzu dálkových letů pro business cestující byl využit pár CDG – SIN, které mají důležitý význam jak z pohledu evropského (evropský letecký uzel), tak asijského (letecký uzel v Jihovýchodní Asii). Autor předpokládá, že mezi těmito uzly bude velký tok business cestujících, kromě toho byl vybrán tento uzel z důvodu působení dvou leteckých společností, které nabízejí přímé lety. Data byla sbírána od 11. 10. 2023 do 10. 11. 2023.

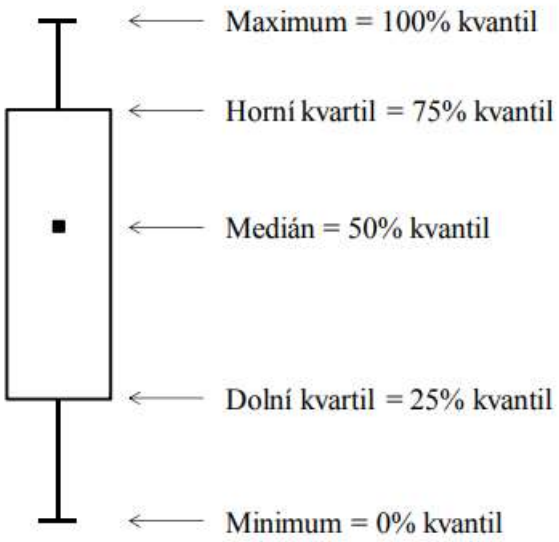
Jako třetí destinace byl vybrán městský pár PRG – HAN pro analýzu sezónní špičky. Jedná se o městský pár, který autor využívá k cestám za rodinou. Sledované období bylo 18. 7. 2023 – 9. 3. 2024.

4.2. Popisná statistika pro zpracování natěžených dat o cenách letenek

Ke zpracovávání naměřených dat o cenách letenek na dlouhých tratích bude využita popisná statistika, kde se ze získaných dat snažíme stručně a výstižně popsat pozorované výsledky. Během zpracování byly použity kvalitativní veličiny a grafické nástroje popisné statistiky, které budou popsány níže [23].

4.2.1. Grafické znázornění o cenách letenek pomocí popisné statistiky

Kvantitativní neboli numerická data lze charakterizovat číselnou hodnotou. Naše měřená data jsou diskrétní data, což jsou data na reálné ose zobrazena pomocí izolovaných bodů. Pro vizualizaci naměřených dat byl použit tzv. krabicový graf (box plot). Jednotlivé prvky krabicového grafu odpovídají významným kvantilům vypočtených na základě pozorovaných dat. Příklad krabicového grafu je uveden na Obrázku 5.



Obrázek 5.: Krabicový graf s vyznačením významných kvantilů pozorovaných dat [24]

Uvnitř obdélníkového tvaru je naznačena pozice mediánu (50 % kvantilu, jedná se o prostřední hodnotu) a obdélník samotný značí polohu dolního a horního kvartilu, tedy 25 % (čtvrtina hodnot je menších nebo rovných) a 75 % kvantilu (čtvrtina hodnot je větší nebo stejná). Tyto dva kvantily odpovídají kvantilovému rozpětí, které ohraničuje 50 % pozorovaných hodnot. Fousky dosahující za hranice obdélníkového tvaru pak signalizují polohu hodnot více vzdálených od mediánu, nejčastěji odpovídají 5 % kvantilu (spodní fousek) a 95 % kvantilu (horní fousek), případně minimu a maximu pozorovaných hodnot [23] [24]. V práci byla využita hlavně hodnota mediánu, která není ovlivněna extrémními hodnotami [25], dolní kvartil oddělující čtvrtinu nejnižších hodnot sloužil k analýze segmentu volnočasových cestujících (angl. leisure traveler), kteří se zaměřují hlavně na cenu. Maximum tedy 100 % kvantil bude sloužit například pro vyřazení letenek nad touto hodnotou, které jsou neúměrně drahé. Pro vytvoření pokročilého modelu byl vybrán segment cestujících citlivých na cenu, proto byla využita hodnota dolního kvartilu a mediánu pro popis těchto cestujících.

Dalším způsob grafického zpracování dat, které byly v této práci použity, byl histogramem nebo pomocí bodového grafu. Histogram dává nahlédnout, jak jsou jednotlivé hodnoty znaku v našem souboru rozloženy. Díky tomu lze zjistit, které hodnoty se často vyskytují a které zřídka nebo vůbec. Histogram se tvoří pomocí intervalu, které je zobrazen ve níže:

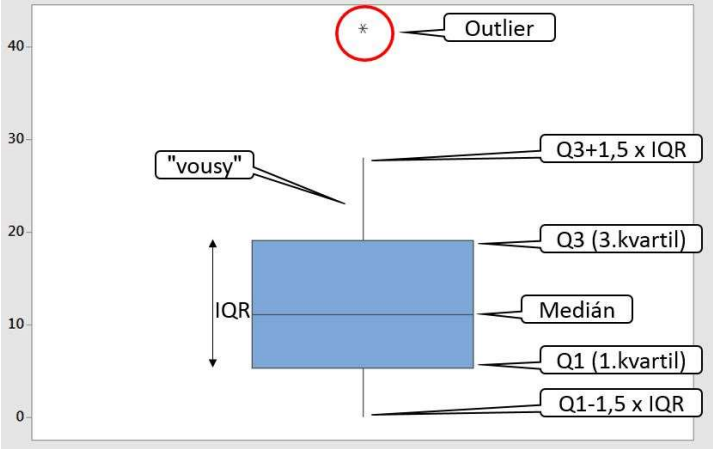
$$I = [a; b] \tag{4.1}$$

Kde *I* je interval, jež pokrývá celé rozmezí, *a* je počáteční hodnota intervalu a, *b* je koncová hodnota intervalu. S výjimkou prvního intervalu jsou všechny zbylé intervaly (sloupce) v histogramu uzavřené zprava. Histogram je tedy grafické znázornění intervalových četností, kde sloupec má výšku počtu pozorování v daném intervalu.

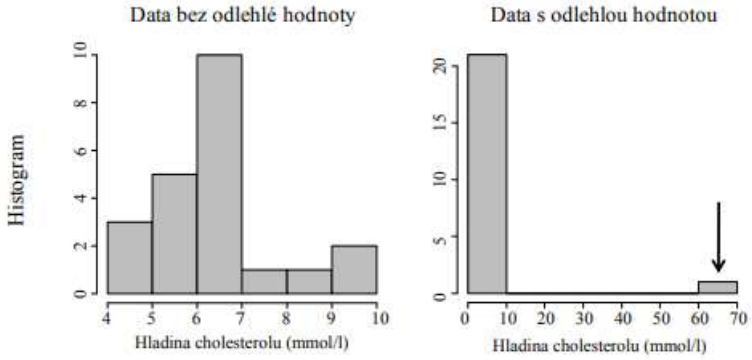
V případě bodový grafu, ten nám znázorňuje body v průsečíku číselných hodnot osy x a y [26]. V naší práci budeme využívat bodový graf k vizuálnímu pozorování shluků dat.

4.2.2. Odstranění přebytečných či nežádoucích cen letenek

Pro odstranění přebytečných hodnot (v dané práci se jedná o data o cenách letenek na dlouhých tratích) lze využít metody popisné statistiky, která bude založena na identifikaci odlehlých hodnot.



Obrázek 6.: Příklad odlehlé hodnoty krabicového grafu [24]



Obrázek 7.: Příklad odlehlé hodnoty histogramu [27]

Ideálními nástroji, pro identifikaci odlehlých hodnot, je pomocí sloupcového (histogram) či krabicového diagramu, které odhalí problematickou hodnotu jako nezvykle vzdálenou od pozorovacích hodnot. Příklad odlehlé hodnoty krabicového grafu je zobrazen na Obrázku 6 a příklad odlehlé hodnoty histogramu je zobrazen na Obrázku 7 [24] [27]. U náhodných veličin se k identifikace odlehlých hodnot používá bodový graf.

Je zřejmé, že zejména na větších datových souborech se nelze v identifikaci odlehlých hodnot obejít bez vizualizace a popisných statistik. Stejně tak se ale nelze obejít bez znalosti daného problému, která nám pomáhá se orientovat v tom, jaký je vůbec obor možných hodnot sledované náhodné veličiny. [28] Jak již bylo zmíněno autor práce bude využívat výše uvedené metody pro popis natěžených dat o cenách letenek na dlouhých tratích.

4.3. Analýza zpracovaných dat rešerše vhodných metod pro tvorbu pokročilého modelu

V této kapitole bude popsán způsob výběr koeficientů pro budoucí tvorbu pokročilého modelu vývoje cen letenek na dlouhých tratích. Tyto koeficienty budou dále testovány pomocí testu významnosti, aby mohly být využity pro vytvoření pokročilého modelu. Pro tuto práci byl zvolen Bayesovský model, který bude detailněji popsán v následující podkapitole 4.4.4.

4.3.1. Testování závislosti veličin ze zpracovaných dat o cenách letenek

Pro vytvoření jakéhokoliv regresního modelu je důležitou součástí test závislosti veličin. Tato analýza nám ukazuje, zdali je náhodná veličina Y ovlivňována náhodnou veličinou X a naopak. Vztah mezi těmito veličinami může mít různou intenzitu, a to od úplné nezávislosti až po úplnou (funkční) závislost.

Pearsonův korelační koeficient patří mezi nejdůležitější míry vztahu dvou náhodných spojitých proměnných x a y. Počítáme je z n párových hodnot (x_i ,y_i) naměřených na n jednotkách náhodně vybraných prvků. Korelační koeficient může nabývat hodnot mezi -1 a 1. Tento koeficient r se počítá pomocí tzv. kovariance s_{xy} a směrodatných odchylek s_x a s_y obou proměnných. Vzorec je vyjádřen v rovnicích (4.2) a (4.3):

$$s = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})}{n - 1} \tag{4.2}$$

$$r_{xy} = \frac{s_{xy}}{s_x * s_y} \tag{4.3}$$

Kde:

s – směrodatná odchylka

s_{xy} – kovariance

$\bar{x}$ – průměrná hodnota x

$\bar{y}$ – průměrná hodnota x

n – počet pozorování

s_x – směrodatná odchylka x

s_y – směrodatná odchylka y

Počítáme je z n párových hodnot (x_i,y_i) naměřených na n jednotkách náhodně vybraných prvků. Korelační koeficient může nabývat hodnot mezi -1 a 1. Tento koeficient r se počítá pomocí tzv. kovariance s_{xy} a směrodatných odchylek s_x a s_y obou proměnných.

Správná interpretace Pearsonova korelačního koeficientu předpokládá, že obě proměnné jsou náhodné veličiny a mají společné dvourozměrné normální rozdělení. Nulový korelační koeficient pak znamená, že veličiny jsou nezávislé. Pokud není splněn předpoklad dvourozměrné normality, nelze z nulové hodnoty korelačního koeficientu usuzovat, zda veličiny jsou či nejsou korelované [29].

Dle Evansovy příručky [30] poznáme korelační vazby jako:

0,00 - 0,19 „velmi slabá“

0,20 - 0,39 „slabá“

0,40 - 0,59 „střední“

0,60 - 0,79 „silná“

0,80 - 1,00 „velmi silná“

Hodnoty korelace se vyskytují v interval mezi -1 a 1.

Autor práce bude považovat statisticky významné veličiny ty, jejichž závislost bude od intervalu 0,4-1 (střední korelace – velmi silná korelace).

4.3.2. Výběr modelů pro vizualizaci vývoje cen letenek na dálkových tratích

Výběr modelu je dalším důležitým krokem pro vytvoření vhodného vzorce, který popíše vývoj cen letenek na dlouhých tratích. Pro letenky je vhodný vybrat lineární regresní model, který naměřená data proloží křivkou, která následně popíšou nasbíraná data.

Při vytváření modelu cen letenek na dálkových tratích se autor inspiroval zdrojem, který se zabýval predikování pomocí lineárních modelů často používané pro přesné předpovězení cen letenek. Protože autor nemá dostatečné zkušenosti s vytvářením regresního modelu založený na strojovém učení. Rozhodl se autor na základě konzultace vytvořit pokročilý model postavený na Bayesově větě pro odhad parametrů modelu, který je dle doc. Ing. Evženie Uglickich, CSc. vhodný pro analýzu cen letenek na dlouhých tratích. Dále byl kromě Bayesovského modelu vytvořen jednoduchý lineární model postavený na Metodě nejmenších čtverců (MNČ), se hojně využíval v předchozích studiích pro predikci cen letenek.

4.3.3. vytvoření jednoduchého modelu pro popis cen letenek na dálkových tratích

Pro zpracovaná data o cenách letenek na dlouhých tratích byl vybrán jednoduchý model postavený na Metodě nejmenších čtverců. Tento model je považován za deterministický, neboť obsahuje deterministickou funkci postavenou pouze s naměřenými daty (Bayesovský model má navíc vliv neurčitosti, který bude popsán v další podkapitole). Metoda je postavena na lineární regresi, kde vztah mezi závislou a nezávislou proměnnou lze vyjádřit pomocí následující rovnice (4.4):

$$y=mx+b \tag{4.4}$$

kde m a b jsou regresní koeficienty, y je nezávislá proměnná a x je závislá proměnná.

Pokud má rovnice více závislých proměnných (v uvedeném případě *n* proměnných) bude vzorec z rovnice (4.5) vypadat následovně:

$$y = b_0+b_1x_1 + b_2x_2 + \ \ldots + b_nx_n \tag{4.5}$$

Metoda nejmenších čtverců minimalizuje součet čtvercových reziduí. Rezidua jsou odchylky mezi měřenou hodnotou a odhadovanou hodnotou. Díky tomu se sestrojí přímka, která bude procházet kartézkou souřadnicí reprezentující [31].

Jedním ze způsobů, jak postupovat při použití Metody nejmenších čtverců je pomocí maticového násobení a je popsán následujícím vztahem

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \; Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \; A = \begin{bmatrix} b \\ m \end{bmatrix} \; E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{4.6}$$

$$Y=XA+E \tag{4.7}$$

Kde matice Y jsou hodnoty y (nezávislá data) ve sloupcích, matice X se skládá z prvního sloupce obsahující 1 a závislých dat v řádcích. E je matice chyb. Ma;tice A je matice obsahující koeficienty odhadované funkce a je vypočítána následujícím způsobem [31]:

$$A = (X^TX)^{-1}X^TY \tag{4.8}$$

Daný model bude autor využívat pro validaci s pokročilým modelem, neboť je tento model vhodný pro analýzu podobných dat [32].

4.3.4. Vytvoření pokročilého modelu pro popis cen letenek na dálkových tratích

Pro vytvoření pokročilého modelu byla zvolena regresní analýza pomocí Bayesova vzorce, který vyjadřuje vztah mezi apriorní (třeba historická data) a aposteriorní hustotou pravděpodobnosti. Bayesův vzorec popisuje vývoj hustoty pravděpodobnosti na parametrech pro průběžně měřená data. Oproti Metodě nejmenších čtverců je Bayesovský model stochastický model, neboť je model pod vlivem neurčitosti (prvek nejistoty). Stochastický model je charakterizován tím, že dovolují poměrně přesnou matematickou manipulaci se vztahy s veličinami, i když ve skutečnosti platí tyto vztahy pouze přibližně.

Vzorec (4.9) pro Bayesovo pravidlo (Bayesův vzorec) říká:

$$f\left(\theta\middle|d(t)\right) \propto f(y_t|\psi_t\theta) \quad f(\theta|d(t-1)) \tag{4.9}$$

Kde:

$f\left(\theta\middle|d(t)\right)$– aposteriorní inverzní Gauss-Wischartovo rozdělení (aposteriorní GiW) v čase *t*

$f(y_t|\psi_t\theta)$ – model

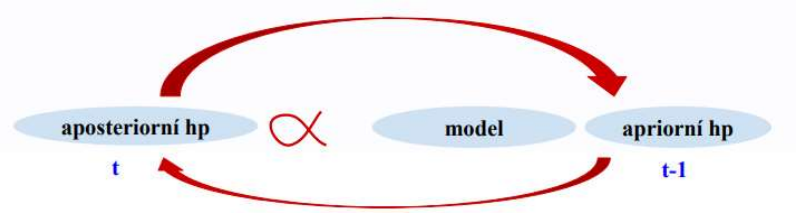
$f(\theta|d(t-1))$ – apriorní inverzní Gauss-Wischartovo rozdělení (apriorní GiW) v čase *t*

Apriorní GiW je hustota pravděpodobnosti, ze které počítáme rekurzivně aposteriorní GiW. Aposteriorní GiW je hustota pravděpodobnosti, která dává kompletní stochastický odhad parametrů založený na počáteční znalosti průběžně měřených datech. Na Bayesovém vzorci je postaveno bayesovské odhadování, kde bodový odhad je číslo (vektor), který vyjadřuje optimální hodnotu parametru.

Bodový odhad je 0-kroková predikce. Bodová predikce je analogie bodového odhadu parametrů modelu, Podle zvoleného kritéria se vypočítá hodnota, která nejlépe vyjadřuje odhad budoucí hodnoty výstupu. Predikce lze určit jednokrokovou nebo vícekrokovou předpovědí s regresním modelem. Vzorec (4.10) střední hodnoty pro spočítání n-krokové prediktivní pravděpodobnosti:

$$\hat{y}_{t+n} = \int_{y_{t+n}} y_{t+n} f(y_{t+n}|d(t))\ dy_{t+n} \tag{4.10}$$

Kde y_i postupně nahrazujeme jejich predikcemi $\hat{y}_i$ spočítané v minulém kroku. [33; 34]



Obrázek 8.: Bayesovo pravidlo [35]

Bayesovský model bude vytvořený na základě Bayesovém pravidle založeném na rekurzi (viz. Obrázek 8), kde se budou reprodukovat apriorní (naměřená data) hustota pravděpodobnosti a aposteriorní hustota pravděpodobnosti (odhadovaná data). Data se budou postupně dosazovat a bude se postupně posunovat čas t, během tohoto procesu se bude model upřesňovat (aktualizovat). Rekurze bude provedena pomocí algoritmu odhadu, kde po proběhnutí rekurze vznikne aktualizovaná matice (tzv. Update statistik). Tato matice obsahuje důležité informace, která se následně rozloží. Výsledkem bude bodový odhad regresních koeficientů, které jsou nezbytné pro odhad vytvořeného modelu [35].

4.4. Validování výsledku získaných z modelů

Pro validaci vytvořených modelů byly vybrány 2 metody, které určují výkonnost statistických modelů. Jedná se o koeficient determinace a odmocnina ze střední kvadratické chyby. Tyto metody vyhodnocování jsou velice populární a budou popsány níže [32].

4.4.1. Určení přínosnosti modelu pomocí Koeficientu determinace (R^2)

Jako první metoda byl zvolen koeficient determinace (také „R kvadrát“ nebo R^2). Tato veličina udává přínosnost modelu. Ukazuje, jakou část z celkové variability v odhadovaných datech bylo vysvětleno díky vytvořenému modelu.

Vzorec (4.11) pro určení koeficientu determinace:

$$R^2 = 1 - \frac{S_e}{S_T} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (4.11)$$

Kde:

S_e je celková variabilita výsledkové proměnné

S_T je reziduální součet čtverců

y_i jsou data nezávisle proměnné

$\bar{y}_i$ průměr nezávisle proměnných dat

$\hat{y}_i$ jsou odhadovaná data

Výsledek koeficientu osciluje mezi 0 a 1, kde, čím blíže se model blíží k hodnotě 1 tím více se model hodí k proměnné, kterou se snažíme vysvětlit. Hodnoty blížící se k nule pak značí nespolehlivost modelu [36] [37].

V autorově práci bude provedeno srovnání hodnoty R^2 mezi vytvořeným pokročilým modelem a modelem vytvořený na Metodě nejmenších čtverců. Dále bude získaná hodnota validována s modely zabývajícím se zpracování dat o cenách letenek a na základě výsledku budou vyvozeny závěry.

4.4.2 Určení úspěšnosti vytvořených modelů pomocí RMSE

Odmocnina ze střední kvadratické chyby známé pod zkratkou RMSE (root mean square error) je nástroj, pomocí který se používá k vyhodnocení úspěšnosti výsledku simulací s chybějícími hodnotami. RMSE představuje průměrnou odchylku odhadovaného parametru od konečného počtu realizací. Vzorec (4.12) pro RMSE je [38] [39]:

$$RMSE(\theta, q) = \sqrt{\frac{1}{N}\sum_{v=1}^{N}\left(\theta - \theta_{vq}\right)^2} \quad (4.12)$$

Kde:

θ – parametr

θ_{vq}- odhad parametru θ metodou q ve výběru v

N – počet pozorování

Cílem je, aby hodnota RMSE byla co nejblíže k 0 (nabývá hodnot 0 až nekonečno). Čím nižší je hodnota, tím lepší je výsledný model. Není určeno, jaká hodnota RMSE je dobrá, ale při použití RMSE u jednotlivých modelů lze pozorovat, který model je přesnější na dané datové sadě. Nelze použít pro porovnání různých datových sad mezi sebou [40].

V autorově práci bude provedeno srovnání hodnoty RMSE mezi vytvořeným pokročilým modelem a modelem vytvořený na Metodě nejmenších čtverců.

5. Analýza městského páru Vídeň (VIE) – Bangkok (BKK)

V průběhu analýzy působilo na městském páru VIE – BKK 2 dopravci, které nabízeli přímý let. Jmenovitě se jednalo o společnosti Austrian Airlines (IATA kód OS) a Eva Air (IATA kód BR), které budou popsány v další podkapitole. Kromě těchto přímých letů byla možnost využít lety s jedním či dvěma přestupy od evropských či asijských leteckých dopravců.

5.1. Austrian Airlines

Jedná se o rakouskou leteckou společnost patřící do skupiny Lufthansa Group. Aerolinka nabízí lety do přibližně 130 destinací, které zajišťuje 360 lety denně. Báze aerolinky je na letišti VIE a je členem Star Alliance [41].

Všechny přímé lety byly v době těžení obsluhovány letouny typu Boeing 777-200, jedná se o širokotrupé letadlo s kapacitou 306 pasažérů (38 sedaček v business třídě, 24 sedaček v premium economy a 244 v ekonomické třídě) [42]. Přímé lety na této lince létaly pod označením:

- Let OS25, který v době sběru létal jednou týdně kromě pátku, a to vždy ve 23:25 z VIE a příletem v následující den v 15:30 v BKK [43].
- Let OS15, který v období sběru odlétal vždy každý pátek a to ve 20:15 z VIE a přílet následující den v 11:15 [43].
- Let OS26, který létal každý den v týdnu kromě soboty a to ve 23h 50min z BKK a příletem následující den v 5:25 [43].
- Let OS16, který probíhal týdně vždy v sobotu s odletem ve 14 h z BKK a příletem na letiště VIE v 19h 35min [43].

Ze serveru Flightradar24 [43] bylo zjištěno, že letadla, která se používala pro tento let měly stáři od 16-24 let.

5.2. EVA Air

Jedná se o tchajwanskou leteckou společnost a zároveň je sesterská společnost od Evergreen Line zaměřující se na kontejnerovou přepravu. Je také členem Star Alliance. Letecká báze se nachází na letišti TPE [44]. Z toho vychází zajímavost využití jejích letadel, který má přesah mimo městský pár VIE–BKK kde působily tyto lety:

- let BR62, který působí nejen na analyzovaném městském páru, ale tento let obsahuje na letišti BKK mezizastávku (trvající 1h 40min). Daný let pak dále pokračuje na letiště TPE [45].

- let BR61 s mezizastávkou v BKK činí 1 hodinu a 20 minut. Na těchto letech byly použity modernější typy letounů než u přímých letů s Austrian Airlines a to letadlo Boeing 787-10 se stářím 3 roky.

Let BR62 probíhal týdně vždy v úterý, ve středu, v pátek a v neděli. s kapacitou 342 pasažérů (34 Royal Laurel class a 308 economy). Z pohledu komfortu se třída Royal Laurel považuje za business class [43] [46].

Let BR61 o stejné konfiguraci sedadel probíhal týdně vždy v úterý, ve středu, v pátek a v neděli s odletem ve 2h 45min z BKK a přistáním ve VIE podle letového plánu v 8:25.

5.3. Analýza letecké poptávky městského páru VIE - BKK

Pro lepší představu byla vytvořena Tabulka 1 obsahující přímé zpáteční lety obou aerolinek. Na které bude analyzována poptávka:

Tabulka 1: Základní informace o přímých letech mezi VIE a BKK

Číslo letu	Městský pár	Aerolinka	Typ letounu	Kapacita (PAX)	počet letů týdně
OS25	VIE - BKK	Austrian Airlines	Boeing 777-200	306	6
OS15	VIE - BKK	Austrian Airlines	Boeing 777-200	306	1
BR62	VIE - BKK	EVA Air	Boeing 787-10	342	4
OS26	BKK - VIE	Austrian Airlines	Boeing 777-200	306	6
OS16	BKK - VIE	Austrian Airlines	Boeing 777-200	306	1
BR61	BKK - VIE	EVA Air	Boeing 787-10	342	4

Z nejnovějších dat od IATA [8] je uveden průměrný load factor 78,9 %, tuto hodnotu tedy použijeme pro náš další výpočet load factoru na páru VIE - BKK. Při náhledech do statisik vládní agentury Eurostat [47] byla zjištěn počet cestujících mezi daným městským párem. Ze statistik byla zjištěna data za období 8/2022–1/2023, kde byly měsíce sběru (11/2022 a 12/2022) jedny z nejvytíženějších. Protože jsme těžili pouze jeden měsíc mezi obdobím 11/2022 až 12/2022, bude zvolen aritmetický průměr těchto dat. Vypočítaná kapacita letů je zobrazena v Tabulce 1 a srovnání přímých letů s daty z Eurostatu [47] je vyobrazeno v Tabulce 2. Z přímých letů nelze zjistit kolik procent cestujících využívá letiště BKK či VIE

jako cílovou destinaci či jako přestupní, neboť je letiště BKK také asijským uzlem. Lze však usoudit, že přepravní kapacita na dané lince odpovídá datům z Eurostatu [47], neboť mnoho cestujících bude také využívat lety s přestupem díky silné pozici blízkovýchodních uzlů.

Tabulka 2: Celková týdenní kapacita

Číslo letu	Celková týdenní kapacita*0.789*(31/7) (PAX)
OS25	6 415
OS15	1 069
BR62	4 780
OS26	6 415
OS16	1 069
BR61	4 780

Z výsledků z **Tabulky** lze vypozorovat, že kapacita sedaček přibližně odpovídá statistikám z Eurostatu [47], autor práce neví, kolik cestujících na těchto přímých letech využívají let jako cílovou destinaci nebo pouze přestupují.

Tabulka 2: Přímé lety vs data z Eurostatu

Odlety a přílety VIE	Měsíční kapacita (PAX)
Přímé lety VIE-BKK	12 264
Přímé lety BKK-VIE	12 264
data z Eurostatu VIE-BKK s tranzitem	15 174 [47]
data z Eurostatu BKK-VIE s tranzitem	13 649 [47]
data z Eurostatu VIE-BKK bez tranzitu	14 827 [47]
data z Eurostatu BKK-VIE bez tranzitu	8 629 [47]

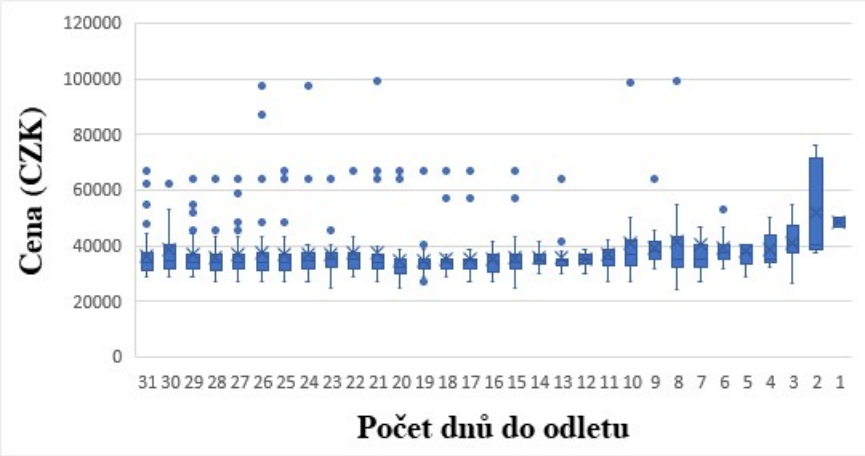
5.4. Zpracování dat přímého letu

Doba přímého letu je z VIE do BKK 10h 05 min (platí pro Austrian Airlines i Eva Air) a při opačné cestě z BKK do VIE 11h 35 min s Austrian Airlines respektive 11h 40min s Eva Air. Celkem bylo za sbírané období natěženo 1 054 možných zpátečních letů.

Dále byly nabízeny lety s jedním či dvěma přestupy. Vzhledem k tomu, že let s přestupem vyžaduje čekání na jednom či dvěma letišti je třeba brát v úvahu to, že se na tranzitních letištích cestující zdrží. Celkem během doby těžení dat nasbíralo 13 593 možných kombinací letů s jedním přestupem a 3 331 letů s 2 přestupy.

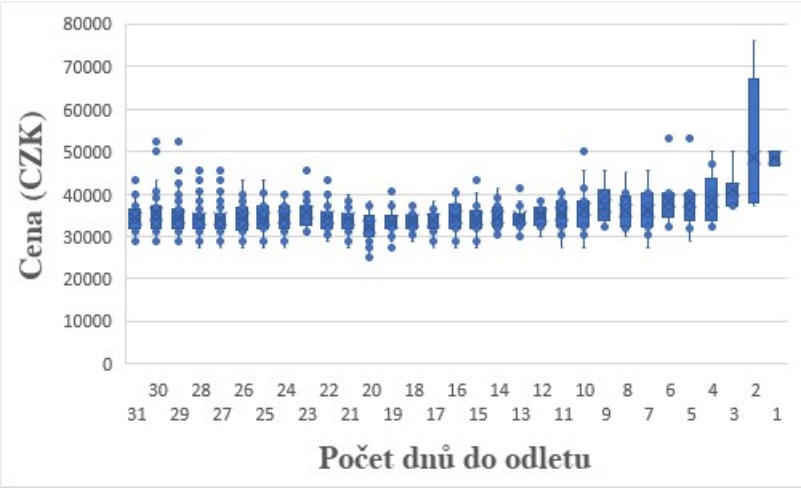
5.5. Analýza přímého letu

V úvodu se bude autor graficky znázornit a popsat průběh vývoje přímých zpátečních letenek pomocí krabicových grafů.

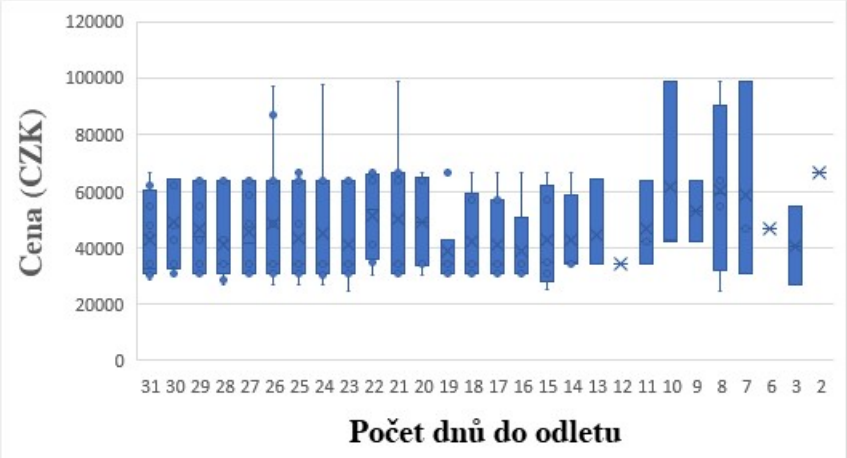


Obrázek 9.: Rozpětí cen přímých zpátečních letů v závislosti na době před odletem VIE – BKK

Tento graf na Obrázku 9 vyjadřuje rozpětí letenek obou přímých linek, kde, pokud se podíváme na pozorněji, můžeme si všimnout několika souvislostí. S tím, jak se blížila doba před odletem se začal objevovat zřejmý výkyv cen, který je patrný od 15. dne před odletem, kdy medián začal nabývat vyšších hodnot se snižujícím se počtem nabízených letů. Pokud se podíváme na graf přímých letů, tak mezi 31. a 15. dnem před odletem byl medián letenek mezi částkou 32 269 Kč a 34 429 Kč, pomineme-li 20. den před odletem (32 269 Kč), tak je rozdíl mediánu pouze 383 Kč. Od 15. dne se cena začíná zvyšovat a posledním dnem byl medián cen letenek již na částce 48420 Kč za přímý let. Pro další analýzu představu byl Obrázek rozdělen zvlášť na další 2 krabicové grafy obsahujíc data z Austrian Airlines, respektive Eva Air.



Obrázek 10.: Rozpětí cen zpátečních letenek Austrian Airlines z VIE do BKK



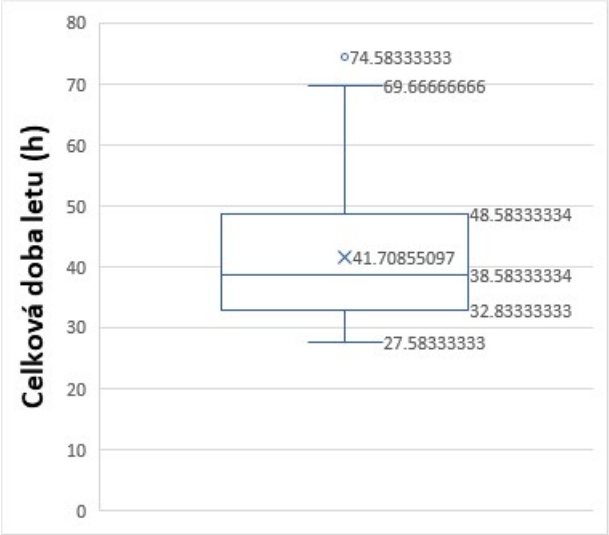
Obrázek 11.: Rozpětí cen zpátečních letenek Austrian Airlines z VIE do BKK

Po rozdělení dat přímých letů obou aerolinek na 2 grafy (Obrázek 10 a Obrázek 11) si můžeme všimnout několika souvislostí. Medián cen letenek nabízený společností EVA Air byla 34 398 Kč (206 nabízených letů) a společností Austrian Airlines byla 34 131 Kč (848 nabízených letů). Austrian Airlines nabízí více termínů, tím se předpokládá, že obsazuje i dny, které jsou méně vytížené. Naopak Eva Air létá pouze čtyřikrát týdně. Vidíme zde podobnost v nabízených cen letenek, kdy se medián skoro neliší. Důvodem může být to, že se aerolinky dohodli na průměrné ceně letenek nebo to, že je na této trase velká konkurence a aerolinky si nedovolí výrazně naceňovat svoje letenky. Autor práce usuzuje, že je na tomto městském páru použita strategie tzv. price-matching, neboť na této lince působí rakouský národní dopravce, který může být pro zákazníky výhodou. Aerolinka EVA Air proto za podobnou cenu nabízí oproti Austrian Airlines novější flotilu (stáří 3 roky ve srovnání s 16-24 let), který by mohl být jedním z důvodu úspěchu této aerolinky na daném městském páru.

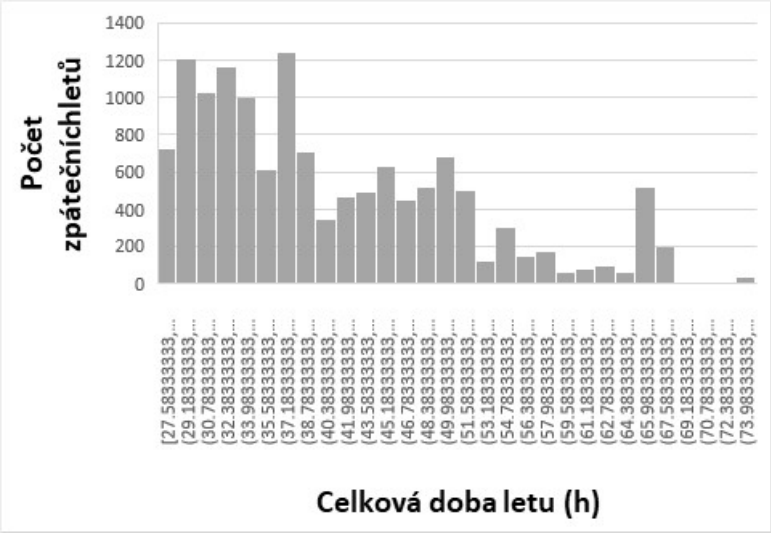
5.6. Zpracování dat letů s jedním přestupem

V této části práce se bude autor zabývat analýzou dat o cenách letenek s jedním nebo dvěma přestupy mezi VIE – BKK. Při analýze ideální doby, kterou je ochoten cestující strávit na letišti se názory různí. Z datového souboru bylo také velmi pracné tyto data získat, proto se autor rozhodl identifikovat vhodné lety pro budoucí analýzu městského páru VIE – BKK pomocí statistických metod. Z těchto metod autor odstraní odlehlá data, která nejsou vhodná pro budoucí analýzu, ať už z důvodu nepřiměřeně vysoké ceny oproti přímým letům nebo příliš dlouhé celkové doby letu.

Pro účely filtrace letů s velice dlouhou dobou (čas strávený během letu + čekání na letišti) byl vypracován krabicový graf a histogram celkové doby letu s jedním přestupem.

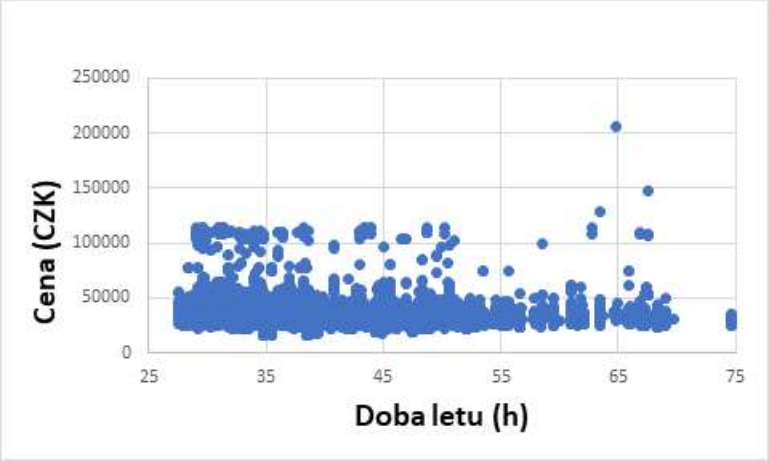


Obrázek 12.: Krabicový graf doby letu



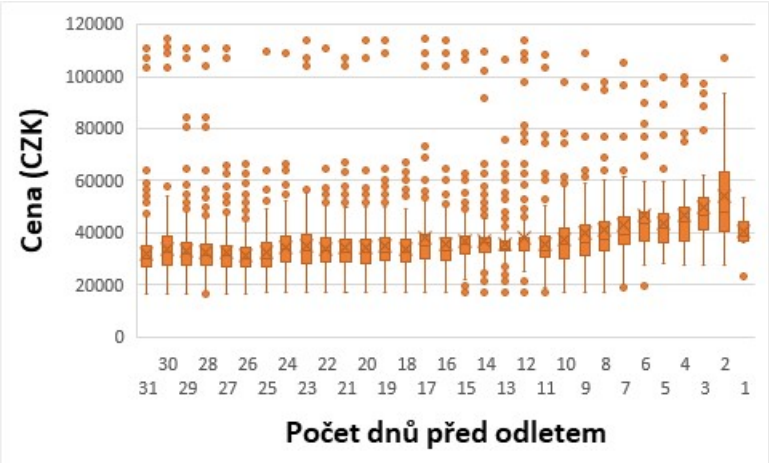
Obrázek 13.: Histogram závislosti počtů zpátečních letů na dobu letu

Bylo zjištěno, že horní kvartil všech letů s jedním přestupem, měl celkovou dobu letu mezi 27,58 hodin (nejkratší zpáteční let s Turkish Airlines) a 48,58 h. Z histogramu (Obrázek 13) je zřejmé, že data z horního kvantilu budou poměrově stačit ve srovnání s počtem dat bez přestupu.



Obrázek 14.: Bodový graf závislosti ceny na dobu letů (před filtrací dat)

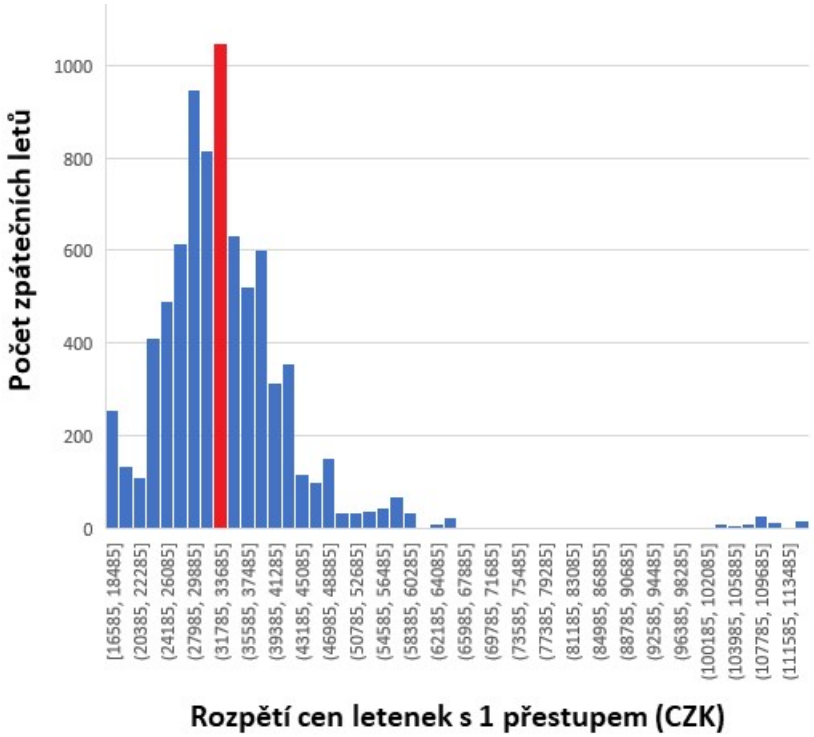
Z bodového grafu (Obrázek 14) je patrné, že lety s jedním přestupem, které trvají déle než 48,58 h neznamenají automaticky nižší cenu letenky za získané pohodlí, naopak vidíme v některých případech i vyšší cenu, než u letenek s nižší celkovou dobou letu a přestupem. Autor se tedy rozhodl pro vyřazení všech zpátečních letů trvající více než 48,58 h.



Obrázek 15.: Krabicový graf letů s jedním přestupem po filtraci celkové doby letu

Z krabicového grafu na Obrázku 15 lze vypozorovat, že filtrace doby letu je stále nedostačující, neboť jsou pozorovatelné odlehlé cenové hodnoty letenek, které je třeba také odstranit. Dalším krokem bude proto tyto odlehlé zpáteční ceny letenek z grafu odstranit. Krabicový graf dále určuje, že od 14 dne dochází k vyšší fluktuaci cen. Autor se rozhodl, že tento, že pro další analýzu rozdělí data na 2 části a to 31 až 15 dnů před odletem a 14 dnů až 1 dnem před odletem. U druhé části lze vypozorovat úbytek letenek za nižší cenu, které se postupně stávají odlehlými hodnotami, a naopak ceny letenek s vyšší cenovkou začínají být relevantní.

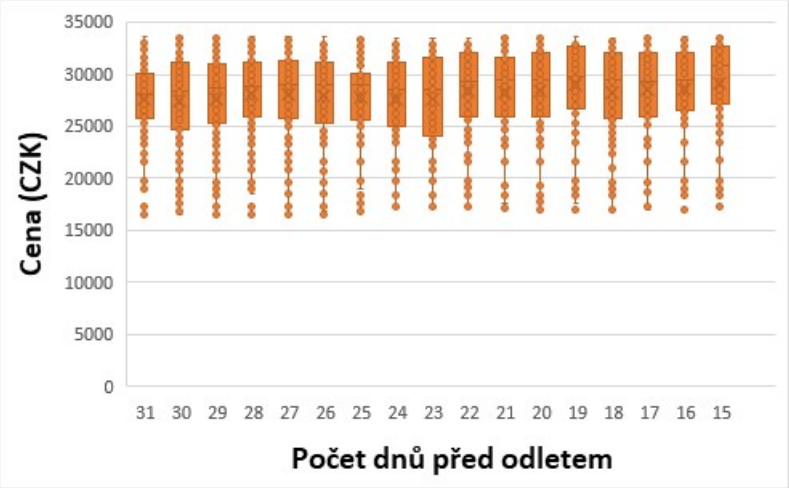
Mezi 31 dnů a 15 dnů před odletem lze vypozorovat, že ceny letenek nad 100 197 Kč lze považovat za odlehlou hodnotu a pro další práci budou odstraněna.



Obrázek 16.: Histogram závislosti počtů zpátečních letů na cenu letenky

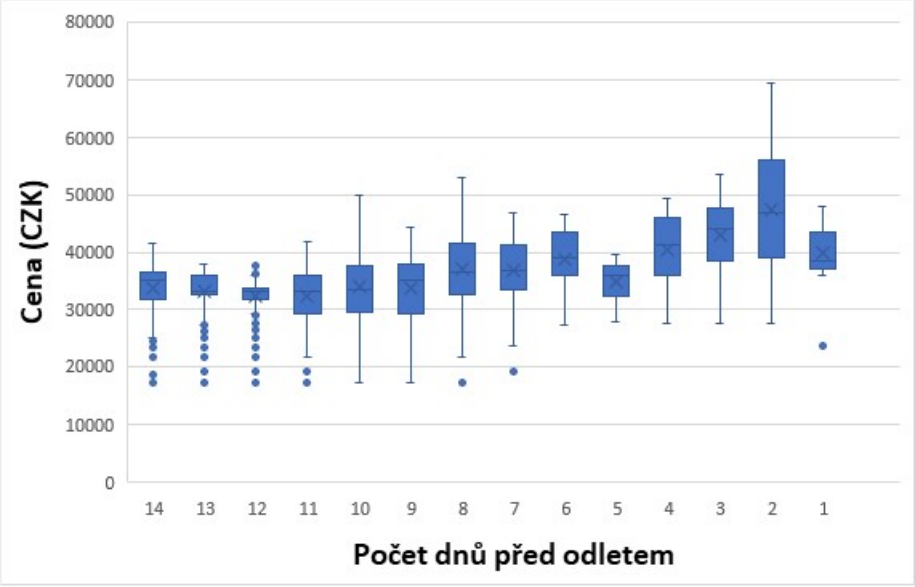
Ze sloupcového grafu (Obrázek 16) mezi 31 dnů a 15 dnů před odletem lze vypozorovat, že ceny letenek nad 100 197 Kč lze považovat za odlehlou hodnotu a pro další práci budou odstraněna.

Další důležitý nález, který je potřeba zmínit je, že počet zpátečních letenek s cenou od 33 685 Kč se začíná dramaticky snižovat (viz. červený sloupec na Obrázek 16). Autor usuzuje z toho, že na této lince působí 2 letecké společnosti nabízející přímé lety, kde cenový medián těchto letenek mezi obdobím 31 až 15 dnů před odletem se pohybuje na ceně mezi 32 269 Kč a 34 429 Kč. Předpokládá se, že cestující nebudou kupovat letenky dražší, než nabízí 2 letecké společnosti, které poskytují komfortnější cestu bez přestupu. Proto byla hranice pro filtrace ceny zvolená hodnota 33 685 Kč. Výsledek po zpracování filtrace je vyjádřený v krabicových grafech níže.



Obrázek 17.: Krabicový graf letů s jedním přestupem kompletním zpracováním 31 až 15 dnů před odletem

Od 15. dne lze vypozorovat větší poměr letenek s vyšší cenovkou, proto je potřeba na období mezi 14 dny a jedním dnem před odletem podívat podrobněji. Hranici pro filtraci cen zvolil autor hranici 100 % kvantilu cen přímých letů (nejdražší zpáteční přímé letenky pro daný počet dnů před odletem). Výsledek po filtraci je pak vyjádřen na Obrázku 18.



Obrázek 18.: Krabicový graf letů s jedním přestupem kompletním zpracováním 14 dnů až den před odletem

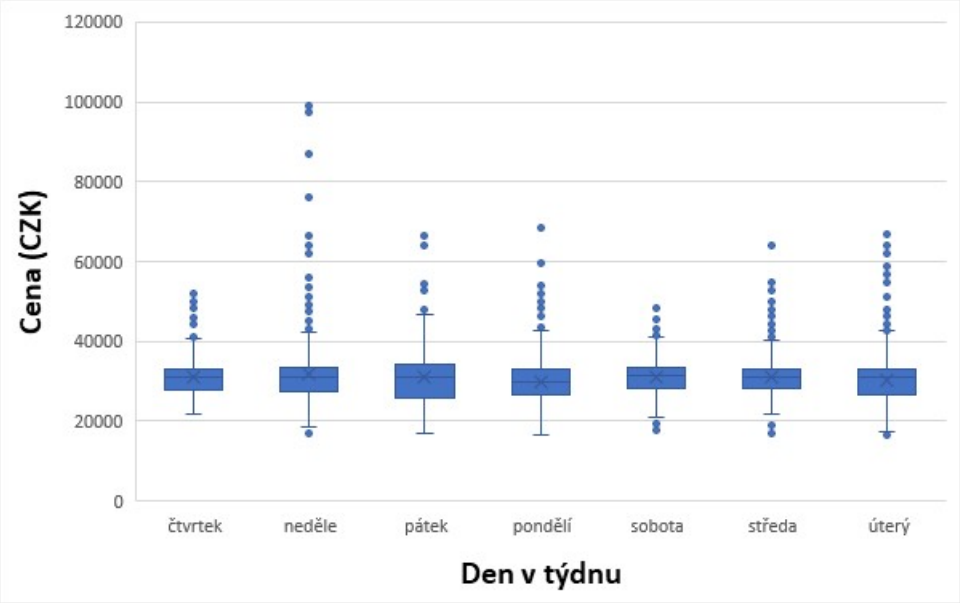
Pokud se podíváme na letenky mezi 14 a 1 dnem před odletem, tak lze z krabicových grafů výše vypozorovat kromě konstantního nárůstu ceny také velkou fluktuaci 2 dny před odletem a následné uklidnění cen v den odletu stejně jako bylo na přímých letech. Z krabicových grafů je dále vidět trend, že letenky s nižší cenou se začínají vyskytovat v čím dál nižším počtu, což zapříčiní zvýšení mediánu.

5.7. Zpracování dat letů s dvěma přestupy

Protože při sběru dat je v nabídce velké množství letů s 1 přestupem, a to i po filtraci dat (10 201 dat), budou relevantní lety s dvěma přestupy brány stejným kritériem, a to celková doba letu nepřekračující 48,58 h a cenou nepřekračující 33 685 Kč mezi 31 dnů a 15 dnů před odletem a mezi 14 dnů a jedním dnem nepřekračující horní kvartil nejdražší přímé letenky v daný den před odletem.

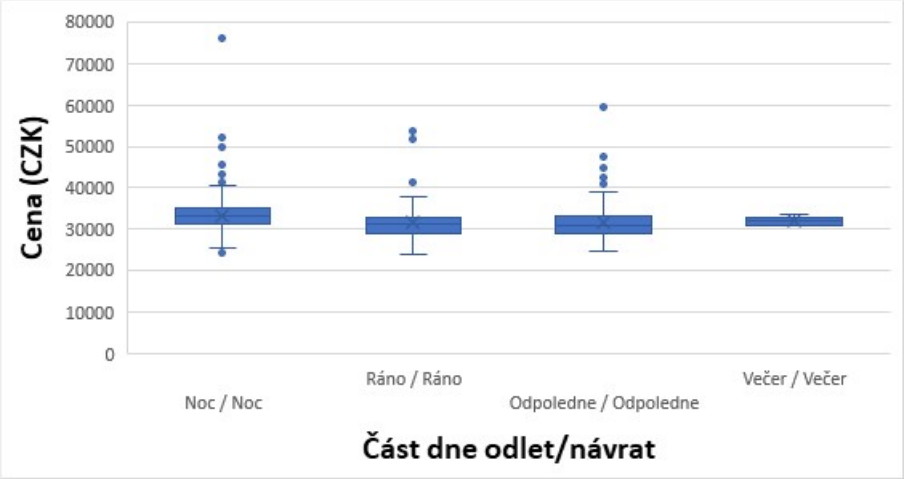
5.8. Výsledky zpracovaných dat

Z výsledků zpracovaných dat poslouží v kapitole 8 k vytvoření pokročilého modelu, který je součástí diplomové práce. Avšak ze zpracovaných dat bylo možné zjistit další zajímavé poznatky, které budou prezentovány níže. Analýza byla inspirována zdrojem [48] zabývajícím se predikcí cen letenek pomocí algoritmu strojového učení. Výsledky jsou popsány níže:



Obrázek 19.: Krabicový graf cen letenek v jednotlivých dnech

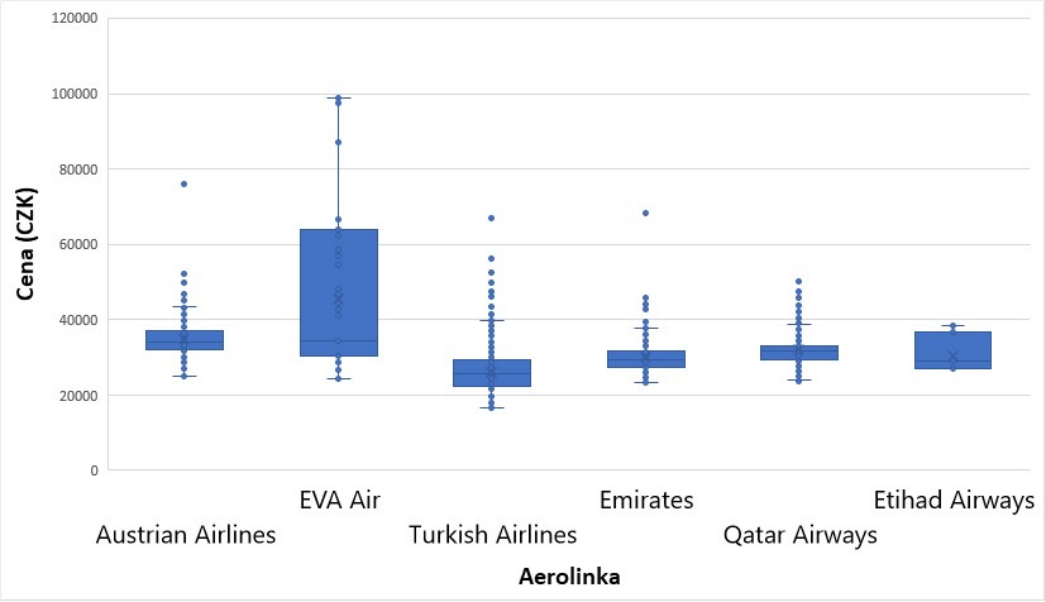
Díky metodě sběru mohl být vytvořen krabicový graf srovnávající dny v týdnu (zpáteční letenka měla odlet z VIE a návrat z BKK ve stejný den v týdnu). Z výsledků vyplynulo, že nejdražší dny pro cestování jsou obvykle soboty (medián 31 291 Kč) a středy (31 202 Kč) a naopak nejlevnějšími dny jsou pondělky (29 936 Kč) a pátky (30 834 Kč).



Obrázek 20.: Krabicový graf cen letenek v jednotlivých dnech

Z krabicových grafů vyplývá, že zpáteční lety v nočních hodinách jsou nejdražší (medián 33 105 Kč), naopak lety zpáteční lety odpoledne byly v době těžení nejvýhodnější (medián 30 989 Kč).

Jak již bylo zmíněno v teoretické části. Na letech mezi Evropou a Asií působí silná blízkovýchodní konkurence nabízející lety s přestupem. Proto je vhodné udělat srovnání mediánů cen letenek, pro zjištění, jaká aerolinka je v porovnání s vybranými aerolinkami nejvýhodnější.



Obrázek 21.: Krabicové grafy ukazující rozpětí cen vybraných aerolinek

Z krabicových grafů bylo dříve zjištěno, že EVA Air (34 398 Kč) a Austrian Airlines (34 131 Kč) mají podobné mediány. Zajímavostí je, že všechny vybrané blízkovýchodní aerolinky nabízejí ceny letenky levněji než obě aerolinky nabízející přímý let. Nejlevnější volba pro let s přestupem byla aerolinka Turkish Airlines (25 763 Kč), jejich letenky byly tedy výrazně levnější než přímé zpáteční letenky.

6. Analýza městského páru Paříž (CDG) – Singapur (SIN)

V průběhu analýzy působilo na městském páru CDG – SIN 2 dopravci, které nabízeli přímý let. Jmenovitě se jednalo o společnosti Air France (IATA kód AF) a Singapore Airlines (IATA kód SQ), které budou popsány v další podkapitole. Kromě těchto přímých letů byla možnost využít lety s jedním či dvěma přestupy od evropských či asijských leteckých dopravců. Jak již bylo zmíněno výše, díky pozici letišť jako letecké uzly byl tento pár zvolen pro analýzu business třídy.

6.1. Air France

Air France je francouzská letecká společnost Air France patřící pod skupinu Air France-KLM. Hlavními leteckými uzly má letišti CDG a AMS. Je členem aliance SkyTeam a létá do více než 200 destinací [49]. Přímé lety na této lince létaly pod označením:

- Let AF256 je provozován letounem Boeing 777-300ER s 48 sedačkami ve třídě business se stářím 14-18 let. Let byl provozován s frekvencí 1x týdně s odletem v 20h 45min a příletem v 16h 35min následujícího dne [43] [50].
- Let AF257 léta mezi SIN a CDG s frekvencí 1 let denně s odletem ve 23h 05min a příletem v 6h 10 min následující den. A je obsluhován stejným letounem jako let AF256 [43].
- Let AF254 odlet každý týden v úterý ve středu v 13:10 a příletem následují den v 9:00 v CDG, tento let byl obsluhován letouny Boeing 777-200ER a Boeing 777-300ER a 21-22 let respektive 18 let u verze 300ER s 40 respektive 48 sedačkami ve třídě business [43].
- AF255 odlet týdně vždy ve středu a ve čtvrtek s odletem v 10:45 a příletem v 17:50. a je obsloužený stejným letadlem jako let AF254 [43].

6.2. Singapore Airlines

Singapore Airlines je singapurský národní dopravce se základnou na nejvytíženějším leteckém uzlu na světě Singapur-Changi, které je podle Skytrax [51] zároveň zatím nejlepším letištěm roku (5/2023). Aerolinka je součástí Star Alliance a dle Skytrax [52]nabízelo v roce 2022 druhou nejlepší business třídu hned po Qatar Airways. Přímé lety na této lince létaly pod označením:

- Let SQ331 byl provozován na páru CDG – SIN v listopadu týdně vždy v úterý, ve čtvrtek a v sobotu a od prosince vypravilo každý den pro tento let 1 spoj a to ve 22:20 s příletem v 17:55 dalšího dne. Na tomto letu byl použity letouny Boeing 777-300ER s 48 sedačkami v business třídě a se stáří letadla bylo mezi 7 a 16 lety [43] [53].
- Let SQ332 byl pro provozován na páru SIN – CDG v listopadu týdně vždy v úterý, ve čtvrtek a v sobotu a od prosince vypravilo každý den pro tento let 1 spoj a to ve 13:55 s příletem v 20:55. Byla použita sejná letadla jako u letu SQ331 [43].

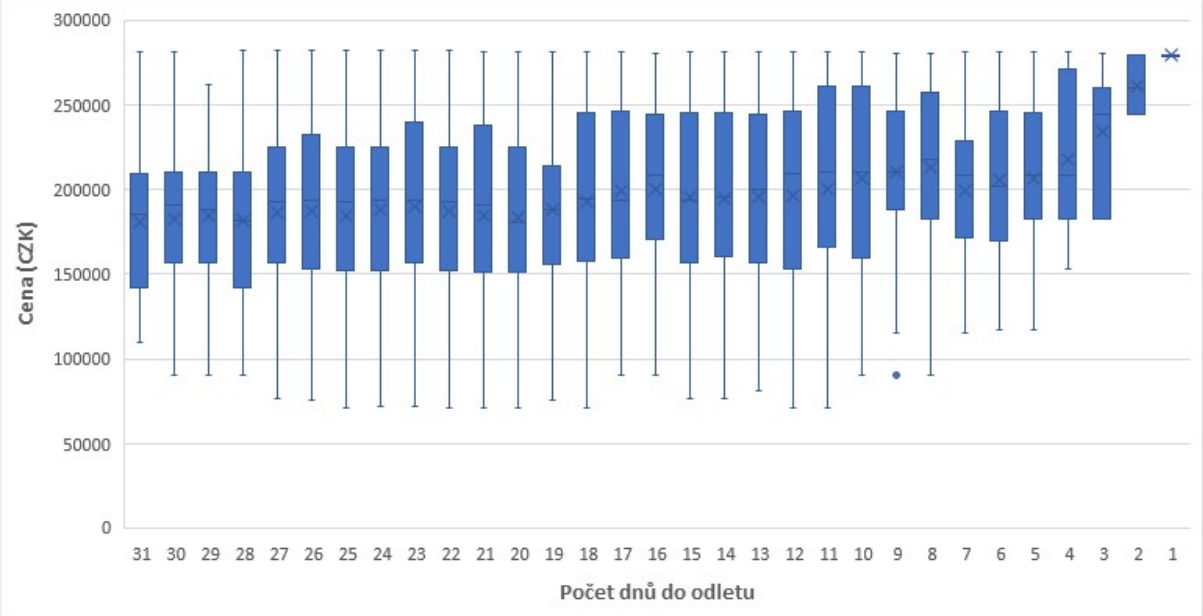
- Let SQ335 byl provozován na páru CDG – SIN s frekvencí 1 let denně a to v 10h 40min z CDG a příletem v 6h 15min následující den. Na tomto páru byl využit letoun Boeing 777-300ER s 48 sedačkami a stářím mezi 8 až 16 lety [43].
- Let SQ336 byl pro provozován na páru SIN – CDG s frekvencí 1 let denně a to v 00h 15min a příletem v 7h 15min. Letouny byly použity stejné jako na letu SQ336 [43].

6.3. Zpracování dat přímého letu

Doba přímého letu je z CDG do SIN 12h 35 min se Singapore Airlines respektive 12h 50 min s Air France a při opačné cestě z SIN do CDG 14h 00 min s Singapore Airlines respektive 14h 05 min s Air France. Celkem bylo za sbírané období natěženo 1434 možných zpátečních letů.

Dále byly nabízeny lety s jedním či dvěma přestupy. Vzhledem k tomu, že let s jedním přestupem vyžaduje čekání na jednom z letiště je třeba brát v úvahu to, že se cestující zdrží.

V úvodní analýze bude autor graficky znázorňovat průběh vývoje přímých zpátečních letenek pomocí krabicových grafů.

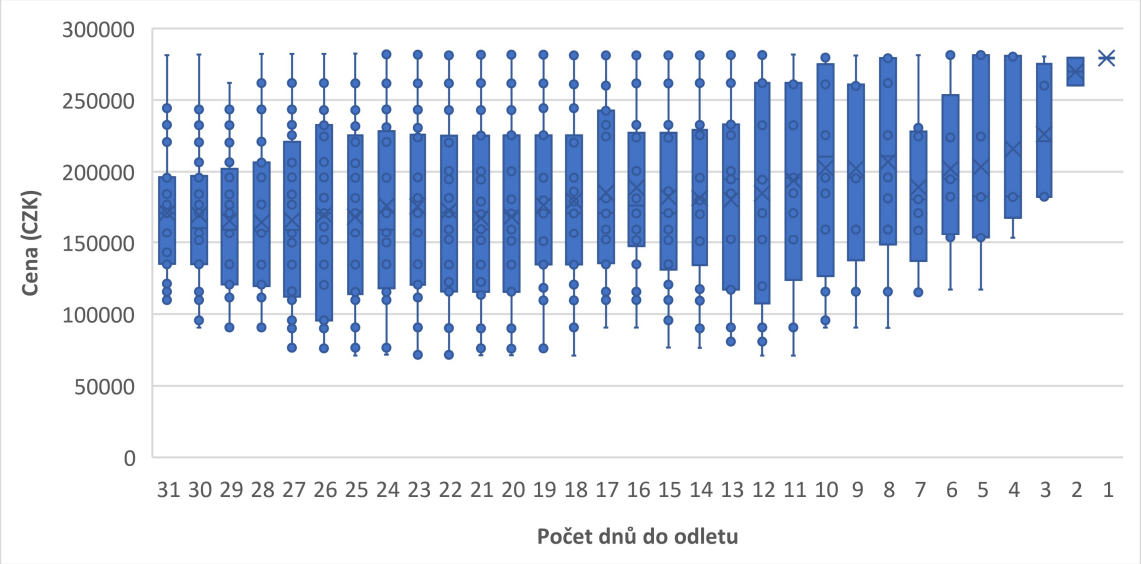


Obrázek 22.: Rozpětí cen přímých zpátečních letů v závislosti na době před odletem

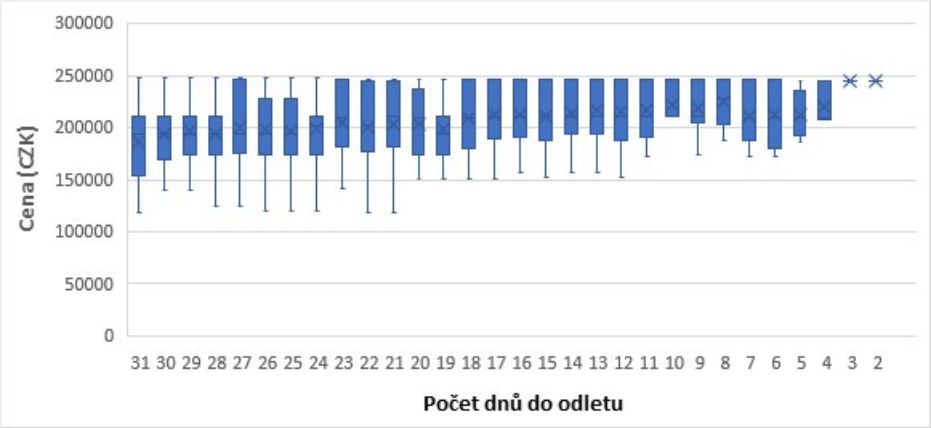
CDG – SGD

Při pohledu na krabicový graf (Obrázek 22) business letenek poskytované aerolinkami Air France a Singapore Airlines si můžeme všimnout zastropování cen letenek jak na horní (ceny cca 282 000 Kč), tak na spodní straně grafu (71 000 Kč). Mezi 31 dny před odletem a 17 dny před odletem se cena letenky pohybuje do 194 573 Kč. Pokud nebudeme počítat 15 a 14 dnů před odletem, tak medián cen letenek již překračuje 200 000 Kč a s ubývajícím počtem levnějších medián cen letenek šplhá až k 279

317 Kč. Pro další analýzu představu byl Obrázek 22 rozdělen zvlášť na další 2 krabicové grafy obsahujíc data z Air France, respektive Singapore Airlines.



Obrázek 23.: Rozpětí cen zpátečních letenek Air France z CDG do SIN



Obrázek 24.: Rozpětí cen zpátečních letenek Singapore Airlines z CDG do SIN

Při pohledu na graf si lze povšimnout zajímavé strategie obou aerolinek. Neplatí zde, že medián cen letenek obou aerolinek je podobný jako při analýze ekonomické třídy mezi VIE a BKK. Je zde výrazný rozdíl mediánu všech nabízených přímých letenek, a to 171 206 Kč (Air France) a 208 475 Kč (Singapore Airlines), což se také projevilo na dvou Obrázcích 23 a 24 výše, kde lze postřehnout větší rozptyl letenek u Air France než u Singapore Airlines. Air France nabízí nejen levnější, ale také dražší letenky (rozmezí mezi cca 71 000 Kč – 282 000 Kč) než Singapore Airlines (od cca 119 000 Kč – 247 000 Kč).

7. Analýza městského páru Praha (PRG) – Hanoj (HAN)

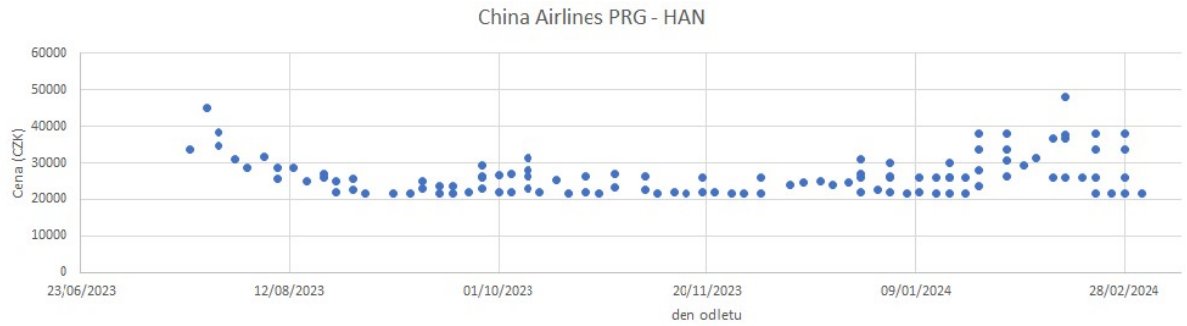
Jako třetí městský pár autor práce zvolil data pro ekonomickou třídu na trase Praha Hanoj, vzhledem k autorově zkušenosti s častými cestami mezi lety z Evropy do Asie (hlavně lety na letiště Hanoj), je tato linka vhodná pro popsání nacenění letenek během různých dob během roku.

Celkem bylo z webu Pelikan.cz natěženo 2 599 kombinací zpátečních letenek, které byly následně analyzovány.

Autor si vybral pro tuto analýzu lety poskytované aerolinií China Airlines. Tato aerolinka dostala v době psaní článku dle ČTK [54] dne 8.3.2023 povolení k provozování přímého spojení mezi letištěm Taipei Taoyuan (TPE) a Letiště Václava Havla Praha (PRG). Letenky se nabízely nejen na tomto městském páru, ale také na městském páru mezi PRG a HAN (s jedním přestupem) za akční ceny, neboť je let nový a neobsazený, což je vhodné pro analýzu sezónnosti (neboť je jisté, že nehrozí, že by byl let vyprodán, což by mohlo navýšit cenu a ovlivnit získaná data při analýze). Letový řád byl stanovený od 18.7.2023, kde autor mezi obdobím 18.7.2023 – 9.3.2024 sbíral data. Protože přímé lety mezi Tchaj-pejí a Prahou budou provozovány dvakrát týdně (z Tchaj-wanu v úterý a sobotu s příletem do Prahy vždy druhý den a tedy odletem z Prahy vždy ve středu a neděli.), autor sbíral každý týden 2 zpáteční lety a to:

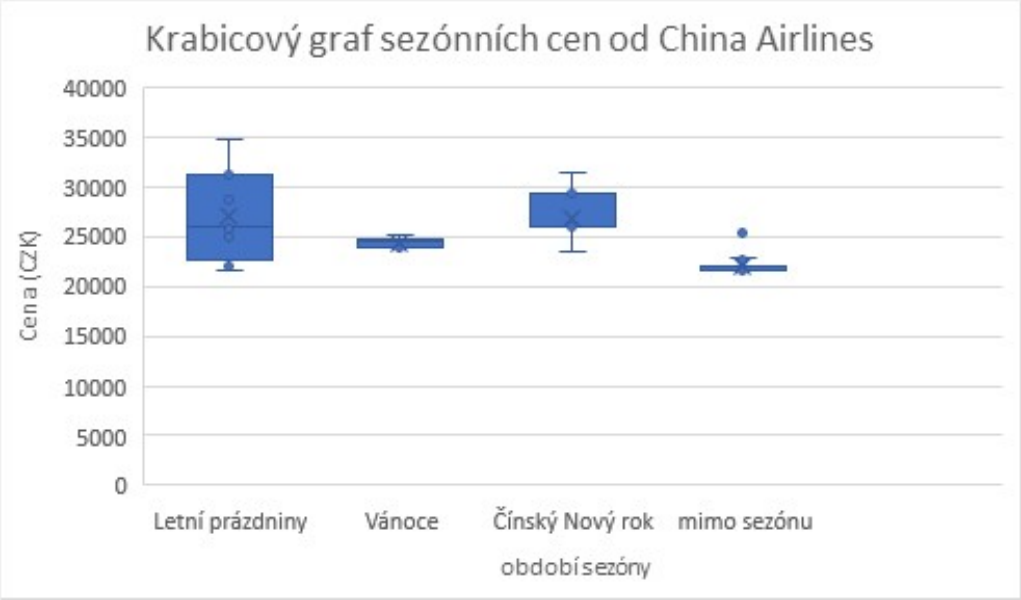
s odletem z Prahy ve středu a návratem v úterý z Hanoje

s odletem z Prahy v neděli a návratem v úterý z Hanoje



Obrázek 25.: Vývoj ceny letenek China Airlines během období 18. 7. 2023 – 9. 3. 2024 na lince PRG - HAN

Při rozboru cen letenek dle sezóny bylo zjištěno několik souvislostí, v době sběru bylo pozorovány 3 období, kde cena zpátečních letenek se zvyšovala nad obvyklou hodnotu, prvním obdobím je období letních prázdnin, které trvalo do konce srpna (letní prázdniny, návštěva příbuzných), dalším obdobím byl měsíc prosinec (období před Vánocemi) a posledním obdobím bylo od 24. 1. 2024 – 18. 2. 2024, kdy se očekává Čínský Nový rok.



Obrázek 26.: Krabicový graf sezónních cen od China Airlines.

Z analýzy bodového grafu (Obrázek 26) byly vypozorovány čtyři období, a to:

- Letní prázdniny od 18. 7. 2023 do 30. 8. 2023
- Období před Vánocemi od 10.12.2023 do 24. 12. 2023
- Období před a po Čínském Novém roku od 24. 1. 2024 do 18. 2. 2024 (Čínský Nový rok připadá na 10. 2. 2024)

Při vytváření krabicového grafu, byly vybrány nejlevnější lety daného dne s China Airlines, aby nedošlo ke zkreslení dat vlivem odlehlých dat. Nejdražší (předpokládá se, že i nejžádanější) období jsou Čínský Nový rok a období letních prázdnin, kde medián zpátečních cen letenek byl shodně 25 962 Kč (o 20 % dražší než mimo sezónu), medián období před Vánočními svátky byl 24 522 Kč (o 14 % dražší než mimo sezónu). Rozptyl zpátečních letenek během letních prázdnin je způsobený postupně klesající cenou letenek s blížícím se koncem prázdnin. Ceny na konci prázdnin se už postupně blížily k cenám mimo sezónu. Medián cen zpátečních letenek mimo sezónu byla na hodnotě 21 522 Kč, kde až na výjimky se zpáteční letenky prodávaly ve skoro všech případech za cenu 21 522 Kč.

8. Vytvoření modelů

8.1. Model cen letenek na městském páru VIE – BKK

Jedním z cílů této práce je vytvoření pokročilého modelu pro zkoumání cen letenek na dlouhých tratích mezi různými typy letišť se zaměřením na různé segmenty cestujících a různé typy destinací.

Model je zpracován na základě datového souboru z období 6.10.2022 do 6.11.2022. Při analýze dat bude tato práce pracovat se statistickým softwarem Scilab. Při tvorbě modelu jsme rozhodli, že využijeme medián (50 % kvartilu) a dolního kvartil (25% kvantil). Z hlediska statistiky má medián výhodu oproti průměru, že se nemusí vyrovnat s problémy, pokud v datovém souboru jsou nějaké extrémní hodnoty. 25% kvantil zas napovídá, že nejméně 25 procent prvků má hodnotu menší nebo rovnu 100 procenty našich dat, vytváříme tedy model, který bude odpovídat vývoj cen letenek na dolním kvartil, což odpovídá vývoji cen letenek pro cestující citlivější na cenu [25].

Na městském páru je několik faktorů, které autor předpokládá, že přímo ovlivňují přepravu. Vzhledem k rozsahu těžených dat (1 měsíc) lze předpokládat, že se za toto období nepromítne cena paliva, které si dopravci hedgeují dopředu, či HDP, které by bylo viditelné, pokud bychom měli data i z předchozích let. Autor se tedy zaměřuje na faktory, které přímo v tu dobu ovlivňuje cenu letenky a dospěl k závěru, že hlavními vlivy by mohl být konkurenční přímý let působící na dané trase či působení jiných dopravců, kteří nabízí lety s jedním přestupem.

Model (vzorec) který bude popsán bude aerolinka Austrian Airlines nabízející přímý spoj mezi tímto párem na kterém bude provedena analýza v závislosti na aerolince EVA AIR (nabízí také přímý spoj = přímý konkurent) a na letech s jedním a dvěma přestupy. Po zjištění testu závislosti bude vytvořený model predikce.

Dalším krokem je tedy potřeba provést test závislosti mezi koeficienty za účelem zjištění statisticky významných či nevýznamných (nízká nebo žádná závislost) koeficientů. To se provedlo pomocí funkce CORELL v programu Excel. Po provedení testu bylo zjištěno, že u cenově dostupnějších letenek (dolní kvartil) je korelace mezi Austrian Airlines a Eva Air silná a u mediánu je korelace střední. Naopak mezi Austrian Airlines a lety s přestupem je korelace velmi slabá a téměř nulová. To znamená, že lety s přestupy na dolní hranici kvartilu a v mediánu při změně ceny neovlivní to, za jakou cenu se bude prodávat letenka při letu s Austrian Airlines.

Tabulka 3: Výsledky kolinearity mezi koeficienty VIE - BKK

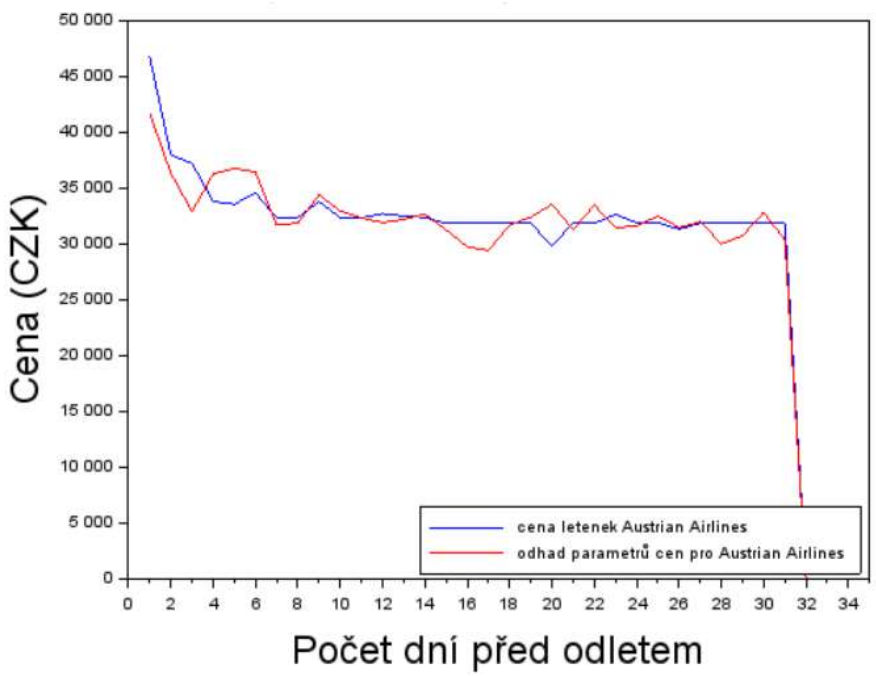
Korelace mezi koeficienty	**Pearsonův korelační koeficient**
25% Austrian Airlines a 25% EVA Air	0.619572
medián Austrian Airlines a medián EVA Air	0.461799
25% Austrian Airlines a 25% 1+2 přestupy	-0.00148
medián Austrian Airlines a medián 1+2 přestupy	-0.02083

Proto při sestavování modelu bude brán v potaz pouze statisticky významného koeficientu. Pro vytvoření modelu dolní hranice kvartilu a modelu mediánu pro ceny letenek bude využit jako závislá proměnná data cen letenek od Austrian Airlines a pro nezávislá data cen letenek od Eva Air.

8.2. Model pro dolní kvartil

Pro vytvoření modelu bude potřeba provést Bayesovský odhad parametru modelu, postup pro vytvoření modelu je uveden v Příloze 2**.**

Bayesovský model začíná odhadovat až od určitého posunutého času, proto bylo potřeba vložit počáteční data tak, aby odhadování probíhalo až po posunutí času. Autor práce tento problém vyřešil tak, že vložil hodnotu cen letenek ve výši 1 Kč 32 dní před odletem. 31 dní před odletem Bayesovský model již začíná vyhodnocovat data až do 1 dne před odletem. Dalším problémem, který musel autor řešit byly chybějící data (v Excelu nemají tyto buňky hodnotu, tzv. NaN). Tento problém vyřešil autor cyklem *for* na řádku 6-9 v Příloze 2, který nahrazuje chybějící hodnotu tou, co byla předtím (např.: pokud 20 dnů před odletem nenabízela aerolinka EVA Air let, byla v této buňce doplněna cena z buňky 19 dnů před odletem).



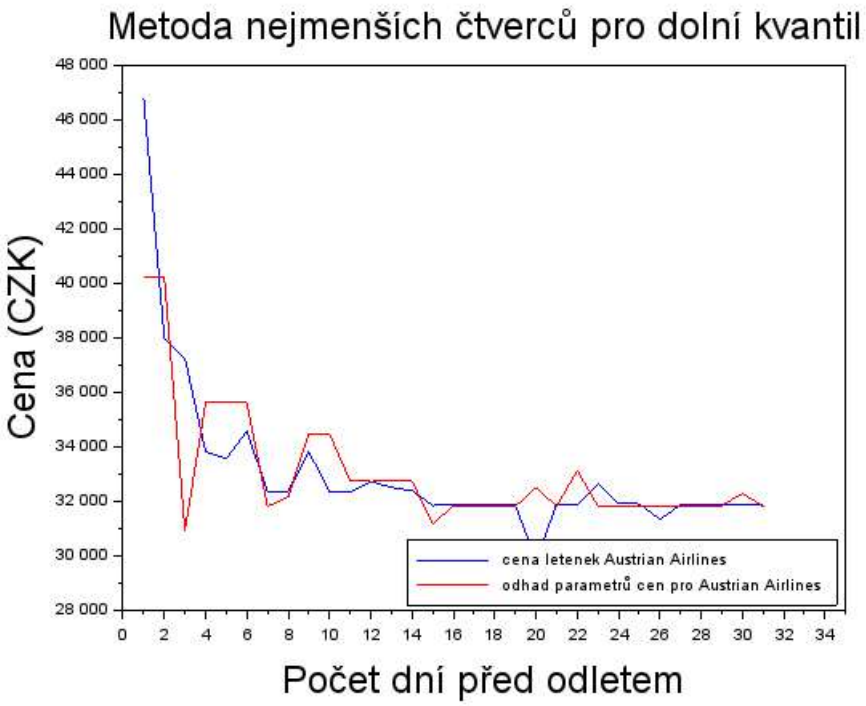
Obrázek 27.: Bayesovský model pro dolní kvartil

Protože je model stochastický (pod vlivem neurčitosti), je po každém přehrání programu výsledek podobný, ale ne stejný (prvek *rand(1,1,'n')* v příloze 2). Autor si po několika pokusech zapsal nejpřesnější výsledek.

```
"Odhad regresních koeficientů"

 0.1815322
-0.1694861
 0.1821614
 25783.571

"Odhad rozptylu"

 2188785.5
R2  =
 0.9046441

"R2: 0.9046441"

"RMSE: 1990.8393"
```

Obrázek 28.: Výsledky Bayesovského modelu – dolní kvartil

Dále byl vytvořený jiný jednoduchý model pro dolní kvartil pomocí Metody nejmenších čtverců. Vzniklá křivka modelu je zobrazena na Obrázku 29 níže.



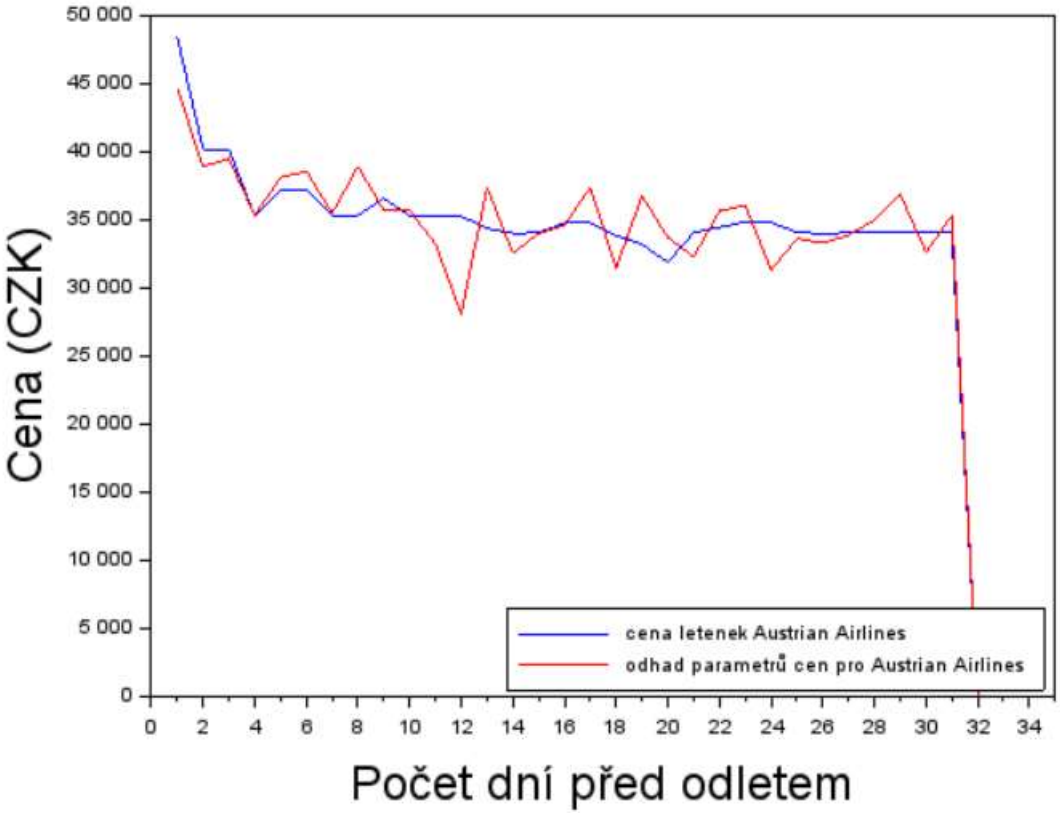
Obrázek 29.: Model založený na Metodě nejmenších čtverců pro dolní kvartil

Tato metoda pracuje pouze s offline daty a není zde prvek neurčitosti (deterministický model), zdroj pro vytvoření tohoto modelu je uveden v Příloze 3. Z modelu pak vychází výsledky, které jsou popsány níže.

```
RMSE  =
  1908.2670

 "Koeficienty:"

  24753.513
  0.2321308

 "RMSE:"

  1908.2670
R2  =
  0.5836422

 "R2: 0.5836422"
```

Obrázek 30.: Výsledky MNČ modelu – dolní kvartil

8.3. Modely pro medián



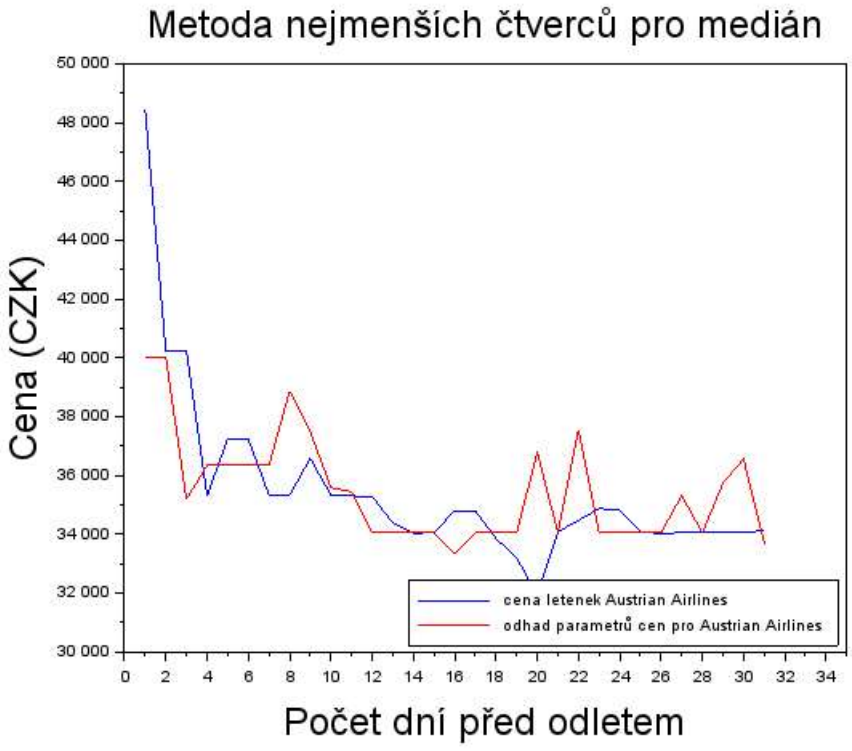
Obrázek 31.: Bayesovský model pro medián

Stejný způsob byl použit k vytvoření Bayesovského modelu pro medián, tzn. byl pozměněn pouze datový soubor. Po několika pokusem (vlivem prvku neurčitosti) byl vybrán nejvhodnější způsob, který je zobrazen na Obrázku 31. Zdrojový kód modelu je zobrazen v Příloze 4. Odhadované výsledky jsou zobrazeny níže včetně statistických testů pro budoucí validaci.

```
"Odhad regresních koeficientů"

 0.1508011
-0.0688113
 0.0917063
 27787.229

"Odhad rozptylu"

 4756929.6

"R2: 0.8925296"

"RMSE: 2231.5716"
```

Obrázek 32.: Výsledky Bayesovského modelu – medián

Stejně jako u jednoduchého modelu pro dolní kvartil pomocí Metody nejmenších čtverců byl vytvořen model pro medián, kde byl zachován postup a změněn pouze datový soubor. Vzniklá křivka modelu je zobrazena na **Obrázku** níže.



Obrázek 33.: Model založený na Metodě nejmenších čtverců pro medián

Zdroj pro vytvoření tohoto modelu je uveden v Příloze 5**.** Z modelu pak vychází výsledky, které jsou popsány níže.

```
"Koeficienty:"

 27693.416
 0.1846180

"RMSE:"

 2295.0422

"R2: 0.3827777"
```

Obrázek 34.: Výsledky z modelu MNČ - medián

8.4. Model cen letenek na městském páru PRG – HAN

Cílem analýzy cen letenek na dálkových tratích je také analýza cen mimo sezónu a během sezónní špičky. Autor se rozhodl, že i na tomto městském páru se pokusí popsat pomocí modelu, který byl použit v předchozí **kapitole.** Mezi tímto městským párem nevede žádná přímá linka, tudíž autor vybral nezávislá data ceny zpátečních letenek od aerolinky China Airlines. Z teorie víme, že lety na letech mezi Evropou a Asií působí velcí dopravci na Blízkém východě, jejich data budou využita jako závislá data. Budeme tedy vytvářet model pro ceny zpátečních letenek od China Airlines. Na tomto městském páru působí několik významných blízkovýchodních dopravců, a to Qatar Airways, Emirates a Turkish Airlines. Vzhledem nízkým počtem natěžených dat od společnosti Turskish Airlines byly pro analýzu použity pouze natěžená data od společnosti Qatar Airways a Emirates.

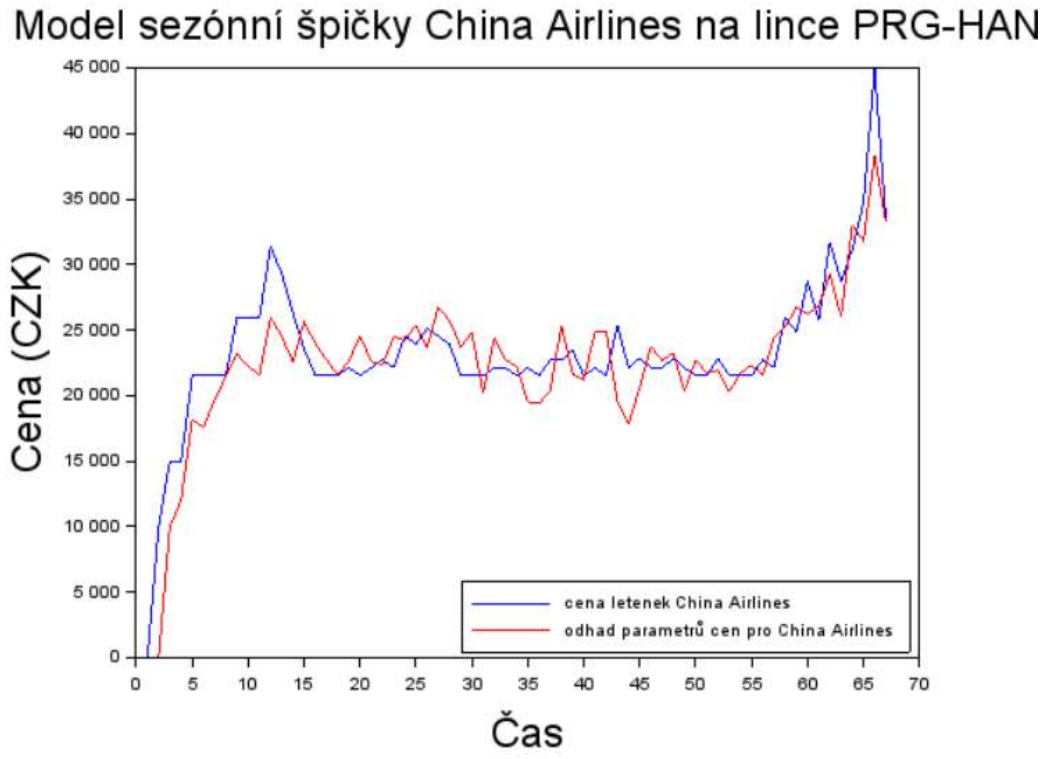
Aby se mohly určit závislé proměnné, je třeba provést test závislosti veličin, na základě kterého budou vybrány pouze vhodné veličiny pro další analýzu. Výsledky jsou zobrazeny v Tabulce 4**.** Z výsledků vyplývá, že jak mezi China Airlines a Qatar Airways tak mezi China Airlines a Emirates je silná korelace. To znamená, že oba datové soubory jsou vhodné k odhadování křivky cen letenek pro China Airlines.

Tabulka 4: Výsledky kolinearity mezi koeficienty PRG - HAN

Korelace mezi koeficienty	**Pearsonův korelační koeficient**
China Airlines a Qatar Airways	0.744524
China Airlines a Emirates	0.70924

8.5. Model pro městský pár PRG – HAN

Stejný postup jako na městském páru VIE – BKK byl použit i zde. Byl vytvořen Bayesovský model, kde datový kód lze najít v Příloze 7**.** Po několika pokusech (stále kvůli vlivu neurčitosti) byly zapsán nejpřesnější model, který je zobrazen na Obrázku. Daný model se liší od modelu na páru VIE-BKK v tom, že potřebuje více času, než začne odhadovat parametry, proto bylo do datového souboru vložena data odpovídající době před odhadováním (*y(t)* až *y(t+2)*) Bayesovského modelu. Dále byl model upraven přehozením dat sběru (sběr od 7/2023 do 3/2024 prohozeny na 3/2024 do 7/2023) neboť období 7/2023 je hlavní sezónní špička a byla by škoda, kdyby tento popis nebyl zobrazen kvůli času, který model potřebuje pro rozjezd.



Obrázek 35.: Bayesovský model pro sezónní špičku

```
"Odhad regresních koeficientů"

 0.1448270
 0.0126986
 0.3471037
 0.0826394
-0.0441947
 0.1486487
 6522.8886

"Odhad rozptylu"

 5066295.9

"R2: 0.7314237"

"RMSE: 2825.1273"
```

Obrázek 36.: Výsledky Bayesovského modelu – sezónní špička

Dále byl vytvořený jednoduchý model pro sezónní špičku pomocí Metody nejmenších čtverců. Vzniklá křivka modelu je zobrazena na Obrázku 37 níže.

Obrázek 37.: Metoda nejmenších čtverců pro sezónní špičku

```
"Koeficienty:"

 14880.287
 0.2611719
 0.0343558

"RMSE:"

 2771.0426

"R2: 0.5531896"
```

Obrázek 38.: Výsledky MNČ – sezónnost

8.6. Validace získaných dat

Pro validaci dat byla vytvořena Tabulka 5 srovnávající výkonnost vytvořených modelů. Tučně byly vyznačeny hodnoty, které jsou v rámci srovnání pokročilého a jednoduchého modelu přesnější. Tabulka napovídá, že ne ve všech případech byl pokročilý Bayesovský model přesnější než jednoduchý model postavený na Metodě nejmenších čtverců. Metoda nejmenších čtverců byla lepší ve dvou případech, kdy se validovalo pomocí RMSE. Pro dolní kvartil byla Metoda nejmenších čtverců o 4,2 % přesnější a v případě sezónní špičky o 1,9 %. Avšak Bayesovský model vykazoval výrazně lepší výsledky při testování pomocí koeficientu determinace R2. Ze statistických testů vyplynulo, že ve všech datových vzorcích je Bayesovský model vhodnější pro modelování dat o cenách letenek.

Tabulka 5: Srovnání výkonnosti různých regresních modelů

Použitelnost modelu	**Model**	**RMSE**	**R^2**
	Spojení VIE - BKK		
Vhodnější	Bayesovský model pro dolní kvantil	1,991	**0.904644**
	Metoda nejmenších čtverců pro dolní kvantil	**1908.267**	0.5836
Vhodnější	Bayesovský model pro medián	2231.5716	**0.89253**
	Metoda nejmenších čtverců pro medián	**2295.0422**	0.382778
	Spojení PRG - HAN		
Vhodnější	Bayesovský model pro sezónní špičku	2825.1273	**0.731424**
	Metoda nejmenších čtverců pro sezónní špičku	**2771.0426**	0.55319

Při srovnání získaných výsledků s podobnou prací zabývajícím se vytvářením modelů pro predikování letenek bylo zjištěno, že hodnota R^2 v našem pokročilém modelu podobných hodnot (v jednom případě

i lepší) jako komplikovanější metody založené na strojovém učení. Hodnoty se navíc blížily hodnotě 1, což znamená, že náš vytvořený model je přínosný. Také se nám podařilo validovat jednoduchý model postavený na Metodě nejmenších čtverců, kde hodnota R^2 byla v rozmezí 0,38-0,58. V jiné práci [32] dosahoval jednoduchý model na bázi lineární regrese hodnot 0,612 a 0,618.

Dále byla pozorována další souvislost, kde v jiné práci [32] měla komplikovanější metoda (Metoda podpůrných vektorů) slabší výsledek RMSE než jednoduchý model a to 1,7 %.

Pro připomenutí hodnota RMSE nelze srovnávat mezi různými datasety, slouží pouze ke srovnání modelů na stejném datasetu.

Diskuze

Na začátku této diplomové práce byla provedena rešerše, na základě které byly popsány poznatky, jež byly v dalších částech práce ověřovány. Prvním poznatkem byla pokračující regulace letecké dopravy pomocí bilaterálních smluv, kdy na každém městském páru působili maximálně dva dopravci, kteří zajišťovali přímé lety. Tito letečtí dopravci byli většinou národními dopravci, ale s postupnou deregulací se začínají objevovat i sedmé svobody vzduchu, což bylo potvrzeno na městském páru Vídeň – Bangkok na letu BR61 respektive BR62. Při analýze flotily, která se používá na dálkových tratích, bylo potvrzeno využití hlavně širokotrupých letadel s dlouhým doletem (verze LR či novější modely úsporných letounů). Dále byly zkoumány praktiky leteckých společností, kde na přímém spojení Vídeň – Bangkok byla zjištěna strategie price-matchingu. Z revenue managementu byla dále potvrzena strategie cenotvorby, která u business třídy převyšovala násobně letenky v ekonomické třídě o podobné letové vzdálenosti. Další poznatek, který byl vypozorován, bylo nacenění letenek s přestupem, kdy ve všech vybraných případech byl medián citelně levnější než u přímých letů na dlouhých vzdálenostech, což se následně projevilo při korelačním testu, kde letenky s jedním přestupem neovlivňují, jak se budou naceňovat letenky na přímých letech (nízká závislost). Při analyzování business třídy byla potvrzena strategie účtování prémiových cen, kdy aerolinka, díky vyššímu standardu a tím odlišení svého produktu od konkurenta, si mohla dovolit účtovat průměrné ceny vyšší než přímý konkurent (Air France). V dalším městském páru byl zkoumán další revenue management založený na historických datech, a to sezónní špička. Ze zdrojů a také vlastní autorovy zkušenosti je známo, že aerolinky se nebojí účtovat výrazně vyšší ceny než v období mimo sezónu. Při těžení dat o cenách letenek pro linku Praha – Hanoj, se projevily tyto sezónní špičky ve všech pozorovaných obdobích a u všech vybraných aerolinií. Co se nepodařilo ověřit, byly týdenní a denní špičky, kdy u týdenní špičky bylo zmíněno, že jsou obvykle letenky dražší o víkendech, což se potvrdilo jen u soboty, neděle byla v určitých případech levnější než pracovní dny. U denních špiček bylo vypozorováno, že lety začínající v noci jsou žádanější, avšak nepodařilo se ověřit tvrzení preference cestujících k ověření tohoto argumentu.

Na základě poznatků z natěžených dat o cenách letenek na dlouhých tratích vytvořil autor modely, které analyzují cenotvorbu v krátkém časovém horizontu (31 dní až jeden den před odletem) a sezónnost. Vytvořený pokročilý model postavený na Bayesově větě byl následně validován s jednoduchým modelem založeným na Metodě nejmenších čtverců. Ve všech případech datasetu byl Bayesovský model vhodnější než model založený na Metodě nejmenších čtverců, přestože ne ve všech případech byl pokročilý model lepší. V hodnotách RMSE ztrácel ve dvou případech pouze jednotky procent, ale při validaci pomocí R^2 byl Bayesovský model lepší o desítky procent. Hodnoty R^2 u Bayesovského modelu dosahoval vysokých hodnot i v porovnání s jinými odbornými pracemi, založenými na

komplikovanějších metodách jako je strojové učení. Hodnota RMSE byla použita jen pro porovnání 2 vytvořených modelů, kde oba modely dosahovaly podobných hodnot.

Závěr

V rámci diplomové práce byla vypracována analýza cen letenek na dlouhých tratích. Tato analýza byla provedena na třech městských párech a to VIE – BKK, CDG – SIN a PRG – HAN, což značí, že byla zaměřená na evropského zákazníka. Pro účel práce byla potřebná data těžena z webu Pelikan.cz. Pro každý městský pár byl natěžen zvolený datový soubor.

V úvodu diplomové práce byly v teoretické části definovány důležité aspekty pro fungování letecké dopravy. Při zkoumání aktuální situace s provozováním letecké dopravy na dálkových linkách bylo zjištěno, že jsou stále závislé na regulaci, která jsou popsána především v leteckých dohodách. Tyto dohody tedy definují fungování leteckých společností na určených trasách. Do dnešní doby jsou dálkové lety tedy stále regulované. Dále byla popsána částečná deregulace v podobě aplikace sedmé svobody, která se vyskytla také v datovém souboru.

Při pohledu strategií leteckých společností byly v práci popsány různé způsoby nacenění letenek. Potvrzeny byly teorie price-matchingu, kdy dvě soupeřící aerolinky nabízely téměř totožné mediány cen letenek, výrazně se dále na těžených vzorcích dat projevily sezónní špičky. Tyto špičky se projevily na všech vybraných linkách. Při zkoumání letenek vyšší třídy, v našem vzorku dat to byla business třída, byla ověřena strategie odlišení produktu, kdy díky vyššímu standardu služeb si účtovala aerolinka Sigapore Airlines průměrně vyšší ceny než konkurenční aerolinka Air France.

V další části práce se autor zabýval zpracováním dat. Během tohoto procesu bylo nutné řešit řadu limitujících vlivů jako jsou neúplné soubory či chybějící údaje, které poté nešly dohledat nebo musely být doplněny ručně. Tento limit se naštěstí nahradil velkým vzorkem dat obsahujícím desítky tisíc kombinací zpátečních letenek. Popisná statistika zajistila, že ze všech natěžených dat zůstaly pouze relevantní zpáteční lety, ze kterých byly vypozorovány první znaky revenue managementu leteckých společností.

Kromě grafického popisu byla data dále využívána pro vytvoření pokročilého modelu založeném na Bayesově větě. Limitací uvedeného modelu byl šum, který je nutno aplikovat, neboť je model stochastický a tento šum se projevil na hodnotě RMSE. Dále tento model potřebuje určitý čas pro rozjezd, než začne odhadovat parametry modelu. Po provedení validace, bylo zjištěno, že výsledky vycházející z pokročilého modelu založeném na Bayesově větě naznačují, že je model vhodný pro modelování vývoje cen letenek na dlouhých tratích, a to jak v krátkém (doba do odletu), tak dlouhém časovém horizontu (sezónnost).

Práce poskytuje užitečné informace internetovým portálům zaměřeným na cestování, které by mohly cestujícím dávat tipy (např. tajné tipy či rady, kdy letět nebo neletět), které nevycházejí jen z analýzy dat, ale také z vypracovaného modelu o cenách letenek. Dále by práce mohla být určená pro cestovní

kanceláře či firmy, pro které mohou být informace získané touto prací přínosné. Další přínos lze spatřovat také pro aerolinie, které se chystají vstoupit, či chtějí mít přehled o daném trhu.

Zdroje

[1] HOLLOWAY, Stephen. *Straight and Level: Practical Airline Economics*. 3rd ed. Hampshire: Ashgate Publishing Limited, 2008. ISBN ISBN 978-0-7546-7256-2.

[2] SHAW, Stephen. *Airline Marketing and Management*. 3rd ed. Hampshire: Ashgate Publishing Limited, 2007. ISBN 978-0-7546-4820-8.

[3] *Air Transport & Travel Industry IATA EDIFACT AND XML CODESET* [online]. 2020 [cit. 2023-05-11]. Dostupné z: https://cdn.document360.io/4b1f93b7-f7b8-4ebb-81be-c17af837a94f/Images/Documentation/Code%20set%20Directory%2020.2.pdf

[4] GRIMME, Wolfgang, Stephan BINGEMER a Sven MAERTENS. An analysis of the prospects of ultra-long-haul airline operations using passenger demand data. *Transportation Research Procedia* [online]. 2020, **51**, 208-216 [cit. 2023-05-11]. ISSN 23521465. Dostupné z: doi:10.1016/j.trpro.2020.11.023

[5] *Introduction to Air Transport Economics*. Third edition. Oxon: Taylor & Francis, 2018. ISBN 9781315299068. Dostupné také z: https://www.google.cz/books/edition/Introduction_to_Air_Transport_Economics/cmYPEAAAQBAJ?hl=cs&gbpv=1

[6] PANDE, Pranjal a HARDIMAN JAKE. What Is An Airline's Load Factor And Why Does It Matter?. In: *Simpleflying.com* [online]. [cit. 2023-05-11]. Dostupné z: https://simpleflying.com/load-factor/

[7] IATA. *Industry Statistics: Fact Sheet*. 2021, 2 s. Dostupné také z: https://www.iata.org/en/iata-repository/publications/economic-reports/airline-industry-economic-performance---october-2021---data-tables/

[8] IATA. *Industry Statistics: Fact Sheet*. 2022, 2 s. Dostupné také z: https://www.iata.org/en/iata-repository/pressroom/fact-sheets/industry-statistics/

[9] KEMMIS, Sam. Two One-Way Flights or One Round-Trip: Which Is Better?. In: *Finance.yahoo.com* [online]. [cit. 2023-05-11]. Dostupné z: https://finance.yahoo.com/news/two-one-way-flights-one-192346838.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACb5F78Mc3FznXCh1VglMdAseiDB3wkGkXgRoTmZ_UsrK9x56FvDZMzLqr-T3NX7QTPA_BLJV5dEeC9Uo_v_hjL179IdwHL8ghsw3kHC

[10] EISELT, David. Cestovní třídy v letadle. Jaké jsou a čím se liší?. In: *Cestujlevne.com* [online]. [cit. 2023-05-12]. Dostupné z: https://www.cestujlevne.com/blog/cestovni-tridy-v-letadle

[11] BINGGELI, Urs a Mathieu WEBER. A short life in long haul for low-cost carriers. In: *Mckinsey.com* [online]. [cit. 2023-05-11]. Dostupné z: https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/a-short-life-in-long-haul-for-low-cost-carriers

[12] MORRELL, Peter. Can long-haul low-cost airlines be successful?. *Research in Transportation Economics* [online]. 2008, **24**(1), 61-67 [cit. 2023-05-11]. ISSN 07398859. Dostupné z: doi:10.1016/j.retrec.2009.01.003

[13] IATA. *Air Connectivity: Measuring the connections that drive economic growth*. 129 s. Dostupné také z: https://www.iata.org/en/iata-repository/publications/economic-reports/air-connectivity-measuring-the-connections-that-drive-economic-growth/

[14] GROSCHE, Tobias, Richard KLOPHAUS a Adam SEREDYńSKI. Competition for long-haul connecting traffic among airports in Europe and the Middle East. *Journal of Air Transport Management* [online]. 2017, **64**, 3-14 [cit. 2023-05-11]. ISSN 09696997. Dostupné z: doi:10.1016/j.jairtraman.2017.06.019

[15] PILTZ, Christopher, Augusto VOLTES-DORTA a Pere SUAU-SANCHEZ. A comparative analysis of hub connections of European and Asian airports against Middle Eastern hubs in intercontinental markets. *Journal of Air Transport Management* [online]. 2018, **66**, 1-12 [cit. 2023-05-11]. ISSN 09696997. Dostupné z: doi:10.1016/j.jairtraman.2017.09.006

[16] How to Save Money on Flights with Skyscanner-But Should You?. In: *Worldtravelfamily.com* [online]. [cit. 2023-05-12]. Dostupné z: https://worldtravelfamily.com/save-money-flights-skyscanner/

[17] DATA COLLECTION. In: *Infare.com* [online]. [cit. 2023-05-12]. Dostupné z: https://infare.com/data/data-collection/

[18] *WEB SCRAPER: Making web data extraction easy and accessible for everyone* [online]. [cit. 2023-05-12]. Dostupné z: https://webscraper.io/

[19] HANKE, Dominik. *POST-COVIDOVÝ VÝVOJ CEN LETENEK V RÁMCI VYBRANÝCH MĚSTSKÝCH PÁRŮ*. Praha, 2022. Dostupné také z: https://dspace.cvut.cz/handle/10467/101958. Diplomová práce. ČVUT v Praze, Fakulta dopravní, Ústav letecké dopravy.

[20] MARIANO, Kristin. Thailand is most popular destination in SEA for North Americans and Europeans. In: *Traveldailymedia.com* [online]. [cit. 2023-05-12]. Dostupné z: https://www.traveldailymedia.com/thailand-is-most-popular-destination-in-sea-for-north-americans-and-europeans-1/

[21] ATAG. *Aviation Benefits report* [online]. 2020, 96 s. [cit. 2023-05-14]. Dostupné z: aw-oct-final-atag_abbb-2020-publication-digital.pdf

[22] Z letiště až do centra. In: *Wien.info* [online]. [cit. 2023-05-15]. Dostupné z: https://www.wien.info/cs/hotely-cestov%C3%A1n%C3%AD/p%C5%99%C3%ADjezd/z-leti%C5%A1t%C4%9B-a%C5%BE-do-centra

[23] DOŠLÁ, Zuzana. *Základy popisné statistiky* [online]. [cit. 12.5.2023]. Dostupné z: https://is.muni.cz/el/sci/podzim2015/M1010/um/

[24] PAVLÍK, Tomáš a Ladislav DUŠEK. *Biostatistika*. Brno: Akademické nakladatelství CERM, 2012. ISBN 978-80-7204-782-6.

[25] Matematika polopatě: Medián. In: *Matweb.cz* [online]. [cit. 2023-05-14]. Dostupné z: https://www.matweb.cz/median/

[26] Grafické zobrazení dat. In: *Wikisofia.cz* [online]. [cit. 2023-05-14]. Dostupné z: https://wikisofia.cz/wiki/Grafick%C3%A9_zobrazen%C3%AD_dat

[27] Interpretace naměřených dat. In: *Lean6sigma.cz* [online]. [cit. 2023-05-14]. Dostupné z: https://lean6sigma.cz/interpretace-namerenych-dat/

[28] PAVLÍK, Tomáš a Ladislav DUŠEK. Biostatistika. *Matematickabiologie.cz*. Vyd. 1. Brno: Akademické nakladatelství CERM, 2012. Dostupné také z: https://www.matematickabiologie.cz/media/3293331/pavlik-biostatistika.pdf

[29] HENDL, Jan. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Vyd. 2., opr. Praha: Portál, 2006. ISBN 80-7367-123-9.

[30] *Pearsonův korelační koeficient: Maths and Stats Support Centre* [online]. In: . [cit. 2023-05-15]. Dostupné z: https://mathstat.econ.muni.cz/media/12657/pear_cor.pdf

[31] MAYO, Matthew. Linear Regression, Least Squares & Matrix Multiplication: A Concise Technical Overview. In: *Kdnuggets.com* [online]. [cit. 2023-05-13]. Dostupné z: https://www.kdnuggets.com/2016/11/linear-regression-least-squares-matrix-multiplication-concise-technical-overview.html

[32] WANG, Tianyi, Samira POUYANFAR, Haiman TIAN, Yudong TAO, Miguel ALONSO, Steven LUIS a Shu-Ching CHEN. A Framework for Airfare Price Prediction: A Machine Learning Approach. In: *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)* [online]. IEEE, 2019, s. 200-207 [cit. 2023-05-13]. ISBN 978-1-7281-1337-1. Dostupné z: doi:10.1109/IRI.2019.00041

[33] NAGY, Ivan. Základní pojmy z bayesovské statistiky. In: *Fd.cvut.cz* [online]. [cit. 2023]. Dostupné z: https://www.fd.cvut.cz/personal/nagyivan/WebLab/Notions.pdf

[34] UGLICKICH, Evženie. Přednáška 4 – Bayesovský odhad parametrů modelu. In: *Staff.utia.cas.cz* [online]. [cit. 2023]. Dostupné z: http://staff.utia.cas.cz/uglickich/pdfka/slajdy_Pr4.pdf

[35] UGLICKICH, Evženie. *Přednáška 4 – Bayesovský odhad parametrů modelu*. 7 s. Dostupné také z: http://staff.utia.cas.cz/uglickich/pdfka/slajdy_Pr4.pdf

[36] Koeficient determinace. In: *Matematickabiologie.cz* [online]. [cit. 2023-05-14]. Dostupné z: https://portal.matematickabiologie.cz/index.php?pg=analyza-a-hodnoceni-biologickych-dat--regresni-modelovani--linearni-regresni-model--jak-definujeme-linearni-regresni-model--koeficient-determinace

[37] Koeficient determinace (R na druhou). In: *Definebusinessterms.com* [online]. [cit. 2023-05-14]. Dostupné z: https://www.definebusinessterms.com/cs/koeficient-determinace-r-na-druhou/

[38] PETRÚŠEK, Ivan. *Analýza chybějících hodnot: srovnání metod při zkoumání determinantů politické znalosti a příjmu*. Praha: Sociologický ústav AV ČR, 2015. ISBN 978-80-7330-267-2.

[39] HAVLÍK, Jindřich. *Hodnocení kvality odhadu stavu stochastických systémů*. Plzeň, 2014. Dostupné také z: https://otik.uk.zcu.cz/bitstream/11025/12330/1/DP.pdf. Diplomová práce. Západočeská

univerzita v Plzni Fakulta aplikovaných věd Katedra kybernetiky. Vedoucí práce Ing. Ondřeji Strakovi, Ph.D.

[40] ALLWRIGHT, Stephen. What is a good RMSE value? Simply explained. In: *Stephenallwright.com* [online]. [cit. 2023-05-14]. Dostupné z: https://stephenallwright.com/good-rmse-value/

[41] About Austrian Airlines. In: *Austrian.com* [online]. [cit. 2023-05-11]. Dostupné z: https://www.austrian.com/us/en/company

[42] Boeing 777-200ER. In: *Austrian.com* [online]. [cit. 2023-05-11]. Dostupné z: https://www.austrian.com/jp/en/boeing-777-200er

[43] *Flightradar24.com* [online]. [cit. 2023-05-11]. Dostupné z: https://www.flightradar24.com/

[44] EVA Values. In: *Evaair.com* [online]. [cit. 2023-05-11]. Dostupné z: https://www.evaair.com/en-global/about-eva-air/about-us/eva-values/

[45] *Flightradar24.com* [online]. [cit. 2023]. Dostupné z: https://www.flightradar24.com/

[46] 787-10. In: *Evaair.com* [online]. [cit. 2023-05-11]. Dostupné z: https://www.evaair.com/en-us/fly-prepare/our-fleets/passenger-airplanes/787-10/

[47] Data Browser: Air passenger transport between the main airports of Austria and their main partner airports (routes data). In: *Ec.europa.eu* [online]. [cit. 2023-05-12]. Dostupné z: https://ec.europa.eu/eurostat/databrowser/view/AVIA_PAR_AT/default/table?lang=en

[48] SUBRAMANIAN, R., Marisetty MURALI, B DEEPAK, P DEEPAK, Hamsinipally REDDY a R. SUDHARSAN. Airline Fare Prediction Using Machine Learning Algorithms. In: *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)* [online]. IEEE, 2022, s. 877-884 [cit. 2023-05-14]. ISBN 978-1-6654-0118-0. Dostupné z: doi:10.1109/ICSSIT53264.2022.9716563

[49] Air France. In: *Centreforaviation.com* [online]. [cit. 2023-05-14]. Dostupné z: https://centreforaviation.com/data/profiles/airlines/air-france-af

[50] Our Boeing 777. In: *Airfrance.cz* [online]. [cit. 2023-05-14]. Dostupné z: https://wwws.airfrance.cz/cs/information/prepare/nos-avions/b777

[51] SKYTRAX. The World's Top 10 Airports of 2023. In: *Worldairportawards.com* [online]. [cit. 2023-05-14]. Dostupné z: https://www.worldairportawards.com/the-worlds-top-10-airports-of-2023/

[52] SKYTRAX. World's Best Business Class Airlines 2022. In: *Worldairlineawards.com* [online]. [cit. 2023-05-14]. Dostupné z: https://www.worldairlineawards.com/worlds-best-business-class-airlines-2022/

[53] Boeing 777-300 Extended Range. In: *Singaporeair.com* [online]. [cit. 2023-05-14]. Dostupné z: https://www.singaporeair.com/saar5/pdf/OurFleet/777-300ER.pdf

[54] AUTOŘI WEBU ČT24. *Prahu a Tchaj-pej má spojit přímá letecká linka, stát ji povolil* [online]. [cit. 2023-05-14]. Dostupné z: https://ct24.ceskatelevize.cz/svet/3570235-prahu-a-tchaj-pej-ma-spojit-prima-letecka-linka-stat-ji-povolil

Přílohy

Příloha 1: Výdaje při operování letounu na dálkových tratích

	Mainline 247 seats[1] 3,200 nautical miles (nm); 740 km/h block speed 85% load factor		Low-cost carrier 291 seats[1] 3,200 nm; 740 km/h block speed 85% load factor		Charter 291 seats[1] 3,200 nm; 740 km/h block speed 95% load factor		
Line item	**$/BH[2]**	**Input cost**	**$/BH**	**Input cost**	**$/BH**	**Input cost**	**Unit**
Aircraft/ insurance	**2,467**	879,000	**2,467**	879,000	**2,467**	879,000	$/month
		12		12		12	BH/day
Fuel	**4,386**	1,700	**4,386**	1,700	**4,386**	1,700	Gallons/BH
		2.5		2.5		2.5	$/gallon
Maintenance	**855**	855	**855**	855	**855**	855	$/BH
		300,000		300,000		300,000	Salary/year/crew, $
Cockpit	**544**	**60%**	**411**	**20%**	**411**	**20%**	Benefit load, %
		10,000		10,000		10,000	Training cost/year/crew, $
		75		75		75	BH/month/crew
		32,000		32,000		32,000	Salary/flight attendant/year, $
Cabin crew	**390**	**40%**	**310**	**10%**	**310**	**10%**	Benefit load, %
		8		8		8	Cabin crew/aircraft
		2,000		2,000		2,000	Training cost/year/crew, $
		80		80		80	BH/month/crew
Hotel accommodation	**167**	**200**	**125**	**150**	**125**	**150**	$/crew member
		2,000		1,500		1,500	$/turn, aircraft
Airport/navigation	**831**	**2,530**	**661**	**2,530**	**679**	**2,530**	$/leg, landing and navigation
		10		5		5	$/passengers, handling
On board	**524**	**20**	**310**	**10**	**173**	**5**	$/passengers
Sales and distribution	**918**	**35**	**310**	**10**	**104**	**3**	$/passengers
General and administrative	**315**	**12**	**155**	**5**	**69**	**2**	$/passengers
Total cost $/BH	**11,396**		**9,989**		**9,579**		
CASK[3]	**6.2**		**4.6**		**4.4**		

[1]Seat counts based on announced configurations by carriers fitting the respective archetype.
[2]Block hour.
[3]Cost per available seat kilometer.

McKinsey&Company | Source: *Aircraft Commerce*; annual reports; Boeing; US Department of Transportation, Form 41; International Air Transport Association; International Civil Aviation Organization; pprune.org; McKinsey analysis

Příloha 2 – Bayesovský odhad parametru modelu dolního kvartilu *pro pár VIE - BKK*

```
clear, clc, mode(0), xdel(winsid())
///Odhad regresního modelu $y_t=b_0u_{1;t}+a_ly_{t-1}+b_lu_{1;t-1}+k+e_t$

A=csvRead('data1.csv');

for i = 2:length(A)
    if isnan(A(i))
        A(i) = A(i-1);
    end
end
y=A(1:32,5); //data Austrian Airlines
u=A(1:32,2); //data Eva Air

nd=size(y,1);   ///počet dat


//////////////////////Odhadování
nps=4+1;    //rozměr informační matice V - musím spočítat ručně: počet koeficientů+1
//V=zeros(nps,nps);    //počáteční informační matice V
V=0,001*eye(nps,nps); //rozměr informační matice o velikosti npsxnps
k=0;                  //počáteční počítadlo k

for t=2:nd;           //od 2 protože je to rovnice 2. řádu
    Ps=[y(t) u(t) y(t-1) u(t-1) 1];//rozšířený regresní vektor
    V=V+Ps'*Ps;       //update statistiky V
    k=k+1;            //update statistiky k
end
//rozklad informační matice V
Vy=V(1,1);        //část Vy
Vyps=V(2:$,1);    //část Vyps
Vps=V(2:$,2:$);   //část Vps

//bodově odhady
theta_odhad=inv(Vps)*Vyps;      //bodové odhady regresních koeficientů
r_odhad=(Vy-Vyps'*inv(Vps)*Vyps)/k; //bodový odhad rozptylu

////////////Simulace s odhadem parametrů pro predikci výstupu
yp(1)=1; yp(2)=1; //počáteční predikce -YP predikovane !

for t=2:nd
    ps=[u(t) yp(t-1) u(t-1) 1];  //regresní vektor s predikcí
    yp(t)=ps*theta_odhad+sqrt(r_odhad)*rand(1,1,'n');//generování predikce
end


////výsledky
disp('Odhad regresních koeficientů',theta_odhad) //zobrazení dat
disp('Odhad rozptylu',r_odhad)

R2 = 1 - sum((y - yp).^2) / sum((y - mean(y)).^2)
disp('R2: '+string(R2))



RMSE = sqrt(mean((yp - y).^2));

disp('RMSE: '+string(RMSE))


x = (32:-1:1)';
plot(x,y(1:nd),'b-')
plot(x,yp(1:nd),'r-')


legend('cena letenek Austrian Airlines','odhad parametrů cen pro Austrian Airlines',4);
title('Bayesův model pro dolní kvantil','fontsize',5)
ylabel('Cena (CZK)','fontsize',5)
xlabel('Počet dní před odletem','fontsize',5)

```

```
 "Odhad regresních koeficientů"

  0.1815322
 -0.1694861
  0.1821614
  25783.571

 "Odhad rozptylu"

  2188785.5
R2  =
  0.9046441

 "R2: 0.9046441"

 "RMSE: 1990.8393"
```

Příloha 3 – Metoda nejmenších čtverců dolního kvartilu *pro pár VIE - BKK*

```
clear, clc, mode(0), xdel(winsid())
A=csvRead('data.csv');
for i = 2:length(A)
    if isnan(A(i))
        A(i) = A(i-1);
    end
end
y = A(1:31,5); // sloupec s závislá proměnnou, Austrian Airlines
x = A(1:31,2); //nezávislá proměnná EVA Air
X = [ones(length(x),1) x]; //matice 63x3´, kde první sloupec jsou jedničky a další jsou data z x
Y = y;
b = inv(X'*X)*X'*Y; //vzorec
y_pred = X*b;       //výpočet odhadovaných hodnot

//Výpočet RMSE
RMSE = sqrt(sum((y-y_pred).^2)/length(y))
x = (32:-1:1)';
//Vypsání koeficientů a RMSE
disp("Koeficienty:")
disp(b)
disp("RMSE:")
disp(RMSE)
a = (31:-1:1)'; //obrácení souřadnic
plot(a,y,'b-')
plot(a,y_pred,'r-')

R2 = 1 - sum((y - y_pred).^2) / sum((y - mean(y)).^2)
disp('R2: '+string(R2))
legend('cena letenek Austrian Airlines','odhad parametrů cen pro Austrian Airlines',4);
title('Metoda nejmenších čtverců pro dolni kvantil','fontsize',5)
ylabel('Cena (CZK)','fontsize',5)
xlabel('Počet dní před odletem','fontsize',5)

```

```
RMSE  =
  1908.2670

 "Koeficienty:"

  24753.513
  0.2321308

 "RMSE:"

  1908.2670
R2  =
  0.5836422

 "R2: 0.5836422"
```

Příloha 4 – Bayesovský odhad parametru modelu mediánu pro pár VIE - BKK

```
clear, clc, mode(0), xdel(winsid())
///Odhad regresního modelu $y_t=b_0u_{1;t}+a_1y_{t-1}+b_1u_{1;t-1}+k+e_t$

A=csvRead('data1.csv');

for i =2:length(A)
    if isnan(A(i))
        A(i) = A(i-1);
    end
end
y=A(1:32,6); //data Austrian Airlines
u=A(1:32,3); //data Eva Air

nd=size(y,1);    ///počet dat (data jsou ve sloupci - určují mi řádek (2 by určovalo sloupce))

//////////////////////Odhadování
nps=4+1;    //rozměr informační matice V - musím spočítat ručně
//V=zeros(nps,nps);    //počáteční informační matice V
V=0,00001*eye(nps,nps); //matice
k=0;                    //počáteční počitadlo k

for t=2:nd
    Ps=[y(t) u(t) y(t-1) u(t-1) 1];//rozšířený regresní vektor
    V=V+Ps'*Ps;         //update statistiky V
    k=k+1;              //update statistiky k
end
//rozklad informační matice V
Vy=V(1,1);          //část Vy
Vyps=V(2:$,1);    //část Vyps
Vps=V(2:$,2:$);  //část Vps

//bodové odhady
theta_odhad=inv(Vps)*Vyps;      //bodové odhady regresních koeficientů
r_odhad=(Vy-Vyps'*inv(Vps)*Vyps)/k; //bodový odhad rozptylu

////////////Simulace s odhadem parametrů pro predikci výstupu
yp(1)=1; yp(2)=1; //počáteční predikce -YP predikované !

for t=2:nd
    ps=[u(t) yp(t-1) u(t-1) 1];  //regresní vektor s predikcí
    yp(t)=ps*theta_odhad+sqrt(r_odhad)*rand(1,1,'n');//generování predikce
end

////výsledky
disp('Odhad regresních koeficientů',theta_odhad)
disp('Odhad rozptylu',r_odhad)

R2 = 1 - sum((y - yp).^2) / sum((y - mean(y)).^2);
disp('R2: '+string(R2))

RMSE = sqrt(mean((yp - y).^2));

disp('RMSE: '+string(RMSE))

x = (32:-1:1)';
plot(x,y(1:nd),'b-')
plot(x,yp(1:nd),'r-')
legend('cena letenek Austrian Airlines','odhad parametrů cen pro Austrian Airlines',4);
title('Bayesův model pro medián','fontsize',5)
vlabel('Cena (CZK)','fontsize',5)
xlabel('Počet dni před odletem','fontsize',5)

```

```
"Odhad regresních koeficientů"

 0.1508011
-0.0688113
 0.0917063
 27787.229

"Odhad rozptylu"

 4756929.6

"R2: 0.8925296"

"RMSE: 2231.5716"
```

Příloha 5 – Metoda nejmenších čtverců mediánu pro pár VIE - BKK

```
clear, clc, mode(0), xdel(winsid())

A=csvRead('data.csv');
for i = 2:length(A)
    if isnan(A(i))
        A(i) = A(i-1);
    end
end
y = A(1:31,6); //sloupec s závislá proměnnou, China Airlines
x = A(1:31,3);
X = [ones(length(x),1) x]; //matice 63x3', kde první sloupec jsou jedničky
Y = y;
b = inv(X'*X)*X'*Y;
y_pred = X*b;


//Výpočet RMSE
RMSE = sqrt(sum((y-y_pred).^2)/length(y));

//Vypsání koeficientů a RMSE
disp("Koeficienty:")
disp(b)
disp("RMSE:")
disp(RMSE)
a = (31:-1:1)'; //obrácení souřadnic
plot(a,y,'b-')
plot(a,y_pred,'r-')

R2 = 1 - sum((y - y_pred).^2) / sum((y - mean(y)).^2);
disp('R2: '+string(R2))
legend('cena letenek Austrian Airlines','odhad parametrů cen pro Austrian Airlines',4);
title('Metoda nejmenších čtverců pro medián','fontsize',5)
ylabel('Cena (CZK)','fontsize',5)
xlabel('Počet dni před odletem','fontsize',5)
```

```
"Koeficienty:"

 27693.416
 0.1846180

"RMSE:"

 2295.0422

"R2: 0.3827777"
```

Příloha 6 – Bayesovský odhad parametru modelu sezónní špičky pro pár PRG - HAN

```
clear, clc, mode(0), xdel(winsid())
/// Odhad regresního modelu $y_t=b_0v_{1;t}+c_0v_{2;t}+a_1y_{t-1}+b_1v_{1;t-1}+c_1v_{2;t-1}+a_2y_{t-2}+k+e_t$

A=csvRead('data2.csv');

for i = 2:length(A)
    if isnan(A(i))
        A(i) = A(i-1);
    end
end
y=A(:,2); // data Chinese Airlines
v1t=A(:,3); // data Qatar Airways
v2t=A(:,4); ///data Emirates

v1=v1t';    //transponování pro násobení matic
v2=v2t';

nd=size(y,1);    ///počet dat (data jsou ve sloupci - určují mi řádek (2 by určovalo sloupce))


//////////////////////Odhadování
nps=7+1;    //rozměr informační matice V - musím spočítat ručně
//V=zeros(nps,nps);    //počáteční informační matice V
V=0,0000001*eye(nps,nps);
k=0;                //počáteční počítadlo k

for t=3:nd
    Ps=[y(t) v1(t) v2(t) y(t-1) v1(t-1) v2(t-1) y(t-2) 1];//rozšířený regresní vektor
    V=V+Ps'*Ps;        //update statistiky V
    k=k+1;            //update statistiky k
end
//rozklad informační matice V
Vy=V(1,1);        // část Vy
Vyps=V(2:$,1);    // část Vyps
Vps=V(2:$,2:$);   //část Vps

//bodové odhady
theta_odhad=inv(Vps)*Vyps;      //bodové odhady regresních koeficientů
r_odhad=(Vy-Vyps'*inv(Vps)*Vyps)/k; //bodový odhad rozptylu

////////////Simulace s odhadem parametrů pro predikci výstupu
yp(1)=1; yp(2)=1; //počáteční predikce -YP predikované !

for t=3:nd
    ps=[v1(t) v2(t) yp(t-1) v1(t-1) v2(t-1) yp(t-2) 1];  //regresní vektor s predikcí
    yp(t)=ps*theta_odhad+sqrt(r_odhad)*rand(1,1,'n');//generování predikce
end

////výsledky
disp('Odhad regresnich koeficientů',theta_odhad)
disp('Odhad rozptylu',r_odhad)

R2 = 1 - sum((y - yp).^2) / sum((y - mean(y)).^2);
disp('R2: '+string(R2))

RMSE = sqrt(mean((yp - y).^2));

disp('RMSE: '+string(RMSE))

plot(y(1:nd),'b-')
plot(yp(1:nd),'r-')
legend('cena letenek China Airlines','odhad parametrů cen pro China Airlines',4);
title('Model sezónní špičky China Airlines na lince PRG-HAN','fontsize',5)
ylabel('Cena (CZK)','fontsize',5)
xlabel('Čas','fontsize',5)

```

```
"Odhad regresnich koeficientů"

 0.1448270
 0.0126986
 0.3471037
 0.0826394
-0.0441947
 0.1486487
 6522.8886

"Odhad rozptylu"

 5066295.9

"R2: 0.7314237"

"RMSE: 2825.1273"
```

Příloha 7 – Metoda nejmenších čtverců mediánu pro pár PRG - HAN

```
clear, clc, mode(0), xdel(winsid())

A=csvRead('data.csv');
for i = 2:length(A)
    if isnan(A(i))
        A(i) = A(i-1);
    end
end
y = A(:,2); // sloupec s závislá proměnnou, China Airlines
x1 = A(:,3);
x2 = A(:,4);// sloupce se nezávislými proměnnými, Emirates a Qatar Airways
X = [ones(length(x1),1) x1 x2]; //matice 63x3', kde první sloupec jsou jedničky
Y = y;
b = inv(X'*X)*X'*Y;
y_pred = X*b;

// Výpočet RMSE
RMSE = sqrt(sum((y-y_pred).^2)/length(y));

// Vypsání koeficientů a RMSE
disp("Koeficienty:")
disp(b)
disp("RMSE:")
disp(RMSE)
plot(y,'b-')
plot(y_pred,'r-')

R2 = 1 - sum((y - y_pred).^2) / sum((y - mean(y)).^2);
disp('R2: '+string(R2))
legend('cena letenek China Airlines','odhad parametrů cen pro China Airlines',4);
title('Model sezónní špičky China Airlines na lince PRG-HAN','fontsize',5)
ylabel('Cena (CZK)','fontsize',5)
xlabel('Čas','fontsize',5)
```

```
"Koeficienty:"

 14880.287
 0.2611719
 0.0343558

"RMSE:"

 2771.0426

"R2: 0.5531896"
```