



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta dopravní
Ústav letecké dopravy

Analýza cen turistických letenek z vybraných leteckých uzlů
Analysis of Tourist Ticket Prices from Selected Airport Hubs

Bakalářská práce

Studijní program: B3710 – Technika a technologie v dopravě a spojích

Studijní obor: PIL – Profesionální pilot

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

Petr Korpas

Praha 2023



K621.....Ústav letecké dopravy

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Petr Korpas

Studijní program (obor/specializace) studenta:

bakalářský – PIL – Profesionální pilot

Název tématu (česky): **Analýza cen turistických letenek z vybraných leteckých uzlů**

Název tématu (anglicky): Analysis of Tourist Ticket Prices from Selected Airport Hubs

Zásady pro vypracování

Při zpracování bakalářské práce se řiďte následujícími pokyny:

- Cílem práce je srovnání cen letenek z vybraných leteckých uzlů do turistických destinací. Je potřeba najít charakteristiky jednotlivých spojení a jejich vliv na ceny.
- Proved'te rešerši vlivu různých faktorů na ceny letenek.
- Prozkoumejte ekonomická, provozní a kontextová data o cenách letenek a stanovte datový set pro analýzu.
- Proved'te sběr dat, jejich zpracování, stanovte metodiku analýzy cen turistických letenek z vybraných leteckých uzlů.
- Proved'te analýzu a interpretujte její závěry.
- Závěry analýzy diskutujte, proved'te validaci výsledků a stanovte výhled pro další zkoumání.

Rozsah grafických prací: dle pokynů vedoucího závěrečné práce


Rozsah průvodní zprávy: minimálně 35 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)

Seznam odborné literatury: Jiang X., - Analysis on price discrimination in airplane tickets 2021
Clemons E., et.al. - Price Dispersion and Differentiation in Online Travel: An Empirical Investigation
Lu J., - Machine learning modeling for time series problem: Predicting flight ticket prices

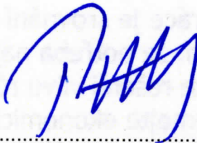
Vedoucí bakalářské práce: **doc. Ing. Peter Vittek, Ph.D.**

Datum zadání bakalářské práce: **7. října 2022**
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

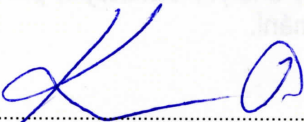
Datum odevzdání bakalářské práce: **7. srpna 2023**
a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia


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




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

Potvrzuji převzetí zadání bakalářské práce.


.....
Petr Korpas
jméno a podpis studenta

V Praze dne.....1. května 2023



Abstrakt

Práce se věnuje analýze cen letenek, jejího průběhu s pomocí využití regresních modelů. Jedná se o linky z letišť Paris Charles de Gaulle, London Heathrow a Munich International do destinací Barcelona El-Prat, Ibiza a Tenerife Sur Reina Sofia. Předmětem analýzy je taktéž studium cen letenek z pohledu různých segmentů zákazníků, které byly v práci definovány jako segmenty týdeních, desetidenních, dvoutýdenních nebo víkendových návštěvníků. Data byla sbírána pomocí HTTP requestů na stránce kiwi.com. Byly zjištěny rozdíly mezi časovými intervaly – často aerolinky naceňují letenky jinak pro každý časový interval strávený v destinaci. Taktéž byl pozorován jiný průběh ceny pro low-cost a legacy aerolinky. Jedná se jak o rozdíly ve výši ceny letenky, tak o průběh naceňování – někdy cena klesá na celém sledovaném časovém úseku (90 dní do odletu), jindy stoupá. Také se může jednat o funkci polynomiální s konvexním i konkávním charakterem.

Klíčová slova: analýza, cenotvorba letenek, revenue management



Abstract

Subject of this thesis is the analysis of flight ticket prices, its respective course (how it is changing with respect to the number of days until departure) with the use of regression models. Origin airports are Paris Charles de Gaulle, London Heathrow and Munich International and the tourist destinations were chosen as Barcelona El-Prat, Ibiza and Tenerife Sur Reina Sofia. The subject of the analysis the study of flight tickets in regards with different segments of customers – one week, ten days, two weeks or a weekend customer. The dataset for this thesis was obtained using HTTP requests on the website kiwi.com. The results were mixed – there are times when airlines price the flight tickets according to the length of stay in the destination, the progress of predicted price was also different for low-cost and legacy airlines and there were differences not only in price but also in the slope (price going up or down) and curvature (convex and concave) of regressions throughout the whole period for which the graphs were created (90 days to departure).

Keywords: analysis, flight ticket pricing, revenue management



Poděkování

Chtěl bych poděkovat svému vedoucímu, panu docentu Peteru Vittekovi za vhled a přínosné informace týkajících se ekonomiky trhu s letenkami. Taktéž bych chtěl poděkovat paní doktorce Pavle Pecherkové za pomoc při tvorbě statistických metod. Taktéž bych chtěl poděkovat firmě Flightera.net za poskytnutí dat na mnou studovaných linkách a v neposlední řadě bych chtěl poděkovat svým rodičům a přátelům za podporu po dobu celého studia na ČVUT.

I would like to express thanks to the supervisor of my Bachelor's Thesis – doc. Ing. Peter Vittek, Ph.D. for the insight on the economics in the flight ticket sector and to the same extent I would like to express my gratitude towards Ing. Pavla Pecherková, Ph.D. for her help in creating the statistical methods. I would also like to thank the company Flightera.net for providing useful data for the studied lines and last but not least I would like to thank my family and friends for their support throughout my studies at CTU.



Čestné prohlášení

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.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č.121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

Praze dne 7. srpna 2023

.....

Podpis



Obsah

Úvod	15
1 Rešerše	16
1.1 Informace o trhu	16
1.2 Letiště odletu	19
1.2.1 Londýn – Heathrow (LHR)	19
1.2.2 Paříž – Charles de Gaulle (CDG)	21
1.2.3 Mnichov (MUC)	23
1.3 Letiště příletu	24
1.4 Výběr linek a časových intervalů	24
2 Metodika	26
2.1 Sběr dat	26
2.2 Zpracování souborů	28
2.3 Zpracování dat	29
2.3.1 Základní zpracování	29
2.3.2 Pokročilé zpracování	31
3 Prezentace výsledků	33
3.1 Obecná data	33
3.2 Podíl aerolinek na trhu	34
3.3 Spojnicové grafy surových dat	34
3.4 Regrese	39
3.4.1 Regrese sdružené podle aerolinky	39
3.4.2 Regrese sdružené podle času	43
3.5 Ostatní data	45
3.5.1 Počet vypravených sedadel	45
3.5.2 Data z Eurostat a CAA.co.uk	47
4 Diskuze výsledků	50



5 Závěr	55
Seznam použité literatury	56
Přílohy	58
A Spojnicové grafy surových dat	58
B Podíl na trhu	66
C Regresní modely	75



Seznam obrázků

1.1	Mezinárodní linky ze Spojených států amerických [9]	17
1.2	Počet přepravených pasažérů - EU a celkem [16]	20
1.3	Spádová oblast letišť LHR a STN [17]	20
1.4	Hrubý domácí produkt na obyvatele v Londýně [18]	21
1.5	Mapa Paříže [20]	22
1.6	Mapa Mnichova [20]	23
2.1	Metody HTTP requestů	26
2.2	Vytváření collection Lety a .json souborů	27
2.3	Vytváření .xlsx souborů	28
2.4	Slučování .xlsx souborů	29
2.5	Tvorba slovníku slovníků	30
2.6	Vývojový diagram postupu v práci	32
3.1	Vývoj ceny letenek na lince CDG-BCN, aerolinka: AF	35
3.2	Vývoj ceny letenek na lince CDG-BCN, aerolinka: U2	36
3.3	Vývoj ceny letenek na lince CDG-BCN, aerolinka: VY	36
3.4	Vývoj ceny letenek na lince CDG-IBZ, aerolinka: AF	37
3.5	Vývoj ceny letenek na lince CDG-TFS, aerolinka: AF	38
3.6	Vývoj ceny letenek na lince CDG-TFS, aerolinka: U2	38
3.7	Regresní modely na lince CDG-BCN, aerolinka: AF	39
3.8	Regresní modely na lince CDG-BCN, aerolinka: U2	40
3.9	Regresní modely na lince CDG-BCN, aerolinka: VY	40
3.10	Regresní modely na lince MUC-IBZ, aerolinka: DE	41
3.11	Regresní modely na lince MUC-IBZ, aerolinka: LH	41
3.12	Regresní modely na lince LHR-IBZ, aerolinka: BA	42
3.13	Regresní modely na lince LHR-TFS, aerolinka: BA	43
3.14	Regresní modely na lince CDG-BCN, časový interval: 10 dní	43
3.15	Regresní modely na lince CDG-BCN, časový interval: 2 týdny	44
3.16	Regresní modely na lince CDG-BCN, časový interval: týden	44
3.17	Počet vypravených sedadel z CDG	45



3.18 Počet vypravených sedadel z LHR	46
3.19 Počet vypravených sedadel z MUC	46
3.20 Množství přepravených cestujících z MUC	47
3.21 Množství přepravených cestujících z CDG	48
A.1 Vývoj ceny na lince CDG-BCN, aerolinka: EC	58
A.2 Vývoj ceny na lince LHR-BCN, aerolinka: BA	58
A.3 Vývoj ceny na lince LHR-BCN, aerolinka: IB	59
A.4 Vývoj ceny na lince LHR-IBZ, aerolinka: BA	59
A.5 Vývoj ceny na lince LHR-IBZ, aerolinka: IB	60
A.6 Vývoj ceny na lince LHR-IBZ, aerolinka: BA	60
A.7 Vývoj ceny na lince MUC-BCN, aerolinka: LH	61
A.8 Vývoj ceny na lince MUC-BCN, aerolinka: VY	61
A.9 Vývoj ceny na lince MUC-IBZ, aerolinka: IB	62
A.10 Vývoj ceny na lince MUC-IBZ, aerolinka: BA	62
A.11 Vývoj ceny na lince MUC-TFS, aerolinka: 1L	63
A.12 Vývoj ceny na lince MUC-TFS, aerolinka: 4Y	63
A.13 Vývoj ceny na lince MUC-TFS, aerolinka: DE	64
A.14 Vývoj ceny na lince MUC-TFS, aerolinka: LH	64
A.15 Vývoj ceny na lince MUC-TFS, aerolinka: MBU	65
A.16 Vývoj ceny na lince MUC-TFS, aerolinka: X3	65
B.1 Podíl na trhu, linka CDG-BCN, 10 dní	66
B.2 Podíl na trhu, linka CDG-BCN, 2 týdny	66
B.3 Podíl na trhu, linka CDG-BCN, týden	66
B.4 Podíl na trhu, linka CDG-BCN, víkend	66
B.5 Podíl na trhu, linka CDG-IBZ, 10 dní	67
B.6 Podíl na trhu, linka CDG-IBZ, 2 týdny	67
B.7 Podíl na trhu, linka CDG-IBZ, týden	67
B.8 Podíl na trhu, linka CDG-IBZ, víkend	67
B.9 Podíl na trhu, linka CDG-TFS, 10 dní	68
B.10 Podíl na trhu, linka CDG-TFS, 2 týdny	68
B.11 Podíl na trhu, linka CDG-TFS, týden	68



B.12 Podíl na trhu, linka CDG-TFS, víkend	68
B.13 Podíl na trhu, linka LHR-BCN, 10 dní	69
B.14 Podíl na trhu, linka LHR-BCN, 2 týdny	69
B.15 Podíl na trhu, linka LHR-BCN, týden	69
B.16 Podíl na trhu, linka LHR-BCN, víkend	69
B.17 Podíl na trhu, linka LHR-IBZ, 10 dní	70
B.18 Podíl na trhu, linka LHR-IBZ, 2 týdny	70
B.19 Podíl na trhu, linka LHR-IBZ, týden	70
B.20 Podíl na trhu, linka LHR-IBZ, víkend	70
B.21 Podíl na trhu, linka LHR-TFS, 10 dní	71
B.22 Podíl na trhu, linka LHR-TFS, 2 týdny	71
B.23 Podíl na trhu, linka LHR-TFS, týden	71
B.24 Podíl na trhu, linka LHR-TFS, víkend	71
B.25 Podíl na trhu, linka MUC-BCN, 10 dní	72
B.26 Podíl na trhu, linka MUC-BCN, 2 týdny	72
B.27 Podíl na trhu, linka MUC-BCN, týden	72
B.28 Podíl na trhu, linka MUC-BCN, víkend	72
B.29 Podíl na trhu, linka MUC-IBZ, 10 dní	73
B.30 Podíl na trhu, linka MUC-IBZ, 2 týdny	73
B.31 Podíl na trhu, linka MUC-IBZ, týden	73
B.32 Podíl na trhu, linka MUC-IBZ, víkend	73
B.33 Podíl na trhu, linka MUC-TFS, 10 dní	74
B.34 Podíl na trhu, linka MUC-TFS, 2 týdny	74
B.35 Podíl na trhu, linka MUC-TFS, týden	74
B.36 Podíl na trhu, linka MUC-TFS, víkend	74
C.1 Regresní modely na lince CDG-IBZ, aerolinka: AF	75
C.2 Regresní modely na lince CDG-TFS, aerolinka: AF	75
C.3 Regresní modely na lince CDG-TFS, aerolinka: U2	76
C.4 Regresní modely na lince MUC-BCN, aerolinka: LH	76
C.5 Regresní modely na lince MUC-BCN, aerolinka: VY	77
C.6 Regresní modely na lince MUC-TFS, aerolinka: DE	77



C.7 Regresní modely na lince MUC-TFS, aerolinka: MBU	78
C.8 Regresní modely na lince MUC-TFS, aerolinka: X3	78
C.9 Regresní modely na lince CDG-IBZ, časový interval: víkend	79
C.10 Regresní modely na lince CDG-TFS, časový interval: 10 dní	79
C.11 Regresní modely na lince CDG-TFS, časový interval: 2 týdny	80
C.12 Regresní modely na lince CDG-TFS, časový interval: týden	80
C.13 Regresní modely na lince CDG-TFS, časový interval: víkend	81
C.14 Regresní modely na lince LHR-BCN, časový interval: 2 týdny	81
C.15 Regresní modely na lince LHR-BCN, časový interval: týden	82
C.16 Regresní modely na lince LHR-IBZ, časový interval: 10 dní	82
C.17 Regresní modely na lince LHR-IBZ, časový interval: 2 týdny	83
C.18 Regresní modely na lince LHR-IBZ, časový interval: týden	83
C.19 Regresní modely na lince LHR-IBZ, časový interval: víkend	84
C.20 Regresní modely na lince LHR-TFS, časový interval: 10 dní	84
C.21 Regresní modely na lince LHR-TFS, časový interval: 2 týdny	85
C.22 Regresní modely na lince LHR-TFS, časový interval: týden	85
C.23 Regresní modely na lince LHR-TFS, časový interval: víkend	86
C.24 Regresní modely na lince MUC-BCN, časový interval: 10 dní	86
C.25 Regresní modely na lince MUC-BCN, časový interval: 2 týdny	87
C.26 Regresní modely na lince MUC-BCN, časový interval: týden	87
C.27 Regresní modely na lince MUC-BCN, časový interval: víkend	88
C.28 Regresní modely na lince MUC-IBZ, časový interval: 10 dní	88
C.29 Regresní modely na lince MUC-IBZ, časový interval: 2 týdny	89
C.30 Regresní modely na lince MUC-IBZ, časový interval: týden	89
C.31 Regresní modely na lince MUC-IBZ, časový interval: víkend	90
C.32 Regresní modely na lince MUC-TFS, časový interval: 10 dní	90
C.33 Regresní modely na lince MUC-TFS, časový interval: 2 týdny	91
C.34 Regresní modely na lince MUC-TFS, časový interval: týden	91



Seznam tabulek

1.1	Množství pasažérů ve spádové oblasti MUC	24
1.2	HDP na obyvatele konkrétní regiony v číslech [23]	24
1.3	Vzdálenosti v kilometrech z počátečních letišť do destinací	25
1.4	Doby trvání letů z počátečních letišť do destinací	25
3.1	Průměrné ceny zpátečních letenek z CDG	33
3.2	Průměrné ceny zpátečních letenek z LHR	33
3.3	Průměrné ceny zpátečních letenek z MUC	34
3.4	Podíl na trhu, linka LHR-IBZ	34
3.5	Podíl na trhu, linka LHR-TFS	34
3.6	Směrodatné odchytky na trati CDG-BCN pro jednotlivé aerolinky	37
3.7	Statistiky pro linky z MUC	47
3.8	Statistiky pro linku z CDG	48
3.9	Load factor LHR [24]	49
4.1	Průběh regresí na lince CDG-BCN v posledních 30 dnech do odletu	52
4.2	Průběh regresí na lince CDG-IBZ v posledních 30 dnech do odletu	52
4.3	Průběh regresí na lince CDG-TFS v posledních 30 dnech do odletu	52
4.4	Průběh regresí na lince LHR-IBZ v posledních 30 dnech do odletu	52
4.5	Průběh regresí na lince LHR-TFS v posledních 30 dnech do odletu	53
4.6	Průběh regresí na lince MUC-BCN v posledních 30 dnech do odletu	53
4.7	Průběh regresí na lince MUC-IBZ v posledních 30 dnech do odletu	53
4.8	Průběh regresí na lince MUC-TFS v posledních 30 dnech do odletu	53



Seznam symbolů a zkratk

1L	CitizenPlane	CitizenPlane
4Y	Eurowings	Eurowings
AF	Air France	Air France
BA	British Airways	British Airways
BCN	Letiště Barcelona-El Prat	Barcelona–El Prat Airport
CI	Kanárské ostrovy	Canary Islands
CDG	Letiště Charlese de Gaulla	Paris Charles de Gaulle Airport
DE	Condor Airlines	Condor Airlines
D-W	Durbin-Watsonův	Durbin-Watson
EC	EasyJet Europe	EasyJet Europe
HDP	Hrubý domácí produkt	Gross Domestic Product
IB	Iberia	Iberia
IBZ	Letiště Ibiza	Ibiza Airport
LH	Lufthansa	Lufthansa
LHR	Letiště London Heathrow	London Heathrow Airport
MBU	Marabu	Marabu
MUC	Letiště Mnichov	Munich International Airport
PPS	Kupní síla	Purchasing Power Strength
TFS	Letiště Tenerife Jih	Tenerife Sur Reina Sofia
U2	EasyJet	EasyJet
VY	Vueling	Vueling
X3	TUI fly Deutschland	TUI fly Deutschland



Úvod

Letecký průmysl jako takový významně přispívá k rozvoji globální ekonomiky a hraje důležitou roli při zkracování času při cestování v mezinárodním měřítku. Vysoce dynamický charakter tohoto ekonomického sektoru poskytuje zajímavou oblast výzkumu, zejména s ohledem na ceny letenek. Jejich cena má velmi složitou strukturu, která je ovlivněna mnoha faktory jako je poptávka, náklady na let a konkurence na lince.

Navzdory významnému pokroku v technologiích, nejen za poslední desetiletí, při služeb vyvstává spotřebiteli stále stejná otázka, již bývá, jaká je cena letenky a co její cenu určuje. Podle práce Windmanna [1] průzkum mezi cestujícími odhalil, že byt požadují vyšší standardy služeb, cena letenek stále zůstává hlavním faktorem. Zkoumání tohoto tématu odhaluje složité nejen ekonomické interakce stojící za cenou letenky, kterou poté platíme jako cestující, ale také poskytuje vhled do obchodních strategií leteckých společností a jejich schopnosti se přizpůsobit stále vyvíjejícímu se trhu. Pochopení naceňování letenek přináší všem zúčastněným stranám nespornou výhodu, protože nejen že spotřebitel zjistí, kdy je pro něj nejvýhodnější si letenku koupit, ale poskytovatel letenek (aerolinky) zároveň zjistí, v jakém jsou postavení na již velmi konkurenčním trhu.

V práci byl proveden experiment s daty o cenách letenek, ve kterém byla prostudována nabídka a její vývoj. Předmětem zájmu byla cena a dostupnost letenek z vybraných letišť do určitých destinací. Práce se zaměřuje mimo jiné i na porovnání jednotlivých časových intervalů, respektive dostupnost letenek pro dané segmenty zákazníků, které jsou definovány jako víkend, 1 týden, 10 dní a 2 týdny. Byly vybrány různé typy destinací nejen pro leisure travel, ale i možný business spojený s leisure. Taktéž byly vybrány oba typy aerolinek, a to legacy a low-cost. Dohromady se jedná o dvanáct aerolinek na celkem devíti linkách ze tří letišť odletu do tří destinací ve čtyřech časových intervalech.

Sběr dat byl proveden pokročilým způsobem. Jednalo se o dotazování se HTTP requesty na server kiwi.com, respektive jejich server k tomu určený – api.tequila.kiwi.com. Nejedná se o programy typu WebScraper, který by si stránku nejdříve načetl a poté z ní ukládal data podle toho, co se na stránce vyskytuje.



1 Rešerše

V teoretickém základu práce bylo vycházeno hlavně ze článků, které se soustředily na predikci cen na jednotlivých linkách, přičemž bylo k předpovídání cen přistupováno několika různými způsoby jimiž byly algoritmus k-nejbližších sousedů [2] a metoda náhodného lesa [3]. Přístup vyhodnocování cen letenek pomocí přístrojového učení jsem v mé práci nezvolil, avšak dozvěděl jsem se mnoho poznatků z prací jím se zabývajícími.

1.1 Informace o trhu

Vzhledem ke stále se rozvíjejícímu se trhu s letenkami a zrychlováním jeho tempa se i ztěžuje odhadování jejich cen. Letenky jako takové se taktéž každoročně vyvíjejí s tím, jak na trh nastoupili nízkonákladoví dopravci. Přiměli tak nejen letiště snížit svoje přistávací poplatky, ale i ostatní "legacy" aerolinky snížit své ceny za letenky vlivem zvýšené konkurence a strachu tradičních dopravců o snížení počtu zákazníků na daných trasách [4, 5].

Mnoho studií se tomuto tématu nevěnuje, protože jde o velmi komplexní tematiku zejména z důvodu střežení obchodního tajemství každé letecké společnosti, na kterém se podílelo mnoho lidí, a zároveň jej každý den vylepšuje, takže nechtějí, aby se veřejnost dozvěděla, jakým způsobem se letenky naceňují a aerolinka tak přišla o své zisky.

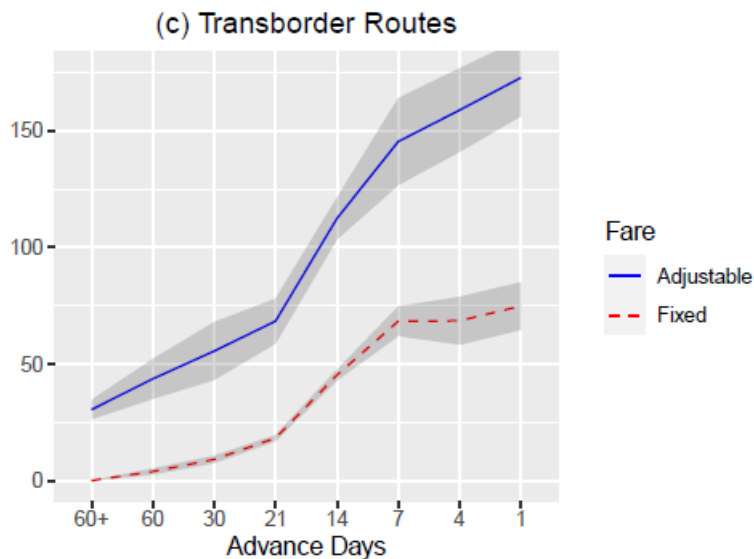
Studie Lewise z roku 2021 [6] taktéž říká, že se ve mnoha studiích přehlíží důležité faktory a nuance, které vznikají na trhu, ve kterém se cenově diskriminuje. I když se jedná o článek z prostředí trhu Spojených států amerických, tak jej můžeme aplikovat i na trh evropský, protože je tu důkaz podobnosti trhu alespoň co se nízkonákladových společností týče – společnost Ryanair, která okopírovala model aerolinky Southwest a je s ním zde velmi úspěšná a tradiční aerolinky s ní dnes již musejí počítat, když své letenky naceňují.

Pokud ale vycházím ze studie Stavinske z roku 2001 [7], tak se dozvídáme, že je cenová diskriminace tím větší, čím je větší koncentrace aerolinek na trhu, což podporuje přirozenou intuici, která říká, že aerolinku drží její unikátní pozice na trhu jako například roster, kterého se drží, a dominance na letišti, což znamená kolik letadel vypraví za den z daného místa, a následně jí to umožní si zachovat svou pozici na trhu ve vztahu k neelastickým, business cestujícím. Může si tak dovolit zvýšit marže na business letenkách a zachovat si tak profit, i když by byla nucena snížit ceny letenek pro cestující elastické (turisty). Pokud by náhodou přišla konkurenční aerolinka na trh s tím,

že bude nabízet prakticky stejné letenky, nicméně s nižší cenovkou, tak to nebude problém, a díky tomuto se fakticky podílí na cenové diskriminaci letenek pro cenově elastický sektor a zároveň si zachovává pozici v dalším segmentu.

Vzhledem ke sběru dat v náhodném čase během dne lze faktor času koupi letenky během dne vyloučit, avšak existuje důkaz [8], že ceny letenek kolísají i během dne, a to zejména kvůli času, kdy si letenku pravděpodobně koupí neelastický (business) zákazník – nejvyšší ceny v dopoledních hodinách a nejnižší navečer, kdy si ji naopak koupí cestující elastický – turista.

Velmi přínosný byl příspěvek z roku 2021 [9], ve kterém jsem se dozvěděl o cenové diskriminaci cen letenek na americkém trhu u vnitrostátních a letech mimo hranice státu. Chování ceny bylo velmi rozdílné mezi oběma lety (viz obrázek 1.1), přičemž když se podíváme na mezinárodní lety, které nás pro potřeby této práce budou zajímat nejvíce, tak jak pro 60dnové, tak pro 90dnové okno existuje velmi agresivní cenová diskriminace v závislosti na počtu dní do odletu.



Obrázek 1.1: Mezinárodní linky ze Spojených států amerických [9]

V tomto článku Chandra pracuje s daty od jedné soukromé aerolinky z Toronta v Kanadě (ve článku se k aerolince referuje jako k North Air, nicméně pravé jméno nebylo zveřejněno), avšak tento dataset přinesl vylepšení oproti datasetům z jiných článků, které používaly standardní DB1B dataset od ministerstva dopravy USA. Jedná se o způsob, jak zaznamenat nějakou cenovou diskriminaci, jímž je den, kdy byla rezervace letenky uskutečněna. Samotný dataset zpracováváný v tomto článku je velmi bohatý – nabízí přes 900000 itinerářů během let 2008-2014, přičemž se jedná o 10 %



celkového objemu prodejů aerolinky. North Air nabízí pouze 3 skupiny "jízdného" – Fixed, Adjustable a Refundable. V případě letenky z kategorie Adjustable má zákazník volnou ruku v tom, kdy chce letět v daný den, na který si letenku koupil – buď dříve nebo později a navíc si může vybrat místo v letadle, na kterém chce cestovat a má taktéž možnost si nechat bezplatně odbavit zavazadlo. Tyto možnosti Fixed letenka nenabízí, a navíc si musí cestující za vše platit. Co se týče letenky z kategorie Refundable, tak tuto Chandra sloučil s kategorií Adjustable s tím, že to nemá prakticky žádný vliv na cenu a její vývoj.

V práci bylo pracováno s daty od areolinek jak low-cost – například Vueling, na jejichž palubu si zákazník může vzít malé příruční zavazadlo většinou o hmotnosti 10 kg a rozměrech 40x20x30 cm [10], tak s aerolinkami "legacy" – AirFrance, na jejichž palubu si ale mohl zákazník vzít jednu malou osobní tašku o maximální hmotnosti 2 kg a rozměrech 40x15x30 cm a jedno příruční zavazadlo o hmotnosti 10 kg a rozměrech 55x25x35 cm [11]. Ve mnoha případech si nešlo ani přilepšit koupí dodatečného odbaveného zavazadla a taktéž nešlo měnit čas odletu, takže se jednalo jistě o kategorii letenky Fixed.

Důležitý byl taktéž bod o určování ceny v závislosti na nedostatku letenek. Tento jev, jemuž se říká inventory management, je výsledkem rozhodování firmy nebo jedince, který má omezené množství něčeho, co potřebuje prodat. Trh s letenkami je velice specifický, protože když se podíváme například na trh s lístky na baseball, jako to bylo provedeno v článku z roku 2012 [12], tak zde lze pozorovat pokles cen s přicházejícím datem hry, pro kterou lístek platí. Podle McAfeeho et al. [13] na druhou stranu letenky ale častokrát ukazují na monopolním trhu opačné chování – potřebují vyšší pravděpodobnost toho, že skutečně určená část letenek zůstane nevyprodaná pro případy zejména business cestujících s nejvyšší finanční výtěžností, kteří se potřebují dostat do své destinace za každou cenu z důvodu potřeby cestovat kvůli obchodu.

Prospekt od IATA z roku 2008 [14] byl též přínosný pro tuto práci, protože v něm určují dopady na poptávku po cestování vzduchem vlivem zvýšení ceny letenky buď jedním nebo všemi dopravci na jedné nebo více linkách. Došli k závěru, že snížení ceny letenky bude nemálo stimulovat celkovou poptávku a zvýší to příjmy – elastická cena. Pokud zvýší cenu všichni dopravci stejným způsobem na menším počtu linek, tak dojde ke dramatickému snížení množství cestujících, stejně jako by se tak stalo v předchozím případě, nicméně dopad na velikost snížení množství nebude tak vysoký, protože se stále bavíme o elastické ceně. Pokud ale cenu zvýší všichni dopravci na většině linek (například z důvodu navýšení daní nebo zvýšení ceny paliva atd.), tak se sice sníží množství



cestujících, ale toto snížení bude ještě méně dramatické než kdyby ke zvýšení došlo všemi dopravci na malém množství linek.

1.2 Letiště odletu

Jako městské páry jsem si vybral jedny z nejvytíženějších letišť v Evropě. Jedná se o letiště nejen v Evropské unii díky Brexitu, který nastal v lednu roku 2020 – letiště Londýn – Heathrow, Paříž – Charles de Gaulle a Mnichov. Důvodem bylo jednak bohatství jednotlivých oblastí, respektive výše HDP na obyvatele, které letiště obsluhují a zároveň bylo očekáváno, že z těchto letišť bude dostatek informací a dat o linkách. Taktéž bylo předpokládáno, že zde bude dostatečné množství dopravců, aby byl vytvořený dataset kompletní.

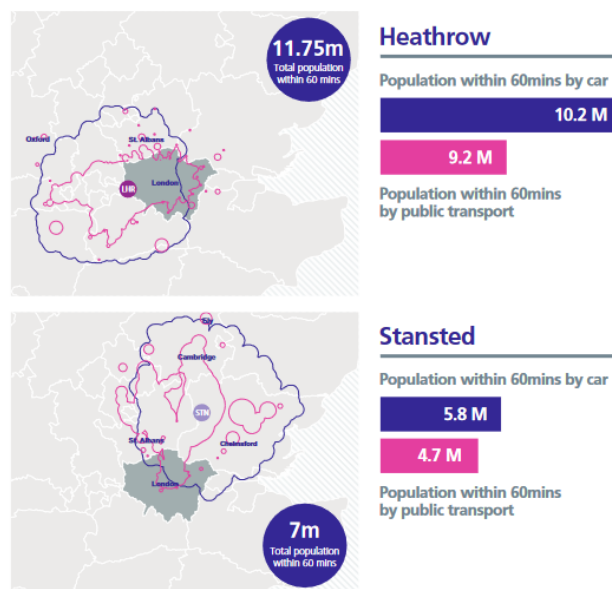
1.2.1 Londýn – Heathrow (LHR)

Letiště London Heathrow v hlavním městě Velké Británie založené v roce 1929 s prvním soukromým letadlem zde přistávajícím v roce 1930 [15] slouží dodnes jako nejvytíženější v celém státě. Jedná se o velký hub pro dopravu pasažérů – před nemocí Covid-19 přes toto letiště procestovalo celkem 80,9 milionu pasažérů za rok, z toho celkem 27,5 milionu jen z Evropské unie. Přes roky 2020 a dále byla tato číslo menší, avšak data o množství pasažérů naznačují, že čísla pasažérů jsou na trajektorii k návratu k těm původním před rokem 2020, jako ostatně většina leteckého sektoru. Pro potřeby této práce mi stačí data o množství přepravených pasažérů, viz graf 1.2.



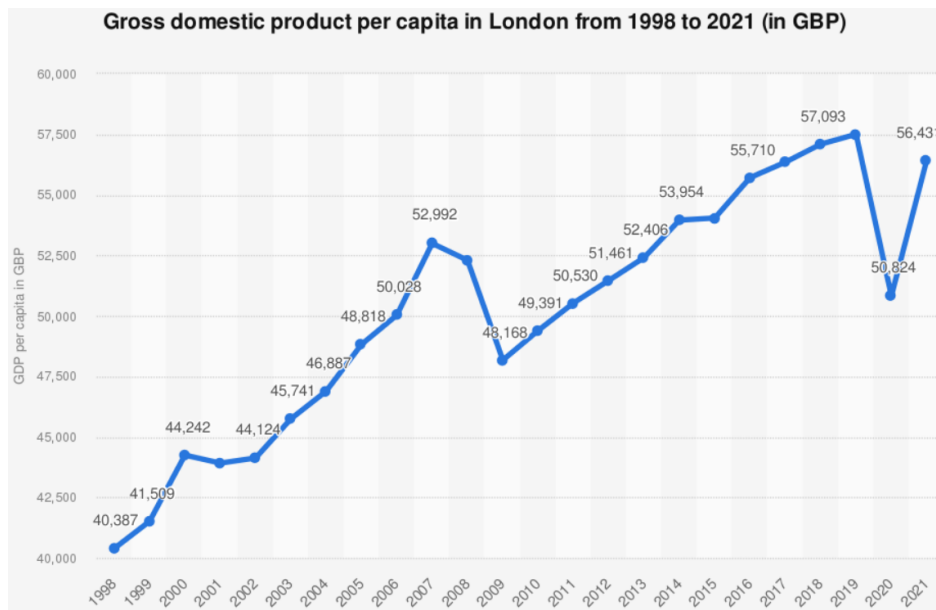
Obrázek 1.2: Počet přepravených pasažérů - EU a celkem [16]

Letiště Heathrow také disponuje nasycenou spádovou oblastí. Když se podíváme na spádovou oblast ve vzdálenosti rovné 60minutové jízdě autem, tak se dostaneme k číslu 11,75 milionu obyvatel žijící v této oblasti, což je v porovnání s dalším letišťem v Londýně (London Stansted) o téměř 50 % vyšší, viz obrázek 1.3 [17].



Obrázek 1.3: Spádová oblast letišť LHR a STN [17]

Londýn je město bohaté, konkrétně HDP na obyvatele z roku 2021 činí 56430 £ (graf 1.4), což bylo v čase psaní práce zaznamenáno ve výši 65940 €, z čehož soudím, že se to projeví i na ceně letenky.



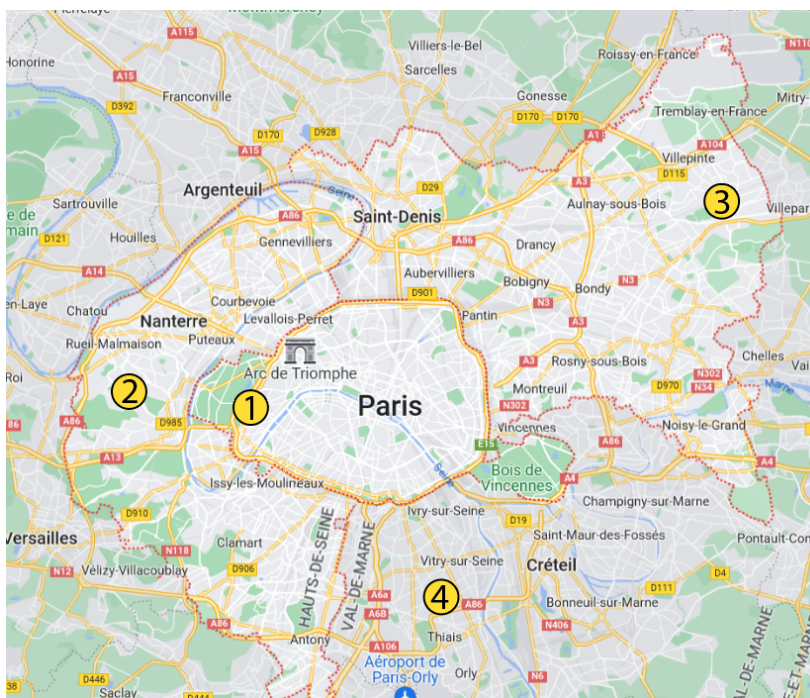
Obrázek 1.4: Hrubý domácí produkt na obyvatele v Londýně [18]

1.2.2 Paříž – Charles de Gaulle (CDG)

V Paříži je určování složitější než v Londýně – na Eurostatu [19] se dozvíme pouze o datech do roku 2020 a vzhledem k situaci na trhu v této době si dovoluji tento údaj nepovažovat za vhodný a pracuji tedy s daty z roku 2019. Další problém je, že tak velká aglomerace jako je Paříž, se dle postupů v Evropské Unii dělí na několik částí NUTS 3. Centrum Paříže se jmenuje Paris (1), další část je Hauts-de-Seine (2) a část, ve které leží letiště Paris Charles de Gaulle se jmenuje Seine-Saint-Denis (3). Každá z těchto oblastí (rozdělení na obrázku 1.5) má jiný hrubý domácí produkt podle obyvatelstva. Část 1 kupodivu tuto statistiku nevede s výší HDP na obyvatele 113900 €, část 2 má HDP na obyvatele nejvyšší a to 115000 €, část 3, která díky přítomnosti letiště CDG velmi pravděpodobně podporuje růst ve zbylých částech, má výši HDP na obyvatele 40700 €, což je přibližně jedna třetina HDP na obyvatele z částí 1 a 2. V části 4 (Val-de-Marne) se nachází letiště Paris-Orly (ORY) a sousedí tak s částmi 1, 2 i 3 a bude konkurenční pro určitý typ provozu z CDG

a svou výší HDP na obyvatele 40100 € celkový průměr jen snižuje, stejně jako část 3. Ten poté bude se započtením všech oblastí 77425 €.

Když se posuneme o úroveň výše – na úroveň NUTS 2, tak se dostaneme k informacím jako jsou například PPS. Ta je pro oblast Paříže a okolo ní (Île de France) ve výši 52300 € a řadí ji mezi špičku v Evropské unii. HDP na obyvatele je v této oblasti na úrovni NUTS 2 56700 €. Pro účely této práce zpracovávám data z regionu NUTS 2.



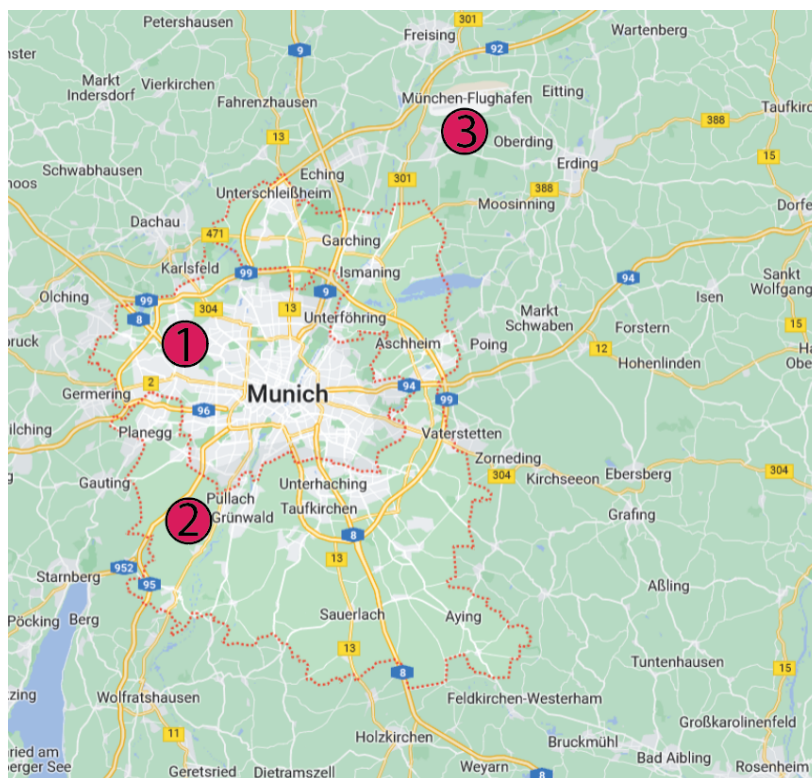
Obrázek 1.5: Mapa Paříže [20]

Spádová oblast byla určena pro CDG jako vzdálenost 200 km od letiště na oficiálních stránkách provozovatele letiště [21]. Jedná se o celkem 25 milionů obyvatel v tomto okruhu. Byť je tato vzdálenost vyšší než u LHR, tak je to pochopitelné z důvodu lepší konektivity zejména kolejovou dopravou, staví zde totiž souprava TGV od dopravce SNCF.

1.2.3 Mnichov (MUC)

Podobně jako v Paříži je určování přesného HDP v Mnichově trochu složitější. Eurostat [19] opět nabízí sektorová data pro jednotlivé části Mnichova (viz obrázek 1.6). Část München Landkreis (1) a München Kreisfrei Stadt (2), pro něž existují data o HDP z roku 2019, a byla zaznamenána výši 116000 €, respektive 84200 €. Letiště München International (3) se ale nachází mimo – dle dělení NUTS 3 leží ve spádové oblasti Freising, pro níž bylo HDP na obyvatele 51100 €. Průměr HDP na obyvatele ze všech oblastí činí 83770 €.

Letiště München International leží na úrovni NUTS 2 v oblasti Oberbayern a pro tu jsou data z roku 2019 následující – PPS = 53800 € a HDP na obyvatele = 59600 €. Vidíme, že se jedná v případě PPS o částku podobnou Paříži a v případě HDP je tomu taktéž – dokonce se blíží i Londýnu. Opět pro potřeby této práce pracuji s HDP na obyvatele v Mnichově na úrovni NUTS 2, a to na hodnotě 59600 €.



Obrázek 1.6: Mapa Mnichova [20]



Pro toto letiště existují přesnější data pro spádovou oblast, než je tomu u CDG – viz tabulka 1.1. Celkem se jedná o 27,2 milionů obyvatel ve spádové oblasti MUC [22].

časová vzdálenost	množství pasažérů
0-60 min	4,3 mil.
61-120 min	7,4 mil.
121-180 min	15,5 mil.

Tabulka 1.1: Množství pasažérů ve spádové oblasti MUC

1.3 Letiště příletu

Nespornou výhodou při pozdější analýze v práci je geografická příslušnost všech tří destinací. Je jí Španělsko, byť Barcelona (region Katalánsko) je zástupcem pevninské části, Ibiza Baleárských ostrovů a Tenerife ostrovů Kanárských. Hrubý domácí produkt na obyvatele v celém Španělsku činil v roce 2021 25498 €, v ostatních regionech pro porovnání v tabulce 1.2.

	HDP na obyvatele (€)
Katalánsko	29,942
Španělsko celkem	25,498
Baleáry	24,866
Kanárské ostrovy	18,990

Tabulka 1.2: HDP na obyvatele konkrétní regiony v číslech [23]

1.4 Výběr linek a časových intervalů

Linky jsem vybíral tak, aby dané destinace poskytovaly více možností vyžití – Barcelona (letiště Barcelona El-Prat – BCN), Tenerife (letiště Tenerife Sur Reina Sofia – TFS) a Ibiza (Aeroport d'Eivissa – IBZ).



Výběr počátků linek byl prováděn na základě vzdáleností mezi jednotlivými letišti, respektive tak, aby bylo do všech destinací ze všech počátků stejně daleko a nemusel se tak brát v potaz i rozdíl ve vzdálenostech. Zároveň mě zajímaly velikosti aglomerací / měst – jejich spádové oblasti, ze kterých se odlétá. Vybral jsem si 3 hubová letiště ve Spojeném království – London Heathrow, ve Francii – Paris Charles de Gaulle a v Německu – Munich International. Vzdálenosti na linkách v kilometrech a čase v tabulkách 1.3 a 1.4.

Vzdálenosti [km]			
	CDG	LHR	MUC
BCN	826	1083	1152
IBZ	1006	1405	1041
TFS	3236	3119	3545

Tabulka 1.3: Vzdálenosti v kilometrech z počátečních letišť do destinací

Doba letu [hh:mm]			
	CDG	LHR	MUC
BCN	01:55	02:10	02:10
IBZ	02:15	02:30	02:20
TFS	04:25	04:25	04:55

Tabulka 1.4: Doby trvání letů z počátečních letišť do destinací

Abych bylo možné analyzovat změnu ceny letenky v závislosti na tom, jakého segmentu je zákazník součástí, tak jsem si zvolil 4 časové intervaly, ve kterých jsem následně sbíral data – jeden týden, dva týdny, deset dní a víkendy.

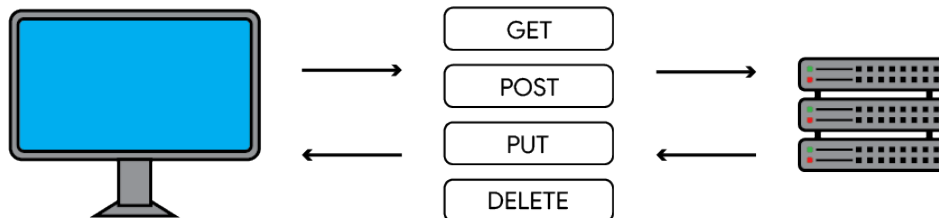
Kapitola 1 se věnovala především rešerši trhu v oblasti ekonomické, provozní a kontextové a definoval jsem pomocí tohoto celou práci z pohledu spojení mezi vybranými letišti.

2 Metodika

V metodice byla porovnávána cena letenek do jednotlivých turistických destinací na určitých časových intervalech. Pro vyřešení tohoto problému bylo potřeba získat kontextuální data, jimiž jsou mimo jiné HDP a délka letu, provozní, které byly považovány jako informace o konkurenci na dané lince / trhu a data ekonomická, což znamenalo pro potřeby práce ceny letenek nabízené jednotlivými společnostmi. K vyhodnocování bylo přistupováno jak pomocí vyhodnocení samotných sesbíraných dat, tak pomocí vytváření regresních modelů ze sesbíraných dat v programovacím jazyce Python.

2.1 Sběr dat

Sbírání dat probíhalo pokročilým způsobem, a to pomocí mnou vytvořených HTTP requestů pomocí metody GET na servery firmy kiwi.com, jmenovitě `api.tequila.kiwi.com`, viz obrázek 2.1. Tyto requesty jsem vytvářel v programu Postman, což je freeware program dostupný na adrese `postman.com`.



Obrázek 2.1: Metody HTTP requestů

Celkem těchto requestů musí být 36 – 3 počátky, 3 destinace a 4 časové intervaly. Každý tento request musel být vyspecifikován příslušnými parametry, kterými bylo například `partner-ID`, jež jsem dostal po registraci na `tequila.kiwi.com`. V parametrech musí být taktéž uvedeno, odkud linka začíná a kde linka končí, na jakou dobu do budoucna se chceme podívat – v jakém časovém intervalu od současného dne chceme dostat nabídky (já jsem si zvolil dle článku [9] 90denní časový interval, abych pokryl možnou variaci ceny letenky).

Dále jsem volil parametry i pro jednotlivé časové intervaly (respektive počet nocí strávených v destinaci). Pro jeden týden v destinaci jsem volil 6-8 nocí, pro dva týdny 13-15 nocí, pro deset dní 9-11 a pro víkendy 2-5. Víkendy byly velmi specifické pro svůj request tím, že jsem musel

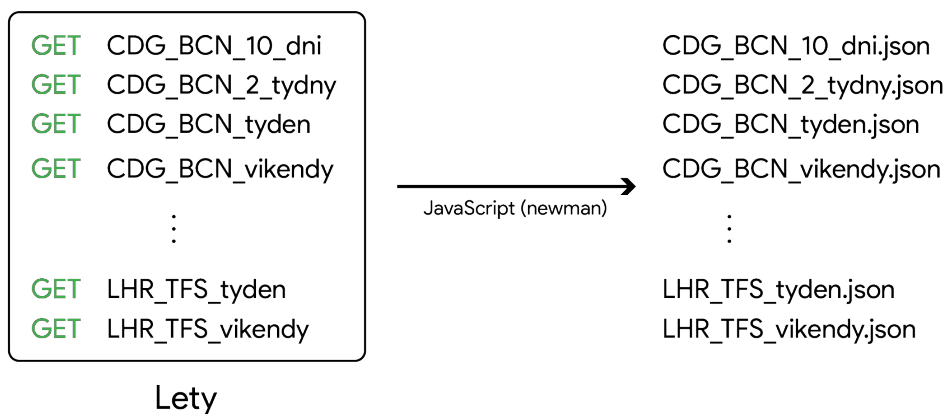


specifikovat i v jakých dnech chceme odlétat – řazení dní je podle amerického formátu – první den týdne je neděle a zároveň je číslováno podle standardního konsenzu v programovacích jazycích jako jsou například C# a Python – začínáme v čísle 0, tudíž jsem zvolil parametr `fly_days` (kdy chci odlétat) v rozmezí 4-5 (čtvrtek a pátek) a parametr `ret_fly_days` (kdy se chci vrátit) v rozmezí 0-2 (neděle až úterý).

Pro potřeby práce jsem taktéž limitoval počet cestujících na jednoho dospělého člověka parametrem `adults` a další limit pro request jsem definoval parametrem `limit` na hodnotu 250, a to z důvodu velkého množství vrácených dat z jednoho requestu – 250 nabídek každý den v časovém intervalu sběru dat od prosince 2022 do června 2023 (celkem jsem nasbíral 2.63 GB dat)

V programu Postman jsem každý z těchto requestů poté uložil do takzvané collection, abych je měl pohromadě a měl s nimi jednodušší manipulaci. Následně jsem si celou collection vyexportoval ve formátu souboru `.json` (JavaScript Object Notation) ve verzi 2.1. Velmi důležité bylo soubory správně pojmenovávat. Součástí jména souboru je den, kdy byl vytvořen, a také počátek linky a destinace a jako poslední ve jméně figuruje časový interval, pro který byl request vytvořen – kolik nocí v destinaci cestující stráví.

Následuje spuštění celé collection, kterou jsem si pojmenoval `Lety`. S využitím modulu `newman`, který slouží na spouštění collection z programu Postman, jsem si v programovacím jazyku JavaScript vytvořil skript k ukládání odpovědí z HTTP GET requestů do souborů `.json` a pojmenoval jej `run_collection_to_JSON.js`. Za pomoci runtime prostředí NodeJS v operačním systému Windows jsem následně tento skript spouštěl každý den pomocí příkazu `node run_collection_to_JSON.js`. Názorné zpracování v obrázku 2.2.

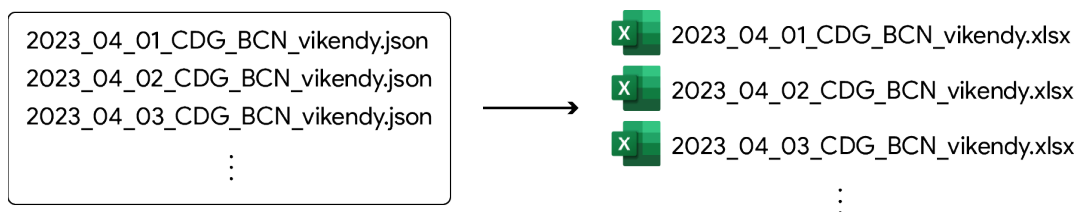


Obrázek 2.2: Vytváření collection Lety a .json souborů

Je taktéž důležité říci, že každý request obsahuje celkovou cenu pro zpáteční letenku, tudíž se na ní mohou podílet oba dopravci zároveň. Je to dáno vlastnostmi odpovědí na requesty ze stránky `api.tequila.kiwi.com`, které neumožňují filtraci dat pouze pro jednu z aerolinek, protože je toto API vytvořené nejspíše pro webové stránky, které tato data taktéž shromažďují, ale již s nimi nepracují a používají je pouze pro potřeby zobrazení výsledků vyhledávání zákazníků hledajících možnosti cestování letadlem. Tudíž, když jsem vzal v potaz tuto skutečnost, tak pracuji pouze s odhadem cen pro daný let – mohou nastat případy, kdy jsou na lince 2 dopravci, ale nemusí, takže v tuto chvíli se jedná pouze o vydělení výsledné ceny dvěma a následnou analýzu provádím pro letenku nabízenou jedním dopravcem.

2.2 Zpracování souborů

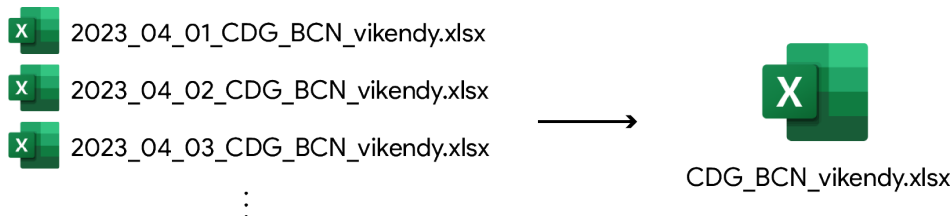
Vzhledem k obsáhlosti získaných dat z `.json` souborů a potřebě tato data vizualizovat bylo třeba je zpracovat. Zvolil jsem programovací jazyk Python kvůli jeho jednoduchosti a uživatelské přívětivosti. V něm jsem napsal skript na zpracování a filtrování dat a následné uložení do `.xlsx` souborů – pro každý `.json` soubor jeden `.xlsx` soubor – kvůli možnosti soubory sloučit jak pomocí programování, tak pomocí vestavěných nástrojů v programu Microsoft Excel. Tyto `.xlsx` soubory jsem pojmenovával stejně jako souboru formátu `.json` z důvodu přehlednosti. Důležitá data pro mne jsou zejména datum, kdy má být uskutečněn let do destinace, číslo letu a aerolinka, která má daný let provést. Způsob zpracování `.json` souborů uveden na obrázku 2.3.



Obrázek 2.3: Vytváření `.xlsx` souborů

K dalším úpravám těchto `.xlsx` souborů jsem zvolil programovací jazyk Python. V něm jsem si vytvořil další skript na slučování souborů a přidávání jejich jména do prvního sloupce, což bude nadále sloužit jako hlavní informace pro výpočet dní do odletu. Další skript provádí zpracování těchto předešle sloučených souborů – sloučí IATA kód aerolinky s číslem letu, které poté uloží do

sloupců `flight_no_1` a `flight_no_2`, a vypočítá počet dní do odletu odečtením dne odletu a dne vytvoření původního `.json` souboru a následně je ukládá do sloupce se jménem `day_to_dep` – data se dny do odletu. Slučování `.xlsx` souborů znázorněno na obrázku 2.4.



Obrázek 2.4: Slučování `.xlsx` souborů

2.3 Zpracování dat

V nejobsáhlejším skriptu, který byl vytvářen taktéž v programovacím jazyce Python, bylo prováděno zpracování dat jako takových. Programování probíhalo v programu Microsoft Visual Studio Code, známém jako VS Code.

2.3.1 Základní zpracování

Na začátku byly naimportovány důležité knihovny jako jsou `pandas`, `glob`, `math`, `numpy`, `statistics` a další. Pomocí knihovny `glob` si skript načte všechny soubory v adresáři, do kterého se předchozími příkazy ukládaly výsledné `.xlsx` soubory, a uloží si je za pomoci knihovny `pandas`. Následně dochází k další filtraci dat – jmenovitě potřebujeme zejména sloupce `day_to_dep`, `price` – data o ceně zpáteční letenky a `airline`, kde jsou uloženy IATA kódy aerolinek poskytující zpáteční letenky.

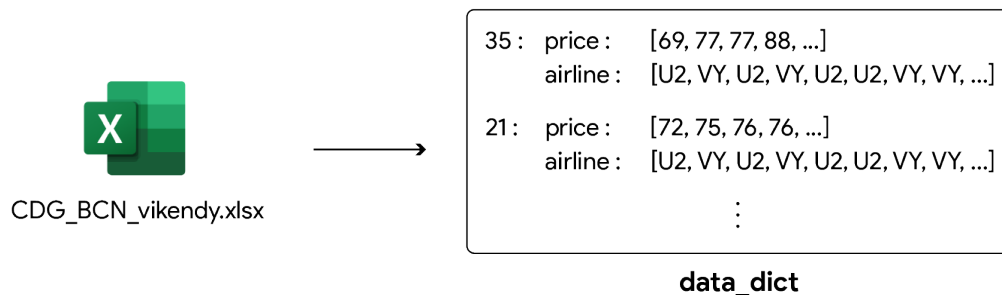
Jednoduchým příkazem si program přečte `.xlsx` soubor, který byl sesumírován dříve, a v dalších řádcích kódu dochází k vytvoření DataFrame `statistics_df` pomocí knihovny `numpy` s prázdnými sloupci pro statistické hodnoty, které ze sesbíraných dat chci získat – jsou jimi například medián, průměr, standardní odchylka a další.

Nejprve byl ale vytvořen slovník – dictionary (`data_dict`) a ten obsazen hodnotami ve formátu `data_dict[day_to_dep] : {'price' : [hodnoty pro cenu], 'airline' : [IATA kódy aerolinek]}` – viz obrázek 2.5. Jedná se v podstatě o slovník slovníků,

kde jsou v nejnižších keys ('price' a 'airline') uloženy listy, kde jsou data, která budeme zpracovávat. Výsledek po přiřazení přečteného prvního řádku .xlsx souboru bude vypadat jeden vstup ve slovníku data_dict kupříkladu takto: 35: {'price': [70], 'airline': ['AF', 'VY']}, avšak po přečtení celého souboru bude v nejnižších keys 'price' a 'airline' seznam cen pro daný den do odletu (v našem případě 35) v následujícím formátu:

```
{35: {'price': [69, 77, 77, 88, ...],  
'airline': ['U2', 'VY', 'U2', 'VY', 'U2', 'U2', 'VY', 'VY', ...]}}
```

Tento slovník data_dict poté uspořádám podle počtu dnů do odletu pro lepší přehlednost, avšak z hlediska funkčnosti to není nutné.



Obrázek 2.5: Tvorba slovníku slovníků

Po vytvoření DataFrame se sloupci pro jednotlivé statistické hodnoty skript pomocí for cyklu prochází každý den do odletu a jednoduchými příkazy z knihovny numpy a statistics počítá statistické hodnoty, přičemž v případě průměru ještě filtruje odlehlá pozorování pomocí standardní odchylky – podle pravidla tří sigma.

Při průchodu slovníku data_dict bylo taktéž spočítáno množství výskytů nabídky aerolinky v každém dni do odletu (bude potřeba při vytváření podílu na trhu) a pro kontrolu jej taktéž ukládám do DataFrame statistics_df. Tento celý tento DataFrame s daty byl následně uložen do nového souboru se jménem, v tomto případě statistics_CDG_BCN.xlsx – jak pro kontrolu, tak pro další zpracování.



2.3.2 Pokročilé zpracování

Pro lepší operaci s daty a další analýzu dat byl naprogramován jiný skript, který iteruje `.xlsx` soubory nejdříve podle letiště odletu, poté podle destinace a nakonec podle časového intervalu, pro který byl `.xlsx` soubor vytvořen. Toto používání for cyklů bylo zvoleno z důvodu zjednodušení psaní kódu pro vytváření a slučování grafů.

Při novém přečtení `.xlsx` souboru byla vytvářena stejná datová struktura jako v případě předešlém (`data_dict`) s tím rozdílem, že byly vytvořeny i 4 seznamy – `file_x_values`, `file_y_values`, `interval_x_values` a `interval_y_values`. Do `file_x_values` a `file_y_values` byly ukládány informace o dnech do odletu a vyfiltrované průměrné ceně a tyto údaje poté připojeny do listů `interval_x_values` a `interval_y_values` a jednoduchým příkazem z knihovny `matplotlib` vytvořeny spojnicové grafy pro jednotlivé časové intervaly na lince – společný jmenovatel je linka a zobrazené jsou průměry všech aerolinií.

Tyto grafy jsou ale bohužel příliš obecné, a tak byly vytvořeny grafy specifitější pomocí dalších skriptů. V jednom z nich byl zvolen jako společný jmenovatel aerolinky na časovém intervalu a v dalším časový interval pro aerolinky. V těchto skriptech se jedná již o složitější časovou náročnost ve skriptu – celkem se jedná o 4 for cykly v obou případech – $O(n^4)$.

V případě společného jmenovatele časového intervalu pro aerolinky byly procházeny sloupce s hodnotami průměrů cen letenek pro jednotlivé dny do odletu. Následně byla tato data testována na normalitu a na základě výsledků testu na normalitu vybrán druh korelačního testu – Spearmanův nebo Pearsonův – pro data pocházející z normálního rozdělení test Spearmanův a pro data rozdělené nenormálně test Pearsonův. V programovacím jazyce Python tyto testy nejsou nativně k nalezení, proto bylo potřeba sáhnout do knihovny `sklearn` a využívat příkazy z ní.

Na základě výsledků testů na normalitu a korelaci byl poté zvolen typ regrese. Pro lineární regrese musí platit, že data pochází z normálního rozdělení a zároveň p-hodnota testu na korelaci musí být menší než hladina významnosti, kterou jsem si pro všechny statistické testy zvolil na hodnotu 0,05. Pokud data nepochází z normálního rozdělení, ale p-hodnota testu na korelaci byla menší než hladina významnosti (0,05), tak se bude jednat o nelineární regresi, která byla vybrána na základě konzultace s paní doktorkou Pecherkovou jako polynomiální druhého řádu.

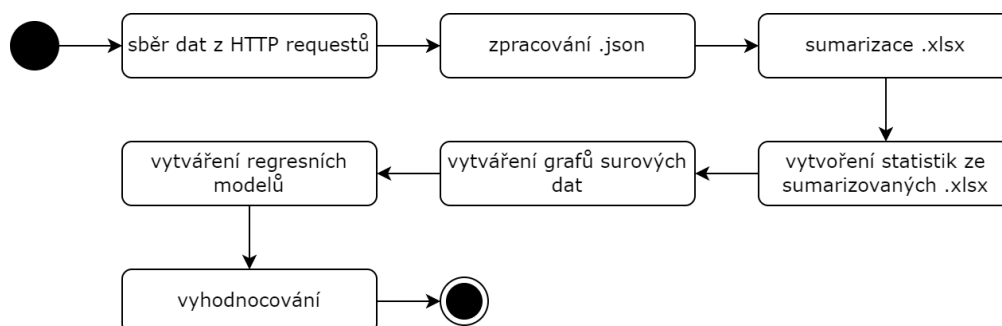
Důležité je taktéž otestovat, zdali byla regrese zvolena správně, tudíž po obou typech regrese následovaly validační testy. V případě lineární regrese se jedná o f-test podílu vysvětleného



a nevysvětleného rozptylu a v případě polynomiální regrese o Durbin-Watsonův test nezávislosti reziduí / autokorelace reziduí. Pro f-test je jednoduché zamítnout nulovou hypotézu, protože se opět používá p-hodnota a zamítáme nulovou hypotézu (zvolená regrese je nevhodná), pokud je p-hodnota menší než hladina významnosti, ale kdybychom měli ověřovat správnost polynomiální regrese, tak Durbin-Watsonův (D-W) test neprodukuje p-hodnotu a musíme se zaměřit na statistiku jinou – výsledek D-W testu bude mít hodnotu vždy na uzavřeném intervalu od 0 do 4, a pokud bychom měli zamítnout nulovou hypotézu (zvolená hypotéza byla zvolena správně – neexistuje autokorelace mezi reziduy), tak výsledek D-W testu musí být zásadně odlišný od čísla 2. Nulovou hypotézu v tomto případě avšak zamítnout nechceme, protože by byla regrese tím pádem zvolená špatně a v tomto má D-W test nevýhodu, protože je pro nás statisticky lepší, když nulovou hypotézu zamítneme a jsme si na 100 % jisti, že je hypotéza vyvrácena a ne že vyvrácena není. Celé zpracování poté už obnáší jen vykreslení jednotlivých regresí do grafu a drobné kosmetické úpravy ve smyslu upravení měřítka, zobrazení mřížky, legendy a správných barev pro jednoduchou orientaci v grafu.

Při zpracovávání dat se společným jmenovatelem aerolinky na časových intervalech dochází v podstatě ke stejné analýze jako v předchozím případě – otestování na normalitu, vytvoření regresního modelu a následně validace regrese ať už jedním nebo druhým testem a vykreslení do grafu a uložení. Zbytek dat je sbírán manuálně a grafy vykreslen buď pomocí nových, ale jednoduchých nebo již použitých částí skriptů.

Celkový postup v práci je uveden ve vývojovém diagramu 2.6. Jedná se o diagram vytvořený podle formátu UML 2.5, což je předposlední verze celosvětově používaného standardu mezi modelovacími jazyky.



Obrázek 2.6: Vývojový diagram postupu v práci



3 Prezence výsledků

Po vykreslení všech potřebných grafů a analýz byla zvolena jako dobrý příklad linku mezi letišti CDG a BCN. Důvodem byla grafická přehlednost a názornost vzniklých grafů. Muselo být zvoleno nestandardní číslování os, kde jsou uvedeny vlevo dole nejmenší hodnoty pro osu y, ale pro osu x to platí naopak, protože chceme pro lepší přehlednost spíše chronologické pořadí než matematické.

3.1 Obecná data

Základní pohled na sesbíraná data a prvotní analýzu mi umožnily hodnoty o průměrných cenách na linkách z CDG, LHR a MUC v tabulkách 3.1, 3.2 a 3.3. Z nich je patrné, že průměrné nejvyšší ceny na všech dnech do odletu jsou nastaveny z LHR, a to do všech destinací ve všech časových intervalech.

	CDG-BCN	CDG-IBZ	CDG-TFS
týden	128.51	344.52	431.81
10 dní	133.08	360.37	421.06
2 týdny	130.57	383.10	436.47
víkendy	149.15	419.78	402.86

Tabulka 3.1: Průměrné ceny (€) zpátečních letenek z CDG

	LHR-BCN	LHR-IBZ	LHR-TFS
týden	218.26	347.10	675.76
10 dní	248.34	335.97	690.56
2 týdny	228.85	374.40	710.85
víkendy	349.10	341.00	734.69

Tabulka 3.2: Průměrné ceny (€) zpátečních letenek z LHR



	MUC-BCN	MUC-IBZ	MUC-TFS
týden	192.01	388.57	554.36
10 dní	190.59	347.28	601.72
2 týdny	191.71	403.05	553.48
víkendy	213.21	398.97	534.16

Tabulka 3.3: Průměrné ceny (€) zpátečních letenek z MUC

3.2 Podíl aerolinek na trhu

Podíl aerolinií na trhu jednotlivých letištních párů v tabulkách 3.4 a 3.5 byl vypočten jako průměr podílu na linkách po jednotlivých časových intervalech. Data o zastoupení aerolinky na trhu byla získána HTTP requesty – dochází ke dvojitmu použití dat – jak pro cenu, tak pro počty nabízených letenek na trhu. Pro potřeby další analýzy jsem uvedl jen dva příklady. V příloze lze najít zbylé grafy s podílem na příslušných trzích.

	LHR-IBZ
BA	95,3 %
IB	4,7 %

Tabulka 3.4: Podíl na trhu, linka LHR-IBZ

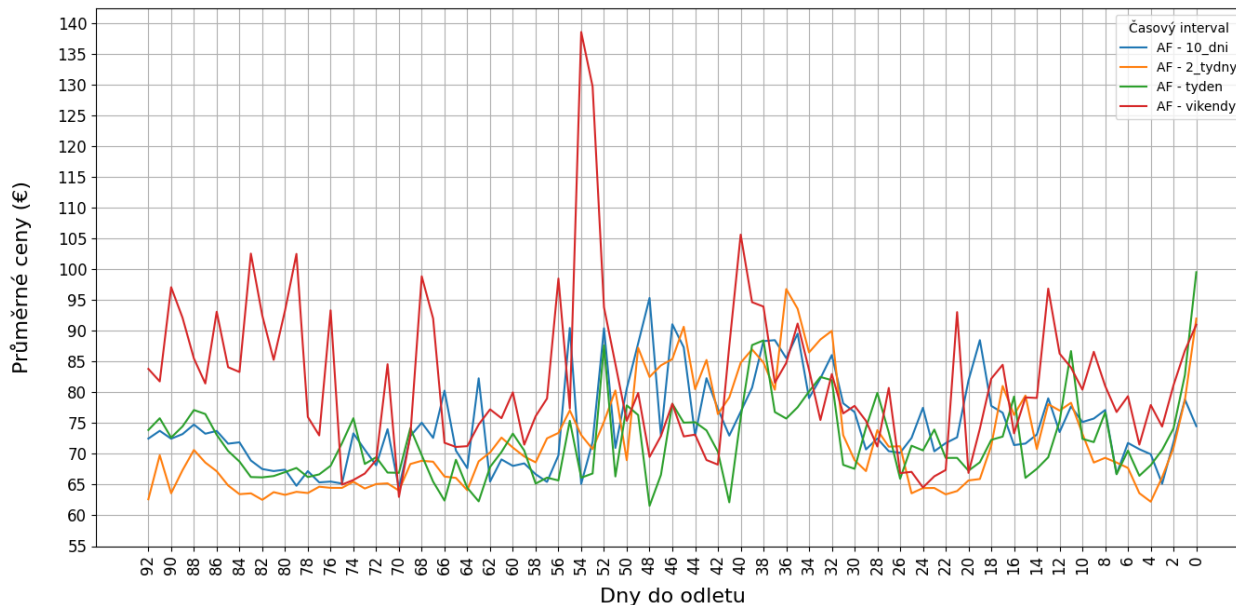
	LHR-TFS
BA	100 %
IB	0 %

Tabulka 3.5: Podíl na trhu, linka LHR-TFS

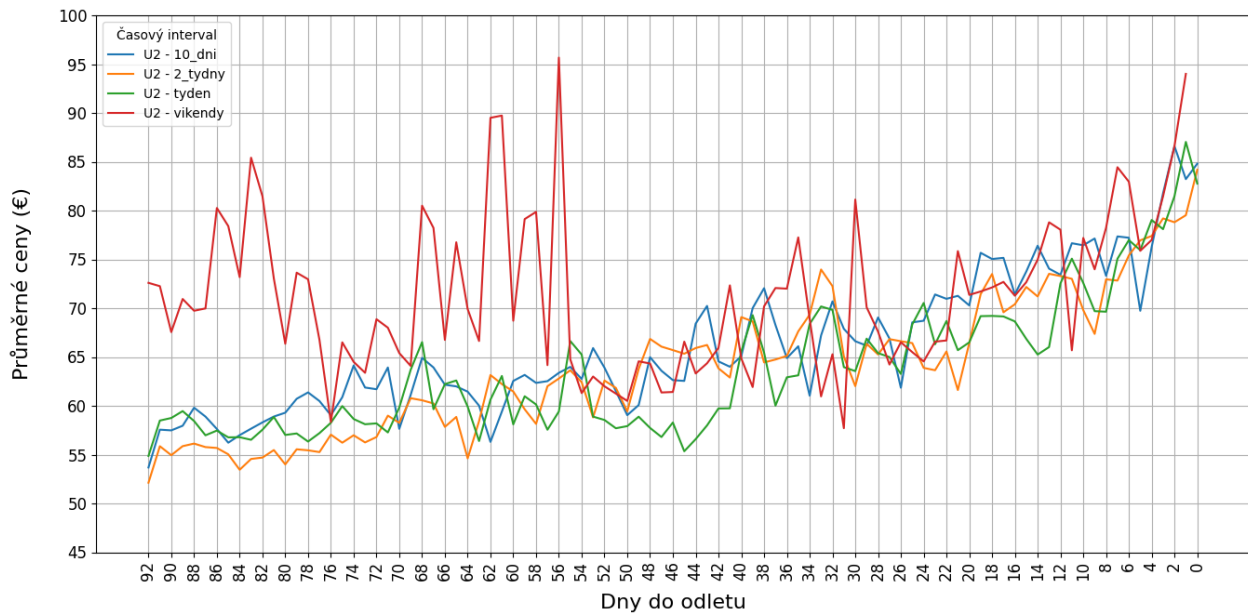
3.3 Spojnicové grafy surových dat

Na obrázcích 3.1, 3.2 a 3.3 je vidět, že všichni tři dopravci mají velmi odlišnou politiku nastavování ceny letenek pro turisty, kteří letí do destinace pouze na víkend. Taktéž je zřejmé, že v okolí 55. dne do odletu aerolinky naceňují na částku výrazně vyšší, než když je delší časové období do odletu. Z grafů 3.2 a 3.3 je patrné, že cena na linkách v segmentu dvou týdnů, týdne a deseti dnů stoupá s ubývajícími dny do odletu. V případě grafu 3.1 můžeme také hovořit o nárůstu, ale není tak dramatický jako v předchozích případech. Zároveň vrcholky ve všech třech grafech v okolí padesátého dne do odletu budou znemožňovat vytváření regresních modelů z důvodu jejich přílišné

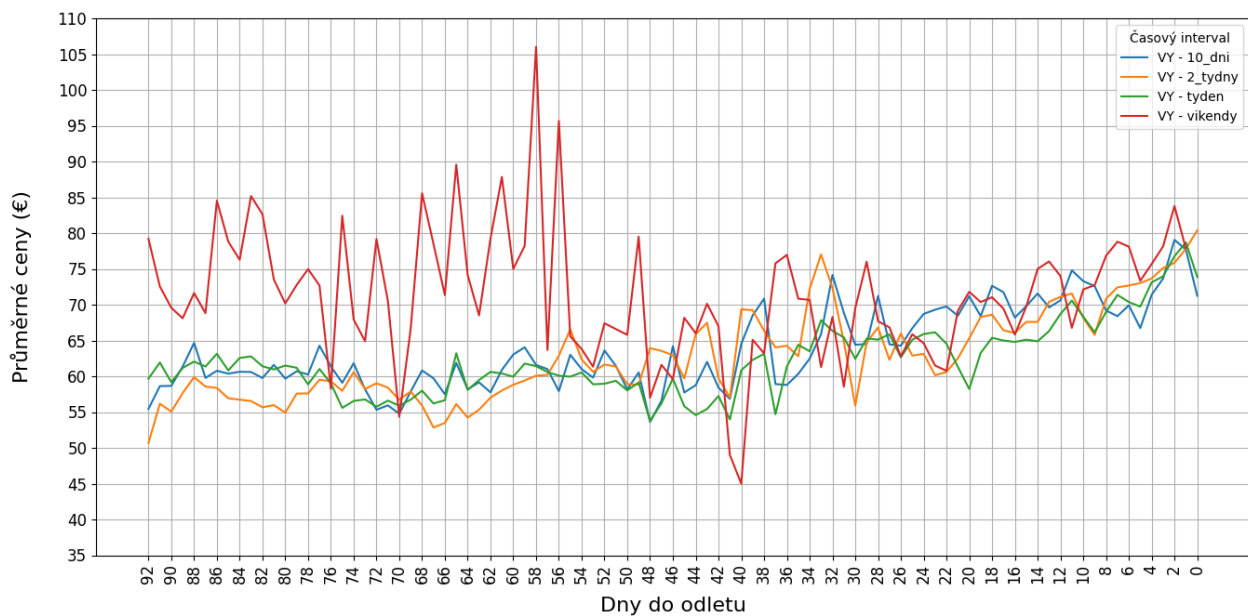
odlišnosti od zbytku dat pro víkendový segment. Nejedná se pouze o odlehlá pozorování, protože mediány u dopravců U2 a VY v 52. a 56. dni do odletu vychází téměř 80 €. Tento argument podporuje i tabulka 3.6, ve které jsou uvedeny směrodatné odchylky pro jednotlivé segmenty. Je zjevné, že největší směrodatné odchylky jsou právě pro segment víkendy. Můžeme si taktéž všimnout, jak odlišně naceňuje "legacy" aerolinka AF oproti nízkonákladovým dopravcům, jako jsou v tomto případě U2 a VY. Cena AF nepatrně stoupá v průměru, ale v případě U2 a VY se jedná o strmý nárůst s ubývajícimi dny do odletu. V případě AF je průměrná cena letenky taktéž zpočátku vyšší než u U2 a VY, nicméně U2 a VY poté AF svým naceňováním "doženou" a během posledního týdne do odletu se jedná o cenu téměř shodnou.



Obrázek 3.1: Vývoj ceny letenek na lince CDG-BCN, aerolinka: AF



Obrázek 3.2: Vývoj ceny letenek na lince CDG-BCN, aerolinka: U2



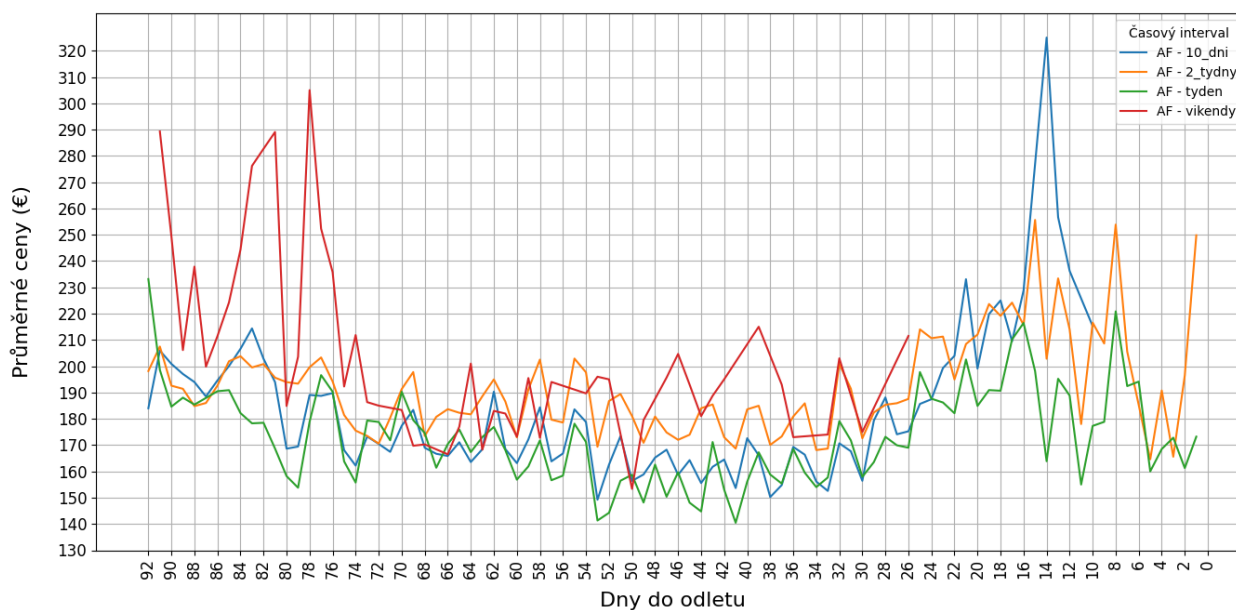
Obrázek 3.3: Vývoj ceny letenek na lince CDG-BCN, aerolinka: VY



	týden	10 dní	2 týdny	víkendy
AF	6.50	7.12	8.53	12.55
U2	6.87	7.03	7.11	7.94
VY	5.09	5.71	6.35	9.14

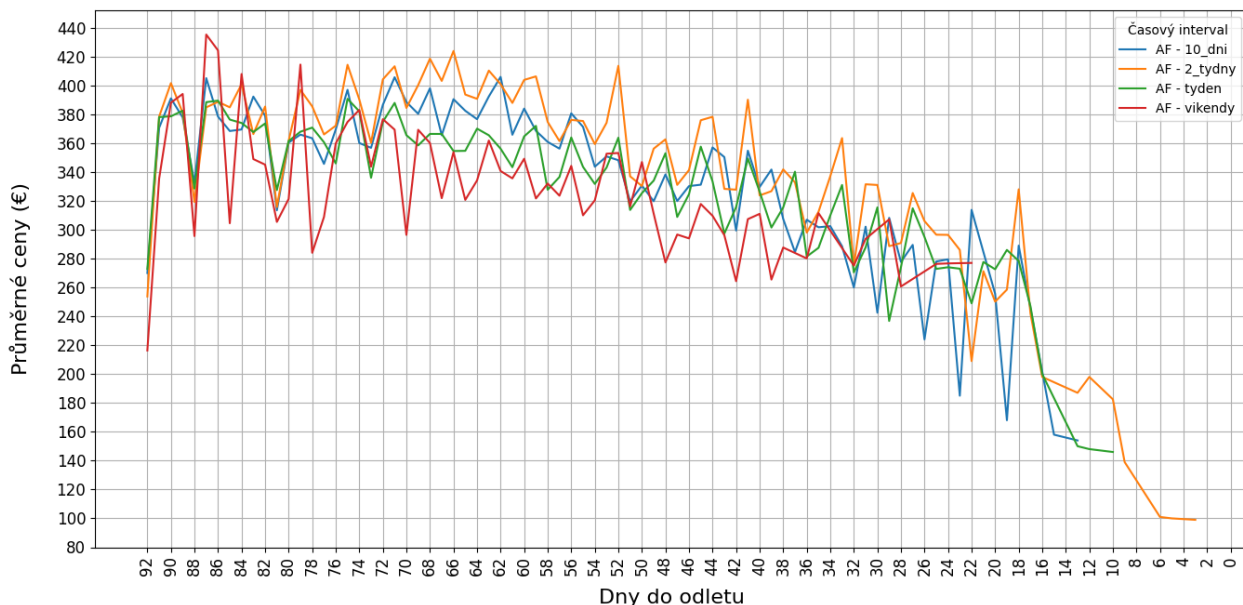
Tabulka 3.6: Směrodatné odchylky na trati CDG-BCN pro jednotlivé aerolinky

Dalším zajímavým porovnáním s linkou CDG-BCN jsou linky CDG-IBZ a CDG-TFS. Na lince CDG-IBZ je Air France jedinou aerolinkou nabízející spojení mezi těmito místy, tudíž průběh ceny v závislosti na počtu dnů do odletu je odlišný od konkurenčního trhu, na kterém bojuje o zákazníky s ostatními aerolinkami, viz graf 3.4.

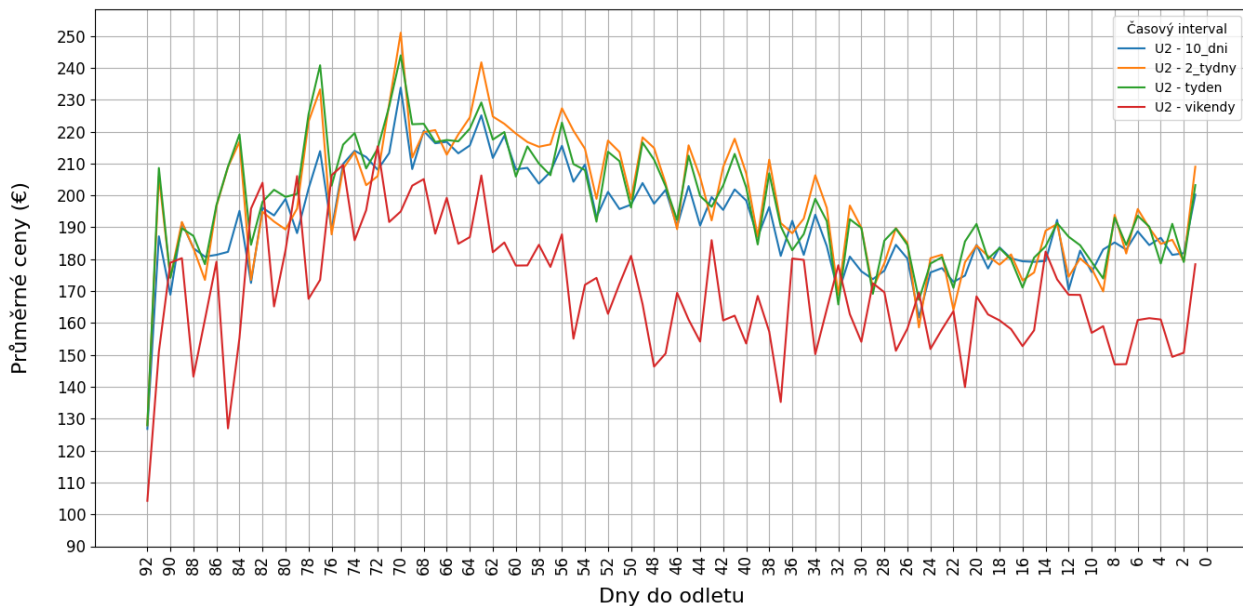


Obrázek 3.4: Vývoj ceny letenek na lince CDG-IBZ, aerolinka: AF

Na lince CDG-TFS (obrázky 3.5 a 3.6) jsou pak aerolinky již dvě (narozdíl od CDG-IBZ), a to AF společně s U2. Průběh ceny je ale opět odlišný od linky CDG-BCN, na které byli dopravci ve větším počtu a zastoupení oběma skupinami – legacy a low-cost.



Obrázek 3.5: Vývoj ceny letenek na lince CDG-TFS, aerolinka: AF



Obrázek 3.6: Vývoj ceny letenek na lince CDG-TFS, aerolinka: U2

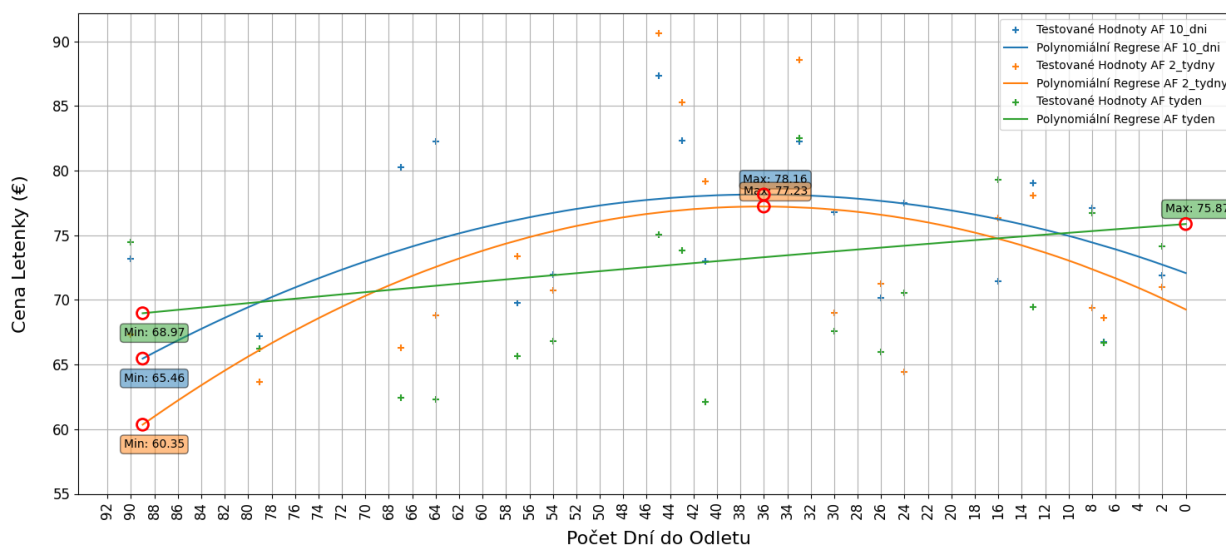


3.4 Regrese

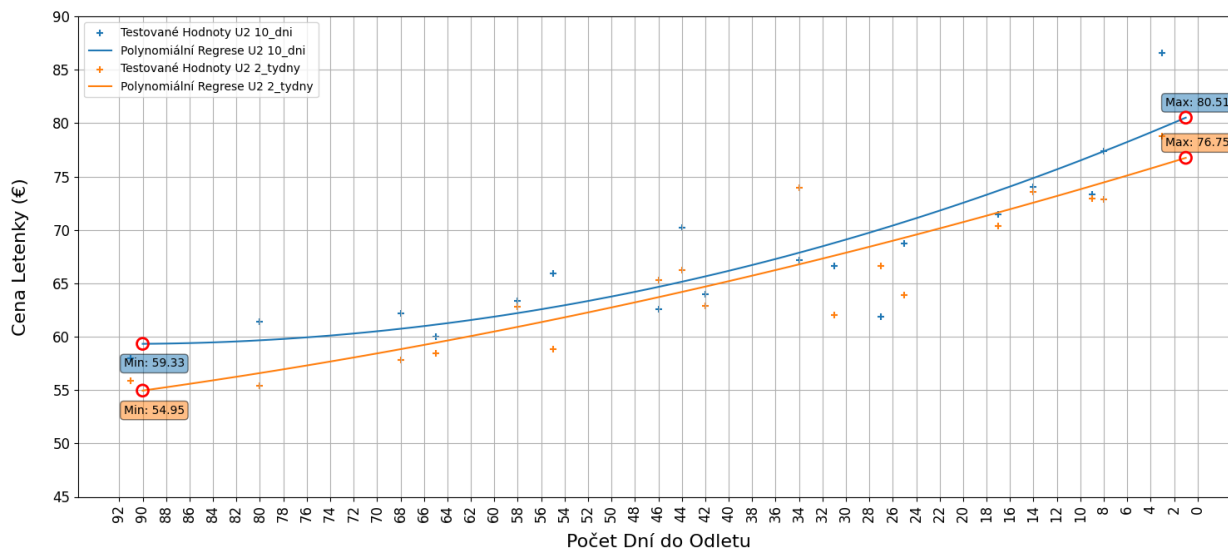
Na základě sesbíraných dat jsem v některých případech vytvořil i regresní křivky pro cenu letenky v závislosti na dnech do odletu. Jedná se o odhad vývoje ceny při současných podmínkách. Regresní modely jsem vytvářel dva – lineární a polynomiální druhého řádu.

3.4.1 Regrese sdružené podle aerolinky

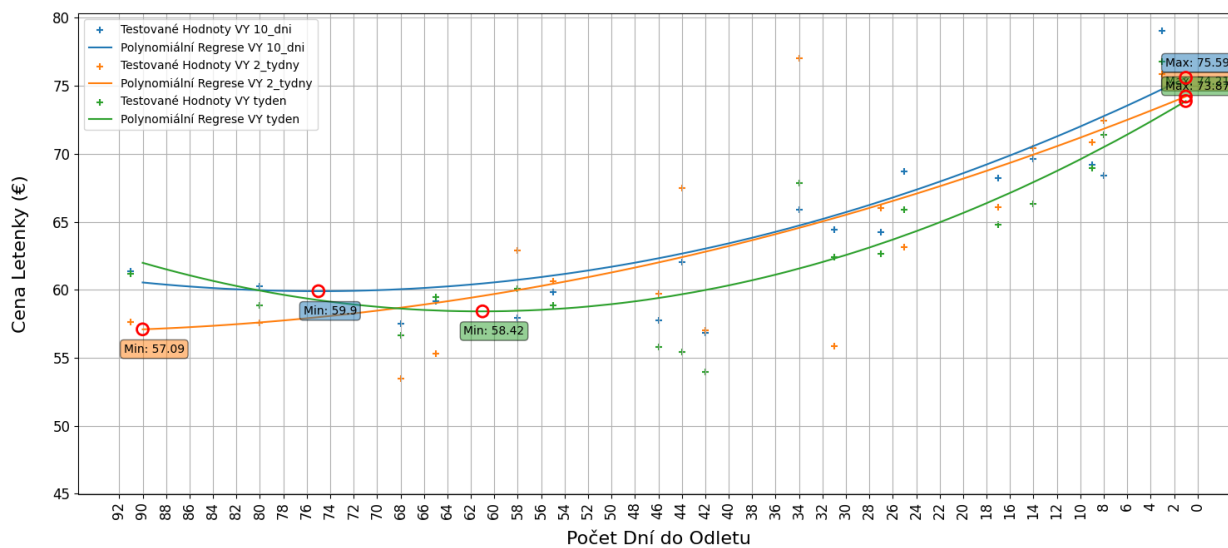
Z grafů 3.7, 3.8 a 3.9 je patrné, že ne vždy se podařilo vytvořit všechny regrese – víkendy jsou problematické zejména tím agresivním naceňováním v okolí 55. dne do odletu.



Obrázek 3.7: Regresní modely na lince CDG-BCN, aerolinka: AF



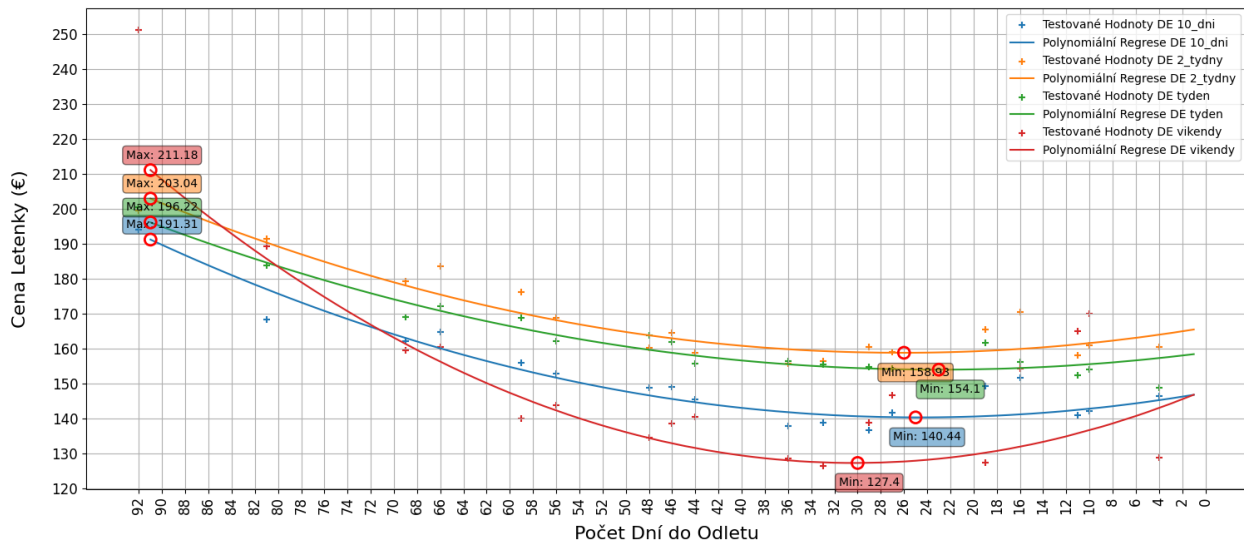
Obrázek 3.8: Regresní modely na lince CDG-BCN, aerolinka: U2



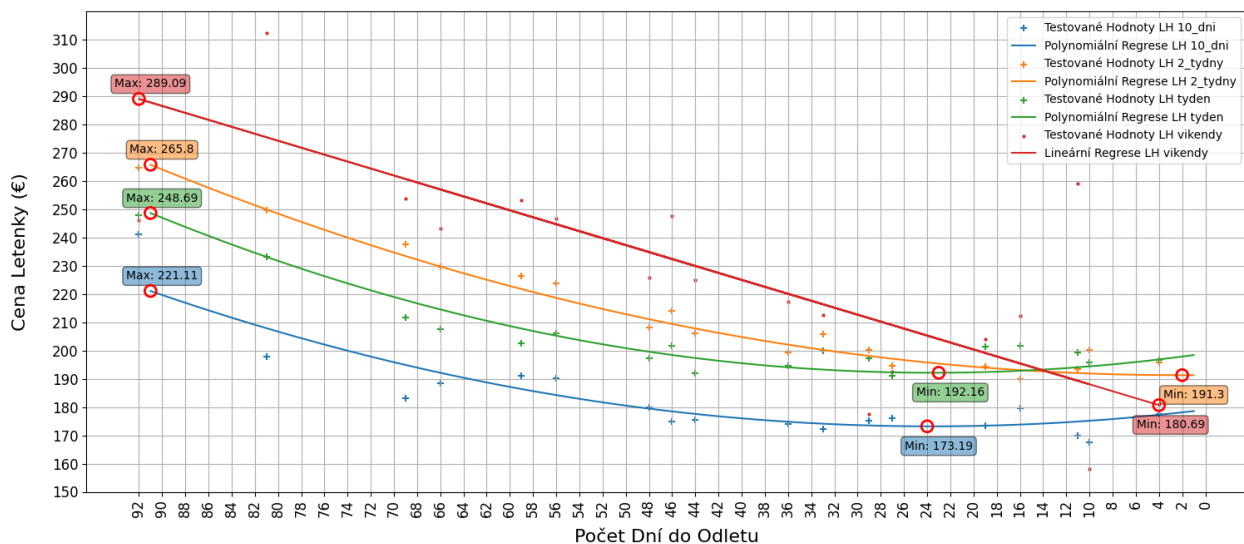
Obrázek 3.9: Regresní modely na lince CDG-BCN, aerolinka: VY

Taktéž je zřejmé, že pro AF je týdenní pobyt z hlediska naceňování ne tak dynamický jako ostatní délky pobytu. U2 a VY mají v naceňování odhadovaný skoro stejný průběh s drobnými odlišnostmi v cenách a v případě VY je zde rozdíl, že nevidíme minimum ve dvou měsících do odletu, ale již na začátku sledování – 90 dní.

Opačný průběh průměrné ceny letenek nastává v případě linky MUC-IBZ (obrázky 3.10 a 3.11), a to jak u nízkonákladové aerolinky DE, tak u "legacy" aerolinky LH. Zde se dokonce podařilo vytvořit regrese pro všechny čtyři časové intervaly.



Obrázek 3.10: Regresní modely na lince MUC-IBZ, aerolinka: DE

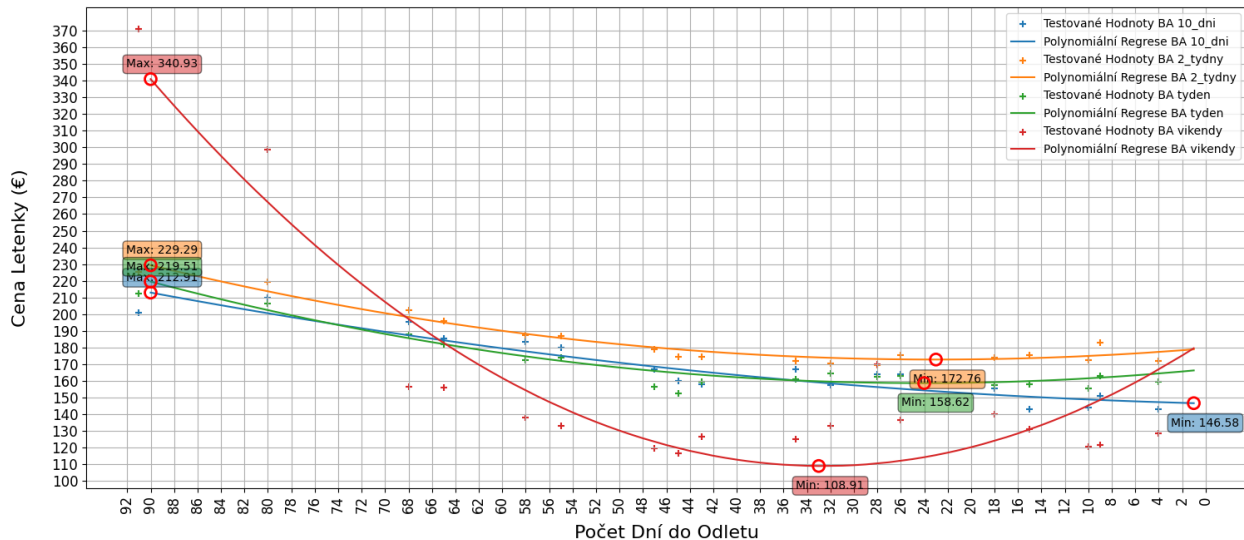


Obrázek 3.11: Regresní modely na lince MUC-IBZ, aerolinka: LH

Je zřejmé, že odhadovaný průběh pro cenu letenky je v případě dvou týdnů, týdne a 10 dní v destinaci téměř shodný s rozdílem výše ceny letenky, která je u nízkonákladového dopravce DE

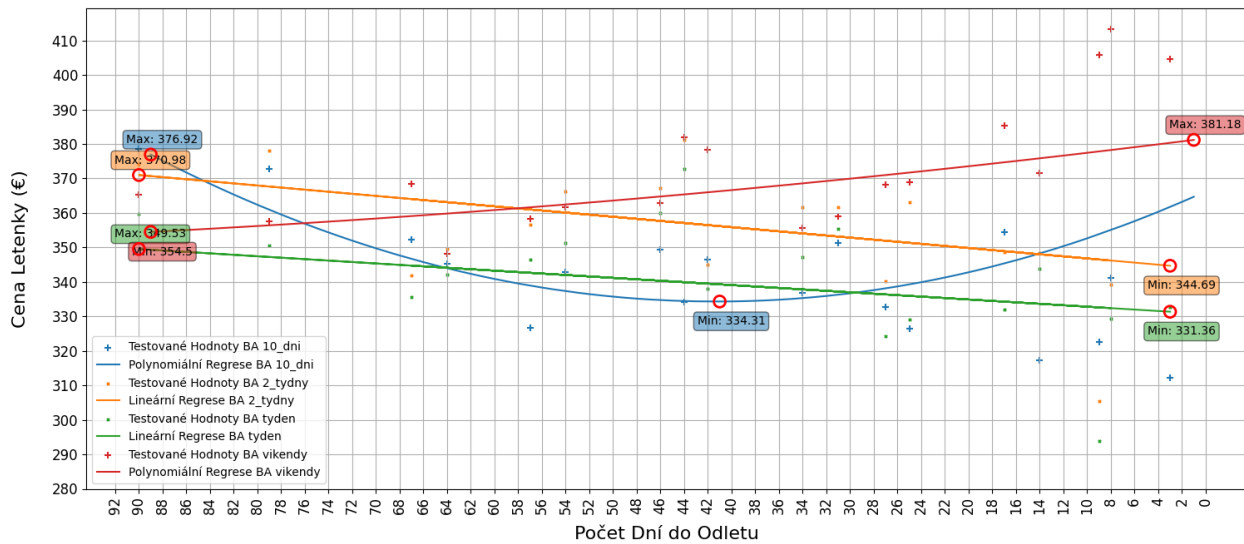
nižší, avšak rozdíl nastává v průběhu odhadované ceny pro pobyt v destinaci na jeden víkend, kdy regresní křivka pro LH následuje lineární trend s minimem v posledním dni, kdy byly letenky nabízeny (4 dny do odletu), zatímco u DE se rozhodli nechat nastavený polynomiální trend s minimem v blízkosti třicátého dne do odletu.

Na lince LHR-IBZ (obrázek 3.12) pro aerolinku BA lze pozorovat různé průběhy regresí pro časové intervaly týden, 2 týdny a 10 dní, u nichž se chová cena velmi podobně s malými odlišnostmi v ceně až do 32. dne do odletu, ale v případě víkendů se jedná o enormní rozdíl jak v ceně, tak v průběhu. Minimum bylo zaznamenáno na výši necelých 109 € ve 33. dni do odletu a maximum na začátku regrese, což znamená 90. den do odletu ve výši necelých 341 €.



Obrázek 3.12: Regresní modely na lince LHR-IBZ, aerolinka: BA

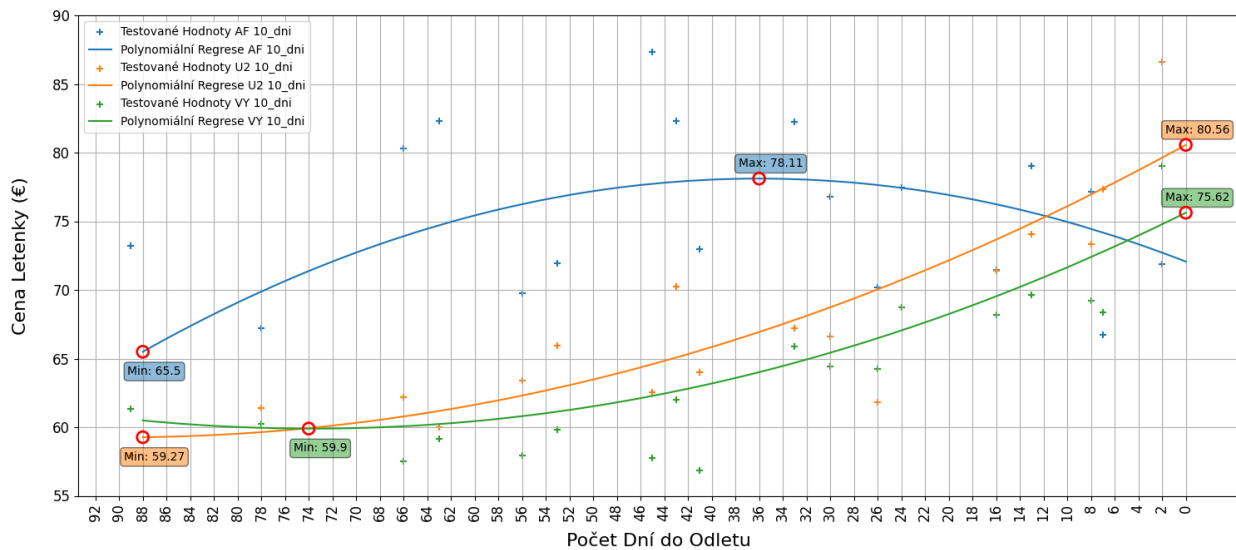
Takto markantní rozdíly v ceně jako v grafu výše avšak nenastávají v přístupu BA k naceňování letenek na lince LHR-TFS (obrázek 3.13) – pro jeden a dva týdny v destinaci se jedná o lineární regrese s tím, že letenka pro jeden týden vyjde levněji. V případě víkendu jde o polynomiální regresi se stoupajícím trendem – minimum na začátku regresní křivky (necelých 355 €) a maximum na konci – 381.2 €. Vymykající se všem těmto trendům je letenka pro 10 dní v destinaci, u níž je průběh konvexní s minimem ve 41. dni do odletu ve výši necelých 335 €.



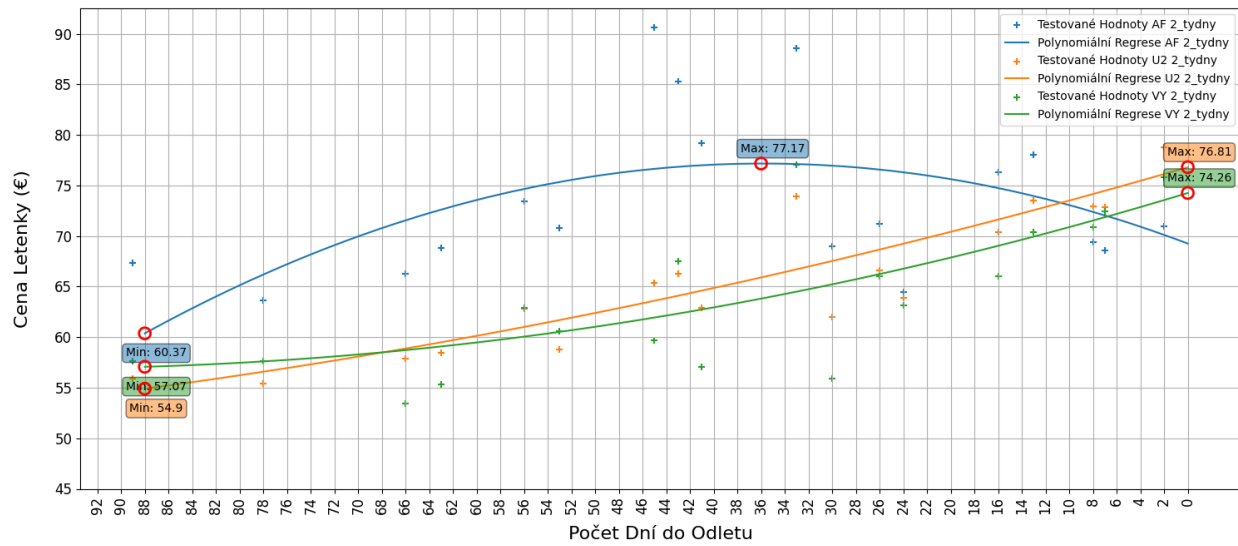
Obrázek 3.13: Regresní modely na lince LHR-TFS, aerolinka: BA

3.4.2 Regrese sdružené podle času

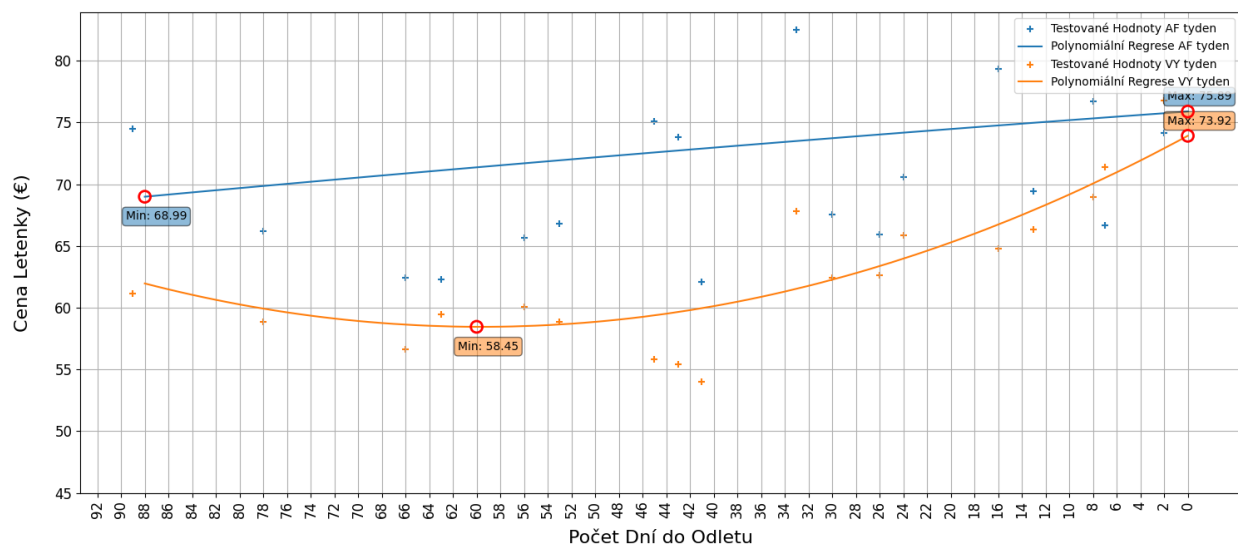
Na grafech 3.14, 3.15 a 3.16 lze spatřit regrese linku CDG-BCN sdružené podle časových intervalů a průběhy regresních křivek pro jednotlivé dopravce a rozdíl v odhadovaném naceňování podle typu aerolinky. Bohužel se mi nepodařilo vytvořit regrese jak pro víkendy od všech dopravců, tak pro týden pro U2.



Obrázek 3.14: Regresní modely na lince CDG-BCN, časový interval: 10 dní



Obrázek 3.15: Regresní modely na lince CDG-BCN, časový interval: 2 týdny



Obrázek 3.16: Regresní modely na lince CDG-BCN, časový interval: týden

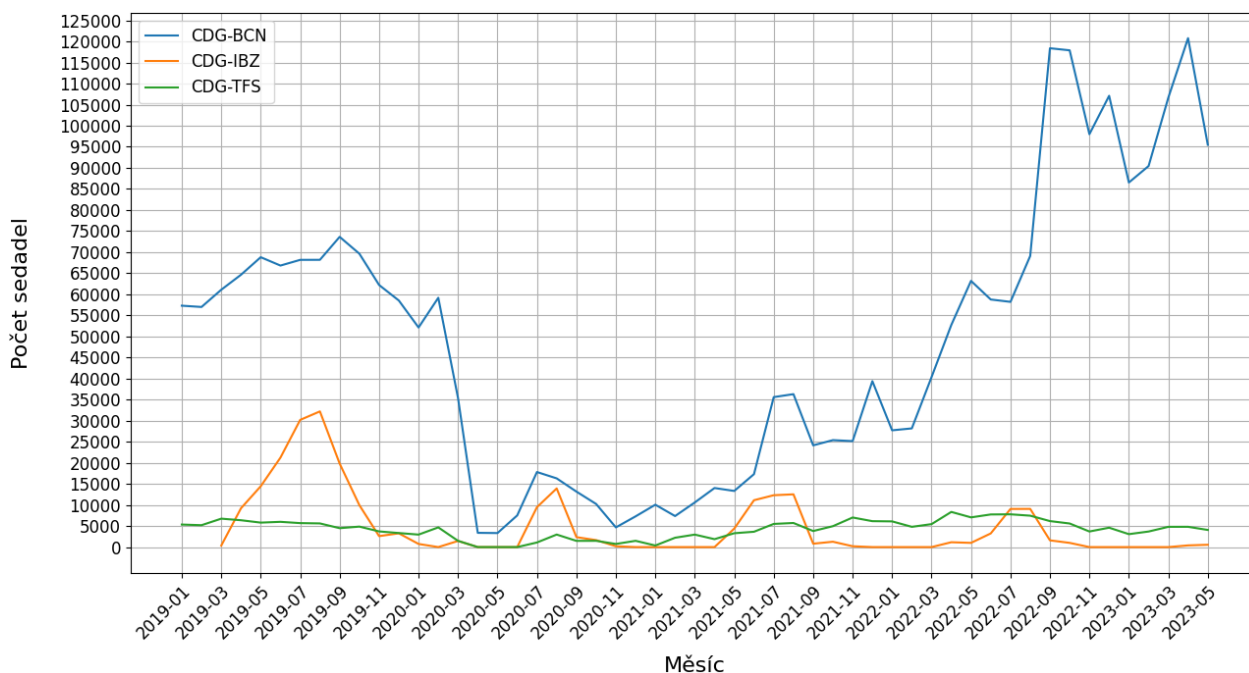


3.5 Ostatní data

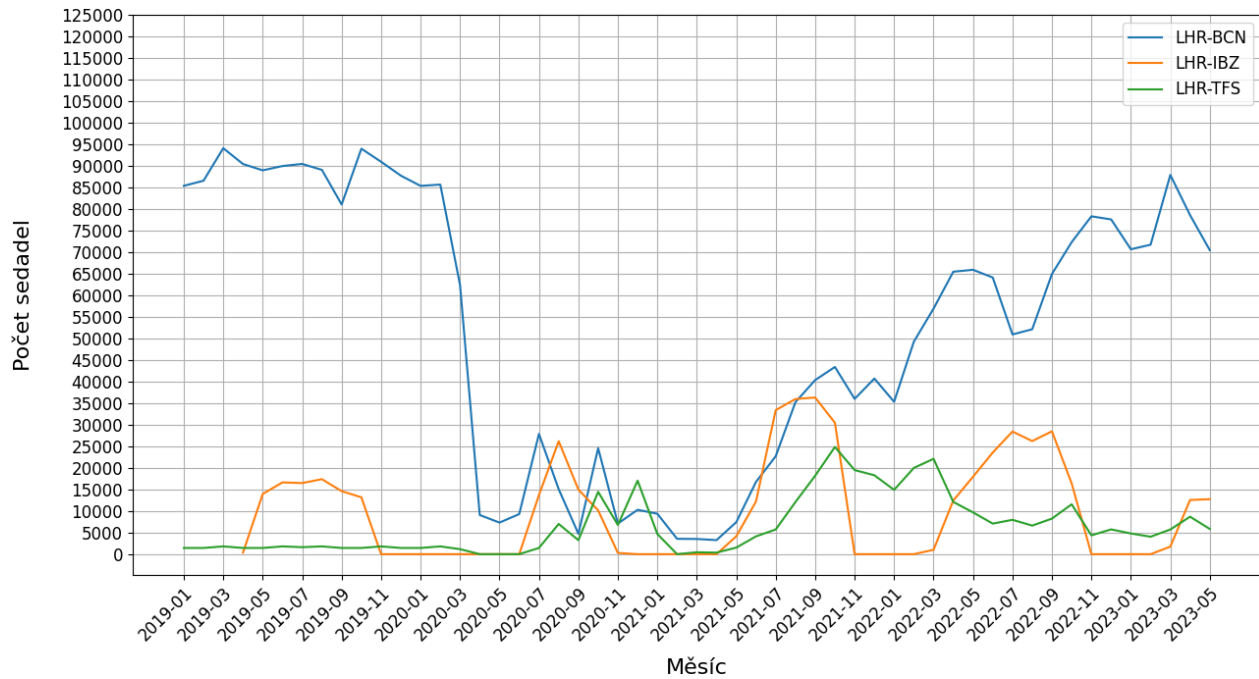
V této kapitole spojuji data z ostatních zdrojů, ať už veřejně dostupných (CAA.co.uk a Eurostat) nebo soukromě získaných pro účely této práce (Flightera.net). Zástupce firmy Flightera.net mi poskytl data o množství vypravených sedadel na linkách z CDG, MUC a LHR. Jedná se o dataset sahající až do roku 2017, avšak mně postačí data od roku 2019 pro nejlepší vykreslení situace.

3.5.1 Počet vypravených sedadel

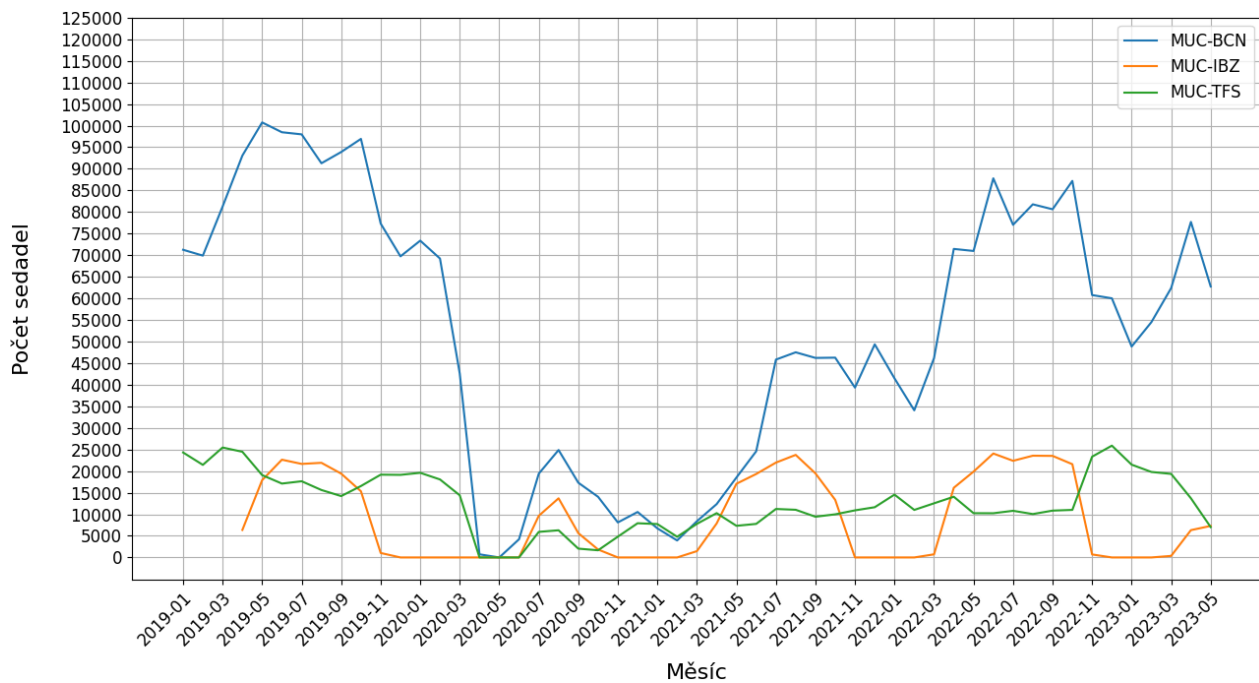
Zde se jedná o data z Flightera.net a z grafů 3.17, 3.18 a 3.19 je patrné, že linky do BCN jsou nejpopulárnější a že linka na IBZ je velmi sezónní – zimní měsíce se na IBZ ani z jednoho letiště nelétá, což vysvětluje problémy se sestavením regresí na linkách na IBZ.



Obrázek 3.17: Počet vypravených sedadel z CDG



Obrázek 3.18: Počet vypravených sedadel z LHR

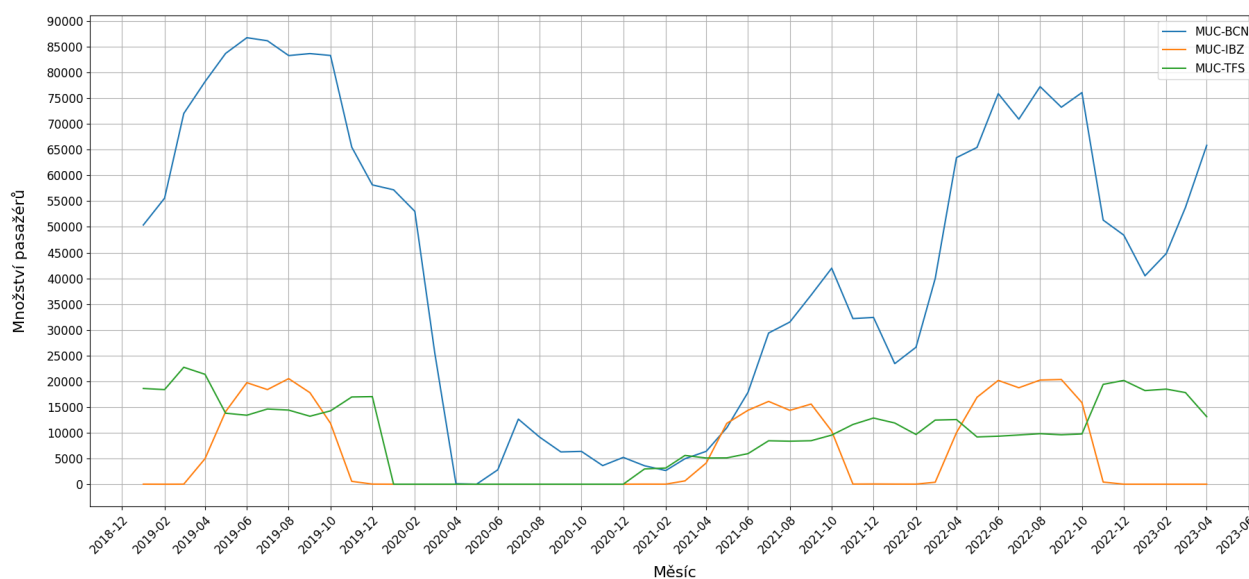


Obrázek 3.19: Počet vypravených sedadel z MUC



3.5.2 Data z Eurostat a CAA.co.uk

Nejlépe reprezentovaná data ze serveru Eurostat jsou na linkách z MUC – viz obrázek 3.20. Pro linky z LHR jsou data z důvodu rozhodnutí Britů v roce 2020 pro nesetrvání v Evropské unii neaktuální a pro linky z CDG z méně neznámého důvodu data neexistují pro linku na Tenerife a pro linku na Ibizu existují data pouze do konce roku 2019, poté se množství vykázaných přepravených cestujících ustálilo na nule, což velmi pravděpodobně není pravda, alespoň nějaký load factor od cestujících tam bude téměř vždy, vzhledem k nabídce na kiwi.com.



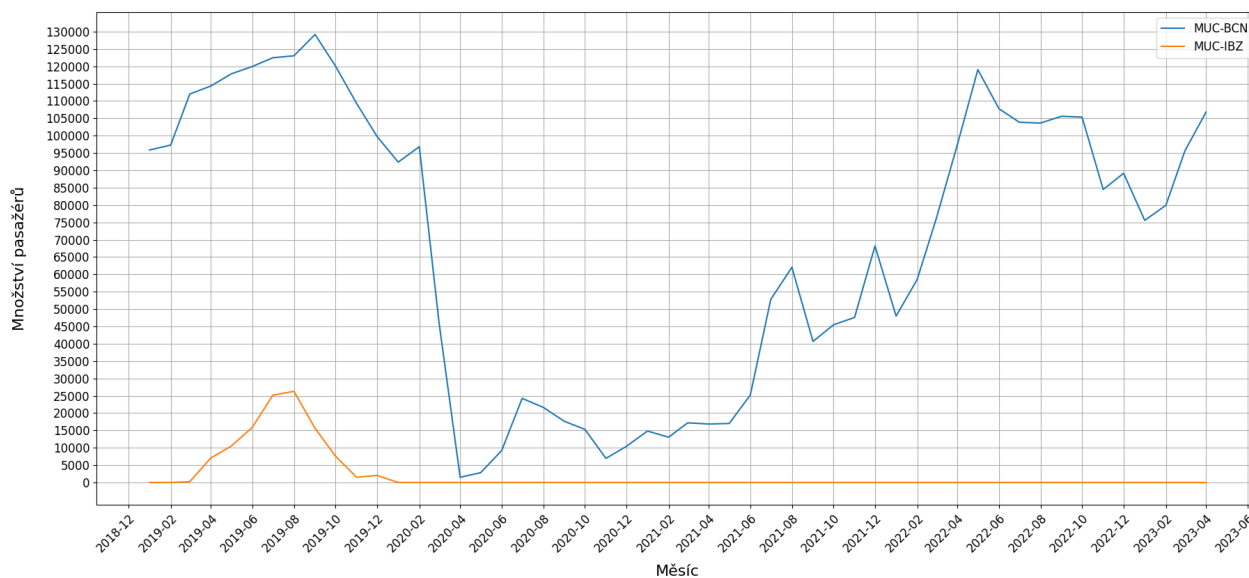
Obrázek 3.20: Množství přepravených cestujících z MUC

Statistiky pro linky z MUC ve spojení z daty z Flightera.net uvedeny v tabulce 3.7.

	MUC-BCN	MUC-IBZ	MUC-TFS
vypravených sedadel	2.780.431	447.793	670.635
přepravených cestujících	2.216.218	318.329	496.726
load factor	79,71%	71,09%	74,07%

Tabulka 3.7: Statistiky pro linky z MUC

Data pro linky z CDG jsou uvedena v grafu 3.21.



Obrázek 3.21: Množství přepravených cestujících z CDG

Vzhledem k tomu, že neexistují na Eurostatu data k lince CDG-TFS a k lince CDG-IBZ existují data pouze pro rok 2019 a méně, tak jediný smysluplný loadfactor ve spojení dat z Eurostatu a Flightera.net je pro linku z CDG uveden v tabulce 3.8.

	CDG-BCN
vypravených sedadel	3.583.872
přepravených cestujících	2.604.226
load factor	72.67%

Tabulka 3.8: Statistiky pro linku z CDG



V tabulce 3.9 je průměrný load factor na odletech a příletech do a z pevninského a ostrovního (Kanárské ostrovy) Španělska z letiště LHR.

Rok	Kvartál	Destinace	Více než 0 a méně méně než 10 %	Více nebo rovno 10 a méně než 50 %	Více nebo rovno 50 %	Celkem letů
2023	Q1	Pevnina	0	10.5	89.5	1954
2023	Q1	CI	0	1.5	98.5	39
2022	Q4	Pevnina	0	10.5	89.5	2092
2022	Q4	CI	0	8	92	57
2022	Q3	Pevnina	0	5	95	2297
2022	Q3	CI	0	2.5	97.5	60
2022	Q2	Pevnina	0	5.5	94.5	2337
2022	Q2	CI	0	0.5	99.5	85
2022	Q1	Pevnina	0	16.5	93.5	1411
2022	Q1	CI	0	16.5	93.5	257
2021	Q4	Pevnina	0.5	17	82.5	1561
2021	Q4	CI	0	15	85	298
2021	Q3	Pevnina	0	20.5	79.5	1857
2021	Q3	CI	0	26.5	73.5	154
2021	Q2	Pevnina	1	59.5	39.5	480
2021	Q2	CI	5.5	89	5.5	19
2021	Q1	Pevnina	2.5	46.5	51	263
2021	Q1	CI	55.5	42.5	2	27

Tabulka 3.9: Load factor LHR [24]

V kapitole 3 jsem se zabýval prezentací výsledků při sběru a analýze dat. Uvedl jsem pouze pár nejzajímavějších příkladů z důvodu vysoké komplexnosti studovaného tématu. Jsou prezentované v grafech 3.1-3.16. Byl ověřován i load factor, při jehož studiu byla spojována data z Eurostatu, Flightera.net a CAA.co.uk. Zbytek grafů a dat je uveden v přílohách A, B a C.



4 Diskuze výsledků

Sledované průměrné ceny z LHR mohou být dané kvůli nízké konkurenci na linkách z LHR, přičemž destinace TFS je obsloužena z LHR pouze BA, do ostatních dvou destinací létá ještě jeden dopravce, kterým je IB. Stejné chování vykazuje i AF na lince CDG-IBZ, na níž je vidět, že dopravce vykazuje chování typické pro samostatného hráče na trhu, které se vyznačuje svévolností v naceňování letenek.

Tyto ceny potvrzují mou hypotézu o výši ceny letenky dle letiště odletu, jež tvrdí, že pro vysoké HDP na obyvatele je pravděpodobné, že si je toho aerolinka vědoma a podle toho tak nastaví cenu letenky. Londýn se svým HDP na obyvatele ve výši téměř 66000 € má ceny vyšší než ceny z Paříže, která má naopak ceny nejnižší společně s HDP na obyvatele ve výši 56700 €.

Na lince CDG-BCN dochází v časovém intervalu k záměrnému zvýšení ceny v okolí 55. dne. Důvody mohou být různé, zejména pravděpodobně ten, že aerolinky cílí na business travel namísto leisure, protože business cestující není tak citlivý na cenu letenky. Implikací z tohoto navyšování může být navýšení ziskovosti při zachování konkurenceschopnosti.

Zajímavé je, že na lince CDG-TFS aerolinka AF naceňuje své letenky s tendencí snižovat cenu s přibližujícím se datem odletu, což je opět jiný přístup než aerolinka U2, která začíná v měřítku své ceny takřka v průměru ceny za letenku, poté cena stoupá, klesá a opět stoupá – časové intervaly 10 dní, týden a 2 týdny kopírují téměř sinusoidu, viz graf 3.6.

Výše jsou uvedena data nasbíraná v průběhu vytváření mé práce. Je z nich patrné, že ne vždy se podařilo vytvořit regresní model pro chování ceny letenek ve všech čtyřech časových intervalech. Je to dáno tím, že ne všechny dny v `data_dict` jsou vyplněny dostatkem dat a data taktéž nemusela být vhodná pro regresi – nezamítnutí nulové hypotézy korelačních testů (jsou nezávislé). Zároveň jsem nevykresloval regresi, když byla dalšími testy označena jako nesprávná.

V regresních modelech uvedených v kapitole 3.4.1 jsou vidět na lince CDG-BCN různé přístupy k naceňování u různých typů leteckých společností – legacy a low-cost. Je pravděpodobné, že U2 spolu s VY budou svoje letenky naceňovat se vzestupnou polynomiální tendencí s blížícím se dnem odletu, přičemž v AF přijali přístup naceňování letenky pro týden v destinaci s lineární tendencí a v případě týdnu a dvou týdnů v destinaci tendencí polynomiální s maximem v 36. dnu. Domnívám se, že za tím stojí charakteristika společnosti a její politika v poskytování služeb.



Na lince MUC-IBZ dochází k odlišnostem v naceňování mezi aerolinkami DE a LH. Když pomineme rozdíl v ceně, který je jasný z charakteristiky společnosti, tak si lze též všimnout odlišností v průběhu regresní křivky v časovém intervalu "víkend". Zde se ale též domnívám, že se jedná o rozdíl v přístupu k zákazníkovi ze strany aerolinky.

Z regresního grafu pro linku LHR-IBZ od aerolinky BA lze pravděpodobně vyvodit závěr takový, že časový interval víkend je pro BA s větším rozptylem. Začínají s naceňováním na téměř 341 €, u čehož je pravděpodobné, že necílí na zákazníky, kteří vybírají dovolenou v časovém předstihu. Doufají, že se někteří "chytí" a koupí si neúměrně drahou letenku. Následně má křivka strmý trend k minimu v 33. dni, kdy si aerolinka uvědomovala, že potřebuje naplnit letadlo na load factor, který jí zajistí ziskovost a zároveň je pravděpodobné, že se cestující dokáže rozhodnout o víkendovém pobytu v destinaci i měsíc dopředu, a právě během 33. dne došlo k cílovému naplnění a aerolinka vykazuje svévolné chování v naceňování letenek, protože téměř nemá konkurenci, a následně ji navyšuje až do dne odletu.

Rozdíl mezi průběhem regresní křivky na linkách LHR-IBZ a LHR-TFS na časovém intervalu jeden víkend je dle mého charakterizován destinací a segmentem cestujících, kteří letí do destinace. Aerolinka zde postupuje jiným způsobem, byť má v podstatě stejné podmínky.

Dalším poznatkem z obrázku 3.13 lze mimo jiné označit rozdíl v cenách mezi jedním a dvěma týdny strávenými v destinaci – aerolinka pravděpodobně předpokládá, že když už letí cestující do destinace na dva týdny, tak disponuje větším kapitálem, a je tudíž schopný si pořídit letenku za více peněz a aerolinka na něm tedy vydělá více.

V práci jsem potvrdil, že letenky na Tenerife jsou v průměru nejdražší, což je velmi pravděpodobně dáno vzdáleností mezi letištěm odletu a destinací. V případě letenek do BCN a IBZ se nejedná o tak markantní rozdíl ve vzdálenosti, ale v ceně. To je nejspíš dáno tím, že aerolinky cílí na vytvoření zisku přes množství provedených letů, viz obrázky 3.19 a 3.18. Na linkách do IBZ se projevuje vysoká sezónnost v počtu vypravených sedadel – nelétá se do této destinace v zimních měsících, kdežto do BCN se létá celoročně a ve vysokém počtu.



Celkově průběhy regresí na linkách z CDG v posledních 30 dnech jsou uvedeny v tabulkách 4.1, 4.2 a 4.3.

AF:	10 dní	2 týdny	týden	víkendy	
	klesá	klesá	roste	-	průběh
U2:	10 dní	2 týdny	týden	víkendy	
	roste	roste	-	-	průběh
VY:	10 dní	2 týdny	týden	víkendy	
	roste	roste	roste	-	průběh

Tabulka 4.1: Průběh regresí na lince CDG-BCN v posledních 30 dnech do odletu

AF:	10 dní	2 týdny	týden	víkendy	veličina
	-	-	-	roste	průběh

Tabulka 4.2: Průběh regresí na lince CDG-IBZ v posledních 30 dnech do odletu

AF:	10 dní	2 týdny	týden	víkendy	veličina
	-	-	-	klesá	průběh
U2:	10 dní	2 týdny	týden	víkendy	veličina
	klesá	klesá	klesá	klesá	průběh

Tabulka 4.3: Průběh regresí na lince CDG-TFS v posledních 30 dnech do odletu

V tabulkách 4.4 a 4.5 jsou uvedeny průběhy vytvořených regresních modelů pro linky z LHR.

BA:	10 dní	2 týdny	týden	víkendy	veličina
	klesá	roste	roste	roste	průběh

Tabulka 4.4: Průběh regresí na lince LHR-IBZ v posledních 30 dnech do odletu



BA:	10 dní	2 týdny	týden	víkendy	veličina
	roste	klesá	klesá	roste	průběh

Tabulka 4.5: Průběh regresí na lince LHR-TFS v posledních 30 dnech do odletu

A v tabulkách 4.6, 4.7 a 4.8 jsou uvedeny průběhy regresí z MUC.

LH	10 dní	2 týdny	týden	víkendy	veličina
	-	-	-	roste	průběh

VY	10 dní	2 týdny	týden	víkendy	veličina
	roste	roste	roste	roste	průběh

Tabulka 4.6: Průběh regresí na lince MUC-BCN v posledních 30 dnech do odletu

DE	10 dní	2 týdny	týden	víkendy	veličina
	roste	roste	roste	roste	průběh

LH	10 dní	2 týdny	týden	víkendy	veličina
	roste	klesá	roste	klesá	průběh

Tabulka 4.7: Průběh regresí na lince MUC-IBZ v posledních 30 dnech do odletu

DE	10 dní	2 týdny	týden	víkendy	veličina
	klesá	roste	klesá	-	průběh

MBU	10 dní	2 týdny	týden	víkendy	veličina
	klesá	klesá	klesá	-	průběh

X3	10 dní	2 týdny	týden	víkendy	veličina
	klesá	klesá	-	-	průběh

Tabulka 4.8: Průběh regresí na lince MUC-TFS v posledních 30 dnech do odletu

Z tabulek 4.1-4.8 lze usuzovat, že linky do Barcelony ať už od legacy nebo od low-cost aerolinek podporují přirozenou hypotézu o zdražování letenek, která tvrdí, že ceny letenek s ubývajícími



dny do odletu rostou. Je to dáno tím, že je zde segment, který je méně cenově elastický a je více business orientovaný a že tato lokalita podporuje rozhodování o dovolené / výletu na poslední chvíli. Bohužel data z LHR nemám kvůli nevhodnosti regresních modelů ať už pro BA nebo IB (společnosti provozující linky z LHR), ale z CDG a MUC tuto hypotézu potvrzují. Současně s linkami do BCN toto potvrzují i linky na Ibizu – cena během posledních 30 dní do odletu ve většině případů roste. Výjimkami jsou LHR-IBZ od BA v časovém intervalu 10 dní, MUC-IBZ od LH v časových intervalech 2 týdny a víkendy.

Tuto hypotézu nicméně vyvrací linky na TFS, na nichž dochází v naprosté většině případů k poklesu ceny během posledních 30 dní, a to opět jak v případě low-cost, tak legacy aerolinky. Nachází se zde taktéž výjimky – LHR-TFS od BA v časových intervalech 10 dní a víkendy a MUC-TFS od DE v časovém intervalu 2 týdny.

Mám za to, že na linkách na TFS dopravci potřebují na poslední chvíli naplnit letadlo do co nejvíce ziskových možností, proto cena klesá (je možné, že není o linky na TFS takový zájem), ale o linky do Barcelony a na Ibizu je zájem takový, že jsou letadla již vyprodána do takového load factoru, že si aerolinky již mohou dovolit letenky nacenit výše a zvýšit tak svůj zisk.

Dataset získaný v této práci považuji za validní vlivem používaných příkazů ve vytváření datasetů na jednotlivých tratích – jedná se o běžně používanou metodu HTTP requestů, považovanou v dnešní době prakticky za standard, na rozdíl od techniky sbírání dat pomocí WebScrapingových aplikací, které si musí data nejdříve přečíst na načtené webové stránce, jež sbírá data pro jejich načtení pomocí HTTP requestů použitím metody GET.

Validitu výsledků podporuje vytváření regresí, které jsou dnes běžně používané v jednom z nejpoužívanějších programovacích jazyků Python. Při rozhodování o použitelnosti dat pro regrese, jejich vytvoření a následnou analýzu jsem využil materiálů od paní docentky Uglickich [25] a odborné konzultace od paní doktorky Pecherkové.



5 Závěr

V práci jsem potvrdil předpoklad o ceně letenek v závislosti na HDP v oblasti odletu – čím vyšší HDP, tím vyšší průměrná cena za letenku. Taktéž bylo potvrzeno, že aerolinky mají v případě slabé nebo žádné konkurence volnou ruku v nastavování ceny letenek – v závislosti na předešle zjištěných parametrech o vykupování kapacity letadel.

Důležité je též pro zákazníka si správně načasovat koupi letenky podle dnů do odletu. Neexistuje však bohužel na základě zjištěných dat všeobecný návod na všechny typy aerolinek a linek a časových intervalů. Do BCN je pravděpodobné, že na nejnižší cenu zákazník narazí alespoň 2 měsíce před odletem, na IBZ je výhodnější si koupit letenku během posledního měsíce do odletu a na TFS existuje návod pouze pro pobyt pro jeden nebo dva týdny – ideálně zakoupit letenku v průběhu posledního měsíce.

Naceňování je odlišné podle typu aerolinky – low-cost přistupuje k tomuto problému jinak než legacy aerolinka, tudíž si na typ dopravce musí dát zákazník taktéž pozor při výběru letenky. Její chování v rámci nabízení zpáteční letenky pro dva odlišně dlouhé pobyty v destinaci vykazovalo tendence naceňovat letenku výše pro delší pobyt v destinaci, pravděpodobně s premisou, že cestující bude mít více finančních prostředků, tudíž nebude tak citlivý na cenu.

Práce je limitována nejen z teoretického hlediska, protože aerolinky nezveřejňují strategii svého revenue managementu. Zároveň se ne vždy podařilo vytvořit regrese pro naceňování, s největší pravděpodobností kvůli nedostatku dat nebo byly výkyvy v ceně příliš vysoké. Taktéž bylo sledováno malé množství linek, abychom mohli na 100 % určit správnost potvrzení / vyvrácení hypotéz. Tímto se mohu zabývat ve své diplomové práci. Výzkum v této práci avšak poodhalil komplexnosti v naceňování letenek a pomáhá tak utvářet obrázek o tom, jak se staví aerolinky k trhu. Taktéž celá část programování mohla být vymyšlena efektivněji, avšak v průběhu vytváření práce docházelo k různým iteracím, tudíž byl častokrát vytvářen kód nový, byť mohly být využity části z předtím napsaného kódu. Práci jako takovou toto nijak neomezuje, ale je zde prostor ke zlepšení.

Navazující vědecká činnost by se mohla skládat ze sběru dat na větším časovém intervalu, aby se podařilo rozšířit dataset o letní a podzimní měsíce, a z kontaktování zástupců z aerolinek, zda-li by neposkytli jejich pohled na toto téma.



Seznam použité literatury

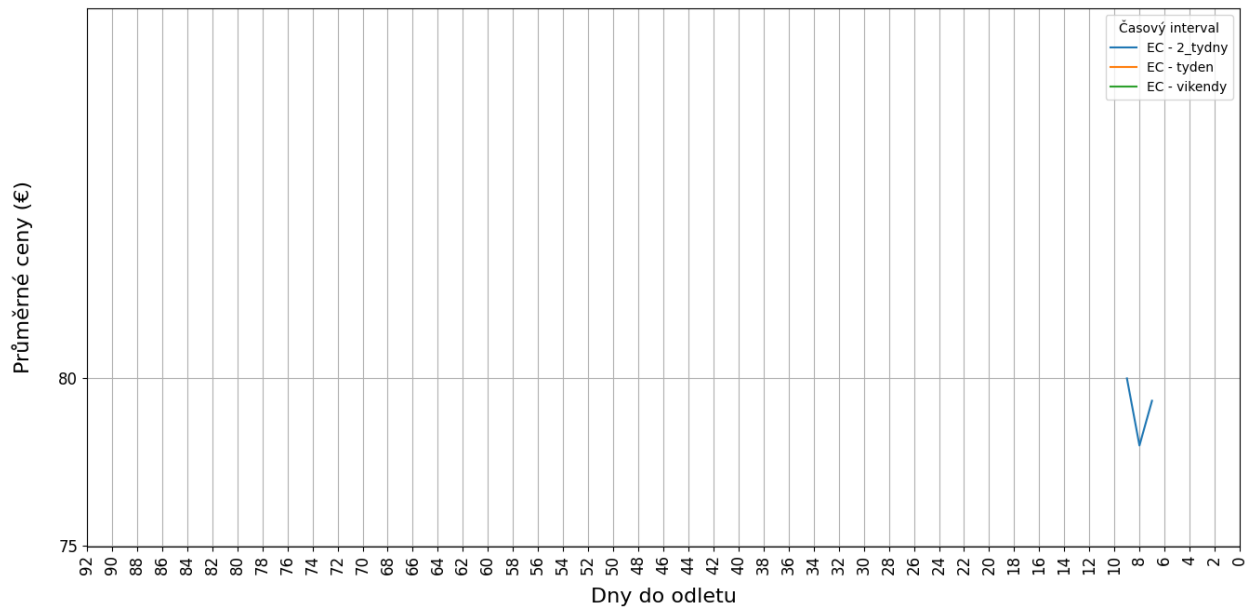
1. WINDMANN, Marcel; PERÉZ DE LA CRUZ, José Manuel. *Competitive Position Analysis of Airlines: Traditional Airlines and Low Cost Carriers - Market Development, Trends and Outlooks based on the European Market*. Plaça Cívica, 08193 Bellaterra, Barcelona, Španělsko, 2015.
2. PRASATH, S. Naveen; M, Sathish Kumar; ELIYAS, Sherin. A Prediction of Flight Fare Using K-Nearest Neighbors. In: *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 2022. Dostupné z DOI: 10.1109/icacite53722.2022.9823876.
3. JANSSEN, Tom; DIJKSTRA, Ton. *A Linear Quantile Mixed Regression Model for Prediction of Airline Ticket Prices*. Houtlaan 4, 6525 XZ Nijmegen, 2014.
4. FRANCIS, Graham; FIDATO, Alessandro; HUMPHREYS, Ian. Airport–airline interaction: the impact of low-cost carriers on two European airports. *Journal of Air Transport Management*. 2003, roč. 9, č. 4, s. 267–273. Dostupné z DOI: 10.1016/s0969-6997(03)00004-8.
5. RENOLD, Manuel; KULJANIN, Jovana; KALIĆ, Milica. The comparison of financial performance of airlines with different business model operated in long-haul market. *Transportation Research Procedia*. 2019, roč. 43, s. 178–187. Dostupné z DOI: 10.1016/j.trpro.2019.12.032.
6. LEWIS, Matthew S. Identifying airline price discrimination and the effect of competition. *International Journal of Industrial Organization*. 2021, roč. 78, s. 102761. Dostupné z DOI: 10.1016/j.ijindorg.2021.102761.
7. STAVINS, Joanna. Price Discrimination in the Airline Market: The Effect of Market Concentration. *Review of Economics and Statistics*. 2001, roč. 83, č. 1, s. 200–202. Dostupné z DOI: 10.1162/rest.2001.83.1.200.
8. ESCOBARI, Diego; RUPP, Nicholas G.; MESKEY, Joseph. An Analysis of Dynamic Price Discrimination in Airlines. *Southern Economic Journal*. 2018. Dostupné z DOI: 10.1002/soej.12309.
9. CHANDRA, Ambarish. *Price Discrimination along Multiple Dimensions: Theory and New Evidence*. 105 St. George Street, Toronto, Ontario, M5S 3E6, Canada: University of Toronto, 2021.



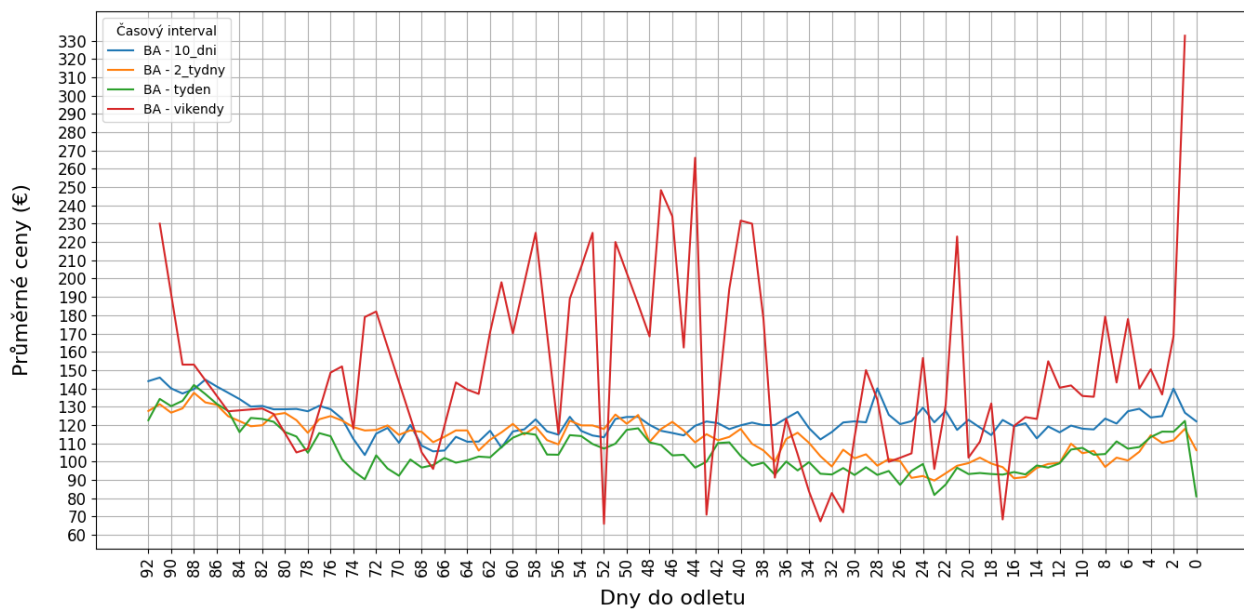
10. *Air Transport Conditions – Vueling*. 2023. Dostupné také z: <https://www.vueling.com/en/customer-services/conditions-of-carriage>.
11. *General conditions of carriage, legal information – Air France*. 2023. Dostupné také z: <https://www.airfrance.cz/information/legal/edito-cg-airfrance>.
12. SWEETING, Andrew. Dynamic Pricing Behavior in Perishable Goods Markets: Evidence from Secondary Markets for Major League Baseball Tickets. *Journal of Political Economy*. 2012, roč. 120, č. 6, s. 1133–1172. Dostupné z DOI: 10.1086/669254.
13. MCAFEE, R Preston; TE VELDE, Vera. Dynamic pricing in the airline industry. *Handbook on economics and information systems*. 2006, roč. 1, s. 527–67.
15. *Airport Expansion Consultation – Heathrow*. 2019. Dostupné také z: <https://shorturl.at/bekV0>.
16. *Traffic Statistics – Heathrow*. 2023. Dostupné také z: <https://www.heathrow.com/company/investor-centre/reports/traffic-statistics>.
17. *Heathrow – Best Placed for Britain*. 2013. Dostupné také z: <https://shorturl.at/pHLNW>.
18. *Gross domestic product per capita in London from 1998 to 2021 (in GBP)*. 2023. Dostupné také z: <https://www.statista.com/statistics/378990/gdp-per-head-london/>.
19. *Eurostat / Regions and Cities Illustrated (RCI)*. 2023. Dostupné také z: <https://ec.europa.eu/eurostat/cache/RCI/#?vis=nuts3.economy&lang=en>.
20. *Google Maps*. 2023. Dostupné také z: <https://www.google.com/maps>.
21. *E-commerce and express freight - Cargo - Paris Aéroport*. 2023. Dostupné také z: <https://shorturl.at/emEPY>.
22. *MUNICH (MUC) - BARIG*. 2023. Dostupné také z: <https://shorturl.at/arD39>.
23. *Gross domestic product (GDP) per capita in Spain in the year 2021, by autonomous community*. 2022. Dostupné také z: <https://shorturl.at/loqAI>.
24. *CAA.co.uk*. 2023. Dostupné také z: <https://shorturl.at/clpMZ>.
25. *Volba testu hypotéz*. 2023. Dostupné také z: <http://staff.utia.cas.cz/uglickich/pdfka/volbaTH.pdf>.

Přílohy

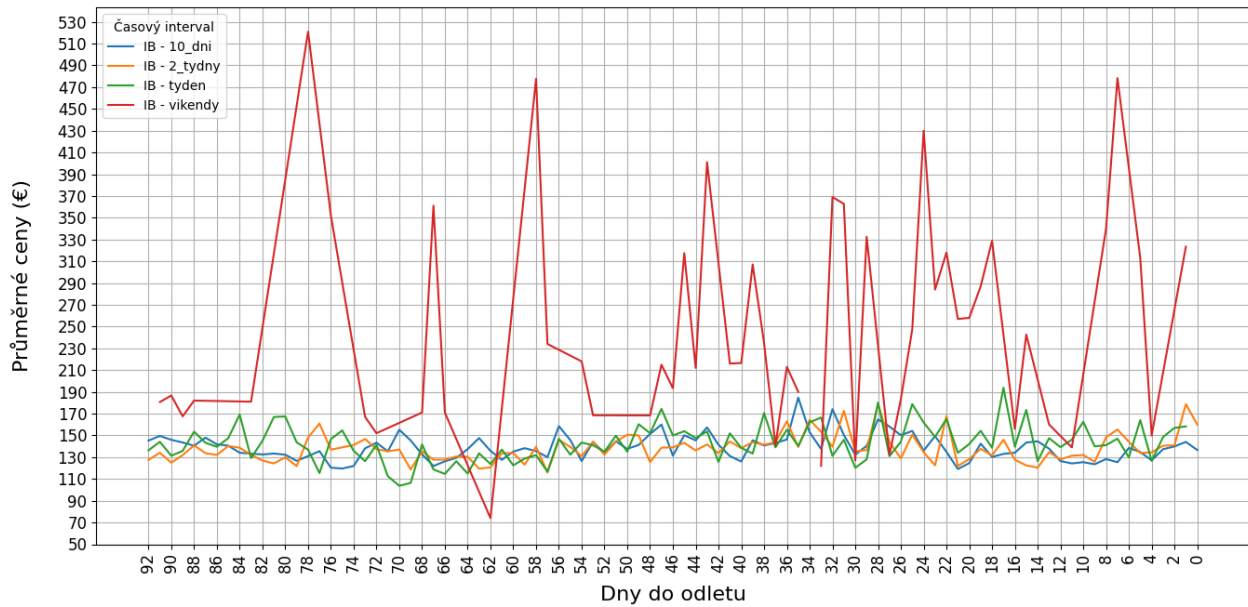
A Spojnicové grafy surových dat



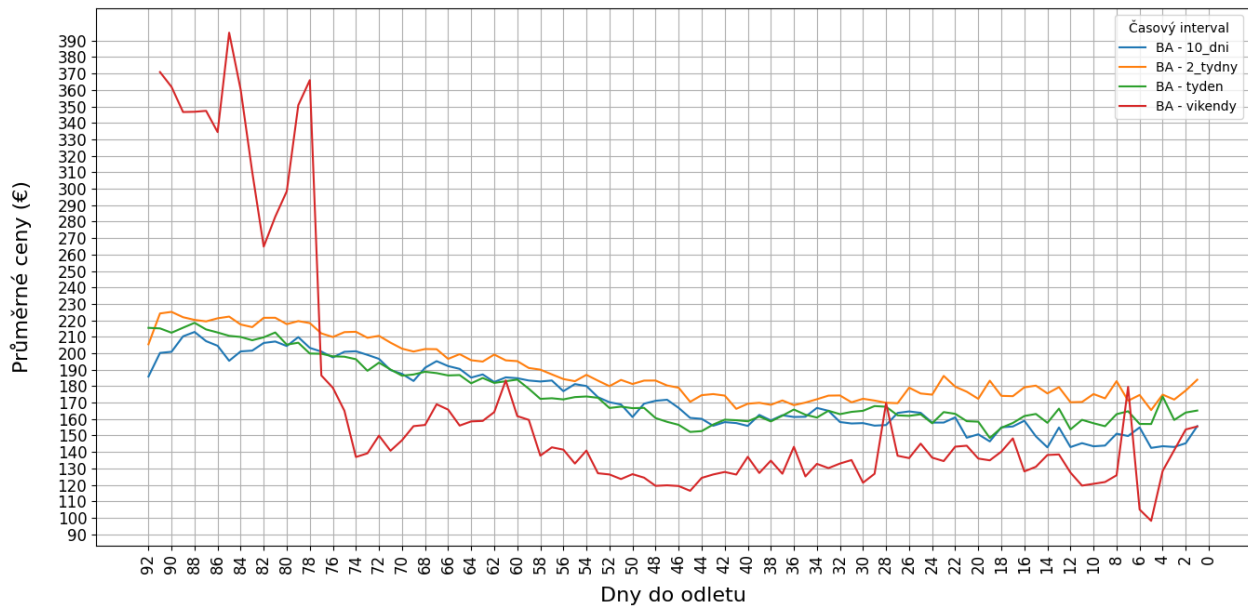
Obrázek A.1: Vývoj ceny na lince CDG-BCN, aerolinka: EC



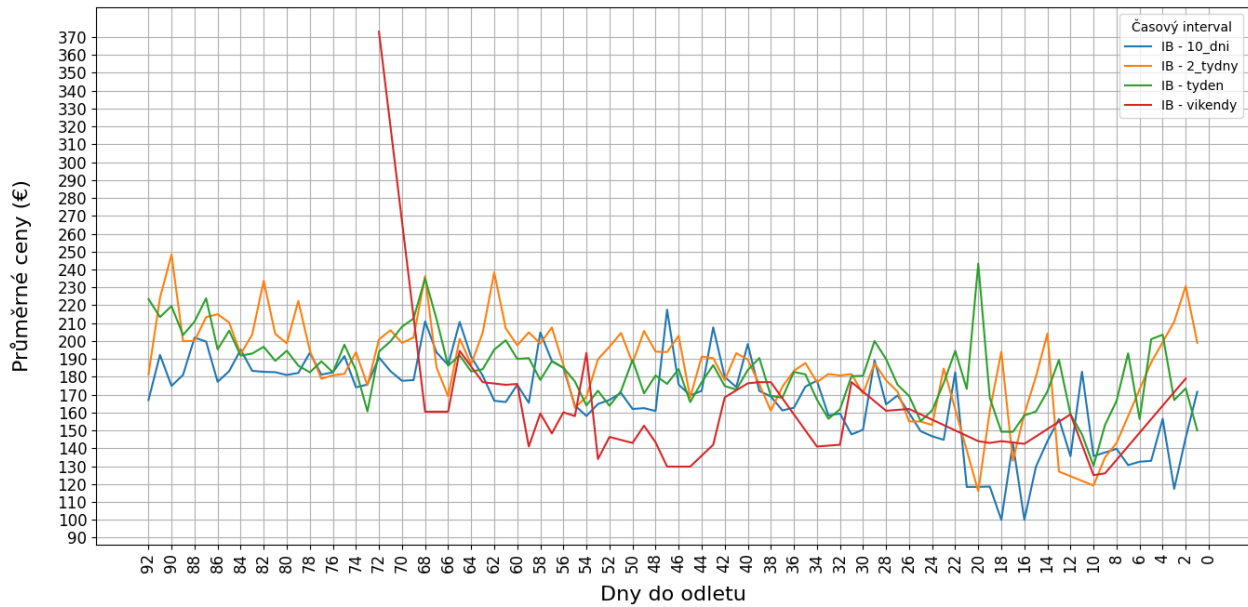
Obrázek A.2: Vývoj ceny na lince LHR-BCN, aerolinka: BA



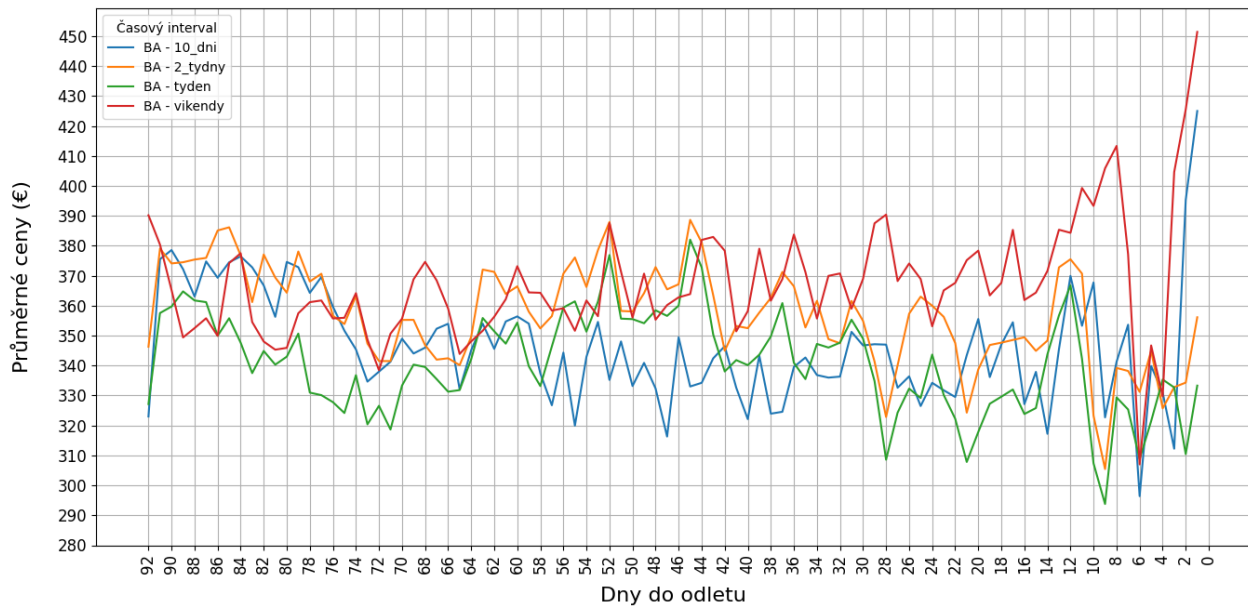
Obrázek A.3: Vývoj ceny na lince LHR-BCN, aerolinka: IB



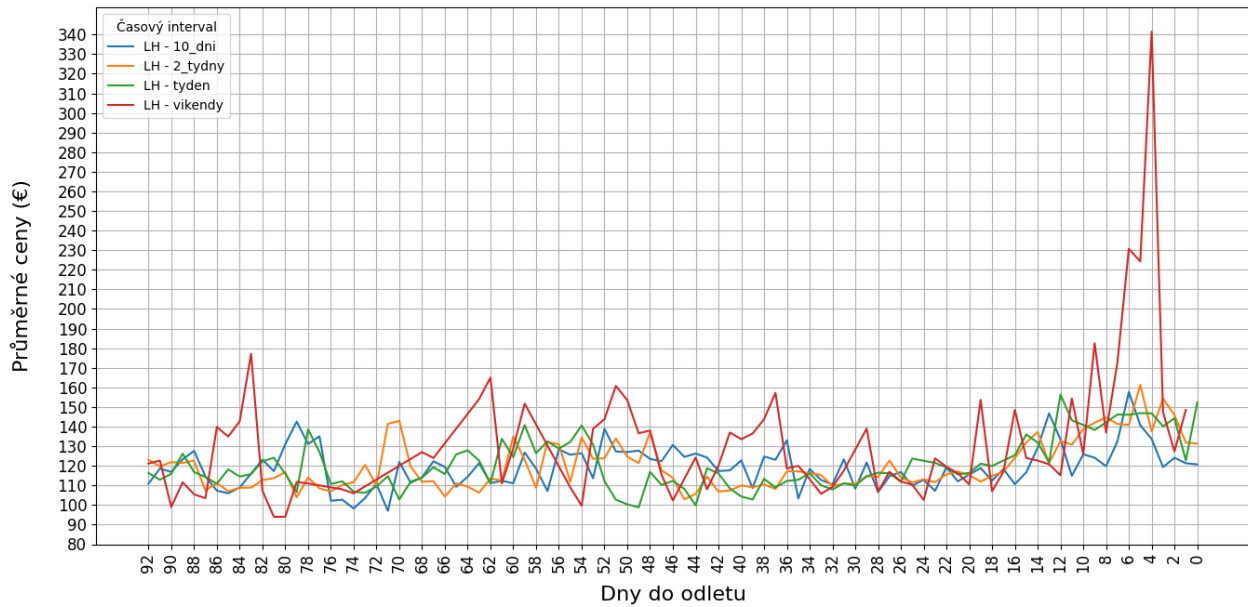
Obrázek A.4: Vývoj ceny na lince LHR-IBZ, aerolinka: BA



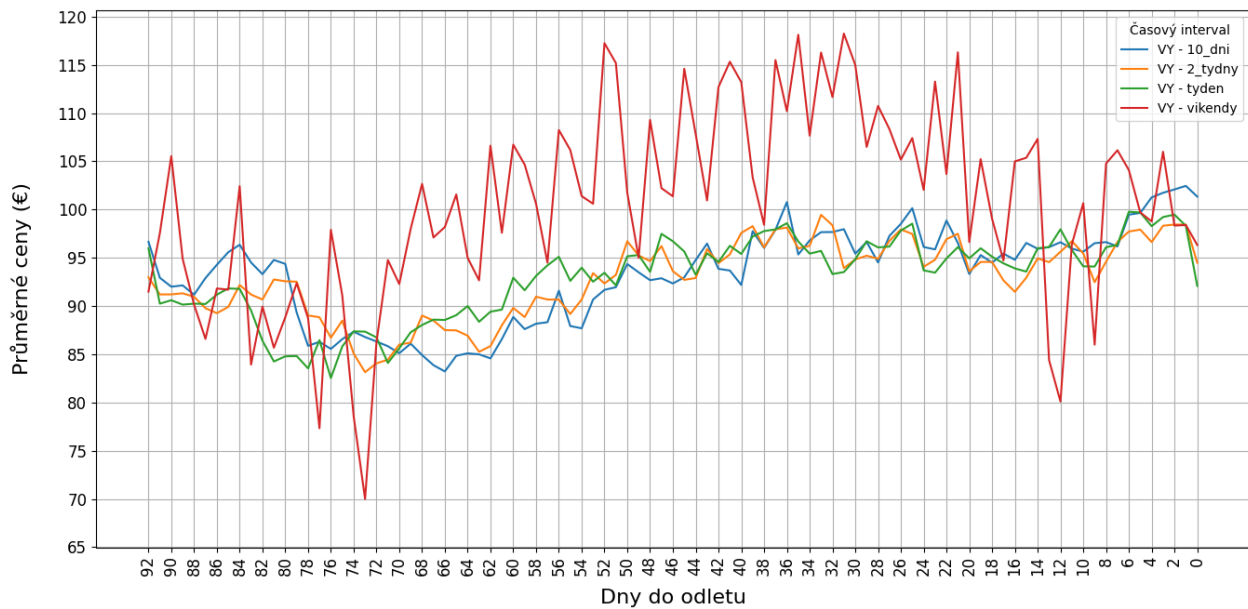
Obrázek A.5: Vývoj ceny na lince LHR-IBZ, aerolinka: IB



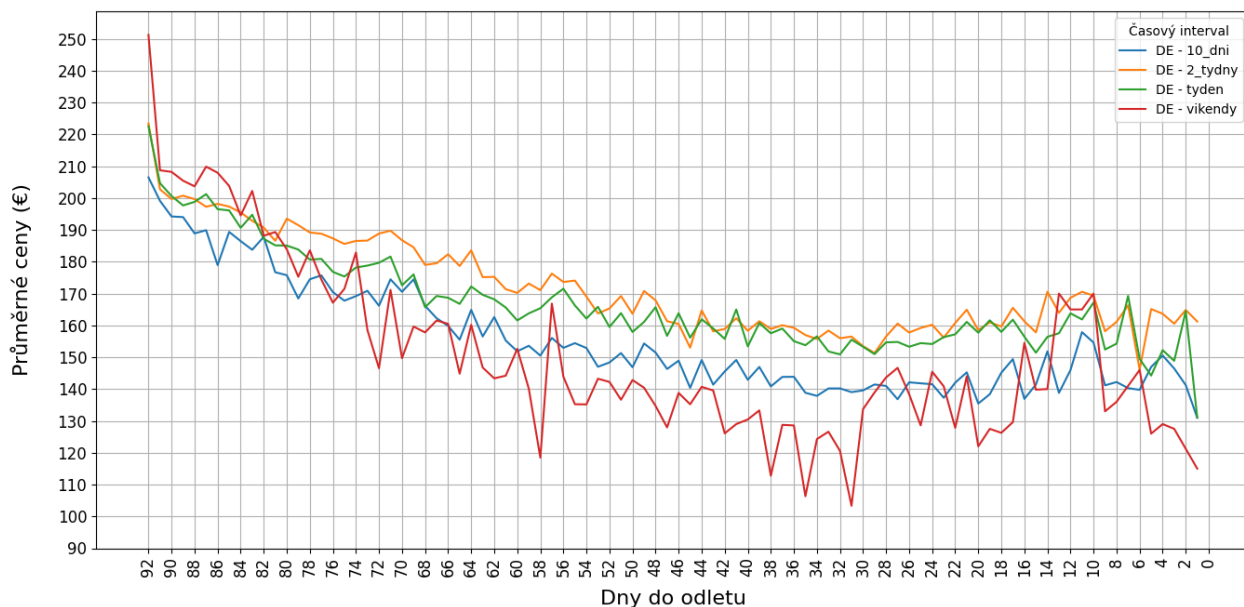
Obrázek A.6: Vývoj ceny na lince LHR-IBZ, aerolinka: BA



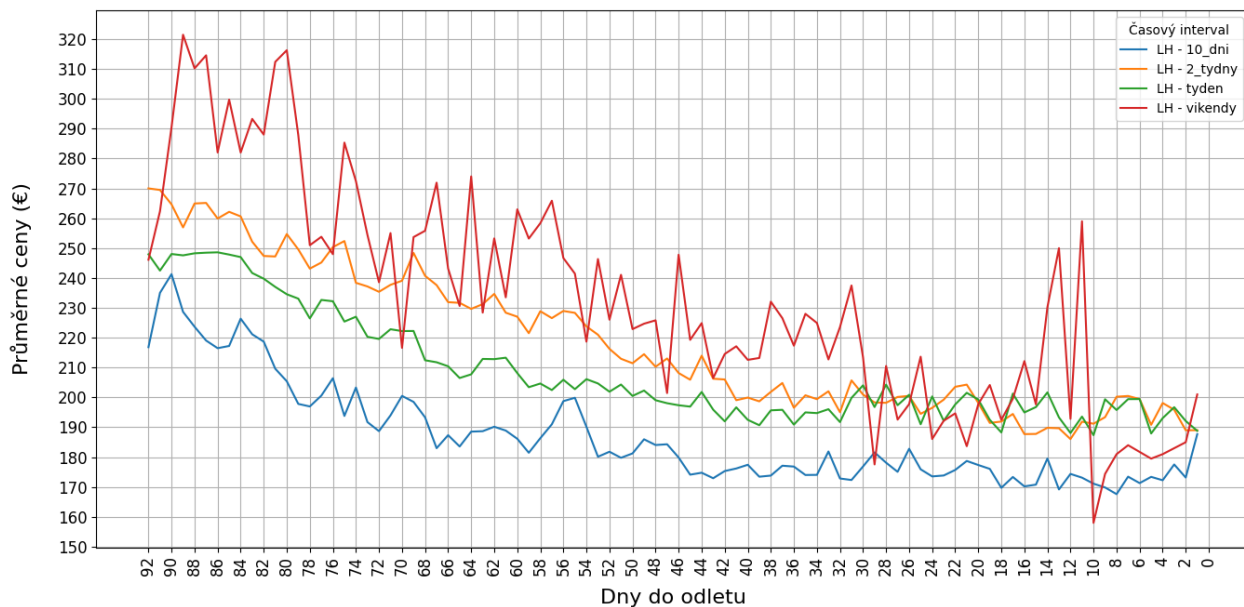
Obrázek A.7: Vývoj ceny na lince MUC-BCN, aerolinka: LH



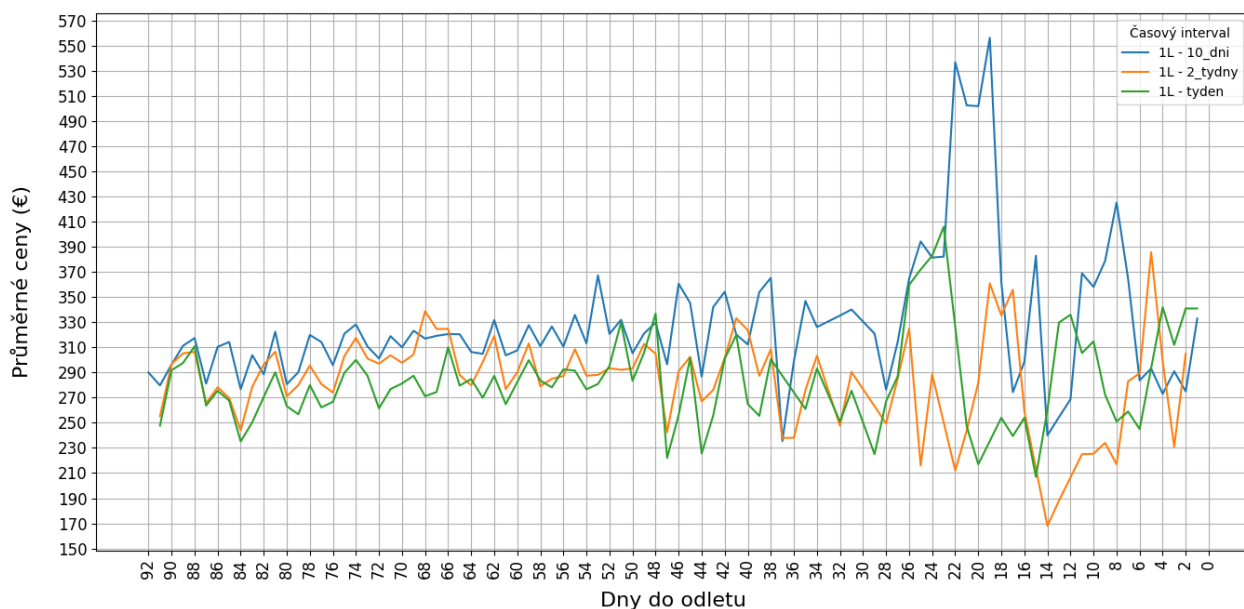
Obrázek A.8: Vývoj ceny na lince MUC-BCN, aerolinka: VY



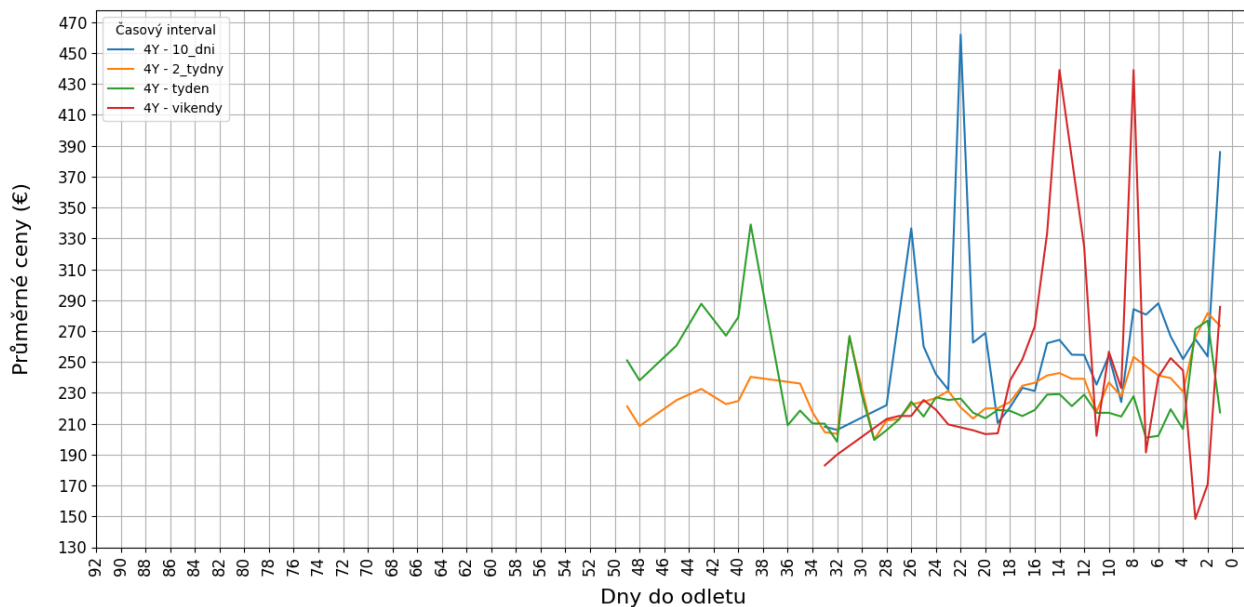
Obrázek A.9: Vývoj ceny na lince MUC-IBZ, aerolinka: IB



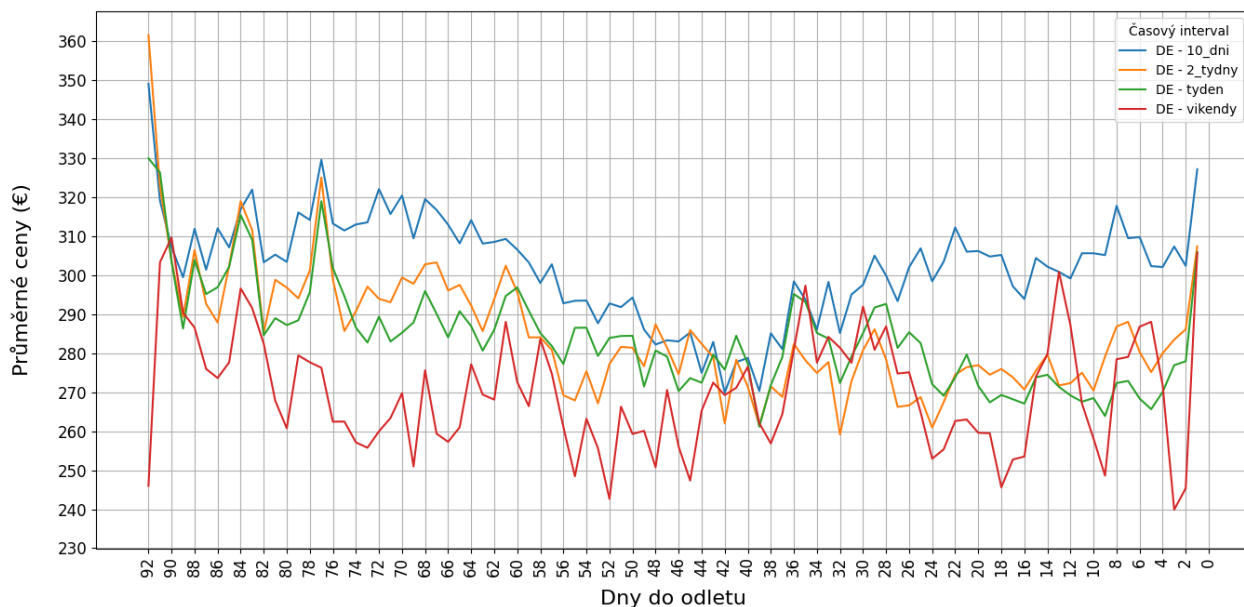
Obrázek A.10: Vývoj ceny na lince MUC-IBZ, aerolinka: BA



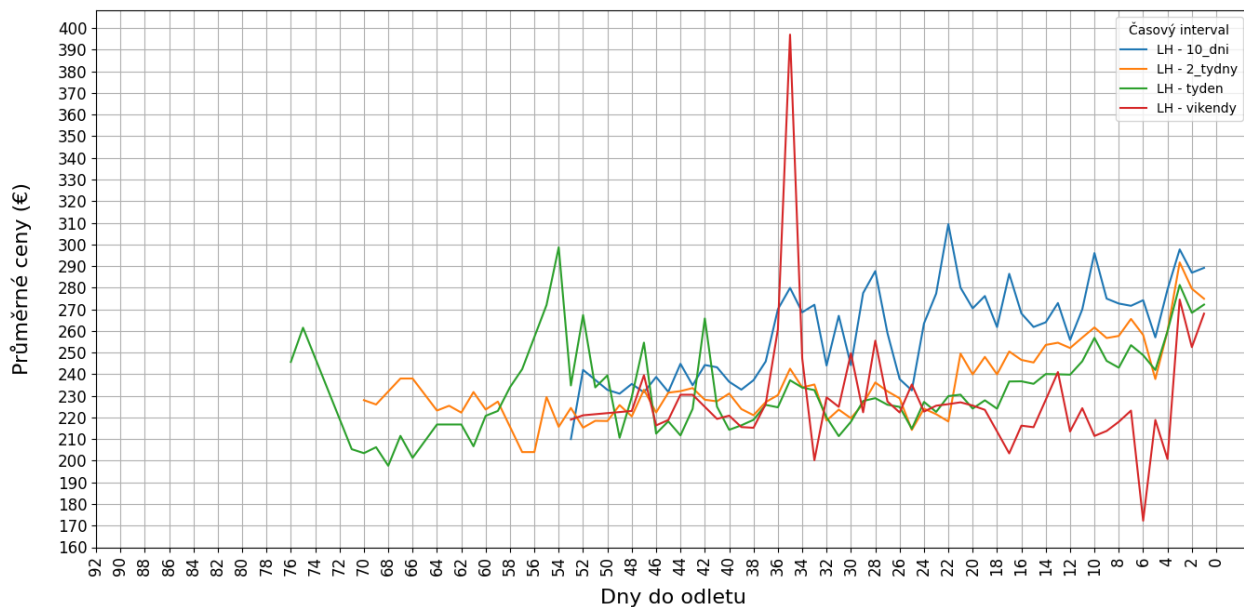
Obrázek A.11: Vývoj ceny na lince MUC-TFS, aerolinka: 1L



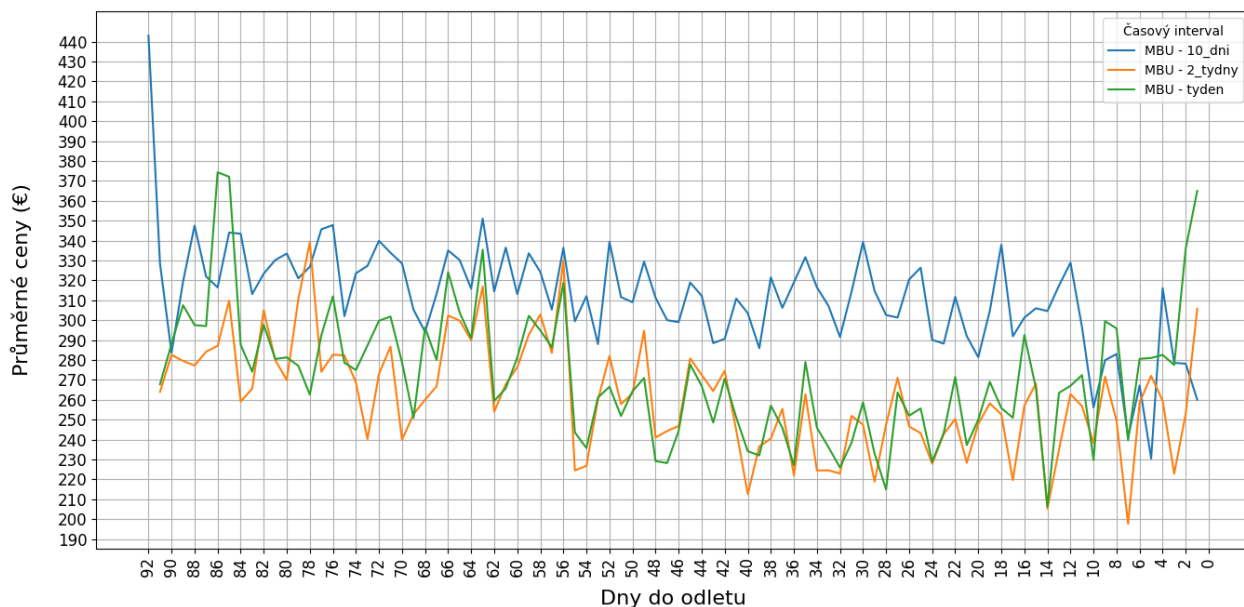
Obrázek A.12: Vývoj ceny na lince MUC-TFS, aerolinka: 4Y



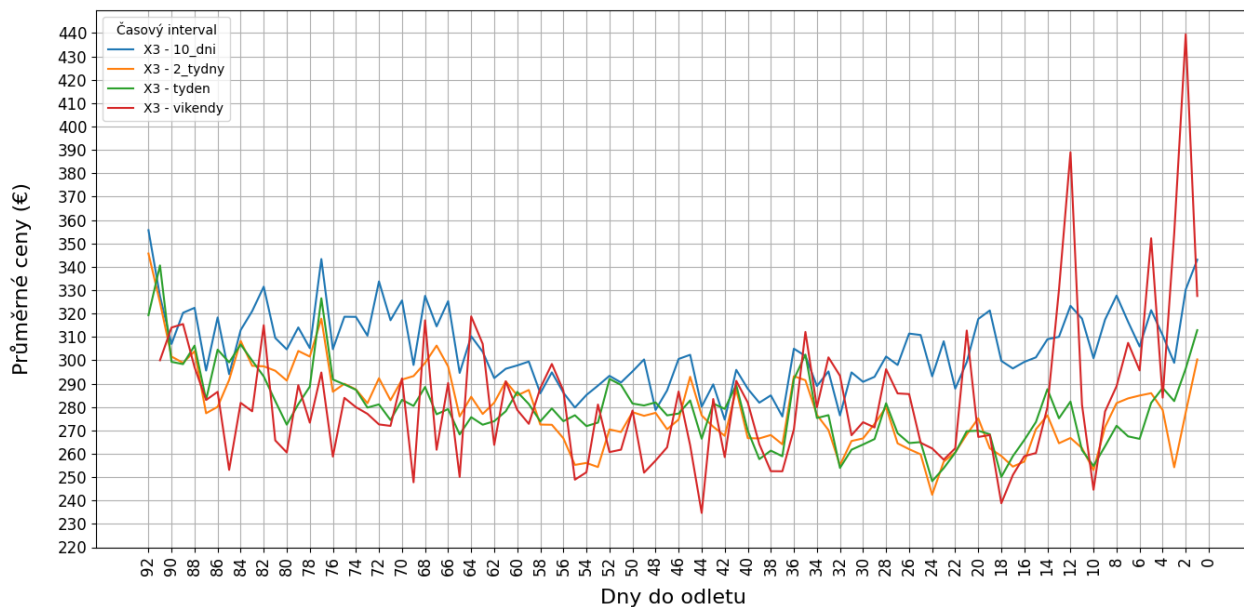
Obrázek A.13: Vývoj ceny na lince MUC-TFS, aerolinka: DE



Obrázek A.14: Vývoj ceny na lince MUC-TFS, aerolinka: LH

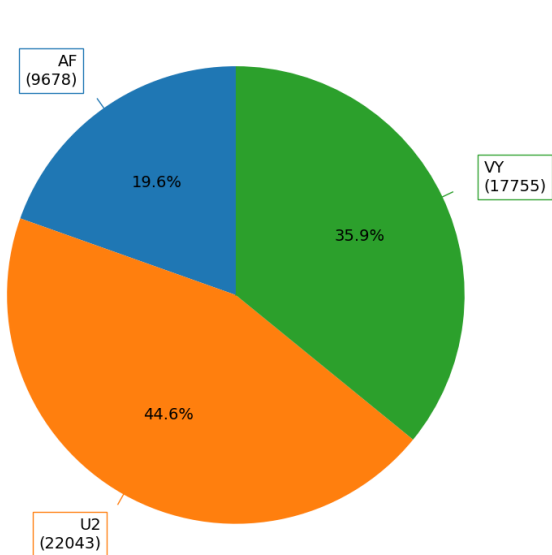


Obrázek A.15: Vývoj ceny na lince MUC-TFS, aerolinka: MBU

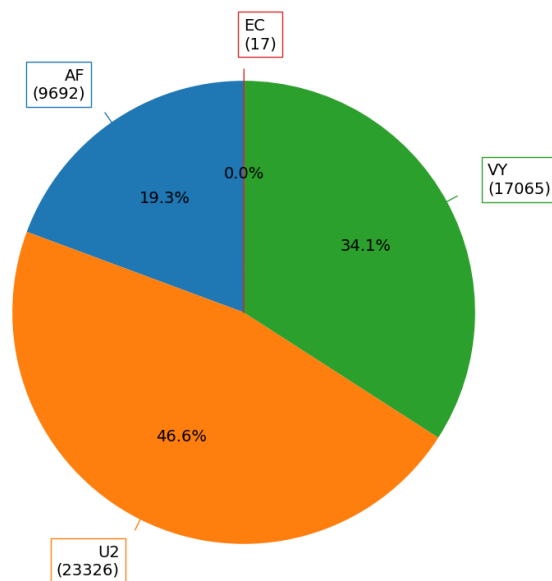


Obrázek A.16: Vývoj ceny na lince MUC-TFS, aerolinka: X3

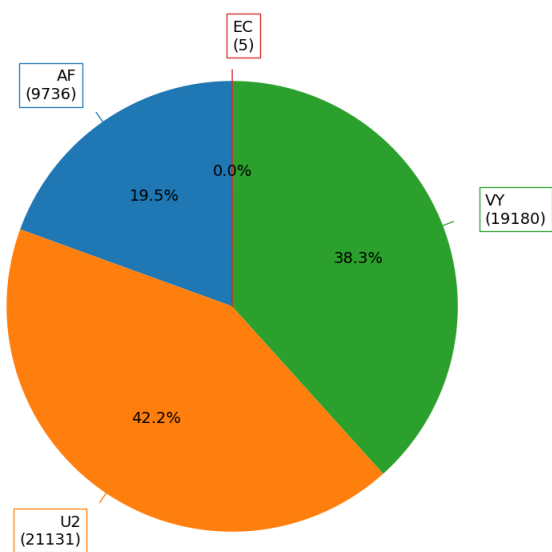
B Podíl na trhu



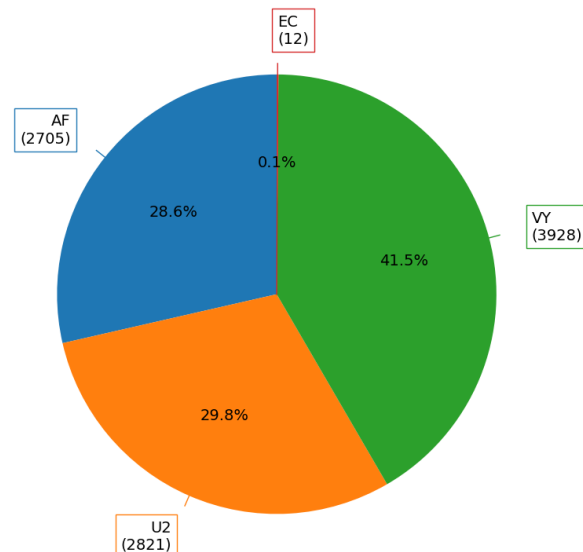
Obrázek B.1: Podíl na trhu, linka CDG-BCN, časový interval: 10 dní, počet letů



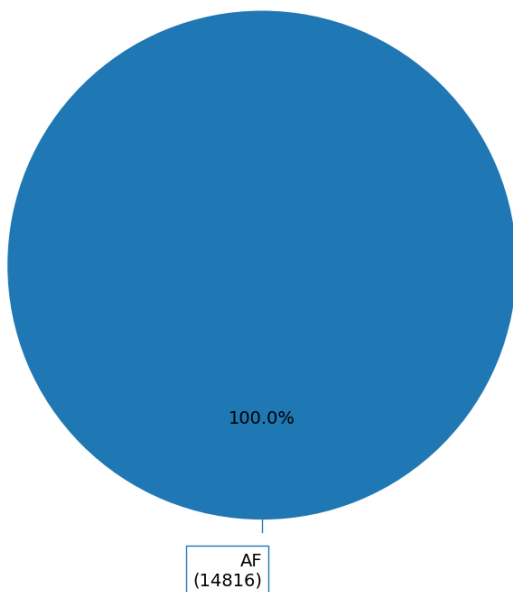
Obrázek B.2: Podíl na trhu, linka CDG-BCN, časový interval: 2 týdny, počet letů



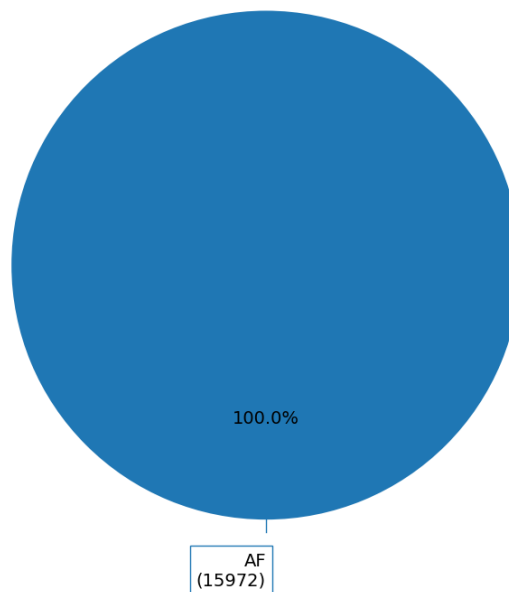
Obrázek B.3: Podíl na trhu, linka CDG-BCN, časový interval: týden, počet letů



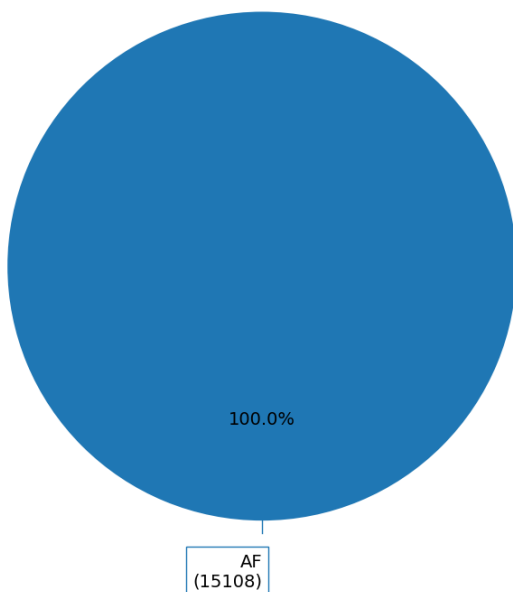
Obrázek B.4: Podíl na trhu, linka CDG-BCN, časový interval: víkend, počet letů



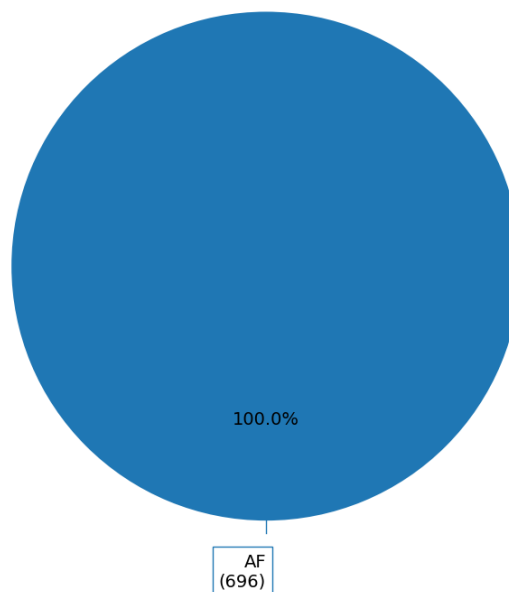
Obrázek B.5: Podíl na trhu, linka CDG-IBZ,
časový interval: 10 dní, počet letů



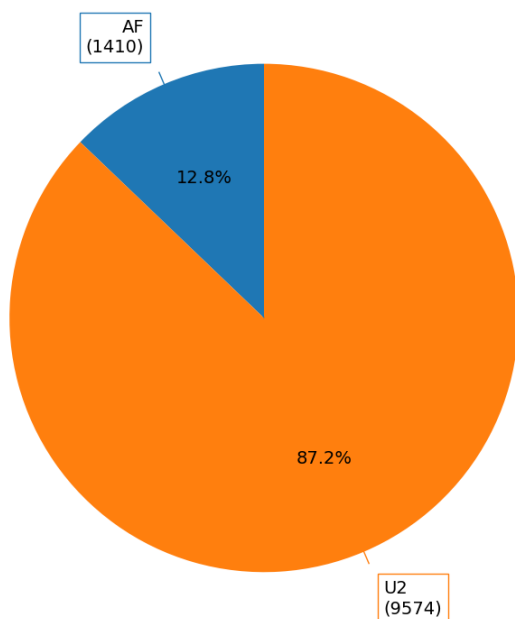
Obrázek B.6: Podíl na trhu, linka CDG-IBZ,
časový interval: 2 týdny, počet letů



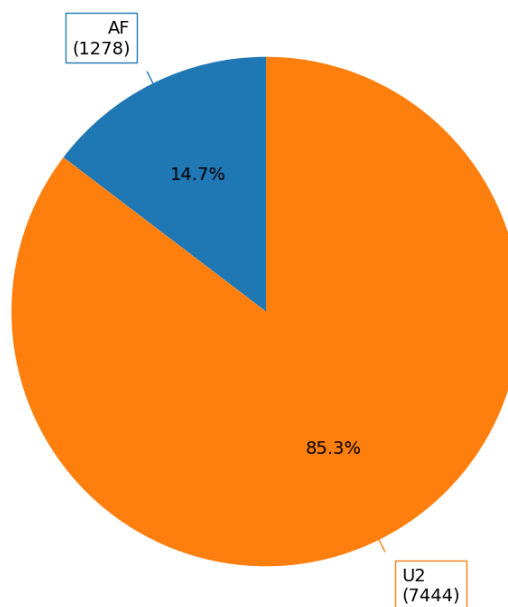
Obrázek B.7: Podíl na trhu, linka CDG-IBZ,
časový interval: týden, počet letů



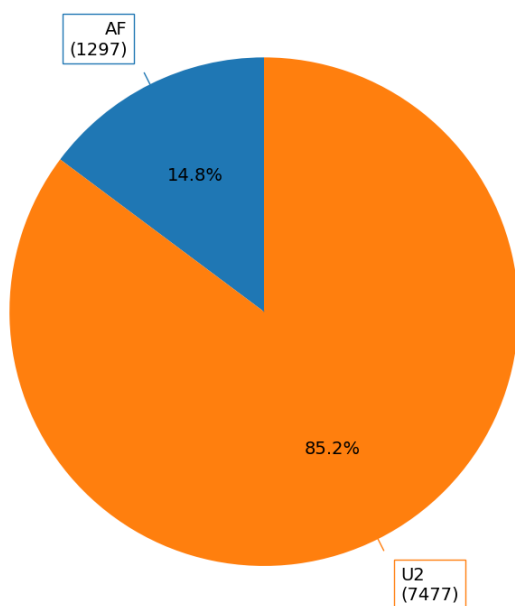
Obrázek B.8: Podíl na trhu, linka CDG-IBZ,
časový interval: víkend, počet letů



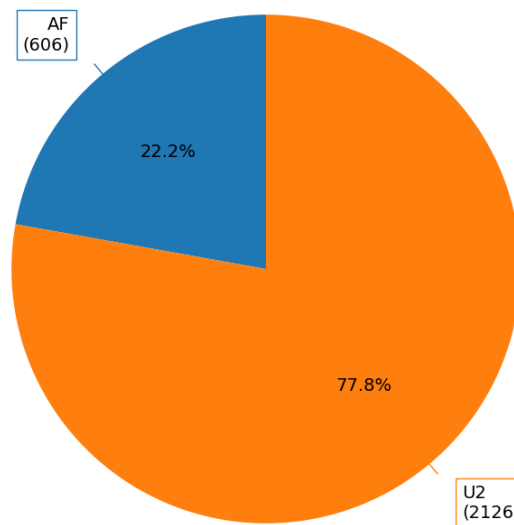
Obrázek B.9: Podíl na trhu, linka CDG-TFS, časový interval: 10 dní, počet letů



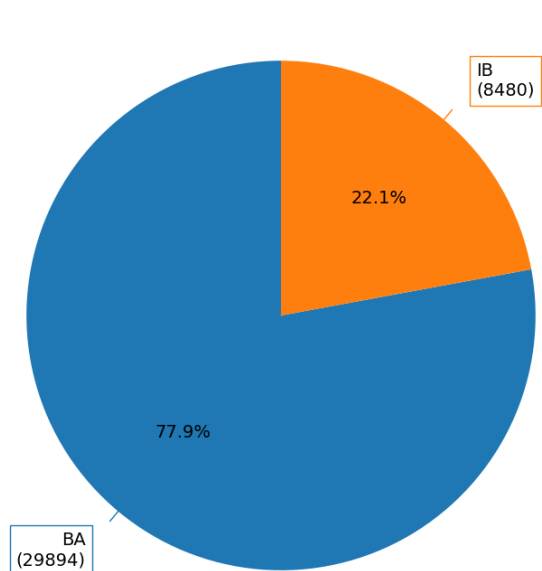
Obrázek B.10: Podíl na trhu, linka CDG-TFS, časový interval: 2 týdny, počet letů



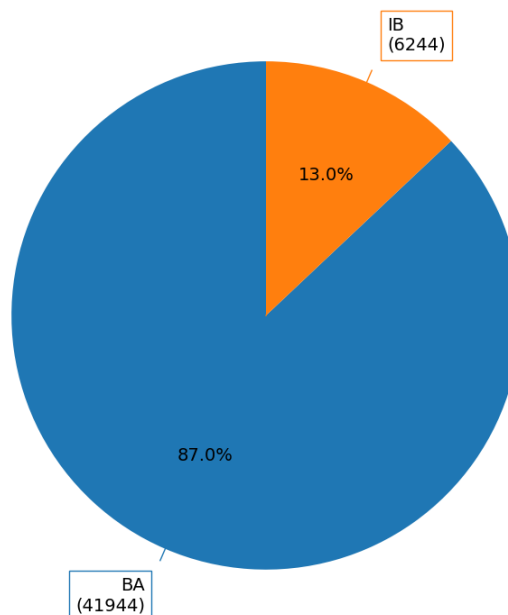
Obrázek B.11: Podíl na trhu, linka CDG-TFS, časový interval: týden, počet letů



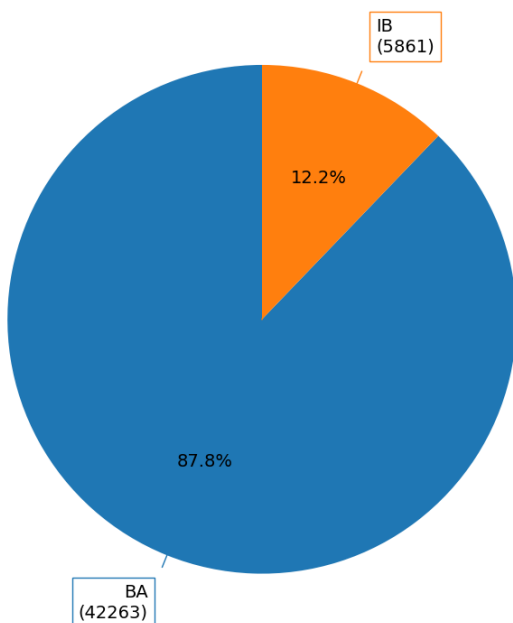
Obrázek B.12: Podíl na trhu, linka CDG-TFS, časový interval: víkend, počet letů



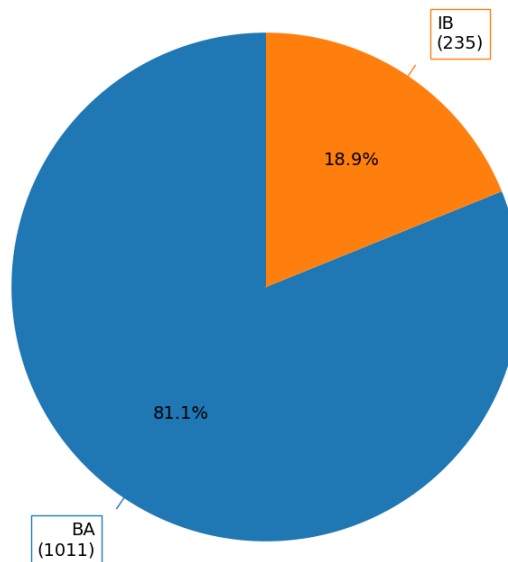
Obrázek B.13: Podíl na trhu, linka LHR-BCN, časový interval: 10 dní, počet letů



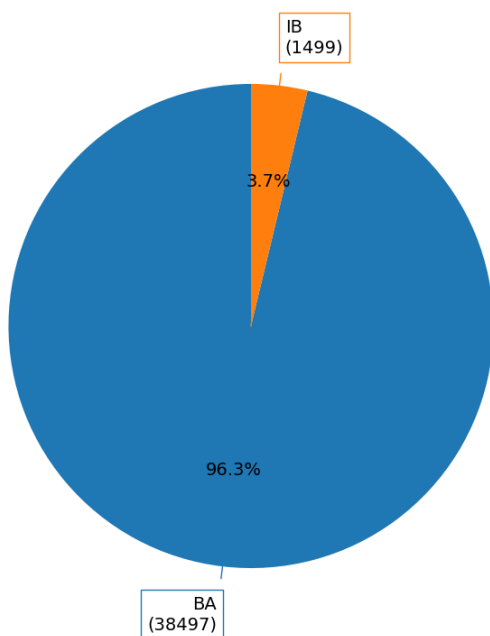
Obrázek B.14: Podíl na trhu, linka LHR-BCN, časový interval: 2 týdny, počet letů



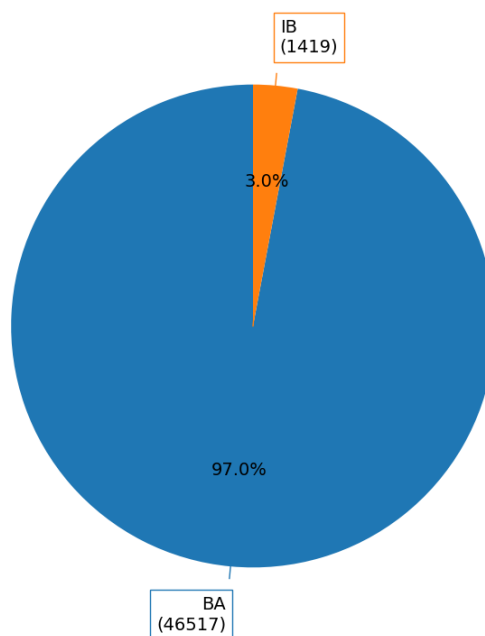
Obrázek B.15: Podíl na trhu, linka LHR-BCN, časový interval: týden, počet letů



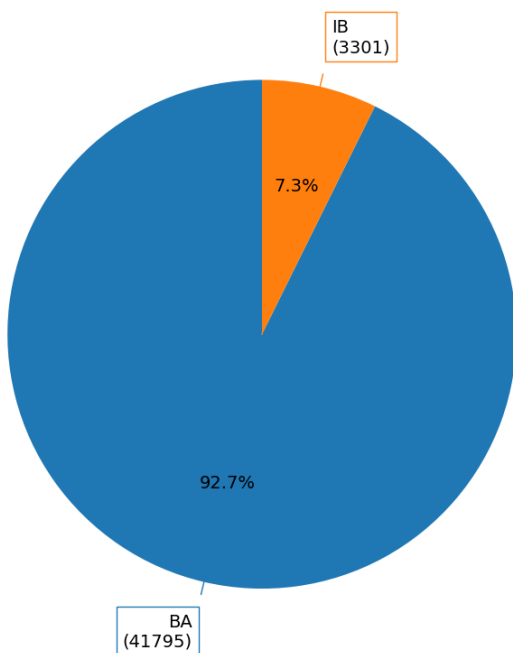
Obrázek B.16: Podíl na trhu, linka LHR-BCN, časový interval: víkend, počet letů



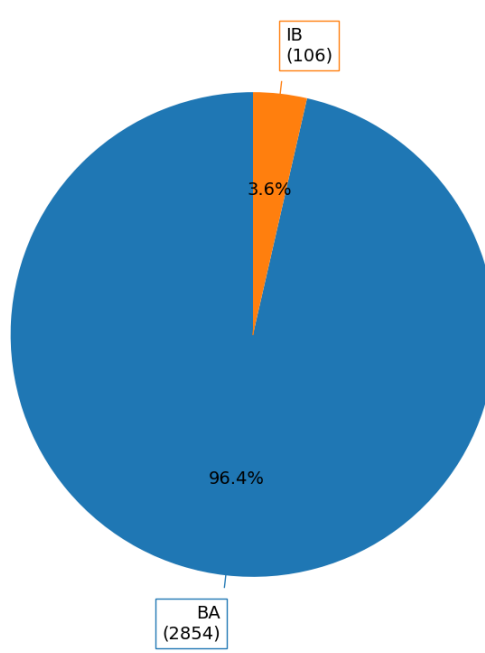
Obrázek B.17: Podíl na trhu, linka LHR-IBZ,
časový interval: 10 dní, počet letů



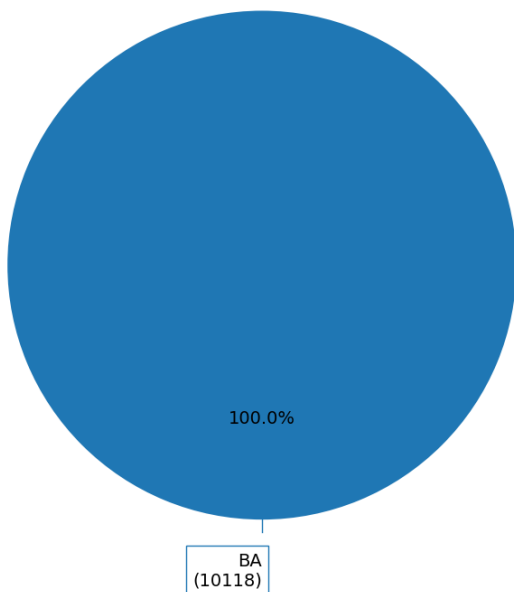
Obrázek B.18: Podíl na trhu, linka LHR-IBZ,
časový interval: 2 týdny, počet letů



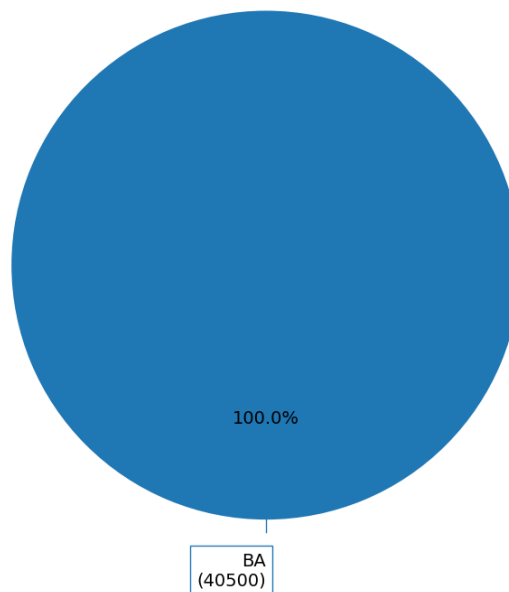
Obrázek B.19: Podíl na trhu, linka LHR-IBZ,
časový interval: týden, počet letů



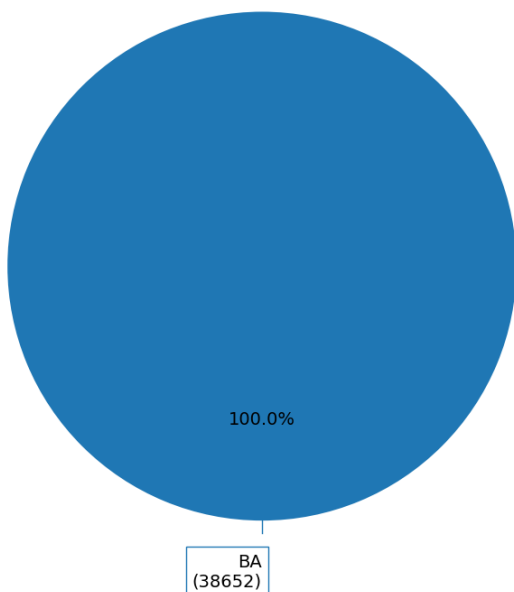
Obrázek B.20: Podíl na trhu, linka LHR-IBZ,
časový interval: víkend, počet letů



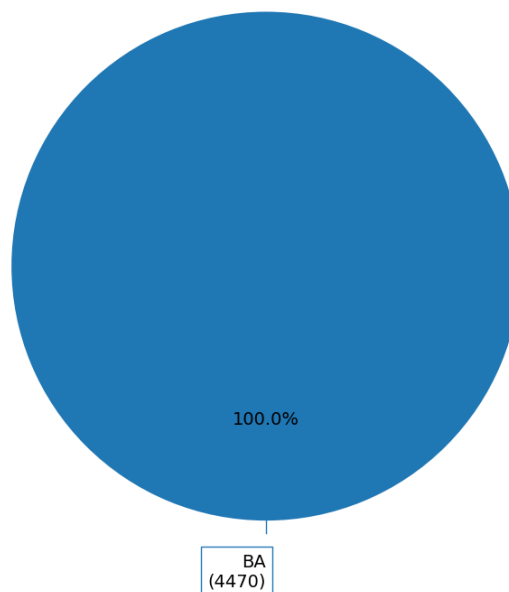
Obrázek B.21: Podíl na trhu, linka LHR-TFS,
časový interval: 10 dní, počet letů



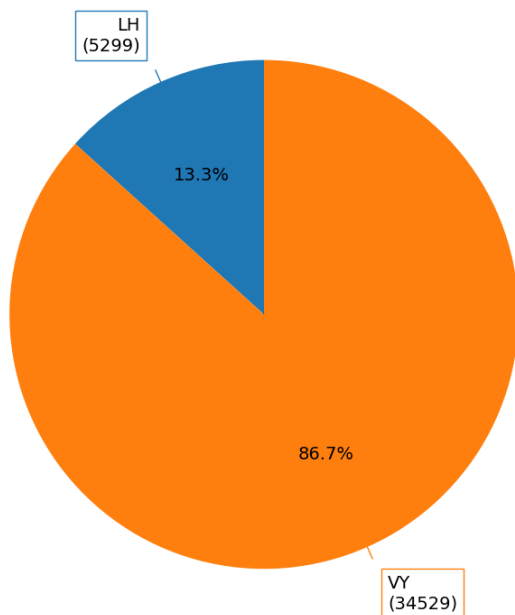
Obrázek B.22: Podíl na trhu, linka LHR-TFS,
časový interval: 2 týdny, počet letů



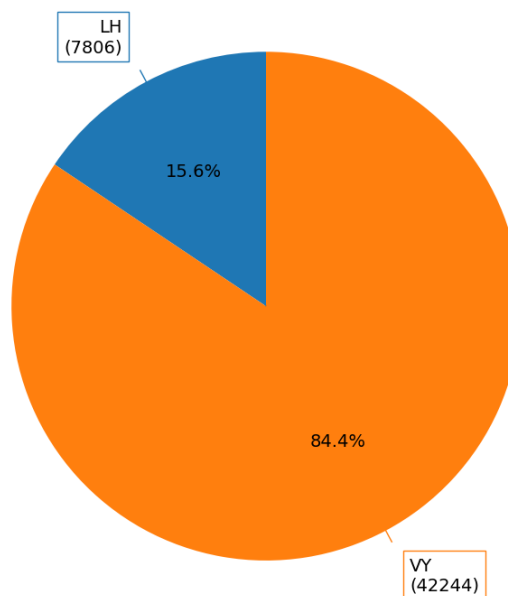
Obrázek B.23: Podíl na trhu, linka LHR-TFS,
časový interval: týden, počet letů



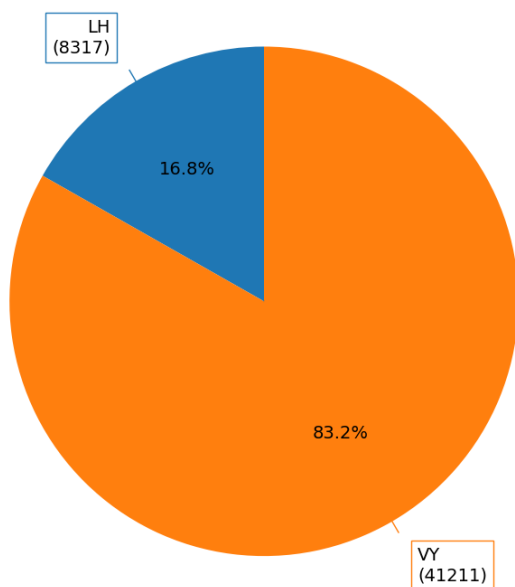
Obrázek B.24: Podíl na trhu, linka LHR-TFS,
časový interval: víkend, počet letů



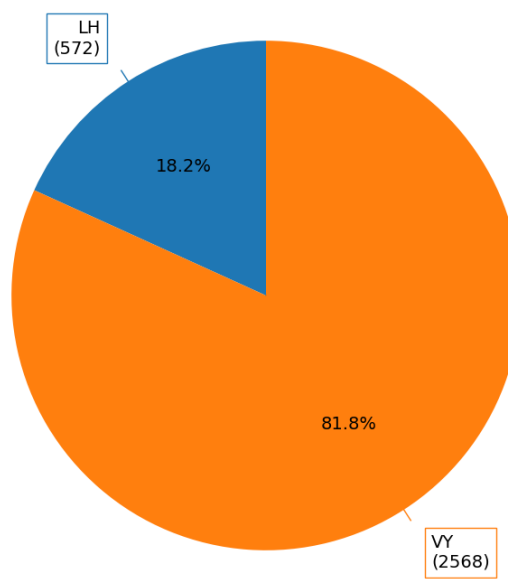
Obrázek B.25: Podíl na trhu, linka MUC-BCN, časový interval: 10 dní, počet letů



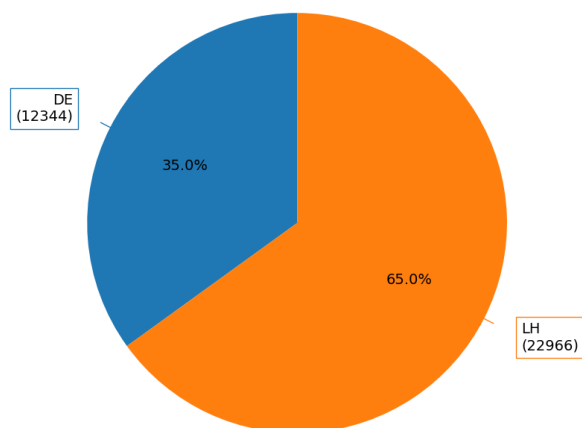
Obrázek B.26: Podíl na trhu, linka MUC-BCN, časový interval: 2 týdny, počet letů



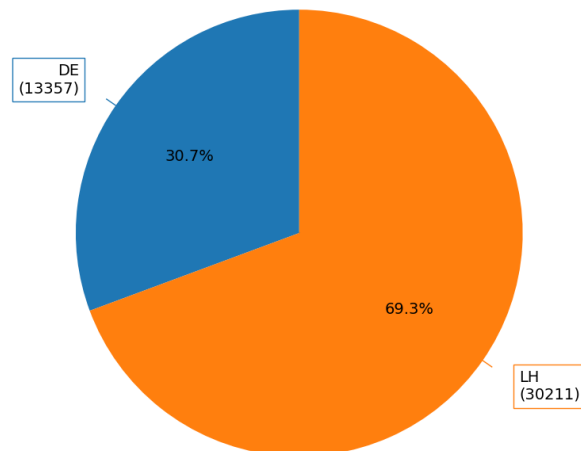
Obrázek B.27: Podíl na trhu, linka MUC-BCN, časový interval: týden, počet letů



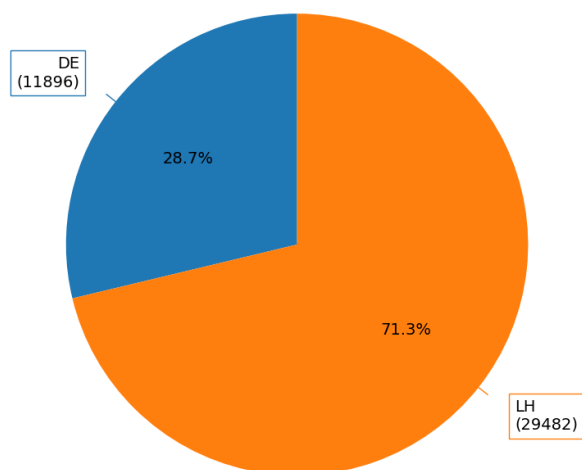
Obrázek B.28: Podíl na trhu, linka MUC-BCN, časový interval: víkend, počet letů



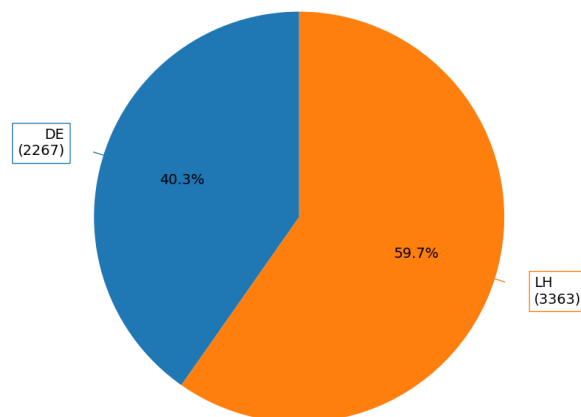
Obrázek B.29: Podíl na trhu, linka MUC-IBZ, časový interval: 10 dní, počet letů



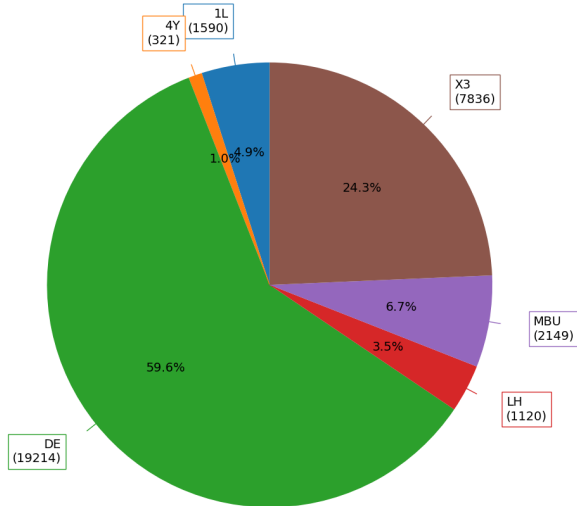
Obrázek B.30: Podíl na trhu, linka MUC-IBZ, časový interval: 2 týdny, počet letů



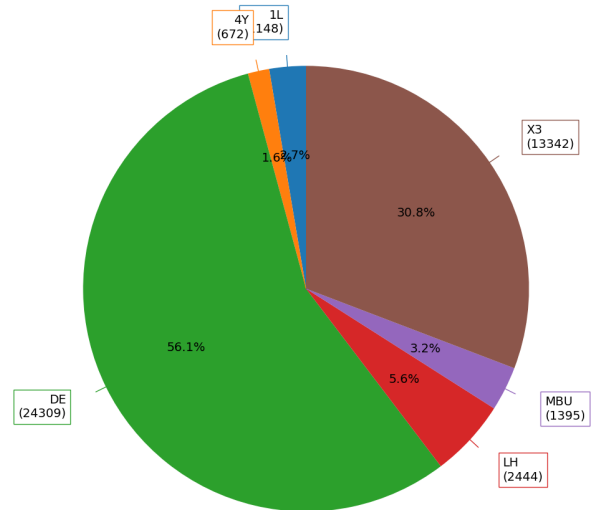
Obrázek B.31: Podíl na trhu, linka MUC-IBZ, časový interval: týden, počet letů



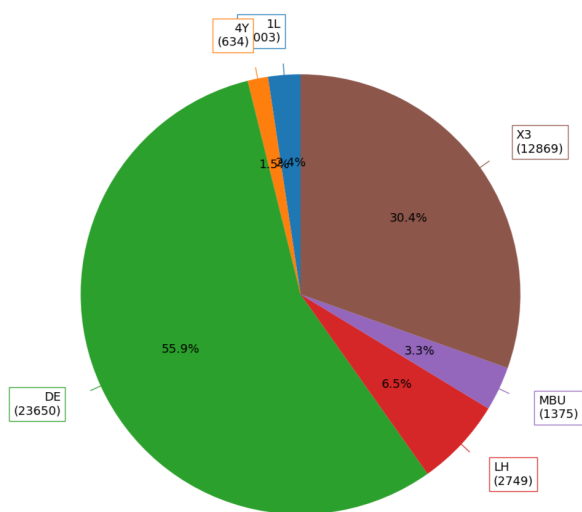
Obrázek B.32: Podíl na trhu, linka MUC-IBZ, časový interval: víkend, počet letů



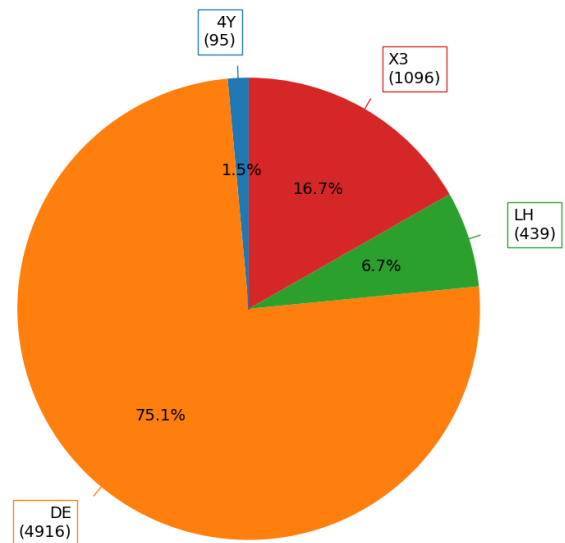
Obrázek B.33: Podíl na trhu, linka MUC-TFS, časový interval: 10 dní, počet letů



Obrázek B.34: Podíl na trhu, linka MUC-TFS, časový interval: 2 týdny, počet letů

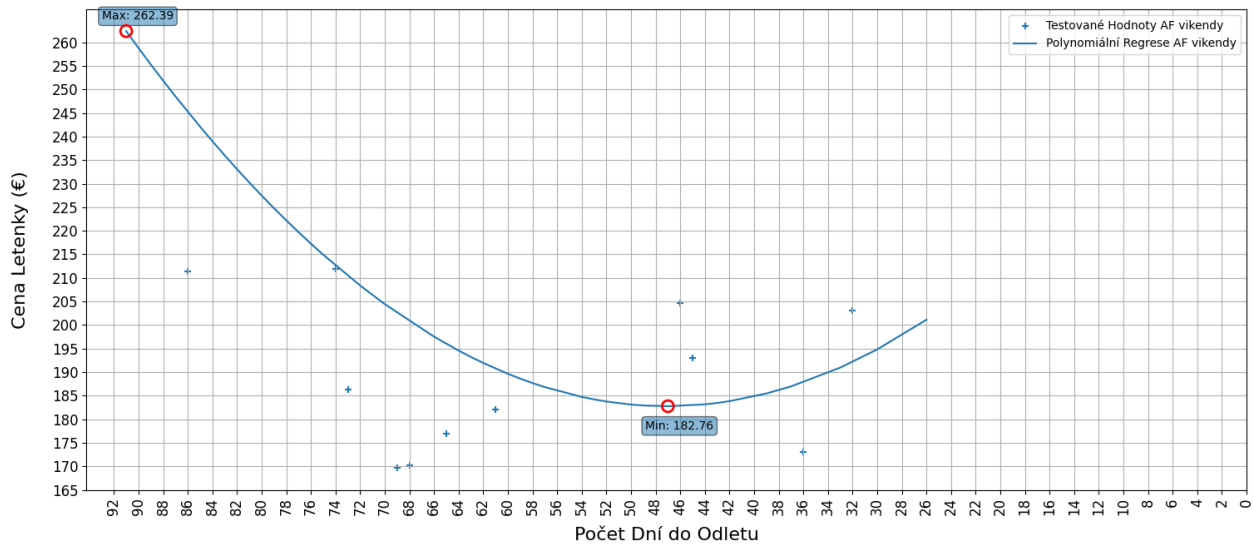


Obrázek B.35: Podíl na trhu, linka MUC-TFS, časový interval: týden, počet letů

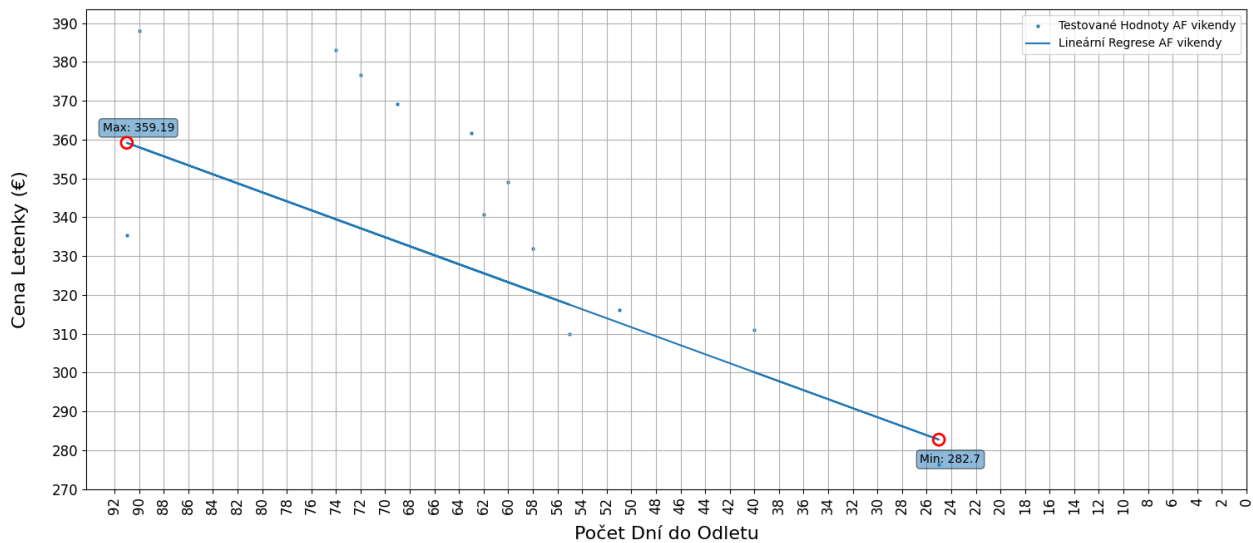


Obrázek B.36: Podíl na trhu, linka MUC-TFS, časový interval: víkend, počet letů

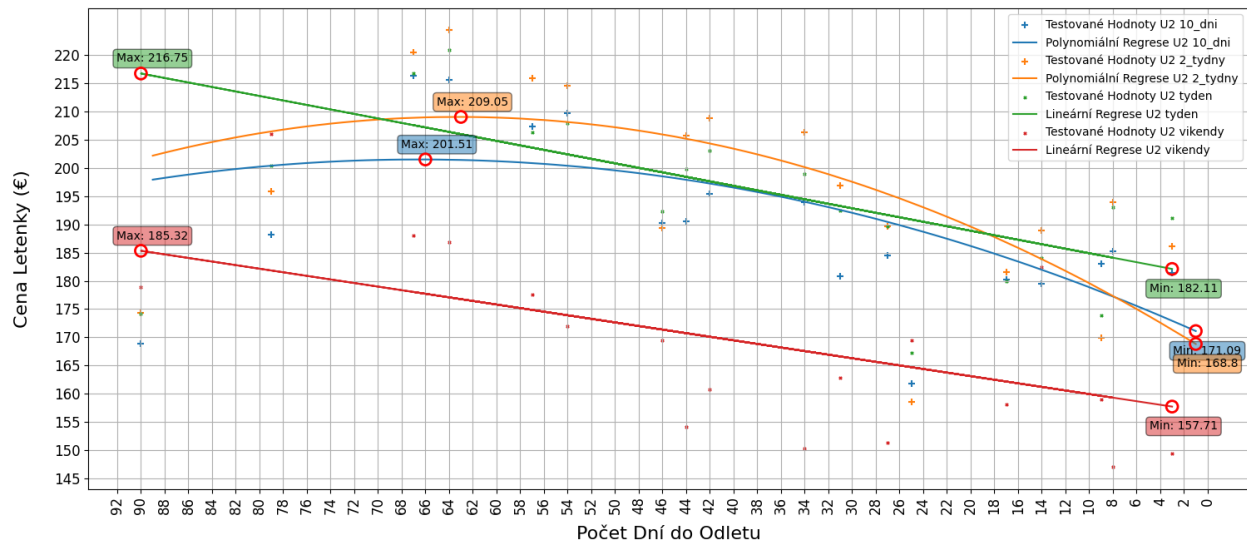
C Regresní modely



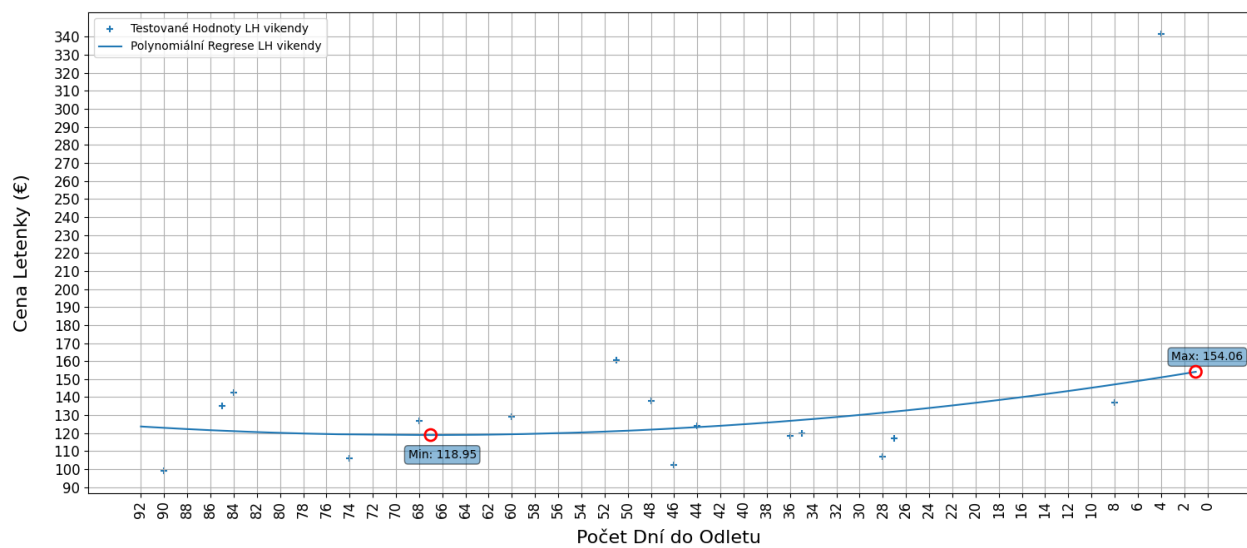
Obrázek C.1: Regresní modely na lince CDG-IBZ, aerolinka: AF



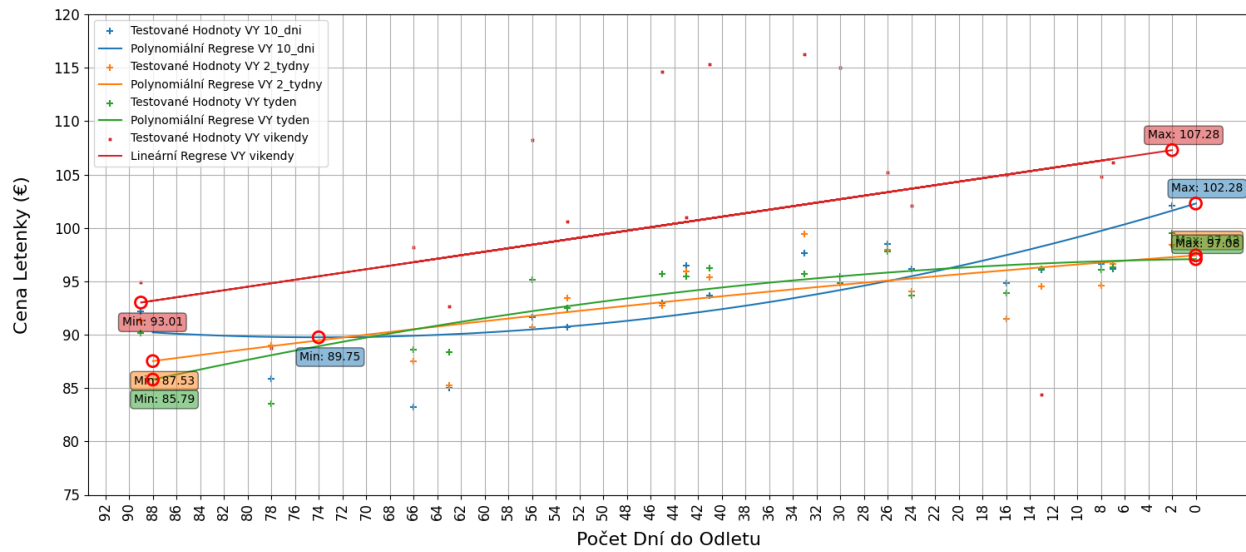
Obrázek C.2: Regresní modely na lince CDG-TFS, aerolinka: AF



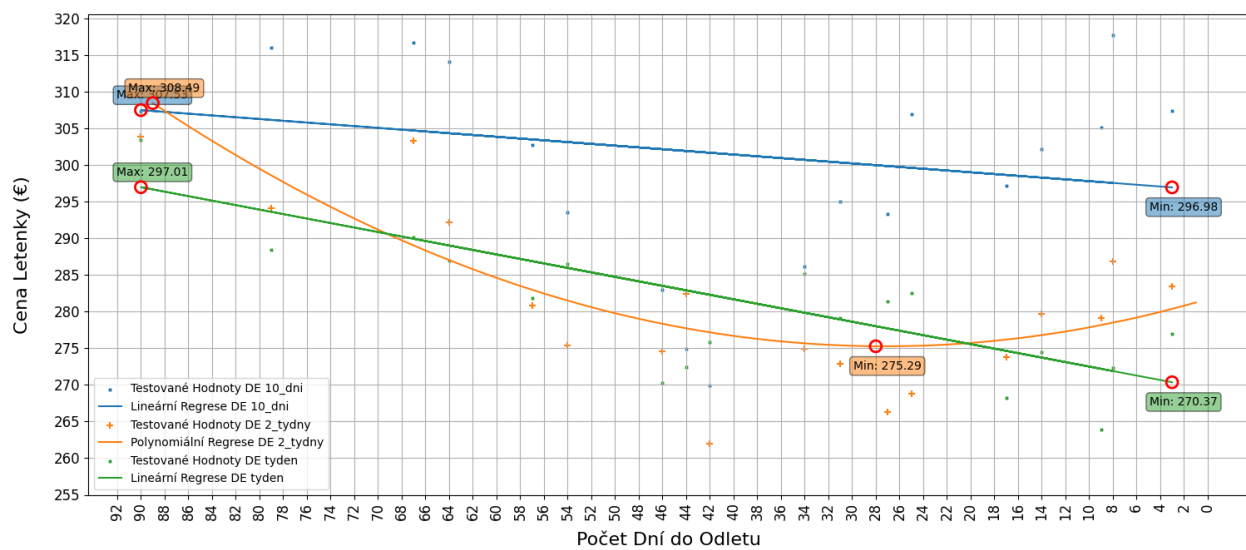
Obrázek C.3: Regresní modely na lince CDG-TFS, aerolinka: U2



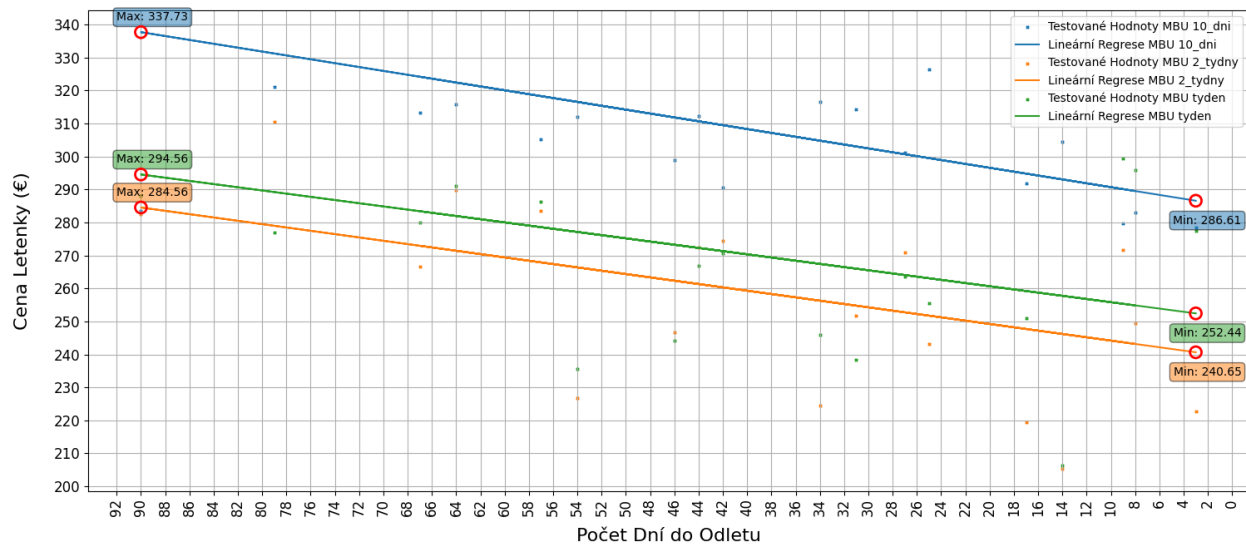
Obrázek C.4: Regresní modely na lince MUC-BCN, aerolinka: LH



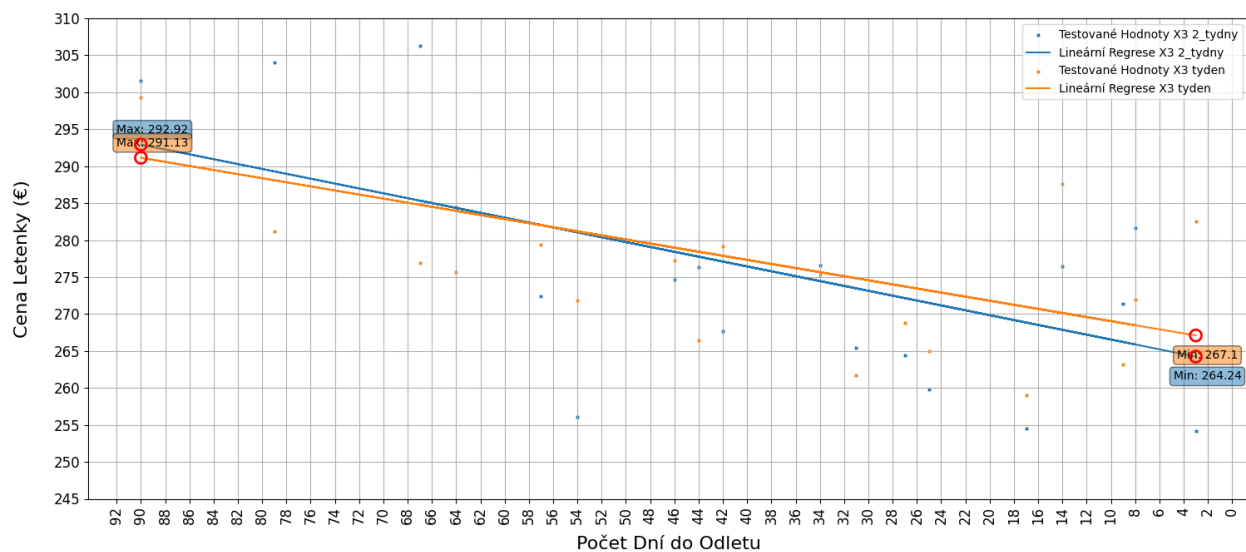
Obrázek C.5: Regresní modely na lince MUC-BCN, aerolinka: VY



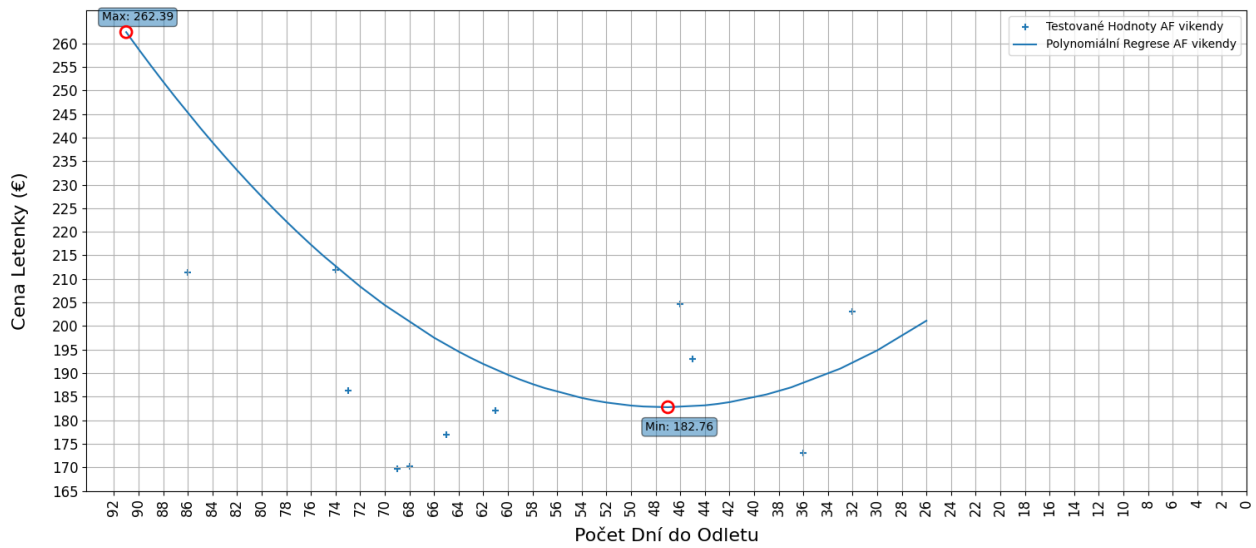
Obrázek C.6: Regresní modely na lince MUC-TFS, aerolinka: DE



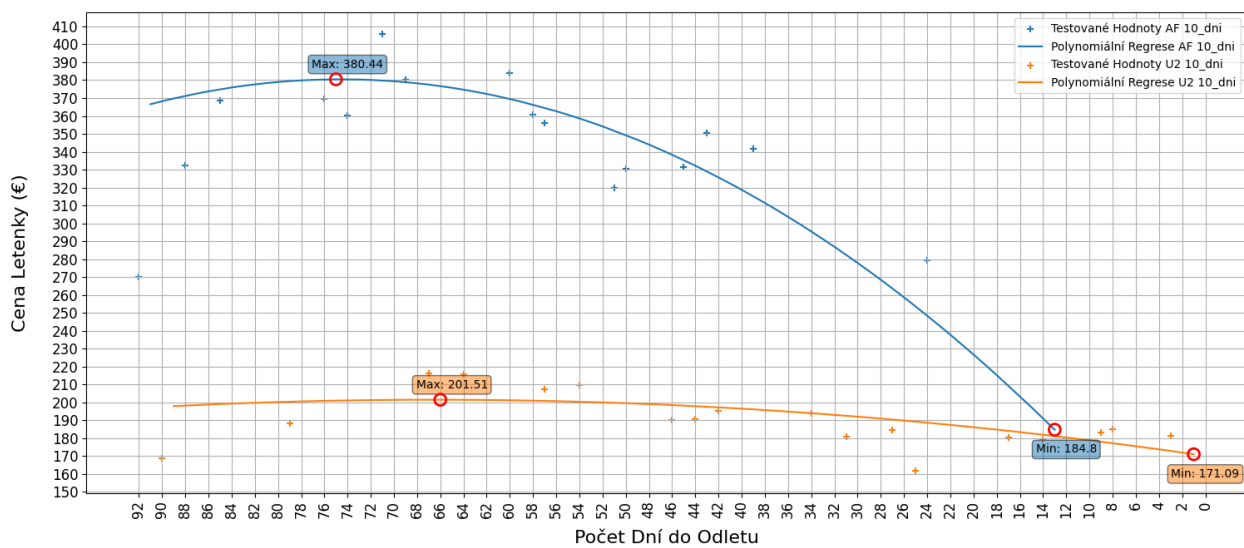
Obrázek C.7: Regresní modely na lince MUC-TFS, aerolinka: MBU



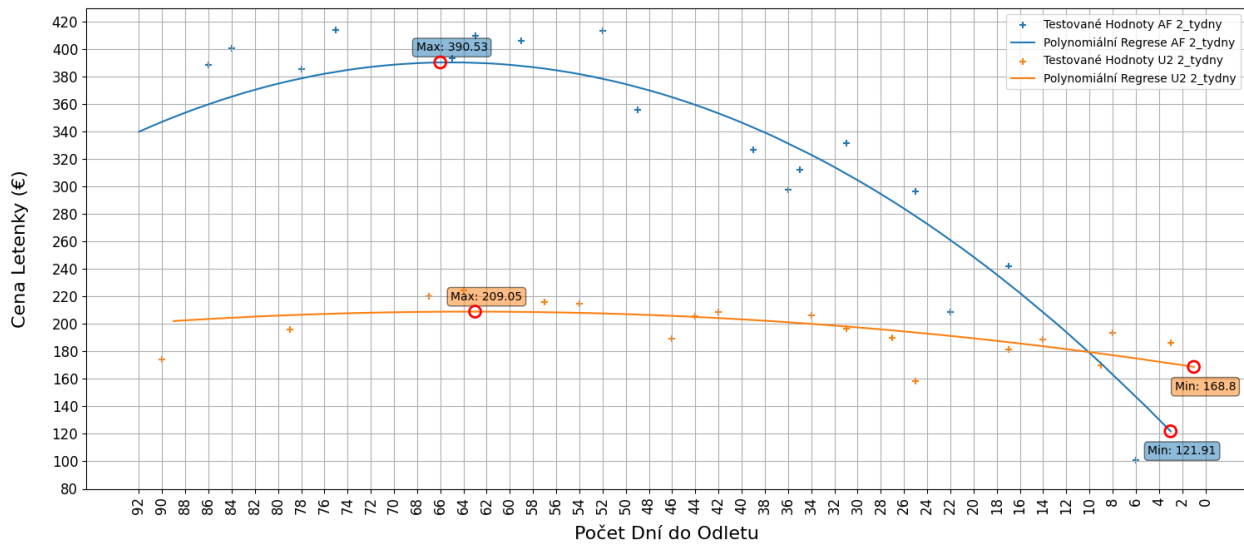
Obrázek C.8: Regresní modely na lince MUC-TFS, aerolinka: X3



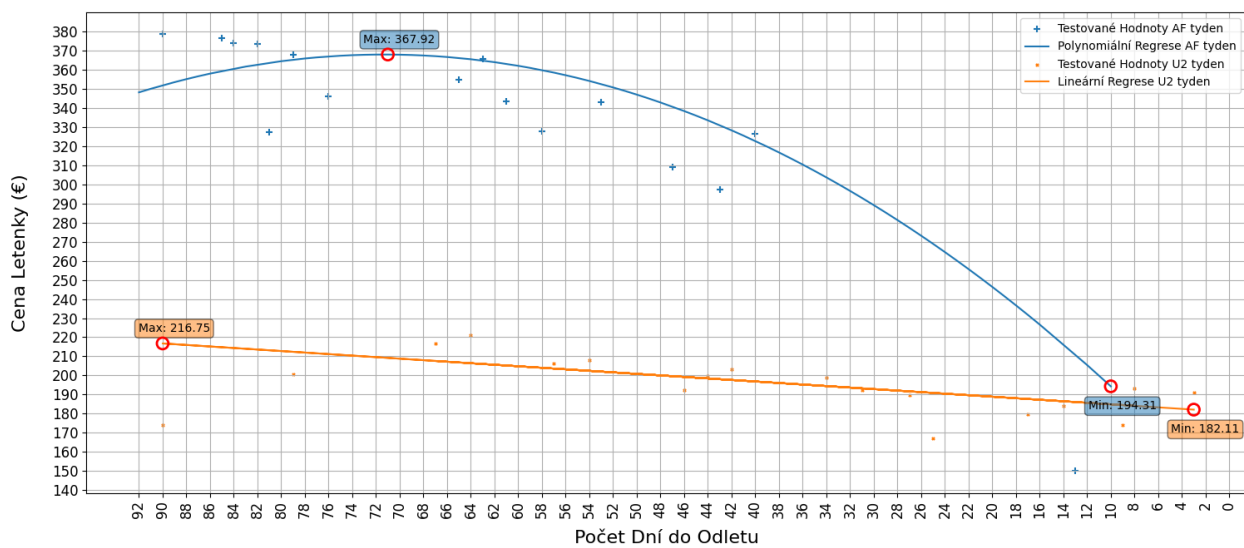
Obrázek C.9: Regresní modely na lince CDG-IBZ, časový interval: víkend



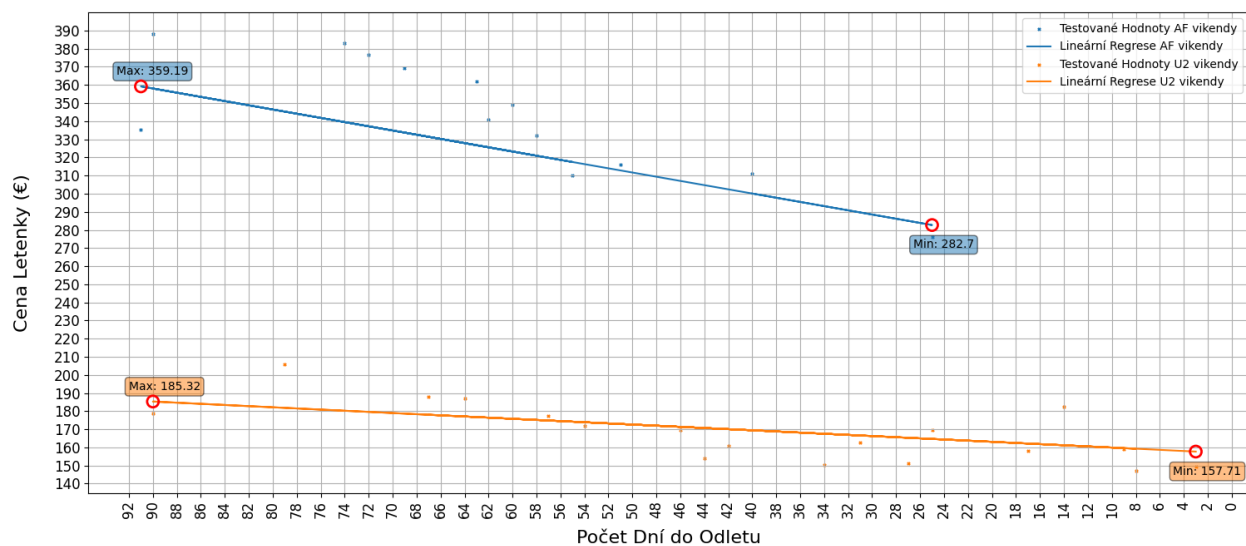
Obrázek C.10: Regresní modely na lince CDG-TFS, časový interval: 10 dní



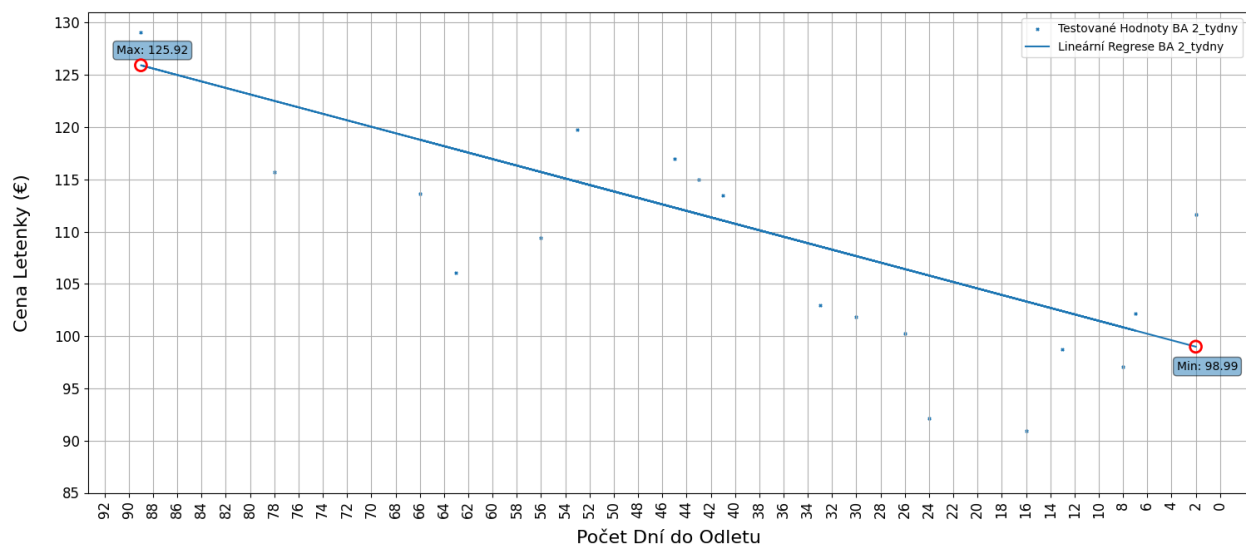
Obrázek C.11: Regresní modely na lince CDG-TFS, časový interval: 2 týdny



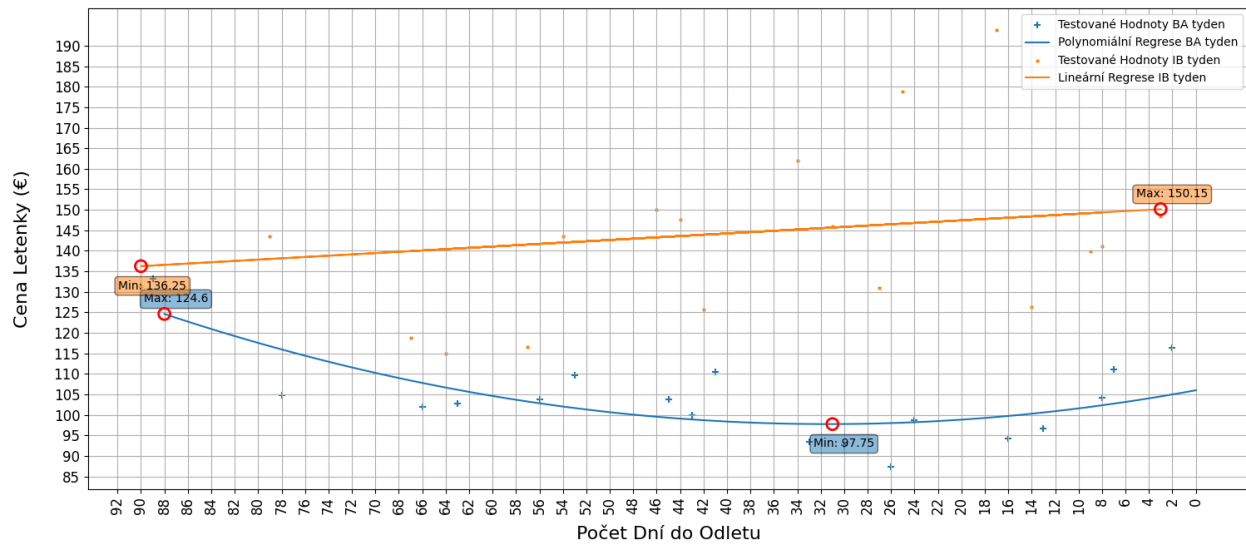
Obrázek C.12: Regresní modely na lince CDG-TFS, časový interval: týden



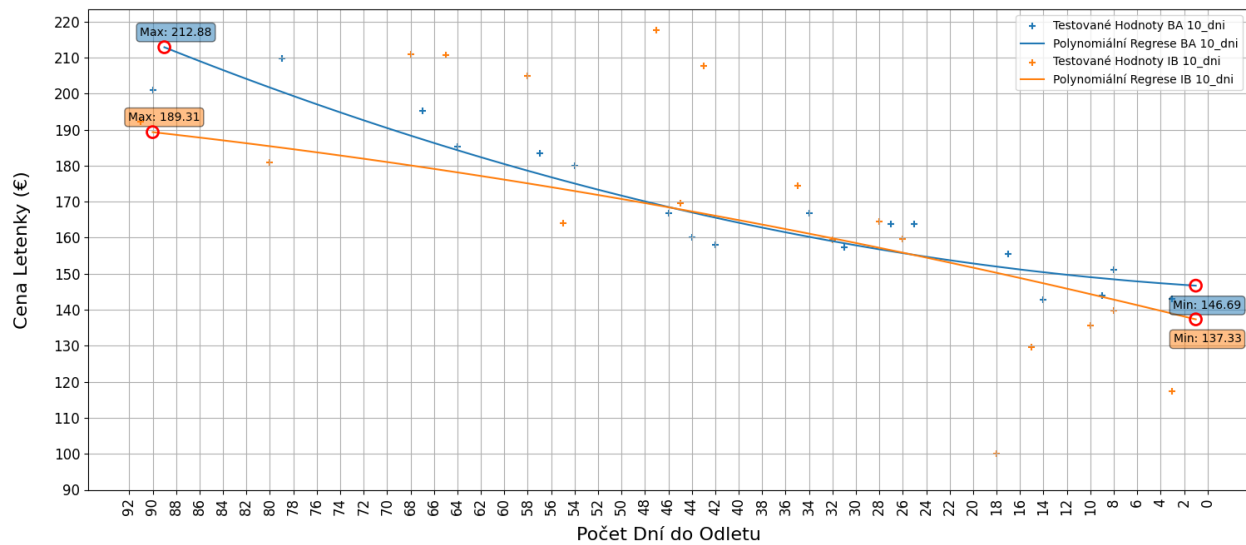
Obrázek C.13: Regresní modely na lince CDG-TFS, časový interval: víkend



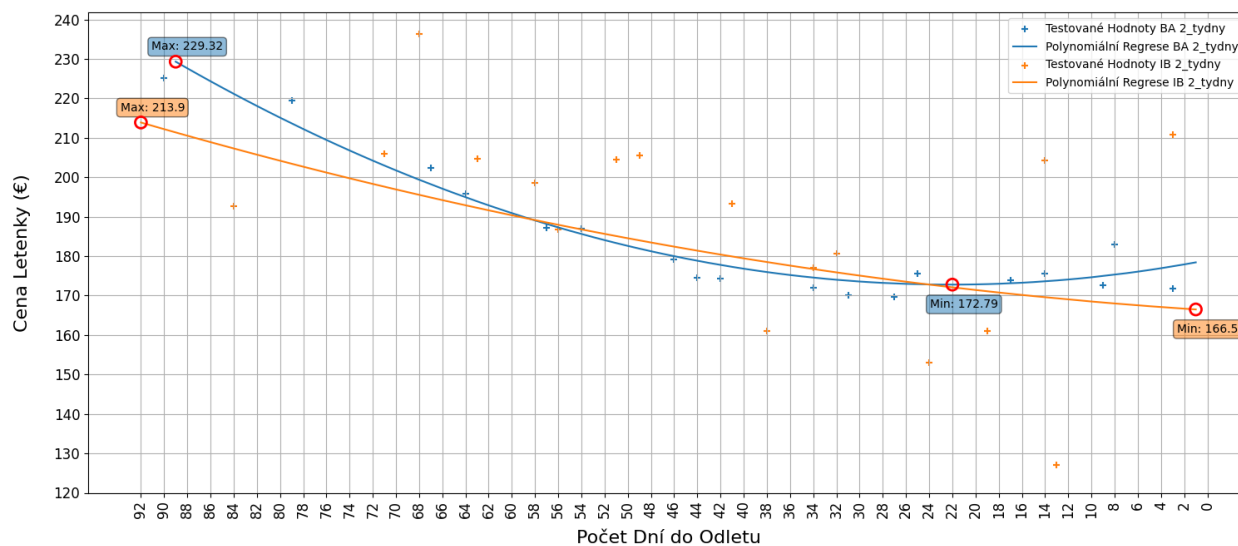
Obrázek C.14: Regresní modely na lince LHR-BCN, časový interval: 2 týdny



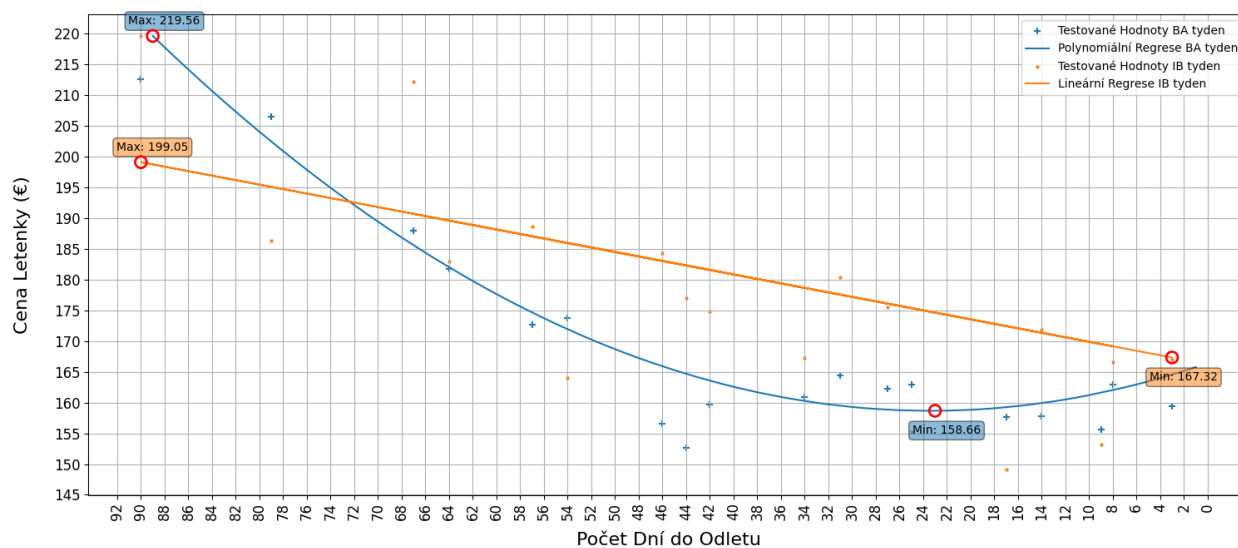
Obrázek C.15: Regresní modely na lince LHR-BCN, časový interval: týden



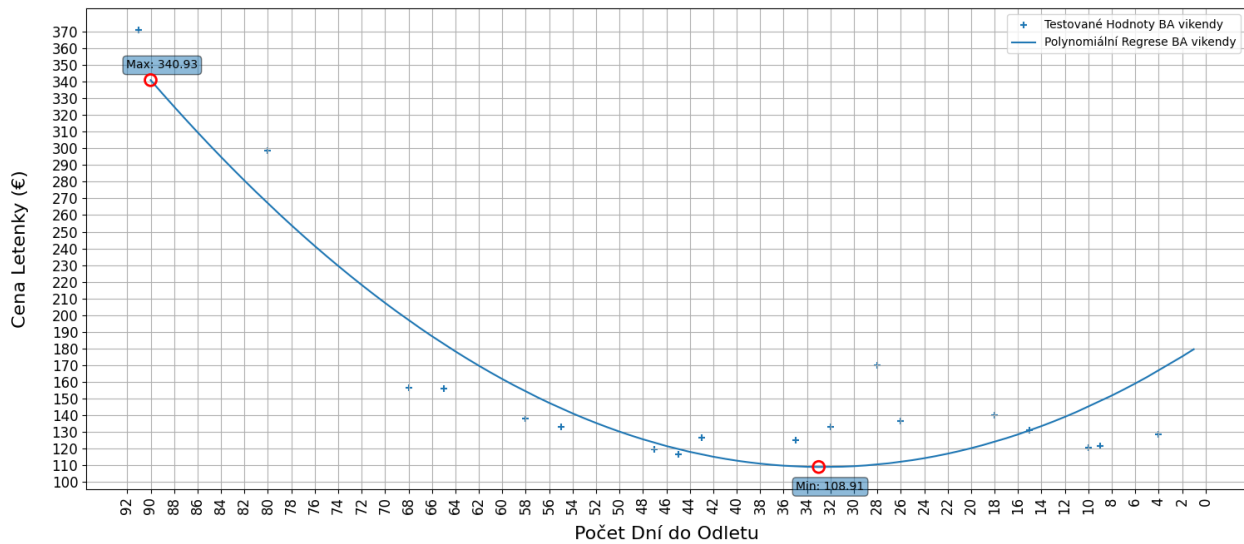
Obrázek C.16: Regresní modely na lince LHR-IBZ, časový interval: 10 dní



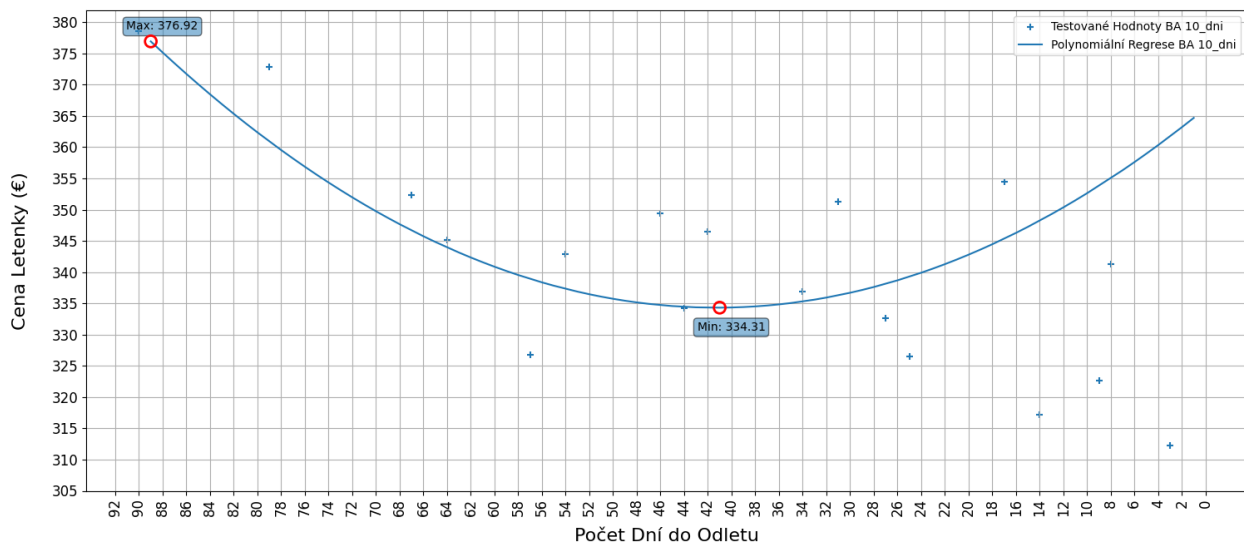
Obrázek C.17: Regresní modely na lince LHR-IBZ, časový interval: 2 týdny



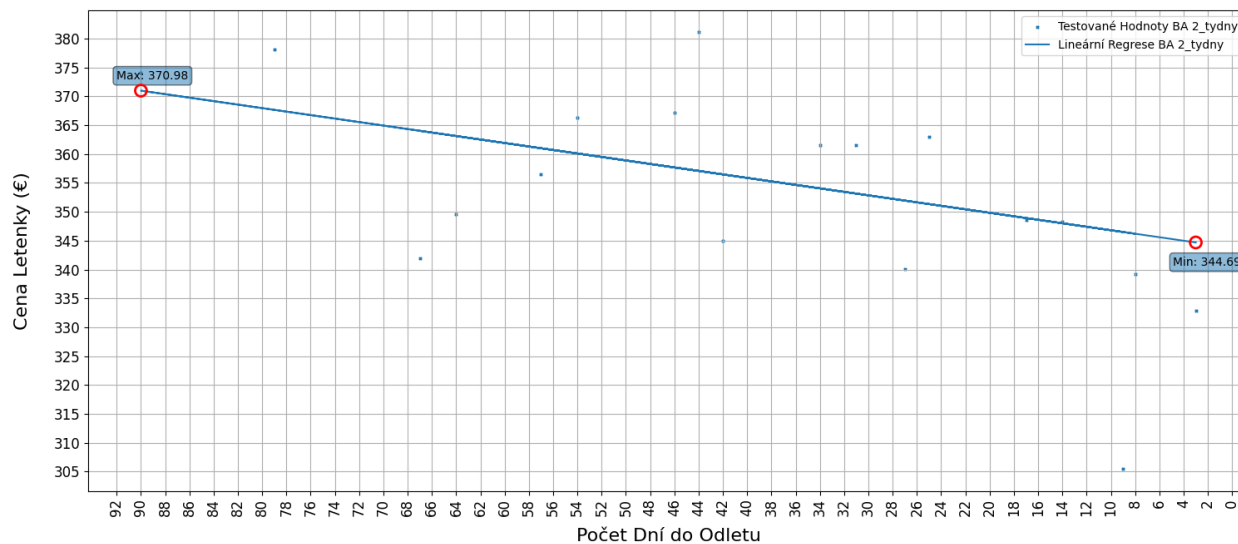
Obrázek C.18: Regresní modely na lince LHR-IBZ, časový interval: týden



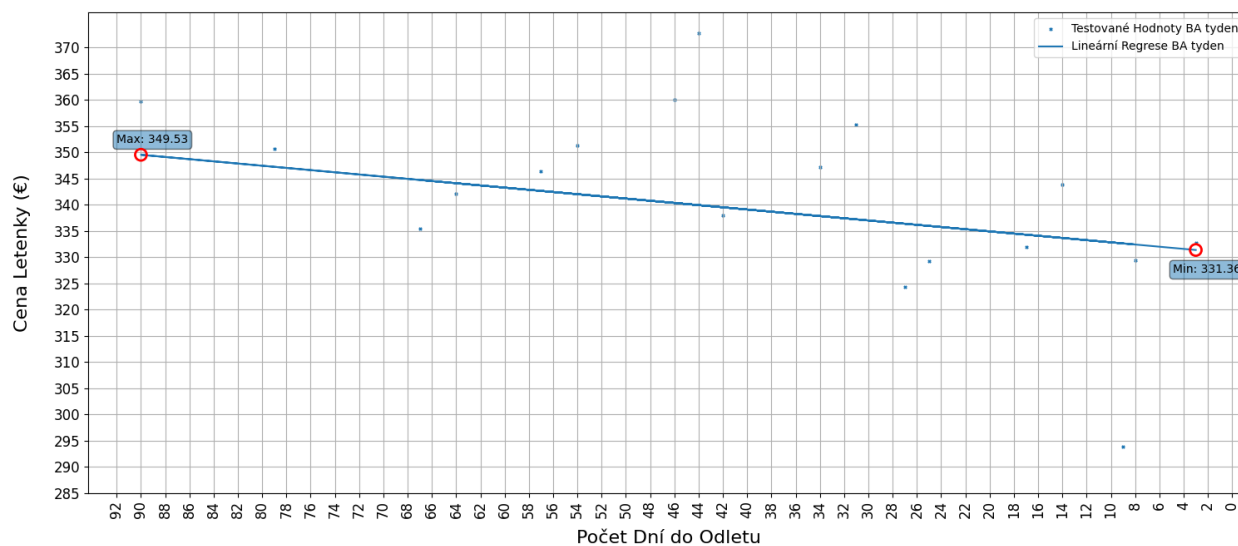
Obrázek C.19: Regresní modely na lince LHR-IBZ, časový interval: víkend



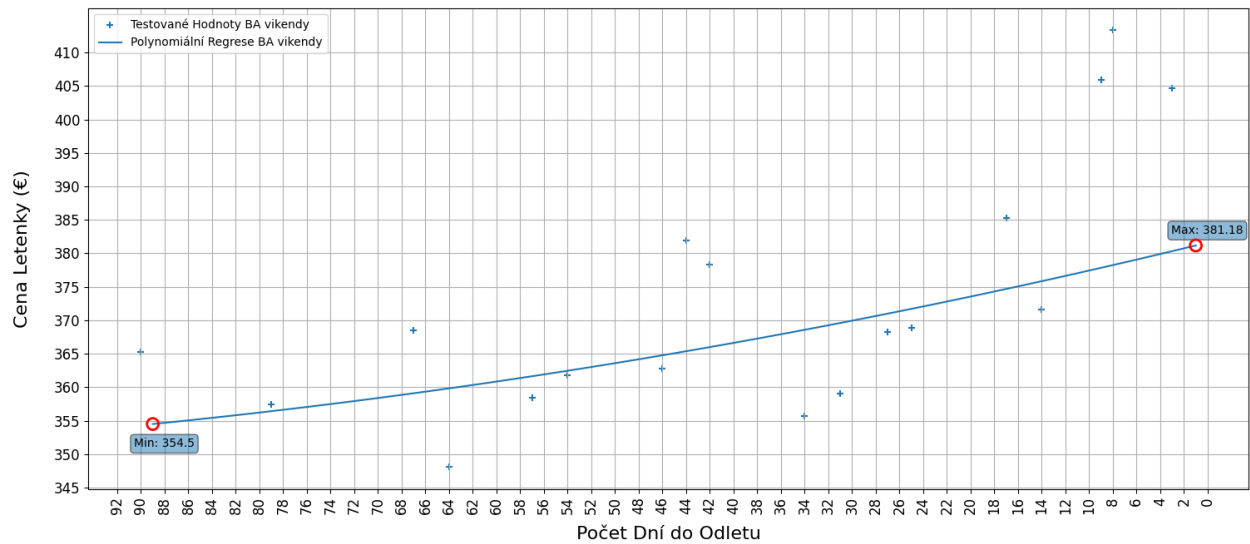
Obrázek C.20: Regresní modely na lince LHR-TFS, časový interval: 10 dní



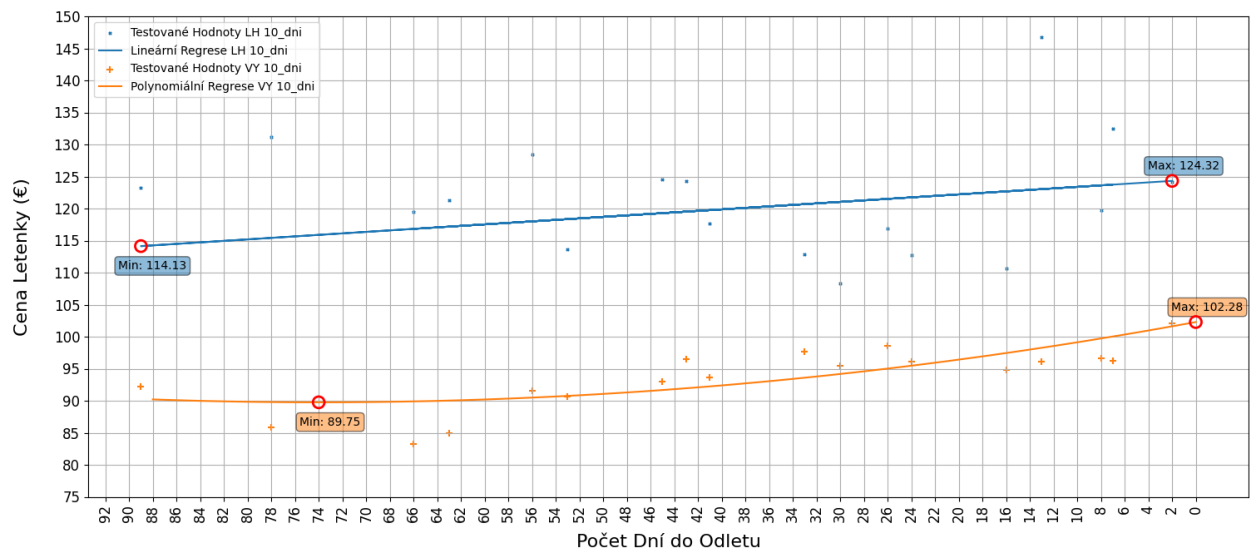
Obrázek C.21: Regresní modely na lince LHR-TFS, časový interval: 2 týdny



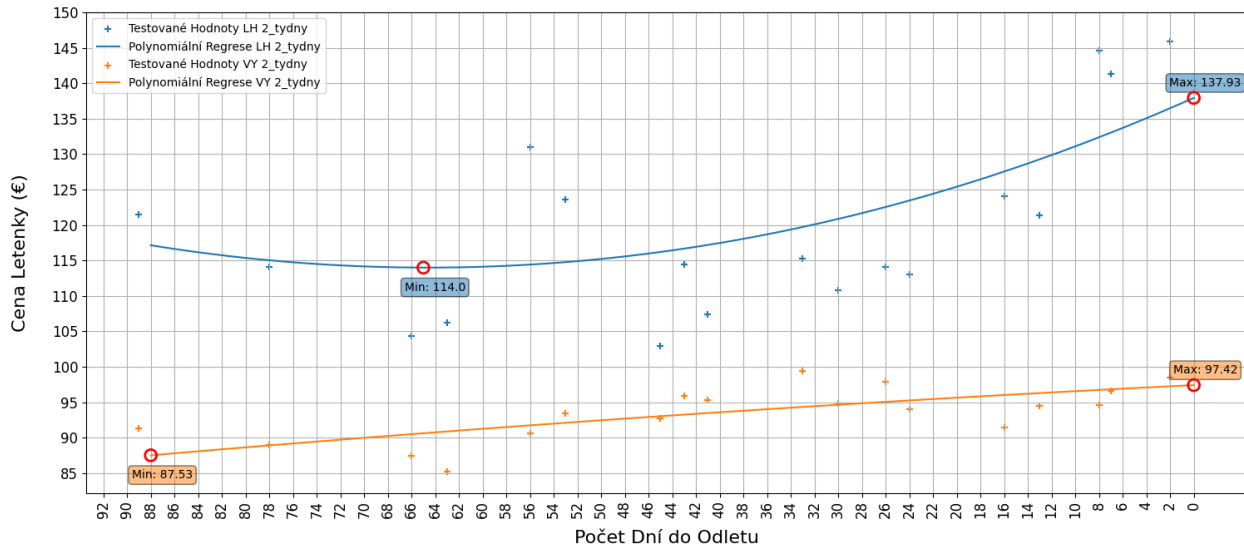
Obrázek C.22: Regresní modely na lince LHR-TFS, časový interval: týden



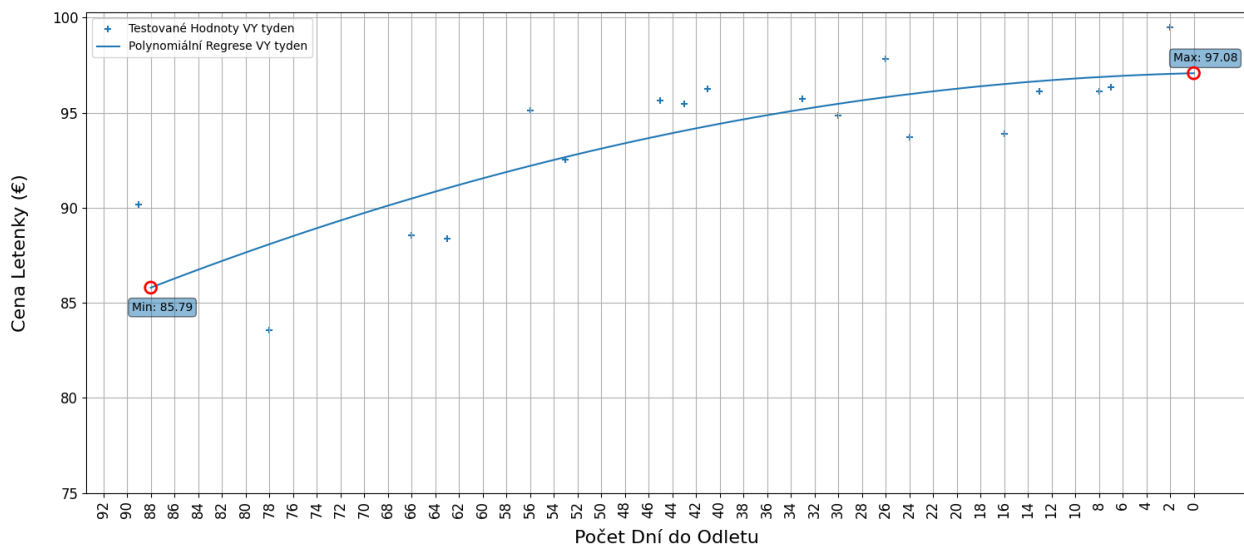
Obrázek C.23: Regresní modely na lince LHR-TFS, časový interval: víkend



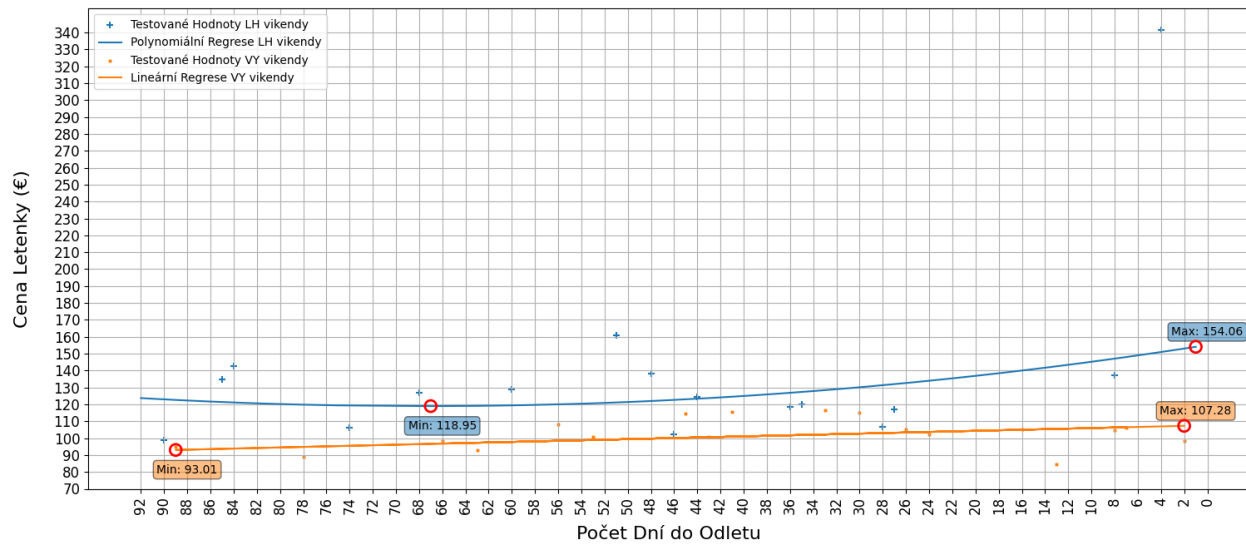
Obrázek C.24: Regresní modely na lince MUC-BCN, časový interval: 10 dní



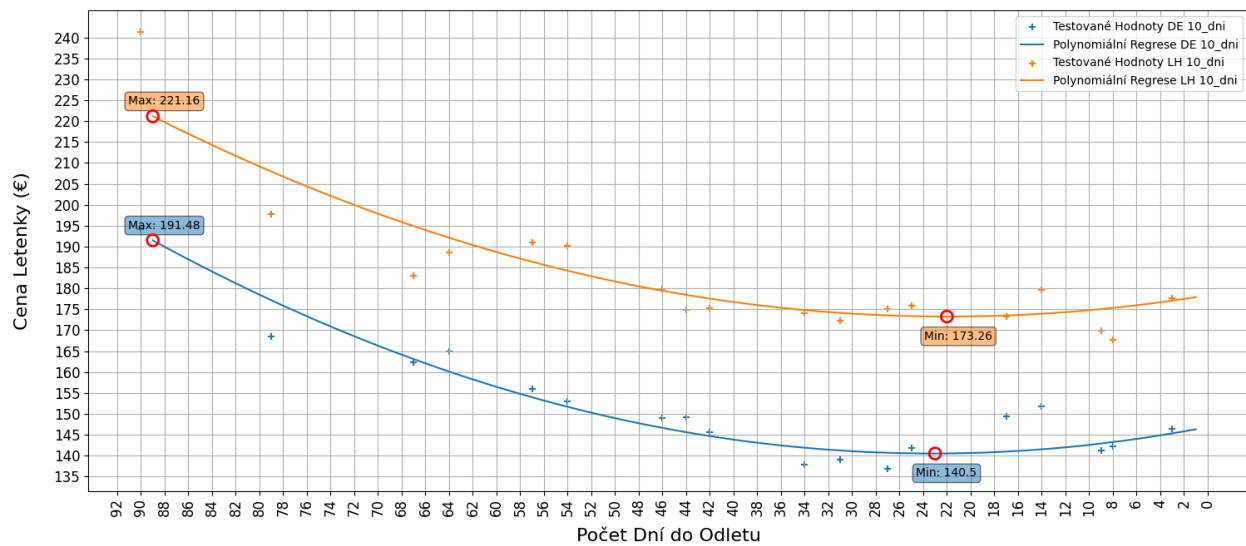
Obrázek C.25: Regresní modely na lince MUC-BCN, časový interval: 2 týdny



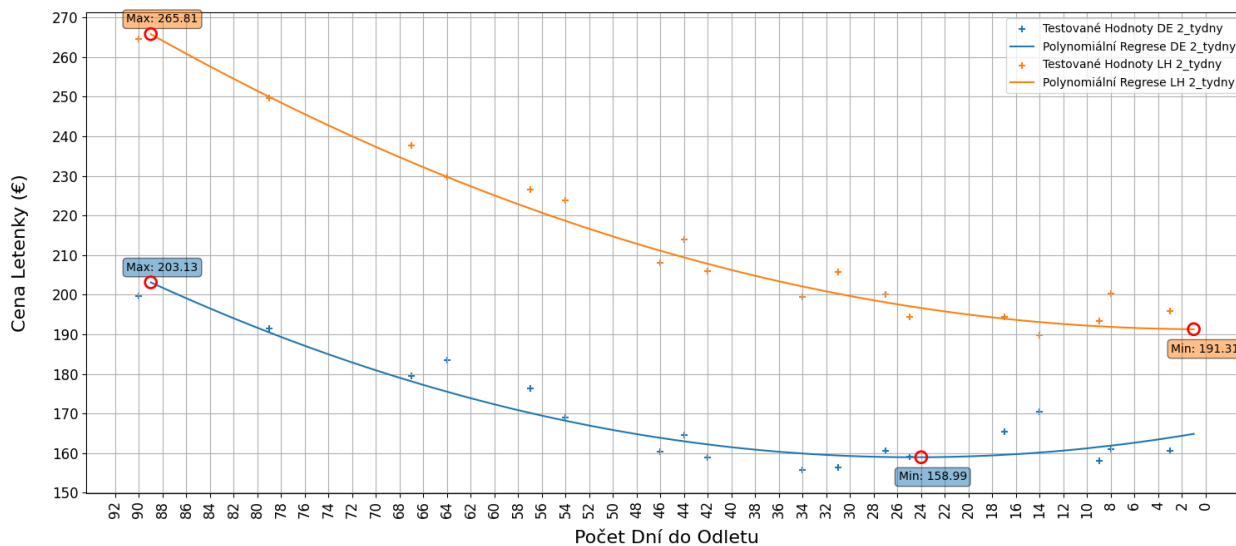
Obrázek C.26: Regresní modely na lince MUC-BCN, časový interval: týden



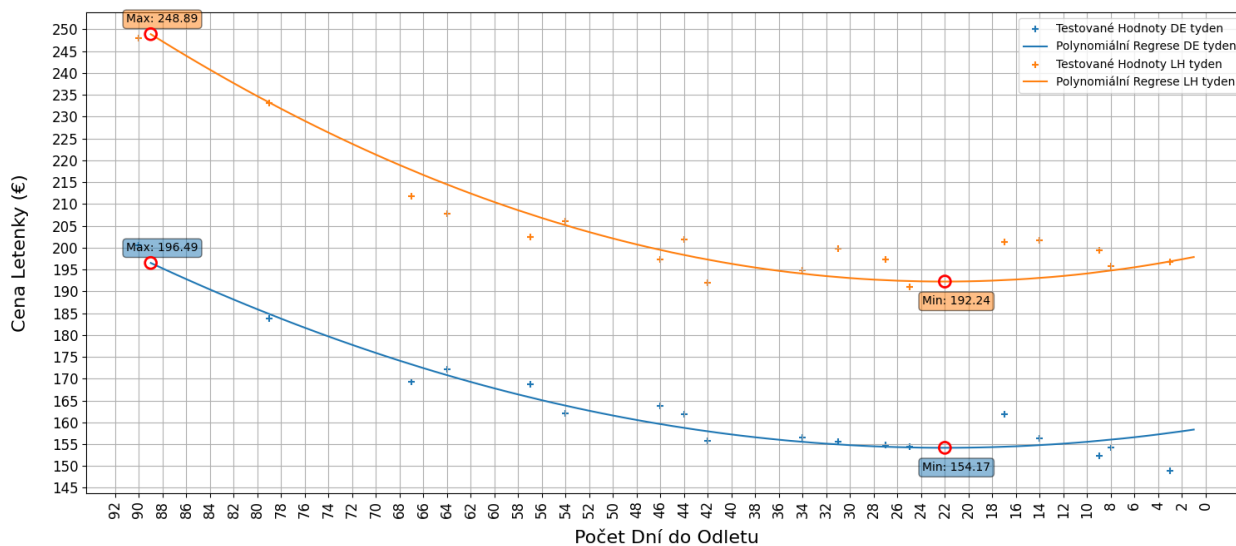
Obrázek C.27: Regresní modely na lince MUC-BCN, časový interval: víkend



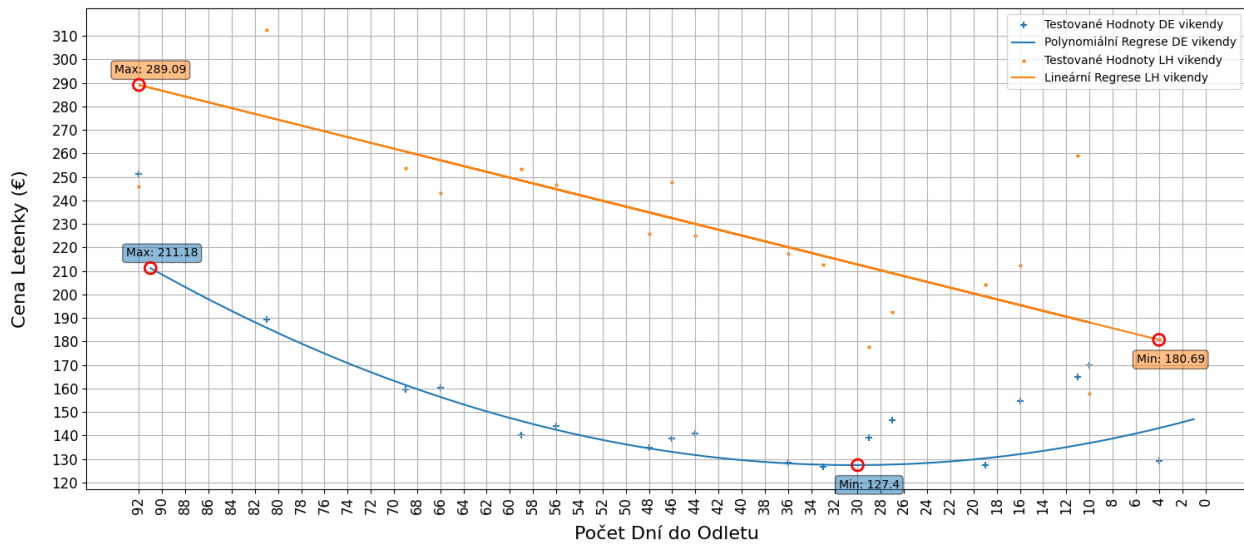
Obrázek C.28: Regresní modely na lince MUC-IBZ, časový interval: 10 dní



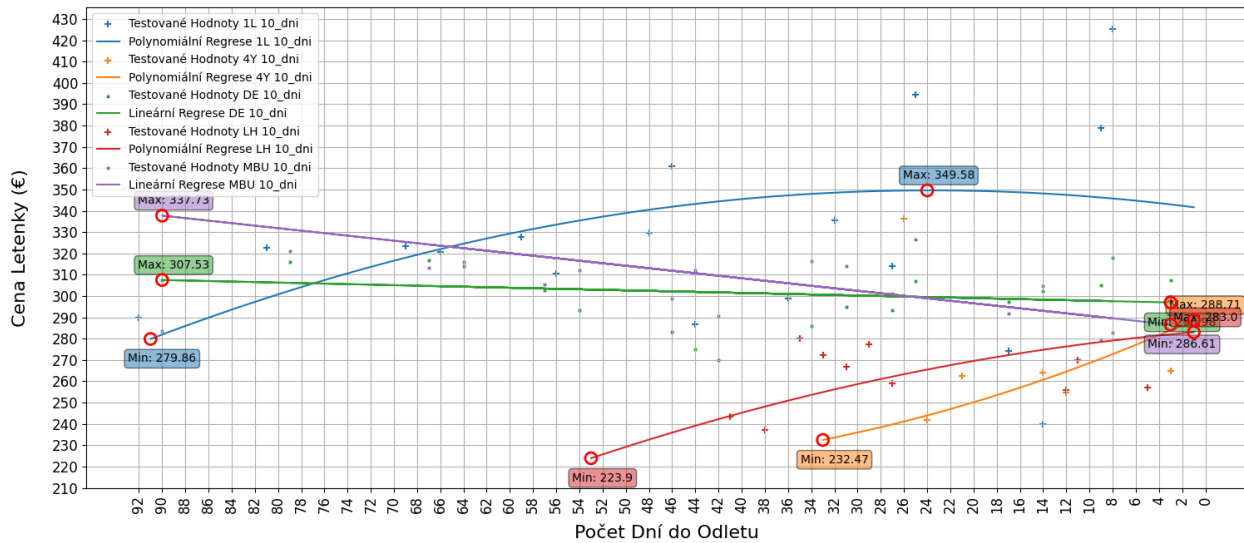
Obrázek C.29: Regresní modely na lince MUC-IBZ, časový interval: 2 týdny



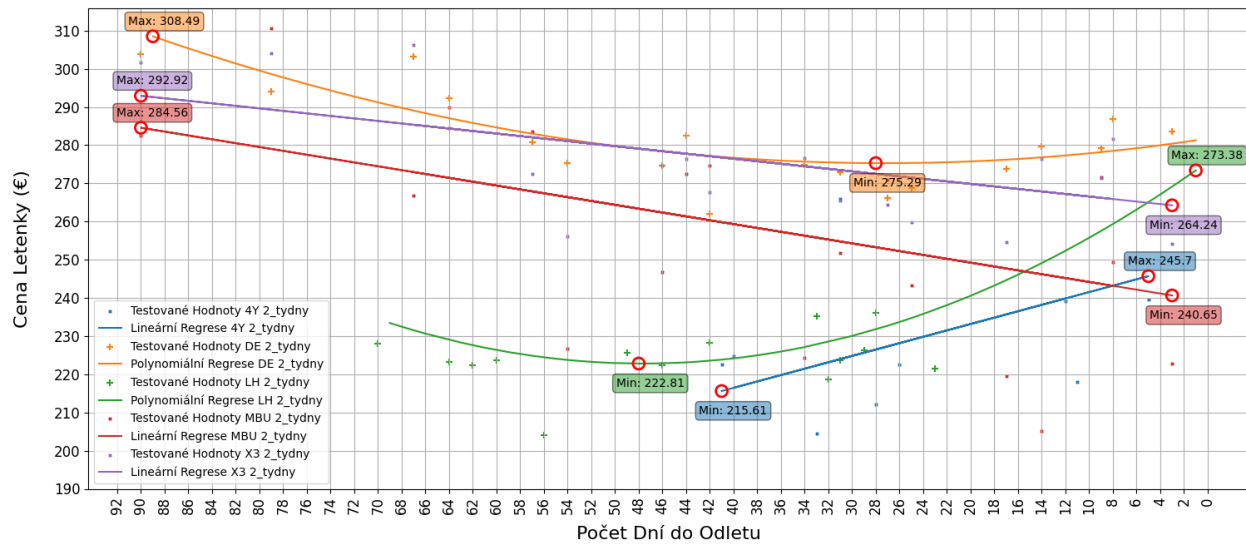
Obrázek C.30: Regresní modely na lince MUC-IBZ, časový interval: týden



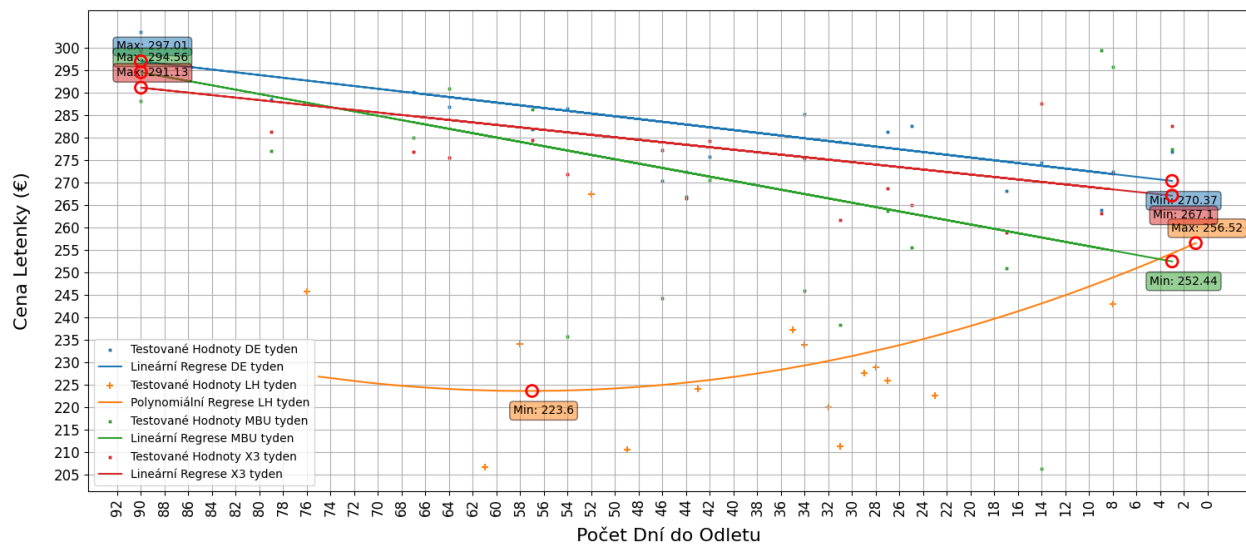
Obrázek C.31: Regresní modely na lince MUC-IBZ, časový interval: víkend



Obrázek C.32: Regresní modely na lince MUC-TFS, časový interval: 10 dní



Obrázek C.33: Regresní modely na lince MUC-TFS, časový interval: 2 týdny



Obrázek C.34: Regresní modely na lince MUC-TFS, časový interval: týden