UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MODELOVANIE ANGAŽOVANOSTI NA SOCIÁLNYCH SIEŤACH FIRIEM FINANČNÉHO SEKTORA

Diplomová práca

2024

Bc. Natália Gibasová

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MODELOVANIE ANGAŽOVANOSTI NA SOCIÁLNYCH SIEŤACH FIRIEM FINANČNÉHO SEKTORA

Diplomová práca

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	1113 Matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Školiteľ:	Mgr. Martin Hurban, PhD.

Bratislava, 2024

Bc. Natália Gibasová





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Natália Gibasová
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Modelovanie angažovanosti na sociálnych sieťach finančného sektora
Modelling of engagement on social networks of finance sector

Anotácia: Komunikácia na sociálnych sieťach je v súčasnosti významnou súčasťou marketingových procesov. Rozhodovanie o štruktúre, spôsobe a čase komunikácie na sociálnych sieťach podložené analýzou správania na jednotlivých sociálnych sieťach je dôležité pre vhodné nastavenie marketingových procesov.

Cieľ: Identifikácia rozdielnych komunikačných stratégii bankového sektora na sociálnych sieťach. Modelovať angažovanosť publika v závislosti od typu komunikácie na sociálnych sieťach. Definovať odporúčania na zlepšenia komunikácie na sociálnych sieťach.

Vedúci: Mgr. Martin Hurban, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.

Dátum zadania: 10.01.2023

Dátum schválenia: 11.01.2023 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Poďakovanie: Veľmi rada by som sa chcela poďakovať môjmu školiteľovi Mgr. Martinovi Hurbanovi, PhD. za všetok čas, ktorý mi venoval, za cenné rady, ktoré mi dal a za vedomosti, ktoré som vďaka nemu získala. Taktiež by som sa chcela poďakovať mojej kamarátke Sise, ktorá mi venovala svoj čas pri finálnom kontrolovaní práce.

Abstrakt

GIBASOVÁ, Natália: Modelovanie angažovanosti na sociálnych sieťach firiem finančného sektora [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; vedúci práce: Mgr. Martin Hurban, PhD., 2024, 109 s.

Táto práca sa zameriava na identifikáciu a vyhodnotenie vplyvu rozdielnych komunikačných stratégií bankového sektora na sociálnych sieťach. Cieľom je modelovať angažovanosť publika v závislosti od typu komunikácie na týchto platformách a definovať odporúčania na zlepšenie komunikácie.

Na základe výsledkov sme identifikovali vplyv rôznych faktorov na počet fanúšikov na sociálnych sieťach, pričom uverejňovanie finančných súťaží ich priláka najviac. Taktiež sme analyzovali, čo a ako vplýva na množstvo interakcií s príspevkami. V tomto prípade publikovanie príspevkov počas sviatkov, venovanie sa témam vplyv makroekonomických faktorov na finančné trhy alebo dôležitosť cestovného poistenia pri cestovaní vyzerajú pre ľudí zaujímavo. Práca poskytuje prehľad o prioritách, na ktoré by sa mali banky sústrediť, aby prilákali viac fanúšikov na svoje sociálne siete a získali väčšie množstvo reakcií na ich príspevky za účelom dostať sa do väčšieho povedomia.

Kľúčové slová: sociálne siete, bankový sektor, komunikačné stratégie, angažovanosť publika, štatistické modely.

Abstract

GIBASOVÁ, Natália: Modelling of engagement on social networks of finance sector [Diploma thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Martin Hurban, PhD., 2024, 109 p.

This thesis focuses on identifying and evaluating the impact of different communication strategies used by the banking sector on social media. The aim is to model audience engagement depending on the type of communication on these platforms and to define recommendations for improving communication.

Based on the results, we have identified the impact of various factors on the number of fans on social networks, with the publication of financial competitions attracting the most. We have also analyzed what and how it affects the number of interactions with posts. In this case, publishing posts during holidays, dealing with topics such as the impact of macroeconomic factors on financial markets or the importance of travel insurance when traveling, seem interesting to people. The thesis provides an overview of the priorities that banks should focus on to attract more fans to their social networks and gain more reactions to their posts in order to increase awareness.

Keywords: social networks, banking sector, communication strategies, audience engagement, statistical models.

Obsah

Úvod **11**

1 Deskriptívna analýza dát na sociálnych sieťach **13**

1.1 Fanúšikovská základňa 13

1.2 Príspevky na sociálnych sieťach 16

1.2.1 Reakcie na príspevky 17

1.2.2 Publikované videá 19

1.2.3 Súťažné príspevky 22

2 Modelovanie prírastku fanúšikovskej základne na sociálnych sieťach **25**

2.1 Príprava dát 26

2.1.1 Prehľad výkonnosti profilov 26

2.1.2 Uverejnené súťaže 27

2.1.3 Ekonomická situácia na Slovensku 27

2.1.4 Závislá premenná 28

2.2 Modelovacie techniky 29

2.2.1 Fixed effects model 30

2.2.2 LDV model 31

2.2.3 Prehľad metrík kvality modelu 32

2.3 Výsledky modelov 33

2.3.1 Fixed effects model 33

2.3.2 LDV model 38

2.3.3 Porovnanie modelov 44

3 Modelovanie interakcií s príspevkami **45**

3.1 Príprava dát 45
3.1.1 Konštrukcia závislej premennej 45
3.1.2 Nezávislé premenné 46
3.2 Modelovacie techniky 48
3.2.1 SlovakBERT model 48
3.2.2 Analýza hlavných komponentov 49
3.2.3 K - means 50
3.2.4 DBScan 51
3.2.5 GAM 52
3.3 Vektorová reprezentácia textu 54
3.3.1 Použitie SlovakBERT modelu 54
3.3.2 Redukcia dimenzie dát 57
3.4 Zhlukovanie tém 58
3.4.1 Použitie metódy DBScan 58
3.4.2 Použitie metódy K–means 60
3.4.3 Výber zhlukovacej metódy 62
3.5 Trénovanie GAM modelu 63
3.6 Výsledky 69
3.6.1 Spojité premenné 71
3.6.2 Kategorické premenné 74
Záver **77**
Modelovanie prírastku fanúšikovskej základne na sociálnych sieťach 77
Modelovanie interakcií s príspevkami 78

Zhrnutie 78

Zoznam použitej literatúry **81**

Príloha A - Predpoklady lineárnej regresie **83**

Príloha B - Závislosť kategórií príspevkov od interakcií **87**

Príloha C - Zoznam všetkých tém, ktorým sa firmy zaoberajú **97**

Úvod

V dnešnom dynamickom svete je komunikácia na sociálnych sieťach neoddeliteľnou súčasťou strategických marketingových procesov, a preto je dôležité rozhodovať o štruktúre a spôsobe komunikácie na základe analýzy správania na konkrétnych sociálnych sieťach. Sociálne siete pomáhajú značkám zvýšiť ich povedomie, priamo sa spájať s ich zákazníkmi, budovať si s nimi vzťah a získavať od nich spätnú väzbu [10]. S nástupom digitálnej éry sa finančný sektor vydal cestou interakcie s verejnosťou prostredníctvom rôznych sociálnych platforiem. V práci sa budeme zameriavať na štúdium komunikačných stratégií bankového sektora na týchto platformách a na modelovanie angažovanosti verejnosti v závislosti od spôsobu komunikácie.

Cieľom tejto práce je nielen identifikovať rozdiely v komunikačných prístupoch bankových inštitúcií na sociálnych sieťach, ale aj vytvoriť model, ktorý by pomohol predpovedať úroveň angažovanosti verejnosti na základe týchto stratégií. S nárastom dôležitosti sociálnych sietí v marketingovom portfóliu bánk sa stáva kritickým poskytnúť vhodné odporúčania na zlepšenie komunikácie a zvýšenie zapojenia publika.

V prvej časti sa budeme venovať grafickému znázorneniu dát, aby sme lepšie pochopili správanie fanúšikov. V ďalšej časti budeme modelovať prírastky fanúšikovskej základne jednotlivých firiem finančného sektora a na koniec sa pozrieme na to, ako zvýšiť interakcie s publikovanými príspevkami.

Budovanie fanúšikovskej základne na sociálnej sieti si vyžaduje strategický prístup. Nie je to len o publikovaní príspevkov a dúfaní, že ľudia si ich všimnú. Je to komplexný proces, ktorý zahŕňa rôzne úlohy. Medzi ne patrí napríklad definovanie cieľového publika, tvorba relevantného a pútavého obsahu, pravidelná publikácia obsahu, interakcia s fanúšikmi či v neposlednom rade analýza a meranie výsledkov [2].

Pri publikácií príspevkov sa marketér musí zamyslieť nad rôznymi otázkami. Aký typ obsahu by som mal publikovať, aby som oslovil cieľové publikum? V aký deň a čas by som mal príspevky zverejniť, aby som dosiahol maximálny dosah? Ako môžem

interagovať so svojimi fanúšikmi a budovať s nimi vzťahy? Na všetky tieto otázky sa v našej práci nájdu odpovede, keďže v nej rozoberieme akým témam by sa mali banky venovať, aký typ obsahu ľudí zaujíma a taktiež aj kedy je vhodná doba na publikovanie príspevku, aby s ním ľudia čo najviac interagovali.

V závere tejto práce poskytneme odporúčania bankám, ktoré by mohli využiť na vylepšenie svojich komunikačných stratégií na sociálnych sieťach pre získanie väčšieho počtu priaznivcov. Týmto spôsobom by sme mohli prispieť k ďalšiemu porozumeniu dôležitosti sociálnych médií v modernom bankovom prostredí a pomôcť vytvárať efektívnejšie marketingové kampane.

1 Deskriptívna analýza dát na sociálnych sieťach

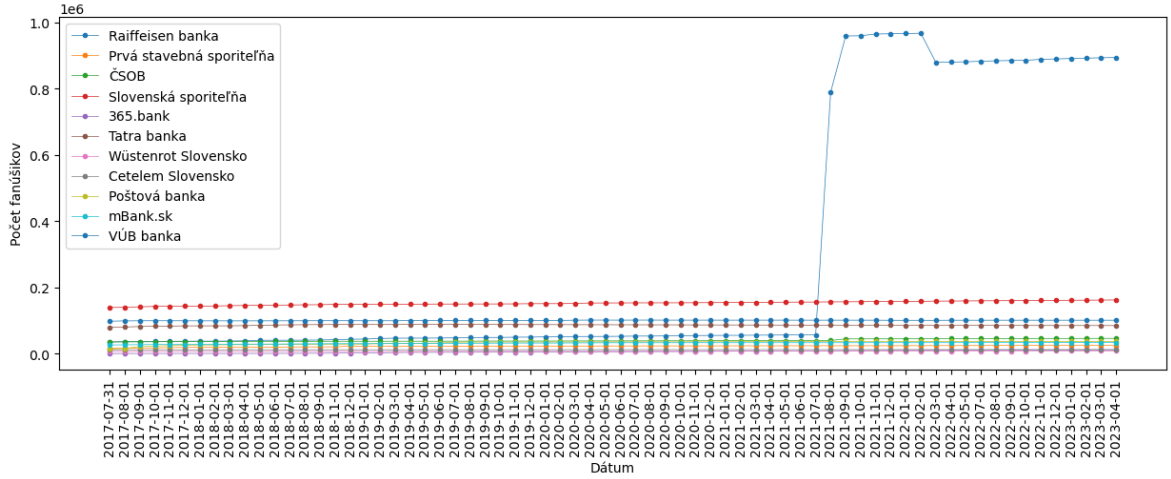
V tejto časti sa budeme bližšie pozerať na naše dáta zo sociálnych sietí firiem finančného sektora, ktorých je 22, z čoho iba pre 11 z nich máme údaje za každý rok. Dôvod, prečo máme všetky dáta iba pre niektoré firmy, je aj ten, že nie všetky banky u nás stále pôsobia.

V nasledujúcej kapitole budeme pracovať v programovacom jazyku *Python*. Na spracovanie dát nám pomôže knižnica *Pandas* a knižnica *NumPy* nám umožní vykonať matematické operácie na veľkom objeme dát. Na štatistické výpočty použijeme knižnicu *statistics* a na vizualizáciu výsledkov nám pomôže knižnica *MatPlotLib*.

1.1 Fanúšikovská základňa

Pod pojmom fanúšik / follower / odoberateľ sa berie jednotlivec, ktorý sa prihlási na odber daného účtu na sociálnej sieti, aby dostával jeho pravidelné aktualizácie a novinky. Veľký počet odoberateľov znamená väčšiu popularitu, dosah a aj vplyv danej značky / osoby, ktorá profil vlastní [11].

Pozrime sa na vývoj množstva fanúšikov v čase pre vybrané firmy. Na obrázku 1 vidíme, že takmer všetkým bankám rastie množstvo fanúšikov pomerne konštantne. Iba Raiffeisen banka mala v auguste a aj v septembri 2021 jeho výrazný nárast. Dôvodom môže byť to, že Raiffeisen banka uverejňovala rozhovory so svetoznámym srbským tenistom Novakom Đokovićom. Ďalším dôvodom mohla byť súťaž o karavan, ktorá sa vyhodnocovala práve v týchto mesiacoch, pričom ľudia mohli dať odber preto, aby im neunikol výherca. Taktiež v ten mesiac uverejňovali aj iné väčšie súťaže, čo opäť mohlo napomôcť k takému razantnému nárastu.



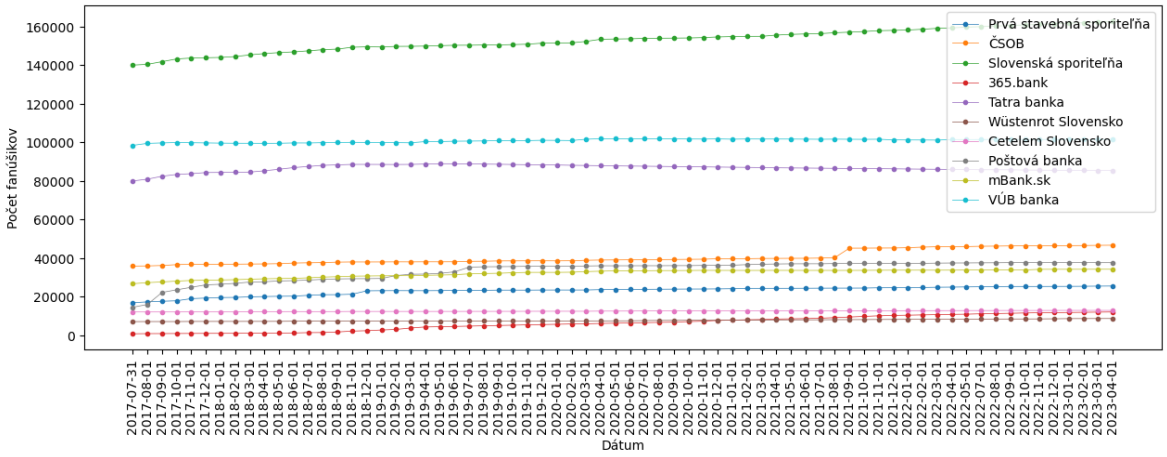
Obr. 1: Vývoj počtu fanúšikov vybraných firiem finančného sektora.

Aby sme sa mohli lepšie pozrieť na vývoj počtu fanúšikov ostatných bánk, vykreslíme si graf ešte raz, no tentokrát to bude bez Raiffeisen banky. Na obrázku 2 vidíme, že poradie bánk s väčším množstvom sledovateľov sa významne nemení, avšak pri menších bankách sa toto poradie mení.

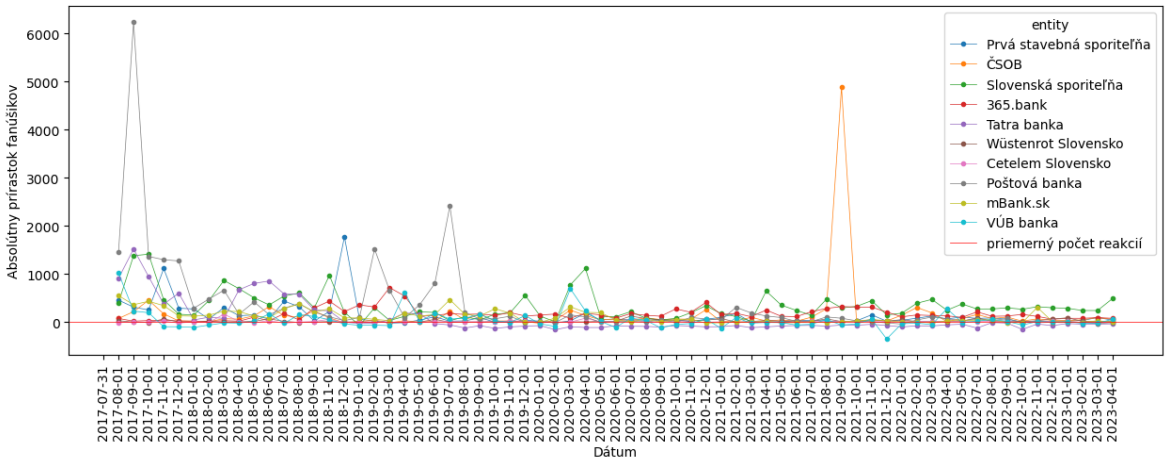
Za povšimnutie stojí, že Poštová banka, ktorá bola štvrtou najmenšou bankou v počte fanúšikov, s prehľadom predbehla Prvú stavebnú sporiteľňu a aj mBank.sk.

Určitý skok v množstve followerov mala aj ČSOB a Poštová banka, ktorý je lepšie vidieť na obrázku 3, na ktorom sú absolútne prírastky fanúšikov, ktoré nám dávajú lepší pohľad na túto metriku.

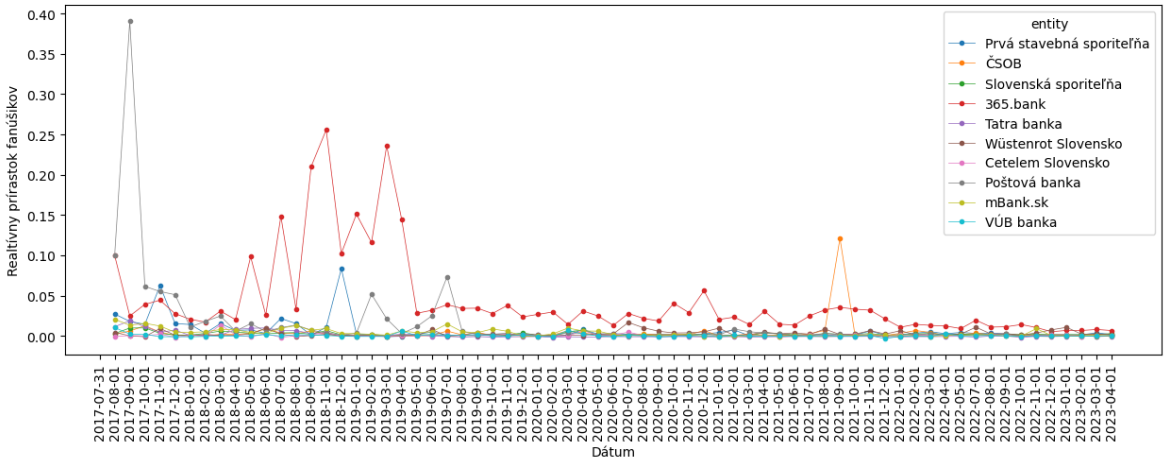
Relatívne prírastky na obrázku 4 nám hovoria, o koľko percent sa zvýšilo množstvo odoberateľov. Vďaka tomuto vidíme, že aj napriek tomu, že mala 365.bank najmenej fanúšikov, tak oproti ostatným bankám celé obdobie najviac napredovala.



Obr. 2: Vývoj počtu fanúšikov vybraných firiem finančného sektora bez Raiffeisen banky.



Obr. 3: Absolútny prírastok fanúšikov vybraných firiem finančného sektora bez Raiffeisen banky.



Obr. 4: Relatívny prírastok fanúšikov vybraných firiem finančného sektora bez Raiffeisen banky.

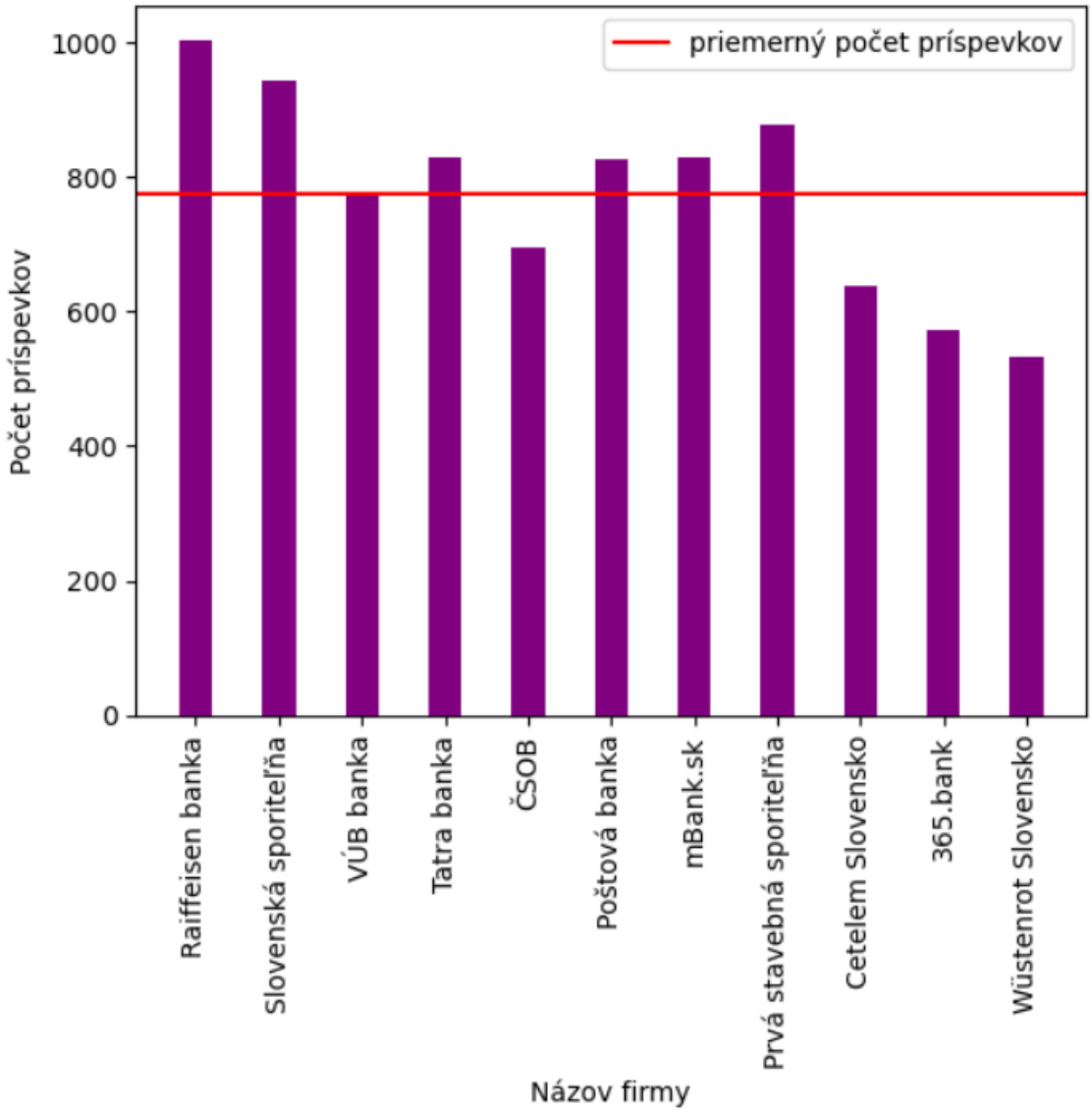
Všetky vykreslené obrázky vypovedajú o tom, že sa počet fanúšikov v priebehu času stále zvyšuje. Z toho môžeme usúdiť, že aj povedomie bánk sa stále zvyšuje. Dôvodom tohto dlhodobého nárastu môže byť zvýšená dostupnosť internetu v čoraz mladšom veku, platené reklamy, množstvo zaujímavého obsahu, poskytovanie kvalitných služieb a iné. My sa budeme neskôr v kapitole 2 venovať práve tomu, čo a ako pôsobí na množstvo fanúšikov, a teda čo by mali banky robiť preto, aby sa im fanúšikovská základňa stále zväčšovala z pohľadu uverejňovania obsahu, ktorý ľudí zaujíma.

1.2 Príspevky na sociálnych sieťach

Dôvodom, pre ktorý sa fanúšik rozhodne dať odber nejakému účtu je práve jeho publikovaný obsah. V tejto kapitole sa budeme pozerať na publikované príspevky jednotlivých bánk na sociálnej sieti *Facebook*.

Ako prvé sa pozrieme na množstvo vypublikovaných príspevkov jednotlivými firmami a skúsime zhodnotiť, či veľa príspevkov musí nutne znamenať veľa fanúšikov.

Na obrázku 5 máme na x - ovej osi zoradené banky v zmysle počtu odoberateľov, od najväčšej po najmenšiu. Na y - ovej osi máme celkový počet vypublikovaných príspevkov jednotlivých firiem. Môžeme si všimnúť, že najviac uverejneného obsahu má práve Raiffeisen banka, ktorá bola aj najväčšia. Tretím najväčším prispievateľom na sociálnej sieti je Prvá stavebná sporiteľňa. Jej počet odoberateľov je avšak podpriemerný oproti ostatným bankám, a preto nie je úplne zjavné, či práve množstvo príspevkov napomáha k vyššiemu počtu odoberateľov. To, či príspevky pôsobia na množstvo sledovateľov si otestujeme v kapitole 2.



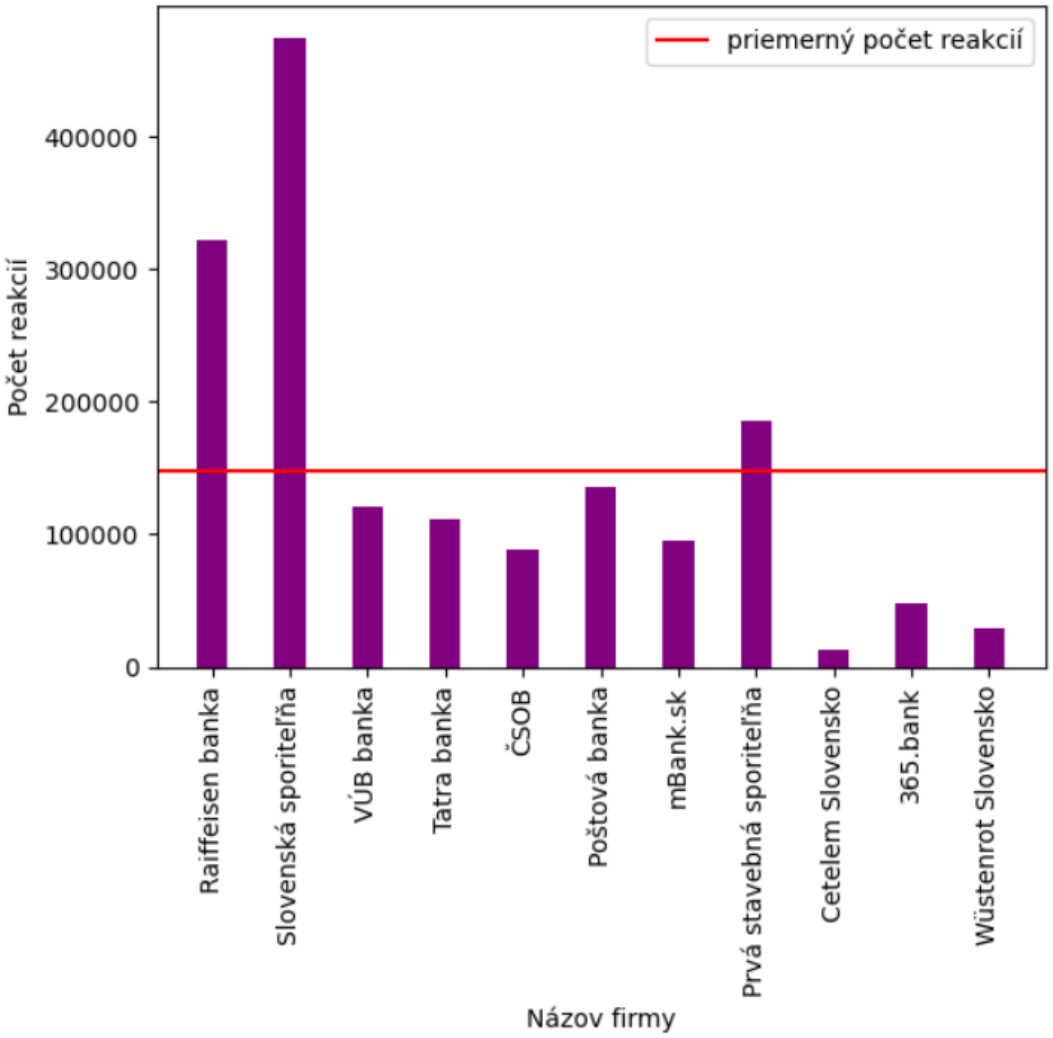
Obr. 5: Početnosť príspevkov publikovaných finančnými subjektmi na sociálnej sieti Facebook.

1.2.1 Reakcie na príspevky

Reakcia je názor, ktorý môžu používatelia sociálnych sietí vyjadriť pod akýkoľvek príspevok. Pod jeho počtom myslíme súčet všetkých komentárov, zdieľaní, označení „páči sa mi to, haha, love, wow, sorry, anger". Nás zaujíma, či vieme z grafov vyjadriť závislosť medzi počtom fanúšikov, príspevkov a počtom reakcií.

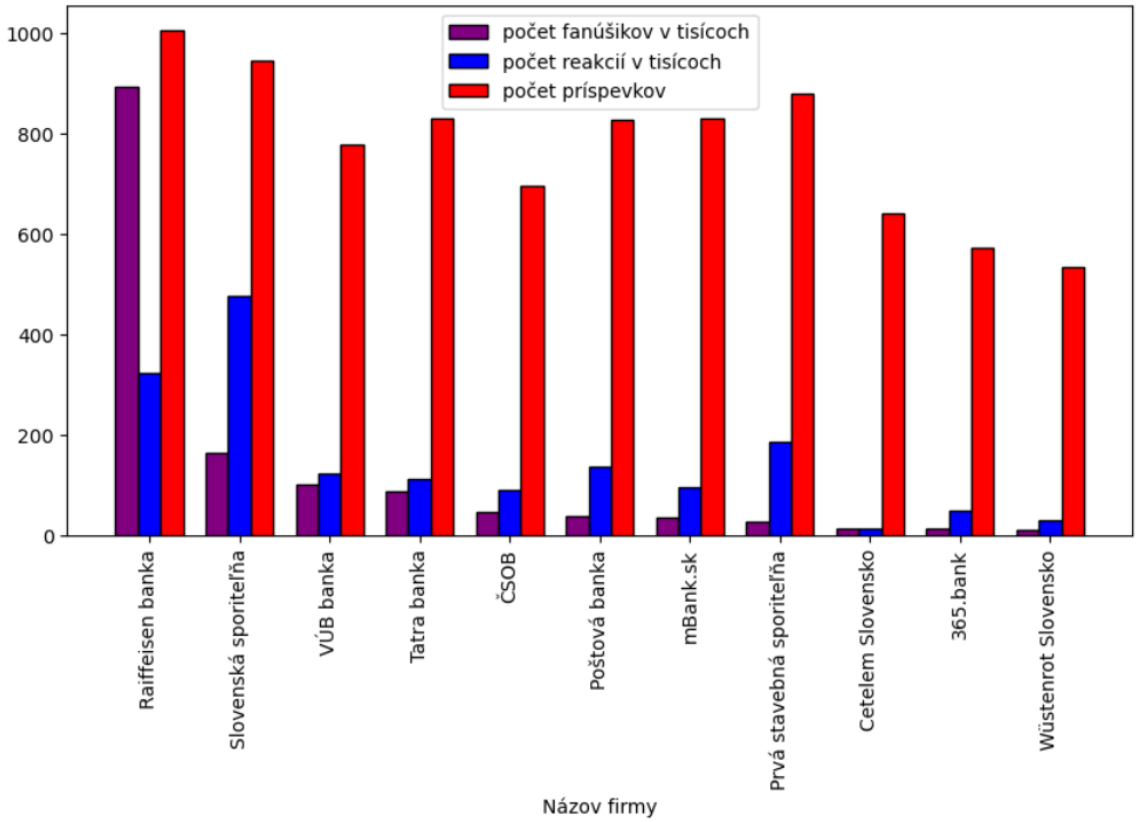
Obrázok 6 nám tentokrát ukazuje, že veľký počet príspevkov nemusí nutne znamenať vyšší počet reakcií. Slovenská sporiteľna síce mala na obrázku 5 veľký počet

príspevkov a na obrázku 6 najväčšie množstvo reakcií, avšak keď sa pozrieme napríklad na Tatra banku alebo Poštovú banku, tak ich nadpriemerný počet príspevkov nezaručil vysoké množstvo reakcií.



Obr. 6: Počet reakcií na príspevky vybraných firiem finančného sektora.

Keď si medzi sebou porovnáme všetky tri spomínané metriky, tak na obrázku 7 vidíme, že počet fanúšikov značne nezávisí od objemu príspevkov alebo od množstva reakcií. Taktiež si môžeme všimnúť, že zatiaľ čo sa počet príspevkov medzi jednotlivými firmami až tak neodlišuje, tak medzi počtami reakcií a fanúšikov sú rozdiely viditeľnejšie.



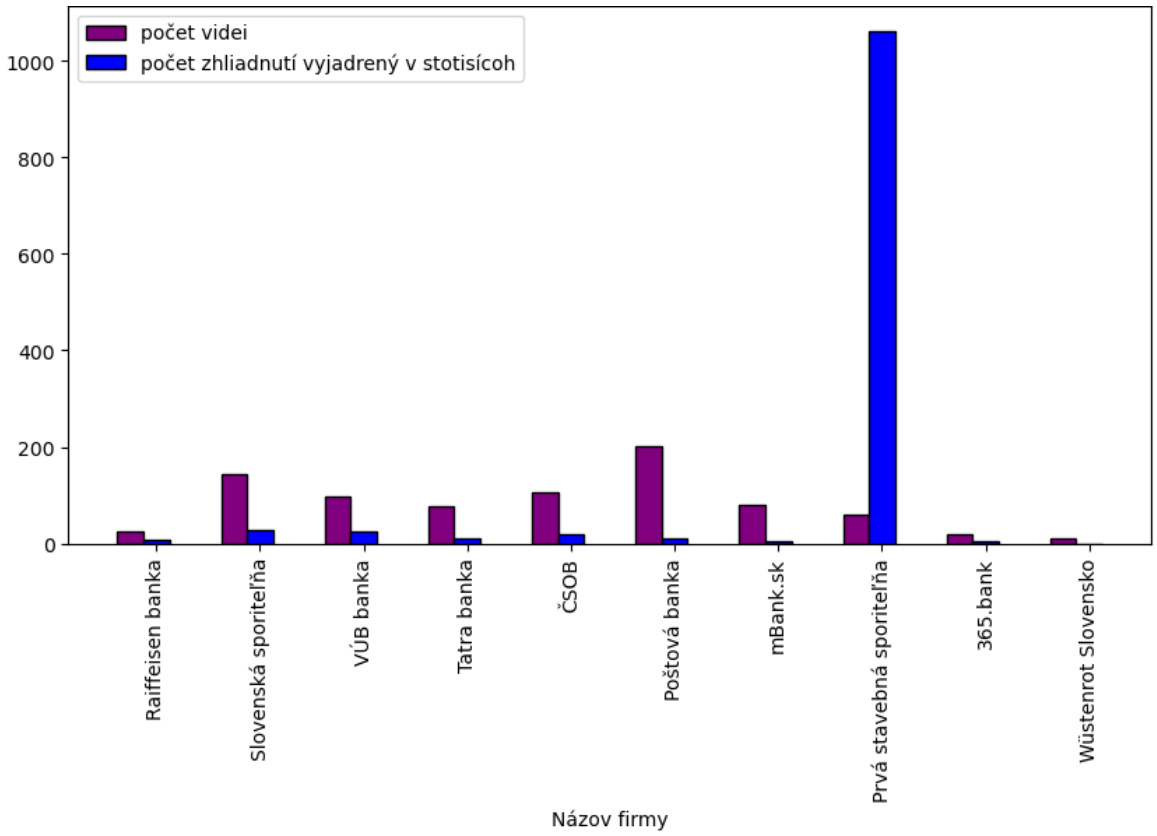
Obr. 7: Porovnanie počtu príspevkov, reakcií a fanúšikov vybraných firiem finančného sektora.

1.2.2 Publikované videá

Ďalšou formou príspevku, ktorý môže byť na fanúšikovskej stránke publikovaný, je video. Celkové spracovanie videa môže byť pre firmy veľmi náročné, a preto je podstatné vedieť, či z takejto formy príspevku firmy profitujú, alebo nie.

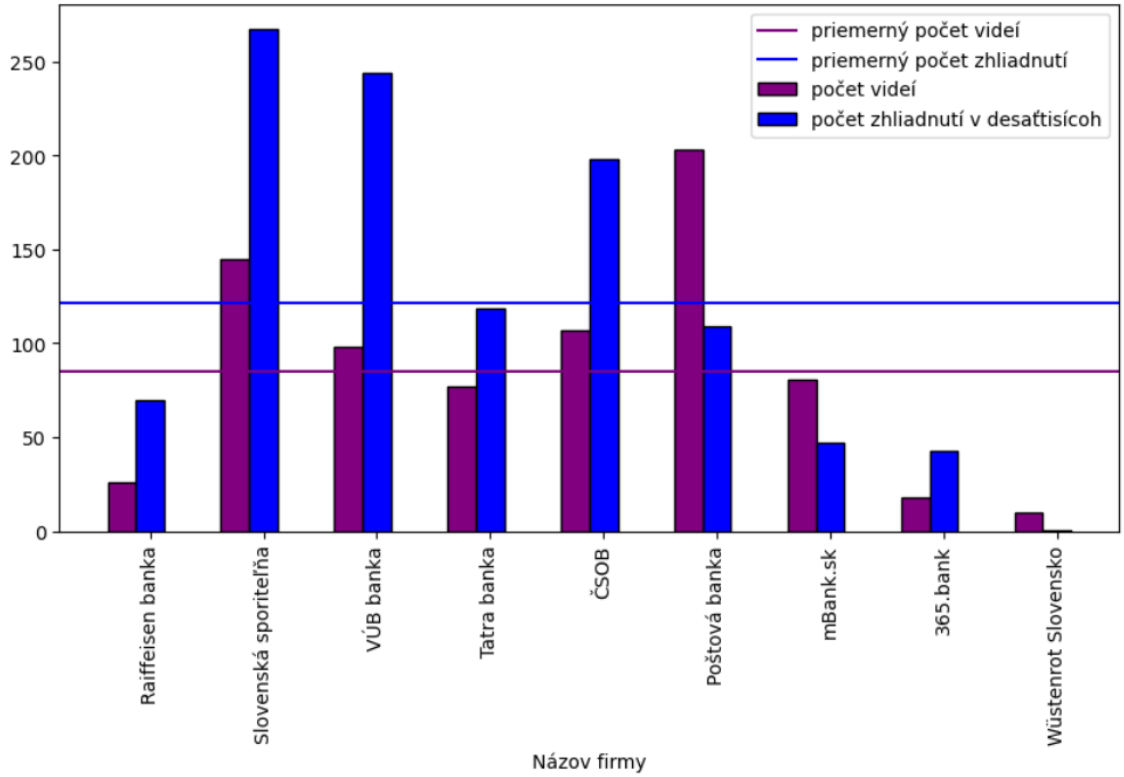
Na obrázku 8 vidíme porovnanie počtu vypublikovaných videí na sociálnej sieti *Facebook* a ich celkové množstvo zhliadnutí. Pri grafickom spracovaní sme vynechali firmu Cetelem Slovensko, pretože nepublikuje videá na YouTube. Najväčšou zaujímavosťou na tomto grafe je to, že Prvá stavebná sporiteľňa má veľmi veľké množstvo zhliadnutí. Zaujímavé je ešte aj to, že aj napriek tomu, že množstvo zhliadnutí je niekoľkonásobne väčšie ako u nejakej inej banky, jej celkový počet videí je podpriemerný. Keď sme sa pozreli bližšie na jej jednotlivé videá, nenašli sme žiadne, ktoré by spôsobilo takýto výrazný nárast zhliadnutí, keďže viacero videí Prvej stavebnej sporiteľne

má viac ako 10 000 000 zhliadnutí.



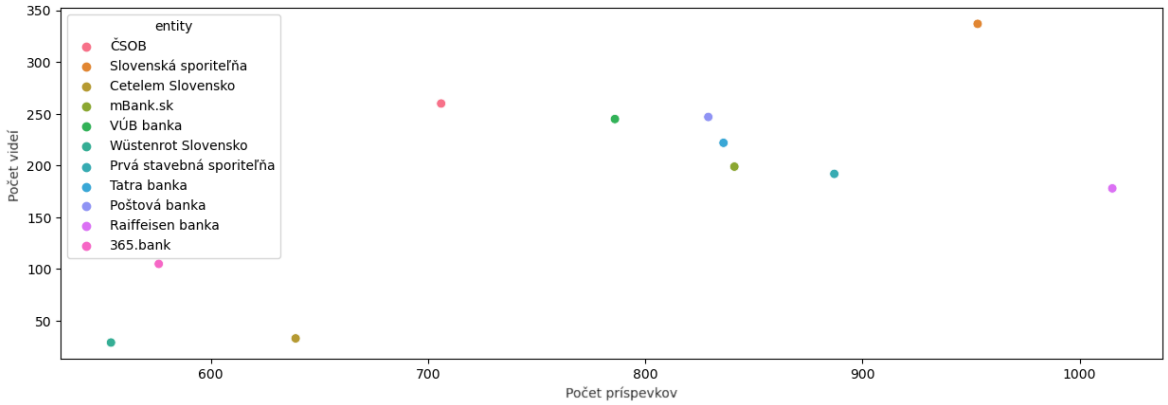
Obr. 8: Porovnanie počtu videí a zhliadnutí vybraných firiem finančného sektora.

Porovnajme ešte raz počty videí a počty zhliadnutí, tentokrát ale bez Prvej stavebnej sporiteľne. Banky máme opäť zoradené od najväčšej po najmenšiu a teda obrázok 9 nám ukazuje, že ani počet videí nemusí úplne priamo ovplyvňovať množstvo fanúšikov na sociálnej sieti.



Obr. 9: Porovnanie počtu videí a zhliadnutí vybraných firiem finančného sektora mimo Prvej stavebnej sporiteľne.

Ak by sme sa chceli pozrieť na závislosť medzi počtom príspevkov a videí, tak na obrázku 10 vidíme, že množstvo videí rastie približne lineárne spolu s publikovanými príspevkami. Taktiež aj vidíme, že Slovenská sporiteľňa má podiel videí najväčší, zatiaľ čo Wüstenrot ho má najmenší.



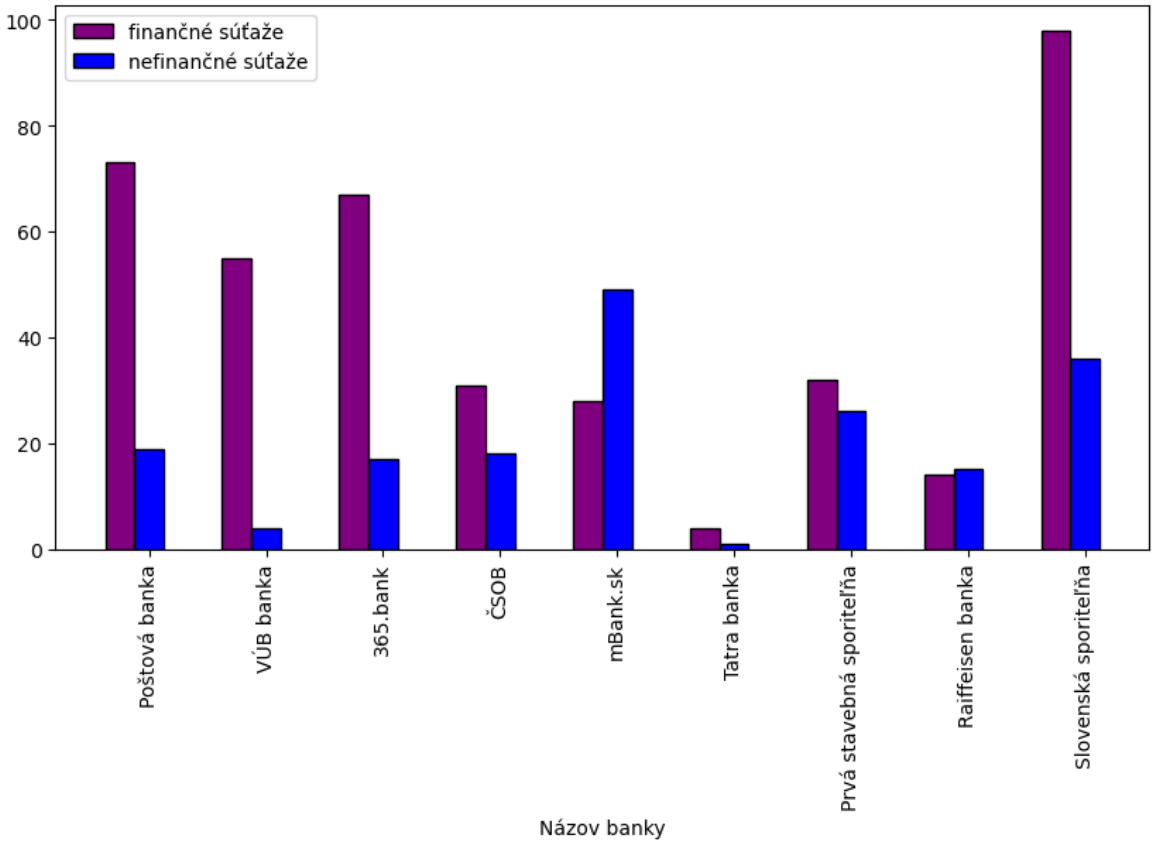
Obr. 10: Porovnanie počtu videí a príspevkov vybraných firiem finančného sektora.

1.2.3 Súťažné príspevky

Mnoho firiem nielen z finančného sektora usporadúva súťaže s cieľom prilákať nových fanúšikov a tým zvýšiť povedomie o svojej značke. Súťaže môžu byť finančného charakteru, kde súťažiaci môže vyhrať peňažnú cenu, alebo nefinančného charakteru, pri ktorých môže byť výhrou produkt alebo služba (dovolenka, auto, mobil a pod.).

Keďže organizovanie súťaží s hodnotnejšími výhrami je finačne nákladnejšie, nebýva pravidlom, že každá firma ich musí uskutočňovať. Každá firma má však inú marketingovú stratégiu a teda vynaloženie väčších finančných prostriedkov nie je jediným dôvodom, prečo firmy ako Cetelem Slovensko alebo Wüstenrot tieto súťaže nerobia. V tejto časti budeme teda zobrazovať iba zvyšných 9 firiem, ktoré už niekedy nejakú súťaž na sociálnej sieti Facebook vyhlásili.

Na obrázku 11 si hneď môžeme všimnúť, že Tatra banka neorganizuje takmer žiadne súťaže. Z grafu ďalej pozorujeme, že väčšina firiem usporadúva podstatne viac súťaží s finančnou výhrou. Výnimkou je okrem Raiffeisen banky najmä mBank.sk, ktorá nielenže organizuje najviac súťaží s nefinančnými výhrami, ale zároveň volí tento typ výhry pri takmer dvoch tretinách súťaží. 365.banka, ktorá je z vykreslených firiem najmenšia (čo sa týka počtu fanúšikov), publikuje nadpriemerne veľa súťaží. Môže sa to brať aj ako forma stratégie na zvýšenie povedomia o značke a tým pádom aj o zvýšenie množstva odoberateľov. Na druhej strane Raiffeisen banka, ktorá je najväčšia, publikovala takmer najmenej súťaží, čo taktiež môže byť preto, že má oproti ostatným firmám finančného sektora niekoľkonásobne vyšší počet fanúšikov a nepotrebuje ho touto cestou zvyšovať.

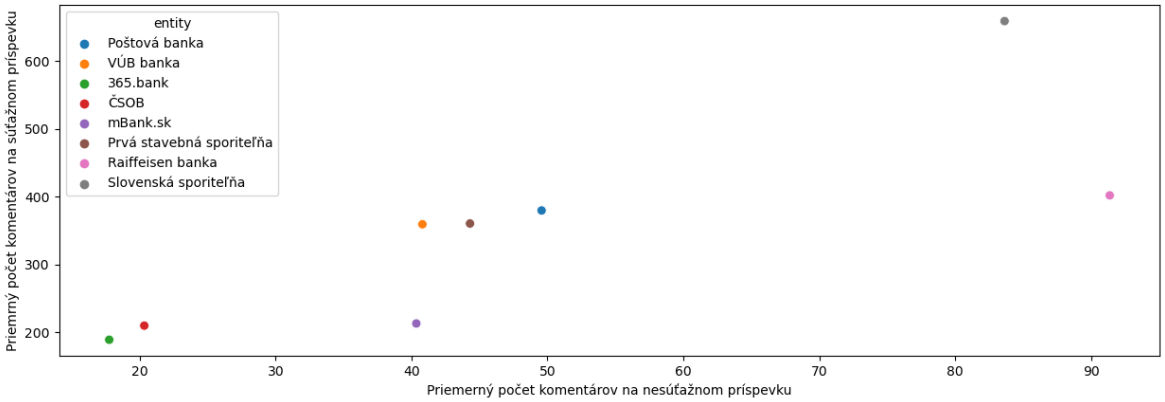


Obr. 11: Porovnanie počtu finančných vs. nefinančných súťaží vybraných firiem finančného sektora.

Posledná vec, na ktorú sa v tejto podkapitole pozrieme, je, či súťažné príspevky zvyšujú angažovanosť na sociálnej sieti.

Na obrázku 12 vidíme na x - ovej osi priemerný počet komentárov na bežnom príspevku a na y - novej priemerný počet komentárov na príspevku súťažnom. V tomto prípade sme opäť vynechali Cetelem Slovensko a Wüstenrot, pretože žiadnu súťaž neorganizovali. Taktiež sme vylúčili Tatra banku, pretože jej priemerný počet komentárov na súťažných príspevkoch bol 12 060,8, čo by nám spôsobilo horšie analyzovanie komentárov pri ostatných bankách, zatiaľ čo na bežných príspevkoch mala iba 8,74.

Môžeme si všimnúť, že ľudia sa v priemere častejšie zapájajú do súťažných príspevkov, čo je logické, keďže väčšinou v súťažných pravidlách je práve komentovanie daného príspevku. To je pre banku výhodou, pretože takýto príspevok sa potom dostane medzi viac ľudí, čo môže firme napomôcť k rozšíreniu jej publika.



Obr. 12: Porovnanie priemerného počtu komentárov na súťažných vs. nesúťažných príspevkoch.

2 Modelovanie prírastku fanúšikovskej základne na sociálnych sieťach

V tejto kapitole budeme skúmať čo a ako pôsobí na zmenu počtu fanúšikov (sledovateľov) banky na sociálnej sieti. Túto premennú sme si vybrali kvôli tomu, že množstvo sledovateľov priamo ovplyvňuje potenciálny dosah príspevkov - teda koľko používateľov môže príspevok uvidieť, čím môže fanúšikovská stránka získať väčší vplyv na rozhodovanie ľudí. Veľký počet fanúšikov zvyšuje pravdepodobnosť, že príspevky budú zdieľané a tak sa dostanú medzi väčší počet ľudí. V našom prípade to znamená, že väčšie množstvo ľudí sa dozvie o bankových produktoch, súťažiach, novinkách a celkovo sa tak zvýši povedomie o danej banke. Podľa článku [9], veľký počet fanúšikov taktiež prispieva k zvýšeniu dôveryhodnosti značky, čo môže mať za následok, že viac ľudí investuje do produktu danej banky.

Pri modelovaní našich dát budeme využívať 2 modely. Ako prvý bude jednoduchý statický *fixed effects model* a potom využijeme dynamický *LDV* model. Každý model použijeme dvakrát, pričom druhýkrát pridáme do našich dát aj dáta obsahujúce ekonomickú situáciu.

Rovnako ako v predchádzajúcej kapitole budeme pracovať v programe *Python* s dátami, ktoré máme uložené v databáze *Oracle*. Naše dáta sú panelové, čo znamená, že sa v nich nachádzajú informácie napríklad o spomínanom počte fanúšikov v priebehu času a pre viacero firiem. Výhodou tohto typu dát je možnosť analyzovať nielen rozdiely medzi jednotlivými premennými, ale aj ich zmeny v čase. To poskytuje rozmanitejší pohľad na sledovanú premennú.

Na načítanie dát použijeme knižnice *pandas*, *sqlalchemy*, *project_lib*. Na vizualizáciu dát použijeme *matplotlib*, *seaborn*. Na štatistické výpočty budeme používať knižnice *sklearn*, *statsmodels*, *scipy*.

Dáta, s ktorými budeme pracovať budú opäť od 1.7.2017 a bude sa jednať o mesačné rezy sledovaných premenných nami vybranými finančnými entitami. Z načítaných

dát vyberieme iba 11 firiem z 22, pre ktoré máme dáta za každý mesiac za spomínané obdobie, aby sme mali pre každý mesiac rovnaký počet záznamov pre všetky banky. Jedným z dôvodov absencie dát je to, že niektoré banky sa medzičasom spojili, zatiaľ čo iné zanikli, respektíve prestali pôsobiť na Slovensku.

2.1 Príprava dát

V našich testoch budeme používať dáta z 3 rôznych tabuliek, pričom prvá obsahuje výkonnosť profilov na sociálnej sieti, druhá hovorí o uverejnených súťažiach na jednotlivých profiloch a posledná obsahuje ekonomickú situáciu na Slovensku.

Vzhľadom na to, že formát dátumu sa v jednotlivých zdrojoch nezhodoval, v každej tabuľke sme ho štandardizovali na formát $YYYY - MM - 01$. Taktiež sme vytvorili aj numerickú premennú *numerical_date*, ktorá slúži na číselnú reprezentáciu dátumu.

2.1.1 Prehľad výkonnosti profilov

Prvá tabuľka obsahuje nezávislé premenné, v ktorých sú názvy bánk spolu s mesačnými informáciami o počte fanúšikov, príspevkov a ich interakcií s nimi. My budeme používať nasledujúce premenné:

- **num_posts** je počet všetkých príspevkov, ktoré banka za daný mesiac publikovala na svojej sociálnej sieti,
- **comments** je počet všetkých komentárov, ktoré boli za daný mesiac uverejnené pod všetkými príspevkami na sociálnej seti banky,
- **shares** je počet všetkých zdieľaných príspevkov na sociálnej sieti banky za daný mesiac,
- **num_of_reactions** je počet všetkých reakcií, ktoré získali príspevky za daný mesiac na sociálnej sieti banky,

- **datum** je mesiac, počas ktorého boli zbierané vybrané metriky.

2.1.2 Uverejnené súťaže

Ďalšia tabuľka obsahuje informácie o súťažiach organizovaných jednotlivými finančnými sektormi. Nás bude ale zaujímať iba stĺpec s názvom $prize_fund$, ktorý má hodnotu buď *null*, ak sa jednalo o nefinančnú súťaž, alebo väčšie ako 0, pokiaľ sa jednalo o súťaž finančnú. Vďaka tejto informácii si vyrobíme nové premenné, ktoré nám budú hovoriť o počte finančných (*pocet_fin*), respektíve nefinančných (*pocet_nefin*), súťažiach za jednotlivé mesiace pre všetky banky.

Ďalšia premennú, ktorú si z týchto dát vytvoríme, bude počet všetkých súťaží za posledné 3 mesiace (*competition_posts_3*) a 6 mesiacov (*competition_posts_3*), aby sme zistili, či súťaže, ktoré boli zverejnené pred dlhšou dobou, majú nejaký vplyv na prírastok fanúšikov. Urobili sme to z toho dôvodu, že niektoré súťaže mávajú v podmienkach zapojenia sa napríklad dať odber stránke, ktorý po skončení súťaže môžu jej participanti zrušiť.

2.1.3 Ekonomická situácia na Slovensku

Keďže predpokladáme, že aj ekonomická situácia môže vplývať na rozhodovací proces investovania peňazí do bankového produktu potenciálneho zákazníka, rozhodli sme sa vybrať niekoľko faktorov, ktoré použijeme v modeloch neskôr.

Dáta o ekonomickej situácii máme k dispozícii až od januára 2018, zatiaľ čo dáta o výkonnosti profilov od júla 2017. Kvôli tomu sme zadefinovali / vytvorili premennú *competition_posts_3*. Tá zabezpečí, že pri analýze budeme dáta o výkonnosti profilov používať taktiež až od roku 2018.

Tabuľka s ekonomickou situáciou na Slovensku obsahovala viacero premenných, z ktorých sme si do nášho modelu vybrali nasledovné:

- **gasoline_price** je priemerná cena benzínu za predchádzajúci mesiac,

- **new_mortage_ptf_m** je počet nových úverov za predchádzajúci mesiac,
- **new_loan_ptf_m** je počet nových pôžičiek za predchádzajúci mesiac,
- **avg_salary_ptf** je priemerný plat klientov ČSOB za predchádzajúci mesiac,
- **change_inflation_g** je zmena inflácie za posledné 3 mesiace.

Pri výbere premenných sme brali do úvahy metriky, ktoré najviac vplývajú na jednotlivcov.

2.1.4 Závislá premenná

Po spracovaní predchádzajúcich 3 zdrojov a ich spojenia do jednej finálnej tabuľky sme pridali závislú premennú *prirastok*, ktorá nám hovorí o zmene počtu fanúšikov jednotlivých firiem od predchádzajúceho mesiaca. Keďže budeme testovať práve prírastok fanúšikovskej základne, ako prvé sme si vykreslili jeho histogram, aby sme zistili, či sa v vo fanúšikovskej základni nenachádzajú nejakí outlieri.

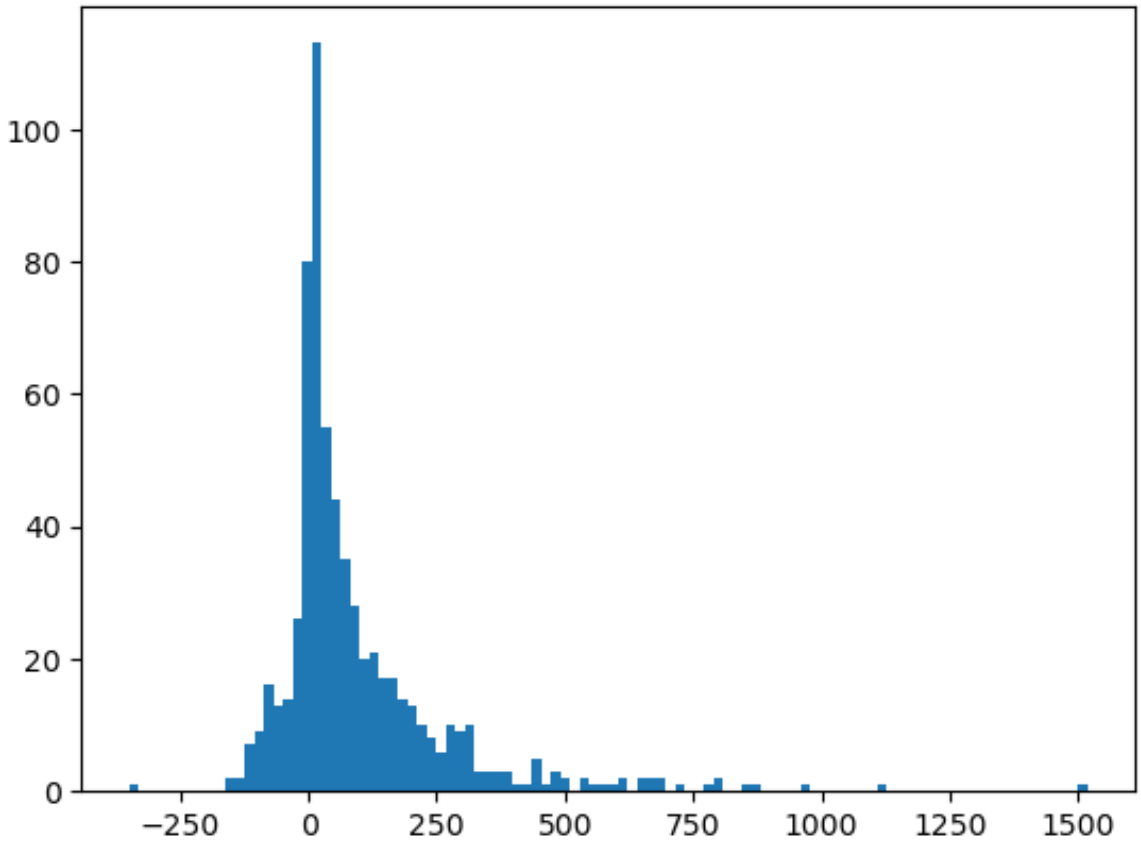
Zistili sme, že naše dáta obsahujú outliera, ktorým je Raiffeisen banka. Už na obrázku 1 sme videli, že táto firma mala od určitého bodu prudký nárast fanúšikov. Z tohto dôvodu nebudeme pracovať s touto entitou, pretože by nám skreslila výsledky testovania.

Po bližšom preskúmaní počtu prírastkov firma ČSOB mala v septembri 2021 prudký náraz fanúšikov. Túto anomáliu pripisujeme zlúčeniu firmy s bankou OTP, ktorá nastala práve v tomto mesiaci [4], a preto dáta pre firmu ČSOB za tento mesiac vyradíme z nášho datasetu.

Ďalším outlierom bola Poštová banka v júli a v septembri roku 2017. Po preskúmaní príspevkov na Facebookovej stránke banky z daného obdobia sme zistili, že v týchto mesiacoch organizovala súťaže s finančnou odmenou. V júli propagovali aj slovenský folklórny festival, ktorého prívrženci sa mohli kvôli tomu prihlásiť na odber tejto stránky na sociálnej sieti. V septembri sa banka na svojej stránke spojila

so slovenským módnym návrhárom a slovenskou moderátorkou a vytvorili novú sériu príspevkov, ktoré boli uverejňované iba v danom mesiaci. Toto mohol byť taktiež dôvod náhleho nárastu fanúšikov a z toho dôvodu nebudeme ďalej pracovať s týmito dvoma záznamami.

Po odstránení ďalších outlierov vidíme na obrázku 13 finálne rozloženie fanúšikovskej základne.



Obr. 13: Rozloženie prírastkov fanúšikovskej základne po odstránení outlierov.

2.2 Modelovacie techniky

V tejto podkapitole predstavíme konkrétne modely, ktoré použijeme pri modelovaní prírastkov fanúšikovských základní firiem finančného sektora. Taktiež vysvetlíme koeficienty, ktoré do jednotlivých modelov vstupujú a spôsoby ako špecificky modelovanie a analýzu dát ovplyvňujú. Aby sme mohli dáta modelovať lineárnou regresiou, musia byť splnené jej predpoklady, ktoré sú v prílohe A.

2.2.1 Fixed effects model

Prvým lineárnym regresným modelom, ktorý budeme používať, je statický *Fixed Effects model*. Jedná sa o najviac využívaný model pri panelových dátach. Prívlastok statický má preto, lebo modeluje súčasný vplyv nezávislej premennej na závislú a predpokladá, že sa tento vplyv v čase nezachová [18]. Jeho predpis je nasledovný:

$$y_{i,t} = \alpha_i + \beta_1 x_{i,t} + u_t + \epsilon_{i,t}, \tag{1}$$

pričom

- $y_{i,t}$ je závislá premenná v čase t pre jednotku i,
- α_i je fixný efekt jednotky i,
- $x_{i,t}$ je nezávislá premenná,
- β_1 je koeficient pre nezávislú premennú $x_{i,t}$,
- u_t je časovo - špecifický fixný efekt,
- $\epsilon_{i,t}$ je náhodná chyba.

Časovo - špecifický fixný efekt bude v našom prípade modelovaný premennou *numerical_ date*. Táto premenná je spojitá a teda namiesto u_t budeme tento efekt označovať iba písmenom u.

Model s fixnými a nie náhodnými efektami budeme používať preto, lebo vieme kontrolovať vzťahy medzi premennými pre všetky banky. Taktiež je tento model menej náchylný na korelácie medzi premennými, keďže odhaduje efekt pre každú banku idividuálne [18].

2.2.2 LDV model

V tejto časti budeme modelovať naše dáta pomocou dynamického modelu, keďže narozdiel od statických modelov, modely dynamické odhadujú krátkodobé (*short - run*), ale aj dlhodobé efekty (*long - run*), ktoré sú rozložené v čase. Tieto typy sú častokrát lepšou voľbou, pretože môžu zachytiť komplexnú dynamickú povahu údajov.

Základnou črtou dynamického modelu je, že obsahuje aj oneskorenú závislú premennú tzv. *lagged dependent variable*. Najzákladnejším dynamickým modelom je práve *lagged dependent variable (LDV)* model, ktorý je taktiež označovaný ako *Koyck model* a vyzerá nasledovne [18]:

$$y_{i,t} = \alpha_i + \phi_1 y_{i,t-1} + \beta_1 x_{i,t} + e_{i,t} \tag{2}$$

pričom

- $y_{i,t}$ je hodnota závislej premennej v čase t pre jednotku i,
- α_i je fixný efekt jednotky i,
- ϕ_1 je koeficient oneskorenej závislej premennej,
- $x_{i,t}$ je nezávislá premenná,
- β_1 je koeficient pre nezávislú premennú $x_{i,t}$,
- $y_{i,t-1}$ je hodnota závislej premennej v čase $t-1$ pre jednotku i,
- $e_{i,t}$ je náhodná chyba.

V LDV modeli vystupuje aj dlhodobý efekt, čo je efekt nezávislej premennej na závislú premennú v dlhodobom horizonte. Umožňuje nám teda vidieť, ako nezávislé premenné ovplyvňujú závislú v dlhodobom horizonte. Jeho výpočet vyzerá nasledovne [18]:

$$\frac{\hat{\beta}_1}{1-\hat{\phi}_1}, \tag{3}$$

Ak je autoregresívny koeficient ϕ blízky 1, potom sa tento efekt rozloží na dlhší časový horizont, zatiaľ čo v prípade, že by jeho hodnota bola blízka 0, potom by sa väčšina long - run efektu prejavila skôr, teda efekt by bol rozložený na kratší časový horizont.

2.2.3 Prehľad metrík kvality modelu

V tejto podkapitole si vysvetlíme základné metriky kvality modelu, ktoré sú bližšie charakterizované v [20] a [17].

- **R - squared** je koeficient determinácie, ktorý je druhou mocninou korelačného koeficientu a vysvetľuje, koľko percent variability závislej premennej (prírastku) ovplyvňujú nezávislé premenné,
- **Adj. R - squared** je upravená verzia *R - squared*, ktorá zohľadňuje počet nezávislých premenných v modeli, a preto poskytuje presnejší obraz o tom, koľko variability nezávislej premennej je vysvetlené závislou premennou,
- **F - statistic** je testovacia štatistika, ktorá sa používa na testovanie hypotézy, že všetky koeficienty v modeli sú nulové. Testuje teda, či majú nezávislé premenné v modeli vplyv na závislú premennú.
- **Prob (F - statistic)** je *p - value* pre *F - statistic*,
- **t** je testovacia štatistika, ktorá sa používa na testovanie významnosti koeficientu v regresnom modeli,
- **P>|t|** je *p - value* pre testovaciu štatistiku t z t - rozdelenia. Hodnota $< 0{,}05$ indikuje štatistickú významnosť koeficientu.

2.3 Výsledky modelov

V ďalších častiach tejto podkapitoly si opíšeme výsledky, ktoré sme dostali vďaka vyššie spomínaným modelom. Výstupom bude taktiež aj odporúčanie pre banku, aký typ príspevkov by mala uverejňovať za účelom prilákania nových fanúšikov svojej Facebookovej stránky.

2.3.1 Fixed effects model

Odhad fixných efektov prebieha 2 spôsobmi. Buď sa vytvoria odchýlky od priemeru pre závislé a nezávislé premenné, alebo sa pre každý prípad zavedú tzv. *„dummy“* premenné.

My budeme odhadovať fixné efekty pomocou sady *dummy* premenných, čo sa taktiež nazýva aj ako metóda *least square dummy variable* (LSDV). Našimi *dummy variable* budú samotné banky. Fixný efekt sa odhadne pre každý prípad, avšak jeden z nich dostane hodnotu 0 a stáva sa tak referenčnou hodnotou, čo bude v našom prípade práve ČSOB.

V našom prípade budú spomínané premenné v podkapitole 2.2.2 nasledovné:

- $y_{i,t}$ je prírastok fanúšikov v čase t pre jednotku pre jednotlivé banky,
- α_i je fixný efekt banky,
- $x_{i,t}$ sú počet príspevkov, reakcií, komentárov, zdieľaní, počet všetkých súťaží a počty finančných a nefinančných súťaží samostatne,
- β_1 je koeficient pre nezávislé premenné,
- u_t poradie mesiaca danej premennej,
- $\epsilon_{i,t}$ je náhodná chyba.

Na obrázku 14 vidíme výsledky fixed effects modelu, ktorý sme použili na naše dáta. Prvá časť tabuľky nám opisuje regresný model, druhá časť obsahuje koeficienty

regresnej rovnice a v poslednej časti vidíme rôzne štatistické testy a metriky, ktoré slúžia na posúdenie primeranosti modelu a predpokladov lineárnej regresnej analýzy. Nás budú momentálne zaujímať prvé dve časti tabuľky a výsledky budeme interpretovať na hladine významnosti 5 %.

V prvej časti hodnota *R - squared* vyšla rovná 44,9 %, čo je dosť a mohli by sme podľa toho tvrdiť, že náš model je dostatočný. Keď sa pozrieme na *Adj. R - squared*, tak jeho hodnota 43,4 % nás presvedčuje o tom, že naozaj môžeme tvrdiť, že náš model dostatočne popisuje prírastok fanúšikov a teda je vhodným. Hodnota *Prob (F - statistic)* nám vyšla veľmi blízka nule a teda môžeme tvrdiť, že nezávislé premenné majú významný vplyv na závislú premennú.

Keďže sme si zvolili ČSOB ako našu referenčnú premennú, tak konštanta (*const*), ktorá má v našom modeli hodnotu $53,2316$ znamená, že pri nulovej hodnote ostatných nezávislých premenných bude priemerný mesačný prírastok tejto banky dosahovať práve túto hodnotu. Môžeme to tvrdiť z toho hľadiska, že P>|t| pre Intercept je menšia ako 5 %.

Premenné ako *num_posts, shares, comments, pocet_fin, numerical_date* sú štatisticky významné. Znamená to, že banky by mali očakávať, že publikované príspevky s vyšším počtom komentárov a zdieľaní budú mať pozitívny vplyv na prírastok nových fanúšikov na ich stránke. Taktiež by sa im malo odporučiť pridávať viac finančných súťaží, namiesto nefinančných, keďže premenná *pocet_nefin* nie je významná a teda nevieme určiť, aký vplyv by mala na počet fanúšikov.

num_of_reactions, competition_posts_3 a *competition_posts_6* nemajú štatisticky významný vplyv na počet fanúšikov. Počet reakcií a počet publikovaných súťaží pred 3 a 6 mesiacmi nevplývajú na to, koľko ľudí sa rozhodne dať odber na fanúšikovskú stránku jednotlivých bánk.

Pri bankách ako *365.bank, Slovenská sporiteľňa, Tatra banka* a *VÚB banka* sú koeficienty štatisticky významné a preto môžeme tvrdiť, že ak by napríklad *Tatra*

banka nepridala za daný mesiac žiaden príspevok, ani by nemala žiadne interakcie s predchádzajúcimi príspevkami, jej priemerný mesačný prírastok by bol:

$$53,2314 - 78,5938 = \textbf{-25,3624} \tag{4}$$

pričom $53,2314$ je hodnota koeficientu pri absolútnom člene a $-78,5938$ je hodnota koeficientu pri Tatra banke. Z tohto nám vyplýva, že ak by Tatra banka nevynaložila žiadnu snahu na jej sociálnej sieti, každý mesiac by prichádzala v priemere o $-25,3624$ fanúšikov. Treba avšak spomenúť, že $-25,3624$ je iba bodový odhad a existuje interval spoľahlivosti, v ktorom sa táto predikcia nachádza.

V tabuľke 1 vidíme, že Slovenská sporiteľňa má v prípade nevynaloženia nejakej snahy najväčší mesačný prírastok fanúšikov, zatiaľ čo Tatra banka spolu s VÚB bankou by mali úbytok odoberateľov na ich stránke na sociálnej sieti.

Názov banky	Prírastok fanúšikov
ČSOB	53,2316
Tara banka	-25,3622
VÚB banka	-23,3753
365.bank	143,9892
Slovenská sporiteľna	162,0698
Cetelem Slovensko	25,6166
Poštová banka	76,8666
Prvá stavebná sporiteľňa	29,2782
Wüstenrot Slovensko	23,368
mBank.sk	37,9224

Tabuľka 1: Prírastky fanúšikov po regresnej analýze pre Fixed - effects model.

Pozrime sa ešte, ako by vyzerali koeficienty v prípade, že by sme zobrali do úvahy aj ekonomické faktory spomenuté v podkapitole 2.1.3. Na obrázku 15 vidíme, že jedine cena benzínu za minulý mesiac má štatistický význam pri našej nezávislej premennej.

V prvej časti tabuľky na obrázku 14 sa nám potvrdilo, že nezávislé premenné majú významný štatistický vplyv na závislú premennú. Taktiež sa nám ukázalo, že náš model vysvetľuje veľkú časť z procesu rastu fanúšikovskej základne firiem finančného

sektora.

Naše odporúčania po vykonaní testu na našich dátach sú nasledovné:

- získavať komentáre a zdieľania na príspevky,
- pravidelne publikovať príspevky,
- v prípade oraganizovania súťaže, uprednostniť finančú odmenu pred nefinančnou.

OLS Regression Results

Dep. Variable:	prirastok	R-squared:	0.449
Model:	OLS	Adj. R-squared:	0.433
Method:	Least Squares	F-statistic:	28.00
Date:	Wed, 20 Sep 2023	Prob (F-statistic):	3.51e-68
Time:	07:10:54	Log-Likelihood:	-4049.6
No. Observations:	638	AIC:	8137.
Df Residuals:	619	BIC:	8222.
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	53.2316	24.803	2.146	0.032	4.523	101.940
num_of_reactions	0.0022	0.005	0.453	0.650	-0.007	0.012
num_posts	5.5016	1.173	4.691	0.000	3.199	7.805
shares	0.0703	0.026	2.659	0.008	0.018	0.122
comments	0.0328	0.006	5.593	0.000	0.021	0.044
pocet_fin	27.3497	6.643	4.117	0.000	14.305	40.395
pocet_nefin	1.4734	11.012	0.134	0.894	-20.152	23.099
365.bank	90.7576	25.456	3.565	0.000	40.768	140.748
Cetelem Slovensko	-27.6150	25.305	-1.091	0.276	-77.310	22.080
Poštová banka	23.6350	25.726	0.919	0.359	-26.885	74.155
Prvá stavebná sporiteľňa	-23.9534	25.492	-0.940	0.348	-74.014	26.107
Slovenská sporiteľňa	108.8382	28.621	3.803	0.000	52.633	165.043
Tatra banka	-78.5938	25.736	-3.054	0.002	-129.134	-28.054
VÚB banka	-76.6069	25.197	-3.040	0.002	-126.089	-27.125
Wüstenrot Slovensko	-29.8636	25.299	-1.180	0.238	-79.545	19.818
mBank.sk	-15.3092	26.475	-0.578	0.563	-67.301	36.683
numerical_date	-1.6108	0.315	-5.115	0.000	-2.229	-0.992
competition_posts_3	6.0163	3.387	1.776	0.076	-0.635	12.667
competition_posts_6	-1.2362	2.022	-0.611	0.541	-5.207	2.735

Omnibus:	414.213	Durbin-Watson:	1.938
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7328.668
Skew:	2.589	Prob(JB):	0.00
Kurtosis:	18.776	Cond. No.	2.82e+04

Obr. 14: Výsledky regresnej analýzy pre fixed effects model.

OLS Regression Results

Dep. Variable:	prirastok	R-squared:	0.453
Model:	OLS	Adj. R-squared:	0.434
Method:	Least Squares	F-statistic:	23.51
Date:	Mon, 09 Oct 2023	Prob (F-statistic):	1.57e-67
Time:	22:02:51	Log-Likelihood:	-4106.8
No. Observations:	648	AIC:	8260.
Df Residuals:	625	BIC:	8363.
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	9.1744	39.828	0.230	0.818	-69.038	87.386
num_of_reactions	0.0009	0.005	0.182	0.856	-0.009	0.011
num_posts	5.4822	1.167	4.699	0.000	3.191	7.773
shares	0.0777	0.027	2.922	0.004	0.025	0.130
comments	0.0329	0.006	5.643	0.000	0.021	0.044
pocet_fin	26.7919	6.625	4.044	0.000	13.781	39.803
pocet_nefin	2.1203	10.913	0.194	0.846	-19.310	23.551
gasoline_price	70.3429	34.394	2.045	0.041	2.802	137.884
new_mortgage_ptf_m	-0.0151	0.016	-0.961	0.337	-0.046	0.016
avg_salary_ptf	-0.0366	0.031	-1.186	0.236	-0.097	0.024
change_inflation_q	-326.4887	522.415	-0.625	0.532	-1352.390	699.413
365.bank	89.5213	25.095	3.567	0.000	40.241	138.801
Cetelem Slovensko	-28.5914	24.954	-1.146	0.252	-77.596	20.413
Poštová banka	22.3136	25.362	0.880	0.379	-27.491	72.118
Prvá stavebná sporiteľňa	-25.0297	25.136	-0.996	0.320	-74.391	24.332
Slovenská sporiteľňa	109.4917	28.296	3.869	0.000	53.925	165.059
Tatra banka	-79.9483	25.327	-3.157	0.002	-129.684	-30.213
VÚB banka	-76.5019	24.830	-3.081	0.002	-125.262	-27.742
Wüstenrot Slovensko	-31.8168	24.932	-1.276	0.202	-80.778	17.145
mBank.sk	-16.7759	26.098	-0.643	0.521	-68.026	34.474
numerical_date	-1.3665	0.354	-3.861	0.000	-2.061	-0.672
competition_posts_3	5.9565	3.367	1.769	0.077	-0.655	12.568
competition_posts_6	-1.2917	2.009	-0.643	0.520	-5.236	2.653

Omnibus:	421.127	Durbin-Watson:	1.970
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7615.837
Skew:	2.588	Prob(JB):	0.00
Kurtosis:	18.977	Cond. No.	2.65e+05

Obr. 15: Výsledky regresnej analýzy pre fixed effects model po zahrnutí ekonomických premenných.

2.3.2 LDV model

Rozdielom pri odhadovaní koeficientov modelu 2 v programe *Python* od modelu 1 je pridanie oneskorenej premennej s názvom *prirastok_ lagged*.

- $y_{i,t}$ je prírastok fanúšikov v čase t pre jednotku pre jednotlivé banky,
- α_i je fixný efekt banky,

- $x_{i,t}$ je počet príspevkov, reakcií, komentárov, zdieľaní, všetkých súťaží dokopy, počet finančných súťaží a počet nefinančných súťaží,
- β_1 je koeficient pre nezávislú premennú $x_{i,t}$,
- ϕ_1 je koeficient pre prírastok fanúšikov v čase $t-1$,
- $y_{i,t-1}$ je hodnota prírastku fanúšikov v čase $t-1$ pre jednotlivé banky,
- $e_{i,t}$ je náhodná chyba.

Na obrázku 16 vidíme tabuľku s výsledkami regresnej analýzy. Z prvej časti tabuľky nás bude zaujímať *Adj. R - squared*, ktoré nám vyšlo 47,9 %, čo je viac ako v prípade modelu *Fixed effects* a hodnota *Prob (F - statistic)* nám vyšla ešte bližšia nule ako v predchádzajúcom modeli. Opäť teda môžeme tvrdiť, že náš model dostatočne popisuje našu závislú premennú a nezávislé premenné majú na ňu významný vplyv.

Opäť sme si pre náš model zvolili ČSOB ako referenčnú premennú, avšak v tomto prípade nám konštanta nevyšla ako štatisticky významná 16 a teda základná úroveň závislej premennej nie je rozlíšiteľná od 0. Z tohto dôvodu sme sa rozhodli zvoliť 365.bank ako referenčnú hodnotu, pri ktorej koeficient už štatisticky významný je 17 a jeho hodnota je 95.0750.

Štatisticky významné nezávislé premenné nám vyšli rovnaké ako pri predchádzajúcom modeli 14. Opäť teda množstvo príspevkov, komentáre, zdieľania a počet finančných súťaží má pozitívny vplyv na rast počtu fanúšikov. Okrem vymenovaných má pozitívny vplyv na počet fanúšikov aj veľkosť prírastku z minulého mesiaca. Výsledky regresie nám hovoria, že keby boli ostatné hodnoty nulové, tak by bol prírastok fanúšikov napríklad pre 365 banku súčet hodnoty konštanty a takmer 30 % z prírastku fanúšikov predchádzajúceho mesiaca.

Čo sa týka zvyšných nezávislých premenných, ani v tomto prípade nemá množstvo reakcií, nefinančných súťaží, súťaží pred 3 a 6 mesiacmi významný vplyv na rast fanúšikov.

Štatisticky významný koeficient pri bankách nemá iba *Slovenská sporiteľňa*. Ostatné banky sú štatistiky významné.

V tabuľke 2 vidíme, aké by mali prírastky fanúšikov štatisticky významné banky po regresnej analýze pre LDV model. Je z nej jasne vidieť, že 365 banka by mala najvyšší prírastok fanúšikov v prípade, že by na svoju sociálnu sieť nič nepublikovala.

Tatra banka a VÚB banka by v takomto prípade mali úbytok odoberateľov a museli by publikovať príspevky, aby nestrácali fanúšikov.

Názov banky	Prírastok fanúšikov
365.bank	95,075
Cetelem Slovensko	16,0284
Poštová banka	32,6639
Prvá stavebná sporiteľňa	10,8179
Tara banka	-41,1726
VÚB banka	-27,2373
Wüstenrot Slovensko	15,2609
mBank.sk	27,4524
ČSOB	30,064

Tabuľka 2: Prírastky fanúšikov po regresnej analýze pre LDV model.

Pozrime sa aj pri tejto analýze na to, ako by vyzerali koeficienty, keby sme k nezávislým premenným pridali aj premenné týkajúce sa ekonomickej situácie. V tomto prípade nám okrem priemernej ceny benzínu za minulý mesiac vyšiel štatisticky významný koeficient aj v prípade priemerného platu klientov.

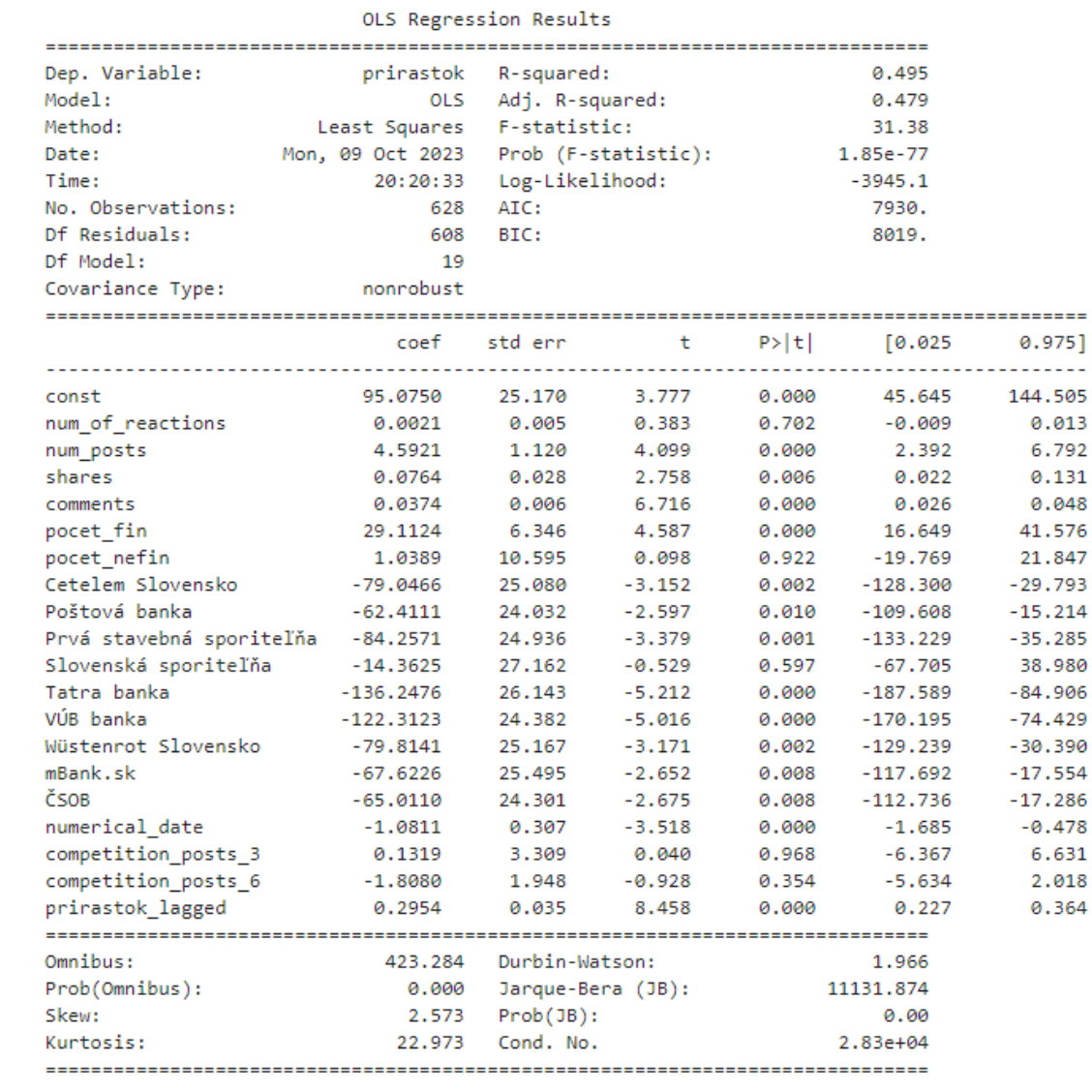
OLS Regression Results

Dep. Variable:	prirastok	R-squared:	0.495
Model:	OLS	Adj. R-squared:	0.479
Method:	Least Squares	F-statistic:	31.38
Date:	Mon, 09 Oct 2023	Prob (F-statistic):	1.85e-77
Time:	20:13:32	Log-Likelihood:	-3945.1
No. Observations:	628	AIC:	7930.
Df Residuals:	608	BIC:	8019.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	30.0640	24.276	1.238	0.216	-17.611	77.739
num_of_reactions	0.0021	0.005	0.383	0.702	-0.009	0.013
num_posts	4.5921	1.120	4.099	0.000	2.392	6.792
shares	0.0764	0.028	2.758	0.006	0.022	0.131
comments	0.0374	0.006	6.716	0.000	0.026	0.048
pocet_fin	29.1124	6.346	4.587	0.000	16.649	41.576
pocet_nefin	1.0389	10.595	0.098	0.922	-19.769	21.847
365.bank	65.0110	24.301	2.675	0.008	17.286	112.736
Cetelem Slovensko	-14.0356	24.052	-0.584	0.560	-61.271	33.199
Poštová banka	2.5999	24.352	0.107	0.915	-45.225	50.425
Prvá stavebná sporiteľňa	-19.2462	24.237	-0.794	0.427	-66.844	28.352
Slovenská sporiteľňa	50.6484	27.930	1.813	0.070	-4.203	105.500
Tatra banka	-71.2366	24.381	-2.922	0.004	-119.118	-23.356
VÚB banka	-57.3013	23.904	-2.397	0.017	-104.246	-10.356
Wüstenrot Slovensko	-14.8031	24.026	-0.616	0.538	-61.986	32.380
mBank.sk	-2.6117	25.176	-0.104	0.917	-52.053	46.830
numerical_date	-1.0811	0.307	-3.518	0.000	-1.685	-0.478
competition_posts_3	0.1319	3.309	0.040	0.968	-6.367	6.631
competition_posts_6	-1.8080	1.948	-0.928	0.354	-5.634	2.018
prirastok_lagged	0.2954	0.035	8.458	0.000	0.227	0.364

Omnibus:	423.284	Durbin-Watson:	1.966
Prob(Omnibus):	0.000	Jarque-Bera (JB):	11131.874
Skew:	2.573	Prob(JB):	0.00
Kurtosis:	22.973	Cond. No.	2.71e+04

Obr. 16: Výsledky regresnej analýzy pre LDV model s referenčnou premennou ČSOB.



OLS Regression Results

Dep. Variable:	prirastok	R-squared:	0.495
Model:	OLS	Adj. R-squared:	0.479
Method:	Least Squares	F-statistic:	31.38
Date:	Mon, 09 Oct 2023	Prob (F-statistic):	1.85e-77
Time:	20:20:33	Log-Likelihood:	-3945.1
No. Observations:	628	AIC:	7930.
Df Residuals:	608	BIC:	8019.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	95.0750	25.170	3.777	0.000	45.645	144.505
num_of_reactions	0.0021	0.005	0.383	0.702	-0.009	0.013
num_posts	4.5921	1.120	4.099	0.000	2.392	6.792
shares	0.0764	0.028	2.758	0.006	0.022	0.131
comments	0.0374	0.006	6.716	0.000	0.026	0.048
pocet_fin	29.1124	6.346	4.587	0.000	16.649	41.576
pocet_nefin	1.0389	10.595	0.098	0.922	-19.769	21.847
Cetelem Slovensko	-79.0466	25.080	-3.152	0.002	-128.300	-29.793
Poštová banka	-62.4111	24.032	-2.597	0.010	-109.608	-15.214
Prvá stavebná sporiteľňa	-84.2571	24.936	-3.379	0.001	-133.229	-35.285
Slovenská sporiteľňa	-14.3625	27.162	-0.529	0.597	-67.705	38.980
Tatra banka	-136.2476	26.143	-5.212	0.000	-187.589	-84.906
VÚB banka	-122.3123	24.382	-5.016	0.000	-170.195	-74.429
Wüstenrot Slovensko	-79.8141	25.167	-3.171	0.002	-129.239	-30.390
mBank.sk	-67.6226	25.495	-2.652	0.008	-117.692	-17.554
ČSOB	-65.0110	24.301	-2.675	0.008	-112.736	-17.286
numerical_date	-1.0811	0.307	-3.518	0.000	-1.685	-0.478
competition_posts_3	0.1319	3.309	0.040	0.968	-6.367	6.631
competition_posts_6	-1.8080	1.948	-0.928	0.354	-5.634	2.018
prirastok_lagged	0.2954	0.035	8.458	0.000	0.227	0.364

Omnibus:	423.284	Durbin-Watson:	1.966
Prob(Omnibus):	0.000	Jarque-Bera (JB):	11131.874
Skew:	2.573	Prob(JB):	0.00
Kurtosis:	22.973	Cond. No.	2.83e+04

Obr. 17: Výsledky regresnej analýzy pre LDV model s referenčnou premennou 365.bank.

OLS Regression Results

Dep. Variable:	prirastok	R-squared:	0.504
Model:	OLS	Adj. R-squared:	0.485
Method:	Least Squares	F-statistic:	26.71
Date:	Mon, 09 Oct 2023	Prob (F-statistic):	8.34e-77
Time:	22:01:34	Log-Likelihood:	-3939.4
No. Observations:	628	AIC:	7927.
Df Residuals:	604	BIC:	8033.
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	71.3735	39.481	1.808	0.071	-6.163	148.910
num_of_reactions	0.0005	0.005	0.091	0.928	-0.010	0.011
num_posts	4.4653	1.126	3.965	0.000	2.253	6.677
shares	0.0865	0.028	3.115	0.002	0.032	0.141
comments	0.0371	0.006	6.688	0.000	0.026	0.048
pocet_fin	28.7692	6.337	4.540	0.000	16.324	41.214
pocet_nefin	0.5092	10.553	0.048	0.962	-20.215	21.233
gasoline_price	129.2125	40.333	3.204	0.001	50.003	208.422
new_mortgage_ptf_m	-0.0157	0.015	-1.063	0.288	-0.045	0.013
avg_salary_ptf	-0.1156	0.044	-2.648	0.008	-0.201	-0.030
change_inflation_q	-679.3260	506.999	-1.340	0.181	-1675.022	316.370
Cetelem Slovensko	-81.6639	24.971	-3.270	0.001	-130.704	-32.624
Poštová banka	-63.1930	23.896	-2.645	0.008	-110.122	-16.264
Prvá stavebná sporiteľňa	-84.2862	24.795	-3.399	0.001	-132.981	-35.591
Slovenská sporiteľňa	-11.3354	27.030	-0.419	0.675	-64.419	41.748
Tatra banka	-139.4258	26.032	-5.356	0.000	-190.549	-88.302
VÚB banka	-123.5412	24.248	-5.095	0.000	-171.161	-75.921
Wüstenrot Slovensko	-81.9728	25.049	-3.273	0.001	-131.166	-32.780
mBank.sk	-67.2172	25.351	-2.651	0.008	-117.004	-17.430
ČSOB	-66.4200	24.170	-2.748	0.006	-113.888	-18.953
numerical_date	-0.7713	0.344	-2.243	0.025	-1.447	-0.096
competition_posts_3	-0.0367	3.300	-0.011	0.991	-6.518	6.444
competition_posts_6	-1.8863	1.940	-0.972	0.331	-5.697	1.924
prirastok_lagged	0.2888	0.035	8.291	0.000	0.220	0.357

Omnibus:	420.892	Durbin-Watson:	2.003
Prob(Omnibus):	0.000	Jarque-Bera (JB):	10881.610
Skew:	2.558	Prob(JB):	0.00
Kurtosis:	22.740	Cond. No.	2.62e+05

Obr. 18: Výsledky regresnej analýzy pre LDV model s referenčnou premennou 365.bank po pridaní ekonomických premenných.

V tabuľke 3 vidíme long - run efekt pre nezávislé premenné, ktoré model vyhodnotil ako štatisticky významné. Tento efekt sme počítali pomocou rovnice 3, pričom nám autoregresný koeficient vyšiel rovný 0,2995. Najväčší long - run efekt majú finančné súťaže. Na 100 takých príspevkov/súťaží privedie takmer 41 nových fanúšikov v dlhodobom horizonte. Druhý najväčší long - run efekt má počet príspevkov, kde 100 príspevkov prinesie v dlhodobom horizonte takmer 7 ďalších fanúšikov.

Premenná	Long - run effect
num_posts	6,46501
shares	0,10756
comments	0,05265
pocet_fin	40,98606

Tabuľka 3: Long - run efekt pre štatisticky významné nezávislé premenné.

Aj pri tomto modeli sa nám na obrázku 17 potvrdilo, že model nám dostatočne popisuje mesačný prírastok fanúšikov a nezávislé premenné majú štatisticky významný vplyv na závislú premennú.

Ukázali sme si aj long - run efekt pre jednotlivé štatisticky významné premenné, kde sa ukázalo, že finančné súťaže majú tento efekt najväčší.

Odporúčania po vykonaní tohto testu zostávajú rovnaké, avšak berúc ohľad na long - run efekt, firmy by sa mali viac zameriavať na finančné súťaže a množstvo príspevkov.

2.3.3 Porovnanie modelov

Obidva modely sa podľa *R - squared* a *F - statistic* zdajú byť dobré. Nedostatkami modelov sú pozitívna autokorelácia rezíduí, ktorú vidíme podľa Durbin - Watson testu či nesplnený predpoklad o normálnom rozdelení, o čom nám hovorí Jarque - Bera. Ak nevezmeme do úvahy vyššie spomínané nedostatky, oba modely dostatočne dobre odhaľujú, čo majú banky robiť, aby prilákali väčšie množstvo fanúšikov.

Ak by sme mali vybrať iba jeden z modelov, podľa ktorého by sme sa riadili, vybrali by sme LDV model, hlavne vďaka tomu, že nám regresná analýza určila koeficienty pri takmer všetkých bankách ako štatisticky významné. Dokonca aj oneskorená hodnota prírastku mala štatistickú významnosť, a teda vieme následne brať do úvahy aj long - run efekt.

3 Modelovanie interakcií s príspevkami

3.1 Príprava dát

Pri modelovaní interakcií budeme používať tabuľku, v ktorej sa nachádzajú informácie o vypublikovaných príspevkoch firiem finančného sektora. Konkrétnejšie sa jedná o firmu, ktorá ho pridala, počet všetkých typov reakcií, ktoré príspevok získal a taktiež aj jeho popis. Tieto informácie sú zbierané s mesačným odstupom. Tentokrát nebudeme robiť žiadnu selekciu, ale použijeme všetky dáta, ktoré k dispozícii máme. Týchto príspevkov je 13 776, čo nám poskytne dostatočne dobrý prehľad o správaní používateľov.

3.1.1 Konštrukcia závislej premennej

Pod pojmom reakcia alebo interakcia sa myslí komentovanie príspevku, zdieľanie a vyjadrenie emócie, čím sa myslí označenie páči sa mi to, haha, love, sorry, anger a wow, o ktorých sme písali aj v kapitole 1.2.1. Reakcie na príspevky sú pre nás dôležitou metrikou, pretože pozitívne ovplyvňujú Facebookový algoritmus tak, aby sa zobrazovali väčšiemu počtu ľudí, čo zvyšuje dosah komunikácie a robí kanál efektívnejším.

Keďže nás nezaujímajú typy reakcií jednotlivo, ale chceme vedieť ako ľudia celkovo interagujú s príspevkami, vytvoríme si závislú premennú, ktorá bude obsahovať všetky interakcie spomenuté vyššie. Vzhľadom na to, že každá reakcia vplýva na zobrazovanie príspevku inak, túto premennú sme vytvorili ako súčet jednotlivých interakcií prenásobených o prislúchajúce konštanty.

O tom, prečo je dôležité chápať správanie ľudí na internete sa píše v článku [19]. Pri výbere vhodných konštánt sme sa preto zamýšľali nad tým, akú veľkú námahu musia fanúšikovia vynaložiť, aby reagovali na nejaký príspevok. Táto problematika je veľmi dobre zhrnutá v článku [12], z ktorého sme čerpali hodnoty jednotlivých váh.

V tabuľke 4 vidíme prehľad zvolených konštánt. „Páči sa mi to“ má váhu 1,

pretože je to najzákladnejšia reakcia, pri ktorej používateľ nemusí vynaložiť žiadnu námahu, iba vyjadrí názor, že sa mu páči príspevok, resp. s ním súhlasí alebo dáva na javo, že berie na vedomie informáciu, ktorá sa v príspevku nachádza. Reakcie ako sú haha, love, sorry, anger a wow sú veľmi podobné tej predchádzajúcej, avšak pri týchto fanúšik vyjadruje svoj názor na daný príspevok konkrétnejší a taktiež musí stráviť nejaký čas pri reagovaní, pretože táto možnosť sa nezobrazí hneď.

Komentovanie príspevkov slúži na konkrétne vyjadrenie názoru na danú problematiku, ktorá sa v príspevku rozoberá. Pre nás je to veľmi dôležitým signálom, aby sme vedeli posúdiť, či používatelia tejto sociálnej siete súhlasia s tým, čo sa publikovalo alebo nie. Taktiež niekedy stačí iba jediný komentár, ktorý vie veľmi dobre rozprúdiť debatu, čo opäť napomáha príspevku, aby sa dostal medzi viacero ľudí.

Poslednou metrikou je zdieľanie, ktorého váha je najväčšia. Keď používateľ zdieľa nejaký príspevok, dáva najavo, že je s ním stotožnený a zobrazí sa aj na jeho profile. To je potom pre banku veľmi dobré, pretože sa tak príspevok môže zobraziť ďalším ľuďom - Facebookovým priateľom používateľa, ktorý príspevok zdieľal, čím sa zvyšuje povedomie o konkrétnej banke, ktorá príspevok publikovala.

Typ reakcie	Hodnota konštanty
Páči sa mi to	1
Like, Haha, Anger, Wow, Sorry	3
Komentár	7
Zdieľanie	14

Tabuľka 4: Typy reakcií s prislúchajúcimi váhami.

3.1.2 Nezávislé premenné

Premenné, ktoré nás budú zaujímať rozdelíme do dvoch kategórií. Prvou sú spojité premenné a druhou kategorické, pričom volíme také premenné, ktoré vieme aj ovplyvniť. **Spojité premené** sú:

- **day_of_year** je poradie dňa v roku, ktorý sme získali z dátumu,

- **cas_min** je čas pridania príspevku v minútach,
- **pocet_znakov** je dĺžka celého popisu príspevku,
- **pocet_slov** je množstvo slov, ktoré boli použité v príspevku,
- **vzdialenost_od_kategórie** je vzdialenosť popisu príspevku od centra konkrétnej kategórie príspevku. Tejto premennej sa budeme venovať v kategórii 3.2.3, keďže tému, respektíve kategóriu, príspevku nemáme vopred danú a musíme ju dodatočne zanalyzovať.

Kategorické premenné sú:

- **sviatok** nám určuje, či je v daný deň štátny sviatok na Slovensku alebo nie,
- **type** je typ príspevku:
  - **album** je priečinok/album, v ktorom sa nachádzajú videá alebo fotky,
  - **link** je internetový odkaz,
  - **photo** je fotografia,
  - **status** je čistý text,
  - **video** je video.

Medzi kategorické premenné patrí aj *chyba_cas*, čo nám hovorí o tom, či disponujeme informáciou o čase pridania príspevku, keďže nie pri všetkých príspevkoch máme aj jeho čas zverejnenia.

Vzhľadom na to, že maximálny počet dní sa líši od toho, či sa jedná o priestupný rok alebo nie, museli sme ešte premennú *day_of_year* dodatočne upraviť. Dáta sme teda štandardizovali na hodnotu 365. Pokiaľ sa jednalo o priestupný rok, predelili sme túto premennú hodnotou 366 a následne prenásobili konštantou 365.

3.2 Modelovacie techniky

V tejto podkapitole si priblížime metódy, ktoré bolo potrebné použiť na to, aby sme vedeli učiť kategóriu príspevku. Kategórie sú ďalším ukazovateľom, ktoré nám dajú prehľad o tom, aký obsah je pre fanúšikov firiem finančného sektora zaujímavý a aký nie. Toto nám pomôže viac pochopiť, čo ľudí zaujíma, akým témam by sa mali banky venovať a aké témy by mali obmedziť, respektíve ich inak podať.

3.2.1 SlovakBERT model

Aby sme texty príspevkov dokázali kategorizovať a následne zhodnotiť ich význam, rozhodli sme sa ich najskôr vektorovo reprezentovať, aby sme s nimi vedeli ďalej pracovať. Na túto reprezentáciu sme použili model SlovakBERT [13], ktorý podľa [21] patrí medzi najkvalitnejšie metódy. Tento model je umelá neurónová sieť, ktorá bola natrénovaná na rôznorodom množstve slovenských textov dostupných na internete. Vďaka nemu teda reprezentujeme popisy príspevkov do priestoru o veľkosti 768 [8].

My budeme používať tento predtrénovaný model [8] a následne urobíme jeho doladenie na našich vlastných dátach, ktorými sú texty príspevkov. Model síce vie rozoznávať veľké a malé písmená, avšak s emotikonmi pracovať nevie, a teda sme ich z popisov museli odstrániť. Konkrétne použitie modelu spolu s našou voľbou parametrov si ukážeme v kapitole 3.3.1.

Architektúra BERT sa skladá z viacerých vrstiev transformátorov, ktoré sú vzájomne prepojené. Transformátory sú špeciálny typ neurónovej siete, ktoré dokážu efektívne zachytiť dlhodobé závislosti medzi slovami v texte. BERT sa predbežne trénuje na obrovskom súbore textov a kódov, čím sa naučí reprezentovať význam slov a fráz v kontexte. Hlavnými parametrami modelu BERT sú:

- **Zásobník kódovačov z Transformeru**: BERT využíva viacero kódovacích blokov z architektúry Transformer, ktoré sú naskladané na seba. Tieto bloky analyzujú a kódujú vstupný text, pričom berú do úvahy vzťahy medzi slovami.

- **Počet vrstiev**: Určuje, koľkokrát sa informácie v texte prenesú cez transformátory. Väčší počet vrstiev vedie k presnejšiemu modelu, ale aj k náročnejšiemu výpočtu.
- **Veľkosť skrytej vrstvy**: Určuje počet neurónov v skrytej vrstve transformátorov. Väčší počet vrstiev umožňuje modelu zachytiť komplexnejšie informácie o texte.
- **Počet pozorných hláv *(attention heads)***: Určuje, koľko rôznych perspektív sa pri analýze textu používa. Väčší počet vedie k presnejšiemu modelu, ale aj k náročnejšiemu výpočtu.

Model sa trénuje v dvoch fázach:

1. **Predbežné trénovanie**: model sa trénuje na dvoch úlohách:
   - **Maskovaný jazykový model (MLM)**: učí predpovedať náhodne zamaskované slová v texte.
   - **Predpoveď nasledujúcej vety (NSP)**: Model sa učí rozlišovať, či dve vety idú za sebou logicky alebo nie.
2. **Jemné doladenie:** predbežne trénovaný model sa doladí pre konkrétnu úlohu NLP, ako je zodpovedanie otázok alebo sentimentálna analýza. To sa dosiahne pridaním novej výstupnej vrstvy a jemným doladením parametrov modelu.

3.2.2 Analýza hlavných komponentov

Vzhľadom na to, že po aplikovaní SlovakBERT dostaneme maticu o veľkosti $n\ x\ p$, použijeme na ňu Analýzu hlavných komponentov (PCA), čo je štatistická metóda, ktorá sa používa na redukciu dimenzie dát a identifikáciu hlavných komponentov [5], pričom počet komponentov bude p_pca. My ju budeme používať z toho dôvodu, aby sme znížili výpočtovú náročnosť pri ďalšom spracovávaní dát a limitovali počet redundatných premenných pred použitím klastrovacích techník.

PCA sa začína štandardizáciou dát (5), kde od každej hodnoty odpočítame ich priemernú hodnotu a predelíme štandardnou odchýlkou. Toto zabezpečí, že každá premenná bude mať priemernú hodnotu 0 a štandardnú odchýlku rovnú 1. Keby sa tento krok vynechal, tak by to mohlo viesť k skresleným hodnotám, keďže je PCA dosť citlivá na odchýlky počiatočných premenných.

$s = \frac{x - \mu}{\sigma}$, kde x je pôvodná hodnota, μ je priemerná hodnota a σ je štandardná odchýlka (5)

Ďalším krokom je vypočítanie vlastných hodnôt (λ_1, λ_2,..., λ_n) a vlastných vektorov (v_1, v_2,..., v_n) kovariačnej matice, ktorá musí byť kladne semidefinitná. Vlastné hodnoty si následne zoradíme od najväčšej po najmenšiu (napr. $\lambda_1 > \lambda_2 > ... > \lambda_k$). Takto získame hlavné komponenty, ktoré prislúchajú vlastným vektorom jednotlivých vlastných hodnôt ($\mathrm{PC}_1 = v_1$, $\mathrm{PC}_2 = v_2$,..., $\mathrm{PC}_n = v_n$). Prvý komponent vysvetľuje najväčšiu varianciu, zatiaľ čo ostatné komponenty v sebe skrývajú čoraz menšie množstvo variancie. Množstvo pokrytej variancie jednotlivými komponentami (EV) sa potom vypočítajú ako:

$$EV_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

Posledným krokom je transformácia pôvodných dát do nového priestoru pomocou vlastných vektorov a následné určenie počtu hlavných komponentov, ktoré nám vysvetlia pre nás dostatočné množstvo variancie.

3.2.3 K - means

K-means je zhlukovací algoritumus, ktorý zhlukuje dáta na základe ich podobnosti [6]. Princíp zhlukovania je, že chceme nájsť rozdelenie n objektov do k disjunktných skupín $C_1, C_2, ..., C_k$ (6).

$$C_1 \cup ... \cup C_k = \{1, ..., n\}, \text{ ak } i \neq j \rightarrow C_i \cap C_j = \varnothing \tag{6}$$

Účelová funkcia tohto algoritmu spočíva v rozdelení množiny n bodov do k zhlukov tak, aby sa minimalizoval súčet štvorcov vzdialeností bodov od ich zhlukových centier (centroidov) (7), kde ρ je euklidovská vzdialenosť, x_r je vektor premennej a $c_i = |C_i|^{-1} \sum_{r \in C_i} x_r$ je centroid C_i. Centroidy sú reprezentované ako priemery bodov priradených do zhluku.

$$\min \sum_{i=1}^{k} \sum_{r \in C_i} \rho^2(x_r, c_i) \tag{7}$$

Na riešenie problému K - means sa za používa interaktívna metóda s názvom **Loydov algoritmus**:

1. *Náhodne sa vyberú počiatočné centroidy.*
2. *Striedavo sa vykonajú nasledovné kroky:*
   a) *Každý bod sa priradí k najbližšiemu centroidu*
   b) *Každý centroid sa aktualizuje na základe vypočítania priemeru bodov priradených k tomuto centroidu.*
3. *Toto opakujeme, až pokiaľ nedostaneme tie isté zhluky, teda prvky sa v ďalšej iterácii priradia do rovnakých zhlukov.*

Najväčšou výhodou K - means je, že je rýchla aj pre veľké dátové sady a je ľahká na implementáciu. Medzi nevýhody patrí predpoklad konvexného zhluku a počet zhlukov, ktorý je potrebný určiť vopred.

3.2.4 DBScan

Density - based scan (DBScan) je klastrovací algoritmus, ktorý je založený na identifikácii zhlukov na princípe hustoty [6]. Hlavné parametre použité v algoritme sú:

- ϵ - polomer okolia,
- **MinPoints** - minimálny počet bodov v zhluku,

vďaka ktorým sú body rozdelené do 3 skupín:

- **Core** - bod, v ktorého okolí je počet bodov > MinPoints,
- **Border** - bod, v ktorého okolí je počet bodov $\leq$ MinPoints,
- **Noise** - bod, v ktorého okolí nie je žiaden bod.

Algoritmus zhlukovania funguje nasledovne:

1. Vyberie sa náhodný bod. Ak je v jeho ϵ - okolí počet bodov $\geq$ MinPoints, vytvorí sa zhluk.
2. Prehĺadajú sa susedia všetkých susedov tohto bodu. Ak má prehliadaný bod v ϵ - okolí počet susedov $\geq$ MinPoints, je priradený k zhluku a jeho susedia sa zaradia do zoznamu bodov na prehliadanie.
3. Ak bod nemá v ϵ - okolí počet bodov $\geq$ MinPoints, označí sa ako šum.
4. Vyberie sa ďalší bod, ktorý ešte nebol prehľadávaný.

Jeho hlavnou výhodou je, že dokáže identifikovať zhluky rôznych tvarov a veľkostí. Taktiež vie odhaliť body (šum / noise), ktoré nepatria do žiadneho zhluku. Nevýhodou je, že je výpočtovo náročnejší a citlivý na nastavenie parametrov.

3.2.5 GAM

Generalized Additive Model (GAM) je štatistická metóda, ktorá sa používa na modelovanie vzťahov medzi premennými [7]. Jeho výhodou je, že je flexibilnejší a dokáže zachytiť aj nelineárne vzťahy medzi premennými. Z regresného hľadiska je jeho forma nasledovná:

$$E(Y|X_1, X_2, ..., X_p) = \alpha + f_1(X_1) + f_2(X_2) + ... + f_p(X_p),$$

kde X_1, X_2,...,X_p predstavujú prediktory (nezávislé premenné), Y je predikovaná hodnota (závislá premenná) a f_1, f_2,...,f_p sú hladké (neparametrické) funkcie.

Všeobecne platí, že podmienený stred $\mu(X)$ premennej Y súvisí s aditívnou funkciou prediktorov prostredníctvom *link function* g:

$$g(\mu(Y)) = \alpha + f_1(X_1) + f_2(X_2) + ... + f_p(X_p)$$

Príklady $linkfunction$ funkcie sú nasledovné:

- $g(\mu) = \mu$ je *identity link*, ktorá sa používa na predikované dáta s Gaussovým rozdelením
- $g(\mu) = \text{logit}(\mu)$ alebo $g(\mu) = \text{probit}(\mu)$ pre binomické rozdelenie, pričom *probit* je inverzná Gaussovská kumulatívná funkcia rozdelenia,
- $g(\mu) = \log(\mu)$ pre dáta z Poissonového rozdelenia.

My budeme v našej práci robiť výpočty v programovacom jazyku *Python*, ktorý obsahuje knižnicu *pyGAM* [16], ktorá slúži práve na tento typ modelu. *pyGAM* podporuje nasledujúce transformácie nezávislých premenných:

- l() - lineárny,
- s() - splajn,
- f() - faktor,
- te() - tensor,
- Intercept.

My budeme používať práve prvé dva typy transformácií, keďže budeme modelovať spojité a kategorické premenné. Konkrétny výber parametrov, ktoré táto knižnica obsahuje a budeme ich aj používať, si predstavíme v kapitole 3.5.

3.3 Vektorová reprezentácia textu

V tejto podkapitole sa budeme venovať vektorovej reprezentácii popisov príspevkov, aby sme ich mohli neskôr matematicky analyzovať. Túto reprezentáciu budeme robiť pomocou SlovakBERT modelu a budeme ju robiť na všetkých našich dátach. Keďže dostaneme veľkú dimenziu, použijeme následne na dáta analýzu hlavných komponentov, aby sme túto dimenziu zredukovali a dali do interpretovateľnejšej podoby.

3.3.1 Použitie SlovakBERT modelu

Ako sme písali v kapitole 3.2.1, tento model budeme používať na vektorovú reprezentáciu popisov príspevkov, aby sme ich mohli následne modelovať. Predtým, ako ho budeme trénovať na našich dátach, musíme importovať nasledujúce knižnice:

```
from transformers import RobertaTokenizer, RobertaModel, TFRobertaModel
from transformers import RobertaTokenizerFast, RobertaConfig,
RobertaForMaskedLM, DataCollatorForLanguageModeling
from transformers import Trainer, TrainingArguments, LineByLineTextDataset
import torch
```

Pred samotným trénovaním sme ako prvé definovali parametre štyroch funkcií. Prvá má názov *data_ collector*, ktorá slúži na načítanie predspracovaných dát. Tam použijeme už natrénovaný *tokenizer*, ktorý konvertuje text na sekvenciu tokenov, ktoré sú potom použité na vstup pre model.

Funkcia *dataset* obsahuje naše dáta, na ktorých budeme trénovať model. V nej opäť zadávame rovnaký *tokenizer* a ďalej cestu k nášmu datasetu (popisy príspevkov). Nakoniec sa určí maximálna dĺžka sekvencie tokenov, ktorú vzhľadom na veľkosť nášho

datasetu nastavíme na odporúčanú hodnotu 512. Čím je táto hodnota väčšia, tým je natrénovanie výpočtovo náročnejšie.

Ďalšia funkcia slúži na definovanie rôznych parametrov a nastavení, ktoré ovplyvňujú proces trénovania modelu. Jej hlavné parametre sú:

- **output_dir:** určuje cestu k adresáru, do ktorého sa uložia výsledky modelu,
- **num_train_epochs:** určuje koľko krát prejde model celý trénovací dataset, kvôli výpočtovej náročnosti sme zvolili hodnotu 3,
- **per_device_train_batch_size:** určuje koľko príkladov sa spracuje pred aktualizáciou váh modelu, hodnotu sme zvolili 10,
- **save_total_limit:** určuje maximálny počet kontrolných bodov modelu, ktoré sa ukladajú počas trénovania, my volíme hodnotu 2,
- **seed:** je náhodný počiatočný stav pre generátor náhodných čísel. Ak je nastavený na konkrétnu hodnotu, opakované trénovanie modelu za rovnakých podmienok prinesie rovnaké výsledky. My sme ho nastavili na hodnotu 1,

Posledná funkcia, ktorá bude na trénovanie potrebná, je *trainer* a je zodpovedná za skutočné trénovanie modelu na trénovacích dátach s použitím špecifikovaných nastavení a parametrov, ktorými sú:

- **model:** určuje model, ktorý sa bude trénovať, v našom prípade to bude už spomínaný SlovakBERT,
- **args:** je objekt, ktorý obsahuje nastavenia tréningu, ktoré sme definovali pomocou funkcie *training_args*,
- **data_collator:** definuje spôsob, akým sa kombinujú a spracovávajú vstupné dáta, kde jeho hodnotou bude nami definovaná funkcia *data_collator*,

- **train_dataset:** je dátová množina, ktorá obsahuje trénovacie dáta (naša funkcia *dataset*).

Nižšie sa nachádza konkrétne použitie parametrov v spomínaných funkciách:

```
tokenizer = RobertaTokenizerFast.from_pretrained('/slovakbert')
model = RobertaForMaskedLM.from_pretrained('/slovakbert')

data_collator = DataCollatorForLanguageModeling(
    tokenizer=tokenizer, mlm=True, mlm_probability=0.15)

dataset = LineByLineTextDataset(
    tokenizer=tokenizer,
    file_path='/popisy_prispevkov.txt',
    block_size=512,)

training_args = TrainingArguments(
    output_dir='/MojBert',
    overwrite_output_dir=True,
    num_train_epochs=3,
    per_device_train_batch_size=10,
    save_steps=5000,
    save_total_limit=2,
    seed=1)

trainer = Trainer(
    model=model,
    args=training_args,
    data_collator=data_collator,
    train_dataset=dataset)
```

Keď už sme nastavili všetky potrebné parametre, spustíme samotné trénovanie modelu. Toto trénovanie nám trvalo necelých 7 hodín, pričom sme používali operačnú

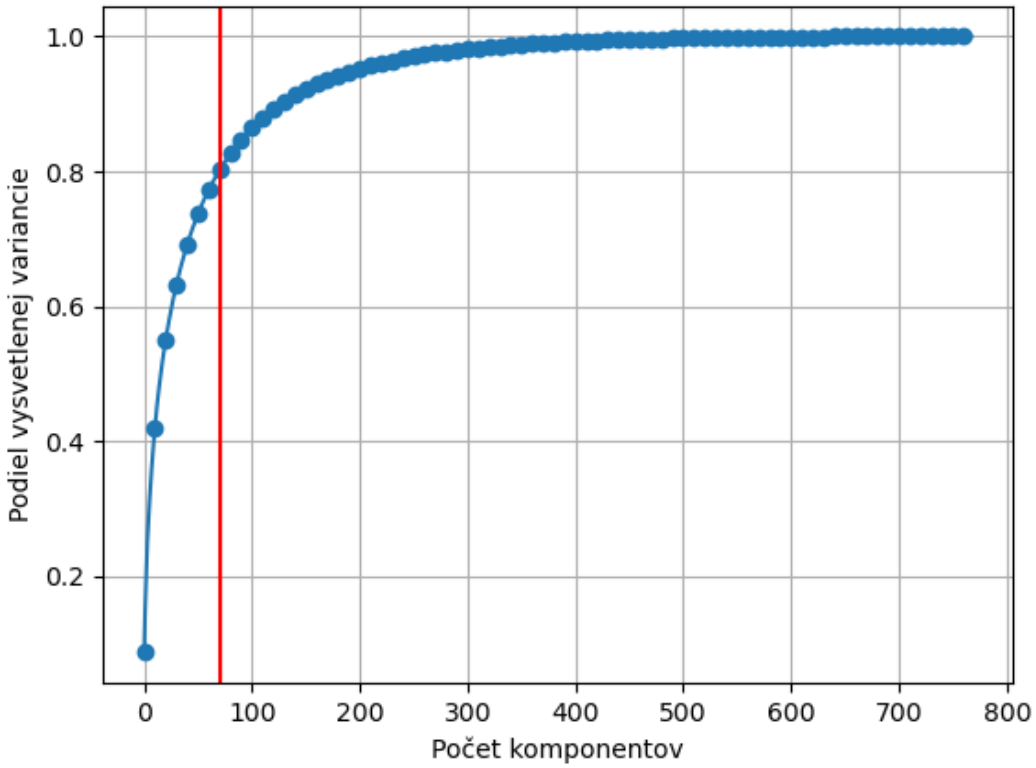
pamäť 16 GB a vCPU = 3. Vďaka nemu sme potom mohli urobiť *embeding* všetkých príspevkov a dostali sme tak ich vektorovú reprezentáciu s dimenziou 768.

3.3.2 Redukcia dimenzie dát

Keďže pri analýze hlavných komponentov nám ide o redukciu dimenzie dát za účelom interpretability a zachovanie čo najväčšieho množstva vysvetlenej variancie, vykreslili sme si lakťový graf, ktorý nám ukazuje závislosť počtu komponentov od spomínanej variancie. Toto sme urobili pomocou funkcie:

```
plt.plot(np.cumsum(pca.explained_variance_ratio_), marker='o',
linestyle='-', markevery=10)
```

Na obrázku 19 nám červená čiara ukazuje, že už pri 70 komponentoch z 768 možných máme vysvetlenej 80 % variancie. Toto percento je pre nás postačujúce, a preto si v ďalšom kroku pretransformujeme naše dáta na 70 rozmerné vďaka funkcii *pca.transform*. Takto zredukované dáta sú pripravené na ďalšiu prácu s nimi.



Obr. 19: Množstvo vysvetlenej variancie pri rôznom množstve hlavných komponentov.

3.4 Zhlukovanie tém

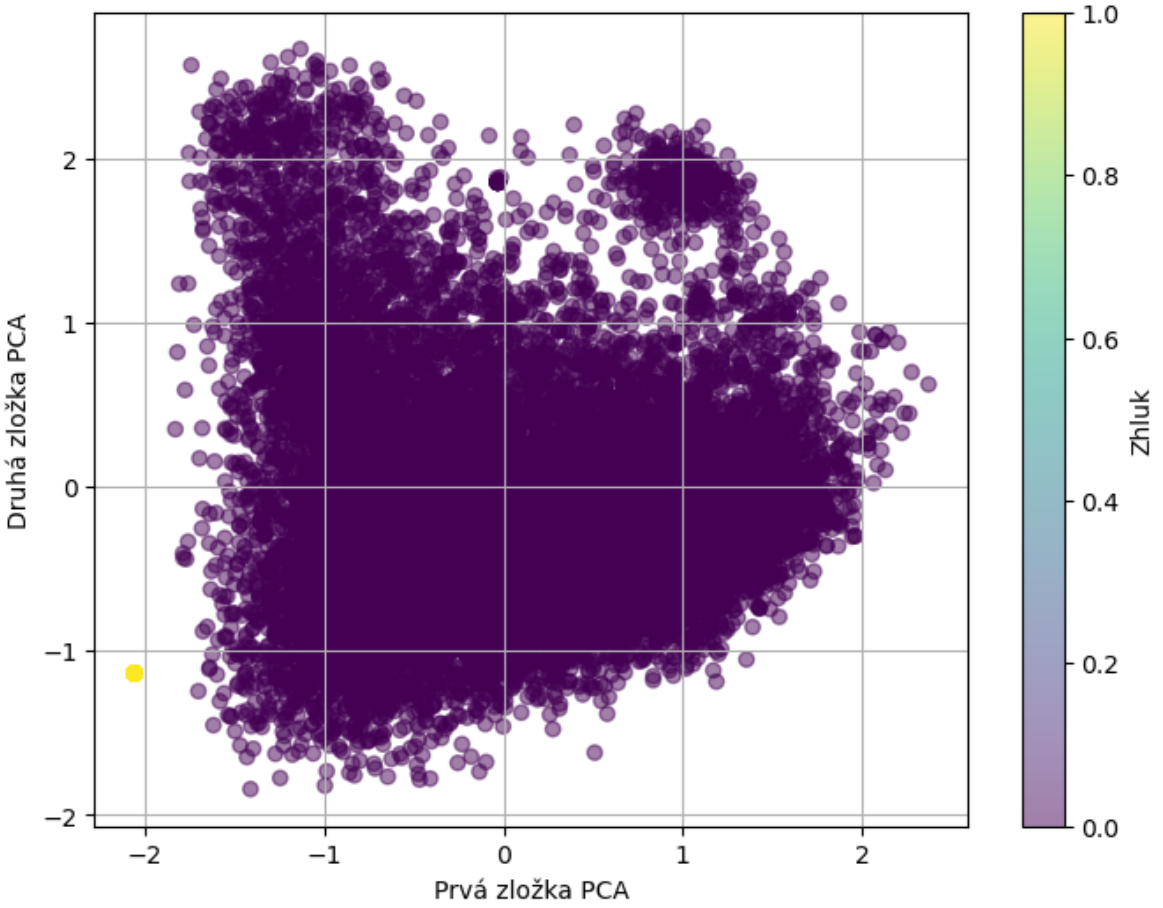
V tejto podkapitole si ukážeme použitie vybraných zhlukovacích tém za účelom rozdelenia popisov príspevkov do rôznych kategórií. Na konci porovnáme obidve metódy a vyberieme tú, ktorá nám lepšie zaradí dáta do zhlukov.

3.4.1 Použitie metódy DBScan

V tejto časti sa budeme snažiť pomocou metódy DBScan rozdeliť príspevky do zhlukov, aby sme vedeli určiť ich kategóriu. Keďže je v tejto metóde veľmi dôležitá voľba parametrov ϵ a *MinPoint*, spomínaných v kapitole 3.2.4, danú metódu sme spustili pre ich rôzne hodnoty.

Ako prvé sme prehľadali parametrický priestor pomocou diskretizácie. Parameter ϵ sme diskretizovali na intervale <1; 3,5>, pričom sa v každom kroku jeho hodnota

zvyšovala o 0,1 a parameter *MinPoints* na intervale <5; 30>, pričom sa zvyšovala o hodnotu 1. Túto metódu sme spúšťali pre všetky kombinácie a chceli sme minimalizovať *noise*, aby sme mali zazhlukovaných čo najviac príspevkov a vedeli im tak určiť kategóriu.



Obr. 20: Zazhlukovanie príspevkov pomocou metódy DBScan pri použití optimálnych parametrov na minimalizáciu šumu.

Optimálne parametre nám vyšli $\epsilon = 3.400000000000002$ a $MinPoints = 5$. Na obrázku 20 vidíme, ako nám táto metóda rozdelila príspevky do zhlukov pomocou prvých dvoch komponentov. Šum minimalizovala na hodnotu 0, avšak až 98 % príspevkov skončilo v jednom zhluku a počet všetkých zhlukov bol dokopy 151.

Aby sme si overili, či boli parametre zvolené správne, vypočítali sme si siluetu, čo je metrika, ktorá hodnotí kvalitu zhlukov. Jej hodnota sa pohybuje od -1 po 1, pričom vyššia hodnota znamená lepšie zaradenie príspevku do jeho zhluku [15]. Hodnota tejto metriky nám vyšla 0,29, z čoho usudzujeme, že príspevky boli zhlukované správne.

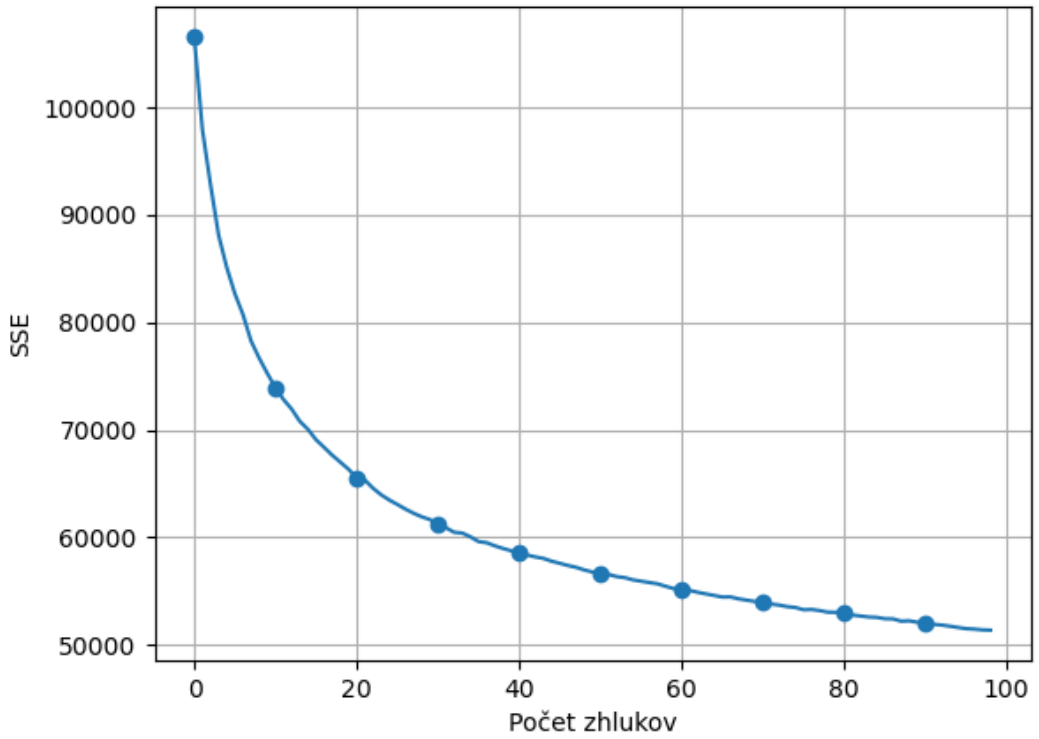
3.4.2 Použitie metódy K–means

Pri tejto zhlukovacej metóde je najdôležitejšia voľba počtu zhlukov. Ako prvé sme spustili túto metódu pre rôzny počet zhlukov z intervalu <1; 100> jedenkrát tak, aby sme minimalizovali súčet štvorcov vzdialeností bodov od ich centroidov.

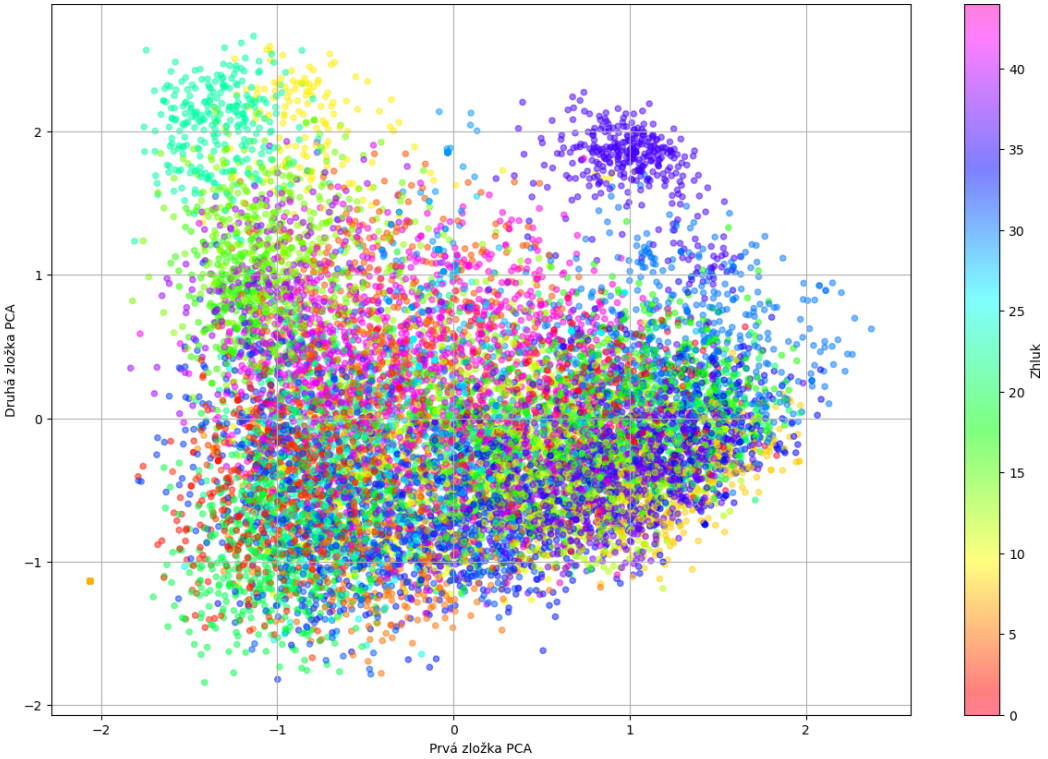
Na obrázku 21 vidíme, že už pri 40 zhlukoch je súčet štvorcov pomerne nízky. Urobili sme multistart metódy pre počet zhlukov od 40 do 60, pričom sme metódu zopakovali 10–krát a maximálny počet iterácií pre každú metódu bol 150. Pri všetkých zhlukoch nám najlepšia silueta vyšla približne 0,10. Počet zhlukov sme preto zvolili 45, aby ich nebolo príliš veľa, ale ani úplne málo.

Na obrázku 22 vidíme, ako nám táto metóda rozdelila príspevky do zhlukov pomocou prvých dvoch komponentov. Síce nevykresľujeme všetky komponenty, ale už na tomto obrázku si môžeme všimnúť, že nám model vytvoril vhodné zhluky.

Keďže jednotlivé príspevky nie sú zamerané iba na jednu konkrétnu tému, ale môžu spadať do viacerých kategórií, vytvorili sme premennú *vzdialenost_ od_ centroid_ k*, pričom $k \in <0; 44>$, ktorá nám hovorí o tom, ako vzdialený je príspevok od centroidu každého zhluku. Tieto premenné nám budú neskôr vstupovať do modelu, aby nám pomohli zistiť, ktoré témy príspevkov vzbudzujú u fanúšikov najväčšiu potrebu interakcie.



Obr. 21: Súčet štvorcov vzdialeností od centroidov v závislosti od počtu zhlukov po aplikovaní metodý K - means.



Obr. 22: Zhlukovanie príspevkov pomocou metódy K - means pri 45 zhlukoch.

3.4.3 Výber zhlukovacej metódy

Výstupom tejto časti je výber správnej zhlukovacej metódy. Vzhľadom na to, že sme chceli rozdeliť popisy príspevkov do zhlukov so zámerom určiť im kategóriu, metóda DBScan nám poskytuje uspokojivé výsledky. Aj napriek tomu, že nám minimalizovala šum na nulu a silueta bola vyššia ako u K-means, rozhodli sme sa ďalej nepracovať s výsledkami tejto metódy, pretože je nezmyselné, aby 98 % príspevkov bolo zameraných na jednu tému, pričom tém bolo spolu zadefinovaných až 151.

Metóda K-means síce nemala veľkú siluetu, avšak bola väčšia ako 0, čo je pre nás v tomto bode postačujúce. Počet zhlukov, ktorý sme pri nej vybrali považujeme za rozumný a vieme vďaka nemu zaradiť témy do rôznych zhlukovacích tém.

Keď sme vybrali vhodnú zhlukovaciu metódu, mohli sme sa pustiť do vytvárania tém, pod ktoré budú príspevky spadať. Z každého zhluku sme si vybrali 10 príspevkov, ktoré boli najbližšie k jeho stredu a pomocou umelej inteligencie od firiem Google AI [1] a OpenAI [3] sme im priradili témy. Postupovali sme pri tom nasledovne:

1. Obidvom umelým inteligenciám (jazykom) sme priradili prompt: *Napíš mi jednu komplexnú tému, ktorou sa zaoberajú nasledujúce príspevky:* + prvých desať príspevkov z daného zhluku.
2. Porovnali sme obidva výsledky a buď sme vybrali jeden z nich alebo sme ich upravili.
3. Toto sme zopakovali pre každý zhluk.

V tabuľke 5 uvádzame prvých 10 tém a zvyšné, spolu aj s príkladmi, sa nachádzajú v Prílohe C.

Téma
Dôležitosť poistenia pre rôzne životné situácie
Podpora komunitných projektov a aktívnej participácie v meste
Vianočné a sviatočné posolstvá
Podpora kultúrnych a spoločenských podujatí
Finančné správy a tipy pre podnikateľov a investorov
Vylepšovanie bankových služieb pomocou mobilných aplikácií
Štatúty súťaží
Príspevok bez textu
Reklamné kampane na pôžičky
Vplyv makroekonomických faktorov na finančné trhy.

Tabuľka 5: Prvých 10 tém, ktorými sa firmy zaoberajú vo svojich príspevkoch.

3.5 Trénovanie GAM modelu

V tejto časti sa budeme venovať už konkrétnemu modelovaniu interakcií fanúšikov s príspevkami firiem finančného sektora. Pre identifikáciu vlastností príspevkov, ktoré vzbudzujú najviac interakcií použijeme GAM model, konkrétne *PoissonGAM*, pretože premenná zodpovedá počtu udalostí a ktorého základy sme vysvetlili v kapitole 3.2.5. Dôvodom, prečo sme si vybrali práve tento model je to, že GAM model uvažuje nelineárny vplyv prediktorov na závislú premennú, čo nám neskôr podporia aj obrázky s našimi výsledkami.

Zvolené nezávislé premenné, ktoré sme použili na predikciu interakcií s príspevkami sme spomenuli v kapitole 3.1.2 a teraz si predstavíme celý postup a voľbu hyperparamentrov pre odhadnutie koeficientov, ktoré sme použili v spomínanom modeli.

Ako prvé sme si museli zvoliť transformácie nezávislých premenných, ktoré budeme na naše dáta používať. My sme zvolili 2 typy, ktorými sú *spline* pre spojité premenné a *linear* pre kategorické premenné.

Ďalším krokom bola voľba parametrov pri týchto transformáciách. Obidvom sme najskôr nechali predvolené hodnoty, ktorými sú:

- n_splines=20 *(počet splajnov použitých na aproximáciu)*,

- spline_order=3 *(rad splajnu)*,
- lam=0.6 *(penalta pre zjemnenie krivky)*,
- penalties='auto' *(typ penalty)*,
- constraints=None *(obmedzenie pre krivku, napr. konkávna, konvexná)*,
- dtype='numerical' *(typ dát, môže byť numerická alebo kategorická)*,
- basis='ps' *(typ bázových funkcií, čo v tomto prípade znamená penalizovaný B-spline)*,
- by=None *(hodnota True sa používa pre periodické krivky)*,
- edge_knots=None *(slúži na určenie vlastných okrajových uzlov)*.

Následne sme vypočítali ich parciálne závislosti, aby sme vedeli určiť, ktorá premenná ako vplýva na množstvo interakcií. Na vykreslenie sme použili predvolenú funkciu, ktorá v sebe obsahuje natrénovaný model:

```
XX = GAM.generate_X_grid(term=i)
plt.plot(XX[:, i], gam_5_splines_skuska.partial_dependence(term=i, X=XX))
plt.plot(XX[:, i], gam_5_splines_skuska.partial_dependence(term=i,
X=XX, width=.95)[1], c='r', ls='--')
```

Táto funkcia nám však neposkytla vizualizácie dát, z ktorých by sme mohli vyvodiť nejaké závery, keďže iba pri prvých dvoch premenných nám vykreslila správne predikcie a pri ostatných nám vykreslila iba vertikálnu čiaru. Z tohto dôvodu sme si vytvorili vlastný kód, ktorý nám vykreslí tieto závislosti pre všetky premenné.

Aby sme mohli vykresliť parciálne závislosti vybranej premennej, museli sme pre ostatné hodnoty zvoliť konštantu. Rozhodli sme, že každá spojitá premenná dostane jej priemernú hodnotu a kategorická premenná bude mať najčastejšie sa vyskytujúcu hodnotu. Premenná, ktorú budeme sledovať bude nadobúdať hodnoty z rozsahu pozorovaných hodnôt. Nami vytvorený kód spolu s poznámkami je nasledovný:

```
num_samples = data_x.shape[0]
num_variables = data_x.shape[1]
title = data.columns

# Cyklus pre premenné
for premenna_index in range(0, num_variables):

    # Štandardizácia dát
    mean = np.mean(data_x)
    std = np.std(data_x)
    standard_data_x = (data_x - mean)/std

    # Inicializácia premenných
    vektor_premennych = np.zeros((num_samples, num_variables))
    avg_premenne = []

    # Výpočet priemerov premenných
    for i in range(0, num_variables):
        avg_premenne.append(np.average(standard_data_x[:num_samples,i]))

    # Priradenie priemerov premenným
    for i in range (0, num_samples):
        vektor_premennych[i,:] = avg_premenne

    # Spracovanie kategórie premenných
    for i in range(0, num_variables):
        if len(np.unique(standard_data_x[:num_samples,i])) == 2:
            vektor_premennych[:,i] =
            st.mode(standard_data_x[:num_samples,i]).mode

    # Priradenie premennej záujmu
    vektor_premennych[:, premenna_index] =
```

```
standard_data_x[:num_samples, premenna_index]

# Predikcia interakcií
predictions = gam_.predict(vektor_premennych)

prediction = pd.DataFrame()
prediction['predictions'] = predictions
prediction['premenna'] = vektor_premennych[:, premenna_index]
prediction = prediction.sort_values(by = 'premenna')

# Vizualizácia výsledkov
fig, ax1 = plt.subplots()

# Transformácia pozorovaných dát na pôvodné hodnoty pred vykreslením
vektor_premennych[:, premenna_index] =
(vektor_premennych[:, premenna_index])*std + mean
prediction['premenna'] = prediction['premenna']*std + mean

# Vykreslenie hranice záujmu
plt.axvline(x = (max(prediction['premenna']) -
min(prediction['premenna']))/5 +
min(prediction['premenna']), color='g')

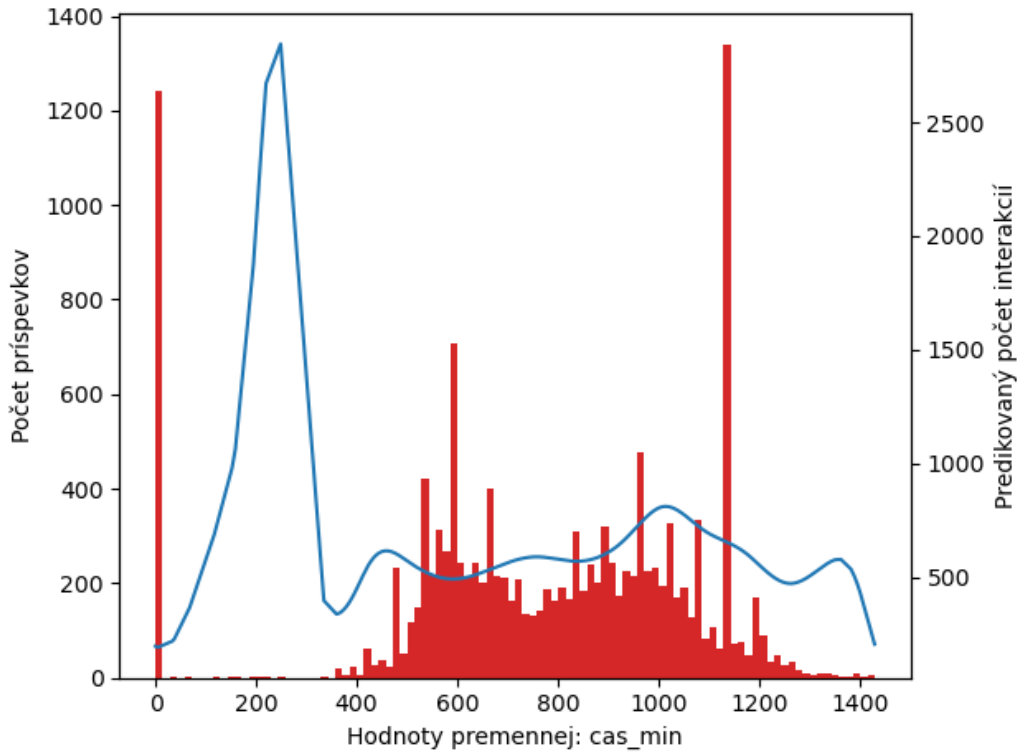
# Vykreslenie histogramu
color = 'tab:red'
ax1.hist(vektor_premennych[:, premenna_index], bins=100, color=color)

# Vykreslenie predikcií
ax2 = ax1.twinx()
color = 'tab:blue'
ax2.plot(prediction['premenna'], prediction['predictions'],
color = color)
```

```
fig.tight_layout())
```

Na obrázku 23 vidíme vykreslenú závislosť premennej *cas_min* od počtu interakcií, ktorému patrí pravá os. Ľavá os patrí histogramu, ktorý nám ukazuje, koľko príspevkov bolo pridaných v konkrétnych častiach dňa. Môžeme si všimnúť, že pri hodnote 0 bolo pridaných veľa príspevkov. Ako sme už spomínali v kapitole 3.1.2, nie pri všetkých príspevkoch disponujeme informáciou o tom, v akom čase bol publikovaný a to vidíme pekne na tomto obrázku, že takých hodnôt bolo až 1200.

Môžeme si taktiež všimnúť, že daná metóda neberie regularizačný parameter *lam* do úvahy. Chceli by sme, aby bola krivka jemnejšia a taktiež, aby mala aj menej efektívnych stupňov voľnosti, ktorých je teraz až 886, respektíve aby mala menšiu komplexnosť modelu.



Obr. 23: Vykreslenie závislosti premennej cas_min.

Transformácia *spline*, ktorú používame pri spojitých premenných, v sebe obsahuje parametre *lam*, *n_splines* a *penalties*, ktoré slúžia k zjemneniu krivky a zmenšeniu

počtu stupňov voľnosti.

Ako prvé sme sa snažili znížiť počet splajnov (*n_spline*). Toto sa nám najskôr nepodarilo, pretože pokiaľ sme zvolili počet menší ako 20, model nám neskonvergoval.

Ďalej sme menili hodnotu parametra *lam*. Tento parameter je sila penalizácie, ktorú na naše hodnoty dáme, aby bola krivka viac jemná. Tento parameter však vôbec nefungoval. Skúšali sme jeho rôzne hodnoty aj na menšom počte dát, no rozdiel vo výsledkoch, keď sme zvolili hodnotu 0,6 alebo 1000 nebol absolútne žiadny. Rozhodli sme teda tento parameter nechať na pôvodnej predvolenej hodnote 0,6 a viac sme s sním už nepracovali.

Parameter *penalties* je typ penalizácie na hladkosť krivky. Aby sme zistili, ktorý typ penalizácie je pre naše dáta najvhodnejší, spustili sme model 3-krát, pričom raz sme nechali predvolenú hodnotu, ďalšíkrát sme zvolili hodnotu *derivative* a poslednýkrát *l2*. Z týchto troch nám hodnota *derivative* najviac zjemnila krivku a naďalej sme používali tú.

Keďže stále nebola krivka tak jemná, ako by sme si predstavovali, skúsili sme štandardizovať dáta pomocou rovnice (8), kde sme od každej hodnoty odpočítali ich priemernú hodnotu a predelili štandardnou odchýlkou, aby sme zistili, či nemáme nejakú numerickú chybu, keďže štandardizácia dát zvyšuje numerickú stabilitu modelu a taktiež zvýši pravdepodobnosť úspešného skonvergovania metódy. Skúsili sme natrénovať model s rovnakými parametrami, no tentokrát na štandardizovaných dátach, avšak tvar krivky sa nezmenil. Usúdili sme, že numerická chyba tam nie je.

$$s = \frac{x - \mu}{\sigma} \tag{8}$$

Ďalšou možnosťou, ako odhaliť problém v dátach, bolo skúšať spúšťať model iba na časť z nich. Pri tejto metóde sme neboli úspešní, keďže model skonvergoval buď na prvých 5000 riadkoch, na riadkoch 7000 - 8000, alebo 8000 - 10 000. Akákoľvek iná kombinácia nefungovala. Skúsili sme ešte natrénovať model na náhodných 5000

riadkoch, avšak ani pri tomto postupe sme neboli úspešní, keďže ani po desiatkach pokusov nám model neskonvergoval.

Poslednou možnosťou bolo zmeniť parameter *tol*, čo je tolerancia pre kritéria zastavenia. Tento parameter sa aplikuje na celý model a v praxi to je hodnota, pri ktorej dosiahnutí model úspešne skonverguje. Jeho predvolená hodnota je 0.0001 a my sme ju chceli o niečo zvýšiť. Už na hodnote 0.0003 nám model úspešne skonvergoval aj pri *n_splines = 20*, čo nám zabezpečilo viac hladké krivky predikcie. Je potrebné ešte spomenúť, že takto nám model skonvergoval iba na štandardizovaných dátach. Keď sme použili pôvodné, model opäť zlyhal.

3.6 Výsledky

Nižšie môžeme vidieť zhrnutie modelu GAM:

Distribution:	PoissonDist	Effective DoF:	200.9387
Link Function:	LogLink	Log Likelihood:	-5952920.0654
Number of Samples:	13776	AIC:	11906242.0082
		AICc:	11906248.0468
		UBRE:	859.8078
		Scale:	1.0
		Pseudo R-Squared:	0.4862

Toto zhrnutie nám vysvetľuje nasledovné parametre:

- Distribution: Typ rozdelenia použitého v modeli.
- Effective DoF: Efektívny počet stupňov voľnosti. Táto hodnota hodnota je o 3/4 menšia ako pri použití pôvodných parametrov v modeli. Hodnota 200,9387, čo naznačuje, že model má primeranú zložitosť.
- Link Function: Väzbová funkcia, ktorá spája parametre modelu s očakávanými hodnotami pozorovaní.

- Log Likelihood: Logaritmická pravdepodobnosť modelu. Hodnota -5952920,0654 naznačuje, že model má dobrú schopnosť vysvetliť pozorované dáta.
- Number of Samples: Počet pozorovaní v dátovej sade.
- AIC: Akaikeova informačné kritérium (AIC).
- AICc: Korigované Akaikeovo informačné kritérium (AICc). Táto hodnota je modifikovanou verziou AIC, ktorá zohľadňuje počet pozorovaní v dátovej sade.
- UBRE: Nepoužitý zvyškový súčet štvorcov (UBRE). Táto hodnota kvantifikuje zostatkovú chybu modelu po zohľadnení efektu pevných efektov.
- Scale: Miera disperzie v modeli. V tomto prípade je hodnota 1,0, čo naznačuje, že disperzia je konštantná.
- Pseudo R-Squared: Táto hodnota kvantifikuje, akú časť variability v dátach model vysvetľuje. V tomto prípade je hodnota 0,4862, čo naznačuje, že model vysvetľuje takmer polovicu variability v dátach.

Na základe uvedených parametrov môžeme konštatovať, že model s Poissonovým rozdelením a logaritmickou väzbovou funkciou má dobrú schopnosť vysvetliť pozorované dáta. Model má primeranú zložitosť a jeho pseudo R-squared koeficient naznačuje, že vysvetľuje značnú časť variability v dátach.

Čo sa týka štatisticky významných premenných, tak všetky sa ukázali ako štatisticky významné. O tento výsledok sa avšak neopierame, keďže použitá knižnica to nemá korektne naprogramované, na čo nás aj sama upozorňuje.

V nasledujúcich podkapitolách si graficky ukážeme závislosti jednotlivých premenných a na základe toho formulujeme odporúčania na spôsob publikovania príspevku. Všetky grafy budú vyzerať nasledovne:

- pravá os patrí krivke, ktorá znázorňuje predikovaný počet interakcií,

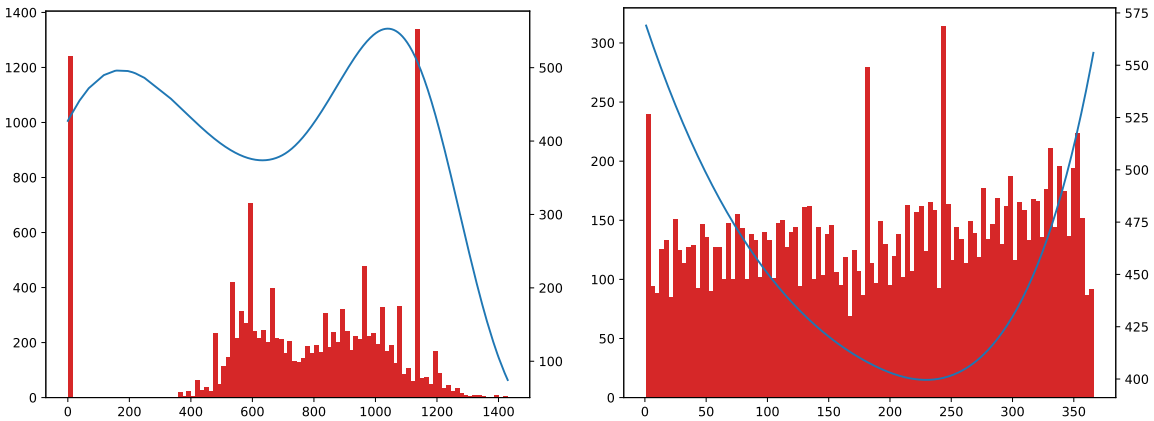
- ľavá os patrí histogramu, ktorý nám dá pohľad na to, koľko príspevkov bolo publikovaných pri jednotlivých hodnotách premenných.

.

3.6.1 Spojité premenné

Na obrázkoch 24 vidíme závislosť času a dňa pridania príspevkov od počtu interakcií. Na obrázku vľavo si môžeme všimnúť, že krivka má dve lokálne maximá. Prvé okolo hodnoty 200 a druhé má hodnotu približne 1000. Z týchto dvoch si vyberieme práve to druhé, keďže prvé maximum je pravdepodobne skreslené veľkým množstvom príspevkov, u ktorých nám chýba informácia o čase pridania a taktiež na histograme vidíme, že okolo tej hodnoty neboli publikované takmer žiadne príspevky. Druhé maximum nám hovorí o tom, že ak chcú mať firmy čo najviac interakcií na príspevkoch, mali by ich pridávať okolo 17:00, čo dáva zmysel, keďže vtedy ľudia buď cestujú z roboty alebo už sú doma a majú čas na oddych.

Čo sa týka závislej premennej *day_of_year*, čo je deň v roku, v ktorom bol príspevok pridaný, najväčší úspech majú príspevky, ktoré boli publikované na konci alebo začiatku roka. Počas leta ľudia zjavne menej interagujú s nimi a počas vianočných sviatkov, keď firmy možno ponúkajú nejaké vianočné balíky, tak ľudí to zaujíma viac.

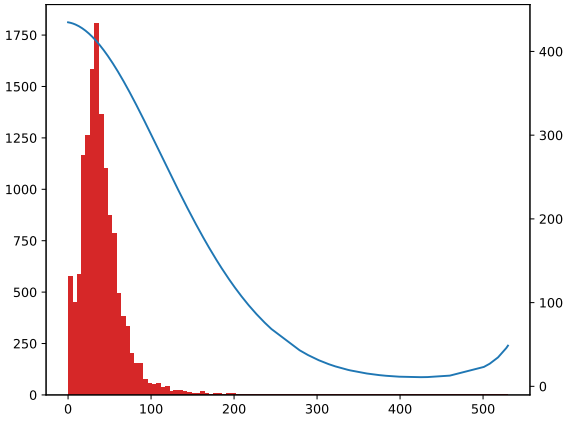


Závislosť od premennej cas_min

Závislosť od premennej day_of_year

Obr. 24: Závislosť premenných cas_min a day_of_year od interakcií.

Ďalšia premenná, na ktorú sa pozrieme, je počet slov, ktoré sa v príspevku nachádzajú. Na obrázku 25 vidíme, že ľudí viac zaujímajú príspevky, ktoré majú menší počet slov.



Obr. 25: Závislosť od premennej pocet_slov

To, ako vplývajú témy príspevkov na interakcie vieme rozdeliť do 3 kategórií:

1. Príspevky, ktoré ľudí zaujímajú.
2. Príspevky, o ktoré ľudia nemajú veľký záujem.
3. Príspevky, ktoré ľudí neprilákajú, ale ani neodlákajú.

Pri interpretácií kategórií príspevkov je potrebné si uvedomiť, že čím je hodnota vyššia, tým viac je príspevok vzdialený od danej témy a naopak. Predstavme si, že máme zhluk tém definovaný centroidmi a chceme zistiť, či daný príspevok patrí do jednej z týchto tém. *Príspevok považujeme za zaradený do témy vtedy, ak je jeho vzdialenosť od centroidu zhluku témy menej ako 1/5 rozsahu danej premennej.*

Majme:

x = príspevok, pre ktorý chceme určiť, či patrí do danej témy,

c_t = centroid zhluku danej témy,

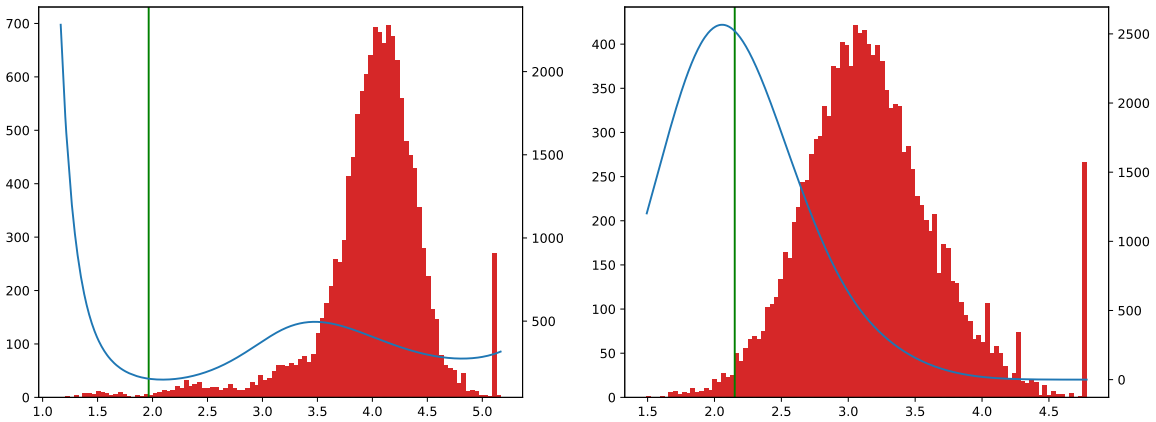
$d(x, c_t)$ = vzdialenosť príspevku x od centroidu c_t,

R = rozsah danej premennej.

Potom príspevok x patrí do danej témy, ak:

$$d(x, c_t) \leq \frac{1}{5}R \tag{9}$$

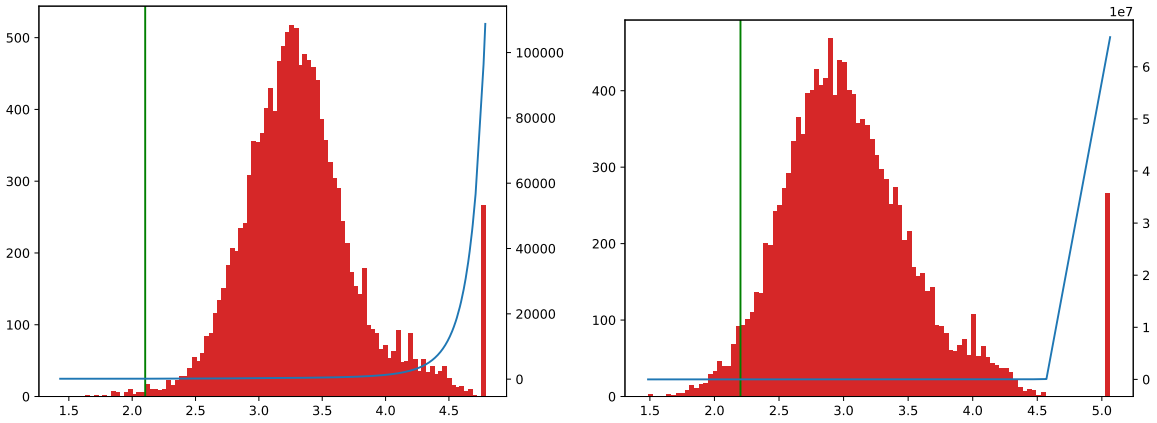
Na obrázku 26 teda vidíme témy príspevkov, na ktoré sa oplatí robiť obsah, keďže prinášajú veľký počet reakcií. Kategória 10 je vplyv makroekonomických faktorov na finančné trhy a kategória 26 sa zaoberá dôležitosťou cestovného poistenia pri cestovaní.



Vzdialenosť od kategórie 10 Vzdialenosť od kategórie 26

Obr. 26: Závislosť vzdialeností od kategórie 10 a 26.

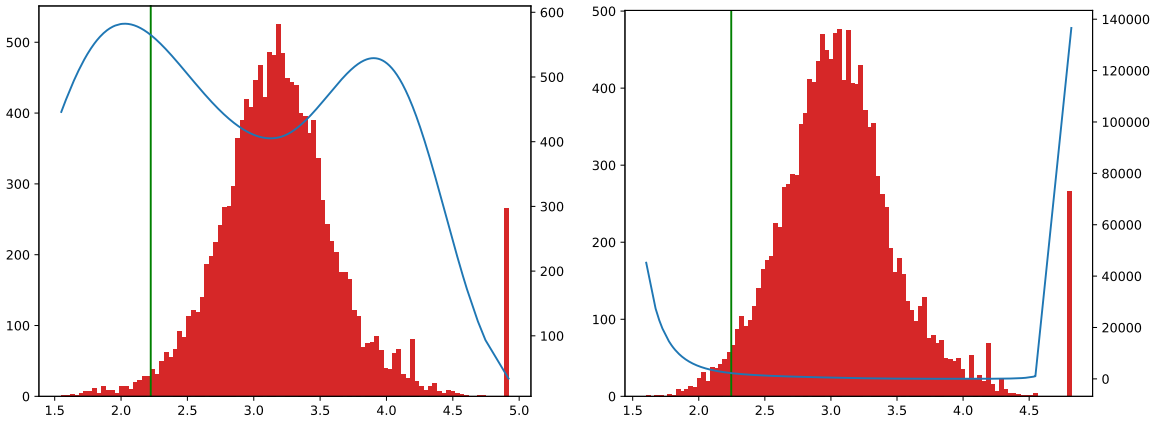
Na obrázku 27 vidíme závislosť vzdialeností od kategórií 20 a 44, ktoré nemajú dobrý vplyv na počet interakcií. Čím viac sa príspevky vzdiaľujú od tejto témy, tým to ľudí viac zaujíma. Usudzujeme teda z toho, že firmy by sa mali vyhnúť príspevkom, ktoré podporujú zdravý životný štýl alebo umenie, vzdelanie či inovácie.



Vzdialenosť od kategórie 20 Vzdialenosť od kategórie 44

Obr. 27: Závislosť vzdialeností od kategórie 20 a 44.

Témy, pri ktorých sme nevedeli určiť, či by sa im firmy mali venovať sú podľa obrázkov 28 napríklad spoločenská zodpovednosť bánk alebo úspechy bánk počas celého roka.



Vzdialenosť od kategórie 42 Vzdialenosť od kategórie 43

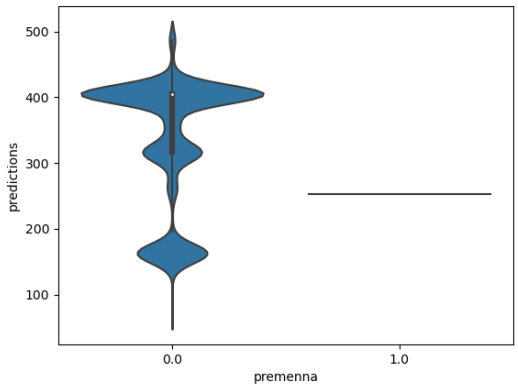
Obr. 28: Závislosť vzdialeností od kategórie 42 a 43.

Závislosti všetkých kategórií príspevkov od interakcií môžeme vidieť v prílohe B.

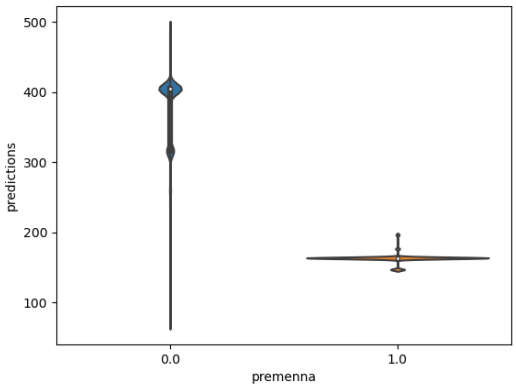
3.6.2 Kategorické premenné

Pri interpretácií kategorických premenných budeme využívať takzvaný *violin plot*, ktorý nám ukáže priemernú predikovanú hodnotu. Na obrázkoch 27 vidíme nami vybrané kategorické premenné a ich predikované hodnoty. Čo sa týka premenných album,

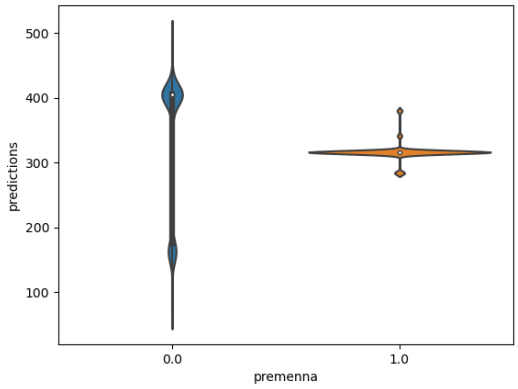
video, status alebo link, tak ak majú hodnotu 1, znamená to, že majú daný typ. Takéto typy príspevkov sa neoplatí publikovať, keďže priemerná predikovaná hodnota je vyššia vtedy, ak príspevok nemá tento typ. Čo sa týka fotky, tak na obrázku jasne vidíme, že príspevky, ktoré obsahujú nejakú fotografiu, teda majú hodnotu 1, majú väčší počet interakcií ako tie, ktoré fotografiu neobsahujú (hodnota 0). Čo sa týka premennej sviatok, aj tam sa oplatí pridávať príspevky práve počas dní pracovného pokoja (hodnota 1) viac ako v bežné dni.



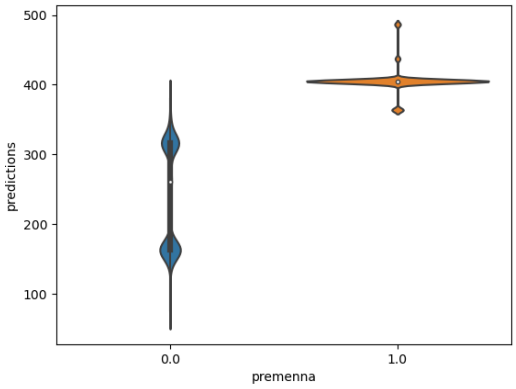
Album



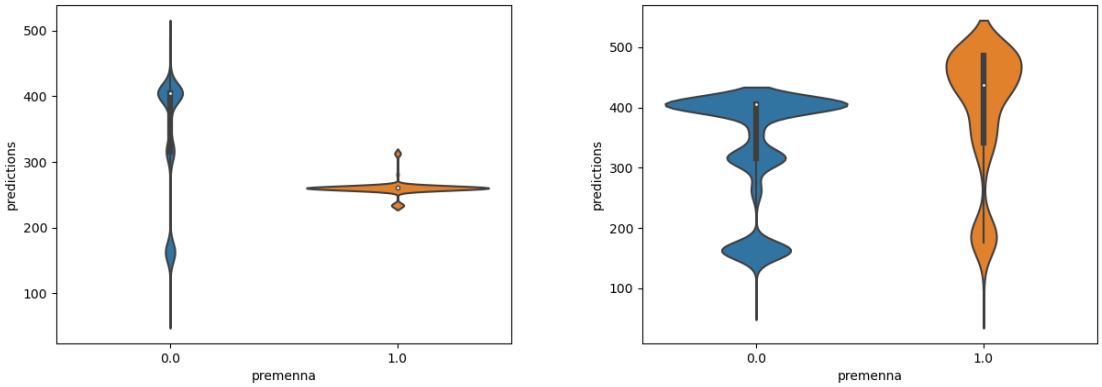
Link



Video



Fotka



Status Sviatok

Obr. 27: Predikované hodnoty kategorických premenných

Záver

V našej práci sme prenikli do podstaty komunikácie bankového sektora na sociálnych sieťach. Analýza dát prostredníctvom grafických vizualizácií nám na začiatku odhalila zaujímavé vzory v správaní fanúšikov a samotných bánk na ich sociálnych sieťach. Príkladom je, že vyšší počet uverejneného obsahu na sociálnej sieti, nezaručuje väčší počet fanúšikov. Obrázky nám taktiež ukázali, že väčšina firiem uprednostňuje publikovať súťaže s finančnou výhrou namiesto hmotnej a taktiež, že fanúšikovia viac interagujú so súťažnými príspevkami namiesto s obyčajnými.

Modelovanie prírastku fanúšikovskej základne na sociálnych sieťach

Pri modelovaní prírastkov fanúšikov sme využívali lineárne modely Fixed Effects model a LDV model, čo nám umožnilo presne identifikovať faktory, ktoré prispievajú k rozvoju fanúšikovskej komunity. Tieto výsledky predstavujú užitočné informácie pre banky, ktoré sa snažia optimalizovať svoje marketingové stratégie za účelom rastu počtu fanúšikov.

Na základe štatistického modelu sme odhadli, že keby banky ako napríklad Slovenská sporiteľna alebo 365.bank nepridávali žiaden obsah na sociálnej sieti, stále by mesačne získavali viac ako 100 fanúšikov. Môže to byť tým, že majú rôzne platené reklamy na internete, v televízii alebo na billboardoch. Na druhej strane banky ako Tatra banka a VÚB banka by museli vynakladať úsilie, čo sa týka publikovania obsahu, pretože ako jediné by v tomto prípade mesačne prichádzali o svojich sledovateľov.

Čo sa týka reakcií na pridávané príspevky, tak najväčší prínos majú pre firmy práve finančné súťaže, pretože tie priemerne prinášajú 27 fanúškov na jednu súťaž. Toto je premenná, ktorú firmy vedia ovplyvniť a teda tento typ súťaží by mali určite preferovať pred nefinančnými. Medzi ďalšie ovplyvniteľné parametre patrí napríklad množstvo príspevkov, čo nám evokuje, že ľudia budú radšej sledovať „aktívne stránky“,

ktoré pravidelne pridávajú aktuálny obsah pred tzv. „mŕtvymi“, na ktorých fanúšikovia ťažko nájdu niečo aktuálne.

Modelovanie interakcií s príspevkami

Podrobnejšia analýza interakcií s príspevkami nám ukázala dôležitosť správneho načasovania publikácie príspevku, či zvolenie vhodnej témy. Vhodným časom sa ukázala 17 - ta hodina večer, kedy ľudia už sú väčšinou z práce doma. Taktiež ľudia viac interagujú s príspevkami, ktoré boli publikované počas štátneho sviatku zjavne z toho dôvodu, že sú doma a majú viac času na oddych a sledovanie sociálnych sietí.

Najvhodnejšie obdobie v roku bolo v okolí začiatku a konca roka. Optimálna dĺžka príspevkov sa javila tá, kde bol použitý menší počet znakov.

Keď sme sa bližšie pozreli na jednotlivé témy, ktorými sa príspevky zaoberali, ako vhodné sa javili tie, ktoré sa zameriavali na vplyv makroekonomických faktorov na finančné trhy a dôležitosť cestovného poistenia pri cestovaní. Témam ako zdravý životný štýl a umenie, vzdelanie či inovácie, by sa mali vyhnúť, pretože ľudí až tak veľmi nezaujímajú.

Čo sa týka typov príspevku, tak ľudia moc nemajú radi tie, ktoré obsahujú nejaký album s fotkami / videami alebo tie, ktore boli čisto textové. Najviac ľudí zaujímajú príspevky, ktoré v sebe obsahujú nejakú fotografiu.

Zhrnutie

Zistenia v našej diplomovej práci nám predstavujú dôležitú informáciu pre banky, ktoré tak môžu efektívnejšie cieliť svoje marketingové kampane a zlepšiť svoju prítomnosť na sociálnych sieťach. Odporúčame napríklad:

- pravidelne pridávať aktuálny obsah,
- venovať sa témam ohľadom vplyvu makroekonomických faktorov na finančné

trhy alebo dôležitosti cestovného poistenia pri cestovaní,

- naplánovať pridanie príspevku vo večerných hodinách, kedy už budú ľudia doma z práce,
- v období Vianoc a nového roka publikovať viac príspevkov,
- publikovať viac finančných súťaží,
- uprednostniť publikáciu príspevku s fotografiou ako bez,
- publikovať príspevky najmä počas sviatkov

Celkovo sme chceli s našou prácou prispieť k hlbšiemu pochopeniu komunikačných vzťahov medzi bankami a ich publikom na sociálnych sieťach. Veríme, že naše odporúčania môžu slúžiť ako inšpirácia pre banky pri formovaní účinnejších a atraktívnejších komunikačných stratégií, čo následne povedie k vyššej angažovanosti fanúšikov a posilneniu online komunity.

Zoznam použitej literatúry

[1] Bard, Google AI. *„Napíš mi jednu komplexnú tému, ktorou sa zaoberajú nasledujúce príspevky"Bard,* Google AI, (12.4.2024).

[2] Buffer.com [online]. *Social Media Strategy* [cit. 1.5.2024] Dostupné na adrese: `https://buffer.com/library/category/social-media-strategy/`.

[3] ChatGPT. „Napíš mi jednu komplexnú tému, ktorou sa zaoberajú nasledujúce príspevky"ChatGPT, OpenAI platform, (12.4.2024).

[4] Finreport.sk [online]. *OTP banka Slovensko sa definitívne stáva členom skupiny ČSOB. Čo to prinesie klientom oboch bánk?* [cit. 5.10.2023] Dostupné na adrese: `https://www.finreport.sk/banky-a-poistovne/otp-banka-slovensko-sa-definitivne-stava-clenom-skupiny-csob-co-to-prinesie-klientom-oboch-bank/`

[5] Harman R. *Multivariate Statistical Analysis.* Faculty of Mathematics, Physics and Informatics, Comenius University (2023).

[6] Harman R. Poznámky k predmetu *Analýza zhlukov a klasifikácia dát.*

[7] Hastie, T., Tibshirani, R., Friedman, J. H. *he Elements of Statistical Learning; Data Mining, Inference, and Prediction* Springer, 295-304 (2008).

[8] Huggingface.co [online]. *slovakbert* [cit. 13.12.2024] Dostupné na adrese: `https://huggingface.co/gerulata/slovakbert/tree/main`

[9] Iienstitu.com [online]. *Is The Number Of Followers Social Media Success?* [cit. 6.12.2023] Dostupné na adrese: `https://www.iienstitu.com/en/blog/is-the-number-of-followers-an-indication-of-social-media-success`

[10] Kaplan, A. M., Haenlein, M. *Users of the world, unite! The challenges and opportunities of social media.* Kelley School of Business, Indiana University (2009).

[11] Klipfolio.com [online]. *Social Media Followers* [cit. 13.12.2023] Dostupné na adrese: `https://www.klipfolio.com/resources/kpi-examples/social-media/social-followers`

[12] Lugo V. *5 Facebook interactions ranked by usage importance.* [cit. 13.12.2023] Dostupné na adrese: `https://vladimirlugo.com/5-facebook-interactions-ranked-by-usage-importance/`

[13] Pikuliak M., et al. *SlovakBERT: Slovak masked language model.* arXiv preprint arXiv:2109.15254, (2021).

[14] Raiffeisen.sk [online]. *Profil banky* [cit. 15.6.2023] Dostupné na adrese: `https://www.raiffeisen.sk/sk/o-banke/o-raiffeisen-banke/`

[15] Rousseeuw, P. J., *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis* University of Fribourg, (1986).

[16] Servén D., Brummitt C. [online]. *pyGAM* [cit. 1.4.2024] Dostupné na adrese: `https://pygam.readthedocs.io/en/latest/index.html` (2018).

[17] Somorčík J. Poznámky k predmetu *Počítačová štatistika.*

[18] Thombs, R. P. *Guide to Analyzing Large N, Large T Panel Data.* Socius, 8, (2022).

[19] Trunfio, M., Rossi, S. *Conceptualising and measuring social media engagement: A systematic literature review.* Ital. J. Mark. 2021, 267–292 (2021).

[20] Zvára K. *Regrese* MATFPRESS, vydavetelství Matematicko-fyzikálni fakulity Univerzity Karlovy v Praze (2008).

[21] Wang, A., Nelson, A., et al. *GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding.* arXiv preprint arXiv:2004.07461 (2019).

Príloha A - Predpoklady lineárnej regresie

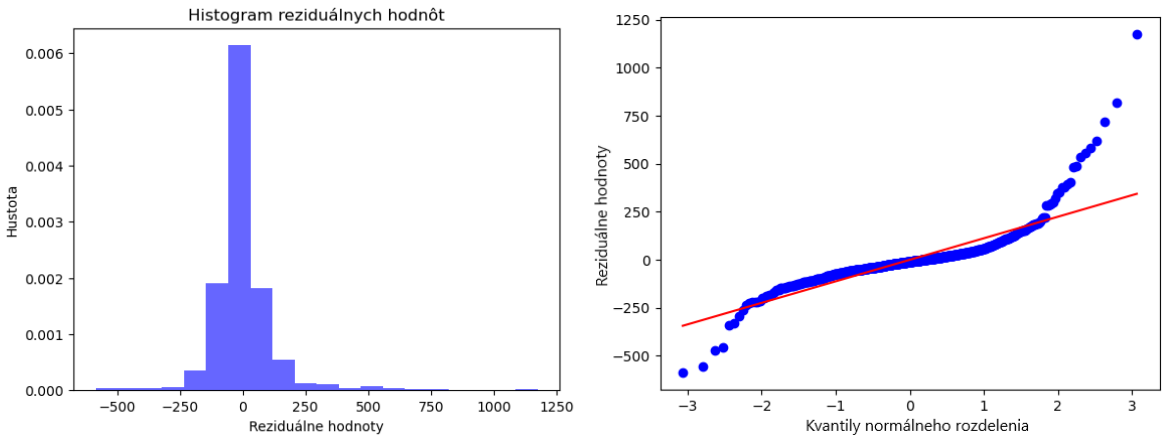
Na obrázku 28 vidíme histogram a Q-Q plot pre overenie normality rezíduí. Histogram nám ukazuje, že rezíduá vyzerajú byť z normálneho rozdelenia. Q-Q plot nám naopak ukazuje, že má ťažký chvost smerom do prava a teda náš model nemusí úplne dobre popisovať väčšie prírastky fanúšikov.

Ako ďalšie sme overovali lineritu pre vybranú premennú. Našou premennou boli komentáre a na obrázku 29 vidíme, že nie všetky komentáre sa nám správajú lineárne.

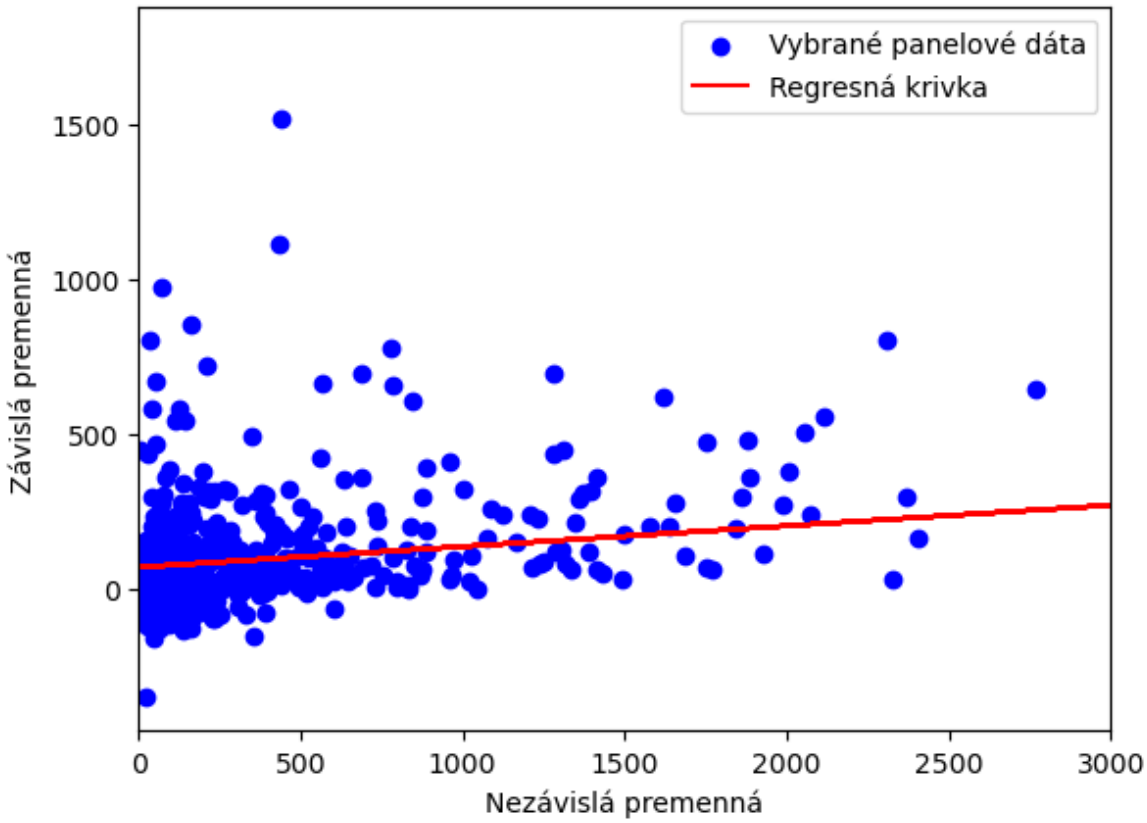
Ďalším predpokladom je overenie nezávislosti dát. Obrázok 30 nám ukazuje korelačnú maticu všetkých nezávislých premenných a rezíduí. Podľa heat mapy vidíme, že medzi rezíduami a jednotlivými premennými nie je významná korelácia a teda predpoklad nezávislosti rezíduí vyzerá neporušený.

Medzi predpoklady patrí aj overenie homoskedasticity a náhodnosti chýb. Na obrázku 31 si môžeme všimnúť, že dáta po hodnotu približne 400 vyzerajú homoskedacitne, keďže sa tvorí mrak okolo červenej čiary. Toto nám evokuje náhodnosť chýb a teda odhady by mohli byť správne aj napriek tomu, že pri hodnote väčšej ako 400 je homoskedasticita mierne porušená.

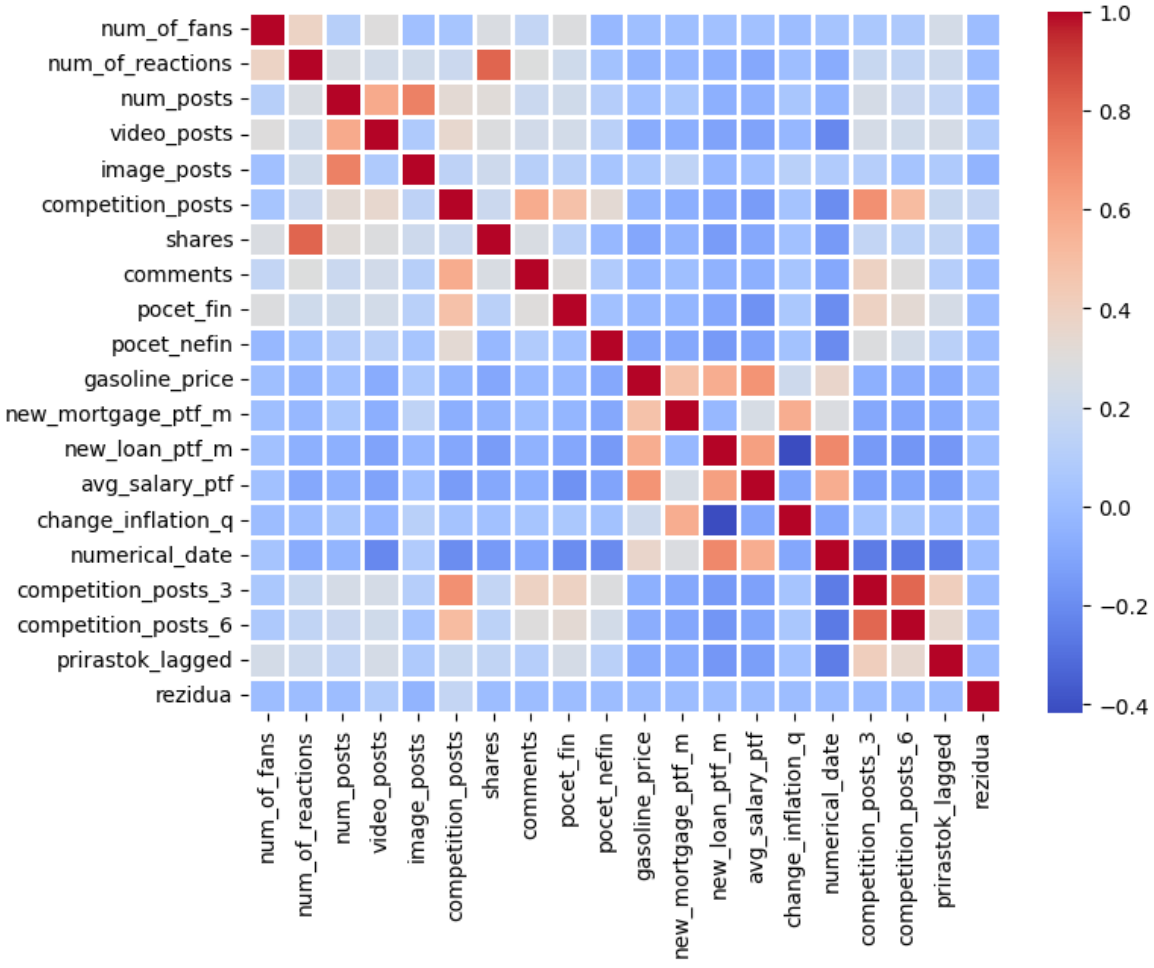
Ako posledné sme overili autokoreláciu pomocou Durbin-Watson. Durbin-Watson štatistika nám vyšla 1,989, čo indikuje negatívnu koreláciu a znamená to, že hodnoty rezíduí sú ovplyvnené ich predchádzajúcou hodnotou negatívne. Keďže je ale táto hodnota veľmi blízko hodnote 2, čo by znamenalo náhodnosť chýb, urobili sme ešte Ljung-Box, ktorého *p-value* nám vyšla vačšia ako 5 %, a teda predpoklad nezávislosti rezíduí nie je v rozpore s našimi pozorovaniami.



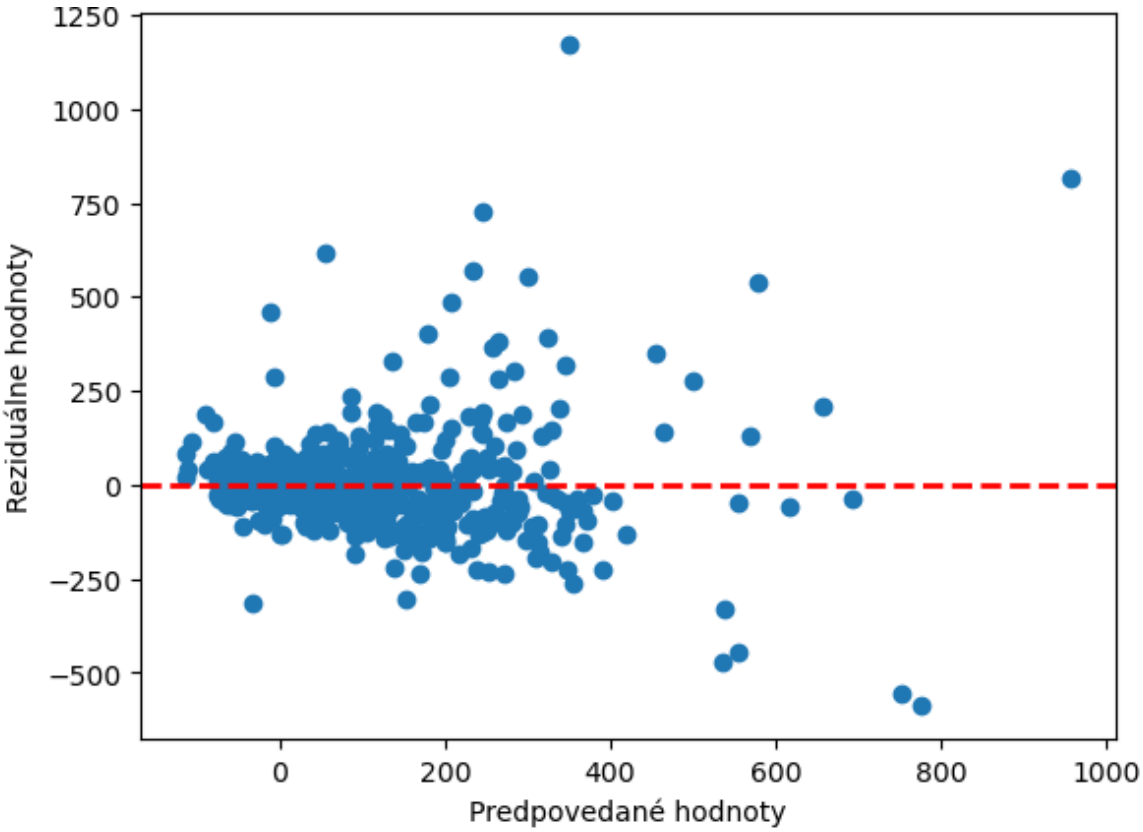
Obr. 28: Histogram a Q-Q plot pre overenie normality rezíduí.



Obr. 29: Overenie linearity pre komentáre.



Obr. 30: Korelačná matica nezávislých premenných a rezíduí.

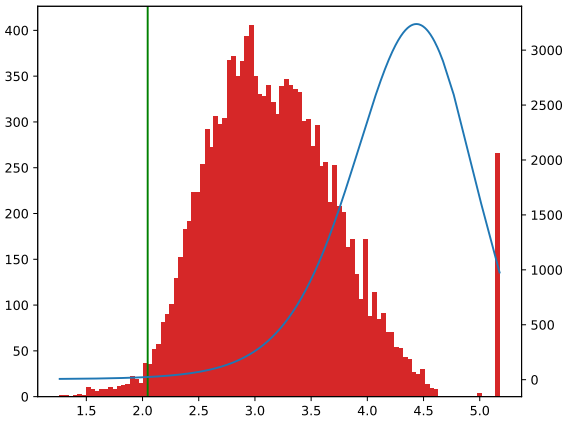


Obr. 31: Závislosť predpovedaných hodnôt a rezíduí

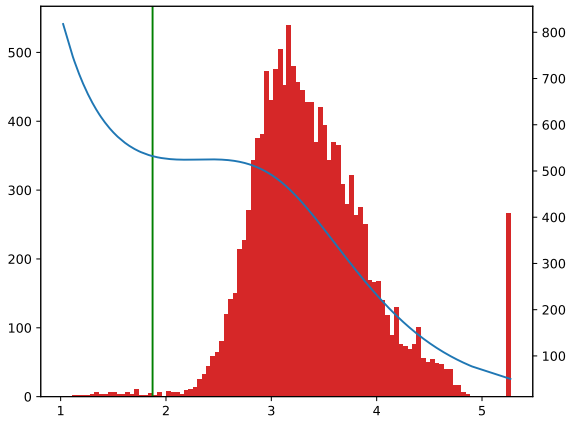
Príloha B - Závislosť kategórií príspevkov od interakcií

Na obrázkoch nižšie vidíme prehľad závislostí všetkých vzdialeností od kategórií príspevkov od počtu interakcií s nimi. Ľavá os patrí histogramu, ktorý znázorňuje počet príspevkov pri jednotlivých hodnotách premennej. Pravá os patrí predikovanému počtu interakcií na príspevok pri rôznych hodnotách premennej, ktorá nás zaujíma na x - ovej osi. Na obrázku sa taktiež nachádza zelená vertikálna čiara, ktorá nám znázorňuje hranicu, ktorá nám ukazuje vzdialenosť, do ktorej musia príspevky spadať, aby patrili do danej témy.

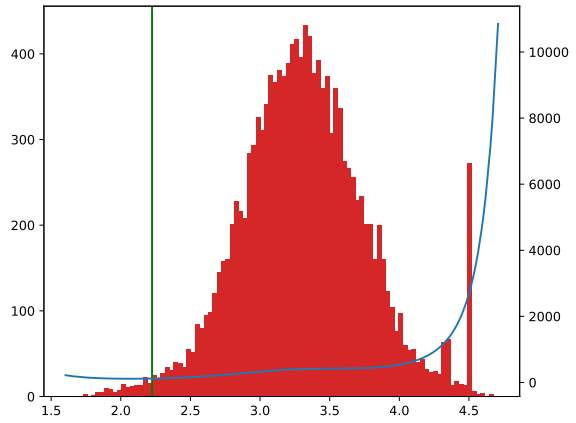
Je potrebné poznamenať, že čím je hodnota vzdialenosti od kategórie nižšia, tým viac patrí príspevok danej kategórii a naopak. Ako príklad si zoberme hneď prvý obrázok, ktorý znázorňuje vzdialenosť od kategórie 1. Vidíme na ňom, že maximálna hodnota predikcie nastala v bode s hodnotou približne 4,6. Z tohto usudzujeme, že firmy by sa mali tejto téme vyhnúť, pretože čím je príspevok bližšie k tejto téme, tým menej ľudí s ním interaguje.



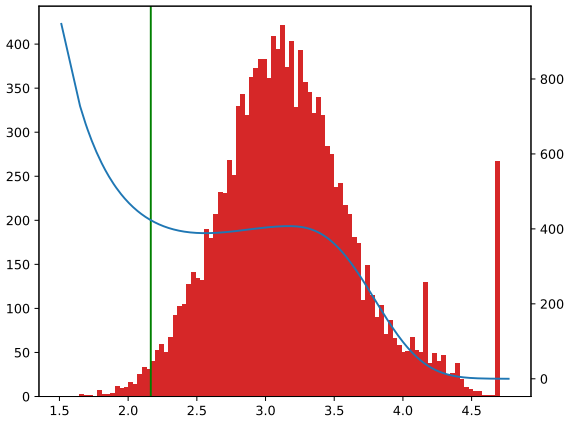
Vzdialenosť od kategórie Dôležitosť poistenia pre rôzne životné situácie



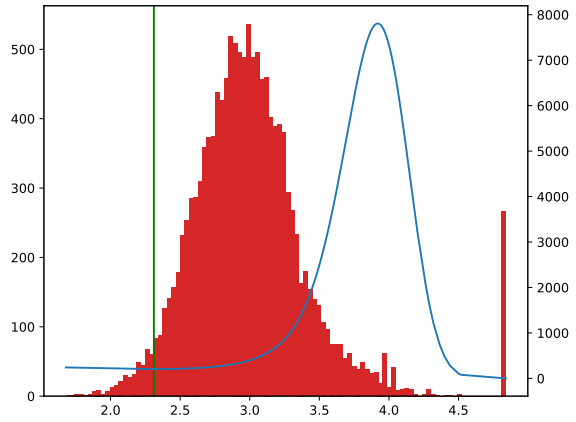
Vzdialenosť od kategórie Podpora komunitných projektov a aktívnej participácie v meste



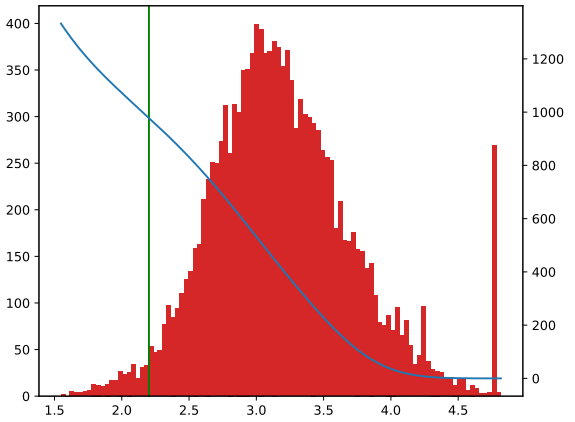
Vzdialenosť od kategórie Vianočné a sviatočné posolstvá



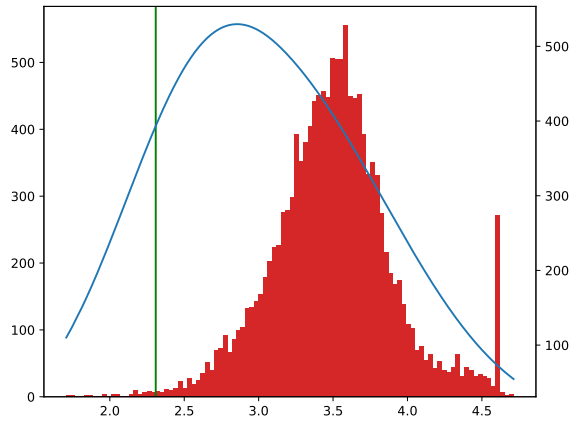
Vzdialenosť od kategórie Podpora kultúrnych a spoločenských podujatí



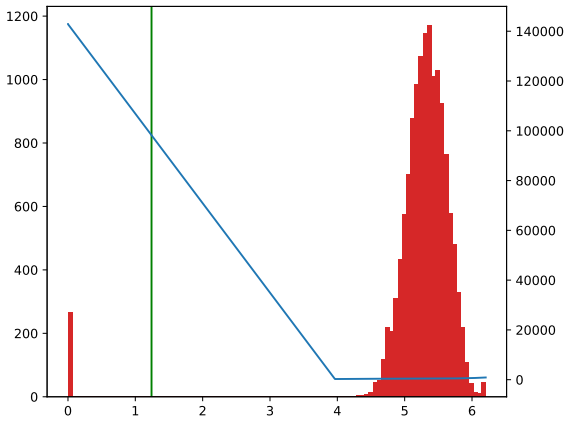
Vzdialenosť od kategórie Finančné správy a tipy pre podnikateľov a investorov



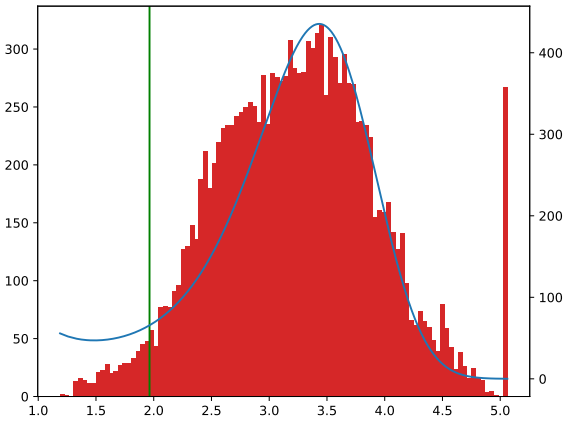
Vzdialenosť od kategórie Vylepšovanie bankových služieb pomocou mobilných aplikácií



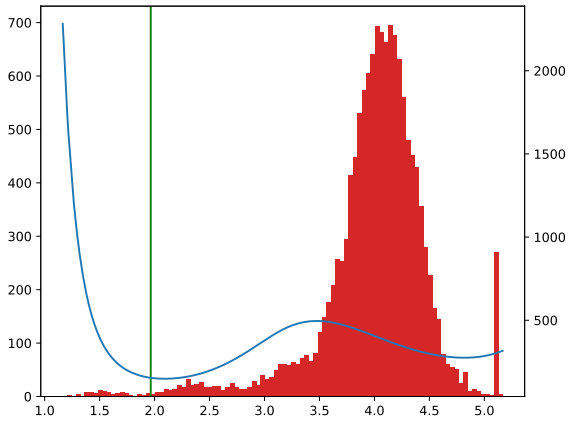
Vzdialenosť od kategórie Štatúty súťaží



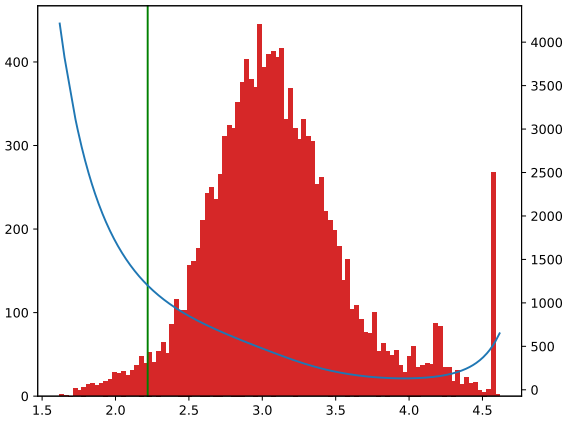
Vzdialenosť od kategórie Príspevok bez textu



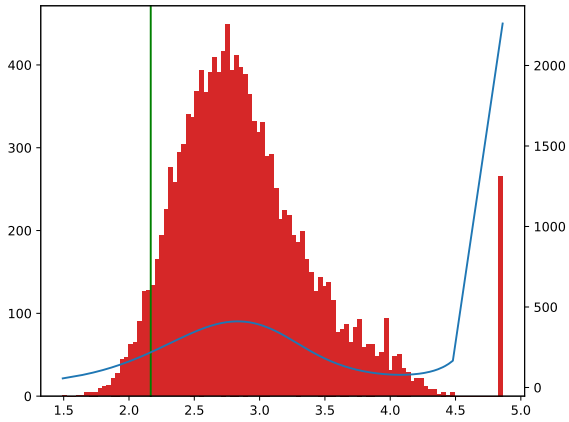
Vzdialenosť od kategórie Reklamné kampane na pôžičky



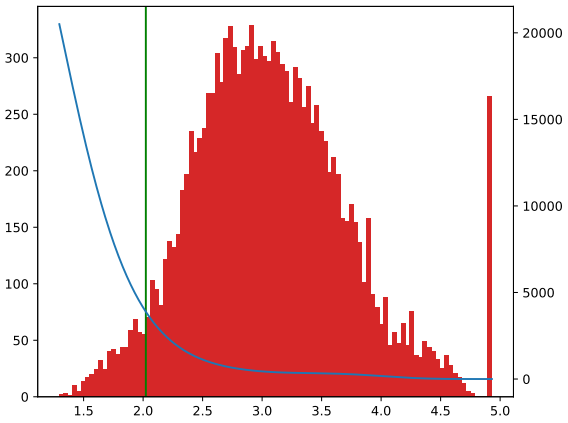
Vzdialenosť od kategórie Vplyv makroekonomických faktorov na finančné trhy



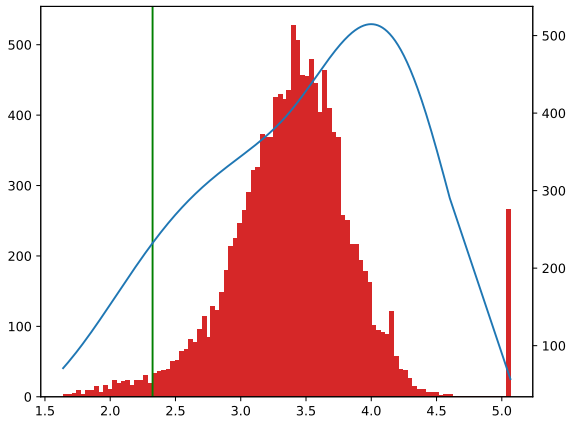
Vzdialenosť od kategórie Kultúra vytvárania pozitívnej komunitnej atmosféry



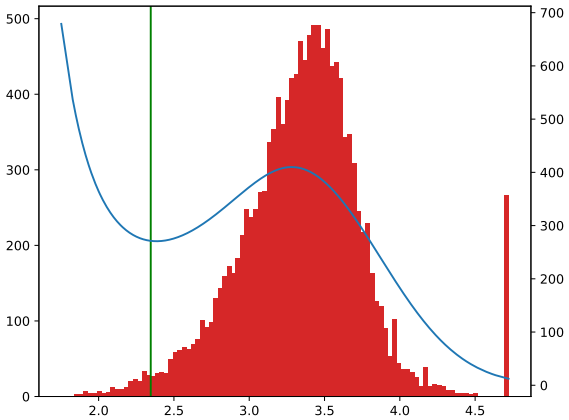
Vzdialenosť od kategórie Bezpečné finančné správanie a moderné bankovníctvo



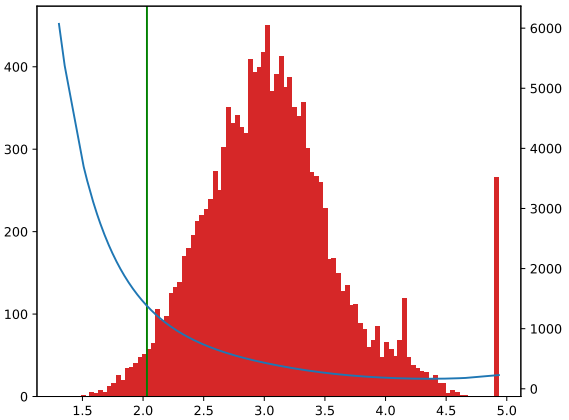
Vzdialenosť od kategórie Súťaže a tipy na úspory



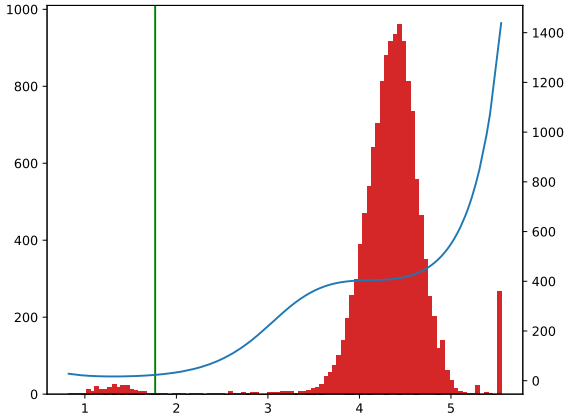
Vzdialenosť od kategórie Podpora umenia a ocenenie umeleckých diel



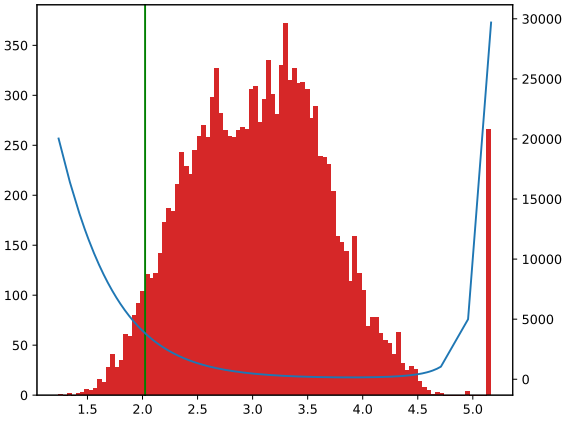
Vzdialenosť od kategórie Rast nerovnosti a koncentrácia bohatstva v globálnom kontexte



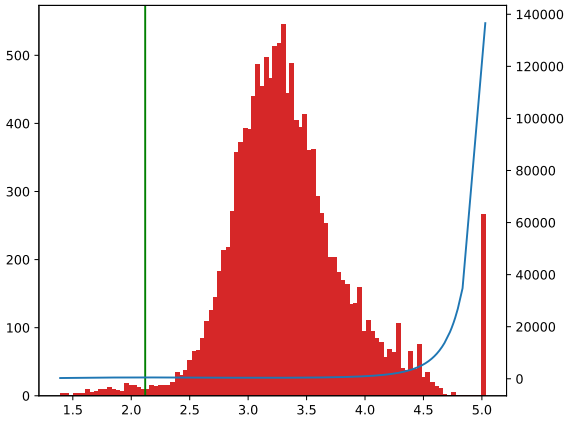
Vzdialenosť od kategórie Refinancovanie úverov a hypoték



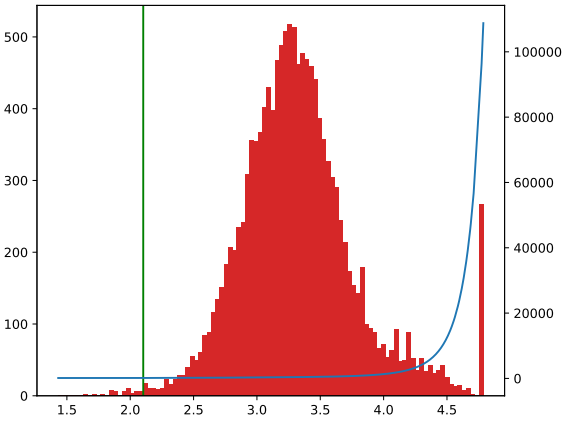
Vzdialenosť od kategórie Citáty o peniazoch a šťastí



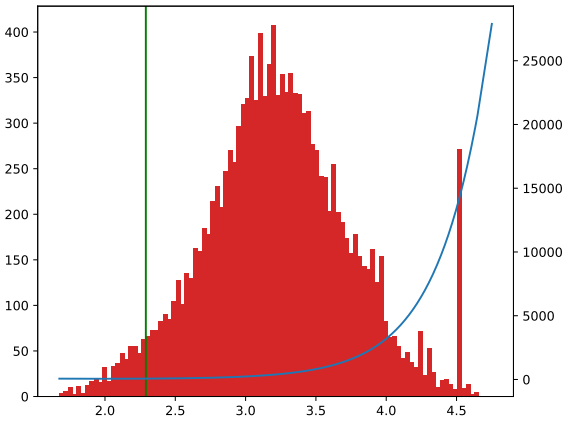
Vzdialenosť od kategórie Finančné plánovanie a správa osobných financií



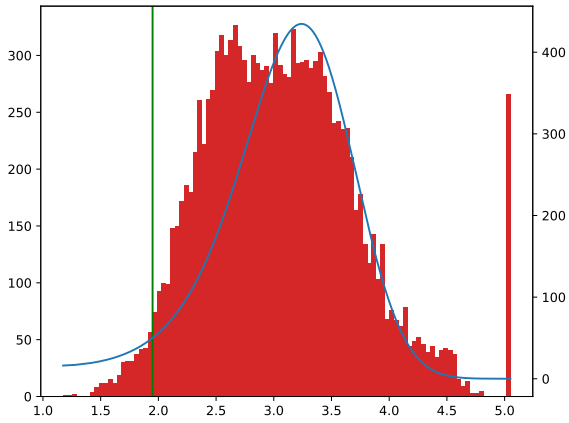
Vzdialenosť od kategórie Plánované technické odstávky v bankových službách



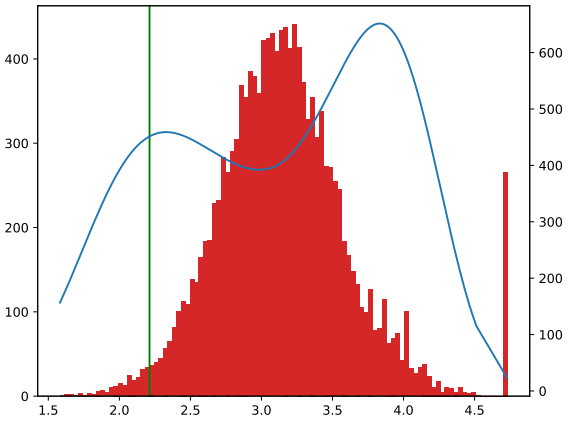
Vzdialenosť od kategórie Podpora športových udalostí



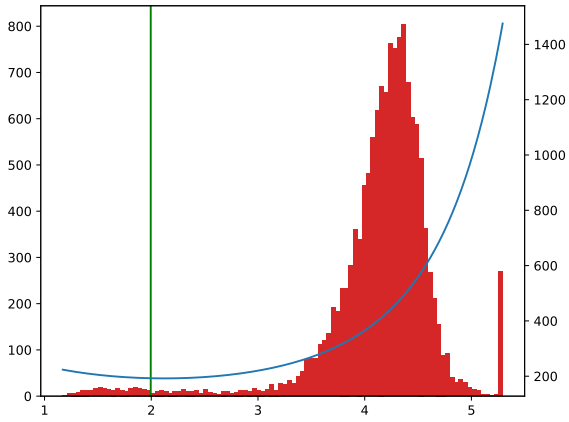
Vzdialenosť od kategórie Rôzne témy z rôznych oblastí života



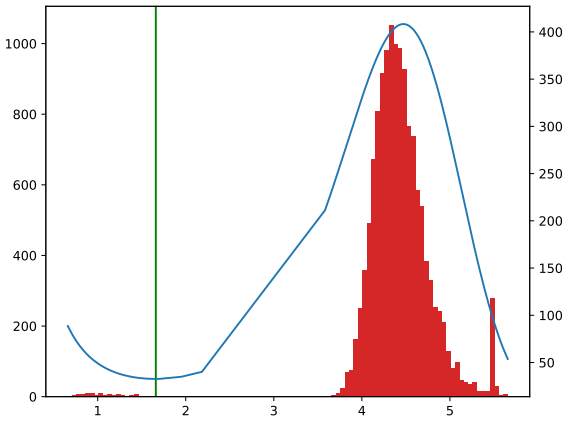
Vzdialenosť od kategórie Marketingové kampane s platobnými kartami



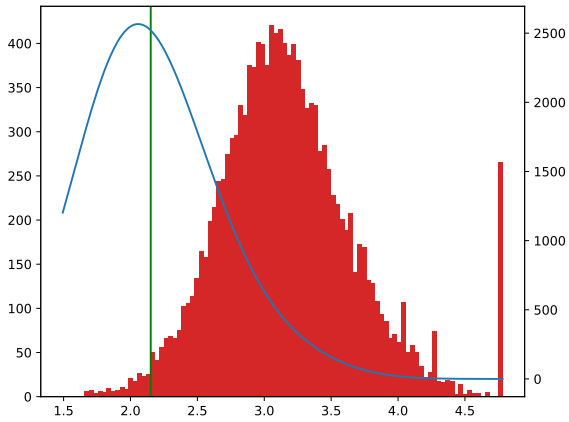
Vzdialenosť od kategórie Sociálne a environmentálne zodpovedné príspevky



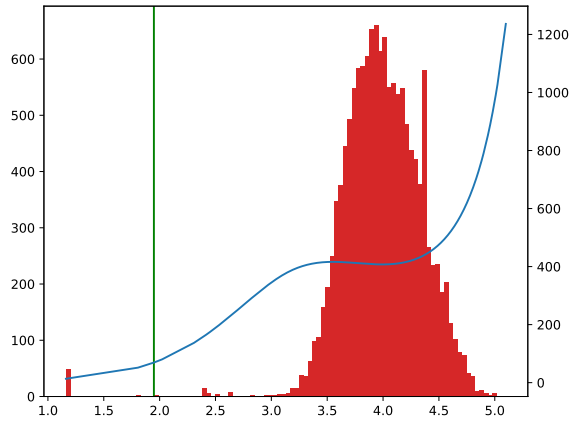
Vzdialenosť od kategórie Vplyv makroekonomických faktorov a geopolitických udalostí na svetové akciové trhy



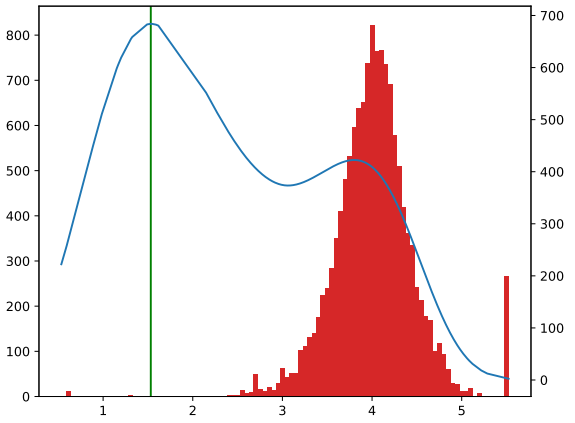
Vzdialenosť od kategórie Bankové služby a finančná bezpečnosť



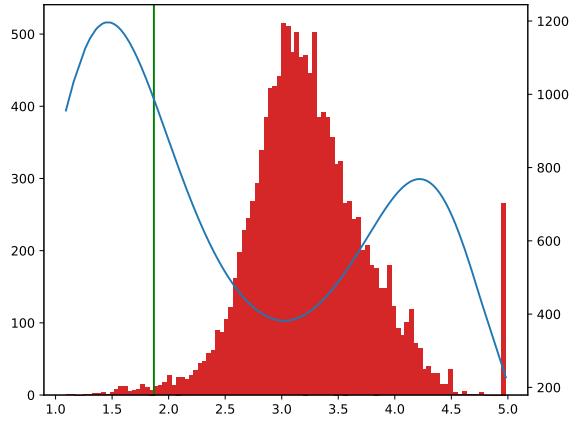
Vzdialenosť od kategórie Zmena titulnej fotografie



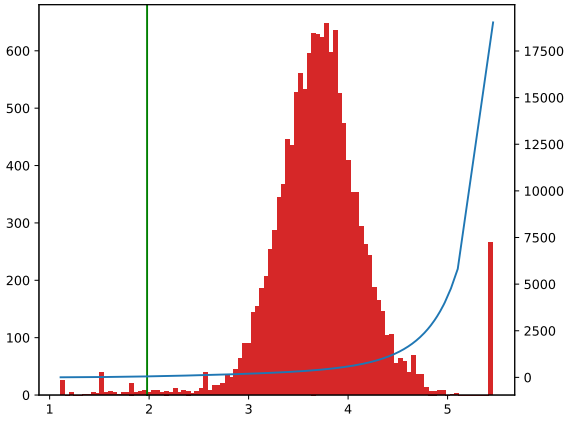
Vzdialenosť od kategórie Bezpečná a pohodová dovolenka s cestovným poistením



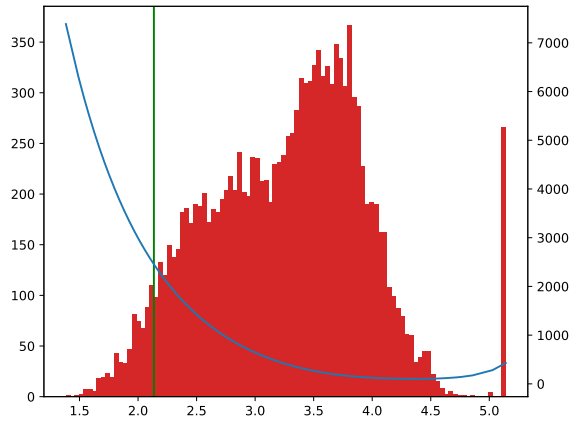
Vzdialenosť od kategórie Alternatívne investičné možnosti a výnosy v porovnaní s tradičnými bankovými produktami



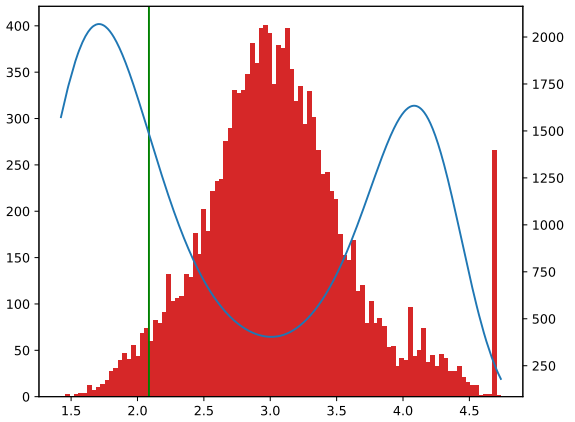
Vzdialenosť od kategórie Ako sa chrániť pred internetovými podvodmi



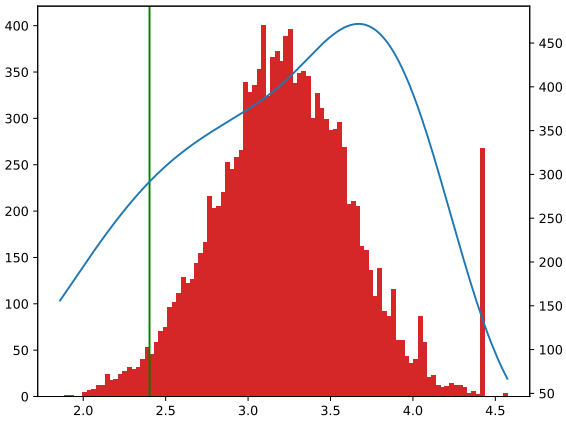
Vzdialenosť od kategórie Pôžičky a ich výhody



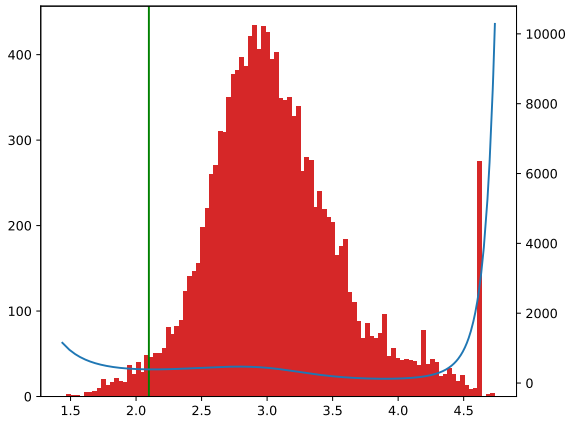
Vzdialenosť od kategórie Rôzne aspekty bankovníctva a finančnej bezpečnosti



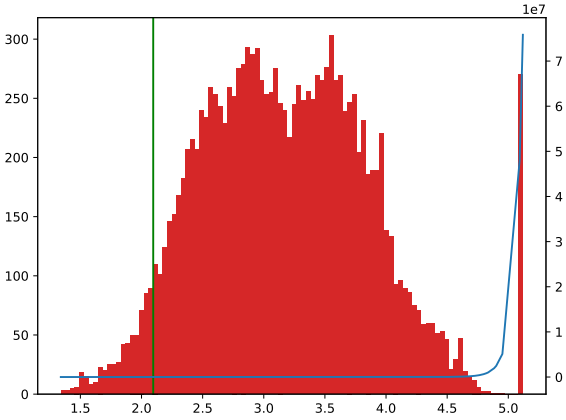
Vzdialenosť od kategórie Marketingové súťaže a promočné akcie v bankovom a finančnom sektore



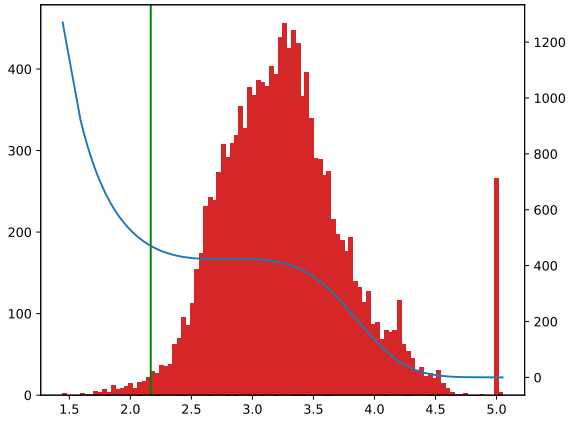
Vzdialenosť od kategórie Finančné návyky a spotreba v Slovenskej republike



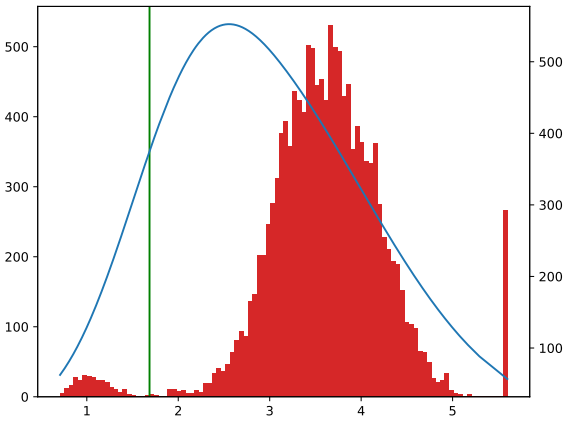
Vzdialenosť od kategórie Finančné služby a zvyky v digitále



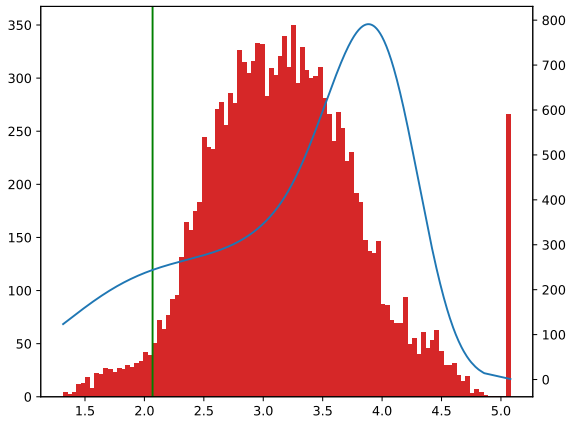
Vzdialenosť od kategórie Inovácie v bankovníctve a digitálne platobné služby



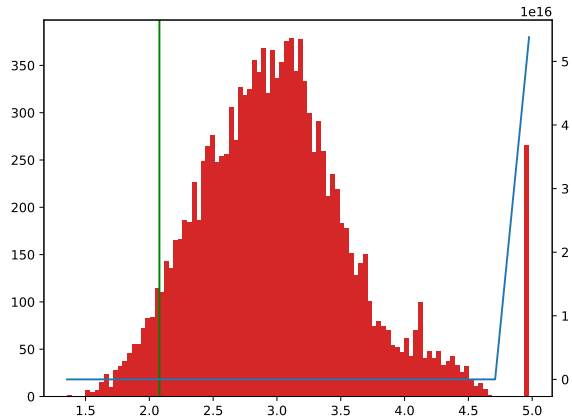
Vzdialenosť od kategórie Aktuálne marketingové aktivity



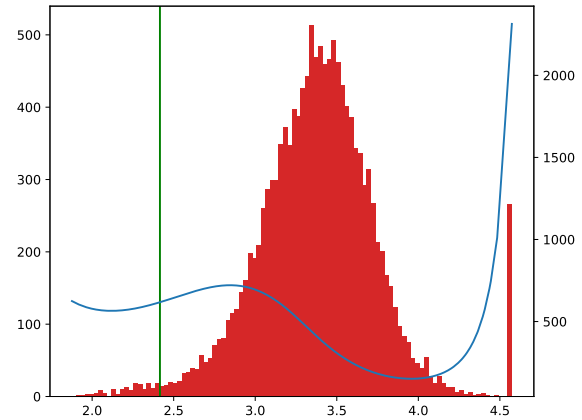
Vzdialenosť od kategórie Fintech novinky zo sveta



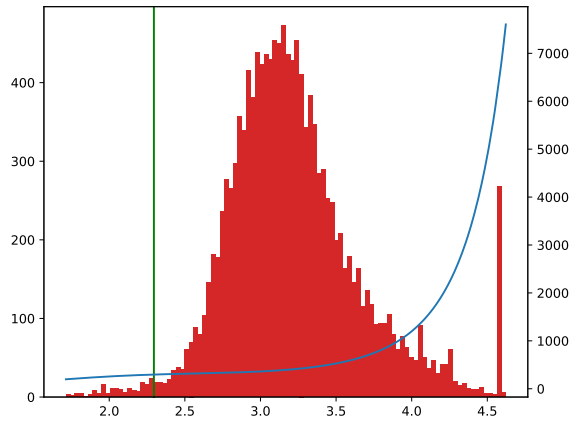
Vzdialenosť od kategórie Reklamné kampane OTP Banky



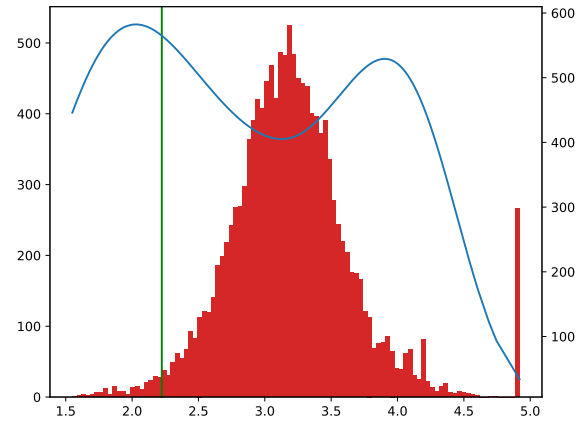
Vzdialenosť od kategórie Finančné produkty a vý-
hody pri nákupe prostredníctvom splátok a získa-
vaní bonusov



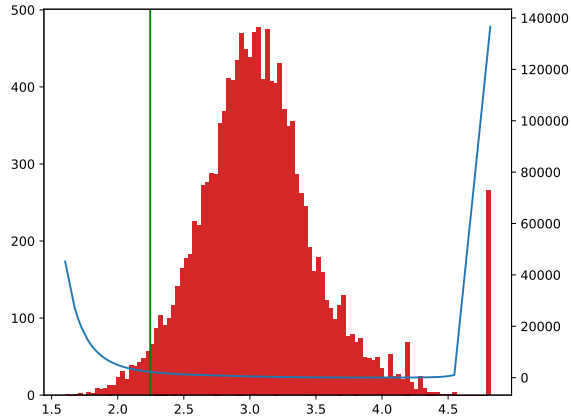
Vzdialenosť od kategórie Luxusné produkty a ex-
kluzívne investície v súčasnej dobe



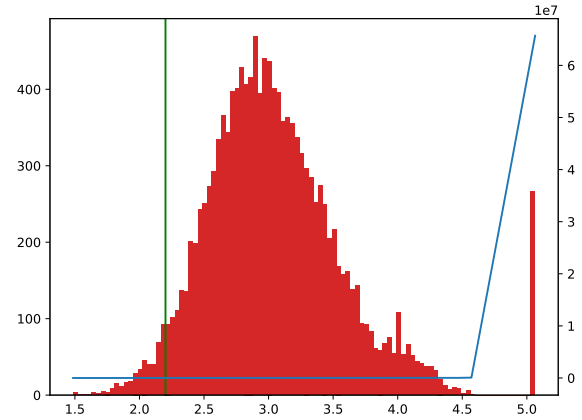
Vzdialenosť od kategórie Finančná gramotnosť a
zodpovedný prístup k peniazom



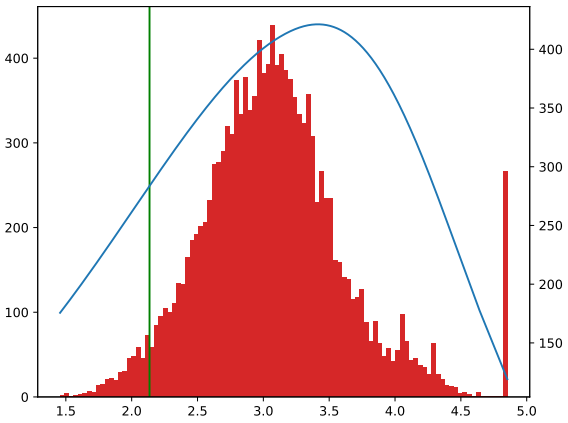
Vzdialenosť od kategórie Úspechy slovenských bánk
počas roka



Vzdialenosť od kategórie Spoločenská zodpoved-
nosť slovenských bánk



Vzdialenosť od kategórie Podpora umenia, vzdelá-
vania a inovácií na Slovensku



Vzdialenosť od kategórie Inovácie vo finančných službách na Slovensku

Obr. 10: Závislosť vzdialenosti od jenotlivých kategórií od interakcií

Príloha C - Zoznam všetkých tém, ktorými sa firmy zaoberajú

V tabuľke nižšie sa nachádza zoznam všetkým tém, ktoré sme pomocou zhlukovacej metódy našli spolu aj s ich príkladmi.

Téma	Príklad príspevku
Dôležitosť poistenia pre rôzne životné situácie	Ten, kto športuje, sa asi nikdy nevyhol úrazom. Niektoré boli menšie, iné si vyžadovali dlhší čas rehabilitácie. Prečítajte si, ako môže byť Úrazové poistenie Spirit dôležité pre váš výkon. Zriaďte si ho jednoducho online na deň alebo dlhodobo bez dokladovania zdravotného stavu. Navyše teraz za uzavretie dlhodobého poistného Spirit môžete získať 10 € poukážku do športového obchodu. Viac info na `www.csob.sk/poistenie-urazu-a-zivota-spirit`
Podpora komunitných projektov a aktívnej participácie v meste	Ďakujeme! Opäť ste prihlásili úžasné projekty - tentokrát v oblasti podpory vzdelávania Pomôžte nám aj vybrať tie najlepšie, ktoré od nás získajú 1 000 €na ich realizáciu Hlasovať môžete na `www.gestopremesto.sk` #Raiffeisenbanka #Gestopremesto
Vianočné a sviatočné posolstvá	Najkrajší vianočný darček nič nestojí. Pretože ním je rodina Vypnite dnes mobily, nechajte doma žiariť iba vianočné svetielka a užite si Vianoce s vašimi najbližšími. Celá Raiffeisen banka na čele s Ralfom vám praje krásne vianočné sviatky

"Podpora kultúrnych a spoločenských podujatí	Tatry budú tento rok kvitnúť o čosi dlhšie a vy môžete byť vďaka SLSP pri tom. Od 1. do 3. septembra sa tu koná prvý ročník festivalu Tatra Flowers. Viete, v ktorom roku bola postavená najstaršia tatranská chata, Rainerova útulňa? Až 20 správnych odpovedí odmeníme dvomi lístkami na koncert Jany Kirschner a SĽUKu, ktorí vystúpia na Tatra Flowers 1. septembra. :) Hráme do 14. 8. 2017.
Finančné správy a tipy pre podnikateľov a investorov	Aj tento rok vám Živnostenský účetTB prináša seriál rozhovorov SZČO? S novým moderátorom Otom Kóňom prenikáme viac do hĺbky podnikania a v sérii rozhovorov prinášame rady a tipy pre živnostníkov. V prvom dieli novej série sme prebrali, ako by mal vyzerať správny cashflow, prvý rozpočet či efektívna práca so subdodávateľmi. Pozvanie prijali hostia Ľudo Franta zo stolárskej dielne Ajnfach a Gabo Baleja z Woollies - slovenské rastúce oblečenie z merino vlny a bio bavlny, ktorí sa zaoberajú výrobou oblečenia z merino vlny. Celý diel si môžte pozrieť, alebo vypočuť, formou podcastu na `bit.ly/tatra-banka-szco`
Vylepšovanie bankových služieb pomocou mobilných aplikácií	Pripravte sa, náš nekonečný kolotoč štartuje. Oddnes máte možnosť si k svojmu účtu #BezPoplatku požiadať o Visa kartu, s ktorou máte všetky výbery z bankomatov po celom svete zadarmo. Nech sa vám z toho nezamotá hlava.
Štatúty súťaží	Štatút súťaže ECCO poukážka
Príspevok bez textu	

Reklamné kampane na pôžičky	Získajte peniaze rýchlo a jednoducho vďaka Parádnej pôžičke. Požičajte si 1 000 €len za 15 €mesačne alebo 2 000 €a za 30 €mesačne. Či už snívate o novom parádnom bicykli alebo by ste sa skôr vybrali na exotickú dovolenku, to všetko máte na dosah. Viac informácií nájdete na `https://www.raiffeisen.sk/sk/pozicka/paradna-pozicka/` #Raiffeisenbanka #ParádnaPôžička
Vplyv makroekonomických faktorov na finančné trhy	Silný začiatok podnikovej výsledkovej sezóny za tretí kvartál na Wall Street znamenal návrat konštruktívnej atmosféry na trhy a dal stopku obavám zo stagflačného prostredia a energetickej krízy. Prečítajte si komentár analytika k 41. týždňu. #trhovysumar #financnetrhy #akcie #komodity #devizy
Kultúra vytvárania pozitívnej komunitnej atmosféry	Dnes oslavujeme medzinárodný deň komplimentov. V súčasnosti asi všetci túžime po lichôtke, ktorá nás rozveselí Čo by ste chceli počuť od vašich blízkych či kolegov? Podeľte sa s nami v komentároch #Raiffeisenbanka #DenKomplimentov
Bezpečné finančné správanie a moderné bankovníctvo	V týchto časoch zažívajú mnohí z nás doslova prievan v peňaženke. Hore išlo takmer všetko. Od potravín, cez plyn a elektrinu až po odvody na sociálne poistenie. Význam a dôležitosť efektívnej práce s peniazmi sa ukazuje najmä teraz. Ako míňať peniaze s rozumom? A ako vie byť banka nápomocná? Viac o tejto téme si môžete vypočuť v podcaste, ktorý nájdete tu: `http://bit.ly/41Qmw2q` #Raiffeisenbanka

Súťaže a tipy na úspory	ADVENTNÁ SÚŤAŽ - 3. deň Pre súťaživé typy sme si pripravili rébus. Spočítajte, koľko má náš vianočný stromček trojuholníkov dokopy. Pozor, je ich tam viac, než sa na prvý pohľad zdá Vaša správna odpoveď bude zaradená do zlosovania o výhernú poukážku Alza v hodnote 10 €. Tešíme sa na vaše odpovede. Pôžičku si ľahko vybavíte TU `https://moneytupozicka.sk/`
Podpora umenia a ocenenie umeleckých diel	Svet umenia nám dodáva inšpiráciu a vieru v krajšiu budúcnosť. Sme preto veľmi radi, že SNG – Slovenská národná galéria prijala náš dar vo forme zbierky 10 dizajnových šiat, 10 slovenských odevných dizajnérov. Šaty vznikali v priebehu 10 rokov pri príležitosti odovzdávania Ceny Nadácie Tatra banky za umenie pre múzu večera Adelu Vinczeovú. Tešíme sa na vašu návštevu v priestoroch Esterházyho paláca (SNG), v čase od 6. 10. 2020 do 10. 1. 2021. Aj vďaka Tatra banke má každý návštevník celoročný vstup do SNG zadarmo. Viac o výstave nájdete na `www.nadaciatatrabanky.sk/vystava-o-saty`
Rast nerovnosti a koncentrácia bohatstva v globálnom kontexte	Prvýkrát v histórii tvorili milionári 1 % svetovej dospelej populácie. Vytváranie bohatstva sa zdá „kompletne oddelené“ od ekonomických ťažkostí počas pandémie
Refinancovanie úverov a hypoték	Splácate priveľa za pôžičky v iných bankách? Spravte s tým niečo! Preneste si ich k nám do OTP Banky! Znížite si úroky až o 50 %. Takto budete splácať menej alebo získajte peniaze navyše pri nezmenenej mesačnej splátke. O prefinancovaní sa dozviete na bit.ly/Pozicka_na_refinancovanie

Citáty o peniazoch a šťastí	"Peniaze vám šťastie nekúpia, ale nedostatok peňazí vám určite kúpi utrpenie.", povedal Daniel Kahneman, americký profesor.
Finančné plánovanie a správa osobných financií	Začnite sporiť investične akúkoľvek sumu a pomôžte svojim peniazom rásť. Navyše, ak si do 30. 11. otvoríte InvestičnéSporenieTB, ste v súťaži o 300 EUR. Urobiť tak môžete jednoducho cez mobilnú aplikáciu alebo Internet banking. Viac informácii nájdete na `www.tatrabanka.sk/sk/novy-investor`
Plánované technické odstávky v bankových službách	Plánované obmedzenie služieb ČSOB V rámci skvalitňovania našich služieb je nevyhnutná technická odstávka. Dovoľujeme si vás preto informovať, že v sobotu 17. marca 2018 od 00:01 h do 12:00 h, budú nedostupné služby elektronického bankovníctva ČSOB. Vybrané platobné operácie budú obmedzené do 14.00 h Všetko dôležité o plánovanej technickej odstávke sa dozviete na `bit.ly/odstavka_elektronickych_sluzieb` Ďakujeme za porozumenie a za prípadné komplikácie sa vopred ospravedlňujeme.
"Podpora športových udalostí	Po prázdninovej prestávke je Detská Tour Petra Sagana späť a už teraz sa tešíme na všetkých malých cyklistov a cyklistky na štarte. Nechaj svoje deti zahviezdiť počas pretekov: 27. augusta v Šali. 18. septembra v Žiline na veľkom finále celej Tour. Nenechajte si to ujsť aj vďaka našej nadácii, ktorá je generálnym partnerom podujatia. Prihlášky tu: `www.detskatour.sk`

Rôzne témy z rôznych oblastí života	V piatkovom zápase #slovenskisokoli hviezdili a my sa už nevieme dočkať ako dopadne ten dnešný proti Anglicku. Obrana totiž môže nastúpiť opäť v plnom zložení. ;) Aký výsledok tipujete vy?
Marketingové kampane s platobnými kartami	Oblečte sa do nového so svojou kreditnou kartou Cetelem. Využite povianočné a novoročné výpredaje a nakúpte si nové oblečenie na celý rok a 100 z Vás odmeníme poukážkou v hodnote 100 €na relaxačný alebo dovolenkový pobyt. Viac informácii o súťaži na: `www.cetelem.sk/kreditna-karta/online/` #spoluzodpovedne

"Sociálne a environmentálne zodpovedné príspevky	Áno, jeseň je sen, ale pre strechy to môže byť nočná mora. Vlhko, dážď, sneh a chladné počasie im dávajú poriadne zabrať. Odborníci odporúčajú robiť údržbu strechy 2x ročne - na jar a na jeseň. Tu sú naše tipy, ako na to: Skontrolujte strechu. Všímajte si praskliny, poškodené časti alebo iné známky opotrebovania. Ak ich nájdete, je na čase jednať. Preč s lístím! Spadnuté listy na streche spôsobujú zadržiavane vody, čo nevyhnutne vedie k poškodeniu materiálu. Vyčistite odkvapové žľaby. Nezabudnite odstrániť všetko, čo by mohlo spôsobiť problémy - suché lístie či konáriky. Predídete tak zvlhnutiu a plesniam. Viete, ako často treba vymieňať strechu? Vo všeobecnosti sa čas určuje na základe materiálu: šindle vymeňte každých 12 - 20 rokov, asfaltové šindle každých 15 - 30 rokov a drevené šindle 20 - 25 rokov. S financovaním opravy strechy vám radi pomôžeme. Informujte sa u našich expertov na financovanie bývania na obchodných zastúpeniach alebo na infolinke +421 2 58 55 58 55.
Vplyv makroekonomických faktorov a geopolitických udalostí na svetové akciové trhy	Wall Street pokračovala v tiahlej rely z jesene minulého roka a prepisovala historické maximá pre očakávania zlatovlasého makrovývoja v Spojených štátoch a pretrvávajúcu eufóriu okolo umelej inteligencie a jej možné pozitívne efekty na produktivitu a ekonomický rast.

Bankové služby a finančná bezpečnosť	Nyugodtan alhatunk, ha áttekintésünk van számláinkról. Mostantól Önnek sem kell álmatlanul forgolódnia, attól félve, hogy elfelejtette időben befizetni a lakbért vagy a telefonszámlát. Az OTP Banka mobil alkalmazása segítségével néhány kattintással elintézhet minden befizetést, akár az ágyban fekve is. Volt már ideje kipróbálni? `http://bit.ly/OTP_mobilna_aplikacia`
Bezpečná a pohodová dovolenka s cestovným poistením	Cestujete na dovolenku vlastným autom? Nezabudnite na prípravu. Bezchybný technický stav je základ, v zahraničí vás aj malá technická porucha vyjde poriadne draho. Skontrolujte si platnosť povinného zmluvného poistenia a havarijného poistenia. Pre prípad neočakávaných situácií na cestách majte vždy v aute dostatok tekutín pre celú posádku a niečo malé na zjedenie. Ako na dovolenku budete cestovať vy?
Zmena titulnej fotografie	
Alternatívne investičné možnosti a výnosy v porovnaní s tradičnými bankovými produktami	Zarábajte ako banky! Na Žltom melóne môžete mať výnos 7,9 % p.a. namiesto 0,8 % na termínovanom vklade.

Ako sa chrániť pred internetovými podvodmi	V posledných dňoch sa s podvodmi cez e-maily a SMSky roztrhlo vrece. Všímajte si e-mailové adresy, z ktorých vám prichádzajú správy. Na prvý pohľad sa síce môžu tváriť ako informácie z banky, môžu vás však vyjsť draho. Zopakujme si základné pravidlo: v banke nikdy od klientov nepožadujeme vypĺňanie citlivých údajov alebo ich diktovanie do telefónu. Ak sa neúmyselne nachytáte, bezodkladne riešte situáciu s našim Klientskym centrom. Viac info o bezpečnosti a radách ako nenaletieť nájdete na našom webe: `www.slsp.sk/sk/ludia/bezpecnost#ako-nenaletiet`
Pôžičky a ich výhody	Na Žltom melóne si požičiavate priamo od iných ľudí a vyhýbate sa nákladnému finančnému systému. Množstvo ponúk od ľudí – investorov, Vám zabezpečí tie najvýhodnejšie podmienky pre Vašu pôžičku.
Rôzne aspekty bankovníctva a finančnej bezpečnosti	Pripravili sme pre vás ďalší krátky rozhovor zo série "Významné ženy v našej banke". Tento raz sme sa rozprávali s Ljiljanou Petrov, ktorá zastáva pozíciu CFO. `www.cetelem.sk/o-spolocnosti/blog/zo-zivota-nasej-spolocnosti/rozhovor-s-ljiljanou-petrov/` #nielenosmehomarca #genderequality #positiveimpact #womenofourbank

Marketingové súťaže a promočné akcie v bankovom a finančnom sektore	Počítajte pomaly, odpovedajte rýchlo a môžete vyhrať 50 €! Aj tento týždeň sme pre vás pripravili súťaž! Úloha je jednoduchá. Spočítajte, koľko darčekových kariet sa ukrýva na obrázku. Správnu odpoveď napíšte do komentára, dajte lajk a ste v hre o dve darčekové platobné karty s kreditom 50 €. Výhercov vyhlásime v pondelok 11.12.2017 o 10.00 hod. Už teraz vám prezradíme, že budúci týždeň na vás čaká ďalšia súťaž.
"Finančné návyky a spotreba v Slovenskej republike	Viete, že Slováci mesačne najviac míňajú na bývanie?
Finančné služby a zvyky v digitále	Nové albumy, filmy, seriály a teplé ponožky. Ak si po aktívnom lete už pripravený/á na dlhé jesenné večery, priprav na ne aj svoju domácnosť. V 365.specials teraz hráme o 45 vychytávok, ktoré ti prinesú zábavu aj relax priamo do obývačky. Vybrali sme kúsky, ktoré potešia dušu kvalitou a oko zas dizajnom. Pozri na to a zapoj sa bit.ly/letnyspecials P. S.: Každá 5. platba 365 kartou od Mastercard alebo aktivácia Syslenia či sporenia ti zvýšia šancu na výhru. Čas máš už len do 15.9., tak smelo do toho! Držíme palce.
Inovácie v bankovníctve a digitálne platobné služby	Vďaka novej funkcii zdieľania platby sa so známymi vyrovnáte ešte jednoduchšie. Priamo v aplikácii VÚB Mobil Banking si vygenerujete platobný link, vďaka ktorému si viete vyžiadať platbu od kohokoľvek. S naším linkom si poradia aj iné banky, nielen tá oranžová. Viac informácií nájdete na vub.sk/payme.

Aktuálne marketingové aktivity	Ak chcete mať istotu, že vaše peniaze nezhltne inflácia, investujte do mFondov. Investičný účet otvoríte jednoducho v apke alebo v internet bankingu a samotné investovanie nemôže byť ľahšie. Jednoducho si nastavíte trvalý príkaz na ľubovoľnú sumu od 15 EUR a ďalej sa o nič nemusíte starať. A ak máte ďalšie otázky, spýtajte sa našich odborníkov 13. októbra na neformálnom brunchi v Panta Rhei Eurovea: Prihláste sa s promo kódom: Brunch na tomto odkaze: `bit.ly/viac_o_mFondoch`
Fintech novinky zo sveta	Fintech spoločnosť PrimaryBid uskutočnila svoje B investičné kolo, v rámci ktorého získala 50 mil. USD na globálnu expanziu. Viac sa môžete dočítať v našich aktualitách: `www.zltymelon.sk/aktuality`
Reklamné kampane OTP Banky	Rozmýšľate nad pôžičkou Cetelem? Aké podmienky musíte spĺňať? Dozviete sa v našom článku: `https://goo.gl/8RDY5k`
Finančné produkty a výhody pri nákupe prostredníctvom splátok a získavaní bonusov	Doprajte si neskrývanú radosť z nákupu. Aj v Datart SK môžete do konca roka využiť Šikovnú splátku s 0 % navýšením pri splatení do 6 mesiacov. Navyše môžete získať aj bonus 100 €za prenos vášho účtu do VÚB banka. To je radosť! Viac info o akcii a reprezentatívny príklad bit.ly/Quatro-radost. Tento šikovný vysávač si môžete kúpiť tu `bit.ly/3rTvgHY`. #QUATRO #onlinenakupovanie #prinasaVUB

Luxusné produkty a exkluzívne investície v súčasnej dobe	Automobilka Aston Martin sa tentoraz spojila s architektonickým štúdiom S3 Architectures, ktoré navrhlo podobu luxusnej rezidencie v štáte New York v blízkosti Manhattanu. Prvotná vila s rozlohou takmer 600 štvorcových metrov bude v predaji za viac ako sedem miliónov dolárov.
Finančná gramotnosť a zodpovedný prístup k peniazom	Dnes je Svetový deň vďačnosti Rodina, priatelia, ale aj zdravie a strecha nad hlavou... Vy ste za čo najviac vďační vo svojom živote? Inak, všimli ste si? Na uľahčenie vyznania vďaky sme pre vás pripravili aj pohľadnicu, ktorá čaká na preposlanie vašim srdcu blízkym #Raiffeisenbanka #Budmevdacni
Úspechy slovenských bánk počas roka	Oranžová banka je dnes strieborná. V ankete Smartbanka 2020 sa naša aplikácia VÚB Mobil Banking umiestnila na 2. mieste. Ďakujeme vám za každý jeden hlas, ktorým motivujete našich skvelých zamestnancov. Tí aj vďaka vám neustále vylepšujú naše online služby.
Spoločenská zodpovednosť slovenských bánk	Sme bankou pre všetko, na čom záleží. A preto sme podporili rekonštrukciu Kliniky detskej hematológie a onkológie, aby vytvorila príjemné a bezpečné prostredie malým pacientom. Veríme, že pozitívna energia a smiech dokáže pomáhať v každej situácii! :)

Podpora umenia, vzdelávania a inovácií na Slovensku	Veríme, že umenie vytvára hodnoty, ktoré sú dôležité za každých okolností. Aby sme dokázali ešte viac, spojili sme sa s Bratislavským samosprávnym krajom. Individuálne dotácie BSK @bratislavskykraj.sk na podporu kultúry vo výške 150 000 EUR navýši Nadácia Tatra banky o ďalších 150 000 EUR. Spolu tak pomôžeme krízou COVID-19 ohrozenému umeniu a kultúre sumou vo výške až 300 000 EUR. Viac informácií nájdete na `bit.ly/nadacia_bsk`.
Inovácie vo finančných službách na Slovensku	Služby našej Infolinky stále rozširujeme. Odteraz si cez telefón s Diskrétnym údajom alebo Hlasovou biometriou toho vybavíte ešte viac, napríklad: získate ePIN na e-mail, zmeníte limity na platobnej karte, zrušíte SMS o pohyboch na účte, nastavíte, zmeníte alebo zrušíte inkaso, vybavíte už aj papierové potvrdenie o vedení a zostatku na účte aktivujete karty na platenie s mobilom alebo hodinkami. Ešte Hlasovú biometriu nevyužívate? Všetky informácie aj postup, keď si chcete Hlasovú biometriu zriadiť nájdete na `https://6500.sk/hlasovabiometria`.