

České vysoké učení technické v Praze
Fakulta stavební



Úvod do zpracování prostorových dat

Ekonomické subjekty

Bc. Tomáš Klemsa

Bc. Petr Poskočil

Bc. Robin Pflug

Bc. Marek Fáber

Obsah

| | | |
|----------|--|-----------|
| 1 | Úvod | 2 |
| 2 | Zdroje dat | 2 |
| 2.1 | ČSÚ | 2 |
| 2.1.1 | ČSÚ - NACE | 2 |
| 2.2 | ČÚZK - RÚIAN | 2 |
| 2.2.1 | VDP | 3 |
| 2.2.2 | Služby | 3 |
| 2.2.3 | VFR | 3 |
| 2.2.4 | CSV | 4 |
| 3 | Software | 4 |
| 3.1 | PostgreSQL | 4 |
| 3.2 | Python | 4 |
| 4 | Praktická část | 5 |
| 4.1 | Stažení dat | 5 |
| 4.2 | Tvorba tabulek | 5 |
| 4.3 | Struktura databáze | 6 |
| 4.3.1 | Tabulka business register | 6 |
| 4.3.2 | Tabulka company nace mapping | 6 |
| 4.3.3 | Tabulka electronic registration of sales | 7 |
| 4.3.4 | Tabulka nace | 7 |
| 5 | Dotazy | 7 |
| 6 | Výstupy projektu | 10 |
| 7 | Závěr | 13 |
| 8 | Reference | 13 |

1 Úvod

Dokumentace k projektu *Ekonomické subjekty* vypracovaného jako součást předmětu 155UZPR Úvod do zpracování prostorových dat. Autory projektu jsou Bc. Tomáš Klemsa, Bc. Petr Poskočil, Bc. Robin Pflug a Bc. Marek Fáber. Vedoucí předmětu je Ing. Martin Landa, Ph. D. Předmět je zaměřen na zpracování (geo)prostorových dat, geodatabáze, správu geoprostorových dat v objektově-relačních databázových systémech a jejich zpracování.

Jako podmínka pro splnění předmětu je vypracování semestrálního projektu, který bude vhodně spjatý se zaměřením předmětu. Jako téma projektu bylo zvoleno propojení ekonomických subjektů s polohovou informací a zjišťování závislosti zániku a vzniku těchto subjektů na spuštění EET.

Téma projektu je zvoleno na základě pracovní zkušenosti Bc. Tomáše Klemsy. Projektem vytvořená databáze umožňuje získávat bezplatně jednoduše informace, které jsou v současné době obtížně dostupné a finančně nákladné. SQL dávky jsou zaměřeny na změny týkající se ekonomických subjektů v návaznosti na zavedení elektronické evidence tržeb.

2 Zdroje dat

2.1 ČSÚ

Český statistický úřad (ČSÚ) je ústředním orgánem české státní správy a byl zřízen už 8. ledna 1969. Hlavní činností úřadu je získání a zpracování údajů pro statistické účely a jejich poskytnutí dalším státním orgánům, veřejnosti a do zahraničí. Zároveň také koordinuje sběr a zpracování dat údajů prováděných pro jednotlivá ministerstva. Jeho základním posláním je vytvářet objektivní a ucelený obraz ekonomického, sociálního, demografického a ekologického vývoje České republiky a jejích částí.

Všechna data a informace jsou na serveru zdarma pro státní správu i běžného uživatele. Data jsou většinou k dispozici ve dvou formátech - ve formátu XML a CVS.[1]

2.1.1 ČSÚ - NACE

NACE je akronym pro statistickou klasifikaci ekonomických činností, kterou používá Evropská unie (resp. Evropská společenství) od roku 1970. NACE vytváří rámec pro statistická data o činnostech v mnoha ekonomických oblastech (např. ve výrobě, zaměstnanosti, národních účtech).

Statistiky, které vzniknou za použití klasifikace NACE, lze srovnávat v celé Evropské unii. S nižší mírou podrobnosti (na vyšších úrovních) je možné srovnání i se světovými statistikami. Používání NACE je povinné pro všechny členské státy Evropské Unie.

2.2 ČÚZK - RÚIAN

V polovině roku 2012 byl úspěšně spuštěn provoz systému základních registrů veřejné správy ČR. Jeden ze čtyř základních registrů je i registr územní identifikace, adres a nemovitostí (RÚIAN). RÚIAN je veřejný seznam, který umožňuje uživatelům z řad veřejné,



Obrázek 1: Logo ČSÚ [1]

ale i komerční a akademické sféry, dálkový přístup přes internet - aplikace Veřejného dálkového přístupu (VDP) k datům RÚIAN je dostupná zdarma a bez registrace.[2]

2.2.1 VDP

Aplikace Veřejný dálkový přístup k datům RÚIAN (VDP) umožňuje nahlížet a získávat data základního registru RÚIAN a také některá data editačního agendového informačního systému územní identifikace (ISÚI) a informačního systému katastru nemovitostí (ISKN). Pro přístup do aplikace VDP není potřeba žádné registrace. Poskytovaná data z VDP jsou zdarma. Data poskytovaná prostřednictvím VDP nejsou referenční, mají pouze informativní charakter.[2]

2.2.2 Služby

V rámci Informačního systému základních registrů (ISZR) fungují čtyři základní registry veřejné správy. Zajišťováním provozu ISZR a správou eGON služeb základních registrů se zabývá Správa základních registrů (SZR). eGon služby základních registrů poskytující referenční údaje ze základních registrů i služby poskytující zprostředkované údaje z jiných registrovaných Agendových informačních systémů (AIS). Webové služby ISZR slouží pouze pro komunikaci registrovaných AIS veřejné správy se základními registry. Služby jsou publikované na vnějším rozhraní systému základních registrů.[2]

2.2.3 VFR

Jednou z forem poskytování dat RÚIAN je jejich předávání ve formě souborů obsahujících data RÚIAN nebo ISÚI ve výměnném formátu RÚIAN (VFR). VFR jsou poskytovány ve formátu GML 3.2.1.

Soubory VFR je možné stahovat:

- prostřednictvím aplikace Veřejný dálkový přístup (VDP) – volně dostupné pro všechny
- z internetových (URL) adres vrácených službami ISZR ruianSouboryDat a ruianSouboryZmen – dostupné pouze orgánům státní správy a samosprávy

[2]

2.2.4 CSV

Další formou poskytování údajů je seznam adresních míst RÚIAN ve formátu CSV. Soubory jsou rozděleny po obcích a jsou generovány měsíčně ze stavového VFR.[2]



Obrázek 2: Logo ČÚZK [2]

3 Software

Pro tvorbu databáze a prostorových dotazů byl využit open-source software PostgreSQL. Ke stažení dat byl vytvořen skript v programovacím jazyce Python.

3.1 PostgreSQL

PostgreSQL je univerzální a objektově relační systém správy databází, nejpokročilejší open-source databázový systém. PostgreSQL byl vyvinut na základě POSTGRES 4.2 na Berkeley Computer Science Department, University of California.

PostgreSQL byl navržen tak, aby fungoval na platformách typu UNIX. PostgreSQL byl však také navržen tak, aby byl přenosný, aby mohl běžet na různých platformách, jako jsou Mac OS X, Solaris a Windows.

PostgreSQL je bezplatný a open source software. Jeho zdrojový kód je k dispozici pod licencí PostgreSQL, liberální licence s otevřeným zdrojovým kódem. Můžete libovolně používat, upravovat a distribuovat PostgreSQL v jakékoli formě.[3]



Obrázek 3: Logo PostgreSQL [3]

3.2 Python

Python je vysokoúrovňový skriptovací programovací jazyk, který v roce 1991 navrhl Guido van Rossum. Nabízí dynamickou kontrolu datových typů a podporuje různá programo-

vací paradigmatu, včetně objektivě orientovaného, imperativního, procedurálního nebo funkcionálního.

Python je vyvíjen jako open source projekt, který zdarma nabízí instalační balíky pro většinu běžných platform (Unix, MS Windows, macOS, Android); ve většině distribucí systému GNU/Linux je Python součástí základní instalace.[4]



Obrázek 4: Logo python [4]

4 Praktická část

4.1 Stažení dat

Data byla získána z Registru Ekonomických Subjektů (RES), klasifikace ekonomických činností (CZ-NACE) a Registru Územní Identifikace, Adres a Nemovitostí (RÚIAN). Skriptem v jazyce Python byly na základě *URL id* iterativně stažena všechna potřebná data z webové aplikace českého statistického úřadu. Data byla ukládána do textového souboru. Celkové množství stažených záznamů je přibližně 7 milionů, některé záznamy jsou prázdné.

```
1 "6";"0000078";"Lesní správa Lány";"331";"1.1.1974";"";"Lány, 27061, Lány, Lesní 140";"541991";"240";"13110"  
2 "15";"00000205";"Vojenské lesy a statky ČR, s.p."; "301";"1.1.1972";"";"Praha 6, 16000, Dejvice, Pod Juliskou 1621/5";"500178";"450";"11001"  
3 "24";"00000337";"Vojenské stavby - státní podnik v likvidaci";"301";"1.1.1972";"";"Praha 8 - Libeň, V Mezihoří 1082/2, PSČ 18000";"500208";
```

Obrázek 5: Ukázka části textového souboru se staženými daty

4.2 Tvorba tabulek

Tvorba databáze byla realizována na serveru *geo102.fsv.cvut.cz* v databázi *pgis_uzpr* pod názvem *uzpr20_g*. Všechny pět tabulek bylo vytvořeno SQL skriptem v příkazovém řádku prostřednictvím PostgreSQL. Plnění dat do tabulek zajišťují příkazy obsažené v souboru typu Linux Shell Commands.

| List of relations | | | |
|-------------------|----------------------------------|-------|----------|
| Schema | Name | Type | Owner |
| uzpr20_g | business_register | table | uzpr20_g |
| uzpr20_g | company_nace_mapping | table | uzpr20_g |
| uzpr20_g | electronic_registration_of_sales | table | uzpr20_g |
| uzpr20_g | employee_size | table | uzpr20_g |
| uzpr20_g | nace | table | uzpr20_g |

Obrázek 6: Přehled vytvořených tabulek

Po vytvoření tabulek s ekonomickými subjekty byly připojeny polohové informace. Polohové informace byly získány z RÚIAN na základě variací společných atributů s ČSÚ. Pouze 2,6% subjektů zůstaly bez přiřazené polohové informace.

4.3 Struktura databáze

4.3.1 Tabulka business register

- **id:** číslo záznamu v tabulce
- **identification_number:** identifikační číslo ekonomického subjektu
- **company_name:** název ekonomického subjektu
- **legal_form:** kód právní formy organizace
- **establishment:** datum vzniku ekonomického subjektu
- **dissolution:** datum zániku ekonomického subjektu
- **adress:** adresa ekonomického subjektu
- **basic territorial unit:** základní územní jednotka ekonomického subjektu
- **employee_size:** kód kategorie počtu zaměstnanců ekonomického subjektu
- **institutional_sector:** institucionální sektor ekonomického subjektu
- **adress_point:** adresní bod ekonomického subjektu

| Table "uzpr20_g.business_register" | | | | | | | |
|------------------------------------|------------------------|-----------|----------|---------|----------|--------------|-------------|
| Column | Type | Collation | Nullable | Default | Storage | Stats target | Description |
| id | integer | | not null | | plain | | |
| identification_number | character(8) | | not null | | extended | | |
| company_name | character varying(500) | | not null | | extended | | |
| legal_form | character(3) | | not null | | extended | | |
| establishment | date | | not null | | plain | | |
| dissolution | date | | | | plain | | |
| adress | character varying(200) | | not null | | extended | | |
| basic_territorial_unit | integer | | not null | | plain | | |
| employee_size | character(3) | | not null | | extended | | |
| institutional_sector | character(5) | | not null | | extended | | |
| adress_point | integer | | | | plain | | |

Obrázek 7: Tabulka business register

4.3.2 Tabulka company nace mapping

- **identification_number:** identifikační číslo ekonomického subjektu
- **nace_code:** kód názvu ekonomické činnosti

| Table "uzpr20_g.company_nace_mapping" | | | | | | | |
|---------------------------------------|-----------------------|-----------|----------|---------|----------|--------------|-------------|
| Column | Type | Collation | Nullable | Default | Storage | Stats target | Description |
| identification_number | character(8) | | not null | | extended | | |
| nace_code | character varying(10) | | not null | | extended | | |

Obrázek 8: Tabulka company nace mapping

4.3.3 Tabulka electronic registration of sales

- **nace:** kód názvu ekonomické činnosti
- **start_date:** datum spuštění EET
- **end_date:** datum ukončení EET

| Table "uzpr20_g.electronic_registration_of_sales" | | | | | | | |
|---|-----------------------|-----------|----------|---------|----------|--------------|-------------|
| Column | Type | Collation | Nullable | Default | Storage | Stats target | Description |
| nace | character varying(10) | | not null | | extended | | |
| start_date | date | | not null | | plain | | |
| end_date | date | | | | plain | | |

Obrázek 9: Tabulka electronic registration of sales

4.3.4 Tabulka nace

- **level:** level typu ekonomické činnosti
- **code:** kód názvu ekonomické činnosti
- **description:** popis ekonomické činnosti

| Table "uzpr20_g.nace" | | | | | | | |
|-----------------------|------------------------|-----------|----------|---------|----------|--------------|-------------|
| Column | Type | Collation | Nullable | Default | Storage | Stats target | Description |
| level | integer | | | | plain | | |
| code | character varying(10) | | not null | | extended | | |
| description | character varying(200) | | not null | | extended | | |

Obrázek 10: Tabulka nace

5 Dotazy

Vytvořené dotazy jsou zaměřené na spuštění evidence elektronických služeb a jejího vlivu ekonomické subjekty. Jsou zde zohledněny 2 vlny spuštění EET (1.12.2016, 3.1.2017) a jejich vliv na ekonomické subjekty s počtem zaměstnanců nižším než 6.


```

1 -----
2 ----- 1. wave of EET -----
3 -----
4 select extract(year from i::date) as year
5         ,extract(week from i::date) as week
6         ,null as dissolutions
7         ,null as establishments
8 into temp table eet1
9 from generate_series('2015-01-01','2019-12-31', '1 week'::interval) i;
10
11 -- update dissolutions
12 update eet1
13 set dissolutions = d.diss
14 from (
15     select extract(year from br.dissolution) as year
16            ,extract(week from br.dissolution) as week
17            ,count(*) as diss
18     from business_register br
19     join company_nace_mapping cnm
20         on cnm.identification_number = br.identification_number
21     join electronic_registration_of_sales eros
22         on eros.nace = cnm.nace_code
23     where eros.start_date = '2016-12-01' -- First wave of EET
24           and br.employee_size = '110' -- 0 employees
25           and br.dissolution is not null
26     group by extract(year from br.dissolution)
27              ,extract(week from br.dissolution)
28     ) d
29 where eet1.year = d.year and eet1.week = d.week;
30
31 -- update establishment
32 update eet1
33 set establishments = e.est
34 from (
35     select extract(year from br.establishment) as year
36            ,extract(week from br.establishment) as week
37            ,count(*) as est
38     from business_register br
39     join company_nace_mapping cnm
40         on cnm.identification_number = br.identification_number
41     join electronic_registration_of_sales eros
42         on eros.nace = cnm.nace_code
43     where eros.start_date = '2016-12-01' -- First wave of EET
44           and br.employee_size = '110' -- 0 employees
45     group by extract(year from br.establishment)
46              ,extract(week from br.establishment)
47     ) e
48 where eet1.year = e.year and eet1.week = e.week;
49
50 select * from eet1
51 order by year, week;
52
53 drop table eet1;

```

Obrázek 11: Atributový dotaz vracející počet ekonomických subjektů nově vzniklých a ukončujících svou činnost v intervalu od 1.1.2015 do 31.12.2019 vypisovaný po jednom týdnu. Tyto subjekty podléhají první vlně zavedení EET.

```

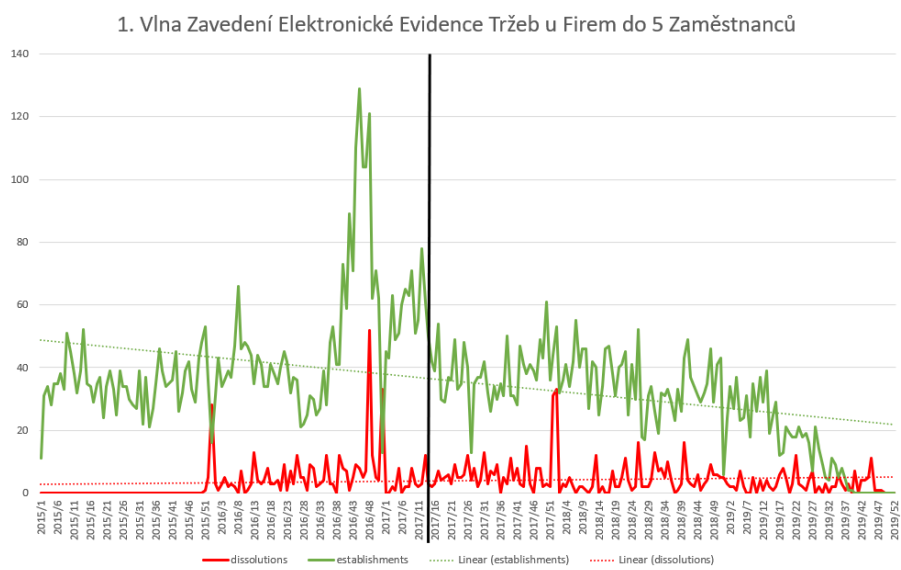
1 -----
2 ----- 1. wave of EET -----
3 -----
4 select extract(year from i::date) as year
5         ,extract(week from i::date) as week
6         ,null as dissolutions
7         ,null as establishments
8 into temp table eet1
9 from generate_series('2015-01-01','2019-12-31', '1 week'::interval) i;
10
11 -- update dissolutions
12 update eet1
13 set dissolutions = d.diss
14 from (
15     select extract(year from br.dissolution) as year
16            ,extract(week from br.dissolution) as week
17            ,count(*) as diss
18     from business_register br
19     join company_nace_mapping cnm
20         on cnm.identification_number = br.identification_number
21     join electronic_registration_of_sales eros
22         on eros.nace = cnm.nace_code
23     where eros.start_date = '2016-12-01' -- First wave of EET
24           and br.employee_size = '110' -- 0 employees
25           and br.dissolution is not null
26     group by extract(year from br.dissolution)
27              ,extract(week from br.dissolution)
28            ) d
29 where eet1.year = d.year and eet1.week = d.week;
30

```

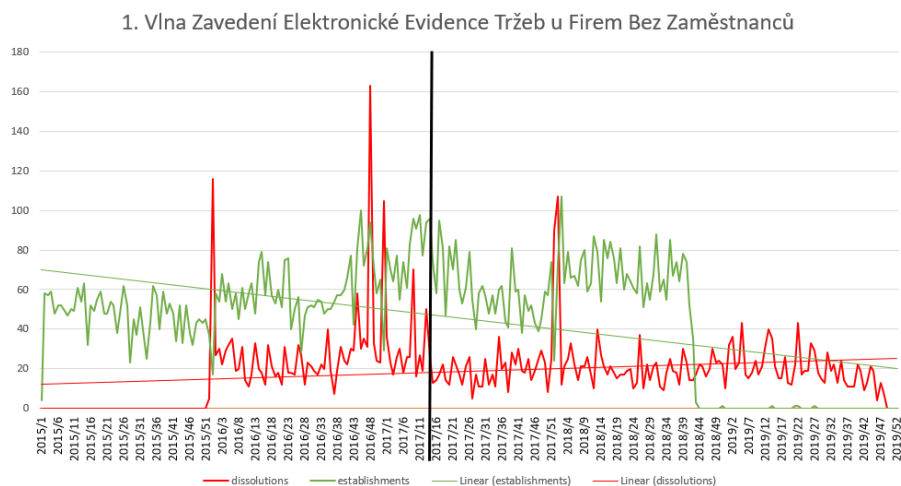
Obrázek 12: Prostorový dotaz vypisující ekonomické subjekty ukončující svoji činnost týden po zavedení první vlny EET.

Stejný postup byl použit i pro druhou vlnu EET, kompletní řešení je obsaženo v SQL dávce.

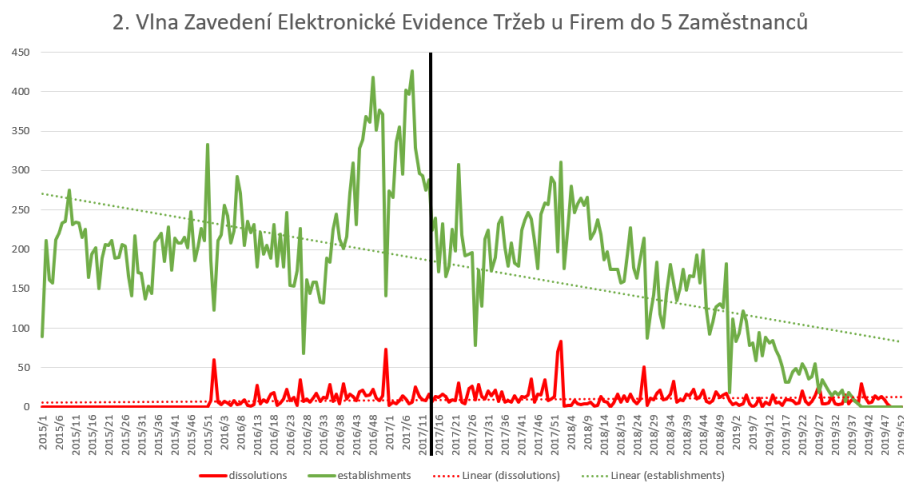
6 Výstupy projektu



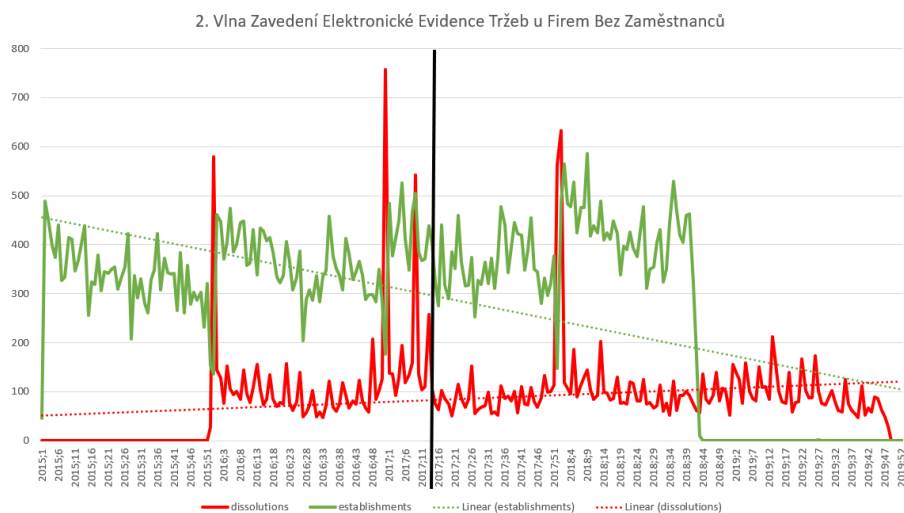
Obrázek 13: Graf zobrazující počet vznikajících a zanikajících ekonomických subjektů. Počet zaměstnanců je 5 a méně.



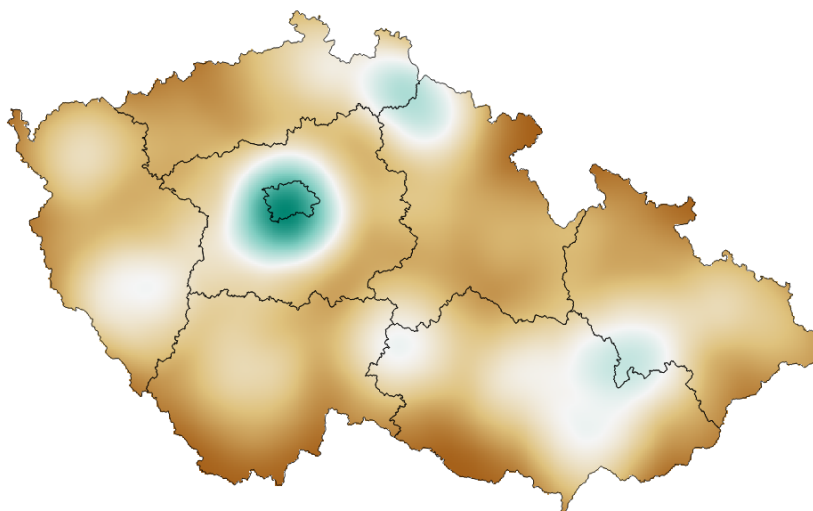
Obrázek 14: Graf zobrazující počet vznikajících a zanikajících ekonomických subjektů spadajících do první vlny EET. Počet zaměstnanců je 0.



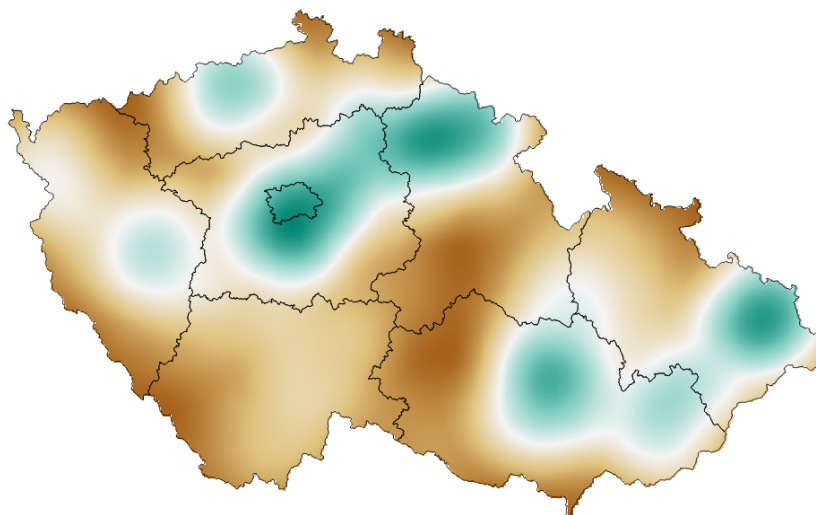
Obrázek 15: Graf zobrazující počet vznikajících a zanikajících ekonomických subjektů spadajících do druhé vlny EET. Počet zaměstnanců je 5 a méně



Obrázek 16: Graf zobrazující počet vznikajících a zanikajících ekonomických subjektů spadajících do druhé vlny EET. Počet zaměstnanců je 0.



Obrázek 17: Celoplošná kernel-heat mapa ČR, zobrazující množství ekonomických subjektů ukončující svou činnost v prvním týdnu po zavedení první vlny EET. Počet zaměstnanců je 0.



Obrázek 18: Celoplošná kernel-heat mapa ČR, zobrazující množství ekonomických subjektů ukončující svou činnost v prvním týdnu po zavedení druhé vlny EET. Počet zaměstnanců je 0.

7 Závěr

Cílem projektu bylo vytvoření nové databáze pro sledování vnějších a vnitřních vlivů na ekonomické subjekty v celé škále měřítka, od drobných podnikatelů až po korporátní společnosti. Databáze kombinuje atributovou a prostorovou informaci, tudíž nabízí širší perspektivu interpretace dat. Mezi hlavní způsoby využití může být sledování dopadů ve vývoji ekonomiky z hlediska evaluace. Zároveň také může sloužit pro plánování obchodních strategií z hlediska konkurenceschopnosti a pohledávky.

Další neméně důležitou částí projektu je způsob jakým byla data získána. Za normálních okolností by data pro tento projekt byla finančně nedostupná. Za pomoci vhodné aplikace byla data získána z veřejných webových aplikací.

8 Reference

1. *ČESKÝ STATISTICKÝ ÚŘAD* [online], [cit. 2020-2-3]. Dostupné z: www.czso.cz
2. *ČÚZK* [online], [cit. 2020-2-3]. Dostupné z: www.cuzk.cz
3. *PostgreSQL Tutorial* [online], [cit. 2020-2-3]. Dostupné z: www.postgresqltutorial.com
4. *python* [online], [cit. 2020-2-3]. Dostupné z: www.python.org