

# Ukládání a příprava dat – Projekt

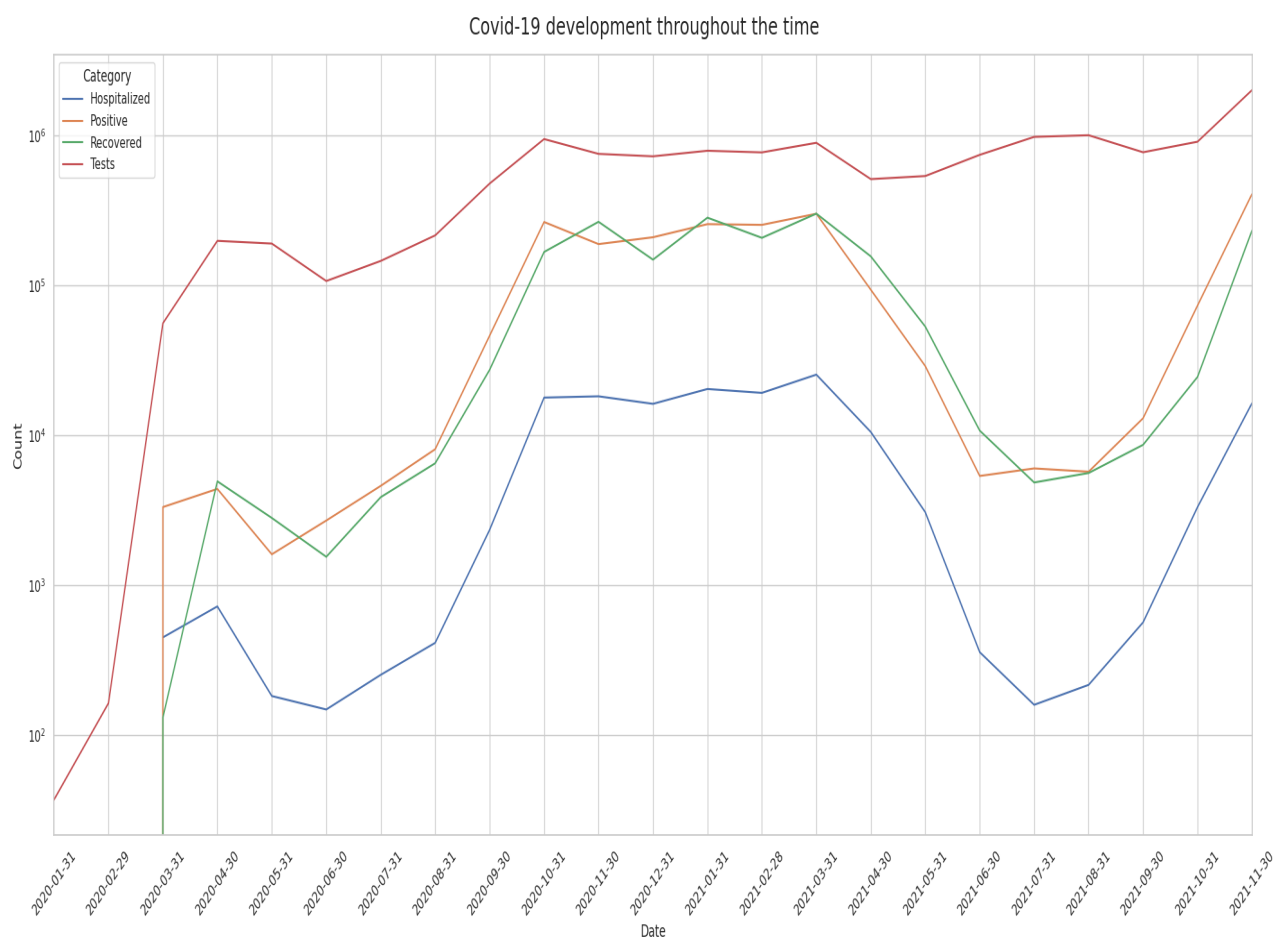
## Výsledky získané v rámci řešení úloh

Matej Berezný, Ondrej Valo, Švenk Adam

xberez03, xvalo00, xsvenk00

### 1 1. dotaz skupiny A

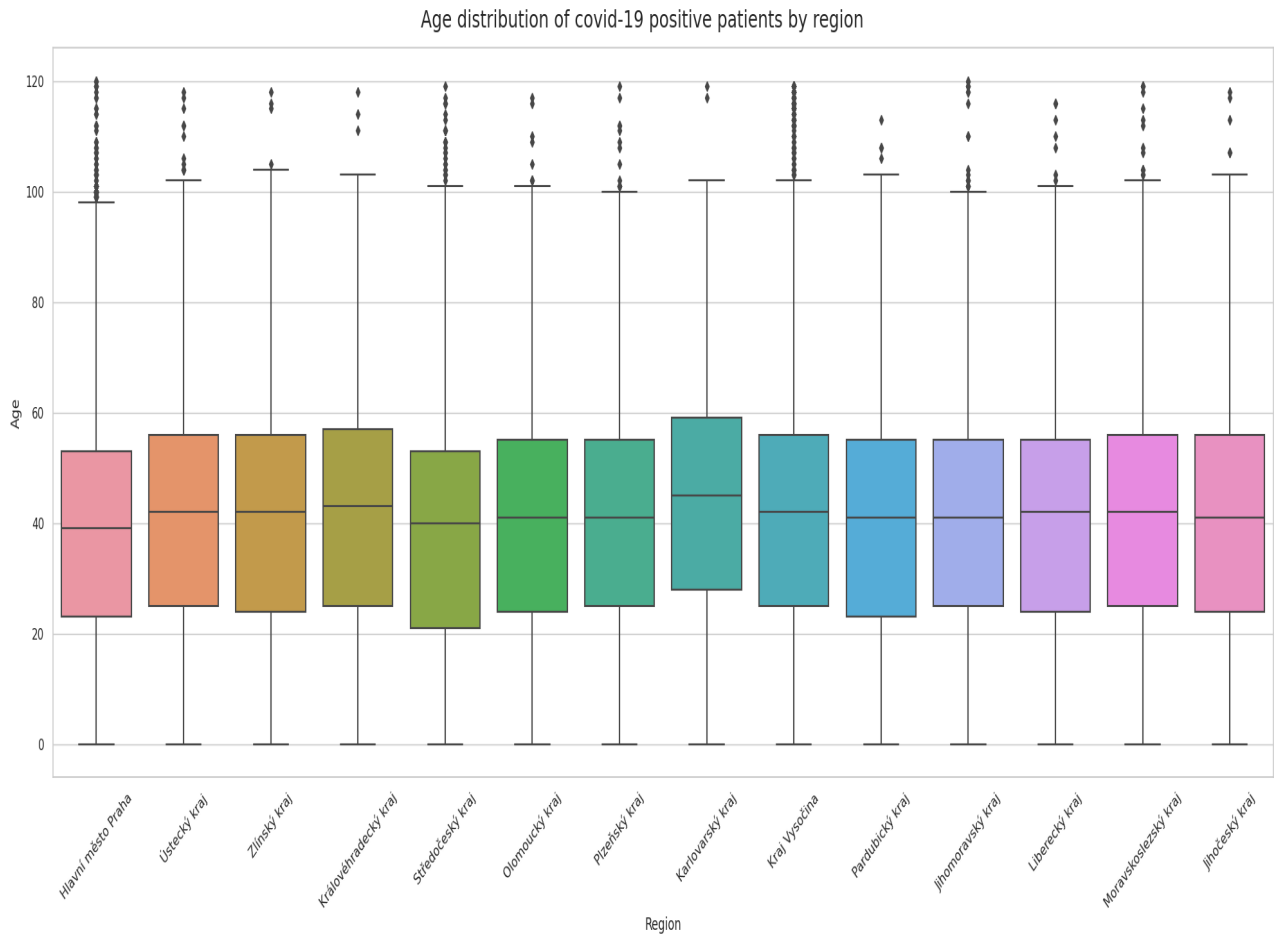
V rámci prvej úlohy bolo treba vygenerovať graf zobrazujúci vývoj covidovej situácie po mesiacoch. V grafe sú zobrazené hodnoty: počet novo nakazených za mesiac(oranžová), počet novo vyliečených za mesiac(zelená), počet novo hospitalizovaných osôb za mesiac(modrá) a počet vykonaných testov za mesiac(červená).



Obrázek 1: Vývoj Covid-19 v priebehu času.

## 2 2. dotaz skupiny A

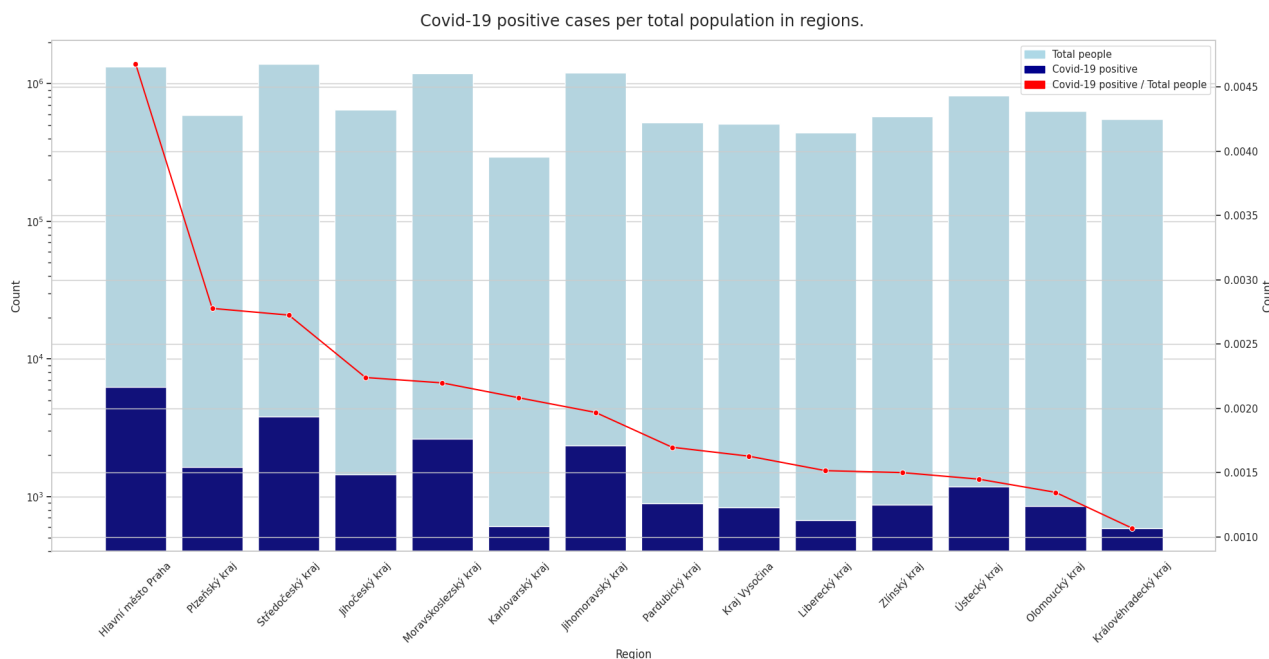
V rámci druhej úlohy bolo treba vytvoriť krabicové grafy zobrazujúce rozloženie veku nakazených osôb v jednotlivých krajoch. Na grafe sú kraje zľava: Hlavné mesto Praha, Ústecký kraj, Zlínský kraj, Královéhradecký kraj, Středočeský kraj, Olomoucký kraj, Plzeňský kraj, Karlovarský kraj, Kraj Vysočina, Pardubický kraj, Jihomoravský kraj, Liberecký kraj, Moravskoslezský kraj, Jihočeský kraj.



Obrázek 2: Vekové rozloženie pacientov pozitívnych na covid-19 podľa regiónu.

### 3 Dotaz skupiny B

V úlohe skupiny B bolo treba zostaviť 4 rebríčky krajov "best in covid" za posledné 4 štvrťroky. Ako kritérium bol braný počet novo nakazených prepočítaný na jedného obyvateľa kraja. A pre jeden štvrt rok vygenerovať graf, ktorý zobrazuje počet novo nakazených (tmavo modrá), celkový počet obyvateľov (svetlo modrá) a počet nakazených na jedného obyvateľa (červená). Počet nakazených na jedného obyvateľa bol vypočítaný ako *počet nakazených v kraji/celková populácia v kraji*. Graf 3 bol vygenerovaný na základe tabuľky 5



Obrázek 3: Pozitívne prípady Covid-19 na celkovú populáciu v regiónoch.

| Region               | index | Date       | infected_per_month | total_people | metric   |
|----------------------|-------|------------|--------------------|--------------|----------|
| Královéhradecký kraj | 66    | 2021-03-31 | 60574              | 552946       | 0,109548 |
| Liberecký kraj       | 67    | 2021-03-31 | 43009              | 438594       | 0,098061 |
| Karlovarský kraj     | 64    | 2021-03-31 | 29072              | 301726       | 0,096352 |
| Plzeňský kraj        | 71    | 2021-03-31 | 54906              | 572687       | 0,095874 |
| Pardubický kraj      | 70    | 2021-03-31 | 49367              | 516440       | 0,095591 |
| Středočeský kraj     | 72    | 2021-03-31 | 120701             | 1291816      | 0,093435 |
| Ústecký kraj         | 74    | 2021-03-31 | 61897              | 826764       | 0,074867 |
| Jihočeský kraj       | 63    | 2021-03-31 | 46635              | 636611       | 0,073255 |
| Hlavní město Praha   | 61    | 2021-03-31 | 90054              | 1246780      | 0,072229 |
| Kraj Vysočina        | 65    | 2021-03-31 | 32832              | 511207       | 0,064224 |
| Olomoucký kraj       | 69    | 2021-03-31 | 40035              | 637609       | 0,062789 |
| Jihomoravský kraj    | 62    | 2021-03-31 | 72686              | 1168650      | 0,062197 |
| Moravskoslezský kraj | 68    | 2021-03-31 | 74408              | 1226602      | 0,060662 |
| Zlínský kraj         | 73    | 2021-03-31 | 31761              | 587693       | 0,054044 |

Obrázek 4: Dáta Pozitívnych prípadov Covid-19 na celkovú populáciu v regiónoch za prvý štvrt' rok.

| Region               | index | Date       | infected_per_month | total_people | metric   |
|----------------------|-------|------------|--------------------|--------------|----------|
| Zlínský kraj         | 88    | 2021-06-30 | 9888               | 587693       | 0,016825 |
| Jihočeský kraj       | 78    | 2021-06-30 | 10403              | 636611       | 0,016341 |
| Kraj Vysočina        | 80    | 2021-06-30 | 7993               | 511207       | 0,015636 |
| Ústecký kraj         | 89    | 2021-06-30 | 11990              | 826764       | 0,014502 |
| Moravskoslezský kraj | 83    | 2021-06-30 | 16971              | 1226602      | 0,013836 |
| Olomoucký kraj       | 84    | 2021-06-30 | 8613               | 637609       | 0,013508 |
| Pardubický kraj      | 85    | 2021-06-30 | 6351               | 516440       | 0,012298 |
| Jihomoravský kraj    | 77    | 2021-06-30 | 13943              | 1168650      | 0,011931 |
| Liberecký kraj       | 82    | 2021-06-30 | 5187               | 438594       | 0,011826 |
| Středočeský kraj     | 87    | 2021-06-30 | 14431              | 1291816      | 0,011171 |
| Hlavní město Praha   | 76    | 2021-06-30 | 13003              | 1246780      | 0,010429 |
| Plzeňský kraj        | 86    | 2021-06-30 | 4811               | 572687       | 0,008401 |
| Královéhradecký kraj | 81    | 2021-06-30 | 2756               | 552946       | 0,004984 |
| Karlovarský kraj     | 79    | 2021-06-30 | 1145               | 301726       | 0,003795 |

Obrázek 5: Data Pozitivních případů Covid-19 na celkovou populaci v regionech za druhý čtvrt' rok.

| Region               | index | Date       | infected_per_month | total_people | metric   |
|----------------------|-------|------------|--------------------|--------------|----------|
| Hlavní město Praha   | 91    | 2021-09-30 | 6248               | 1246780      | 0,005011 |
| Středočeský kraj     | 102   | 2021-09-30 | 3811               | 1291816      | 0,00295  |
| Plzeňský kraj        | 101   | 2021-09-30 | 1643               | 572687       | 0,002869 |
| Jihočeský kraj       | 93    | 2021-09-30 | 1442               | 636611       | 0,002265 |
| Moravskoslezský kraj | 98    | 2021-09-30 | 2623               | 1226602      | 0,002138 |
| Karlovarský kraj     | 94    | 2021-09-30 | 611                | 301726       | 0,002025 |
| Jihomoravský kraj    | 92    | 2021-09-30 | 2356               | 1168650      | 0,002016 |
| Pardubický kraj      | 100   | 2021-09-30 | 888                | 516440       | 0,001719 |
| Kraj Vysočina        | 95    | 2021-09-30 | 830                | 511207       | 0,001624 |
| Liberecký kraj       | 97    | 2021-09-30 | 672                | 438594       | 0,001532 |
| Zlínský kraj         | 103   | 2021-09-30 | 871                | 587693       | 0,001482 |
| Ústecký kraj         | 104   | 2021-09-30 | 1184               | 826764       | 0,001432 |
| Olomoucký kraj       | 99    | 2021-09-30 | 852                | 637609       | 0,001336 |
| Královéhradecký kraj | 96    | 2021-09-30 | 589                | 552946       | 0,001065 |

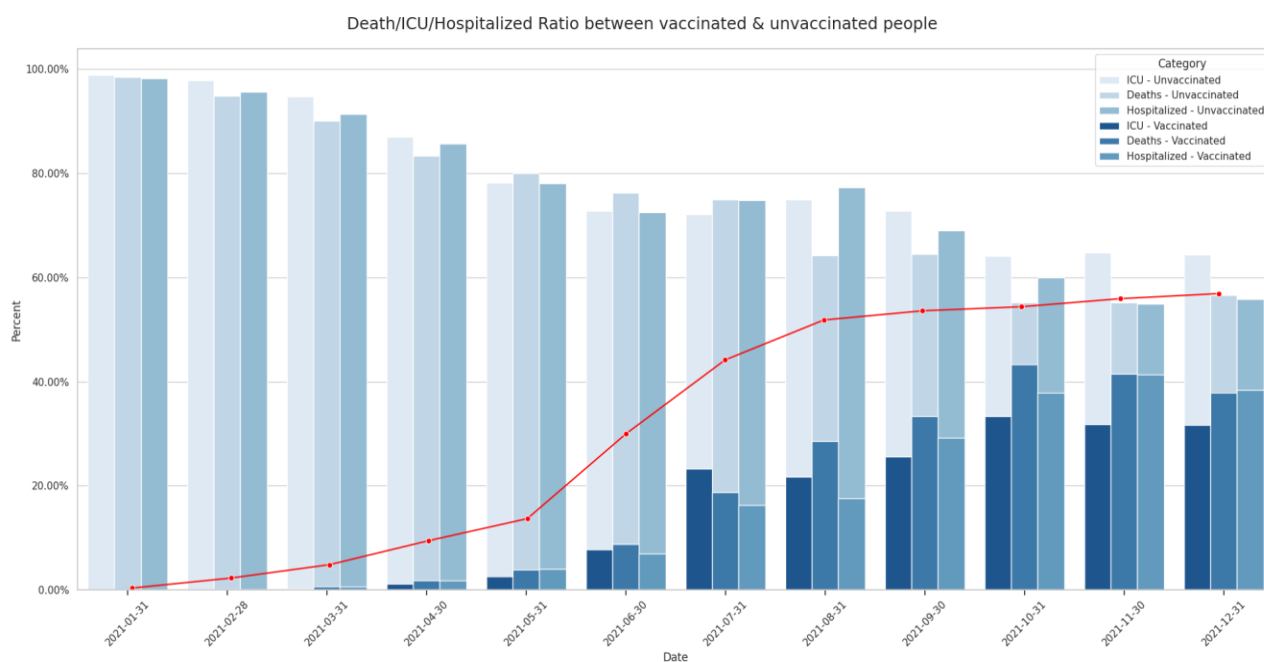
Obrázek 6: Data Pozitivních případů Covid-19 na celkovou populaci v regionech za třetí čtvrt' rok.

| Region               | index | Date       | infected_per_month | total_people | metric      |
|----------------------|-------|------------|--------------------|--------------|-------------|
| Olomoucký kraj       | 114   | 2021-12-31 | 53647              | 637609       | 0,084137771 |
| Jihomoravský kraj    | 107   | 2021-12-31 | 89805              | 1168650      | 0,076845078 |
| Zlínský kraj         | 118   | 2021-12-31 | 44713              | 587693       | 0,07608224  |
| Moravskoslezský kraj | 113   | 2021-12-31 | 89717              | 1226602      | 0,073142715 |
| Jihočeský kraj       | 108   | 2021-12-31 | 44399              | 636611       | 0,069742747 |
| Pardubický kraj      | 115   | 2021-12-31 | 33805              | 516440       | 0,065457749 |
| Hlavní město Praha   | 106   | 2021-12-31 | 79246              | 1246780      | 0,063560532 |
| Středočeský kraj     | 117   | 2021-12-31 | 81962              | 1291816      | 0,063447116 |
| Kraj Vysočina        | 110   | 2021-12-31 | 30789              | 511207       | 0,060228049 |
| Plzeňský kraj        | 116   | 2021-12-31 | 32473              | 572687       | 0,056702876 |
| Liberecký kraj       | 112   | 2021-12-31 | 20968              | 438594       | 0,047807312 |
| Královéhradecký kraj | 111   | 2021-12-31 | 25880              | 552946       | 0,046803847 |
| Ústecký kraj         | 119   | 2021-12-31 | 38695              | 826764       | 0,046802957 |
| Karlovarský kraj     | 109   | 2021-12-31 | 7235               | 301726       | 0,023978709 |

Obrázek 7: Data Pozitivních případů Covid-19 na celkovou populaci v regionech za čtvrtý čtvrtý rok.

## 4 Vlastný dotaz 1

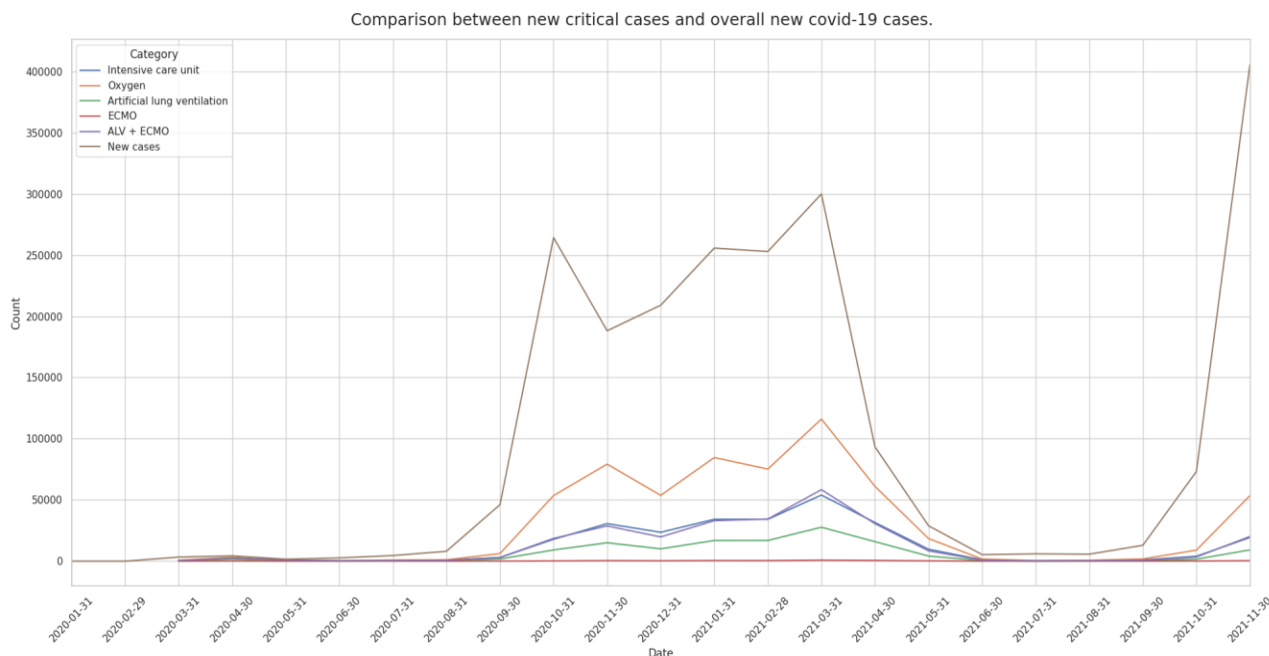
Prvý dotaz zobrazuje pomer medzi očkovanými (tmavo modré stĺpčeky) a neočkovanými osobami (svetlo modré stĺpčeky) z hľadiska počtu úmrtí (druhý stĺpec), počtu pacientov na JIP (prvý stĺpec), počtu hospitalizovaných pacientov (tretí stĺpec) a ich vzťahu k celkovej zaočkovanosti populácie (červená čiara) počas jednotlivých mesiacov. Pomer aj celková zaočkovanosť je vyjadrená v percentách z dôvodu veľkých rozdielov dát pre jednotlivé mesiace.



Obrázek 8: Pomer úmrtí, JIP a hospitalizácie medzi očkovanými a neočkovanými osobami.

## 5 Vlastný dotaz 2

Druhý dotaz zobrazuje vývoj počtu pacientov s veľmi ťažkým priebehom ochorenia nachádzajúcich sa na špeciálnych prístrojoch/oddeleniach (JIP, pľúcna ventilácia etc.) ku celkovému počtu nových prípadov ochorenia Covid-19 (hnedá farba) počas jednotlivých mesiacov.



Obrázek 9: Porovnanie medzi novými kritickými prípadmi a celkovo novými prípadmi COVID-19.

## 6 Dotaz skupiny C

Cieľom úlohy skupiny C bolo vytvorenie súboru vo formáte CSV, ktorý bude možné predať dolovaciemu algoritmu. Predmetom dolovacieho algoritmu je vyhľadávanie miest (okresov), v ktorých je vývoj aktuálnej situácie týkajúcej sa šírenia ochorenia Covid-19 podobný. Normalizácia prebiehala výpočtom percentuálneho podielu počtu očkovanej populácie za obdobie posledných 4 štvrťrokov (čiže období posledného roka) vzhľadom na celkovú populáciu danej regionálnej oblasti. Diskretizáciu dát vychádza z kategorizácie percentuálneho počtu zaočkovanej populácie. Okrem dodaného súboru CSV, na ktorom je vykonaná normalizácia a diskretizácia, je dodávaný aj súbor vo formáte CSV bez týchto vykonaných úkonov. Zadanie vyžadovalo detekciu a nahradenie odľahlých hodnôt. Vzhľadom na vybratú tému a použité dátové kolekcie však nebolo možné detekovať a nahradiť odľahlé hodnoty, nakoľko takéto vykonanie nie je prospešné v prípade zvolenej témy. Výsledný súbor vo formáte CSV (konkrétne jeho časť) je možné vidieť na obrázku 10).

| Region                    | Age [0-15] | Age [15-55] | Age [55+] | Infected | Vaccinated | Vaccination_percentage | Vaccination     |
|---------------------------|------------|-------------|-----------|----------|------------|------------------------|-----------------|
| <u>Benešov</u>            | 16757      | 50079       | 32906     | 41496    | 21360      | 21.4152513484791       | partial         |
| <u>Beroun</u>             | 17267      | 50535       | 28601     | 30940    | 15458      | 16.0347707021566       | partial         |
| <u>Blansko</u>            | 18098      | 55124       | 35882     | 42224    | 28635      | 26.2456005279366       | about_a_quarter |
| <u>Brno-město</u>         | 59790      | 198385      | 124230    | 364      | 129930     | 33.9770661994482       | about_a_quarter |
| <u>Brno-venkov</u>        | 41455      | 117858      | 67129     | 68068    | 47724      | 21.0755955167328       | partial         |
| <u>Bruntál</u>            | 13292      | 46182       | 31497     | 24388    | 23614      | 25.9577227907795       | about_a_quarter |
| <u>Břeclav</u>            | 17930      | 60052       | 38500     | 22932    | 27877      | 23.9324530828798       | partial         |
| <u>Čeb</u>                | 14089      | 46926       | 30419     | 14560    | 19862      | 21.7227727103703       | partial         |
| <u>Chomutov</u>           | 19703      | 66020       | 38877     | 16016    | 31515      | 25.292937399679        | about_a_quarter |
| <u>Chrudim</u>            | 16512      | 52934       | 35262     | 39312    | 20906      | 19.9660006876265       | partial         |
| <u>Domažlice</u>          | 9887       | 31768       | 20129     | 27664    | 6973       | 11.2860934869869       | partial         |
| <u>Děčín</u>              | 20711      | 65321       | 42417     | 18928    | 23686      | 18.440003425484        | partial         |
| <u>Frýdek-Místek</u>      | 34311      | 109980      | 70296     | 26208    | 39670      | 18.4866744024568       | partial         |
| <u>Havlíčkův Brod</u>     | 14938      | 47121       | 32553     | 43680    | 21703      | 22.9389506616497       | partial         |
| <u>Hodonín</u>            | 22176      | 78118       | 53313     | 29848    | 43682      | 28.4375061032375       | about_a_quarter |
| <u>Hradec Králové</u>     | 26050      | 82158       | 56227     | 37856    | 60715      | 36.9234043847113       | about_a_quarter |
| <u>Jablonec nad Nisou</u> | 14851      | 45653       | 29694     | 12740    | 21382      | 23.7056254018936       | partial         |
| <u>Jeseník</u>            | 5327       | 18794       | 13588     | 8736     | 7121       | 18.8840860272084       | partial         |
| <u>Jihlava</u>            | 18551      | 58249       | 36864     | 44772    | 18403      | 16.1907024211712       | partial         |
| <u>Jindřichův Hradec</u>  | 13625      | 45006       | 31648     | 38584    | 19311      | 21.3903565613266       | partial         |
| <u>Jičín</u>              | 12496      | 40248       | 27390     | 40404    | 20890      | 26.0688347018744       | about_a_quarter |

Obrázek 10: Pripravené dáta pre dolovací algoritmus